

IBM Deep Computing Institute (ワトソン研究所)

9月15日9:00~13:00)

計算生物学センター概況

説明者:

- Dr. Joseph Jasinski
Senior Manager,
Computational Biology
Center



ここは仮想的組織であり、メンバーの大部分はこのビルディングに居るが、そのほか世界各地のIBM研究所からも研究者が参加している。

IBMはここ9ヶ月に2つの発表をした。

その一つは9ヶ月前のブルージーンプロジェクトの発表。5年間で100M\$を投じて1000テラフロップ、即ち1ペタフロップのコンピュータを作り、それを用いて生物分子のシミュレーションを行ない、蛋白を構成するアミノ酸の連鎖が折りたたまれて最終構造をとるメカニズムを理解する。

もう一つは1ヶ月前の発表で、3年間で100M\$を投じて生命科学会社と提携する。

〔当センターの主要プロジェクト〕

- パターンディスカバリとバイオインフォマティクス
- 構造生物学 蛋白構造の理解
- 蛋白力学 (ブルージーン)
- 機能遺伝子学
- 知識管理 膨大なデータベースの構造

ゲノムや蛋白に関する大量のデータが得られるようになり、それらのデータを理解するために計算力が必要になった。IBMのような会社は高速計算をし、大量データを保存し、扱う能力がある。

ディープコンピューティングとは巨大スケールの科学・ビジネス問題と巨大計算能力とのアルゴリズムによる結合であり、問題を解く方法を見つけるのがデータ量の増大と共に指数関数的に困難になる。

蛋白を結晶化しX線で照射、あるいはNMRを用いて構造を知ることが出来る。しか

蛋白構造予測の戦略には、「Homology(相同性)」、「Fold Recognition(折畳み認識)」、「Ab Initio (アベニシオ)」の 3 つがある。

「Homology」は蛋白を構成するアミノ酸のシーケンスの相同性を通じて構造を予測する。構造が予測できればそれに伴って機能も予測できる。この方法の予測計算時間は数時間である。

構造の知られている蛋白のデータベース中のアミノ酸のシーケンス情報からその構造を“ Splash プログラム ”を用いて予測し、結果を実際の構造と比較する研究をやっている。シーケンスからパターンを自動的に発見し、パターンを分類したり階層的なパターンを見つけたりしてこれらの情報を用いる。

「Fold Recognition」は蛋白の折畳みの情報を用いる。計算時間は 1 日のオーダーである。シーケンスがまったく異なり構造は非常に類似しているものもある。

「Ab Initio」は原子間の相互作用を力学的に解いて構造を予測するもので、極めて長い計算時間を要するが、非常によい予測結果が得られる可能性がある。これをブルージーンでおこなう。

ブルージーン・プロジェクト

説明者

・ Dr. Bob Germain

Protein Dynamics Group , Computational Biology Center

本プロジェクトの目標は大型シミュレーションによって蛋白科学に貢献すること、ハード・ソフトの設計に貢献することである。

巨大シミュレーションで蛋白の折畳みが解明できれば、蛋白形成の機構を理解して、医薬と蛋白との相互作用、即ちドッキングプロブレムが分り、触媒のデータベースの構造をリファインできる。

採り得る形の数膨大であるにもかかわらず蛋白はある形にすばやく折畳まれていく。一般にアミノ酸列は必ずしも速くは折畳まれないが、体内の蛋白は機能だけでなく素早く折畳まれていくという点でも（人体を構成するアミノ酸として）選択されている。折畳まれる毎に自由エネルギーレベルが低下して、次に取り得る形の数も急速に減少していくらしい。

伝統的分子ダイナミクスでシミュレーションを行なう。つまり多数のパラメータから

なる原子間力のモデルを作り、ニュートン力学でその問題を解く。量子力学ではないので原子間の化学結合を切ることは出来ない。

1ペタフロップスの処理速度のマシンで1年かけて3.2万個の原子のシミュレーションを1千億～1兆タイムステップ行なう。実際の折畳みのプロセスのかなりの部分をシミュレートできる。一分子を初めから終わりまでシミュレートするのでは得られる情報は少ない。特定の部分に焦点を絞ってわずかに折畳まれたところでまっすぐにして再び折畳まれていくのを繰り返しシミュレートしたり、部分的に安定したところ、しばらくその形に留まるところに焦点を置く。

ブルージーン・コンピュータ

説明者

・Dr. Marc Snir

Blue Gene Project Manager ,

このプロジェクトは製品開発ではなく研究である。バイオインフォマティクスとコンピュータサイエンスとの両方が目標である。問題は、現在あるマシンより百倍速い1ペタフロップスのマシンを如何にして作るかということである。



チップには標準的なシリコン技術を用いる。今の高速コンピュータは、日本のベクトルマシンは別として、PC や小型ワークステーションを多数接続するのが普通である。しかし、それだけではPC 百万個を用いることになってあまりに高価になるし、消費電力も100メガワットになる。ちなみにこのビルの電力容量は2メガワットである。又、故障も多くなる。IBM で現在最速のマシンは12テラフロップスで、ワークステーションから構成されている。

標準のマイクロプロセッサは一つのプログラムに最適化するために1チップに多数のトランジスタを用いている。各トランジスタとしては最適化されていないので、トランジスタの数を増やしても速度は比例しては上がらない。そのためプログラムをスレッドに分割して、多数の単純なプロセッサに計算させる。このことにより各トランジスタ当たりの速度が上がる。又、メモリのアクセス速度がプロセッサの処

理速度ほど速くならず、両者の処理速度のミスマッチを生じるというフォンノイマン・ボトルネックで、標準のマイクロプロセッサではチップの半ばがキャッシュになり、キャッシングが非常に複雑になる。

標準の OS ではソフトウェアが複雑になるという問題もある。システムの性能をフロップスで計るが、バスの速度や通信能力のことは言わない。

これらの問題を解決するため、最近の技術を用いてメモリをプロセッサと同じチップに搭載する。それでもデータのやり取りに数サイクルの時間を要するので、プログラムをマルチスレッディングし、必要なデータが届いていないプロセッサには別のプログラムをやらせるマルチプログラム方式を採る。又、速度が速くなるとクロックサイクル中に信号をプロセッサ間で動かすロジックが複雑になる。そのため、所要面積が小さい単純なマイクロプロセッサを 1 チップに 32 ~ 36 個搭載する計画である。1 チップ中で 1 プロセッサが故障しても残りのチップ全体は作動する構成にしてコストを低減する。この構成は固まってはいるが。

1 プロセッサが 1 ギガフロップスエンジンだから、9 チップを 1 枚のボードに搭載して 4 ボードで、あるいは 36 チップを 1 枚のボードに搭載してそれで 1 テラフロップになる。このボードを積み上げて 1 ベタフロップにする。1 チップは 6 箇所接続でき、3 次元接続できる。各チップがプロセッサ、メモリ、スイッチを持っている。

このプログラムのスタートの発表が昨年 12 月、プロトタイプの完成が来年末、全体の完成が 3 ~ 4 年先になる。

ソフトウェア問題に故障の問題がある。ソフト的にチップの故障を救済するセルフヒーリング方式を採る。パリティ、エラーコレクションコード等で周期的にエラーをチェックし、エラーが出たら故障部分を見つけて切り離し、元に戻って計算しなおす。このソフトウェアはこれからの研究次第である。

言語については、低級言語でプログラムを全て書くのは耐えがたいから、高級言語で書きコンパイラ、ライブラリメソッドを使うが、進歩は遅くすばらしいアイデアはない。