

学際開拓プログラム、物質科学分野

オーダーN法第一原理計算プログラム CONQUESTの開発と応用計算

物質・材料研究機構、計算科学センター
宮崎 剛

(MIYAZAKI.Tsuyoshi@nims.go.jp)

構成メンバー：

物質・材料研究機構 ICYS-IMAT

A. S. Torralba

物質・材料研究機構、計算科学センター

大塚 教雄（理研）、M. Todorovic

- **オーダーN法第一原理計算プログラムCONQUEST
”超大規模系に対する第一原理計算”**

- **本研究課題**

- 並列化効率、演算効率の向上
- 生体系への応用計算

**H20後期：
CONQUESTをT2K-Tsukubaへ移植。**

オーダーN法第一原理計算プログラム CONQUEST

- 通常の第一原理計算手法は、演算量が系の原子数 N の3乗で増加する。
 - 1千原子以上の系を含む系に対する第一原理計算はかなり困難。
- **オーダーN法第一原理計算プログラムCONQUEST**
 - 英国University College London (UCL)と物質・材料研究機構(NIMS)で共同開発。
 - 計算に必要なメモリ量、演算量が原子数 N に比例。
 - 数万、数十万以上の原子を含む系に対する計算をターゲットにしている。
 - 現在、共同研究を基本とした限定 β 版を公開。一般公開に向けてプログラム開発中。

CONQUEST：計算手法

- 密度行列最適化手法

- 密度行列 $\rho(\mathbf{r}, \mathbf{r}') = \sum_{\nu} f_{\nu} \psi_{\nu}^*(\mathbf{r}) \psi_{\nu}(\mathbf{r}')$

- $\{\psi_{\nu}(\mathbf{r})\}$: Kohn-Sham ハミルトニアン固有関数、波動関数)

- 密度行列の局所性: $\rho(\mathbf{r}, \mathbf{r}') \rightarrow 0$ for $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$

カットオフ半径を導入

- 密度行列を局在軌道 $\phi_{i\alpha}(\mathbf{r})$ で表す。

- $\rho(\mathbf{r}, \mathbf{r}') = \sum_{i\alpha, j\beta} \phi_{i\alpha}^*(\mathbf{r}) K_{i\alpha, j\beta} \phi_{j\beta}(\mathbf{r}')$

- サポート関数 $\phi_{i\alpha}(\mathbf{r})$: 原子 i の軌道 α に対する関数.

LNV法: $K=3LSL-2LSLSL$

L : auxiliary density matrix

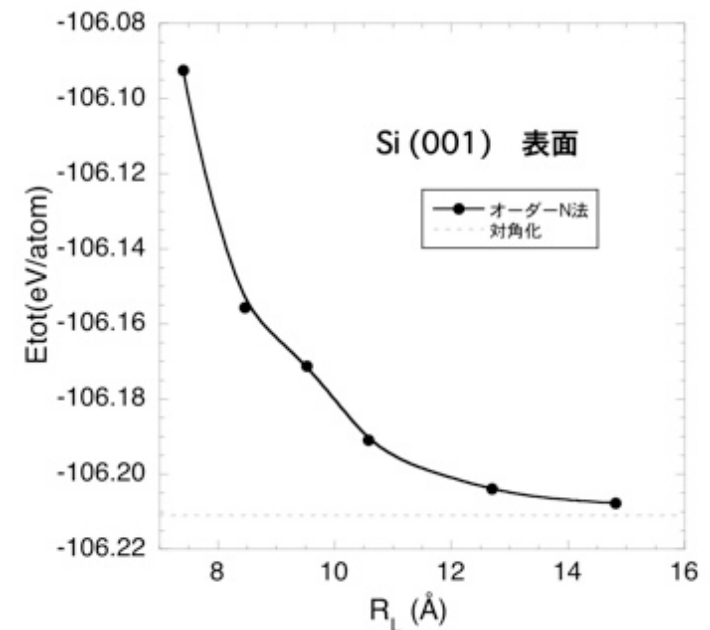
S : overlap matrix

Minimise total energy with respect to

$L_{i\alpha, j\beta}$ and $\phi_{i\alpha}(\mathbf{r})$

密度行列のカットオフ半径を長くすると、

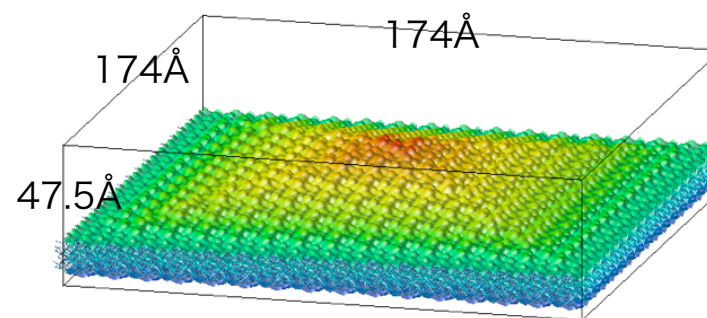
1. 全エネルギーが正確になる。
2. 計算量は増える。



CONQUEST: 適用例

- Si(001)上のGe 3次元島構造
 - 約23,000原子を含む系に対する構造最適化に成功
 - 吸着Geの被覆率が3より大きいと、2次元から3次元構造に安定構造が変わる。（実験と一致）

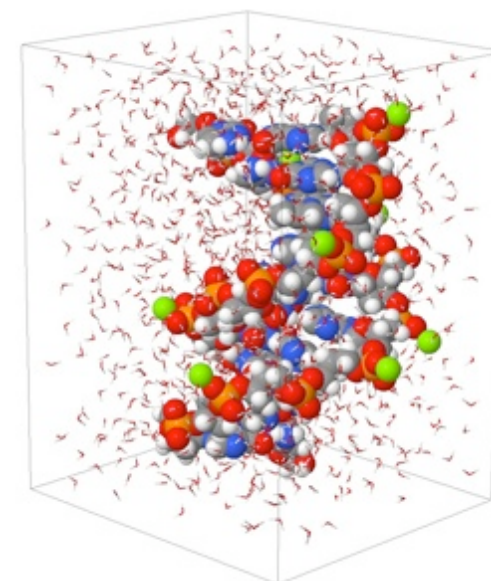
T. Miyazaki et al., J. Phys. Soc. Jpn 77, 123706(2008).



地球シミュレータ512 proc.で12hour x 3

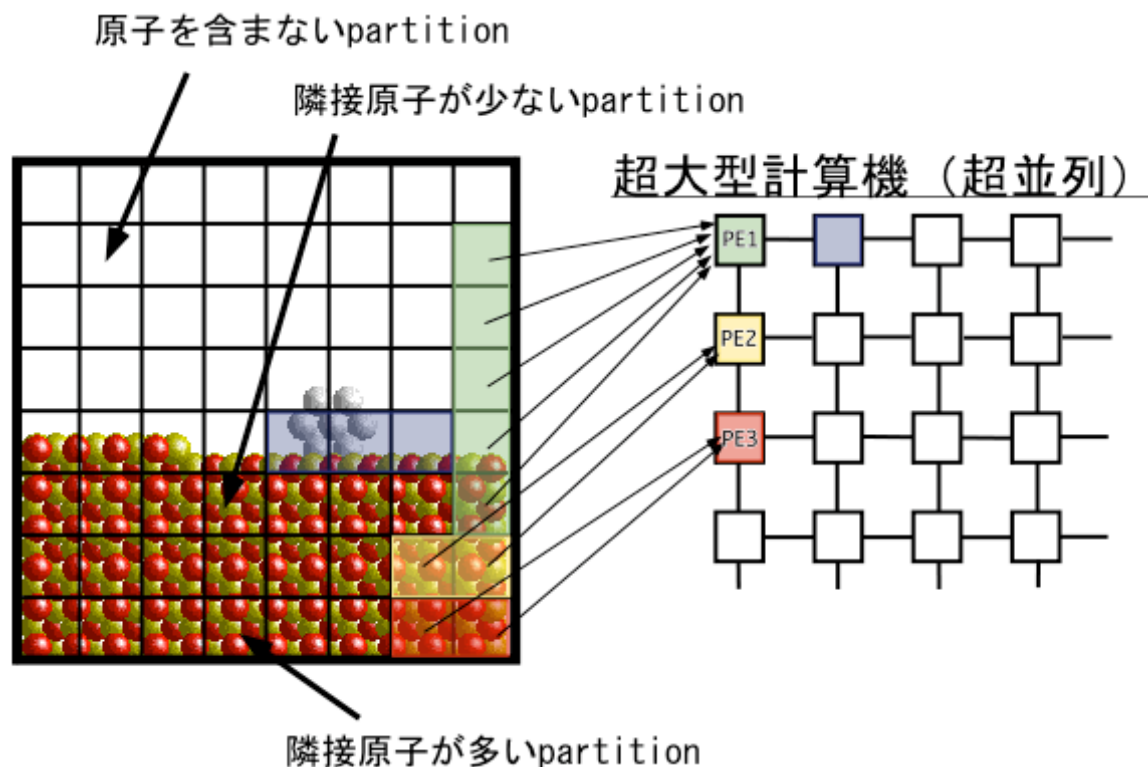
- 水溶液中DNA
 - 水和水を多数含む10塩基対の系（約3,400原子）の全エネルギーと力の計算
 - 密度行列最適化手法によるオーダN法が極めて高精度であることを示した。
 - AMBERとCONQUESTの力の計算の比較。

T. Otsuka et al., J. Phys.: Condens. Matter 20, 294201 (2008).
宮崎剛、大塚教雄、固体物理 Vol. 43, p.849(2008).



SR11K 128 proc. で12hourx5

粗行列の行列積



$$C_{i\alpha, j\beta} = \sum_{k\gamma} A_{i\alpha, k\gamma} B_{k\gamma, j\beta}$$

原子 軌道(s,p,...)

例えば、(LS)=LS

行列Lと行列Sは異なる
カットオフ半径を持つ。
(行列LSも)

$$K = 3(LSL) - 2(LSL)(SL)$$

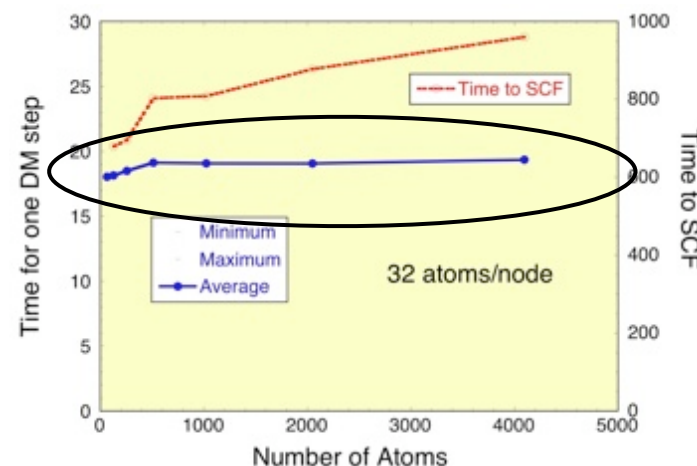
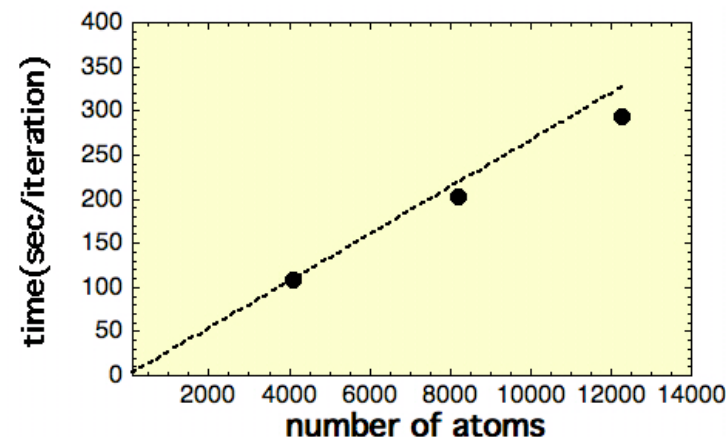
行列KのカットオフはSと同程度でいい。

生体系の場合：

- ✦ 軌道の数が原子ごとに大きく違う
- ✦ 場所で結合の種類も違う (隣接原子数が違う)

CONQUEST:実行効率、並列化効率

- オーダーN：計算時間がNに比例
- 並列化効率
 - 2種類の並列化効率
 - 計算する系が同じで、PE数を増やした時の効率
 - 1 PE当たりの原子数を同じにして、PEの数と系を同時に大きくする。
- 通常ユーザー版と並列化効率最適化版の二つが存在。
- （粗行列の行列積において、演算数とデータ転送量を見積もる事によって、通信の局所性と均等な負荷分散を同時に満たす分割の仕方を計算する別プログラムを用いる。）



半導体表面の場合うまくいく！

生体系は??

平成20年度後期

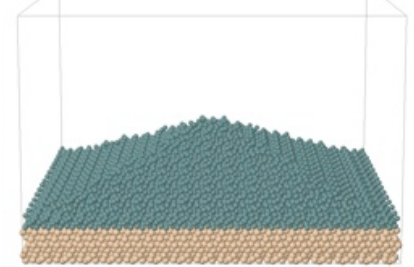
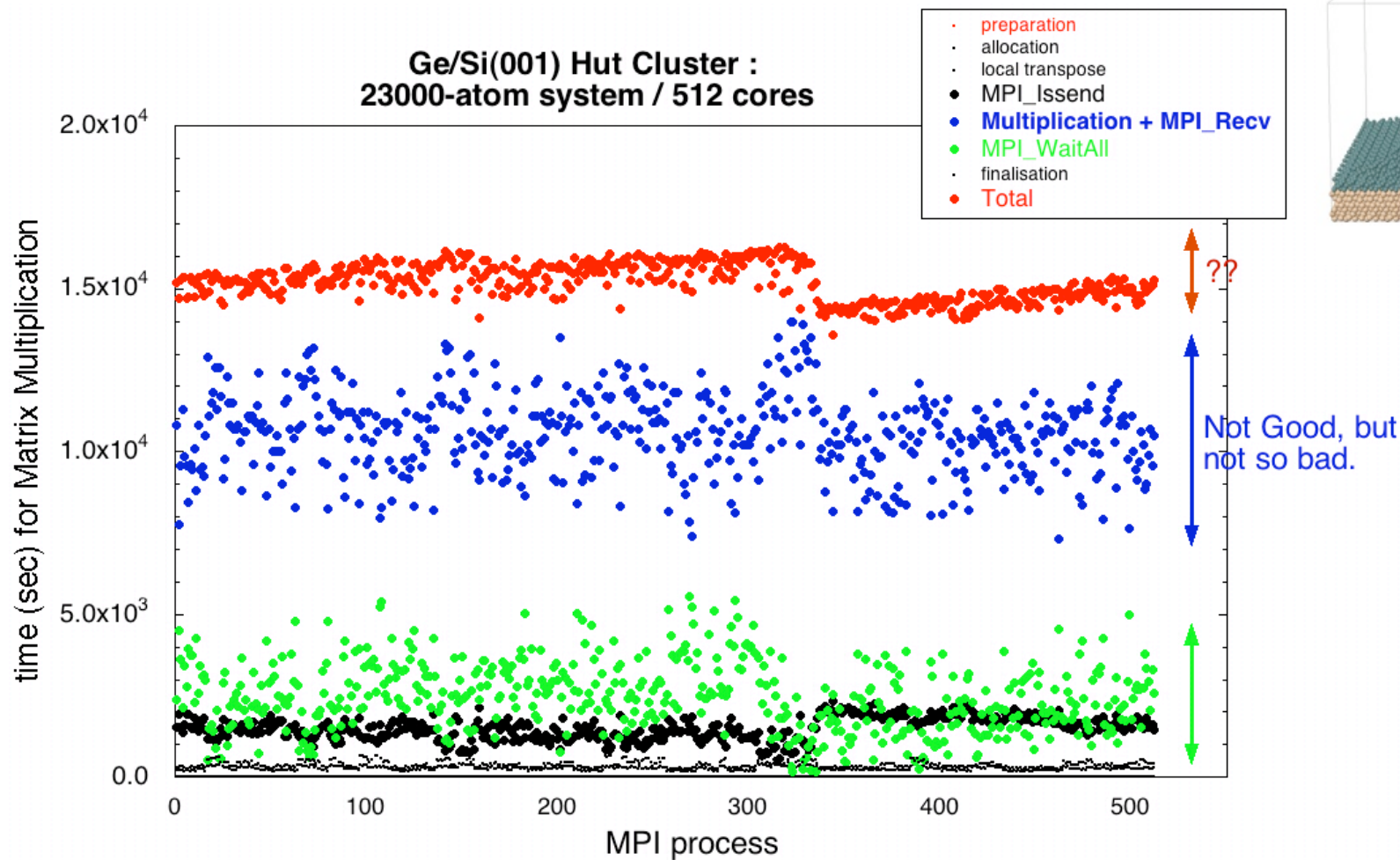
- ◆ CONQUESTのT2Kへの移植
 - ◆ Scalapack 完了 (performanceは?)
 - ◆ MPI2 必要? これからテスト要
 - ◆ 結果の不一致 FFTの問題: 完了
- ◆ CONQUESTによる並列計算
 - ◆ Ge hut cluster on Si(001)
 - ◆ bulk Si
 - ◆ H₂O molecules
 - ◆ 生体系(DNA、イオンチャネル等)
- ◆ 生体系への適用
 - ◆ イオンチャネル gramicidin-A (<-> 舘野研)

CONQUEST on T2K

- ◆ 典型的な系でSR11Kよりも2、3倍速い
 - ◆ 550-atom system、対角化：9.8min(T2K) vs. 26.7min. (SR11K), 44min(T2K) vs. 130.5 min (SR11K)
 - ◆ 他の系でも、同様の傾向。
- ◆ 並列化効率は今のところ良く分からない。
 - ◆ 512 core で、十分正確な計算条件で2、3万原子の第一原理計算が出来そう。（ただし、スピードが問題になる。）
 - ◆ MVAPICH2で不安定だったので、multirailが試せなかった
 - ◆ weak scalingもうまくいかない。
 - ◆ scalapackの効率は？（超大規模計算では使わない）
 - ◆ すべてこれから。。。

CONQUEST on T2K

Ge/Si(001)hut cluster:

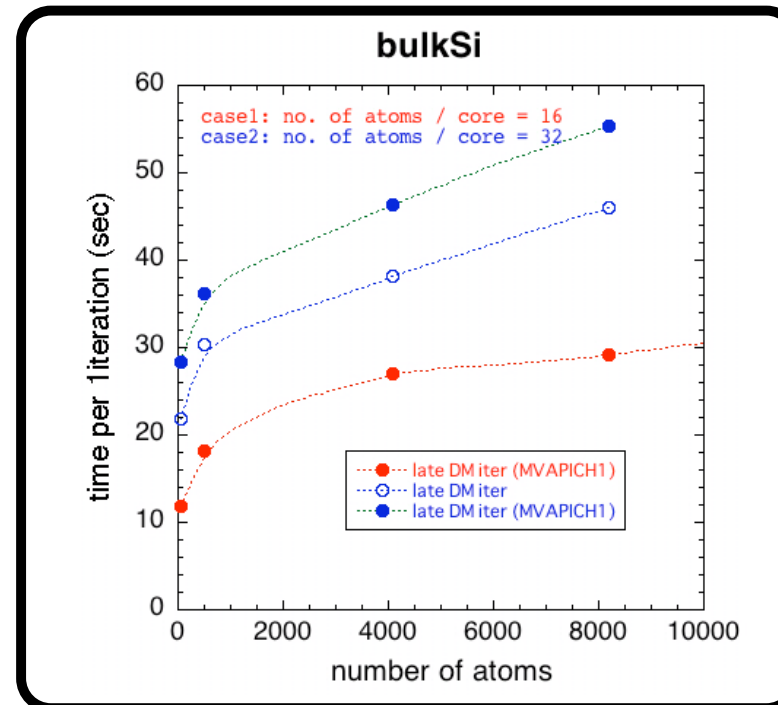
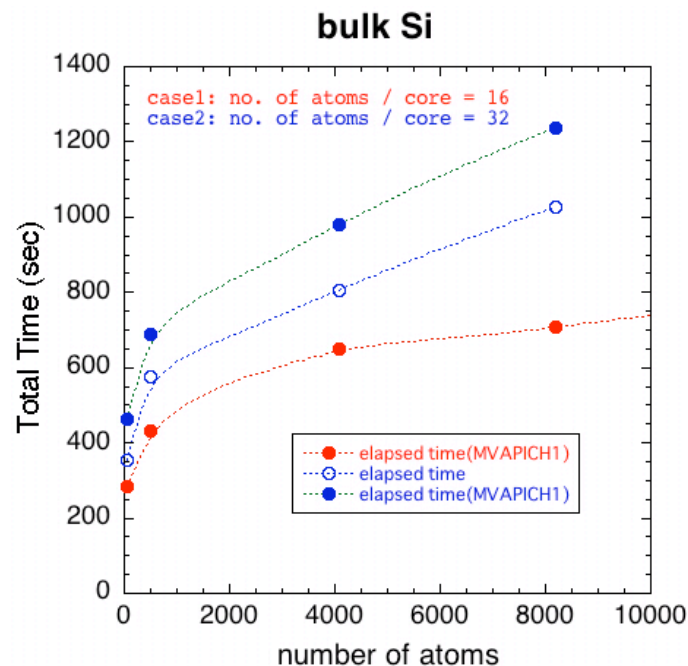


- 悪くない。but...
- ESより10倍遅い

- 通信が問題？
- MPI2の片方向通信は？

CONQUEST on T2K

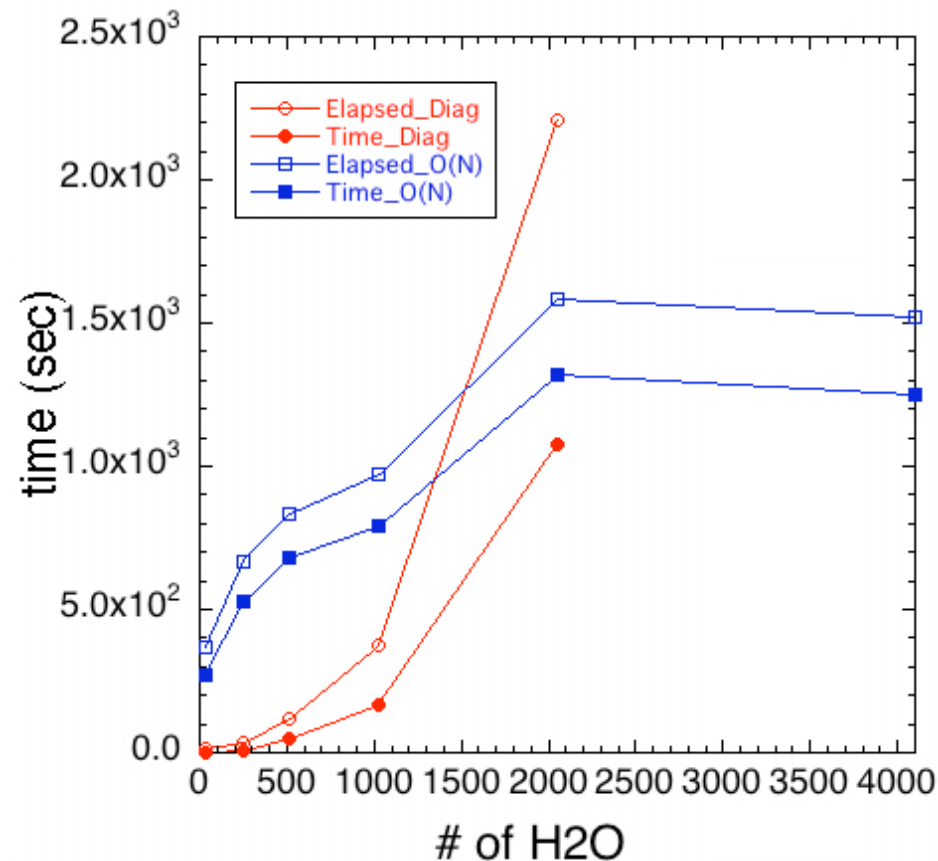
bulk Si



- 小さいところは仕方がないが、flatにならない
- strong scalingは良さそう
- 原子の配り方が良くなかったので再測定が必要
- multirail の効果もチェック要

CONQUEST on T2K

H2O molecules



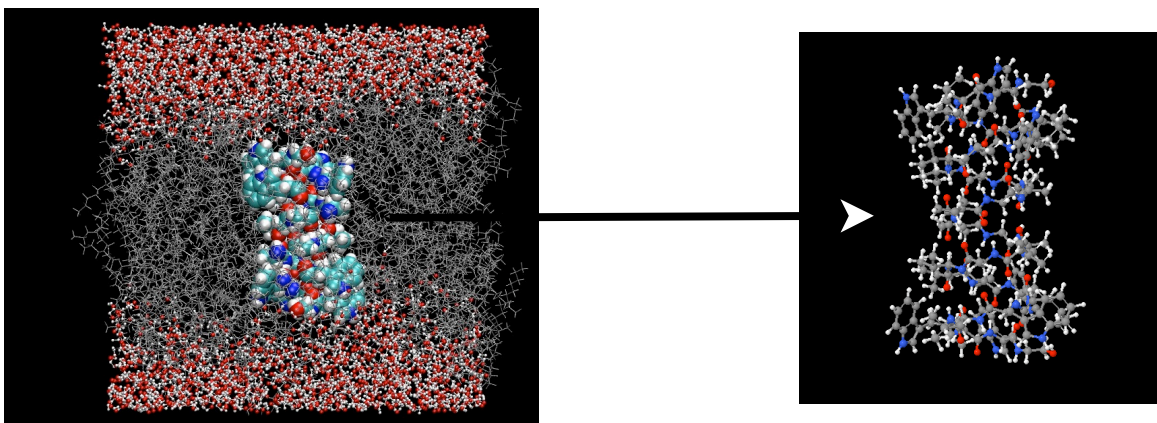
- 原子位置がランダムになるので load balanceが少し困難になるはず。

- O(N)の方も scalingが flatにならない。（対角化の方は元々 flatになるはずがない）

- O(N)は収束条件が厳しいために遅い。（系が大きい方が iterationの回数が多くなっている。）

2-b. gramicidin-A system

Present Study: isolated gramicidin-A



Aim of the study:

- mechanism of the ion selectivity
 - ▶ structural properties of the system
 - ▶ interactions between lipid molecules (membrane) and gramicidin-A
 - the role of the side chains
 - ▶ interactions between gramicidin-A, ions and water molecules.

Study of the isolated gramicidin-A:

- structural stability of the isolated molecule
- the effects of the side chains
 - ▶ on the structural stability
 - ▶ on the electronic structure
- feasibility of the O(N) method

まとめ

- 1点計算ならば、2048 coreで10万原子の系に対する第一原理計算が可能に思える。
- 並列化効率を高め、計算時間を速くしないと構造最適化などは困難。
- なぜ並列化効率が悪いのかを確認し、その後の最適化が必要。