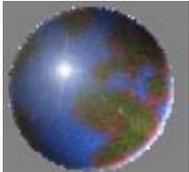


# 有限要素法による 第一原理 分子動力学シミュレーション

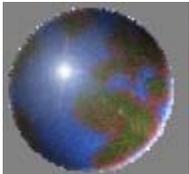
産業技術総合研究所 計算科学研究部門

土田 英二

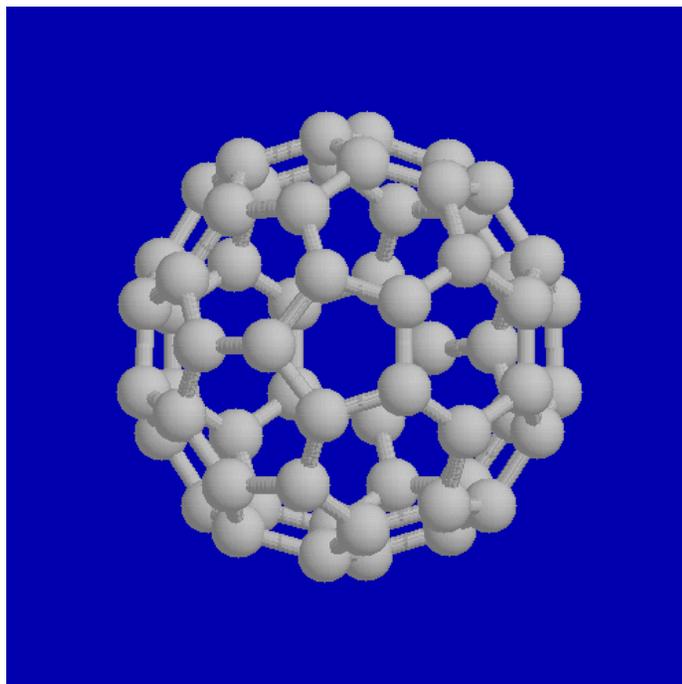


# 第一原理計算とは

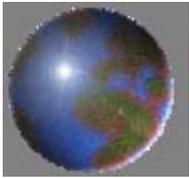
- 原子・分子から固体・液体まで、様々な系を同じ枠組みの中で扱うことができる
- 古典的な経験ポテンシャルと比べて信頼性の面で大幅に優れている
- 計算は非常に重いですが、アルゴリズムの発展と計算機の進歩に伴って急速に発展してきた



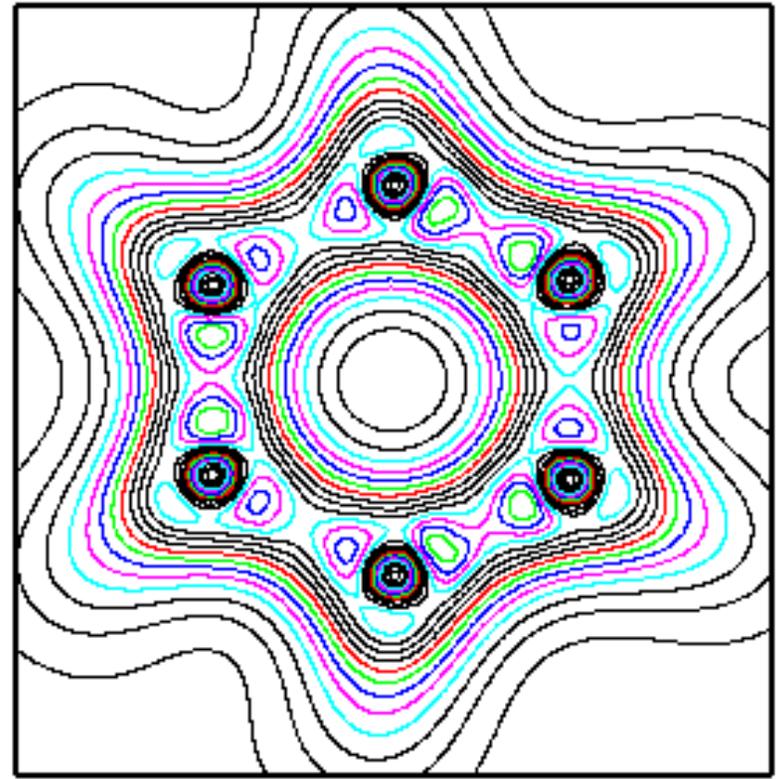
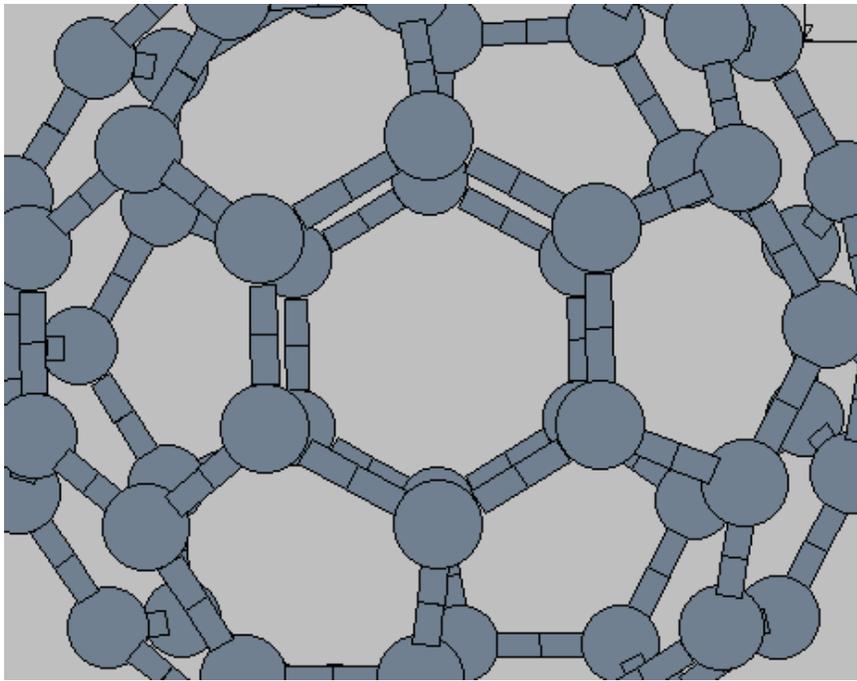
# フラーレン ( $C_{60}$ ) の構造

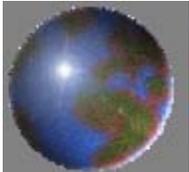


	This work	Other	Exp.
Bond length [Å]	$1.39 \pm 0.003$	1.39	$1.40 \pm 0.015$
	$1.45 \pm 0.002$	1.44	$1.45 \pm 0.015$
Bond angle [degrees]	$108.0 \pm 0.17$	108	
	$120.0 \pm 0.22$	120	



# フラーレンの電子密度





# *Kohn-Sham 方程式 (I)*

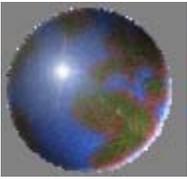
$$\mathcal{H} \psi_i = \epsilon_i \psi_i, \quad i = 1, \dots, M$$

ハミルトニアン

$$\mathcal{H} = -\frac{1}{2} \nabla^2 + V(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \mu_{xc}(n(\mathbf{r}))$$

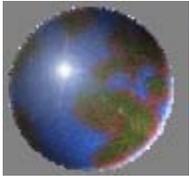
電子密度

$$n(\mathbf{r}) = \sum_{i=1}^M |\psi_i(\mathbf{r})|^2$$



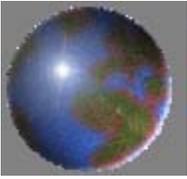
## *Kohn-Sham 方程式 (II)*

- 基本的には、一電子に対するシュレーディンガー方程式と同形
  - 一種の固有値問題に相当
- 原子配置や他の電子の影響は有効ポテンシャルとして反映される



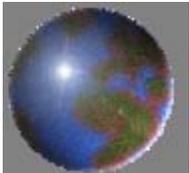
# 第一原理分子動力学計算の流れ

- 1: 原子配置を決める
- 2: ハミルトニアンを計算
- 3: ハミルトニアンに対応する波動関数を計算
- 4: 波動関数を使って原子に働く力を計算する
- 5: 力に応じて原子を動かす → 1へ

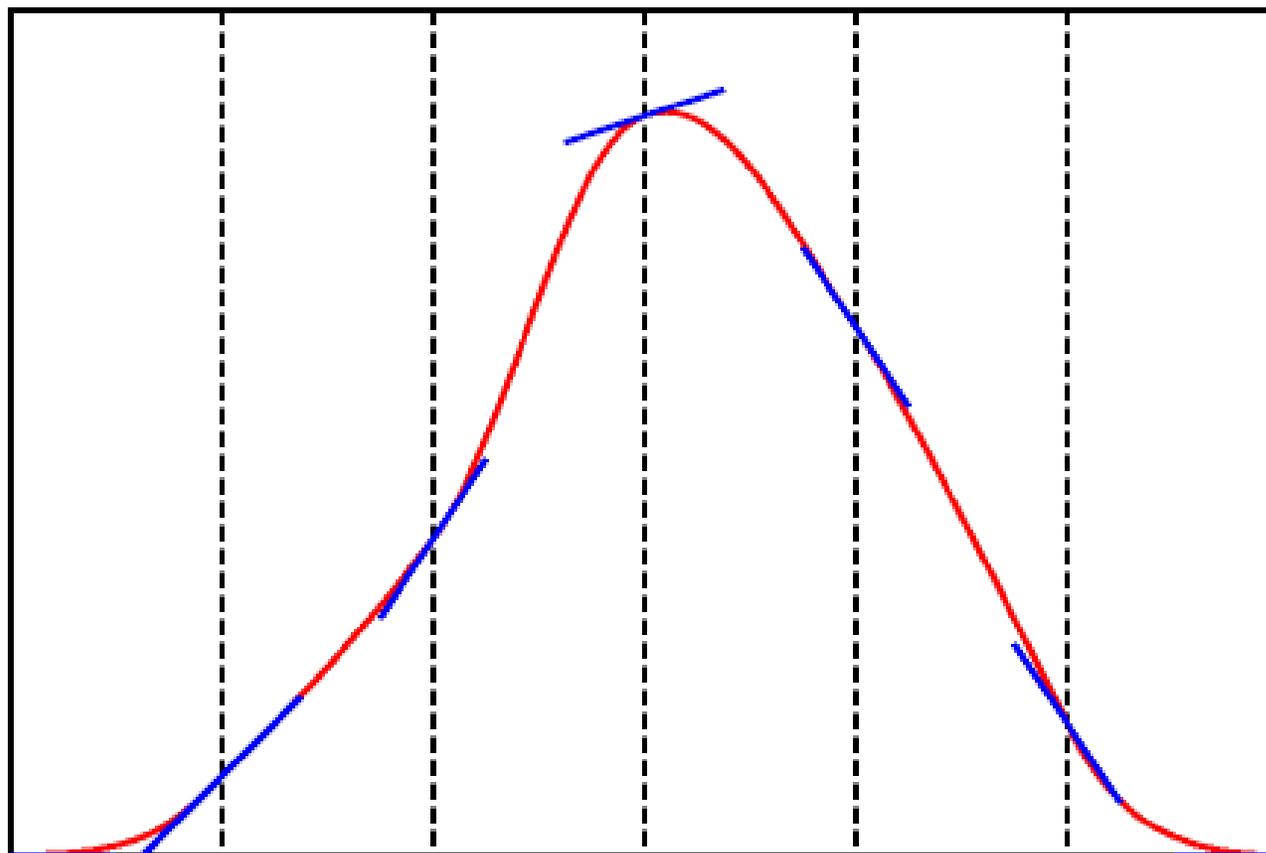


# 基底関数

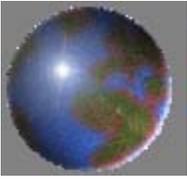
- Kohn-Sham 方程式を解くためには、波動関数  $\psi$  を基底関数で展開する必要がある
  - ハミルトニアンは「行列」になる
- 本研究では**有限要素法**を用いて波動関数を展開
  - 高い並列性を持つ



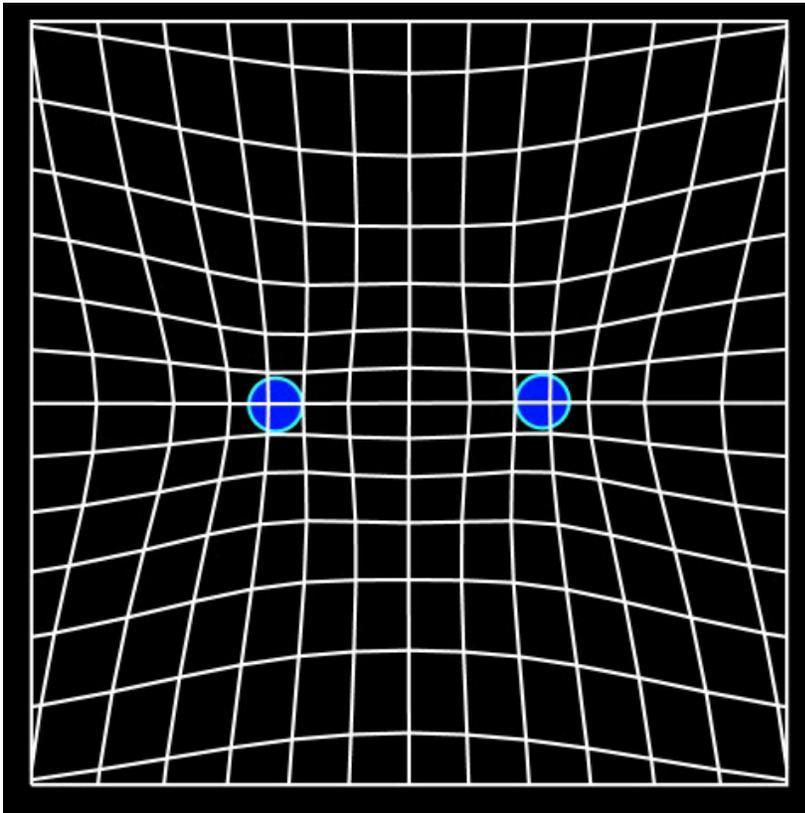
# 三次の有限要素展開



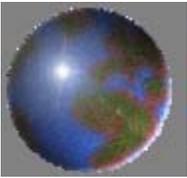
- 各節点での値と微分で決まる(一種のスプライン)



# メッシュの取り方: 曲線座標系

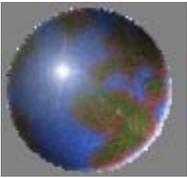


- 原子が複数あるときは線形に足合わせる
- 原子が移動すると座標系も追従する
- 負荷分散を維持できるため、並列化し易い



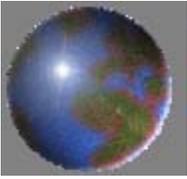
# 電子状態計算の方法 (I)

- ハミルトニアンは巨大な疎行列
  - バンド幅は200程度
- ハミルトニアンの次元を $N$ として、 $N/100 \sim N/1000$  程度の固有値を(下から)求める必要がある
  - $N$  は100万を超えることもある
- 有限要素法の場合は基底関数が非直交なため、一般化固有値問題になる



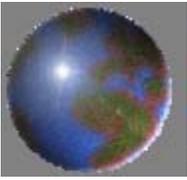
## 電子状態計算の方法 (II)

- 直接ハミルトニアンを変形するのはほぼ不可能なため、反復法で計算する
  - 固有値問題を最適化問題に変換して解く
  - 準ニュートン法の省メモリ版を使用
    - Limited memory BFGS (Liu and Nocedal)
  - 非線形問題に関しては共役勾配法より高速
  - 10~20回程度で収束



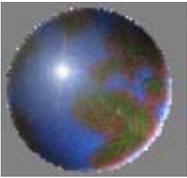
# 並列化の方法

- 実空間を分割し、各領域をプロセッサ毎に割り当てている
  - 必要な通信は主に隣接プロセッサ間での、波動関数等の通信
  - グローバルな通信は最小限で済む
  - 負荷分散もほぼ完全に等しい

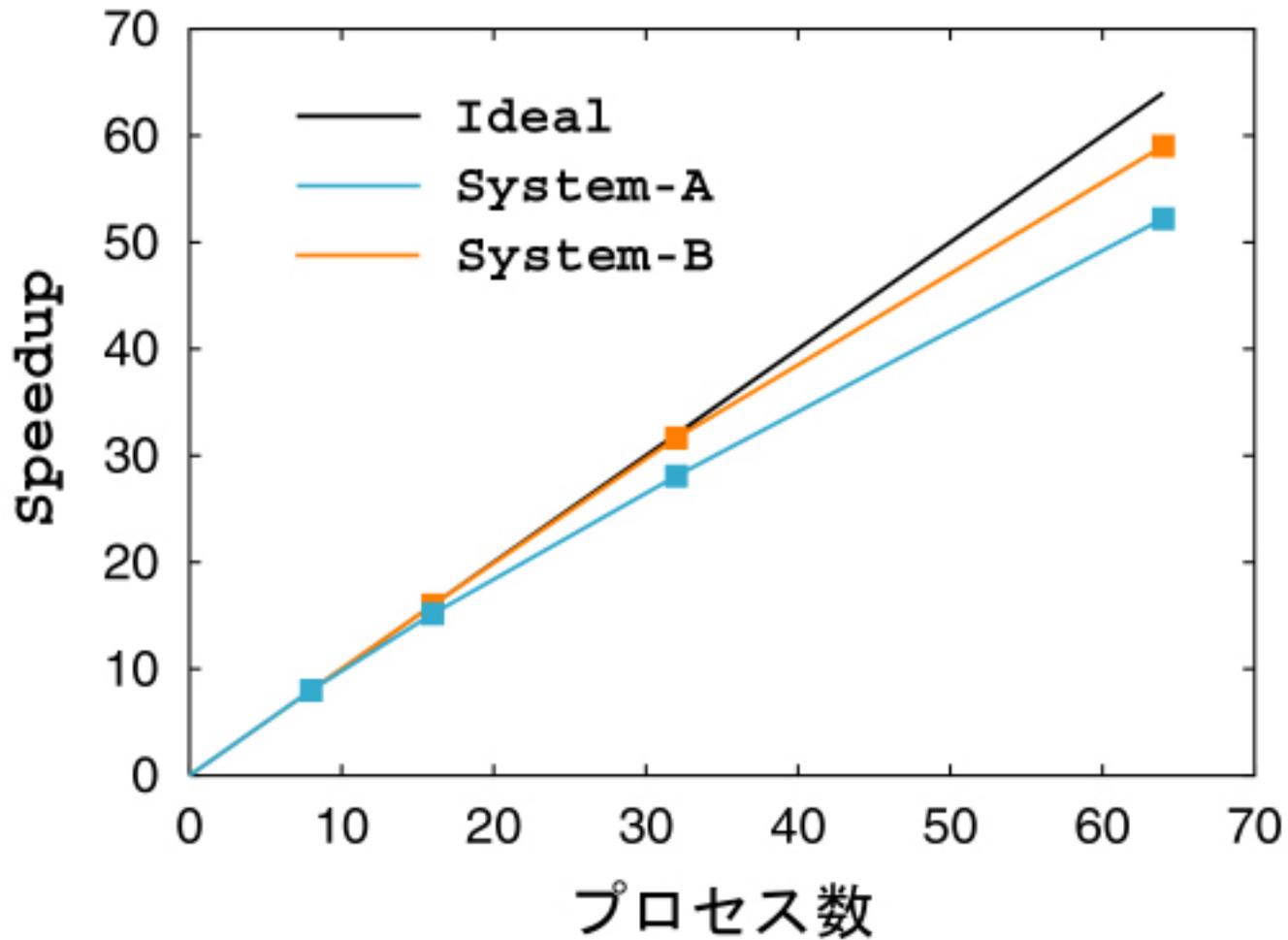


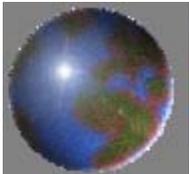
# 並列効率 (I)

- 分子性液体のMD 1ステップにかかる時間を色々な台数の上で測定した
  - NAREGI 分散システムを使用
  - Intel Xeon を Myrinet 接続したPCクラスタ
  - 通信ライブラリは SCore+MPI を使用
  - 1ノード2プロセス→分散メモリとして扱う
  - System-A: 200原子系
  - System-B: 400原子系

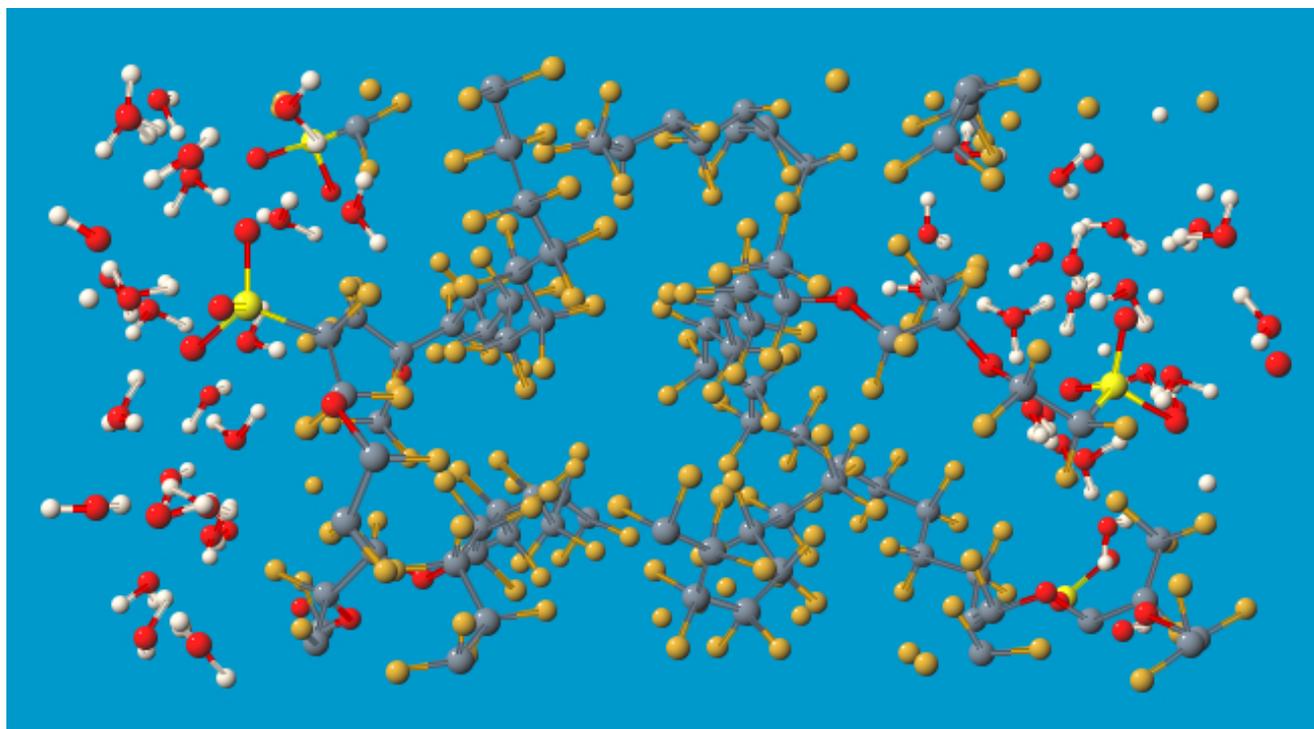


# 並列効率 (II)

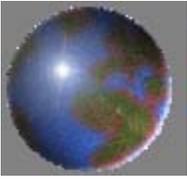




# 最近の計算例(燃料電池)

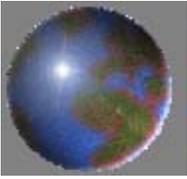


- 産総研(崔・土田・池庄司) + 豊田中研(山川・兵頭)の共同研究



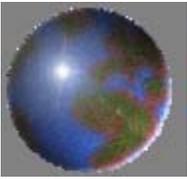
# 計算の詳細

- ナフィオン2本＋水48分子(周期境界条件)
  - 原子数:424、波動関数の数:1022
  - 基底関数の数～50万個
- Opteron クラスタ×128プロセッサを使用
  - 理論ピーク性能 = 512 Gflops
- 約90°Cでプロダクション・ラン
  - 1ステップ毎に約5億個の未知数について最適化
  - 1ステップ当たり5分程度かかる
  - ～20000 ステップの分子動力学計算



# PCクラスタ上での最適化 (I)

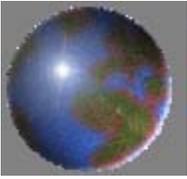
- メモリのバンド幅がスパコンより厳しい
  - なるべく連続アクセス & ブロック化
  - ループは極力レベル3の BLAS で置換
  - SR8Kではピーク性能の60%以上出ていたが、PCではその半分程度？



# PCクラスタ上での最適化 (II)

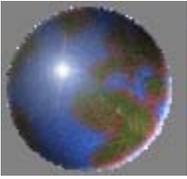
## ● MPI通信による並列化

- 非同期通信 (ISEND/Irecv)、集団通信 (ALLREDUCE/ALLGATHER) が中心
- 行列演算は ScaLAPACK を使用
- レイテンシー／バンド幅はカタログ性能ほど出していない印象 (Myrinet)
- 1対1通信が時々極端に(10倍以上)遅くなる



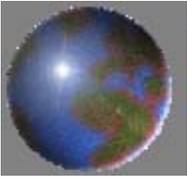
## PCクラスタ上での最適化(III)

- SR-2201 では 1024 並列までそれなりに性能が出たが、PCクラスタでは未確認(最大256並列まで)
- 以前は共有メモリと分散メモリの二階層で並列していたが、制約が多いので現在はMPIで分散並列のみ
  - OpenMP+MPI の方が良い？



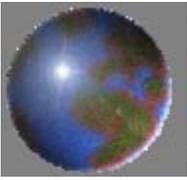
# ASC上での運用(I)

- 現在、AISTスーパークラスタでは一人当たり64ノード(128CPU)までの上限付きで、一週間まで連続して流せるような運用
  - 8～16ノード程度のジョブを複数流すユーザーが多い
  - 本格的な計算は64ノード×数ヶ月程度の資源が必要
  - 申請すればノード数は増やせる



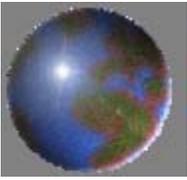
## ASC上での運用(II)

- 基本的にヘビーユーザーはあればあるだけ使うので、「紳士的利用」ではあまりうまく行かない
  - 「公平な資源配分」と「稼働率」の兼ね合いが難しい
- ディスク I/O が弱い: 誰かが大量にファイル入出力を行うと、全員に影響する
  - コンパイルやファイルの編集まで遅くなる...



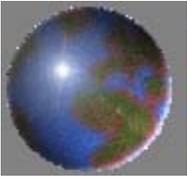
## ASC上での運用(III)

- OS・ミドルウェア・コンパイラ・数値ライブラリ・通信ライブラリの整合性が重要
  - ▣ バージョンを上げる際に問題が起き易い
  - ▣ 独立に開発しているのが問題？
- SCore/MPICH-GM/MPICH-MX を使用
  - ▣ 一週間連続して稼動する確率は低い
  - ▣ ノード数が多いほど落ち易い印象
  - ▣ 通信内容が化けることも...



## ASC上での運用(IV)

- ジョブが落ちた場合にそのままキューに残ることが多い→素直に落ちた方がベター
  - 再実行されるのも不都合(上書きされる)
- 「チェックポイント・リスタート」機能は(個人的には)不要
  - メモリを上限近くまで使うことが多いため、コストの方が大きい
  - 通常はアプリの方で対応している



# T2K上での大規模計算への期待

- 10TFlops × 1ヶ月程度専有できれば、4～5000原子程度の本格的な第一原理分子動力学計算が可能になると期待
  - 前例のないスケール
  - 燃料電池やナノテクに関連した新素材の開発に貢献: 実験サイドの期待も大きい
  - 安定性や並列性能が課題