SUPERCOMPUTING PROGRAMS AND ACTIVITIES FOR COMPUTATIONAL SCIENCE AND ENGINEERING IN JAPAN

Genki Yagawa[1,2], Mitsuo Yokokawa[2], Kenji Higuchi[2], Hideo Kaburaki[2] and Toshio Hirayama[2]

[1]The University of Tokyo
and
[2]Japan Atomic Energy Research Institute

Corresponding Address
yagawa@q.t.u-tokyo.ac.jp;m.yokokawa@aist.go.jp; higuchi@koma.jaeri.go.jp;
kaburaki@sugar.tokai.jaeri.go.jp; hirayamt@koma.jaeri.go.jp

ABSTRACT

Two major developments in the infrastructure of the computational science and engineering research in Japan are reviewed. Both of these developments, resulting from the recent construction of a high-speed backbone network and a huge vector parallel computer, will surely change the scene of the computational science and engineering researches. The first one is the ITBL (Information-Technology Based Laboratory) project, where R&D are made to realize a virtual research environment over the network. Here, basic software tools for distributed environments have been developed to solve science and engineering problems. The second one is the Earth Simulator project. In this project, a huge SMP-cluster vector parallel system was developed, which will undoubtedly give a great impact on the numerical simulations in the areas, for example, the climate modeling. Furthermore, activities in large-scale numerical simulations, which are carried out in various application fields and have a potential for further integration of the above systems, are presented.

1. INTRODUCTION

In Japan, computational science and engineering research environments have become both highly heterogeneous and homogeneous due to the construction of a high-speed backbone network, SuperSINET, and a huge SMP-cluster vector parallel system, the Earth Simulator. In the former case of a heterogeneous environment, researchers in the field of computational science and engineering face complex situations, such as execution of programs and mutual

collaborations over the distributed systems. Software tools for distributed communication have been developed under the ITBL project for the purpose of mitigating the complexity of the system. In the latter case of homogeneous environment, researchers pursue to attain high-performance and to solve large-scale problems using the Earth Simulator (ES). These two contrasting projects are presented.

The Japanese ITBL (Information-Technology Based Laboratory) project[1], which is a construction of a virtual research environment, was started in 2001. The project is now conducted through the collaboration of six organizations under the Ministry of Education, Culture, Sports, Science and Technology (MEXT); Japan Atomic Energy Research Institute(JAERI), RIKEN, National Institute for Materials Science (NIMS), National Aerospace Laboratory of Japan (NAL), National Research Institute for Earth Science and Disaster Prevention (NIED), and Japan Science and Technology Corporation (JST). In this report, an outline of the purposes, and the status of the ITBL project and features of the software infrastructure are summarized.

The Earth Simulator project was launched in order to comprehensively understand and forecast various global changes. The development of the Earth Simulator (ES) was commenced as a joint project of the National Space Agency of Japan (NASDA), the Japan Atomic Energy Research Institute (JAERI), and the Japan Marine Science and Technology Center (JAMSTEC) in 1997. The ES was successfully completed and put into operational use at the Yokohama Institute for Earth Science (YES/JAMSTEC) at the end of February 2002. Along with the establishment of this epoch-making hardware system, an atmospheric general circulation model (AGCM) called AFES has also been developed as a test-bed application code for the ES. In this paper, a general system configuration of the ES and a world-record performance attained by AFES on the ES are presented.

2. THE ITBL AND THE EARTH SIMULATOR PROJECTS

2.1 THE ITBL PROJECT

The advancements made in computers and high-speed networks allow for efficient research work performance by bridging together computational, data and experimental resources, which are physically distributed over multiple sites. This will also allow for the sharing of information among research collaborators located in various regions and/or organizations. In

order to support the advanced research environment, the notion of GRID[2] has been proposed. Many projects have been started to implement this idea of GRID computing (see *e.g.* [3], [4], [5]). ITBL is one of such projects. The target of the project is to connect computational resources including database situated mainly in the national laboratories and universities in Japan. At present, six organizations under MEXT have joined in on the ITBL project. JAERI is responsible for the development of a software infrastructure consisting of an authentication mechanism and a basic tool kit for supporting the development of various meta-applications. RIKEN will provide a secure network infrastructure based on the VPN (Virtual Private Network) technology. As for the actual applications, following are planned to be developed by the six organizations; Integrated aircraft simulation system (NAL), full cell simulation system (RIKEN), three-dimensional full-scale earthquake simulation system (NIED), materials design simulation system (NIMS and JST), and regional environmental simulation system (JAERI). The project is currently in the beginning phase. Several developments have been made to date. For instance, a prototype software infrastructure on a test-bed across six organizations has been constructed by JAERI. Furthermore, a secure VPN (Virtual Private Network) for the ITBL has been developed by RIKEN. Once the evaluations and improvements of these prototypes made by the above organizations are complete, the system will be ported to other non-member organizations under MEXT.

In designing the infrastructure, user-side mentalities were always considered so that they will benefit from the ITBL system. The potential users who are believed to benefit by using the ITBL system are;

(1) those who execute their computing jobs by selecting a light-loaded computer among those for which the user has accounts
(2) those who perform a very large or complicated simulation which cannot be executed on a single computer due to resource restrictions, or cannot be executed efficiently due to their complexity,
(3) those who execute a task which consists of programs and data on distributed resources, for example, processing experimental data on a supercomputer which is apart from the equipment,
(4) those who share information with cooperative research group members in different orgranizations.

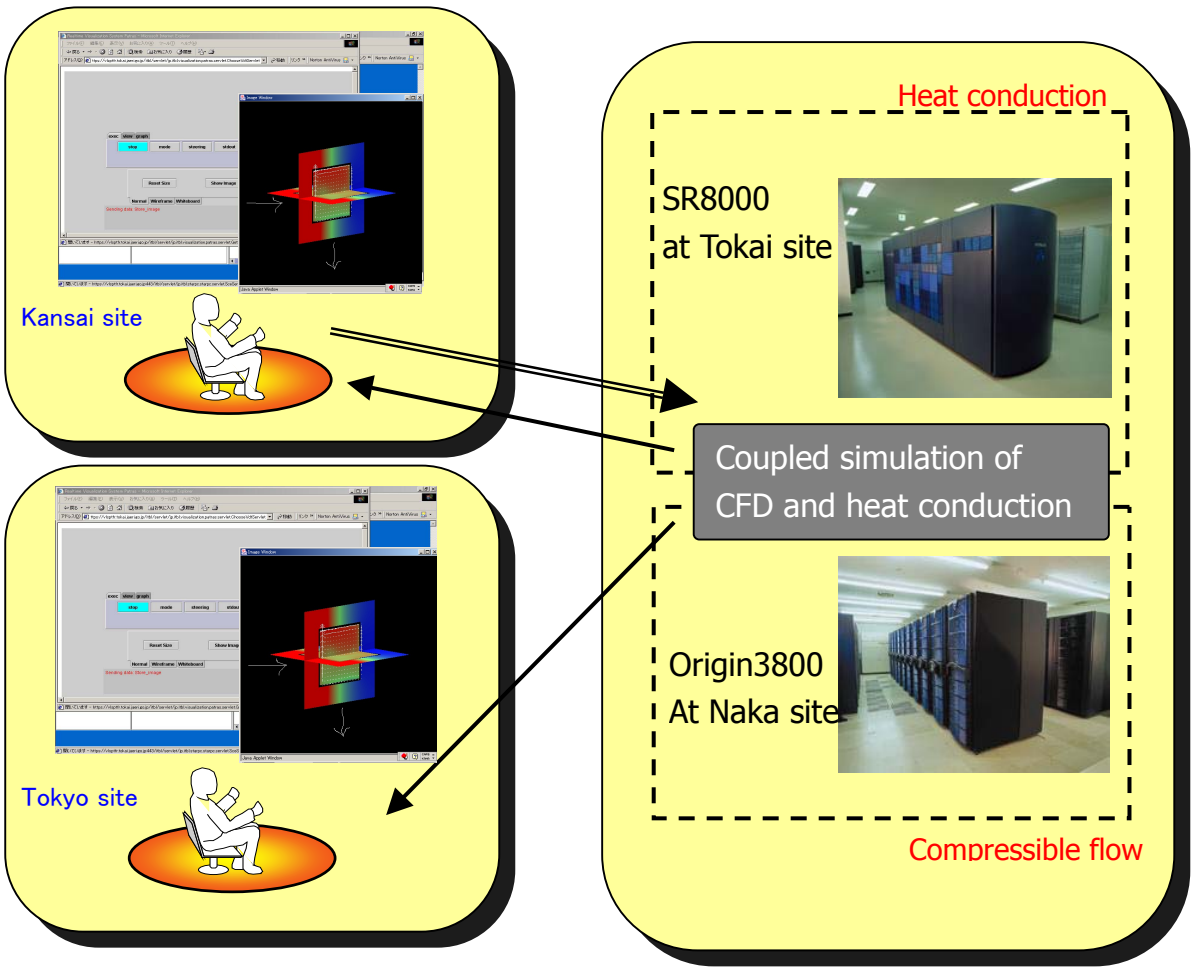

Figure 1. Coupled simulation of CFD and heat conduction and visualization using ITBL system

The users are supported by a component-programming environment and global data space, called a community, on a security system. Components are programs written in existing language and are considered to be the basic-building block of a meta-application. The component-programming environment is a tool kit that supports: the development of components on distributed computers, the assemblage of components into a meta-application, and the selection of computing resources on which programs are executed. A Community is a unified data repository that realizes a transparent access to the distributed data for members, while rejecting access from non-members. Security system provides a single sign-on mechanism based on the PKI (Public Key Infrastructure) X.509 certificates. It establishes easy and secure access to resources distributed among multiple sites.

The infrastructure is implemented by enhancing the functions of the STA system[6] that JAERI had developed in 1995. The system is already equipped with various tools that support component programming. The development costs of the ITBL infrastructure were considerably reduced by adopting the STA system as its basis. A prototype of software infrastructure has been completed, and its evaluation and improvements in functions such as security and communication are currently underway. The evaluations and improvements are being made mainly over four sites, Tokyo, Tokai, Naka, and Kansai sites of JAERI, which are connected through a high-speed backbone network called the SuperSINET. For instance, a coupled simulation of thermal-fluid analysis has been demonstrated, as shown in Figure 1. In the demonstration, simulations of fluid and heat flows were carried out on SGI Origin3800 at Naka site and Hitachi SR8000 at Tokai site, respectively. At the same time, the numerical results were visualized at Tokyo and Kansai sites simultaneously. In addition, the software infrastructure developed by JAERI is being ported to the VPN network infrastructure developed by RIKEN and will be implemented into the six member organizations.

2.2 THE EARTH SIMULATOR PROJECT

The ES is an SMP-cluster (Symmetric Multi-Processor) vector parallel system. It consists of 640 processing nodes interconnected by 640 x 640 single-stage crossbar switches. Each node contains 8 vector processors (AP) with a peak performance of 8 Gflops each, a shared memory system (MS) of 16 GB, a remote access control unit (RCU), and an I / O processor (IOP). This configuration, number of processors totaling to 5120 and memory amounting to 40TB, gives this machine a theoretical maximum performance of 40 Tflops [7, 8]. The extension of the NEC's SUPER-UX UNIX is implemented in the ES as its operating system. Each vector processor, AP, is equipped with a vector operation unit (VU), a 4-way super-scalar operation unit (SU), and a main memory access control unit, all of which are implemented on a single-chip LSI. The super-scalar operation unit, SU, has a 64KB instruction cache, a 64KB data cache, and 128 general-purpose scalar registers which are capable of executing such instructions as branch prediction, data pre-fetching, and out-of-order instructions. The VU consists of 8 logical sets of vector pipelines, vector registers, and some mask registers. Vector pipelines can handle 6 types of operations, i.e., addition/shift, multiplication, division, logical operations, masking, and load/store operations. Eight pipelines of the same type work together on a single vector instruction and different types of pipelines can be operated concurrently. In total, there are 72 vector registers, each having 256 vector elements. Both VUs and SUs support the IEEE754 floating point data format. The memory within each node (MS) is equally shared by 8 APs and is configured with 32 main

memory package units (MMU). As described in the previous passage, the memory capacity of each node is 16 GB. Here, each AP has a data transfer rate of 32 GB/s with the memory devices amassing the throughput to 256 GB/s per node.

The single-stage crossbar network (IN) consists of two units: inter-node crossbar control unit (XCT) and inter-node crossbar switch (XSW). Inter-node crossbar control unit (XCT) coordinates the switch operations and the inter-node crossbar switch (XSW) is an actual data path. XSW is composed of 128 separate switches, each of which has 1 byte bandwidth operated independently at a clock cycle of 8 nsec. Every pair - node and switch - is connected via electric cables. The theoretical data transfer rate between any two nodes is 12.3 GB / s x 2 ways.

The ES provides three-levels of parallel processing environments; i.e., vector processing on an AP, shared-memory parallel processing within a node, and SMP-cluster parallel processing among distributed nodes via the crossbar network. A message passing programming model implementing the MPI (Message Passing Interface) libraries is supported both within and among the nodes as a basic programming environment so that the three-level parallel processing environment can be used efficiently. The performance of MPI_Put functions is essential in attaining high performance. The performance of the MPI_Put function, which is beyond the scope of this study, was measured [9]. The maximum throughput and latency of MPI_Put are 11.63 GB/s and 6.63 μsec, respectively. It should be noted that the time for barrier synchronization is only 3.3 μsec because the system has a hardware system dedicated for barrier synchronization among nodes.

3. ACTIVITIES FOR LARGE SCALE SIMULATIONS IN VARIOUS FIELDS

3.1 HIGH PERFORMANCE SIMULATIONS OF THE AGCM CODE ON THE EARTH SIMULATOR

AFES is an optimized code developed by the Earth Simulator Research and Development Center (ESRDC). The software implements an atmospheric general circulation model NJR-SAGCM [10] and was developed with an intention to be run on the ES. The model NJR-SAGCM is based on the CCSR/NIES AGCM jointly developed by the Center for Climate System Research (CCSR) of the University of Tokyo and the National Institute for Environmental Studies, Japan (NIES) [11]. The original model is based on the three-

dimensional global hydrostatic primitive equations. The spectral transform method [12] is applied to discretize in the horizontal direction and a finite-difference method in the vertical direction with the use of sigma coordinates. AFES predicts such variables as horizontal winds, temperatures, ground-level pressure, specific humidity, and cloud water at grid points generated around the entire process.

Table 1: Scalable performance of AFES[13]

Total number of Aps	Elapsed time (sec)	Speed up	Tflops(ratio to peak)
80 = 80 nodes *1AP	238.037	1.000	0.52 (81.1%)
160 = 160 nodes *1AP	119.260	1.996	1.04 (81.0%)
320 = 320 nodes *1AP	60.516	3.933	2.04 (79.8%)
640 = 80 nodes *8AP	32.061	7.425	3.86 (75.3%)
1280 = 160 nodes *8AP	16.240	14.657	7.61 (74.3%)
2560 = 320 nodes *8AP	8.524	27.926	14.50 (70.8%)
5120 = 640 nodes *8AP	4.651	51.180	26.58 (64.9%)

The scalable performance of AFES with the T1279L96 resolution (grid interval of approximately 10 km around the equator) for different node configurations is presented in Table 1. It should be noted that these performance data were obtained for particular 10 time-integration steps during the simulation of 1 model day. As may be witnessed, AFES shows excellent sustained performance (65 to 75 %) and scalability for different node configurations. Table 1 indicates the speedup with respect to the number of processors used for the T1279L96-resolution simulation. AFES running on the ES shows remarkably high performance sustained over a wide range of processor configuration with 26.58 Tflops for the 640 full nodes (5,120 APs). It also achieved 23.93 Tflops for a 1-day model run by exploiting the entire ES system. These performance figures correspond to 64.9 % and 58.4 % of the theoretical peak performance, respectively, thus, surpassing the computational efficiency of the currently available weather and climate simulation models (typically 25 % - 50 %).

3.2 NUMERICAL SIMULATIONS IN COMPUTATIONAL SCIENCE AND ENGINEERING

With the development of the infrastructure described above, *e.g.* high-speed network and supercomputers, it is expected that the various fields of science and engineering will be studied numerically, capitalizing on the potential of the newly developed software and hardware. In the final analysis, tools must be verified through these applications. In the

Center for Promotion of Computational Science and Engineering of the Japan Atomic Energy Research Institute, researches on large-scale numerical simulations of material science, superconductors, thermal, hydraulic, and structural phenomea and their coupling, multiphase flow, and quantum bioinformatics have been conducted.

4. CONCLUSIONS

Prototype of software infrastructure has been constructed in ITBL project as a result of first stage. The project is ready to move to next stage. In other words, the software infrastructure is being sophisticated combining ITBL secure VPN from the viewpoint of practice. Various kinds of scientific applications on organizations of MEXT are being employed as testbeds. After the evaluation and improvement, the software and hardware infrastructure developed in this project will be offered for practical use to scientific researchers in Japan.

The developments of the ES and AFES were successfully completed as initially planned achieving extremely high performance of 26.58 Tflops (64.9 % of peak performance) for a very high-resolution global climate model with the full exploitation of 640 nodes of the ES. Along with the world-record performance of 35.86 Tflops (87.5 % of peak performance) measured on the LINPACK benchmark suite, these encouraging performance results can act as a driving force for the comprehensive understanding of the challenging issues, such as the impacts of anthropogenic emission of greenhouse gases. Today, improved computing capabilities are considered crucial to the scientific understanding and policy making required for sustainable habitation of our planet earth. The authors are expecting that this innovative system can serve as an intermediary between geosciences and computational sciences, and will be capable of providing solutions to possible scenarios of symbiotic global environmental and anthropogenic changes.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the members of the ITBL task force committee for their valuable comments on infrastructure design. We also would like to offer sincere condolence to late Hajime Miyoshi who initiated the Earth Simulator project with outstanding leadership.

REFERENCES

1. ITBL project Web Site: http://www.itbl.jp
2. Foster, I., and Kesselman, C., eds.: The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann (1998).
3. Johnston, W.E.: Using Computing and Data Grids for Large-scale Science and Engineering, the International Journal of High Performance Computing Applications, Vol. 15, No.3 (2001), pp.223-242.
4. Unicore project Web Site: http://www.unicore.de
5. Segal, B.: Grid Computing: The European Data Grid Project, IEEE Nuclear Science Symposium and Medical Imaging Conference (2000).
6. Takemiya, H. Imamura, T., Koide, H., Higuchi, K., Tsujita, Y., Yamagishi, N., Matsuda, K., Ueno, H., Hasegawa, Y., Kimura, T., Kitabata, H., Mochizuki, Y., and Hirayama, T.: Software Environment for Local Area Metacomputing, SNA2000 Proceedings, (2000)
7. M. Yokokawa, S. Shingu, S. Kawai, K. Tani, and H. Miyoshi, "Performance Estimation of the Earth Simulator", Proceedings of 8th ECMWF Workshop, pp. 34-53, World Scientific (1998).
8. K. Yoshida, S. Shingu, "Research and Development of the Earth Simulator", Proceedings of 9th ECMWF Workshop, pp. 1-13, World Scientific (2000).
9. H. Uehara, M. Tamura, and M. Yokokawa, "An MPI Benchmark Program Library and Its Application to the Earth Simulator", LNSC 2327 (2002).
10. Y. Tanaka, N. Goto, M. Kaei, T. Inoue, Y. Yamagishi, M. Kanazawa, H. Nakamura, "Parallel Computational Design of NJR Global Climate Models", High Performance Computing ISHPC'99 Proceedings, pp281-291, Springer (1999).
11. A Numaguti, S. Sugata, M. Takahashi, T. Nakajima, and A. Sumi, "Study on the Climate System and Mass Transport by a Climate Model", CGER's Supercomputer Monograph Report, **3**, Center for Global Environmental Research, National Institute for Environmental Studies (1997).
12. I. T. Foster, and P. H. Worley, "Parallel Algorithms for the Spectral Transform Method", ORNL/TM-12507 (1994).
13. S.Singu, H.Takahara, H.Fuchigami, M.Yamada, Y.Tsuda, W.Ohfuchi, Y.Sasaki, K.Kobayashi, T.Hagiwara, S.Habata, M.Yokokawa, H.Itoh, K.Otsuka, "A 26.58 Tflops Global Atmospheric Simulation with the Spectral Transform Method on the Earth Simulator", Proceedings of the IEEE/ACM SC2002 Conference