

Multiresolution Methods

Aiichiro Nakano

*Collaboratory for Advanced Computing & Simulations
Department of Computer Science
Department of Physics & Astronomy
Department of Quantitative & Computational Biology
University of Southern California*

Email: anakano@usc.edu

Divide-&-conquer in continuum simulations



Discrete vs. Continuum Applications



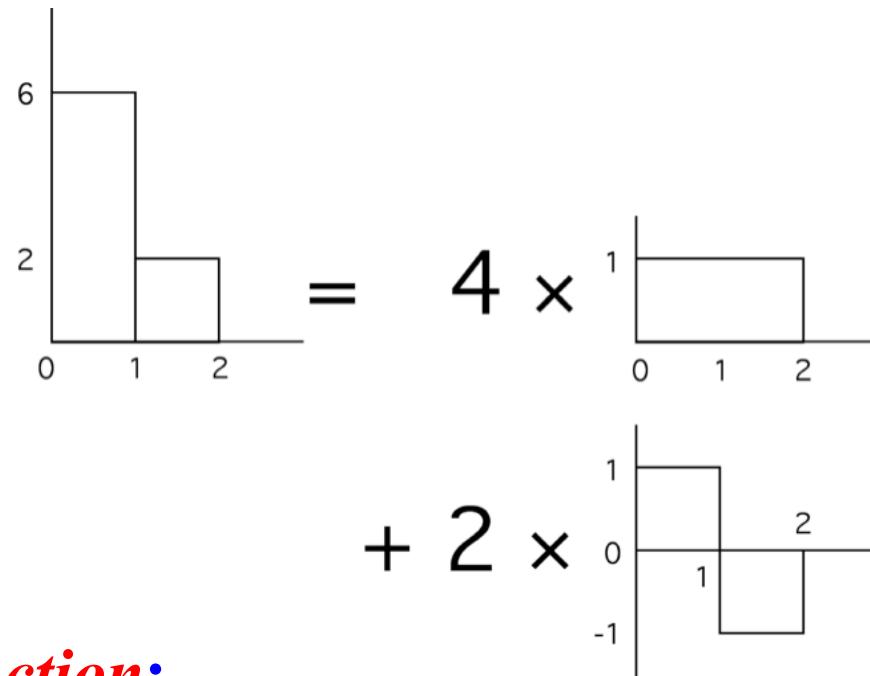
Figure 2. Newton and Schrödinger's cat. Previously, classical physics and quantum chemistry belonged to rivalling worlds. The Nobel Laureates in Chemistry 2013 have opened a gate between those worlds and have brought about a flourishing collaboration.

<https://www.nobelprize.org/prizes/chemistry/2013/summary/>

- 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

Haar Wavelet Basis

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



- Haar *scaling function*:
 $\phi(x) = 1 (0 \leq x < 2); 0 (\text{otherwise})$
- Haar *wavelet function*:
 $\psi(x) = 1 (0 \leq x < 1); -1 (1 \leq x < 2); 0 (\text{otherwise})$

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:

$$s[i] = (I[2*i] + I[2*i+1]) / 2 \quad (i=0, \dots, 7)$$

- Detailed component:

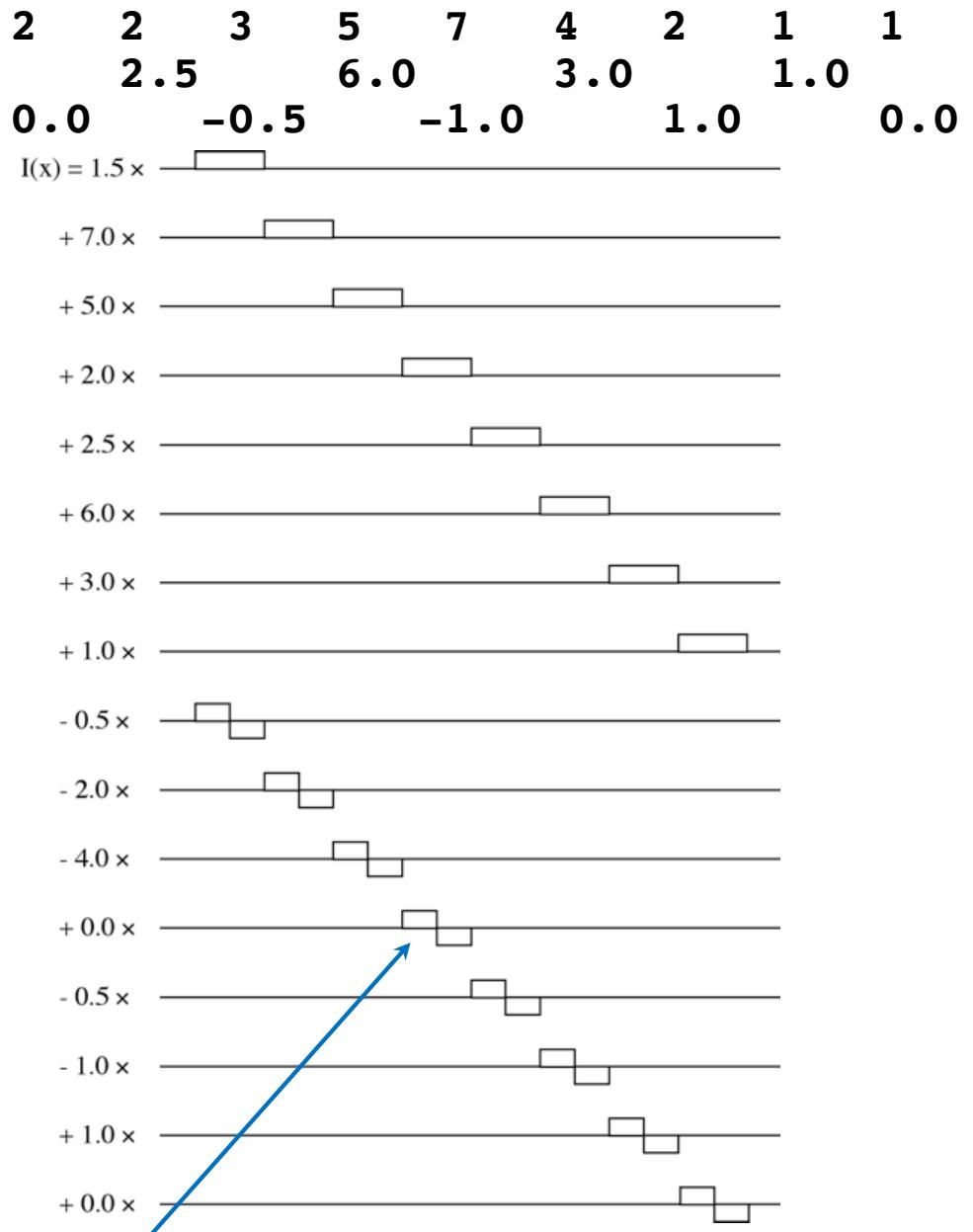
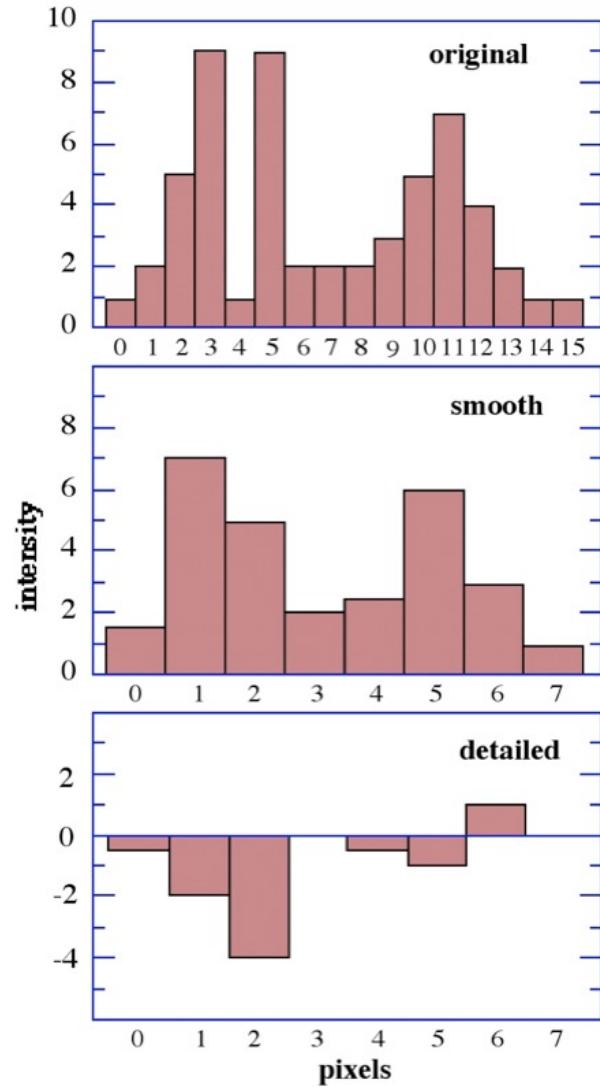
$$d[i] = (I[2*i] - I[2*i+1]) / 2 \quad (i=0, \dots, 7)$$

- Wavelet decomposition:

I[16]	1	2	5	9	1	9	2	2	2	3	5	7	4	2	1	1
s[8]	1.5		7.0		5.0		2.0		0.0		2.5		6.0		3.0	
d[8]		-0.5		-2.0		-4.0				-0.5		-1.0		1.0		0.0

Wavelet Decomposition

I[16]	1	2	5	7.0	9	1	5.0	9	2	2	2	3	5	7	4	2	1	1
s[8]	1.5								2.0	2.5	6.0					3.0	1.0	
d[8]																		0.0



Wavelets = spatially localized waves also localized in wavenumber

Multiresolution Analysis

- Recursive wavelet decomposition:

$$\mathbf{I}[16] \rightarrow \mathbf{s}[8], \mathbf{d}[8]$$

$$\mathbf{s}[8] \rightarrow \mathbf{ss}[4], \mathbf{sd}[4]$$

$$\mathbf{ss}[4] \rightarrow \mathbf{sss}[2], \mathbf{ssd}[2]$$

$$\mathbf{sss}[2] \rightarrow \mathbf{ssss}[1], \mathbf{sssd}[1]$$

- Recursive vector-subspace decomposition:

$$\mathbf{V}^0 = \mathbf{V}^1 + \mathbf{W}^1; \mathbf{s}[8] \in \mathbf{V}^1; \mathbf{d}[8] \in \mathbf{W}^1$$

$$\mathbf{V}^1 = \mathbf{V}^2 + \mathbf{W}^2; \mathbf{ss}[4] \in \mathbf{V}^2; \mathbf{sd}[4] \in \mathbf{W}^2$$

...

- Multiresolution representation

$$\mathbf{I}[16] \rightarrow \mathbf{d}[8], \mathbf{sd}[4], \mathbf{ssd}[2], \mathbf{sssd}[1], \mathbf{ssss}[1]$$

or

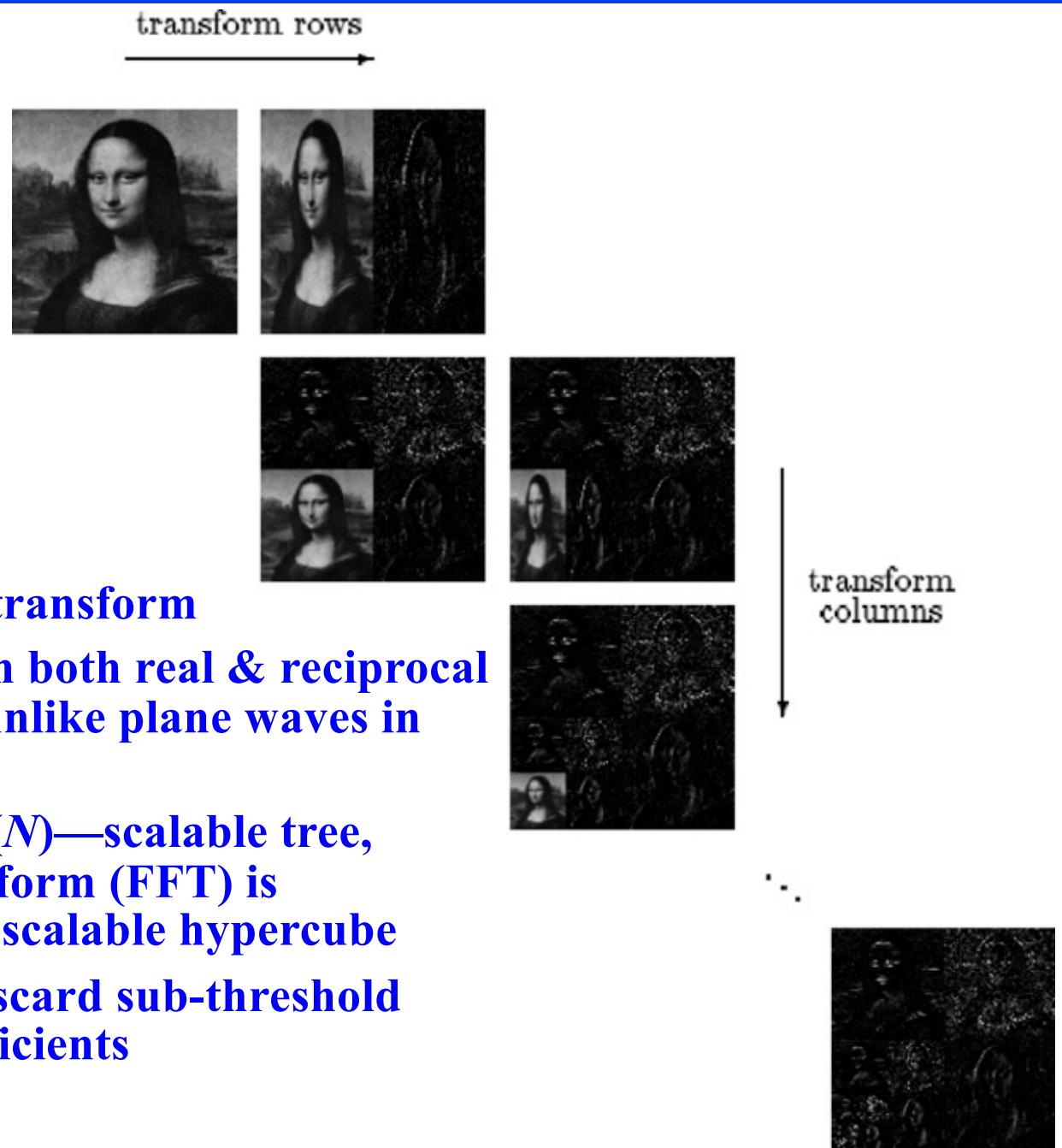
$$\begin{aligned}\mathbf{V}^0 &= \mathbf{V}^1 + \mathbf{W}^1 \\ &= \mathbf{V}^2 + \mathbf{W}^2 + \mathbf{W}^1 \\ &= \mathbf{V}^3 + \mathbf{W}^3 + \mathbf{W}^2 + \mathbf{W}^1 \\ &= \mathbf{V}^4 + \mathbf{W}^4 + \mathbf{W}^3 + \mathbf{W}^2 + \mathbf{W}^1\end{aligned}$$

Very smooth

Progressively more details

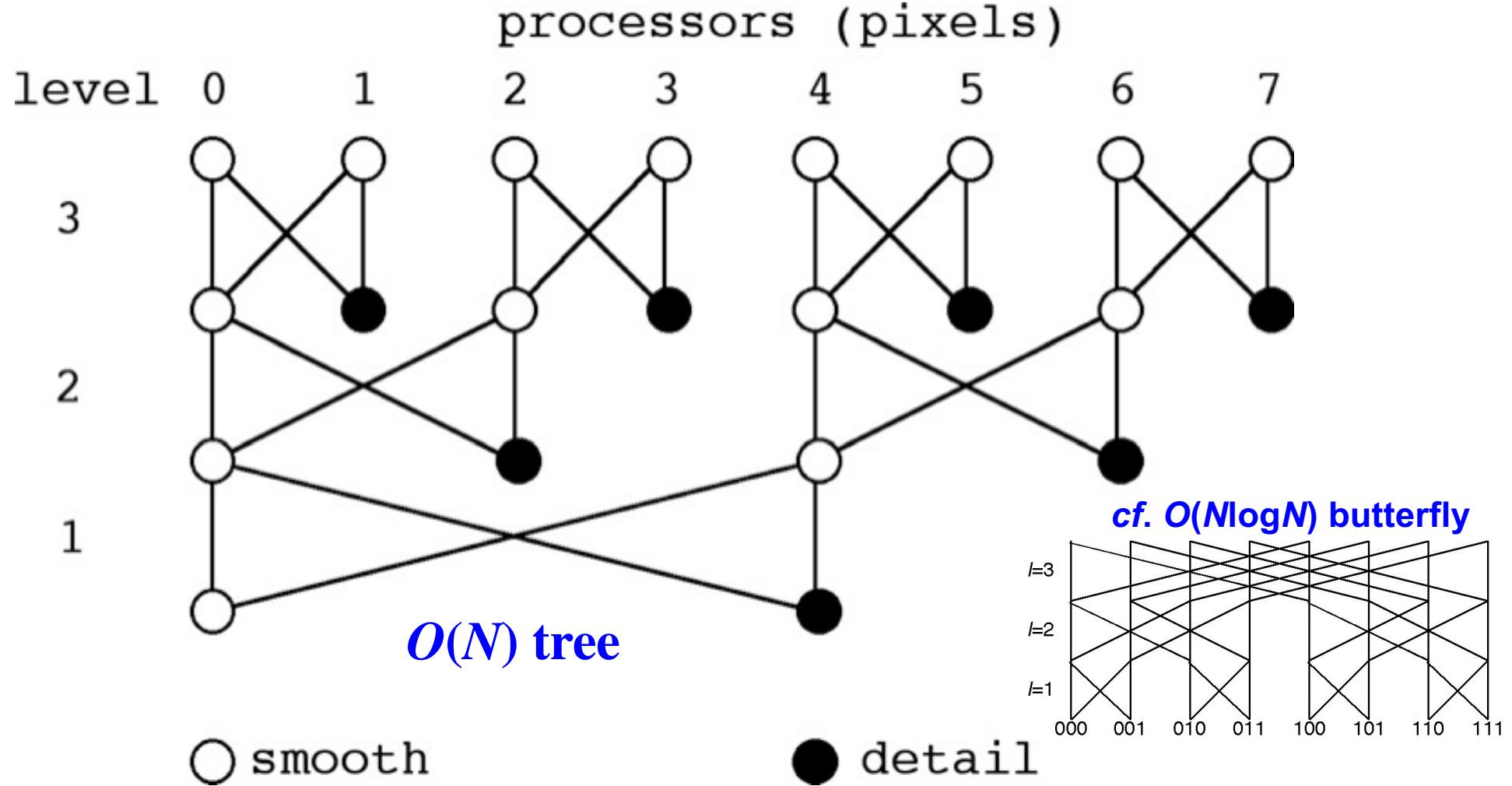
Wavelet Image Decomposition

- Alternate row & column transformations



Parallel Multiresolution Analysis

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



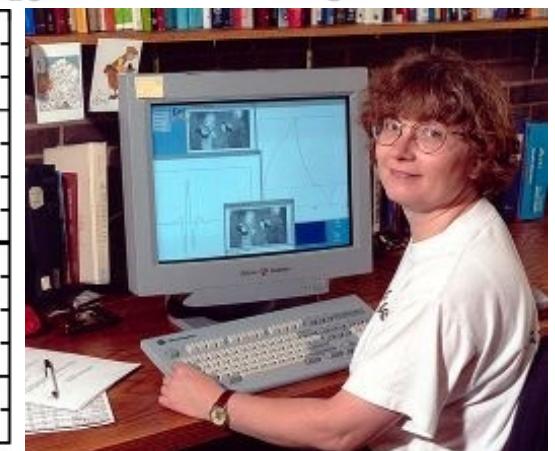
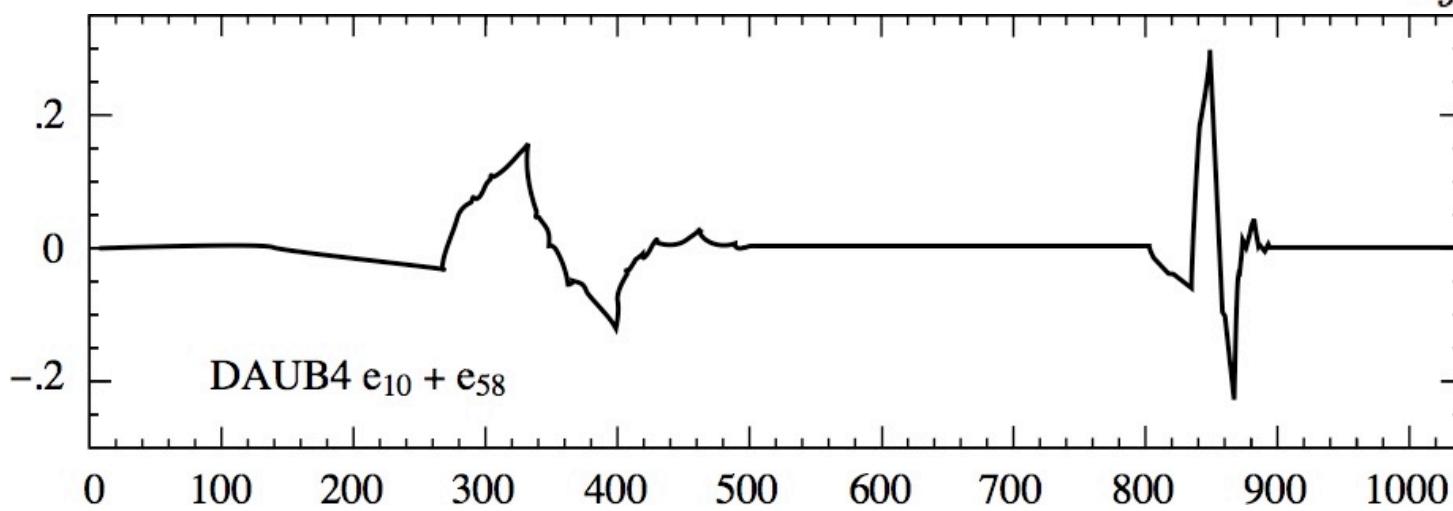
- Retain detailed components? Let subtree slaves do (cf. hypercube quicksort)

Daubechies Wavelets

- Moving window for smooth & detailed filters

$$\begin{array}{c}
 \left[\begin{array}{cccc} c_0 & c_1 & c_2 & c_3 \\ c_3 & -c_2 & c_1 & -c_0 \\ & c_0 & c_1 & c_2 & c_3 \\ & c_3 & -c_2 & c_1 & -c_0 \\ \vdots & \vdots & & & \ddots \\ & & & & \begin{array}{cccc} c_0 & c_1 & c_2 & c_3 \\ c_3 & -c_2 & c_1 & -c_0 \\ & c_0 & c_1 & \\ & c_3 & -c_2 & \end{array} \end{array} \right] \longrightarrow \left[\begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \end{array} \right] \quad \left[\begin{array}{c} s_1 \\ d_1 \\ s_2 \\ d_2 \\ s_3 \\ d_3 \\ s_4 \\ d_4 \\ s_5 \\ d_5 \\ s_6 \\ d_6 \\ s_7 \\ d_7 \\ s_8 \\ d_8 \end{array} \right]
 \end{array}$$

$c_0 = (1+\sqrt{3})/4\sqrt{2} = 0.4829629131445341$
 $c_1 = (3+\sqrt{3})/4\sqrt{2} = 0.8365163037378079$
 $c_2 = (3-\sqrt{3})/4\sqrt{2} = 0.2241438680420134$
 $c_3 = (1-\sqrt{3})/4\sqrt{2} = -0.1294095225512604$



Ingrid Daubechies

Iterative Solution of Linear Systems

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$$
$$\begin{bmatrix} X & & \\ X & X & \\ X & X & X \end{bmatrix} + \begin{bmatrix} X & & \\ & X & \\ & & X \end{bmatrix} + \begin{bmatrix} & X & X & X \\ & X & X & \\ & & X \end{bmatrix}$$

- Fixed-point equation

$$\mathbf{x} = \mathbf{D}^{-1}[-(\mathbf{L}+\mathbf{U})\mathbf{x} + \mathbf{b}]$$

$$(\mathbf{D} + \mathbf{L} + \mathbf{U})\mathbf{x} = \mathbf{b}$$
$$\mathbf{D}\mathbf{x} = -(\mathbf{L} + \mathbf{U})\mathbf{x} + \mathbf{b}$$

- Jacobi iteration

$$\mathbf{x}^{(n+1)} = \mathbf{D}^{-1}[-(\mathbf{L}+\mathbf{U})\mathbf{x}^{(n)} + \mathbf{b}]$$

$$x_i^{(n+1)} = \frac{1}{a_{ii}} \left(- \sum_{\substack{j=1 \\ (j \neq i)}}^N a_{ij} x_j^{(n)} + b_i \right)$$

More Iterative Smoothing

$$x_i^{(n+1)} = \frac{1}{a_{ii}} \left(- \sum_{\substack{j=1 \\ (j \neq i)}}^N a_{ij} x_j^{(n)} + b_i \right) = x_i^{(n)} + \frac{1}{a_{ii}} \overbrace{\left(- \sum_{j=1}^N a_{ij} x_j^{(n)} + b_i \right)}^{\text{residual}}$$

over|under relaxation $\rightarrow x_i^{(n)} + \frac{\Delta}{a_{ii}} \left(- \sum_{j=1}^N a_{ij} x_j^{(n)} + b_i \right)$

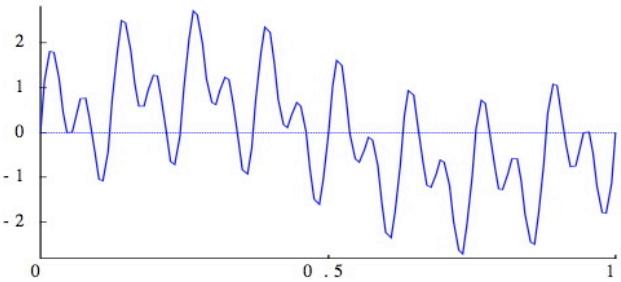
Over ($\Delta > 1$) or under ($\Delta < 1$) relaxation to accelerate convergence

- More general fixed-point iteration

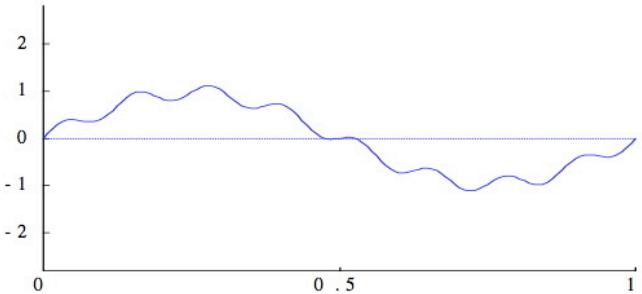
$$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{Z} \overbrace{\left(-\mathbf{A}\mathbf{x} + \mathbf{b} \right)}^{\text{residual}} = (\mathbf{I} - \mathbf{Z}\mathbf{A})\mathbf{x} + \mathbf{Z}\mathbf{b}$$

- High-frequency residual (error) dies out quickly

- Initial error:



- Error after 35 iteration sweeps:



Multigrid Method

- **Residual equation:**

$\mathbf{A}^{(l)}$: l -th level matrix

\mathbf{v} : Current guess

\mathbf{e} : error vector

\mathbf{r} : residual vector

$$\begin{aligned} & \text{exact} \\ & \mathbf{A}^{(l)} \overbrace{(\mathbf{v} + \mathbf{e})} = -4\pi e^2 \mathbf{n} \\ & \quad \rightarrow \mathbf{A}^{(l)} \mathbf{v} = -4\pi e^2 \mathbf{n} + \mathbf{r} \\ & \mathbf{A}^{(l)} \mathbf{e} = -\mathbf{r} \end{aligned}$$

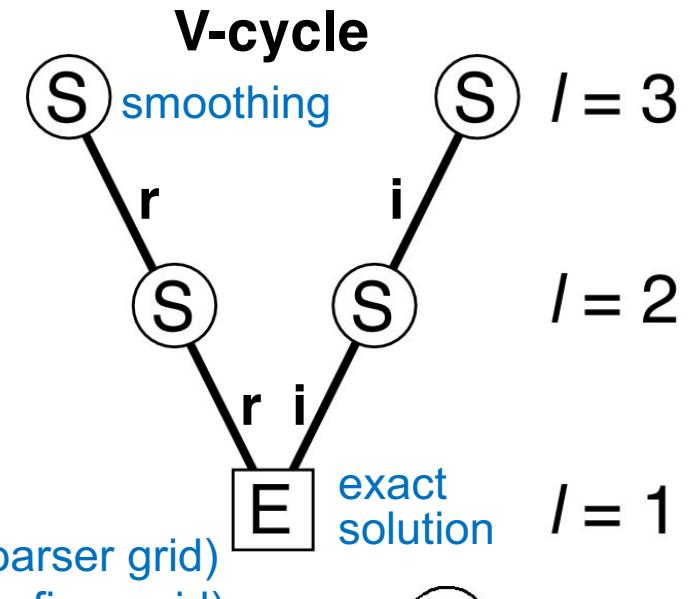
- **Smoothing:**

$$\mathbf{e} \leftarrow [1 + \mathbf{Z}^{(l)} \mathbf{A}^{(l)}] \mathbf{e} + \mathbf{Z}^{(l)} \mathbf{r}$$

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

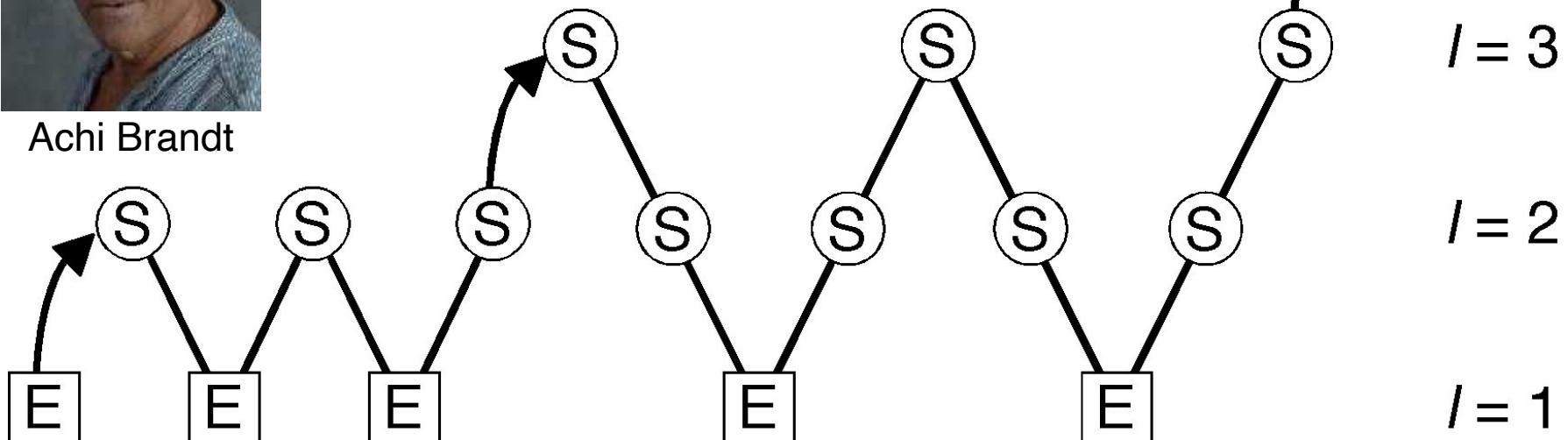


Achi Brandt



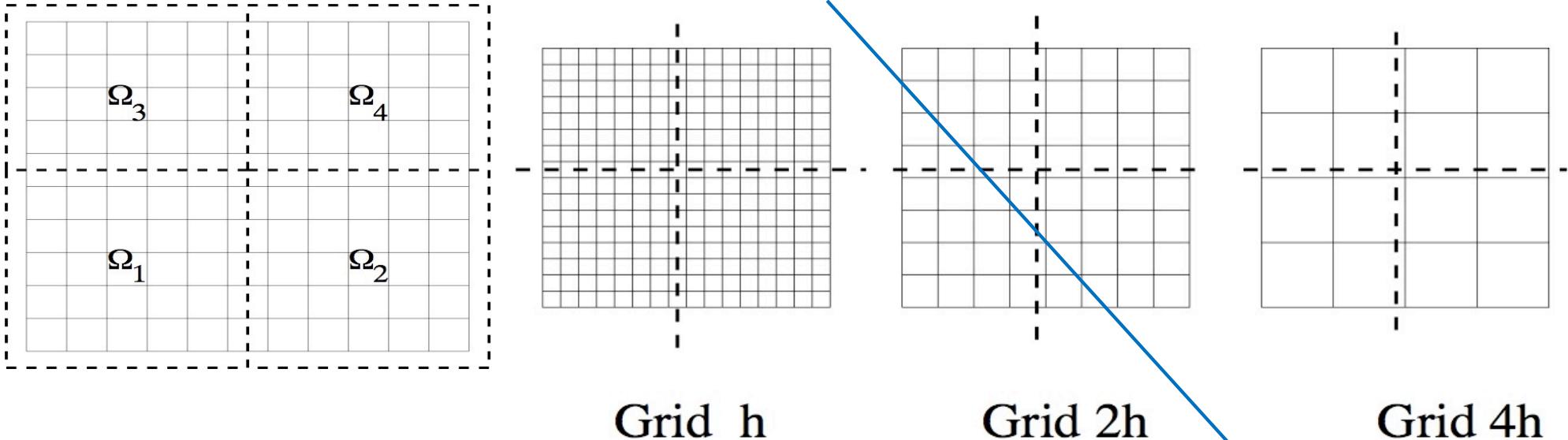
r: restrict (to coarser grid)
i: interpolate (to finer grid)

Full multigrid



Parallel Multigrid Method

- Domain decomposition with boundary-layer caching



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

$N \times N$ grids each on $P \times P$ processors: $T(N^2 P^2, P^2) = a \log NP + bN + cN^2$

Weak-scaling speedup & efficiency

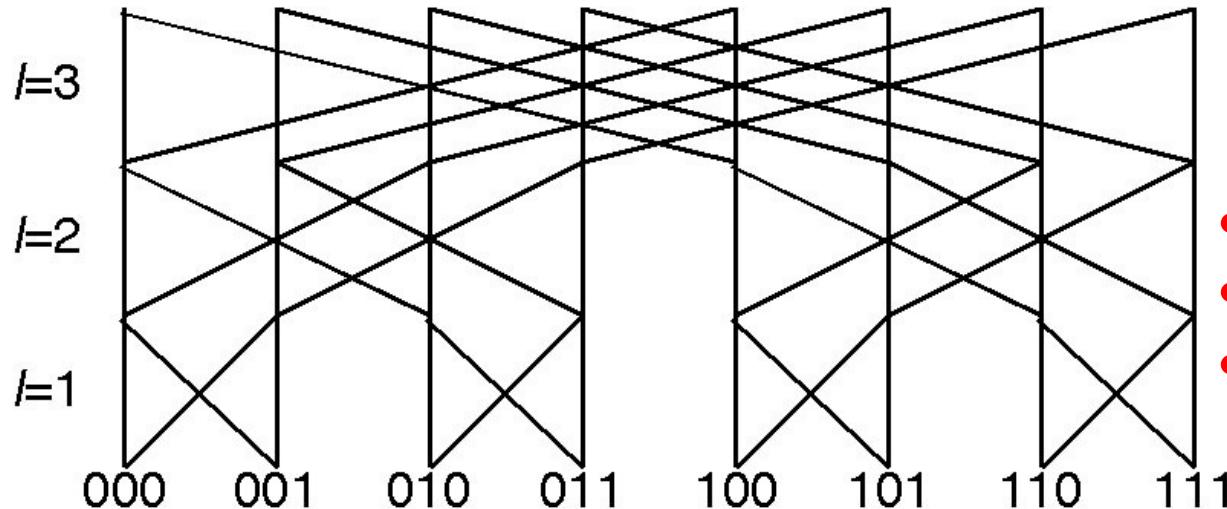
$$S_{P^2} = \frac{N^2 P^2 T(N^2, 1)}{N^2 T(N^2, P^2)} = \frac{P^2 (cN^2)}{a \log NP + bN + cN^2} = \frac{P^2}{1 + \frac{b}{cN} + \frac{a}{cN} \log NP}$$

$$E_{P^2} = \frac{S_{P^2}}{P^2} = \frac{1}{1 + \frac{b}{cN} + \frac{a}{cN} \log NP}$$

local V-cycle level $\underbrace{\log N}_{\log N}$ + idle processor $\underbrace{\log P}_{\log P}$

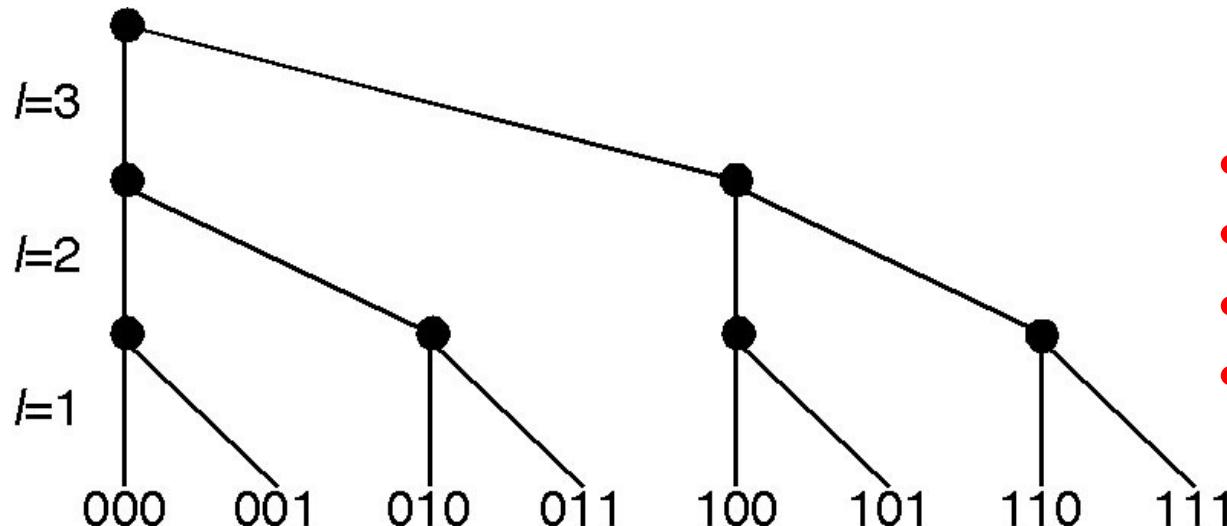
Global Communications

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



- All-to-all reduction
- Quicksort
- Fast Fourier transform

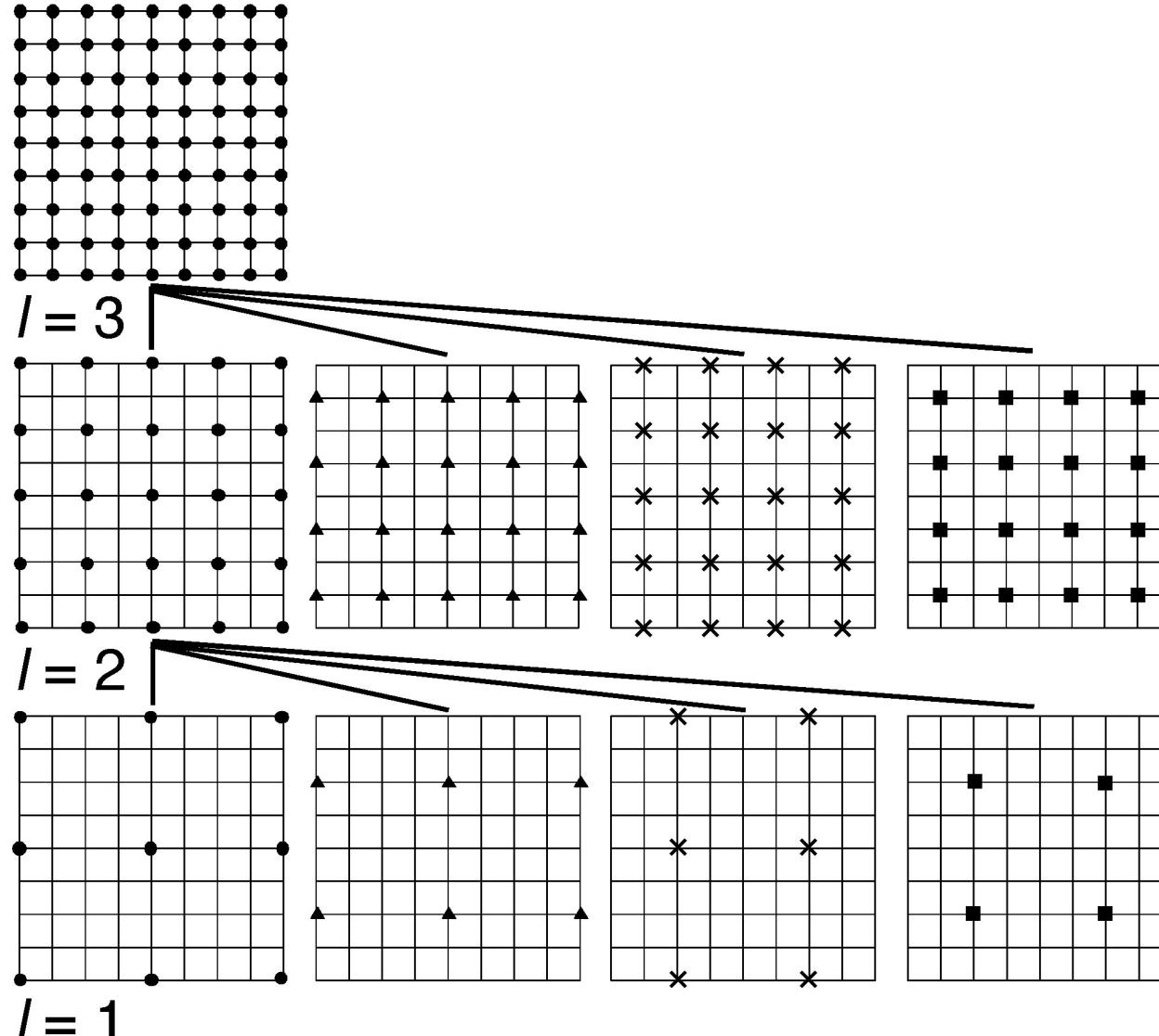
All-to-one (tournament): $O(N)$



- Global reduction
- Fast multipole method
- Multigrid method
- Wavelets

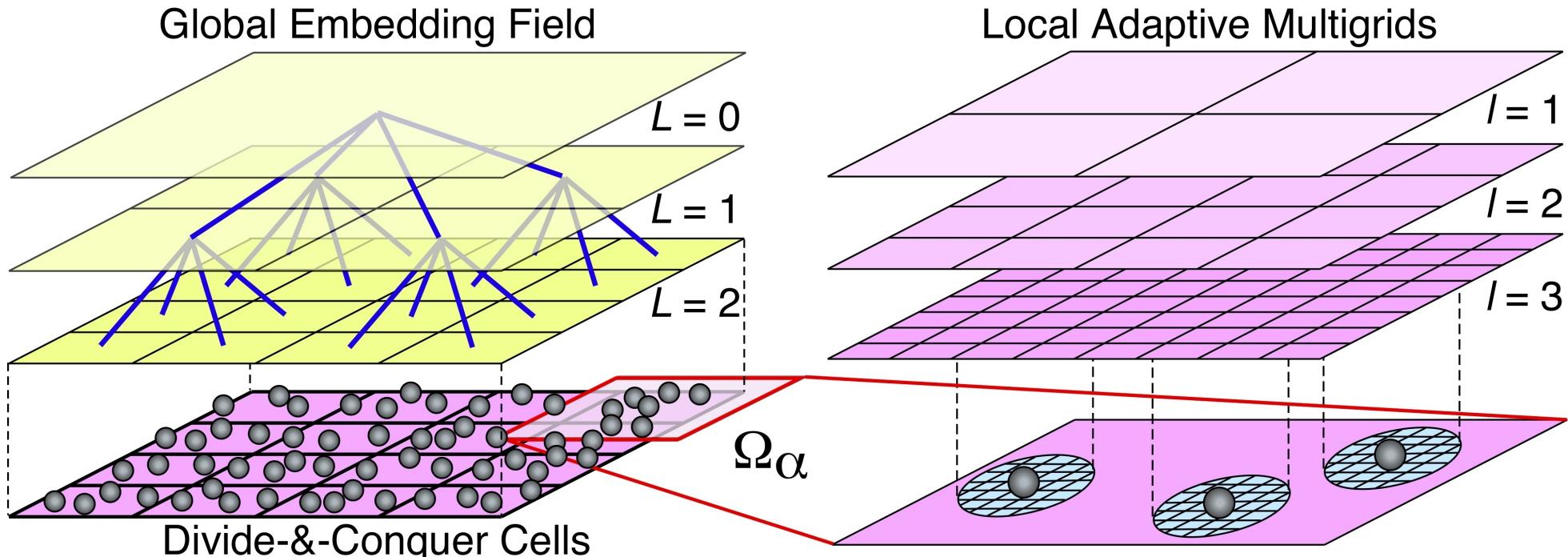
Solving the Idle Processor Problem

- Parallel superconvergent multigrid: Solve multiple coarse problems to accelerate the convergence Frederickson & McBryan, '88



Nakano et al., *Comput. Phys. Commun.* 83, 181 ('94)

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

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

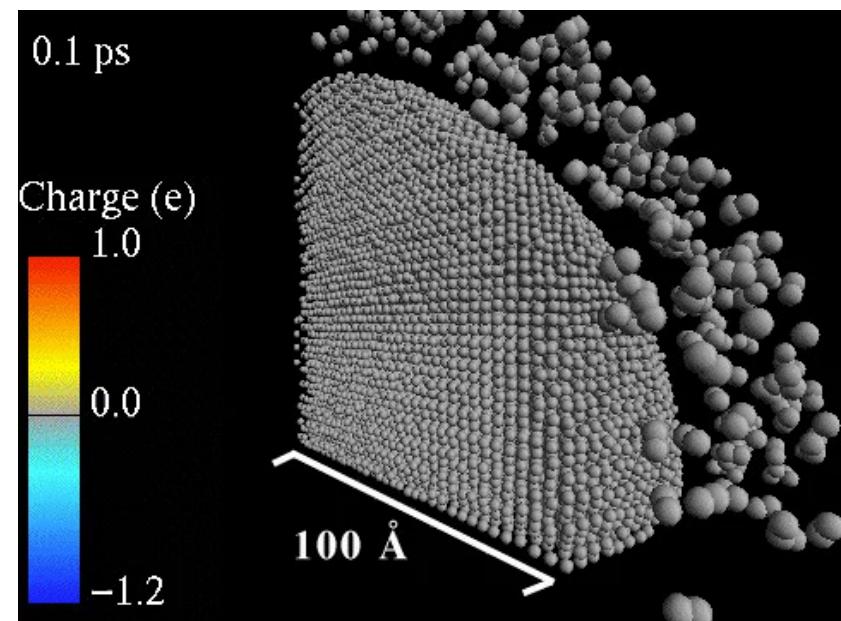
- Reliable interatomic potential

$$E_{\text{MD}} = \sum_{i < j} u_{ij}(r_{ij}) + \sum_{i, j < k} v_{jik}(\mathbf{r}_{ij}, \mathbf{r}_{ik})$$

- N -body problem

Long-range electrostatic interaction — $O(N^2)$

Evaluate $V_{\text{es}}(\mathbf{x}) = \sum_{j=1}^N \frac{q_j}{|\mathbf{x}-\mathbf{x}_j|}$ at $\mathbf{x} = \mathbf{x}_i$ ($i = 1, \dots, N$)



- $O(N)$ space-time multiresolution MD (MRMD) algorithm

1. Fast multipole method (FMM) Greengard & Rokhlin, '87

2. Symplectic multiple time stepping (MTS) Tuckerman *et al.*, '92

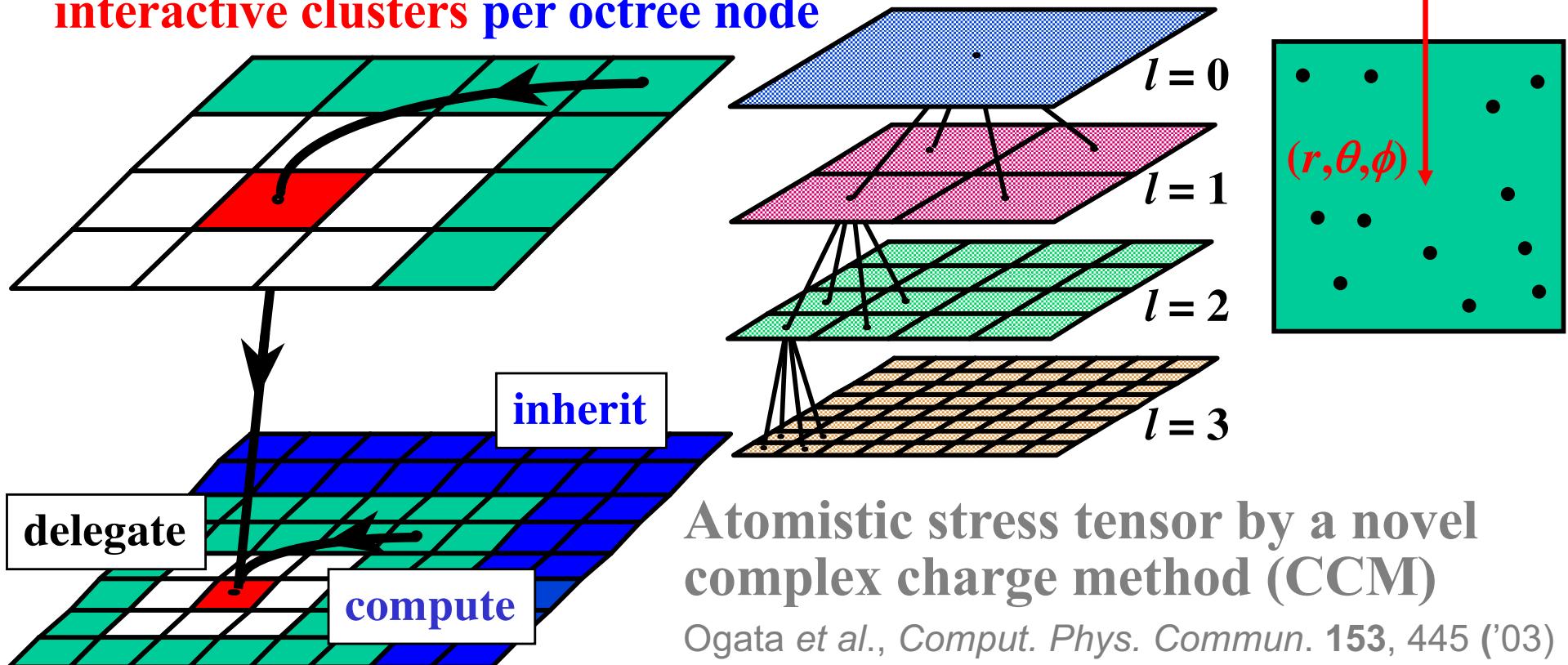
Spatial Locality: Fast Multipole Method

1. Clustering: Encapsulate far-field information using multipoles

$$V(\mathbf{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left\{ \sum_{i=1}^N q_i r_i^l Y_l^{*m}(\theta_i, \phi_i) \right\} \frac{Y_l^m(\theta, \phi)}{r^{l+1}}$$

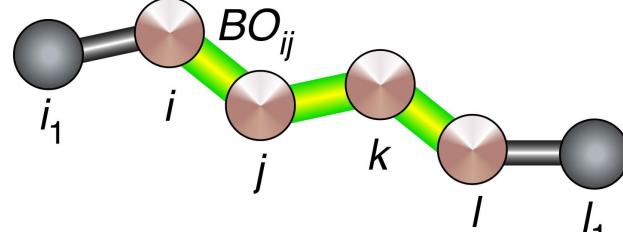
2. Hierarchical abstraction: Octree data structure

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

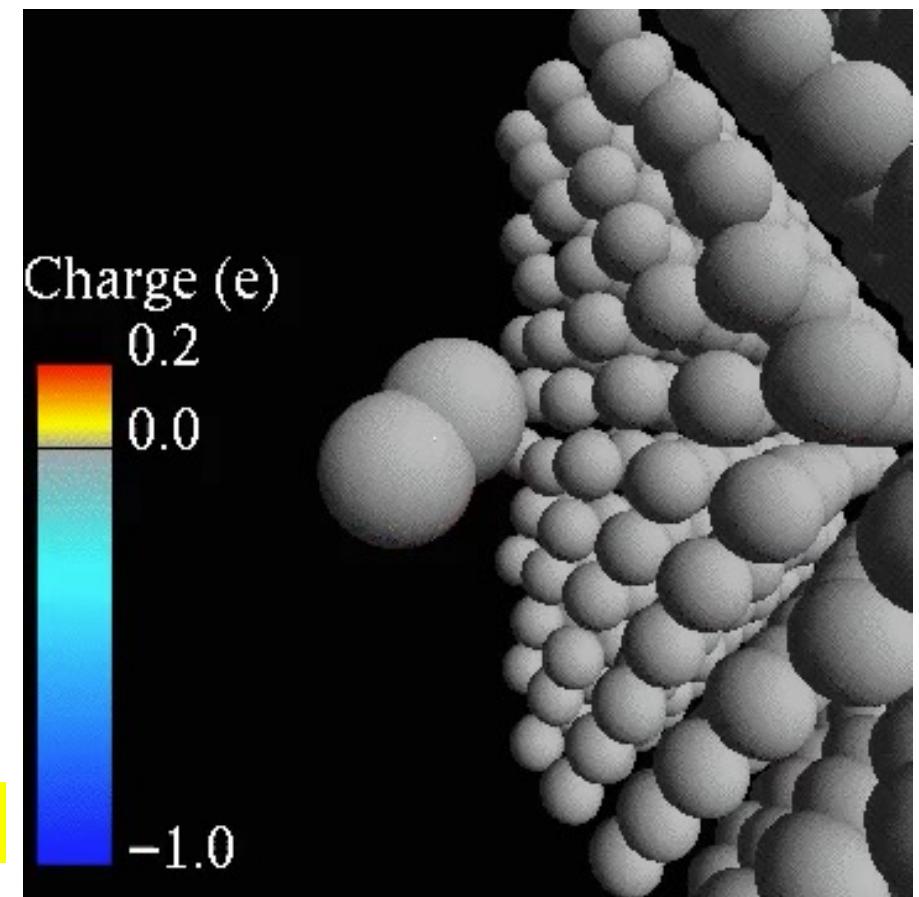


Reactive Force-Field (ReaxFF) MD: Variable N-Charge Problem

- Reactive bond order potential energy: $E_{\text{bond}}(\{r_{ij}\}, \{r_{ijk}\}, \{r_{ijkl}\}, \{BO_{ij}\})$
→ Bond breakage & formation



- Charge-equilibration (QEeq)
→ Charge transfer
Determine atomic charges $\{q_i | i = 1, \dots, N\}$ every MD step
to minimize $E_{\text{ES}}(\mathbf{r}^N, q^N)$ with
charge-neutrality constraint:
 $\sum_i q_i = 0$
— Dense linear system: $M q = -\chi$
 $O(N^3)!$



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

Fast Reactive Force-Field Algorithm

- $O(N)$ fast reactive force-field (F-ReaxFF) algorithm

1) Fast multipole method

2) Temporal locality, $q_i^{(\text{init})}(t+\Delta t) = q_i(t)$

$$M q = -\chi$$

- Multilevel preconditioned conjugate gradient (MPCG) method

1) Split Coulomb matrix: $M = M_{\text{near}} + M_{\text{far}}$

2) Sparse near-field preconditioner: $M_{\text{near}}^{-1} M q = -M_{\text{near}}^{-1} \chi$

$$V_{\text{es}}(\mathbf{x}_i) = \sum_{j=1}^N \frac{q_j}{|\mathbf{x}_i - \mathbf{x}_j|}$$

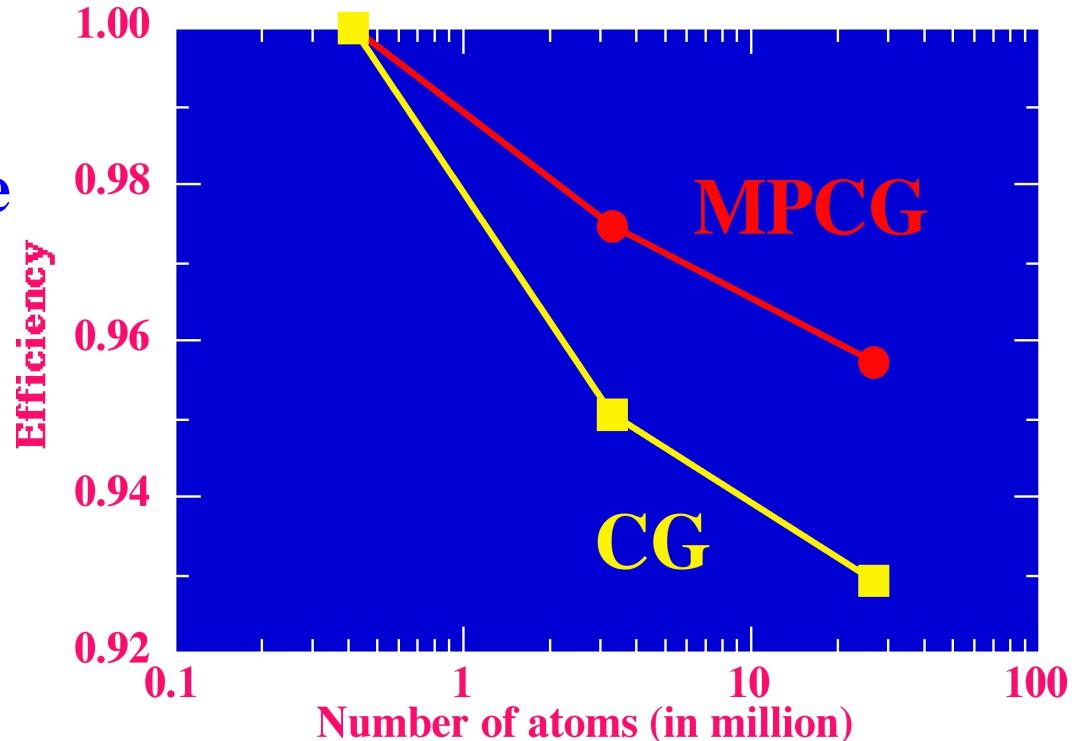
Results:

- 20% speed up of convergence

• Enhanced data locality:

Improved parallel efficiency

0.93 → 0.96 for 26.5M-atom
 Al_2O_3 on 64 Power nodes



A. Nakano, *Comput. Phys. Commun.*
104, 59 ('97)

Extended-Lagrangian RMD (XRMD)

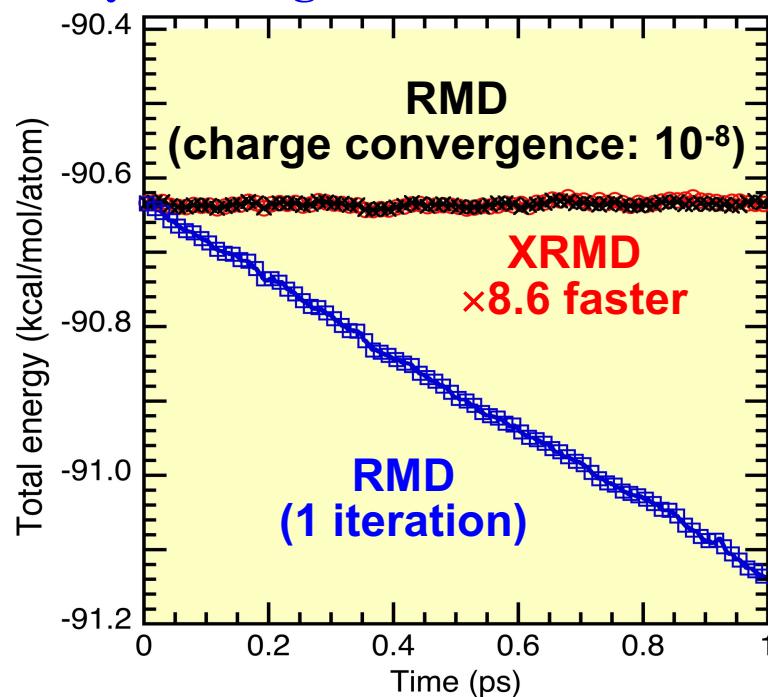
- Eliminated speed-limiting iteration for charge-equilibration in reactive molecular dynamics (RMD) by adapting an extended-Lagrangian scheme proposed for QMD

Souvatzis & Niklasson, *J. Chem. Phys.* **140**, 044117 ('14)

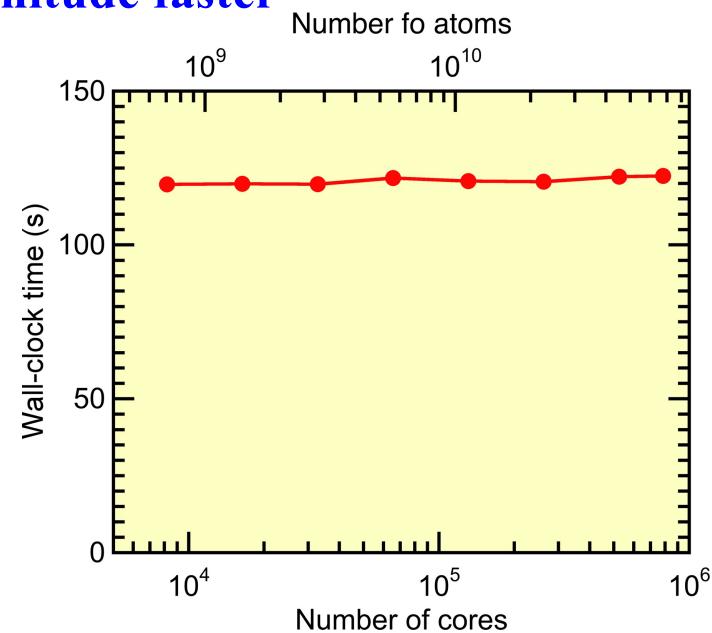
$$L_{\text{XRMD}} = L_{\text{RMD}} + \frac{\mu}{2} \sum_i \dot{\theta}_i^2 - \frac{\mu\omega^2}{2} \sum_i (\theta_i - q_i)^2$$

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

cf. Shadow dynamics Niklasson & Negre, *J. Chem. Phys.* **158**, 154105 ('23)

Rather than poorly solving the exact dynamics, exactly solve a constrained proxy (*cf.* backward error analysis)

Quantum Molecular Dynamics (QMD)

$$M_I \frac{d^2}{dt^2} \mathbf{R}_I = -\frac{\partial}{\partial \mathbf{R}_I} E[\{\mathbf{R}_I\}, \psi(\mathbf{r}_1 \dots, \mathbf{r}_N)] \quad (I = 1, \dots, N_{\text{atom}})$$

First molecular dynamics using an empirical interatomic interaction

A. Rahman, *Phys. Rev.* **136**, A405 ('64)



Atoms

$$\psi(\mathbf{r}_1 \dots, \mathbf{r}_N) \leftarrow \operatorname{argmin} E[\{\mathbf{R}_I\}, \psi(\mathbf{r}_1 \dots, \mathbf{r}_N)]$$

Electrons

Complexity reduction

Density functional theory (DFT)

Hohenberg & Kohn, *Phys. Rev.* **136**, B864 ('64)

W. Kohn, *Nobel chemistry prize*, '98

$$O(C^N) \rightarrow O(N^3)$$

1 N -electron problem N 1-electron problems
intractable tractable

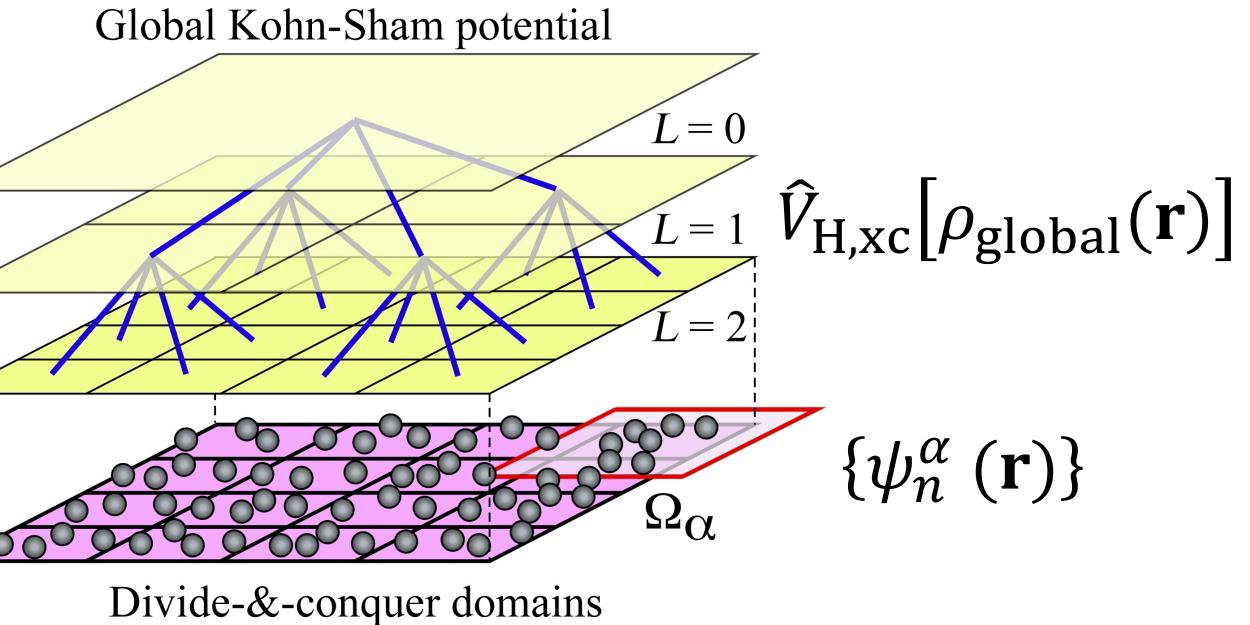
$$\psi(\mathbf{r}_1 \dots, \mathbf{r}_N) \quad \{\psi_i(\mathbf{r}) | i = 1, \dots, N\}$$

$O(N)$ 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)

Physical data locality!

Divide-&-Conquer Density Functional Theory



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

Global-local
self-consistent
field (SCF)
iteration

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

- Global & domain electron densities

$$\rho_{\text{global}}(\mathbf{r}) = \sum_\alpha p_\alpha(\mathbf{r}) \rho_\alpha(\mathbf{r})$$

Domain support function

$$\sum_\alpha p_\alpha(\mathbf{r}) = 1$$

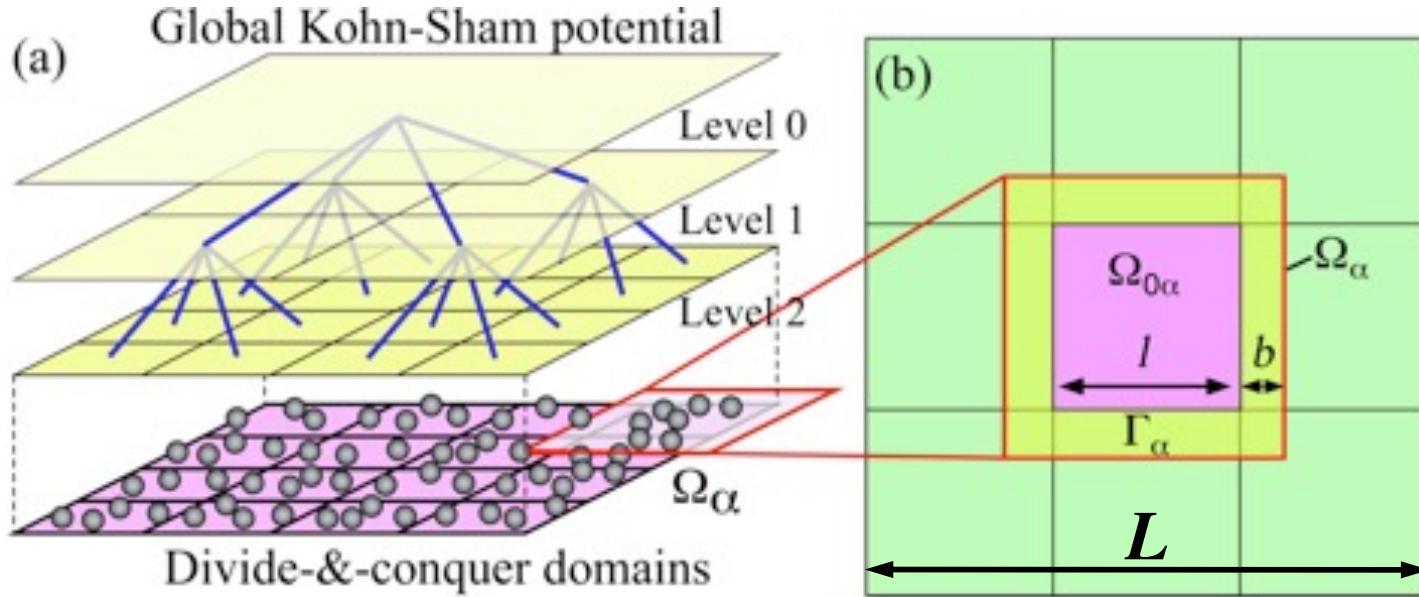
$$\rho_\alpha(\mathbf{r}) = \sum_n |\psi_n^\alpha|^2 \Theta(\mu - \epsilon_n^\alpha)$$

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

$$l_* = \operatorname{argmin}(T_{\text{comp}}(l)) = \operatorname{argmin} \left(\left(\frac{l}{L}\right)^3 (l + 2b)^{3\nu} \right) = \frac{2b}{\nu - 1}$$

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

- Error analysis: Buffer thickness b is dictated by the accuracy requirement

$$b = \lambda \ln (\max \{ |\Delta \rho_\alpha(\mathbf{r})| \mid \mathbf{r} \in \partial \Omega_\alpha \}) / \varepsilon \langle \rho_\alpha(\mathbf{r}) \rangle$$

Decay length

$\rho_\alpha(\mathbf{r}) - \rho_{\text{global}}(\mathbf{r})$

Error tolerance

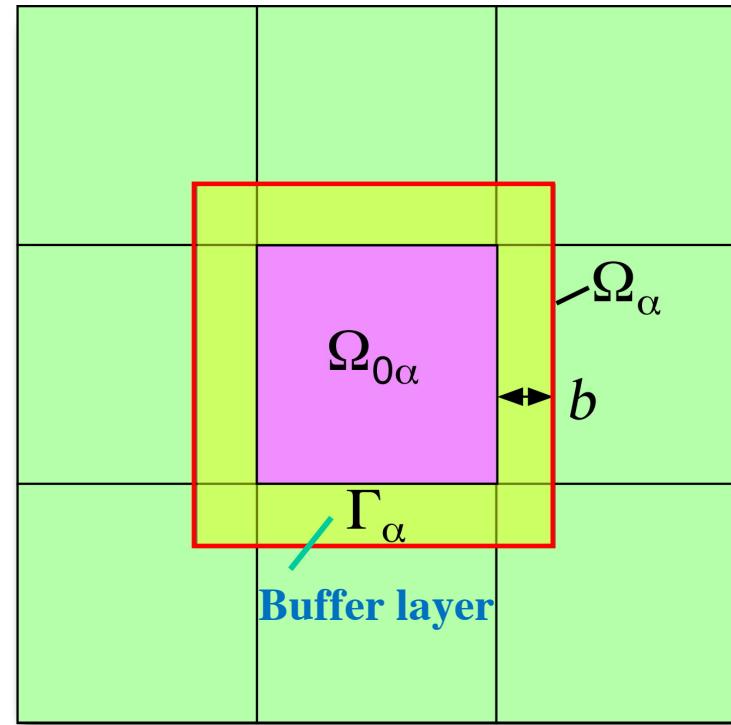
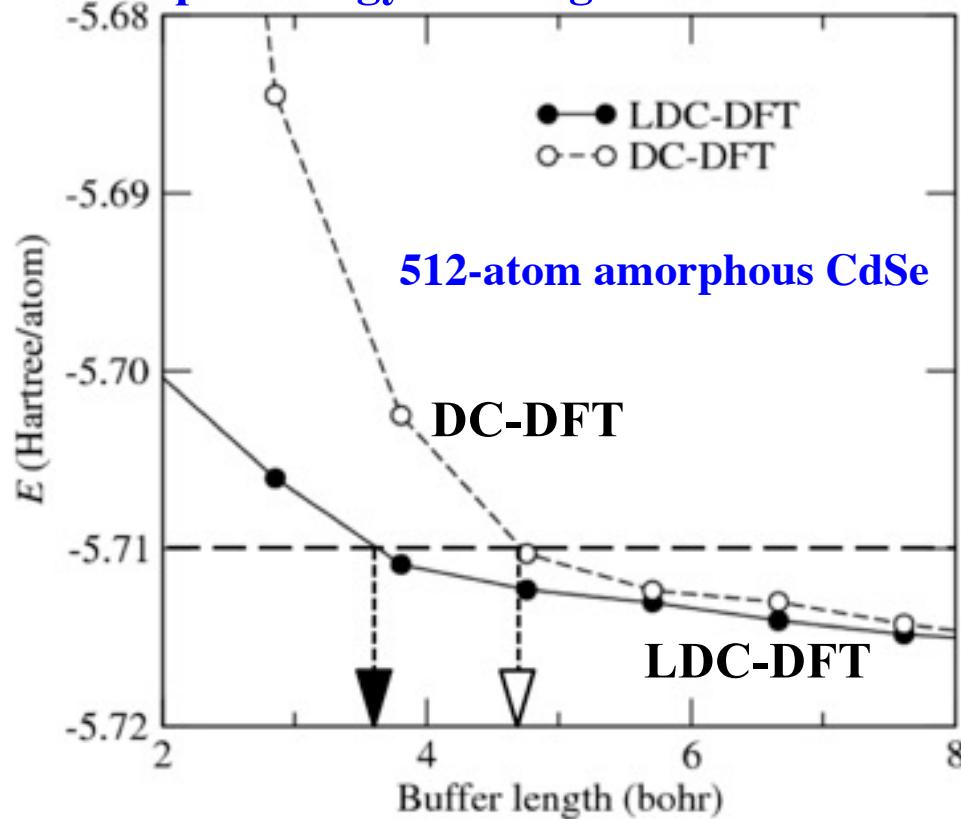
cf. quantum nearsightedness Kohn, Phys. Rev. Lett. 76, 3168 ('96)

Lean Divide-&-Conquer (LDC) DFT

- Density-adaptive boundary potential to reduce the $O(N)$ prefactor

$$\nu_{\alpha}^{\text{bc}}(\mathbf{r}) = \int d\mathbf{r}' \frac{\partial \nu(\mathbf{r})}{\partial \rho(\mathbf{r}')}\left(\rho_{\alpha}(\mathbf{r}) - \rho_{\text{global}}(\mathbf{r})\right) \cong \frac{\rho_{\alpha}(\mathbf{r}) - \rho_{\text{global}}(\mathbf{r})}{\xi}$$

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

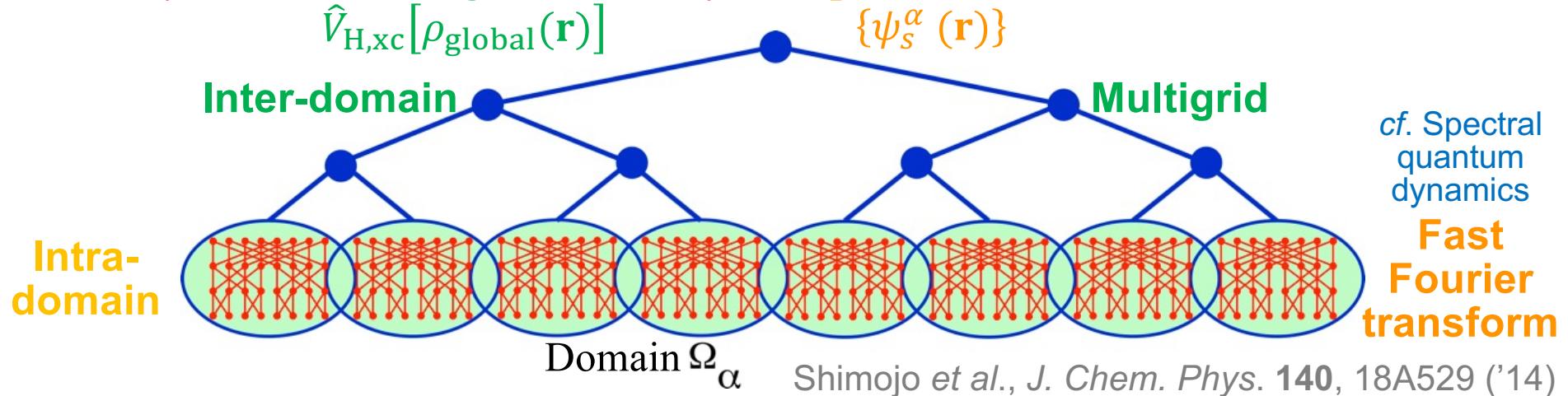


- Factor 2.03 (for $\nu = 2$) ~ 2.89 (for $\nu = 3$) reduction of the computational cost with an error tolerance of 5×10^{-3} a.u. (per-domain complexity: n^{ν})

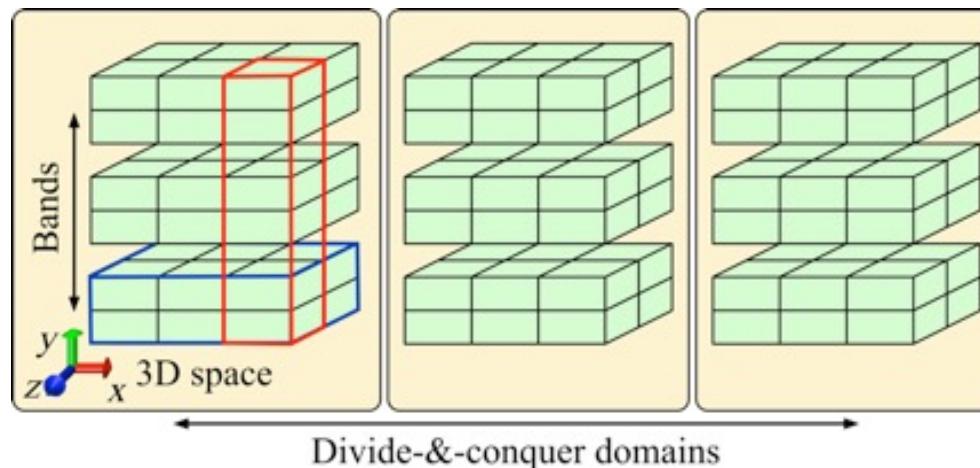
F. Shimojo *et al.*, *J. Chem. Phys.* **140**, 18A529 ('14);
Phys. Rev. B **77**, 085103 ('08); *Comput. Phys. Commun.* **167**, 151 ('05)

Hierarchical Computing

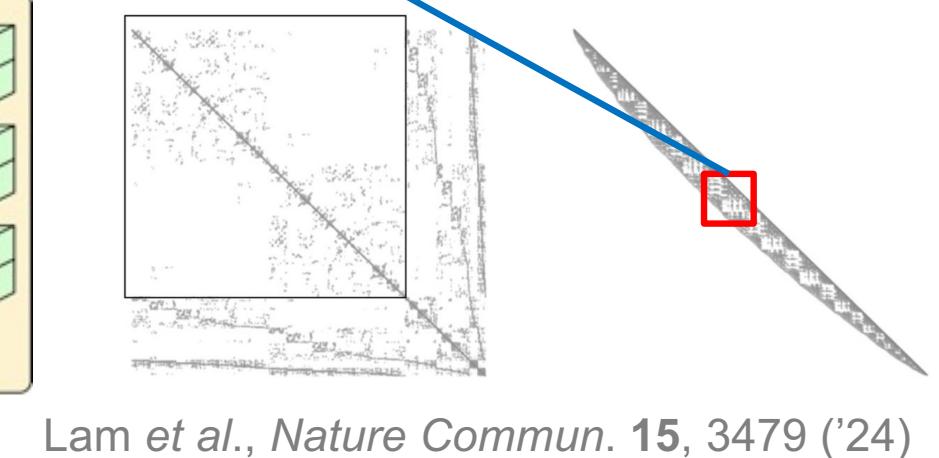
- Globally-scalable (multigrid) & locally-fast (plane wave) (GSLF) electronic solver



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



cf. Globally-sparse yet locally-dense (GSLD) eigen-solver
Physical data locality (quantum nearsightedness)
Kohn, PRL 76, 3168 ('96)



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

D. Vanderbilt, *Phys. Rev. B* 41, 7892 ('90)

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



$$\Psi = [|\psi_1^\alpha\rangle, \dots, |\psi_{N_{\text{band}}}^\alpha\rangle] \quad \widetilde{\mathbf{B}}(i) = [|\beta_{i,1}\rangle, \dots, |\beta_{i,N_{\text{atom}}}\rangle] \quad [\widetilde{\mathbf{D}}(i,j)]_{I,J} = D_{ij,I} \delta_{IJ}$$

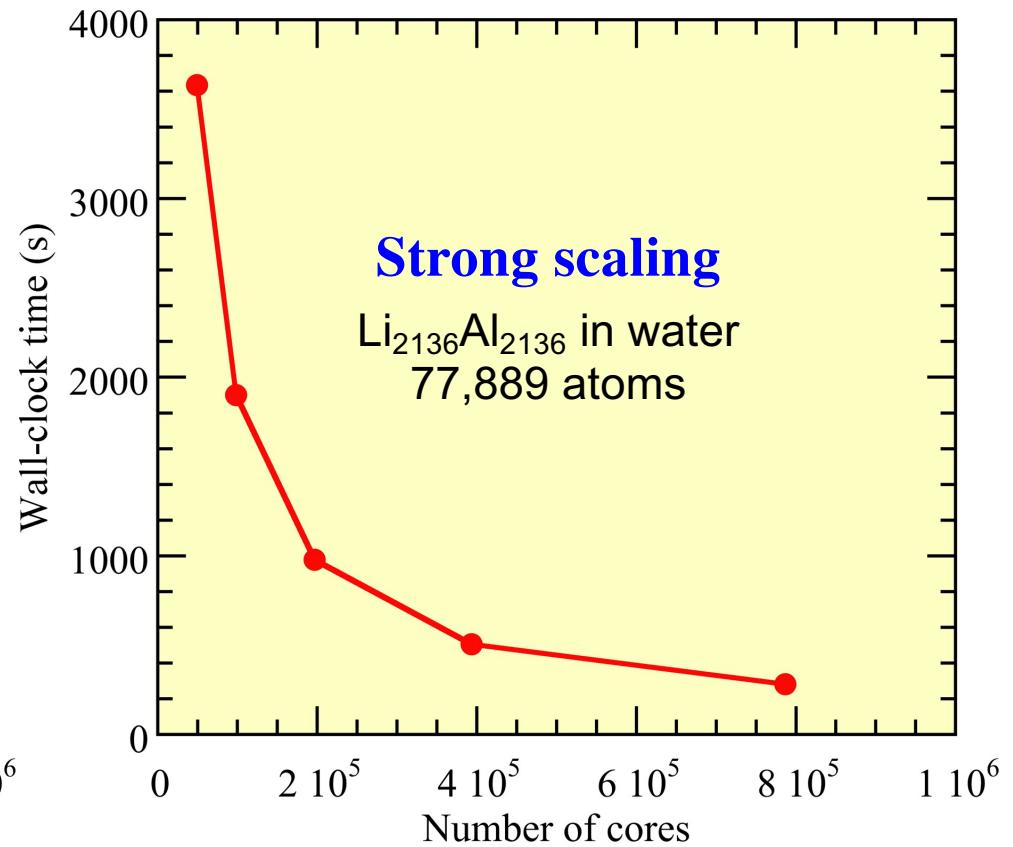
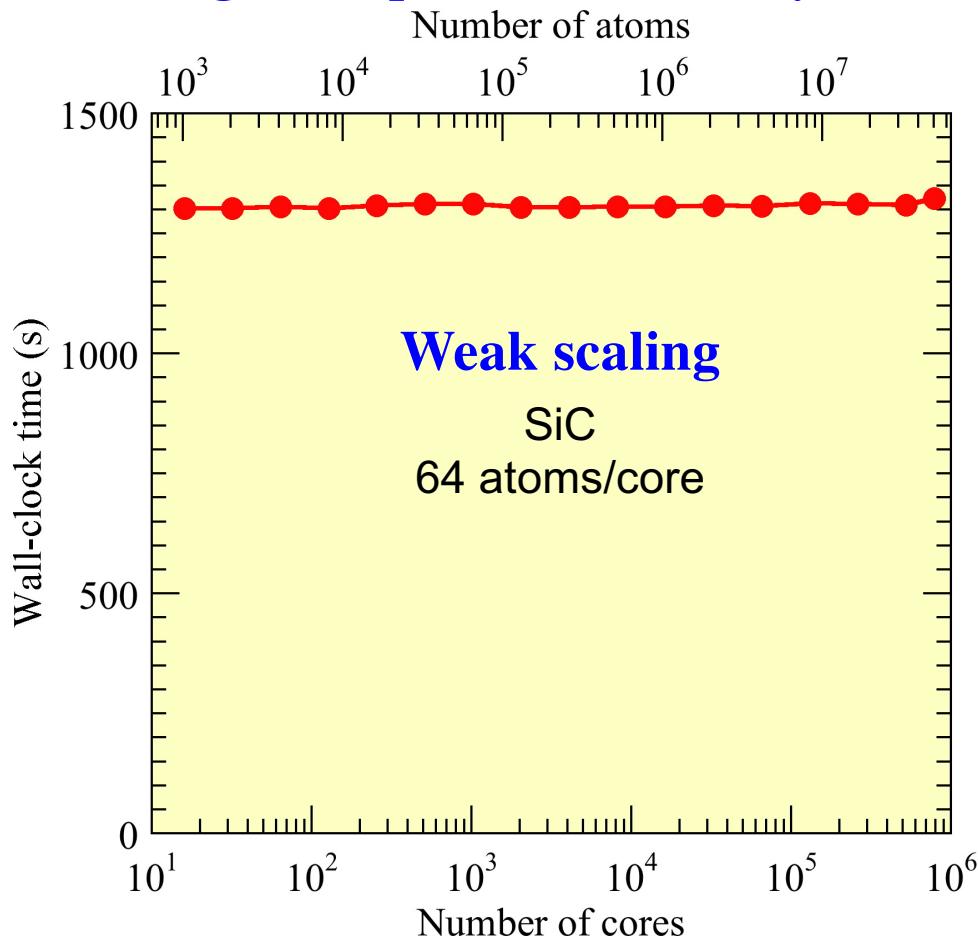
$$\hat{v}_{\text{nl}}\Psi = \sum_{i,j}^L \widetilde{\mathbf{B}}(i) \widetilde{\mathbf{D}}(i,j) \widetilde{\mathbf{B}}(j)^T$$

- 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)

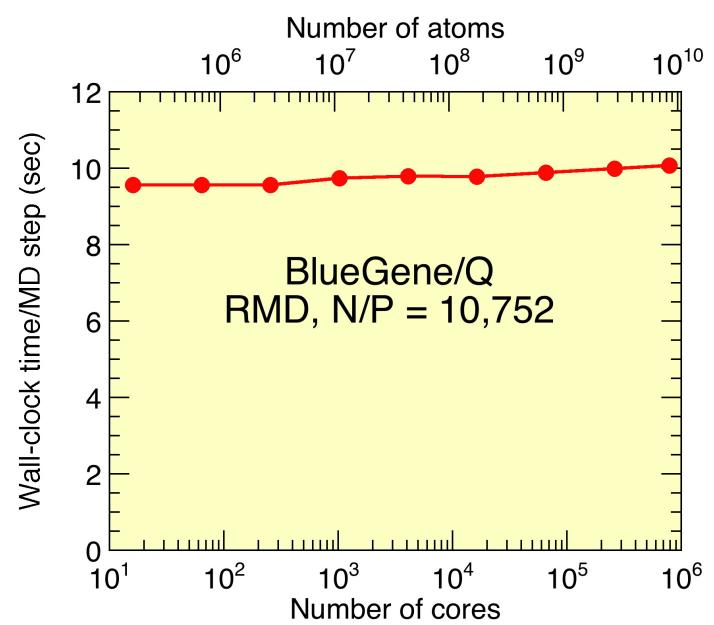
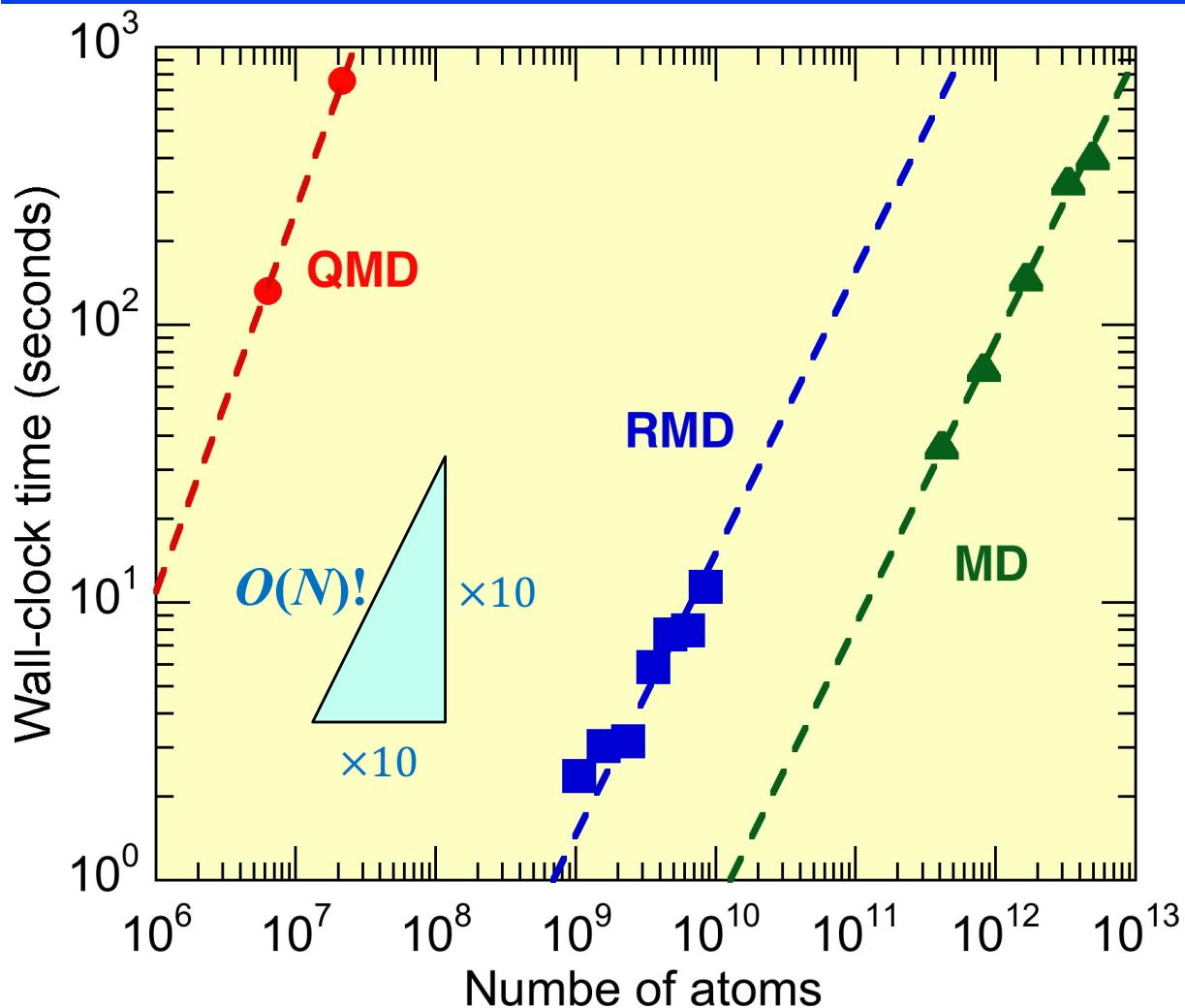
Parallel Performance

- 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]

Scalable Simulation Algorithm Suite



QMD (quantum molecular dynamics): DC-DFT

RMD (reactive molecular dynamics): F-ReaxFF

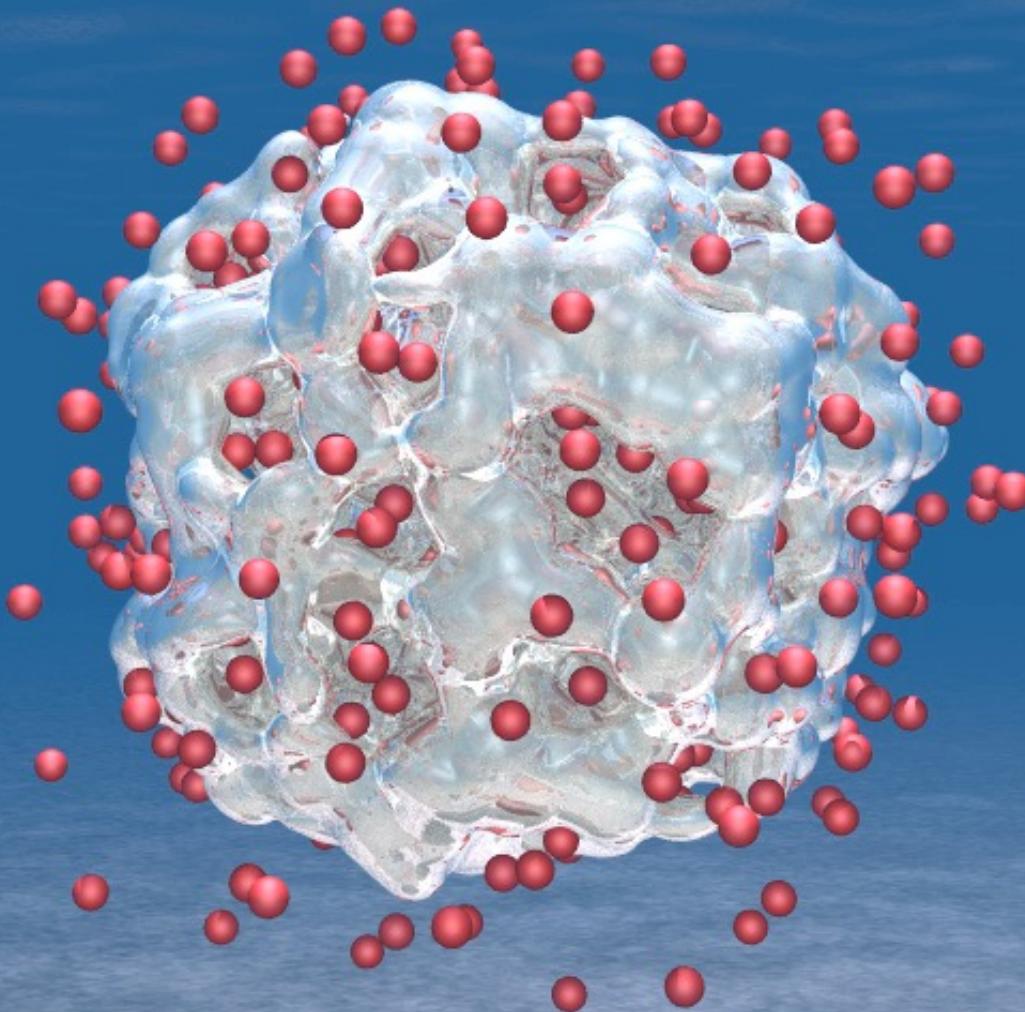
MD (molecular dynamics): MRMD

- 4.9 trillion-atom space-time multiresolution MD (MRMD) of SiO_2
 - 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

H₂ Production from Water Using LiAl Particles

16,661-atom QMD simulation of Li₄₄₁Al₄₄₁ in water
on 786,432 IBM Blue Gene/Q cores

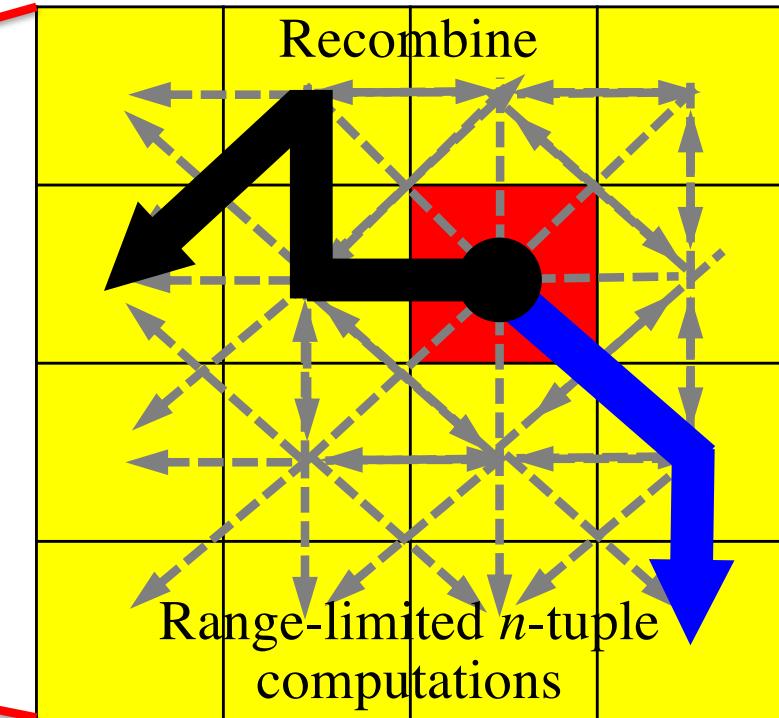
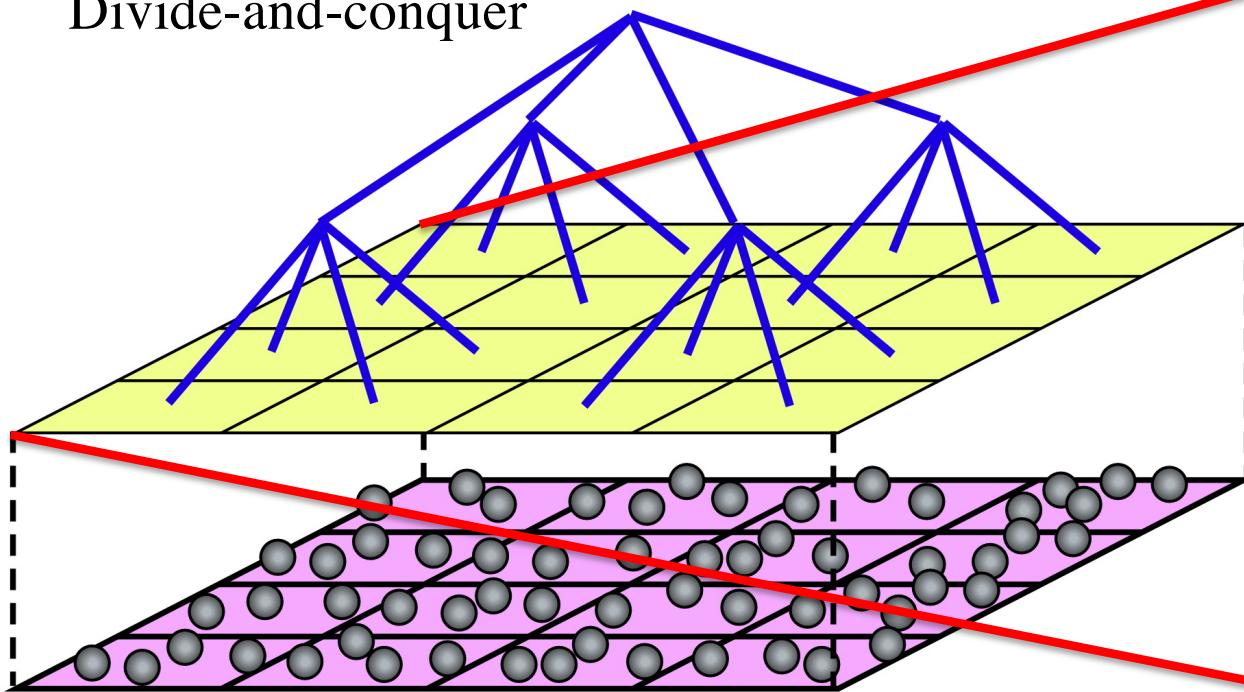
K. Shimamura *et al.*,
Nano Lett. **14**, 4090 ('14)



21,140 time steps (129,208 self-consistent-field iterations)

Divide-Conquer-Recombine Algorithms

Divide-and-conquer



M. Kunaseth et al., ACM/IEEE SC13 ('13)

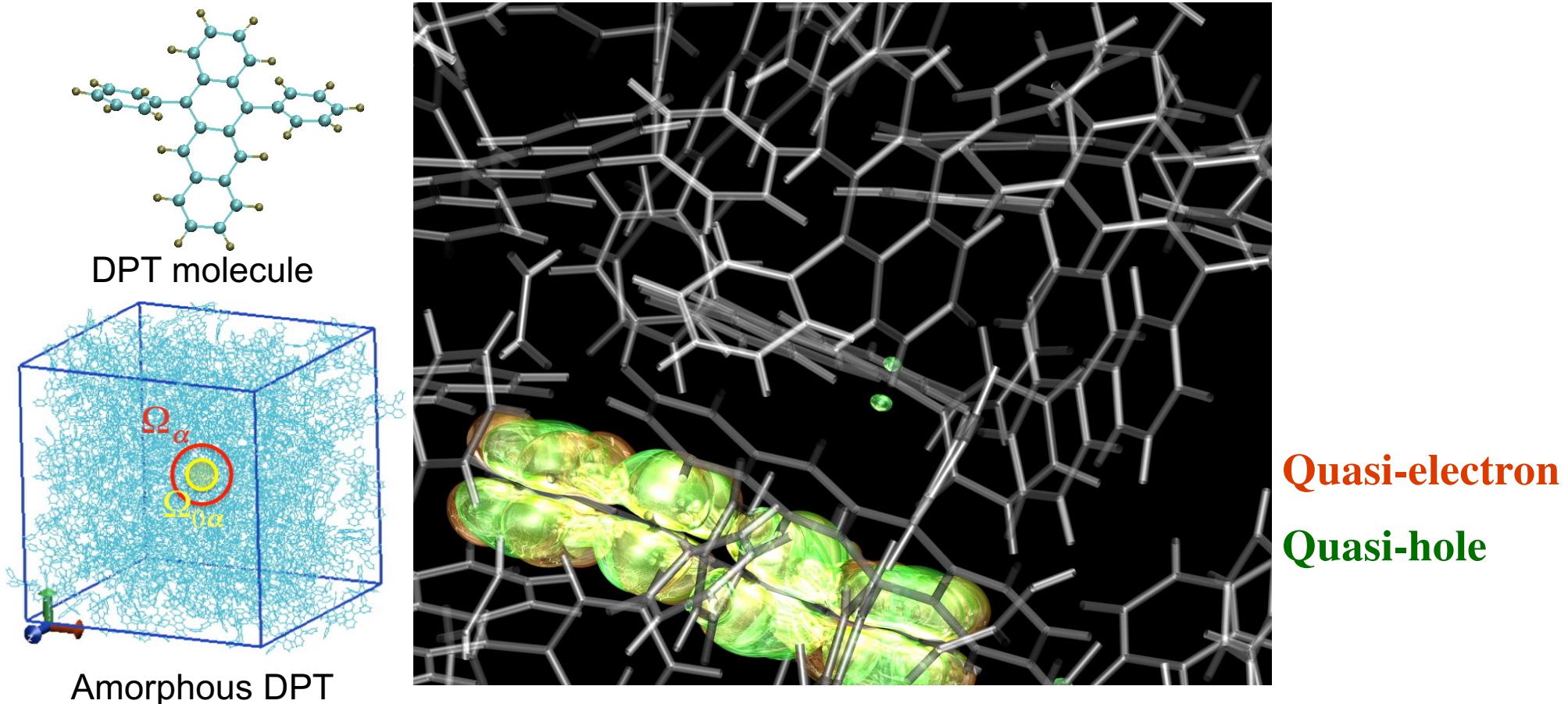
Globally informed local DC-DFT solutions are used in the recombine phase as compact bases to synthesize global properties in broad applications:

- 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

F. Shimojo et al., J. Chem. Phys. 140, 18A529 ('14); K. Nomura et al., IEEE/ACM SC14 ('14)

Singlet Fission in Amorphous DPT

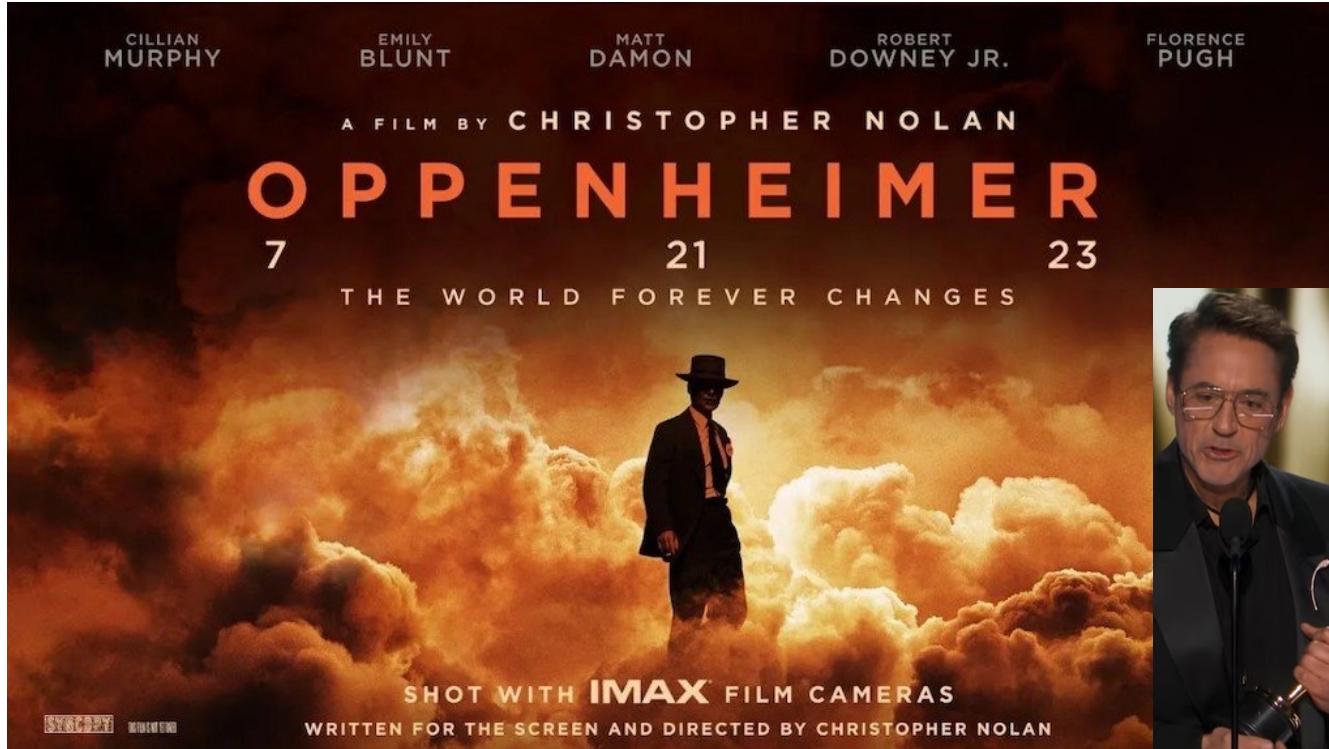
- Photo-current doubling by splitting a singlet exciton into 2 triplet excitons
- Singlet fission in mass-produced disordered organic solid → efficient low-cost solar cells
- Exp'l breakthrough: SF found in amorphous diphenyl tetracene (DPT)



- 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

Born-Oppenheimer Approximation

Basis of adiabatic quantum molecular dynamics



1927 Electrons stay in the ground state № 20

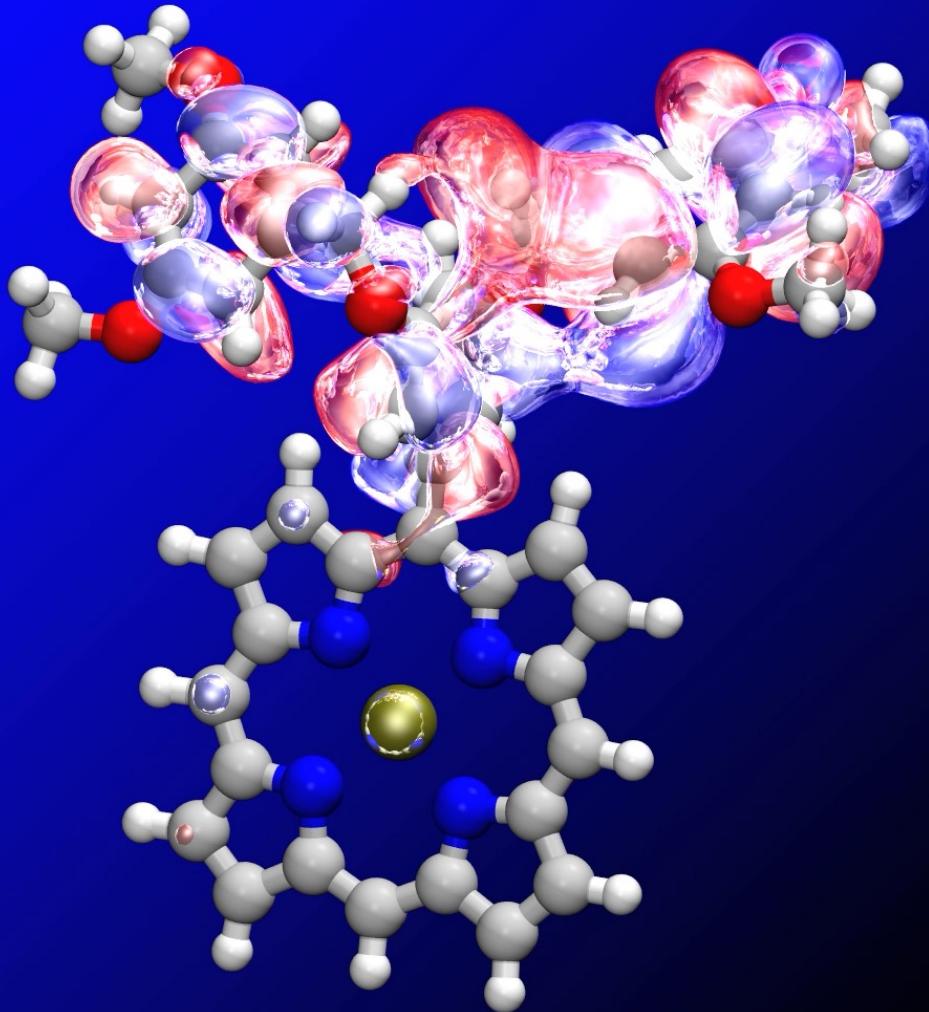
ANNALEN DER PHYSIK
VIERTE FOLGE. BAND 84

1. Zur Quantentheorie der Moleküle;
von M. Born und R. Oppenheimer



Nonadiabatic Quantum Molecular Dynamics

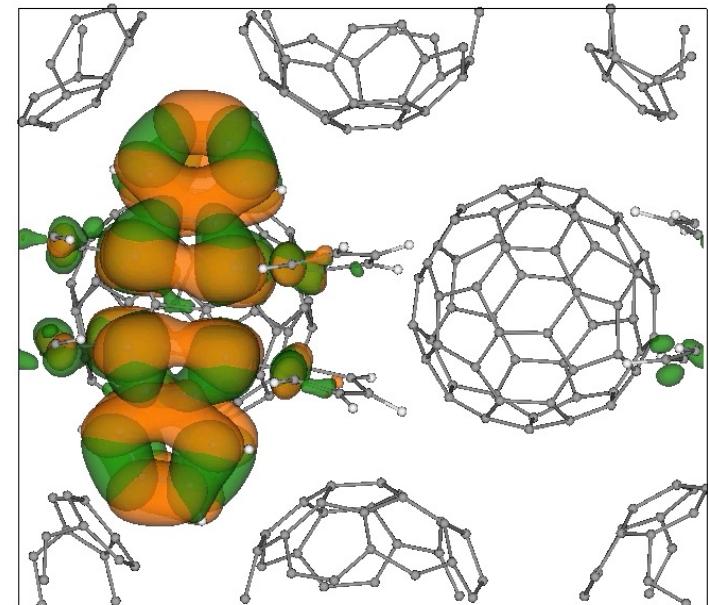
Beyond Born-Oppenheimer



Appl. Phys. Lett. **98**, 113302 ('11); *ibid.* **100**, 203306 ('12); *ibid.* **102**, 173301 ('13); *Comput. Phys. Commun.* **184**, 1 ('13); *J. Chem. Phys.* **140**, 18A529 ('14); *IEEE Computer* **48**(11), 33 ('15); *Sci. Rep.* **5**, 19599 ('16); *Nature Commun.* **8**, 1745 ('17); *Nano Lett.* **18**, 4653 ('18); *Nature Photon.* **13**, 425 ('19); *Science Adv.* **8**, eabk2625 ('22); *ibid.* **10**, eadp1890 ('24)

Zn porphyrin

Rubrene/C₆₀



quasi-electron; quasi-hole

- **Excited states:** Linear-response time-dependent density functional theory Casida, '95
- **Interstate transitions:** Surface hopping Tully, '90; Jaeger, Fisher & Prezhdo, '12

Attosecond Light-Matter Interaction

The Nobel Prize in Physics 2023



© Nobel Prize Outreach. Photo:

Clément Morin

Pierre Agostini

Prize share: 1/3



© Nobel Prize Outreach. Photo:

Clément Morin

Ferenc Krausz

Prize share: 1/3



© Nobel Prize Outreach. Photo:

Clément Morin

Anne L'Huillier

Prize share: 1/3

Attosecond = 10^{-18} seconds

The Nobel Prize in Physics 2023 was awarded to
Pierre Agostini, Ferenc Krausz and Anne L'Huillier
"for experimental methods that generate
attosecond pulses of light for the study of electron
dynamics in matter"

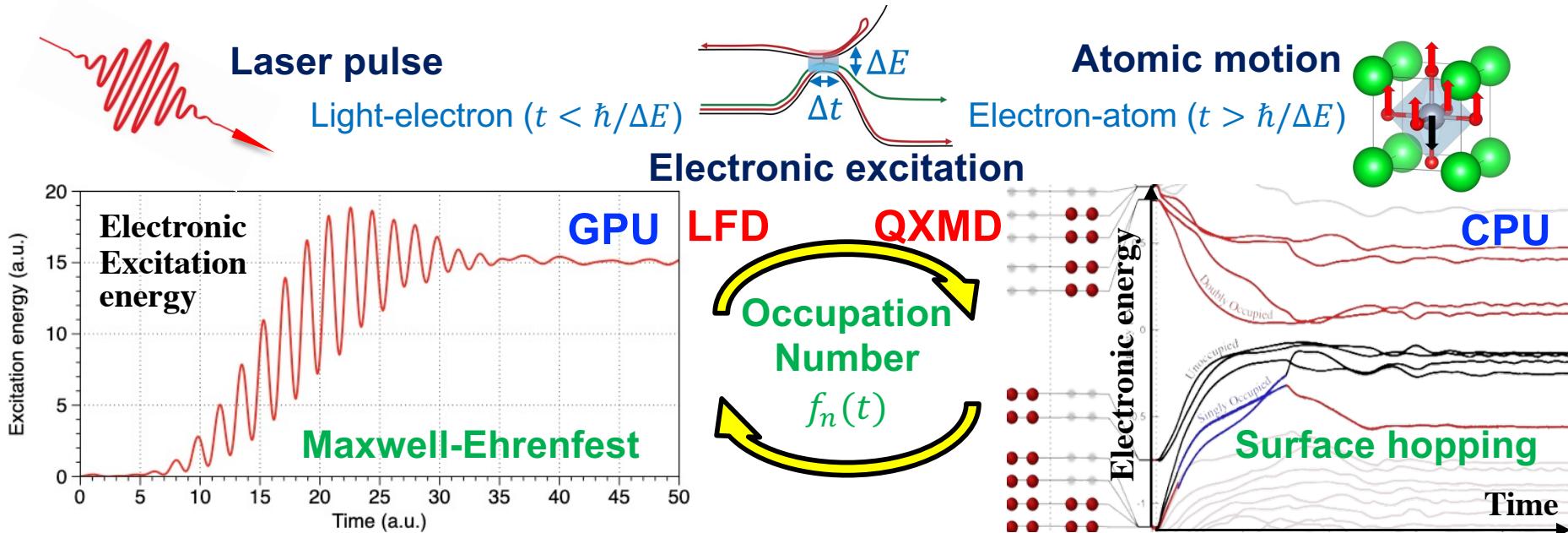
- Attosecond physics could revolutionize information technology by enabling ultrafast computing & sensing devices

cf. Petahertz electronics? Heide et al., Nat. Rev. Phys. 6, 648 ('24)

$10^6 \times$ faster than current GHz CMOS technology

Nonadiabatic Quantum MD: DC-MESH

- **DC-MESH (divide-&-conquer Maxwell + Ehrenfest + surface-hopping):** $O(N)$ algorithm to simulate photo-induced quantum materials dynamics
- **LFD (local field dynamics):** Maxwell equations for light & real-time time-dependent density functional theory equations for electrons to describe light-matter interaction
Yabana *et al.*, *Phys Rev B* **85**, 045134 ('12); Jastadt *et al.*, *Adv. Phys.* **68**, 225 ('19)
- **QXMD (quantum molecular dynamics with excitation):** Nonadiabatic coupling of excited electrons & ionic motions based on surface-hopping approach
- “Shadow” LFD (GPU)-QXMD (CPU) handshaking *via* electronic occupation numbers with minimal CPU-GPU data transfer



Linker *et al.*, *Science Adv.* **8**, eabk2625 ('22); *IEEE-PDSEC* ('24); *JPCL* **16**, 9267 ('25)

Paradigm Shift: DCR/MSA

- Solved the multiscale/multiphysics/heterogeneity/low-precision challenge by harnessing heterogeneity & low-precision arithmetic
- *Divide-conquer-recombine (DCR)*^{*} algorithms divide a problem into *not only spatial but also physical subproblems* of different computational characteristics, which are solved using appropriate methods on best-matching hardware units before recombined into a total solution

*K. Fukui, *Creation of Academics* (in Japanese, '87); F. Shimojo et al., *JCP* 140, 18A529 ('14)

- *Metamodel-space algebra (MSA)*[†] lets subproblems to *reside in respective hardware units*, while minimizing communication & precision requirements

[†]K. Morokuma et al., *J. Mol. Struct.* 1, 461 ('99); A. Nakano et al., DOI: 10.1142/9789812701084_0005 ('05)

- DCR/MSA delineates subproblems with *small dynamic ranges & minimal mutual information* with *parameterized precision*, which map well onto AI accelerators that support a spectrum of hybrid precision modes

Hardware supports precision/speed trade-off

Aurora allows 1|2|3 BF16 values to be accumulated in FP32 to provide varying precision/speed

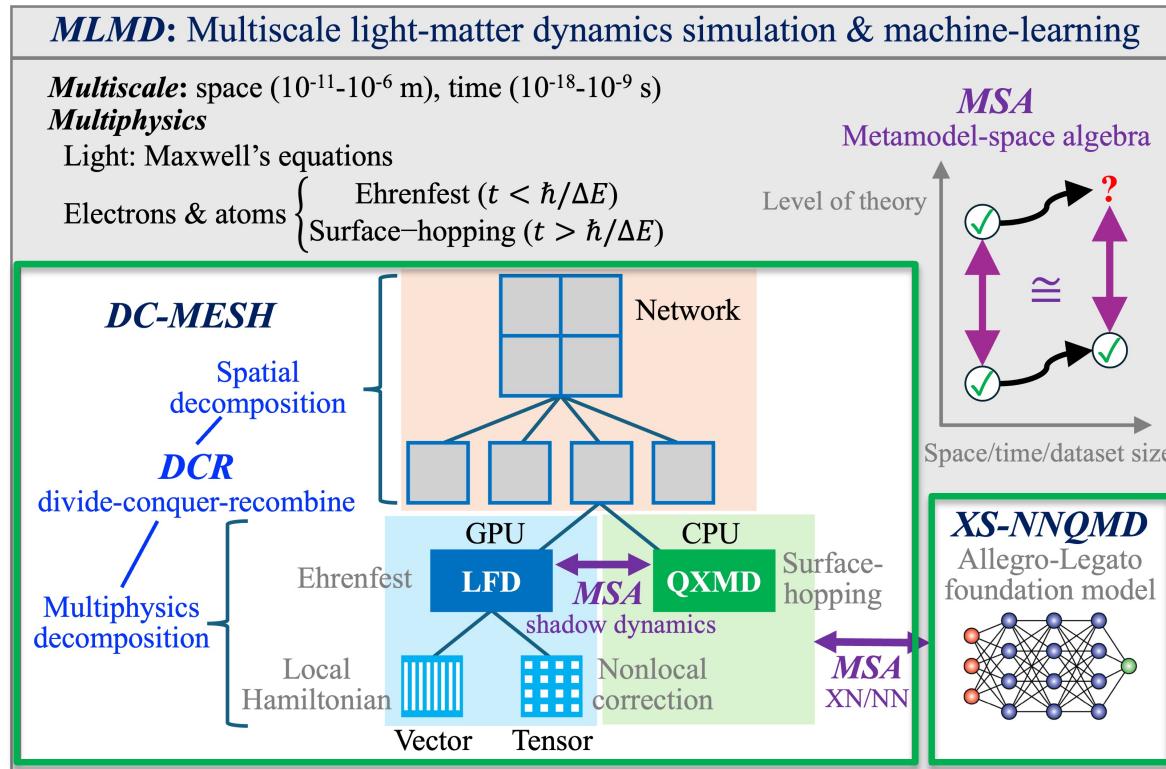
Metascalable (i.e., design-once, scale-to-future) algorithm-hardware co-design opportunities in the post-exascale era

T. M. Razakh et al., *IEEE/ACM SC* ('25)

AI-Enhanced Multiscale Simulation

- *Multiscale light-matter dynamics (MLMD): First-principles nonadiabatic quantum molecular dynamics (NAQMD) is boosted by AI-accelerated excited-state neural-network quantum molecular dynamics (XS-NNQMD)*

DCR:
Divide-
conquer-
recombine



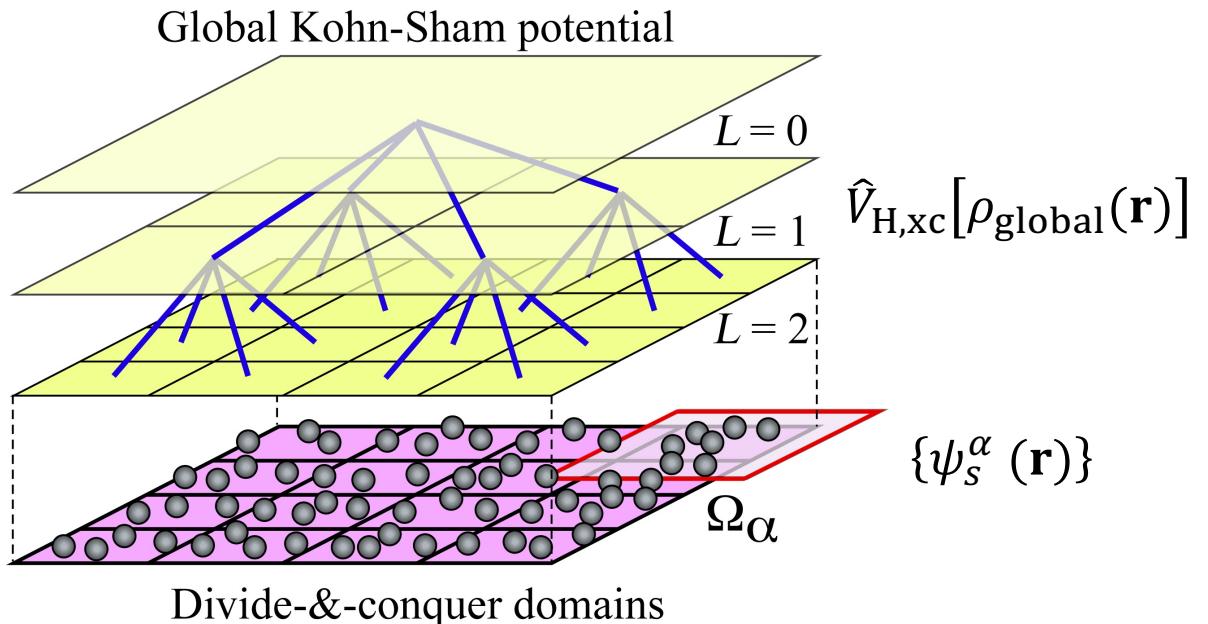
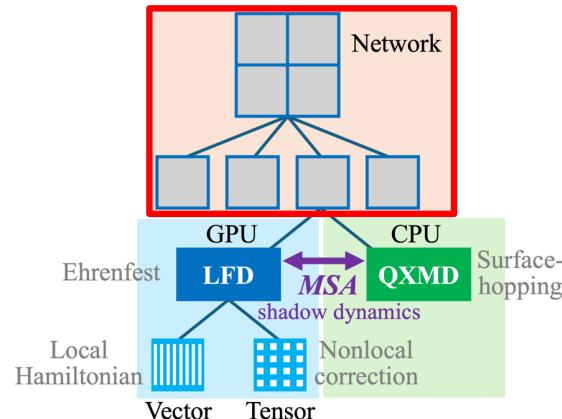
DC-MESH module:
Divide-&-conquer Maxwell-Ehrenfest-
surface hopping NAQMD

XS-NNQMD module:
Excited-state neural-network
quantum molecular dynamics

- **2024 Nobel physics & chemistry heralded the new era, where AI is embedded in the very fabric of science; this is an exemplar**

DCR1: Spatial Divide-&-Conquer DFT

Divide-conquer-recombine 1



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

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

Global-local
self-consistent
field (SCF)
iteration

- Global & domain electron densities

$$\rho_{\text{global}}(\mathbf{r}) = \sum_\alpha p_\alpha(\mathbf{r}) \rho_\alpha(\mathbf{r})$$

Domain support function

$$\sum_\alpha p_\alpha(\mathbf{r}) = 1$$

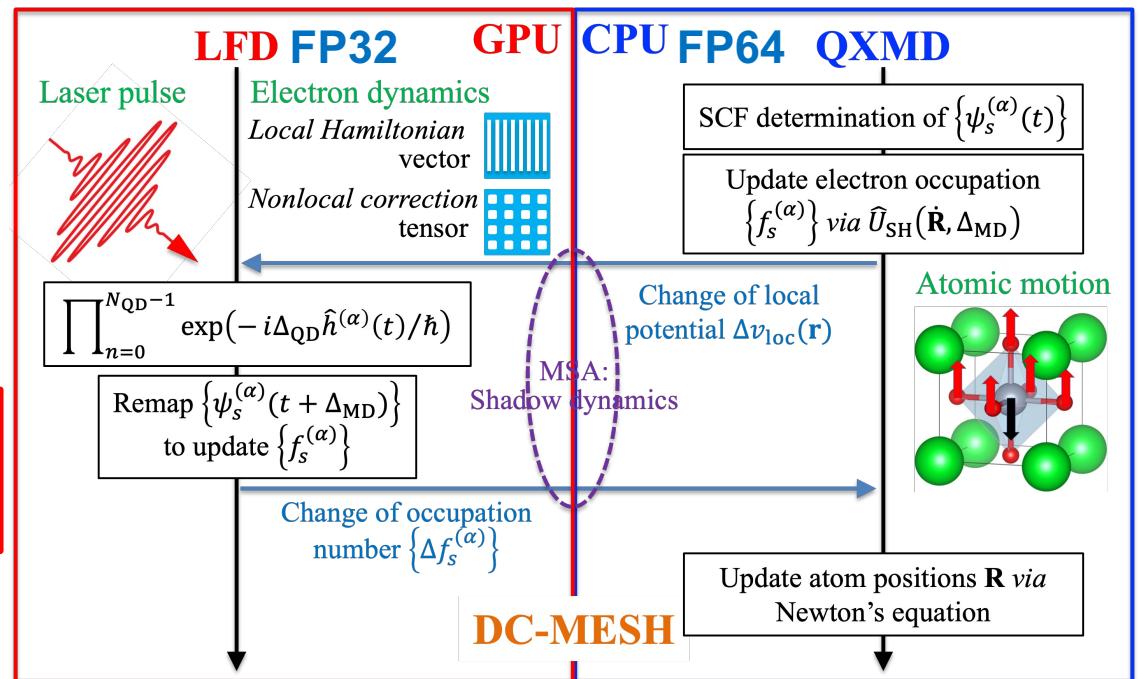
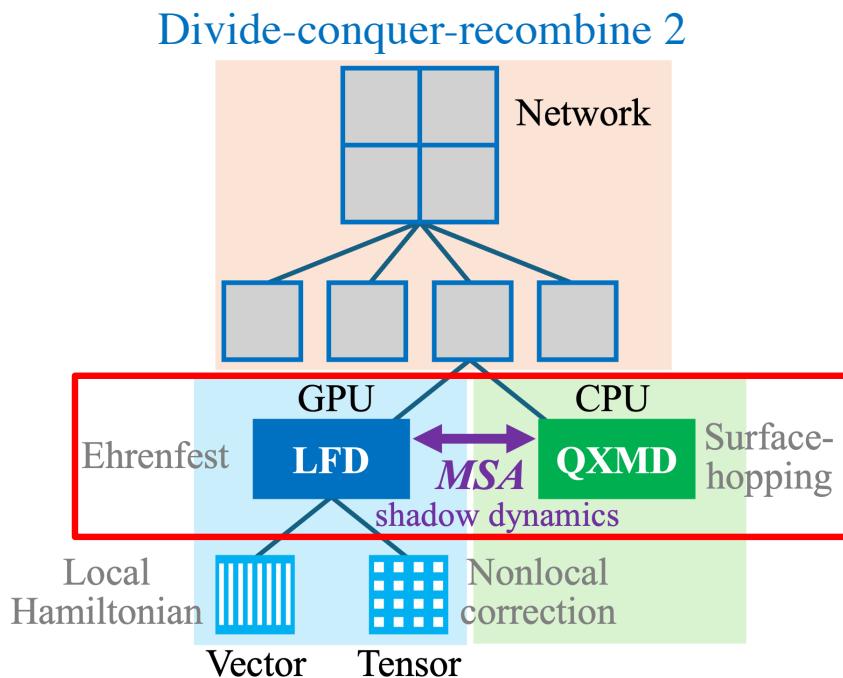
Global chemical potential

$$N = \int d\mathbf{r} \rho_{\text{global}}(\mathbf{r})$$

Yang, Phys. Rev. Lett. 66, 1438 ('91)

DCR2/MSA1: LFD-QXMD Decomposition

- **Local Field Dynamics (LFD):** Fine-level partial differential equations for light & electrons are data-parallel & fit naturally to hardware accelerators such as GPU
- **Quantum eXcitation Molecular Dynamics (QXMD):** Complex chemical interaction in coarse electron-atom description can take advantage of complex instruction sets in CPU



- **Shadow dynamics:** GPU-resident proxy captures effective action of LFD on QXMD through minimal information, i.e., electronic occupation numbers, $f_s^{(\alpha)} \in [0, 1]$

Niklasson et al., J. Chem. Phys. 158, 154105 ('23)

Small dynamic-range
~ low precision

cf. Mixed-precision occupation by Das, Gavini et al., Gordon-Bell prize ('23)

DCR3: Local-Nonlocal Split Operator

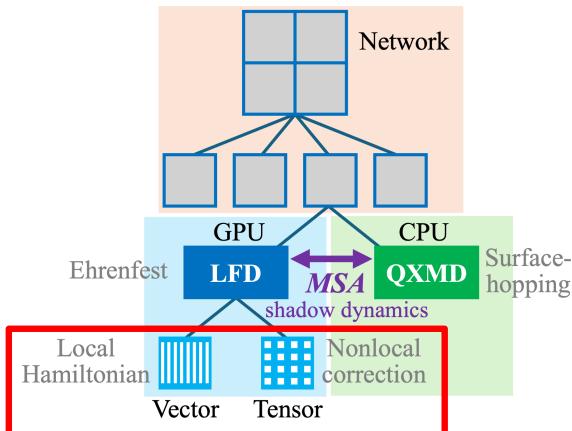
Trotterization
for quantum-dynamics steps

$$|\psi_{s\sigma}(t + \Delta)\rangle = \exp\left(-\frac{i\Delta}{\hbar} \hat{h}\right) |\psi_{s\sigma}(t)\rangle$$

Kohn-Sham (KS)
wave functions

Unitary time-
propagator

Divide-conquer-recombine 3



Electronic Hamiltonian operator

$$\hat{h} = \frac{1}{2m} \left(\frac{\hbar}{i} \nabla + \frac{e}{c} \mathbf{A} \right)^2 + \hat{v}_{\text{ion}} + v_{\text{H}} + \hat{v}_{\text{xc}} = \hat{h}_{\text{loc}} + \hat{v}_{\text{nl}}$$

Basis
switching

Local Hamiltonian
Finite difference basis,
Data parallel (vector)

Nonlocal correction
KS basis
Dense matrix (tensor)

Nakano, CPC 83, 181 ('94) Wang et al., JPCM 31, 214002 ('19)

Local-nonlocal split-operator (LNSO)

Vector-
tensor
splitting

Hybrid FP32/BF16 GEMM

$$\exp\left(-\frac{i\Delta}{\hbar} \hat{h}\right) = \frac{1 - \frac{i\Delta}{2\hbar} \hat{v}_{\text{nl}}}{\left\| 1 - \frac{i\Delta}{2\hbar} \hat{v}_{\text{nl}} \right\|} \exp\left(-\frac{i\Delta}{\hbar} \hat{h}_{\text{loc}}\right) \frac{1 - \frac{i\Delta}{2\hbar} \hat{v}_{\text{nl}}}{\left\| 1 - \frac{i\Delta}{2\hbar} \hat{v}_{\text{nl}} \right\|}$$

Perturbatively-small, bandgap-reproducing

Vlcek et al., JCP 150, 184118 ('19)

Wang et al., JPCM 31, 214002 ('19)

Stencil

$$\exp\left(-\frac{i\Delta}{\hbar} \hat{h}_{\text{loc}}\right) \frac{1 - \frac{i\Delta}{2\hbar} \hat{v}_{\text{nl}}}{\left\| 1 - \frac{i\Delta}{2\hbar} \hat{v}_{\text{nl}} \right\|}$$

Self-consistent propagation

Sato et al., JCP 143, 224116 ('15)

Implementation Innovations

- **Open programming:** Portable across supercomputers through MPI (message passing) + OpenMP (multithreading) + OpenMP target (GPU) + PyTorch (NNQMD)
- **Data/loop reordering, blocking & hierarchical parallelization:** Reuse spatial stencil operators for an array of electronic wave functions as structure of arrays (SoA), along with blocking for cache utilization & hierarchical parallelization over teams & threads
- **GPU resident kernels:** Shadow dynamics (MSA) makes large electronic wave functions GPU-resident to minimize CPU-GPU data transfer, facilitated by custom OpenMP target data allocator/destructor to ease programming
- **Parametrized mixed-precision computation** through class template with parameterized precision for FP64 QXMD, FP32 LFD to compute small occupation number ($\in [0, 1]$) correction, and hybrid FP32 (accumulation)/BF16 \times 1|2|3 (matrix multiplication) to compute perturbatively small nonlocal correction in LFD

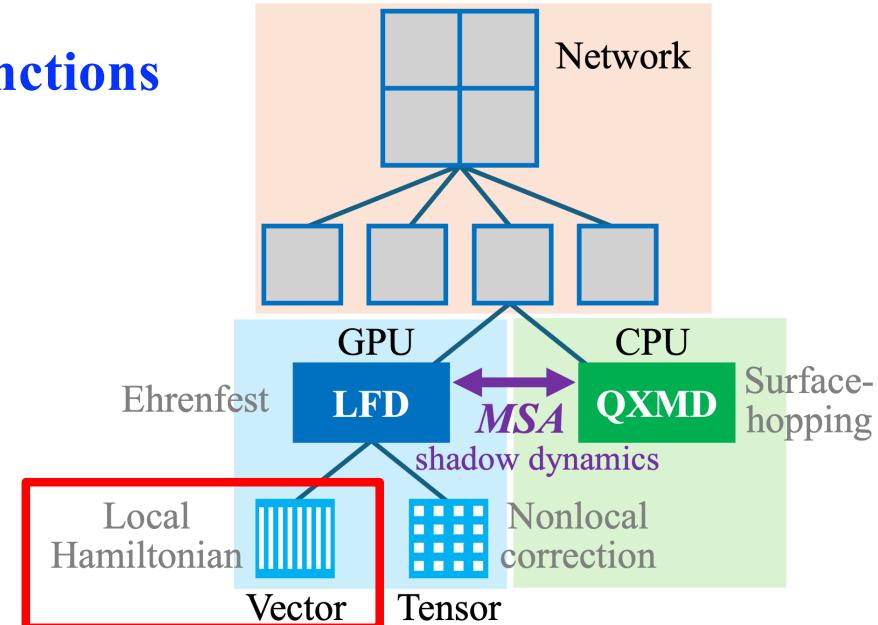
Local Time-Propagator Optimization

- Data/loop reordering, blocking & hierarchical parallelization

1. Reuse spatial stencil operators for an array of electronic wave functions each on N_{grid} spatial grid points as structure of arrays (SoA)

2. Blocking for cache utilization

3. Hierarchical parallelization over teams & threads on GPU



Runtime of the *kin_prop()* function for local time-propagation
in the LFD subprogram of the DC-MESH module

Implementation	Target	Runtime (s)	Speedup
Baseline	CPU	8.655	1
Data & loop re-ordering	CPU	2.356	3.67
Blocking/tiling	CPU	0.939	9.22
Hierarchical parallel regions	GPU	0.026	338

GEMMification of Nonlocal Computation

- Switch of representation from finite-difference to Kohn-Sham (KS) orbital bases makes nonlocal correction into dense complex-matrix multiplications

$$\underbrace{\Psi(t)}_{\text{FP32}} = \delta \underbrace{\Psi(0)}_{\text{BF16}} \underbrace{\Psi^\dagger(0)}_{\text{BF16}} \underbrace{\Psi(t)}_{\text{BF16}}$$

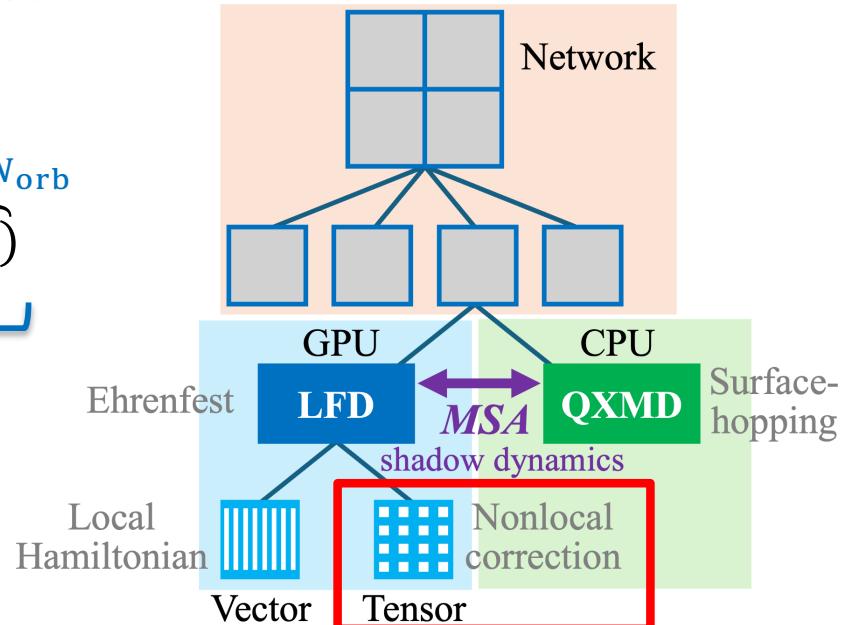
$N_{\text{grid}} \times N_{\text{orb}}$ $N_{\text{grid}} \times N_{\text{orb}}$ $N_{\text{orb}} \times N_{\text{grid}}$ $N_{\text{grid}} \times N_{\text{orb}}$

- FP32 to accumulate perturbatively small BF16 matrix multiplications with sufficient precision

N. Piroozan et al., SCW24, DOI: 10.1109/SCW63240.2024.00187

Single GPU-tile performance of DC-MESH for several problem sizes

Number of KS orbitals	TFLOP/s	% of FP64 peak (23 TFLOP/s)
256	5.22 (FP32)	22.69
864	9.74 (FP32)	42.35
1024	14.98 (FP32)	65.16
1024	17.95 (FP32/BF16)	78.03
1024	7.69 (FP64)	33.43



Single GPU-tile performance of hot-spot kernels

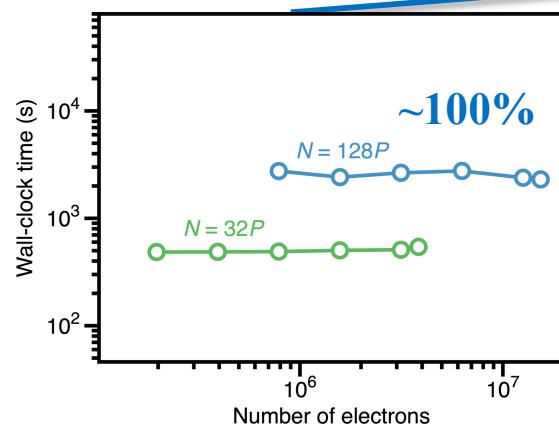
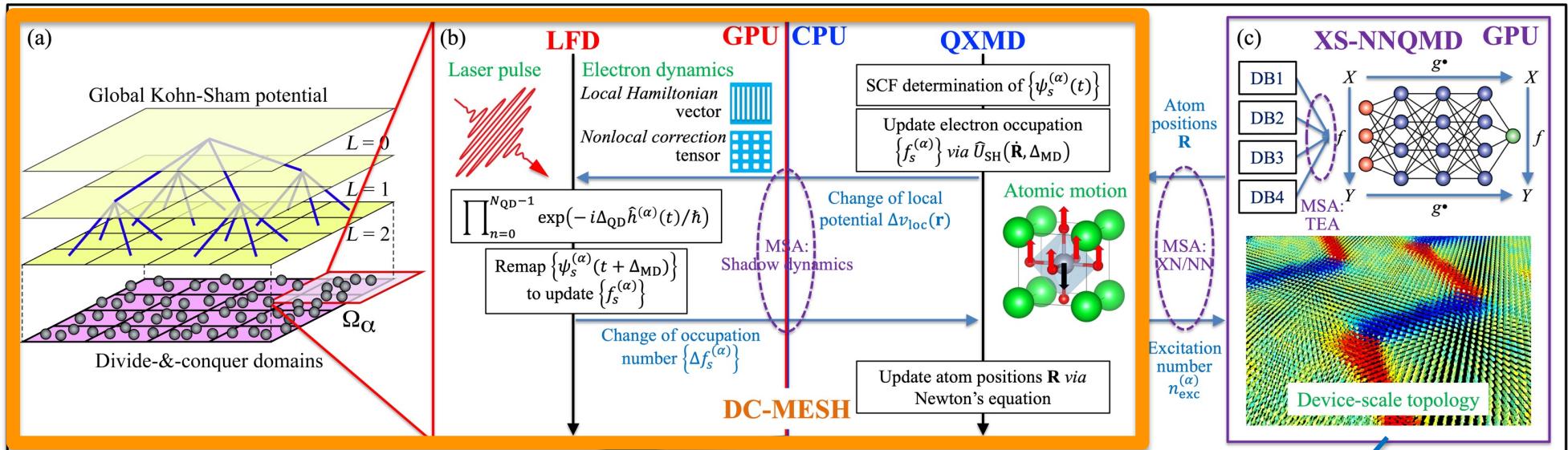
Kernels	TFLOP/s	% of peak
CGEMM (1)	18.72 (FP32)	81.39
CGEMM (2)	21.66 (FP32)	94.17
nlp_prop()	16.02 (FP32)	69.65
kin_prop()	3.51 (FP32)	15.26

Local Nonlocal

Exascale Light-Matter Dynamics

- Breaking the Exaflop/s barrier:** On 60,000 GPUs of Aurora: **1.87 Exaflop/s**; **152 \times** & **3,780 \times** improvements of time-to-solution over state-of-the-art for **15.4M-electron DC-MESH NAQMD** & **1.2T-atom excited XS-NNQMD**; **nearly perfect parallel efficiency**

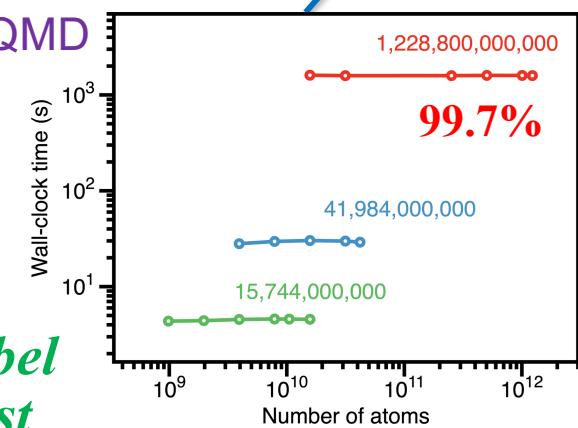
Exaflop/s = 10^{18} mathematical operations per second



DC-MESH NAQMD

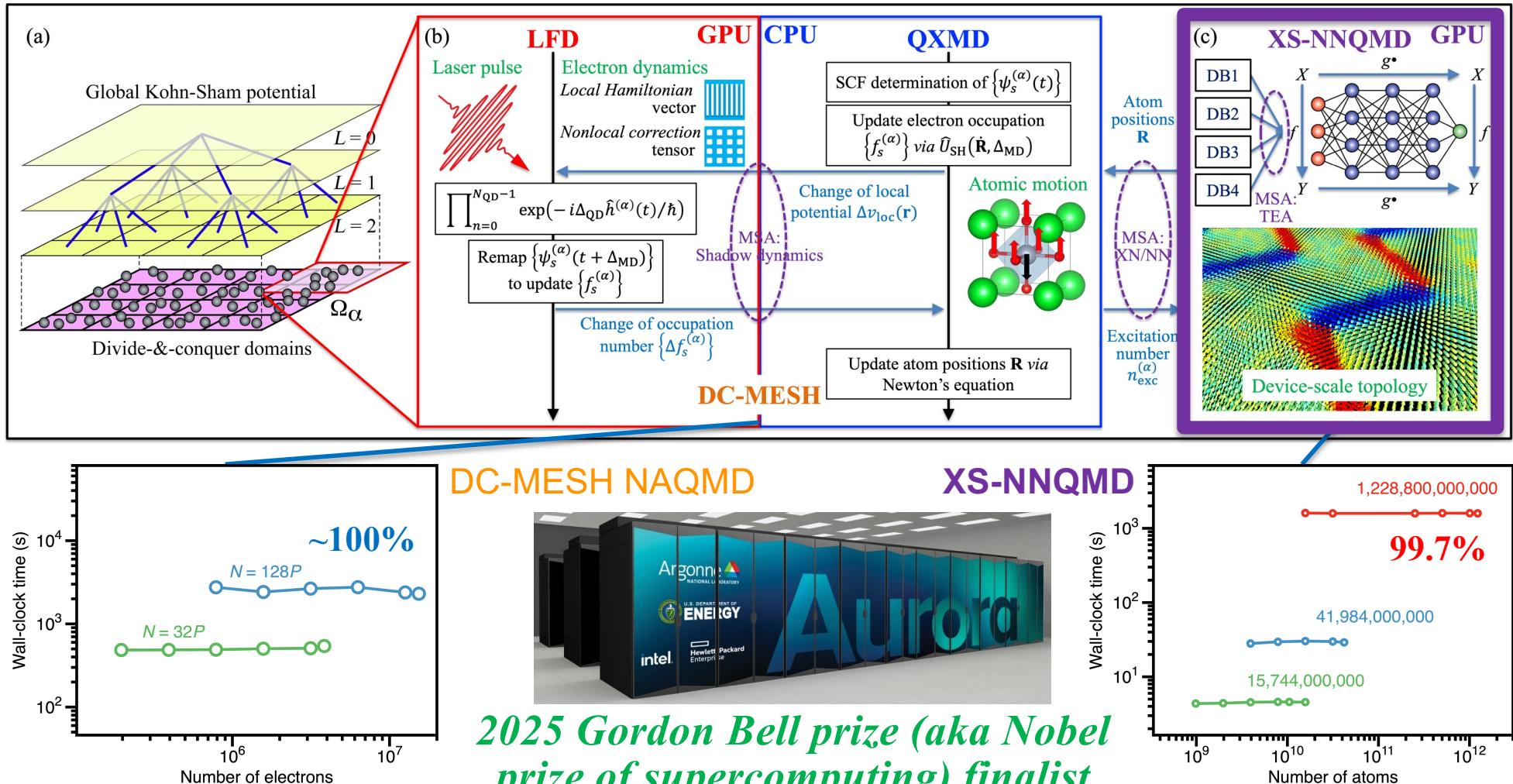


2025 Gordon Bell prize (aka Nobel prize of supercomputing) finalist

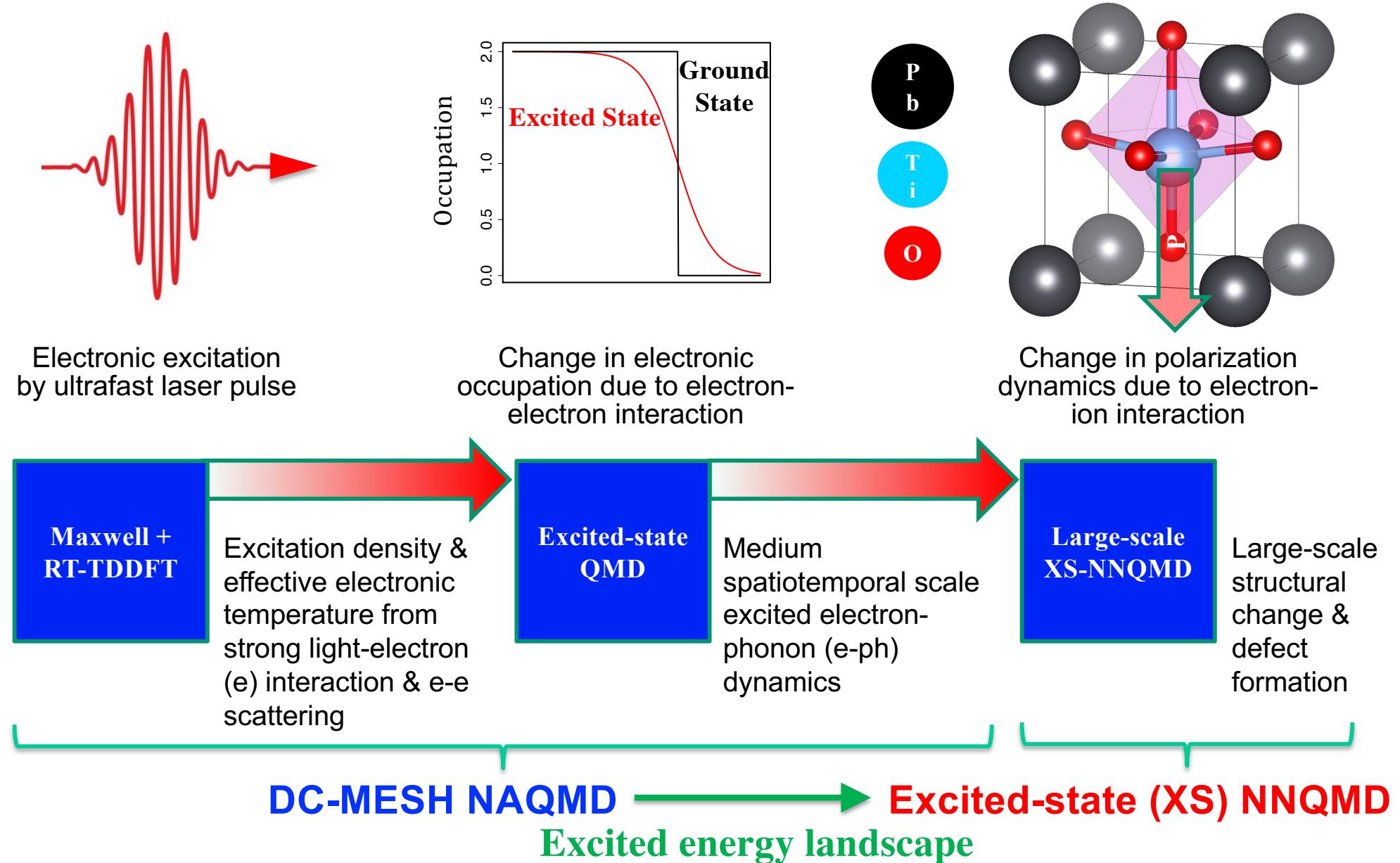


Excited-State Neural-Network QMD

- Breaking the Exaflop/s barrier: On 60,000 GPUs of Aurora: 1.87 Exaflop/s; 152× & 3,780× improvements of time-to-solution over state-of-the-art for 15.4M-electron DC-MESH NAQMD & 1.2T-atom excited XS-NNQMD; nearly perfect parallel efficiency*
- Exaflop/s = 10^{18} mathematical operations per second

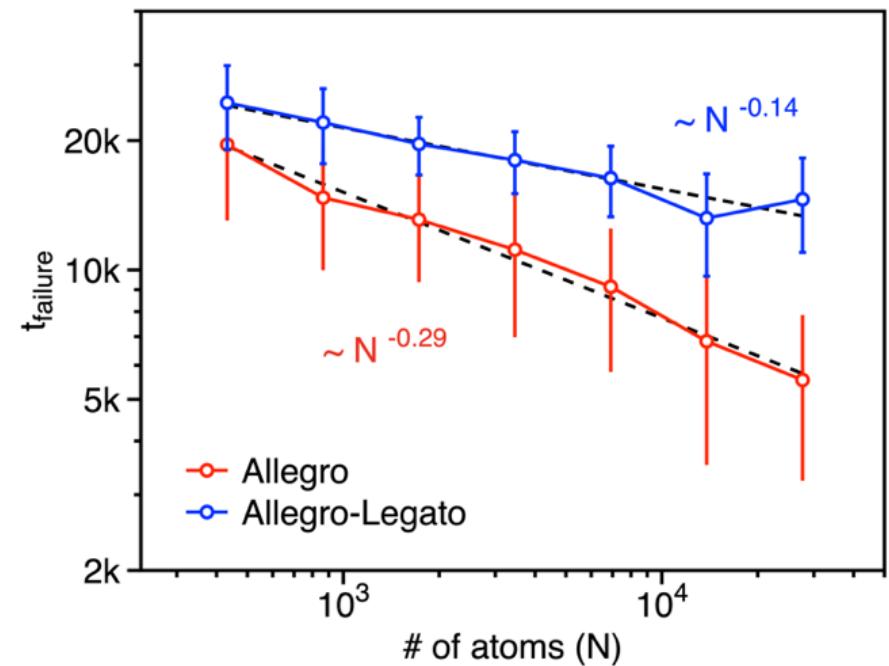
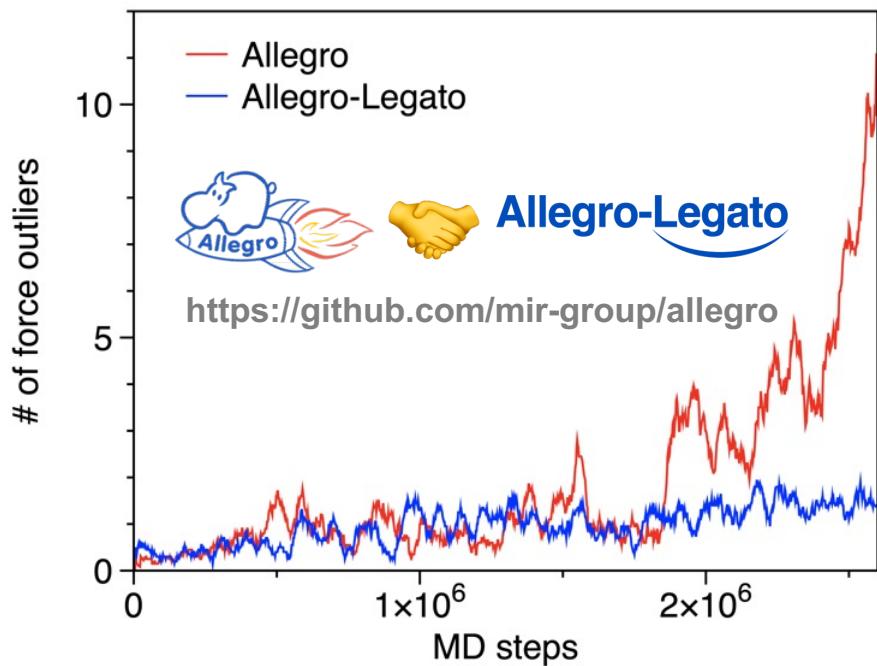


Multiscale DC-MESH + XS-NNQMD



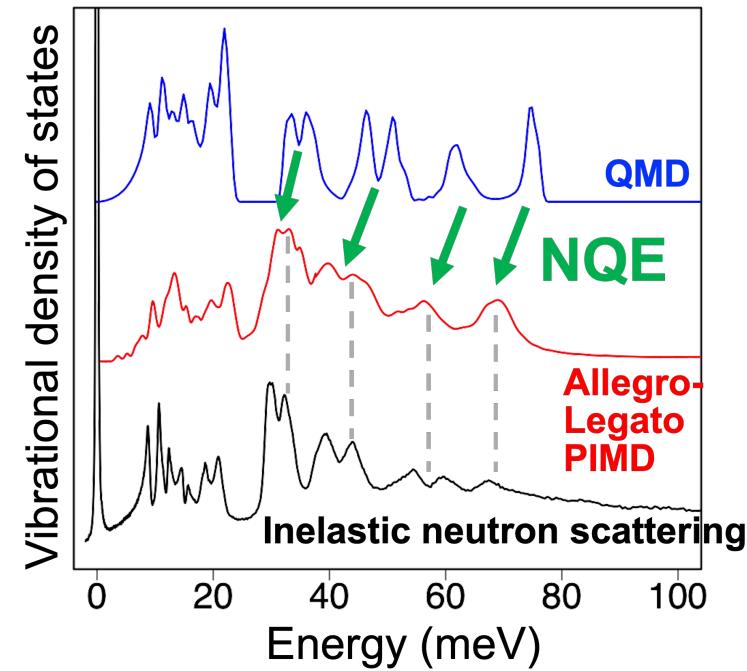
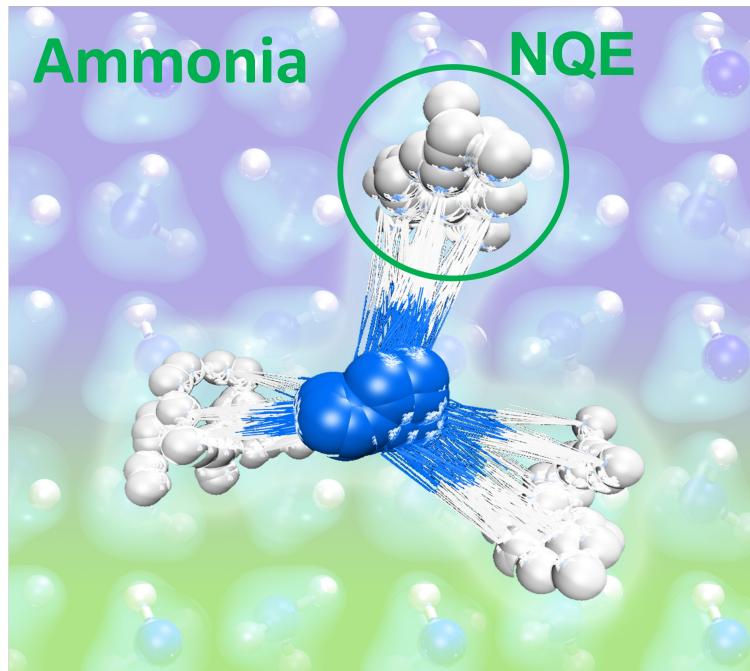
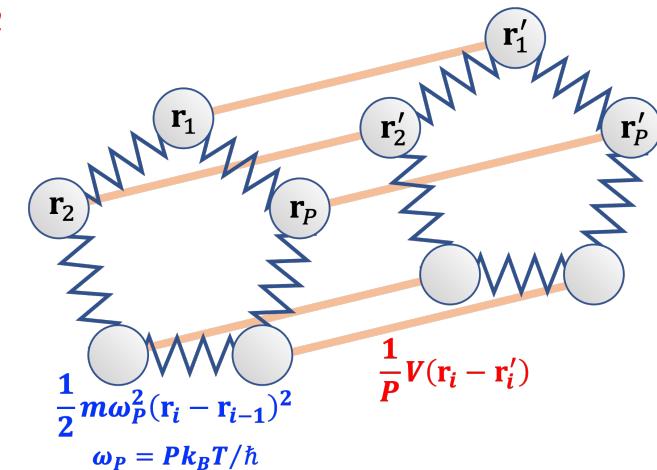
Fast & Robust NNQMD: Allegro-Legato

- **Allegro (fast) NNQMD:** State-of-the-art *accuracy & speed* founded on *group-theoretical equivariance* & local descriptors Musaelian et al., *Nat. Commun.* **14**, 579 ('23)
- **Fidelity-scaling problem:** On massively parallel computers, growing number of unphysical (adversarial) force predictions prohibits simulations involving larger numbers of atoms for longer times *It's symmetry & data locality!*
- **Allegro-Legato (fast and “smooth”):** *Sharpness aware minimization (SAM)* enhances the *robustness* of Allegro through improved smoothness of loss landscape
 $w_* = \operatorname{argmin}_w [L(w) + \max_{\|\epsilon\|_2 \leq \rho} \{L(w + \epsilon) - L(w)\}]$ (L : loss; w : model parameters)
- **Elongated time-to-failure scaling, $t_{\text{failure}} = O(N^{-\beta})$,** without sacrificing accuracy or speed, thereby achieving spectroscopically stable long-time Hamiltonian trajectory



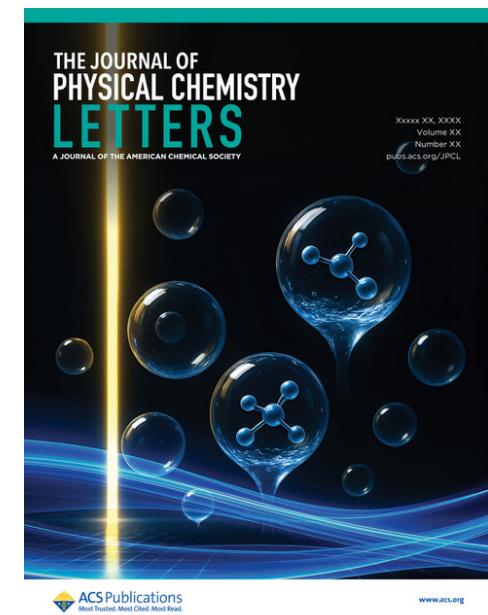
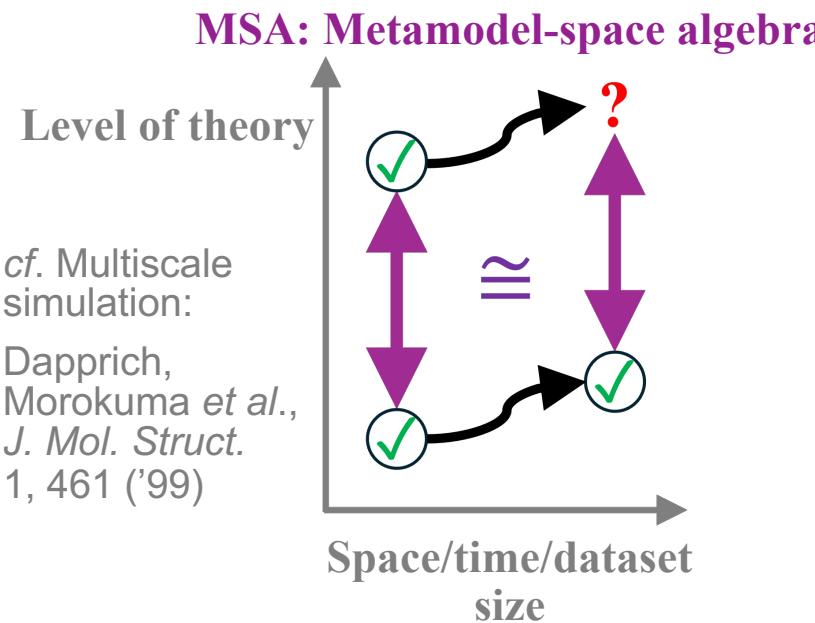
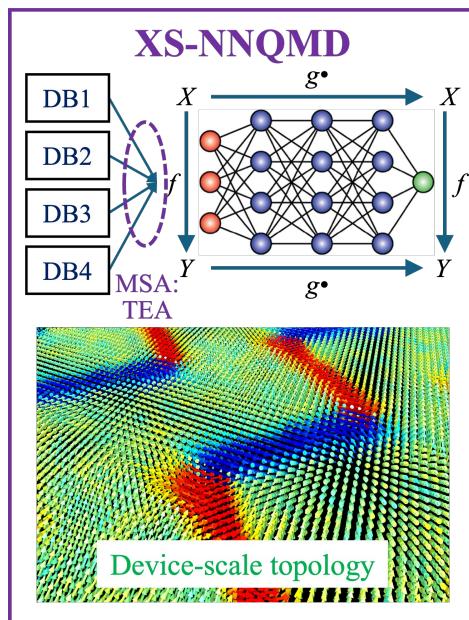
Nuclear-Quantum NNQMD

- **Allegro-Legato-PIMD:** Incorporate *nuclear quantum effect (NQE)* through path-integral molecular dynamics (PIMD)
- NNQMD trained by QMD achieves the required large number (P) of replicas at low temperature & long-time Hamiltonian dynamics to resolve fine vibrational structures
- NQE down-shifts inter-molecular vibrational modes in ammonia to explain high-resolution inelastic neutron scattering experiments *Spectroscopically-stable long-time NNQMD!*

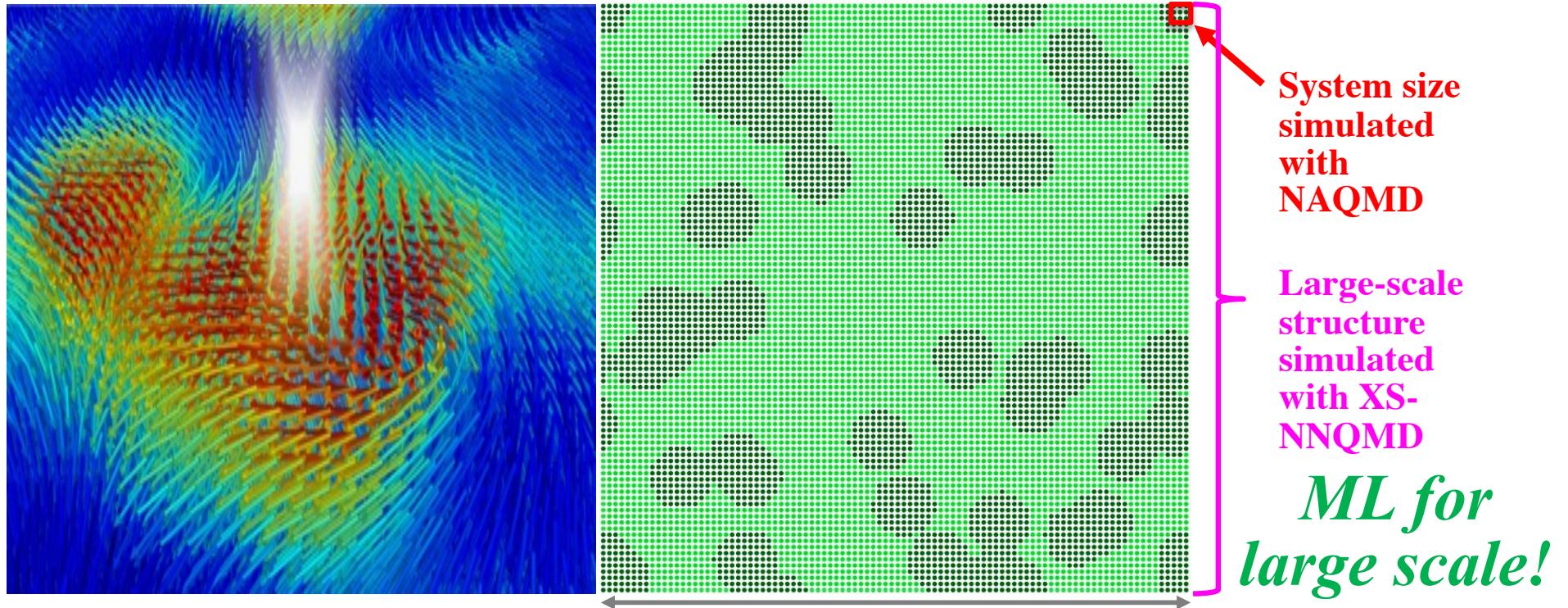


MSA2: Allegro-FM (Foundation Model)

- Foundation models are a paradigm shift in AI, where a single universal model acquires sufficient generalizability for diverse downstream tasks
- *Allegro-FM* describes many material properties & processes covering 89 elements in the periodic table, exhibiting *emergent capabilities* for which the model was not trained
- Universality was achieved by unifying large databases of multiple fidelity through affine (shift & scale) transformations in a metamodel-space: *total energy alignment (TEA)* Shiota et al., arXiv:2412.13088 ('24)

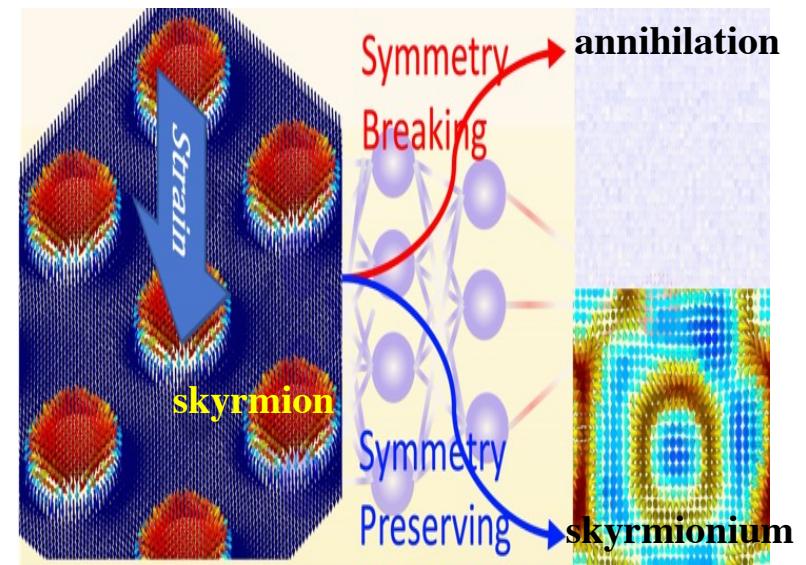


Application: Ferroelectric Opto-Toptronics



- Quantized ferroelectric topology is protected against thermal noise → future ultralow-power opto-electronics applications
- Billion-atom NNQMD revealed photo-induced topological phase-transition dynamics (*cf.* Kibble-Zurek mechanism in cosmology)
- Symmetry-controlled skyrmion-to-skyrmionium^{*} switching *Composite of skyrmions with opposite topological charges

Linker *et al.*, *Science Adv.* **8**, eabk2625 ('22);
JPCL **13**, 11335 ('22); *Nano Lett.* **23**, 7456 ('23)



What We Have Learned

- Multiresolution algorithms based on $O(N)$ *divide-&-conquer tree* will continue to scale on future computer architectures (i.e., *metascalable*)
e.g., wavelets, multigrid, fast multipole method
- *Divide-conquer-recombine (DCR)* algorithms divide a problem into *not only spatial but also physical subproblems* of different computational characteristics, which are solved using appropriate methods on best-matching hardware units before recombined into a total solution, thus providing *algorithm-hardware co-design* pathways at the nexus of post-exascale computing, quantum computing & AI

Use in your application domain!