

CMSI計算科学技術特論A

古典分子動力学計算法の高速化

2015年7月2日

名古屋大学大学院工学研究科
附属計算科学連携教育研究センター
吉井範行



MD計算の要素 各要素の高速化、並列化

MD計算とは

- MD計算の流れ、取り扱う原子数と周期境界条件、相互作用(力)、分子モデルと運動方程式
- MD計算の例、歴史、現状

非並列のMD計算

- 短距離相互作用 (bookkeeping, cell index法)
- 長距離相互作用(クーロン相互作用) Ewald, PME法
- 効率化のための方法
 - 拘束動力学 (Δt を大きくする)
 - Multiple time step (Δt を大きくする)
 - On the flyによる最適化 (収束計算をなくす)

高速化のために、キャッシュの有効利用、キャッシュミス

MD計算の並列化

- 高速多重極展開法(FMM)

データ構造 メタデータ法、通信量の最小化、データ局在化による計算効率の向上

ベンチマーク

MODYLAS

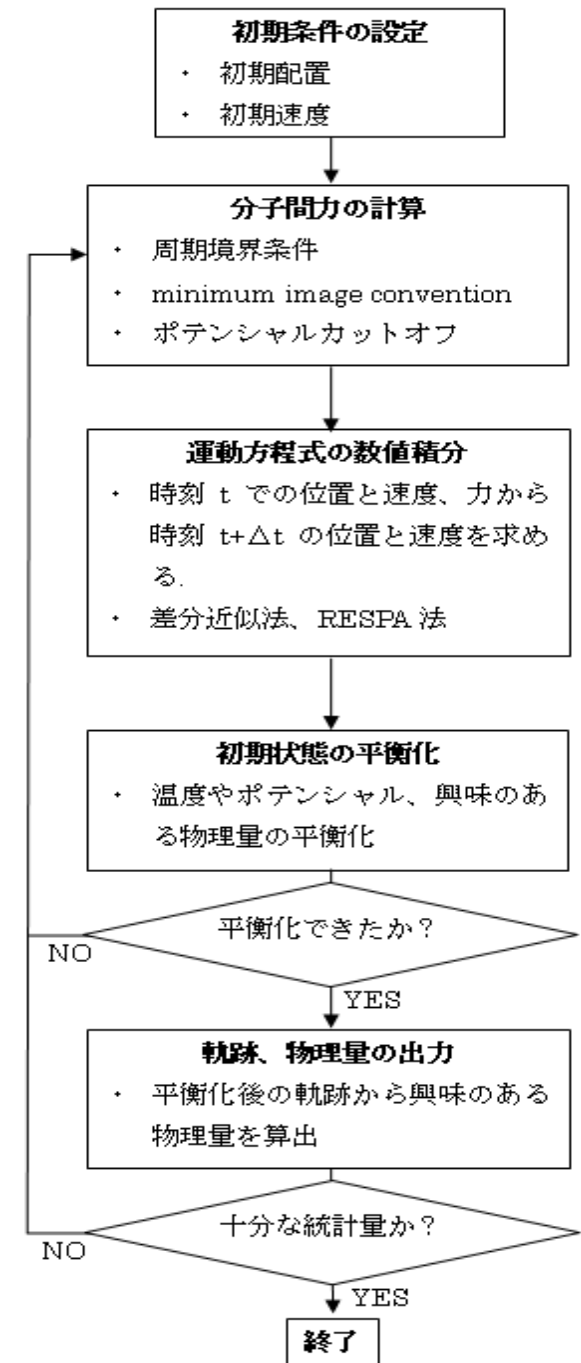
分子動力学(MD)計算とは



- 多数の原子や分子が集まってできた分子集団系
- 原子、分子、もっと大きな塊の力を計算
- 運動方程式を数値積分し、運動を追跡
- 得られた原子、分子、もっと大きな塊の軌跡から熱力学、統計力学的性質を抽出

MD計算の流れ

- 初期条件の設定
- 分子間力の計算
- 運動方程式の数値積分
時間刻み Δt
- 初期状態の平衡化
- 平衡化できたか？
- 軌跡、物理量の出力
- 十分に統計はとれたか？
- 最終座標・速度の出力



取り扱う原子数と周期境界条件

- ・原子数

100~100,000,000 個（長距離相互作用のある場合）
対象系の規模や物理現象による。

- ・基本セルの一辺の長さ

- ・周期境界条件

基本セル（立方体、直方体、平行六面体etc）
イメージセルを基本セルの周囲に配置。

→表面効果の除去

3次元のみでなく、2, 1次元方向にイメージセルを配置

2次元: 界面、平面膜、2枚の平板にはさまれた状態

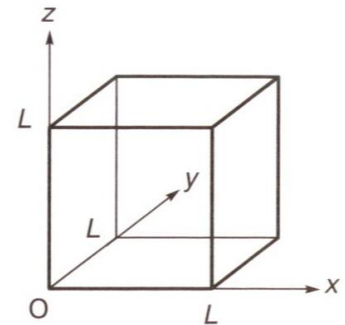
1次元: 細孔、棒状分子

- ・ポテンシャルカット

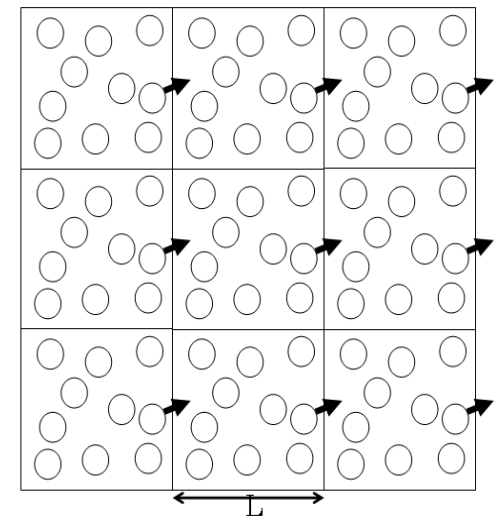
Lennard-Jones相互作用

Ewald法の実空間部分

通常、原子直径の3倍(3σ)程度。



立方体の基本セル



相互作用(力)

- 分子間相互作用(力)

2体力の和 $V_{\text{intermolecule}}(\mathbf{r}^N) = \sum_{i>j} \sum V(r_{ij})$

Lennard-Jones(LJ)相互作用

$$V_{\text{LJ}}(r) = 4\varepsilon \left\{ \left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right\}$$

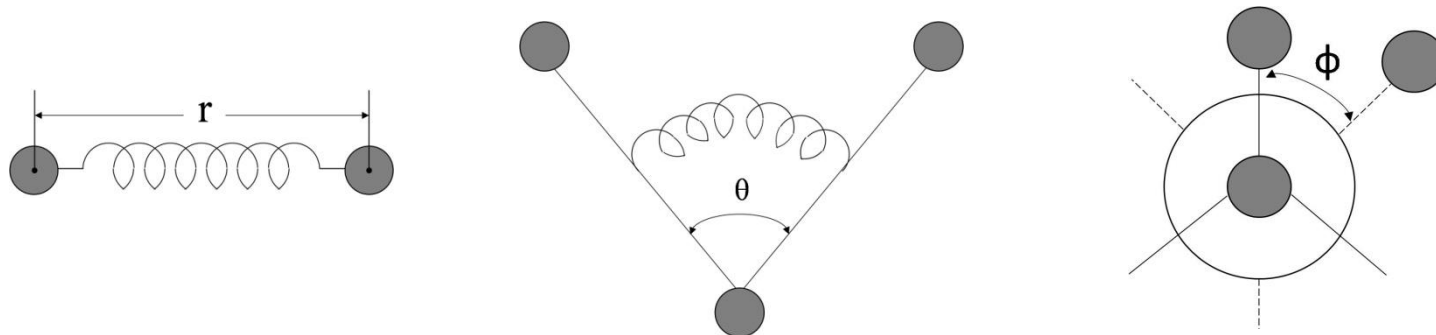
クーロン相互作用 $V_{\text{Coulomb}}(r) = \frac{q_1 q_2}{r}$

- 分子内相互作用(力)

分子内の原子を剛体的に取り扱わず、フレキシブルに運動できるとする。(フレキシブルモデル)
分子内ポテンシャルを伸縮、変角、torsionなどで表現する。

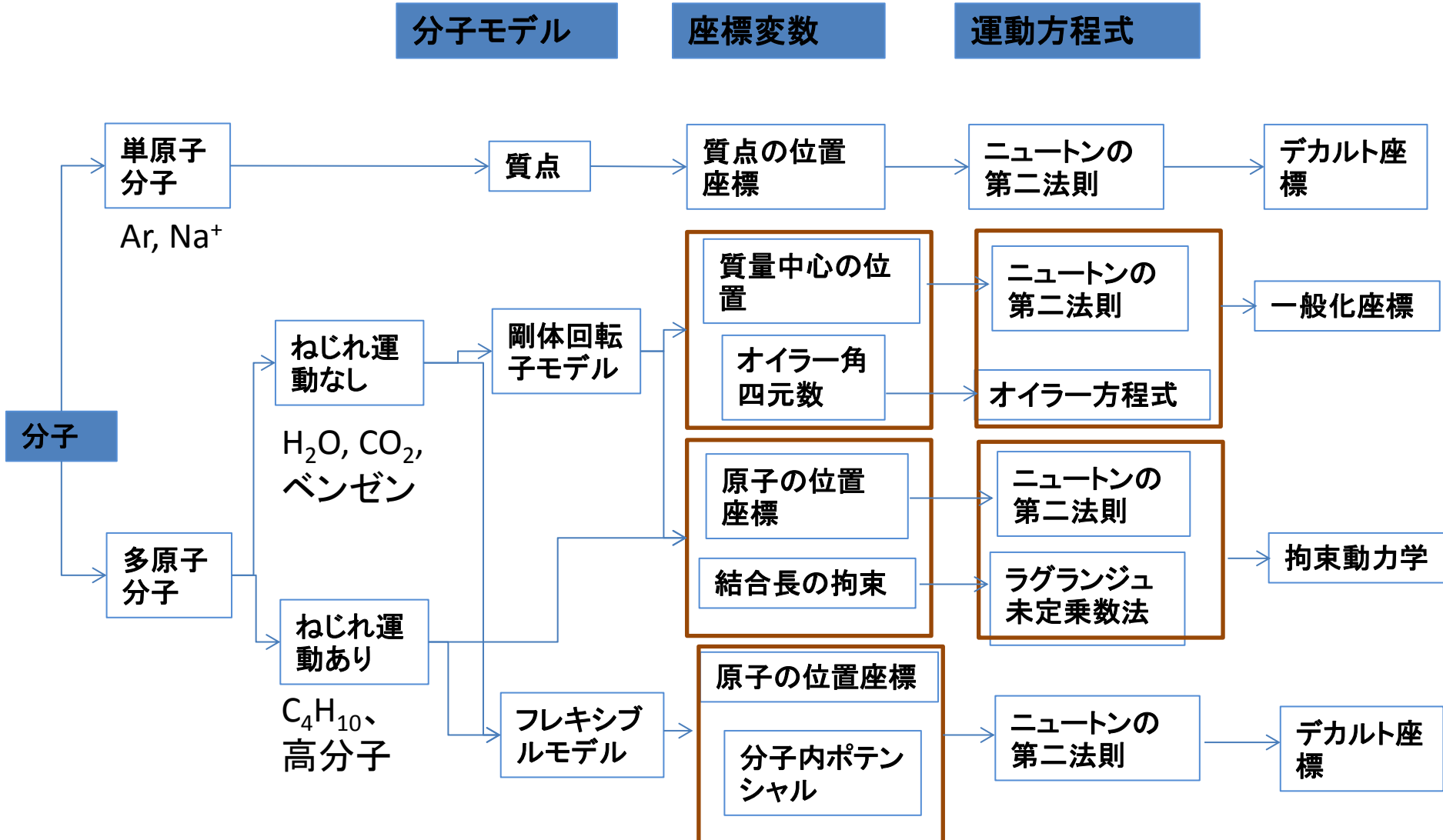
$$V_{\text{intramolecule}} = V_{\text{st}} + V_{\text{be}} + V_{\text{tor}}$$

$$= \sum k^{\text{st}} (r - r_0)^2 + \sum k^{\text{be}} (\theta - \theta_0)^2 + \sum_n \frac{1}{2} k_n^{\text{tor}} \{1 + \cos(n\phi - \phi_0)\}$$

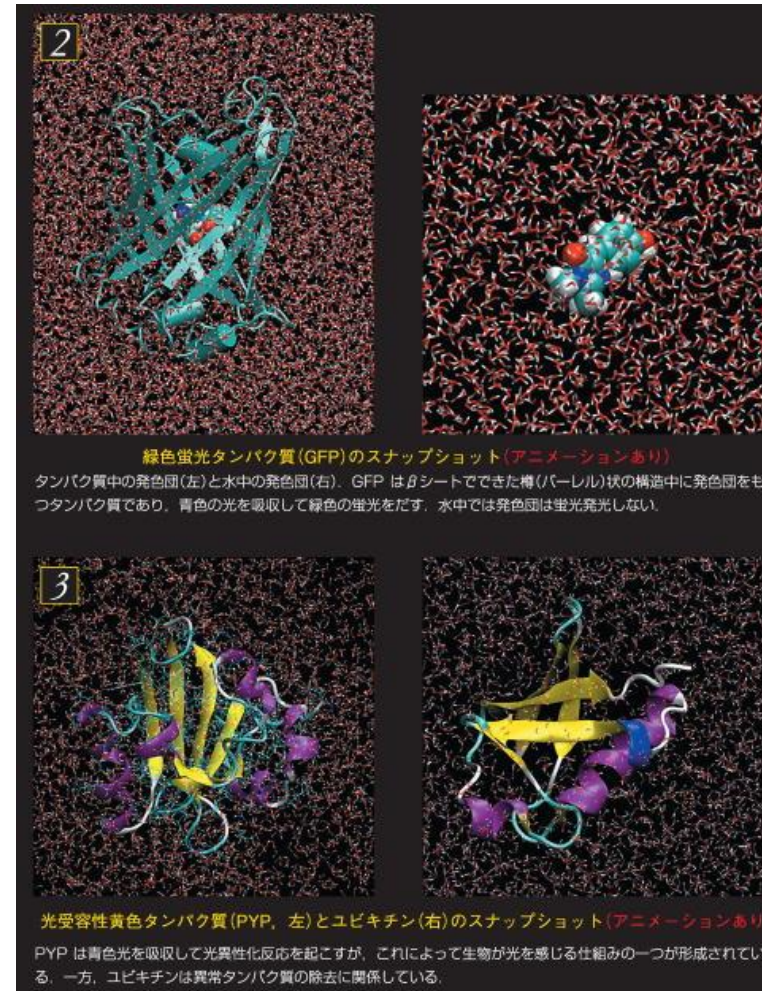
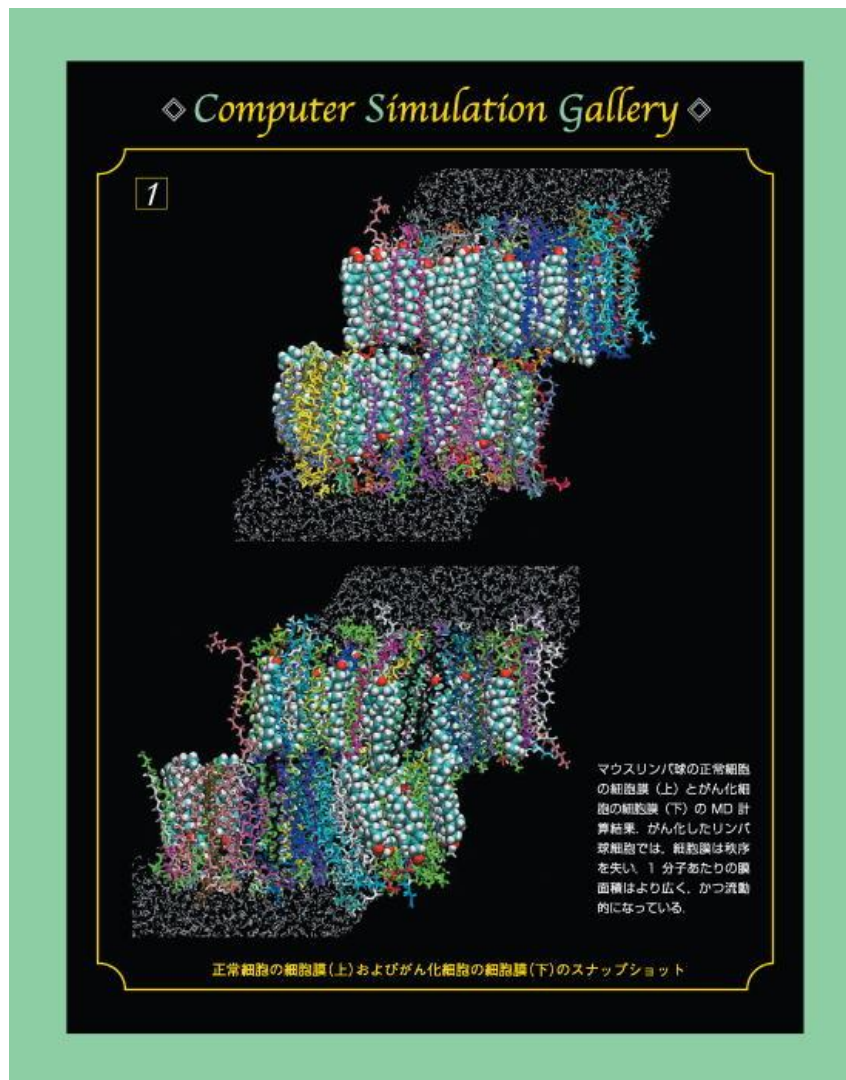


分子間相互作用の精度がMD計算の質を決める

分子モデルと運動方程式



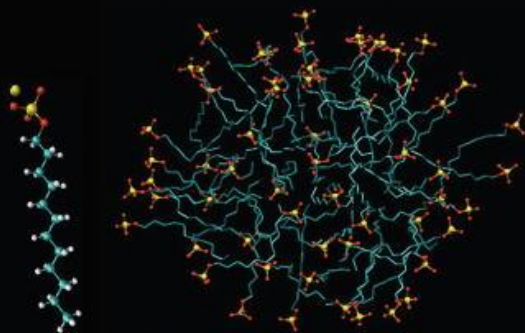
MD計算の例



MD計算の例



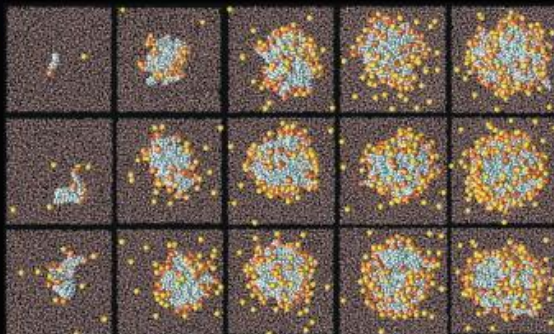
4



乱れたミセル構造のスナップショット

ドデシル硫酸ナトリウム(SDS)の分子構造(左)および SDS ミセルの MD 計算結果(右)。親水基を赤と黄色の球で、また疎水基を青色で示す。

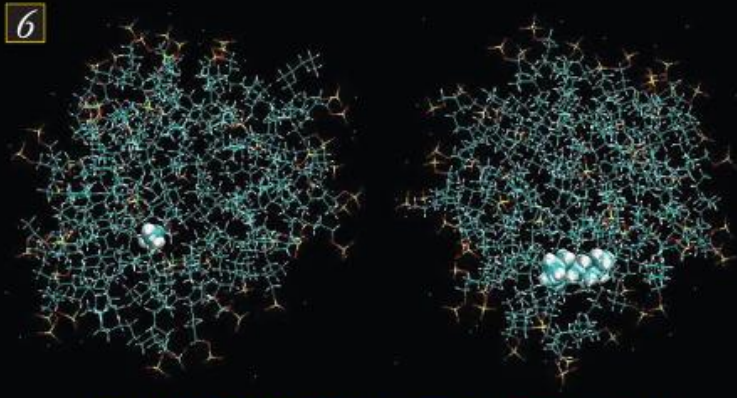
5



さまざまな大きさの球状ミセルのスナップショット

球状ミセル(SDS)に対する自由エネルギーの MD 計算から、水溶液中では約 60 個の分子からなる球状ミセルが安定に形成されることがわかった(下段中央)。

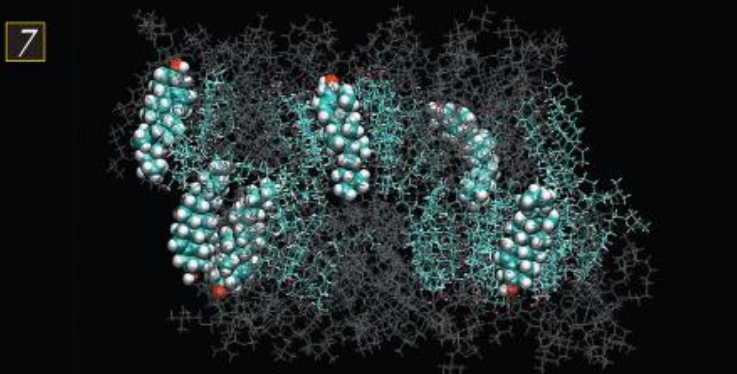
6



ミセルによる不溶性溶質の可溶性シミュレーション (アニメーションあり)

SDS 中に可溶性された不溶性溶質のメタン分子 (左) とオクタン分子 (右)。自由エネルギーの MD 計算から、水中よりもミセル中のほうが自由エネルギー的に安定であることが示される。

7

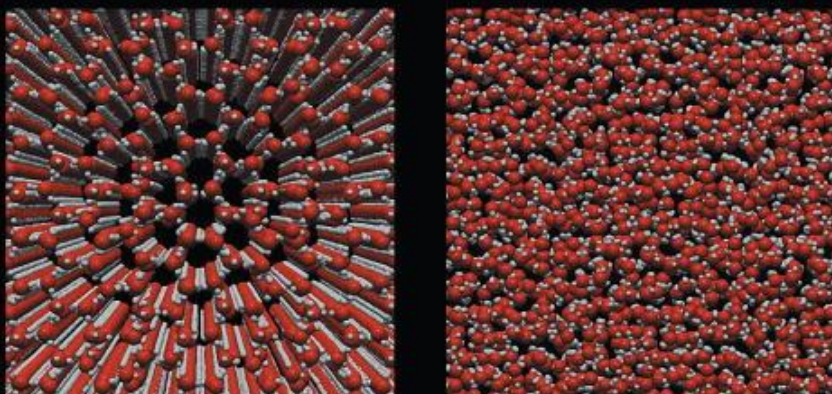


混合脂質膜のスナップショット (アニメーションあり)

ジパルミトイルフォスファチジルコリン (DPPC) + コレステロール (空間充填モデルで表示) 脂質混合系の MD 計算結果。コレステロールの添加により、膜はより秩序化し固くなる。

MD計算の例

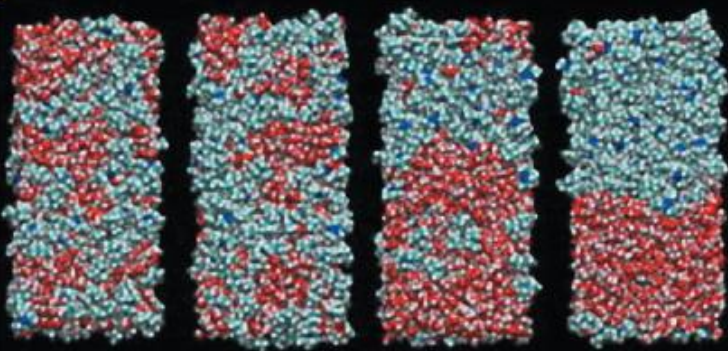
8



100万原子系のMD計算 (アニメーションあり)

氷(左)と水(右)。ここに示された原子数は少ないが、アニメーションは100万原子系に対するものである。現時点で、100万原子系のMD計算は一般的ではないが、数年後には日常的に計算されるようになるだろう。

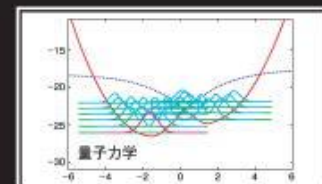
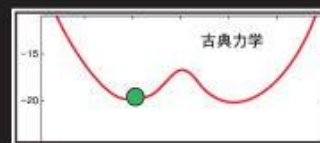
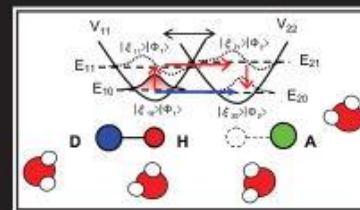
9



トリエチルアミン水溶液の相分離のスナップショット (アニメーションあり)

トリエチルアミン水溶液は下部臨界溶解温度をもつ二成分系であり、低温で完全に混合していた溶液(左端)は、時間の経過(右方向)とともに相分離する(右端)。水(赤色)、トリエチルアミン(水色)。

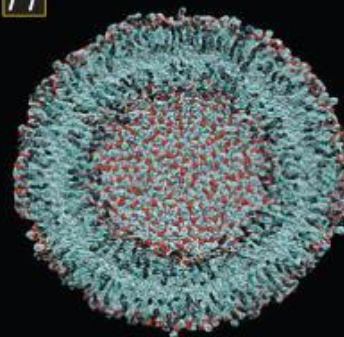
10



プロトン移動の量子動力学シミュレーション (アニメーションあり)

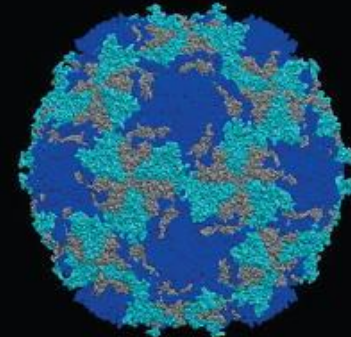
溶液中におけるマロンアルデヒドの分子内プロトン移動。量子効果が大きく、トンネル移動や振動励起、振動緩和など、運動は量子力学的に取り扱わなければならない。

11



リポソームの全原子MD計算によるスナップショット

水中に生成される球状の脂質二重層膜。多層のラメラ相と異なり、一枚膜である。計算は進行中で、300万原子系のMD計算が必要となる。



ウイルスカプシドの全原子MD計算によるスナップショット

小児マヒウイルスカプシド(タンパク質でできた球状の殻)。このなかにRNAが収められている。1000万原子系のMD計算が必要となる。現在、「次世代スパコン」を用いた計算が計画されている。



MD計算の例

対象	系・現象	対象	系・現象
低分子液体、溶液	水、水溶液、 溶媒和、化学反応	タンパク質	折りたたみ、酵素、イオンチャネル、阻害剤、タンパク質複合体
固体	相転移、フォノン	DNA、RNA	
不均一系、界面	結晶成長、吸着、トライボロジー、蒸発、凝縮、透過、分配、界面増強スペクトル、界面張力	生体膜	脂質膜、ラフト、抗生物質、薬物透過、吸収
両親媒性分子水溶液 ミセル、単分子膜、二分子膜、バイセル	球状ミセル、棒状ミセル、 L膜、LB膜、ベシクル	ウイルス	カプシドタンパク質
化学反応	溶液内化学反応、 電解反応、 自由エネルギー曲面	ガラス	ガラス化、スローダイナミクス
超臨界流体	分子系、混合系、クラスターの生成	金属、半導体	電極、液体金属
高分子	分離膜、電解質膜、複雑系、スローダイナミクス	非平衡マクロ動力学	対流、ずり、核生成、流体力学

MD計算の歴史



発表年	手法	対象	発表者
1953	MC	初めてのMC、剛体球	Metropolis et al.
1957	MC	球状分子	Wood and Parker
1957	MD	初めてのMD、剛体球	Alder and Wainwright
1964	MD	球状分子	Rahman
1969	MC	水	Barker and Watts
1971	MD	水(回転の運動方程式)	Rahman and Stillinger
1971	MC	イオン	Woodcock
1977	MD	拘束動力学	Rychaert et al.
1977	MD	タンパク質分子	McCammon et al.
1980	MD	圧力一定	Andersen
1980	MD	圧力テンソル一定	Parrinello and Rahman
1981	MD	経路積分、量子系	Chandler and Wolynes
1984	MD	温度一定	能勢
1985	MD	第一原理MD	Car and Parrinello
1986	MD	溶液内化学反応	Hynes et al.
1987	MD, MC	FMM	Greengard
1988	MD, MC	汎用ポテンシャル	Jorgensen et al.
1991	MD	非断熱量子動力学	Rosky et al.

MDの現状



- ・汎用ポテンシャルセットの開発
対象系の分子のポテンシャルが簡単に構築できる
- ・粗視化モデル
大規模系が取扱い可能に

ポテンシャル名	開発者
AMBER	Kollman
CHARMM	Karplus
OPLS	Jorgensen
GROMOS	Berendsen

プログラム名	開発者
AMBER	Kollman
CHARMM	Karplus
GROMACS	Hess
TINKER	Ponder
NAMD	Schulten
DESMOND	Shaw
LAMMPS	Plimpton
MODYLAS	Okazaki
GENESIS	Sugita
MRABLE	Ikeguchi



非並列のMD計算

シングルコアで高速化を進める

MD計算手法の選択

短距離力の計算 (bookkeeping, cell index)

長距離力の計算 (Ewald, PME, FMM)

拘束動力学の利用

Multiple time stepの導入

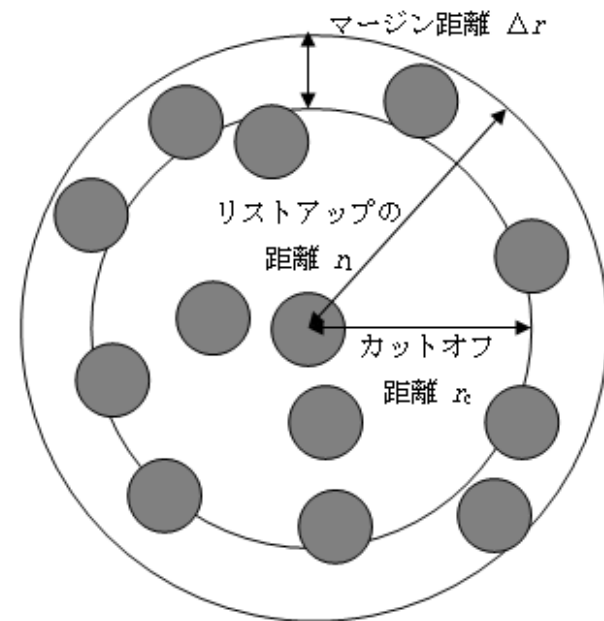
on the fly計算の利用

- 高速化のための工夫

短距離相互作用 bookkeeping法



- ある分子からの距離が r_c よりも短い相手分子の番号すべてをリストアップ。
- リストの作成は何ステップかに1回毎。
固体のような流動性のないものなら更新不要
- r_c よりも Δr だけ長い距離のものまで余分にリストに載せる。
- リストを用いて相互作用を計算する際には、リストを複数のノードやスレッドに割り振って並列計算させるループ分割法が容易に実装可能。



帳簿法.

L	N	N'	L/r_1	N'/N
9σ	600	150	2.6	0.25
12σ	1500	150	3.4	0.10
15σ	3000	150	4.3	0.05
18σ	5000	150	5.1	0.03



```
k=0
do i=1,n-1
  do j=i+1,n
    i, j間の距離が $r_l$ 以下
    k=k+1
    list(1,k)=i   リスト作成
    list(j,k)=j   リスト作成
  enddo
enddo
```

```
do i=1,k
  リストをもとに相互作用計算
enddo
```

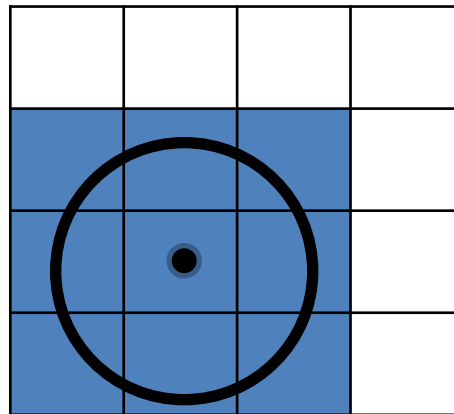
MPI並列なら
相互作用の計算結果をノード間で通信
スレッド並列
通信不要

- 数ステップに一度相互作用対のリスト(list)を更新する
- 毎ステップリストを用いて相互作用計算する。

短距離相互作用 Cell index法



- 帳簿を作成, 更新する際の $N(N-1)/2$ 通りの距離計算が大きな負荷. 大きなメモリを要する.
- L/r_i が4以上の場合には, 一辺の長さが r_i よりも大きくなる範囲内で基本セルそのものをx, y, zそれぞれの方向に4分割以上の分割を行う.
- 注目している分子が属しているセルとそれに直接接しているセルあわせて27個の分割セル内にある分子との間だけでbookkeeping法を行う.
- 分割セルの一辺の長さをよりも短くする場合には, 第二近接のセルまで考慮に入れる
- Cell index法についても作成したリストを複数のノードやスレッドに割り振って相互作用の並列計算を行えばよい.



長距離相互作用 クーロン相互作用



- 直接計算 $O(N^2)$
→計算方法はLJ相互作用と同じ
カットオフの悪影響大のため用いない。
- エwald法 $O(N^2)$ あるいは
Particle Mesh Ewald法 $O(N\log N)$
→小、中規模系向け
- 高速多重極展開法 (Fast Multipole Method: FMM) $O(N)$
→大規模系向け

長距離相互作用 Ewald法



Ewald法の要点

- ・減衰の遅いクーロン相互作用を、減衰の速い関数と遅い関数に分ける。
- ・遅い関数はフーリエ変換することにより逆格子ベクトルを用いて評価。
- ・速い関数は実空間で評価。

$$V_N(\mathbf{r}^N) = \frac{1}{2} \sum_n \sum_i \sum_j' \frac{1}{4\pi\epsilon_0} \frac{Q_i Q_j}{|\mathbf{r}_i - \mathbf{r}_j + L\mathbf{n}|}$$

(減衰の遅い $1/r$)

$$= \begin{array}{l} \text{(減衰の速い関数)} \\ + \text{(減衰の遅い関数)} \end{array} \quad \frac{1}{r} = \frac{\text{erf}(\alpha r)}{r} + \frac{\text{erfc}(\alpha r)}{r}$$

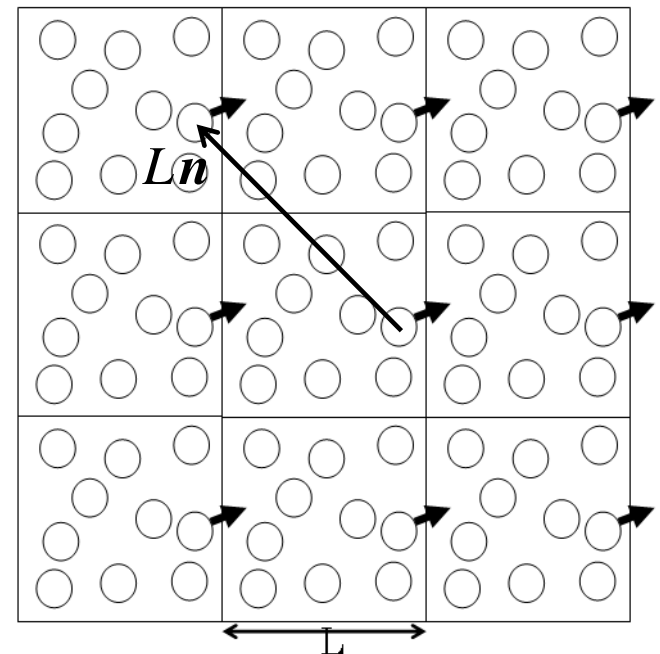
減衰の速い関数 ... この関数形で評価

$$\frac{\text{erfc}(\alpha r)}{r}$$

減衰の遅い関数 ... フーリエ変換したもので

$$\frac{\text{erf}(\alpha r)}{r}$$

評価



\mathbf{n} 整数ベクトル $(-1, 2, 3)$

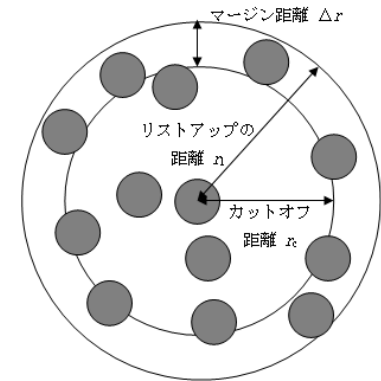
長距離相互作用 Ewald法



• Ewald法によるポテンシャル

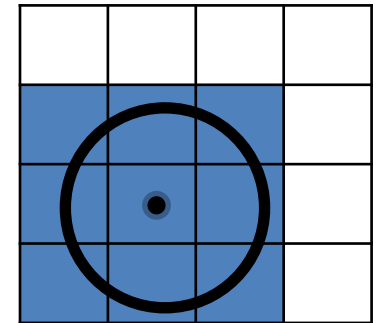
$$V_N = V_N^{(1)} + V_N^{(2)} + V_N^{(3)}$$

bookkeeping法
セルインデックス法



$$V_N^{(1)} = \frac{1}{2} \sum_n \sum_i \sum_j \frac{Q_i Q_j \operatorname{erfc}(\alpha |r_i - r_j + Ln|)}{4\pi\epsilon_0 |r_i - r_j + Ln|}$$

実空間での相互作用
Lと一緒に計算する



$$V_N^{(2)} = \frac{1}{2\epsilon_0 L^3} \sum_{\mathbf{G} \neq 0} \frac{\exp(-|\mathbf{G}|^2 / 4\alpha^2)}{|\mathbf{G}|^2} \left[\left\{ \sum_i Q_i \cos(\mathbf{G} \cdot \mathbf{r}_i) \right\}^2 + \left\{ \sum_i Q_i \sin(\mathbf{G} \cdot \mathbf{r}_i) \right\}^2 \right]$$

Gについてのループ分割

→ PMEによる高速化

$$V_N^{(3)} = -\sum_i \frac{Q_i^2}{4\pi\epsilon_0} \frac{\alpha}{\sqrt{\pi}}$$

Particle mesh Ewald (PME) 法



逆格子ベクトルの項は

$$\begin{aligned} V_N^{(2)} &= \frac{1}{2\varepsilon_0 L^3} \sum_{\mathbf{G} \neq 0} \frac{\exp\left(-\frac{|\mathbf{G}|^2}{4\alpha^2}\right)}{|\mathbf{G}|^2} \sum_i Q_i \exp(i\mathbf{G} \cdot \mathbf{r}_i) \sum_j Q_j \exp(-i\mathbf{G} \cdot \mathbf{r}_j) \\ &= \frac{1}{2\varepsilon_0 L^3} \sum_{\mathbf{G} \neq 0} H(\mathbf{G}) S(\mathbf{G}) S(-\mathbf{G}) \end{aligned}$$

$$H(\mathbf{G}) = \frac{\exp\left(-\frac{|\mathbf{G}|^2}{4\alpha^2}\right)}{|\mathbf{G}|^2}$$

$$S(\mathbf{G}) = \sum_i Q_i \exp(i\mathbf{G} \cdot \mathbf{r}_i)$$

$S(\mathbf{G})$ を毎ステップ計算しなければならない。
 $S(\mathbf{G})$ を格子点 \mathbf{u} における補間を用いて表現すると

$$S(\mathbf{G}) = \sum_{\mathbf{u}} \hat{S}(\mathbf{u}) \exp(i\mathbf{G} \cdot \mathbf{u})$$

となる。ここで $S(\mathbf{u})$ は $S(\mathbf{G})$ の \mathbf{u} における重み因子。これにより

$$V_N^{(2)} = \frac{1}{2\varepsilon_0 L^3} \sum_{\mathbf{u}} \hat{H} * \hat{S}(\mathbf{u}) \hat{S}(\mathbf{u})$$

$S(\mathbf{G})$ の計算に補間法を利用(格子点で評価)
高速フーリエ変換(FFT)利用 $O(N \log N)$

効率化のための方法



Δt を大きくする

- 拘束動力学 (⇔剛体回転子)
- Multiple time step

収束計算をなくす

- On the fly による最適化

拘束動力学

• 拘束条件付きの運動方程式

拘束条件

$$g = (\mathbf{r}_1 - \mathbf{r}_2)^2 - d_{12}^2 = 0$$

運動方程式

$$m_1 \ddot{\mathbf{r}}_1 = \mathbf{F}_1 + \lambda \frac{\partial g}{\partial \mathbf{r}_1}$$



ラグランジアン

$$L = \sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 - V(\mathbf{r}^N)$$

オイラー・ラグランジュ方程式

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{r}}_1} \right) - \frac{\partial L}{\partial \mathbf{r}_1} = 0$$

運動方程式

$$m_1 \ddot{\mathbf{r}}_1 = \mathbf{F}_1$$

拘束力のある場合

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{r}}_1} \right) - \frac{\partial L}{\partial \mathbf{r}_1} = \lambda \frac{\partial g}{\partial \mathbf{r}_1} \quad \text{ラグランジュの未定乗数}\lambda$$

$$\frac{\partial g}{\partial \mathbf{r}_1} = \begin{cases} 2\{\mathbf{r}_1(t) - \mathbf{r}_2(t)\} & \text{(拘束に原子1が関与するとき)} \\ 0 & \text{(そうでないとき)} \end{cases}$$

運動方程式

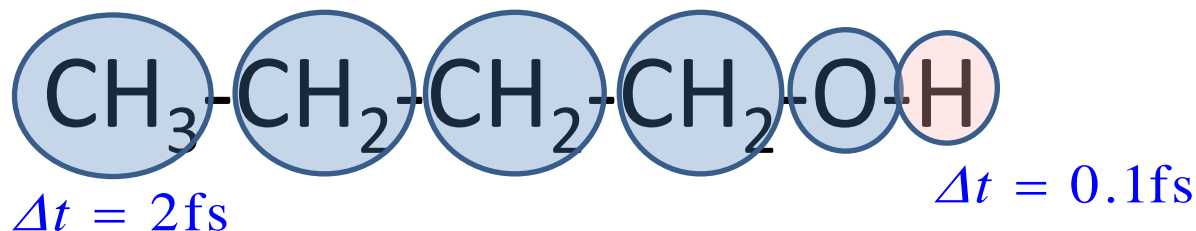
$$m_1 \ddot{\mathbf{r}}_1 = \mathbf{F}_1 + 2\lambda \{\mathbf{r}_1(t) - \mathbf{r}_2(t)\}$$

Multiple time step

- 同じMD計算のなかで変化の速いものと遅いものがある場合には、速い変化のほうに対応した**小さな Δt を選択しなければならない**.

例

United atomのブチルアルコールのフレキシブルモデル



- Hに対する運動方程式とほかの重い原子に対する運動方程式を区別して、**異なった Δt で解くことができれば非常に都合のよい(部分的に Δt を大きくできる)**ことになる。この考え方に基づいた方法がMultiple time step法である。



Multiple time step

- 時間発展演算子法により、重い(遅い)運動と軽い(速い)運動とを分離して数値積分する。
- 時間発展演算子法

リウビル演算子 L

$$iL = iL_r + iL_p = v \frac{\partial}{\partial r} + \frac{F}{m} \frac{\partial}{\partial v}$$

時間発展演算子

$$e^{iL\Delta t} \quad q_i(\Delta t) = e^{iL\Delta t} q_i(0)$$

異なる自由度など2つの部分に分離できると

$$e^{iL\Delta t} = e^{iL_p\Delta t/2} e^{iL_r\Delta t} e^{iL_p\Delta t/2} + O(\Delta t^3)$$

$$r_i(\Delta t) = e^{iL_p\Delta t/2} e^{iL_r\Delta t} e^{iL_p\Delta t/2} r_i(0)$$



Multiple time step

- 時間発展演算子法により、重い(遅い)運動と軽い(速い)運動とを分離して数値積分する。

リウビル演算子 L

$$iL = iL_r + iL_{heavy} + iL_{light} = v \frac{\partial}{\partial r} + \frac{F_{heavy}}{m_{heavy}} \frac{\partial}{\partial v_{heavy}} + \frac{F_{light}}{m_{light}} \frac{\partial}{\partial v_{light}}$$

時間発展演算子

$$e^{iL\Delta t} q_i(\Delta t) = e^{iL\Delta t} q_i(0)$$

$$\begin{aligned} e^{iL\Delta t} &\approx e^{iL_{heavy}\Delta t/2} e^{i(L_r + L_{light})\Delta t} e^{iL_{heavy}\Delta t/2} \\ &\approx e^{iL_{heavy}\Delta t/2} \left\{ e^{iL_{light}\Delta t \delta t/2} e^{i(L_r)\delta t} e^{iL_{light}\delta t/2} \right\}^n e^{iL_{heavy}\Delta t/2} \end{aligned}$$

On the fly による最適化



- 第一原理MD (Car-Parrinello法)
- 電荷揺らぎモデル (TIP4P-FQ、SPC-FQモデル)
- 分極モデル (点双極子モデル)
- 最適値を得るために繰り返し計算を必要する変数を含む
- 繰り返し計算を行わず、最適値近傍を動的に揺らがせる近似解で満足できる場合
- その変数に仮想的な慣性量、運動エネルギーを与えて運動方程式を解く。
- 繰り返し計算がなくなるため高速化が可能。
- 運動エネルギーの相当する分だけ最適値からずれて、誤差が生じる。

分極モデル

$$U = U_0 + U_{\text{pol}}$$

$$U_{\text{pol}} = - \sum_i \mathbf{E}_i \cdot \boldsymbol{\mu}_i + \frac{1}{2\alpha} \sum_i \boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_i$$

$$\mathbf{E}_i = \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j$$

$$\boldsymbol{\mu}_i = \alpha \mathbf{E}_i$$

拡張系のラグランジアン

粒子と双極子モーメントの自由度がある。

$$L(\mathbf{r}^N, \boldsymbol{\mu}^N) = \frac{1}{2} \sum_{i=1}^N m \dot{\mathbf{r}}_i^2 + \frac{1}{2} \sum_{i=1}^N M \dot{\boldsymbol{\mu}}_i^2 - U$$

双極子モーメントの運動方程式

$$M \dot{\boldsymbol{\mu}}_i = \frac{\partial}{\partial \boldsymbol{\mu}} L(\mathbf{r}^N, \boldsymbol{\mu}^N) = -\frac{\boldsymbol{\mu}_i}{\alpha} + \mathbf{E}_i$$

高速化のために

遅い演算（除算と組み込み関数）

- 除算は他の四則演算より遅い。
- 組み込み関数も時間のかかる演算である。
- ホットスポット内ではできるだけこれらの演算を減らすように工夫する。

```
do i=1,n  
  b(i)=a(i)/x  
enddo
```

```
y=1.d0/x  
do i=1,n  
  b(i)=a(i)*y  
enddo
```

除算をループの外へ移動

```
do i=1,n  
  x(i)=a(i)/b(i)/c(i)  
enddo
```

```
do i=1,n  
  x(i)=a(i)/(b(i)*c(i))  
enddo
```

除算を乗算に変更

```
do i=1,n  
  x(i)=a(i)/b(i)  
  y(i)=c(i)/d(i)  
enddo
```

```
do i=1,n  
  tmp=1.d0/(b(i)*d(i))  
  x(i)=a(i)*d(i)*tmp  
  y(i)=c(i)*b(i)*tmp  
enddo
```

通分の利用

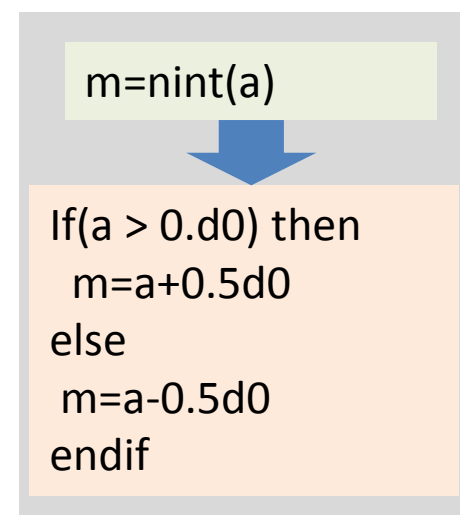
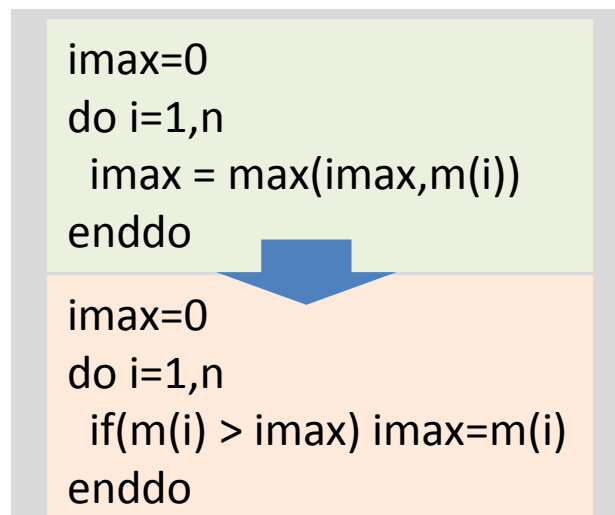
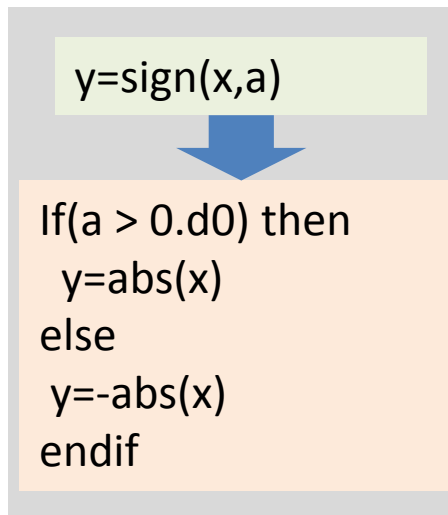


高速化のために

遅い演算(組み込み関数)

組み込み関数のfortranコード化 環境によって速くなる場合とそうでない場合がある。両者の比較が必要。

- $e^x \times e^y \longrightarrow e^{x+y}$
- $\log_a x + \log_a y \longrightarrow \log_a xy$
- $\sin\theta \times \cos\theta \longrightarrow 0.5\sin 2\theta$
- $\text{If}(\text{sqrt}(x^*x+y^*y) < R) \longrightarrow \text{If}(x^*x+y^*y < R^2)$





高速化のために

分子シミュレーションへの適用

- Lennard-Jones相互作用

$$V = \left(\frac{1}{r}\right)^{12} - \left(\frac{1}{r}\right)^6$$

$$r^6 = r^2 * r^2 * r^2$$

$$r^{12} = r^6 * r^6$$

$$V = 1.d0/r^{12} - 1.d0/r^6$$



$$r^6 = r^2 * r^2 * r^2$$

$$r^{12} = r^6 * r^6$$

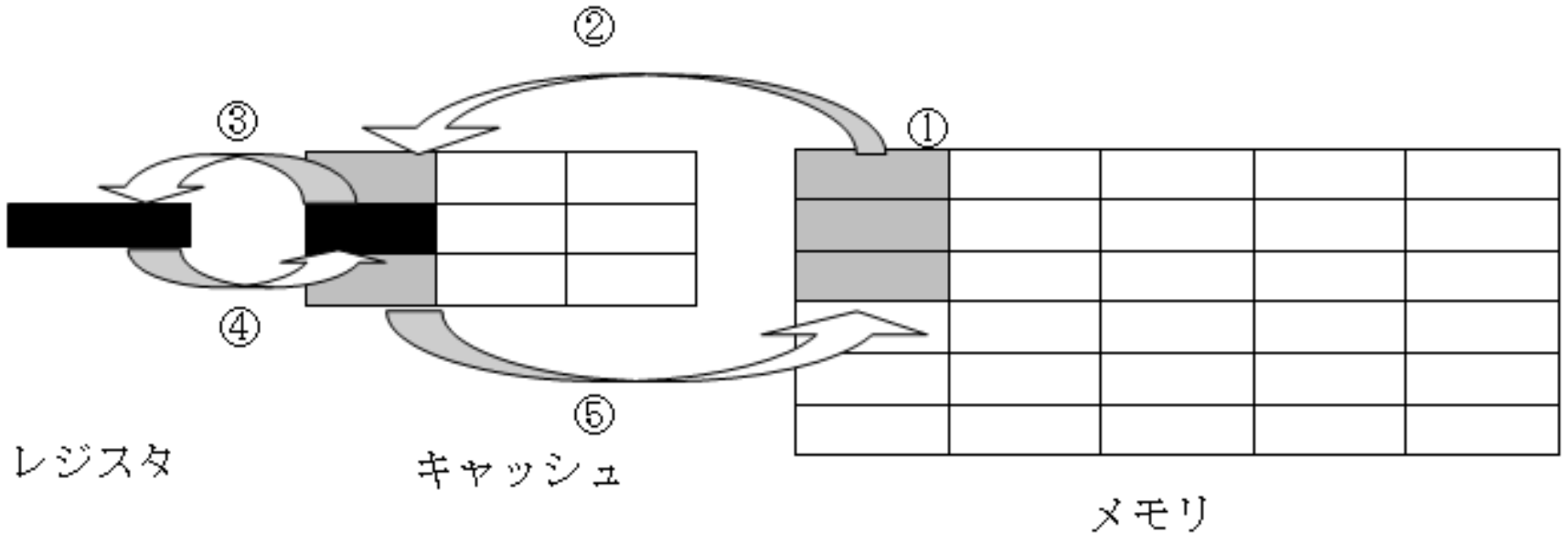
$$V = (1.d0 - 1.d0 * r^6) / r^{12}$$

LJ相互作用への適用

キャッシュの有効利用

CPUの必要なデータに高速アクセス

- ① データはメモリに格納
- ② 必要に応じて隣接するデータと共にキャッシュ上に
- ③ 演算で必要となるとレジスタにコピー
- ④ 演算された結果は再びキャッシュに上書き
- ⑤ キャッシュ上で用いられなくなるとメモリ上に上書きコピーされる。



キャッシュミス

- データ: レジスタ ← キャッシュ ← メモリ
- レジスタ ← キャッシュ: 高速アクセス
- キャッシュ ← メモリ: ← 低速アクセス
- 演算を行う時にデータがキャッシュ上にないときはメモリまで取りに行く。キャッシュミス。
- データがキャッシュに乗っているように工夫することが重要。

キャッシュミスを防ぐ



- 1次元配列の変数: 引数の順にキャッシュに格納
- 1次元配列の変数へのアクセス→配列の引数について連続的にアクセスするように心がける。
- 2次元以上の配列の変数: Fortranでは配列左側の引数の順にキャッシュに格納。
(C言語は右側から)
- 2次元以上の配列の変数へのアクセス→Fortranでは配列の左側の引数について連続的にアクセスするように心がける。A(l,j)
(C言語は右側)

MD計算の並列化

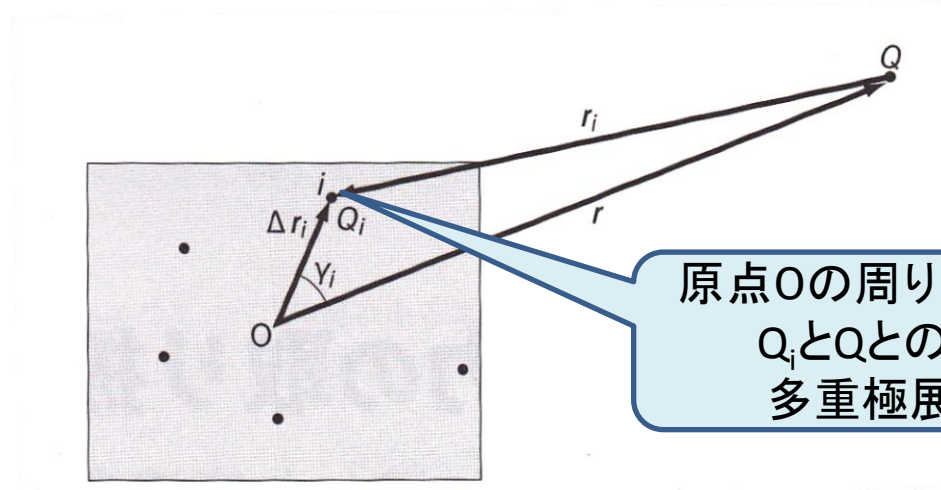
- ループ分割: 計算負荷の大きなdo ループを並列化
 - ⇒並列化が容易
 - ⇒全ノード間で通信(MPI)
 - ⇒MPIの高並列になると効率が悪い
 - 領域分割: 基本セルを複数の領域に分割
 - ⇒並列化が大変。
 - ⇒通信は隣接領域間のみ。
 - ⇒超並列でも効率がよい。
- データ構造を大幅に変更
ほぼ最初から書き直し

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

高速多重極展開法(FMM)



- 多極子モーメントを用いた多重極展開



原点Oの周りに電荷 Q_i が分布
 Q_i と Q との相互作用を
多重極展開で表す。

余弦定理

$$\frac{1}{r_i} = \frac{1}{\sqrt{r^2 - 2r\Delta r_i \cos \gamma_i + \Delta r_i^2}} = \frac{1}{r \sqrt{1 - 2\frac{\Delta r_i}{r} \cos \gamma_i + \left(\frac{\Delta r_i}{r}\right)^2}}$$

テイラー展開

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{1}{2}x + \frac{3}{8}x^2 - \frac{5}{16}x^3 + \dots \quad |x| < 1$$

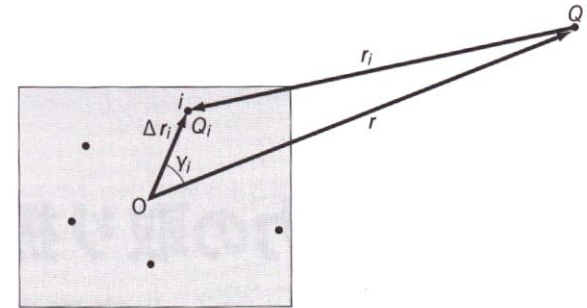
$$\frac{1}{r_i} = \frac{1}{r} + \frac{\Delta r_i}{r^2} \cos \gamma_i + \frac{\Delta r_i^2}{2r^3} (3 \cos^2 \gamma_i - 1) + \frac{\Delta r_i^3}{2r^4} (5 \cos^3 \gamma_i - 3 \cos \gamma_i) + \dots$$

高速多重極展開法(FMM)



- Q_i と Q との相互作用の多極子モーメントを用いた表式

$$\begin{aligned}
 V &= \frac{1}{4\pi\epsilon_0} \sum_i \frac{Q_i Q}{r_i} \\
 &= \frac{1}{4\pi\epsilon_0} Q \sum_i Q_i \left[\frac{1}{r} + \frac{\Delta r_i}{r^2} \cos \gamma_i + \frac{\Delta r_i^2}{2r^3} (3 \cos^2 \gamma_i - 1) + \dots \right] \\
 &= \frac{1}{4\pi\epsilon_0} Q \left[\frac{\sum_i Q_i}{r} + \frac{1}{r^3} \mathbf{r} \cdot \sum_i Q_i \Delta \mathbf{r}_i + \frac{1}{r^5} \mathbf{r} \cdot \sum_i Q_i \frac{1}{2} \{ 3 \Delta \mathbf{r}_i \Delta \mathbf{r}_i - (\Delta r_i)^2 \mathbf{I} \} \cdot \mathbf{r} + \dots \right] \\
 &= \frac{1}{4\pi\epsilon_0} Q \left[\frac{Z}{r} + \frac{\mathbf{r} \cdot \boldsymbol{\mu}}{r^3} + \frac{\mathbf{r} \cdot \mathbf{Q} \cdot \mathbf{r}}{r^5} + \dots \right]
 \end{aligned}$$



$$\begin{aligned}
 Z &= \sum_i Q_i \\
 \boldsymbol{\mu} &= \sum_i Q_i \Delta \mathbf{r}_i \\
 \mathbf{Q} &= \sum_i \frac{Q_i}{2} \{ 3 \Delta \mathbf{r}_i \Delta \mathbf{r}_i - (\Delta r_i)^2 \mathbf{I} \}
 \end{aligned}$$

Legendre多項式を用いた表式

$$\begin{aligned}
 \frac{1}{r_i} &= \frac{1}{r} \sum_n P_n(\cos \gamma_i) \left(\frac{\Delta r_i}{r} \right)^n \\
 V &= \frac{1}{4\pi\epsilon_0} \frac{Q}{r} \sum_i Q_i \sum_n P_n(\cos \gamma_i) \left(\frac{\Delta r_i}{r} \right)^n
 \end{aligned}$$

展開のしかた

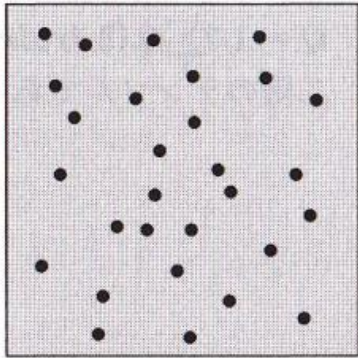
- 多極子モーメント
- Legendre多項式
- 球面調和関数

等々

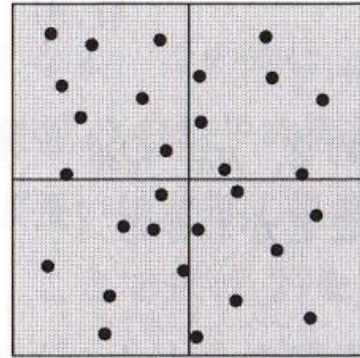
高速多重極展開法(FMM)



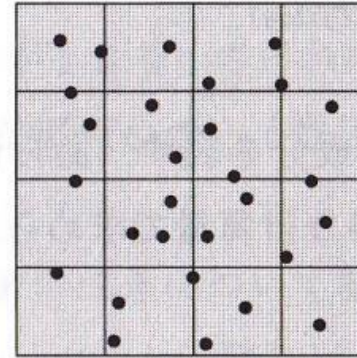
- 空間分割



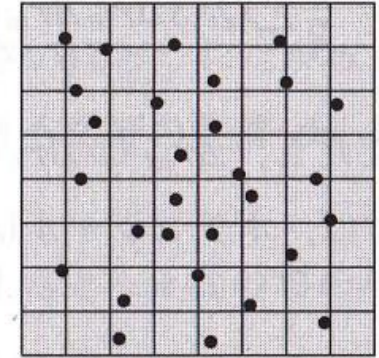
(a) レベル 0



(b) レベル 1



(c) レベル 2



(d) レベル 3

2次元: 4分木構造
3次元: 8分木構造

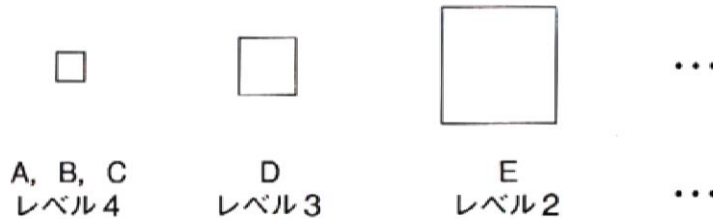
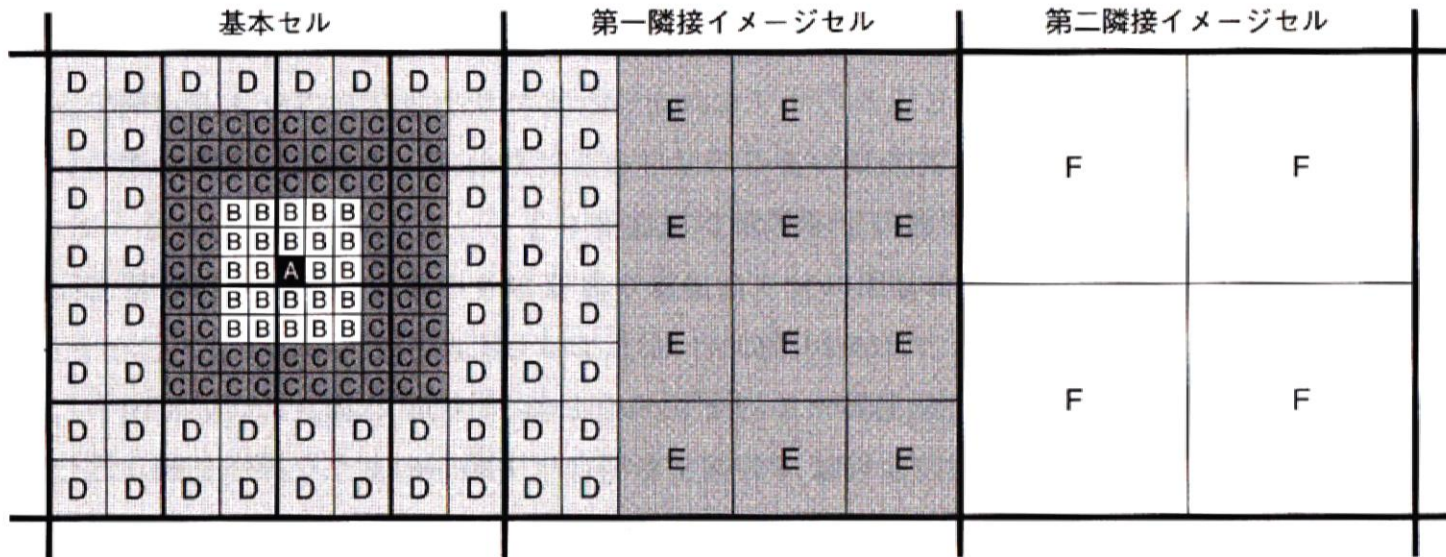
レベル	サブセル一辺の長さ
0	L
1	$L/2$
2	$L/2^2$
3	$L/2^3$

最も細かく分割したレベルのサブセルについて領域分割を適用

高速多重極展開法(FMM)



- 粒子と分割されたサブセルとの相互作用 周期境界条件あり



自セルAの粒子iのポテンシャル V_i

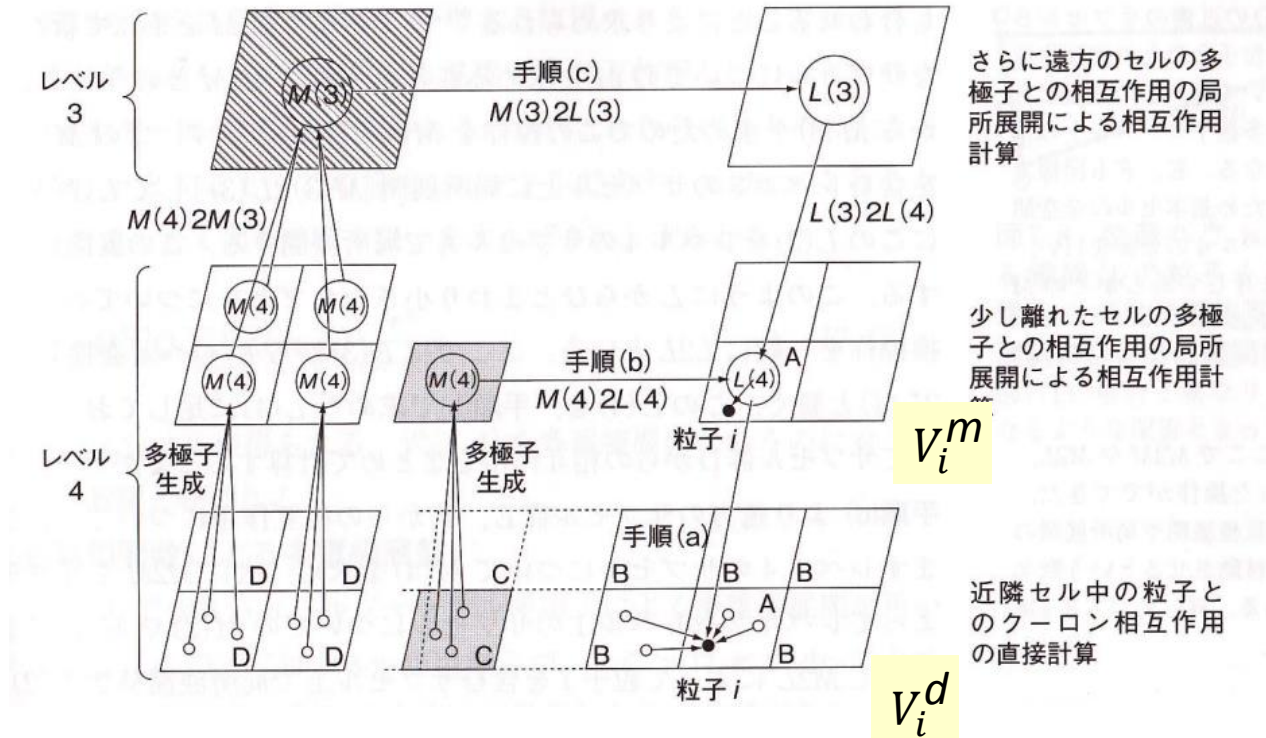
$$V_i = V_i^d + V_i^m$$

距離に応じてレベルの異なる多極子モーメントを利用する

高速多重極展開法(FMM)



多重極展開と局所展開



手順(d) より遠いサブセル群E,FについてもDと同様。M(2), M(1)を求め、M2L, L2LによってL(4)を求める。

手順(e) 第3隣接以遠無限遠に至るイメージセルにM(0)があるものとして、基本セル上に作るL(0)を求める。L2LによりL(4)を求める。

手順(f) L(4)を用いてポテンシャルを評価する。

これを効率的に行うために手順(c), (d), (e)で得られたLそれぞれのレベルで足し合わせてからより小さなサブセルにL2L変換する。

高速多重極展開法(FMM)



計算精度

- 静電+Lennard-Jones相互作用

原子数 10,125,000 個
体積 46.8³ nm³
セル数 64³ (=262,144) 個 (セルあたり39個)

	V / 10 ⁻¹³ J	F _x / 10 ⁻¹⁰ N	F _y / 10 ⁻¹⁰ N	F _z / 10 ⁻¹⁰ N
correct	-218735559	-4.357177	6.380146	-1.477348
4次	-218735742	-4.356659	6.380802	-1.475152
8次	-218735559	-4.357201	6.380145	-1.477340

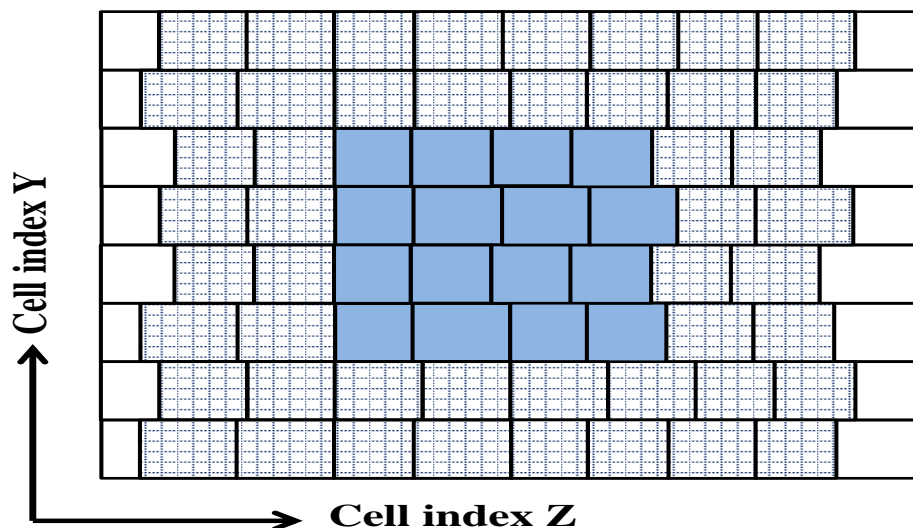
展開
エネルギー 4次 6~7桁
8次 8~9桁
力 4次 4~5桁
8次 6~7桁

- 計算精度を展開次数で制御



原子のデータ構造

- 原子座標、速度、力、セグメントデータ → メタデータ化
- プロセスに属するサブセルを割り当てて固定（図では4×4のサブセルを1つのプロセスに割り当て）
- サブセル内の原子のデータをプロセスに局在化させる。グローバルに持たない→省メモリ。
- データ領域に空きを用意 → サブセル内の原子数の増減に対応。各サブセル内の原子数に合わせて要素数に変化。要素数情報もあわせて保持。
- メタデータ配列のまま演算も通信も行う。サブセルを表す3次元配列を持っているので、取り扱いが容易
- バッファリングやindirect accessがない。
- サブセル単位のメタデータがキャッシュに載るようにサブセル内の原子数を調整。
- 原子の運動に伴って原子の帰属するサブセルを更新。

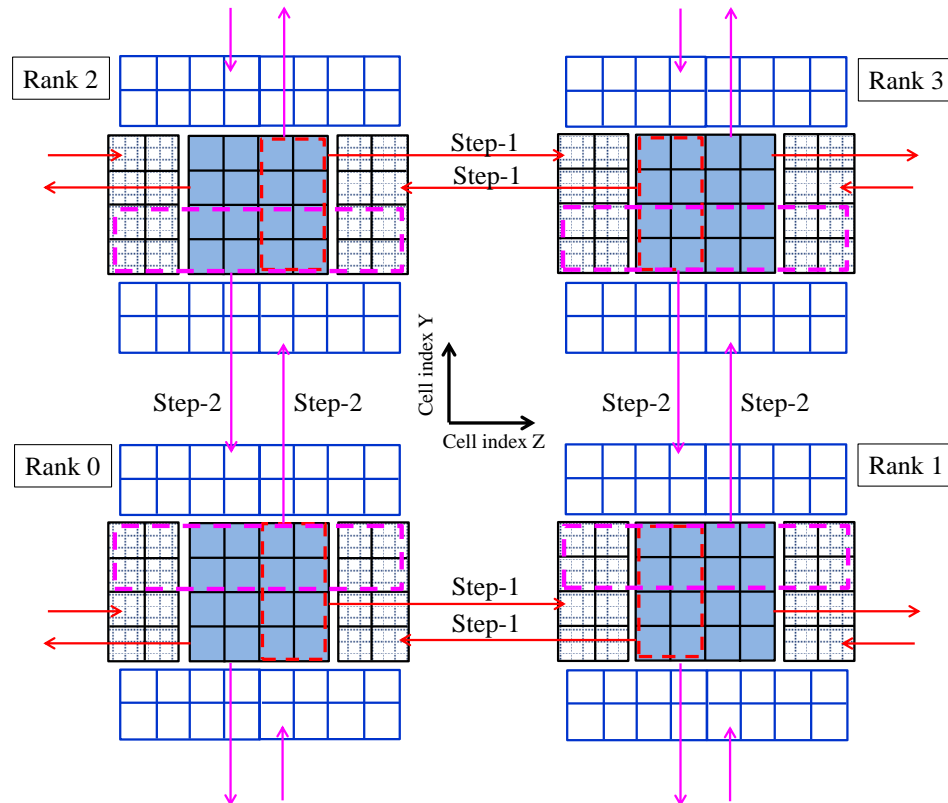


・セグメントデータ
一塊の原子団のデータ
CH₃、H₂Oなど。
原子間距離拘束の情報



自セルと相互作用する原子情報を集める。

- 通信回数と通信量の最小化
- 隣接セルと通信して1次元(z軸)方向にデータを結合し棒状のデータを作る。
- 棒状データをy軸方向に通信し結合させて、平面状データにする。
- x軸方向に平面状データを通信し結合させて、立方体状のデータにする。
- Kースパコンの多重同時通信(各軸ごとに+、-方向)



データ局在化による計算効率の向上



キャッシュの利用効率の向上

- L1キャッシューL2キャッシューメモリ
- L2ーメモリ間、L1ーL2間のデータ通信の最小化

最重要ホットスポット

- 2体間相互作用計算
- FMMのM2L計算

2体間相互作用計算

- 3重ループ構造

```
do 5つのサブセルを指定するインデックス
```

```
do 注目する1つのサブセル内の粒子
```

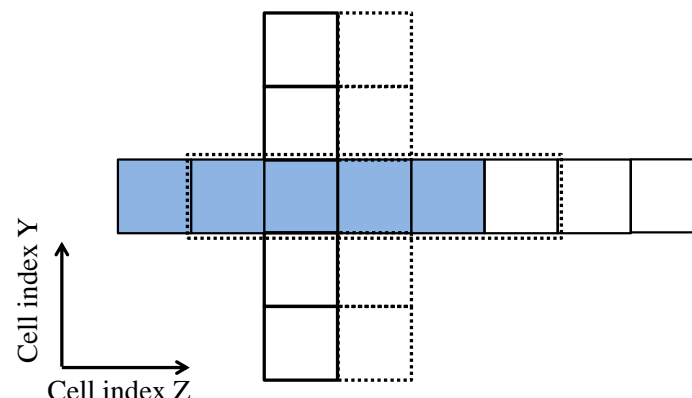
```
do 隣接する5つのサブセル内の原子  
相互作用計算
```

```
enddo
```

```
enddo
```

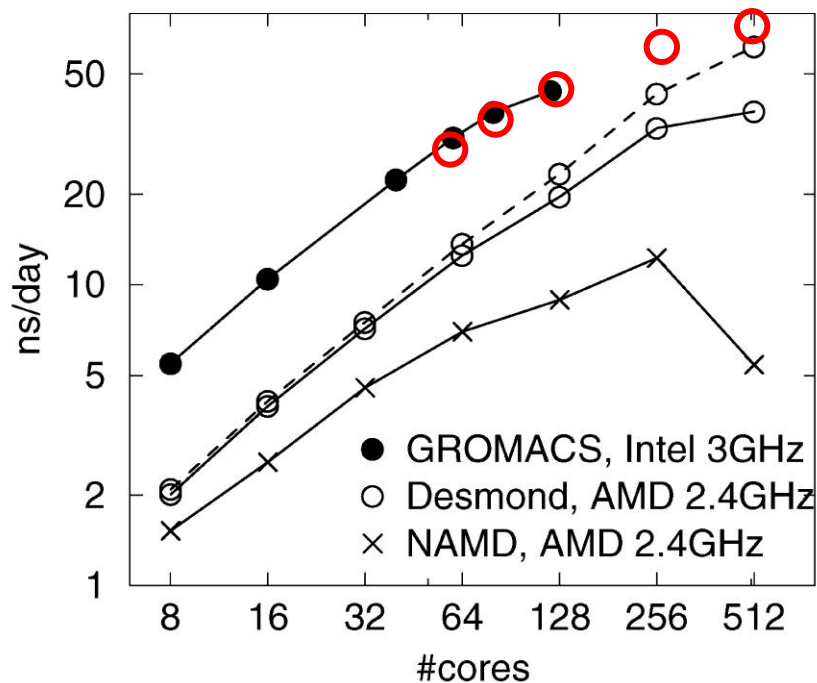
```
enddo
```

- これにより、5つのサブセルのデータが効率的に再利用される。
- ただし5つのサブセルの座標データをL1キャッシュに載せておくことが重要





PCクラスタを用いたベンチマーク



Hess et al. (2008)

⊙ : Desmond (2008)

DHFR in water
23,558 原子系

$r_c = 9.6 \text{ \AA}$ GROMACS
 9 \AA Desmond
 9 \AA NAMD

FFT : $64 \times 64 \times 64$

$\Delta t = 1 \text{ fs}$

8コア(1ノード)の使用と比較して512コア(64ノード)で、すでに30%程度の並列化効率

24,000 原子系

100~500並列

1~2 ms/step

(50~200 atoms / processors)

CMSI計算科学技術特論A

ANTONのパフォーマンス (専用マシン)



DHFR 23,558 原子系 force error 1.5×10^{-4}

512 ノード並列 14.5 $\mu\text{s}/\text{day}$

400 MHz

1 ms / 70 days

24,000原子系

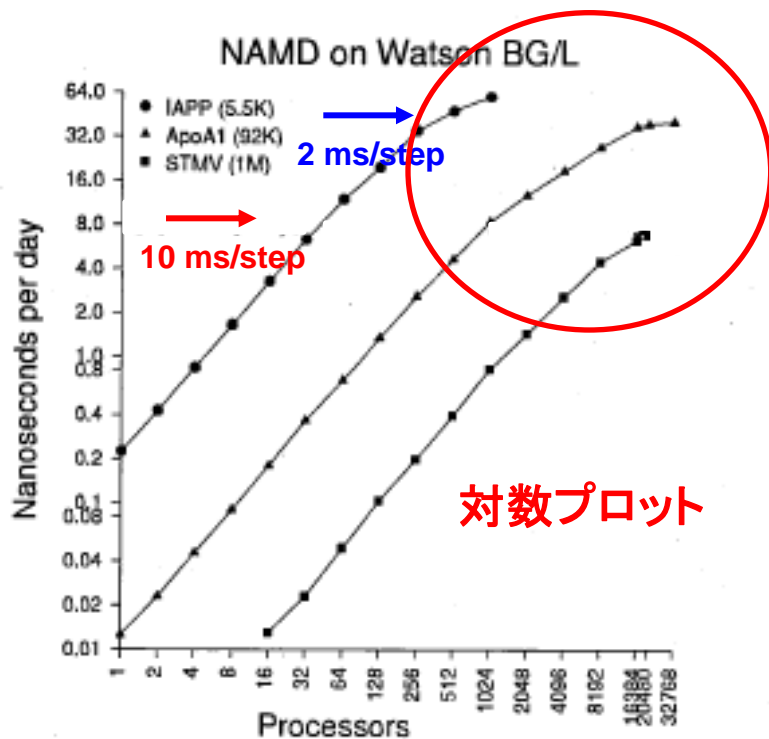
512並列

$\sim 6 \mu\text{s}/\text{step}$



Shaw et al. (2008)

NAMDのパフォーマンス on Blue Gene/L



(coprocessor mode)

Schulten et al. (2008)

IAPP 5,570 原子系

cutoff 12 Å $\Delta t = 2$ fs

ApoA1 92,224 原子系

cutoff 12 Å $\Delta t = 1$ fs

STMV 1,066,628 原子系

cutoff 12 Å $\Delta t = 1$ fs

1,000,000原子系

~20,000(40,000)並列

~15 ms/step

(50 atoms / processors)

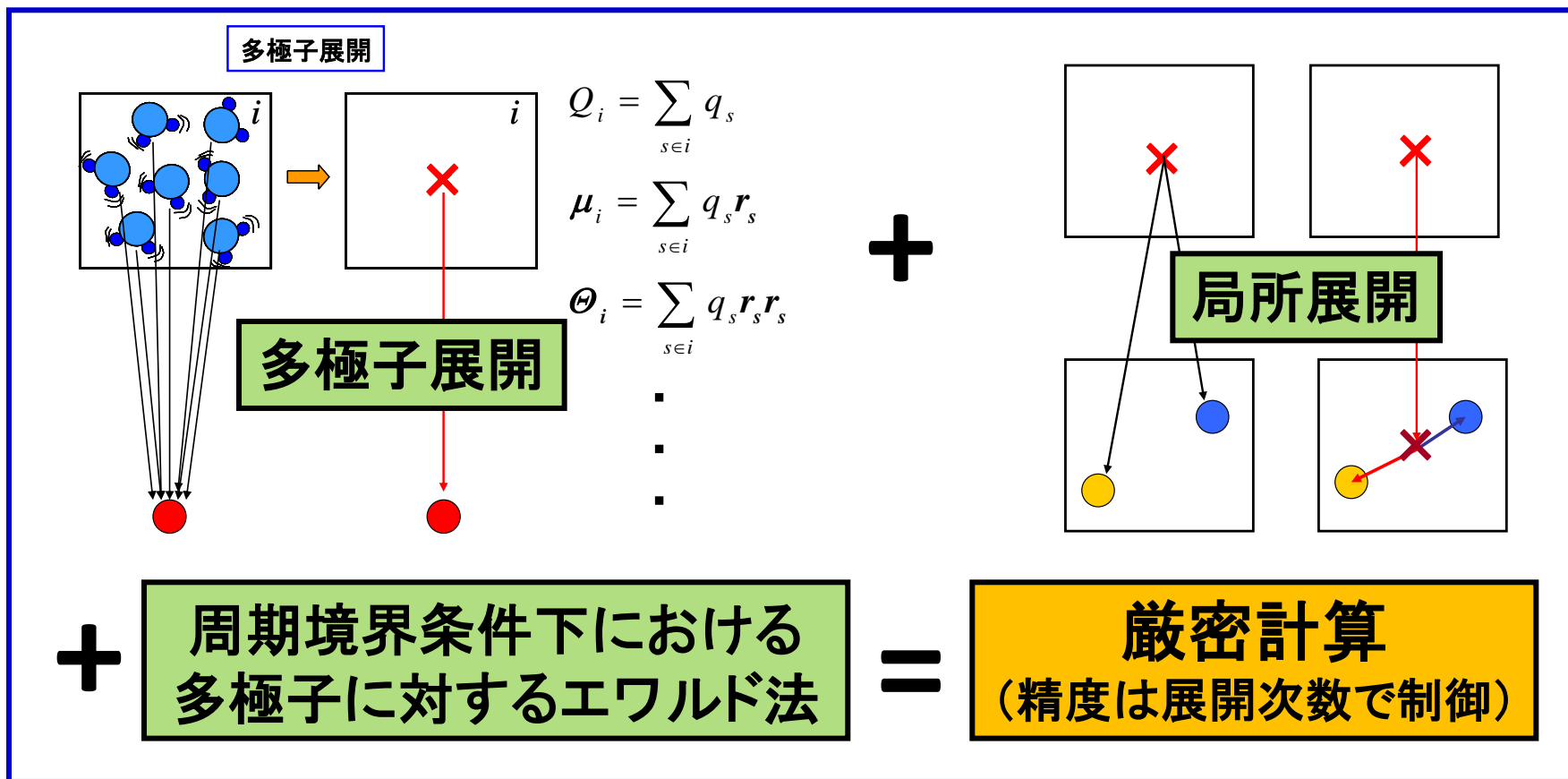
CMSI計算科学技術特論A

並列化

- ・完全領域分割
- ・多階層隣接通信化

実質的に FFT free !

大きな通信はすべて隣接通信化



ウイルスカプシドの分子動力学計算



全原子計算

マイクロ秒程度の軌跡

~1000万原子系

- ・周期境界条件下
- ・長距離相互作用



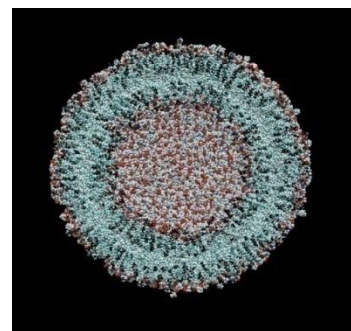
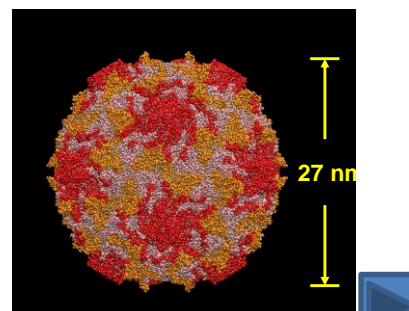
京スパコン

10⁰ms/step

- ・大規模並列が不可欠
- ・通信もms



隣接通信化が不可欠
(FFTは不可)



小中規模系の計算

- ・2~3万原子系
- ・Ewald法
領域分割
FFT

専用計算機

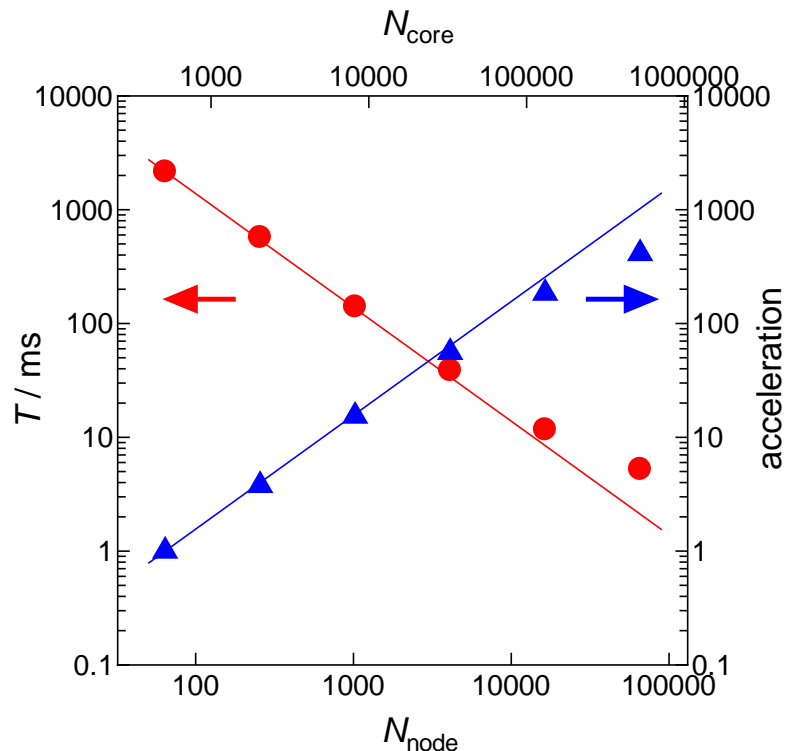
- ・ANTON
512並列
数 μ s/step
- ・MDGRAPE

従来のスパコン

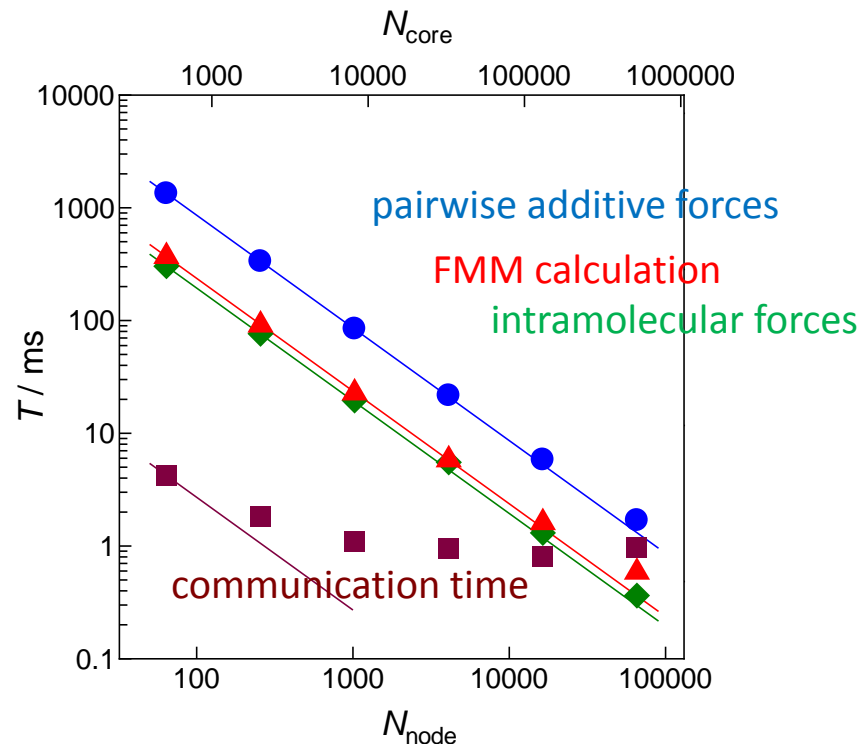
1マイクロ秒: 10年(解析不可能)
(実効性能 0.1ペタフロップス)



PYP 水溶液、1000万原子系



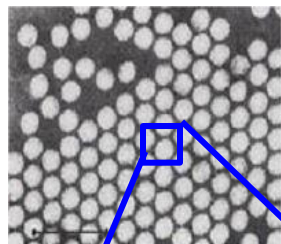
The measured overall calculation time per one MD step(Dt) for the 10-million atomistic system, the PYP solution, and the acceleration ratio with respect to the 64-nodes calculation as a function of degree of parallelism, N_{node} .



The measured direct calculation time of the pairwise additive forces(circle), the FMM calculation(triangle), the intramolecular forces(diamond), and the communication time(brown) per one MD step(Dt) for the 10-million atomistic system, the PYP solution, as a function of degree of parallelism, N_{node} .

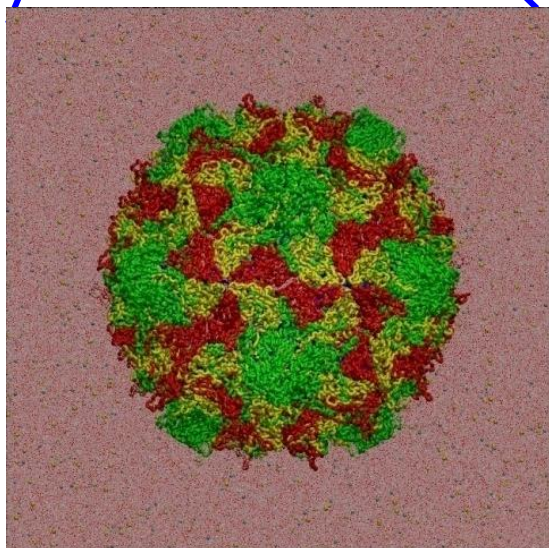
感染機構を解き明かす

ウイルスとレセプターの特異な相互作用



小児麻痺
ウイルス
"Medical Virology",
edited by D. O. White and
F. Fenner, Academic Press

10,000,000原子

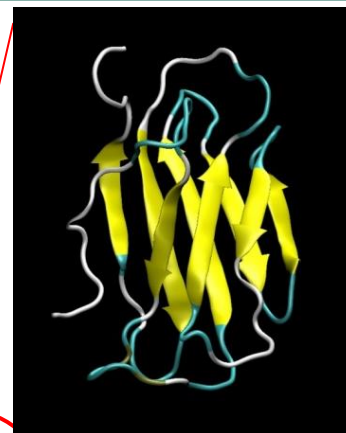


小児マヒウイルス
P=1 atm, T=310.15 K

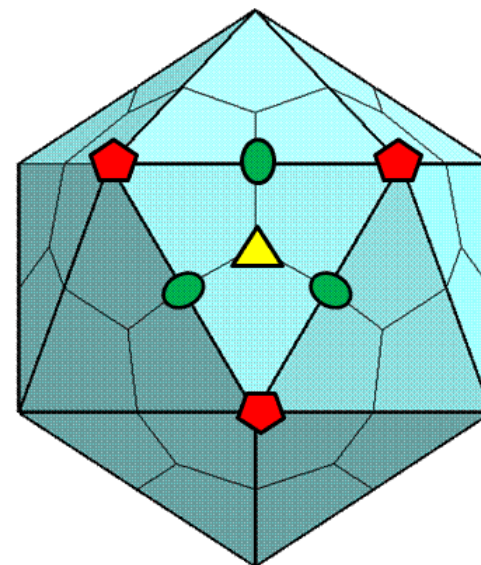
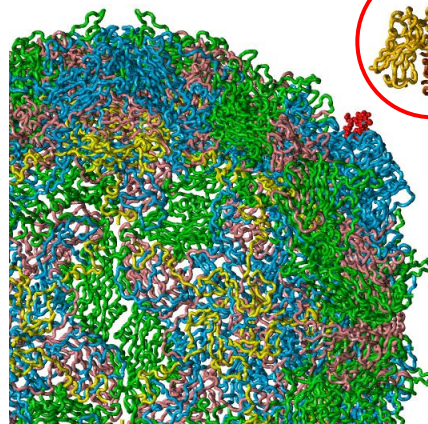
「京」スパコンが不可欠



ウイルス全体1000万原子系の
分子動力学計算が必要



レセプター



今日のまとめ



MD計算の要素 各要素の高速化、並列化

MD計算とは

- MD計算の流れ、取り扱う原子数と周期境界条件、相互作用(力)、分子モデルと運動方程式
- MD計算の例、歴史、現状

非並列のMD計算

- 短距離相互作用 (bookkeeping, cell index法)
- 長距離相互作用(クーロン相互作用) Ewald, PME法
- 効率化のための方法
 - 拘束動力学 (Δt を大きくする)
 - Multiple time step (Δt を大きくする)
 - On the flyによる最適化 (収束計算をなくす)

高速化のために、キャッシュの有効利用、キャッシュミス

MD計算の並列化

- 高速多重極展開法(FMM)

データ構造 メタデータ法、通信量の最小化、データ局在化による計算効率の向上

ベンチマーク

MODYLAS

参考文献



MD全般

- D. Frenkel, B. Smit, "Understanding Molecular Simulation," Academic Press, San Diego (1996).
- 上田 顕, 「コンピュータシミュレーション」, 朝倉書店(1990).
- M. P. Allen, D. J. Tildesley, "Computer Simulation of Liquids," Oxford Science, Oxford(1987).
- J. P. Hansen, I. R. McDonald, "Theory of Simple Liquids," Academic Press, London(1986).
- 岡崎 進, 吉井 範行, コンピュータ・シミュレーションの基礎(第2版), 化学同人(2011).

力学

- H. Goldstein, 「ゴールドスタイン新版古典力学(上), (下)」(瀬川富士, 矢野 忠, 江沢康生 訳), 吉岡書店(1983).

数値積分

- G. J. Martyna, M. E. Tuckerman, D. J. Tobias, M. L. Klein, Mol. Phys., 87, 1117(1996).

Particle mesh Ewald法

- T. Darden, D. York, L. Pedersen, J. Chem. Phys., 98, 10089(1993).

FMM

- L. Greengard, V. Rokhlin, J. Comput. Phys., 73, 325(1987).

SHAKE

- G. Ciccotti, J. P. Ryckaert, Comp. Phys. Rep., 4, 345(1986).
- P. Gonnet, J. Comput. Phys. 220, 740(2007).

MODYLAS

- Y. Andoh, et al., *J. Chem. Theory Comput.* 9, 3201 (2013)
- Y. Andoh, et al., *J. Chem. Phys.* 141, 165101 (2014).