Linear Scaling Methods and Applications with the Divide-and-Conquer Method and with Localized Orbitals

Weitao Yang

Department of Chemistry

Duke University

Taisung Lee(MSI)   Darrin York(Minnesota)
James Lewis (BYU)   Steven Burger

Yingkai Zhang (UCSD)   Haiyan Liu (U.Sci.Tech.)
Zhenyu Lu   Amy Boone
Andres Cisneros   Jason Donald

Our Two Approaches to the Modelling of Large Systems

1. Linear Scaling Quantum Mechanical Methods: Divide-and-conquer Method Phys. Rev. Lett., 66, 1438(1991)

- For the first time, it shows that linear scaling is possible for electronic structure calculations

- Applications of the divide-and-conquer method to enzymes and protein dynamics

2. QM/MM: –Combined QM and Molecular Mechanics

JCP, 110, 46(1999). JCP, 112, 3483(2000). JACS, 122, 6560(2000).

$O(1)$ in practice!

- A Pseudobond Approach to Combining Quantum Mechanical and Molecular Mechanical Methods

- Optimization and a Free-energy Perturbation Method

- Applications to enzyme reaction mechanism

–Before our work on the Divide-and_Conquer, electronic structure calculations scaled at least as N^3.

–Our divide-and-conquer approach (1991) is linear scaling, $O(N)$.

–Computational effort $\propto$ the size of molecule

It broke the bottleneck for quantum mechanical calculations of large systems.

Key to **linear scaling algorithms**: the use of the *localized* electronic degrees of freedom.

Review**:** Yang and Perez-Jorda, "Linear Scaling Methods for Electronic Structure Calculations", Encyclopedia of Computational Chemistry, edited by Schleyer, John Wiley & Sons (1998).

For a latest review: Wu and Jayanthi, Phys. Rep., 2002

The Divide-and-Conquer Approach

–the first linear scaling method for electronic structure calculations

Yang, Phys. Rev. Lett., 66, 1438 (1991),
Lee and Yang, J. Chem. Phys., 163, 5674(1995).

- **semiempirical QM approaches**

Lee York and Yang, J. Chem. Phys. 105, 2744(1996).
Dixon, and Merz, J. Chem. Phys. 104, 6643 (1996).

- **solids and surfaces**

Zhu, Pan and Yang, Phys. Rev. B., 53, 12713(1996).
Warschkow, Dyke, & Ellis, J. Comp. Phys., 143, 70(1998)

- **electrostatic problems**

Gallant and St-Amant, Chem.Phys. Lett. 256, 569 (1996)

Diagonalization: there are three general categories:

- Local Hamiltonian Approaches

  – Divide-and-conquer

  Yang (1991), Yang & Lee (1995)

  – Recursion

  Baroni (1992)
  Jayanthi, Wu, Cocks, Luo, Xie, Menon and Yang (1997)

  – Chebyshev polynomial expansion

  Godecker & Colombo (1994)
  Baer & Head-Gordon, (1997)

- Variational Principles Approaches

  – Density Matrix

  Li, Nunes & Vanderbilt (1993), Daw (1993)
  Kohn (1996)
  Hernandez, E. & Gillan (1994)
  Xu & Scuseria (1996)
  Challacombe (1999)

  – Localized Molecular Orbitals

  Mauri, Galli & Car (1993)
  Ordejon, Drabold, Grumbach & Martin (1993)
  Kim, Mauri & Galli (1995)

  – Non-orthogonal Localized Molecular Orbitals

  Hierse & Stechel (1994)
  Yang (1997)

- Other: Pseudo-diagonalization, Stewart (1996), Density matrix, Helgaker (2000), Mazzioti(2001).

Martix Construction

Head_Gordon

Scuseria

Challacombe

SIESTA

Perez-Jorda and Yang

  $O(N)$ for matrix element integration: CPL, 241,469(1995).

  $O(N)$ for electrostatic interaction

   CPL, 247, 484(1995), JCP, 104, 8003(1996),JCP, 107, 1218(1997)

In DFT, the **first-order reduced density operator** $\hat{\rho}$ determines the electron density $\rho(\mathbf{r})$,

$$\rho(\mathbf{r}) = 2 \left\langle \mathbf{r} \left| \hat{\rho} \right| \mathbf{r} \right\rangle, \quad \rho(\mathbf{r}', \mathbf{r}) = \left\langle \mathbf{r}' \left| \hat{\rho} \right| \mathbf{r} \right\rangle = \sum_{i=1}^{N/2} \psi_i(\mathbf{r}')\psi_i^*(\mathbf{r}) = \left\langle \mathbf{r}' \right| \eta(\mu - \mathbf{h}) \left| \mathbf{r} \right\rangle$$

The total electronic energy functional of $\hat{\rho}$ is

$$E[\hat{\rho}] = T_s[\hat{\rho}] + \int v_{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}$$

$$+ E_{xc}[\rho(\mathbf{r})] + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}',$$

$E_{xc}[\rho(\mathbf{r})]$, the exchange-correlation energy functional

$T_s[\hat{\rho}] = 2 \, \mathrm{Tr}(\hat{t}\hat{\rho})$, the kinetic energy functional

The conventional approach represents $\hat{\rho}$ in terms of a set of $N/2$ **orthogonal and normalized orbitals** $\{\psi\}$:

$$\hat{\rho} = \sum_{i}^{N/2} |\psi_i\rangle\langle\psi_i| \, .$$

The Kohn-Sham equation:

$$\hat{h}\psi_i(\mathbf{r}) = [-\frac{1}{2}\nabla^2 + v_{eff}(\mathbf{r})]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}),$$

$v_{eff}(\mathbf{r})$: is the Kohn-Sham effective potential,

$$v_{eff}(\mathbf{r}) = v_{ext}(\mathbf{r}) + v(\mathbf{r}) + v_{xc}(\mathbf{r}),$$

$v(\mathbf{r})$: the electrostatic potential

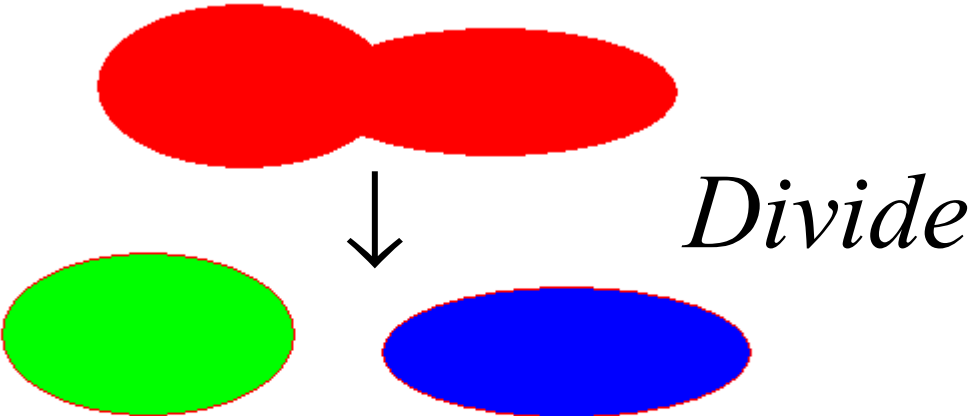$v_{xc}(\mathbf{r})$ the exchange-correlation potential

In LCAO

$$\psi_i(\mathbf{r}) = \sum_{\gamma}^{N_{basis}} C_{\gamma i} \chi_\gamma(\mathbf{r})$$

$$\rho(\mathbf{r}', \mathbf{r}) = \sum_{i=1}^{N/2} \psi_i(\mathbf{r}') \psi_i^*(\mathbf{r})$$

$$= \sum_{\gamma\delta}^{N_{basis}} \rho_{\gamma\delta} \chi_\delta(\mathbf{r}') \chi_\gamma^*(\mathbf{r})$$

which calculates the ρ for the *entire* molecule with the same expression on the Hamiltonian matrix $\mathbf{h}$.

$$\rho_{\gamma\delta} = \sum_{i=1}^{N/2} (c_{\gamma i})^* c_{\delta i}$$

$$= \sum_i f_\beta(\mu - \epsilon_i)(c_{\gamma i})^* c_{\delta i}$$

The idea of divide and conquer

Divide-and-Conquer Method-exploring the *local* nature of physical systems.

Molecular properties of a part of a molecule, described by the density matrix ρ associated with that part, depends *locally* on the atoms nearby.

one subsystem at a time.

- Partition functions are used to divide the density matrix into subsystem contributions (Mulliken population analysis)

- Each subsystem density matrix is determined separately with a local Hamiltonian.

- There is a common μ for the entire system to enable the electron flow between different subsystems and to ensure proper normalization of the density matrix.

Divide

We can define a partition matrix $p^\alpha_{\gamma\delta}$ for the subsystem α in the space of atomic orbitals. The partition matrix $p^\alpha_{\gamma\delta}$ needs to satisfy the normalization condition

$$\sum_\alpha p^\alpha_{\gamma\delta} = 1.$$

There is a simple way to construct such matrices; namely,

$$p^\alpha_{\gamma\delta} = \begin{cases} 1 & \text{if} \quad \gamma \in \alpha, \delta \in \alpha \\ 1/2 & \text{if} \quad \gamma \in \alpha, \delta \notin \alpha \\ 0 & \text{if} \quad \gamma \notin \alpha, \delta \notin \alpha \end{cases},$$

The density matrix, can be divided into subsystem contributions as follows:

$$\rho_{\gamma\delta} = \sum_\alpha p^\alpha_{\gamma\delta} \rho_{\gamma\delta} = \sum_\alpha \rho^\alpha_{\gamma\delta},$$

The Mulliken population analysis is thus used here in a reversed way—to construct the density matrix from subsystem contributions.

Approximate

The subsystem can be approximated efficiently with a *local* Hamiltonian that depends on the local physical space.

$$\rho_{\gamma\delta}^{\alpha} = p_{\gamma\delta}^{\alpha} \sum_i f_\beta(\mu - \epsilon_i^\alpha)(c_{\gamma i}^\alpha)^* c_{\delta i}^\alpha,$$

$c_{\gamma i}^\alpha$ and ϵ_i^α are local eigenvectors and eigenvalues of the subsystem:

$$(\mathbf{h}^\alpha - \mathbf{s}\epsilon_i^\alpha)c_i^\alpha = 0$$

The density matrix of the entire system is then,

$$\rho_{\gamma\delta} = \sum_\alpha \rho_{\gamma\delta}^\alpha = \sum_\alpha p_{\gamma\delta}^\alpha \sum_i f_\beta(\mu - \epsilon_i^\alpha)(c_{\gamma i}^\alpha)^* c_{\delta i}^\alpha,$$

and the chemical potential μ is determined by the normalization condition:

$$N = 2\operatorname{Tr}(\rho\mathbf{s}) = 2\sum_{\gamma\delta}^{N_{basis}} \rho_{\gamma\delta}\mathbf{s}_{\gamma\delta}$$

Use of the local Hamiltonian matrices $\mathbf{h}^\alpha$ for a subsystem α, introduces a truncation error.

In numerical implementation, $\mathbf{h}^\alpha$ consist of the Hamiltonian matrix elements associated with the atoms in the subsystem and the neighboring atoms. These neighboring atoms are called *buffer atoms.*

- Buffer atoms create a buffer zone for the better representation of the density matrix.

- A truncation radius R_b can be defined. Increasing R_b improves systematically the accuracy of the method.

The energy and gradient errors of divide-and-conquer method for the RP71955 protein molecule. The heat of formation of the original MOPAC calculation is -743.13604 Kcal/mole.

R_h is the distance used to cutoff the matrices $\mathbf{h}$.

R_b is the distance used to define the buffer atoms.

R_b	R_h =6.0	R_h =7.0	R_h =8.0	R_h =9.0
4.0	7.33E-03	-1.75E-01	-3.87E-01	-7.43E-01
	8.80E-01	2.14E+00	2.12E+00	2.11E+00
5.0	1.38E-02	-2.10E-02	-4.28E-02	-4.62E-02
	2.34E-01	2.40E-01	2.06E-01	3.04E-01
6.0	4.23E-03	3.20E-03	3.11E-03	3.10E-03
	8.95E-02	9.07E-02	9.08E-02	9.08E-02
7.0	2.01E-03	7.69E-04	6.39E-04	6.23E-04
	1.76E-02	1.30E-02	1.29E-02	1.29E-02
8.0	1.78E-03	5.02E-04	3.56E-04	3.37E-04
	1.38E-02	6.97E-03	6.77E-03	6.77E-03

Features

- Use the Mulliken population analysis in a reverse way to synthesize density matrix from fragments. Conventional usage is to analyze a density matrix.

- Linear scaling: the computational cost of a subsystem is independent of size of the entire system.

- can be applied to methods based on density matrix, like the Hartree-Fock, semiempirical, and DFT..
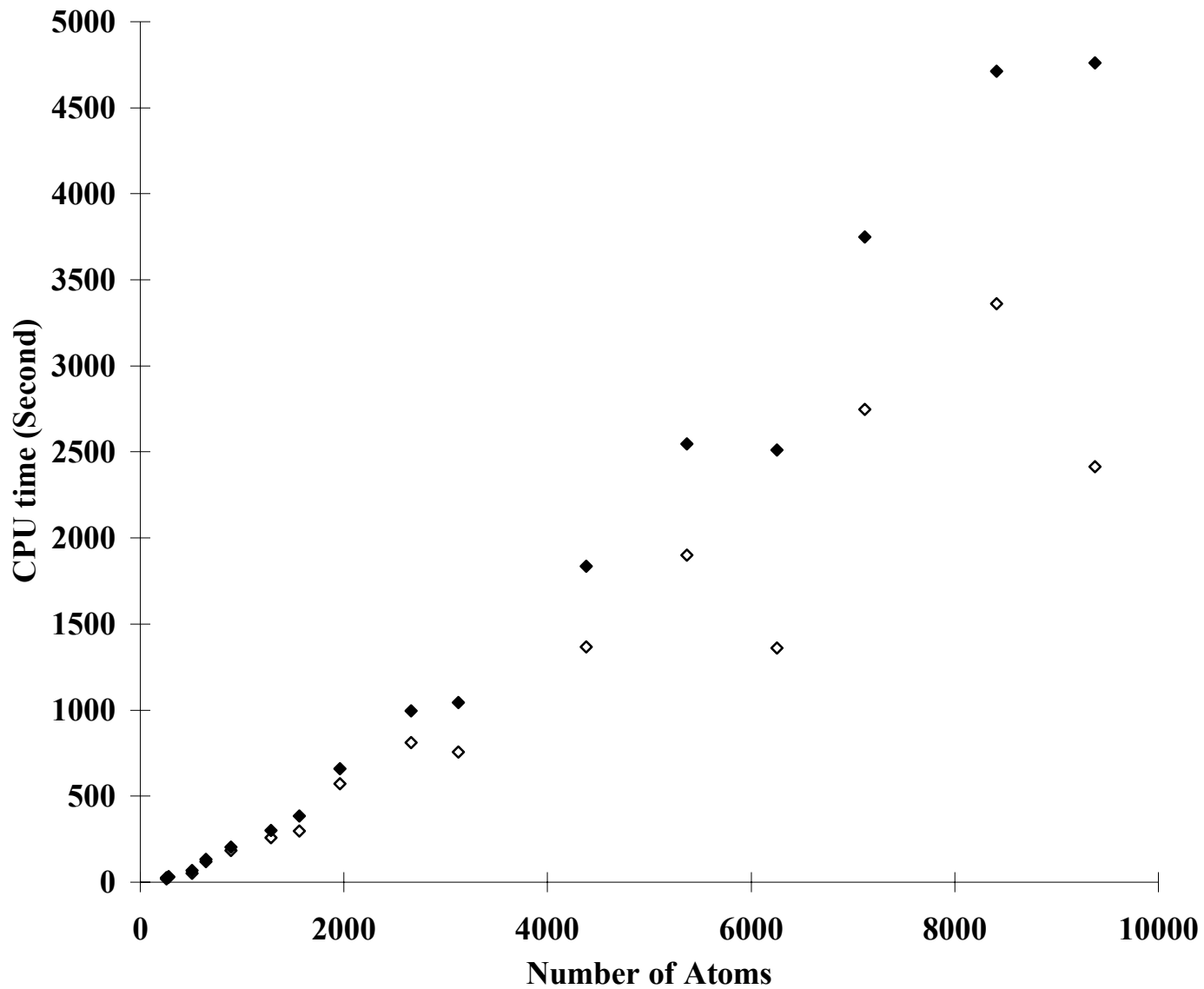
Applications with two models of solvent

- Modelling solvent with a continuum dielectric model

  - Calculation of solvation free energies of proteins and DNAs: Solvent polarization
    York, Lee and Yang, JACS Comm.,1996.
  - Solvent effects on the electronic Structure of Proteins and DNA
    York, Lee and Yang, Phys. Rev. Lett., 1998.
  - Catalytic reaction mechanism of cytidine deaminase
    Lewis, Carter, Hermans and Yang ,JACS,1998.

- Modeling solvent with molecular mechanics

 - Long Time Molecular dynamics simulations of proteins in water

Quantum Mechanics Simulation of Protein Dynamics on Long Time Scale

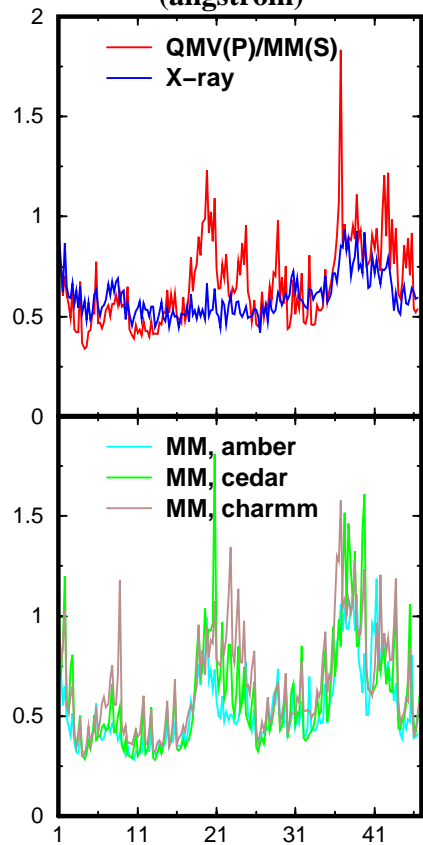Haiyan Liu, Marcus Elstner, Efthimios Kaxiras, Thomas Frauenheim, Jan Hermans and Weitao Yang

Protein structure and dynamics are the keys to a wide range of problems in biology. A molecular dynamics simulation of a protein, crambin, in solution for 350 ps has been performed.
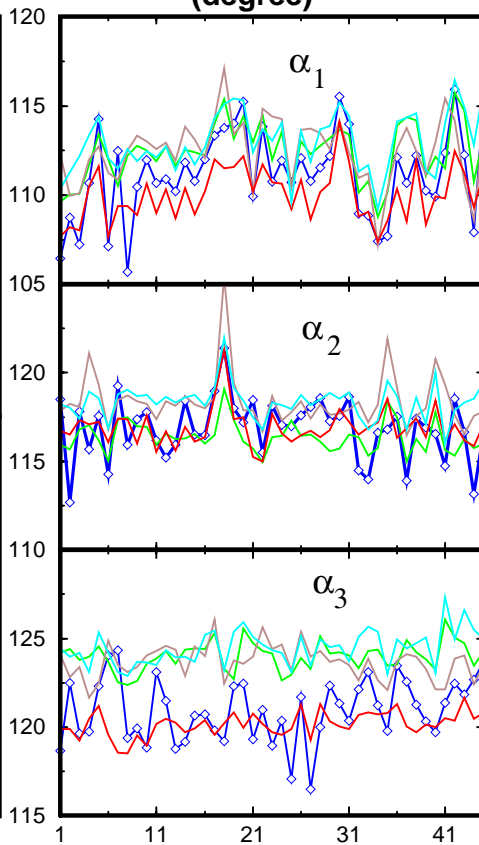
- The divide-and-conquer linear scaling approach.

- A hybrid quantum mechanical/molecular mechanical scheme that allows use of an explicit solvation model without an increase of the size of the quantum mechanical system.

- An efficient and accurate self-consistent-charge density functional theory-based tight-binding scheme (SCC-DFTB).

- The explicit incorporation of the long-range van der Waals energy and forces left out by the quantum mechanical model.
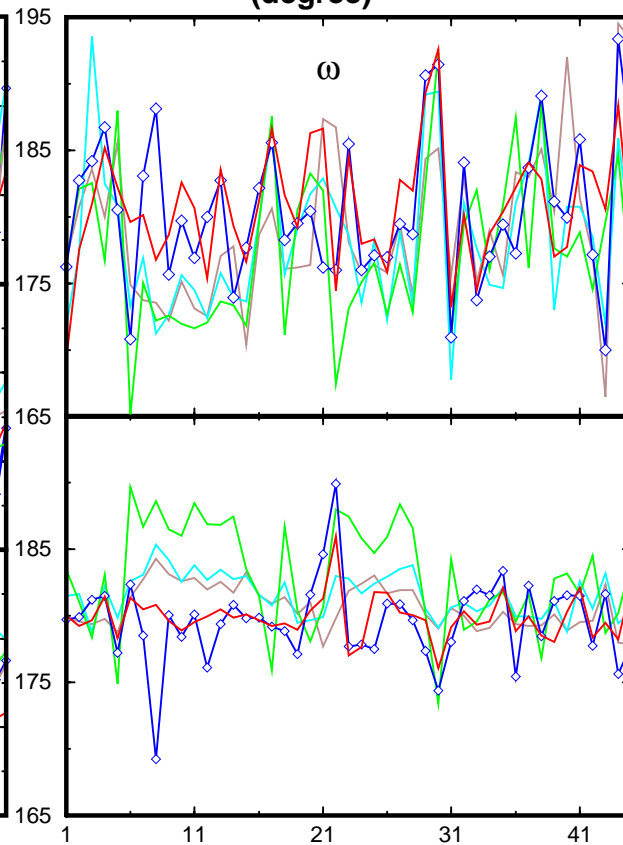
a RMSD from X–ray structure
(angstrom)

QM(P)/MM(S)

QMV(P)/MM(S)

time (ps)

b
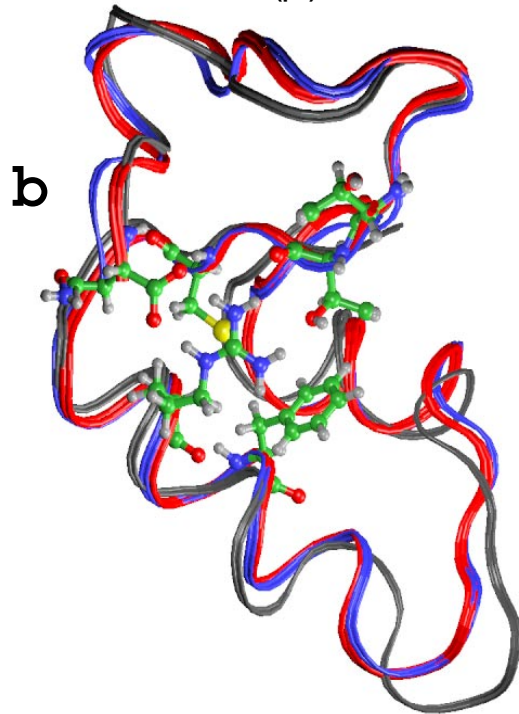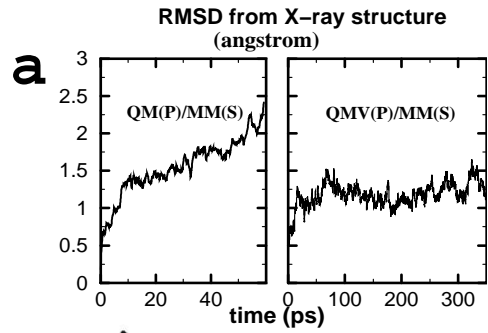
a. RMS fluctuations
(angstrom)

QMV(P)/MM(S)
X−ray

MM, amber
MM, cedar
MM, charmm

b. bond angles
(degree)

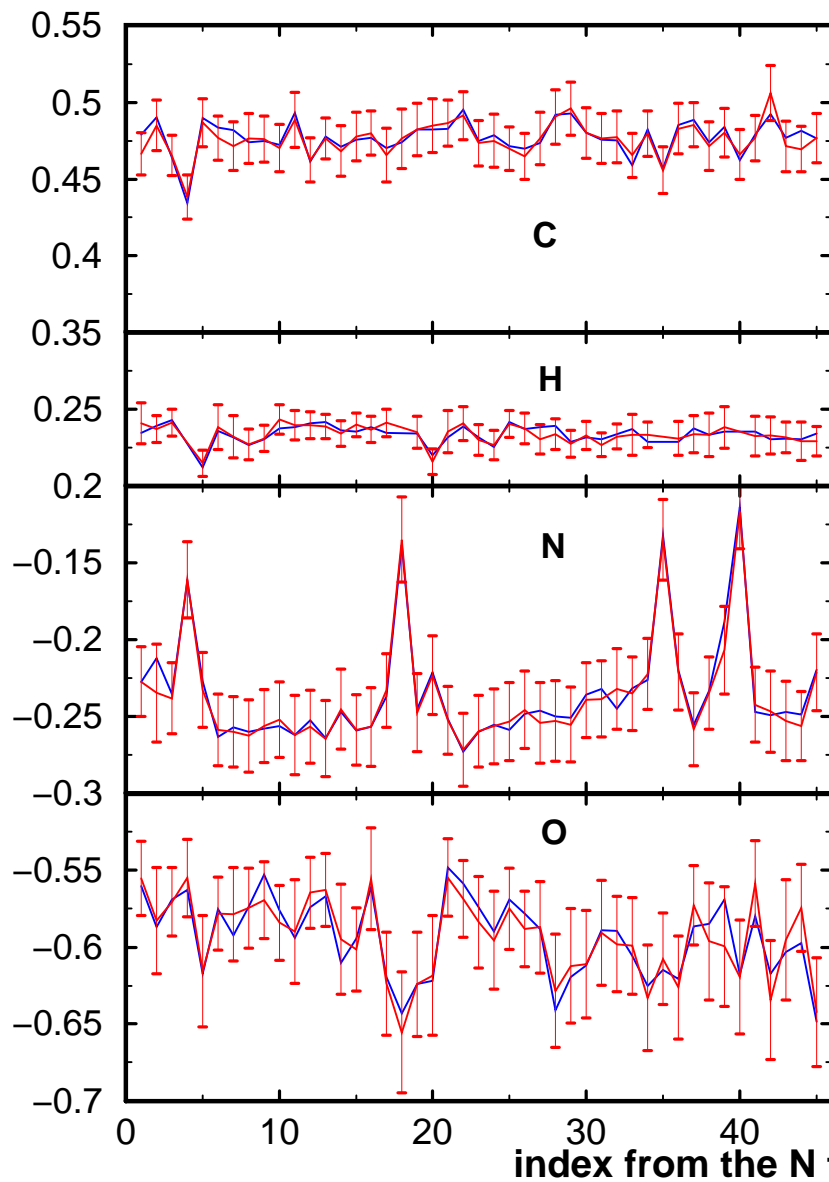α_1
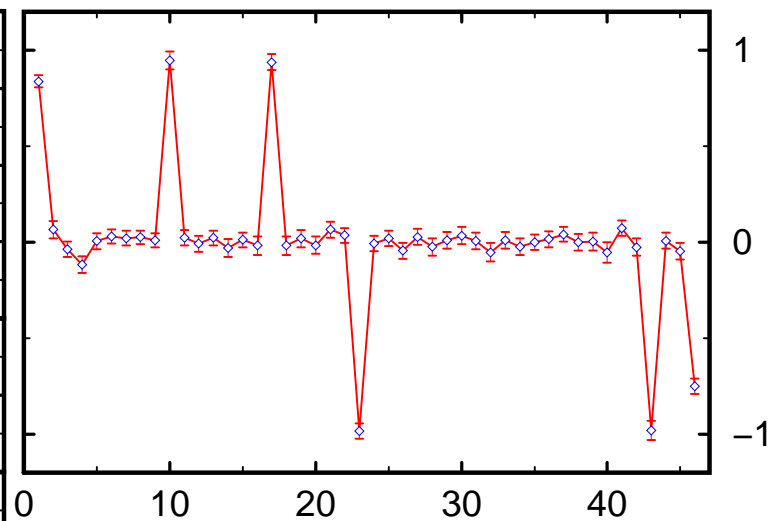
α_2

α_3

c. dihedral angles
(degree)

ω

index from the N terminal to the C terminal

a. backbone atomic charge

b. residue charge

c. cumulative charge

index from the N to the C terminal
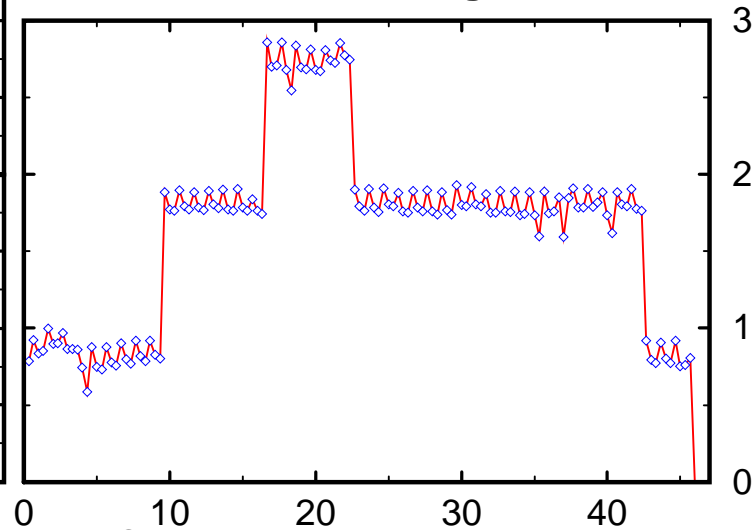
- Comparison with a recent very high resolution crystal structure of crambin (Jelsch et al., Proc Natl Acad Sci USA 2000;102:2246-2251) shows that **geometrical detail is better reproduced** in this simulation than when several alternate molecular mechanics force fields are used to describe the entire system of protein and solvent, even though the structure is no less flexible.

- Individual **atomic charges** deviate in both directions from "canonical" values, and some charge transfer is found between the N and C-termini.

- Importance of **van der Waals attractions**.

- The capability of simulating protein dynamics on and beyond the few hundred ps time scale with a demonstrably accurate quantum mechanical model will bring new opportunities to extend our understanding of a range of basic processes in biology such as molecular recognition and enzyme catalysis.

Non-orthogonal Localized Molecular Orbital (NOLMO) Approach to $O(N)$

The DFT ground state energy is given by **the absolute energy minimum principle**,

$$E(N) = \min_{\hat{\rho}} E[\hat{\rho}]$$

$\hat{\rho}$ has to satisfy the following three conditions:

$$\text{Hermitian}: \quad \hat{\rho} = \hat{\rho}^{\dagger}$$

$$\text{idempotency}: \quad \hat{\rho}^2 = \hat{\rho},$$

$$\text{normalization}: \quad 2\operatorname{Tr}(\hat{\rho}) = N.$$

Or

$$\hat{\rho} = \eta(\mu - \hat{h})$$

The conventional approach represents $\hat{\rho}$ in terms of a set of $N/2$ **orthogonal and normalized orbitals** $\{\psi\}$:

$$\hat{\rho}_{ortho} = \sum_{i}^{N/2} |\psi_i\rangle\langle\psi_i| .$$

The density operator $\hat{\rho}_{ortho}$ obeys all the three required conditions. Thus the minimization of energy over $\hat{\rho}$ can be achieved by minimization over the $N/2$ MO orbitals $\{\psi\}$, under the orthonormal constraint for the orbitals,

$$\langle\psi_i|\psi_j\rangle = \delta_{ij}.$$

One Way to Achieve Linear Scaling

1. **Truncation of the density matrix in the real space representation**.

$$\rho(\mathbf{r}', \mathbf{r}) = 2 \sum_{\gamma\delta}^{N_{basis}} \rho_{\gamma\delta}\chi_\delta(\mathbf{r}')\chi_\gamma^*(\mathbf{r})$$

$$\rho(\mathbf{r}', \mathbf{r}) \sim 0, \text{if } |\mathbf{r}' - \mathbf{r}| > R_c$$

$$\rho_{\gamma\delta} \sim 0, \text{if two atoms } \gamma \text{ and } \delta \text{ are far apart.}$$

The number of non zero elements in $\rho_{\gamma\delta}$ are only $O(N)$.

2. **Reformulation of the variational principle to eliminate the idempotency constraint**, $\hat{\rho}^2 = \hat{\rho}$.

The original variational principle has the idempotency constraint—difficult to impose.

3. **Conjugate gradient minimization.**

Local Energy Minimum Principle: Density Matrix

Li, Nunes and Vanderbilt, and Daw

McWeeny transformation $\hat{\rho}_m = 3\hat{\rho}^2 - 2\hat{\rho}^3$

Input a trial density operator $\hat{\rho}$, output a purified version $\hat{\rho}_m$, which is more idempotent.

For a **noninteracting** N-electron system with a Hamiltonian $\hat{h}$:

$$E_{LNV}[\hat{\rho}] = 2\,\mathrm{Tr}\left[\hat{\rho}_m\left(\hat{h} - \mu\right)\right] + \mu N = 2\,\mathrm{Tr}\left[\left(3\hat{\rho}^2 - 2\hat{\rho}^3\right)\left(\hat{h} - \mu\right)\right] + \mu N,$$

If $\epsilon_{N/2} < \mu < \epsilon_{N/2+1}$, then this functional has two properties:

(i) with the exact ground state $\hat{\rho}_o$, $E_{LNV}[\hat{\rho}_o]$ delivers the ground state energy

$$E^o(N) = 2\sum_i^{N/2} \epsilon_i,$$

(ii) $\Omega_{LNV}[\hat{\rho}]$ has a **local minimum** at $E^o(N)$ for variation with a Hermitian $\hat{\rho}$.

Local Energy Minimum Principle: Localized Orbitals

Mauri, Galli & Car, Ordejón, Drabold, Martin & Grumbach, Kim, Mauri & Galli

The functional is defined for a N-electron system described by a non-self-consistent Hamiltonian $\hat{h}$ and a chemical potential μ.

In terms of a set of M (possibly **linearly-dependent**) orbitals $\{\psi_i, i = 1, \ldots, M\}$, where $M \geq N/2$, it is

$$E_{KMG}[\{\psi\}] = 2\operatorname{Tr}[(\mathbf{H} - \mu\mathbf{S})(2 - \mathbf{S})] + \mu N,$$

where $\mathbf{H}$ and $\mathbf{S}$ are the $(M \times M)$ Hamiltonian and overlap matrix defined in the representation of the $\{\psi\}$,

$$H_{ij} = \langle \psi_i \mid \hat{h} \mid \psi_j \rangle, \ S_{ij} = \langle \psi_i | \psi_j \rangle$$

- The chemical potential is chosen such that $\epsilon_{N/2} < \mu < \epsilon_{N/2+1}$, and the energy functional has a **local minimum** at the ground state energy.

- Uses more localized orbitals than occupied states and has the chemical potential μ to allow the transfer of electrons. Free of the **multi-minimum problem**.

Improvements Needed

- All deal with **non-interacting electron systems**–repeated minimization needed for a self-consistent calculations.

- Local minimum principles only for the localized orbitals approaches

Absolute energy minimum principles

Yang, Phys. Rev. B. 56, 9294-9297 (1997).

- Two **absolute energy minimum principles** are developed for **first-principle** linear-scaling electronic structure calculations.

- One is with a normalization constraint and the other without any constraint.

- The density matrix is represented by a set of non-orthogonal localized orbitals and an auxiliary matrix.

- The number of localized orbitals is allowed to exceed the number of occupied orbitals.

Well Known:

Given a set of $N/2$ *linearly-independent* orbitals $\{\psi\}$, which are **not necessarily orthonormal**, then

$$\hat{\rho}_{N/2} = \sum_{i,j}^{N/2} |\psi_i\rangle \left(S^{-1}\right)_{ij} \langle\psi_j|,$$

where $\mathbf{S}^{-1}$ is the inverse of the overlap matrix $\mathbf{S}$, $S_{ij} = \langle\psi_i \mid \psi_j\rangle$.

$\hat{\rho}_{N/2}$ obeys all the above three constraints.

New:

Build a density operator from a set of arbitrary M (**possibly linearly-dependent**) orbitals $\{\psi_i, i = 1, \ldots, M\}$, where $M \geqslant N/2$.

$$\hat{\rho} = \sum_{i,j}^{M} |\psi_i\rangle \left(S^-\right)_{ij} \langle \psi_j |,$$

- $\mathbf{S}^-$ is the (1)-inverse of the overlap matrix $\mathbf{S}$, defined by

$$\mathbf{S}\mathbf{S}^-\mathbf{S} = \mathbf{S}.$$

The (1)-inverse is the simplest one of the several generalized inverses for a **singular** matrix.

- $\hat{\rho}$ defined is **idempotent** and **Hermitian** (two of the three constraints).

The only remaining constraint for the density operator is the normalization condition

$$2\,\text{Tr}(\hat{\rho}) = 2\,\text{Tr}(\mathbf{S}^-\mathbf{S}) = N,$$

or

$$2\,\text{rank}(\mathbf{S}) = N$$

$\text{rank}(\mathbf{S})$ is the rank of the $\mathbf{S}$ matrix

and $\text{rank}(\mathbf{S}) = \text{Tr}(\mathbf{S}^-\mathbf{S})$

rank is the number of independent column or row vectors of $\mathbf{S}$.

Consider non-interacting electrons here,

$$E(N) = \min_{\{\psi\}, \ 2 \ \mathrm{rank}(\mathbf{S})=N} 2 \operatorname{Tr}\left[\hat{h}\hat{\rho}\right]$$

with

$$\hat{\rho} = \sum_{i,j}^{M} |\psi_i\rangle \left(S^-\right)_{ij} \langle\psi_j|$$

Problem with $\mathbf{S}^-$: difficult to determine.

Solution: A variational approach.

$$\mathrm{Tr}(\mathbf{B}\mathbf{S}^-) = \min_{\mathbf{X}=\mathbf{X}^\dagger} \mathrm{Tr}[\mathbf{B}(2\mathbf{X} - \mathbf{X}\mathbf{S}\mathbf{X})]$$

That is, $\mathbf{S}^-$ is replaced by its variational estimate, quadratic in $\mathbf{X}$

$$(2\mathbf{X} - \mathbf{X}\mathbf{S}\mathbf{X})$$

- any *negative definite* matrix $\mathbf{B}$ such that

$$B_{ij} = \langle \psi_i \mid \hat{b} \mid \psi_j \rangle$$

with some operator $\hat{b}$.

- The auxiliary $M \times M$ matrix $\mathbf{X}$ is Hermitian.

Then, take η as a constant such that $\mathbf{B} = (\mathbf{H} - \eta\mathbf{S})$ is a *negative definite* matrix.

We obtain **the first variational principle**:

$$E(N) = \min_{\{\psi\},\ 2\ \mathrm{rank}(\mathbf{S})=N} 2\,\mathrm{Tr}\left[\hat{h}\hat{\rho}\right]$$

$$= \min_{\{\psi\},\ 2\ \mathrm{rank}(\mathbf{S})=N} 2\,\mathrm{Tr}\left[(\hat{h} - \eta)\hat{\rho}\right] + \eta N$$

$$\hat{\rho} = \sum_{i,j}^{M} |\psi_i\rangle \left(S^-\right)_{ij} \langle\psi_j|$$

$$E(N) = \min_{\{\psi\},\ 2\ \mathrm{rank}(\mathbf{S})=N} 2\,\mathrm{Tr}\left[(\mathbf{H} - \eta\mathbf{S})\mathbf{S}^-\right] + \eta N$$

$$= \min_{\{\psi\},\ 2\ \mathrm{rank}(\mathbf{S})=N}\ \min_{\mathbf{X}=\mathbf{X}^\dagger} 2\,\mathrm{Tr}\left[(\mathbf{H} - \eta\mathbf{S})(\mathbf{2X} - \mathbf{XSX})\right]$$

$$+ \eta N$$

using $\mathbf{2X} - \mathbf{XSX}$ for $\mathbf{S}^-$.

The only constraint is

$$2\ \text{rank}(\mathbf{S}) = N$$

implemented also by the variational expression:

$$\text{rank}(\mathbf{S}) = \text{Tr}(\mathbf{S}\mathbf{S}^-) = -\min_{\mathbf{X}=\mathbf{X}^\dagger} \text{Tr}[(-\mathbf{S})(2\mathbf{X} - \mathbf{X}\mathbf{S}\mathbf{X})]$$

The **first variational principle** is

$$E(N) = \min_{\{\psi\},\, 2\,\mathrm{rank}(\mathbf{S})=N}\; \min_{\mathbf{X}=\mathbf{X}^{\dagger}} 2\,\mathrm{Tr}\left[(\mathbf{H}-\eta\mathbf{S})(2\mathbf{X}-\mathbf{X}\mathbf{S}\mathbf{X})\right]$$
$$+\,\eta N$$

The **second variational principle** is for fixed-μ, obtained by releasing the constraint on the number of electrons.

$$E(N) = \min_{\{\psi\}}\;\Big\{ \min_{\mathbf{X}=\mathbf{X}^{\dagger}} 2\,\mathrm{Tr}\left[(\mathbf{H}-\eta\mathbf{S})(2\mathbf{X}-\mathbf{X}\mathbf{S}\mathbf{X})\right]$$
$$+\; 2(\eta-\mu)\mathrm{rank}(\mathbf{S})\Big\} + \mu N$$

Features:

- **absolute minimum principles**: can provide more robust minimization algorithms.

- extension to **self-consistent calculations**.

- the number of localized orbitals is permitted to exceed the number of occupied orbitals.

- the minimizing orbitals are also allowed to be non-orthogonal, which makes optimal localization.

- for accurate calculations, NLMO has much less number of variables than density matrix.

- In the special case of $M = N/2$, the orbitals $\{\psi_i, \ i = 1, \ldots, N/2\}$ are *linearly independent*. we obtain the variational principle of Mauri *et al.* and Ordejon *et al.*

- ignoring the minimization procedures, the second variational principle becomes the local-minimum variational expression of Kim *et al.* with generalization allowing the orbitals to be non-orthogonal.

Results for alanine 24

Results for alanine 24

The spread of the Nonorthogonal Localized Molecular Orbitals (NOLMO) ϕ_i

Liu and Yang, J. Chem. Phys., 112, 1634–1644(2000).

$$\Omega = \sum_i^{N/2} \langle \phi_i | (\hat{\mathbf{r}} - \mathbf{r}_i)^2 | \phi_i \rangle$$

where

$$\mathbf{r}_i = \langle \phi_i | \hat{\mathbf{r}} | \phi_i \rangle.$$

–minimize Ω while keeping the orbitals linear independent.

Molecule	Spread ($a.u.$)		
	Canonical MO	OLMO (%)	NOLMO (%)
CH_4	17.50	10.58(39.5)	9.54(9.8)
NH_3	14.13	11.12(21.1)	8.40(24.5)
C_2	22.88	11.83(48.3)	9.47(19.9)
H_2O	10.11	8.18(19.1)	6.35(22.1)
C_2H_2	30.87	15.47(49.9)	13.24(14.4)
C_2H_4	41.94	18.42(56.1)	16.63(9.8)
C_2H_6	55.44	19.55(64.7)	16.70(14.6)
N_2H_2	29.98	15.18(49.4)	13.10(13.7)
N_2H_4	43.31	18.73(56.7)	14.48(22.7)
CH_3OH	34.94	16.40(53.1)	11.38(30.5)
CH_3NH_2	38.63	17.95(53.5)	14.36(20.0)
C_6H_6	231.4	54.58(76.4)	41.33(24.3)
CO	14.02	12.41(11.5)	9.62(22.5)
CO_2	57.21	18.45(67.8)	14.90(19.2)
LiF	8.74	7.69(12.0)	6.57(14.6)
N_2	19.71	11.73(40.5)	9.32(20.5)
B_2H_6	59.76	19.14(68.0)	17.59(8.9)
C_3H_6(cycle)	73.13	25.88(64.6)	22.25(14.0)
SF_6	275.47	46.12(83.3)	40.31(12.6)
$CH_3CH_2CH_3$	109.23	29.04(73.4)	24.26(16.5)
$CH_3CH_2CH_2CH_3$	182.83	39.59(78.3)	33.20(16.1)
$CH_3CONHCH_3$	168.56	40.30(76.1)	35.21(12.6)
Average %		54.5	17.2