Multiresolution Methods

Aiichiro Nakano

Collaboratory for Advanced Computing & Simulations
Department of Computer Science
Department of Physics & Astronomy
Department of Chemical Engineering & Materials Science
Department of Biological Sciences
University of Southern California

Email: anakano@usc.edu

Divide-&-conquer in continuum simulations
Discrete vs. Continuum Applications

- **Discrete simulation**: e.g. molecular dynamics (Newtonian mechanics)
- **Continuum simulation**: e.g. quantum dynamics (Schrödinger equation); *cf.* image processing
- **Multiresolution methods in the context of image processing**

*Figure 2*. Newton and Schrödinger’s cat. Previously, classical physics and quantum chemistry belonged to rivalling worlds. The Nobel Laureates in Chemistry 2013 have opened a gate between those worlds and have brought about a flourishing collaboration.
Haar Wavelet Basis

- One-dimensional “image”: $I[2] = (6, 2)$
- Smooth component: $s = (6 + 2)/2 = 4$
- Detailed component: $d = (6 - 2)/2 = 2$
- Wavelet decomposition: $I[] = (6, 2) = 4 \times (1,1) + 2 \times (1,-1)$

- Haar scaling function: 
  $\phi(x) = 1 (0 \leq x < 2); 0$ (otherwise)
- Haar wavelet function: 
  $\psi(x) = 1 (0 \leq x < 1); -1 (1 \leq x < 2); 0$ (otherwise)
Wavelet Decomposition

• **One-dimensional “image”:**
  \[ I[16]=(1,2,5,9,1,9,2,2,2,3,5,7,4,2,1,1) \]

• **Smooth component:**
  \[ s[i] = (I[2*i]+I[2*i+1])/2 \ (i=0,...,7) \]

• **Detailed component:**
  \[ d[i] = (I[2*i]-I[2*i+1])/2 \ (i=0,...,7) \]

• **Wavelet decomposition:**

<table>
<thead>
<tr>
<th>I[16]</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>9</th>
<th>1</th>
<th>9</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>4</th>
<th>2</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>s[8]</td>
<td>1.5</td>
<td>7.0</td>
<td>5.0</td>
<td>2.0</td>
<td>2.5</td>
<td>6.0</td>
<td>3.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d[8]</td>
<td>-0.5</td>
<td>-2.0</td>
<td>-4.0</td>
<td>0.0</td>
<td>-0.5</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Wavelet Decomposition

Wavelets = spatially localized waves also localized in wavenumber
Multiresolution Analysis

- **Recursive wavelet decomposition:**
  \[ I[16] \rightarrow s[8], d[8] \]
  \[ s[8] \rightarrow ss[4], sd[4] \]
  \[ sss[2] \rightarrow ssss[1], ssssd[1] \]

- **Recursive vector-subspace decomposition:**
  \[ V^0 = V^1 + W^1; s[8] \in V^1; d[8] \in W^1 \]
  \[ V^1 = V^2 + W^2; ss[4] \in V^2; sd[4] \in W^2 \]
  \[ \ldots \]

- **Multiresolution representation**
  \[ I[16] \rightarrow d[8], sd[4], ssd[2], sssd[1], sssss[1] \]
  or
  \[ V^0 = V^1 + W^1 \]
  \[ = V^2 + W^2 + W^1 \]
  \[ = V^3 + W^3 + W^2 + W^1 \]
  \[ = V^4 + W^4 + W^3 + W^2 + W^1 \]

Very smooth → Progressively more details
Wavelet Image Decomposition

- Alternate row & column transformations

- Wavelet transform is invertible, like Fourier transform

- Wavelets are localized in both real & reciprocal (wavenumber) spaces, unlike plane waves in Fourier transform

- Wavelet transform is $O(N)$—scalable tree, while fast Fourier transform (FFT) is $O(N\log N)$—slightly less scalable hypercube

- Image compression? Discard sub-threshold wavelet-expansion coefficients
Parallel Multiresolution Analysis

- Local with spatial decomposition at fine scales
- Subtree masters own coarse smooth components

• Retain detailed components? Let subtree slaves do (cf. hypercube quicksort)
Daubechies Wavelets

- Moving window for smooth & detailed filters

\[
\begin{bmatrix}
    c_0 & c_1 & c_2 & c_3 \\
    c_3 & -c_2 & c_1 & -c_0 \\
    & c_0 & c_1 & c_2 & c_3 \\
    c_3 & -c_2 & c_1 & -c_0 \\
    & & & & c_0 & c_1 & c_2 & c_3 \\
    & & & & c_3 & -c_2 & c_1 & -c_0 \\
    & & & & & & & & c_0 & c_1 & c_2 & c_3 \\
    & & & & & & & & c_3 & -c_2 & c_1 & -c_0 \\
\end{bmatrix}
\]

\[C_0 = (1+\sqrt{3})/4\sqrt{2} = 0.4829629131445341\]
\[C_1 = (3+\sqrt{3})/4\sqrt{2} = 0.8365163037378079\]
\[C_2 = (3-\sqrt{3})/4\sqrt{2} = 0.2241438680420134\]
\[C_3 = (1-\sqrt{3})/4\sqrt{2} = -0.1294095225512604\]
Iterative Solution of Linear Systems

\[ Ax = b \]

\[ A = L + D + U \]

- **Fixed-point equation**
  \[ x = D^{-1}[-(L+U)x + b] \]

- **Jacobi iteration**
  \[ x^{(n+1)} = D^{-1}[-(L+U)x^{(n)} + b] \]

\[ x_{i}^{(n+1)} = \frac{1}{a_{ii}} \left( - \sum_{j=1}^{N} a_{ij} x_{j}^{(n)} + b_{i} \right) \]
Multigrid Method

- **Residual equation:**
  \[ A^{(l)} (v + e) = -4\pi e^2 n \]
  \[ \rightarrow A^{(l)} v = -4\pi e^2 n + r \]
  \[ A^{(l)} e = -r \]

- **Smoothing:**
  \[ e \leftarrow \left[ 1 + Z^{(l)} A^{(l)} \right] e + Z^{(l)} r \]

- **Coarsening of residual & interpolation of error**

Achi Brandt

\[ A^{(l)} \]

\[ v \] Current guess
\[ e \] error vector
\[ r \] residual vector

V-cycle

\[ l = 3 \]
\[ l = 2 \]
\[ l = 1 \]
\[ l = 4 \]

Full multigrid

\[ l = 3 \]
\[ l = 2 \]
\[ l = 1 \]
Parallel Multigrid Method

- Domain decomposition with boundary-layer caching

- 2D computational & communication costs (isogranular or weak scaling)

\[ N \times N \text{ grids each on } P \times P \text{ processors: } T (N^2 P^2, P^2) = a \log NP + bN + cN^2 \]

Weak-scaling speedup & efficiency

\[ S_{P^2} = \frac{N^2 P^2 T (N^2, 1)}{N^2 T (N^2, P^2)} = \frac{P^2 (cN^2)}{a \log NP + bN + cN^2} = \frac{P^2}{1 + \frac{b}{cN} + \frac{a}{cN} \log NP} \]

\[ E_{P^2} = \frac{S_{P^2}}{P^2} = \frac{1}{1 + \frac{b}{cN} + \frac{a}{cN} \log NP} \]
Global Communications

All-to-all (hypercube): $O(N\log N)$

- Quicksort
- Fast Fourier transform

All-to-one (tournament): $O(N)$

- Global reduction
- Fast multipole method
- Multigrid method
- Wavelets
Solving the Idle Processor Problem

- Parallel superconvergent multigrid: Solve multiple coarse problems to accelerate the convergence [Frederickson & McBryan, '88]
Divide-\&-Conquer Algorithms

- **N-body problem**: $O(N^2) \rightarrow O(N)$
  > Space-time multiresolution molecular dynamics (MRMD): Fast multipole method \& symplectic multiple time stepping

- **Variable N-charge problem**: $O(N^3) \rightarrow O(N)$
  > Fast reactive force-field (F-ReaxFF) MD: Multilevel preconditioning

- **Quantum N-body problem**: $O(C^N) \rightarrow O(N)$
  > DC density functional theory (DC-DFT): Adaptive multigrids

Molecular Dynamics: N-Body Problem

- Newton’s equations of motion

\[ m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -\frac{\partial E_{MD}(\mathbf{r}^N)}{\partial \mathbf{r}_i} \quad (i = 1, \ldots, N) \]

- Reliable interatomic potential

\[ E_{MD} = \sum_{i<j} u_{ij}(r_{ij}) + \sum_{i,j<k} v_{ijk}(\mathbf{r}_{ij}, \mathbf{r}_{ik}) \]

- N-body problem
Long-range electrostatic interaction — \( O(N^2) \)

Evaluate \( V_{es}(x) = \sum_{j=1}^{N} \frac{q_j}{|x-x_j|} \) at \( x = x_i \) (\( i = 1, \ldots, N \))

- \( O(N) \) space-time multiresolution MD (MRMD) algorithm
1. Fast multipole method (FMM) [Greengard & Rokhlin, ’87]
2. Symplectic multiple time stepping (MTS) [Tuckerman et al., ’92]

1. **Clustering:** Encapsulate far-field information using multipoles

\[
V(x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left\{ \sum_{i=1}^{N} q_i r_i^l Y^m_l(\theta_i, \phi_i) \right\} \frac{Y^m_l(\theta, \phi)}{r^{l+1}}
\]

2. **Hierarchical abstraction:** Octree data structure

3. \(O(N)\) algorithm: Constant number of interactive clusters per octree node

---

**Atomistic stress tensor by a novel complex charge method (CCM)**

Reactive Force-Field (ReaxFF) MD: Variable $N$-Charge Problem

• Reactive bond order potential energy: $E_{\text{bond}}(\{r_{ij}\},\{r_{ijk}\},\{r_{ijkl}\},\{BO_{ij}\})$

→ Bond breakage & formation

• Charge-equilibration (QEq)

→ Charge transfer

Determine atomic charges

$\{q_i \mid i = 1, ..., N\}$ every MD step

to minimize $E_{\text{ES}}(\mathbf{r}^N, \mathbf{q}^N)$ with charge-neutrality constraint:

$\sum_i q_i = 0$

— Dense linear system: $M \mathbf{q} = -\mathbf{\chi}$

$O(N^3)$!

$$E_{\text{ES}}(\mathbf{r}^N, \mathbf{q}^N) = \sum_i \left( \chi_i q_i + \frac{1}{2} J_i q_i^2 \right) + \sum_{i<j} \int d\mathbf{x} \int d\mathbf{x}' \frac{\rho_i(q_i; \mathbf{x} - \mathbf{r}_i)\rho_j(q_j; \mathbf{x}' - \mathbf{r}_j)}{|\mathbf{x} - \mathbf{x}'|}$$
Fast Reactive Force-Field Algorithm

• $O(N)$ fast reactive force-field (F-ReaxFF) algorithm
  1) Fast multipole method
  2) Temporal locality, $q_i^{(\text{init})}(t+\Delta t) = q_i(t)$

• Multilevel preconditioned conjugate gradient (MPCG) method
  1) Split Coulomb matrix: $M = M_{\text{near}} + M_{\text{far}}$
  2) Sparse near-field preconditioner: $M_{\text{near}}^{-1} M q = -M_{\text{near}}^{-1} \chi$

Results:

• 20% speed up of convergence
• Enhanced data locality:
  Improved parallel efficiency
  $0.93 \rightarrow 0.96$ for 26.5M-atom Al$_2$O$_3$ on 64 Power nodes

Extended-Lagrangian RMD (XRMD)


\[ L_{\text{XRMD}} = L_{\text{RMD}} + \frac{\mu}{2} \sum_i \dot{\theta}_i^2 - \frac{\mu \omega^2}{2} \sum_i (\theta_i - q_i)^2 \]

- Auxiliary charge: dynamic variable
- Physical charge

- Extended-Lagrangian RMD (XRMD) achieves the same energy conservation as fully converged RMD but an order-of-magnitude faster

Nomura et al., Comput. Phys. Commun. 192, 91 ('15)

- Parallel efficiency 0.977 on 786,432 Blue Gene/Q cores for 67.6 billion atoms
Quantum Molecular Dynamics (QMD)

First molecular dynamics using an empirical interatomic interaction
A. Rahman, Phys. Rev. 136, A405 (’64)

Density functional theory (DFT)
Hohenberg & Kohn, Phys. Rev. 136, B864 (’64)
W. Kohn, Nobel chemistry prize, ’98

\[ O(C^N) \quad \rightarrow \quad O(N^3) \]

1 N-electron problem intractable
\[ \psi(r_1, \ldots, r_N) \]
\{\psi_i(r)|i = 1, \ldots, N\}

\[ \psi(r_1, \ldots, r_N) \leftarrow \arg\min E[\{R_I\}, \psi(r_1, \ldots, r_N)] \]

\[ M_I \frac{d^2}{dt^2} R_I = - \frac{\partial}{\partial R_I} E[\{R_I\}, \psi(r_1, \ldots, r_N)] (I = 1, \ldots, N_{\text{atom}}) \]

\(O(N)\) DFT algorithms

- **Divide-\&-conquer DFT** [W. Yang, Phys. Rev. Lett. 66, 1438 (’91);
  F. Shimojo et al., Comput. Phys. Commun. 167, 151 (’05); Phys Rev. B 77, 085103 (’08);
  Appl. Phys. Lett. 95, 043114 (’09); J. Chem. Phys. 140, 18A529 (’14)]

- **Quantum nearsightedness principle** [W. Kohn, Phys. Rev. Lett. 76, 3168 (’96)]

- **A recent review** [Bowler & Miyazaki, Rep. Prog. Phys. 75, 036503 (’12)]
Divide-\&-Conquer Density Functional Theory

- Overlapping spatial domains: $\Omega = \bigcup_\alpha \Omega_\alpha$
- Domain Kohn-Sham equations

$$\left( -\frac{1}{2} \nabla^2 + \hat{V}_{\text{ion}} + \hat{V}_{\text{H,xc}}[\rho_{\text{global}}(\mathbf{r})] \right) \psi_n^\alpha(\mathbf{r}) = \epsilon_n^\alpha \psi_n^\alpha(\mathbf{r})$$

- Global \& domain electron densities

$$\rho_{\text{global}}(\mathbf{r}) = \sum_\alpha \rho_\alpha(\mathbf{r}) \rho_\alpha(\mathbf{r})$$
$$\rho_\alpha(\mathbf{r}) = \sum_n [\psi_n^\alpha]^2 \Theta(\mu - \epsilon_n^\alpha)$$

Domain support function
$$\sum_\alpha \rho_\alpha(\mathbf{r}) = 1$$

Global chemical potential
$$N = \int d\mathbf{r} \rho_{\text{global}}(\mathbf{r})$$
Optimization of Divide- &- Conquer DFT

- **Computational parameters of DC-DFT** = domain size \( (l) \) + buffer thickness \( (b) \)

- **Complexity analysis to optimize the domain size** \( l \)

\[
l_* = \arg\min(T_{\text{comp}}(l)) = \arg\min\left(\left(\frac{L}{l}\right)^3 (l + 2b)^{3\nu}\right) = \frac{2b}{\nu - 1}
\]

Per-domain computational complexity of DFT = \( O(n^\nu) \): \( \nu = 2 \) or \( 3 \) \( (n < \text{or} > 10^3) \)

- **Error analysis**: Buffer thickness \( b \) is dictated by the accuracy requirement

\[
b = \lambda \ln \left( \max \{|\Delta \rho_\alpha(\mathbf{r})|; |\mathbf{r} \in \partial \Omega_\alpha\} \right) / \varepsilon \langle \rho_\alpha(\mathbf{r}) \rangle
\]

Decay length \( \rho_\alpha(\mathbf{r}) - \rho_{\text{global}}(\mathbf{r}) \) Error tolerance

cf. quantum nearsightedness [Kohn, Phys. Rev. Lett. 76, 3168 ("96)]
Lean Divide-&-Conquer (LDC) DFT

- Density-adaptive boundary potential to reduce the $O(N)$ prefactor
  \[ v_{ab}^{bc}(\mathbf{r}) = \int d\mathbf{r}' \frac{\partial v(\mathbf{r})}{\partial \rho(\mathbf{r}')} \left( \rho_\alpha(\mathbf{r}) - \rho_{\text{global}}(\mathbf{r}) \right) \approx \frac{\rho_\alpha(\mathbf{r}) - \rho_{\text{global}}(\mathbf{r})}{\xi} \]

- More rapid energy convergence of LDC-DFT compared with nonadaptive DC-DFT

512-atom amorphous CdSe

- Factor 2.03 (for $\nu = 2$) $\sim$ 2.89 (for $\nu = 3$) reduction of the computational cost with an error tolerance of $5 \times 10^{-3}$ a.u. (per-domain complexity: $n^\nu$)

Hierarchical Computing

- Globally scalable (real-space multigrid) + locally fast (plane wave) electronic solver
  \[ \hat{V}_{\text{H,xc}}[\rho_{\text{global}}(\mathbf{r})] \quad \{\psi_n^\alpha(\mathbf{r})\} \]

- Hierarchical band (i.e. Kohn-Sham orbital) + space + domain (BSD) decomposition

- Inter-domain

- Intra-domain

- Multigrid

- Fast Fourier transform

- Domain \( \Omega_\alpha \)

- Spiral FFT library
  [M. Puschel et al., *Proc. IEEE* 93, 232 (’05)]

[Diagram showing hierarchical computing with multigrid and Fourier transform techniques]
Floating Point Performance

- Transform from band-by-band to all-band computations to utilize a matrix-matrix subroutine (DGEMM) in the level 3 basic linear algebra subprograms (BLAS3) library
- Algebraic transformation of computations

Example: Nonlocal pseudopotential operation

\[ \hat{v}_{nl}|\psi_n^\alpha\rangle = \sum_{l}^{N_{\text{atom}}} \sum_{ij}^{L_{\text{max}}} |\beta_{i,l}\rangle D_{ij,l} \langle \beta_{j,l}|\psi_n^\alpha\rangle \quad (n = 1, ..., N_{\text{band}}) \]

\[ \Phi = [|\psi_1^\alpha\rangle, ..., |\psi_{N_{\text{band}}}^\alpha\rangle] \begin{pmatrix} \mathbf{A} \end{pmatrix} = [|\beta_{i,1}\rangle, ..., |\beta_{i,N_{\text{atom}}}\rangle] [\mathbf{D}(i,j)]_{I,J} = D_{ij,l} \delta_{IJ} \]

\[ \hat{v}_{nl}\Phi = \sum_{i,j}^{L} \tilde{\mathbf{B}}(i) \tilde{\mathbf{D}}(i,j) \mathbf{B}(j)^T \]

- 50.5% of the theoretical peak FLOP/s performance on 786,432 Blue Gene/Q cores (entire Mira at the Argonne Leadership Computing Facility)
- 55% of the theoretical peak FLOP/s on Intel Xeon E5-2665

K. Nomura et al., IEEE/ACM Supercomputing, SC14 ('14)
Parallel Performance

- Weak-scaling parallel efficiency is 0.984 on 786,432 Blue Gene/Q cores for a 50,331,648-atom SiC system
- Strong-scale parallel efficiency is 0.803 on 786,432 Blue Gene/Q cores

- 62-fold reduction of time-to-solution [441 s/SCF-step for 50.3M atoms] from the previous state-of-the-art [55 s/SCF-step for 102K atoms, Osei-Kuffuor et al., PRL ’14]

K. Nomura et al., IEEE/ACM Supercomputing, SC14 (’14)
Scalable Simulation Algorithm Suite

- 4.9 trillion-atom space-time multiresolution MD (MRMD) of SiO₂
- 8.5 billion-atom fast reactive force-field (F-ReaxFF) RMD of RDX
- 1.9 trillion grid points (21.2 million-atom) DC-DFT QMD of SiC

parallel efficiency 0.98 on 786,432 BlueGene/Q cores
H$_2$ Production from Water Using LiAl Particles

16,661-atom QMD simulation of Li$_{441}$Al$_{441}$ in water on 786,432 IBM Blue Gene/Q cores

K. Shimamura et al., Nano Lett. 14, 4090 ('14)

21,140 time steps (129,208 self-consistent-field iterations)
Globally informed local DC-DFT solutions are used in the recombine phase as compact bases to synthesize global properties in broad applications:

- **High-order inter-molecular-fragment correlation** [S. Tanaka et al., ’13]
- **Global frontier orbitals (HOMO & LUMO)** [S. Tsuneyuki et al., ’09, ’13]
- **Global charge migration** [H. Kitoh-Nishioka et al., ’12; C. Gollub et al., ’12]
- **Global exciton dynamics** [W. Mou et al., ’13]

Singlet Fission in Amorphous DPT

- Photo-current doubling by splitting a singlet exciton into 2 triplet excitons
- Singlet fission in mass-produced disordered organic solid $\rightarrow$ efficient low-cost solar cells
- Exp’l breakthrough: SF found in amorphous diphenyl tetracene (DPT)

- Divide-conquer-recombine nonadiabatic QMD (phonon-assisted exciton dynamics) + time-dependent perturbation theory (singlet-fission rate) + kinetic Monte Carlo calculations of exciton population dynamics in 6,400-atom amorphous DPT