

## SC07 レポート

慶應大学医学部医化学教室 木下綾子

理化学研究所メンバーの一員（レポーター）として、スパコン界最大を誇る学会、SC07 がネバダ州リノで行われ、11月12日から15日までの日程で参加しました。同じくレポーターとして派遣され、今回の旅程をご一緒したのは、東京大学地震研究所で地震のシミュレーションをテーマに研究している齋藤さん、海洋研究開発機構地球シミュレーターセンターで雲のシミュレーションを行っている島さんです。私自身は、医学部という組織で、代謝を分子レベルでシミュレーションし、情報科学と実験を組み合わせることで臓器レベルでの知見を得ようとする研究を行っています。シミュレーションのスケールアップによる計算の大規模化のためにスパコンを利用したいと考えていますが、興味はそれを使って初めて発見できる生物学的な事象にあります。これまで純粋なハイパフォーマンスコンピューティング分野に接したことがなく、そのため、今回の学会の内容を理解できるだろうか、他のレポーターの方々と話が合うだろうか、とかなりの不安を抱えて出発しました。しかし、後者の不安は、往路トランジットで降りたサンフランシスコ空港で消え去りました。長い待ち時間に、眠気をこらえるために入ったカフェで、お互いの分野での問題点や解決策について熱く語り合い、すっかり打ち解けることができました。

学会会場に到着してまず驚いたのは **Registration** のスマートさでした。例年1万人近くが参加するという学会では、名前の確認にさぞかし長蛇の列ができていたのだらうと思っていたのですが、PCにログインすることで参加が確認され、隣の窓口でスムーズにネームプレートと **proceeding** を受け取るシステムになっていました。ここで配布されるものは、かなり使い勝手の良いリュックと小さなプログラム、学会のホームページへのリンクが入った **USB** フラッシュメモリのみで、その合理性にも感心しました。

初日は展示会場を廻りながらのレセプションパーティーで、展示会場各所に設けられた軽食をつまみながら、NEC、Cray、IBMなどのブースを見て回りました。また学会のテクニカルプログラムは、大小組み合わせて約5-10会場で同時に様々な形式の発表が行われる大規模なものでした。

13日には、**Computational Biology** というセッションタイトルのついた **Paper** 枠の3演題を聴講しました。

**"A Preliminary Investigation of a Neocortex Model Implementation on the Cray XD1**

(Kenneth L. Rice et al.)”では、Bayesian ネットワークに基づいて開発された既存の脳新皮質アルゴリズム (George、Hawkinsn による) を、画像を認識するための視覚野モデルに焦点を当てて Cray XD1 (432 dual core、2.0GHz Opteron processor、144 Virtex II ProFPGAs. というスペック) に載せ、高速化することに成功したという話でした。学習と認知を司る脳新皮質の均質性を利用した構造と機能のモデリングは、ここ十年間で多くなされており、特に構造的 Bayesian ネットワークに基づいて、構造的なデバイスとして脳を捉え、洗練されたパターンマッチングによって計算する方法が提唱されているようです。発表された paper は、この方法について掘り下げた研究ではなく、あくまで高速化なようですが、その高速化の要となったのはリコンフィギュラブル・ロジックの工夫、ということでした。

続く”Anatomy of a Cortical Simulator (Ananthanarayanan R. et al.)”もキーワードは脳新皮質でしたが、こちらは神経細胞を離散ユニットとして捉え、その発火と情報伝達メカニズムを組み込み、より実物に忠実な方法を実装した上で、それを高速に、大規模に解くことを目指したものでした。

脳新皮質の機能単位としての神経細胞は、樹状突起を介して周囲の神経細胞から数千というインプットを受け取り、神経細胞の膜電位を閾値以上に上昇させるような刺激が入ったときに”発火”して軸索を通じて刺激を伝導し (このとき、軸索伝導による時間おくれを伴う)、軸索と、別の神経細胞の樹状突起とが接する部位 (シナプス) を介して、接する神経細胞に刺激という情報を伝達します。神経細胞の重要な性質は、発火のタイミングに依存した情報の伝達と、シナプス可塑性です。このモデルでは、発火前後で pre-spiking と post-spiking の2つの状態に神経細胞を分類しており、ネットワークトポロジーと神経細胞のモデル化にいくつかの仮定を置いて、この種の発火神経細胞を扱うモデルの高速化を実現していました。442x10<sup>9</sup> のシナプスがラット脳新皮質に存在し、1Hz の平均速度で神経細胞が発火し、それぞれの神経細胞は1秒に1回ターゲットのシナプスに伝達するという仮定を置くと、442 billion の情報が1秒間に扱われることになるそうです。これを、1ミリ秒か、それより小さい時間分解能でシミュレーションするために、大規模並列処理の最適化と新たなシミュレータのデザインが必須です。この研究で開発されているシミュレータ(C2)は、IBM との共同開発ということですが、IBM は「Blue Brain プロジェクト」という一大プロジェクトも立ち上げており、こちらはより詳細な生物学的情報を考慮した分子レベルの脳モデルを構築することで、生物学的に脳の機能を理解しようとするものだそうです。そして、C2 プロジェクトの大目標は、脳がどのように機能して大量の情報処理を行っているかを理解することで、新たな認知処理、コンピューティングアーキテクチャ、

プログラミングパラダイムを構築し、将来的に企業アプリケーションとしてビジネスに発展させることにあるようです。

「ネットワークは解剖学的な神経可塑性を表現していないが、その確率的な特徴は再現しており、発表者らが開発した C2 は、最先端のハードウェアと新しいアルゴリズムとソフトウェアデザインを融合させることで、前例のない大きなサイズの問題（ラットの脳をほぼ real-time で）を解くことに成功した。しかし、ヒト ではラットの約 500 倍もの神経細胞が存在する。これをシナプス間のコミュニケーションを勘案して解くのは、まだまだ長い道のりである。」と語っていました。

“Large-scale Maximum Likelihood-based Phylogenetic Analysis on the IBM BlueGene/L (Ott M. *et al.*)” は、DNA シーケンスデータに基づく系統樹の予測を行うための問題点（トポロジー探索空間が、シーケンスの長さに対して指数的に増大すること、likelihood function の計算が難しいこと、メモリー消費が激しいこと）について、新しいアルゴリズムの提唱、モデル近似法の改良、ハイパフォーマンスコンピューティングで解決しようと試みた発表でした。複数生物種における DNA やタンパクのシーケンスデータをアライメントして系統樹を書く手法はよく目にします。生命科学の分野では、次世代シーケンス法の出現が今まさに大きな話題になっており、今後数年のうちに飛躍的にシーケンスデータの取得速度が向上するものと見込まれています。こういったシーケンス技術の革新によって指数関数的に蓄積されていくシーケンスデータを基に、例えば 1000 種以上といった、大きな系統樹を書く要求が増えているそうです。種が多くなるほど、また高等生物でシーケンスが長くなるほど効果的な分散処理が必要であり、系統発生の予測は、バイオインフォマティクスの大きなチャレンジの 1 つとなっています。Maximum likelihood(最尤) tree を見つけるためには、可能性のある全ての系統樹を試す必要がありますが、比較する種の多さに従って指数的に系統樹の選択肢は増加してしまいます。そこで、探索空間を減らす、ヒューリスティックな系統樹サーチアルゴリズムが必要であり、今までにもいくつか開発されているものの、どれも「maximum-likelihood tree トポロジーを見つけるものではなく、best-known な Maximum likelihood-based tree をとってくるだけだった。」という問題点があったようですが、発表中には今までの手法の問題点をよく理解することができませんでした。ここでは、題材として、今まで ML で解かれたものの中で最大の問題を選び、開発した最先端の Maximum Likelihood (ML) プログラムが IBM BlueGene/L (BG/L) 上でどの程度効果的に動くかを研究してデータが発表されていました。キャッシュ効率を増やすことで非常にリニアなスピードアップを得ることができるということでした。

ここで、聴衆の一人が「その研究から得られた、または得られるはずの生物学的な発見は何だ？」という趣旨の質問を投げかけていました。どのような情報をどのような速度で処理すると、生命科学にどのくらいのインパクトがあるかを知りたいと思っていましたが、ここまでの3演題の議論は全て、どのくらいコンピュータの性能が出たかということに終始していたため、私も思わず身を乗り出しました。しかし、困惑した発表者から返ってきたのは、「私は biologist ではないので」という回答でした。それぞれに研究の専門があり、それぞれの目的があることは明確ですが、どのような意識を持って異分野間の融合に向かうかということを変更して考えさせられました。

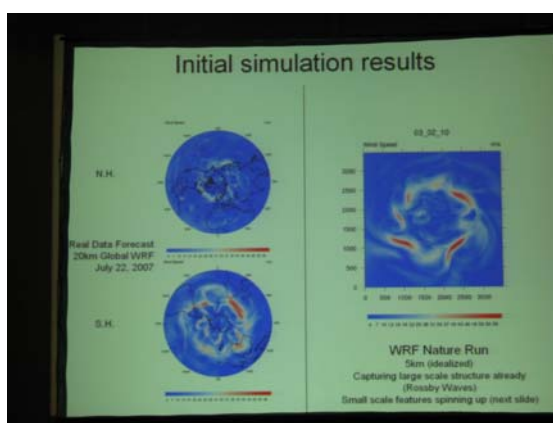
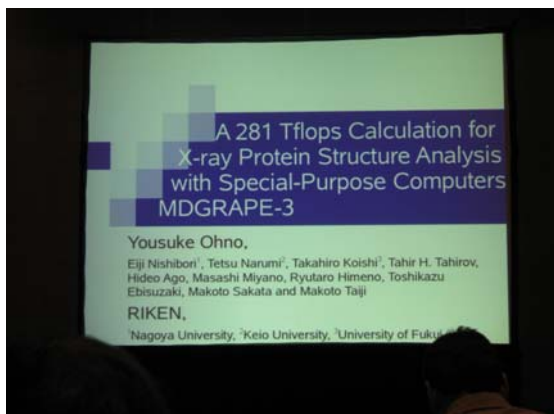


(Paperセッション (Computational Biology) の風景)

Masterworks はハイパフォーマンスコンピューティングを実際の産業に応用している様々な例を紹介するセッションです。“**A Grand Challenge: Multicore and Industrial Modeling and Simulation (Appa J.)**” では、巨大な航空宇宙企業 BAE Systems の Technology and Engineering Services 部門のリーダーである Appa 氏が、産業分野におけるモデリングとシミュレーションの重要性と、そのために必要な技術革新について、サマライズしていました。。次世代 Computing Challenge のキーワードとして、マルチコアプロセッサ、FPGA(Field Programmable Gate Array : プログラム書き換え可能な特殊 LSI)などを挙げていました。

14日には、Gordon Bell Prize のファイナリスト4件の発表が行われました。Gordon Bell Prize は、最初の SC が開催された年から設置されている重みと名誉のある賞です。テクニカルプログラムの最終日に受賞式が行われ、受賞者には \$ 10,000 の賞金が授与されます。SC に投稿されたテクニカル原著論文の中から、厳しい審査によって4件に絞られたそうですが、その評価基準はハイパフォーマンスコンピューティングの科学への応用性であり、

より具体的には、”real scientific computation”への適応性、洗練されたアプローチ、ハイパフォーマンスコンピューティングを用いて行った研究の科学的重要性、によって評価されたようです。4件中2件が日本からの論文で、1つは MDGRAPE-3 というタンパク質3次元構造解析専用コンピュータ、もう1つは地球シミュレータで走る第一原理シミュレータである PHASE の論文でした。Gordon Bell Prize は、科学的発見とアプローチの斬新さが評価された”Extending Stability Beyond CPU-Millennium: Micron-Scale Atomistic Simulation of Kelvin-Helmholtz Instability”(Blue Gene/L を用いており、結果のヴィジュアルも非常に美しかった)が獲得し、惜しくも賞を逃しましたが、日本のハイパフォーマンスコンピューティング分野の驚くべきレベルの高さを改めて認識しました。

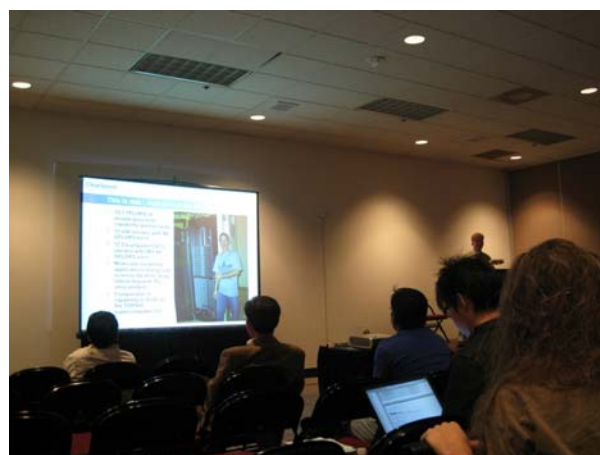


HPC をインタラクティブ・シミュレーションに応用することを目指した同志が集まって開催された BOF(Birds of a Feather)セッションを聴講しました。5人の演者がそれぞれ講演し、陸軍訓練への応用などが紹介されました。



-----

FPGAs-application をテーマにした Exhibitor Forum では、これまで多大なお金と時間を必要としてきた暗号解読において、FPGAs (Field Programmable Gate Arrays)の利用が進んでおり、これまでの数百倍のパフォーマンスが出ることなどが紹介されていました。最終日のポータブルクラスタ開発に関する Exhibitor Forum にも出席し、「ポータブルというからにはもしかして、将来 iPod くらいになったりするのかな？でもそれはクラスターなのか？」と期待しましたが、下にゴロゴロをつけて持ち運べるほどになった、という報告でした。しかし、従来の冷却装置の工夫などで容積はもちろん、消費電力も抑えられるということで、将来的にこのような技術が大きく効いてくるのだろうと予感する Forum でした。発表された ClearSpeed Advance Card は東京工業大学の TSUBAME に導入して、38 テラフロップスか 56.4 テラフロップスへ、48%のパフォーマンスアップを実現する一方、消費電力は 2%増、コストは 10%増にとどまったということでした。



15日は、David E. Shaw 博士の Plenary Lecture “Toward Millisecond-scale Molecular Dynamics Simulations of Proteins” がありました。この講演は、Computational Biology の Paper セッションを聴いて以来の「HPC が生物学に与えるインパクトは結局何だったのだろう？」というわだかまりを払拭する内容でした。今回の学会で驚いたことの一つとして、HPC を利用した分子動力学シミュレーションによるタンパク質 3次元構造予測の研究が非常に盛んに行われていることがあります。企業や研究所の展示ブースでも、タンパク質がディスプレイの中で回転する映像がいたるところに登場していました。しかし、今の MD では一日にナノ秒単位のシミュレーションが限度であるため、タンパク質が構造変化して「機能」をする過程を見ることはできず、生物学的に意味のある結果を得るにはミリ秒単位でのシミュレーションが必要だというのが Shaw 博士の motivation です。Anton と呼ばれ

る専用機を構築中であり(2008年に完成予定)、フレキシビリティと、プログラマビリティを本当に必要とされる箇所のみ限定し、計算機の性能(計算時間)のボトルネックになるオフチップ・メモリへのアクセスを省くというようなことで大幅なスピードアップを実現しているそうです。すでに開発したMD専用ソフト Desmond をこの Anton に載せて 23,558 原子のタンパク(ジヒドロ葉酸還元酵素)のシミュレーションをすると、これまで発表されている MD 専用ソフトや Blue Gene/L を持ってしても一日数ナノ秒という実績だったものが、数ミリから十数ミリ秒のシミュレーションまで可能になると見積もられています(ただし、条件の違いにより単純な比較はできないとのこと)。また既に出来上がっている Desmond と既存のクラスターを組み合わせ(Desmond を使うだけでもこれまでの 10 倍以上の速度が出る)、Na<sup>+</sup>/H<sup>+</sup> 交互輸送体のシミュレーションをしていました。Na<sup>+</sup> と H<sup>+</sup> の交換は、酸塩基平衡を調節し、細胞の恒常性を維持する重要な機構で、大腸菌において構造が既に決定されていますが、その詳細な輸送動態については十分に理解されていませんでした。しかし、(この時点では数日かけて)ミリ秒での MD シミュレーションが可能になったことで、タンパクの機能を原子のレベルから再構築して理解することが可能になりました。実験的には、環境(pH)の違いや輸送体タンパクでの突然変異がどのような機能の違いを起こすか、がこれまでに John Kuriyan 博士の研究室で測定されてきました。MD シミュレーションの結果は、これらの実験によく一致しており、更にシミュレーションモデルを使って、輸送体タンパク質を構成するどのアミノ酸がイオン交換に必須であるかを予測し、イオンの交換輸送機構を詳細に同定することに成功しており、Science に掲載されました。論文は生物学的な発見のみに焦点を当て、シミュレーション自体はそれを導くための実験道具として扱っており、更に HPC はそのシミュレーションに必須の道具として記述されています。しかし、HPC 性能の革新なしには得られなかった研究結果であることは明らかであり、理想的な HPC、シミュレーション、実験科学の融合を見ることができました。



-----

プログラム後も、NEC 主催のパーティー、テクニカルプログラム聴講者が招待されるブルーマンのショー(青い顔の人たちが出てくるまでその存在を知らなかったが、日本でもチケットが取れないほど人気だそうです)と楽しむことができました。HPC に関する講演の内容もさることながら、展示会場の雰囲気、学会全体の充実したファシリティ、開会式の熱気と主催者のユーモアなど、学術学会というよりはスポーツイベントのような、初めて体験する温度の高さがありました。

最後になりましたが、このような貴重な機会を与えてくださった理化学研究所と次世代スーパーコンピュータ開発本部の皆様、手続や学会旅程を全てサポートしてくださった内田紀子さん、レポーターのお二人に厚く感謝申し上げます。

---