Computational Science Roadmap

-Overview-

Social Contributions and Scientific Outcomes Aimed for by Innovations through Large-Scale Parallel Computing

May, 2014

Feasibility Study on Future HPC Infrastructures

(Application Working Group)

Table of contents

Introduction

Super computers are essential to acquire and discover knowledge in contemporary science and technology. At the same time, supercomputers work behind the scenes to support the everyday lives of general citizens. Our country is currently faced with a mountain of difficult problems, such as the recovery after the Great East Japan Earthquake, the containment and environmental clean up after the Fukushima nuclear disaster, energy issues, a decreasing birth rate and an aging population, and fiscal stringency. Large-scale simulations using supercomputers will not only be a driving force for science and technology but also play an important role in solving these complex issues we are facing. Supercomputers are a fundamental technology essential for the solid underpinning of Japanese society and ushering in the age of tomorrow.

Research outcomes and methods obtained through the supercomputers, which have been introduced at the initiative of the central government, further extend the forefront of science and technology and can lead to industrial vitalization in the next era through the adoption of supercomputers by enterprises. Moreover, it will potentially improve skills and services such as medical practice or weather forecasting. These technologies developed with the help of these supercomputers must ultimately be used in industry and society. In other words, spreading the use of computational science will be increasingly important, and simulations and large-scale data processing will play a progressively pivotal role in bringing substantial benefits to society.

In the context of the increase in the contributions of supercomputers to society, the “Working Group for the Investigation of HPC Technology R&D of the Future” was established in 2011 under the “Council for HPCI Plan and Promotion” which is advisory commission of the Research Promotion Bureau of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). It aims to investigate Japan’s requirements for future HPC research and development as part of the HPCI plan promotion. This working group suggested setting up an Applications Working Group and a Computer Architecture/Compiler/System Software Working Group, and these working groups collaborated closely to create a White Paper for a Computational Science Roadmap[1] that was published in March 2012. Furthermore, to examine the working group’s discussed topics in more detail, MEXT sponsored the “Feasibility Study on Future HPC Infrastructures (Application Working Group)” in July 2012. In this study, computational science’s potential contributions to social problems and scientific breakthroughs were identified, and a new Computational Science Roadmap was drawn up as a result. Coming together to draw up this roadmap were not only computational scientists but also experimental, observational, and theoretical scientists included approximately 100 researchers working on the frontline of various scientific communities from universities, research institutions, and

[1] http://www.open-supercomputer.org/workshop/sdhpc/

corporations. Profound discussions were held not only on computing performance but also on suitable performance to balance the overall computing system required for solving social problems and for achieving expected scientific breakthroughs.

This overview of the Computational Science Roadmap outlines the objectives of computational science, and introduces concrete examples of the social problems that computational science may contribute to in the next five to ten years. It also introduces the new scientific problems that could arise from cooperation between various branches of science that have conventionally been regarded as distinct areas. To solve the social and scientific problems introduced in this overview, more intense involvement in the various research fields of computational science is essential. Details for this are given in chapter 4 of the Computational Science Roadmap, but this overview also outlines the individual research tasks for the respective fields in chapter 4.

The Future HPCI Plan and Promotion - Intermediate Report[2] was published on June 25, 2013 by the Working Group for the Study of the Future HPCI Plan and Promotion. This report refers not only to exascale computers taking the lead in supercomputing, but also to the necessity of a second level of supercomputers. The possibilities of large-scale computational science outlined in this report are discussed and clarified purely in terms of what they require for the future of science, and certainly not with only the highest-performing machines in mind.

From chapters 2 to 4, we assessed the extent of the computing resources that will be required by the group of applications designed to process “the tasks to be solved by the next generation.” We have summarized the findings in a requirement table. This table includes data such as computing method, effective computing performance (floating point operation or integer operation) and required memory as estimated from a profile based on current programs and existing machines. It should be noted that this data generally differs from the theoretical performance of actual computers, and that since the data represents a future prediction, it disregards fluctuations of a several-fold range.

The content of this roadmap is valid as of May, 2014. We plan to carry out a more detailed assessment in the future. The website will show the most up-to-date version of the Computational Science Roadmap. [3]

[2] http://www.mext.go.jp/b_menu/shingi/chousa/shinkou/028/gaiyou/1337595.htm

[3] http://hpci-aplfs.aics.riken.jp/top.html

1. Computational Science – Background

1.1 Developments to date in computational science and future prospects

In computational science, the use of supercomputers as a third science and technology technique next to theoretics and experiments is an essential research method that is indispensable to strengthening cutting-edge science and technology and industrial competitiveness. Through the development of computational science, we are now able to predict the future by modeling natural or social events and simulating (computing) them. For instance, weather forecasts or predictions on global warming cannot be achieved without computational science. Moreover, computational science is essential to our understanding of the unexplored field of fundamental physics, to the validation of elaborate theories, and to the application of theories on complex systems. It can be said that the Large Hadron Collider run by CERN, famed for the discovery of the Higgs particle, is the result of Big Science, but this too does not function without the power of computational science. The power of computational science is also essential to analyzing genetic information or economic, financial, or transport conditions.

With the dramatic increase in computer performance in recent years, the areas of application of computational science are increasing in both depth and range. When the first supercomputers saw the light in the 1970s, their range of application was limited to areas, such as strength calculations for buildings or comparatively simple-flow analyses, but it is now becoming possible to perform analyses of the structure, properties, and function of a wide variety of substances and materials. It is also now possible to do analyses at a genetic level and for the human body in its entirety.

Against this background, an innovative High Performance Computing Infrastructure (HPCI) is being built with RIKEN's K computer and Computing and Communications Centers of nine universities nationwide at its core, the purpose being to create outcomes that lead to solutions for a variety of social problems and to accelerate industrial use by offering a computing environment of the world's highest standards to researchers and engineers across a variety of disciplines.

1.2 K computer research outcomes in key areas

K computer is the supercomputer at the core of HPCI. Initiatives focusing on five identified strategic fields of "HPCI Strategic Programs for Innovative Research (SPIRE)" started in 2009 to create world-class research outcomes by leveraging K computer's capacity to the fullest, and to support initiatives to build organizations for the promotion of computational science in those fields.

K computer has been in full operation since fall 2012 and performs calculations using detailed models and calculations for phenomena in their entirety, which were hitherto unavailable due to insufficient computing power.

- Strategic Field 1 – Supercomputational Life Science: Fundamental research to gain a detailed understanding of life phenomena on a cellular level and an understanding of the essence of life, as well as research to obtain fundamental knowledge to solve social problems, including substantial acceleration of drug development and the realization of personalized medicine.
- Strategic Field 2 – New Materials and Energy Creation: Research to elucidate and predict the functions of substances and materials and the electric functions of nanostructured devices, based on basic theory; expected to be important technology for the exploration of high-temperature superconducting materials, high-efficiency thermoelectric elements, catalysts for fuel cells, etc.
- Strategic Field 3 – Advanced Prediction Researches for Natural Disaster Prevention and Reduction: Research for more precise simulations of global-scale environmental change, and research leading to predicting localized heavy rainfall immediately before the event; also, research leading to damage prediction for earthquakes or tsunamis.
- Strategic Field 4 – Industrial Innovation: Research leading to enhanced manufacturing, such as simulation technologies to substantially increase the speed and decrease costs of design and development processes for advanced fluidic devices and nanocarbon devices, as well as entirely earthquake-proof nuclear-reactor-plant simulations.
- Strategic Field 5 – The Origin of Matter and the Universe: Simulation-based research to understand the origin of matter and the universe and the laws governing them through particle accelerator experiments and through the observation of extreme astronomical events, such as black holes or supernova explosions.

1.3 Plan for the next generation HPCI

Supercomputer processing power is expected to rapidly progress, and this evolution in computing the range of areas, where computational science takes on an active role, is expected to get increasingly larger, creating outcomes that relate directly to a variety of problems in society and to industrial competitiveness.

Reflecting on these circumstances, the plan for the next-generation HPCI is now being investigated knowledge that Japan needs to drive research and development in a strategic fashion and from a continuing, long-term perspective.

For the next-generation HPCI plan, more importance than ever is placed on social contributions from computational science, and in line with that awareness, the following basic agreement was reached.

- Build a system not to simply boast the peak-performance achievement, but to thoroughly investigate, scrutinize, and implement the envisioned scientific breakthroughs and the social problems that must be solved by leveraging HPCI.
- Distil as many research tasks as possible that will potentially require HPCI in the medium to long term without restricting it to just those users currently using HPCI, such as the K computer.
- Depending on the type of application, computing can be faster and far more efficient with a specially designed machine than a general-purpose one. We presume the introduction of multiple architectures in addition to a general-purpose architecture to allow for the effective utilization of computational resources.
- Applications requiring HPCI technology shall not be limited to large-scale simulations, but we will also consider how to efficiently analyze large-scale data that are collected from large-scale experimental (observational) facilities being used in the various research areas.

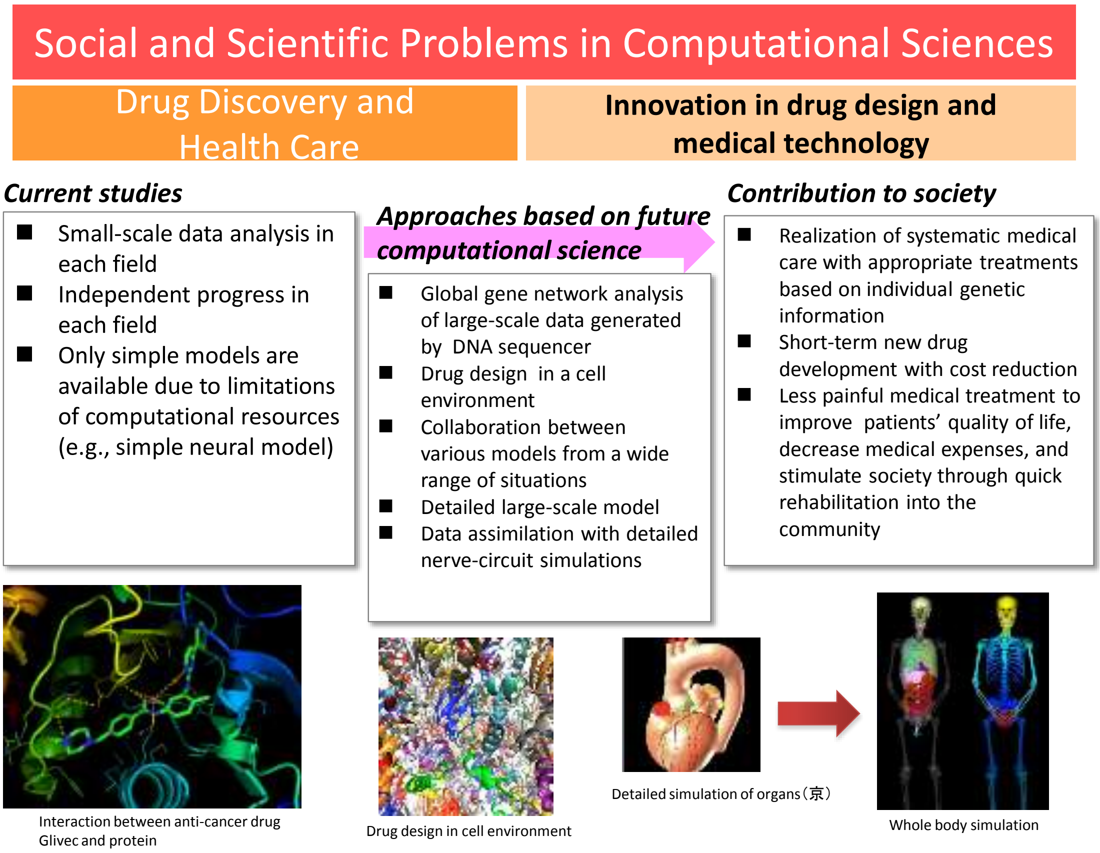
2. Social problems to which computational science may contribute in the future

It is indisputable that large-scale numerical computing makes an essential contribution to the industrial and economic activity supporting our society's current way of life, and the results of further improvements in supercomputing performance will contribute to solving a variety of issues our society is currently faced with. We will describe the concrete contributions that computational science aims to achieve for future social problems in the following four areas: drug discovery and healthcare, disaster prevention and mitigation, energy and environmental problems, and social and economic predictions.

Social problem	Concrete contribution
Drug discovery and healthcare	Innovative drug design and medical technologies.
Disaster prevention and mitigation	Systemization of disaster prediction based on scientific knowledge.
Energy and environmental problems	Harmony between energy technologies and the environment.
Social and economic predictions	Forecasting systems that respond flexibly to social and economic activity.

2.1 Drug discovery and health care

Japan is a rapidly aging society, and the promotion of health is an exceedingly important national issue. To create the innovative drug-discovery methods and medical technologies that will contribute to the promotion of health, it is essential that we have an understanding of the basic life phenomena of the human body. However, too many elements of these life phenomena are intricately entangled with each other, and there is an essential need for large-scale data analysis, such as genetic information, for simulations linking life science and materials science and for medical applications of multi-scale simulations ranging from molecular-scale to cellular-scale, organ-scale, brain-scale and whole-body scale simulations. In practical terms, massive sets of individual genomic data such as genome sequences obtained through next-generation DNA sequencing, which allows for the reading of genomic data at ultra-high speed, will be used to analyze genetic networks representing relationships between multiple genes. This will shed light on the causes of diseases affected by compound factors, such as cancer, and allow us to aim for tailor-made medical treatments providing optimal individual therapies based on individual genetic information. Moreover, using highly reliable simulation techniques to predict protein-drug interactions, for instance, as well as using simulations in whole environments, including cells and viruses, will substantially reduce the cost of new-drug development and significantly shorten lead times. These simulations also aim to contribute to the development of nano-molecular materials that possess new functionalities through the application of biomolecules. Multi-scale simulations, including molecular-scale, cellular-scale, organ-scale, brain-scale, and whole-body-scale simulations, are also instrumental in our understanding of complex disease mechanisms, such as blood clot formation in the heart or brain infarctions, and will be effective in improving patients' Quality of Life (QOL) through the development of minimally invasive treatments, which only pose a slight burden to the patient, and of the medical devices required for these treatments. It will further be effective in revitalizing society through patients' early re-entry into the community and in reducing costs of medical treatment.



The supercomputer's vast computational power will undoubtedly greatly contribute to the development of various aspects in the field of life science, such as detailed neural and cellular simulations, simulations over extended periods of time and space, and almost real-time assimilation[4] of those data. Eventually it could form an important scientific basis for innovative drug design and medical technologies.

The table below lists the computational performance required in the future for the respective areas of drug discovery and healthcare.

[4] One of the methods to merge different observational and experimental data into a numerical model at a high degree.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Personal Genome Analysis	0.0054	0.0001	1.6	0.1	0.7	200000	2700	Sequence matching	Cancer Genome Analysis: Short read mapping and mutation identification of 200,000 people's genome	1 case = 1 person Integer operations are dominant. "Total operation count"= total instruction count (Total FLOP = 46 EFLOP)
Gene Network Analysis	25	89	0.08	0.016	0.34	26000	780000	Baysian network estimation and L1-regularization	40,000 transcripts x 26,000 data sets consisting of 2,800,000 arrays	
MD and Free-energy calculation for drug design and so on	1000	400	0.0001		0.0012	1000000	4300000	Molecular dynamics simulation with all-atom model	Number of Cases: 100,000 ligands X 10 target proteins	B/F=0.4. Supposed to run 100-1000 cases simultaneously. Memory size per case is estimated for a 100 node run.
MD simulations under cellular environments or MD simulations of Virus	490	49	0.2	1.2	48	10	850000	Molecular dynamics simulations with all-atom / coarse-grained model	100,000,000 particles	B/F=0.1
Simulations of cellular signaling pathways	42	100	10	10	240	100	3600000	Microscopic lattice reaction-diffusion simulation	1,000 to 10,000 cells	integer operations
Precise Structure-Based Drug Design	0.83	0.14	1	0.001	1	100	300	Ab initio quantum chemical calculations on the interactions between proteins and drugs	proteins (500 residues) + ligands in solution	1TB/s IO speed required to dump 1TB dataset per second
Design of Biological Devices	1.1	0.19	1	0.001	1	100	400	Spectroscopic analyses of proteins (200-500 residues)	more than 100,000 orbitals	1TB/s IO speed required to dump 1TB dataset per second
Multi-scale simulation of a blood clot	400	64	1	1	170	10	2500000	Semi-implicit FDM simulation of fluid-structure interaction with chemical factors	Length:100mm, D:100um, Calculation Time:10s, Grid size:0.1um, Velocity:10^-2m/s, Delta T:1us	
High Intensity Focused Ultrasound	380	460	54	64	240	10	3300000	Explicit FDM simulation of sound wave and heat transfer	Calculation Area:400mm^3, Grid: 225x10^12, Steps: 1459200, FLOP/grid/step: 1000	
Simulations of Brain and Neural Systems	* 6.9	* 7.6	* 56	* 3600	0.28	100	700	Single compartment model	100 billion neuons, 10000 synapses/neuron, 10^5steps	
Data assimilation of whole insect brain via communication between a phisological experiment and a simulation, Parameter estimator in insect brain simulation	* 71	* 60	* 0.2	* 20	28	20	140000	Multi-compartment HH model with local Crank-Nicolson method, evolutionary algorithm	1000 neurons, 10^6 genes, 100 generations	Supposing 100 MB/s communication to external environment will be required

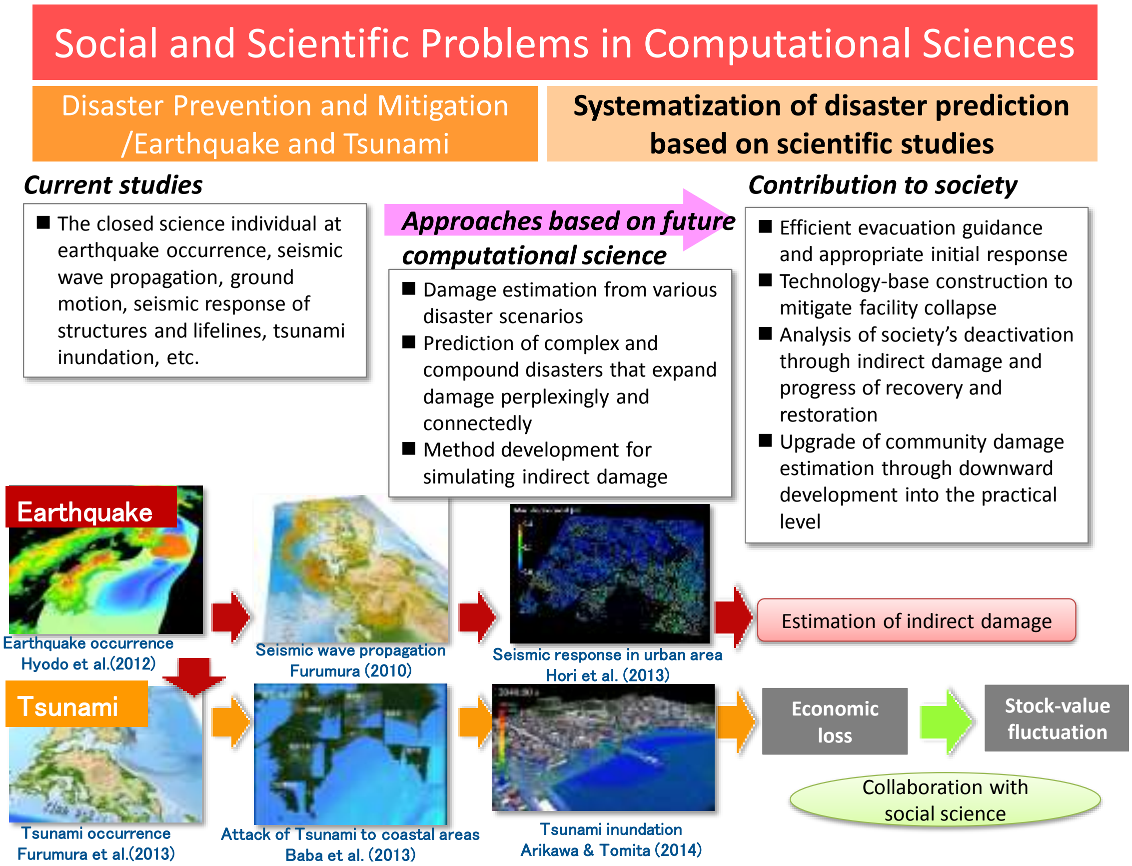
*Figures marked with a * are still under examination. The website will show more accurate figures as they become available.*

2.2 Disaster prevention and mitigation

(1) Earthquake and tsunami prevention and mitigation

Needless to say, since the Great East Japan Earthquake struck our country on March 11, 2011, disaster prevention and mitigation has been a pressing issue for Japan, and it is our destiny to always be prepared for major earthquakes, be it a massive earthquake in the Nankai Trough or an earthquake in the Tokyo metropolitan area. Highly accurate damage estimations are necessary to enable us to reasonably prepare, requiring rational and scientific earthquake estimations and predictions; simulations using large-scale numerical computing for earthquakes and tsunamis and accompanying disasters could be the trump card for this.

Currently, predicting earthquakes and tsunamis is difficult, but large-scale numerical computing using computational science enables us to make a variety of damage estimations based on earthquake- or tsunami-disaster scenarios and to predict so-called complex emergencies, where forces external to the disaster itself further multiply and spread the damage. This is immediately useful in more efficient post-disaster evacuation guidance and in preparing the appropriate initial response immediately after an earthquake. Furthermore, predicting the damage caused by a disaster is useful in improving the strength of buildings, facilities, and coastal defenses, and in building a social technology base to reduce as much as possible the probability of collapse. In practical terms, cutting-edge computers and the computational science technology deploying them will enable us to compute over 1,000 different disaster scenarios and estimate the damage for each, and by creating a database, we aim to provide a system that can be put to use immediately.



In addition to the direct damage to structures and towns, with the ever-increasing globalization of economic activity, earthquake disasters in the future will more seriously damage the economic activity in the disaster-affected towns and regions. The development of simulations of indirect damage is, therefore, also an important task to analyze the drop in economic activity immediately after an earthquake and the recovery of economic activity depending on the recuperation progress from the disaster.

These simulations leading to damage estimates based on earthquake scenarios are extremely important for risk management in Japan. However, tracking complex human behavior in natural settings entirely with computers is very difficult, and we must, using the most advanced supercomputers available at the time, continuingly develop and enhance the analytic techniques and models. At the same time, with the speed increase of supercomputers every five or ten years, it is also important to gradually deploy cutting-edge simulation tools developed on the supercomputers in the previous period at both the research and practical levels in research institutions and universities. This will allow for continuing enhancement of damage estimation not only at the national level but also at the local authority or corporation level, and for a continuing increase in reliability.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Coupled simulation of natural hazard and disaster (prediction of earthquake disaster right after earthquake occurrence) (1) – (6)	7	15	0.1	9	3		310000			(1): 5 domains, 1000 cases for each domain. (2): 20 cases are selected from 1000 cases based on observation (for each domain). (3),(4),(5),(6): 1000 cases for each domain; (parameter studies about the ground structures, the deterioration of building and the submarine topography) * (several large city areas) * (selected case in (2)) Required BF ratio = 8.0
(1) earthquake generation			0.00086	0.00086		5000	48	earthquake cycle simulation using boundary integral method	number of elements: 10ˆ7	Required B/F ratio = 4
(2) seismic wave propagation			0.1	0.5		100	1400	computation of elastic wave propagation using finite difference method	1200 x 1000 x 200 kmˆ3 (grid of 125 x 125 x 62.5 m), 240,000 steps	Original version: Required B/F ratio = 2.14, 14EFLOP per case New version on K by Dr. Maeda: Required B/F ratio =1.4, 20EFLOP per case
(3) ground motion amplification			0.01	4		5000	130000	computation of seismic wave propagation using finite element method	number of nodes: 300,000,000,000 (300 x 250 x 10 kmˆ3)	Required B/F ratio = 8
(4) ground motion amplification			0.01	4		5000	130000	computation of seismic wave propagation using finite element method	number of nodes: 300,000,000 (30 x 25 x 1 kmˆ3)	Required B/F ratio = 8
(5) structure seismic response			0.05	0.05		5000	500		number of buildings: 1,000,000	Estimated from the profile data. Required B/F ratio = 0.26
(6) tsunami inundation			0.002	0.5		5000	50000	Computation of Navier-Stokes equations (hydrostatic state approximation, non-hydrostatic state approximation, Volume of Fluid Method)	Concurrent execution of 7 cities with composite grid, which domains are 3 x 3 x 0.08 (1 city region, grid width: 1m) to 1400 x 1100 x 10 km (grid width: 5.4km), 720,000 steps.	Estimated from the profile data. Required B/F ratio = 10
mass evacuation simulation	3.3	0.28	0.3	0.006	1	5000	60000	human behavior simulation using multi agent model	300,000 agents, 18,000 steps (1 hour simulation), 1,000 Monte-Carlo members	For total number of the operations, instruction is used instead of FLOP (IPS is 40 times larger than FLOPS). Estimated from the profile data on K.

Note: (1) – (6) represents breakdown of "Coupled simulation of natural hazard and disaster".

(2) Meteorological disasters

Japan is prone to various meteorological disasters such as localized heavy rainfall. With 70 percent of the country consisting of mountainous or hilly areas, there is a high risk of mudslides, landslides, and slope failures. The country's high concentration of population and assets in the plains leaves it at risk of floods and high tides and exposes it to the peril of storm and flood damage. Events constituting meteorological disasters are not only typhoons and localized heavy rainfall, which last for comparatively short periods of time, or local high-impact events such as tornados, but also longer-lasting events such as snow cover and droughts. Predictions over a broad range of timescales, from tens of minutes to several months in advance, are essential to reducing economic losses caused by these events. Predicting how these weather and climate events will change with the progress of global warming, as well as increasing predictive accuracy, are pressing issues for our country when considering the future of disaster-prevention measures. The increasingly serious transboundary air pollution, such as PM2.5, poses another current environmental issue. It is important for future disaster mitigation that we engage with these issues and promote nationwide cutting-edge research and development by using the highest-performing computers available.

In accordance with these circumstances, we can list the following research subjects that must be carried out using the next-generation supercomputer for the purpose of meteorological disaster mitigation: prediction studies for high-impact weather events based on more sophisticated techniques to prepare initial conditions, climate-change prediction studies using high-resolution global cloud-resolving computations, and studies on detailed cloud physics processes and turbulence processes.

Firstly, an important problem in weather forecasting that uses numerical models is initial-value accuracy. In short-term weather predictions, the initial condition is paramount, and the study of data assimilation methods, such as the 4D-VAR or the ensemble Kalman filter, is an important target for computing environments of the next era.

Social and Scientific Problems in Computational Sciences

Disaster Prevention and Mitigation /Meteorological Disaster

Measures against meteorological/climatological disasters by advanced prediction

Current studies

- Prediction with insufficient resolution and initialization
- Prediction without objective information on reliability
- Warning system for disasters based mainly on nowcasting

Approaches based on future computational science

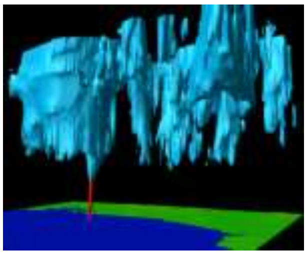
- Projection for climate change and typhoons by global cloud resolving simulations
- High-resolution modeling and improved initialization based on advanced schemes
- Prediction with information on reliability
- Model Improvement through sophisticated physical processes for cloud and turbulence

Contribution to society

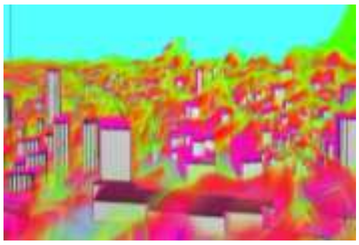
- Typhoon prediction with a much longer lead time
- Reduction of uncertainties in risk assessment for meteorological disasters by climate change
- Short range prediction for local heavy rains and meso-cyclones causing tornadoes
- Prediction by super-high resolution models

Figures offered by: Ryuji Yoshida/ AICS (LEFT), Wataru Mashiko/ Meteorological Research Institute (CENTER), Guixing Chen/ Tohoku Univ. (RIGHT)

Typhoon prediction, climate projection



Torrential rains, tornadoes



Super-high resolution simulation

Secondly, the spatial resolution of the numerical models used for the weather and climate are still too low. In numerical models, approximations applying parameterizations are used for detailed weather forecasts, but these sometimes form a major cause for the unreliability of a forecast. Historically, cumulus convection was parameterized. This technique is sufficiently effective for predicting mid-latitude synoptic scale phenomena, but not applicable for the direct expression of cumulonimbus that cause the high-impact events relating to meteorological disasters. Therefore, future predictions of high-impact events that cause meteorological disasters, such as typhoons, are in need of climate-change predictions using high-resolution global cloud-resolving computations that directly express cumulus convection without parameterization.

Furthermore, the development and validation of more sophisticated numerical models is also quickly progressing for not only cumulus convection but also the bulk parameterization physics of cloud particles and raindrops, and turbulence at the atmospheric boundary layer.

Expectations are high for the above-mentioned issues. The outcomes of the trailblazing attempts by the K computer and the next-generation supercomputers in the meteorological and climate fields will most certainly contribute to future meteorological prediction and climate projection. The appearance of the Earth Simulator in 2002, for instance, allowed us to obtain more reliable information on the intensity of typhoons and frequent rainfalls accompanying climate change. In addition to directly contributing to IPCC reports, it has also provided important technical information for the development of operational numerical forecasting systems at the Japan Meteorological Agency.

In next-generation supercomputing, large-scale computing of the K computer promises to be applicable to many cases and to allow for feasibility study, including, for instance, typhoon prediction with longer lead times than currently possible and the the very short-range prediction of mesocyclones that bring localized, severe rainfall or tornados. With the implementation of storm-scale meteorological disaster predictions, it is essential to conduct further research into the technical problem of assimilating high-frequency, high-density observational data and to build systems that swiftly accumulate such data, forming one of the keys to the effective use of big data. Moreover, quantitative evaluations of forecast errors on all scales through ensemble forecasting will substantially increase the value of information obtained by numerical weather prediction for the purpose of disaster mitigation. If we can improve disaster alerts, which are currently mainly based on nowcasting, with probabilistic predictions that have a longer lead-time or indications of worst-case scenarios, we can take disaster-prevention measures based on risk and cost management, potentially substantially reducing the damage caused by meteorological disasters.

The computational requirements to pursue these subjects are summarized in the table below.

Subject	Perfor- mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
High resolution global weather prediction	130	360	3	58	340	1	150000	Model: NICAM, FVM	1,000,000,000,000grids, 5,200,000steps(dt=1sec, 2months integration)	The number of the assumed total nodes is 100,000 (side-by-side comm.:1GB/s).
High resolution regional weather prediction	33	* 33	0.09	0.3	0.5	2700	160000	Model: asuca, FVM	7,500x7,500x500grids, 130,000steps(dt=1sec, 36hours integration)	FLOP and memory usage are estimated from a profile data obtained on SR16000. The memory throughput is estimated by an assumption that B/F is one. 25 variables are output every 10 minutes. 22,500 nodes are assumed for estimation of communication. (side-by side comm; 40GB/s)
Heavy rain prediction and tropical weather forecast	220	270	0.7	5	580	2	900000	NICAM + Data assimilation: Local Ensemble Transform Kalman Filter(LETKF)	3.5km horizontal resolution, 100 vertical layers, 1,000 ensemble members, 3 hourly assimilation cycle, 2 months integration	The number of the assumed total nodes is 100,000 (side-by-side comm.:1GB/s). FLOP, memory throughput and memory usage are estimated based on a profile data obtained on K.

*Figure marked with a * is still under examination. The website will show more accurate figures as they become available.*

2.3 Energy and environmental problems

Japan, which has few energy resources, must strive to be a low-carbon, energy-efficient society by enhancing energy-use technologies for its sustainable development. To that end, each step in the energy lifecycle (i.e., its generation, conversion/storage/transmission, and usage) must be re-examined in a form that is in harmony with the environment.

From the viewpoint of energy generation, the first goal is to find a more efficient use of renewable energy (natural energy), and there are great expectations with solar-power generation, wind-power generation, and the use of biomass. For instance, there are two requirements to increase the energy-conversion efficiency of photovoltaic cells and artificial photosynthetic devices, which are key to solar-power generation technology, as well as to increase the energy-conversion efficiency of thermoelectric transducers, which convert heat into energy: 1) an understanding of the correlation between the submicron structures of the composite materials, which make up the devices as a whole, and the energy-conversion efficiency, and 2) the elucidated and predicted degradation mechanisms for performance of the used materials. Computational-science techniques, such as large-scale electronic-structure calculations for organic and non-organic materials based on quantum mechanics, are essential for this. Moreover, a prior environmental assessment of local site conditions is needed to ensure a stable power supply based on renewable energy, and effective technologies for predicting power generation need to be put in place when the site is in actual operation.

Any combination of sun-power/wind-power generation and assessment/prediction requires meteorological models with higher accuracy and higher resolution than the currently available regional climate models and weather-prediction t models. Moreover, the clarification of plasma turbulence events that affect the confinement performance of fuel plasma is important for the scientific and technological underpinning of nuclear fusion reactors, another long-term alternative energy source candidate. However, because it is difficult to evaluate this in actual experiments, the role of computational science is indispensable.

2. Social and Scientific Problems in Computational Sciences

Energy and Environmental Problem | **Promoting advanced energy technologies in harmony with environment**

Current studies

- Consists of three independent fields: Materials science, manufacturing innovation, and meteorology & climatology
- Approaches based on theory and experiments

Approaches based on future computational science

- Understanding of correlation between structure and energy conversion efficiency of complex material, elucidation and prediction of deterioration mechanism of material performance
- Developing climate model for refining physical implementation and increasing resolution of grid space
- Elucidation of plasma turbulence
- Elucidation of electrochemical processes and search for scarce element substitutes for use as catalyst or electrode
- Manufacturing using numerical simulation
- Understanding current climate situation and projection using high-reliability climate system model

Contribution to society

- Creation of stable and highly efficient renewable energy
- Demonstration of fusion reactor scientifically and technologically
- Development of technology for secondary battery and fuel battery storage and efficient power supply
- Energy conservation in electronic devices and transportation equipment
- Monitoring geoenvironmental impact by energy utilization

Determination of suitable regions for wind power generation by high-resolution wind evaluation

The development of technology for efficiently storing power in and generating power from secondary batteries or fuel cells is essential from the perspective of energy conversion, storage, and transmission; material design using large-scale simulations is now becoming mainstream both in clarifying electrochemical processes and in the quest for alternatives for rare earth elements used in catalysts and electrodes.

Concerning energy usage, it is important to discover how to reduce the energy consumption of electronic devices (such as semiconductors that drive information—the software) or motor vehicles and aeroplanes, which drive objects—the hardware. To this end, it is necessary to transition from traditional theory and experiment-based development processes to innovative development processes based on computational science, such as numerical simulations, as well as new approaches utilizing those ideas. The application of computational science to development allows for the explanation of complex physical phenomena that have remained unexplained to date, thus making way for product development with an understanding of the physical mechanisms. It also enables the utilization of the optimum design technology where the various design parameters that are theoretically required are determined by trial and error.

In achieving a low-carbon and energy-efficient society, our energy usage will more or less affect the global environment. Measures to deal with issues such as global warming—already an international problem—require us to have a more accurate grasp of the current global environment while addressing the complex factors in, for instance, coupled atmosphere-ocean general circulation models, which involve biological and chemical processes; these measures also require us to link this information to future predictions.

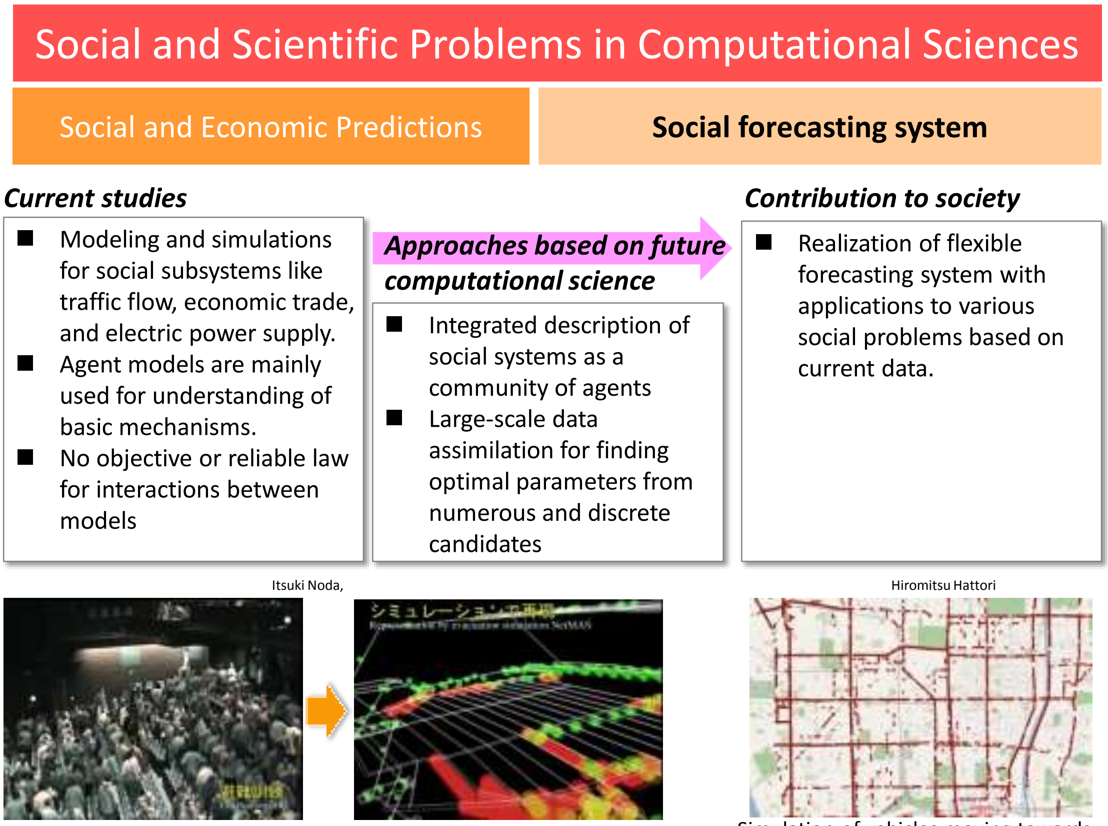
In the past, materials science, manufacturing, and weather and climate research fields have been conducting researches independently of each other. However, as long as the mathematical structures are the same, the simulation technology can be shared across different fields of study. From this perspective, in addition to developing individual basic theories and creating higher-resolution models, we also strive to improve simulation technology and data-assimilation technology while sharing common techniques across many fields and learning from mathematical models in other fields. Through this practice, it becomes possible to comprehend the cycle of energy generation, conversion, and usage comprehensively as an integrated grand field of science; it also becomes possible to contribute to energy and environmental problems on a global scale.

The table below lists the future required computational performance in the respective energy and environmental fields.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Electronic state calculation of electronic materials: method 1	100	20	5	15	240	10	860000	First-principles molecular dynamics simulation	Number of atoms: 100 million, 10^4 steps	
Electronic state calculation of electronic materials: method 2	100	10	1.2	12	96	10	350000	Real space basis O(N^3) first-principles molecular dynamics simulation	Number of atoms: 100 thousand, 100 steps	20 SCF cycles x 100 steps
Functional elucidation of strongly correlated electron systems	1900	2700	0.2		8	100	5500000	Variational quantum Monte Carlo method	Number of atoms: 10,000	The memory usage scales with the number of MPI processes. The maximum memory usage is shown.
Plasma turbulence calculation/ multi-scale turbulence	100	200	0.5	0.1	24	50	430000	Five-dimensional calculation by Boltzmann equation (spectral method + finite difference method)	10^12 grids, 10^6 steps	
Plasma turbulence calculation/ transient evolution of global turbulence	100	200	0.5	1	170	10	610000	Five-dimensional calculation by Boltzmann equation (finite difference method)	10^12 grids, 10^7 steps	
LES simulation (automobile, actual design, optimization problem)	110	230	0.04	4	1	100	41000	LES calculation of Re=10^6～10^7	10^10 grids	Calculated as B/F=2
LES simulation (automobile, high-end benchmarks)	120	230	0.5	48	24	10	100000	LES calculation of Re=10^6～10^7	Number of grid points ：10^12	Estimated by structured grid BF=2、output for 1,000 shots in 30 min.
Assessment of location condition for wind farm	29	89	0.01	0.07	72	100	760000	Model: High resolution LES, FVM	3,300x3,300x300grids (resolution:30x30x10m), 1,230,000steps (dt=0.21sec, 72hours integration including 24hours spin-up)	100 cases (200 days) are necessary for the assessment of each location.
The system for predicting global environment of the near future	56	110	0.6	80	600	1	120000	Model: MIROC-ESM (spectral model for atmosphere, FDM for ocean)	2000x1000x200grids, 53,000,000steps（dt=60sec, 100years integration)	Estimated by atmospheric model. 100 cases of calculations must be completed in one month. It is expected that All-to-all network communications are used in 1000 nodes (1TB/s per node). FLOP, memory throughput and memory usage are estimated based on a profile data obtained on K.

2.4 Social and economic predictions

Despite the fact that social and economic phenomena are the results of people's own behavior, they are often difficult to predict. This is due to the high numbers of individuals and the limits of the behavior and information held by each individual. Our social economy is an aggregate of various networks and systems: wide-area networks delivering transport, power, water and wastewater, gas, and communication services; social groups formed by the family, corporations, local government, and the state or government; industries that operate on the premise of a social system, such as agriculture, forestry, fisheries, manufacturing, and the production systems where these industries interconnect; and economic systems, such as commerce and finance. Predictions of socioeconomic phenomena were broken down into predictions for these individual systems, but these are by no means isolated phenomena; rather, they are intimately interconnected. Therefore, to solve various social problems, we need to regard our human socioeconomic activity as a community of agents and merge them with existing physical simulation techniques.



Multi-agent simulation of human evacuation

Simulation of vehicles moving towards specific destinations.

Furthermore, we aim to apply this not only to socioeconomic activity but also to connect this to predictions for geophysical events relating to either weather and climate or earthquakes and volcanic eruptions.

As it stands, no objective or reliable basic laws have been established for socioeconomic phenomena that meet the physical laws for natural phenomena. Therefore, for socioeconomic predictions, we need to closely study actual phenomena and continue to explore objective simulation models that match the diversity of the actual phenomena. We need to continue adjusting prediction models while closely monitoring events at all times. The continued pursuit of such fine data accumulation and analysis at a large scale has been made possible by the development of computers and networks allowing for socioeconomic data to be gathered, and this type of analysis is a common technique presently known as big data mining.

To understand why economic change occurs , theoretical research uses: agents modeling the main economic activities separately, agent models that possess an expressive power that can deal with the diversity of reality, and agent models to use on the phenomena uncovered by data mining. By continuing to flexibly assimilate data from actual events into model parameters, this kind of research is realizing simulation models that generate socioeconomic predictions.

Computational science will apply a variety of models to various social problems in a flexible manner and strives to realize systems that can make near-future predictions as well as distant-future predictions based on curent data that we have.

The table below lists the future required computational performance in the socioeconomic field.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Real time simulation of traffic flow	1000	* 100	* 0.00011	0.001	2.8E-08	1000	0.1	Agent simulations of global-scale traffic flow (1 billion vehicles, total road length of 34 million Km), (The number of operating cars are estimated to be 100 million)	10^8 vehicles x 10^3 operations x 10^3 steps x 10^3 cases (Simulation of 10 seconds) This should be simulated within 0.1 sec.	These amount of storage and total operations are required for each day. Total operations are estimated to be 10^3FLOP per one vehicles.
Optimization of the rules of stock trading	2100	0.0001	0.00000001		0.0024	10000	180000	Multi-agent Monte Carlo simulations of stock trading. (Duration: one day, Number of different stocks: 1000, Number of markets: 1)	Total number of operations: 5 hours x 3600sec/h x 1000 orders/sec x 10^4 operations/order x 10 traders x 10^4 cases x 10^3 individual stocks = 1.8 x 10^19 operations This should be simulated within 24hours.	Integer operations are dominant. "Total operation count"= total instruction count

*Figures marked with a * are still under examination. The website will show more accurate figures as they become available.*

3. Creating new science through interdisciplinary collaboration

Organically combining various fields that have traditionally been regarded as separate research areas promises to create new fields of science. In this section, we will comment on the "Unified understanding through integration of basic sciences" that pushes the frontiers and opens up new academic areas, such as "Effective use of big data," which is expected to grow in importance in the near future; we will also comment on the "Collaboration with large experimental facilities," where the link with computatinal science will be essential for large-scale data processing.

Area	Collaboration examples
Unified understanding through integration of basic sciences	Collaboration in fundamental physics Planetary science through collaboration in space and earth sciences Interdisciplinary collaboration between life sciences, materials science, and manufacturing
Effective use of big data	Creation and refinement of computational science infrastructure Example of effective use 1: Effective use of satellite and observational data Example of effective use 2: Genome analysis
Collaboration with large experimental facilities	Collaboration with large experimental facilities such as X-ray-free electron laser facility SACLA

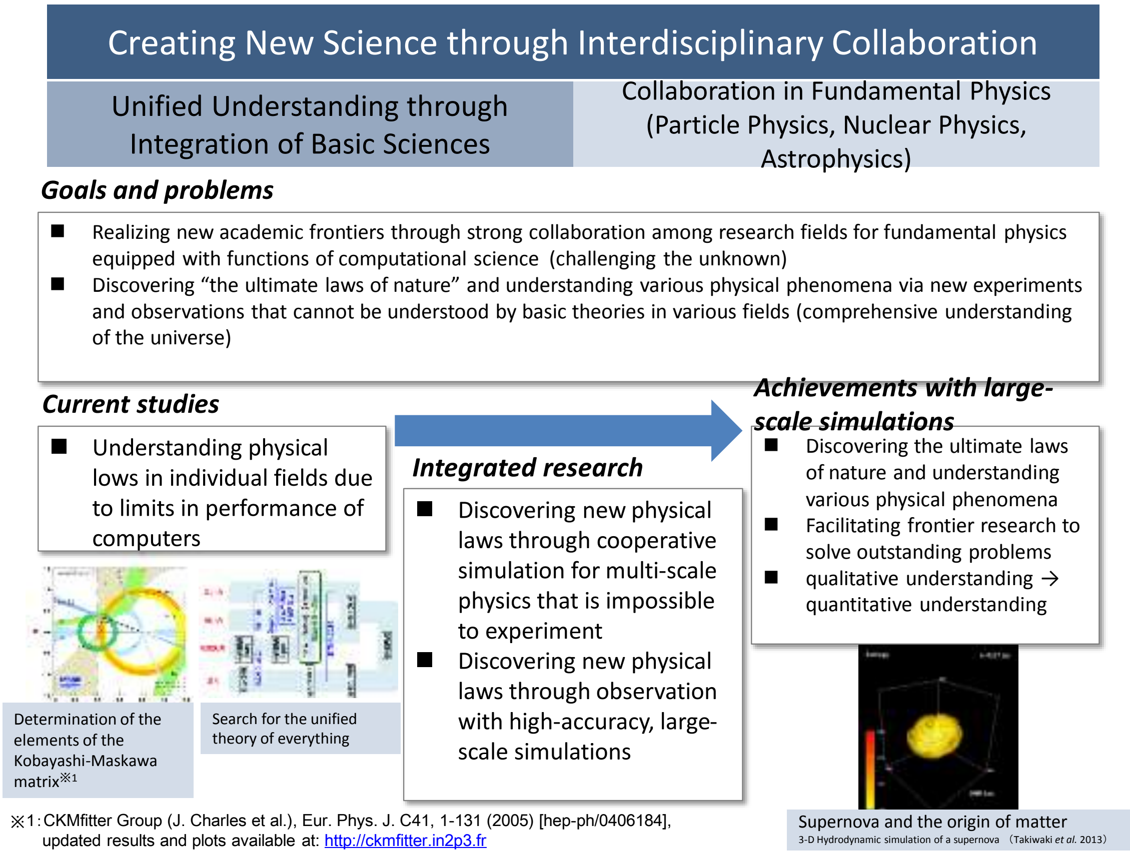
3.1 Unified understanding through integration of basic sciences

The relentless pursuit of knowledge in fundamental science has always been a strong motivator in the development of science and technology. There are many examples where fundamental scientific research, which does not seem to be connected to any practical applications, has led to next-generation applications and has enriched our lives. Through the use of computational science techniques, a stronger collaboration between areas in fundamental science, which thus far have developed in isolation from each other, is expected to push the frontiers and open up new academic areas. Here, we will outline three representative examples of collaboration relating to the universe, materials, and life.

(1) Collaboration in fundamental physics

Areas in fundamental physics, such as particle physics, nuclear physics, or astrophysics, strive to discover the ultimate laws of physics and form an understanding of a variety of physical phenomena.

Conventionally, partly due to the limitations of computer performance, basic laws were established in each separate field isolatedly. However, with the recent advances in science and technology, new experimental and observational evidences, which cannot be understood in terms of the basic laws of individual fields alone, are being revealed. Therefore, by making stronger links between the various developed and growing fields through computational science techniques, we aim to open up new academic areas (the challenge of the unknown), and to understand the new experimental and observational evidences (a unified understanding of nature). To this end, we need to find new theories to overcome a variety of problems, and these need to be computed using large-scale numerical simulations by supercomputers. In recent years, large-scale computations have been enabling the shift to a quantitative understanding of evidence of which we previously only had a qualitative understanding.



By using computational science techniques as a bridge linking the basic laws and interdisciplinary collaboration, we will drive research at the intellectual frontier to solve vital problems leading to both the clarification of the ultimate laws of physics and an understanding of a variety of physical phenomena.

The table below lists the required computational performance for collaboration in fundamental physics.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Determination and application of effective Bryon interactions based on chiral symmetry and QCD	510	390	0.066	0.5	880	10	16000000	Lattice QCD(Chiral fermions in 5-dimensional representation), Hybrid Monte Carlo algorithm, CG solver	Problem size: Lattice size 128^4x32, lattice spacing < 0.1 [fm]	16^4 nodes are assumed. on-chip memory size 200MB, on-chip memory bandwidth 6TB/s, network latency 1usec, network bandwidth 128GB/s par node.
No-core shell-model calculations for atomic nuclei	100	10	0.1	* 0.0001	28	100	1000000	nuclear structure calculation of light nuclei using Monte Carlo shell model	Space is expanded by Harmonic oscillator basis. 7 or 8 major shells are taken as the model spac of shell-model calculations.	The memory capacity is estimated as 10000 node times 10GB.
Exploring supernova explosion by relativistic radiative hydrodynamics	18	70	1.6	1.3	1200	10	780000	Neutorino radiation transport (supernova explosion)	1 sec simulation with the size of space 512x64x128 and momentum space 24^3 grids	100Tflops/node ×10000node, communication b/w 60GB/s/node

*Figure marked with a * is still under examination. The website will show more accurate figures as they become available.*

(2) Planetary science through collaboration in astronomy and earth science

"Where is here, and what are we?" The ultimate goal of astronomy and planetary science is to pursue this age-old universal question. The exploration of the solar system and observation of extrasolar planets in recent years has allowed us to consider this question as a concrete and scientific problem. Moreover, a comprehensive understanding of other planetary systems from the perspectives of cosmology and earth sciences will lead to an understanding of the planet we inhabit. A great number of planetary exploration projects will be planned and implemented both nationally and internationally, and simulation technology (i.e., computational science) and data-assimilation technology, which is applied to modeling the impact on the real world, are both essential to creating and implementing these plans and are expected to make substantial contributions.

Collaboration between astronomy, earth science, and computational science enables simulations over a wide range, including extended periods of time and multiple phenomena, thus allowing for an integral understanding. Furthermore, numerical simulations also enable us to verify the validity of our understanding, theories, and models, bringing us a big step closer to understanding the planet we inhabit and knowing the origin of life.

The table below lists the future required computational performance for planetary science through collaboration in astronomy and earth sciences.

Creating New Science through Interdisciplinary Collaboration

Unified Understanding through Integration of Basic Sciences

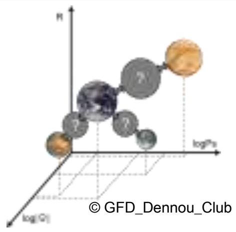
Planetary Science through Collaboration in Astronomy and Earth Sciences

Goals and problems

- To answer old and common questions of mankind: "Where is here? What are we? ". (To know origin of life)
- To understand planetary systems comprehensively from both cosmological and geophysical aspects. (To understand our planet)

Current studies

- Understanding of individual phenomena (simulation of phenomena with limited spatiotemporal scales)
- Difficulty in validation due to limitation of observations



Variation of planetary surface environments

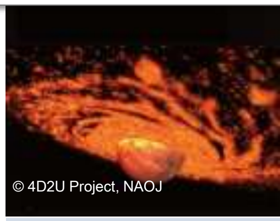
Strategy of the discipline

- Numerical simulations to understand the state of planets, their origin, and evolution
- Numerical simulations to understand the data obtained by observations and experiments

Planets in our solar system

Achievements with large-scale simulations

- Comprehensive understanding by simulations with large domain, long term, multiple phenomena
- Validation of our understanding, theories and models by numerical simulations



Simulation of Earth-Moon formation by the giant impact

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Planetary system formation	4.2	0.021	0.00001	0.05	1000	100	1500000	N-body simulation	100 million particles, 100 million years (10G steps),	This estimation is based on an algorithm reported by a paper and profile data obtained on GRAPE. 15,000FLOP/1step/1particle. 128 particles in a group. Amount of memory access is 32 byte per 6,000 FLOP.
Simulation of Earth and planet formation	520	29	0.001	1	24	100	4500000	SPH method	1 billion particles, few months (100M steps), NlogN FLOPS	The magnitude of FLOP, memory usage and memory throughput are estimated based on a profile data obtained on TSUBAME.
Simulation of formation and evolution of planetary surface environment	5.6	25	0.01	4	100	1000	2000000	Fluid dynamics + radiation simulation (spectrum method + difference method)	3840x1920x192 grids, 100 cases x 10 planets, 10 years (30M steps), number of operations per step: 50K	The magnitude of FLOP and memory usage are estimated based on a profile data obtained on TSUBAME.

(3) Interdisciplinary collaboration in the fields of life sciences, materials science, and manufacturing

In life sciences, biomolecules, such as proteins and DNA, are seen as the building blocks that form life, but in materials science, they are extremely complex materials with little symmetry, possessing aspects of both "life" and "material." Consequently, research of biomolecules and especially analyses based on its steric structure takes place at the interface between the two great fields of life and materials sciences, and we can expect important breakthroughs by combining and using methods that have been produced in the respective disciplines across these areas.

Creating New Science through Interdisciplinary Collaboration

Unified Understanding through Integration of Basic Sciences	Interdisciplinary Collaboration in the Fields of Life Sciences, Materials Science, and Manufacturing

Example: Analysis based on three-dimensional structure of biological molecules, complexes

Goals and problems

- To describe life sciences in terms of basic physics/chemistry
- To rebuild/understand life sciences at the atomic/molecular resolution
- Manifestation of not only application to drug discovery or medical technology but also fundamental understanding of life science from basic physics/chemistry

Current studies

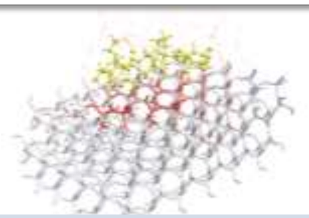
- Commercialization of new technology takes more than 10 years after initiation of development
- Recently, the development period has become longer than before

Integrated research

- Simulation for understanding issues of boundaries between nano-technology and bio-technology
- Molecular dynamics of large-scale systems

Contribution to society

- Dynamical analysis of biomolecules like proteins and DNAs as well as of viruses or large systems in cell environments
- Elucidation of proteins at the level of atomic structure/function and electronic state



Interaction of biomolecules on a material surface

Life science and materials science complement each other, and merging the two enables dramatic progress with important social problems, such as drug discovery or biomolecule-based manufacturing. For instance, the mechanisms of the biological activity of proteins targeted by drugs are the object of study in life sciences. However, drug and protein binding are physicochemical interactions, which is where materials science methodology is effective. With the increase of supercomputer speeds, we can expect breakthroughs in drug discovery through the application of highly accurate computational methods developed in the field of materials science that lead to better protein-drug interaction analyses. We furthermore anticipate the implementation of kinetic analyses for not only isolated biomolecules but also ultra-large systems, such as viral or cellular environments; we also anticipate the explication of the electron state of proteins or functional structures at the electron level.

The table below lists the future required computational performance for interdisciplinary collaboration in the fields of life science, materials science, and manufacturing.

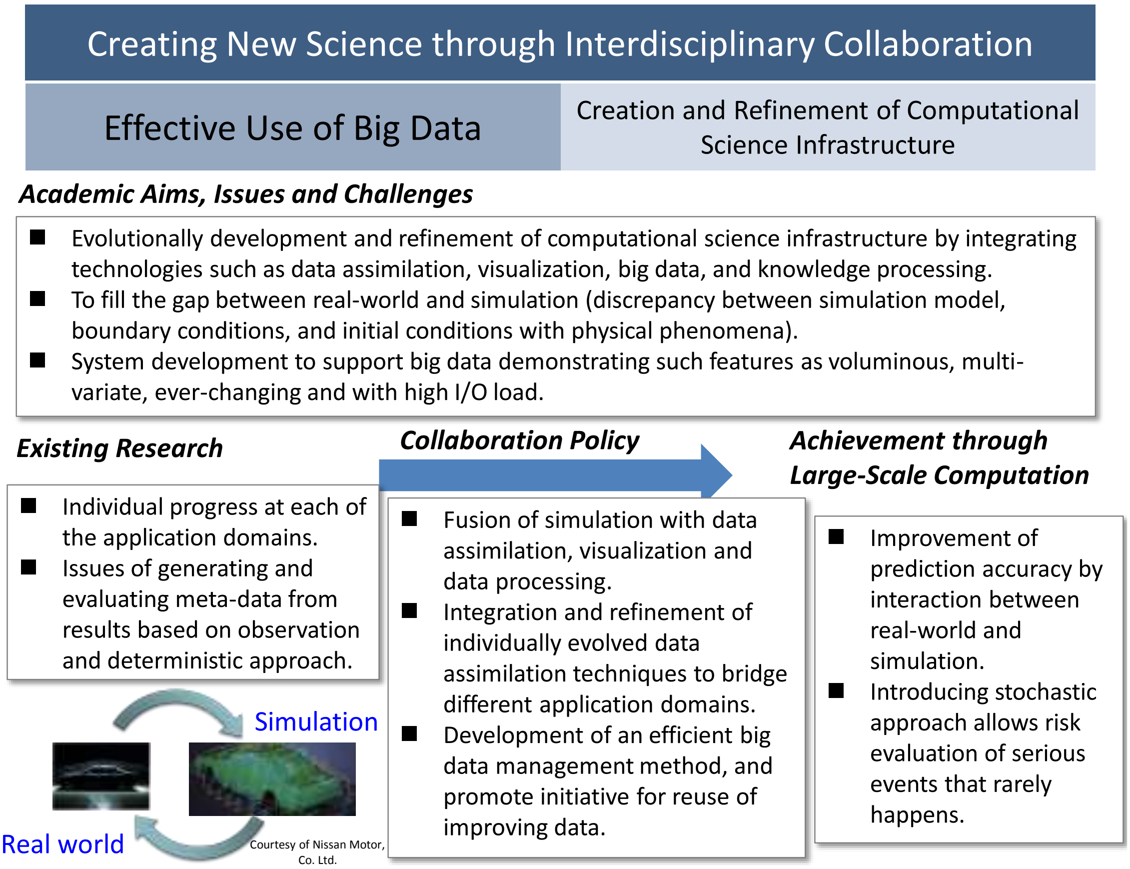
Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
MD and Free-energy calculation for drug design and so on	1000	400	0.0001		0.0012	1000000	4300000	Molecular dynamics simulation with all-atom model	Number of Cases: 100,000 ligands X 10 target proteins	>16GB/node, B/F=0.1
Precise Structure-Based Drug Design	0.83	0.14	1	0.001	1	100	300	Ab initio quantum chemistry calculations on the interactions between protein and ligands	proteins (500 residues) + ligands in solution	1TB/s IO speed required to dump 1TB dataset per second
Design of Biological Devices	1.1	0.19	1	0.001	1	100	400	Spectroscopic analyses of proteins (200-500 residues)	more than 100,000 orbitals	1TB/s IO speed required to dump 1TB dataset per second
MD simulations under cellular environments or MD simulations of Virus	490	49	0.2	1.2	48	10	850000	Molecular dynamics simulations with all-atom / coarse-grained model	100,000,000 particles	Static information of all particles is distributed to all the computer nodes. Therefore, at least 8GB x number of nodes are necessary for the whole system.

3.2 Effective use of big data

Next-generation computational science core technologies are integral to the research needed for solving social problems and the search for knowledge that surpasses the boundaries of conventional computational technologies. Especially in recent years, there has been a dramatic rise in experimental data that can be obtained from large-scale numerical simulations, satellite and observational data, individual genome information, and large experimental facilities; therefore, computational science core technologies are required, including the technology for the efficient use of these so-called big data. In this section, we will address what these computational science core technologies are, discuss their increasing sophistication, and then outline two typical collaboration examples.

(1) Creation and refinement of computational science infrastructure

Events occurring in the real world are complex in themselves, and it is difficult to accurately ascertain any causal relationships. In most cases, they cannot be modeled as they are in the real world, and simulations compute idealized versions of real-world events. Consequently, there is more than a little gap between events occurring in the real world and their simulated versions.



Modeling is the process of selecting an equation-group format that can successfully reproduce the characteristics of a real-world problem and determining the related parameters. Modeling is based on knowledge that is obtained from insights into observational and experimental results and is an extremely important process that affects the accuracy and precision of a simulation.

In the past, modeling and its various related technologies evolved separately for each field in each application, although there were also many shared aspects. For instance, data assimilation, a technology that applies real-world effects to modeling, is studied and used in fields like meteorology, oil drilling, management, manufacturing, drug discovery, and molecular simulation. Furthermore, visualization technologies that help interpret results are also used in various research areas, and research has been conducted to display the methods that are suited for a particular phenomenon.

The table below lists the required future computational performance for the creation and enhancement of computational science core technologies.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Parallel rendering	200	61	0.8	10	0.5	1	360	Volume rendering(ray cast, file based)		The problem size depends on the object to be rendered. The values in this row are estimated from typical examples.
Parallel rendering	200	61	2	1	0.5	1	360	In-situ volume rendering		The problem size depends on the object to be rendered. The values in this row are estimated from typical examples.
Data comporession	500	25	8	10	0.5	1	900	Data compresion by Proper Orthogonal Decomposition		The problem size depends on the object to be compressed. The values in this row are estimated from typical examples.

(2) Example of effective use of big data 1: Effective use of satellite and observational data

Observational data relating to the physical, chemical, and biological environments of the atmosphere, oceans, and land play an important role as basic data for monitoring and extracting environmental change and forecasting impact. Observational data contribute to the appropriate handling of wide-ranging environmental problems from daily weather forecasting to global climate change. These data are obtained in a variety of ways, such as direct in-situ observations and remote sensing by satellites; also, their application areas extend far and wide. However, because these observational data are harvested at different times and locations and because they vary both in quantity and accuracy, it is difficult to use them by simply compiling them with existing observational data. They are therefore processed into easy-to-use data sets through data assimilation, a method that integrates varying observational data using numerical models.

Large-scale data assimilation has been prominent in technology development in the fields of meteorology and climate. Recently, however, its application is also being attempted in other fields, including design management, oil drilling, and molecular simulations; also, there is currently a movement with which data assimilation techniques that had previously been limited to the communities in each application field are now being shared across those communities. Due to advances in observation technologies, data assimilation used for handling big data continue to innovate in order to deal with observational data of increasing resolution, and large amounts of satellite observation data are made possible by diversifying observation variables. They also continue to innovate in order to deal with the extraordinary increase in volume of simulations using both of these types of data.

With the evolution in data assimilation technology, increasingly high-resolution and high-precision simulations are allowing us to deal head-on with predictions of localized weather events that have a high social impact, such as sudden heavy rain or tropical-weather forecasts.

Creating New Science through Interdisciplinary Collaboration

Example of Effective Use of Big Data - 1 - | Effective Use of Satellite and Observational Data

Goals and problems

- Accommodate to the various types of satellite observation data and new high-resolution observation data
- Deriving spatiotemporally homogeneous Earth observation data from a data assimilation (DA) system

Current studies

- Effective use of various observation data
- Qualitative limits of the DA products owing to the model resolution
- Diverse variables and high resolution in advanced observational data

Integrated research

- DA with high-resolution/complex numerical model (weather/climate model)
- DA for better handling of large and high-frequency data
- Advanced DA techniques: non-linear processes, non-Gaussian probability distribution

Achievement through large-scale computation

- Integration of various satellite data by DA
- Advanced prediction of heavy rain with next-generation observation data
- Entering a new era of "Tropical weather forecast"

Prediction of typhoon formation

The table below lists the required future computational performance for the effective use of satellite and observational data.

Subject	Perfor–mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Heavy rain prediction and tropical weather forecast	220	270	0.7	5	580	2	900000	NICAM + Data assimilation: Local Ensemble Transform Kalman Filter(LETKF)	3.5km horizontal resolution, 100 vertical layers, 1,000 ensemble members, 3 hourly assimilation cycle, 2 months integration	100,000 nodes are assumed (side–by–side comm; 1GB/s for atmospheric model) The magnitude of FLOP, memory throughput and memory usage are estimated based on a profile data obtained on K.
Integrated reanalysis of global environment	3.1	13	0.018	0.022	18	240	48000	Data assimilation: 4D−var	640x320x150(atmosphere), 3600x1800x150(ocean) grids. 1min (atmosphere), 30sec (ocean), and 10min (coupling) temporal resolution. 100 iterations, 3 month integration	B/F: 4.66 (atmosphere), 4.24 (ocean) FLOP and memory usage are estimated based on a profile data on ES2. That of memory throughput is estimated from FLOP and application B/F in source code (does not take account of cache).

(3) Example of effective use of big data 2: Genome analysis

The Human Genome Project is one of the milestones in molecular biology; it took many years and the collaboration of many molecular biologists world wide to decode the DNA of just one person. In recent years, however, next-generation DNA sequencers, devices that decode DNA at ultra-high speed and low cost, have been developed, enabling us to perform genome analyses for a variety of cell types for individual genome information.

The volume of data obtained in experiments is increasing dramatically in molecular biology due to the emergence of new technologies such as DNA microarrays, which can simultaneously measure the expression levels of large numbers of genes within cells. Bioinformatics is a research method established using computational science for the purpose of analyzing biological discoveries from these high volumes of data or high-throughput data, and this collaboration with computational science has been recognized as an important and necessary area.

Life science data analysis combines large volumes of widely different data, including genome sequence data, gene expression data, DNA modification data (epigenomic data), or protein interactions. Through future developed observational techniques, more data are expected to be accumulated, and with these advances in observational technologies, a variety of analysis software will be combined and used.

The research advances in this field have allowed for developed methods of searching and effectively analyzing genomic data and will allow us to bring the current extremely high costs of optimal medical therapies based on individual genomic information (personal genomic medicine) down to the general medical treatment level.

Creating New Science through Interdisciplinary Collaboration

Example of Effective Use of Big Data - 2 - | Genome Analysis

Goals and problems

- In collaboration with computer science, development of approaches to reveal biological phenomena using large data obtained in recently developed high-throughput measurements
- Development of gene network analysis for inference and prediction of interactions between genes from the gene expression data

Current studies

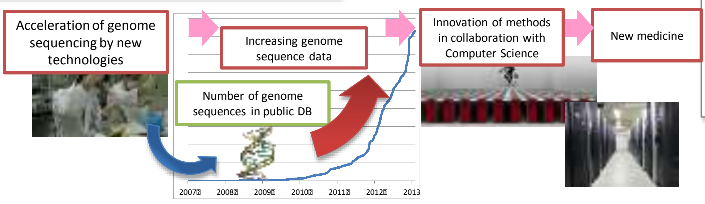
- Bioinformatics for massively large data sets obtained using recent high-throughput measurements
- Inference for gene network from relatively small number of data sets measured in a specific condition

Integrated research

- Analyses using combinations of big data from different fields

Contribution to society

- Development of algorithms for efficient search and analysis of genome information
- Cost reduction of medicine personalized for individual genome information
- Exhaustive applications of inference calculation from hundreds of data sets (tens of thousands of samples) obtained from the whole genome expression data



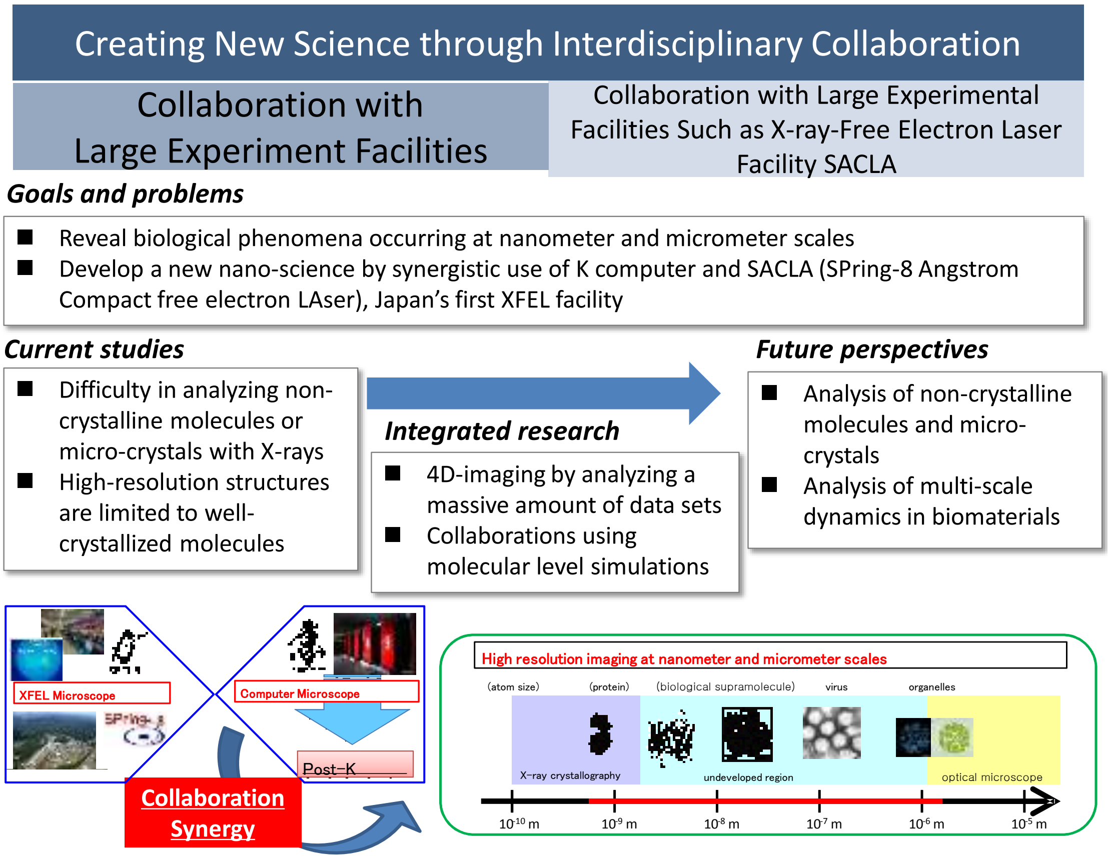
The table below lists the required future computational performance for genome analysis.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Personal Genome Analysis	0.0054	0.0016	1.6	0.1	0.7	200000	2700	Sequence matching	Cancer Genome Analysis: Short read mapping and mutation identification of 200,000 people's genome	To detect mutation, it is necessary to analyze at least 1000 people's genome. Amount of necessary memory is 140K node X 64GB.
Statistical Analysis for human disease-associated genes.	9.9	0.0002	200	2	140	5	25000	Genome-wide association study	Human genome 3Gbp x 200,000 people in total. 40,000 people for each study.	Memory size: 800GB/node. Number of nodes: 250,000

3.3 Collaboration with large experimental facilities

(1) Collaboration with large experimental facilities such as X-ray-free electron laser facility SACLA

XFEL (X-ray-free electron laser) is a light source that can generate ultra-bright X-rays combining the features of laser beams. It is expected to be powerful in analyzing the structure of amorphous particles with sub-micrometer size and small crystals ranging from a few micrometers to a few hundred nanometers, which have so far been difficult to solve. SACLA (SPring-8 Angstrom Compact free electron Laser), the XFEL facility completed in March 2011, requires large volumes of data processing (up to approximately five million diffraction patterns obtained per day) and has stated the importance of collaborating with computational science technologies. Expectations are high in particular for 4D imaging that also incorporates dynamic states of bioparticles.



Amorphous sample imaging experiments by SACLA can capture bioparticles sized 10 nanometers to a micrometer at a nanometer resolution. These spatial scales form a good target for molecular simulations, such as coarse-grained MD methods[5].

[5] A computational technique using a model that represents a few to a few tens of atoms as one particle. The number of atoms that form a coarse-grained unit is determined by to the objective.

The analysis of SACLA experimental data using large-scale molecular simulations enables progress in the research of hierarchical dynamics in bioparticles and leads to an understanding of life phenomena that occur on these spatial scales.

Due to the collaboration between large experimental facilities and computational science, new areas of nanoscience have been cultivated, leading to the elucidation of life phenomena in high resolutions at a scale extending from nanometers to micrometers.

The table below lists the required future computational performance for the collaboration with large experimental facilities.

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
4-dimensional imaging from "big" experimental data	2	0.01	0.000001	0.000001	2.8E-11	1E+12	200	Structural classification, 3-dimensional structure modeling, Statistical analysis on time frames	1 – 10 million images	
Dynamic Structural Modeling based on Experimental data	490	49	0.2	1.2	48	10	850000	Molecular dynamics simulations with all-atom / coarse-grained model	100,000,000 particles	Static information of all particles is distributed to all the computer nodes. Therefore, at least 8GB x number of nodes are necessary for the whole system.

4. Social and Scientific Problems in Computational Sciences

This overview introduces concrete examples of the social problems that computational science may contribute to in the future. It also introduces new scientific problems that could arise from cooperation between various branches of science that have conventionally been regarded as distinct areas. To solve the social and scientific problems introduced in this overview, it is essential to tackle research problems in the various computational science fields that will form the basis for such collaboration. Details for this are given in Chapter 4 of the Computational Science Roadmap, but this overview outlines, as a reference, the new research that has become possible thanks to advances in current research and computing, and the requirements regarding the computational performance that will be needed in the future.

Please note that the requirements for computational performance shown here are valid as of May, 2014. There is still room for more detailed examination, and so we plan to provide more accurate figures in the most up-to-date version of the Computational Science Roadmap.

4.1 Life Science

Social and Scientific Problems in Computational Sciences

Life Science — **Predictable Life Sciences and Applications**

Current studies

- Short timescale dynamics for isolated biomolecules
- Drug design for simple soluble proteins
- Computational screening of a small number of compounds
- Intracellular signaling reaction network
- Simulation for thrombosis, blood flow, and high-intensity focused ultrasound
- Visual cortex simulation in point neuron model
- Moths' olfactory motor system simulation in multi-compartment model
- Gene network analysis on small gene expression data
- Analysis of next-generation sequencing for 2,000 people

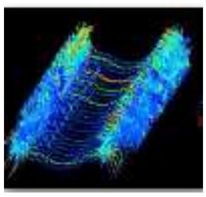
Breakthroughs with computational science

- High-speed replica MD simulations
- Ultra large-scale MD simulations
- High-performance cell simulation software
- Coupling of the multi scale model for time-space
- Larger scale and more detailed biophysical features in brain model
- Data assimilation in neural circuit simulation
- Gene network analysis on large-scale expression data
- Genome analysis by next-generation sequencer for large-scale data

Large-scale computing studies

- Large-scale dynamics of working biomolecules
- Large-scale dynamics of cells , viruses, etc.
- Drug design for complex proteins, e.g., membrane-proteins
- Computational screening of a larger number of compounds
- Elucidating cell fate decision processes at the cell population level
- Blood flow simulation at organ level, coupled with metabolism simulation at cell level
- High-intensity focused ultrasound for drug delivery system
- Whole human brain simulation
- Real-time whole insect brain simulation
- Exhaustive gene network analysis for 100 times the size of present data sets
- Analysis of 4th next generation sequencer for 200,000 people

Relenza (red) bound to influenza virus protein (blue)



Insect's premotor center simulation (NEURON)

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Analysis of Biomolecular Function	29	12	0.0084	1.2	240	10	250000	Molecular dynamics (all-atom, QM/MM, coarse-grain)	1 million atoms, 100 replicas	Low netwark latency less than a sub-micro second is required. Memory size per case is estimated for a 100,000 node run.
MD simulations under cellular environments or MD simulations of Virus	490	49	0.2	1.2	48	10	850000	Molecular dynamics simulations with all-atom / coarse-grained model	100,000,000 particles	B/F=0.1
MD and Free-energy calculation for drug design and so on	1000	400	0.0001		0.0012	1000000	4300000	Molecular dynamics simulation with all-atom model	Number of Cases: 100,000 ligands X 10 target proteins	B/F=0.4. Supposed to run 100-1000 cases simultaneously. Memory size per case is estimated for a 100 node run.
Simulations of cellular signaling pathways	42	100	10	10	240	100	3600000	Microscopic lattice reaction-diffusion simulation	1,000 to 10,000 cells	integer operations
Simulations of cellular signaling pathways	420	0.01	0.001	0.001	240	100	36000000	Green's function reaction dynamics	millions of molecules	low network latency
Multi-scale simulation of a blood clot	400	64	1	1	170	10	2500000	Semi-implicit FDM simulation of fluid-structure interaction with chemical factors	Length:100mm, D:100um, Calculation Time:10s, Grid size:0.1um, Velocity:10^-2m/s, Delta T:1us	
High Intensity Focused Ultrasound	380	460	54	64	240	10	3300000	Explicit FDM simulation of sound wave and heat transfer	Calculation Area:400mm^3, Grid: 225x10^12, Steps: 1459200, FLOP/grid/step: 1000	
Whole human brain simulation by simple firing model	6.9	7.6	56	3600	0.28	100	700	Single compartment neuron mode	100 billion neurons, 1000 synapses/neuron and 100000 steps	
Whole human brain simulation by detailed compartmental model	71	78	250	25000	39	1	10000	Multi-compartment HH model with local Crank-Nicolson method	100 billion neurons, 1000 synapses/neuron and 100000 steps	
Realtime whole insect brain simuation by a detailed compartmental model	71	60	0.002	0.2	0.028	100	720	Multi-compartment HH model with local Crank-Nicolson method	One million neuron, 500 synapse/neuron	
Parameter estimator in insect brain simulation	71	60	0.2	20	28	10	72000	Multi-compartment HH model with local Crank-Nicolson method, evolutionary algorithm	1000 neurons, 10^6 genes, 100 generations	
Data assimilation of whole insect brain via communication between a phisological experiment and a simulation	71	60	0.2	20	28	10	72000	Multi-compartment HH model with local Crank-Nicolson method, evolutionary algorithm	1000 neurons, 10^6 genes, 100 generations	Supposing 100 MB/s communication to external enviroment will be required
Gene Network Analysis	2900	1500	0.08	0.016	0.34	26000	94000000	Baysian network estimation and L1-regularization	40,000 transcripts x 26,000 data sets consisting of 2,800,000 arrays	

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

4.2 Materials Science

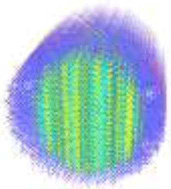
Social and Scientific Problems in Computational Sciences

Materials Science

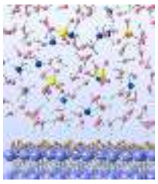
Creation of Next Generation Matter in Collaboration with Theory and Experiment

Current studies

- Electronic state calculation and structure optimization based on DFT
- Molecular simulation by classical MD
- First-principle MD simulation of interfaces
- Different simulation methods for individual scales
- Strongly correlated systems by exact diagonalization and QMC
- Quantum chemistry calculation for nano-scale molecules



Electric state calculation of carbon nano-wire (A. Oshiyama (Univ. Tokyo))



First-principle calculation of electrode-electrolyte Interface (O. Sugino (Univ. Tokyo))

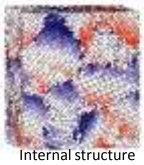
Problems to solve in computational science

- Calculation of thermal and dynamical properties and excited states through large-scale and high-precision first-principle algorithm
- Cultivation of algorithms for systems with electron-photon coupling
- Structure optimization by QM method, elucidation of coupling process through extended ensemble MD
- All-atom simulation of giant viruses
- High-precision quantum chemistry calculation method combined with relativistic theory and multi-reference theory
- Linear scaling algorithms for delocalized electron systems
- Multi-scale method combining quantum chemistry calculation, molecular simulation, coarse-graining modeling, and continuum models
- Exhaustive simulations for technological application
- First-principle methods for strongly-correlated systems

Long-term objective

- Development of next-generation high-speed, large-capacity, and low-power-consumption quantum devices
- Computational design of opto-electronic functional devices
- Theoretical proposal of novel chemical conversion
- Development of next-generation high-efficiency energy conversion devices
- Elucidation of stability, fine structure, and toughness of internal structure of materials
- Search for novel strongly correlated materials, such as high-Tc superconductors and high-efficiency thermoelectric devices
- Quantum chemistry calculation with higher-order electron correlation of macromolecule systems

All-atom simulation of Virus (S. Okazaki (Nagoya Univ.)



Internal structure of iron (M. Kohyama (AIST))

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Next generation advanced devices	100	100	1.2	10	96	10	350000	First-principles calculation RSDFT (pseudopotential method, real space basis set)	Number of atoms: 100,000	
Next generation advanced devices	100	100	2	15	60	100	2200000	First-principles calculation PHASE (pseudopotential method, plane-wave basis set, O(N^3) method)	Number of atoms: 10,000, simultaneous execution of 100 MD tasks	
Next generation advanced devices	100	100	2	15	60	100	2200000	First-principles calculation xTAPP (pseudopotential method, plane-wave basis set, O(N^3) method)	Number of atoms: 10,000, simultaneous execution of 100 MD tasks	
Next generation advanced devices	100	20	5	10	240	10	860000	First-principles calculation CONQUEST (O(N) method based on density matrix optimization)	Number of atoms: 100 million. Considering nanosecond order time scale simulation with 2 fs time step. Be aware of the computation time. Calculation methods used is the same as in "Electronic state calculation of electronic materials: method 1", but this subject requires more calculation speed for each case, thus more efficient network performance is required. The difference in data storage arises from frequency of writing out the output data.	
Optical and electronic devices	1000	10	10	0.1	1	100	360000	High-precision molecular orbital method	20,000 basis sets (atomic orbitals), 1 million numerical quadrature points	100-1000 array jobs in consideration
Molecular function	300	18	4	0.0001	15	10	160000	Large-scale molecular orbital method	Number of atoms: 10,000	
Molecular function	1.1	0.19	1	0.001	1	100	400	Fragment molecular orbital method	Proteins with few hundreds of amino acid residues, eigenvalue problem for dense matrix with dimensions more than 10 million	

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Improvement of safety and characteristic analysis of heat transfer devices	20	6.4	51	44	24	10	17000	Short-range classical molecular dynamics method	Number of particles: 400 billion	
Molecular function and chemical conversion	1000	100	2	1000	150	10	5400000	Long-range classical molecular dynamics method	Number of atoms: 1 billion	
Optical and electronic materials	600	200	200	33	14	10	300000	Electron and electromagnetic field coupled dynamics method for nanostructures	Number of atoms: 960 thousand, computation time per one step is 1 second and operation count is 0.63EFLOP. Computation time for 50,000 steps is about 14 hours.	
Functional elucidation of strongly correlated electron systems	3	390	10	10	10	100	11000	Cluster algorithm quantum Monte Carlo method	Number of atoms: 100 million	Mainly integer operations
Functional elucidation of strongly correlated electron systems	1000	300	0.2		8	100	2900000	Variational quantum Monte Carlo method	Number of atoms: 10,000	The memory usage scales with the number of MPI processes. The maximum memory usage is shown.
Chemical and energy conversion	500	50	0.008	6.4	2.8	10	50000	Quantum molecular dynamics method	100 replicas, 1 million steps	The performance requirement are estimated based on benchmark data from electronic structure calculation of xTAPP, classical MD calculation of MODYLAS, and I/O from short-range classical MD
Chemical and energy conversion	690	69	2	3.2	300	10	7400000	Chemical reaction dynamics and quantum molecular dynamics method (molecular orbital or QM/MM calculation)	10,000 replicas of 1,000 QM atoms and 100,000 MM atoms, 10,000 steps (roadmap)	The performance requirement are estimated based on benchmark data from electronic structure calculation of xTAPP, classical MD calculation of MODYLAS, and I/O from short-range classical MD
Chemical and energy conversion	410	41	0.02	0.05	20	10	300000	Chemical reaction dynamics and quantum molecular dynamics method (first-principles calculation)	Tens of thousands of replicas	The performance requirement are estimated based on benchmark data from electronic structure calculation of xTAPP, classical MD calculation of MODYLAS, and I/O from short-range classical MD
Molecular structure and function	1000	0.5	0.04		24	1	86000	Molecular dynamics method (analysis of frequency dependence of permittivity of relaxor ferroelectrics)	512x512x512	Array jobs with no inter-node network communications
Exploration of new materials	4100	41	20		0.5	1	7400	Cluster expansion method (first-principles calculation)	Number of atoms: 10,000, simultaneous execution of tasks on 100 ion configurations	
Exploration of new materials	0.1	0.02	0.00012		24	10000	86000	First-principles calculation (frozen phonon method)	Number of atoms: 10,000	Simultaneous execution is not considered in this table because the sizes of problem are ten times smaller than that in the case of PHASE applications
Functional elucidation of strongly correlated electron systems	82	130	82	41	42	10	120000	Exact diagonalization method (Lanczos method)	Spin system with 54 sites(S z = 0)	
Exploration of new materials	690	1600	1.5	20	24	20	1200000	Phase-field method	10^13 of space meshs, 10^7 time steps	Calculation of 10000 node parallelization with 100 TFLOPS per node in consideration

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

4.3 Geoscience

(1) Meteorology and Climatology

Social and Scientific Problems in Computational Sciences

Geoscience/Meteorology and Climatology | **Weather Prediction, Global Change Projection, Scale Interactions**

Current studies

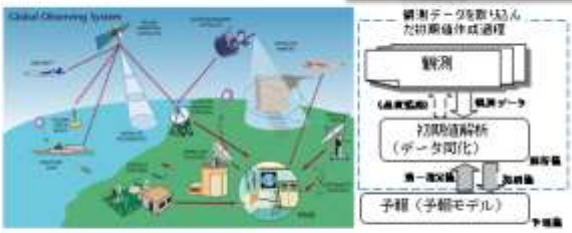
- Difficulty to perform ensemble experiments for resolving cumulonimbus, which causes localized torrential rain
- Great uncertainties in processes in earth system models

Breakthroughs with computational science

- Prediction of localized torrential rain based on ensemble and very-high-resolution experiments
- Development of a monitoring and projection system for global environment
- Development of coupler software and computational library for coupling component models

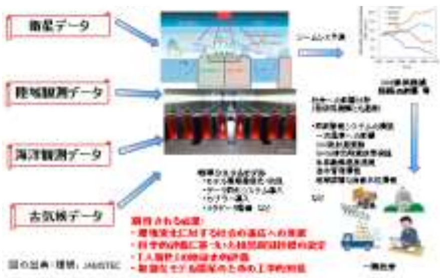
Contribution to society

- Reduction of damage from meteorological disaster by predicting extreme events such as typhoons and torrential downpour
- Contribution to decision making on global changes and enhancement of societal awareness using a monitoring and projection system for global environment



図の出典：世界気象機関 (http://www.wmo.int/pages/prog/www/OSY/GOS.html) および 気象予報士ハンドブック(日本気象予報士会, 2008年, オーム社)の図を改変

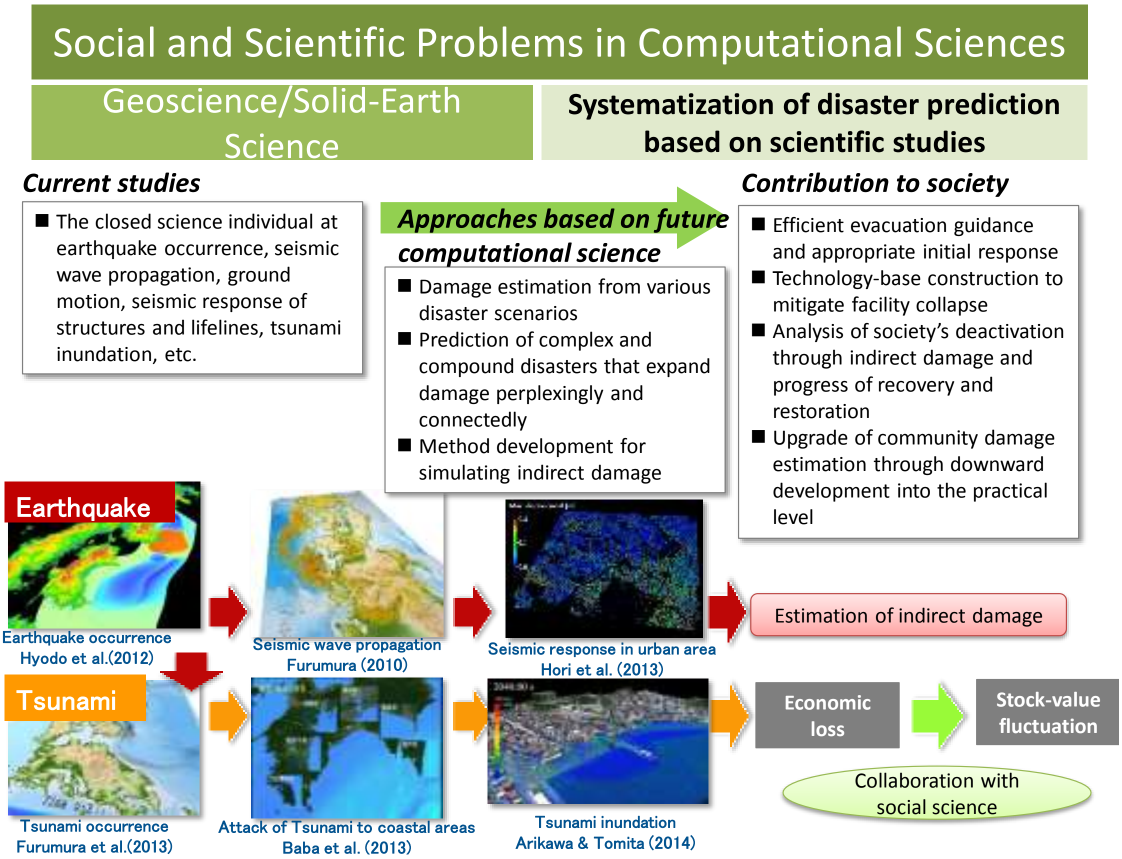
Monitoring and projection system for global environment



Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
High resolution global weather prediction	130	360	3	58	340	1	150000	Model: NICAM, FVM	1,000,000,000,000grids (220m horizontal resolution and 94 vertical layers), 5,200,000steps (dt=1sec, 2months integration)	The number of the assumed total nodes is 100,000 (side-by-side comm.:1GB/s).
High resolution regional weather prediction	33	33	0.09	0.3	0.5	2700	160000	Model: asuca, FVM	7,500x7,500x500grids, 130,000steps (dt=1sec, 36hours integration)	FLOP and memory usage are estimated from a profile data obtained on SR16000. The memory throughput is estimated by an assumption that B/F is one. 25 variables are output every 10 minutes. 22,500 nodes are assumed for estimation of communication. (side-by side comm; 40GB/s)
Prediction of global environment changes	56	110	0.6	80	600	1	120000	Model: MIROC-ESM (spectral model for atmosphere, FDM for ocean)	2,000x1,000x200grids, 53,000,000steps (dt=60sec, 100years integration)	Estimated by atmospheric model. 100 cases of calculations must be completed in one month. It is expected that All-to-all network communications are used in 1000 nodes (1TB/s per node). The magnitude of FLOP, memory throughput and memory usage are estimated based on a profile data obtained on K.
Improvement of the weather prediction skill using the data assimilation	2.5	5	4.8	0.0003	0.5	6100	28000	System: JNoVA (4D-var for assimilation)	4,000x3,000x150grids, 2,700steps, 50 iterations.	FLOP and memory usage are estimated from a profile data obtained on SR16000. The memory throughput is estimated by an assumption that B/F is 2.

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

(2) Solid-Earth Science



Social and Scientific Problems in Computational Sciences

Geoscience/Solid-Earth Science

Mechanism of Geomagnetic Field Variation

Current studies

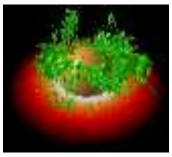
- Advances in understanding the fundamental physics of core convection and the dynamo process in the Earth's core
- Unknown behavior of the actual turbulence and the dynamo process in the Earth's core due to the impossibility of adopting the physical properties of the Earth's core
- Difficulty in investigating the actual geomagnetic field variation

Breakthroughs with computational science

- Adopting the closer or real physical properties in models to study the actual Earth's core convection, dynamo process, and geomagnetic field variation
- Taking account of recent research results of experiments and seismology and influences outside the core in order to studying geomagnetic field variation

Contribution to society

- Precise simulation of the geomagnetic field strength for determining habitability of planets
- Forecasting future changes in geomagnetic field through models that can reproduce past geomagnetic field variations
- Development of space climate research through future prospects of geomagnetic field variation



Three dimensional numerical simulation of the Earth's core convection and geodynamo (Miyagoshi & Kageyama)

If realistic physical properties can be treated using high-performance supercomputers in the future, the forecast of geomagnetic field variations will be possible.

Social and Scientific Problems in Computational Sciences

Geoscience/Solid-Earth Science

Material Circulation by Mantle Dynamics

Current studies

- Strong trade-offs in modeled spatiotemporal scales
- Imprecise modeling of interactions between plates/core
- Expensive solutions for flow field with extremely slow motion
- Difficult solutions for flow fluids with extremely different rheologies (mantle, magmas, plates)

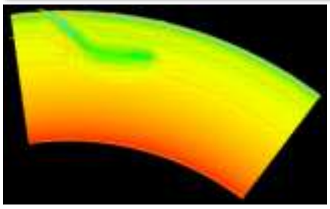
Breakthroughs with computational science

- Efficient/accurate solution for ill-conditioned elliptic problems
- Efficient techniques for coupled systems of differential equations
- Modeling techniques with hierarchy of spatial resolutions
- Combination of fluid-dynamic techniques and particle-based ones

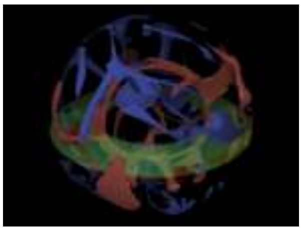
Long-term objective

- Self-consistent modeling of interactions among plates, core, and mantle
- Dynamic feedback from changes in surface (topography, climates) and/or internal (temperature, composition) environments
- Influences of planetary sizes on their habitability

Local modeling near core-mantle boundary(CMB)



Regional modeling of plate subduction zone



Self-consistent global modeling

Figures from Dr.Kameyama

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Coupled simulation of natural hazard and disaster (prediction of earthquake disaster right after earthquake occurrence) (1) – (6)	7	15	0.1	9	3		310000			(1): 5 domains, 1000 cases for each domain. (2): 20 cases are selected from 1000 cases based on observation (for each domain). (3),(4),(5),(6): 1000 cases for each domain; (parameter studies about the ground structures, the deterioration of building and the submarine topography) * (several large city areas) * (selected case in (2)) Required BF ratio = 8.0
(1) earthquake generation			0.00086	0.00086		5000	48	earthquake cycle simulation using boundary integral method	number of elements: 10^7	Required B/F ratio = 4
(2) seismic wave propagation			0.1	0.5		100	1400	computation of elastic wave propagation using finite difference method	1200 x 1000 x 200 km^3 (grid of 125 x 125 x 62.5 m), 240,000 steps	Original version: Required B/F ratio = 2.14, 14EFLOP per case New version on K by Dr. Maeda: Required B/F ratio =1.4, 20EFLOP per case
(3) ground motion amplification			0.01	4		5000	130000	computation of seismic wave propagation using finite element method	number of nodes: 300,000,000,000 (300 x 250 x 10 km^3)	Required B/F ratio = 8
(4) ground motion amplification			0.01	4		5000	130000	computation of seismic wave propagation using finite element method	number of nodes: 300,000,000 (30 x 25 x 1 km^3)	Required B/F ratio = 8
(5) structure seismic response			0.05	0.05		5000	500		number of buildings: 1,000,000	Estimated from the profile data. Required B/F ratio = 0.26
(6) tsunami inundation			0.002	0.5		5000	50000	Computation of Navier-Stokes equations (hydrostatic state approximation, non-hydrostatic state approximation, Volume of Fluid Method)	Concurrent execution of 7 cities with composite grid, which domains are 3 x 3 x 0.08 (1 city region, grid width: 1m) to 1400 x 1100 x 10 km (grid width: 5.4km), 720,000 steps.	Estimated from the profile data. Required B/F ratio = 10
mass evacuation simulation	3.3	0.28	0.3	0.006	1	5000	60000	human behavior simulation using multi agent model	300,000 agents, 18,000 steps (1 hour simulation), 1,000 Monte-Carlo members	For total number of the operations, instruction is used instead of FLOP (IPS is 40 times larger than FLOPS). Estimated from the profile data on K.
mantle convection	1000		0.01		0.083	1	300	multigrid method using cartesian grid	number of grids: 290 x 4000 x 2000 (4 variables per node)	
Earth's dynamo			0.053	4		1		explicit/implicit method using yin-yang grid	number of grids: 2000 x 2000 x 6000: (2 or 8 variables per node)	

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

4.4 Manufacturing Innovation

Social and Scientific Problems in Computational Sciences

Manufacturing Innovation — **Innovative Design Using a High-Performance Computing**

Current studies

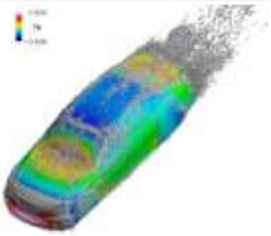
- Trial and error designs by empirical methodologies with primitive models
- Limited use of simplified simulations for design evaluation or understanding of phenomena
- Component-level simulations
- Small-scale parametric study
- Simple data analysis methods such as 2D visualization

Breakthroughs with computational science

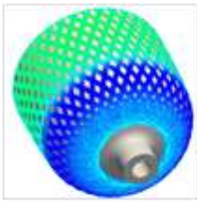
- First-principle, whole system and multidisciplinary simulations using large-scale models (high detail and high fidelity)
- Large number of simulations to enable optimal design and design exploration
- Techniques for information extraction from large data

Contribution to society

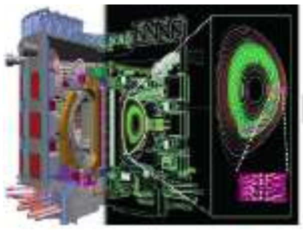
- Achieving high reliability and cost reduction by substitution of experiments
- Strict evaluation of structural reliability and safety
- Creation of new breed of energy devices
- Efficient discovery of design knowledge using data analysis framework



LES simulation
of an automobile



Strength assessment of high-pressure hydrogen tank



Plasma simulation
of a fusion reactor

(1) Thermal Fluid

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Thermo-fluid, vibration and acoustic analysis on turbomachinery	18	100	5	10	120	20	160000	Finite element method	10^13 grid	Calculated as B/F=0.5
LES simulation (automobile, industrial design, optimization problem)	280	560	0.04	4	1	100	100000	LES calculation of Re=10^6 ~10^7	10^11 grid	Calculated as B/F=2
Thermo-fluid and acoustic analysis on electronic devices	0.46	2.5	0.1	1.6	12	1000	20000	Finite element method	10^10 grid	
Aerodynamic design for wing and fuselage of aircrafts, and aeroacoustic analysis on engine and airframe	7.9	20	0.092	8	24	1000	680000	Difference method	10^11 grid	
LES design for spacecraft, propulsion analysis, all system analysis Thermo-fluid design for spacecraft (propulsion system and whole system)	40	99	0.92	80	240	10	340000	Difference method	10^12 grid	
Air flows inside/outside of buildings in cities, and pollutant diffusion analysis	120	490	4	160	96	10	430000	Finite element method	10^13 grid, 10^4 steps	

(2) Structure Analysis

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Vehicle crash simulation	540	27	1	100	24	10	470000	Finite element method (explicit method)	10^11 node	
Steel-sheet stamping/ elasto-plastic analysis	54	2.7	1	1	24	10	47000	Finite element method (implicit method)	10^10 node	
Detail analysis on whole nuclear reactor	540	27	10	10	24	10	470000	Finite element method (implicit method)	10^11 node	

(3) Machine Materials Study

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Strength assessment of functional metal for electronic parts	31	38	0.2	500	10	10	11000	Accelerated Molecular Dynamics Simulation	tensile simulation of copper polycrystal using 40nm grain diameter, 1 micro second, 1000 replicas	Acceleration rate using replica is assumed 666 times per 1000 parallel
Development of carbon fiber reinforced plastics	3.3	160	0.03	500	2	30	720	Nonlinear finite element method	Implicit solver simulation of the test specimen of 30cm, defect size 50 μ m, 10000 steps	—

(4) Plasma/ Fusion Science

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Plasma turbulence calculation/ multi-scale turbulence	100	200	0.5	0.1	24	50	430000	Five-dimensional calculation by Boltzmann equation (spectral method + finite difference method)	10^12 grid, 10^6 steps	Calculated as B/F=2
Plasma turbulence calculation/ transient evolution of global turbulence	100	200	0.5	1	170	10	610000	Five-dimensional calculation by Boltzmann equation (finite difference method)	10^12 grid, 10^7 steps	Calculated as B/F=2

(5) Electromagnetic Field Analysis

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Analysis on all levels of server equipment	3.2	5.3	0.072	0.6	1	20	230	Mix of explicit method and implicit method	10^12 grid	

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

(6) Visualization / Data processing

	File based visualization	In situ visualization
Estimated performance (per node)	100TFLOPS	100TFLOPS
Network bandwidth (per node)	500GB/s	2TB/s
Memory size (per node)	50GB	100GB
Memory bandwidth (per node)	0.2TB/s	20TB/s
Storage size	2 x The size of the maximum simulated results	The size of the maximum simulated results
Storage bandwidth	0.1 PB/s	0.01 PB/s

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

4.5 Fundamental Physics

(1) Astrophysics

Social and Scientific Problems in Computational Sciences

Fundamental Physics/Astrophysics | **History of the Universe**

Current studies

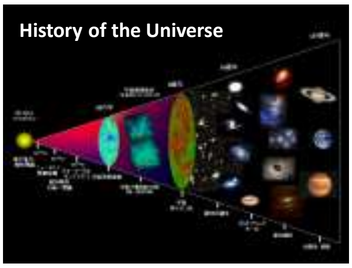
- Establishment of the Big Bang theory and light nucleosynthesis, discovery of the presence of dark matter and dark energy via high-precision cosmic microwave background measurements
- Unknown mechanisms and relationship among large-scale structure formation, galaxy formation, and dark matter and dark energy.
- Structure formation simulations using N-body/Hydrodynamics with a large step size and short time evolution
- Approximate radiation-hydrodynamics simulations for star formation, black hole formation, and inter-galaxy gas distribution
- Low-resolution simulations for large-scale structure formation and particle acceleration in plasma shock wave based on the Boltzmann equation
- Simulations (scale and period limited) for formation of astronomical objects and hydrogen distribution in the early universe

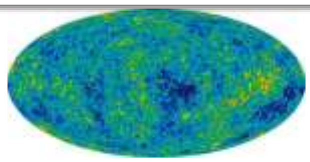
Breakthroughs with computational science

- Simulations with more particles and a longer evolution time
- Improving algorithms and methods in the radiation transfer equation simulations for the large-scale structure formation, galaxy formation, and star formation
- Galaxy formation simulations combining dark matter effects and gravity in the early universe, unified simulations for the large-scale structure, galaxy, and star formations incorporating inter-galaxy gas effects

Contribution to society

- Development of outlook on nature by understanding and tracing back the history of the universe
- Expanding the knowledge of human beings





Cosmic microwave background radiation
(Credit: NASA / WMAP Science Team)

Large-scale structure of the universe (galaxies)
(Credit: M. Blanton and the Sloan Digital Sky Survey.)

Social and Scientific Problems in Computational Sciences

Fundamental Physics/Astrophysics | **Search for New Physics**

Current studies

- Discovery of and unknown origin of Gamma ray bursts, the most high-energy-density phenomenon in the universe
- Unknown phenomena resulting from black hole formation
- Unknown mechanism of supernova explosions
- Unknown origin of dark matter and dark energy
- Unknown origin of ultra-high-energy cosmic rays
- 3D simulations with general-relativity and relativistic hydrodynamics (approximated) for accretion flows and jets in astronomical objects
- Simulations with approximated neutrino transfer for supernova explosions
- Low-resolution electrodynamical plasma particle simulations

Breakthroughs with computational science

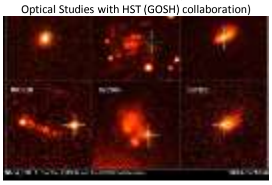
- Simulations for the gamma ray burst and supernovae with exact neutrino transfer, general relativity, and relativistic hydrodynamics
- High-resolution electrodynamical plasma particle simulations to uncover the origin of ultra-high-energy cosmic rays

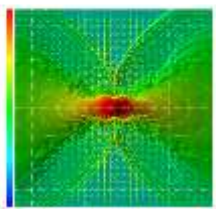
Contribution to society

- Discovering and establishing new laws of physics by understanding various phenomena using fundamental physics
- Potential benefits and everyday applications in the long term through the study of fundamental physics

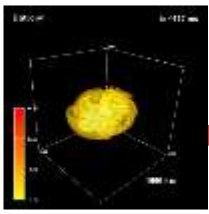
Supernova remnant
(Cassiopeia A, false color)
(Credit: NASA/JPL-Caltech/STScI/CXC/SAO)

Gamma ray burst images by Hubble space telescope
(Credit: NASA, ESA, Andrew Fruchter (STScI), and the GRB Optical Studies with HST (GOSH) collaboration)





3D relativistic simulation of gamma ray bursts
(Sekiguchi, Shibata *et al.* 2012)



3D hydrodynamic simulation of a supernova
(Takiwaki *et al.* 2013)

Social and Scientific Problems in Computational Sciences

Fundamental Physics/Astrophysics

Planetary Formation and Astrobiology

Current studies

- Planetary formation via dust accumulations in the protoplanetary disk
- Simulations for the formation of planets in a narrow ring
- Simulations for the Moon formation via a giant impact of proto-planets and for the internal structure of Earth
- Discovery of outer solar planets and candidate habitable planets (2nd Earth), and evolution of astrobiology
- Simulations for the earth's surface environment and its application to outer solar planets
- Quantum chemical simulations for the origin of the asymmetry of optical isomers of amino acids and origin of life outside Earth (Panspermia hypothesis)

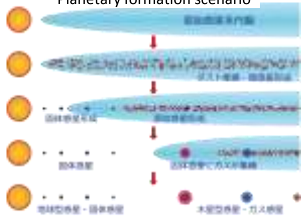
Breakthroughs with computational science

- Simulating planetary formation for a long evolution time and with a wider range ring over 10 au in the proto-planetary disk
- Simulating the simultaneous formation of the Moon, the inter structure of Earth, and the surface environment of Earth via a giant impact of proto-planets
- Simulating the surface environment of outer solar planets under various initial conditions
- Direct quantum many-body chemical simulations for the production of the non-racemic amino acids by circularly polarized light in space

Long-term objective

- Resolving questions about the existence of other earths in the universe, and the existence of life outside of earth.
- Inquisition into mystery, meaning, and genesis of life in the Universe

Planetary formation scenario



Habitable Planets (2nd Earth)
(Credit: Kepler/NASA)

Social and Scientific Problems in Computational Sciences

Fundamental Physics/Astrophysics

Promotion of Space Environmental Science

Current studies

- Understanding of the solar system and of the relationship between climate changes and the activity of the Sun
- Mechanisms of solar magnetic activities, space environment, and solar flare
- Simulations with 3D-diffusive magnetohydrodynamics with low-resolutions and low-Reynolds numbers
- Particle acceleration and shockwave generation by solar winds
- Simulations for shock wave particle accelerations in the solar system using plasma particle-in-cell codes at a low number (10^{14}) of particles
- Developments and experiments for space weather forecasting

Breakthroughs with computational science

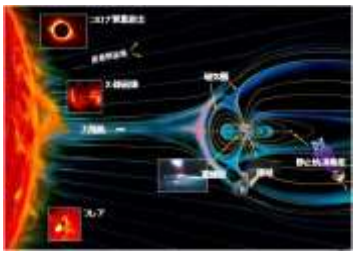
- Simulating solar turbulent flows in a sufficiently long evolution time at a large Reynolds number and high resolution using 3D-diffusive magnetohydrodynamics
- Simulating shock wave particle accelerations using plasma particle-in-cell codes at a large number (10^{16}) of particles covering supernova phenomena
- Space weather forecasting using data-driven simulations with precision measurements of the solar surface and corona

Contribution to society

- Long-term predictions for space environment changes and the solar activity
- Increase the safety and reliability of satellites, spaceships, space activities, and availabilities via the space weather forecast


X-ray image by "Hinode"

JST project (Nagoya University)



Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
N-body/hydrodynamic simulations for the formation of the structure of the Universe	420	1.4	5	100	1000	1	1500000	Hybrid algorithm of tree and an individual timestep	10^14 particles	100Tflops/node x 10000node, communication b/w 30GB/s/node
Radiation hydrodynamic simulations on the formation of galaxies and massive black holes	50	0.63	2	1.2	550	1	98000	Tree-based Radiative Transfer + SPH	4096^3 particles + 6x10^7 radiation sources	100Tflops/node x 10000node, communication b/w 100GB/s/node
6-dimensional Boltzmann equation for the collisionless particle dynamics	45	34	2	2	3.3	10	5400	Finite volume method	Position 256^3, velocity 256^3	100Tflops/node x 10000node, communication b/w 1000GB/s/node
Evolution of the dark matter universe in its dark ages	420	1.4	1	2	20	1	30000	Particle-Mesh + FFT	10^13 particles + 10^5 light sources and 10000 timesteps	100Tflops/node x 10000node, communication b/w 128GB/s/node
Self-gravity radiation hydrodynamic simulations of galaxy scale interstellar gas	1000	0.31	2	10	1000	10	36000000	Tree-Based Radiation Transfer + mesh hydrodynamics (AMR)	8192^3 mesh ＋10^8 light sources	100Tflops/node x 10000node, communication b/w 128GB/s/node
Radiation magnetohydrodynamic simulations of accretion flows and outflows	100	20	0.2	200	1000	2	720000	Approximate Riemann solver for relativistic magnetohydrodynamic equations + 6-dimensional radiative transfer	512^3 grids, 1000 directions of light beams, 100 frequency bins, 3.6x10^7 time steps	100Tflops/node x 10000 node, communication b/w 30GB/s/node
Black hole formation and strong gravity by numerical relativity	1000	100	0.04	50	28	10	1000000	4-dimensional RK, Rad-HRSC	1000^3,10^7steps	100Tflops/node x 10000node, communication b/w 2.88GB/s/node
Supernova simulations exploring supernova explosion by relativistic radiative hydrodynamics	18	70	1.6	1.3	1200	10	780000	Neutorino radiation transport (supernova explosion)	1 sec simulation with the size of space 512x64x128 and momentum space 24^3 grids	100Tflops/node ×10000node, communication b/w 60GB/s/node
High energy density physics and particle accelerations by relativistic particle simulations	310	92	96	1000	200	2	450000	Particle-in-Cell method	4096^3 grid points with 10^15 particles and 10^5 timesteps	100Tflops/node x 10000 nodes, communication b/w 1GB/s/node
6D Vlasov simulations of non-thermal plasma accelerations	24	1.5	50	500	1400	2	240000	semi-Lagrangian method	6D simulation with 1024^3 (configuration space) x 265^3 (velocity space) grid points	100Tflops/node x 10000 nodes, communication b/w 1GB/s/node
Quantum mechanical calculations on cosmic Amino Acid	1000	0.1	1		600	1	2200000	Quantum dynamic calculations with Surface hopping method	20 Amino Acid * 10 initial condition * 3000 surface hopping	100Tflops/node x 10000 nodes, communication b/w 100GB/s/node
Solar and stellar dynamo by radiative magnetohydrodynamics	100	88	7	13	410	1	150000	Reduced speed of sound technique (RSST) + Yin-Yang grid	1024x8192x24576x2 grids, 5x10^7 timesteps	100Tflops/node x 10000nodes, communication b/w 1GB/s/node
Collisionless shocks in space and astrophysical plasmas	160	46	96	1000	1400	2	1600000	Particle-in-Cell method	72000x3072^2 grid points with 10^16 particles	100Tflops/node x 10000 nodes, communication b/w 1GB/s/node
Advanced Simulation Study of Solar-Terrestrial Environmental Science	1000	2	2		100	1	360000	magnetohydrodynamic finite element and finite volume scheme, plasma hybrid scheme, particle-in-cell scheme, etc.	3000^3 grids	100Tflops/node x 10000node, communication b/w 100GB/s/node

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

(2) Elementary Particle Physics

Social and Scientific Problems in Computational Sciences

Fundamental Physics/Elementary Particle Physics	Multi-scale Physics and Physics under Extreme Conditions Based on Lattice Quantum Chromodynamics (Lattice QCD)

Current studies

- Determination of the Standard Model (SM) parameters
- High-precision computation of hadron masses
- Computations at single energy scale
- Simulating quark matter at finite temperature
- Developments on computational methods for hadron-hadron interactions
- Simulations with chiral symmetry in a small volume with a coarse lattice spacing
- Simulations at finite density under algorithmic difficulties

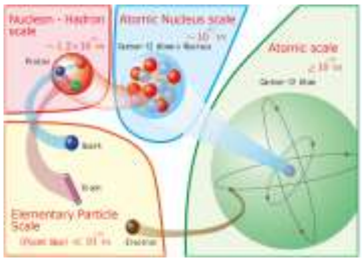
Breakthroughs with computational science

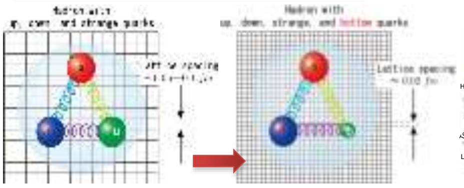
- Computing multi-hadron interactions in a large volume with chiral symmetry
- Simulating bottom quarks with finer lattices for precision measurements
- Algorithmic developments for simulations at finite temperature and finite density

Long-term objective

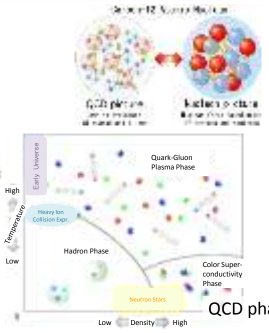
- Understanding multi-scale physics: quarks ⇒ nucleons ⇒ nuclei
- Search for new physics beyond the Standard Model via bottom quark physics
- Cosmology and Astrophysics based on particle physics in extreme environments

Multiple energy (length) scales
Quarks (elementary particle) ⇒ Hadrons (nucleon) ⇒ Atomic nuclei





Computation including bottom quark



QCD phase diagram expected

Social and Scientific Problems in Computational Sciences

Fundamental Physics/Elementary Particle Physics	Understanding the Origin of Matter

Current studies

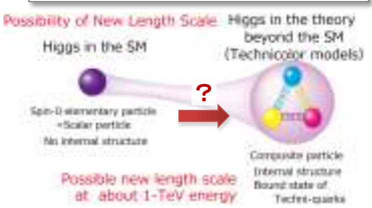
- Discovery of the Higgs particle and establishment of the SM
- Solving hierarchy problem in the SM
- Theoretical study on Technicolor models
- High precision QED calculations for muon anomalous magnetic moment (g-2)
- Automated calculations for Feynman diagrams at one-loop
- Theoretical analysis on the superstring theory and its numerical simulations with matrix models

Breakthroughs with computational science

- Nonperturbative study on the Technicolor models using the lattice QCD technique
- High precision QED calculations for muon g-2 against high precision experiments
- Automated calculations for two-loop Feynman diagrams with multi final states for the ILC experiment
- Development of simulation methods for the matrix models

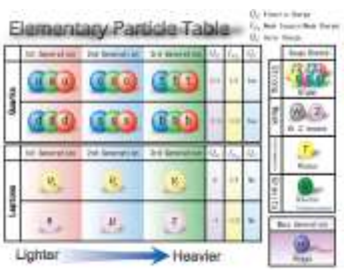
Long-term objective

- Discovery of new physics beyond the SM and the origin of the hierarchy in the SM through combination of precision experiments, theory, and numerical simulations
- Exploration into the origin of the universe, space-time, and matter beyond the SM



Origin of the Higgs particle

Precision calculation for the muon anomalous magnetic moment (g-2)



Elementary particles in the standard model (SM)

Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Determination and application of effective Bryon interactions based on chiral symmetry and QCD	510	390	0.066	0.5	880	10	16000000	Lattice QCD(Chiral fermions in 5-dimensional representation), Hybrid Monte Carlo algorithm, CG solver	Problem size: Lattice size 128^4x32, lattice spacing < 0.1 [fm]	16^4 nodes are assumed. on-chip memory size 200MB, on-chip memory bandwidth 6TB/s, network latency 1usec, network bandwidth 128GB/s par node.
Heavy quark physics	510	370	0.021	1	880	10	16000000	Lattice QCD(Wilson type fermions), Hybrid Monte Carlo algorithm, CG solver, BiCGStab solver	192^4	12^4 nodes are assumed. on-chip memory size 200MB, on-chip memory bandwidth 18TB/s, network latency 1usec, network bandwidth 128GB/s par node.
Particle physics for microscopic hierarchical structure of matters in extreme environments	510	1200	0.066	0.2	880	10	16000000	Lattice QCD(Wilson type fermions), Hybrid Monte Carlo algorithm, CG solver, BiCGStab solver	256^4	16^4 nodes are assumed. on-chip memory size 200MB, on-chip memory bandwidth 18TB/s, network latency 1usec, network bandwidth 128GB/s par node.
Non-perturbative dynamics in Technicolor theories	510	1200	0.46	0.05	880	10	16000000	Lattice QCD(Chiral fermions in 5-dimensional representation), Hybrid Monte Carlo algorithm, CG solver	96^4x32	16^4 nodes are assumed. on-chip memory size 200MB, on-chip memory bandwidth 18TB/s, network latency 1usec, network bandwidth 128GB/s par node.
Higher order corrections in Quantum Electrodynamics (QED) (multi-precision arithmetic)	1.8	1.3	0.00012		24	220	34000	Multi-dimensional integrations by Monte Carlo methods	Over ten-thousand multi-dimensional integrations (8～18 dimensions for each)	Array job consists of ten-thousand single node jobs. SIMD and core parallelization are required. Compilation speed is also important as the integrant in program source level is extremely huge. We require 2～3 years for the completion of all computation. The flop counts are converted to double precision flop counts from multi-precision flop counts. Multi-precision flop count is larger than that of double precision flop count by a factor 30.
Automated computation and evaluation of Feynman diagrams (quadruple precision arithmetic)	3.2	0.13	2E-09	0.0005	24	1000	280000	Multi-dimensional integrations by Monte Carlo methods	About 350,000 diagrams containing Two-loop diagrams. This is required for the theoretical computation of, Bhabha, ZH processes at 250GeV, and Bhabha, ZH and top quark pair production processes at 350 GeV at ILC.	Parallelization is embarrassingly parallelizable. High speed quadruple precision arithmetic with 15bits exponents (binary 128 format of IEEE754-2008) are required. The program source size is extremely large and the compiling speed is also important. Multi-precision arithmetic higher than quadruple precision are required in some cases. We require 0.5～1 year for the completion of a single elementary particle reaction process. The total flop and performance requirements are estimated for quadruple precision. 350 diagrams are computed in each case.

(3) Nuclear Physics

Social and Scientific Problems in Computational Sciences

Fundamental Physics /Nuclear Physics | **Ab Initio Computation of Nuclear Structure**

Current studies

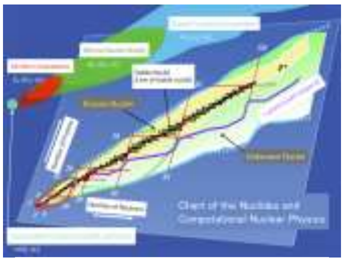
- Developments in microscopic description for various aspects of nuclear structure
- Many-body quantum system of protons and neutrons, with complicated nature of nuclear force
- *Ab initio* computation for light nuclei with about 10 nucleons
- Nuclear shell-model computation in a small model space assuming the frozen closed shell configurations
- Theoretical developments in the effective interaction in nuclei
- Large uncertainties in three-body nuclear forces

Breakthroughs with computational science

- Computing atomic nuclei with about 30 nucleons using Monte Carlo shell model method without assuming an inert core
- Extending the model space to one with 6-8 major shells
- Developing techniques to reduce computational costs and accurate effective interactions.

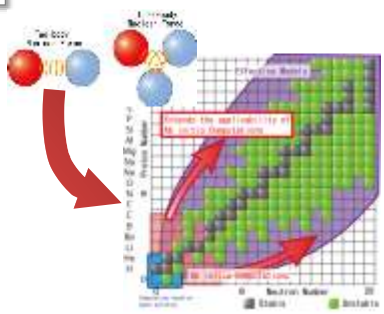
Long-term objective

- Developments of microscopic computational nuclear physics
- Prediction and discovery of new phenomena in nuclei and hadrons
- Uncovering nuclear properties, structure and reaction, in response to social demands
- Application to astrophysics and nuclear engineering

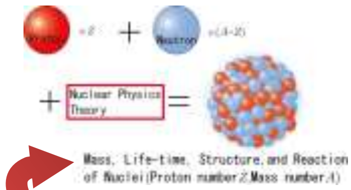


Computational Nuclear Physics and Chart of the Nuclides

"G.F.Bertsch, D.J.Dean, W.Nazarewicz, SciDAC Review (UNEDF collaboration)"



Nucleus on demand



- *Ab initio* calculations of structure and reaction of light nuclei
- Strict test of effective nuclear models, leading to reliable computation of the structure and reaction for heavy nuclei

Social and Scientific Problems in Computational Sciences

Fundamental Physics /Nuclear Physics | **Unified Elucidation of Nucleus Structure and Reactions**

Current studies

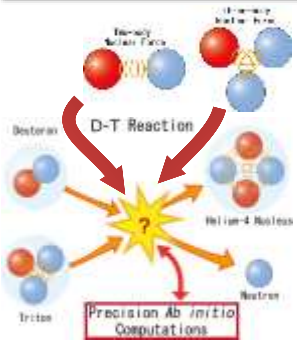
- Developments in microscopic description for various aspects of nuclear structure
- Many-body quantum system of protons and neutrons, with complicated nature of nuclear force
- Effective unified models for nuclear structure and reaction
- Developments in density functional calculation for heavy nuclei and its extension to nuclear reaction
- Description of nuclear fission in terms of macroscopic models

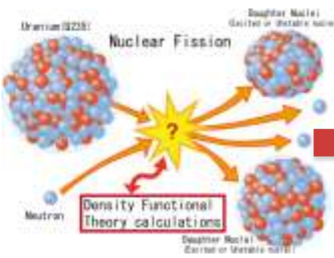
Breakthroughs with computational science

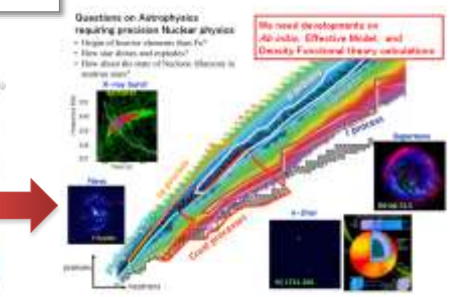
- Computing structure and reaction of light nuclei with realistic finite-range nuclear forces
- Large-scale real-time simulation for heavy-ion reaction
- Determining nuclear equation of state at low densities
- Microscopic description of nuclear fission

Long-term objective

- Developments of microscopic computational Nuclear Physics
- Prediction and discovery of new phenomena in nuclei and hadrons
- Uncovering nuclear properties, structure and reaction, in response to social demands
- Application to Astrophysics and nuclear engineering







Picture, courtesy of H.Schatz and W.Nazarewicz.

Social and Scientific Problems in Computational Sciences

Fundamental Physics /Nuclear Physics

Understanding Quark-Gluon-Plasma of the Early Universe

Current studies

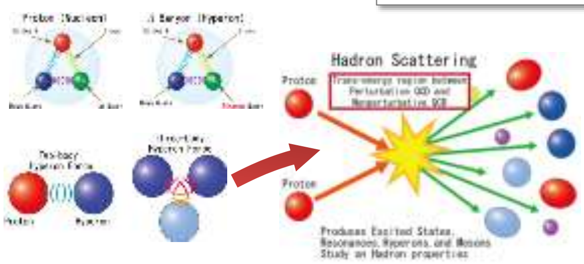
- Developments in description of nucleons and nuclei in terms of quarks and gluons
- *Ab initio* approaches for atomic nuclei using lattice QCD simulations
- Nuclear equation of state at high densities based on effective models
- Large uncertainties in three-body nuclear forces and hyperon forces
- Quark-gluon-plasma production in heavy ion collision experiments and difficulties of simulation in non-equilibrium dynamics
- Computation of scattering amplitude in transition energy region between QCD and hadron physics

Breakthroughs with computational science

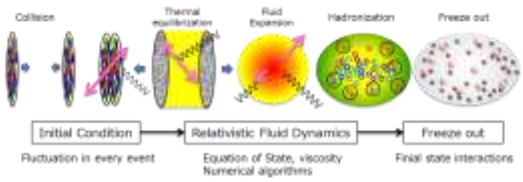
- Nuclear equation of state based on precision nuclear forces computed with lattice QCD
- Hyperon forces with lattice QCD
- Unified and total simulations for relativistic heavy-ion collision events incorporating fluctuation effects
- Constructing basic theories for scattering-amplitude calculation incorporating strangeness productions

Long-term objective

- Developments of microscopic computational Nuclear Physics
- Prediction and discovery of new phenomena in nuclei and hadrons
- Uncovering nuclear properties, structure and reaction, in response to social demands
- Application to Astrophysics and nuclear engineering



Physics in Heavy-Ion Collisions



Subject	Perfor-mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
First-principle calculation of nuclear structure	100	10	0.1	0.0001	28	100	1000000	Nuclear structure calculation of light nuclei using no-core Monte Carlo shell model	7-8 major shells in the harmonic oscillator basis are adopted as the model space of shell-model calculations.	The memory capacity is estimated as 10000 node times 10GB.
Nuclear shell-model calculations for medium-heavy nuclei	14	0.69	0.32	0.0001	10	1000	500000	Nuclear structure calculation of medium-heavy nuclei using Monte Carlo shell model	The model space of shell-model calculations is taken as full two major shells or beyond.	The memory capacity is estimated as 10000 node times 32GB.
Unified description of nuclear structure and reaction	53		0.03		100	50	950000	First-principle CI calculation using the generator coordinate method	Number of spatial mesh points:10,000. Number of configurations: about 100.	
Systematic description of nuclear response functions and construction of computational nuclear data	46	0.22	0.03	0.1	0.1	10000	160000	Diagonalization of linear-response matrices in the quasiparticle basis with the real-space representation	Systematic computation of response function for 10,000 nuclides with respect to each one-body field.	The computational nuclear data becomes realistic when the computational time for each nucleus becomes less than 10 minutes. Currently, innovative iterative methods are being improved, which may replace the diagonalization in future.
Microscopic description of nuclear fission phenomena	42	0.021	0.04	10	24	100	360000	Real-space, real-time calculation	Number of spatial mesh points, number of quasiparticles, and number of time steps are all about 100,000.	3x10^21 FLOP for a single case of the time evolution.
Phase structure and equation of state of nuclear matter	20	2.1	2.4	0.02	24	100	170000	Calculation of thermal equilibrium in terms of AMD methods	Calculation of equation of state for nuclear systems of 3,200 nucleons.	20,000 cases with different density/temperature/asymmetry and interactions. 300,000 time-step calculation for each of these.
Structure and reaction of light hypernuclei	57000	17000	180	0.00001	24	200	980000000	Exact calculation for quantum few-body systems using the gauss expansion method	Applications to seven-body systems (Generalized eigenvalue problems for a 64M x 64M dense matrix)	Based on performance of an eigenvalue calculation library, "EigenExa" (100Mx100M and 10Mx10M).
Relativistic heavy-ion collisions and properties of quark-gluon plasma								Hydrodynamical simulations of high-energy heavy-ion-collision experiments	Importance of fluctuations in the initial state are realized in the experiment. There are significant improvements in theories and models. Aiming at calculation based on the current best model which takes into account established physical situations.	(Current status) Development of algorithm for solution of the relativistic equation for viscous flow with shock waves. Examination of initial conditions. (Necessary developments) Numerical instability in cases of finite viscosity at low temperature.
Investigation of hadron resonances and interactions through multi-particle-production reaction	1.1	0.24	0.0002	0.000005	720	10	29000	Calculation for excited baryons, compared to massive scattering data. Microscopic calculation with multi-channel reaction models.	To calculate chi-2 values, we need to calculate 6,000 complex inverse matrices of 1,000 dimension. Then, it must be repeated by about 2.5 x 10^7 times.	(Current status) Convergence in chi-square optimization with error evaluation. (Necessary developments) Efficient method for the minimum search in the multi-dimensional parameter space. Speed up in the chi-square calculation. (Product run) Expected in 4-5 years with a current team.

The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

4.6 Social Science

Social and Scientific Problems in Computational Sciences

Social science

Social Forecasting System

Current studies

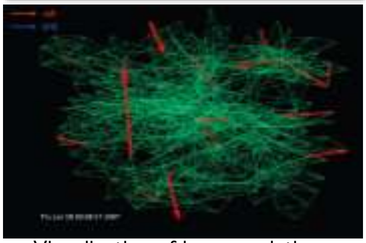
- Phenomenological ad-hoc modeling and simulations for each social subsystem like traffic flow, economic trade, community formation, etc.
- No reliable law even for simplest social phenomenon

Breakthroughs with computational science

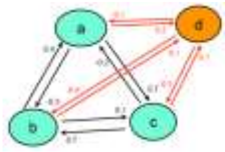
- Integrated model of social system with various subsystems
- Diverse but simple agent modeling
- Big-data mining of social information, and massive parallel search of vast combinatorial parameter-space

Contribution to society

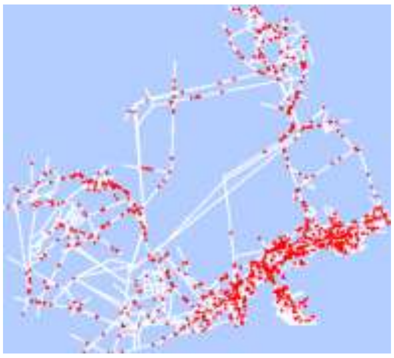
- Realization of social forecasting system for various issues



Visualization of human relations with mobile phone



Preservation and management of social diversity



Simulation of city traffic

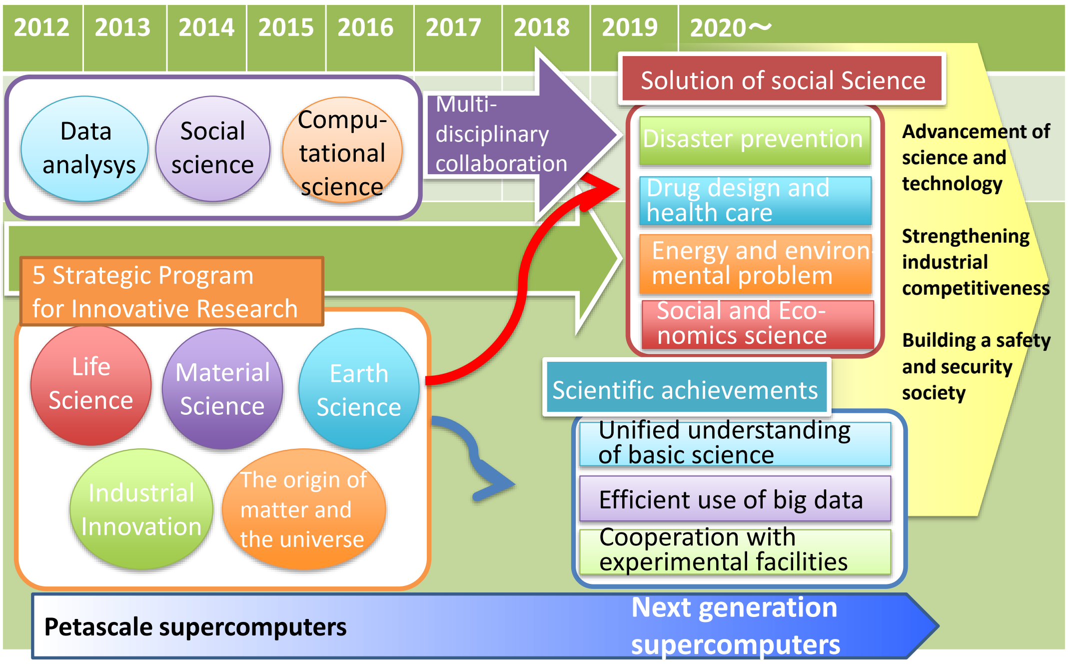
Subject	Perfor–mance (PFLOPS)	Memory bandwidth (PB/s)	Memory size per case (PB)	Storage size per case (PB)	Elapse Time /Case (hour)	Number of Cases	Total operation count (EFLOP)	Summary and numerical method	Problem size	Notes
Real time simulation of traffic flow	1000	100	0.00011	0.001	2.8E–08	1000	0.1	Agent simulations of global-scale traffic flow (1 billion vehicles, total road length of 34 million Km), (The number of operating cars are estimated to be 100 million)	10^8 vehicles x 10^3 operations x 10^3 steps x 10^3 cases (Simulation of 10 seconds) This should be simulated within 0.1 sec.	These amount of storage and total operations are required for each day. Total operations are estimated to be 10^3FLOP per one vehicles.
Optimization of the rules of stock trading	2100	0.0001	0.00000001		0.0024	10000	180000	Multi-agent Monte Carlo simulations of stock trading. (Duration: one day, Number of different stocks: 1000, Number of markets: 1)	Total number of operations: 5 hours x 3600sec/h x 1000 orders/sec x 10^4 operations/order x 10 traders x 10^4 cases x 10^3 individual stocks = 1.8 x 10^19 operations This should be simulated within 24hours.	Simulations mainly consists of integer operations. The figures in the table are number of instructions, not the number of floating operations.
Simulation of human relations								Agent simulation of 10^10 agents which shows a characteristic behavior depending on its community size.		Currently, it is not possible to estimate the requirements because simulation models and algorithms have not been established yet.

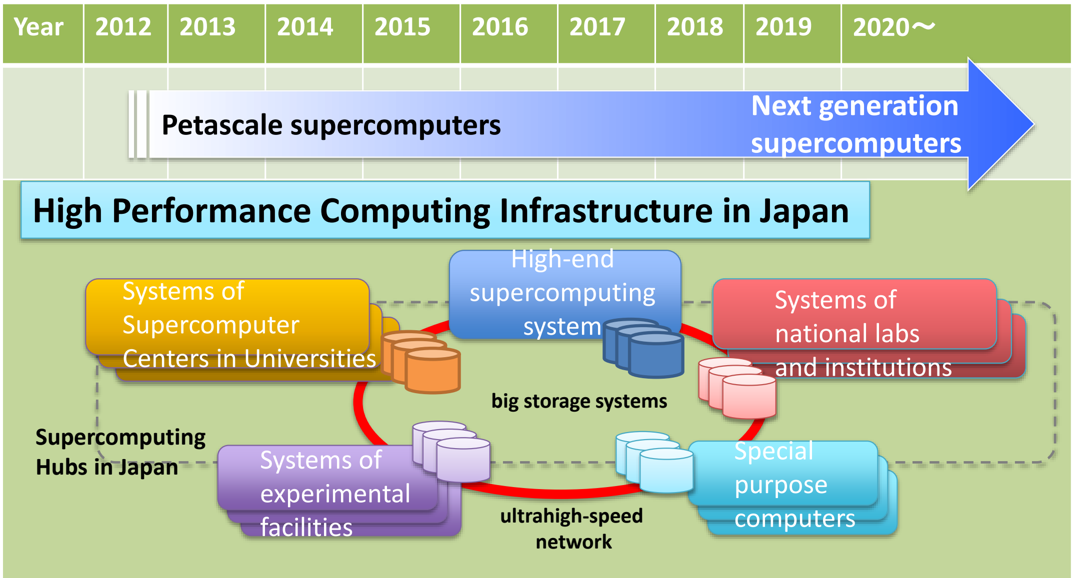
The estimates shown in this table are still under examination. The website will show more accurate figures as they become available.

Conclusion – Toward further development of computational science

With dramatic advances in computational science, we are now achieving computations of detailed models and computations for entire events that were previously not possible due to insufficient computing power. In the fall of 2012, the K computer went into full-scale operation, and through five areas of HPCI Strategic Programs, a wide array of research and development is currently being done using petascale computing. This research and development has provided valuable outcomes that enhance the socioeconomic activities we engage in, across the fields of drug discovery and healthcare, energy, manufacturing, and disaster prevention. However, there are also great expectations for the higher-performance computing environments provided by next-generation supercomputers with respect to the problems that our increasingly complex and globalized society is faced with today.

In keeping with the current research and development in the strategic fields, we have, in this report, approached the problems that we must solved in the next generation from the perspectives of both "social problems that computational science may contribute to in the future" and "the creation of new science through interdisciplinary collaboration." There is a strong demand for research and development that strives to achieve the technological innovations required to solve societal problems; there is also a demand for a continued mutually close collaboration with practical fundamental science to further develop computational science, solve the accumulating social problems, improve the quality of life, ensure safety and security, and establish a technology foundation that forms the groundwork for a developing and prospering industry.





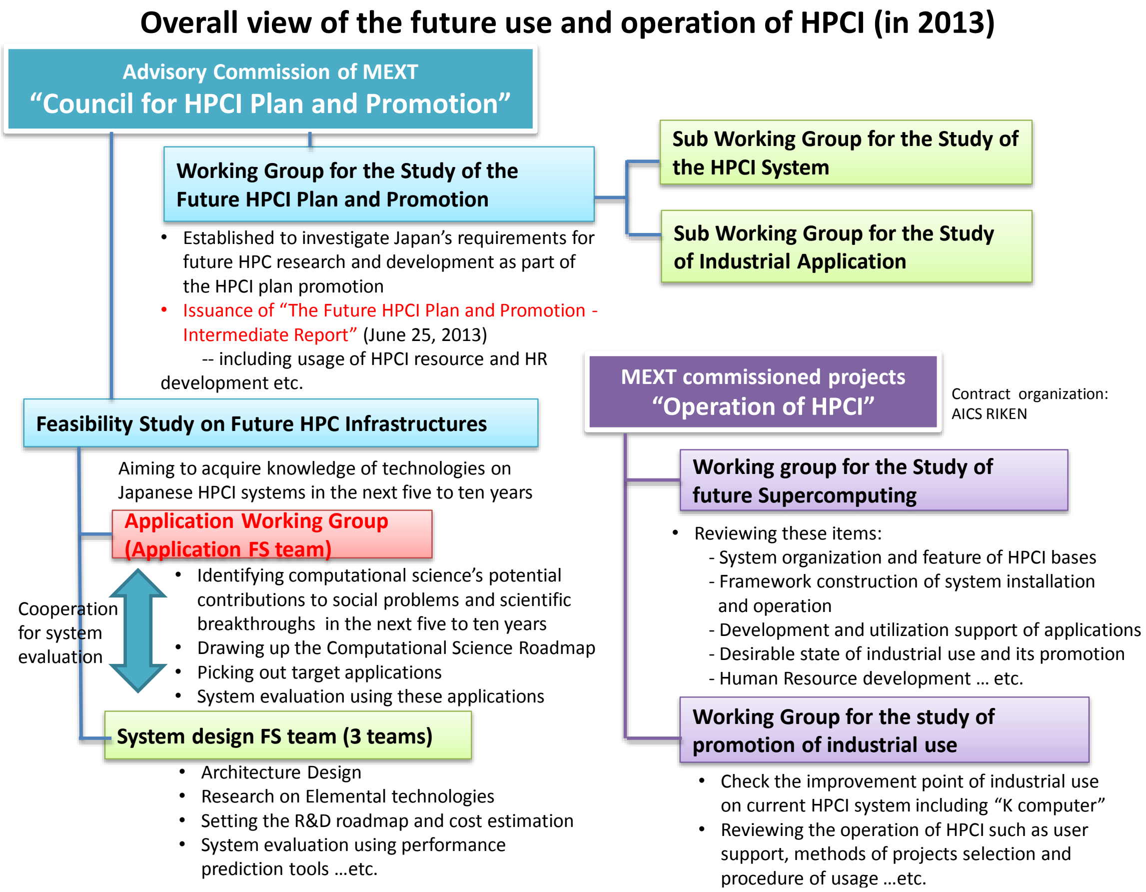
Besides the high-performance computing environments that single supercomputers provide, there are high expectations from the collaboration with large experimental facilities across a wide range of fields. This report talks about the collaboration between SACLA and computational science that promises important results in structural analysis and imaging, which require large-scale data processing for life sciences. Collaboration with large experimental facilities also promises further developments in science and technology in a wide array of areas, including polymers.

Merely building high-performance, large-scale computers will not guarantee that computational science will solve societal problems. We must also train excellent staff that can make efficient use of the high-performance computers, ensure sufficient human resources for model development and operation, and take care of the software side in preparing the appropriate organizations. The Future HPCI Plan and Promotion - Intermediate Report[6] (published on June 25, 2013, by the Working Group for the Study of the Future HPCI Plan and Promotion) has reported on a comprehensive policy for this kind of large-scale numerical computation, including HR development.

In July 2013, we asked members of the public for feedback on the roadmap. Many of the comments stressed the importance of developing computational science further and making the developed software available to a greater range of users, universities and SMEs, by providing access to the software across multiple platforms from PCs to supercomputers and by providing continual support and ample information online and in printed form. There were also comments emphasizing the importance of giving full consideration to the ethical issues regarding in the development of new technologies, given the fact that large-scale computational science entails risks that could potentially bring great damage to socioeconomic activity depending on how it is managed.

[6] http://www.mext.go.jp/b_menu/shingi/chousa/shinkou/028/gaiyou/1337595.htm

As described above, multifaceted research taking in both hardware and software is required to develop the future of computational science, and apart from the Feasibility Study on Future HPC Infrastructures (Application Working Group), which has produced this report, various research organizations are set up to study its future use and operation of HPCI, HR development, apps development, and promotion of use in the industry. The figure below gives an overall view of the future use and operation of HPCI.



This report is a result of an innovative initiative where experimental, observational, and theoretical scientists from the field of computational science, as well as approximately 100 researchers working on the front line of the various scientific communities from universities, research institutions, and corporations, engaged in discussions together. In order to keep developing computational science, we will continue to periodically engage in this kind of initiative. Starting with the system design field, the hardware and computational science is designed to produce results and further strengthen the collaboration between organizations that engage in this study from a variety of perspectives.

Computational Science Roadmap - Overview

Social Contributions and Scientific Outcomes aimed for by Innovations through Large Scale Parallel Computing.

May, 2014

MEXT commissioned projects
“Feasibility Study on Future HPC Infrastructures (Application Working Group) “
“Operation of HPCI”

http://hpci-aplfs.aics.riken.jp/