Frontier Simulation Software for Industrial Science

Project Representative
Chisachi Kato Collaborative Research Center of Frontier Simulation Software for Industrial Science
Institute of Industrial Science (IIS), the University of Tokyo

Authors
Makoto Tsubokura *1, Yuji Mochizuki *2, Katsumi Yamashita *3,
Tadashi Murase *3, Kenji Yamagishi *2, Hiroaki Tokiwa *2, Tatsuya Nakano *4,
Takenori Yamamoto *5 and Takahisa Ohno *6

*1 The University of Electro-Communications
*2 Rikkyo University
*3 NEC Soft, Ltd.
*4 National Institute of Health Sciences
*5 AdvanceSoft Corporation
*6 National Institute for Materials Science

In this project, we have chosen three important sub-projects- turbulence simulation, protein simulation and nano-scale simulations- as representative topics from the fields of engineering, life science, and nanotechnology and achieved the following results.

In the turbulence simulation, we conducted the world's largest engineering large eddy simulation (LES) of flow around a formula car (known as one of the most severe subjects in industry) to emphasize the validity of using high performance computing (HPC) technique to apply large-scale LES to industrial applications . The unstructured-grid LES code "FrontFlow/red" based on a finite volume (FV) scheme was intensively optimized for execution on the Earth Simulator, large-scale vector-parallel super computer, and we successfully simulated the unsteady turbulent flow around the formula car by using over 100 million elements.

In the protein simulation, we recently developed a parallelized integral-direct algorithm for the second-order Møller-Plesset perturbation theory (MP2), and implemented it into the ABINIT-MP program of the fragment molecular orbital (FMO) scheme. We conducted a series of benchmark FMO-MP2 calculations on the Earth Simulator of realistic proteins consisting of the tens of thousands of basis functions. The calculations were found to be highly accurate indicating that the ABINIT-MP program is easily applicable to realistic systems.

In the nanoscale simulation, we present very large-scale first-principles simulations using the PHASE code. PHASE is a plane-wave pseudopotential simulation package within the generalized gradient approximation, and is designed for large-scale vector-parallel super computers such as the Earth Simulator. We calculated the As-donor state in Si by using a super-cell model that contains 5,832 atoms, and revealed its fine structure that is due to intervalley interaction. These results are indispensable for designing future silicon nano-devices.

Keywords: large eddy simulation, aerodynamic force, formula car, FMO, MP2, silicon nano-device, first principles calculation, donor state, intervalley interaction

1. Assessment of aerodynamic force on a formula car by using a large scale LES

The objective of this sub-project is to emphasize the validity of applying large eddy simulation (LES) by applying high performance computing (HPC) technique to industrial applications, as a tool for engineering turbulence simulation. For this purpose, we conducted the world's largest LES of flow around a formula car, which is one of the most severe subjects in industry from the viewpoint of both the demand of very high accuracy and pressed development time. The unstructured-grid LES code FrontFlow/red based on a finite volume (FV) scheme was intensively optimized for execution on the Earth Simulator, a large-scale vector-parallel super computer. We achieved high vectorization and parallelization of more than 96.4% and 99.9%, respectively, in the simulation on 100 nodes (800 processors), and this

made possible conducting the large-scale LES by using over 100 million elements.

1.1 Validation of the large-scale LES on the ASMO model[1)]

Before the formula simulation, we first conducted the LES on the aerodynamic model ASMO (25 million elements) to estimate the validity of our LES method. The pressure distribution on the vehicle body surface obtained by our LES was compared with existing wind tunnel experimental data, and a good agreement was found (see Fig. 1). The most notable feature was the reproduction of the base pressure of the vehicle body, which, in contrast, was very poorly estimated by our RANS (Reynolds Averaged Navier-stokes Simulation, the conventional stimulation currently used in the automotive industry), such as the k-epsilon (k-e) or Reynolds Stress Model (RSM).

1.2 Unsteady 3-D turbulence simulation of the flow around a formula car (LOLA)

The LES of flow around the wind tunnel test model (1/2 scale) of the LOLA B03/51 was conducted using 120 million elements, which is, to the best of our knowledge, one of the largest engineering LES simulations. The large-scale flow structures that dominate turbulent flow field were directly simulated, and time-series behaviors of separated flows and vortices at the wake of the vehicle were identified (see Figs. 2 and 3). In the near future, greater attention will be paid to the unsteady aerodynamic forces acting on automobiles that are generated by sudden steering action in overtaking, or cross wind conditions. These are difficult to measure using conventional wind tunnel tests or RANS models. In addition to its high accuracy, an advantage of LES is its possibility of capturing such unsteady forces, and this will contribute to aerodynamic design innovation in the automotive industry.

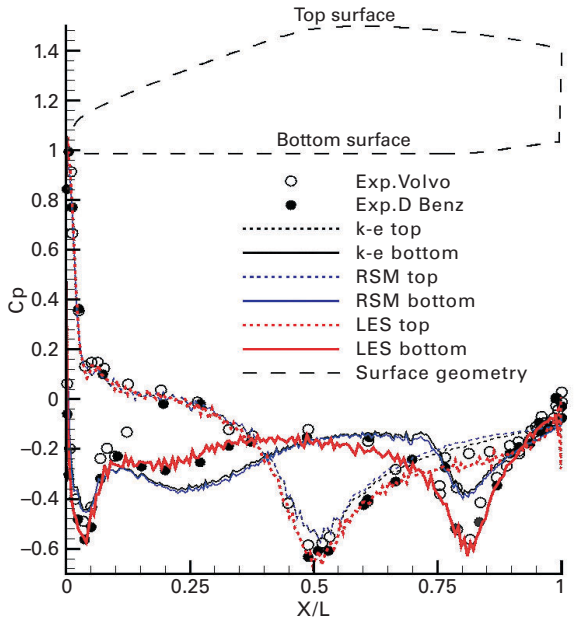


Fig. 1 Surface pressure distribution of ASMO model (Comparison with LES, RANS, and experimental data).

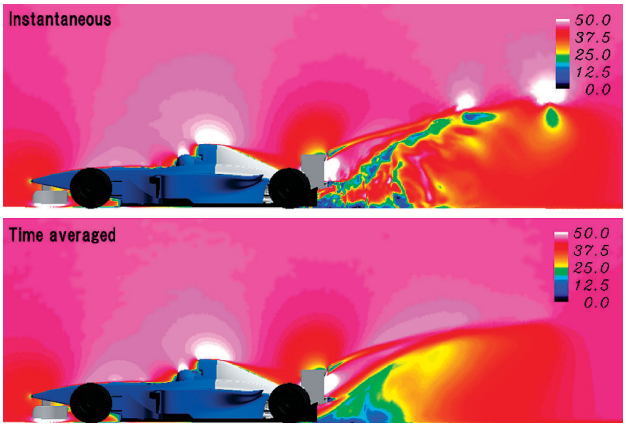


Fig. 2 Instantaneous and time averaged velocity distribution around a formula car obtained with the large-scale LES.

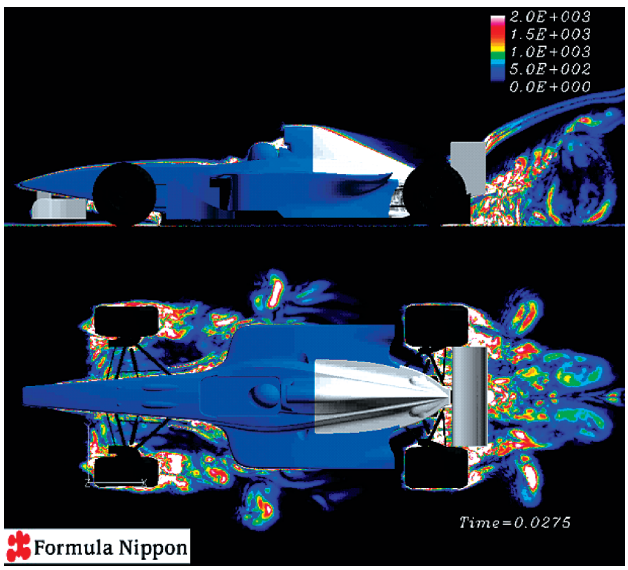


Fig. 3 Instantaneous vorticity distribution around a formula car obtained with the large-scale LES.

2. Protein Simulation

2.1 FMO-MP2 Method

Kitaura's fragment molecular orbital (FMO) method is a promising way of treating the issues of real-scale proteins in an ab initio fashion[2–4)]. In the FMO scheme, calculating the entire protein at once is made unnecessary by dividing it into the fragments of amino acid residues. That is to say, the FMO calculations proceed in a fragment-by-fragment manner. The interactions among fragments are, of course, quantum-mechanically incorporated. The energies and densities of each fragment are summed up finally.

The ABINIT-MP program that was originally developed by Nakano et al. for the parallelized Hartree-Fock (HF) calculations of proteins with FMO method[5, 6)] has been extended in the Frontier Simulation Software for Industrial Science and the Revolutionary Simulation Software project at the

Institute of Industrial Science, the University of Tokyo since FY2002. One of the extensions is the post-HF calculation for introducing the electron correlation effect. These correlated calculations are essential for describing weak interactions of hydrogen bonding (HB) and van der Waals (vdW) interactions both of which are crucial in proteins.

The second order Møller-Plesset perturbation theory (MP2) has an advantage over the density functional theory (DFT) in treating such delicate interactions safely. Thus, we have recently developed a parallelized integral-direct MP2 algorithm and implemented it into ABINIT-MP[7, 8]. In the FMO-MP2 calculations with ABINIT-MP, a flexible parallelization is possible by combining the fragment indices (upper level) and two-electron integral indices (lower level) on distributed computational resources. No communications across worker processes are required during the integral processing for the MP2 correlation energy. The MP2 density can also be evaluated in our implementation.

To vectorize the FMO-MP2 method requires three crucial steps, 1) integral generation, 2) Fock matrix construction, and 3) quarter transformation. The integral generation with the Obara algorithm[11] is vectorized by using a loop expansion technique. The initial integral calculation over s functions for the Obara algorithm— which has a four-term Taylor expansion and asymptotic expansion method that contains one if statement and square-root operation— could be vectorized on the Earth Simulator.

To vectorize the direct Fock matrix construction, we employed a working Fock array algorithm[9] that was originally proposed by Nishiguchi et al.[10] for cell-wise dynamics, and is known as the "particle-pusher" method in the field of plasma simulations. Since the DAXPY and the DDOT operations[12] are used in the most inner loop of each quarter transformation step of MP2 calculation[7, 8], the vectorization of MP2 energy and density calculation is straightforward.

2.2 Benchmark Calculations

We conducted a series of benchmark calculations on the Earth Simulator. Here, we introduce two examples. The first is the FMO-HF calculation of $(Gly)_{160}$ alpha helix conformation at the HF/STO-3G level of theory. The total numbers of atoms and basis functions are 1,123 and 3,687, respectively. We used 128 nodes (1024 CPUs) of the Earth Simulator. The sustained performance was about 1.3 TFLOPS, and that is about 16% of the peak performance.

The second involves a drug design for the vitamin D receptor (VDR) protein, which is a member of the nuclear receptor (NR) superfamily and works as a transcription factor. The agonist of the VDR acts to elicit many activities and this makes it a promising candidate drug for the treatment of a number of diseases including cancer and osteoporosis. The active form of vitamin D, 1α, 25-dihydroxyvitamin D_3, is a typical example of such molecules, and an X-ray structure of its complex with the VDR ligand binding domain (253 residues) is available (PDB ID: 1DB1, resolution 1.8Å)[13, 14]. After hydrogen attachment and reoptimization, the complex system consists of 4,130 atoms and 22,617 basis functions (6-31G basis set). This is the largest size to which the Post-HF calculation has been applied.

The turn-around time for the MP2 energy was only 3.3 hours using 64 nodes (512 CPUs) of the Earth Simulator. A medicinal investigation including comparison with other candidate drugs is in progress. Benchmark calculations have demonstrated the practicality of the FMO-MP2 calculations. Now, we can routinely handle a variety of protein systems with a few hundreds amino acids with FMO-MP2 calculations on the Earth Simulator.

3. Nanoscale Simulation

3.1 As Donor State in Silicon

Although detailed knowledge of shallow donor state is important for the design of future semiconductor nano-devices, no first principles calculation has been made as the donor state expands to a wide area, and requires very large computational resources beyond those of the standard super-computer. Using Earth Simulator, we have succeeded in calculating the As donor state in crystalline silicon. We used a cubic super cell consisting of 5,832 atoms, and substituted an As atom for the center Si atom. The calculation was made using the PHASE code, which is based on the density functional theory (DFT) with its gradient approximation and uses pseudo-potentials and plane wave expansions. With the PHASE code we achieved an effective calculation of 13.6 TFLOPS by using 3,072 CPUs, which attain 57% of the peak performance of Earth Simulator with the same number of CPUs.

The six folded As-donor levels in the present effective-mass approximation were found to split into three levels, an A_1 singlet, a T_2 triplet, and an E doublet. The energy splitting between the A_1 and T_2 levels, and that between the T_2 and E levels were 20 meV, and 1 meV, respectively, and this precisely reproduced the observed values. These results show that first principles calculations are useful for the design of such future nano-devices as those using the fine structure of shallow donors.

3.2 Optimization of High-End Simulation Software to the Earth Simulator

In the course of recent advancements in the capability of high-end computers and super computers, researchers and engineers have been required to exploit skillful simulation technology to accelerate their jobs and produce valuable results within specific time periods. This is the reason these simulations could provide some provisions or insights to experimentalists and theorists beyond the time and space

spent on their problems. With this background, it has been gradually accepted by them as a smart tool for reaching the solutions of their problems. These smart tools, however sometime force them to use elaborate high-performance computing to achieve cost effectiveness for research and development. As is well known, high-end computers have advanced their performance rapidly with increasing complexity in parallel software, algorithms and architectures. However, a wide gap still exists between the peak speed as a theoretical value and sustained speed in practical work. To reduce this gap and achieve cost effectiveness, researchers and engineers have been forced to increase computing speed by tuning their simulation codes for high-end computers by laborious work following their experience and intuition without a reasonable optimization methodology.

In this project, our group is aiming at building up an effective optimizing methodology for the simulation codes that were developed for leading science and technology. Within this framework, we have optimized PHASE, the first principle electron structure code for nano science and technology, to the Earth Simulator by parallelization, vectorization, and modification of parallel algorithms. As a result, PHASE has been parallelized up to 99.98% and vectorized to 99.8% through restructuring the program. For increasing the sustained speed in large-scale simulations, the residual unparalleled subroutines were modified by applying a new parallel algorithm. The speed-up reached approximately 4-7 times that of the original PHASE code.

For validity of the physics, the optimized PHASE was also verified through simulation results by confirming that PHASE enables simulating a typical carbon diamond structure correctly. Results made clear that PHASE is well optimized and the new parallel algorithm is effective for increasing the speed and reducing the time to solution.

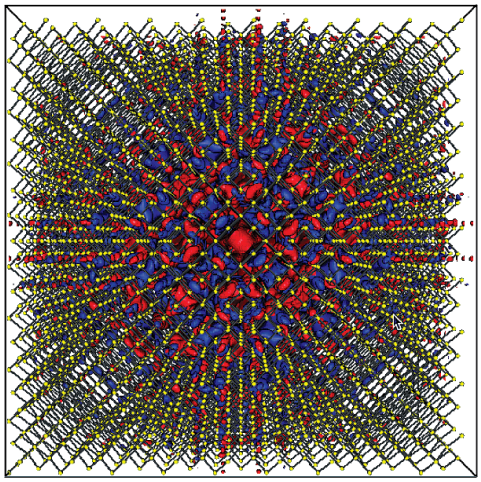
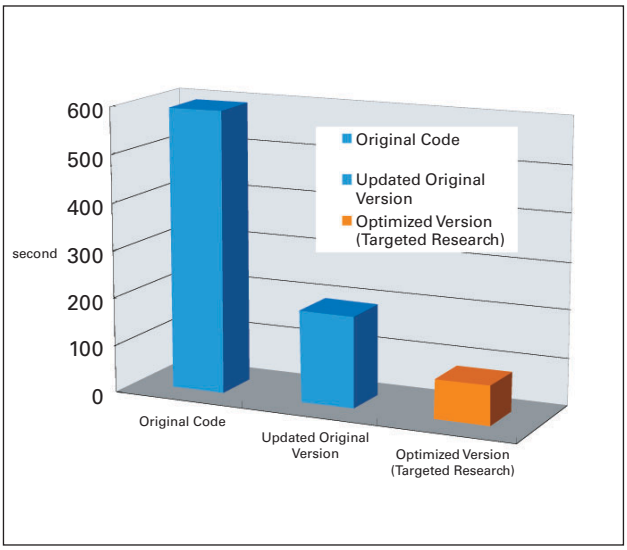


Fig. 4 The wave function of As-donor state in silicon. An As atom is substituted for the center Si atom, and the super-cell contains 5,832 atoms. The blue and red regions show the sign of the wave function.

References

1) M. Tsubokura, K. Kitoh, N. Oshima, H. Zhang, K. Onishi, T. Tominaga, T. Kobayashi, and T. Walker, to be presented at the 31st FISITA World Automotive Congress, 2006.
2) K. Kitaura, T. Sawai, T. Asada, T. Nakano, M. Uebayasi, "Pair interaction molecular orbital method: an approximate computational method for molecular interactions", *Chem. Phys. Lett.* **312**, 319–324 (1999).
3) K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, "Fragment molecular orbital method: an approximate computational method for large molecules", *Chem. Phys. Lett.* **313**, 701–706 (1999).
4) K. Kitaura, S. Sugiki, T. Nakano, Y. Komeiji, M. Uebayasi, "Fragment molecular orbital method: analytical energy gradients", *Chem. Phys. Lett.* **336**, 163–170 (2001).
5) T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, K. Kitaura, "Fragment molecular orbital method: application to polypeptides", *Chem. Phys. Lett.* **318**, 614–618 (2000).
6) T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, K. Kitaura, "Fragment molecular orbital method: use of approximate electrostatic potential", *Chem. Phys. Lett.* **351**, 475–480 (2002).
7) Y. Mochizuki, T. Nakano, S. Koikegami, S. Tanimori, Y. Abe, U. Nagashima, K. Kitaura, "A parallelized integral-direct MP2 method with fragment molecular orbital scheme", *Theor. Chem. Acc.* **112**, 442–452 (2004).
8) Y. Mochizuki, S. Koikegami, T. Nakano, S. Amari, K. Kitaura, "Large scale MP2 calculations with fragment molecular orbital scheme", *Chem. Phys. Lett.* **396**, 473–479 (2004).
9) Y. Mochizuki, M. Matsumura, T. Yokura, Y. Hirahara,



Effect of Performance Improvement

and T. Imamura, "Vectorization of Direct Fock Matrix Construction in DIRAC-DHF Calculations", *J. Nucl. Sci. Tech*. **39**, 195–199 (2002).

10) A. Nishiguchi, S. Orii, and T. Yabe, "Vector calculation of particle code", *J. Comput. Phys*. **61**, 519–522 (1985).

11) S. Obara and A. Saika, "Efficient recursive computation of molecular integrals over Cartesian Gaussian functions", *J. Chem. Phys*. **84**, 3963–3974 (1986).

12) Basic linear algebra subroutines, http://www.netlib.org/blas/

13) N. Rochel, J. M. Wurtz, A. Mitschler, B. Klaholz, and D. Moras, "The Crystal Structure of the Nuclear receptor for Vitamin D bound to Its Natural Ligand", *Mol. Cell* **5**, 173–179 (2000).

14) Protein Data Bank (PDB), http://www.pdb.org/

戦略的基盤ソフトウェアの開発

プロジェクト責任者
加藤　千幸　　東京大学生産技術研究所　計算科学技術連携研究センター

著者
坪倉　　誠*1，望月　祐志*2，山下　勝美*3，村瀬　　匡*3，山岸　賢司*2，常盤　広明*2，中野　達也*4，
山本　武範*5，大野　隆央*6，中村　　壽*7，南　　一生*7

*1　電気通信大学
*2　立教大学
*3　NECソフト(株)
*4　国立医薬品食品衛生研究所
*5　アドバンスソフト(株)
*6　(独) 物質・材料研究機構
*7　(財) 高度情報科学技術研究機構

本プロジェクトでは、工学、生命科学、ナノテクノロジーの分野で核となる流体シミュレーション、タンパク質シミュレーション、ナノシミュレーションを取り上げそれぞれ次の成果を得た。まず、流体シミュレーションでは、産業界の生産現場において、ますますその活用が期待されているデジタルエンジニアリング技術の一つとして、ハイパフォーマンスコンピューティング(HPC)を用いた大規模ラージエディシミュレーション(LES)による流体解析に着目し、その有用性の検討、実証を目的とする。解析対象として産業界で最も過酷な設計開発要件の一つといえるフォーミュラカーに着目し、その非定常空力特性評価を行った。空力評価モデル(ASMO)を用いてHPC/LESの精度の高さを実証し、既存のレイノルズ平均乱流モデルに対する有用性を示すと共に、1億2千万要素を用いた世界最大規模フォーミュラカー(LOLA B03/51)のLES解析を実現した。タンパク質シミュレーションでは、非経験的フラグメント法に基づいた、巨大分子系のための電子状態計算プログラム(ABINIT-MP)を利用した。生体系を構成する分子間の相互作用において重要な水素結合やvan der Waals相互作用を精度よく計算するためには、電子相関を考慮することが必須であり、そのためABINIT-MPには望月により開発された高速な並列化MP2計算エンジンが組み込まれている。本研究では、ABINIT-MPのベクトル化を行い、地球シミュレータ128ノードを用い1.3TFLOPSを達成した。またビタミンD受容体タンパク質(253残基)のFMO-MP2/6-31G計算を64ノードを用い3.3時間で行った。ナノシミュレーションでは、シリコン中の砒素が作るドナー準位の第一原理計算を行った。ドナーの構造はシリコン・デバイスを設計する上で重要な物理量であるにもかかわらず、波動関数の広がりが大きいために、非常に大きなサイズの計算が必要で、これまで第一原理解析を拒んできた。今回、地球シミュレータ384ノードを使用して、5832原子の解析に成功し、ドナーの微細構造、とくに、バレー間相互作用による準位の分裂を明らかにした。これらの結果は将来のナノデバイスの設計に役立つものである。

キーワード：ラージエディ・シミュレーション，空力，自動車, FMO, MP2，シリコン・デバイス，ドナー準位，第一原理計算，
バレー間相互作用