

将来(2010年前後を想定)の スーパーコンピューティング環境について

平成16年11月

泰地 真弘人

理化学研究所 ゲノム科学総合研究センター
ゲノム解析用コンピュータ研究開発チーム

はじめに：計算機と生物学の今後

- 2010年には
 - タンパク質の基本構造: 解明終了
 - ゲノムネットワーク情報
 - イメージング技術の向上
 - 表現型情報

 - 構造やネットワークを基にした創薬
 - タンパク質の機能解明、特に酵素反応
 - 新規タンパク質の設計、バイオナノテクノロジー
 - マイクロアレイ等からの情報抽出、モデリング、シミュレーション
 - イメージからの情報抽出、モデリング
- 計算生物学・bioinformaticsに求められるものは一変
 - デジタルからアナログへ
- 豊富な実験データを基により精密なモデルを扱う必要性
- シミュレーションの需要急増：HPCの必要性

2010年の生体シミュレーション

- 分子シミュレーション
 - 分子動力学
 - 量子化学計算・第一原理分子動力学
 - 創薬への応用
- 遺伝子・生体分子ネットワーク
 - モデリングとシミュレーション
- 連続体シミュレーション
 - 細胞内・細胞間
 - 外科領域

どのシミュレーション・解析にも巨大な計算需要あり

2. 将来想定される研究内容について(ライフサイエンス)

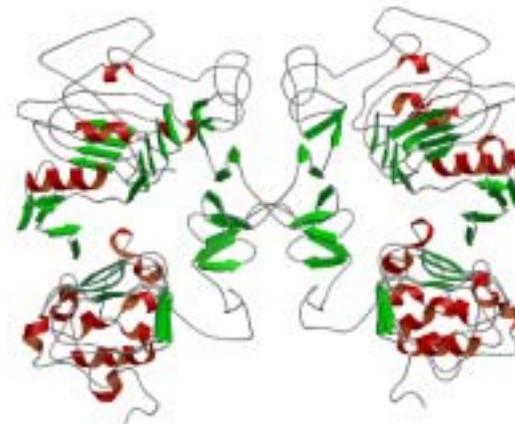
アプリケーション	概要	目的	期待されるブ レークスルー	波及効果	必要な 実行性能
高精度バー チャルスクリー ニング	分子動力学計算 によりタンパク質 と薬剤の結合エネ ルギーを精度よく 求める	医薬品等の効 率的な分子設 計	構造生物学の成 果を活かした計 算機による創薬、 耐性菌などへの システムティック な対策	我が国医薬品 業界の国際競 争力がトップク ラスになる。	10Pflops級
タンパク質の機 能解析	分子動力学計算、 第一原理分子動 力学計算、これら の連成によるタン パク質の動作原 理の解明	分子機械として のタンパク質の 解明、酵素反応 の解析	新規タンパク質 の設計、新しい 創薬ターゲット の発見	学術レベルの 向上、医薬品・ ナノエンジニア リングへの貢 献	10- Pflops級
フォールディ ング	分子動力学計算 によるタンパク質 立体構造予測	タンパク質構造 の予測および フォールディ ング過程の研究	新規タンパク質 の設計、タンパ ク質構造の解明	医薬品・ナノエ ンジニアリング への貢献	10-1000Pflops 級

理研GSCにおけるHPC

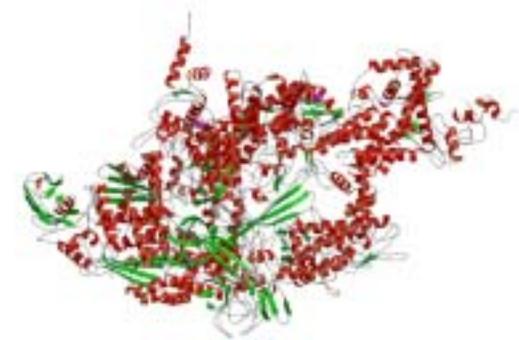
- Protein 3000 Project:
タンパク質の構造を2006年までに3000個決める文部科学省のプロジェクト
- 分子シミュレーションの役割
 - Drug Design
 - Bio-nanotechnology
- 高性能の専用計算機
MDGRAPE-3
で計算量の問題を解決



RIKEN NMR Park



EGFR outer membrane

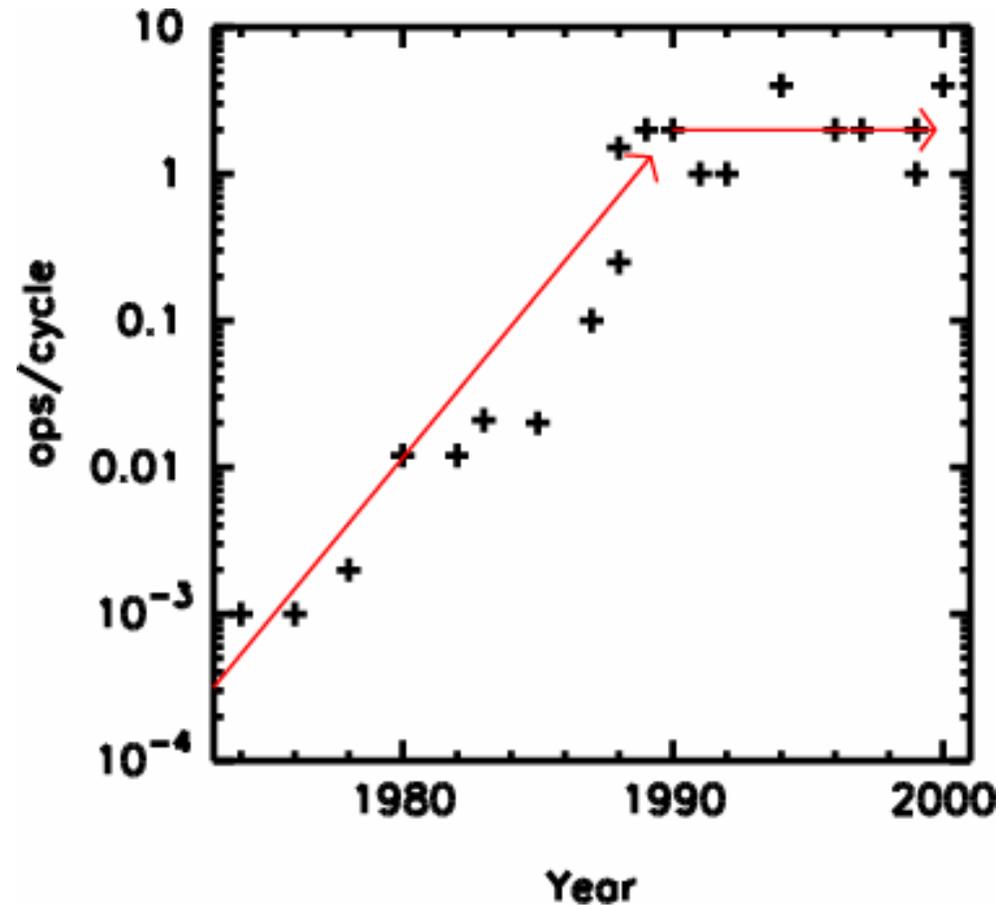


TT RNA Polymerase

なぜ専用計算機が重要か

- 汎用計算機の問題点
 - 半導体集積度の向上を活かしきれない
 - 発熱の問題
- これらの問題は今後ますます深刻になる
- 専用計算機では
 - 集積度の問題 **根本的に解決可能!**
 - 発熱の問題 **かなり緩和可能**

マイクロプロセッサの同時演算数



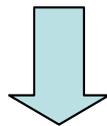
J. Makino, Proc. Toyota Symposium, Elsevier (2001)

ブロードキャスト並列化

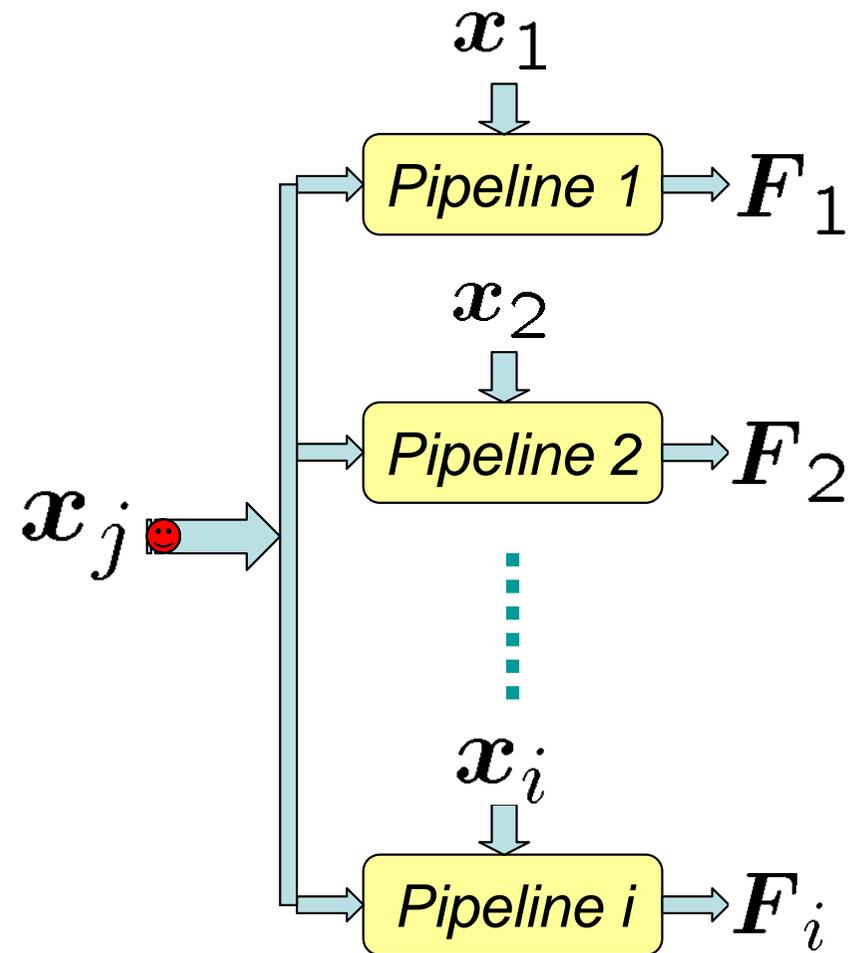
- 分子動力学の場合
- 二体力

$$F_i = \sum_j f(x_i, x_j)$$

- F_i を並列計算するとき、同じ x_j を使える



- ブロードキャスト並列化
- バンド幅問題を解決



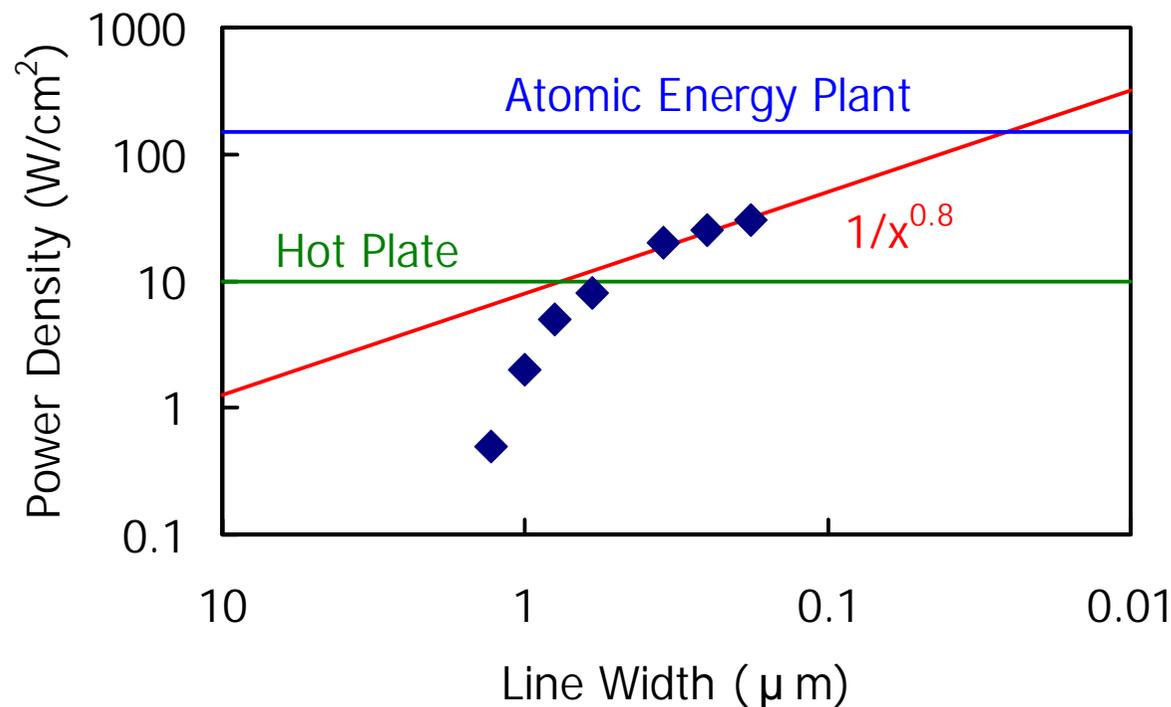
専用化による高い並列度

- 専用機では
 - 1000演算以上の同時実行が可能
 - 汎用CPUでは高々4程度で**すでに飽和**
 - 困難なく、さらに増やせる **今後より有利に**
 - **性能向上のペース ~ 20倍/5年 = ムーアの法則**
- 分子動力学計算の場合

MDGRAPE	600nm	1 pipeline	1Gflops
MDGRAPE-2	250nm	4 pipelines	16Gflops
MDGRAPE-3	130nm	20 pipelines	165Gflops

消費電力の問題

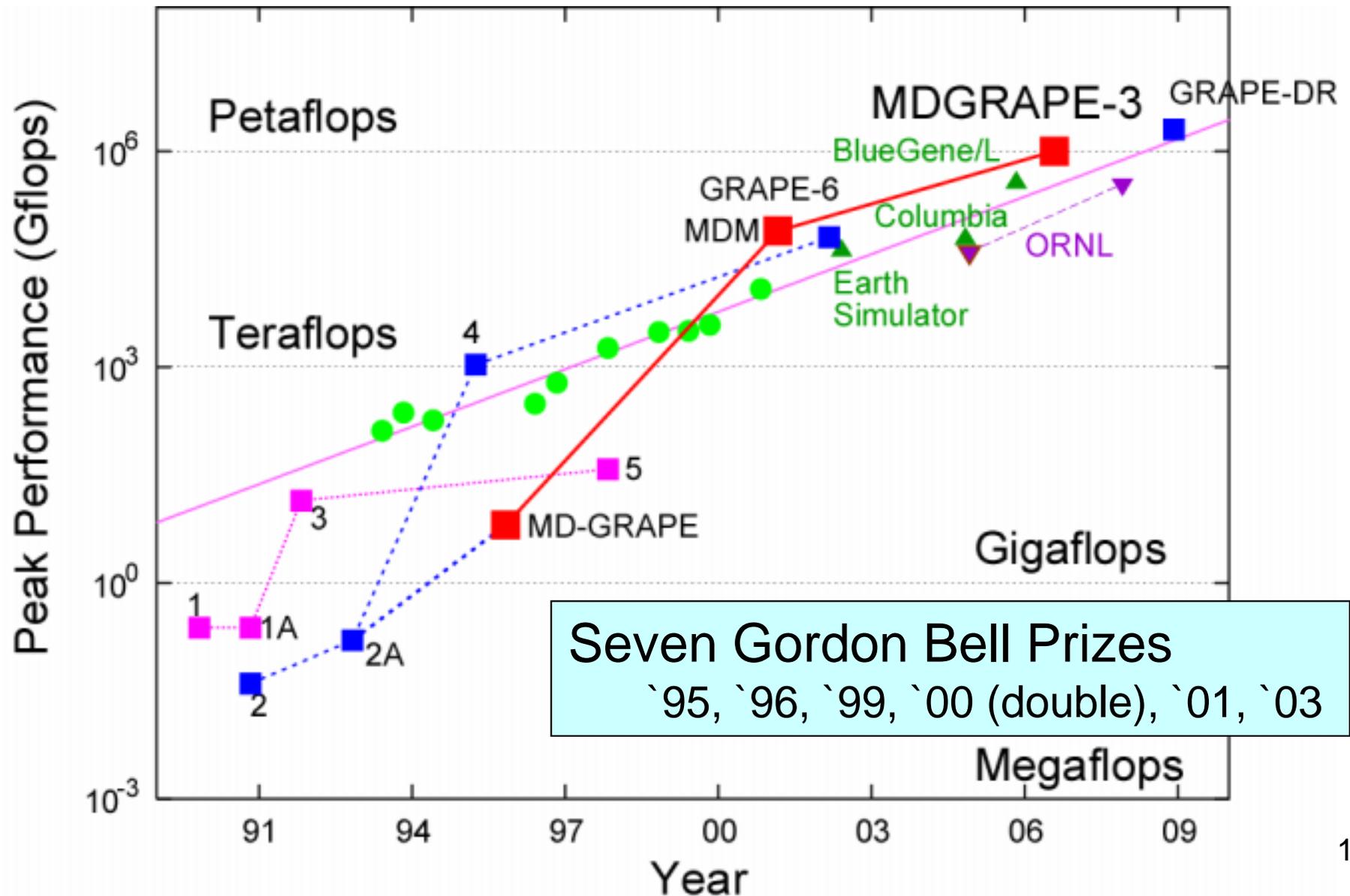
- 半導体の線幅 x が狭くなるにつれ...
- 電力/性能比($x^{2.8}$)はもちろん減る、が
- **電力/最大性能 発熱密度 ($x^{-0.8}$) は増大**



専用機は発熱が小さい

- 低めの周波数で並列度を上げる
 - cf. Columbia QCDSF, IBM BlueGene/L
- 精度を落とす
 - 分子動力学の場合 ... 単精度
- 不要な回路がなくなる
- 同等技術(130nm)での比較
 - Pentium 4 ... 13W/Gflops
 - MDGRAPE-3 ... 0.1W/Gflops

History of GRAPE (GRAVity PipE)



世界的な状況との比較

- 汎用計算機

地球シミュレータ	40Tflops	(2002)
SGI Columbia	60Tflops	(2004)
IBM BlueGene/L	90Tflops	(2004)
	360Tflops	(2005)
IBM BlueGene/P	1000Tflops	(2006)
ORNL/Cray	40Tflops	(2004)
	350Tflops	(2007)

- 専用計算機

理研MDGRAPE-3	600Tflops	(2005)
	1000Tflops	(2006)

東大GRAPE-DR	2000Tflops	(2008)
------------	------------	--------

MDGRAPE-3 (Protein Explorer)計画

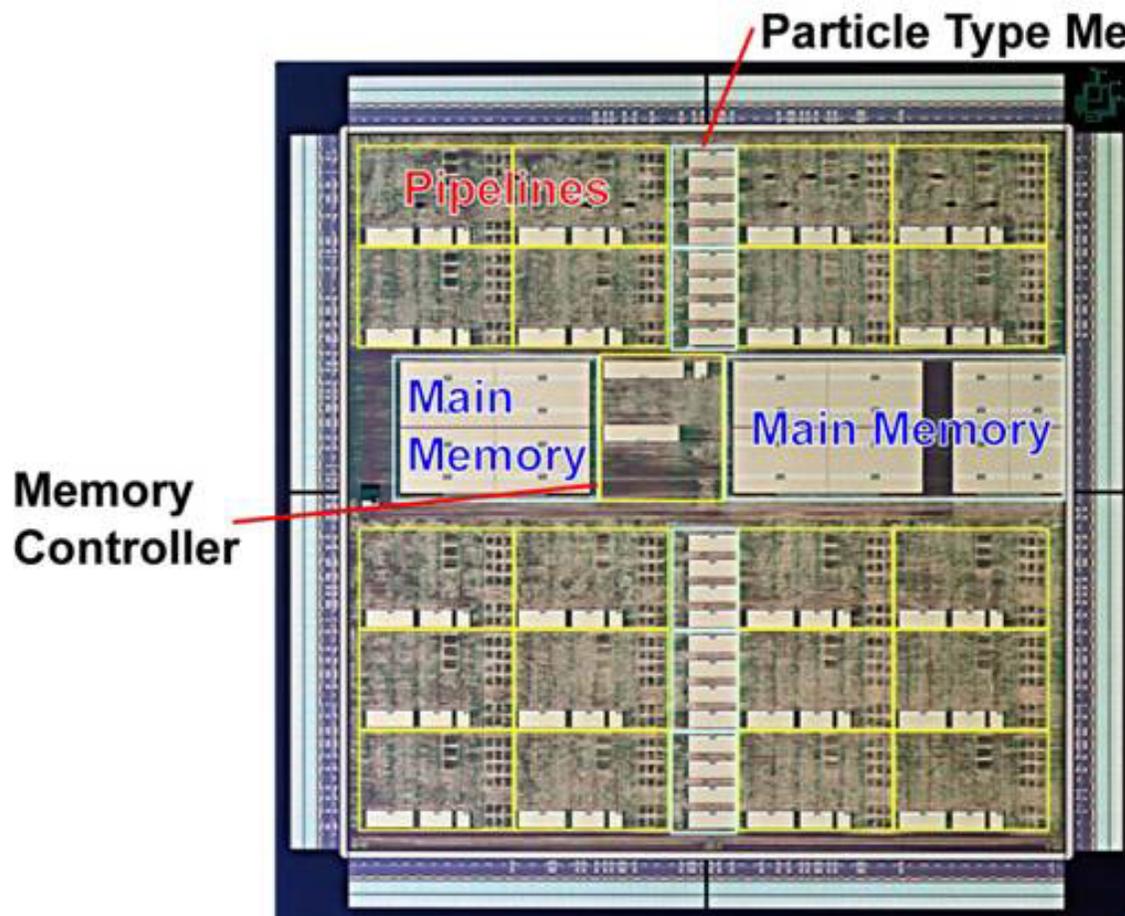
- **Petaflops** の分子動力学専用計算機
- タンパク3000プロジェクトの一環
- 2002年度開発開始、2006年度完成目標
- 現在までにハードウェアの基本要素は開発をほぼ完了



MDGRAPE-3 chip

- 力の計算LSI
- 130 nm technology
- Hitachi HDL4N
- 165 Gflops/chip at 250 MHz
- Sample: 230Gflops @ 350MHz

MDGRAPE-3 chip



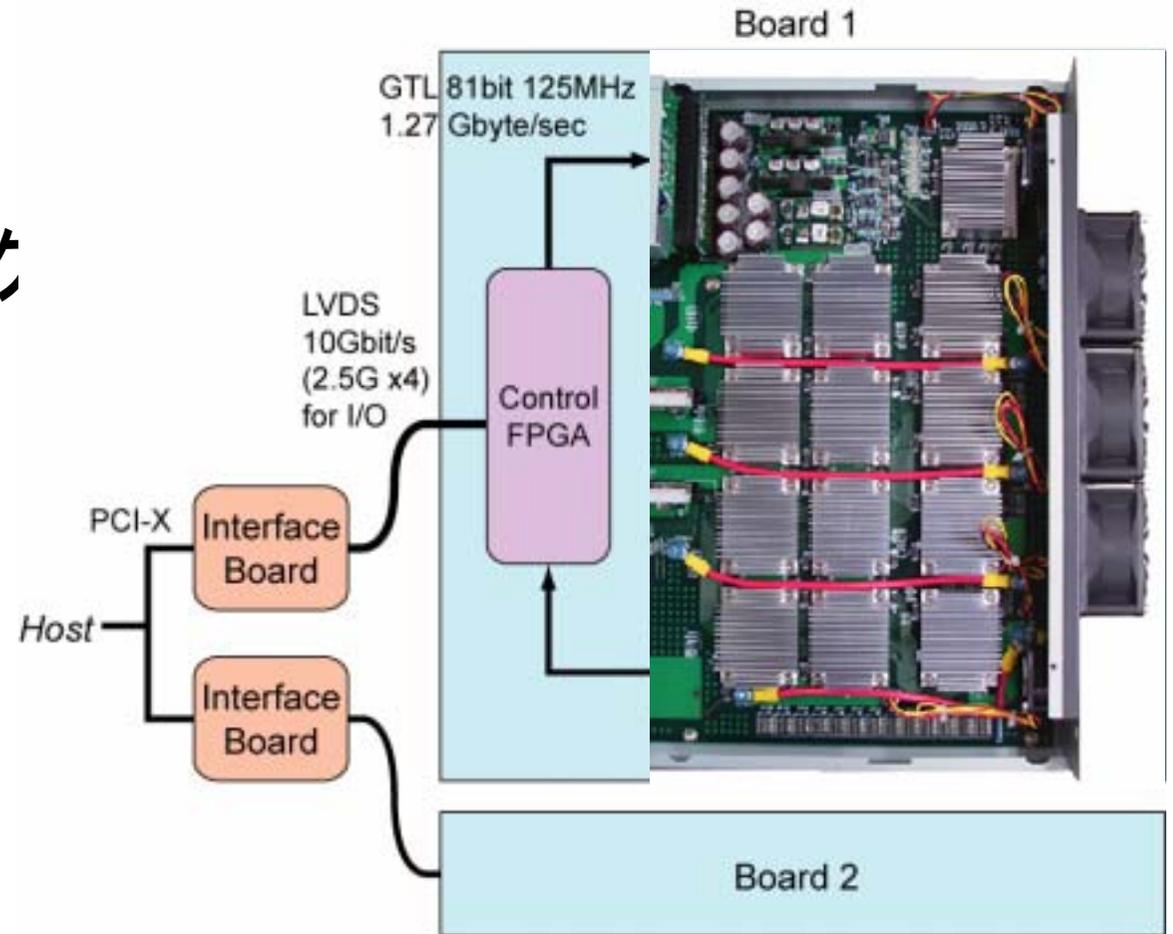
- メモリ内蔵(32,768体分)
- 130 nm (日立HDL4N)
- Vcore=+1.2V
- 15.7 mm X 15.7 mm
- 250MHz (worst)
- 660同時演算

- 19W at 350 MHz
- 1444 pin FCBGA

Chip / Host

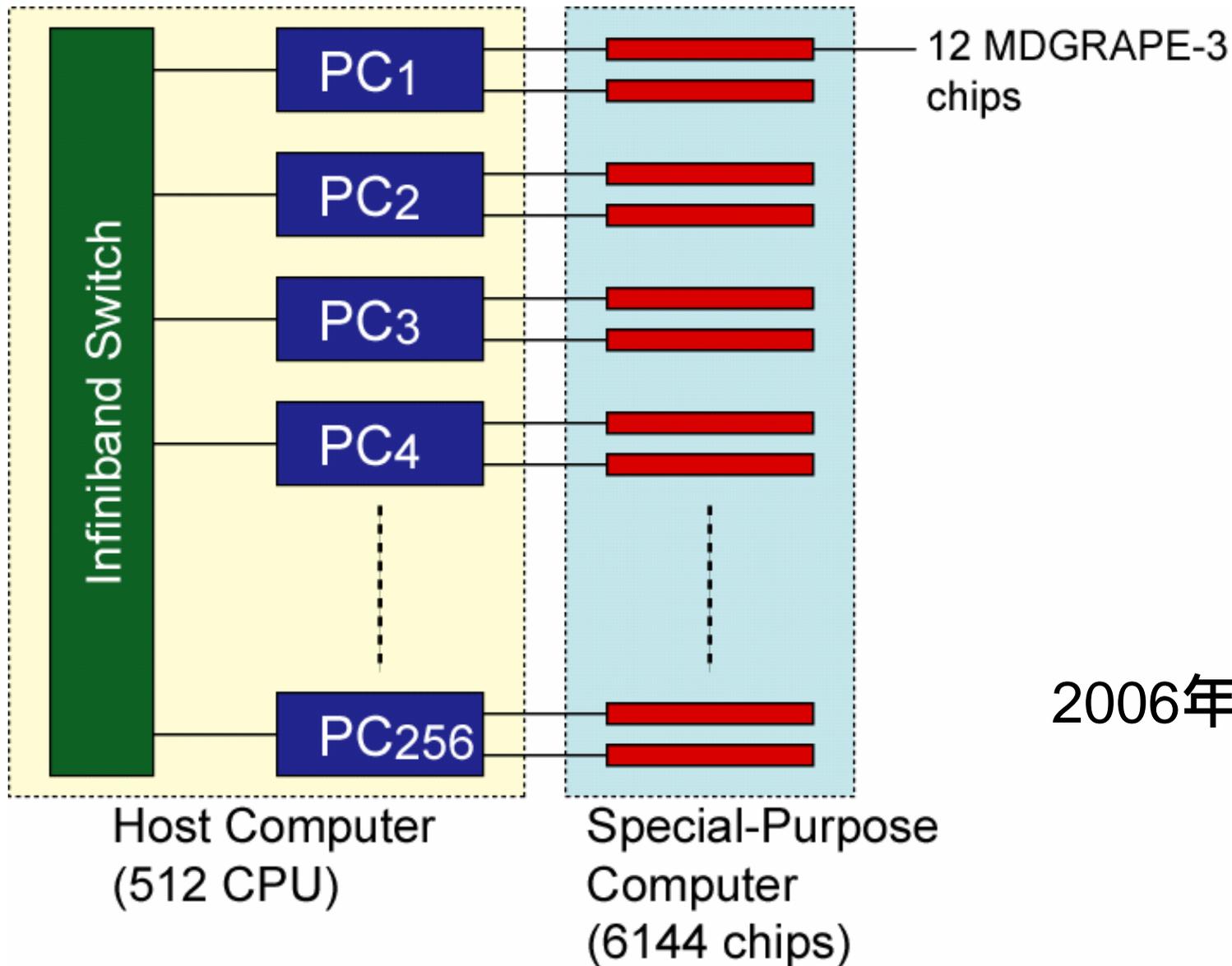
MDGRAPE-3システム

- 12チップ/ボード
- 2ボード/2Uサブラック
= 4 Tflops
- PCI-X接続
- 2チップのPCI-X
ボードも製作



PCI-X board:
460 Gflops

システム構成



2006年3月: 600Tflops

価格・電力性能比

	\$ / Gflops	W / Gflops
MDGRAPE-3	15	0.2
BlueGene/L	300	6
PC Clusters	300	25
Earth Simulator	8000	128
MDGRAPE-2	150	1.5

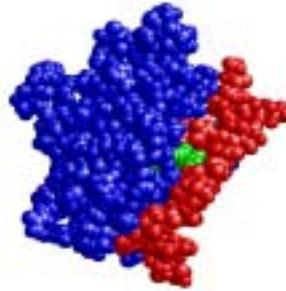
ソフトウェア

- パッケージソフトを移植中
 - Amber 8 (<http://amber.scripps.edu>)
 - CHARMM
 - Gromacs
- ユーザープログラムからはサブルーチンパッケージを通して利用
- 移植はそれほど難しくない

MDGRAPE-3で可能になること

タンパク質構造

X線結晶構造解析
の加速

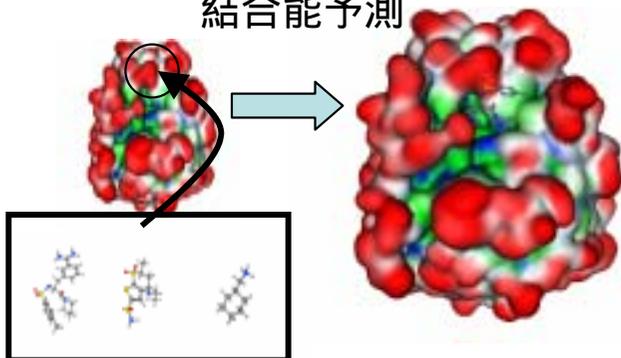


フォールディング

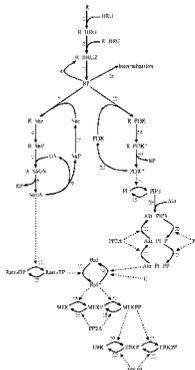
ペタフロップス計算機による分子シミュレーション

創薬

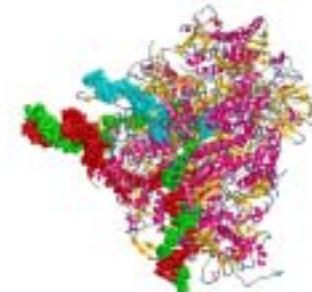
結合能予測



生体分子
ネットワーク

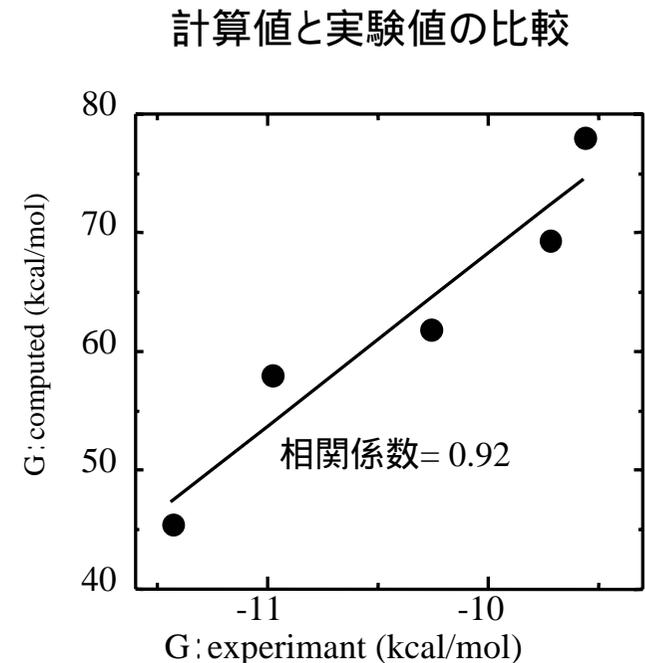


機能解析
発病メカニズム
バイオ・ナノテク



MDGRAPE-3による薬剤設計

- 薬剤とタンパク質の結合エネルギーを精度よく見積もれる
- 通常のドッキングプログラムでは、分子配位は予測できるが結合能は困難な場合も多い
- MDでは、タンパク質・基質・溶媒の動きを取り入れられる
- 原子レベルでの結合様式を見れる
- 薬剤のスクリーニング
ペタフロップスあれば
～500対サンプル/日の能力
- 特に薬剤の最適化に有効
－合成の前に結合能を評価できる



どんなものが専用化に向くか

- 「密な問題」— 計算量がデータ量より多く、
同じデータを使いまわして計算できるもの
 - 古典粒子計算
 - 配列解析(Dynamic Programming)
 - 密行列計算
 - 境界要素法
 - DFT
 - ルジャンドル変換
 - 量子化学計算(一部)
- すごく精度が落とせるもの
- 少ないメモリで高い並列度で計算できるもの

今後の展望：汎用化の方向

- FPU Array / FPGA
- SIMD with broadcast memory (Vector)
 - MACE(NEDO, 統計数理研・理研)
 - GRAPE-DR(振興調整費, 東大中心)
 - Vector Processor Matrix (Proposal)
- 汎用機の限界：今後のHPCにおいて何らかの専用計算機的なものが入ってくるのは必須

Vector Processor Matrix

- PU+レジスタファイルをSIMDで動かす
- そこそこの規模のローカルメモリ
- Target: 特に分子軌道法・密度汎関数法
- 45nmで作って

倍精度512演算/chip, 2GHz, 1Tflops

程度が目標

これで汎用機ではいけないか？

- 汎用機の場合、新デバイス以外決定的な解決策見あたらず
- どうせ汎用機の性能はメモリバンド幅で決まってくる
- 無理しない範囲で、極力バンド幅は上げる
- 基本的には、マルチコアのCPUとして動作
- SIMD並列化の部分は「おまけ」と考える
- CPUの作りこみは少し大変かも
- 動作周波数が問題 ... 専用機の場合、低い周波数で並列度を上げたほうが有利
- 中～長期的な検討が必要

将来(2010年)の専用スーパーコンピュータシステム

- 分子動力学
10Pflops級で、分子動力学による薬剤スクリーニングが実用的に行える
- 量子化学計算
10Pflops級の計算機で、数百残基のタンパク質のナノ秒第一原理分子動力学が可能に
- BlueGene/Q (3-4Pflops?)等に対抗するには、10Pflops以上の性能が必要
- 汎用機10Pflops
- 専用機40Pflops
 - VPM 20Pflops
 - MDGRAPE 20Pflops

汎用機と専用機のバランス

- アプリケーションに依存する
- MDでは汎用:専用=1:100~1:1000
- 汎用:専用 = 1 : 1 ~ 1 : 10ぐらいが現実的か
- コストでは
汎用:専用 = 100 : 1 ~ 10 : 1

開発に向けた課題

- 準汎用機VPM
倍精度1Tflops/chip
- MDGRAPE-4
10Tflops/chip
- これらは半導体プロセス技術が順調に推移すれば実現可能
- 専用機をなるべく汎用機の近くに組み込む