Overview of Japanese National Research Grid Initiative (NAREGI) Project

● Kenichi Miura

(Manuscript received September 2, 2004)

The Japanese National Research Grid Initiative (NAREGI) is a 5-year project covering the period 2003 to 2007. The goals of the project are to 1) conduct R&D on high-performance and scalable Grid middleware to provide a future computational infrastructure for scientific and engineering research in Japan, 2) Grid-enable a specific target application, in this case nanoscience, to prove the usefulness of Grids in future scientific and industrial applications, and 3) perform experimental deployment of 100 TFLOPS-scale Grids based on the middleware and Grid-enabled applications that have been constructed. The National Institute of Informatics (in charge of Grid middleware R&D), the Institute for Molecular Science (in charge of Grid-enabling nanoscience applications), major universities, and industries will participate in this project. The Super SINET, an all-optical 10 Gbps and beyond backbone network for academic research, will be used as the network infrastructure for this project. This paper outlines the NAREGI project.

1. Introduction

Research and development of the Grids for scientific use, which we call "research Grids," have been conducted mainly by universities and national research institutes in both Japan and other countries. The Japanese national projects for the research on Grids preceding the project to be described in this paper are the Information Technology Based Laboratory (ITBL) project[1] (initiated in fiscal 2001), Grid research in the Super Science Information Network (Super SINET) project[2] (initiated in fiscal 2001), BioGrid project[3] (initiated in fiscal 2001), and VizGrid project[4] (initiated in fiscal 2002), all of which have been initiated by the Ministry of Education, Culture, Sports, Science and Technology.

In the NAREGI (Japanese National Research Grid Initiative) project,[5] the computational research environment is expected to be as fast as several hundred TFLOPS (Tera-FLoating-point Operations Per Second) within five years. This project conducts research and development of the practical Grid middleware required for an operational and robust computational research environment. It is also intended to prove the usefulness of the Grid computing environment in actual applications (e.g., in the application software and large-scale simulation in cutting-edge nanoscience and nanotechnology fields), which are useful for developing new materials and next-generation ultrafine devices, and finally establish a Grid computing environment that is also accessible by users in industrial communities. The major purposes of the project are to 1) conduct research on the Grid middleware and integrate the developed software, 2) conduct research on the application of high-speed network technology to the Grid environment, and 3) conduct research and development on the large-scale simulation programs for use in nanoscience fields and study their optimization in the Grid environment.

2. Organization of NAREGI project

To attain the intended goals, the NAREGI project has a research organization built on two centers: 1) the Center for Grid Research and Development (National Institute of Informatics, Research Organization of Information and Systems), which is in charge of research and development of Grid middleware, Grid network, and Grid-enabled nanoscience applications, and 2) the Computational Nanoscience Center (Institute for Molecular Science, National Institutes of Natural Sciences), which is in charge of research and development of the most advanced application software in nanoscience fields (**Figure 1**). Each center also collaborates with collaborating organizations in industry, academia, and the government. The project plans to use Super SINET, which is managed by the National Institute of Informatics, as the network infrastructure (**Figure 2**).

2.1 Subjects of research at the Center for Grid Research and Development

The NAREGI project does not aim at developing ultrafast computer hardware but focuses on

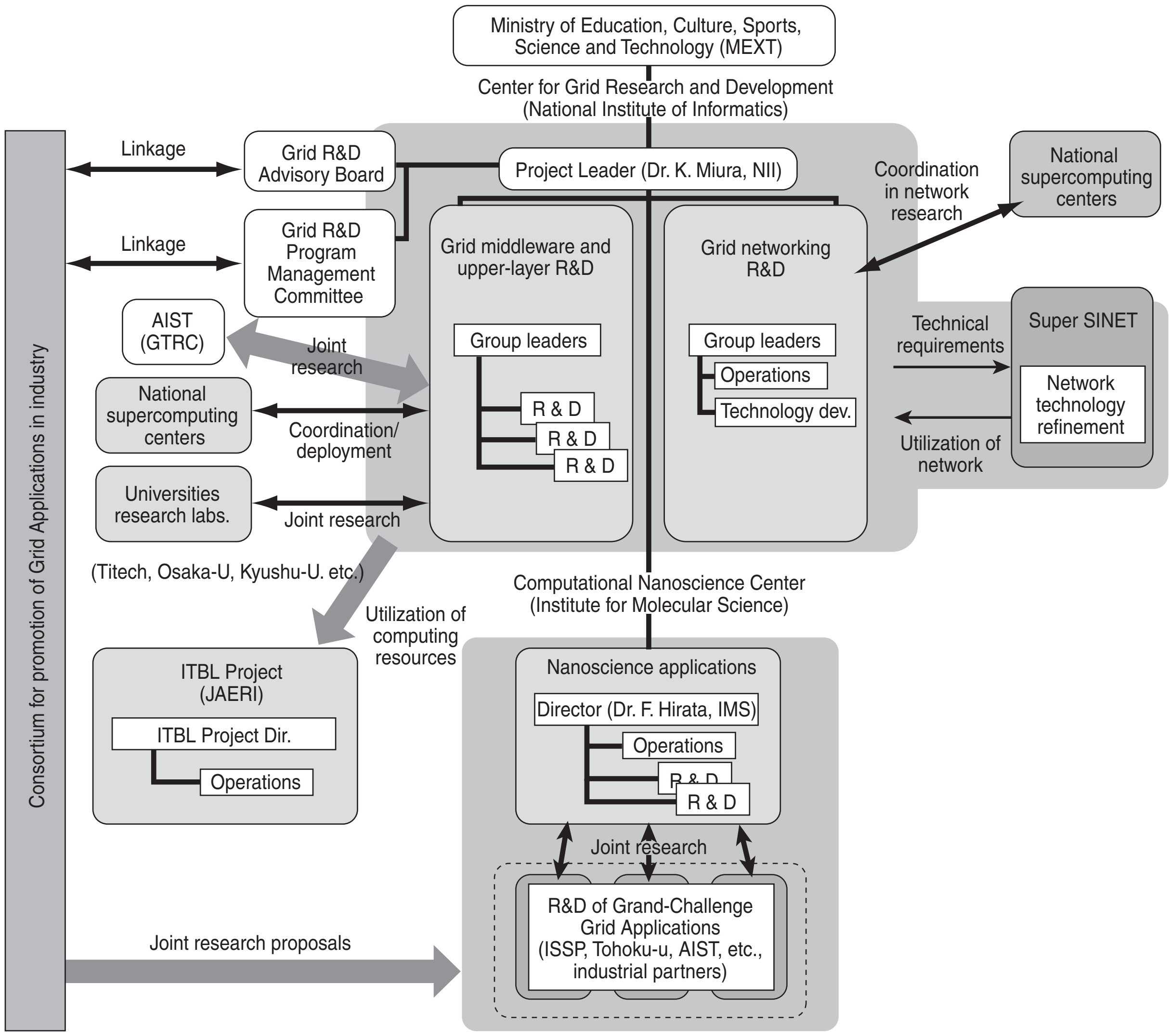

Figure 1
Research organization of NAREGI project.

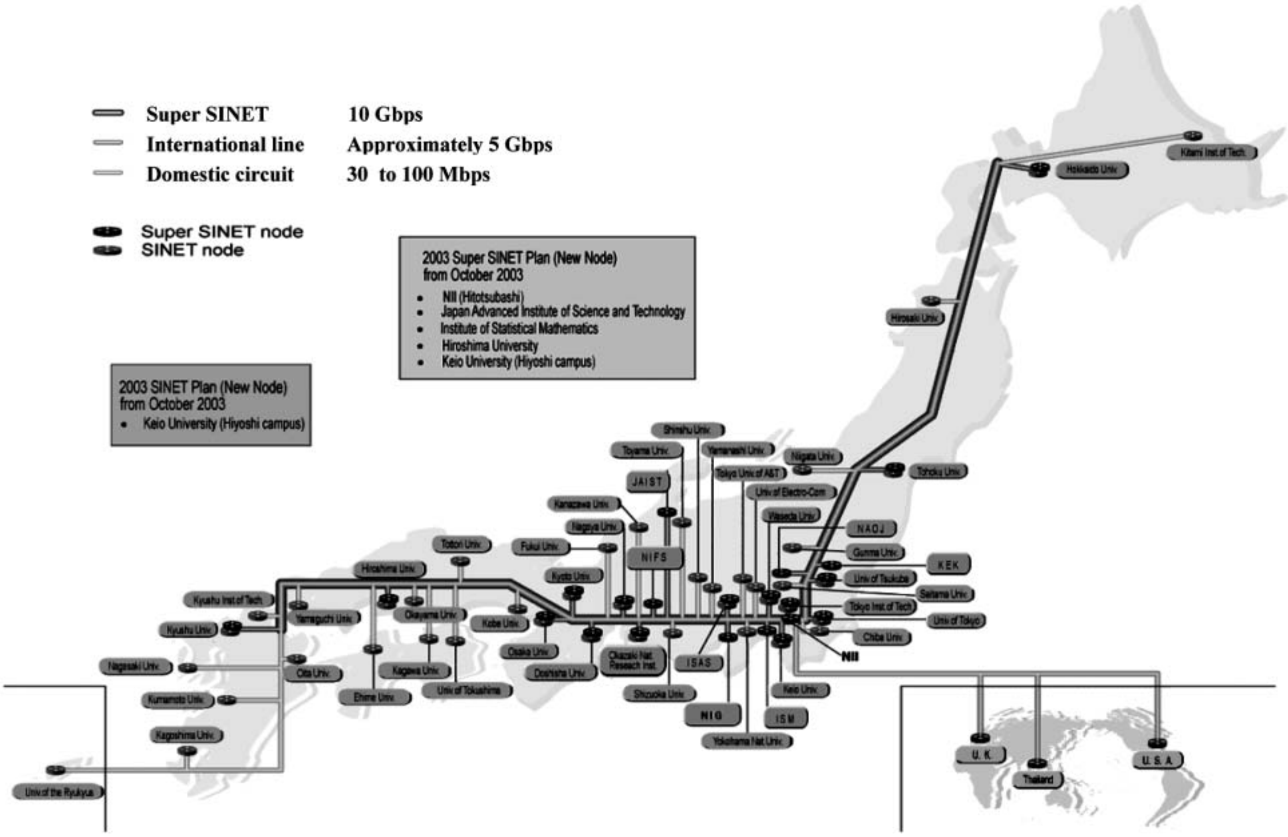

Figure 2
Network topology map of Super SINET (source: National Institute of Informatics).

research and development of the software technology (Grid middleware) that links existing heterogeneous computer resources with each other and presents them as seamlessly integrated resources. Under the NAREGI project, the following research subjects are to be conducted at the Center for Grid Research and Development:

1) Resource management in a Grid environment

A super-scheduler to control the entire scheduling over a geographically distributed area, operation management and accounting for Grid resources, virtualization of Grid computing resources (GridVM), and linkage of the workload management system (Condor) to perform high-throughput scheduling on Grids with European Grid middleware (Uniform Interface to Computing Resources [UNICORE]).

2) Grid programming environment

Grid-enabled remote procedure call (GridRPC) (the method of remotely calling procedures that enable parallel computation using distributed computation resources) and Grid message passing interface (GridMPI) (the communication library using MPI that enables parallelized programs to operate in a Grid environment) for programming on the Grid.

3) Software and environment for development of Grid applications

This consists of the following: 1) a workflow graphical user interface (GUI) and a flow description language to describe job execution in a Grid environment as a flow graph, 2) an application program development environment and simulation execution support environment (problem solving environment [PSE]) to pool and deploy application software to enable collaborations among researchers, and 3) visualization tools suitable for the Grid environment.

4) Network communication infrastructure

This consists of 1) a network traffic measurement and dynamic route optimization based on traffic for Grid computing to be fed back to the super-scheduler, 2) communication protocols, and 3) a public key infrastructure (PKI).

5) Grid-enabling nano-simulation software

This consists of tools and libraries to make parallelized nanoscience/nanotechnology application programs adaptable to the Grid environment and enable execution of coupled simulations on the Grid.

The subjects also include development of the technology to integrate, verify, and deliver the software achievements of the project. **Figure 3** shows the relationships (hierarchical structure) of the above research subjects.

In the project, research and development of Grid middleware do not start from scratch with development of each function. In the first phase, the Grid middleware, for example, UNICORE,[6] Globus Toolkit,[7] and Condor,[8] which are already popular as de facto standards, are used appropriately as the basis and research and development will be conducted on the upper-layer of the middleware. The base software will evolve into the Open Grid Services Architecture (OGSA)[9] when development of OGSA has progressed. Because Dr. David Snelling of Fujitsu Laboratories of Europe developed the main parts of UNICORE, the researchers of the NAREGI project communicate with the UNICORE project team about the internal specifications and future roadmap of UNICORE and the UNICORE project team is expecting feedback and contributions from the NAREGI project.

2.2 Subjects of research at the Computational Nanoscience Center

From the viewpoint of nanoscience, proving the concept that Grid computing is useful as a research tool for nanoscience is one of the major objectives of this project. This computational aspect of nanoscience is regarded as complemen-

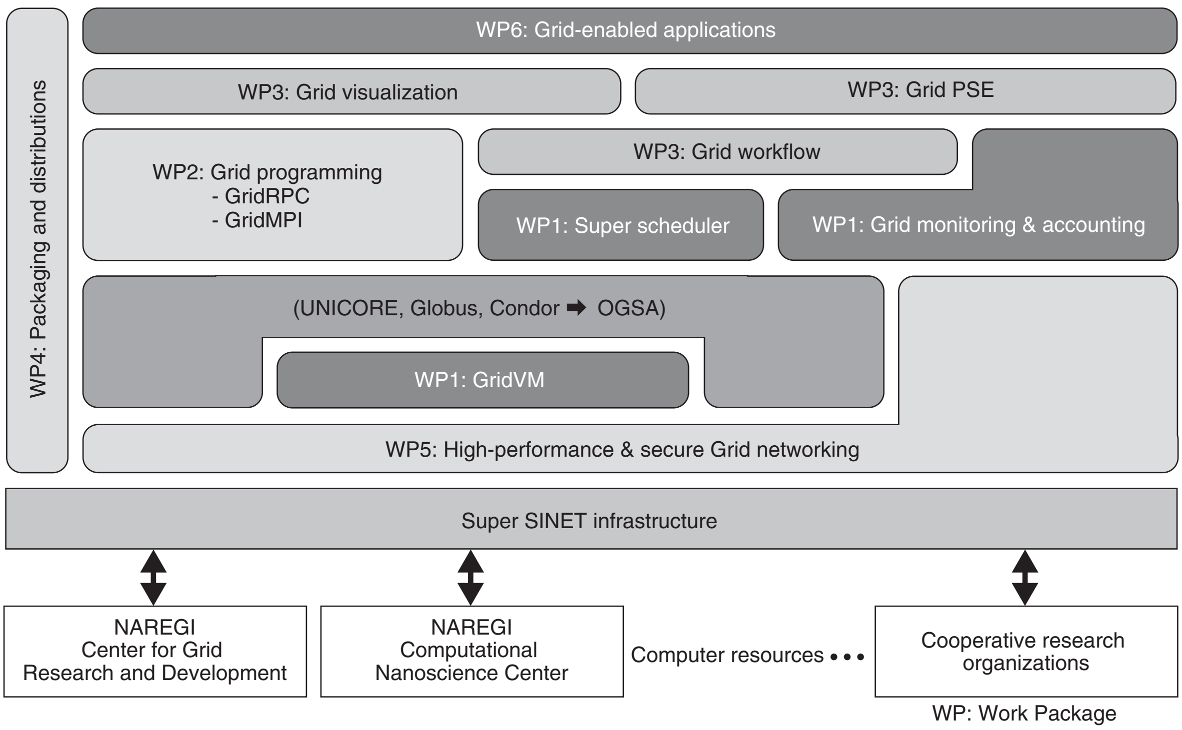

Figure 3
Research topics of NAREGI Grid middleware R&D and software stack.

tary to the experimental aspects in the "wet-lab."

It is to be hoped that, by taking advantage of computational science technology such as large-scale simulations on Grids, a new methodology can be established to solve the types of nano-scale problems that cannot be solved by the conventional theories and methodologies.

The research and development of most advanced nanoscience applications are led and conducted by the Institute for Molecular Science in collaboration with research institutions such as the Institute for Solid State Physics of the University of Tokyo, Institute for Materials Research of Tohoku University, Institute for Chemical Research of Kyoto University, National Institute of Advanced Industrial Science and Technology, and industry. The research subjects are as follows:

1) Functional nano-scale molecules

This research subject involves the establishment of technology to practically design and analyze nano-scale molecules that have desired structures, physical properties, and functions. This research will be conducted with simulations based on quantum chemistry computations such as the molecular orbital (MO) method, density functional theory (DFT), and the fragment molecular orbital method (FMO)[10] as a combination of MO method and DFT. This research will include suggested applications of the computational technology to develop new high-efficiency catalysts, molecular-scale electronics, and molecular magnetism.

2) Nano-scale molecular assembly

This involves the clarification of the structural formation of nano-scale molecular assemblies and single nano-scale molecules by self-organization and the theory of function representation based on the structures by use of the molecular dynamics method, generalized ensemble algorithms, and statistical mechanics computation based on the reference interaction site model (RISM) theory. The research subjects here include the establishment of a simulation methodology to estimate the structures and functions of proteins and bio-membranes.

3) Nano-scale electronic system

This involves the execution of high-accuracy, large-scale numerical computation based on the solid-state electron theory for the purpose of controlling the internal degrees of freedom (charge, spin, and orbit) of electrons by analyzing their behavior at each degree of freedom individually and the correlations among them. The research subjects here include the establishment of the basis for developing next-generation electronic technologies, for example, nano-scale electronic devices that have the functionality of optical switches and other new functions.

4) Nano-scale magnetism

This involves the development of a new computational method concerning 1) the spin state based on the non-empirical calculation of electron state and quantum Monte Carlo method; 2) the clarification of the mechanism for presenting the magnetic and dielectric properties of nano-scale substances by large-scale numerical computation; and 3) the establishment of a fundamental research platform for developing functionally new devices, especially nano-scale magnetic devices, including superdense magnetic memory.

5) Design of nano-scale complex systems

This involves the clarification of the properties of the individual nano-scale components of a complex system that are required to apply nanotechnology by using a simulation based on the fragment molecular orbital method (FMO) and the first principle molecular dynamics methods and the properties of a complex system that contains a macro circuit substrate and nano-scale components, for example, quantum wires and quantum dots, and a phase-separated alloy system from various viewpoints.

6) Integrated nano-scale simulation system

This involves the development of an integrated nano-simulation system by conducting research on how to integrate the programs and data of the nonempirical molecular orbital method, the density functional theory, the molecular dynamics method, and the statistical mechanics computation to be used for nanoscience and nanotechnology research in this project.

7) Verification of nano-scale design

Research proposals from industry have been solicited with applications that meet the demands from the industrial world in order to verify the concept and obtain useful feedback to the above research and development. Through such efforts, it will be possible to construct a Grid-enabled simulation system based on highly efficient methodologies and easy-to-use software.

Figure 4 outlines the research subjects at the Computational Nanoscience Center and their expected contributions to industry.

3. Grid and nanoscience/ nanotechnology simulation

It is expected that applying the Grid technology to computations in nanoscience and nanotechnology will enable ultra large-scale distributed simulations, large-scale linked computations and coupled simulations requiring complicated workflows, and simulations using extensively distributed data, which have been difficult to accomplish on a single supercomputer because of limited computing power and storage capacity. Some examples of expected applications of Grid technology are as follows:

1) Ultra large-scale distributed simulation on the Grid

Tools will be developed based on the Remote Procedure Call and the high-performance communication libraries on the Grid. These tools will enable large-scale simulations to be executed utilizing the distributed resources on the Grid.

2) Seamless data exchange and multi-scale linked/coupled simulations

The Grid middleware as a common platform will be developed to efficiently exchange data between nanoscience application programs running on different computers on a Grid.

Complex, large-scale linked computation and

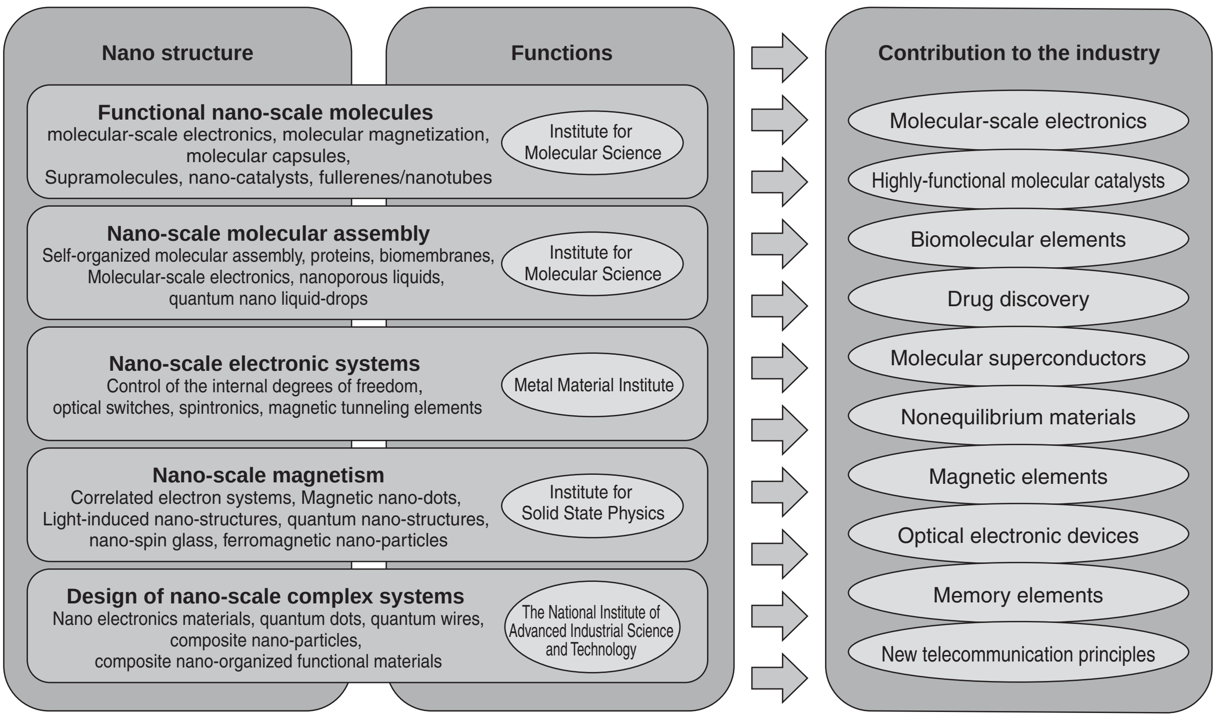

Figure 4
Research topics of NAREGI Computational Nanoscience R&D and their targeted applications in industry.

large-scale coupled simulations can be performed when the Grid middleware is built in with the data conversion capabilities, which are applicable to the wider multi-scale problems and multi-physics problems. In multi-scale problems, we need to analyze physical systems consisting of various models with different scales of time and length. In multi-physics problems, on the other hand, we need to analyze physical systems that are described as a combination of different physical models (e.g., the combination of classical mechanics and quantum mechanics models or of fluid and structural models). In solving multi-physics problems, to satisfy the conservation laws, the physical quantities at the boundary surfaces must be kept consistent as the numerical simulations proceed.

Figure 5 illustrates one example of coupled simulation. In this case, two application programs, FMO and RISM (Reference Interaction Site Model)[11] coupled to simulate the quantum behavior of a molecule under the influence of a solvent by using a Grid middleware called Mediator.

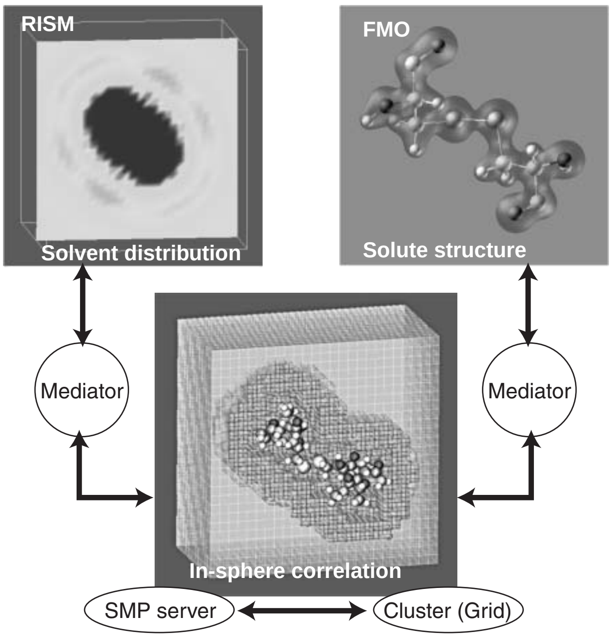

Figure 5
Example of coupled simulation.

3) Workflow Scenario in the Grid environment

When conducting large-scale parameter surveys and linking computing resources to form pipelines, and also when conducting collaborative research with remotely located researchers, the PSE will greatly enhance the efficiency of research and development activities. (The PSE is a simulation environment providing flexible access, linkage, and deployment of the nanoscience application programs, computing modules, and libraries that have already been registered by researchers on a Grid.)

4. International cooperation and standardization

Grid research activities have been spreading worldwide not only in the US and Europe but also in the Asia Pacific region, where a research organization named the Asia Pacific Partnership for Grid Computing has been formed. The leading international organization for Grid research is the Global Grid Forum (GGF),[12] which has various working groups to hold discussions about necessary Grid functions and perform the tasks and documentation required for Grid standardization. GGF is based on the idea that the software components for Grids should be open and interoperable and individual Grid projects should not independently develop all the necessary software from scratch, but share useful products with each other. Therefore, the NAREGI project has a policy of actively participating in the conferences held by GGF and other organizations and deepening its cooperative relations with international Grid projects, for example, the Teragrid project[13] and CyberInfrastructure project[14] in the US, UNICORE forum[6] in Europe, and eInfrastructures project[15] in the EU. The NAREGI project also intends to disclose the achievements of software development as open sources.

5. Fujitsu's role in the NAREGI project

Because the NAREGI project is based on the combined efforts of academia, industry, and government, contributions from industry are highly expected. Fujitsu has sent its researchers on loan to the Center for Grid Research and Development of the National Institute of Informatics. The researchers are actively involved in the research and development of Grid software for assigned research themes, for example, a super-scheduler UNICORE-Condor linkage, PSEs, and Grid-enabled nanoscience applications, in cooperation with the internal research and development team of Fujitsu. As described above, the main parts of UNICORE, which is used as the basis for research and development of Grid middleware in the first phase of the project, have been developed by the former Fujitsu European Center for Information Technology (FECIT) (the predecessor of Fujitsu Laboratories of Europe) and Fujitsu Laboratories of Europe under Dr. David Snelling's guidance. Furthermore, Fujitsu is one of the key members of the Consortium for Industrial Application of Research Grid, which was founded to encourage the application of Grid technology in industry. With regard to nanoscience and nanotechnology fields, Fujitsu Laboratories has submitted a research proposal in response to a public invitation by the consortium and the proposal has been accepted.

6. Conclusion

The NAREGI project has just entered the second year and is now fully engaged in the software development phase. The research projects under the NAREGI project have made considerable progress in the development of prototype software, and some of them have already produced initial results, including GridRPC and GridMPI. These achievements were demonstrated in the international conference, SC2003,[16] held in November 2003 in Phoenix, Arizona in the US and also at the NAREGI Symposium in Tokyo in February 2004. In March 2004, a 10-TFLOPS-class computer facility was installed in the Computational Nanoscience Center of the Institute for Molecular Science and a 5-TFLOPS-class computer facility was installed at the Center for Grid Research and Development of the National Institute of Informatics. These computer facilities will be connected to Super SINET to establish the NAREGI testbed. The NAREGI project is one of the economy revitalization projects participated in by Japan's representative organizations in industry, academia, and government research laboratories. It is hoped that the research Grid will also spread to industry, bring a large economic benefit, and enhance the international competitiveness of Japan.

Acknowledgement

The author expresses his thanks to the researchers in the NAREGI project for offering valuable information and materials for this paper.

References

1) ITBL: Information Technology Based Laboratory. *http://www.itbl.riken.go.jp*
2) Super SINET: Super Science Information Network. *http://www.sinet.ad.jp/english/index.html*
3) BioGrid. *http://www.biogrid.jp*
4) VizGrid. *http://www.vizgrid.org/en/index.htm*
5) NAREGI: National Research Grid Initiative. *http://www.naregi.org/index_e.html* (National Institute of Informatics) *http://ccinfo.ims.ac.jp/nanogrid/index_e.html* (Institute for Molecular Science)
6) UNICORE: UNiform Interface to COmputing Resources. *http://www.unicore.org*
7) Globus Toolkit. *http://www.globus.org*
8) Condor. *http://www.cs.wisc.edu/condor/*
9) OGSA: Open Grid Services Architecture. *http://www.globus.org/ogsa/*
10) K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi: Fragment molecular orbital method: An approximate computational method for large molecules. *Chem. Phys. Lett.*, **313**, p.701-706 (1999).
11) H. Sato, F. Hirata, and S. Kato: *J. Chem. Phys.*, **105**, p.1546 (1996).
12) GGF: Global Grid Forum.

http://www.globalgridforum.org
13) Teragrid.
http://www.teragrid.org
14) CyberInfrastructure.
http://www.cise.nsf.gov
15) eInfrastructures.
http://www.einfrastructures.org/
16) SC2003: Supercomputing Conference 2003.
http://www.sc-conference.org/sc2003/

Dr. Kenichi Miura received the Ph.D. degree in Computer Science from the University of Illinois, Urbana-Champaign in 1973. He joined Fujitsu in 1973, where he has been engaged in high-end computing. From 1992 to 1996, he was Vice President and General Manager of the Supercomputer Group at Fujitsu America, Inc., where he was responsible for supercomputing in the U.S. Since June 2002, he has been a Fellow of Fujitsu Laboratories, Ltd. He also served as a visiting professor at the Computer and Communications Center of Kyushu University from 1990 to 1993. Currently, he is a professor in High-End Computing at the National Institute of Informatics (NII) and the project leader of the Japanese National Research Grid Initiative Project.