Quantum-Mechanical Molecular Dynamics for Distributed Computing and AI Hardware

IPAM UCLA March 27-31

Anders M. N. Niklasson

Theoretical Division, LANL

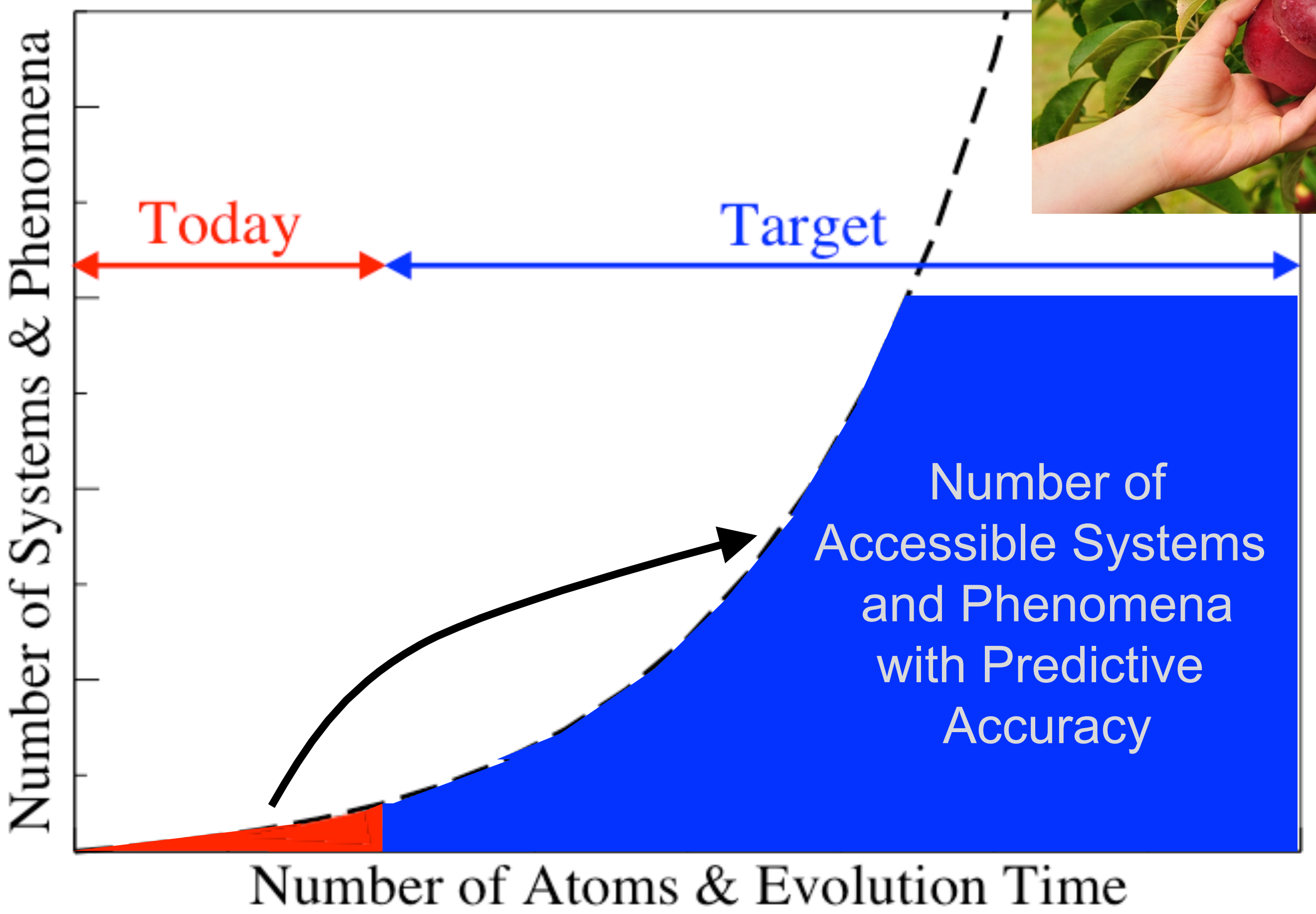
Acknowledgements:
Abrikosov, Albaugh, Aradi, Barros, Bergman, Bjorgaard, Bock, Challacombe, Cawkwell, Coe, Delin, Djidjev, Frauenheim, Germann, Goff, Grånäs, Head-Gordon, Holmström, Johansson, Karplus, Kulichenko, Kolesov, Kroonblawd, Martinez, Minszewski, Negre, Mohd-Yosuf, Odell, Rubensson, Sanville, Skylaris, Smith, Souvatzis, Steneteg, Swart, Tymczak, Tretiak, Velizahnen, Vitale, Voter, Wall, Weber, Zhang, Zheng, …
DOE funding: Office of Basic Energy Sciences CTC, ECP, LANL-LDRD, IPAM

Christian Negre, Joshua Finkelstein, Mike Wall, Susan Minszewski

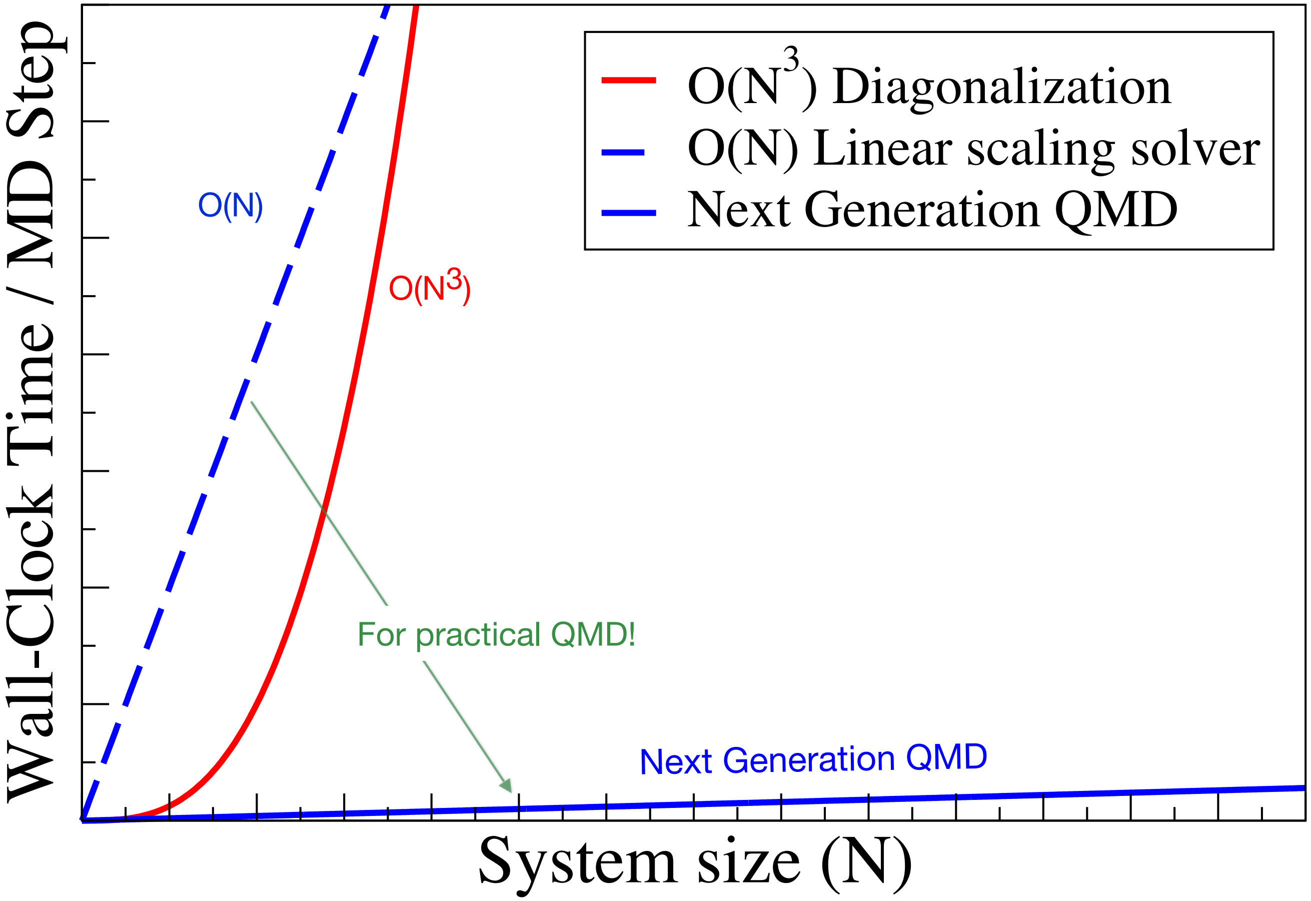
Los Alamos
NATIONAL LABORATORY
EST.1943

LA-UR-23-22859

Motivation (Low-hanging Fruit?)

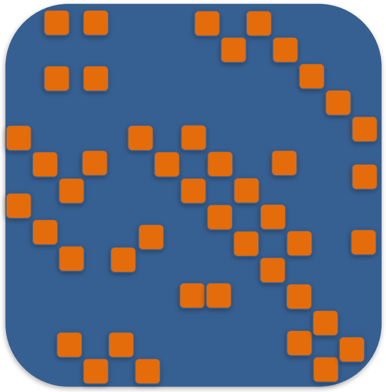


Next Generation QMD

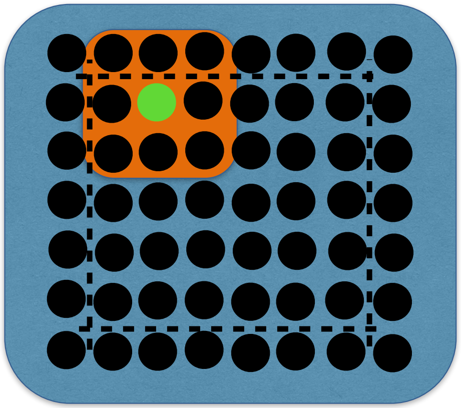


Linear Scaling Electronic Structure Theory

Sparse Matrix Algebra



Divide and Conquer



Sparse Matrix Algebra

Non-linear eigenvalue equation

$$H[\rho]\Psi_i = \epsilon_i \Psi_i$$

$$\rho(\mathbf{r}) = \sum_{i \in \text{occ.}} |\Psi_i|^2$$

$$E_s = \sum_{i \in \text{occ.}} \epsilon_i$$

$$\#\text{SCF} \times \mathcal{O}(N^3)$$

Density matrix calculation

$$P(H)\Psi_i = \theta(\mu - \epsilon_i)\Psi_i$$

$$\rho(\mathbf{r}) = P(\mathbf{r}, \mathbf{r}')|_{\mathbf{r}=\mathbf{r}'}$$

$$E_s = Tr\,[PH]$$

$$\#\text{SCF} \times \mathcal{O}(N^3)$$

Fermi operator expansion

$$P(H) = \theta\,(\mu I - H[n]) \approx \sum_n c_n T_n(H)$$

$$\#\text{SCF} \times \mathcal{O}(N)$$

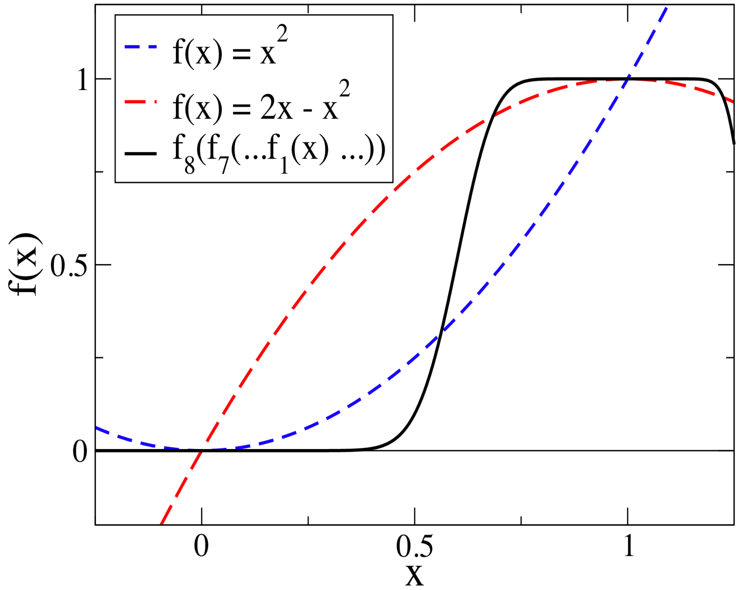
Thresholded Sparse Matrix Algebra

Los Alamos
NATIONAL LABORATORY
EST. 1943

SP2 *O(N)* solver

Recursive Fermi operator expansion

$$\mathbf{P} = \theta\left(\mu\mathbf{I} - \mathbf{H}\right) = \lim_{n\to\infty} f_n(f_{n-1}(\ldots f_0(\mathbf{H})\ldots))$$



$$\rho(\mathbf{r}) = P(\mathbf{r},\mathbf{r})$$

$$E_s = Tr\left[PH\right]$$

$$\mathbf{X}_0 = f_0(\mathbf{H}) = \frac{\varepsilon_{\max}\mathbf{I} - \mathbf{H}}{\varepsilon_{\max} - \varepsilon_{\min}}$$

$$\mathbf{X}_{n+1} = \mathbf{X}_n{}^2 \quad \text{if} \quad Tr[\mathbf{X}_n] > N_{occ}$$

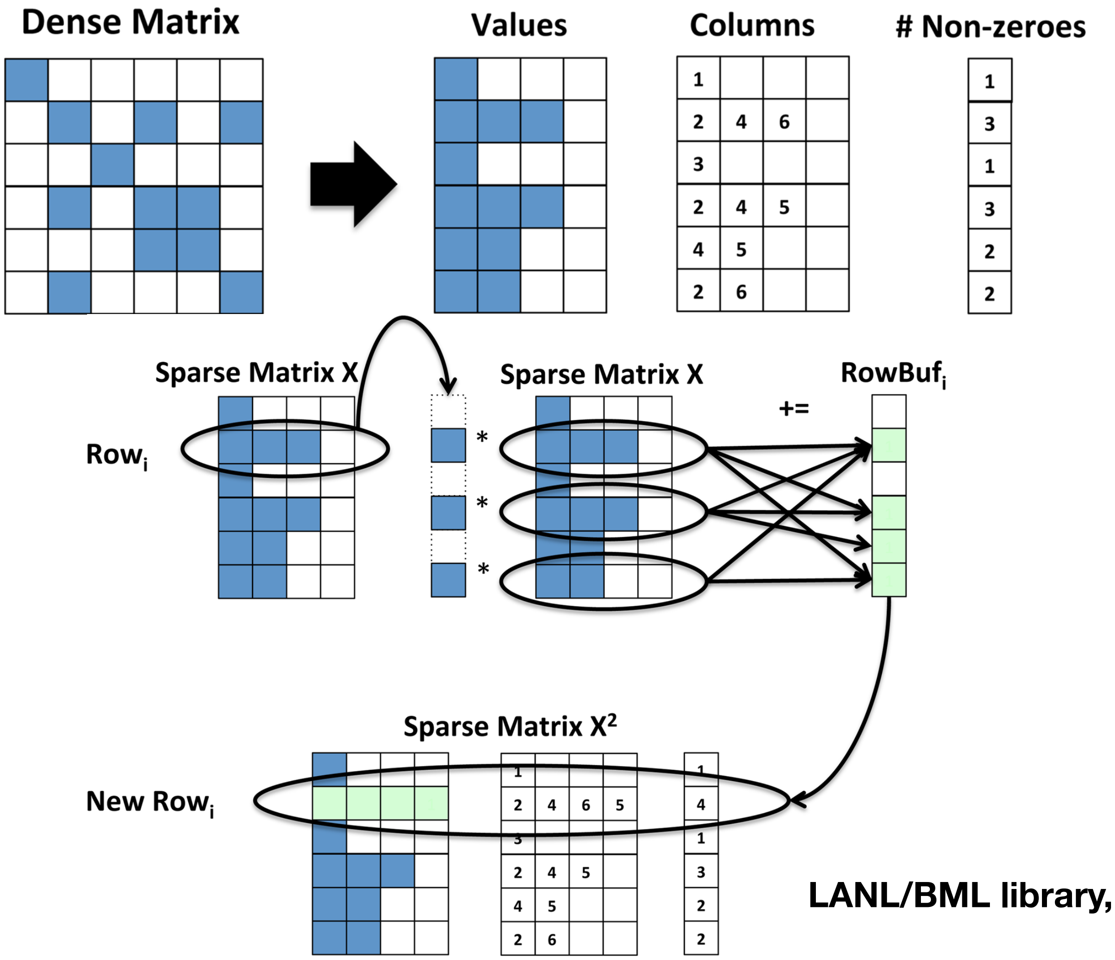
$$\mathbf{X}_{n+1} = 2\mathbf{X}_n - \mathbf{X}_n{}^2 \quad \text{else}$$

$$\mathbf{P} = \lim_{n\to\infty} \mathbf{X}_n \qquad T_e = 0$$

30 multiplications
gives an expansion order > 1 Billion!
No Gibbs oscillations!

A.M.N. Niklasson, Phys. Rev. B ***66****, 155115 (2002)*

Parallel O(N) Matrix-Matrix Multiplication



Columns

1			
2	4	6	
3			
2	4	5	
4	5		
2	6		

# Non-zeroes

1
3
1
3
2
2

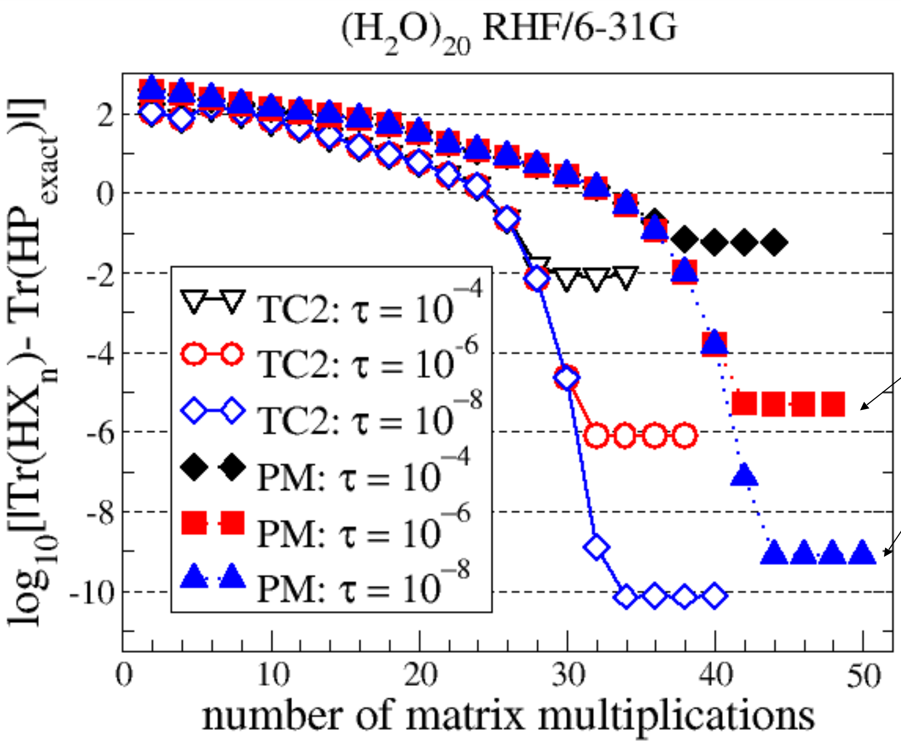
Sparse Matrix X^2

1				1
2	4	6	5	4
3				1
2	4	5		3
4	5			2
2	6			2

LANL/BML library, on Github

Los Alamos
NATIONAL LABORATORY
EST. 1943

Tunable energy convergence

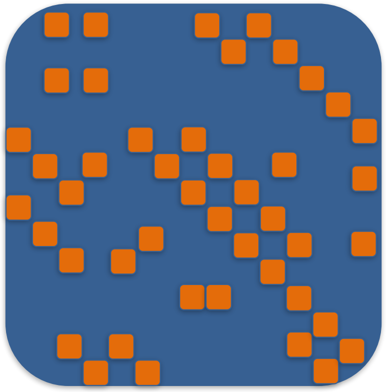


Numerically stable!

*SP2 (TC2): Niklasson, Phys. Rev. B, **66,** 155115 (2002)*
*PM: Palser and Manolopolus, Phys. Rev. B, **58**, 12704 (1998)*
Comparisons: E.H. Rubensson and E. Rudberg,
J. Phys.: Condens. Matter, 23, 075502 (2011)

Linear Scaling Electronic Structure Theory

Sparse Matrix Algebra

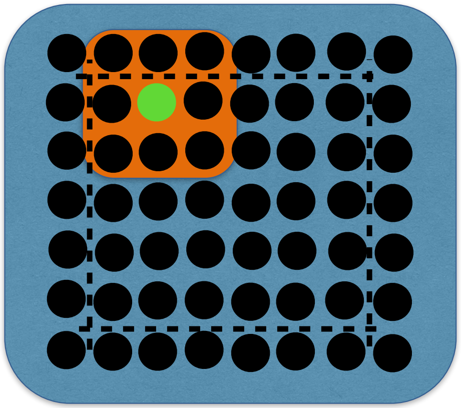


a) Easy to control error

b) Hard to parallelize on distributed memory

c) Small overhead

Divide and Conquer



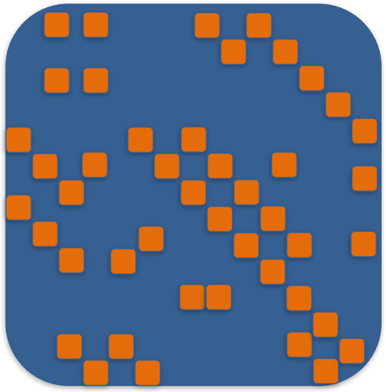
a) Hard to control error

b) Easy to parallelize on distributed memory

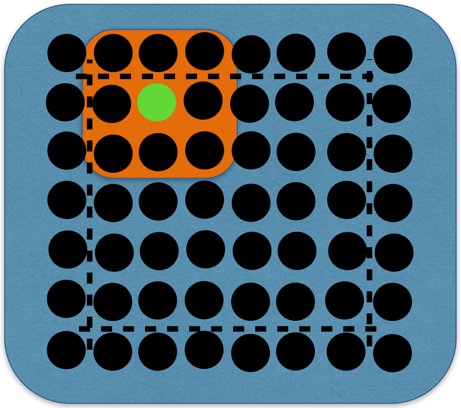
c) Large overhead

Linear Scaling Electronic Structure Theory

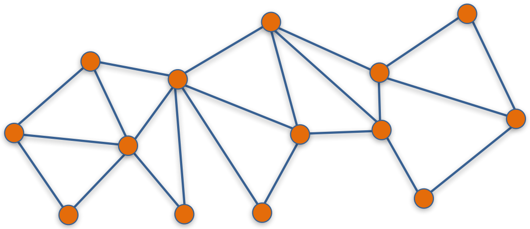
Sparse Matrix Algebra



Divide and Conquer



Graph Theory



a) Easy to control error?

b) Easy to parallelize on distributed memory?

c) Small overhead?

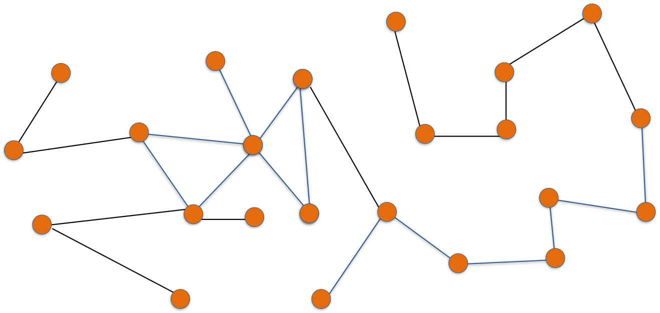
AMNN et al. *“Graph-based linear scaling electronic structure theory”*, J. Chem. Phys. 144, 234101 (2016)

Graph-based Electronic Structure Theory

Data dependency Graph G_τ

$$G_\tau \leftarrow \lfloor \text{Fermi Operator Expansion} \rfloor_{\tau(\text{global})} = D_\tau$$

H



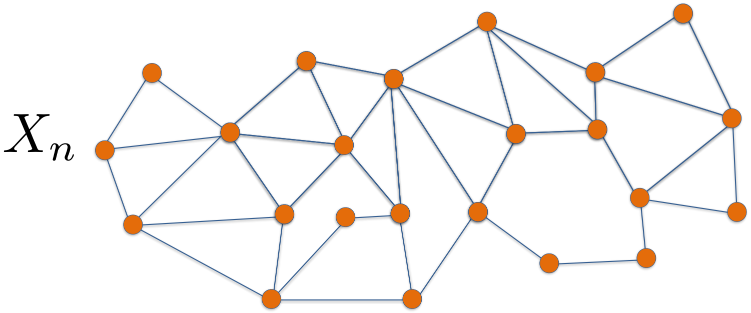
$$D = \theta(\mu I - H) \approx \sum_n c_n T_n(H)$$

AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

Graph-based Electronic Structure Theory

Data dependency Graph G_τ

$$G_\tau \leftarrow \lfloor \text{Fermi Operator Expansion} \rfloor_{\tau(\text{global})} = D_\tau$$

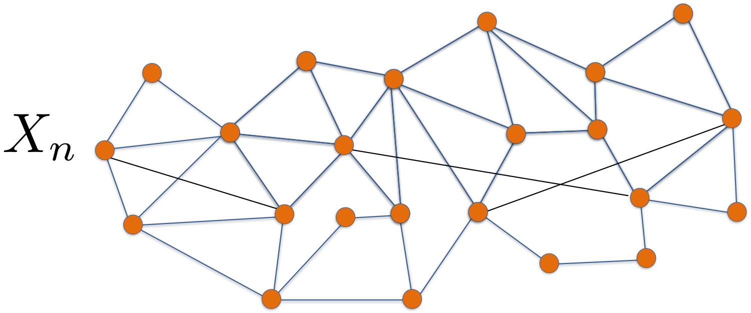


AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

Graph-based Electronic Structure Theory

Data dependency Graph G_τ

$$G_\tau \leftarrow \lfloor \text{Fermi Operator Expansion} \rfloor_{\tau(\text{global})} = D_\tau$$

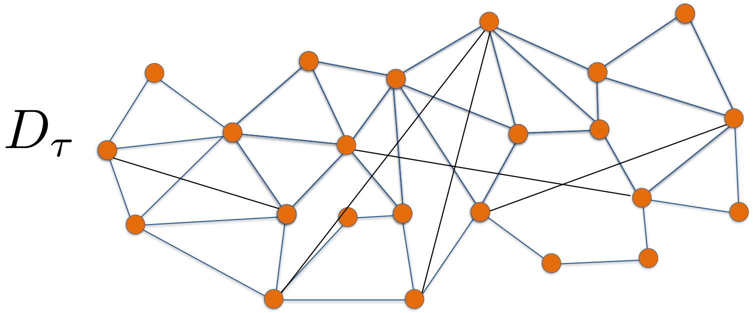


AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

Graph-based Electronic Structure Theory

Data dependency Graph G_τ

$$G_\tau \leftarrow \lfloor \text{Fermi Operator Expansion} \rfloor_{\tau(\text{global})} = D_\tau$$

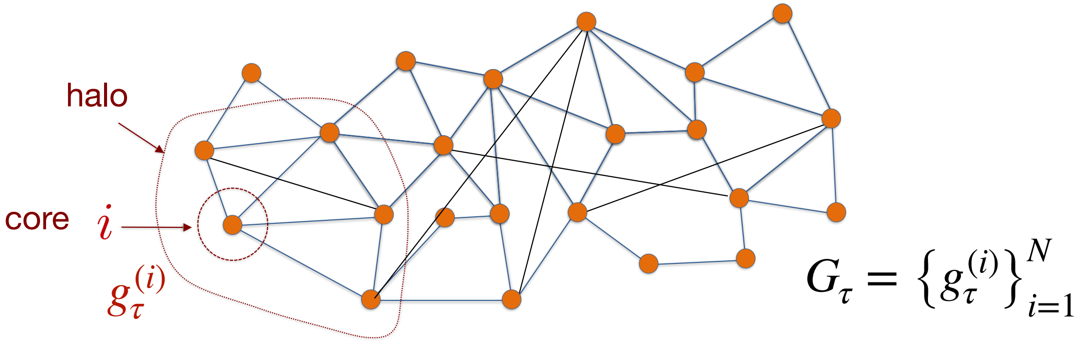


AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

Graph-based Electronic Structure Theory

Data dependency Graph G_τ

$$G_\tau \leftarrow \lfloor \text{Fermi Operator Expansion} \rfloor_{\tau(\text{global})} = D_\tau$$

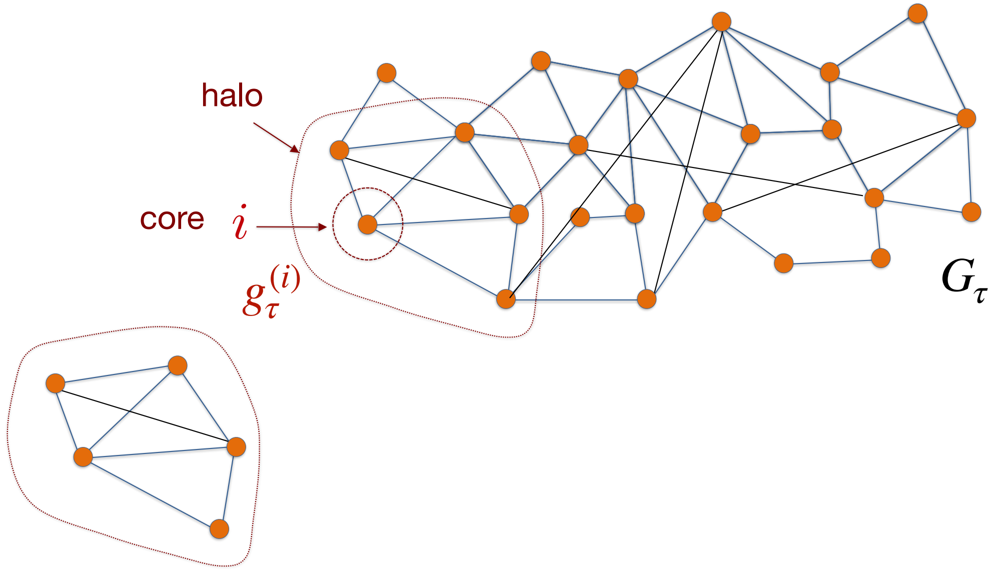


AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

Graph-based Electronic Structure Theory

Data dependency Graph G_τ

$$G_\tau \leftarrow \lfloor \text{Fermi Operator Expansion} \rfloor_{\tau(\text{global})} = D_\tau$$



$$G_\tau = \left\{ g_\tau^{(i)} \right\}_{i=1}^{N}$$

$$H = \left\{ h[g_\tau^{(i)}] \right\}_{i=1}^{N}$$ Principal submatrices of H

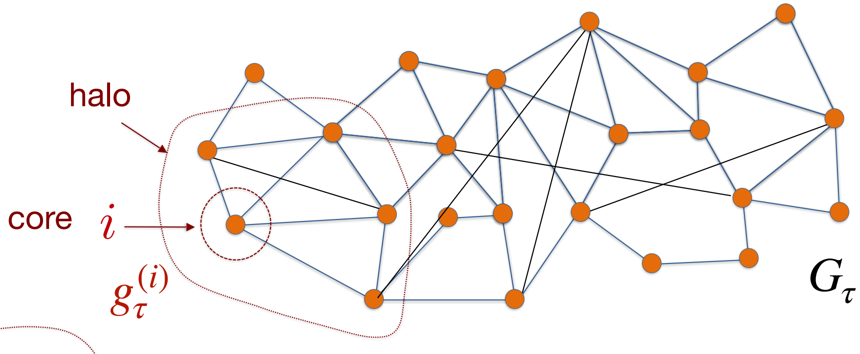
Los Alamos
NATIONAL LABORATORY
EST. 1943

AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

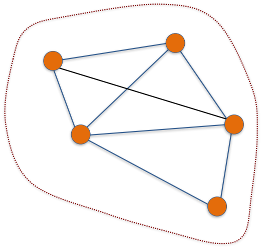
Graph-based Electronic Structure Theory

Data dependency Graph G_τ

$$G_\tau \leftarrow \lfloor \text{Fermi Operator Expansion} \rfloor_{\tau(\text{global})} = D_\tau$$



$$G_\tau = \left\{ g_\tau^{(i)} \right\}_{i=1}^{N}$$



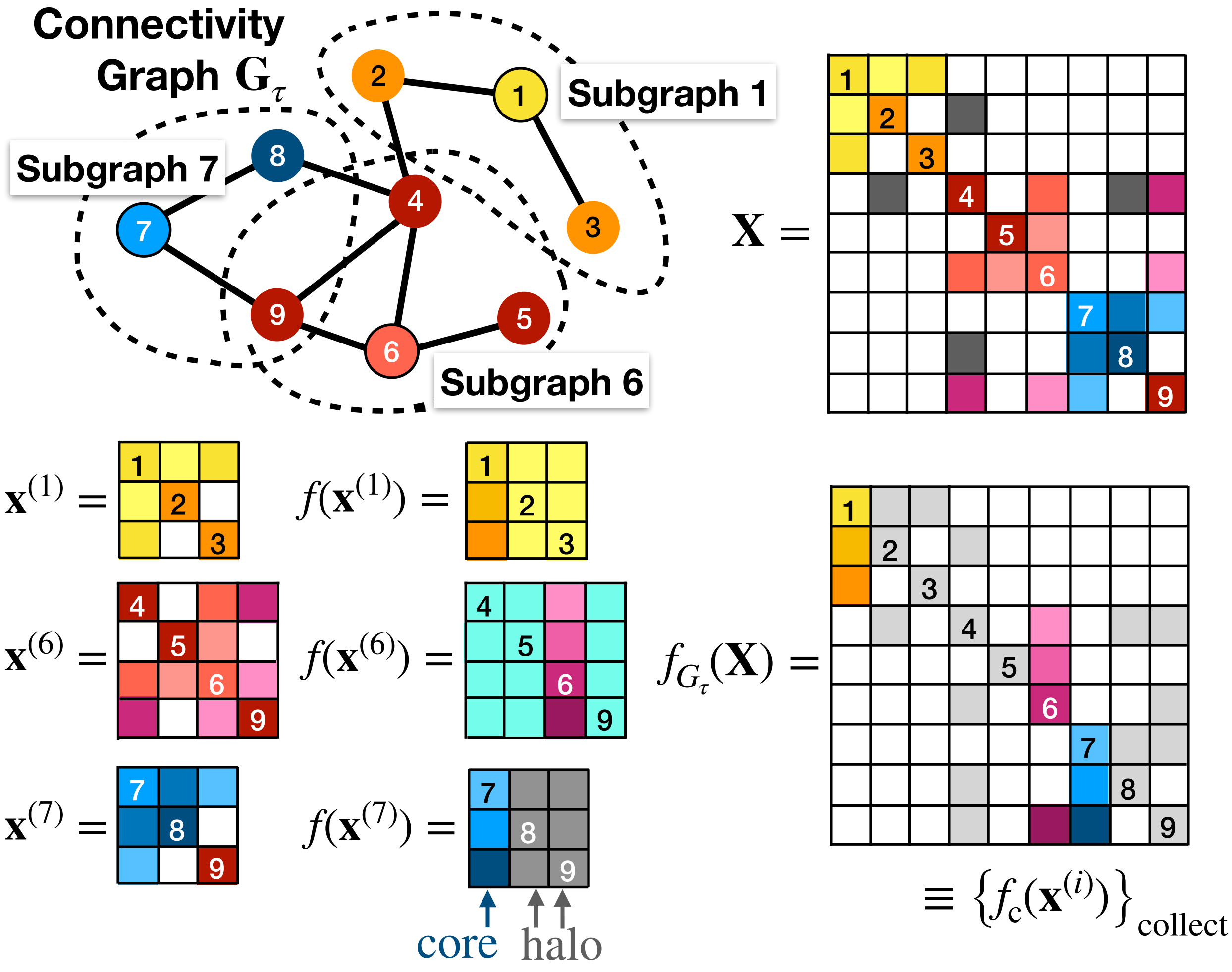
$$H = \left\{ h[g_\tau^{(i)}] \right\}_{i=1}^{N}$$

Recursive Fermi-operator expansion

$$D_\tau = \left\{ \lim_{n \to \infty} f_n(f_{n-1}(\ \dots \ f_0(h[g_\tau^{(i)}]) \ \dots \ )) \right\}_{i=1}^{N}$$

Equivalence between a numerically thresholded Fermi-operator expansion and a partitioned subgraph expansion!

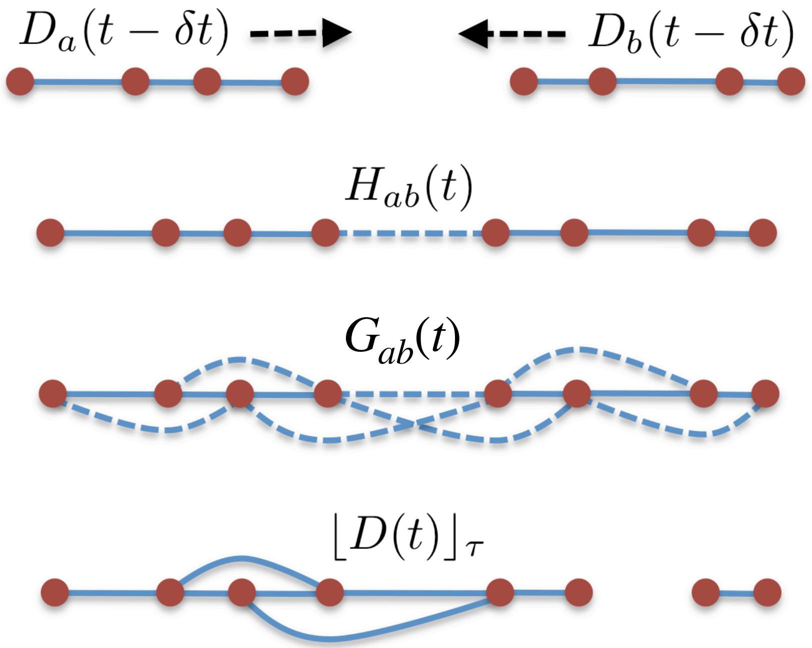
AMNN et al. *“Graph-based linear scaling electronic structure theory”*, J. Chem. Phys. 144, 234101 (2016)



Los Alamos
NATIONAL LABORATORY
EST. 1943

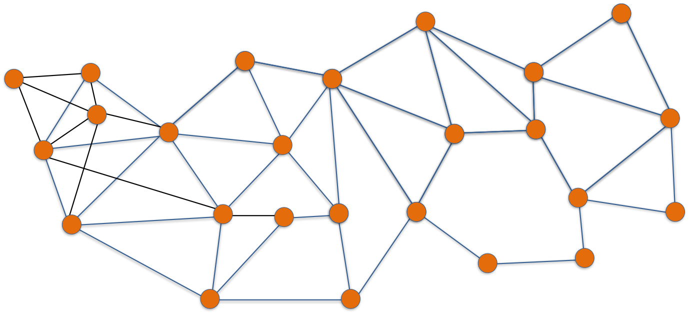
Graph-based Electronic Structure Theory

$$G_\tau(t) \leftarrow \left(\lfloor D(t - \delta t) \rfloor_\tau + H(t) \right)^2$$



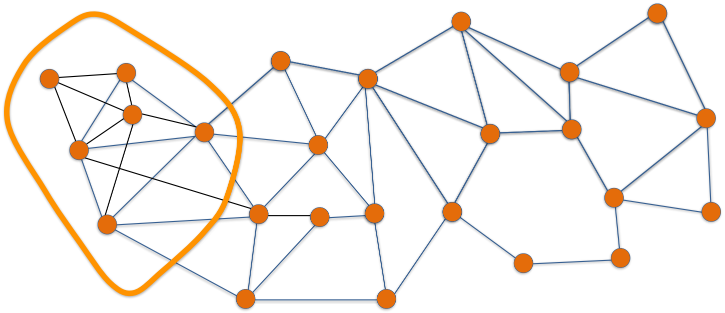
AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

Graph-based Electronic Structure Theory

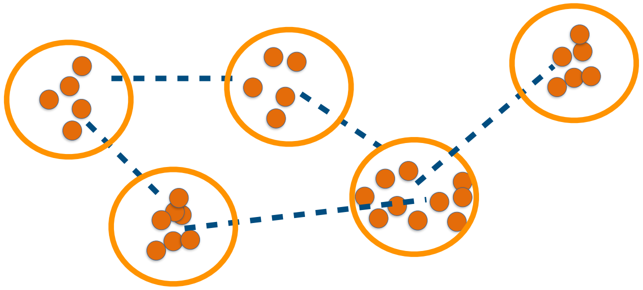


Reduce overhead (redundant overlap) by finding communities using off-the-shelf graph-partitioning schemes

Graph-based Electronic Structure Theory

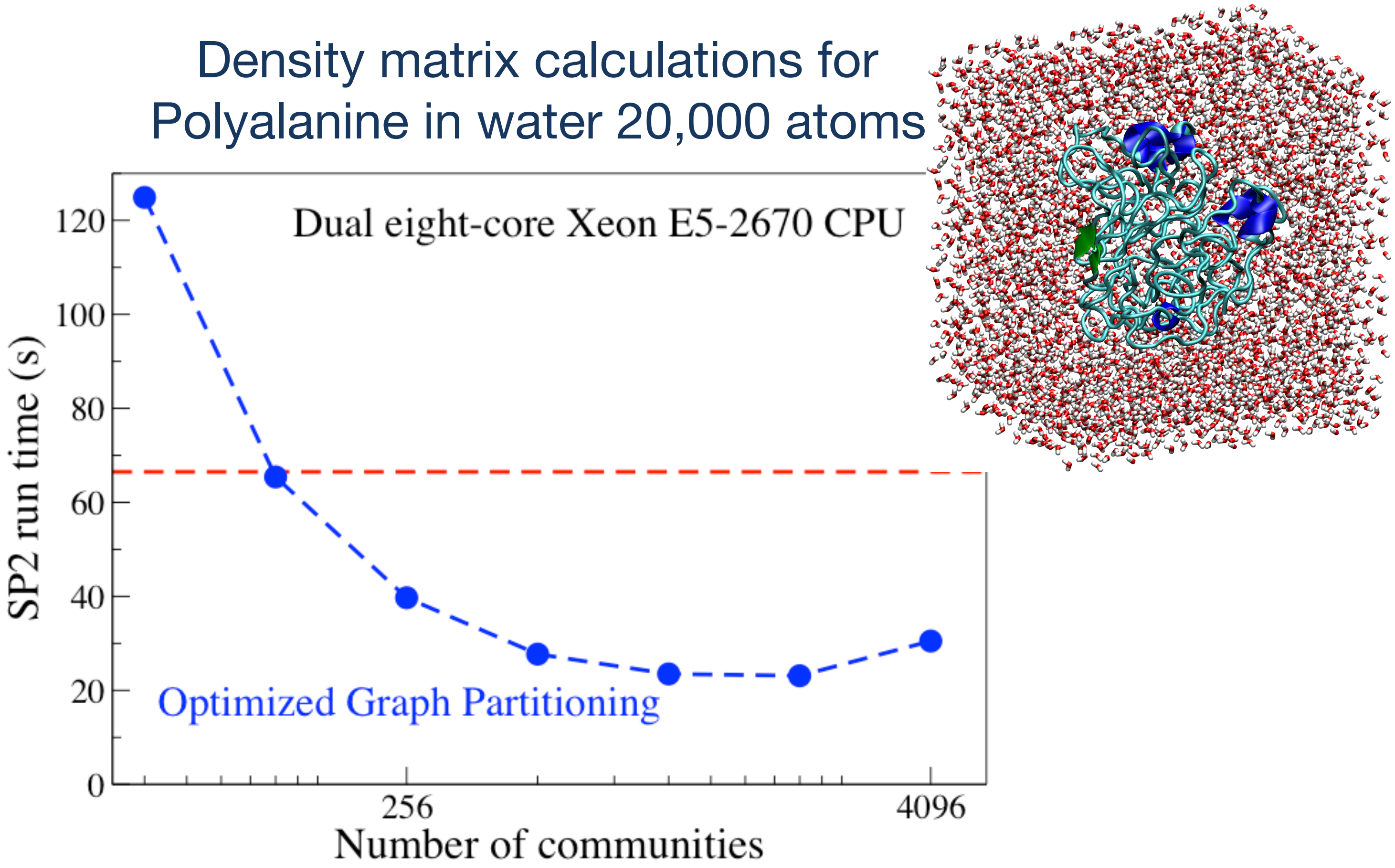


Reduce overhead (redundant overlap) by finding communities using off-the-shelf graph-partitioning schemes



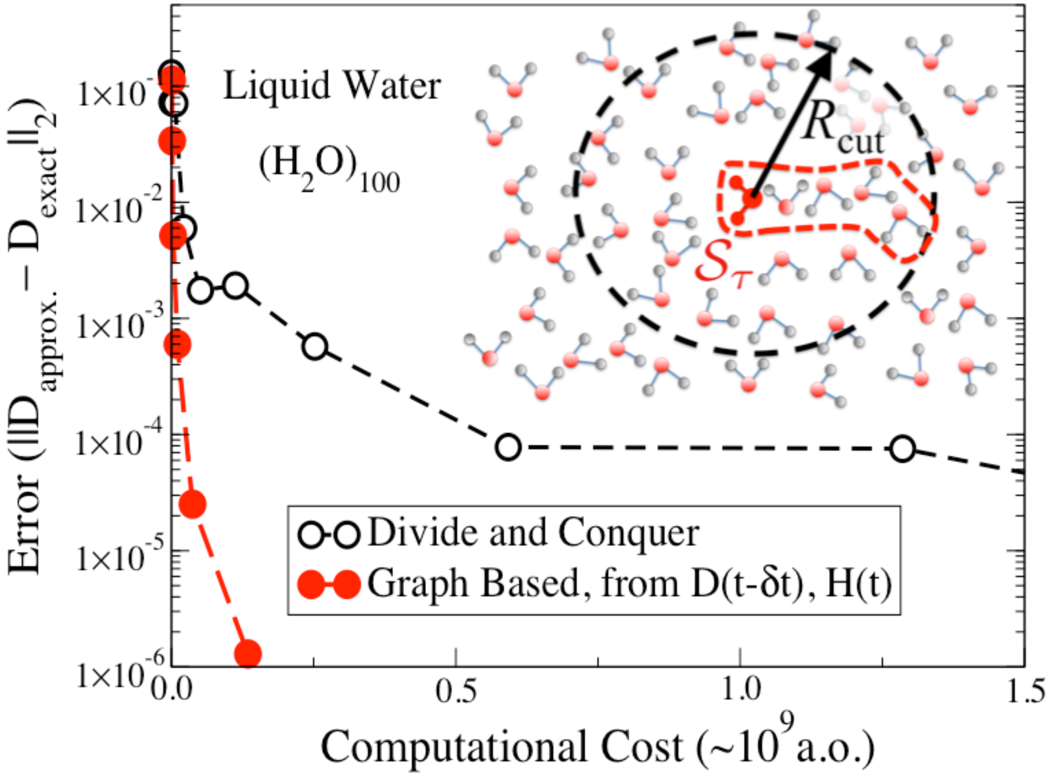
Graph-based Electronic Structure Theory

Density matrix calculations for Polyalanine in water 20,000 atoms



Graph-based Electronic Structure Theory

Density matrix calculations
from snapshot of water simulation



AMNN et al. *"Graph-based linear scaling electronic structure theory"*, J. Chem. Phys. 144, 234101 (2016)

Electronic Structure Methods for Molecular Dynamics Simulations

$$U(\mathbf{R}) = \min_{\rho \in \text{constr.}} \left\{ \Omega\left[\mathbf{R}, \rho(\mathbf{r})\right] + \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \,\middle|\, \int \rho(\mathbf{r}) d\mathbf{r} = N_e \right\} + V(\mathbf{R})$$

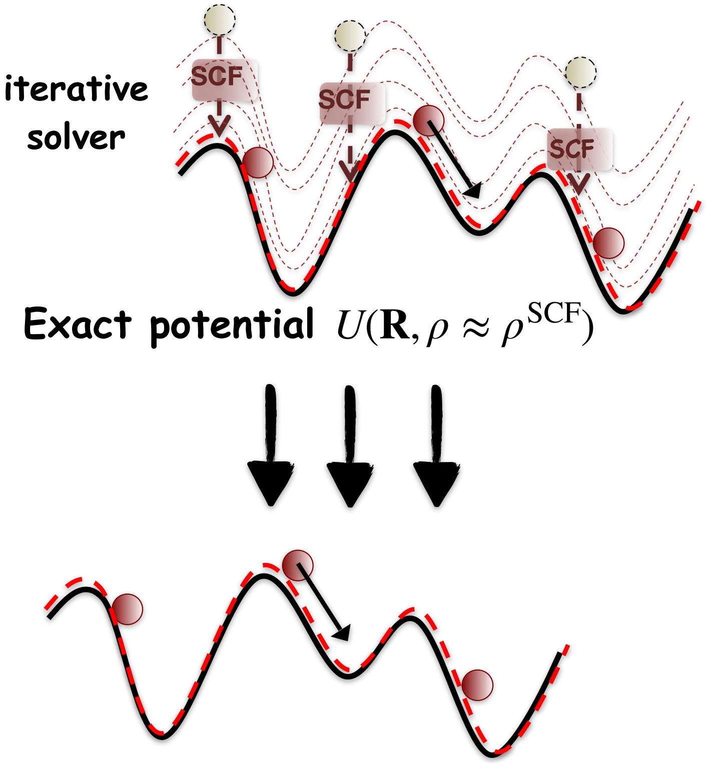
$$M_I \ddot{\mathbf{R}}_I = -\nabla_I U(\mathbf{R})$$

$$\left\{ U(\mathbf{R}),\ \nabla_I U(\mathbf{R}) \right\}$$

Non-linearities → Numerical sensitivities → ~~O(N)~~ ~~Parallelism~~

1) Iterative/sequential solver (non-parallelizable and often hard to converge)
2) Inconsistencies between potential and forces (irreversible with systematic energy drift)
3) Numerically sensitive (linear scaling approaches are not possible, -> N^3 scaling)
4) Hard to distribute calculations without major errors (parallelism is very limited)

Backward error analysis
or shadow molecular dynamics

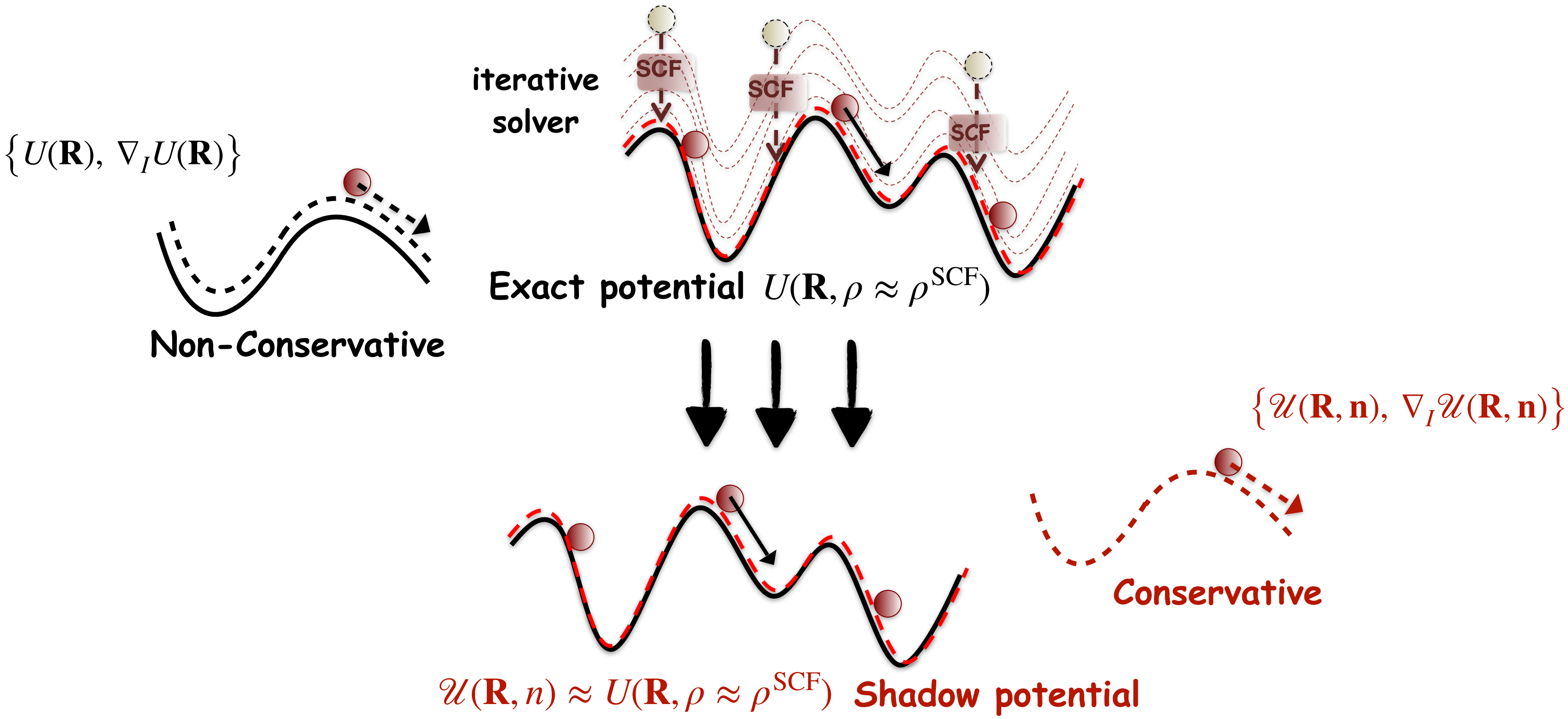


iterative solver

Exact potential $U(\mathbf{R}, \rho \approx \rho^{\rm SCF})$

$\mathcal{U}(\mathbf{R}, n) \approx U(\mathbf{R}, \rho \approx \rho^{\rm SCF})$ **Shadow potential**

Backward error analysis or shadow molecular dynamics



iterative solver

$\{U(\mathbf{R}),\ \nabla_I U(\mathbf{R})\}$

Non-Conservative

Exact potential $U(\mathbf{R}, \rho \approx \rho^{\mathrm{SCF}})$

$\{\mathcal{U}(\mathbf{R}, \mathbf{n}),\ \nabla_I \mathcal{U}(\mathbf{R}, \mathbf{n})\}$

Conservative

$\mathcal{U}(\mathbf{R}, n) \approx U(\mathbf{R}, \rho \approx \rho^{\mathrm{SCF}})$ Shadow potential

Los Alamos
NATIONAL LABORATORY
EST. 1943

DFT-based Born-Oppenheimer MD
or other SCF based models

$$\rho^{\mathrm{SCF}}(\mathbf{R},\mathbf{r}) = \arg\min_{\rho\in\mathcal{N}} \left\{ F[\rho] + \int v(\mathbf{R},\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right\}$$

$$U(\mathbf{R},\rho^{\mathrm{SCF}}) = F[\rho^{\mathrm{SCF}}] + \int v(\mathbf{R},\mathbf{r})\rho^{\mathrm{SCF}}(\mathbf{r})d\mathbf{r} + V_{nn}(\mathbf{R})$$

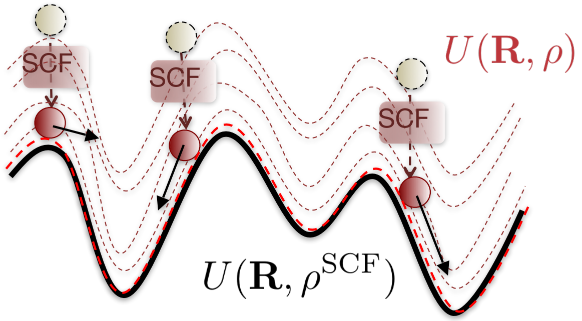
$$M_I\ddot{\mathbf{R}}_I = -\frac{\partial U(\mathbf{R},\rho^{\mathrm{SCF}})}{\partial R_I}$$

DFT-based Born-Oppenheimer MD
or other SCF based models

$$\rho^{\mathrm{SCF}}(\mathbf{R},\mathbf{r}) = \arg\min_{\rho\in\mathcal{N}} \left\{ F[\rho] + \int v(\mathbf{R},\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right\}$$

$$U(\mathbf{R},\rho^{\mathrm{SCF}}) = F[\rho^{\mathrm{SCF}}] + \int v(\mathbf{R},\mathbf{r})\rho^{\mathrm{SCF}}(\mathbf{r})d\mathbf{r} + V_{nn}(\mathbf{R})$$

$$M_I\ddot{\mathbf{R}}_I = -\frac{\partial U(\mathbf{R},\rho^{\mathrm{SCF}})}{\partial R_I}$$



Cost & Error

$$H[\rho]\Psi_i = \varepsilon_i\Psi_i$$

SCF

$$\rho = \sum_{\mathrm{occ.}} |\Psi_i|^2$$

$\rightarrow U(\mathbf{R},\rho^{\mathrm{SCF}})$

$\#\mathrm{SCF} \times \mathcal{O}(N^3)$

Non-conservative forces

DFT-based Born-Oppenheimer MD
or other SCF based models

$$\rho^{\rm SCF}(\mathbf{R},\mathbf{r}) = \arg\min_{\rho\in\mathcal{N}} \left\{ F[\rho] + \int v(\mathbf{R},\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right\}$$

$$U(\mathbf{R},\rho^{\rm SCF}) = F[\rho^{\rm SCF}] + \int v(\mathbf{R},\mathbf{r})\rho^{\rm SCF}(\mathbf{r})d\mathbf{r} + V_{nn}(\mathbf{R})$$

$$M_I\ddot{\mathbf{R}}_I = -\frac{\partial U(\mathbf{R},\rho^{\rm SCF})}{\partial R_I}$$

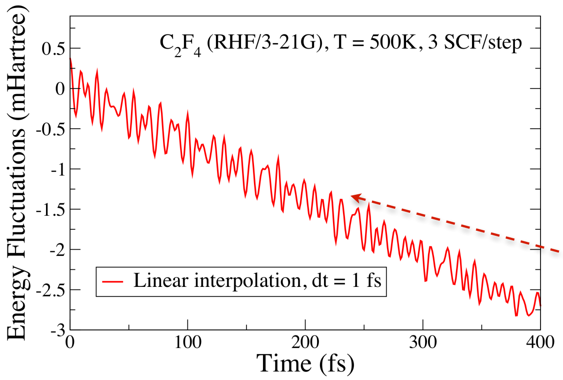
Cost & Error

$$H[\rho]\Psi_i = \varepsilon_i\Psi_i$$

SCF

$$\rho = \sum_{\rm occ.} |\Psi_i|^2 \qquad \rightarrow U(\mathbf{R},\rho^{\rm SCF})$$

$$\#{\rm SCF} \times \mathcal{O}(N^3)$$



Non-conservative forces
Energy drift!

Linearized “Shadow” Potential Energy Surface

$$\mathcal{F}[\rho, n] = F[n] + \int \left.\frac{\delta F[\rho]}{\delta \rho}\right|_{\rho=n} (\rho(\mathbf{r}) - n(\mathbf{r}))\, d\mathbf{r} = F[\rho] + \mathcal{O}\left((\rho - n)^2\right)$$

Linearized “Shadow” Potential Energy Surface

$$\mathcal{F}[\rho,n] = F[n] + \int \left.\frac{\delta F[\rho]}{\delta \rho}\right|_{\rho=n} (\rho(\mathbf{r}) - n(\mathbf{r}))\, d\mathbf{r} = F[\rho] + \mathcal{O}\left((\rho-n)^2\right)$$

$$\varrho[n](\mathbf{r}) = \arg\min_{\rho\in\mathcal{N}} \left\{ \mathcal{F}[\rho,n] + \int v(\mathbf{R},\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right\}$$

$$\mathcal{U}(\mathbf{R},n) = \mathcal{F}[\varrho,n] + \int v(\mathbf{R},\mathbf{r})\varrho(\mathbf{r})d\mathbf{r} + V_{nn}(\mathbf{R})$$ **Shadow BO potential**

Linearized “Shadow” Potential Energy Surface

$$\mathcal{F}[\rho,n] = F[n] + \int \left.\frac{\delta F[\rho]}{\delta \rho}\right|_{\rho=n} (\rho(\mathbf{r}) - n(\mathbf{r}))\, d\mathbf{r} = F[\rho] + \mathcal{O}\left((\rho-n)^2\right)$$

$$\varrho[n](\mathbf{r}) = \arg\min_{\rho\in\mathcal{N}} \left\{ \mathcal{F}[\rho,n] + \int v(\mathbf{R},\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right\}$$

$$\mathcal{U}(\mathbf{R},n) = \mathcal{F}[\varrho,n] + \int v(\mathbf{R},\mathbf{r})\varrho(\mathbf{r})d\mathbf{r} + V_{nn}(\mathbf{R})$$

Small Cost, No SCF

$$H[n]\Psi_i = \varepsilon_i \Psi_i$$

$$\varrho[n] = \sum_{\text{occ.}} |\Psi_i|^2$$

Error in the potential

$$\mathcal{U}(\mathbf{R},n) \approx U(\mathbf{R}) + \mathcal{O}\left((\varrho[n]-n)^2\right) = U(\mathbf{R}) + \mathcal{O}\left((\Delta\rho^{\mathrm{SCF}})^2\right)$$

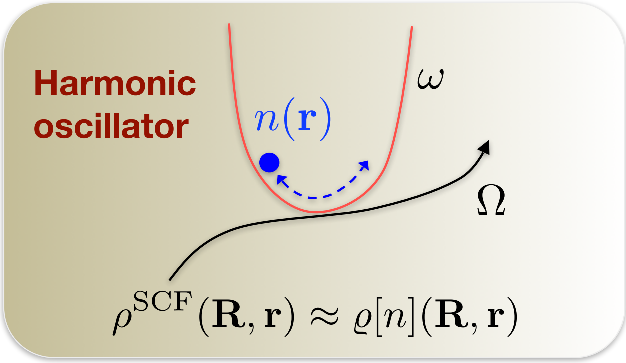
SCF residual $= \varrho[n] - n$

Extended Lagrangian Born-Oppenheimer MD

Niklasson & Cawkwell, JCP 141, 164123 (2014); Niklasson JCP 147, 054103 (2017)

$$\mathcal{L}(\mathbf{R},\dot{\mathbf{R}},n,\dot{n}) = \frac{1}{2}\sum_I M_I \dot{R}_I^2 - \mathcal{U}(\mathbf{R},n) \qquad \text{(Shadow potential)}$$

$$+\frac{\mu}{2}\int \dot{n}^2(\mathbf{r})d\mathbf{r} - \frac{\mu\omega^2}{2}\iint \left(\varrho[n](\mathbf{r}) - n(\mathbf{r})\right) T(\mathbf{r},\mathbf{r}') \left(\varrho[n](\mathbf{r}') - n(\mathbf{r}')\right) d\mathbf{r}d\mathbf{r}'$$



$$\rho^{\rm SCF}(\mathbf{R},\mathbf{r}) \approx \varrho[n](\mathbf{R},\mathbf{r})$$

Next Generation Extended Lagrangian First Principles Molecular Dynamics

Extended Lagrangian

$\mathcal{U}(\mathbf{R}, n)$: Shadow Potential, $T = K^\dagger K$

$$\mathcal{L}(\mathbf{R}, \dot{\mathbf{R}}, n, \dot{n}) = \frac{1}{2}\sum_I M_I \dot{R}_I^2 - \mathcal{U}(\mathbf{R}, n)$$

$$+\frac{\mu}{2}\int \dot{n}^2(\mathbf{r})d\mathbf{r} - \frac{\mu\omega^2}{2}\iint \left(\varrho[n](\mathbf{r}) - n(\mathbf{r})\right) T(\mathbf{r}, \mathbf{r}') \left(\varrho[n](\mathbf{r}') - n(\mathbf{r}')\right) d\mathbf{r}d\mathbf{r}'$$

Equations of Motion in Adiabatic Limit $\begin{pmatrix} \omega \to \infty \\ \mu \to 0 \end{pmatrix}$

$$M_I \ddot{R}_I = -\left.\frac{\partial \mathcal{U}(\mathbf{R}, n)}{\partial R_I}\right|_n$$

$$\ddot{n}(\mathbf{r}) = -\omega^2 \int K(\mathbf{r}, \mathbf{r}') \left(\varrho[n](\mathbf{r}') - n(\mathbf{r}') \, d\mathbf{r}'$$

Kernel

$$K(\mathbf{r}, \mathbf{r}') = -\left(\frac{\delta\varrho[n](\mathbf{r})}{\delta n(\mathbf{r}')} - \frac{\delta n(\mathbf{r})}{\delta n(\mathbf{r}')}\right)^{-1}$$

Low-rank Approximation

$$\mathbf{K} = (\mathbf{K}_0\mathbf{J})^{-1}\mathbf{K}_0 \approx - c\delta(\mathbf{r} - \mathbf{r}')$$

$$c \in [0,1]$$

$$(\mathbf{K}_0\mathbf{J})^{-1} \approx \sum_{i,j=1}^{m} \mathbf{v}_i M_{i,j} \mathbf{f}_{\mathbf{v}_i}$$

$$\mathbf{M} = \mathbf{O}^{-1}, \; O_{i,j} = \mathbf{f}_{\mathbf{v}_i}^T \mathbf{f}_{\mathbf{v}_j}$$

$$\mathbf{f}(\mathbf{n}) = \mathbf{q}[\mathbf{n}] - \mathbf{n} \qquad \mathbf{v_i}, \mathbf{f}_{\mathbf{v_i}} \in \mathcal{K}^\perp$$

$$\mathbf{f}_\mathbf{v} \equiv \left.\frac{\partial \mathbf{f}(\mathbf{n} - \lambda\mathbf{v})}{\partial\lambda}\right|_{\lambda=0} = \mathbf{J}\mathbf{v}$$

Los Alamos
NATIONAL LABORATORY
EST. 1943

Electronic Structure Methods for Molecular Dynamics Simulations

$$U(\mathbf{R}) = \min_{\rho \in \text{constr.}} \left\{ \Omega\left[\mathbf{R}, \rho(\mathbf{r})\right] + \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \,\middle|\, \int \rho(\mathbf{r}) d\mathbf{r} = N_e \right\} + V(\mathbf{R})$$

$$M_I \ddot{\mathbf{R}}_I = - \nabla_I U(\mathbf{R})$$

$$\left\{ U(\mathbf{R}),\ \nabla_I U(\mathbf{R}) \right\}$$

Non-linearities → Numerical sensitivities → ~~O(N)~~

~~Parallelism~~

Los Alamos
NATIONAL LABORATORY
EST. 1943

Electronic Structure Methods for Molecular Dynamics Simulations

$$U(\mathbf{R}) = \min_{\rho \in \text{constr.}} \left\{ \Omega\left[\mathbf{R}, \rho(\mathbf{r})\right] + \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \,\middle|\, \int \rho(\mathbf{r})d\mathbf{r} = N_e \right\} + V(\mathbf{R})$$

$$M_I \ddot{\mathbf{R}}_I = -\nabla_I U(\mathbf{R})$$

$$\left\{U(\mathbf{R}),\ \nabla_I U(\mathbf{R})\right\}$$

~~Non-linearities~~ → ~~Numerical sensitivities~~ → O(N)

Parallelism

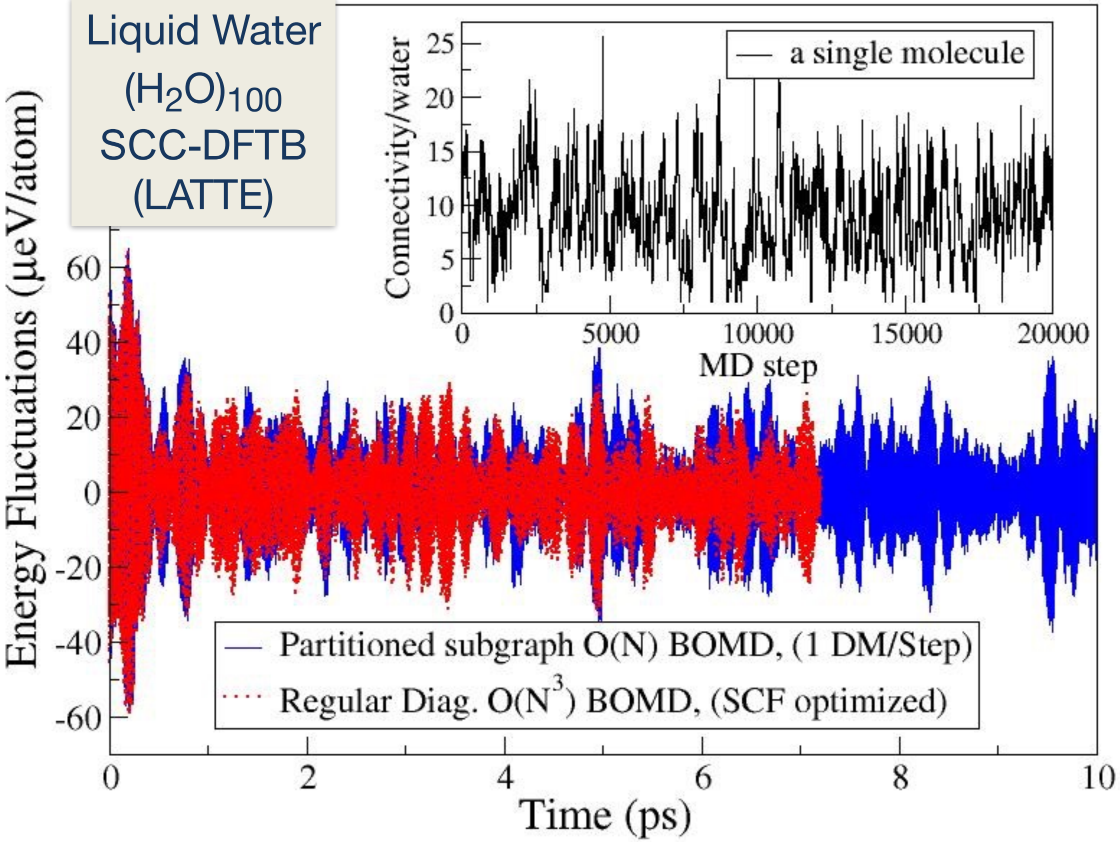
Shadow Molecular Dynamics

$$\left\{\mathscr{U}(\mathbf{R},\mathbf{n}),\ \nabla_I \mathscr{U}(\mathbf{R},\mathbf{n})\right\}$$

Graph-based parallelism
is possible!

Graph-based XL Born-Oppenheimer Molecular Dynamics

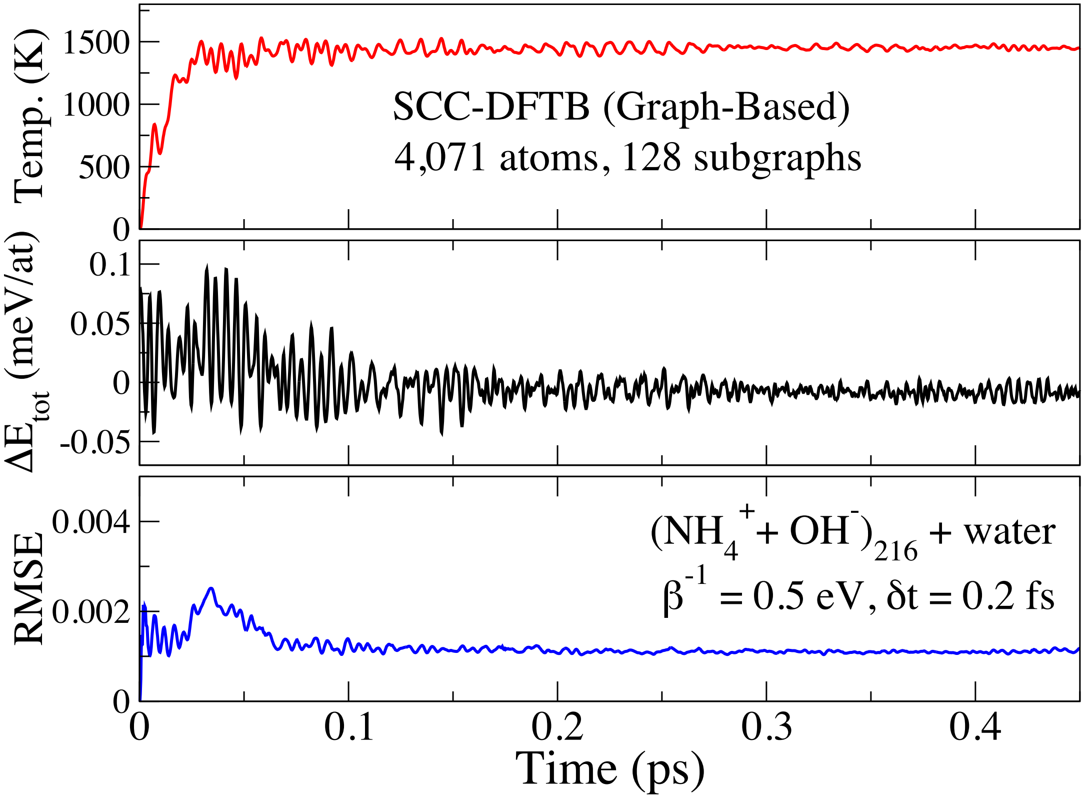
Liquid Water
$(H_2O)_{100}$
SCC-DFTB
(LATTE)



AMNN “Next Generation Extended Lagrangian First Principles Molecular Dyabnamics" J. Chem. Phys. **147**, 054103 (2017)

AMNN et al. *“Graph-based linear scaling electronic structure theory”*, J. Chem. Phys. 144, 234101 (2016)

Graph-based XL Born-Oppenheimer Molecular Dynamics



Low-rank Approximation

$$\mathbf{K} = (\mathbf{K}_0\mathbf{J})^{-1}\mathbf{K}_0 \approx -c\delta(\mathbf{r}-\mathbf{r}')$$

$$c \in [0,1]$$

$$(\mathbf{K}_0\mathbf{J})^{-1} \approx \sum_{i,j=1}^{m} \mathbf{v}_i M_{i,j} \mathbf{f}_{\mathbf{v}_i}$$

$$\mathbf{M} = \mathbf{O}^{-1}, \; O_{i,j} = \mathbf{f}_{\mathbf{v}_i}^T \mathbf{f}_{\mathbf{v}_j}$$

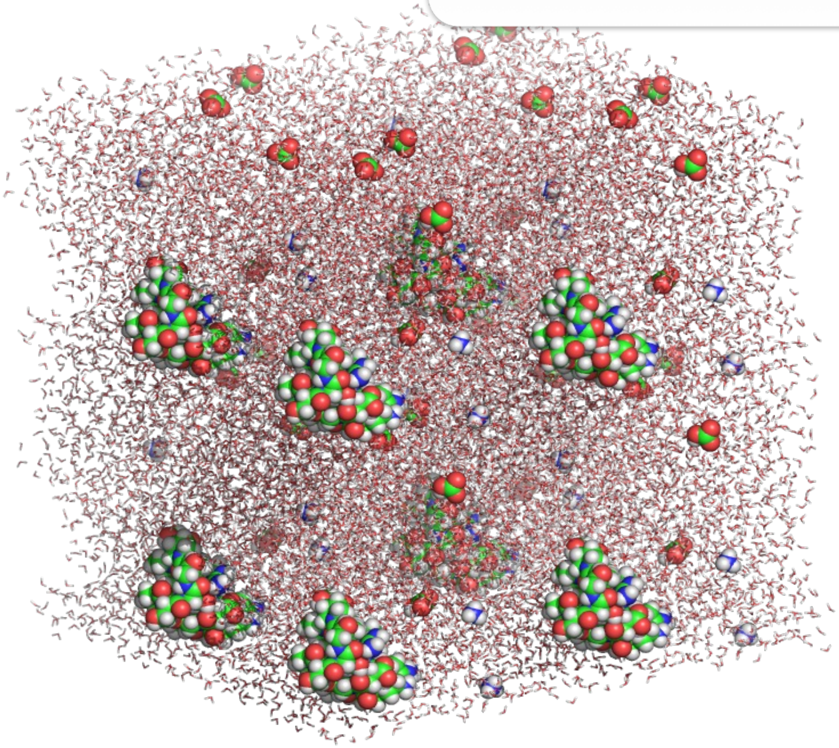
$$\mathbf{f}(\mathbf{n}) = \mathbf{q}[\mathbf{n}] - \mathbf{n} \qquad \mathbf{v_i}, \mathbf{f}_{\mathbf{v_i}} \in \mathcal{K}^{\perp}$$

$$\mathbf{f}_{\mathbf{v}} \equiv \left.\frac{\partial \mathbf{f}(\mathbf{n} - \lambda\mathbf{v})}{\partial \lambda}\right|_{\lambda=0} = \mathbf{J}\mathbf{v}$$

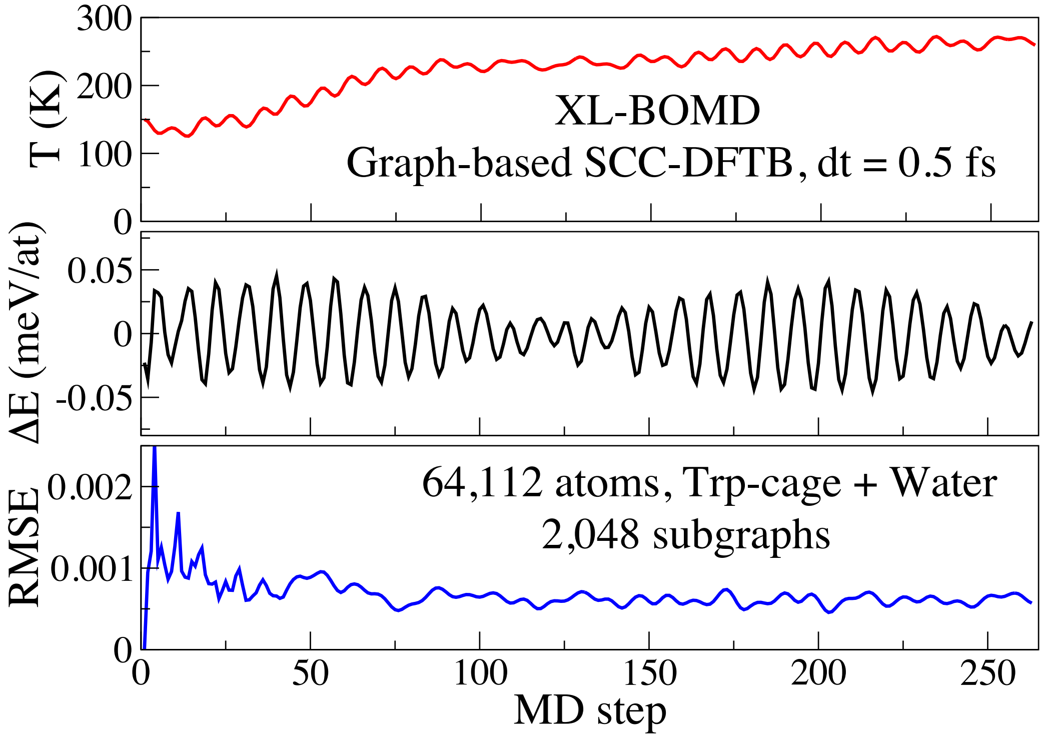
Los Alamos
NATIONAL LABORATORY
EST. 1943

Negre, Wall & Niklasson, J. Chem. Phys. **158**, 074108 (2023)

Graph-based XL Born-Oppenheimer Molecular Dynamics



Trp-cage in ammonium bicarbonate solution
64,112 atoms



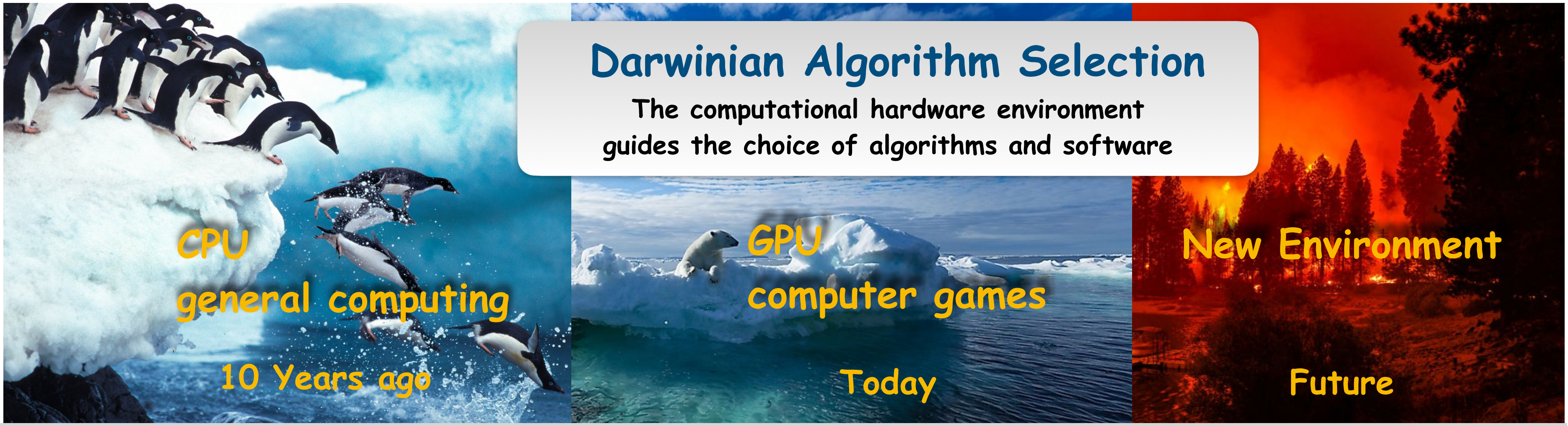
Running on 32 CPUs

Los Alamos NATIONAL LABORATORY EST. 1943

Negre, Wall & Niklasson, J. Chem. Phys. **158**, 074108 (2023)

Darwinian Algorithm Selection

The computational hardware environment
guides the choice of algorithms and software



~10 Years ago
Moving some general scientific
computing tasks from CPU's to GPU's

New opportunity!

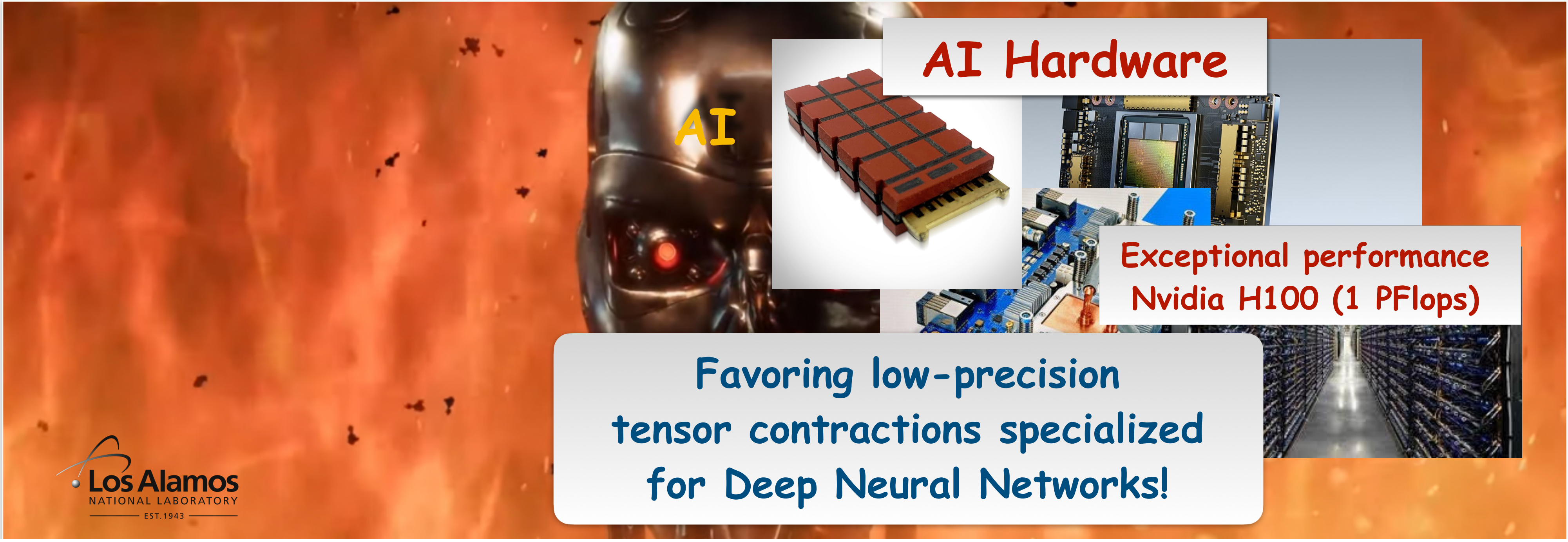
Darwinian Algorithm Selection

The computational hardware environment
guides the choice of algorithms and software

CPU
general computing
10 Years ago

GPU
computer games
Today

New Environment
Future



AI

AI Hardware

Exceptional performance
Nvidia H100 (1 PFlops)

Favoring low-precision
tensor contractions specialized
for Deep Neural Networks!

Los Alamos
NATIONAL LABORATORY
EST. 1943

Darwinian Algorithm Selection

The computational hardware environment guides the choice of algorithms and software

CPU

general computing

10 Years ago

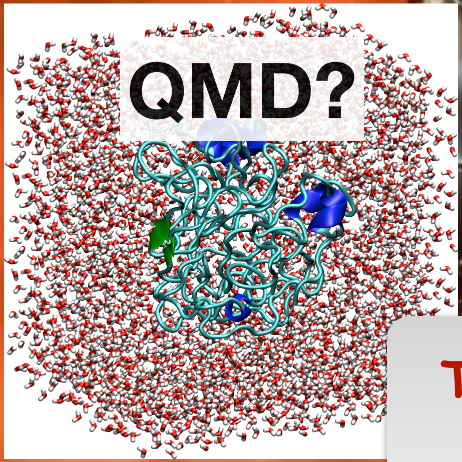
GPU

computer games

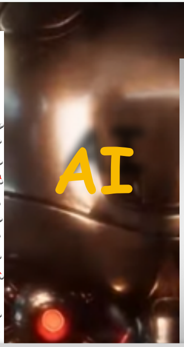
Today

New Environment

Future



QMD?

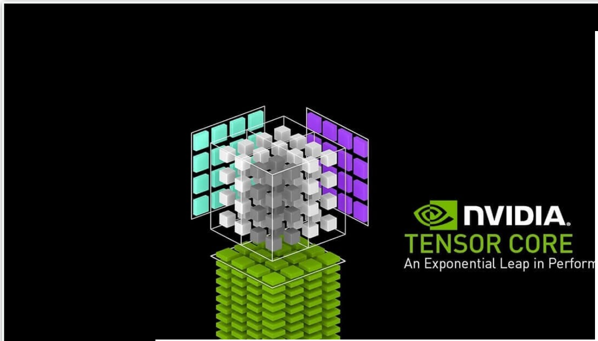


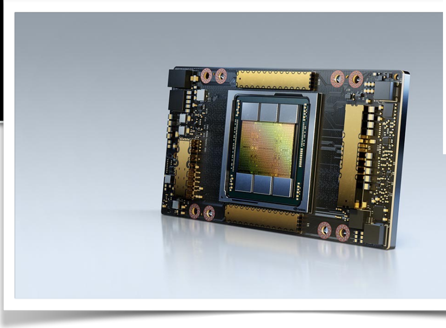
AI

AI Hardware

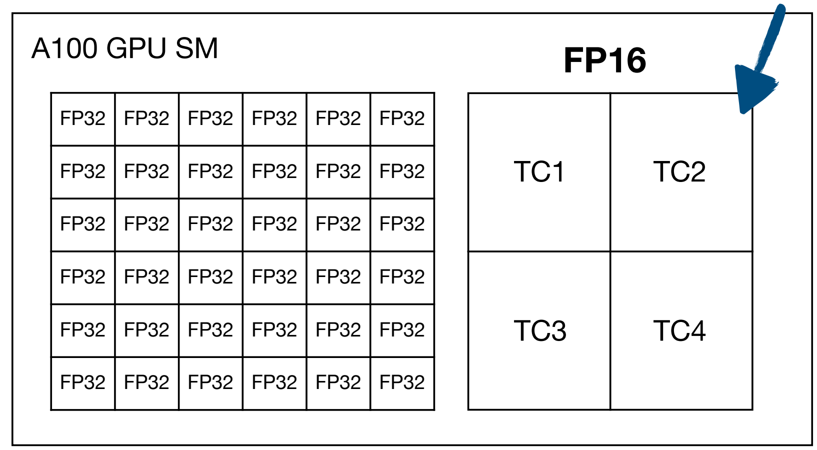
To stay competitive, how do we use AI hardware for more general scientific computing tasks such as QMD?

What is AI hardware?
Tensor cores, by Nvidia





The Tensor cores!


A100 GPU SM

FP32	FP32	FP32	FP32	FP32	FP32
FP32	FP32	FP32	FP32	FP32	FP32
FP32	FP32	FP32	FP32	FP32	FP32
FP32	FP32	FP32	FP32	FP32	FP32
FP32	FP32	FP32	FP32	FP32	FP32
FP32	FP32	FP32	FP32	FP32	FP32

FP16

TC1	TC2
TC3	TC4

Half-precision matrix multiplications
with single-precision accumulation

$$\mathbf{X} = \alpha\mathbf{A} \times \mathbf{B} + \beta\mathbf{C}$$

Ideal for convolutional deep neural networks!

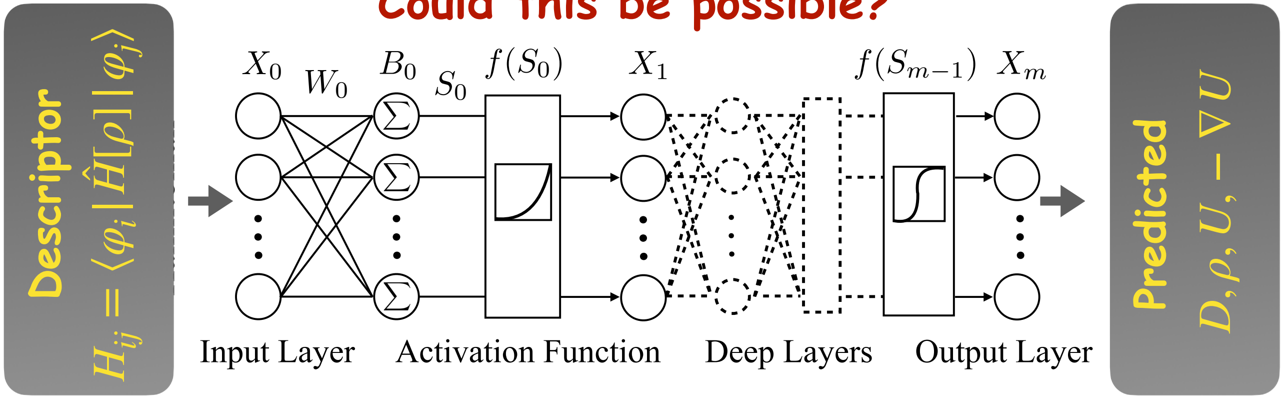
$$A = f_m(\ldots f(W_2 f(W_1 f(W_0 X_0 + B_0) + B_1) + B_2)\ldots)$$

Hundreds of TFlops on a single set of Tensor cores!
(1 PFlops on the latest H100)

Los Alamos
NATIONAL LABORATORY
EST. 1943

Quantum-based Molecular Dynamics

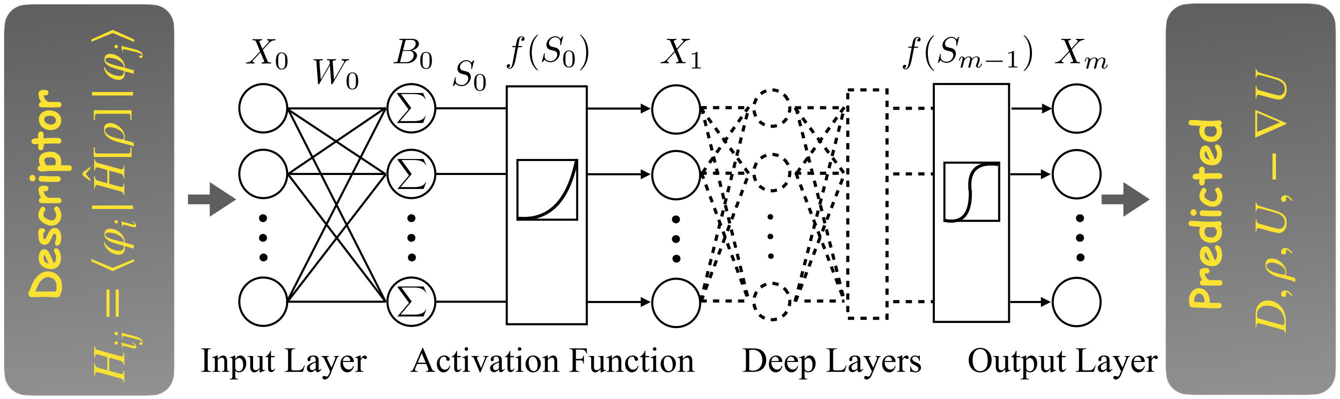
Could this be possible?



$$\rho = f_m(\ldots f(W_2 f(W_1 f(W_0 X_0 + B_0) + B_1) + B_2)\ldots)$$

Tensor operations in half precision!

SP2: Niklasson, PRB 66, 155115 (2002)



Descriptor

$H_{ij} = \langle \varphi_i | \hat{H}[\rho] | \varphi_j \rangle$

X_0 W_0 B_0 S_0 $f(S_0)$ X_1 $f(S_{m-1})$ X_m

Input Layer

Activation Function

Deep Layers

Output Layer

Predicted

$D, \rho, U, -\nabla U$

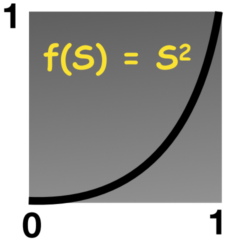
Deep-NN SP2 (2nd-order spectral projections)

$$D = \lim_{n\to\infty} f(\ldots f(W_2 f(W_1 f(W_0 X_0 + B_0) + B_1) + B_2)\ldots) \quad \approx \Theta(\mu I - H)$$

$$W_0 = -I(\varepsilon_{\max} - \varepsilon_{\min})^{-1}, \;\; B_0 = \varepsilon_n I(\varepsilon_{\max} - \varepsilon_{\min})^{-1}$$ **Initialization**

$$S_n = W_n X_n + B_n, \;\; W_n = \pm 1, \;\; B_n = (1 - W_n) S_{n-1}, \;\; n > 0$$

$$X_n = f(S_{n-1}) = S_{n-1}^2$$ **Tensor operation S² in activation function**



Mixed Precision Fermi-Operator Expansion on Tensor Cores from a Machine Learning Perspective,
Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC, 17 2256 (2021);
SP2: Niklasson, PRB 66, 155115 (2002)

Quantum-based Molecular Dynamics

What do we do about the low-precision?

$$\begin{tikzcd}
 & a = 0.123456789 \arrow[dl] \arrow[dr] & \\
a_{\text{high}} = 0.123000 & & a_{\text{low}} = 0.000456
\end{tikzcd}$$

$$a \approx a_{\text{high}} + a_{\text{low}}$$

Quantum-based Molecular Dynamics

What do we do about the low-precision?

$$X_n = f(S_{n-1}) = S_{n-1}^2 \quad \text{Activation function}$$

Double mixed precision matrix representation

$$S \approx (S_{\text{high}} + S_{\text{low}}) \qquad \begin{aligned} S_{\text{high}} &= \text{FP16}[S] \\ S_{\text{low}} &= \text{FP16}[S - S_{\text{high}}] \end{aligned}$$

$$X = f(S) = S^2 \approx \left(S_{\text{high}} + S_{\text{low}}\right)^2 = S_{\text{high}}^2 + S_{\text{high}}S_{\text{low}} + S_{\text{low}}S_{\text{high}} + S_{\text{low}}^2$$

$$\approx S_{\text{high}}^2 + S_{\text{high}}S_{\text{low}} + \left(S_{\text{high}}S_{\text{low}}\right)^{\text{T}}$$

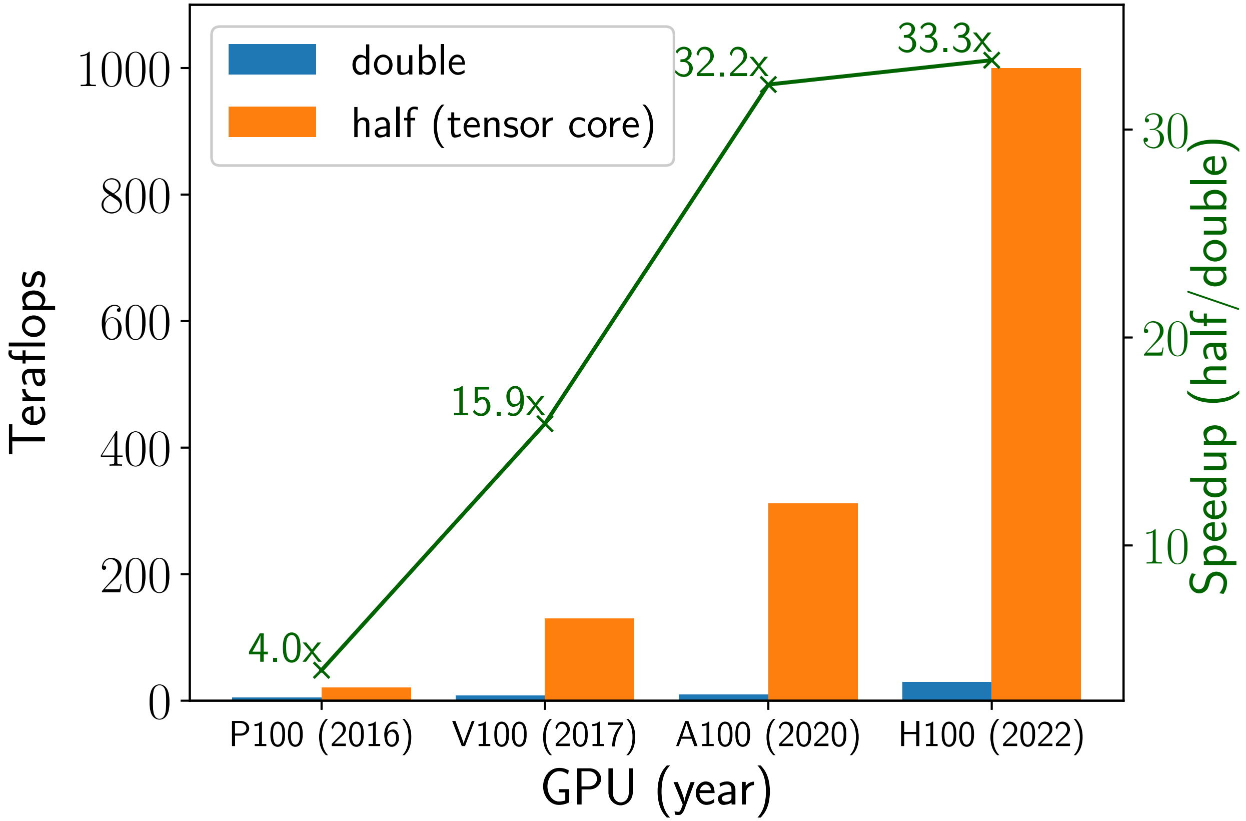
Only Two Matrix Multiplications!

Thanks to Simplicity and Symmetry

Doubles the cost, but
Speed-up x 32
Compared to double precision!

Why half precision?

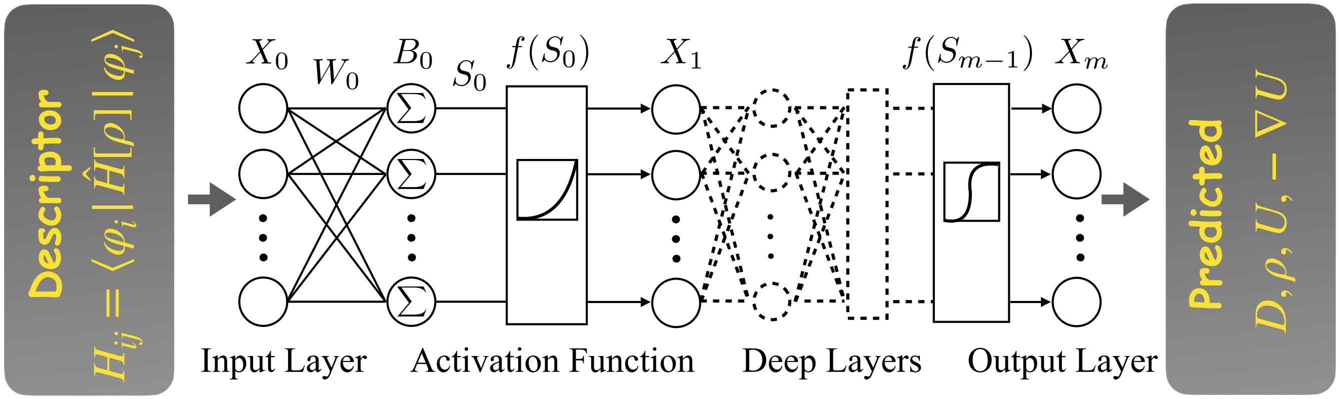
Development of arithmetic performance

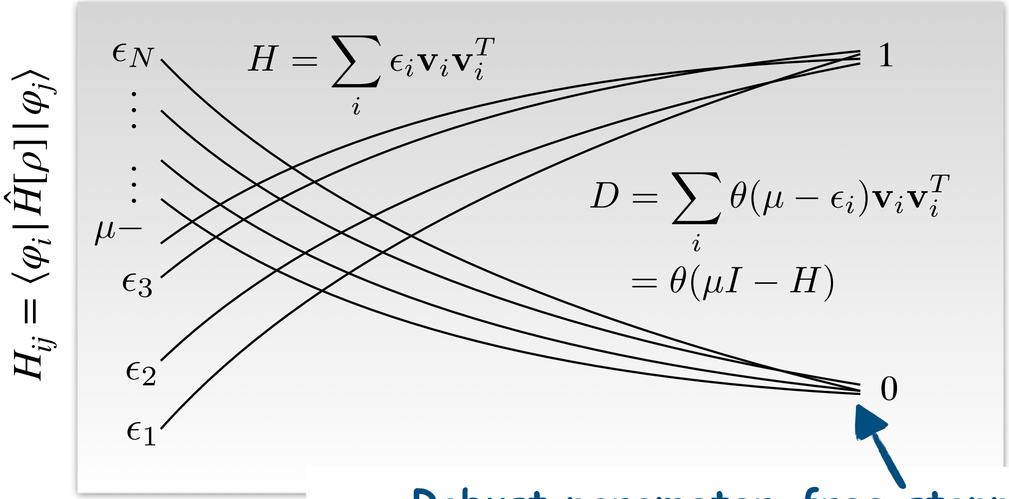


Los Alamos
NATIONAL LABORATORY
EST. 1943

SP2: Niklasson, PRB 66, 155115 (2002)

Deep-NN SP2




$H_{ij} = \langle \varphi_i | \hat{H}[\rho] | \varphi_j \rangle$
ϵ_N
$\mu -$
ϵ_3
ϵ_2
ϵ_1
1
0

$$H = \sum_i \epsilon_i \mathbf{v}_i \mathbf{v}_i^T$$

$$D = \sum_i \theta(\mu - \epsilon_i) \mathbf{v}_i \mathbf{v}_i^T$$
$$= \theta(\mu I - H)$$

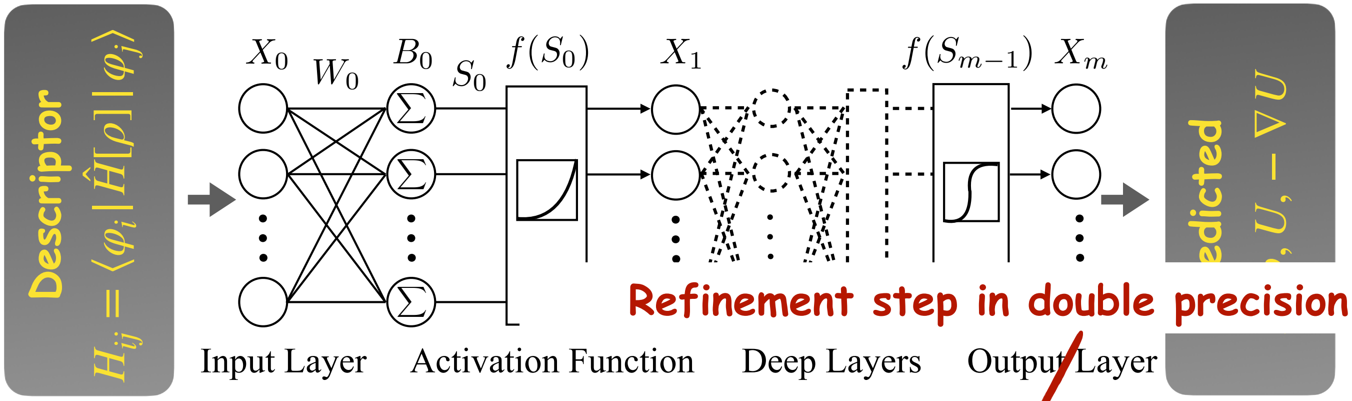
$$\sum_{ij} D_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r})$$
$$= \rho(\mathbf{r})$$

$$E_{\text{band}} = \text{Tr}\,[HD]$$

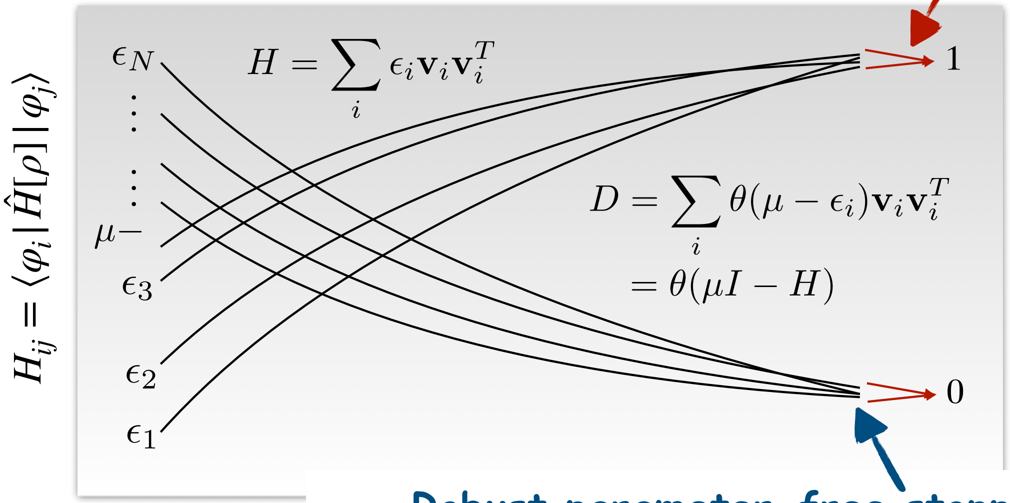
Robust parameter-free stopping criterion

SP2: Niklasson, PRB 66, 155115 (2002)

Deep-NN SP2



Refinement step in double precision



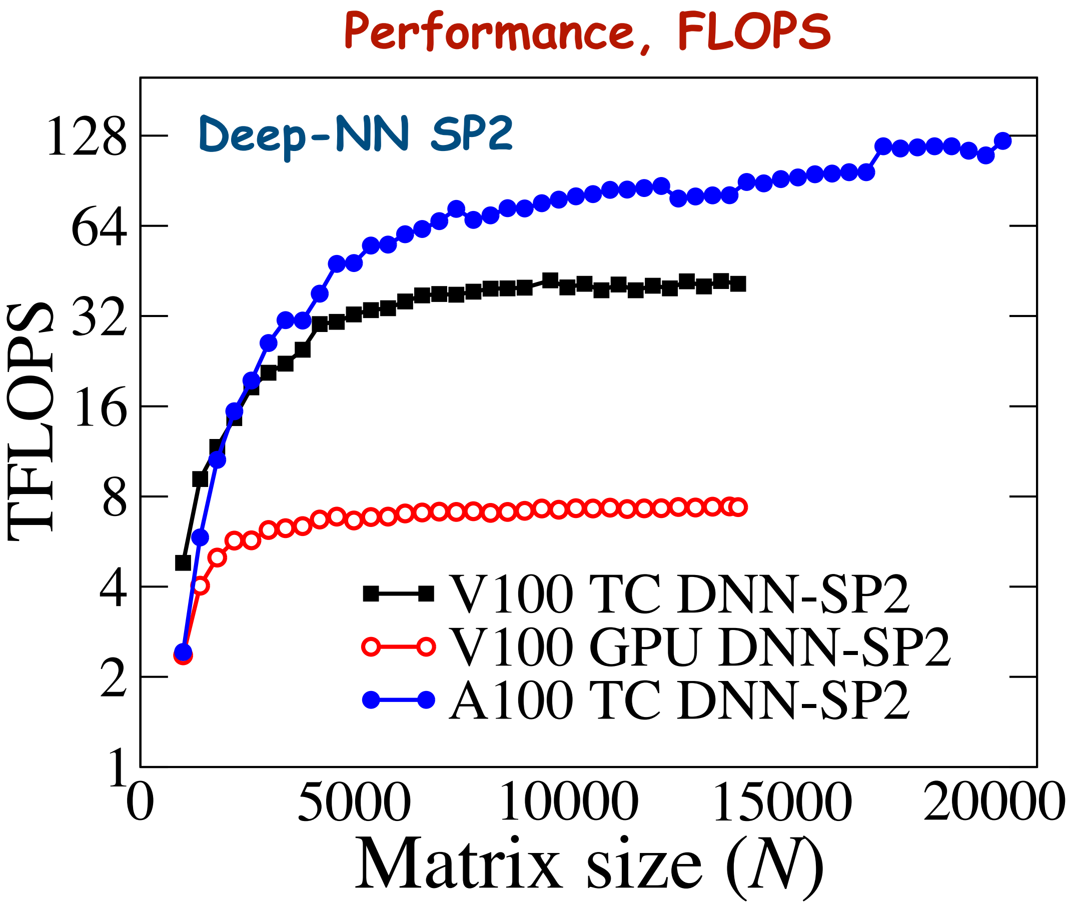
$H_{ij} = \langle \varphi_i \,|\, \hat{H}[\rho] \,|\, \varphi_j \rangle$
ϵ_N
$\mu-$
ϵ_3
ϵ_2
ϵ_1
$H = \sum_i \epsilon_i \mathbf{v}_i \mathbf{v}_i^T$
$D = \sum_i \theta(\mu - \epsilon_i) \mathbf{v}_i \mathbf{v}_i^T$
$= \theta(\mu I - H)$
1
0

$$\sum_{ij} D_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) = \rho(\mathbf{r})$$

$$E_{\text{band}} = \text{Tr}\,[HD]$$

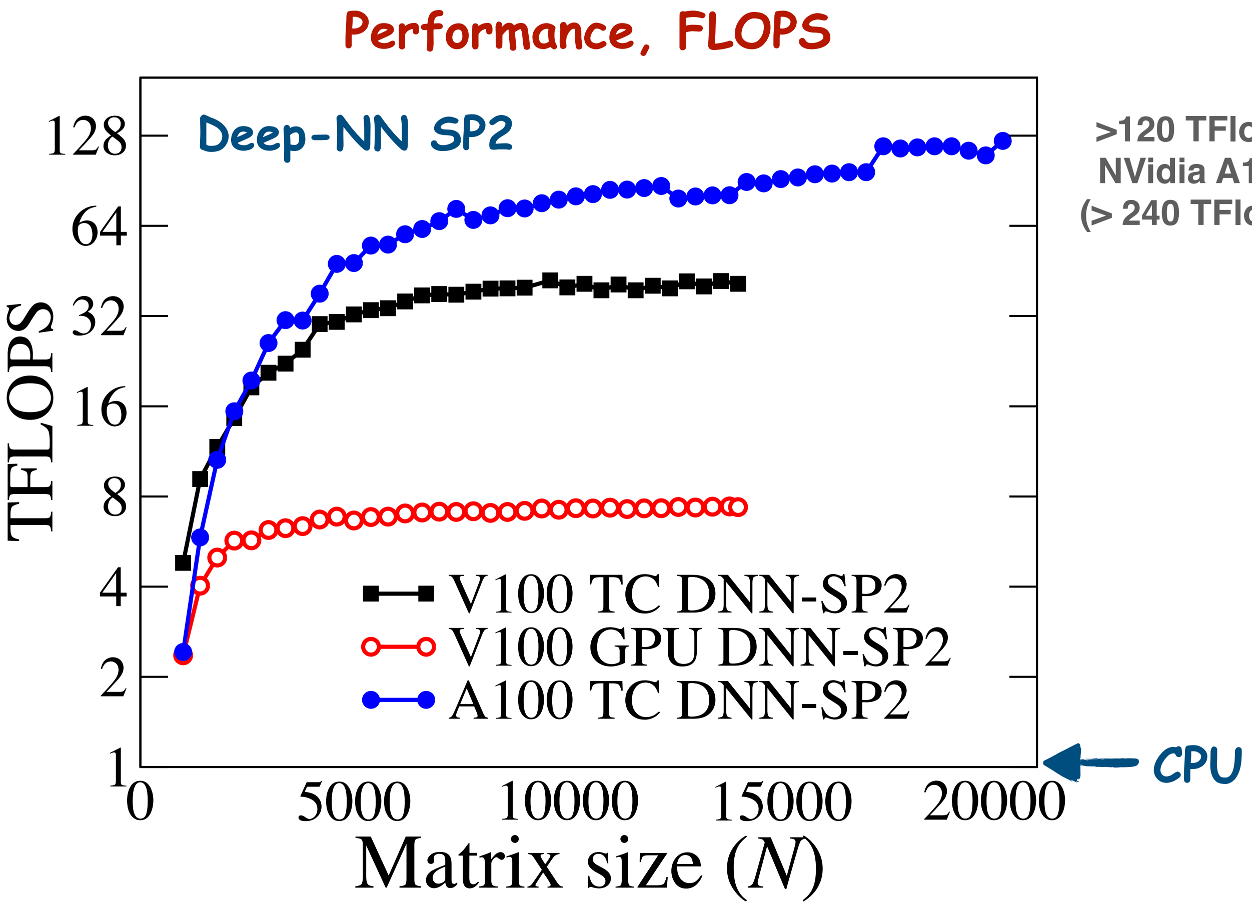
Robust parameter-free stopping criterion

The Electronic Structure from a Deep Neural Network



>120 TFlops
NVidia A100
(> 240 TFlops)

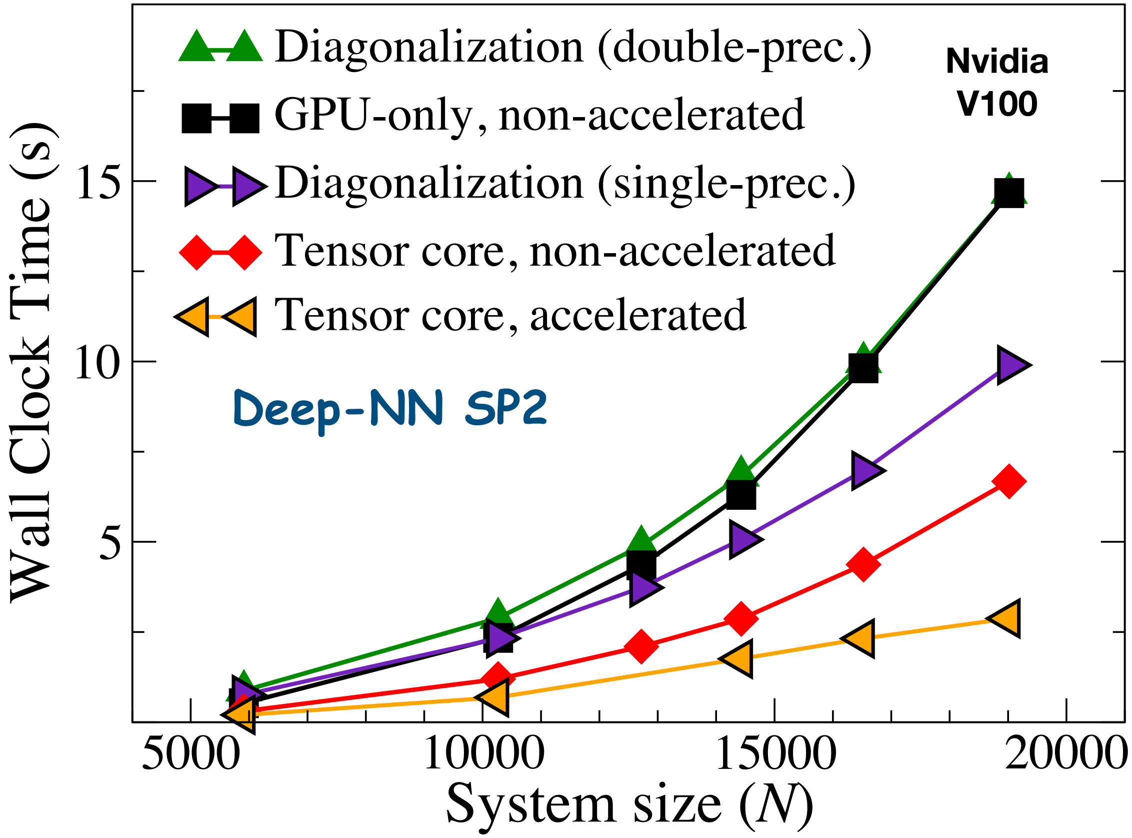
The Electronic Structure from a Deep Neural Network



>120 TFlops
NVidia A100
(> 240 TFlops)

The Electronic Structure from a Deep Neural Network

Performance, Wall Clock



40-60 x CPU

Electronic Structure Methods for Molecular Dynamics Simulations

$$U(\mathbf{R}) = \min_{\rho\in\text{constr.}} \left\{ \Omega\left[\mathbf{R},\rho(\mathbf{r})\right] + \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \,\middle|\, \int \rho(\mathbf{r})d\mathbf{r} = N_e \right\} + V(\mathbf{R})$$

$$M_I\ddot{\mathbf{R}}_I = -\nabla_I U(\mathbf{R})$$

$$\left\{U(\mathbf{R}),\ \nabla_I U(\mathbf{R})\right\}$$

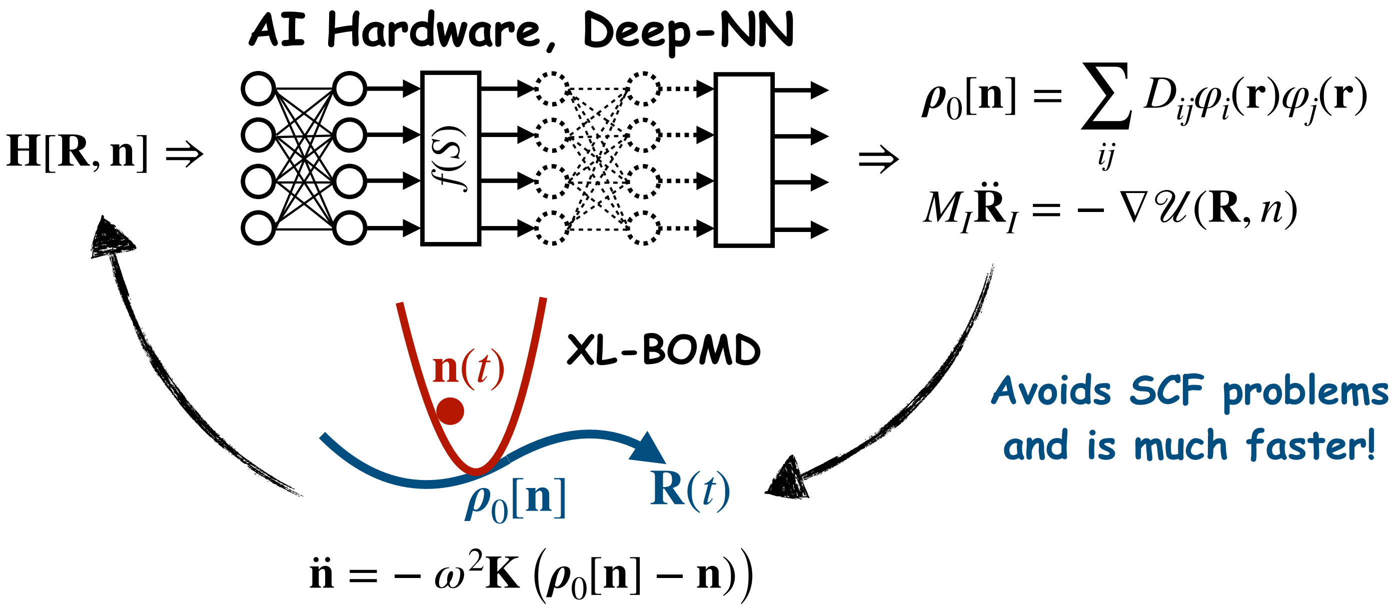
~~Non-linearities~~ → Numerical ~~sensitivities~~ → Low-Precision

Shadow Molecular Dynamics

Using AI-hardware
is possible!

Los Alamos
NATIONAL LABORATORY
EST. 1943

Quantum-based Shadow Molecular Dynamics using Deep-NN and AI Hardware



$$\mathbf{H}[\mathbf{R},\mathbf{n}] \Rightarrow$$

$$\Rightarrow \quad \rho_0[\mathbf{n}] = \sum_{ij} D_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})$$

$$M_I\ddot{\mathbf{R}}_I = -\nabla\mathcal{U}(\mathbf{R},n)$$

$$\ddot{\mathbf{n}} = -\omega^2\mathbf{K}\left(\rho_0[\mathbf{n}] - \mathbf{n}\right)$$

Avoids SCF problems and is much faster!

Dual mixed precision

$$X = f(S) \approx S_{\text{high}}^2 + S_{\text{high}}S_{\text{low}} + \left(S_{\text{high}}S_{\text{low}}\right)^T$$

$$S_{\text{high}} = \text{FP16}[S]$$

$$S_{\text{low}} = \text{FP16}[S - S_{\text{high}}]$$

XL-BOMD

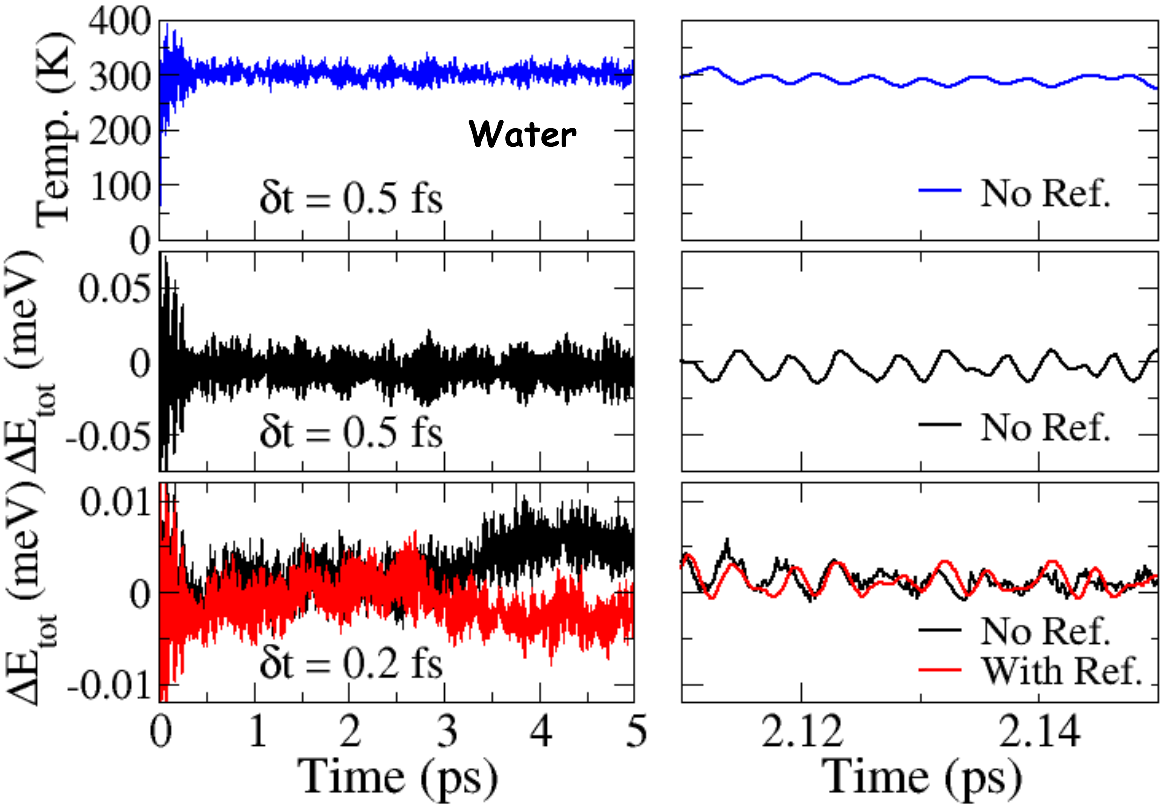
$$M_I\ddot{\mathbf{R}}_I = -\nabla\mathcal{U}(\mathbf{R},n)\Big|_n \quad \textbf{Shadow potential}$$

$$\ddot{n}(\mathbf{r}) = -\omega^2\int K(\mathbf{r},\mathbf{r}')\left(\rho_0[n](\mathbf{r}') - n(\mathbf{r}')\right)d\mathbf{r}'$$

Los Alamos NATIONAL LABORATORY EST. 1943

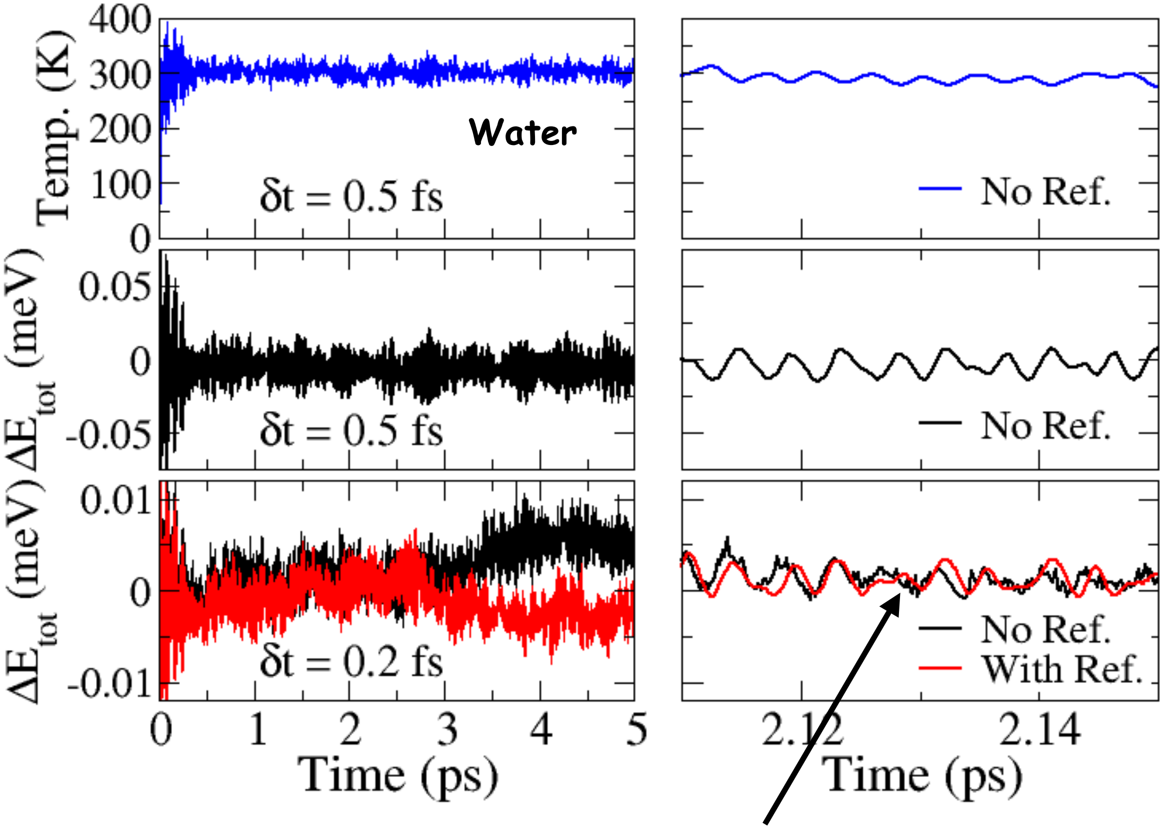
Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

Water XL-BOMD simulations (SCC-DFTB) using a Deep-NN and Tensor cores (AI hardware)



Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

Water XL-BOMD simulations (SCC-DFTB) using a Deep-NN and Tensor cores (AI hardware)



What about the noise?

Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

Include the numerical noise from the low-precision arithmetics as a natural part of a Langevin dynamics in a canonical NVT simulation

XL-BOMD within a Langevin dynamics

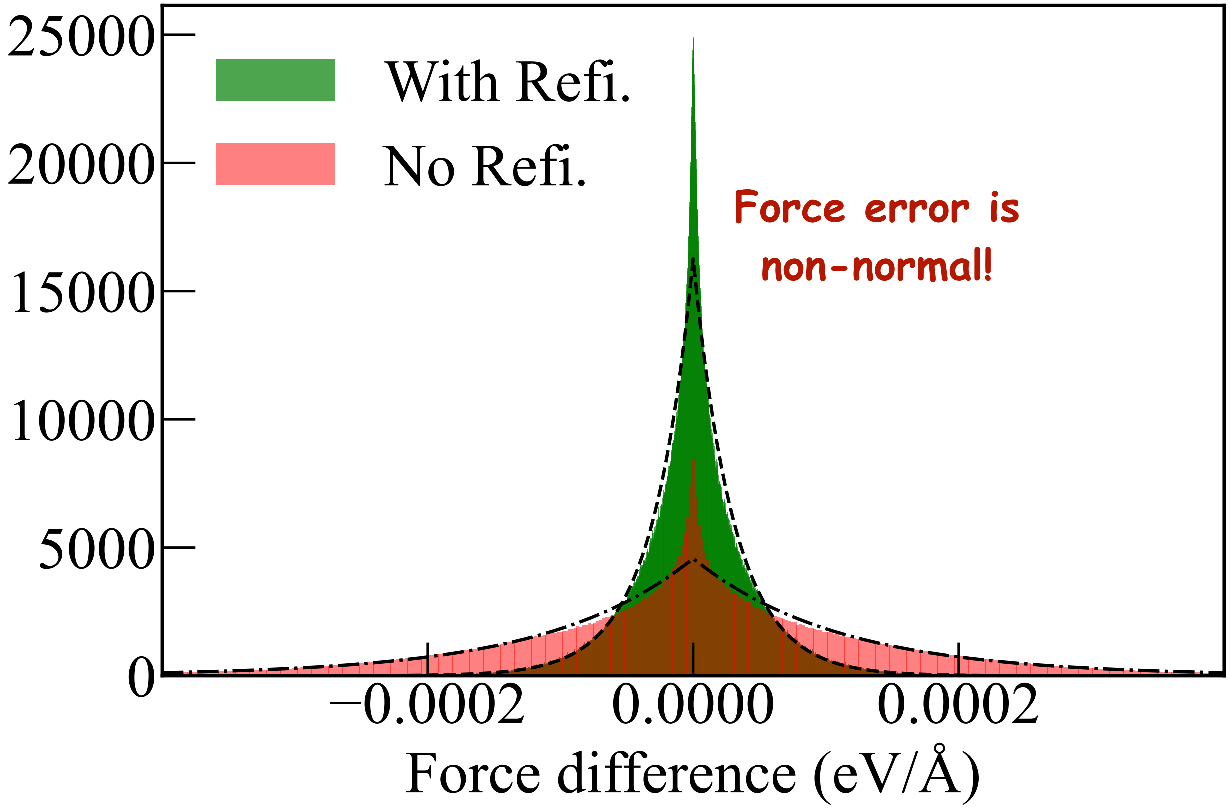
$$dR_I = \dot{\mathbf{R}}_I dt,$$

$$d\dot{\mathbf{R}}_I = -\frac{1}{M_I}\left(\left.\frac{\partial \mathcal{U}(\mathbf{R},n)}{\partial \mathbf{R}_I}\right|_n + \overset{\substack{\text{Numerical Noise} \\ \text{(internal)}}}{\xi_t^I} + \overset{\text{Dissipation}}{\gamma_I \dot{\mathbf{R}}_I} \right) dt,$$

$$\ddot{n}(\mathbf{r}) = -\,\omega^2 \int K(\mathbf{r},\mathbf{r}')(\rho_0[n](\mathbf{r}') - n(\mathbf{r}'))d\mathbf{r}',$$

Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

The Unexpected Problem!



Force error is non-normal!

Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

Our Solution!

Novel Canonical Integration Scheme

$$V_{k+1/4} = V_k + \frac{\delta t}{2M}\left(F_k^{\mathrm{TC}} - \gamma_{\mathrm{TC}} V_k\right)$$

$$R_{k+1/2} = R_k + \frac{\delta t}{2} V_{k+1/4}$$

$$V_{k+3/4} = c_L\, V_{k+1/4} + \sigma_L\, \eta_k$$

$$R_{k+1} = R_{k+1/2} + \frac{\delta t}{2} V_{k+3/4}$$

$$V_{k+1} = V_{k+3/4} + \frac{\delta t}{2M}\left(F_{k+1}^{\mathrm{TC}} - \gamma_{\mathrm{TC}} V_{k+1}\right)$$

Non-standard Fluctuation-Dissipations

$$\sigma_L = \sqrt{k_{\mathrm{B}} T (1 - c_L^2)/M}$$

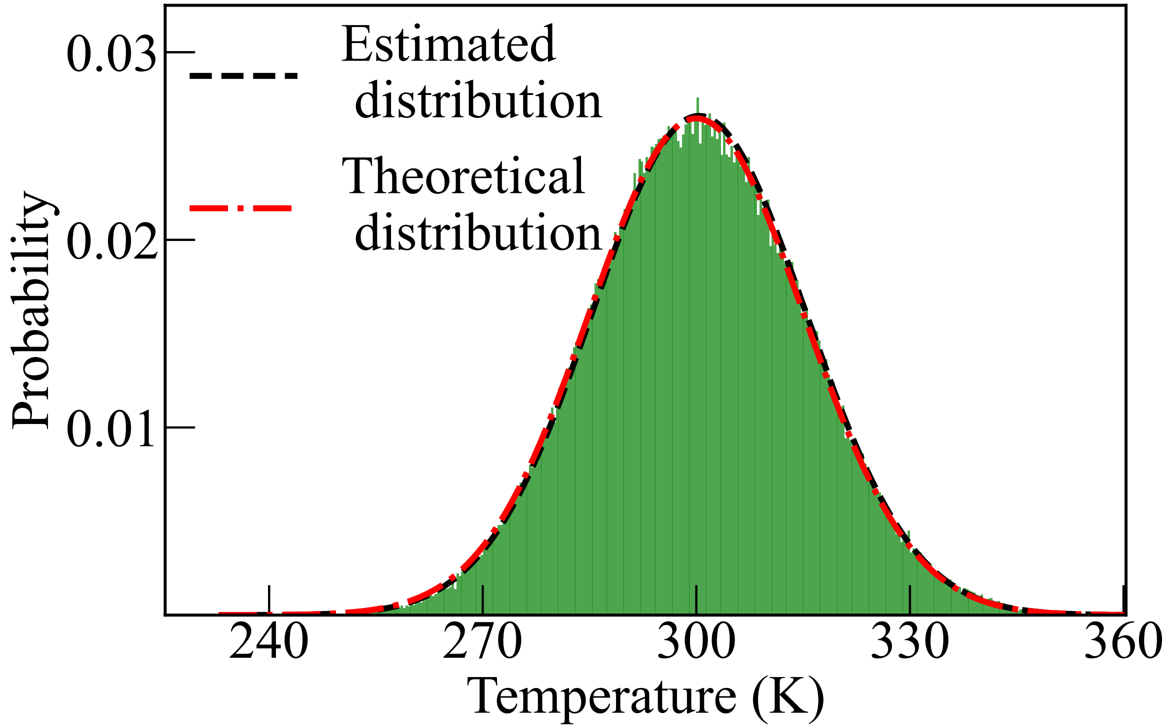
$$\sigma_{\mathrm{TC}} = \sqrt{2 k_{\mathrm{B}} T \gamma_{\mathrm{TC}} / \delta t}$$

$$\eta_k \in \mathcal{N}(0,1)$$

$$c_L = \frac{1 - \gamma_L \delta t/2}{1 + \gamma_L \delta t/2}$$

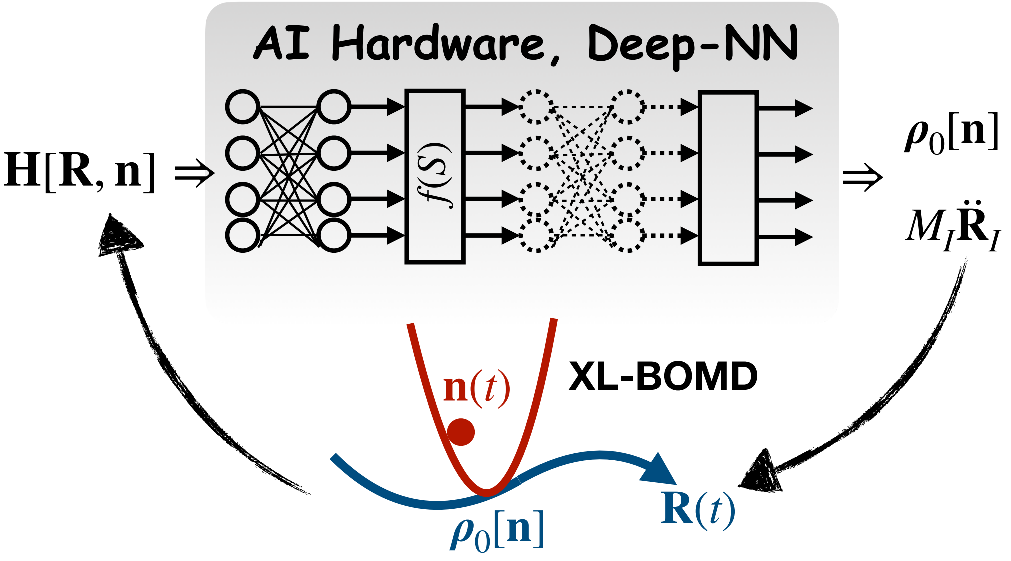
Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

XL-BOMD with a "Langevin-like" dynamic (Water, SCC-DFTB)



Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

Quantum-based Molecular Dynamics using Deep-NN and AI Hardware



$$\rho_0[\mathbf{n}] = \sum_{ij} D_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})$$

$$M_I\ddot{\mathbf{R}}_I = -\nabla\mathcal{U}(\mathbf{R},n)$$

Langevin-like dynamics

$$\sigma_L = \sqrt{k_{\mathrm{B}}T(1-c_L^2)/M}$$

$$\sigma_{\mathrm{TC}} = \sqrt{2k_{\mathrm{B}}T\gamma_{\mathrm{TC}}/\delta t}$$

Dual mixed precision

$$X = f(S) \approx S_{\mathrm{high}}^2 + S_{\mathrm{high}}S_{\mathrm{low}} + \left(S_{\mathrm{high}}S_{\mathrm{low}}\right)^T$$

$$S_{\mathrm{high}} = \mathrm{FP16}[S]$$

$$S_{\mathrm{low}} = \mathrm{FP16}[S - S_{\mathrm{high}}]$$

XL-BOMD

$$M_I\ddot{\mathbf{R}}_I = -\nabla\mathcal{U}(\mathbf{R},n)\Big|_n$$ **Shadow potential**

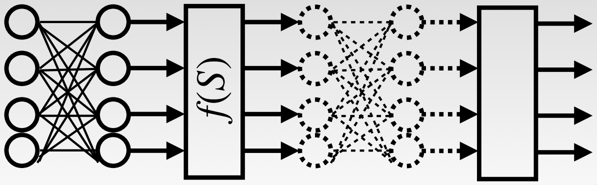
$$\ddot{n}(\mathbf{r}) = -\omega^2\int K(\mathbf{r},\mathbf{r}')\big(\rho_0[n](\mathbf{r}') - n(\mathbf{r}')\big)\,d\mathbf{r}'$$

Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)

Quantum-based Molecular Dynamics using Deep-NN and AI Hardware

AI Hardware, Deep-NN

$\mathbf{H}[\mathbf{R},\mathbf{n}] \Rightarrow$



$$\Rightarrow \rho_0[\mathbf{n}] = \sum_{ij} D_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})$$

$$M_I\ddot{\mathbf{R}}_I = -\nabla\mathcal{U}(\mathbf{R},n)$$

Coordinated Design Approach

to reach the full potential of AI hardware

Langevin-like Dynamics

$$\sigma_L = \sqrt{k_{\rm B}T(1-c_L^2)/M}$$

$$\sigma_{\rm TC} = \sqrt{2k_{\rm B}T\gamma_{\rm TC}/\delta t}$$

Dual mixed precision

$$X = f(S) \approx S_{\rm high}^2 + S_{\rm high}S_{\rm low} + \left(S_{\rm high}S_{\rm low}\right)^T$$

$$S_{\rm high} = {\rm FP16}[S]$$

$$S_{\rm low} = {\rm FP16}[S - S_{\rm high}]$$

XL-BOMD

$$M_I\ddot{\mathbf{R}}_I = -\nabla\mathcal{U}(\mathbf{R},n)\Big|_n$$ **Shadow potential**

$$\ddot{n}(\mathbf{r}) = -\omega^2\int K(\mathbf{r},\mathbf{r}')\big(\rho_0[n](\mathbf{r}') - n(\mathbf{r}')\big)\,d\mathbf{r}'$$

Finkelstein, Smith, Mniszewski, Barros, Negre, Rubensson, Niklasson, JCTC 17, 6180 (2021)