SC09 参加報告

松永康佑*
理化学研究所 次世代計算科学研究開発プログラム 分子スケール研究開発チーム

2009 年 12 月 16 日

1 はじめに

2009 年 11 月 16 日から 19 日までの間にアメリカ合衆国オレゴン州ポートランドで行われた SC09[*1]に理化学研究所のレポータとして参加しました。これはその時のレポートです。様々な分野の発表を聞きましたが、著者の知識不足のために理解できなかったものが多く、著者の専門である分子動力学 (Molecular Dynamics:MD) シミュレーションに偏ったレポートになってしまっていることをどうかご了承ください。

SC は「The internaltional conference for High Performance Computing, Networking, Storage and Analysis」[*2]という正式名称が付いており、主に High Performance Computing (HPC) 分野の最新技術や動向が伺える会議です。IEEE と ACM が主催しています。1988 年の初回から年々参加者は増加しており、この数年は参加者数およそ 1 万人の大規模な会議になっています。参加者をみると、HPC ベンダの技術者・研究者はもちろん、政府系・大学系の役人や研究者まで多岐に渡っています。SC のメインは、HPC ベンダを中心としたブースでの展示会であり、SC の開催とともに各ベンダから新製品が展示・発表され、その模様が各メディアにおいて取り上げられるなど社会的にも注目されています。展示会場を歩いていて、取材風景を何回か見かけました。

各ベンダの展示会と平行して、HPC の最新技術や科学成果に関する講演が催されます。講演の種類には、最新のトピックについての講習会である Tutorials、事前の論文審査を通った Technical Papers、Plenary speakers、博士研究の成果を発表する Doctoral Showcase、各分野の大御所が講演する Masterworks、ゴードンベル賞候補が講演する Gordon Bell Finalists、あるトピックについて皆でディスカッションする Birds-of-a-Feather(BOF) など[*3]があります。Tutorials は展示会に先立った日程で行われるとともに追加料金が必要なので参加しにくいのですが、各トピックについて実践的な事例が学べてお勧めです。参加できなくても Tutorials で用いられた全てのスライドが入った USB メモリスティックが $100 で売られているので、これを購入するのもありかと思います。なお、Technical Papers や Gordon Bell Finalists の論文は全て

図 1 レジストレーション時にもらったグッズ

ACM Digital Library[*4]から入手することができます。

毎年 SC では、推進するテーマ (Thrust areas) があり、今年は「Computing for a Changing World」の標語の下で、17 日を「3D Internet」、18 日を「Bio-Computing」、19 日を「Sustainability」と設定し、各日でそれに関連した講演が行われました (図 2)。特に 19 日の「Sustainability」では「不都合な真実」で有名な元アメリカ合衆国副大統領 Al Gore 氏を Plenary speaker として招いており、HPC の環境・エネルギー分野への比重を上げようとする姿勢が伺えました。実際、会議の運営自体もグリーンであることを心がけているようで、レジストレーション時にもらったグッズの中に、水を飲むためのボトルが入っており、会場のあちこちに設置されたディスペンサーで水を補給し reuse できるようになっていました (図 1)。[*5]。

私が見聞きした限りでは、今回の SC09 での主なニュースは：SC に合わせて発表される Top500 で、ついにオークリッジ国立研究所の Jaguar(ベンダ：クレイ) がロスアラモス国立研究所の Roadrunner(ベンダ:IBM) を追い抜き 1 位になったこと。また中国の National Super Computer Center に新設された Tianhe-1(ベンダ：NUDT)[*6]が 5 位にまで食い込んだこと。IBM の Blue Water に使われる Power7 の展示。Cray の XT6 の発表。NVIDIA の次世代 GPU アーキテクチャ Fermi の発表。(既に別の会議等で発表されていたようですが) D. E. Shaw リサーチの分子動力学シミュレーション用計算機 Anton

* ymatsunaga@riken.jp
*1 http://sc09.supercomputing.org/
*2 数年前から Analysis の語が増えたとのこと
*3 他には、Workshops、Panels、Posters、Exhibitor Forum、Seymour Cray 賞、Sidney Fernbach 賞、Ken Kennedy 賞受賞講演がある
*4 http://portal.acm.org/dl.cfm
*5 ただし、ディスペンサーの横に紙コップも用意してあり、多くの人はそちらで飲んでいたようです
*6 NUDT とは国立国防技術大学のこと。Tianhe-1 は Intel Xeon と ATI Radeon のハイブリッド構成で Infiniband で接続されている。

図 2　17 日の Opening address で紹介された SC09 における 3 つの Thrust areas

図 3　長崎大・浜田剛助教の Gordon Bell Finalist 講演

が、BPTI で 1 ミリ秒のシミュレーションを達成したこと (Gordon Bell 賞、Best Paper 賞受賞)。などが挙げられると思います。

日本勢からは：国立情報学研究所・三浦謙一教授の Seymour Cray 賞受賞。長崎大・浜田剛助教の Gordon Bell 賞受賞 (図 3)。富士通の次世代スパコン筐体 (プロトタイプ) の展示と発表 (図 4)。また東工大が TSUBAME 2.0 で NVIDIA Fermi を搭載し 2010 年 10 月に 3PFlops になるとのニュース。などがありました。

SC09 に参加して全体的に感じたことでは、やはり GPGPU コンピューティング関連の講演は人が多く、注目されていると思いました。また、個人的には D. E. Shaw リサーチの Anton の結果に圧倒されました。今回私が聴講した講演は表 1 のとおりです。以下では、私の能力と時間の都合から、聴講した中からごく一部を報告させていただきます。後日、時間が許せば他の講演についても追加させていただきます。時系列順で、「です・ます」調ではなく「である」調で書いていきます。URL へのリンクは脚注に、論文などへの参照は、末尾の参考文献に載せます。皆様に SC09 の雰囲気が少しでもお伝えできれば幸いです。

日時	タイトル	種類
11/16 8:30	Hybrid MPI and OpenMP Parallel Programming	Tutorials
11/16 13:30	Python for High Performance and Scientific Computing	Tutorials
11/17 8:30	(Opening Address) The Rise of the 3D Internet: Advancements in Collaborative and Immersive Sciences	Invited Speakers
11/17 10:30	Terascale Data Organization for Discovering Multivariate Climatic Trends	Papers
11/17 11:30	Auto-Tuning 3-D FFT Library for CUDA GPUs	Papers
11/17 13:00	GPU computing and TSUBAME (正確なタイトルを失念。。)	NVIDIA Booth
11/17 13:30	Big Data and Biology: The Implications of Petascale Science	Masterworks
11/17 14:15	The Supercomputing Challenge to Decode the Evolution and Diversity of Our Genomes	Masterworks
11/17 15:30	Beyond Homogeneous Decomposition: Scaling Long-Range Forces on Massively Parallel Architectures	Gordon Bell Finalists
11/17 16:00	NVIDIA GPU Acceleration of Molecular Dynamics Simulations in AMBER	NVIDIA Booth
11/17 17:00	Fermi 説明会	NVIDIA Booth
11/17 17:30	Art of Performance Tuning for CUDA and Manycore Architectures	BOF
11/18 8:30	Plenary Speaker: Systems Medicine, Transformational Technologies and the Emergence of Predictive, Personalized, Preventive and Participatory (P4) Medicine	Invited Speakers
11/18 10:30	Cray Award Recipient: Kenichi Miura	Invited Speakers
11/18 11:15	Fernbach Award Recipients: Roberto Car and Michele Parrinello	Invited Speakers
11/18 13:30	A 32x32x32, Spatially Distributed 3D FFT in Four Microseconds on Anton	Papers
11/18 14:30	42 TFlops Hierarchical N-body Simulations on GPUs with Applications in both Astrophysics and Turbulence	Gordon Bell Finalists
11/18 16:00	The Cat is Out of the Bag: Cortical Simulations with 10^9 Neurons, 10^{13} Synapses	Gordon Bell Finalists
11/18 16:30	Millisecond-Scale Molecular Dynamics Simulations on Anton	Gordon Bell Finalists
11/18 17:30	Python for High Performance and Scientific Computing	BOF
11/19 8:30	Keynote Address: Building Solutions: Energy, Climate and Computing for a Changing World	Keynote
11/19 10:30	Toward Climate Modeling in the ExaFlop Era	Masterworks
11/19 11:00	Managing HPC Clouds	Exhibitor Forum
11/19 11:30	Common Application Benchmarks on Current Hardware Platforms	Exhibitor Forum
11/19 12:15	HDF5: State of the Union	BOF
11/19 13:30	Tesla: Fastest Processor Adoption in HPC History	Exhibitor Forum
11/19 14:00	Fujitsu's Technologies for Sustained Petascale Computing	Exhibitor Forum
11/19 15:45	GPGPU and Cloud Computing for DNA Sequence Analysis	Doctoral Research Showcase
11/19 16:15	Simulation and Animation of Complex Flows on 10,000 Processor Cores	Masterworks

表 1　聴講した講演のテーブル。

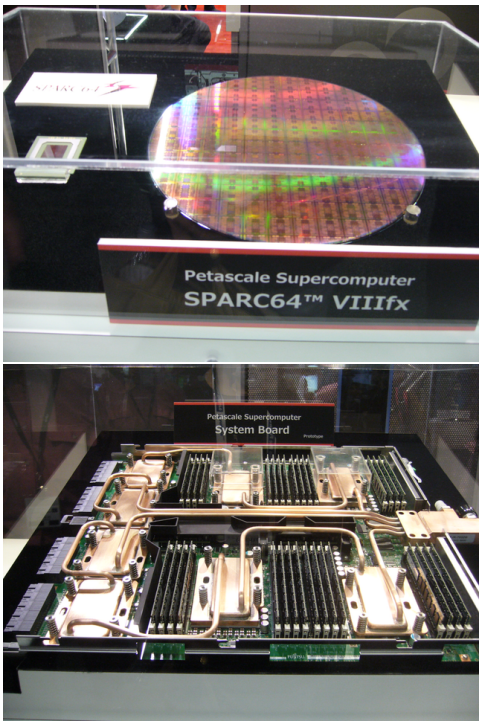


図 4　富士通ブースでの SPARC64 VIIIfx とボード (筐体の写真を撮るのを失念。。)

2　2009 年 11 月 16 日 Tutorials

朝 7 時 30 分にポートランド国際空港に着いたその足で SC09 の会場であるオレゴンコンベンションセンターに向かった。成田空港を 16 時に出ておよそ 9 時間のフライト中ほとんど眠れなかったので、この日は睡眠不足と慣れない英語とで正直しんどかった。会場に着いてすぐにベンダの展示会場を覗いてみると、明日からの展示のための設営が行われていた。大きな研究所のブース設営ではクレーン車まで導入していた (図 5)。展示会場では、夜にテクニカルプログラム参加者向けの Opening Gala があったが、あまりに眠かったので、おいとましてホテルに向かった。

2.1　Python for High Performance Tutorials and Scientific Computing

[*7] Python は、インタプリタ言語であるにもかかわらず NumPy[*8]や SciPy[*9]といった数値計算用パッケージが充実しており、科学技術計算にも多く応用されている。MD 計算においては特に PyMOL という分子可視化ソフトが有名である[*10]。この Tutorial では、Python を用いて HPC を行うコツが実例とともに説明された。まず Python の実装系がいくつか紹介

[*7] スライドは以下から入手可能。～ に注意。
http://www.mcs.anl.gov/~wscullin/python/tut/sc09/Site/Slides.html
[*8] http://numpy.scipy.org/
[*9] http://www.scipy.org/
[*10] 残念なことに、最近開発者である Warren DeLano 氏が若くして亡くなったが、開発は引き継がれるようだ

図 5　ブース設営の様子@NASA ブース。

された。普通、Python の実装系と言えば C 言語で書かれた CPython だが、他に Java で書かれた Jython、.NET 環境で動く IronPython、Python 自身で書かれた PyPy などがあることを知った。 中でも、Google で開発されている Unladen Swallow[*11]は CPython に比べて 5 倍の速度向上を目標に作られているとのことで HPC 用途に使えそうだ。その後は、Python の言語仕様から NumPy、Scipy の使い方が説明された。よく言われるように、Python のリストはメモリ上で連続しておらず遅いから、必ず NumPy の array を使うようにということや、Python は全て参照渡しなので、値渡しをしたい時には、copyrow=arow.copy() とすること、などがあった。

後半では、並列処理のためのモジュールや、C による拡張についての説明があった。並列処理では、共有メモリ環境下で threading、subprocess、multiprocessing モジュールを用いて、また分散メモリ環境下で mpi4py モジュールを用いて π 計算の例が示された。threading モジュールでは、GIL 問題[*12]があり、マルチコア内のロードバランスが悪くなり性能を発揮できないこと、 また subprocess モジュールは GIL 問題を回避しているが、低レベルの命令なので使いにくいこと。結果として、今後はこれらを解決した multiprocessing モジュールを使うのが良さそうに思われた (python3.0 から標準モジュールとして提供される)。C による拡張では、KS-DFT 計算用アプリである GPAW[*13]の開発例が紹介された。GPAW では、コードのほとんどが Python で書かれており、C 言語は 10% のみである。Python は単純に C のラッパーとして用いられているのではなく、高度なアルゴリズムを要する部分は Python で担当し、実行時間の 90% を占める重い数値計算部分は C が担当するということをしている。NumPy の np.dot などは BLAS

[*11] http://code.google.com/p/unladen-swallow/
[*12] GIL（Global Interpreter Lock）とは Python 上で 1 つのスレッドだけが動作するよう強制する仕組み。このロックによって，同時に同じメモリにアクセスする二つのスレッドが存在しないと保証するのでプログラムを単純な構造にできる反面、マルチコア内のロードバランスが低下する問題がある。
[*13] https://wiki.fysik.dtu.dk/gpaw/index.html

図 6　宿泊したホテルの中庭。東京よりも一足早く紅葉が見ごろだった。

を使っているのだが、NumPy の BG/P への移植は難しいらしくて、BLA(C)S や (Sca)LAPACK へのインタフェースを自前で書く例が参考になった。Python-C 混合下でのプロファイリングには Tau[*14]が使えるとのこと。最後に蛇足的に、グラフ描画パッケージである matplotlib[*15]と web フレームワークである web2py[*16]を用いた web アプリケーション作成例の紹介があった。

3　2009 年 11 月 17 日 3D-internet

前日寝不足で頑張ったおかげで非常によく眠れ、時差がほぼ解消された。この時期のポートランドは雨が多いのか天気予報を見ると雨や曇りが多かった。この日も曇りで昼食時ににわか雨。

3.1　(Opening Address) The Rise of the 3D Internet: Advancements in Collaborative and Immersive Sciences

Intel の Chief Technology Officer(CTO) である Justin Rattner 氏のオープニング講演である。講演では、まずアメリカでの HPC 分野の資金の Growth Curve を見せて、伸びが小さいことを指摘した。それもそのはずで、アメリカでも HPC は主に政府からの予算で行われているからだと思われる。そこで氏は、これから HPC 分野の資金を増やすには、HPC での Killer アプリケーションが必要だと主張した (図 7)。その Killer アプリとして氏が考えて入るのが、HPC で行う 3D-Internet 環境であり、その具体的な例として科学のコラボレーション環境 (OpenSimulator[*17]と ScienceSim[*18]) や、アパレル業界での (度重なる) 服の試作を HPC で仮想的に行うことなどが提案された。HPC で実現される 3D のデモとして、ビロードの布が上から机に落ちるときの、皺の様子 (布の間の流体力学的相互作用や机との摩擦まで計算してるらしい) や、水が流れる様子を音まで含めてシミュレーションした結果のムービーを見せた。これらは HPC を導入することで、Internet 環境がよりリッチになることを説得するには十分であったが、私には果てしてコラボレーションツールが Killer アプリに成り得るかはわからない。もっとエンターテイメント性と絡めたほうが流行る気もする。Killer アプリが必要という主張は素晴らしいのだが。

最後に Intel の新 CPU Nehalem や新メニーコア CPU(GPU?)[*19]Larrabee の紹介があった。私の不見識で Larabee を知らなかったのだが、話を聞くと、Larabee とは GPU のストリーミングプロセッサ部分を Intel 製の CPU で置き換えたようなすごいものらしい。パフォーマンスとして Larabee をオーバークロックして (ネジを回して？)SGEMM で 1TFlops に達した様子を見せていた[*20]。

3.2　GPU computing and TSUBAME

NVIDIA ブースでの東工大の松岡聡教授による GPGPU と TSUBAME に関する講演だった。まず冒頭に、CUDA プログラミングについて初の日本語の書籍[*21] が発売されたことがアナウンスされた。これは日本に帰ってから早速購入した。講演では、GPGPU コンピューティングの特徴が説明され、東工大額田氏による GPU 上の FFT ライブラリ (Nukada-FFT) や、GPU を搭載した TSUBAME1.2 上のタンパク質ドッキング計算 (東工大秋山教授による) の紹介があった。ドッキング計算ではエネルギー評価の畳み込み積分に Nukada-FFT を使っていると予想するが、このように計算機科学と分子科学が協力して問題に取り組む例は素晴らしいと思う。その後、TSUBAME のアップデートの歴史が紹介され、如何に工夫して効率的に Top500 の上位に長期間残っているかが説明された。最後に TSUBAME1.2 の詳細と姫野ベンチマーク結果などがあった。GPU を積んだ複数ノード使用でピークが約 260GFlops、CPU のみだと約 30GFlops だった。

3.3　Big Data and Biology: The Implications of Petascale Science

[*22] Amazon で EC2 や S3 の開発を担当している Deepak Singh 氏による、大容量のデータを処理するためにクラウドをどのように準備すればよいかの講演だった。たかだか 30 分足らずの講演のうちに、スライドを 180 枚も用いてものすごいペースで進めて行ったのに驚いてしまった。講演の内容は、まず、任意に増大する科学データに対する Scalable で Fault Tolerance なシステムの重要性を強調し、それらを満足するものとして Amazon EC2 と S3 を紹介した。その後、科学計算のプラットフォームとして既に EC2 上で Mathematica と MATLAB が動いていることを述べ、EC2 上で行われているデータドリブンな科学応用例として、ボランティア参加型の銀河分類プロジェクトである Galaxy Zoo[*23]、Pfizer のドラッグ

[*14] http://www.cs.uoregon.edu/research/tau/home.php
[*15] http://matplotlib.sourceforge.net/
[*16] http://www.web2py.com/
[*17] http://opensimulator.org/wiki/Main_Page
[*18] http://sciencesim.com/wiki/doku.php

[*19] SC09 後の 2009 年 12 月時点のニュースによると、GPU として供給する計画は仕切り直しになったらしい
[*20] 動画が以下にある。
http://www.youtube.com/watch?v=ynjYuS1J3jI
[*21] http://www.kohgakusha.co.jp/books/detail/978-4-7775-1477-9
[*22] スライドは以下から入手可能。
http://www.slideshare.net/mndoci/talk-given-at-cloud-computing-for-systems-biology-workshop
[*23] http://galaxyzoo.org/

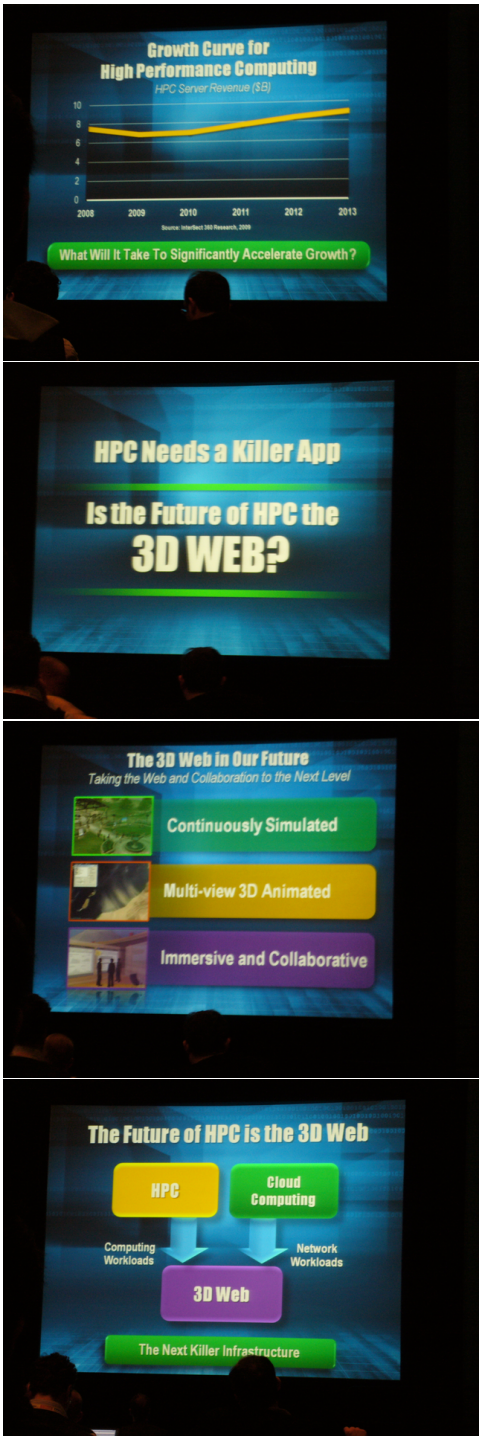


図 7　Intel の CTO である Justin Rattner 氏のオープニング講演でのスライド。

ドッキング計算などが紹介された。

クラウドでの並列処理の話になるといつも出てくるのが、Google の mapreduce[*24] フレームワークによる分散処理環境で、ここでもそのオープンソース版である Hadoop[*25]が紹介されていた。データ処理演算の中でも並列化方法が明らかなものは EC2 上の Hadoop ですぐに使えそうだった。最近では、clustering アルゴリズムも mapreduce 上で実装されており、近い将来に MD トラジェクトリ解析にも利用できるだろう。ただし、自己相関などの時間依存の統計量解析には去年の SC08 で D. E. Shaw リサーチが発表していた HiMach[*26]のように、mapreduce の key-value 部分を独自に拡張する必要があるだろう。

3.4　NVIDIA GPU Acceleration of Molecular Dynamics Simulations in AMBER

[*27] San Diego Supercomputer Center の R. C. Walker 氏による Amber [*28]を GPU に対応させ高速化したという内容の講演である[*29]。実装の内容については触れず、主にベンチマーク結果の話だった。現在のところ Amber10 向けのパッチが公開されており、Amber11 で正式に GPU をサポートするとのこと。ベンチマーク結果では、TRPCage の計算においてシングル GPU で 8CPU よりもおよそ 2 倍速くなっていた。また、Myoglobin では 7 倍ほど、Nucleosome では、およそ 8.5 倍速くなっていた (分子サイズが大きいほど GPU のほうが速くなる)。他に興味深かったのが、GPU なので基本的に単精度で計算しているのだが、ダイナミックレンジが必要とされる部分では倍精度で計算しているそうで、Amber の倍精度での計算結果とエネルギーを比較してよく一致していた[*30]。

現在はシングル GPU で implicit 溶媒である Generalized Born 計算のみをサポートしており、とりあえずユーザにとって小規模な PC クラスターの代わりにシングル GPU を使うという選択肢を与えるものだと説明していた。将来的に PME(Particle Mesh Ewald) 使った explicit 溶媒計算と、multiple GPU をサポートする予定だそうだ (図 8)。

3.5　Fermi 説明会

Fermi の説明会が NVIDIA ブースであるというので、急遽予定を変更して出てみた。が、私が着いた時には既に多くの人が集まっており、遠巻きにしか聞けなかった (図 9)。結果、レポートするだけの情報が得られなかった[*31]。申し訳ありません。。GPGPU に対する HPC 業界の期待がよく分かった一幕

[*24] http://labs.google.com/papers/mapreduce.html
[*25] http://hadoop.apache.org/
[*26] http://www.deshawresearch.com/resources_himach.html
[*27] Amber on GPU に関する詳しい情報は以下にある
http://ambermd.org/gpus/
[*28] Amber とは MD 計算用アプリのスタンダードのひとつである。他に Charmm、NAMD、Gromacs などがある。
[*29] 不見識で知らなかったのだが、知り合いに聞いた話では、Walker 氏よりも先に日本人の方が Amber の GPU 対応を行っていたらしい。また最近では、Charmm の GPU 対応も日本人の方が行っているとのこと。(両者ともお名前を失念してしまいました。。すいません)
[*30] 尤も、短いタイムステップ (500 ステップ程) での比較だったので、長時間の挙動は分からない
[*31] MD 計算に重要な、倍精度・ECC はサポートされるようです

図 8　(Walker 氏のスライドから)Amber の今後の GPU への対応予定

図 9　NVIDIA ブースでの Fermi 説明会に集まる人々

だった。

4　2009 年 11 月 18 日 Bio-computing

朝起きると珍しく晴れていたので、川沿いをジョギングした。ホテル (University Place) から川沿いに 15 分程走ると対岸に会場であるコンベンションセンターが見える (図 10)。他にもジョギングしている人が多くいた。

4.1　Cray Award Recipient: Kenichi Miura

国立情報学研究所・三浦謙一教授の Seymour Cray 賞受賞講演だった (図 11)。受賞理由には「For leadership in developing groundbreaking vector supercomputing hardware and software」とある。三浦教授の講演は、ILLIAC IV の並列計算 (モンテカルロ法) に関する研究時代から始まり、富士通の VP シリーズ時代、NAREGI プロジェクトから最新の次世代スパコンプロジェクトまで説明された。全体として三浦教授の人柄が伝わってくる素晴らしい講演だった。富士通でのベクトルマシンの開発においては、ILLIAC IV 時代の経験がとても役に立ったことを強調されていた。印象的だったのは、ILLIAC IV 時代の Radiation Transport モンテカルロコード並列化に関するスライドの中で、ソースコードから並列化するのではなく、背後にある物理に戻って並列性を考えることが重要で、この手の問題解決は大人には難しいが子供なら簡単に出来る、と指摘していたことだった。

図 10　川沿いの風景。(下) 対岸に見えるツノが 2 本生えた建物が SC09 の会場であるオレゴンコンベンションセンター

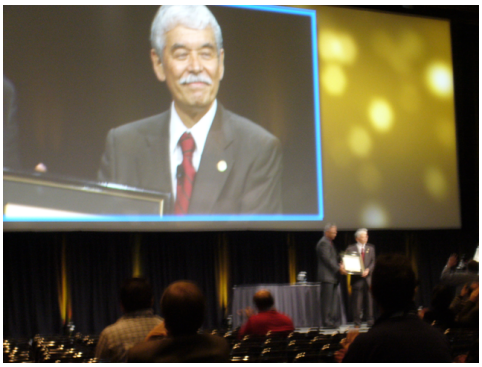


図 11　三浦謙一教授の Seymour Cray 賞受賞式

4.2 Fernbach Award Recipients: Roberto Car and Michele Parrinello

Car-Parinello MD で有名な、Roberto Car 教授と Michele Parrinello 教授それぞれによる Sidney Fernbach 賞受賞講演だった (図 12 上)。受賞理由には、「For leadership in creating the modern theoretical and practical foundations for modeling the chemistry and physics of materials. The software resulting from this work is one of the enabling tools for materials science modeling」とある。講演では、お二方とも最近は Car-Parinello MD の経路積分版に取り組んでいるようで、その最新結果が紹介されていた。また Parrinello 教授のほうは、氏が最近推し進めている Metadynamics についても触れていた。また、Cray 賞の三浦教授と同じく、背後にある物理の並列性を生かす (Exploiting Intrinsic Parallelism) ことが重要だと指摘していたのが印象的だ。面白かったのが、Parrinello 教授のスライドの中に、分子のミリ秒など長時間の挙動を普通の MD シミュレーションで捉えることは無理である、と書いたスライドがあった (図 12 下)。それに対して、講演の最後で (D.E. Shaw リサーチ所属だと思われる人が)Anton ならばミリ秒のシミュレーションができるようになってきたが、それでも長時間挙動を捉えることは無理だろうか、という質問があった。その回答は上手く聞き取れなかったのだが、、それでも無理だという主張をされていた。私にはその主張の真意を推し量ることは出来ないが、Parrinello 教授は分子科学がパワープレイ一辺倒へなることへの危惧を感じていたのではないかと思う。

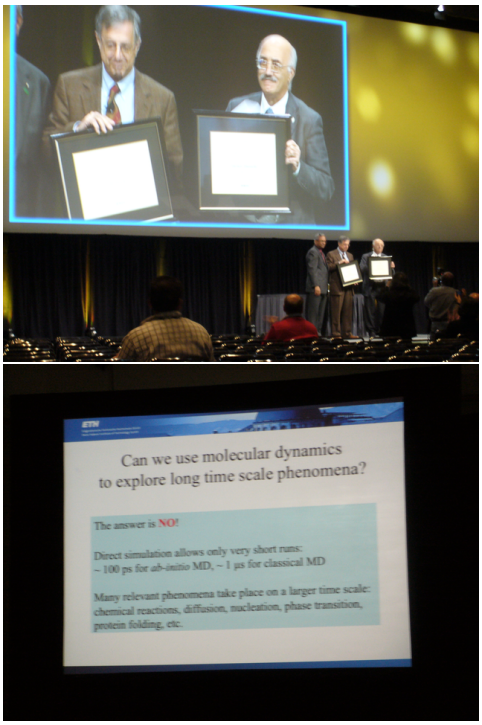


図 12　(上)Car 教授、Parrinello 教授の Sidney Fernbach 受賞式。(下)Parrinello 教授の講演のスライド

4.3 Millisecond-Scale Molecular Dynamics Simulations on Anton

*32 分子動力学用計算機*33 Anton についての D. E. Shaw 氏による講演である。Anton については、2007 年の設計段階 [9] から話題に上がっていた。実機が完成したのが 2008 年 10 月で、2009 年に入ってその成果が発表されはじめた。その内容は、例えば京大の高田彰二准教授により「Anton ショック」として報告されている*34。Anton に関する簡単な読み物としては、英語では [1] がある。日本語では、私の知る限り高田准教授の報告しか見当たらないので、この場を借りて可能な限り詳しく書くことにする。

まず、Shaw 氏の経歴を紹介すると、元々は IT(並列計算) の専門家で、それを金融に応用することをしていた。1986 年にモルガンスタンレーに入り、2 年後に自身のヘッジファンド D. E. Shaw*35を立ち上げる。並列計算により定量的に市場を分析し裁定機会を見つけ出す独自アルゴリズムを用いることで、D. E. Shaw ファンドは高いパフォーマンスを上げ続け、現在では 1,000 人規模の従業員を抱える会社に成長している。ファンド経営中にも科学に精通していたらしく、1994 年からクリントン政権の President's Council of Advisors on Science and Technology という職を 7 年努めたという異色の経歴の持ち主である。2001 年頃、姉妹が病に冒されたことがきっかけで、MD 計算に興味を持つようになり、2002 年頃に MD 計算で医療へ貢献することを目的とした私的な研究所、D. E. Shaw リサーチを立ち上げた。現在ではおよそ 60 人ほどの研究者が雇われているらしく、計算機科学者から MD 専門家まで揃っている [1]。 現在 Shaw 氏は、自身の D. E. Shaw リサーチとコロンビア大学 Center for Computational Biology and Bioinformatics の職を兼任している。

Anton とは、その D. E. Shaw リサーチで開発された分子動力学用計算機である。MD-GRAPE*36 のようなホストコンピュータに取り付ける加速器型とは異なり、Anton 単体で全ての MD 計算 (力の計算だけでなく、積分なども) を行う。*37ボトルネックとなる部分の計算をパイプライン処理で計算するという点では MD-GRAPE と似てはいるが、Anton 内のノード間ネットワーク部分も特注するなど徹底している。Anton の名前は、" 顕微鏡の父" と言われるオランダの科学者 Anton van Leeuwenhoek に由来しており、Anton が"computational microscope" となる期待を込めて付けられた。Anton は、典型的には 512 ノード (8 × 8 × 8 の 3D-torus トポロジー) から成っており [9, 10]、各ノードに

*32 論文は以下から入手可能
http://scyourway.supercomputing.org/conference/view/gb116
*33 英語では、Special-Purpose Machine for Molecular Dynamics Simulation
*34 http://kuchem.kyoto-u.ac.jp/hikari/yuragi/newsletter/F&F_NL9.pdf
*35 http://www.deshaw.com/
*36 http://mdgrape.gsc.riken.jp/
*37 厳密に言えば、Anton もファイル I/O のためにホストコンピュータと接続している

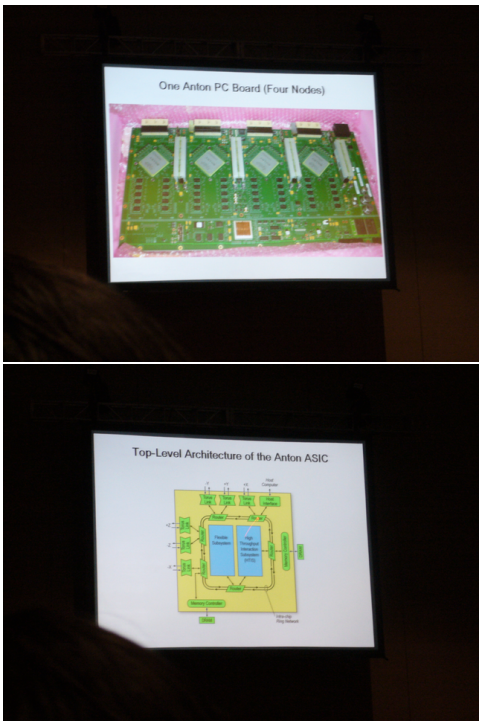


図 13　(Shaw 氏の講演スライドから)Anton の (上) ボードと (下)ASIC

ASIC(Aplication Specific Integrated Circuit)[*38]がひとつある (図 13)。ASIC は、特定の計算用パイプラインが詰まった HTIS(high-throughput interaction subsystem)、プログラマブルな Flexible Subsystem の 2 つのサブシステムから成っており、リング状のネットワークと他のノードへの $\pm x, \pm y, \pm z$ 方向へのリンク、ホストコンピュータへのインタフェース、DRAM が付いている。各サブシステム内には、SRAM と L1 キャッシュが搭載されている。MD 計算に使われるプログラムや原子の座標・速度データ保持は全てこの内部の SRAM・L1 キャッシュ上で賄っており、小・中規模のタンパク質の MD 計算中に DRAM へアクセスすることはほとんどないと思われる [9][*39](図 14)。

著者が理解する限り、Anton の速さは、MD 計算のアルゴリズムとハードウェアが協調的にデザイン (co-design) されていることから来る。以下では、Anton で用いられている MD 計算のアルゴリズムとハードウェアについて説明する。MD 計算では以下のようなポテンシャルエネルギーから相互作用が計算される。

$$V = \sum_{bonds} k_b(r_{ij} - r) + \sum_{angles} k_a(\theta_{ijk} - \theta_0) + \sum_{torsions} A_t[1 + \cos(n\tau - \psi_{ijkl})]$$

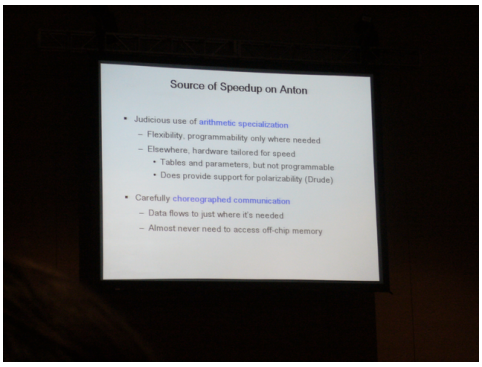


図 14　(Shaw 氏の講演スライドから)MD 中に off-chip メモリへアクセスすることはほとんどないと書かれている

$$+\sum_i \sum_{j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}}. \qquad (1)$$

右辺の第 5 項は静電相互作用であり、粒子の数を N とすれば $O(N^2)$ 個の相互作用からなる重い計算である。通常、この項は Ewald 法によって解かれる。これは仮想的な電荷分布を導入することにより、第 5 項を力の届く有効距離の短い項 (Direct Sum 項) と、有効距離が長く力が滑らかに変化する項 (Reciprocal Sum 項) にわけることで、前者をカットオフ近似で求め、後者を 3D-FFT により逆空間で解く方法である。右辺第 4 項は vdW 項と呼ばれ、有効距離が短いので、これも Direct Sum 項と同じくカットオフ近似して求める。

並列計算によりこの vdW 項と Direct Sum 項 (以下、まとめて力 (ちから) と呼ぶ) を求めるには、まず、原子の座標保持と力の計算を行う作業をプロセッサに割り振る必要がある。これにはおよそ 3 つの方法がある。一つは、粒子分割 (atomic decomposition) という方法で、これは N 個の原子を平等に各プロセッサに割り振って計算する方法である。これは、プロセッサの数を P とすると、プロセッサ当たりの原子数が N/P であり、ロードバランス的には良いのだが、ひとつの力 F_i を計算するのに $O(N)$ の座標通信が必要となり、バンド幅的には良くない。二つ目は、領域分割 (domain decomposition) という方法で、これは原子が現在居る領域にあわせてプロセッサ上に割り振る方法である。この方法だと、カットオフ近似の利点が生かすことが出来て、ある領域を担当するプロセッサはカットオフ長 R 内の領域を担当するプロセッサのみと通信すれば良い (通信料は $O(R^3)$)。しかしながら、プロセッサ数 P をいくら増やしても通信量が減る (スケールする) わけではないところに不満がある[*40]。最後は、力分割 (force decomposition) という方法である [5, 6]。これは、原子 i に働く力が $F_i = \sum F_{ij}$ と j 方向に分割して書けること、つまり i, j というペアの視点でみれば力の計算が分割できることに着目したものである。この方法では、まず原子の座標の保持については、粒子分割と同様にして平等にプロセッサに割り振られる (プロセッサ当

[*38] 特定用途向けの集積回路の総称
[*39] Anton の高速化に最も寄与しているのはこの事実かもしれない
[*40] 厳密には、P が増えるとともに、プロセッサが担当する領域が小さくなり、カットオフ長 R 内の領域が球状に近づくので通信量は少し減る

たり N/P 個を保持)。一方で、力の計算については、F_{ij} を行列とみなし、行列の各ブロックの計算を各プロセッサに割り振る[*41]。ここで、各ブロックのサイズは $N/\sqrt{P} \times N/\sqrt{P}$ となる。この方法では、ブロック内の F_{ij} を計算するために、ブロック間を i 方向について通信して座標 r_i を得て (通信量は $N/\sqrt{P}$)、その後ブロック間を j 方向について座標 r_j を得る (通信量は $N/\sqrt{P}$)。こうして F_{ij} を計算し、保持している N/P 個の原子について力 F_i を計算するために、ブロック内で j 方向について F_{ij} の和をとった後に、ブロック間で j 方向について和をとる (通信料は $N/P \times \sqrt{P} = N/\sqrt{P}$)。結局、通信量は $O(N/\sqrt{P})$ になり、プロセッサ数 P とともに通信料が減らせる (スケールする) ことが特徴である。しかしながら、一方で領域分割のようにカットオフ近似の利点を生かしていない所にまだ問題がある。

Anton で用いられているのは、NT 法 (Neutral Territory 法) と呼ばれる新しい方法であり、M. Snir 氏 [11] によって 2004 年に、また独立に D. E. Shaw 氏により 2005 年 [8] に提案された[*42]。NT 法とは簡単に述べれば、領域分割法と力分割法のハイブリッド版であり、両者の利点を生かして通信量は $O(R^{3/2}/\sqrt{P})$ となる。まず原子の座標の保持については、領域分割と同様にして原子が居る領域にあわせてプロセッサ上に割り振られる。力の計算については、力分割と同じようにペアで分割する。この際に、カットオフ近似の利点を生かせるよう、近い距離にいるペア同士で分割するのが理想であるが、NT 法では、カットオフ長 R 内に居る可能性のある領域の原子ペア i, j の F_{ij} 計算を、原子 i の x 座標と y 座標、かつ原子 j の z 座標に重なる領域の原子座標保持を担当しているプロセッサに割り振る。このように割り振ることで、発生する通信量を幾何学的に表現したものを図 15 に示す。プロセッサ数 P が増えるとともに、領域分割法比べて通信量が少なくなっていくのがわかる。Anton の HTIS は、この NT 法に特化しており、HTIS 内に 32 個ある PPIPs(pairwise point interaction piplines) が力の計算を行う (図 16)。[3]、図 17)。NT 法の弱点として、選んだペアがカットオフ長 R 内にいないことが頻繁にあるという点があるが、このために各 PPIP に距離をチェックする専用の 8 個の match unit というのが付いている。また、32 個の PPIPs のバンド幅は 10Tbit/s を越えるらしく、そのバンド幅を存分に利用して PPIPs を busy にさせるために、PPIPs をアレイ上に繋げて、systolic-array[*43] ライクなアーキテクチャを作り、原子 i, j ペアの座標を 2 次元的に入力させて並列処理を行っているとのことである (direct product selection resuction operations:DPSRs と名づけている [3]、図 17)。

Reciprocal Sum 項の計算では、逆にハードウェアにあわせてアルゴリズムが選ばれている。普通、Reciprocal Sum 項の計算では 3D-FFT を行う前にまず原子の電荷をグリッド点上の点に置き換える操作が必要となる (charge spreading)。また一方で、逆空間で (Poisson 方程式を) 解いて実空間に戻した後で、グリッド点上の力を、近くの原子へ置き換える操作が必要である (force interpolation)。両者はどちらも、グリッド点とその近くの原子の” 相互作用” と見なすことができ、SPME(Smooth Particle Mesh Ewald) では B-spline 関数を使って計算される。これに対して、B-spline の代わりに Gaussian を用いる kGSE(k-space Gaussian SplitEwald)[7] という手法を独自に開発することで、” 相互作用” の計算をほぼハードの変更なしに、HTIS を再利用して行うということを

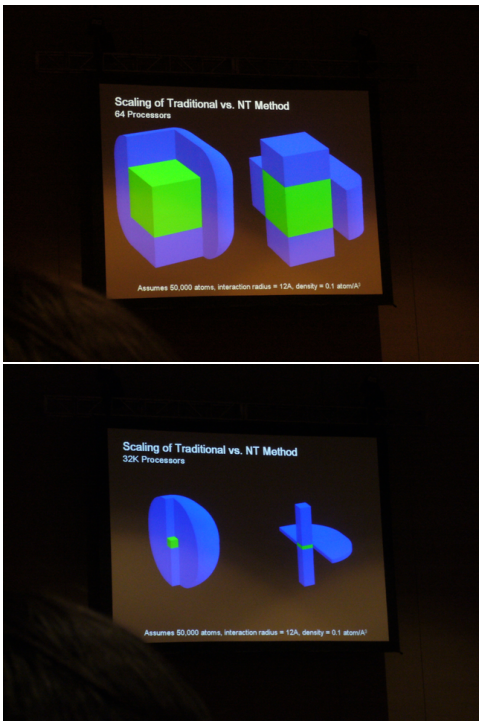


図 15　(Shaw 氏の講演スライドから) 左が領域分割法で、右が NT 法。青色が通信される領域を示す。(上)64 プロセッサ使用の場合。(下)32K プロセッサ使用の場合。” 半球” 状の形になるのは、力の作用・反作用を考慮してである。

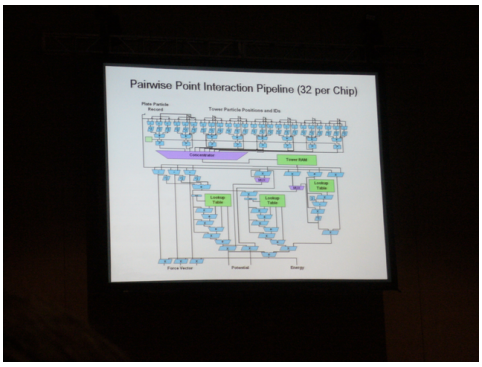


図 16　(Shaw 氏の講演スライドから)PPIPs の説明

[*41] 細かいことを言うと、この際サブブロックの i 方向、j 方向のプロセッサに、力の計算に必要とする座標が保持されているように割り振るのが重要である

[*42] それ以前に、NAMD2 開発グループによって似た手法が提案されていたという話もある

[*43] http://en.wikipedia.org/wiki/Systolic_array

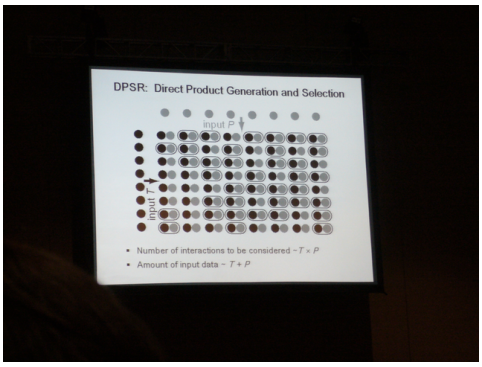
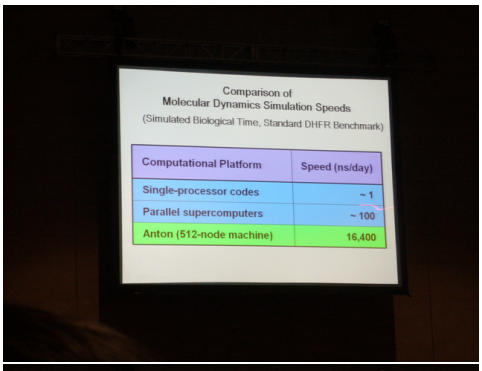


図 17　(Shaw 氏の講演スライドから)DPSRs の説明

実現している。一方、3D-FFT の計算は Flexible Subsystem で行われる。Anton 上の 3D-FFT については、SC09 の別の講演で詳しく説明されていた [12]。あいにく私には難しすぎて理解できなかったのだが、Anton は特注しているネットワークが高速らしく[*44]、3D-FFT 計算ではそのネットワークの高速性を利用した独自アルゴリズム (通常のスパコンでは避けられるワード単位の細かい通信を大量に用いるアルゴリズム) を開発することにより、$32 \times 32 \times 32$ の 3D-FFT が 3.7 マイクロ秒、$64 \times 64 \times 64$ のが 13.3 マイクロ秒で計算できるとのことである[*45]。

他の細かい特徴として、Anton では、基本的に単精度 (32bit) で計算が行われている。興味深いのは、単精度計算を、浮動小数点ではなく、固定小数点を用いて行っており、これもまた演算速度を上げるのに寄与していると思われる。単精度 MD 計算の精度を調べるために、D. E. Shaw リサーチで作っている Desmond という MD アプリを倍精度で PC クラスタ上で走らせた結果との比較スライドがあったが、いまのところ単精度に起因する大きなエラーなどは出てないようである。

Shaw 氏の講演に入ろう。講演では、Anton の簡単な紹介と、ベンチマーク結果、BPTI という小タンパク質での 1.03 ミリ秒 MD 結果の報告、また他のタンパク質での MD 結果も少し紹介された。ベンチマーク結果では、DHFR という中タンパク質の結果が紹介された。通常のスパコン上では、一日あたり 100 ナノ秒の MD 計算ができるのに対して、Anton では 16,400 ナノ秒という結果であり、およそ 100 倍の速さである (図 18 上)。次に BPTI での 1.03 ミリ秒の MD 計算が終了したことが報告された (図 18 下)。それまでの MD シミュレーションの世界最高記録が、イリノイ大学の K. Schulten 氏のグループが WW ドメインで行った 10 マイクロ秒シミュレーション [2] であったので、この結果は驚異的である。BPTI は世界で最初に MD シミュレーションが行われた記念すべきタンパク質であるが [4]、インヒビターとして機能するタンパク質であり、構造変化はほとんどしないと考えれられている。実際、BPTI にはタンパク質のチェーン間を” 硬く” 結びつけるジスルフィド結合が 3 個存在する。なぜこのような面白みのないタンパク質を最初の 1 ミリ秒のターゲットに選んだのかだが、MD に詳しい知人の話では、ミリ秒までシミュレーションをすると、現在我々が MD に用いている力場パラメータの粗 (あら) が出てくる可能性があって、柔らかいタンパク質では壊れてしまう可能性がある、そこで間違いなく壊れないであろう BPTI を最初のターゲットに選んだのではないか、とのことである。



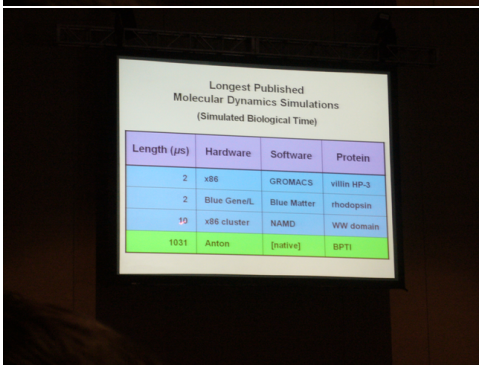


図 18　(Shaw 氏の講演スライドから)Anton の (上) ベンチマーク結果と (下)BPTI での 1.03 ミリ秒達成の報告

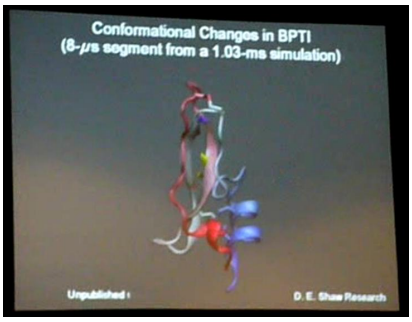
BPTI の結果だが、1.03 ミリ秒のうちで小規模な構造変化が観測された 8 マイクロ秒のムービーを見せていた。硬いタンパク質と考えられている BPTI でも構造内部の 3、4 個の水が出入りすることによって滑らかに構造変化する様子が見てとれた[*46](図 19)。しかもすごいのが、暫くするときちんと元の構造に戻っていた。構造の違いを定量化するためのクラスター解析の結果では、綺麗に複数の安定構造があるのが見えた (図 20)。また RMSD 値[*47]でみると、クラスター分類にしたがって RMSD 値が異なる値を示していた。この構造変化が、果たして BPTI が平衡状態下で本来持っている揺らぎのひとつな

[*44] 論文によれば、ひとつのメッセージを送る際の latency が、チップ上のコア間通信で 85 ナノ秒、異なるチップ上の通信で 55-75 ナノ秒だそうだ

[*45] Anton では、vdW 項と Direct Sum 項の HTIS で行われる計算が速いので、カットオフ長を敢えて大きくとって普通の MD 計算より粗いメッシュが用いられる

[*46] あまりに滑らかな動きなので、もしかしたら Low-pass filter 的な処理を施しているのかもしれないが、そこまで細かい英語が聞き取れなかった。。

[*47] RMSD とは Root Mean Square Deviation の略で、Native 構造からの各原子座標の差の根二乗平均である



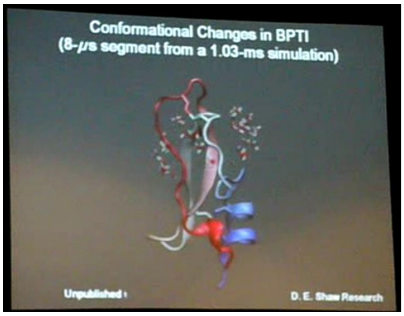


図 19　(Shaw 氏の講演スライドから)BPTI のムービー。(上) 構造変化前。(下) 構造変化後

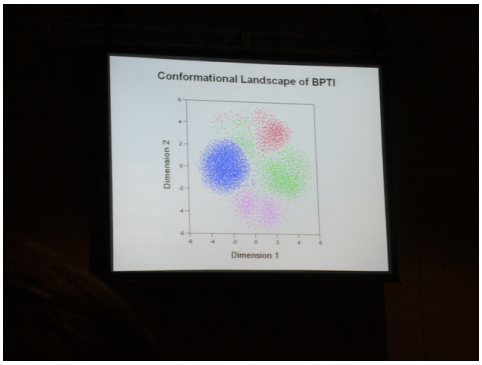


図 20　(Shaw 氏の講演スライドから)BPTI トラジェクトリのクラスター解析

のか、もしくは力場のパラメータが悪いために単に構造が壊れているためなのか、私には今のところ判断がつかない[*48]。ただし、今後このような長いシミュレーションをいくつか行っていけば、現在我々が用いている力場パラメータをより厳しく精査することが可能となるのは間違いない。また、もし今回観測された BPTI の構造変化が力場パラメータの粗 (あら) によるものでないとしたら、もっと” 柔らかい” タンパク質では一体どれだけ揺らいでいるのか見当もつかない。先の知人の話では、それはきっと我々が今まで漠然と信じている、タンパク質

[*48] RMSD 値のプロットは 300 マイクロ秒までで終わっており、その先の 1 ミリ秒付近の RMSD 値がどうなっているか気になった

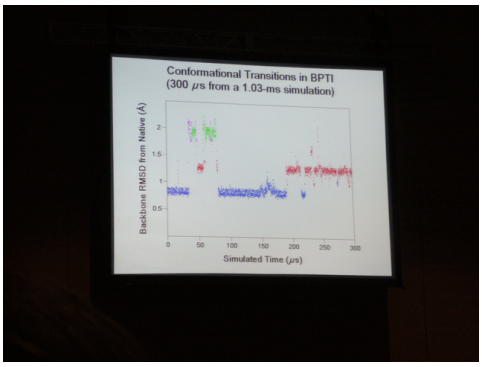


図 21　(Shaw 氏の講演スライドから)BPTI トラジェクトリの RMSD プロット

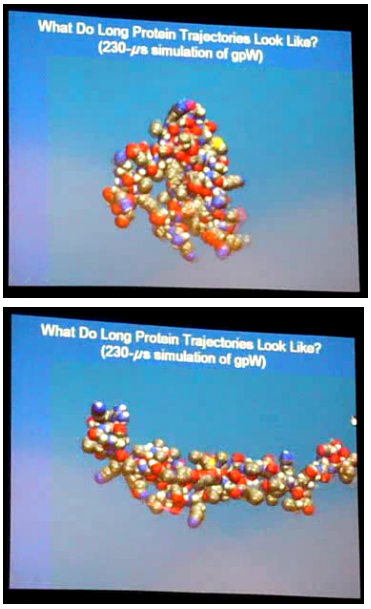


図 22　(Shaw 氏の講演スライドから)gpW の結果。転移温度でのシミュレーション。(上)Fold 状態と (下)Unfold 状態

は Native 状態” 付近” で揺らいでいる、という描像を覆すのではないか、ということだ。

他のタンパク質の結果だが、gpW という小規模タンパク質のフォールディング (図 22)、Kv 1.2 Potassium Channel(図 23) の dwetting、WW ドメインのフォールディングについての結果 (図 24) があった。gpW や WW ドメインで興味深かったのが、どちらも Fold から Unfold した後で、リバーシブルに Fold へ戻っているように見えることだった。

講演の最後に、「Anton はいくらで買えるのか？」という質問が挙がった。それに対して Shaw 氏は「Anton は学術用途で作っているから、売ることは考えていない。ただ、政府から要請があれば (売ることも) 考える」と答えていた。先のレポートで高田准教授が指摘されているように、Anton が世に出回ると、少なくとも 1 マイクロ秒以上の MD 計算でなければ論文

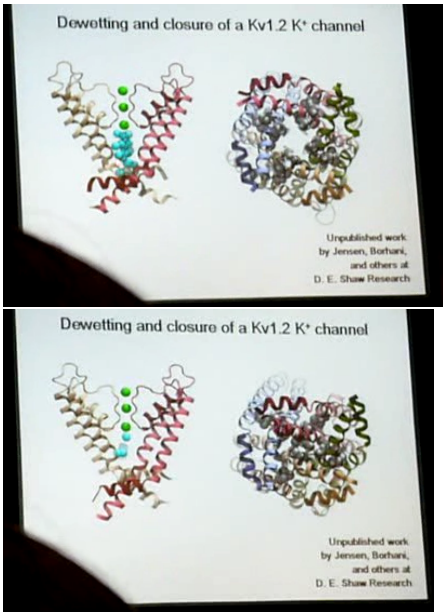


図 23　(Shaw 氏の講演スライドから)Kv 1.2 Potassium Channel の dwetting。(上) 水 (水色の分子) がいる状態と (下)dwetting に伴い構造変化した状態

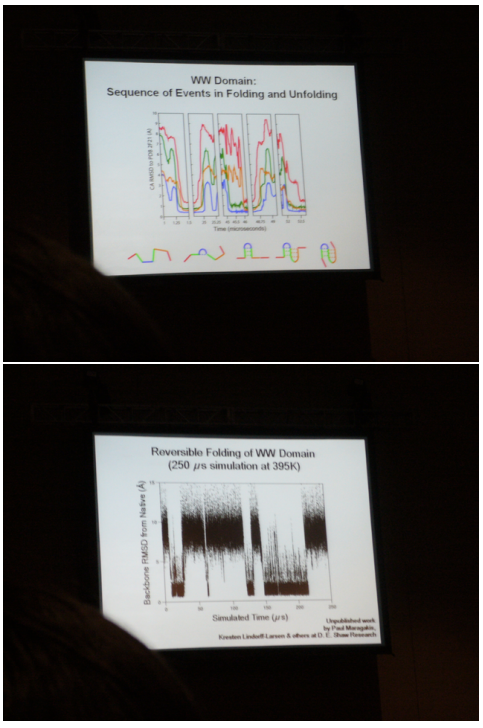


図 24　(Shaw 氏の講演スライドから)WW ドメインのフォールディングシミュレーション結果

がアクセプトされない時代がやってくるかもしれない。もちろん科学の進展のために Anton はできるだけ広まったほうが良いが、完全にパワープレイの時代になってしまうのか、MD 計算を行っている身として不安を感じた一日だった。

5　2009 年 11 月 19 日 Sustainability

朝会場に着くと、Al Gore 氏の Keynote のため警備が厳しくなっていた。Gore 氏の講演は写真撮影一切禁止だった。夜にはテクニカルプログラム参加者向けの Reception があった。場所は、ダウンタウンの Portland Center for the Performing Arts で、フロアの 1 階から 4 階を使用し、各階にいろいろな地元料理がビュッフェスタイルで置いてあった[*49]。日本の学術系会議の懇親会では味わったことのないゴージャスさでとても楽しんだ。去年の Reception は野外で寒く、ひどかったらしいが、今年はかなり良かったと思う。

5.1　Keynote Address: Building Solutions: Energy, Climate and Computing for a Changing World

Al Gore 氏による地球環境と HPC に関する講演であった。さすがに有名人で、他の Keynote よりも明らかに聴衆が多かった。私は「不都合な真実」を視聴したことがあって、そこから Gore 氏のことを単にパフォーマンスの上手い人という印象を持っていたのだが、それは良い意味で裏切られた。実際に講演を聞くと Gore 氏は思慮深い・博識な政治家だった。講演の中で Gore 氏は、「環境問題というのは、その原因が多岐に渡っており、簡単にその原因を直接実感することができない。環境問題に対してコンセンサスを作り、政治的に重大な決断を行うには、HPC の力が必須である」という趣旨の発言を繰り返し述べており、環境問題に対する HPC の力を非常に重要視していた。

驚いたことに、1 時間半の講演中、Gore 氏はスライドを 1 枚も使わずに一人で話しつづけた。それなのに、よいリズムでユーモアを挟み込み、聴衆を飽きさせることがなく、講演終了後にはスタンディングオベーションが起こったのには恐れ入った。政治家といえば、丁度前の週に行政刷新会議による次世代スパコンプロジェクトへの仕分け作業があったばかりであったが、仕分け結果の是非はさておいて、Gore 氏のように科学へ興味を持ってくれて、理解を示してくれる政治家がいるアメリカを羨ましく思った。ただし講演の最後に Gore 氏は、民主主義なのだから科学者ももっと政治に発言してほしい、と釘をさしてもいた。

その他、謝辞

毎日、お昼時には展示会場を廻っていた。そこで撮った各ブースの写真を以下に載せる (図 26-33)。

[*49] 料理以外にも 1 階ではビール・カクテル、2 階・3 階にはビール・ワインが、4 階まで上がるとご褒美のデザートやコーヒーがあった

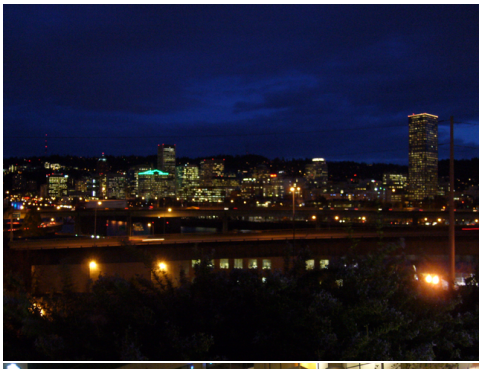


図 25　(上) 全ての聴講を終えた後で SC09 会場 (オレゴンコンベンションセンター) からポートランドの夜景。(中)Reception への移動中、街中を走る路面電車 MAX の最新型に乗れた。(下)Reception 会場である Portland Center for the Performing Arts

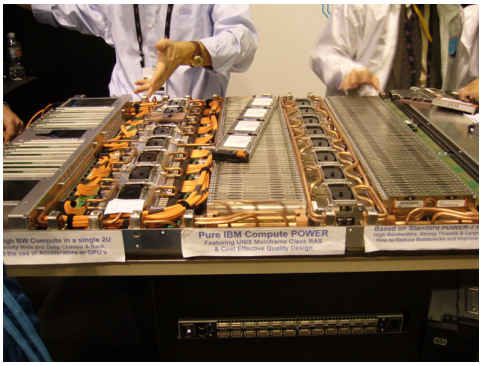


図 26　IBM ブースでの Power7 サーバーノード展示。かなり賑わっていた

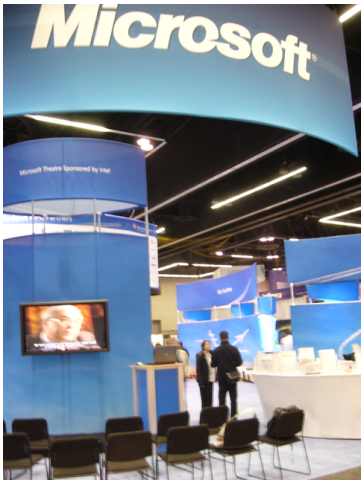


図 27　Microsoft ブース

図 28　Oak Ridge 国立研究所ブース。去年理研に来た J. Smith 氏がビデオに映っている

図 29　Argonne 国立研究所ブース

図 32　AMD ブース

図 30　Sun ブース

図 33　KAUST 大学ブース (サウジアラビアの王立大学)。バーチャルリアリティ装置があった。知人いわく、タンパク質の中に” 入れた” らしい。

図 31　NASA ブース

図 34　理研スパコン開発実施本部のブース

図 37　東工大ブース

図 35　理研 GRAPE プロジェクトのブース

図 38　筑波大ブース

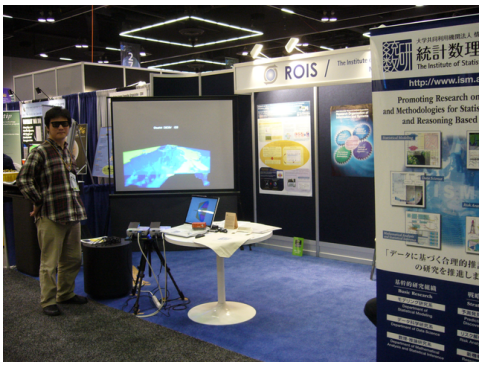


図 36　統計数理研究所ブース

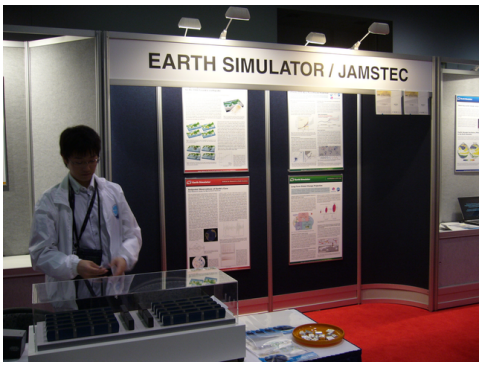


図 39　JAMSTEC ブース

最後に、SC09 への参加の機会をいただいた理化学研究所次世代スーパーコンピュータ開発実施本部の皆様にお礼申し上げます。特に、出張のお世話をしていただいた内田様には大変お世話になりました。また会場や Reception などでお話した皆さんには HPC について色々と教えていただきました。重ねてお礼申し上げます。

参考文献

[1] B. Borrell. Chemistry: power play. *Nature*, Vol. 451, pp. 240–243, 2008.

[2] P.L. Freddolino, F. Liu, M. Gruebele, and K. Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophysical Journal*, Vol. 94, No. 10, pp. 75–77, 2008.

[3] R.H. Larson, J.K. Salmon, R.O. Dror, M.M. Deneroff, RC Young, JP Grossman, Y. Shan, J.L. Klepeis, and D.E. Shaw. Highthroughput pairwise point interactions in Anton, a specialized machine for molecular dynamics simulation. In *Proc. 14th International Symposium on High-Performance Computer Architecture (HPCA 08), Salt Lake City, UT*, pp. 331–342, 2008.

[4] J.A. McCammon, B.R. Gelin, and M. Karplus. Dy-

図 40　T2K ブース

図 41　埼玉工業大学ブース

図 42　AIST ブース

図 43　NiCT ブース

namics of folded proteins. *Nature*, Vol. 267, No. 5612, pp. 585–590, 1977.

[5] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, Vol. 117, No. 1, pp. 1–19, 1995.

[6] S.J. Plimpton and B.A. Hendrickson. Parallel molecular dynamics with the embedded atom method. In *MATERIALS RESEARCH SOCIETY SYMPOSIUM PROCEEDINGS*, Vol. 291, pp. 37–37. Citeseer, 1993.

[7] Y. Shan, J.L. Klepeis, M.P. Eastwood, R.O. Dror, and D.E. Shaw. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *The Journal of chemical physics*, Vol. 122, p. 054101, 2005.

[8] D.E. Shaw. A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. *Journal of computational chemistry*, Vol. 26, No. 13, p. 1318, 2005.

[9] D.E. Shaw, M.M. Deneroff, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, K.J. Bowers, J.C. Chao, et al. Anton, a special-purpose machine for molecular dynamics simulation. *ACM SIGARCH Computer Architecture News*, Vol. 35, No. 2, p. 12, 2007.

[10] D.E. Shaw, R.O. Dror, J.K. Salmon, JP Grossman, K.M. Mackenzie, J.A. Bank, C. Young, M.M. Deneroff, B. Batson, K.J. Bowers, et al. Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pp. 1–11. ACM, 2009.

[11] M. Snir. A note on n-body computations with cutoffs. *Theory of Computing Systems*, Vol. 37, No. 2, pp. 295–318, 2004.

[12] C. Young, J.A. Bank, R.O. Dror, JP Grossman, J.K. Salmon, and D.E. Shaw. A 32x32x32, spatially distributed 3D FFT in four microseconds on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pp. 1–

11. ACM, 2009.