Development of efficient multiscale methods and extrapolation techniques for multiphysics molecular chemistry

Étienne Polack

Doctoral research project defence, January 2022

Chemistry

Chemistry science of matter and its elements;

Chemistry

Chemistry science of matter and its elements; *understanding* it to

Chemistry

Chemistry science of matter and its elements; *understanding* it to

- Synthesise materials (*e.g.*, plastics, 2D materials)

Chemistry

Chemistry science of matter and its elements; *understanding* it to

- Synthesise materials (*e.g.*, plastics, 2D materials)
- Optimise industrial processes (*e.g.*, sewage plants)

Chemistry and drug design as a practical application

Chemistry science of matter and its elements; *understanding* it to

- Synthesise materials (*e.g.*, plastics, 2D materials)
- Optimise industrial processes (*e.g.*, sewage plants)
- Discover new drugs (*e.g.*, opioid antagonist naloxone)

Chemistry and drug design as a practical application

Chemistry science of matter and its elements; *understanding* it to

- Synthesise materials (*e.g.*, plastics, 2D materials)
- Optimise industrial processes (*e.g.*, sewage plants)
- Discover new drugs (*e.g.*, opioid antagonist naloxone)

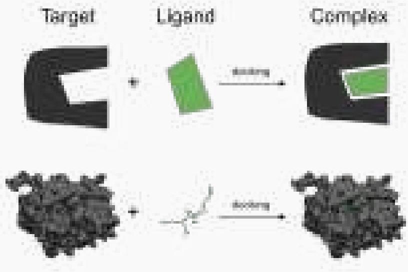
Interest for drug design applications

Chemistry and drug design as a practical application

Chemistry science of matter and its elements; *understanding* it to

- Synthesise materials (*e.g.*, plastics, 2D materials)
- Optimise industrial processes (*e.g.*, sewage plants)
- Discover new drugs (*e.g.*, opioid antagonist naloxone)

Molecular docking



Interest for drug design applications

- Need for predictive models

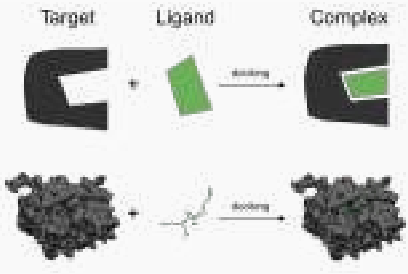
https://commons.wikimedia.org/wiki/File:Docking_representation_2.png

Chemistry and drug design as a practical application

Chemistry science of matter and its elements; *understanding* it to

- Synthesise materials (*e.g.*, plastics, 2D materials)
- Optimise industrial processes (*e.g.*, sewage plants)
- Discover new drugs (*e.g.*, opioid antagonist naloxone)

Molecular docking



Interest for drug design applications

- Need for predictive models
- Wide size range of interesting compounds

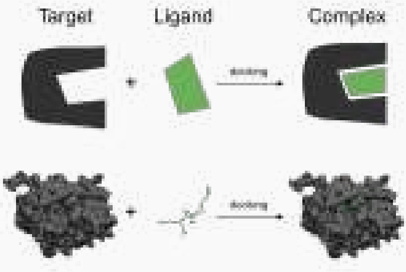
https://commons.wikimedia.org/wiki/File:Docking_representation_2.png

Chemistry and drug design as a practical application

Chemistry science of matter and its elements; *understanding* it to

- Synthesise materials (*e.g.*, plastics, 2D materials)
- Optimise industrial processes (*e.g.*, sewage plants)
- Discover new drugs (*e.g.*, opioid antagonist naloxone)

Molecular docking



Interest for drug design applications

- Need for predictive models
- Wide size range of interesting compounds

→ Quick increase in complexity

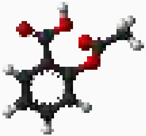
https://commons.wikimedia.org/wiki/File:Docking_representation_2.png

Molecular dynamics simulations and multiscale physics

Time

System size

Molecular dynamics simulations and multiscale physics

Time	fs–ps
	Quantum mechanics 
System size	fm–nm

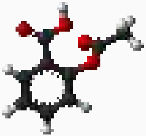
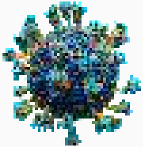
SCHRÖDINGER equation: *ab-initio*, high accuracy, nuclei and electrons

- Density functional theory (DFT)
- (post-)HARTREE–FOCK (HF) methods

Thousand of atoms

https://commons.wikimedia.org/wiki/File:Aspirin-B-3D-balls.png

Molecular dynamics simulations and multiscale physics

Time	fs–ps	ns–µs
	Quantum mechanics	Classical mechanics
		
System size	fm–nm	nm–µm

NEWTON equation: parameterised through force fields, medium accuracy, only nuclei

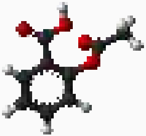
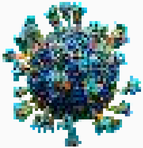
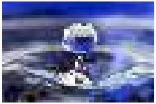
- AMBER
- CHARMM
- AMOEBA

Millions of atoms

https://commons.wikimedia.org/wiki/File:Aspirin-B-3D-balls.png

https://commons.wikimedia.org/wiki/File:Coronavirus._SARS-CoV-2.png

Molecular dynamics simulations and multiscale physics

Time	fs–ps	ns–µs	ms–s
	Quantum mechanics	Classical mechanics	Continuum solvation
			
System size	fm–nm	nm–µm	µm–mm

Interaction with an electric field to model solvent

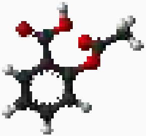
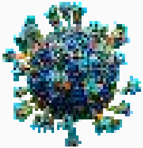
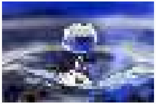
- COSMO

https://commons.wikimedia.org/wiki/File:Aspirin-B-3D-balls.png

https://commons.wikimedia.org/wiki/File:Coronavirus._SARS-CoV-2.png

https://commons.wikimedia.org/wiki/File:Water_drop_001.jpg

Molecular dynamics simulations and multiscale physics

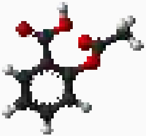
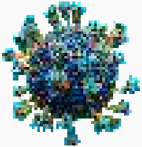
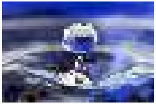
Time	fs–ps	ns–µs	ms–s
	Quantum mechanics	Classical mechanics	Continuum solvation
			
System size	fm–nm	nm–µm	µm–mm

Focus

Context QM/MM(AMOEBA), multiphysics quantum mechanics/molecular mechanics coupling

https://commons.wikimedia.org/wiki/File:Aspirin-B-3D-balls.png
https://commons.wikimedia.org/wiki/File:Coronavirus._SARS-CoV-2.png
https://commons.wikimedia.org/wiki/File:Water_drop_001.jpg

Molecular dynamics simulations and multiscale physics

Time	fs–ps	ns–µs	ms–s
	Quantum mechanics	Classical mechanics	Continuum solvation
			
System size	fm–nm	nm–µm	µm–mm

Focus

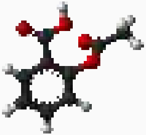
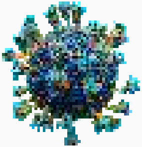
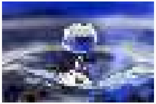
Context QM/MM(AMOEBA), multiphysics quantum mechanics/molecular mechanics coupling

- Part of system described with quantum mechanics (*e.g.*, active site)

https://commons.wikimedia.org/wiki/File:Aspirin-B-3D-balls.png

https://commons.wikimedia.org/wiki/File:Coronavirus._SARS-CoV-2.png

https://commons.wikimedia.org/wiki/File:Water_drop_001.jpg

Time	fs–ps	ns–µs	ms–s
	Quantum mechanics	Classical mechanics	Continuum solvation
			
System size	fm–nm	nm–µm	µm–mm

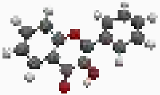
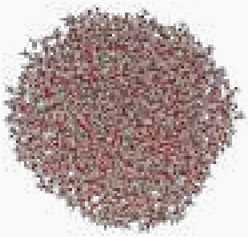
Focus

Context QM/MM(AMOEBA), multiphysics quantum mechanics/molecular mechanics coupling

I Part of system described with quantum mechanics (*e.g.*, active site)

II Part of system described with classical mechanics (*e.g.*, solvent)

https://commons.wikimedia.org/wiki/File:Aspirin-B-3D-balls.png
https://commons.wikimedia.org/wiki/File:Coronavirus._SARS-CoV-2.png
https://commons.wikimedia.org/wiki/File:Water_drop_001.jpg

 + 2503×  = 

Outline

+

2503× 

=

Goal Speeding-up QM/MM(AMOEBA) multiscale ab-initio molecular dynamics

+ 2503×

=

Goal Speeding-up QM/MM(AMOEBA) multiscale ab-initio molecular dynamics

Outline

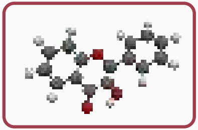
0 General introduction to quantum mechanics and classical mechanics

I Speeding-up quantum mechanics simulations

II Speeding-up classical mechanics simulations

Not discussed: *Polarisable classical molecular dynamics*, *Multiscale ab-initio molecular dynamics* and *Extrapolation using the magic points method* (chapters II, III and V)

Outline



+ 2503×

=

Goal Speeding-up QM/MM(AMOEBA) multiscale ab-initio molecular dynamics

Outline

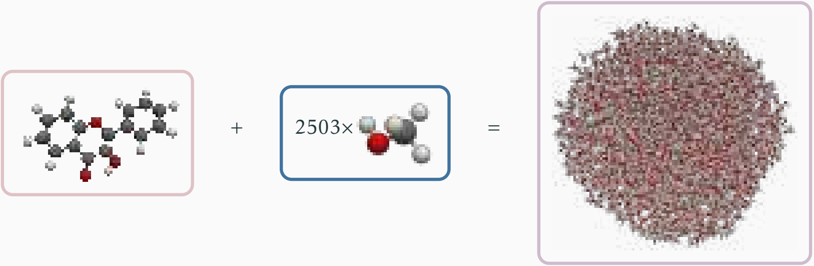
0 General introduction to quantum mechanics and classical mechanics

I Speeding-up quantum mechanics simulations

II Speeding-up classical mechanics simulations

Not discussed: *Polarisable classical molecular dynamics*, *Multiscale ab-initio molecular dynamics* and *Extrapolation using the magic points method* (chapters II, III and V)

Outline



Goal Speeding-up QM/MM(AMOEBA) multiscale ab-initio molecular dynamics

Outline

0 General introduction to quantum mechanics and classical mechanics

I Speeding-up quantum mechanics simulations

II Speeding-up classical mechanics simulations

Not discussed: *Polarisable classical molecular dynamics*, *Multiscale ab-initio molecular dynamics* and *Extrapolation using the magic points method* (chapters II, III and V)

The time-independent Schrödinger equation

- Molecular system[1] with
  - M nuclei of charge Z_k at positions $\mathbf{R}_k$
  - N electrons of charge -1 au

[1] No particles' spins, Born–Oppenheimer approximation.

The time-independent Schrödinger equation

- Molecular system[1] with
  - M nuclei of charge Z_k at positions $\mathbf{R}_k$
  - N electrons of charge -1 au
- Electrons described by antisymmetric[2] wave functions $\psi \in \mathsf{L}^2(\mathbb{R}^{3N}, \mathbb{C})$

[1]No particles' spins, Born–Oppenheimer approximation.
[2]Pauli's exclusion principle

The time-independent SCHRÖDINGER equation

- Molecular system[1] with
  - M nuclei of charge Z_k at positions $\mathbf{R}_k$
  - N electrons of charge -1 au
- Electrons described by antisymmetric[2] wave functions $\psi \in \mathrm{L}^2(\mathbb{R}^{3N}, \mathbb{C})$
- Ground state of SCHRÖDINGER equation is of interest (system of lowest energy)

[1] No particles' spins, BORN–OPPENHEIMER approximation.
[2] PAULI's exclusion principle

The time-independent SCHRÖDINGER equation

- Molecular system[1] with
  - M nuclei of charge Z_k at positions $\mathbf{R}_k$
  - N electrons of charge -1 au
- Electrons described by antisymmetric[2] wave functions $\psi \in \mathsf{L}^2(\mathbb{R}^{3N}, \mathbb{C})$
- Ground state of SCHRÖDINGER equation is of interest (system of lowest energy)

SCHRÖDINGER equation (ground state)

$$\inf\{\langle \psi, \mathcal{H}\psi \rangle \mid \psi \in \mathsf{L}^2_a(\mathbb{R}^{3N}, \mathbb{C}),\ |\psi| = 1\}$$

[1] No particles' spins, BORN–OPPENHEIMER approximation.

[2] PAULI's exclusion principle

The time-independent Schrödinger equation

- Molecular system[1] with
  - M nuclei of charge Z_k at positions $\mathbf{R}_k$
  - N electrons of charge -1 au
- Electrons described by antisymmetric[2] wave functions $\psi \in \mathrm{L}^2(\mathbb{R}^{3N}, \mathbb{C})$
- Ground state of Schrödinger equation is of interest (system of lowest energy)

Schrödinger equation (ground state)

$$\inf\{\langle \psi, \mathcal{H}\psi \rangle \,\big|\, \psi \in \mathrm{L}^2_a(\mathbb{R}^{3N}, \mathbb{C}),\ |\psi| = 1\}$$

$$\mathcal{H} := \underbrace{-\frac{1}{2}\sum_{i=1}^{N} \Delta_{\mathbf{r}_i}}_{\text{kinetic energy}} + \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{M} \frac{-Z_k}{|\mathbf{r}_i - \mathbf{R}_k|}}_{\text{interaction n-e}} + \underbrace{\sum_{1 \leqslant i < j \leqslant N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\text{interaction e-e}}$$

[1] No particles' spins, Born–Oppenheimer approximation.
[2] Pauli's exclusion principle

The time-independent SCHRÖDINGER equation

- Molecular system[1] with
  - M nuclei of charge Z_k at positions $\mathbf{R}_k$
  - N electrons of charge -1 au
- Electrons described by antisymmetric[2] wave functions $\psi \in \mathrm{L}^2(\mathbb{R}^{3N}, \mathbb{C})$
- Ground state of SCHRÖDINGER equation is of interest (system of lowest energy)

SCHRÖDINGER equation (ground state)

$$\inf\{\langle \psi, \mathcal{H}\psi \rangle \mid \psi \in \mathrm{L}^2_a(\mathbb{R}^{3N}, \mathbb{C}),\ |\psi| = 1\}$$

$$\mathcal{H} := \underbrace{-\frac{1}{2}\sum_{i=1}^{N} \Delta_{\mathbf{r}_i}}_{\text{kinetic energy}} + \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{M} \frac{-Z_k}{|\mathbf{r}_i - \mathbf{R}_k|}}_{\text{interaction n-e}} + \underbrace{\sum_{1 \leqslant i < j \leqslant N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\text{interaction e-e}}$$

$$\mathrm{L}^2_a(\mathbb{R}^{3N}, \mathbb{C}) := \bigwedge_{1}^{N} \mathrm{L}^2(\mathbb{R}^3, \mathbb{C})$$

[1] No particles' spins, BORN–OPPENHEIMER approximation.

[2] PAULI's exclusion principle

Curse of dimensionality

Schrödinger equation is great

Curse of dimensionality

SCHRÖDINGER equation is great, but ψ defined on $L^2_a(\mathbb{R}^{3N}, \mathbb{C})$

Curse of dimensionality

SCHRÖDINGER equation is great, but ψ defined on $L^2_a(\mathbb{R}^{3N}, \mathbb{C})$

$\Rightarrow$ curse of dimensionality!

Curse of dimensionality

SCHRÖDINGER equation is great, but ψ defined on $L^2_a(\mathbb{R}^{3N}, \mathbb{C})$

$\Rightarrow$ curse of dimensionality!

Acetylsalicylic acid $C_9H_8O_4$ (aspirin)

Contains 21 nuclei and 94 electrons

Curse of dimensionality

SCHRÖDINGER equation is great, but ψ defined on $L^2_a(\mathbb{R}^{3N}, \mathbb{C})$

$\Rightarrow$ curse of dimensionality!

Acetylsalicylic acid $C_9H_8O_4$ (aspirin)

Contains 21 nuclei and 94 electrons
SCHRÖDINGER problem (BORN–OPPENHEIMER) of dimension greater than $2^{3\times 94} \approx 10^{85}$

Curse of dimensionality

SCHRÖDINGER equation is great, but ψ defined on $L^2_a(\mathbb{R}^{3N}, \mathbb{C})$

$\Rightarrow$ curse of dimensionality!

Acetylsalicylic acid $C_9H_8O_4$ (aspirin)

Contains 21 nuclei and 94 electrons
SCHRÖDINGER problem (BORN–OPPENHEIMER) of dimension greater than $2^{3\times 94} \approx 10^{85}$
Larger than 10^{80}, the EDDINGTON number (protons in observable universe)

Curse of dimensionality

SCHRÖDINGER equation is great, but ψ defined on $\mathrm{L}^2_a(\mathbb{R}^{3N}, \mathbb{C})$

$\Rightarrow$ curse of dimensionality!

Acetylsalicylic acid $C_9H_8O_4$ (aspirin)

Contains 21 nuclei and 94 electrons
SCHRÖDINGER problem (BORN–OPPENHEIMER) of dimension greater than $2^{3\times 94} \approx 10^{85}$
Larger than 10^{80}, the EDDINGTON number (protons in observable universe)

Need for approximations

Density functional theory

Density functional theory Considers electronic density ρ defined on $\mathbb{R}^3$

Density functional theory

Density functional theory Considers electronic density ρ defined on $\mathbb{R}^3$

Electronic density

$$\rho_\psi(\mathbf{r}) := N \int_{\mathbb{R}^{3(N-1)}} \left|\psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)\right|^2 \mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N$$

with $\psi \in \mathrm{L}^2_a(\mathbb{R}^{3N}, \mathbb{C})$

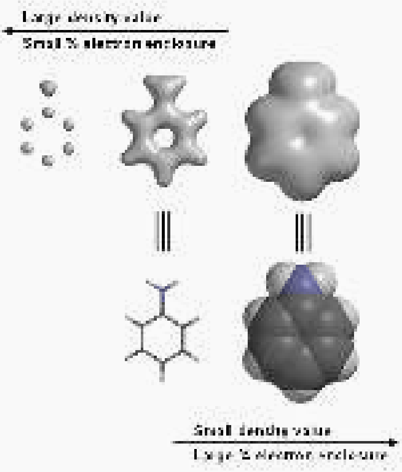
Density functional theory

Density functional theory Considers electronic density ρ defined on $\mathbb{R}^3$

Electronic density

$$\rho_\psi(\mathbf{r}) := N \int_{\mathbb{R}^{3(N-1)}} \left|\psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)\right|^2 \mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N$$

with $\psi \in \mathrm{L}^2_a(\mathbb{R}^{3N}, \mathbb{C})$



https://commons.wikimedia.org/wiki/File:ElectronDensityAniline.PNG

Equation with density *unknown*

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham functional

$$\inf_{\rho} \left\{ \mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \,\middle|\, \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N \right\}$$

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham functional

$$\inf_{\rho}\left\{\mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \,\middle|\, \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N\right\}$$

- External potential (interaction with nuclei)

KOHN–SHAM model

Equation with density *unknown*, but existence of a functional by HOHENBERG–KOHN *theorem*

KOHN–SHAM functional

$$\inf_{\rho}\left\{\mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \,\middle|\, \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N\right\}$$

- External potential (interaction with nuclei)
- Approximation of the kinetics energy

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham functional

$$\inf_{\rho}\left\{\mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \;\middle|\; \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N\right\}$$

- External potential (interaction with nuclei)
- Approximation of the kinetics energy
- Approximations of the repulsive electronic energy

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham functional

$$\inf_{\rho}\left\{\mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \,\middle|\, \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N\right\}$$

- External potential (interaction with nuclei)
- Approximation of the kinetics energy
- Approximations of the repulsive electronic energy
- Exchange-correlation energy functional (approximations)

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham functional

$$\inf_{\rho}\left\{\mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \,\middle|\, \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N\right\}$$

Rewritten nonlinear eigenproblem[1]

Find N lowest eigenvalues and molecular orbitals solving the nonlinear equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}), \quad \text{for } i \in [\![1\,..\,N]\!]$$

in $\mathcal{W}_N := \left\{\Phi := (\phi_1, \dots, \phi_N) \in \mathrm{L}^2(\mathbb{R}^3, \mathbb{C})^N \,\middle|\, \langle \phi_i, \phi_j \rangle = \delta_{ij}, 1 \leq i, j \leq N\right\}$

[1] Euler–Lagrange on energy functional

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham functional

$$\inf_{\rho}\left\{\mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \,\middle|\, \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N\right\}$$

Rewritten nonlinear eigenproblem[1]

Find N lowest eigenvalues and molecular orbitals solving the nonlinear equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \quad \text{for } i \in [\![1\,..\,N]\!]$$

in $\mathcal{W}_N := \left\{\Phi := (\phi_1, \ldots, \phi_N) \in \mathrm{L}^2(\mathbb{R}^3, \mathbb{C})^N \,\middle|\, \langle\phi_i, \phi_j\rangle = \delta_{ij}, 1 \leq i, j \leq N\right\}$

N problems on $\mathrm{L}^2(\mathbb{R}^3, \mathbb{C})$ instead of one problem on $\mathrm{L}^2(\mathbb{R}^{3N}, \mathbb{C})$

[1] Euler–Lagrange on energy functional

Kohn–Sham model

Equation with density *unknown*, but existence of a functional by Hohenberg–Kohn *theorem*

Kohn–Sham functional

$$\inf_{\rho}\left\{\mathcal{E}_{KS}(\rho) := \underbrace{T_{KS}(\rho) + J(\rho) + E_{xc}(\rho)}_{\mathcal{F}(\rho)} + \int \rho V \,\middle|\, \rho \geq 0,\ \rho^{1/2} \in \mathrm{H}^1(\mathbb{R}^3),\ \int \rho = N\right\}$$

Rewritten nonlinear eigenproblem[1]

Find N lowest eigenvalues and molecular orbitals solving the nonlinear equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \quad \text{for } i \in [\![1\,..\,N]\!]$$

in $\mathcal{W}_N := \left\{\Phi := (\phi_1, \dots, \phi_N) \in \mathrm{L}^2(\mathbb{R}^3, \mathbb{C})^N \,\middle|\, \langle\phi_i, \phi_j\rangle = \delta_{ij}, 1 \leq i, j \leq N\right\}$

N problems on $\mathrm{L}^2(\mathbb{R}^3, \mathbb{C})$ instead of one problem on $\mathrm{L}^2(\mathbb{R}^{3N}, \mathbb{C})$

Electronic density $\rho(\mathbf{r}) = 2\sum_{i=1}^{N}|\phi_i(\mathbf{r})|^2$

[1] Euler–Lagrange on energy functional

Discretisation of the KOHN–SHAM equations

GALERKIN approximation of $\mathsf{L}^2(\mathbb{R}^3, \mathbb{C})$: $\mathcal{V} := \mathrm{Span}(\chi_1, \dots, \chi_{\mathcal{N}}) \subset \mathsf{L}^2(\mathbb{R}^3, \mathbb{C})$, with χ_μ Gaussian type orbitals, centered on the atoms

Discretisation of the KOHN–SHAM equations

GALERKIN approximation of $\mathsf{L}^2(\mathbb{R}^3, \mathbb{C})$: $\mathcal{V} := \mathrm{Span}(\chi_1, \dots, \chi_{\mathcal{N}}) \subset \mathsf{L}^2(\mathbb{R}^3, \mathbb{C})$, with χ_μ Gaussian type orbitals, centered on the atoms

Solutions as linear combination of atomic orbitals (LCAO)

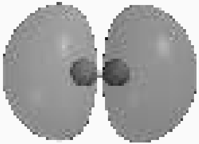
$$\Phi \in \mathcal{V}^N \Rightarrow \phi_i(\mathbf{r}) := \sum_{\mu=1}^{\mathcal{N}} C_{\mu i} \chi_\mu(\mathbf{r}), \quad i \in [\![1 \,..\, N]\!]$$

Discretisation of the KOHN–SHAM equations

GALERKIN approximation of $\mathsf{L}^2(\mathbb{R}^3, \mathbb{C})$: $\mathcal{V} := \mathrm{Span}(\chi_1, \dots, \chi_{\mathcal{N}}) \subset \mathsf{L}^2(\mathbb{R}^3, \mathbb{C})$, with χ_μ Gaussian type orbitals, centered on the atoms

Solutions as linear combination of atomic orbitals (LCAO)

$$\Phi \in \mathcal{V}^N \Rightarrow \phi_i(\mathbf{r}) := \sum_{\mu=1}^{\mathcal{N}} C_{\mu i} \chi_\mu(\mathbf{r}), \quad i \in [\![1 \,..\, N]\!]$$



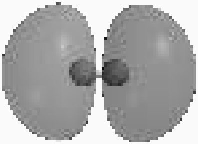
https://commons.wikimedia.org/wiki/File:Dihydrogen-phase-3D-balls.png

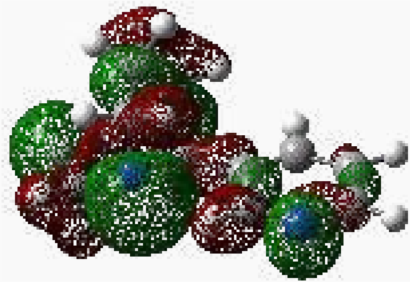
Discretisation of the KOHN–SHAM equations

GALERKIN approximation of $L^2(\mathbb{R}^3, \mathbb{C})$: $\mathcal{V} := \mathrm{Span}(\chi_1, \ldots, \chi_{\mathcal{N}}) \subset L^2(\mathbb{R}^3, \mathbb{C})$, with χ_μ Gaussian type orbitals, centered on the atoms

Solutions as linear combination of atomic orbitals (LCAO)

$$\Phi \in \mathcal{V}^N \Rightarrow \phi_i(\mathbf{r}) := \sum_{\mu=1}^{\mathcal{N}} C_{\mu i} \chi_\mu(\mathbf{r}), \quad i \in [\![1 \,..\, N]\!]$$





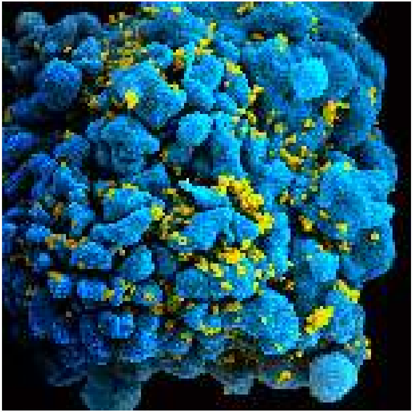
---

https://commons.wikimedia.org/wiki/File:Dihydrogen-phase-3D-balls.png

https://commons.wikimedia.org/wiki/File:Nicotine_(HOMO-LUMO).png

Classical molecular dynamics

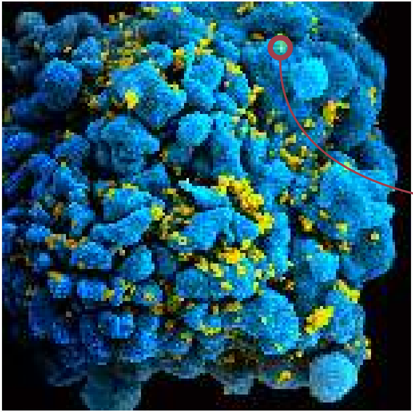
Human immunodeficiency virus 1 (≈ 100 nm) on T cell (≈ 5 µm)

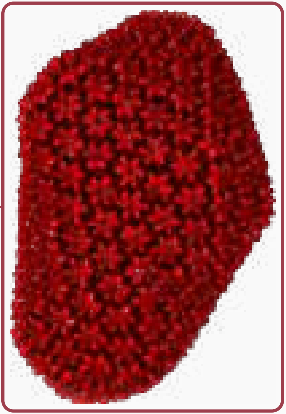


https://commons.wikimedia.org/wiki/File:HIV_H9_T-cell.jpg

Classical molecular dynamics

Human immunodeficiency virus 1 (≈ 100 nm) on T cell (≈ 5 µm), capsid (≈4 mil. atoms)



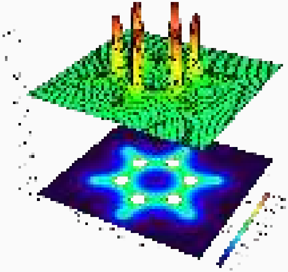


https://commons.wikimedia.org/wiki/File:HIV_H9_T-cell.jpg

https://commons.wikimedia.org/wiki/File:Protein_Imager_high_quality_illustration_example_1_HIV_capsid.png

Classical molecular mechanics

Cannot account for electrons

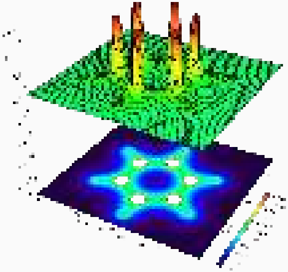


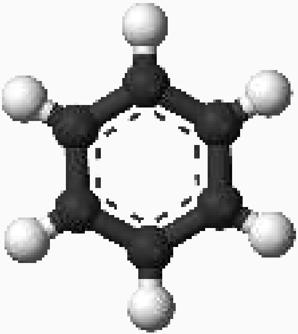
Benzene C_6H_6

https://commons.wikimedia.org/wiki/File:Elektronendichtedarstellungen_von_Benzol.png

Classical molecular mechanics

Cannot account for electrons; replaced by notion of (parameterised) chemical bounds



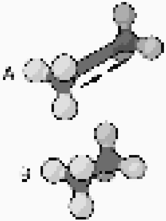


Benzene C_6H_6

https://commons.wikimedia.org/wiki/File:Elektronendichtedarstellungen_von_Benzol.png
https://commons.wikimedia.org/wiki/File:Benzene-aromatic-3D-balls.png

Force fields

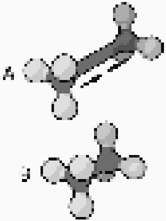
Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

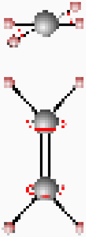


Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

https://commons.wikimedia.org/wiki/File:Bond_stretching_energy.svg

Force fields



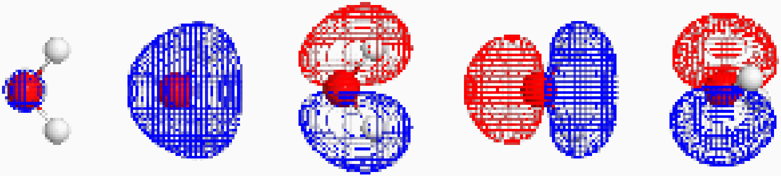


Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

https://commons.wikimedia.org/wiki/File:Bond_stretching_energy.svg

https://commons.wikimedia.org/wiki/File:Torsionsschwingung_ethylen.png

Molecular orbitals of water (quantum mechanics)

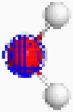


Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

https://commons.wikimedia.org/wiki/File:Molecular_Orbitals_for_Water.png

Force fields

Molecular orbitals of water (quantum mechanics)



Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

First order: point charges

https://commons.wikimedia.org/wiki/File:Molecular_Orbitals_for_Water.png

Force fields

Molecular orbitals of water (quantum mechanics)

Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

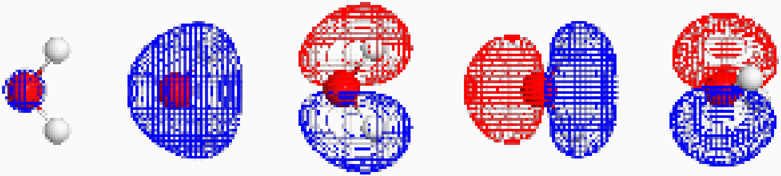
First order: point charges

For anisotropy: static multipoles

https://commons.wikimedia.org/wiki/File:Molecular_Orbitals_for_Water.png

Force fields

Molecular orbitals of water (quantum mechanics)



Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

First order: point charges

For anisotropy: static multipoles

Second order: polarisation (changes of properties with the geometry)

https://commons.wikimedia.org/wiki/File:Molecular_Orbitals_for_Water.png

Force fields

Force fields Choices of representation (intra- and intermolecular terms) along with parameterisations

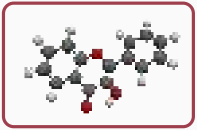
First order: point charges

For anisotropy: static multipoles

Second order: polarisation (changes of properties with the geometry)

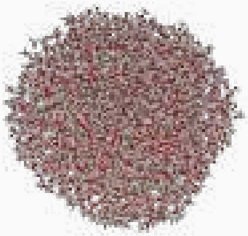
AMOEBA model Polarisable force field with static multipoles up to quadrupoles, implemented in `TINKER`

Speeding-up quantum mechanics simulations



$+ \; 2503\times$

$=$



Kohn–Sham equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \quad \text{for } i \in [\![1 \,..\, N]\!].$$

KOHN–SHAM equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \quad \text{for } i \in [\![1 \,..\, N]\!].$$

After discretisation, molecular system with

- N electrons
- $\mathcal{N}$ basis elements (atomic orbitals)

KOHN–SHAM equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \quad \text{for } i \in [\![1\,..\,N]\!].$$

After discretisation, molecular system with

- N electrons
- $\mathcal{N}$ basis elements (atomic orbitals)

Self-consistent field equations

Find $C \in \mathcal{M}$ and diagonal matrix $E \in \mathbb{R}^{N\times N}$ such that

$$\begin{cases} H_{KS}(D)C = SCE \\ C^\mathsf{T}SC = \mathrm{Id}_N \\ D = CC^\mathsf{T} \end{cases},$$

with $\mathcal{M} := \left\{C \in \mathbb{R}^{\mathcal{N}\times N} \,\middle|\, C^\mathsf{T}SC = \mathrm{Id}_N\right\}$[a]

[a] $S_{\mu\nu} := \int_{\mathbb{R}^3} \chi_\mu\chi_\nu$

KOHN–SHAM equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}), \quad \text{for } i \in [\![1 \,..\, N]\!].$$

After discretisation, molecular system with

- N electrons
- $\mathcal{N}$ basis elements (atomic orbitals)

Self-consistent field equations

Find $C \in \mathcal{M}$ and diagonal matrix $E \in \mathbb{R}^{N \times N}$ such that

$$\begin{cases} H_{KS}(D)C = SCE \\ C^{\mathsf{T}}SC = \mathrm{Id}_N \\ D = CC^{\mathsf{T}} \end{cases},$$

with $\mathcal{M} := \left\{C \in \mathbb{R}^{\mathcal{N} \times N} \mid C^{\mathsf{T}}SC = \mathrm{Id}_N\right\}$[a]

[a] $S_{\mu\nu} := \int_{\mathbb{R}^3} \chi_\mu \chi_\nu$

KOHN–SHAM equations

$$\underbrace{\left(-\frac{1}{2}\Delta + v_{KS}[\rho](\mathbf{r})\right)}_{\mathcal{H}_{KS}[\rho]}\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \quad \text{for } i \in [\![1\,..\,N]\!].$$

After discretisation, molecular system with

- N electrons
- $\mathcal{N}$ basis elements (atomic orbitals)

Self-consistent field equations

Find $C \in \mathcal{M}$ and diagonal matrix $E \in \mathbb{R}^{N\times N}$ such that

$$\begin{cases} H_{KS}(D)C = SCE \\ C^{\mathsf{T}}SC = \mathrm{Id}_N \\ D = CC^{\mathsf{T}} \end{cases},$$

with $\mathcal{M} := \left\{C \in \mathbb{R}^{\mathcal{N}\times N} \mid C^{\mathsf{T}}SC = \mathrm{Id}_N\right\}$[a]

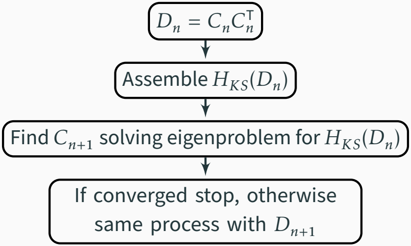
[a] $S_{\mu\nu} := \int_{\mathbb{R}^3} \chi_\mu\chi_\nu$

Initial value C_0, with $C_0^\mathsf{T} S C_0 = \mathrm{Id}_N$ and construct a sequence $(C_n)_{n\in\mathbb{N}}$ solving

$$\begin{cases} H_{KS}(D_n)C_{n+1} = SC_{n+1}\mathsf{E}_{n+1} \\ C_{n+1}^\mathsf{T} S C_{n+1} = \mathrm{Id}_N \\ D_{n+1} = C_{n+1}C_{n+1}^\mathsf{T} \end{cases}$$

Initial value C_0, with $C_0^{\mathsf{T}} S C_0 = \mathrm{Id}_N$ and construct a sequence $(C_n)_{n\in\mathbb{N}}$ solving

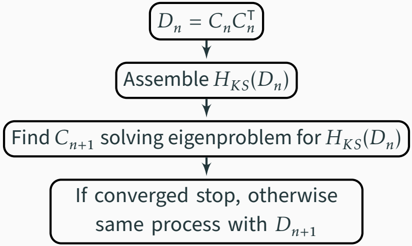
$$\begin{cases} H_{KS}(D_n) C_{n+1} = S C_{n+1} \mathsf{E}_{n+1} \\ C_{n+1}^{\mathsf{T}} S C_{n+1} = \mathrm{Id}_N \\ D_{n+1} = C_{n+1} C_{n+1}^{\mathsf{T}} \end{cases}$$



Fixed-point algorithm for the self-consistent field equations

Initial value C_0, with $C_0^\mathsf{T} S C_0 = \mathrm{Id}_N$ and construct a sequence $(C_n)_{n\in\mathbb{N}}$ solving

$$\begin{cases} H_{KS}(D_n)C_{n+1} = SC_{n+1}\mathsf{E}_{n+1} \\ C_{n+1}^\mathsf{T} S C_{n+1} = \mathrm{Id}_N \\ D_{n+1} = C_{n+1}C_{n+1}^\mathsf{T} \end{cases}$$

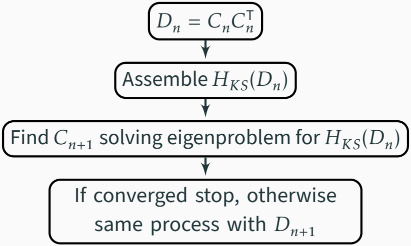


Algorithm stops when residue $|D_{n+1} - D_n|$ smaller than desired threshold

Fixed-point algorithm for the self-consistent field equations

Initial value C_0, with $C_0^\mathsf{T} S C_0 = \mathrm{Id}_N$ and construct a sequence $(C_n)_{n\in\mathbb{N}}$ solving

$$\begin{cases} H_{KS}(D_n)C_{n+1} = SC_{n+1}\mathsf{E}_{n+1} \\ C_{n+1}^\mathsf{T} S C_{n+1} = \mathrm{Id}_N \\ D_{n+1} = C_{n+1}C_{n+1}^\mathsf{T} \end{cases}$$



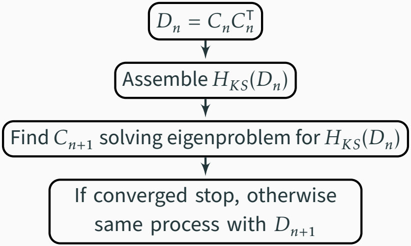
Algorithm stops when residue $|D_{n+1} - D_n|$ smaller than desired threshold

To improve (or even *insure*) convergence, more subtle ways to update the density matrices D_n (damping, DIIS, etc.)…

Fixed-point algorithm for the self-consistent field equations

Initial value C_0, with $C_0^\mathsf{T} S C_0 = \mathrm{Id}_N$ and construct a sequence $(C_n)_{n\in\mathbb{N}}$ solving

$$\begin{cases} H_{KS}(D_n)C_{n+1} = SC_{n+1}\mathsf{E}_{n+1} \\ C_{n+1}^\mathsf{T} S C_{n+1} = \mathrm{Id}_N \\ D_{n+1} = C_{n+1}C_{n+1}^\mathsf{T} \end{cases}$$



Algorithm stops when residue $|D_{n+1} - D_n|$ smaller than desired threshold

To improve (or even *insure*) convergence, more subtle ways to update the density matrices D_n (damping, DIIS, etc.)…

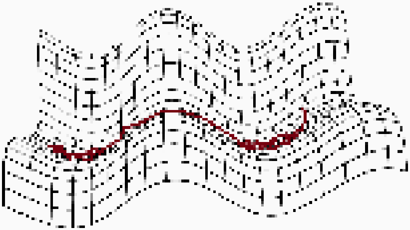
…or provide algorithm with a good initial guess C_0.

- Self-consistent field iterations costly and bottlenecks in density functional theory

- Self-consistent field iterations costly and bottlenecks in density functional theory
- Computations can take months, no small gain

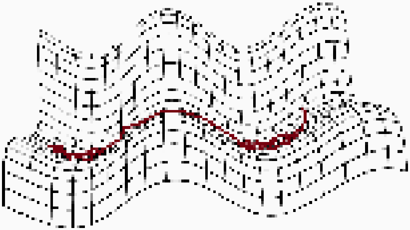
- Self-consistent field iterations costly and bottlenecks in density functional theory
- Computations can take months, no small gain
- With molecular dynamics we expect density matrices to be *close* and want to reuse them from previous steps

- Self-consistent field iterations costly and bottlenecks in density functional theory
- Computations can take months, no small gain
- With molecular dynamics we expect density matrices to be *close* and want to reuse them from previous steps



Simple scheme thanks to LCAO: at time-step t^{k+1}, use previous converged density D_n^k

- Self-consistent field iterations costly and bottlenecks in density functional theory
- Computations can take months, no small gain
- With molecular dynamics we expect density matrices to be *close* and want to reuse them from previous steps



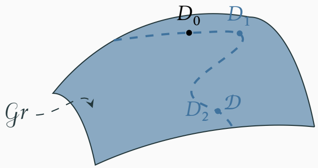
Simple scheme thanks to LCAO: at time-step t^{k+1}, use previous converged density D_n^k

Aim

Fewer self-consistent field iterations on same molecular system using appropriate initial guess

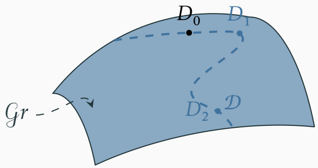
Up to normalisation, density matrices are points on a Grassmannian manifold

$$S^{1/2} D S^{1/2} \in \mathcal{G}r := \left\{ D \in \mathbb{R}^{\mathcal{N}\times N} \,\middle|\, D = D^{\mathsf{T}}, D^2 = D, \mathrm{Tr}(D) = N \right\}.$$



Up to normalisation, density matrices are points on a Grassmannian manifold

$$S^{1/2} D S^{1/2} \in \mathcal{G}r := \left\{ D \in \mathbb{R}^{\mathcal{N} \times N} \,\middle|\, D = D^\mathsf{T}, D^2 = D, \mathrm{Tr}(D) = N \right\}.$$

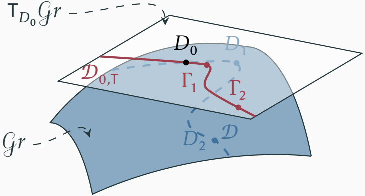


Classical linear combinations extrapolation scheme: objects not on manifold!

$$\sum c_i D_i \notin \mathcal{G}r$$

Up to normalisation, density matrices are points on a Grassmannian manifold

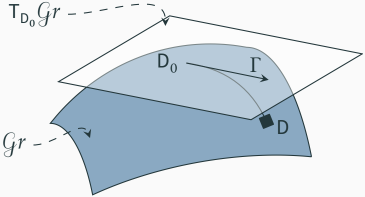
$$S^{1/2}DS^{1/2} \in \mathcal{G}r := \left\{ D \in \mathbb{R}^{\mathcal{N}\times N} \,\middle|\, D = D^\mathsf{T}, D^2 = D, \mathrm{Tr}(D) = N \right\}.$$



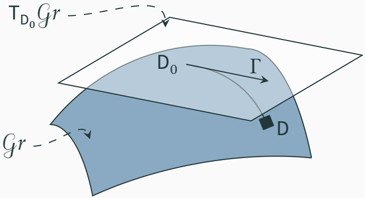
Classical linear combinations extrapolation scheme: objects not on manifold!

$$\sum c_i D_i \notin \mathcal{G}r$$

Let's use the tangent space!



Local diffeomorphism between manifold $\mathcal{G}r$ and affine space, the tangent space, cheap to compute

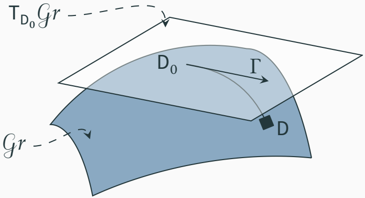


Local diffeomorphism between manifold $\mathcal{G}r$ and affine space, the tangent space, cheap to compute

- Exponential function

$$\begin{aligned}\mathrm{Exp}_{\mathcal{G}r,0} : \mathsf{T}_{D_0}\mathcal{G}r &\to \mathcal{G}r \\ \Gamma &\mapsto CC^{\mathsf{T}},\end{aligned}$$

where $C = [C_0 V \cos(\Sigma) + U \sin(\Sigma)]V^{\mathsf{T}}$, with $\Gamma = U\Sigma V^{\mathsf{T}}$ the singular value decomposition of Γ



Local diffeomorphism between manifold $\mathcal{G}r$ and affine space, the tangent space, cheap to compute

- Exponential function

$$\begin{aligned}\mathrm{Exp}_{\mathcal{G}r,0} : \mathsf{T}_{D_0}\mathcal{G}r &\to \mathcal{G}r \\ \Gamma &\mapsto CC^{\mathsf{T}},\end{aligned}$$

  where $C = [C_0 V \cos(\Sigma) + U \sin(\Sigma)]V^{\mathsf{T}}$, with $\Gamma = U\Sigma V^{\mathsf{T}}$ the singular value decomposition of Γ

- Logarithm function

$$\begin{aligned}\mathrm{Log}_{\mathcal{G}r,0} : \mathcal{G}r &\to \mathsf{T}_{D_0}\mathcal{G}r \\ D &\mapsto \mathrm{Log}_{\mathcal{G}r,0}(D)\end{aligned}$$

  defined similarly

Method's constraints

- Providing a *good* initial guess

Method's constraints

- Providing a *good* initial guess
- Useful for large multiphysics molecular systems

Method's constraints

- Providing a *good* initial guess
- Useful for large multiphysics molecular systems
- Negligible overhead

Method's constraints

- Providing a *good* initial guess
- Useful for large multiphysics molecular systems
- Negligible overhead
- Flexible and as simple as possible for end users

The G-Ext method — Molecular descriptors

Each time-step, nuclei's positions for *free*; we want the corresponding density matrices $D_{\mathbf{R}}$

Molecular descriptors

$$\begin{aligned} \mathbb{R}^{3M} &\to \mathcal{G}r(N, \mathcal{N}) \\ \mathbf{R} &\mapsto D_{\mathbf{R}} \end{aligned}$$

The G-Ext method — Molecular descriptors

Each time-step, nuclei's positions for *free*; we want the corresponding density matrices $D_{\mathbf{R}}$

We want linear combinations on tangent space, and density matrices as initial guesses

Molecular descriptors

$$\begin{aligned} \mathbb{R}^{3M} &\to \mathsf{T}_{D_{\mathbf{0}}}\mathcal{G}r \to \mathcal{G}r(N,\mathcal{N}) \\ \mathbf{R} &\mapsto \quad \Gamma_{\mathbf{R}} \quad \mapsto D_{\mathbf{R}} = \mathrm{Exp}_{\mathcal{G}r,0}(\Gamma_{\mathbf{R}}) \end{aligned}$$

$$\mathbf{R} \mapsto \Gamma_{\mathrm{app}}(\mathbf{R}) = \sum_{i=1}^{N_t} c_{\mathbf{R},i}\,\Gamma_{\mathbf{R}(t_i)} \in \mathsf{T}_{D_{\mathbf{0}}}\mathcal{G}r$$

The G-Ext method — Molecular descriptors

Each time-step, nuclei's positions for *free*; we want the corresponding density matrices $D_{\mathbf{R}}$

We want linear combinations on tangent space, and density matrices as initial guesses

Map should take advantages of properties of the density matrices (*e.g.*, translation invariance)

Molecular descriptors $d_{\mathbf{R}}$

$$\begin{aligned} \mathbb{R}^{3M} \to \mathcal{M} \to \mathsf{T}_{D_0}\mathcal{G}r \to \mathcal{G}r(N,\mathcal{N}) \\ \mathbf{R} \mapsto d_{\mathbf{R}} \mapsto \quad \Gamma_{\mathbf{R}} \quad \mapsto D_{\mathbf{R}} = \mathrm{Exp}_{\mathcal{G}r,0}(\Gamma_{\mathbf{R}}) \end{aligned}$$

Notion from machine learning

The G-Ext method — Molecular descriptors

Each time-step, nuclei's positions for *free*; we want the corresponding density matrices $D_{\mathbf{R}}$

We want linear combinations on tangent space, and density matrices as initial guesses

Map should take advantages of properties of the density matrices (*e.g.*, translation invariance)

Molecular descriptors $d_{\mathbf{R}}$

$$\begin{aligned} \mathbb{R}^{3M} \to \mathcal{M} \to \mathsf{T}_{D_0}\mathcal{G}r \to \mathcal{G}r(N,\mathcal{N}) \\ \mathbf{R} \mapsto d_{\mathbf{R}} \mapsto \quad \Gamma_{\mathbf{R}} \quad \mapsto D_{\mathbf{R}} = \mathrm{Exp}_{\mathcal{G}r,0}(\Gamma_{\mathbf{R}}) \end{aligned}$$

Coulomb matrix at time-step t_k

$$(d_{\mathbf{R}})_{ij} = \begin{cases} 0.5 Z_i^{2.4} & \text{if } i = j \\ \dfrac{Z_i Z_j}{|\mathbf{R}^i(t_k) - \mathbf{R}^j(t_k)|} & \text{otherwise} \end{cases}$$

The G-Ext method — Molecular descriptors

Each time-step, nuclei's positions for *free*; we want the corresponding density matrices $D_{\mathbf{R}}$

We want linear combinations on tangent space, and density matrices as initial guesses

Map should take advantages of properties of the density matrices (*e.g.*, translation invariance)

Molecular descriptors $d_{\mathbf{R}}$

$$\begin{array}{ccccccc} \mathbb{R}^{3M} & \to & \mathcal{M} & \to & \mathrm{T}_{D_0}\mathcal{G}r & \to & \mathcal{G}r(N,\mathcal{N}) \\ \mathbf{R} & \mapsto & d_{\mathbf{R}} & \mapsto & \Gamma_{\mathbf{R}} & \mapsto & D_{\mathbf{R}} = \mathrm{Exp}_{\mathcal{G}r,0}(\Gamma_{\mathbf{R}}) \end{array}$$

Coulomb matrix at time-step t_k

$$(d_{\mathbf{R}})_{ij} = \begin{cases} 0.5 Z_i^{2.4} & \text{if } i = j \\ \dfrac{Z_i Z_j}{|\mathbf{R}^i(t_k) - \mathbf{R}^j(t_k)|} & \text{otherwise} \end{cases}$$

G-Ext method we developed

- Use of molecular descriptors
- Linear approximation on tangent space
- Retraction on manifold $\mathcal{G}r$

The G-Ext method — Least-squares

We want coefficients $c_{\mathbf{R},i}$ that can approximate the density matrices on the tangent space

$$\mathbf{R} \mapsto \Gamma_{\mathrm{app}}(\mathbf{R}) = \sum_{i=1}^{N_t} c_{\mathbf{R},i}\, \Gamma_{\mathbf{R}(t_i)} \in \mathsf{T}_{D_{\mathbf{0}}}\mathcal{G}r$$

The G-Ext method — Least-squares

We want coefficients $c_{\mathbf{R},i}$ that can approximate the density matrices on the tangent space

$$\mathbf{R} \mapsto \Gamma_{\mathrm{app}}(\mathbf{R}) = \sum_{i=1}^{N_t} c_{\mathbf{R},i} \, \Gamma_{\mathbf{R}(t_i)} \in \mathsf{T}_{D_0} \mathcal{G}r$$

Least-squares method to solve

$$\min_{c_{\mathbf{R}} \in \mathbb{R}^{N_t}} \left| d_{\mathbf{R}} - \sum_{i=1}^{N_t} c_{\mathbf{R},i} d_{\mathbf{R}(t_i)} \right|^2$$

The G-Ext method — Least-squares

We want coefficients $c_{\mathbf{R},i}$ that can approximate the density matrices on the tangent space

$$\mathbf{R} \mapsto \Gamma_{\mathrm{app}}(\mathbf{R}) = \sum_{i=1}^{N_t} c_{\mathbf{R},i}\, \Gamma_{\mathbf{R}(t_i)} \in \mathsf{T}_{D_0}\mathcal{G}r$$

Least-squares method to solve

$$\min_{c_{\mathbf{R}} \in \mathbb{R}^{N_t}} \left| d_{\mathbf{R}} - \sum_{i=1}^{N_t} c_{\mathbf{R},i} d_{\mathbf{R}(t_i)} \right|^2$$

We use the same coefficients for density matrices on the tangent space and use the retracted density matrix

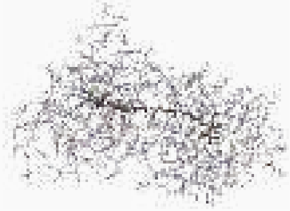
$$D_{\mathrm{app}}(\mathbf{R}) = \mathrm{Exp}_{\mathcal{G}r,0}\left(\sum_{i=1}^{N_t} c_{\mathbf{R},i}\, \Gamma_{\mathbf{R}(t_i)} \right)$$

as an initial guess to the self-consistent field algorithm

Molecular system	Quantum atoms	Classical atoms	$\mathcal{N}$
OCP	129	4915	1038
APPA	31	16 449	309
DMABN	21	6843	185
3HF	28	15 018	290

OCP: orange carotenoid protein

Molecular system	Quantum atoms	Classical atoms	$\mathcal{N}$
OCP	129	4915	1038
APPA	31	16 449	309
DMABN	21	6843	185
3HF	28	15 018	290



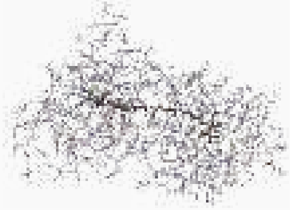
OCP: orange carotenoid protein

- 10^{-5} root-mean-square
- 1 ps dynamics
- 0.5 fs time-step
- ref. method

	OCP		DMABN		APPA		3HF	
Method	Average	σ	Average	σ	Average	σ	Average	σ
XLBO	3.82	0.66	3.98	0.16	3.00	0.03	4.00	0.14
XLBO/MW	2.95	0.31	3.76	0.56	3.00	0.34	3.96	0.31
G-Ext(3)	2.57	0.84	3.54	0.78	2.95	0.50	3.09	0.41
G-Ext(4)	2.48	0.88	3.14	0.62	2.51	0.50	3.25	0.68
G-Ext(5)	2.25	0.96	3.23	0.75	2.51	0.50	3.30	0.72
G-Ext(6)	2.20	0.96	2.99	0.02	2.51	0.50	3.14	0.56

Results — Performances

Molecular system	Quantum atoms	Classical atoms	$\mathcal{N}$
OCP	129	4915	1038
APPA	31	16 449	309
DMABN	21	6843	185
3HF	28	15 018	290



OCP: orange carotenoid protein

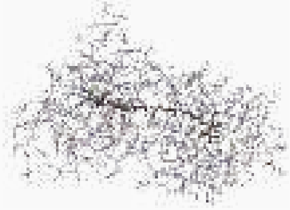
- 10^{-5} root-mean-square
- 1 ps dynamics
- 0.5 fs time-step
- ref. method

	OCP		DMABN		APPA		3HF	
Method	Average	σ	Average	σ	Average	σ	Average	σ
XLBO	3.82	0.66	3.98	0.16	3.00	0.03	4.00	0.14
XLBO/MW	2.95	0.31	3.76	0.56	3.00	0.34	3.96	0.31
G-Ext(3)	2.57	0.84	3.54	0.78	2.95	0.50	3.09	0.41
G-Ext(4)	2.48	0.88	3.14	0.62	2.51	0.50	3.25	0.68
G-Ext(5)	2.25	0.96	3.23	0.75	2.51	0.50	3.30	0.72
G-Ext(6)	2.20	0.96	2.99	0.02	2.51	0.50	3.14	0.56

Decrease in self-consistent field iterations

Results — Performances

Molecular system	Quantum atoms	Classical atoms	$\mathcal{N}$
OCP	129	4915	1038
APPA	31	16 449	309
DMABN	21	6843	185
3HF	28	15 018	290



OCP: orange carotenoid protein

- 10^{-5} root-mean-square
- 1 ps dynamics
- 0.5 fs time-step
- ref. method

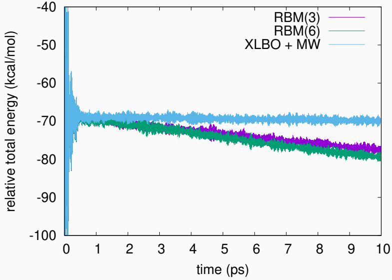
	OCP		DMABN		APPA		3HF	
Method	Average	σ	Average	σ	Average	σ	Average	σ
XLBO	3.82	0.66	3.98	0.16	3.00	0.03	4.00	0.14
XLBO/MW	2.95	0.31	3.76	0.56	3.00	0.34	3.96	0.31
G-Ext(3)	2.57	0.84	3.54	0.78	2.95	0.50	3.09	0.41
G-Ext(4)	2.48	0.88	3.14	0.62	2.51	0.50	3.25	0.68
G-Ext(5)	2.25	0.96	3.23	0.75	2.51	0.50	3.30	0.72
G-Ext(6)	2.20	0.96	2.99	0.02	2.51	0.50	3.14	0.56

Decrease in self-consistent field iterations; adds up with long molecular dynamics!

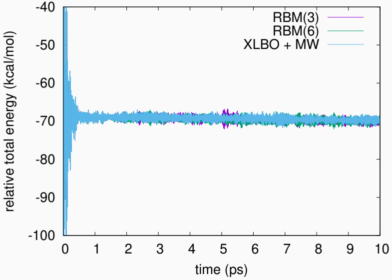
Results — Total energy conservation

- 3HF system
- 10 ps dynamics
- 0.5 fs time-step
- +505 000 kcal/mol shift
- 10^{-5} and 10^{-7} root-mean-square

Relaxed parameters (10^{-5})



Conservative parameters (10^{-7})



Perspectives

Conclusion

Lower number of self-consistent field iterations

Perspectives

Conclusion

Lower number of self-consistent field iterations

Future work

- Time reversibility

Perspectives

Conclusion

Lower number of self-consistent field iterations

Future work

- Time reversibility
- Geometry optimisation

Perspectives

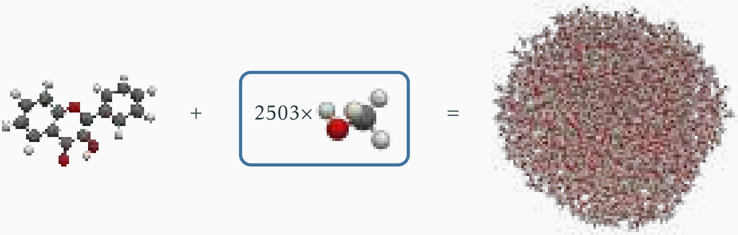
Conclusion

Lower number of self-consistent field iterations

Future work

- Time reversibility
- Geometry optimisation
- Bypass self-consistent field algorithm (*e.g.*, empirical interpolation method for Fock matrices, minimisation on tangent space)

Speeding-up classical mechanics simulations



Molecular system with N nuclei of point charge M_{0i} at positions $\mathbf{r}_i$

N-body problem of computing the electrostatic energy

$$\mathcal{E}_{\mathrm{elec}} := \frac{1}{2} \sum_{1 \leqslant i \neq j \leqslant N} \frac{\mathrm{M}_{0i}\mathrm{M}_{0j}}{|\mathbf{r}_j - \mathbf{r}_i|}.$$

Bottleneck, scales as $\mathcal{O}(N^2)$

Molecular system with N nuclei of point charge M_{0i} at positions $\mathbf{r}_i$

N-body problem of computing the electrostatic energy

$$\mathcal{E}_{\text{elec}} := \frac{1}{2} \sum_{1 \leqslant i \neq j \leqslant N} \frac{\mathrm{M}_{0i}\mathrm{M}_{0j}}{|\mathbf{r}_j - \mathbf{r}_i|}.$$

Bottleneck, scales as $\mathcal{O}(N^2)$

For AMOEBA force field, generalised to (static and non-static) multipoles $\mathbf{M}_n$ with derivation operators $\mathbf{D}_i^n$

$$\hat{\mathbf{L}}_i := \mathrm{M}_{0i} + \mathbf{M}_{1i} \cdot \mathbf{D}_i + \mathbf{M}_{2i} \cdot \mathbf{D}_i^2$$

Molecular system with N nuclei of point charge M_{0i} at positions $\mathbf{r}_i$

N-body problem of computing the electrostatic energy

$$\mathcal{E}_{\text{elec}} := \frac{1}{2} \sum_{1 \leqslant i \neq j \leqslant N} \frac{\mathrm{M}_{0i}\mathrm{M}_{0j}}{|\mathbf{r}_j - \mathbf{r}_i|}.$$

Bottleneck, scales as $\mathcal{O}(N^2)$

For AMOEBA force field, generalised to (static and non-static) multipoles $\mathbf{M}_n$ with derivation operators $\mathbf{D}_i^n$

$$\hat{\mathbf{L}}_i := \mathrm{M}_{0i} + \mathbf{M}_{1i} \cdot \mathbf{D}_i + \mathbf{M}_{2i} \cdot \mathbf{D}_i^2$$

So

$$\mathcal{E}_{\text{elec}} := \frac{1}{2} \sum_{1 \leqslant i \neq j \leqslant N} \hat{\mathbf{L}}_i \hat{\mathbf{L}}_j \left(\frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} \right).$$

Molecular system with N nuclei of point charge M_{0i} at positions $\mathbf{r}_i$

N-body problem of computing the electrostatic energy

$$\mathcal{E}_{\mathrm{elec}} := \frac{1}{2} \sum_{1 \leqslant i \neq j \leqslant N} \frac{\mathrm{M}_{0i}\mathrm{M}_{0j}}{|\mathbf{r}_j - \mathbf{r}_i|}.$$

Bottleneck, scales as $\mathcal{O}(N^2)$

For AMOEBA force field, generalised to (static and non-static) multipoles $\mathbf{M}_n$ with derivation operators $\mathbf{D}_i^n$

$$\hat{\mathbf{L}}_i := \mathrm{M}_{0i} + \mathbf{M}_{1i} \cdot \mathbf{D}_i + \mathbf{M}_{2i} \cdot \mathbf{D}_i^2$$

So

$$\mathcal{E}_{\mathrm{elec}} := \frac{1}{2} \sum_{1 \leqslant i \neq j \leqslant N} \hat{\mathbf{L}}_i \hat{\mathbf{L}}_j \left(\frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} \right).$$

Cuttoff methods not accurate for molecular dynamics because *slow* decrease of the inverse function

Smooth particle mesh EWALD

- Reference method
- Periodic problem
- Smooth function splitting the potential

Smooth particle mesh EWALD

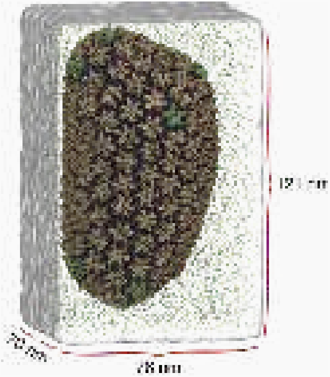
- Reference method
- Periodic problem
- Smooth function splitting the potential

With fast FOURIER transforms: $\mathcal{O}(N \log N)$ instead of $\mathcal{O}(N^2)$

Smooth particle mesh EWALD

- Reference method
- Periodic problem
- Smooth function splitting the potential

With fast FOURIER transforms: $\mathcal{O}(N \log N)$ instead of $\mathcal{O}(N^2)$



$>$ 64 million atoms, months of computations on thousands of GPUs

Juan R. Perilla and Klaus Schulten. "Physical Properties of the HIV-1 Capsid from All-Atom Molecular Dynamics Simulations". In: *Nature Communications* 8 (July 19, 2017), p. 15959

Fast multipole method

- Linear computational complexity hierarchical method

Fast multipole method

- Linear computational complexity hierarchical method
- Common in physics

Fast multipole method

- Linear computational complexity hierarchical method
- Common in physics; less so in chemistry

Fast multipole method

- Linear computational complexity hierarchical method
- Common in physics; less so in chemistry
- *E.g.*, `ScalFMM` solves N-body problem for point charges

Implementation status

Modification of ScalFMM[1] and TINKER for AMOEBA

[1] P2M, L2P and P2P operators

Implementation status

Modification of ScalFMM[1] and TINKER for AMOEBA

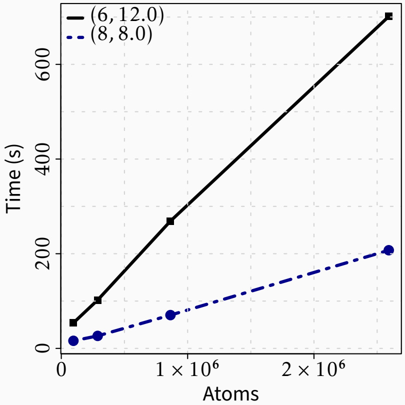
Terms	Potential	Energy	Forces
Charges	•	•	•
Dipoles	✓	✓	✓
Quadrupoles	✓	✓	✓
General multipoles	✓	✗	✗

• present ✓ implemented ✗ not implemented

[1] P2M, L2P and P2P operators

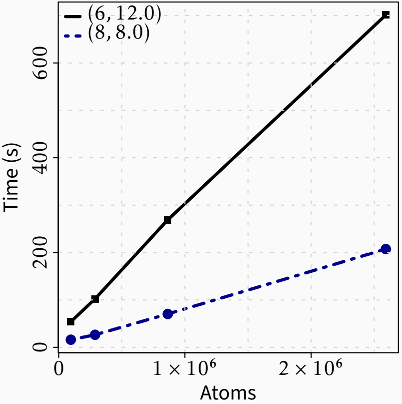
Results — Linear scaling

- Single time-step
- 10^{-5} for polarisation
- 216 cores
- Total execution time of `TINKER(AMOEBA)-ScalFMM`
- Conservative parameters
- Relaxed parameters



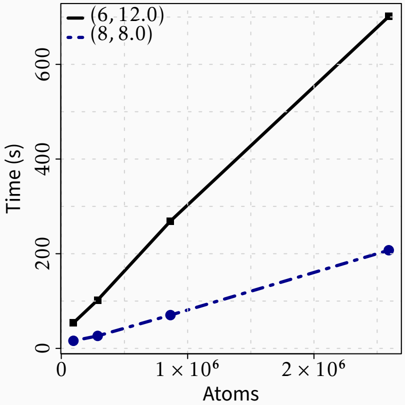
Results — Linear scaling

- Single time-step
- 10^{-5} for polarisation
- 216 cores
- Total execution time of `TINKER(AMOEBA)-ScalFMM`
- Conservative parameters
- Relaxed parameters



Results — Linear scaling

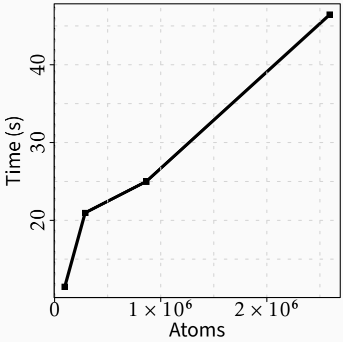
- Single time-step
- 10^{-5} for polarisation
- 216 cores
- Total execution time of TINKER(AMOEBA)-ScalFMM
- Conservative parameters
- Relaxed parameters



(but expected)

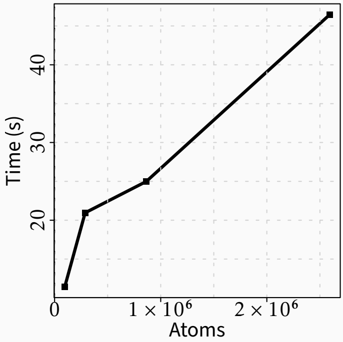
Results — Weak scalability

- No polarisation
- Triple system sizes along cores
- Relaxed parameters
- Total execution time of TINKER(AMOEBA)-ScalFMM



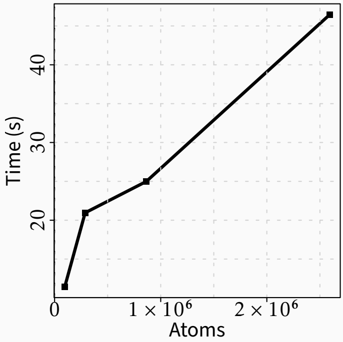
Results — Weak scalability

- No polarisation
- Triple system sizes along cores
- Relaxed parameters
- Total execution time of TINKER(AMOEBA)-ScalFMM



Results — Weak scalability

- No polarisation
- Triple system sizes along cores
- Relaxed parameters
- Total execution time of `TINKER(AMOEBA)-ScalFMM`



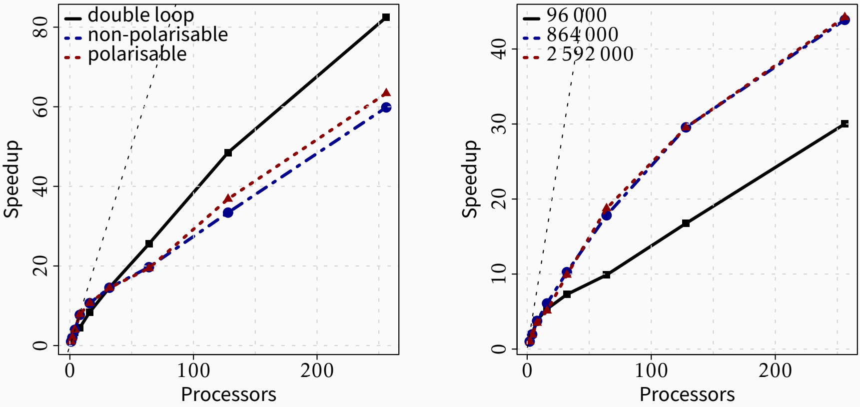
(but expected)

Results — Strong scalability

- Single time-step
- Relaxed set of parameters

Left 96 000 atoms, double loop with `TINKER`

Right Scalability starts at four cores

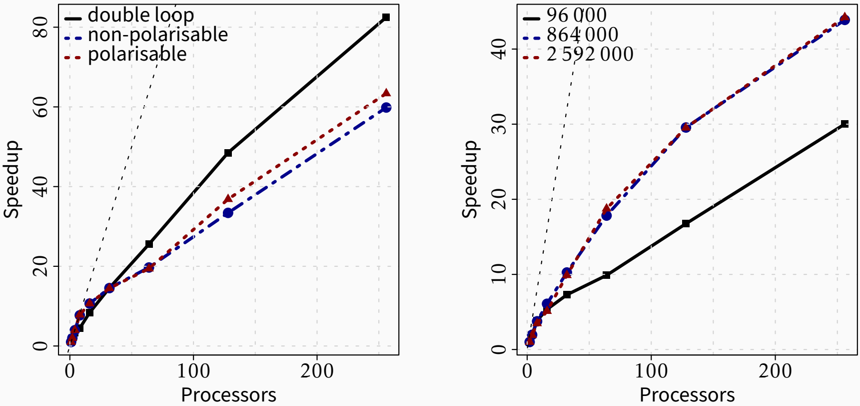


Results — Strong scalability

- Single time-step
- Relaxed set of parameters

Left 96 000 atoms, double loop with `TINKER`

Right Scalability starts at four cores

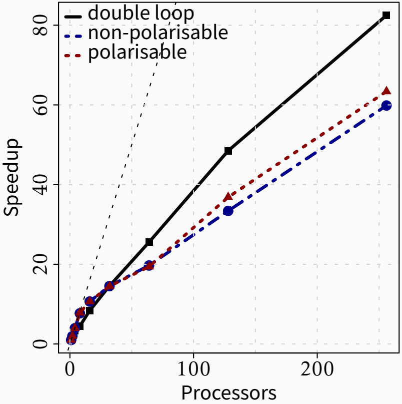


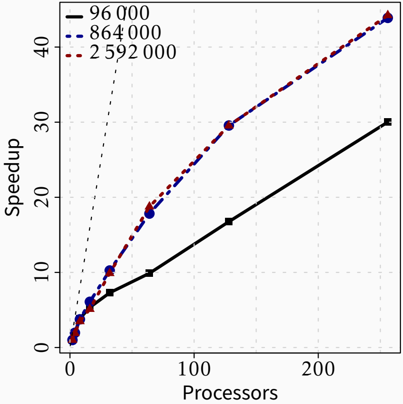
Results — Strong scalability

- Single time-step
- Relaxed set of parameters

Left 96 000 atoms, double loop with `TINKER`

Right Scalability starts at four cores

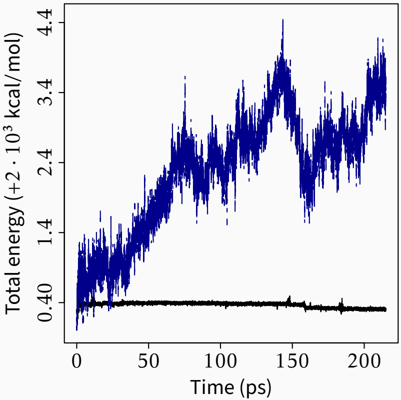




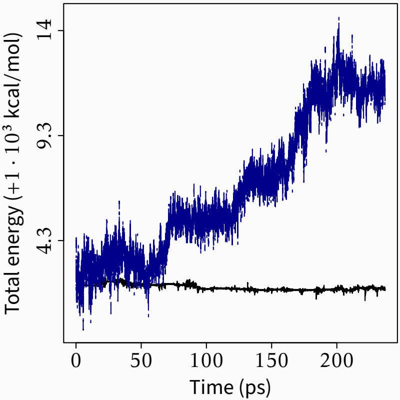
(but not a deal-breaker)

- DHFR system (2489 atoms)
- 0.10 fs time-steps
- Conservative parameters
- Relaxed parameters

Point charges



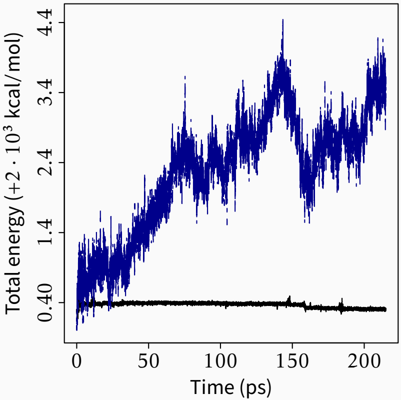
Static multipoles up to quadrupoles



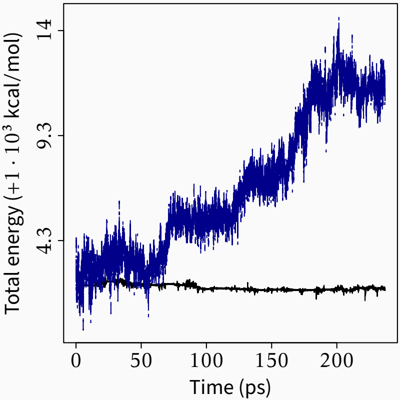
Results — Energy conservation of non-solvated DHFR

- DHFR system (2489 atoms)
- 0.10 fs time-steps
- Conservative parameters
- Relaxed parameters

Point charges



Static multipoles up to quadrupoles

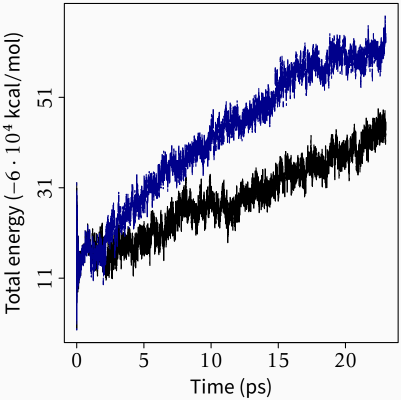


Energy conserved

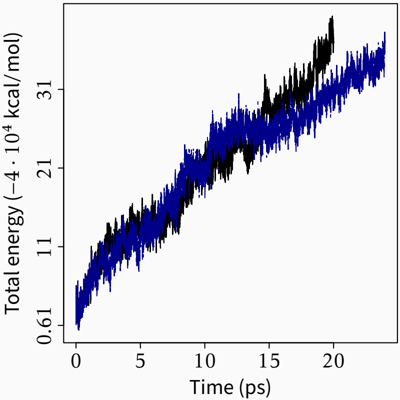
Results — Energy conservation of solvated DHFR

- Solvated DHFR system (23 558 atoms)
- 0.25 fs time-steps
- 10^{-8} for polarisation
- 10^{-5} for polarisation

Conservative parameters



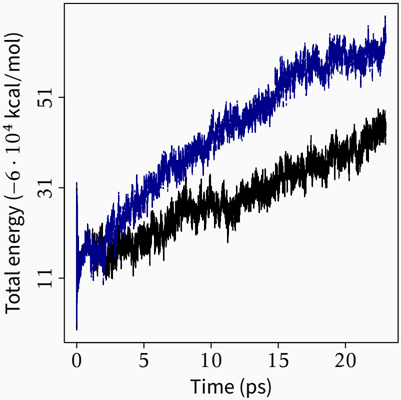
Relaxed parameters



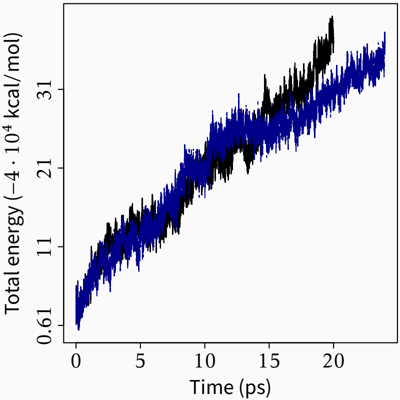
Results — Energy conservation of solvated DHFR

- Solvated DHFR system (23 558 atoms)
- 0.25 fs time-steps
- 10^{-8} for polarisation
- 10^{-5} for polarisation

Conservative parameters



Relaxed parameters

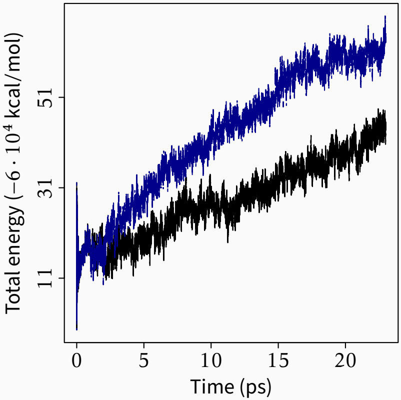


Energy drift

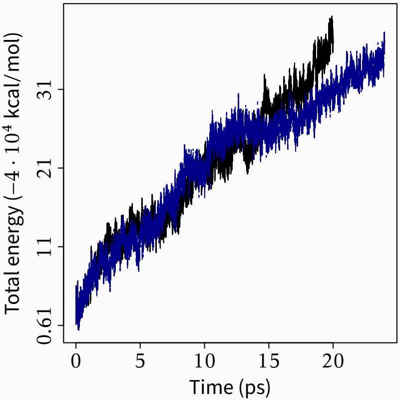
Results — Energy conservation of solvated DHFR

- Solvated DHFR system (23 558 atoms)
- 0.25 fs time-steps
- 10^{-8} for polarisation
- 10^{-5} for polarisation

Conservative parameters



Relaxed parameters



Energy drift (needs investigation)

Perspectives

Conclusion

- Linear scaling for non-periodic systems

Perspectives

Conclusion

- Linear scaling for non-periodic systems
- Straightforward and high-performance implementation

Perspectives

Conclusion

- Linear scaling for non-periodic systems
- Straightforward and high-performance implementation

Future work

- Adaptive grid

Perspectives

Conclusion

- Linear scaling for non-periodic systems
- Straightforward and high-performance implementation

Future work

- Adaptive grid
- Periodic case for systematic comparisons to smooth particle mesh EWALD

Perspectives

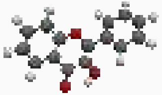
Conclusion

- Linear scaling for non-periodic systems
- Straightforward and high-performance implementation

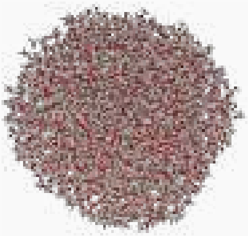
Future work

- Adaptive grid
- Periodic case for systematic comparisons to smooth particle mesh EWALD
- Real-case tests

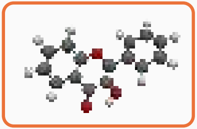
Conclusion



+ 2503×  =



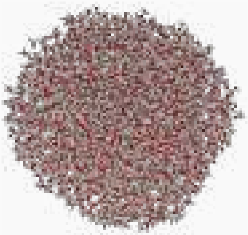
Conclusion



$+$

$2503\times$ 

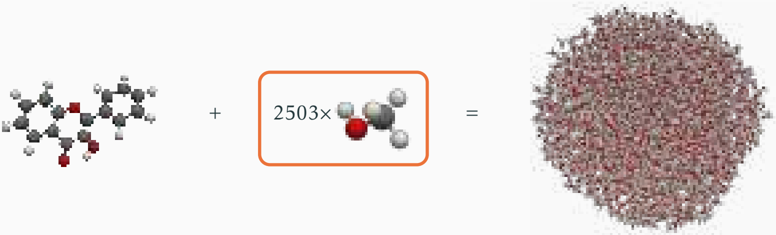
$=$



Major results

Decrease in time for densities of quantum atoms

Conclusion



Major results

Decrease in time for densities of quantum atoms and model classical atoms, solvent in particular, with linear complexity on supercomputers

Conclusion

+ 2503×

=

Major results

Decrease in time for densities of quantum atoms and model classical atoms, solvent in particular, with linear complexity on supercomputers
⇒ Allows for longer multiphysics QM/MM(AMOEBA) simulations, bigger systems or more solvent for more accurate physics

Thank you for your attention!

Publications

Lagardère, Louis et al. "Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: II. Toward Massively Parallel Computations Using Smooth Particle Mesh Ewald". In: *Journal of Chemical Theory and Computation* 11.6 (June 9, 2015), pp. 2589–2599.

Lindgren, Eric B. et al. "An Integral Equation Approach to Calculate Electrostatic Interactions in Many-Body Dielectric Systems". In: *Journal of Computational Physics* 371 (Oct. 15, 2018), pp. 712–731.

Loco, Daniele et al. "A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations". In: *Journal of Chemical Theory and Computation* 12.8 (Aug. 9, 2016), pp. 3654–3661.

Narth, Christophe et al. "Scalable Improvement of SPME Multipolar Electrostatics in Anisotropic Polarizable Molecular Mechanics Using a General Short-Range Penetration Correction up to Quadrupoles". In: *Journal of Computational Chemistry* 37.5 (2016), pp. 494–506.

Polack, Étienne, Yvon Maday, and Andreas Savin. *FLEIM: A Stable, Accurate and Robust Extrapolation Method at Infinity for Computing the Ground State of Electronic Hamiltonians*. Dec. 24, 2021. URL: `http://arxiv.org/abs/2112.13139`.

Polack, Étienne et al. "An Approximation Strategy to Compute Accurate Initial Density Matrices for Repeated Self-Consistent Field Calculations at Different Geometries". In: *Molecular Physics* 118.19-20 (Oct. 17, 2020), e1779834.

Polack, Étienne et al. "Grassmann Extrapolation of Density Matrices for Born–Oppenheimer Molecular Dynamics". In: *Journal of Chemical Theory and Computation* 17.11 (Nov. 9, 2021), pp. 6965–6973.