Optimization for Electronic Structure Calculation

Zaiwen Wen

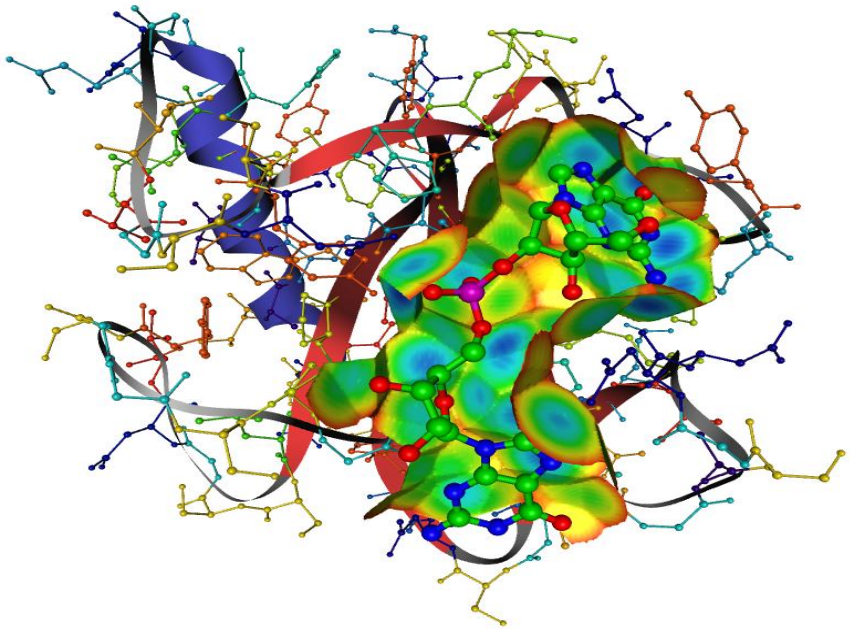
Beijing International Center For Mathematical Research
Peking University

Thanks: Yaxiang Yuan, Xin Liu, Xiao Wang, Xin Zhang, Jinwei Zhu, Aihui Zhou, Michael Ulbrich, Chao Yang

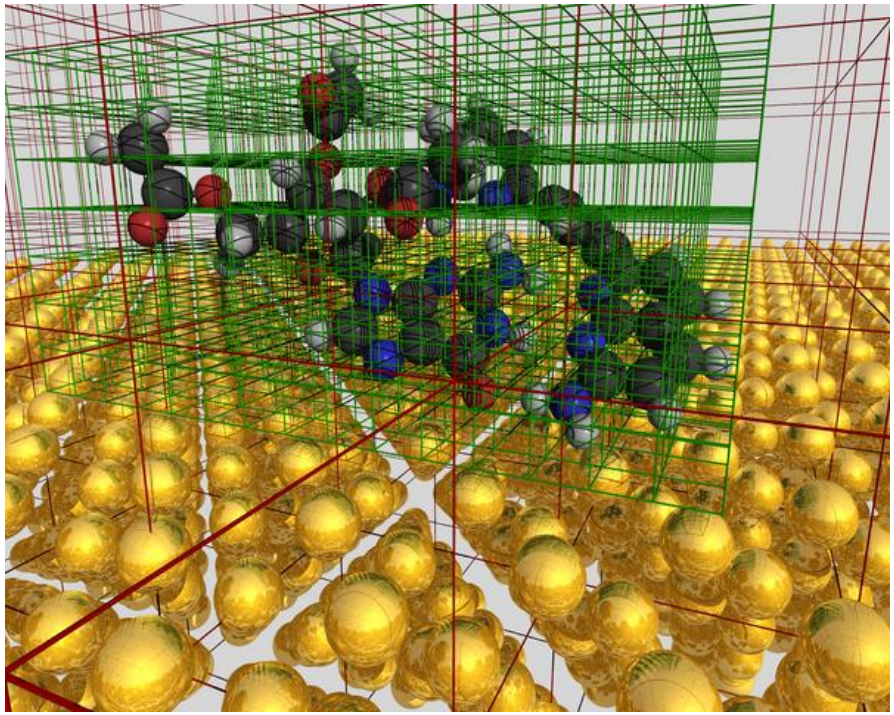
BICMR - PKU, 2014

Electronic Structure Calculation

N particle Schrodinger equation: Physics of material systems — atomic and molecular properties, almost correct (nonrelativistic) phyiscs is quantum mechanics



(a) Thanks: Hege et. al. ZIB Berlin



(b) Thanks: Reinhold Schneider

Numerical simulation of material on atomic and molecular scale

Electronic Structure Calculation

- Main goal: Given atomic positions $\{R_\alpha\}_{\alpha=1}^M$, compute the ground state electron energy $E_e(\{R_\alpha\})$.
- Ground state electron wavefunction $\Psi_e(r_1, \ldots, r_N; \{R_\alpha\})$:

$$\left(-\frac{1}{2} \sum_{i=1}^{N} \Delta_i - \sum_{\alpha=1}^{M} \sum_{j=1}^{N} \frac{Z_\alpha}{|r_i - R_\alpha|} + \frac{1}{2} \sum_{i,j=1, i \neq j}^{N} \frac{1}{|r_i - r_j|} \right) \Psi_e = E_e(\{R_\alpha\}) \Psi_e$$

- Curse of dimensionality: Computational work goes as 10^{3N}, where N is the number of electrons

Density Functional Theory (DFT)

- The unknown is simple — the electron density ρ
- Hohenberg-Kohn Theory (1964)
  - There is a unique mapping between the ground state energy from Schrödinger equation and the electron density
  - Exact form of the functional is unknown
- Independent particle model
  - Electrons move independently in an average effective potential field
  - Add correction for correlation
- Best compromise between efficiency and accuracy. Most widely used electronic structure theory for condensed matter systems.

Kohn-Sham Formulation

- Replace many-particle wavefunctions, Ψ_i, with single particle wavefunction, ψ_i
- Write Kohn-Sham total energy as

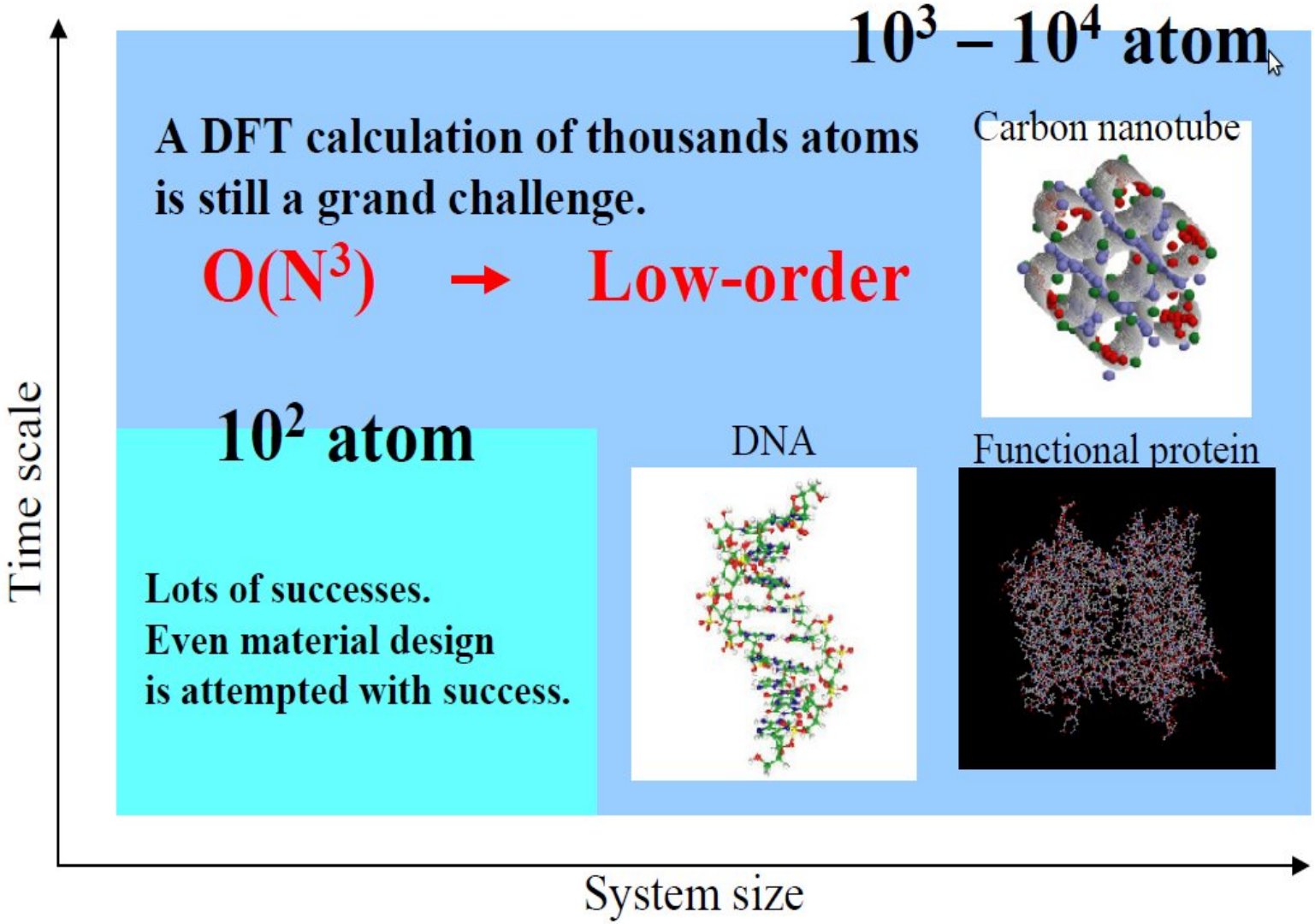
$$\begin{aligned} E_{KS}(\{\psi_i\}) &= \frac{1}{2}\sum_{i=1}^{n_e}\int_\Omega |\nabla\psi_i|^2 + \int_\Omega V_{ion}(\rho) \\ &\quad + \frac{1}{2}\int_\Omega \frac{\rho(r)\rho(r')}{|r-r'|}drdr' + E_{xc}(\rho) \end{aligned}$$

$$\rho(r) = \sum_{i=1}^{n_e} |\psi_i(r)|^2, \int_\Omega \psi_i\psi_j = \delta_{i,j}$$

- Exchange-correlation term, E_{xc}, contains quantum mechanical contribution, plus, part of K.E. not converged by first term when using single-particle wavefunctions

Towards Large-scale Simulation

Thanks: Taisuke Ozaki



Discretized Kohn-Sham Formulation

- Goal: find ground state energy/density by minimizing E_{KS}.
- Finite dimensional problem:

$$\min_{X^*X=I} E_{KS}(X) := E_{kinetic}(X) + E_{ion}(X) + E_{Hartree}(X) + E_{xc}(X),$$

where $X \in \mathbb{C}^{K\times N}$,

$$\begin{aligned}
E_{kinetic}(X) &= \tfrac{1}{2}\mathrm{tr}(X^*LX) \\
E_{ion}(X) &= \mathrm{tr}(X^*V_{ion}X) + \sum_i \sum_l |x_i^* w_l|^2 \\
E_{Hartree}(X) &= \tfrac{1}{2}\rho(X)^\top L^\dagger \rho(X) \\
E_{xc}(X) &= e^\top \epsilon_{xc}(\rho(X)), \quad e = (1,\ldots,1)^\top \\
\rho(X) &= \mathrm{diag}(XX^*) = (\sum_{j=1}^{N} |x_{ij}|^2)_{1\le i\le K}
\end{aligned}$$

KKT Conditions

- Lagrange function: $\mathcal{L}(X,\Lambda) = E_{KS}(X) - \frac{1}{2}\mathrm{tr}(\Lambda(X^*X - I))$
- First-order optimality conditions:

$$\begin{cases} \nabla_X \mathcal{L}(X,\Lambda) = 0. \\ X^*X = I, \end{cases} \Longrightarrow \begin{cases} H(X)X = X\Lambda, \\ X^*X = I. \end{cases}$$

- $\Lambda = X^*H(X)X$, not necessarily a diagonal matrix
- Kohn-Sham Hamiltonian:

$$H(X) := \tfrac{1}{2}L + V_{ion} + \sum_l w_l w_l^* + \mathrm{diag}\left(\Re(L^\dagger)\rho(X) + \partial_\rho \epsilon_{xc}(\rho(X))^\top e\right).$$

Orbital Free DFT (OFDFT)

- Expresses the system by only using the charge density
- Avoids computing N eigenpairs
- Pros: main group elements and nearly-free-electron-like metals
- Cons: not for covalently bonded and ionic systems
- Orbital Free total energy:

$$E_{OF}(\rho) = T_{OF}(\rho) + E_{ext}(\rho) + E_H(\rho) + E_{xc}(\rho) + E_{II}$$

$T_{OF}(\rho)$: kinetic energy density functional (KEDF)

$$\begin{aligned} T_{TFW}(\rho) &= C_{TF} T_{TF}(\rho) + \mu T_{vW}(\rho), \\ T_{LR}(\rho) &= T_{TF}(\rho) + \mu T_{vW}(\rho) + C_{TF} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} K(r - r') \rho^{\alpha}(r) \rho^{\beta}(r') \mathrm{d}r \mathrm{d}r' \end{aligned}$$

Other terms are the same as KSDFT

Orbital Free DFT (OFDFT)

- Variational problem

$$\inf\ E_{OF}(\rho)\ \text{s.t.}\ \rho \in L^1(\mathbb{R}^3),\ \rho^{\frac{1}{2}} \in H^1(\mathbb{R}^3),\ \rho \geq 0,\ \int_{\mathbb{R}^3} \rho(r)\mathrm{d}r = N.$$

- KKT Conditions:

$$\begin{cases} \mathcal{H}_{OF}\varphi \triangleq \left(-\frac{\mu}{2}\Delta + \frac{\delta\left(T_{OF}(\rho) - \mu T_{vW}(\rho)\right)}{\delta\rho} + V_{eff}(\rho) \right)\varphi = \lambda\varphi, \\ \int_{\mathbb{R}^3} \varphi^2 = N, \end{cases}$$

- Discretized Form:

$$\min_{c\in\mathbb{R}^n} E_{OF}(\rho(c)), \quad \text{s.t.} \quad c^\top Bc = 1.$$

Self Consistent Field Iteration (SCF)

SCF Algorithm:

1. Find the p-smallest eigenpairs (X, Λ):

$$\begin{aligned} H(\rho_k)X &= X\Lambda \\ X^*X &= I \end{aligned}$$

2. Calculate $\rho_{out}(X) = \mathrm{diag}(XX^*)$.
3. $\rho_{k+1} = (1-\alpha)\rho_k + \alpha\rho_{out}$.
4. Increment k and go to step 1 until $\rho_{k+1} - \rho_k$ is small enough.

Our motivation:

- Computation for the linear eigenvalue problem can be expensive
- Convergence of SCF is not clear
- Optimization Algorithms for solving DFT directly?

Gradient-Type Approach: Wen and Yin '12

Consider

$$\min E(X), \quad \text{subject to} \quad X^\top X = I.$$

At iteration i

$$\boxed{X^{(i+1)} \leftarrow \text{Orthogonalize} \leftarrow X^{(i)} - \sigma W^{(i)} X^{(i)}}$$

$$\Downarrow$$

$$\boxed{X^{(i+1)} \leftarrow \text{solution } Y(\tau) \text{ of } Y = X^{(i)} + \tfrac{\sigma}{2} W^{(i)}(X^{(i)} + Y)}$$

- $W^{(i)}$ is a skew-symmetric matrix defined by

$$W^{(i)} = \nabla E(X^{(i)}) \left(X^{(i)}\right)^* - X^{(i)} \left(\nabla E(X^{(i)})\right)^*$$

- $[Y^{(i)}]'(0) = -W^{(i)} X^{(i)} =$ tangential part of $-\nabla E(X^{(i)})$

Understanding SCF: the Hessian of E_{KS}

Convenient scaling: $E_s(X) := \frac{1}{2}E_{KS}(X)$.

Gradient: $\nabla E_s(X) := H(X)X$.

Exact Hessian

Suppose that $\epsilon_{xc}(\rho)$ is twice differentiable with respect to ρ. Given a direction $S \in \mathbb{C}^{K\times N}$, the Hessian-direction product for $E_s(X)$ is

$$\nabla^2(E_s(X))[S] = H(X)S + \mathrm{diag}\left(J\left((\bar{X}\odot S + X\odot \bar{S})e\right)\right)X,$$

where $J = \Re L^\dagger + \partial_\rho^2(\epsilon_{xc}^\top e)$.

Note: The second part corresponds to $(H'(X)[S])X$.

Good news: the Hessian-direction product is not too expensive.

SCF from the Viewpoint of Optimization

see also: Yang Meza, Wang '07

The linear eigenvalue problem in each SCF iteration is equivalent to:

$$\min \quad q(X) := \tfrac{1}{2}\langle H(X_k)X, X\rangle \qquad \text{s.t.} \quad X^*X = I.$$

On the other hand, a direct calculation reveals:

$$\begin{aligned}\tfrac{1}{2}\langle H(X_k)X, X\rangle &= \Re\langle H(X_k)X_k, X - X_k\rangle \\ &+ \tfrac{1}{2}\Re\langle H(X_k)(X - X_k), X - X_k\rangle + \text{const.}\end{aligned}$$

- The second part $H'(X_k)[X - X_k]X$ is omitted in SCF.
- similar to Gauss-Newton methods

Our Goals: Provable global convergence + fast local convergence.

Levenberg-Marquardt Type Regularization

- SCF iteration is similar to Gauss-Newton (GN) method.
- Regularization of SCF by Levenberg-Marquardt type approach:

$$\begin{aligned} &\min\ m_k^L(X) := \tfrac{1}{2}\langle H(X_k)X, X\rangle + \frac{\tau_k}{2}\|X - X_k\|_F^2 \\ &\text{s.t. } X^*X = I, \end{aligned}$$

  with regularization parameter $\tau_k > 0$.

- First-order optimality conditions:
  The solution $X = X_{k+1}$ satisfies

$$(H(X_k) + \tau_k I)\, X_{k+1} = X_{k+1}\Lambda_{k+1} + \tau_k X_k \text{ and } X_{k+1}^* X_{k+1} = I,$$

  where $\Lambda_{k+1} = \Lambda_{k+1}^* \in \mathbb{C}^{N\times N}$ is a Lagrange multiplier.

Exact Hessian + Adaptive Regularization

- using the exact Hessian:

$$\begin{aligned} m_k^N(X_k + S) := & \Re\langle H(X_k)X_k, S\rangle + \tfrac{1}{2}\Re\langle H(X_k)S, S\rangle \\ & + \tfrac{1}{2}\Re\left\langle S, \mathrm{diag}(J((\bar{X}_k \odot S + X_k \odot \bar{S})e)X_k\right\rangle \\ & + \frac{\tau_k}{\nu}\|S\|_F^\nu, \end{aligned}$$

- $\frac{\tau_k}{\nu}\|S\|_F^\nu$: trust region like strategy for ensuring global convergence.
- Compute the regularized Newton step:

$$\begin{aligned} \min \quad & m_k^N(X) \\ \text{s.t.} \quad & X^*X = I. \end{aligned}$$

- Cartis, Gould, Toint '10, '11, '12 on cubic regularization

Convergence Results

- Assumption:
  The gradient $\nabla E_s(X) = H(X)X$ is Lipschitz on the convex hull of the Stiefel manifold $\{X\,;\, X^*X = I\}$.
- Let $G_k = \nabla E_s(X_k) = H(X_k)X_k$ and define

$$W_k = G_k X_k^* - X_k G_k^*$$

- Global Convergence Result:

$$W_l = 0 \;\text{ for some }\; l \geq 0 \quad \text{or} \quad \lim_{k\to\infty} \|W_k\|_F = 0.$$

Note: $W_k X_k$ = tangential part of G_k in the canonical inner product.

Formulating the KS Equation as a Fixed Point Map

- Nonlinear equations with respect to ρ as

$$\rho = \mathrm{diag}(X(\rho)X(\rho)^{\mathrm{T}}).$$

- X is determined by the eigenvalue problem:

$$\begin{cases} \hat{H}(\rho)X = X\Lambda, \\ X^{\mathrm{T}}X = I, \end{cases}$$

- the Hamiltonian matrix

$$\hat{H}(\rho) := \frac{1}{2}L + V_{ion} + \mathrm{Diag}(L^{\dagger}\rho) + \mathrm{Diag}(\mu_{xc}(\rho)^{\mathrm{T}}e)$$

Formulating the KS Equation as a Fixed Point Map

- The Hamiltonian matrix

$$H(V) := \frac{1}{2}L + V_{ion} + \mathrm{Diag}(V)$$

- The potential

$$V := \mathcal{V}(\rho) = L^{\dagger}\rho + \mu_{xc}(\rho)^{\mathrm{T}}e$$

- Nonlinear equations with respect to

$$\begin{cases} V = \mathcal{V}(F_{\phi}(V)), \\ F_{\phi}(V) = \mathrm{diag}(X(V)X(V)^{\mathrm{T}}). \end{cases}$$

The Jacobian of the Fixed Point Maps

- Let $\{\lambda_i(V), q_i(V)\}$ be the eigenpairs of $H(V)$:

$$\lambda_1(V) \leq \ldots \leq \lambda_p(V) \leq \lambda_{p+1}(V) \leq \ldots \leq \lambda_n(V).$$

- The eigenvalue decomposition of $H(V)$:

$$H(V) = Q(V)\Pi(V)Q(V)^{\mathrm{T}},$$

- The function $F_\phi(V)$ in (19) is equivalent to

$$F_\phi(V) = \mathrm{diag}(Q(V)\phi(\Pi(V))Q(V)^{\mathrm{T}}),$$

where $\phi(\Pi) = \mathrm{Diag}(\phi(\lambda_1(V)), \phi(\lambda_2(V)), \ldots, \phi(\lambda_n(V)))$ and

$$\phi(t) := \begin{cases} 1 & \text{for } t \leq \frac{\lambda_p(V)+\lambda_{p+1}(V)}{2}, \\ 0 & \text{for } t > \frac{\lambda_p(V)+\lambda_{p+1}(V)}{2}. \end{cases}$$

The Jacobian of the Fixed Point Maps

- Suppose that $\lambda_{p+1}(V) > \lambda_p(V)$. Then the directional derivative of $F_\phi(V)$ at V is

$$\partial_V F_\phi(V)[Z] = \operatorname{diag}\left(Q(V)\left(g_\phi(\Pi(V)) \circ \left(Q(V)^{\mathrm{T}}\operatorname{Diag}(z)\, Q(V)\right)\right) Q(V)^{\mathrm{T}}\right),$$

where $g_\phi(\Pi(V)) \in \mathbb{R}^{n\times n}$ is defined as

$$(g_\phi(\Pi(V)))_{ij} = \begin{cases} \frac{1}{\lambda_i(V)-\lambda_j(V)} & \text{if } i \in \alpha_k,\ j \in \alpha_l,\ k \le r_p(V),\ l > r_p(V), \\ \frac{-1}{\lambda_i(V)-\lambda_j(V)} & \text{if } i \in \alpha_k,\ j \in \alpha_l,\ k > r_p(V),\ l \le r_p(V), \\ 0 & \text{otherwise.} \end{cases}$$

- The Jacobian of $\mathcal{V}(F_\phi(V))$ at V is

$$\partial_V \mathcal{V}(F_\phi(V))[Z] = J(F_\phi(V))\partial_V F_\phi(V)[Z], \quad \text{for all } Z \in \mathbb{R}^n.$$

Convergence of the SCF iteration

- SCF recursively computes:

$$H(V^i)X(V^{i+1}) = X(V^{i+1})\Lambda(V^{i+1}),$$
$$X(V^{i+1})^{\mathrm{T}}X(V^{i+1}) = I,$$

  and then the potential is updated as

$$V^{i+1} = \mathcal{V}(F_\phi(V^i)).$$

- The simple mixing scheme replaces (22) by updating

$$V^{i+1} = V^i - \alpha(V^i - \mathcal{V}(F_\phi(V^i)))$$

Convergence of the SCF iteration

- Assumption: the second-order derivatives of $\epsilon_{xc}(\rho)$:

$$\|\partial \mu_{xc}(\rho) e\|_2 \leq \theta, \quad \text{for all } \rho \in \mathbb{R}^n.$$

- It holds

$$\|\partial_V F_\phi(V)\|_2 \leq \frac{1}{\delta} \quad \text{and} \quad \|\partial_V \mathcal{V}(F_\phi(V))\|_2 \leq \frac{\|L^\dagger\|_2 + \theta}{\delta}.$$

Local Convergence of the SCF iteration

- Let V^* be a solution of the KS equation. Suppose that the eigenvalue gap satisfies

$$\delta > -\lambda_{\min}^* := -\min\{0, \lambda_{\min}(J(F_\phi(V^*)))\}.$$

There exists an open neighborhood Ω of V^*, such that the sequence $\{V^i\}$ generated by the simple mixing scheme using $V^0 \in \Omega$ and a step size

$$\alpha \in \left(0, \frac{2\delta}{||L^\dagger||_2 + \theta + \delta}\right)$$

converges to V^* with R-linear convergence rate no more than

$$\max\left\{\left(1 - \alpha\frac{\delta + \lambda_{\min}^*}{\delta}\right), \left(\alpha\frac{||L^\dagger||_2 + \theta + \delta}{2\delta} - 1\right)\right\}.$$

Local Convergence of the SCF iteration

- The condition $\delta > -\lambda^*_{\min}$ holds if

$$\max(\theta - \lambda_{\min}(L^\dagger), 0) < \delta.$$

- when $J(F_\phi(V^*))$ is positive semidefinite, we have $\lambda^*_{\min} = 0$ and the condition $\delta > -\lambda^*_{\min}$ holds.
- Convergence of using Fermi-Dirac distribution:

$$\frac{4}{\beta} > -\lambda^*_{\min},$$

where $\lambda^*_{\min} := \min\{0, \lambda_{\min}(J(F_\phi(V^*)))\}$.

Approximate Newton Methods

- The Jacobian of $\mathcal{V}(F_\phi(V))$ at V is

$$\partial_V \mathcal{V}(F_\phi(V))[Z] = J(F_\phi(V))\partial_V F_\phi(V)[Z], \quad \text{for all } Z \in \mathbb{R}^n,$$

where

$$\partial_V F_\phi(V)[Z] = \operatorname{diag}\left(Q(V)\left(g_\phi(\Pi(V)) \circ \left(Q(V)^{\mathrm{T}}\operatorname{Diag}(Z)\,Q(V)\right)\right)Q(V)^{\mathrm{T}}\right).$$

- Newton method

$$V^{i+1} = V^i - \alpha\left(I - J(F_\phi(V^i))\partial_V F_\phi(V^i)\right)^{-1}\left(V^i - \mathcal{V}\left(F_\phi(V^i)\right)\right),$$

- Approximate Newton method

$$V^{i+1} = V^i - \alpha\left(I - D^i\right)^{-1}\left(V^i - \mathcal{V}\left(F_\phi(V^i)\right)\right),$$

Approximate Newton Methods

- Setting $D^i := \tau^i J(\rho)$:

$$V^{i+1} = V^i - \alpha \left(I - \tau^i J(F_\phi(V^i)) \right)^{-1} \left(V^i - \mathcal{V} \left(F_\phi(V^i) \right) \right).$$

- Setting $D^i = \tau^i L^\dagger$:

$$V^{i+1} = V^i - \alpha \left(I - \tau^i L^\dagger \right)^{-1} \left(V^i - \mathcal{V} \left(F_\phi(V^i) \right) \right).$$

The Kerker preconditioner (pointed out by Lin Lin)

- Setting $D^i = L^\dagger W^i$ for a diagonal W^i:

$$V^{i+1} = V^i - \alpha \left(I - L^\dagger W^i \right)^{-1} \left(V^i - \mathcal{V} \left(F_\phi(V^i) \right) \right).$$

The method of elliptic preconditioner of Lin and Chao.

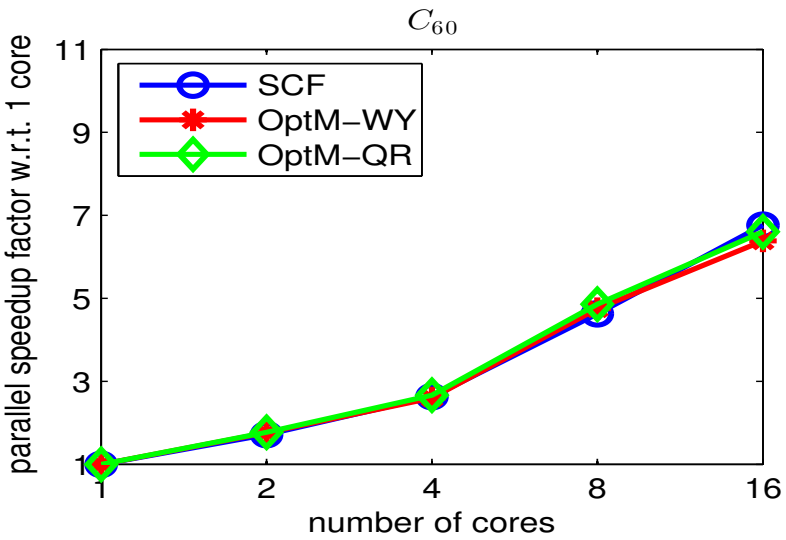
Speedup factors

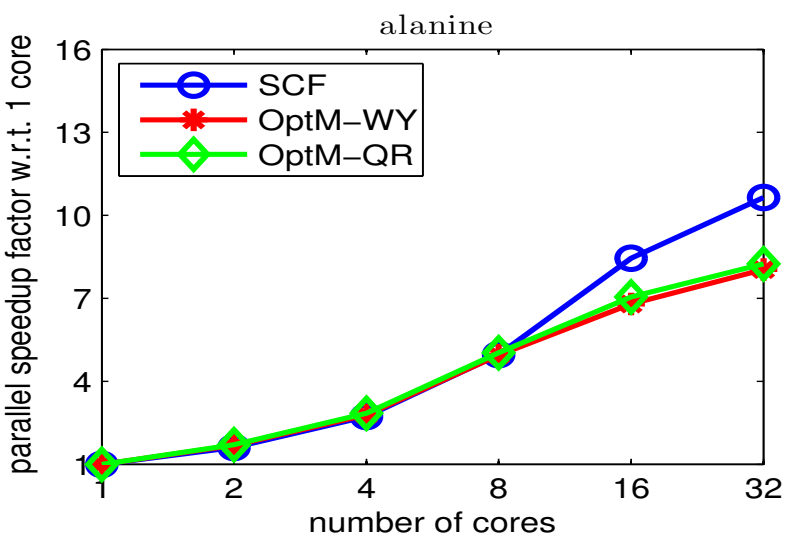
- Definition

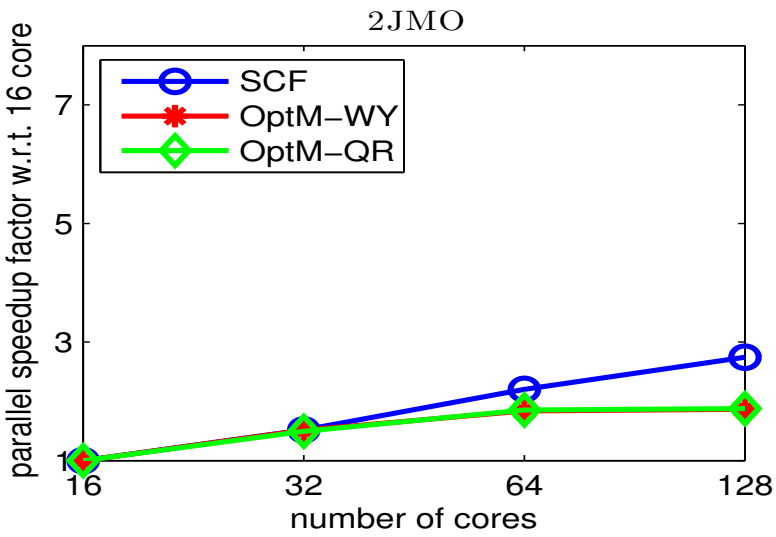
$$\text{speedup-factor}(k_0, k) = \frac{\text{wall clock time for a } k_0\text{-core run}}{\text{wall clock time for a } k\text{-core run}}.$$

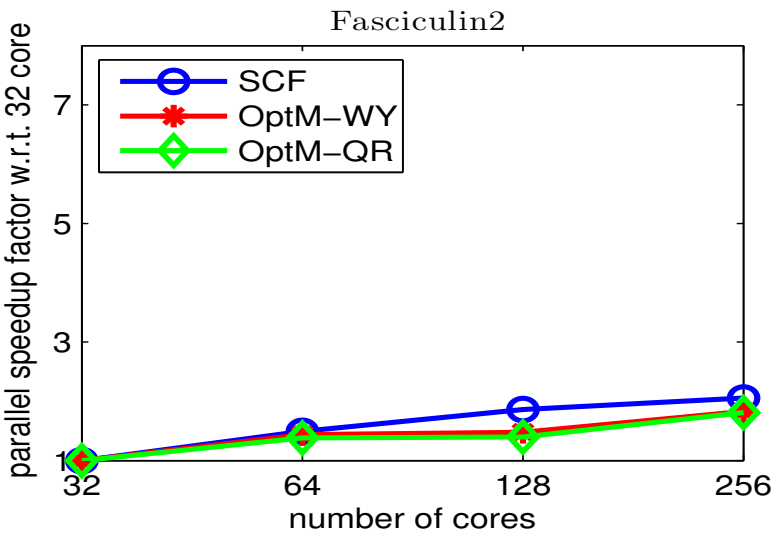
- T_0: the calculation of the total energy $E(X)$ and its partial derivative E_X.
- T_1: all other wall clock time in each algorithm.

Speedup factor with respect to T_0

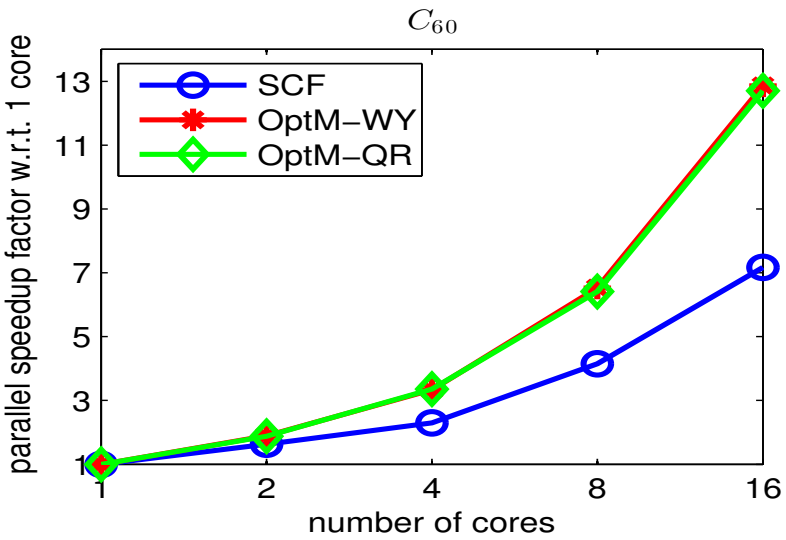


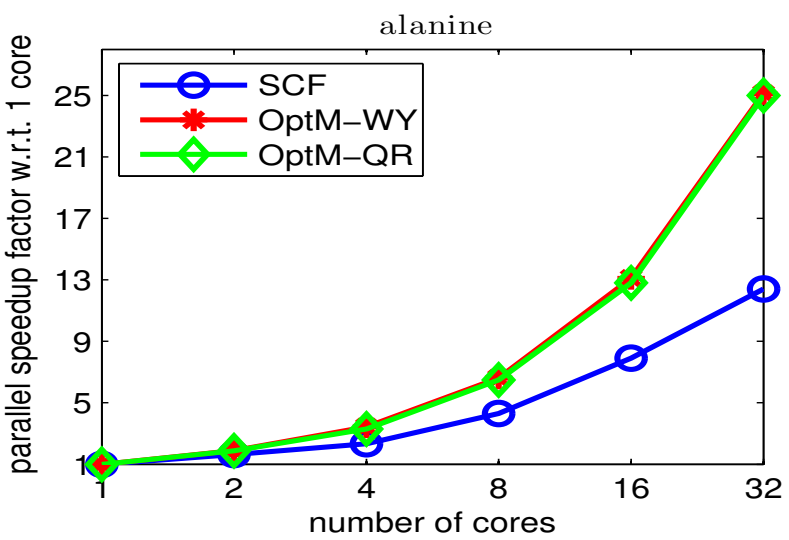


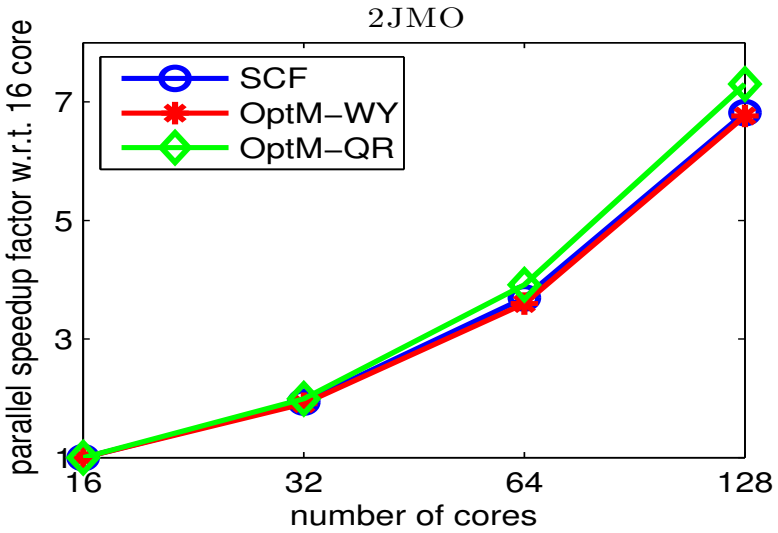


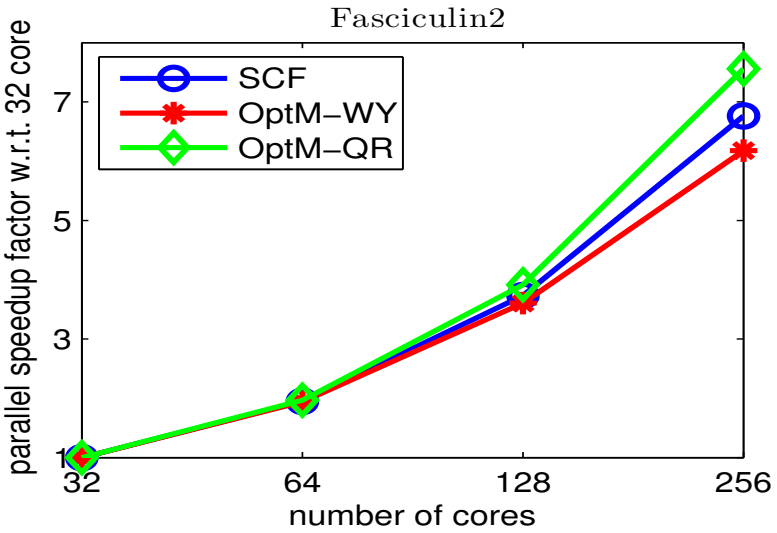


Speedup factor with respect to T_1

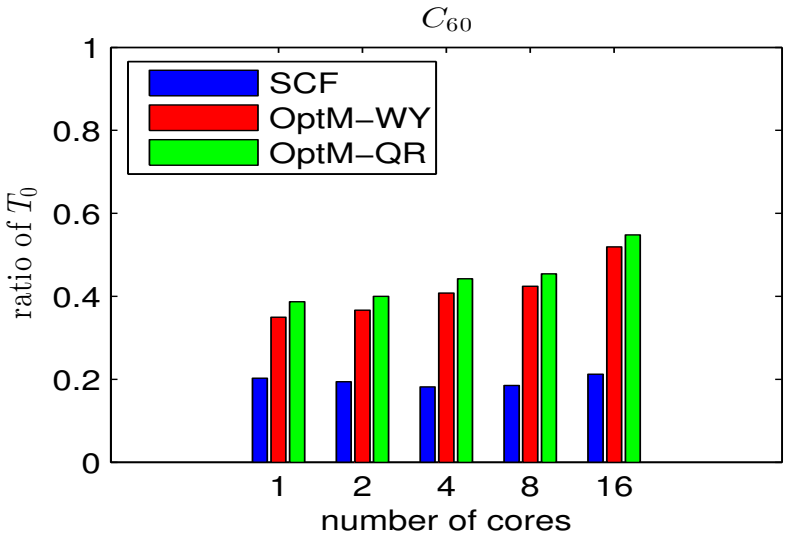


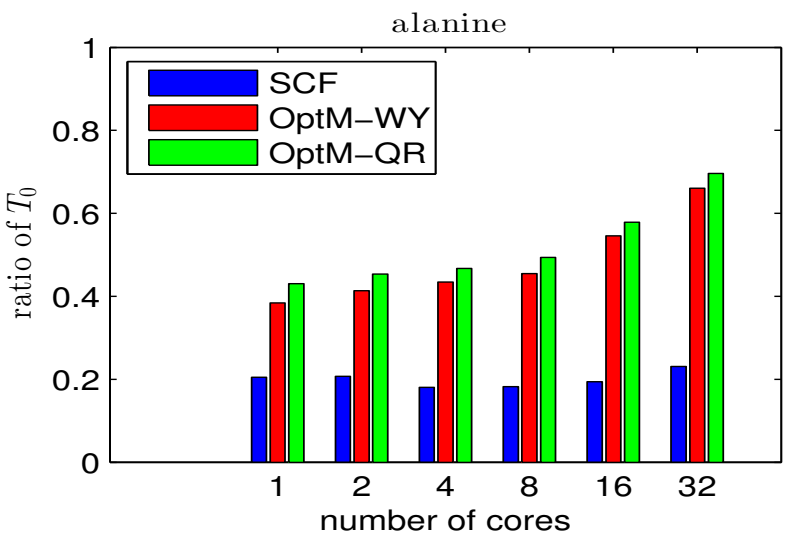


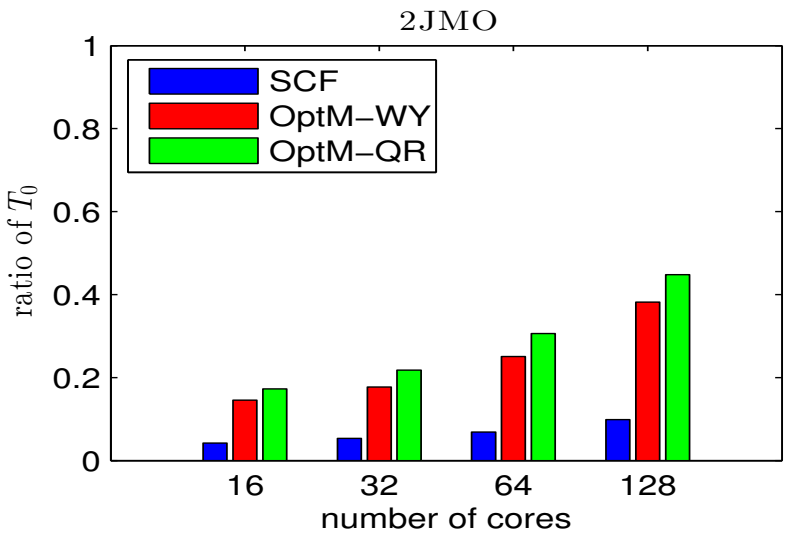


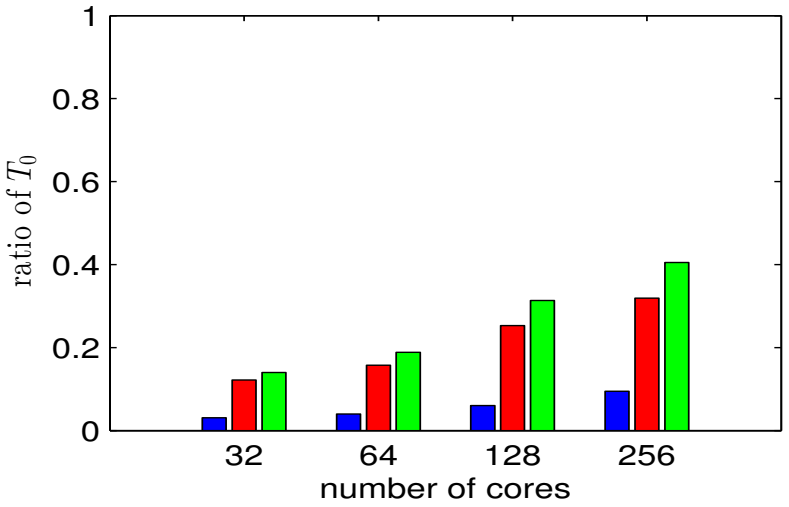


Ratio: $T_0/(T_0+T_1)$









Numerical Results on OFDFT

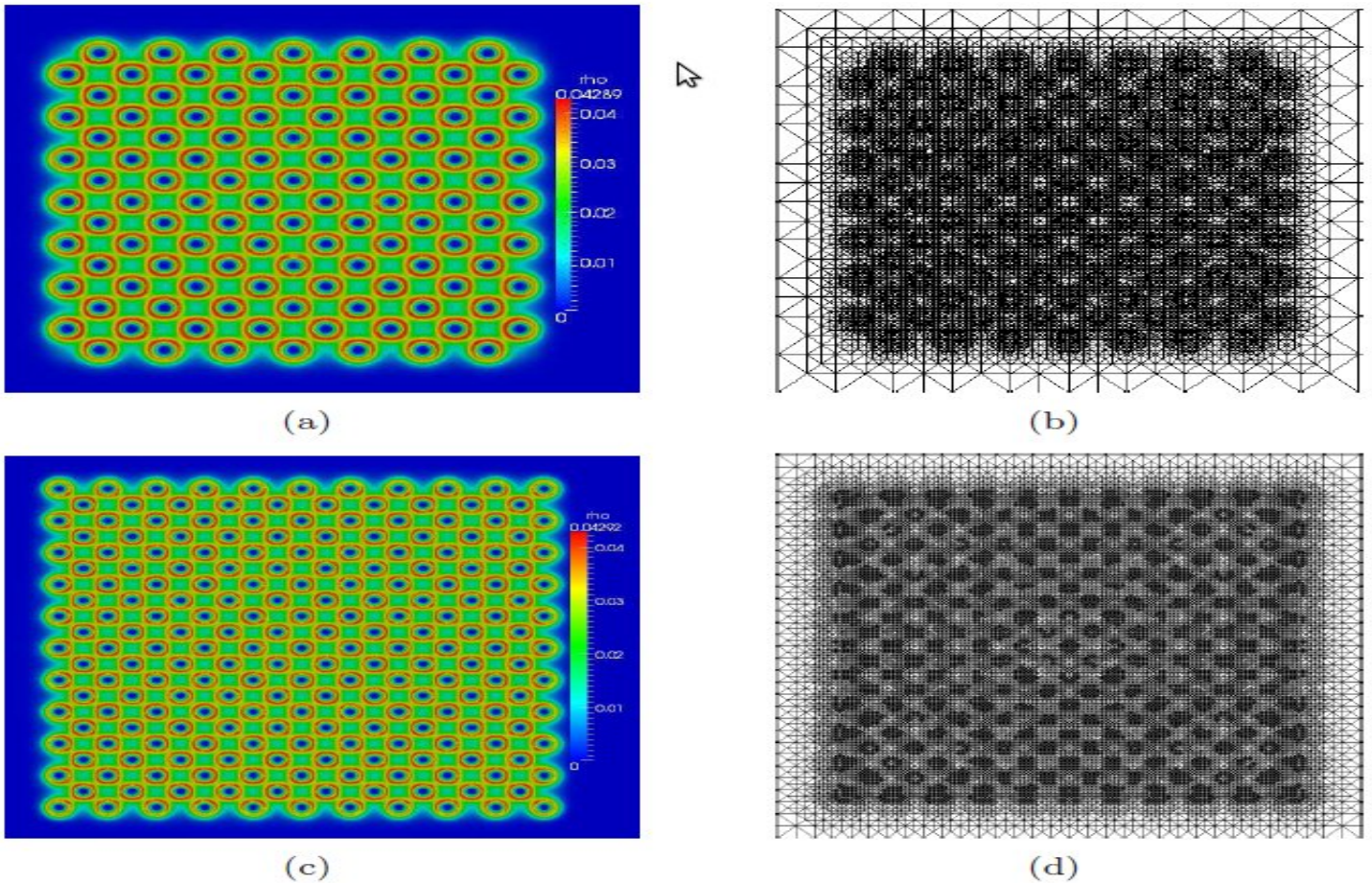
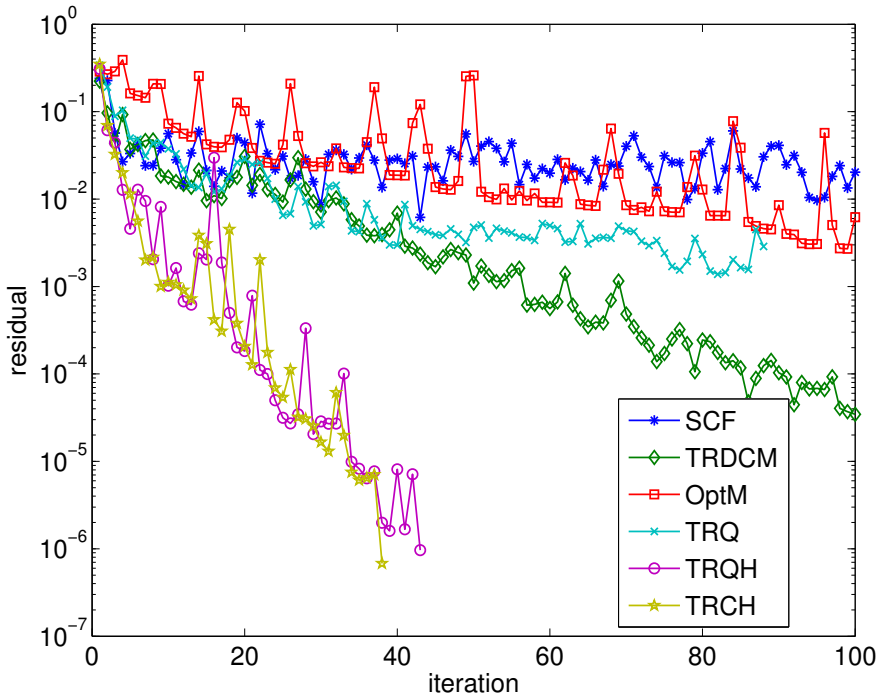
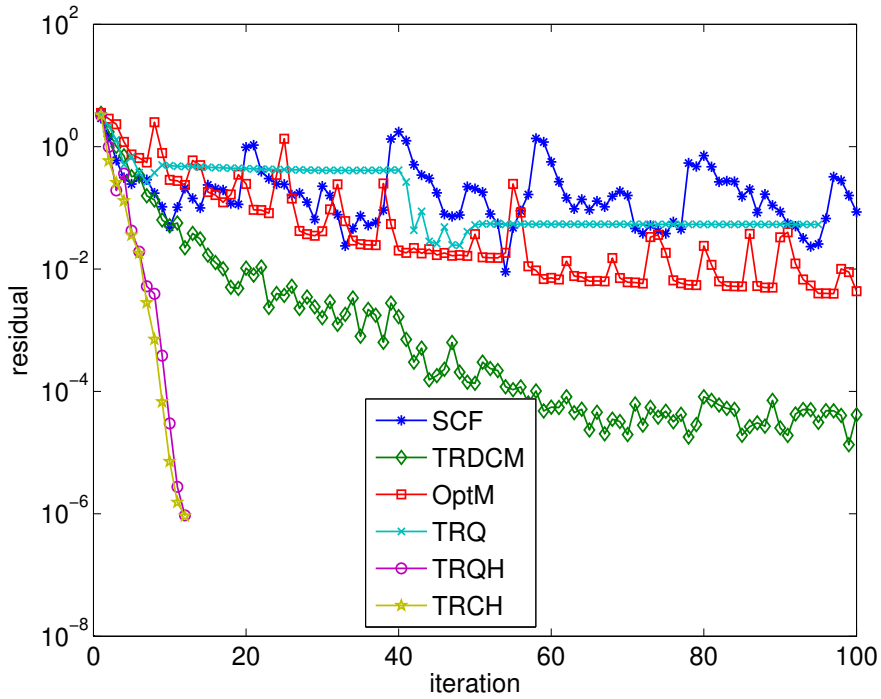


Figure: (a) and (c) are the contours of the ground state charge density at plane $z = 0$ for Al_{1688} and Al_{4631}, respectively. (b) and (d) are the corresponding adaptive mesh distribution of (a) and (c), respectively.

Numerical Results: Residual

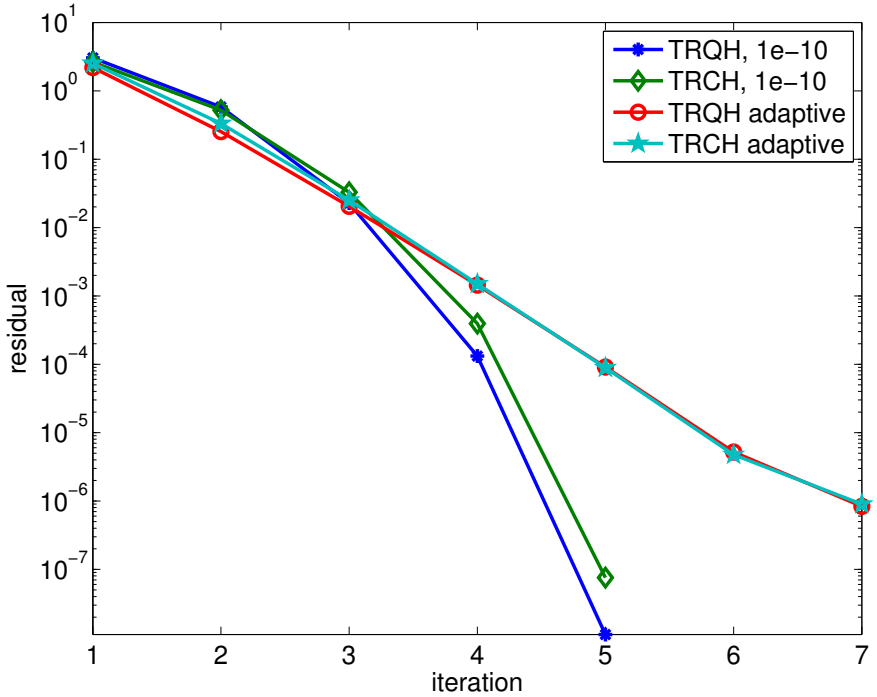


(a) al

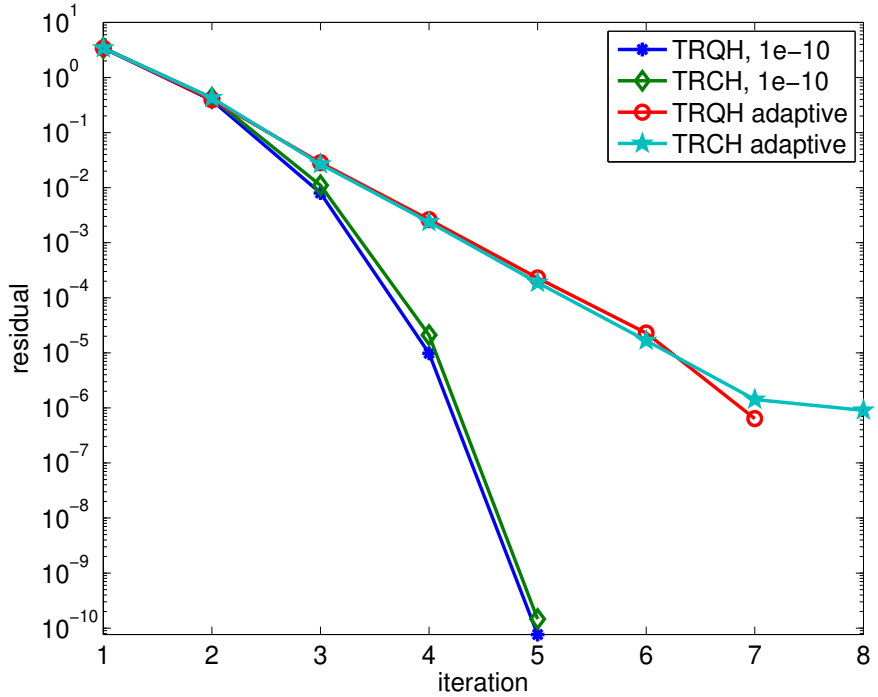


(b) graphene30

Quadratic Convergence is Observable



(a) alanine



(b) c12h26