FrAMonC

Fragment-based Ab Initio Monte Carlo simulations of structural and thermodynamic properties of amorphous molecular materials

Ctirad Červinka

Department of Physical Chemistry

University of Chemistry and Technology, Prague, Czechia



UNIVERSITY OF
CHEMISTRY AND TECHNOLOGY
PRAGUE

Motivation

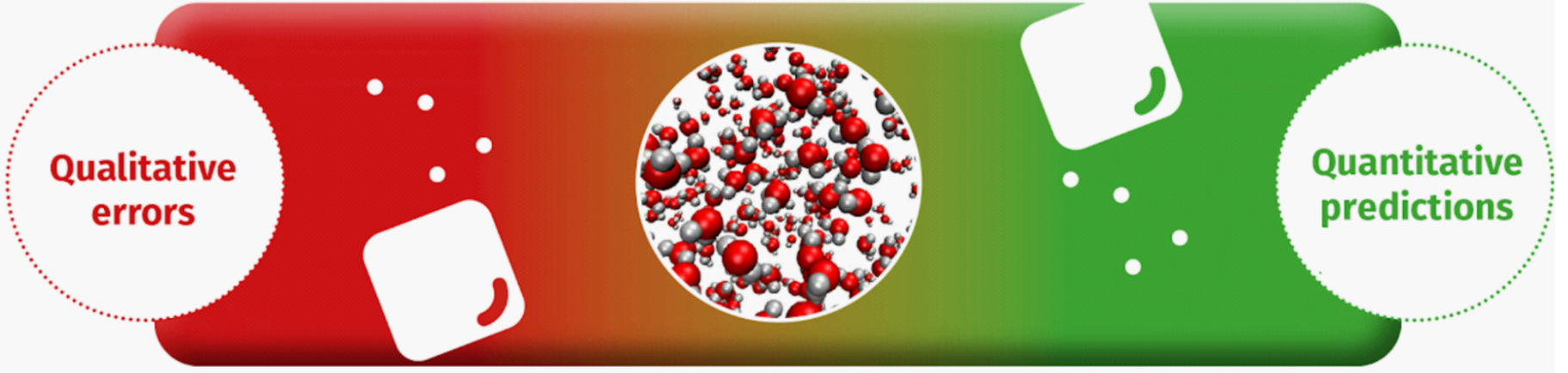
- Why to focus on the **amorphous** state of **molecular materials**?
  - **Melting temperature** cannot be predicted without any knowledge about the liquid
  - **Solvents** are ubiquitous and solvents matter in **crystallization**
  - **Molecular glasses** very beneficial for pharmaceuticals or inevitable for semi-conductors

- **Ab initio** simulations for **liquids** and **glasses** long considered **unfeasible**
  - *Post-DFT* or *wave function* methods are meant by the *ab initio*
  - Exploit the know-how developed for crystals and expand it
  - At first rely on **physical principles** and not solely on data-based approaches

⇒ Achieve exciting **methodology development** & extend the computational state of the art

Goals

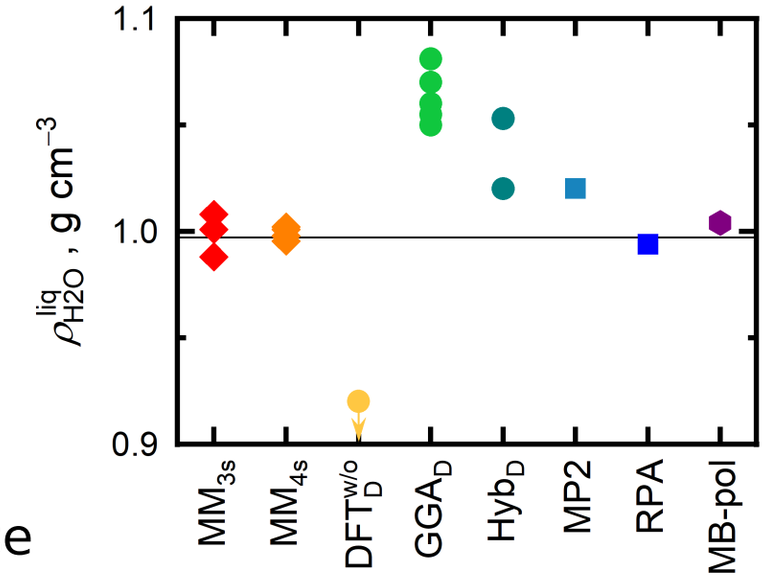
- Develop a truly **ab intio** simulation method for **amorphous molecular materials**
  - Current ab initio methods applicable to molecules, gases and crystals



	Past	Present	FrAMonC Aims
Simulations	Classical	DFT	Ab initio
Empiric parameters	Hundreds	Dozens	Units
Accuracy	Qualitative 10 kJ/mol	Chemical 4 kJ/mol	Sub-chemical <1 kJ/mol
First-principles applicability	Molecules	Crystals	Liquids

Challenges

- **Non-covalent interactions** govern properties of molecular materials
  - Need for **high-level QM** theories due to a delicate NCI interplay
  - **Slow convergence** of bulk properties with respect to QM theory



- Things to consider when dealing with the **amorphous state**
  - **Lack** of any long-range **symmetry** in the amorphous state
  - Permanent **time-evolution** of molecular ensembles

- Thermodynamic properties largely sensitive to computational noise
  - **Large** enough **ensembles** to suppress computational noise
  - **Phase equilibria** governed by **Gibbs energies** of phases

credit: Červinka, J. Chem. Theory Comput., 2025, Just accepted.

Computational methodology I

- **Minimize** the **empiric inputs** for molecular simulations
  - Considering initial absence of any machine-learned potential

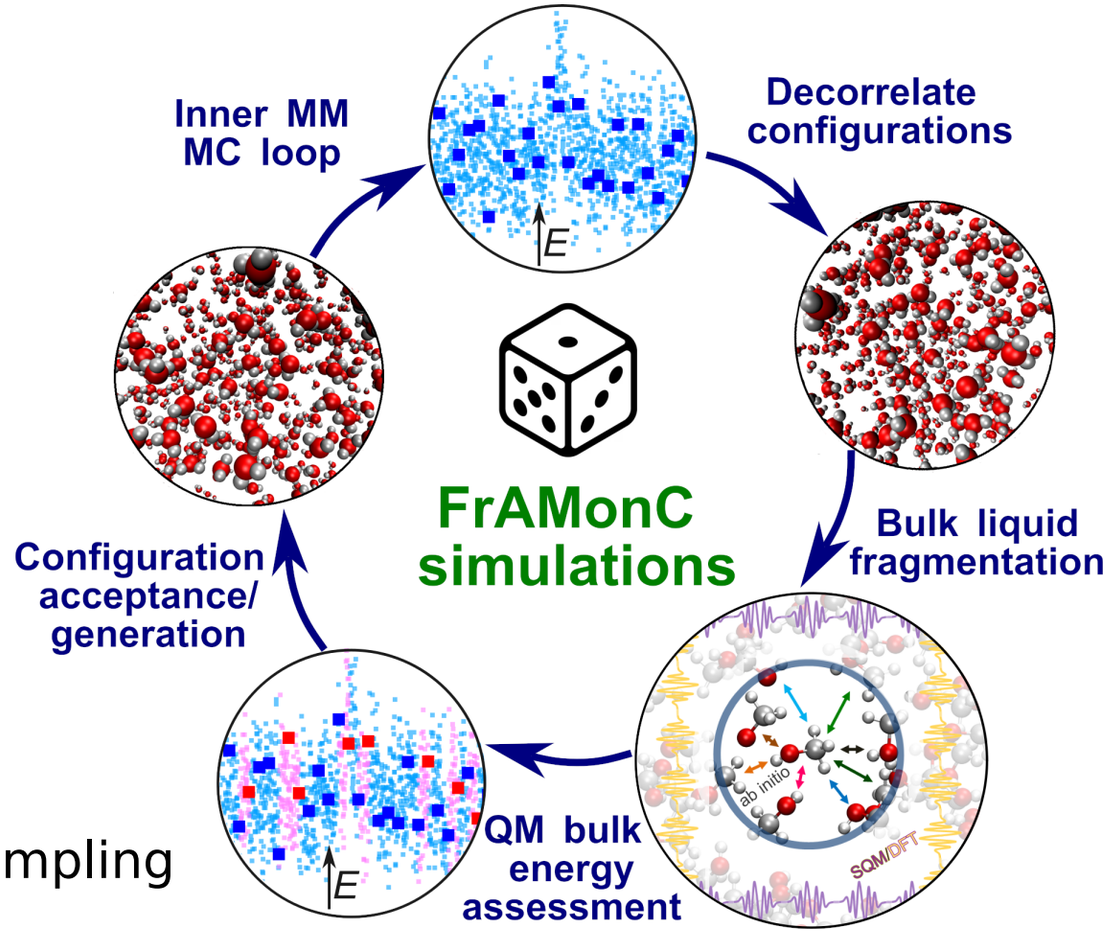
  ⇒ Use **ab initio potentials** and direct sampling

- Ab initio methods often **lack analytic gradients**
  - Impractical to use ab initio molecular dynamics

  ⇒ Rely on **Monte Carlo** simulations

- Many trial configurations rejected in Monte Carlo
  - Costly ab initio energy evaluations of rejected configurations *in vain*

  ⇒ Employ **multi-potential nested** Monte Carlo sampling

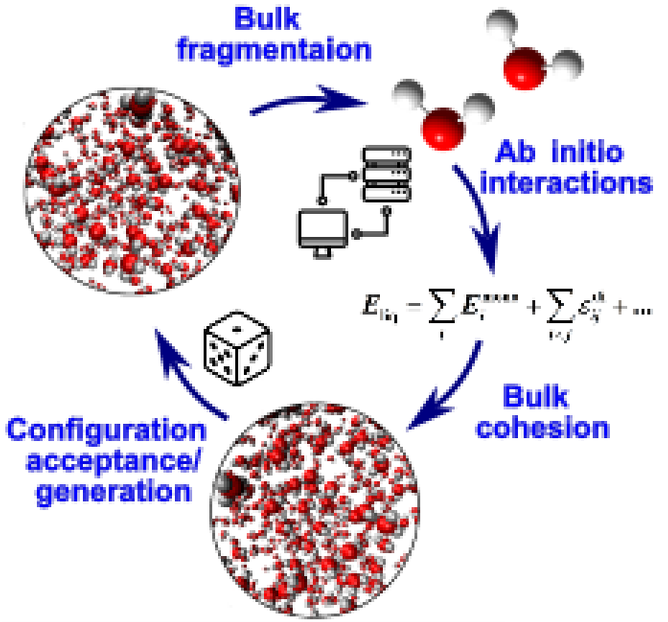


credit: Červinka, J. Chem. Theory Comput., 2025, 21, 9868.

Computational methodology II

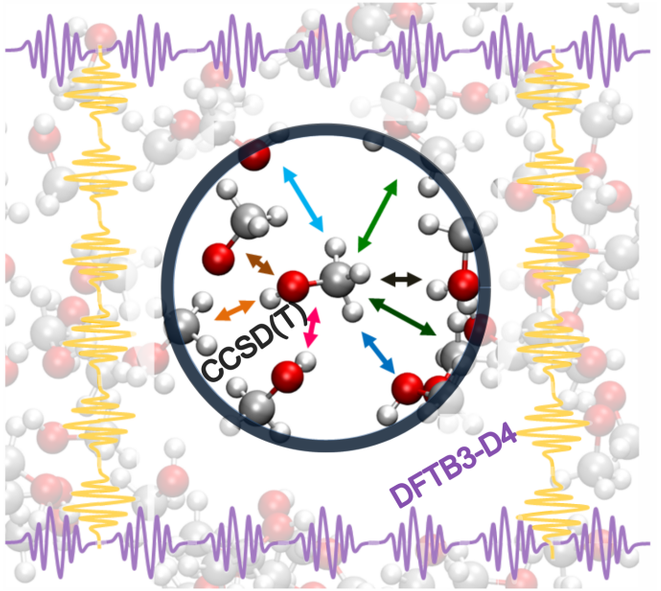
- Complex **ab initio** theories **unaffordable** for bulk materials
  - Low-level QM methods too inaccurate for NCI treatment

  ⇒ Use **many-body energy expansions** enabling importance sampling



- High numbers of unique interactions in the bulk
  - **Lack of symmetry** and time evolution particularly painful

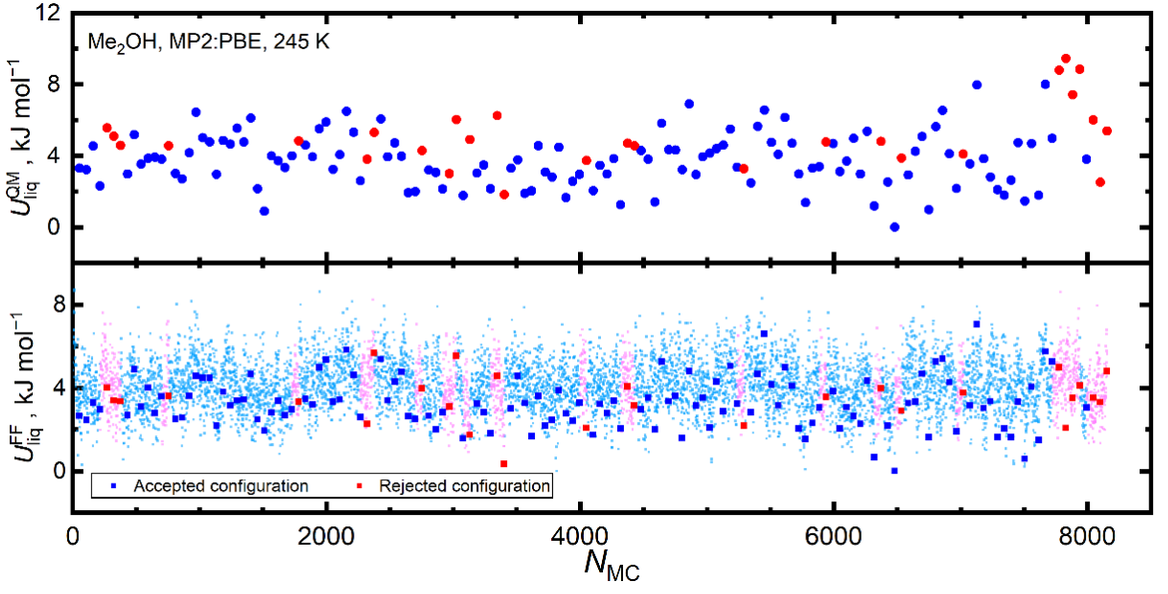
  ⇒ Use embedding schemes and apply high-tier methods for the **closest interactions**

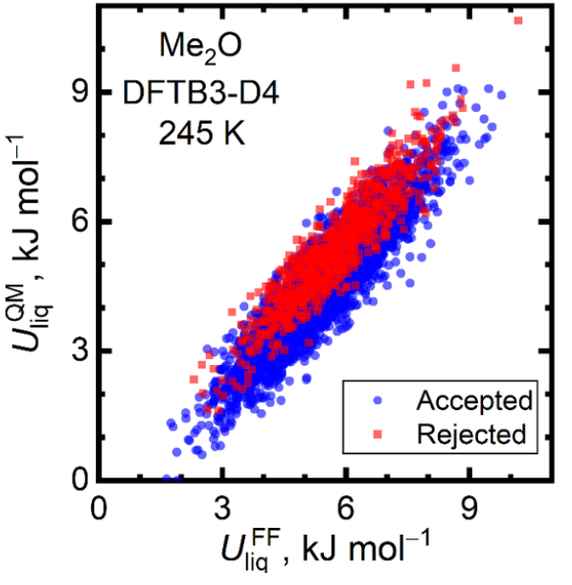


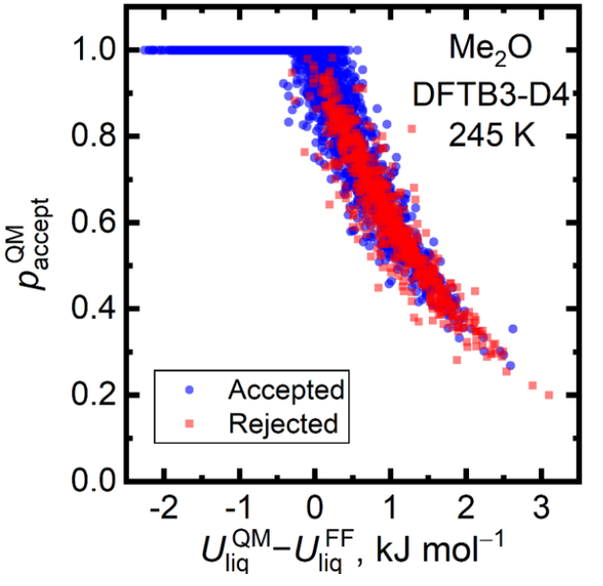
Introduction of **FrAMonC** as *Fragment-based Ab initio Monte Carlo* simulations for amorphous molecular materials

Nested MC sampling

- How to **preselect configurations** with **fair acceptance** chances in QM energy evaluations?
  - Generate inner MC sequences treated with cheap models
  - **QM assessment** only for **terminal configurations** of the inner sequences
  - High acceptance rates thanks to **decorrelating the configurations**
- How important is a good choice of the cheap model for the inner MC sequences?
  - Actually not much, a simple force field suffices
  - Need to adapt the Metropolis criterion to use energy **similarity between the potentials**

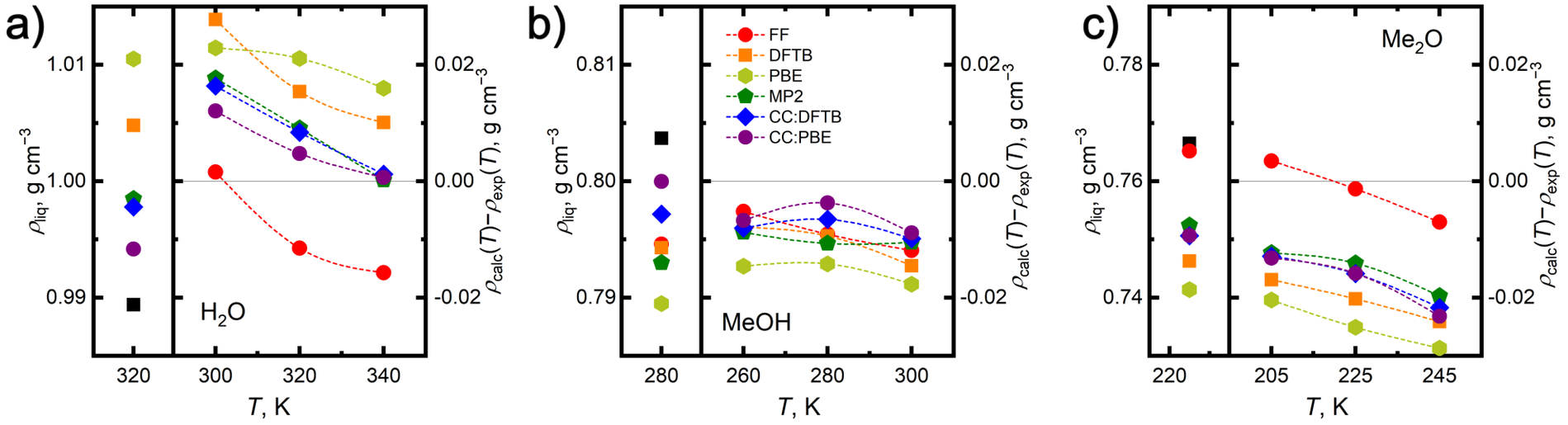


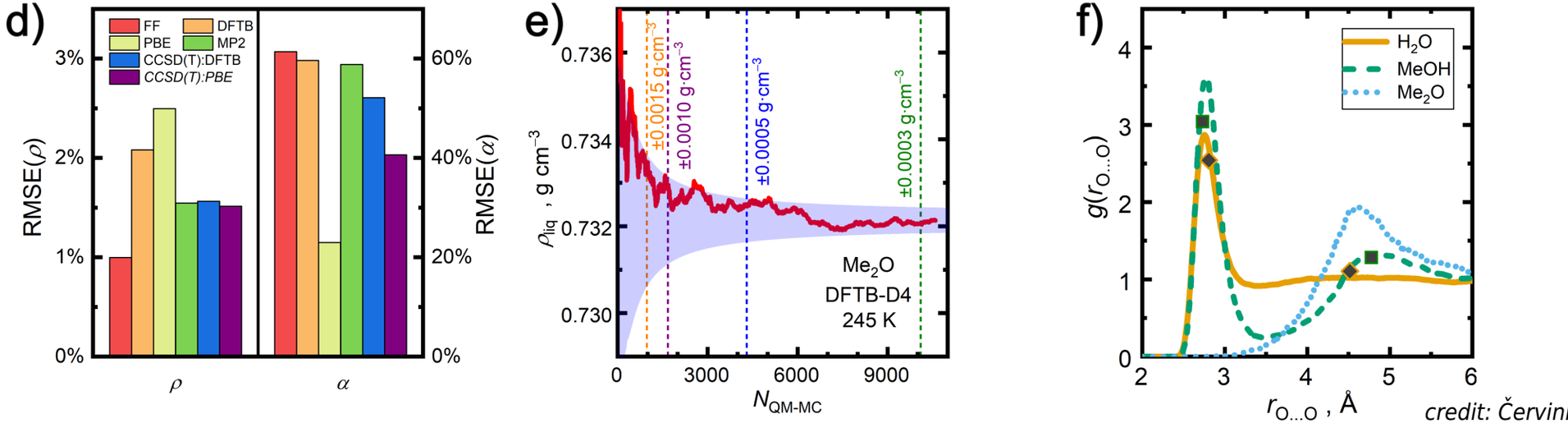


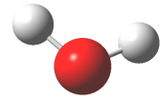


FrAMonC Results – Structure of the Liquids

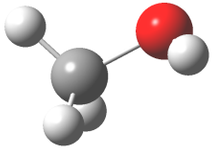
- Rather **systematic convergence** of predicted density and thermal expansivity
  - Short pairwise cut-off suffices to capture the structure



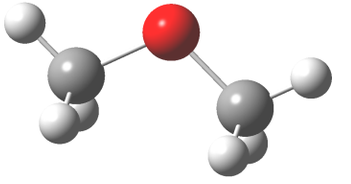




100 molecules
14 Å box



50 molecules
15 Å box

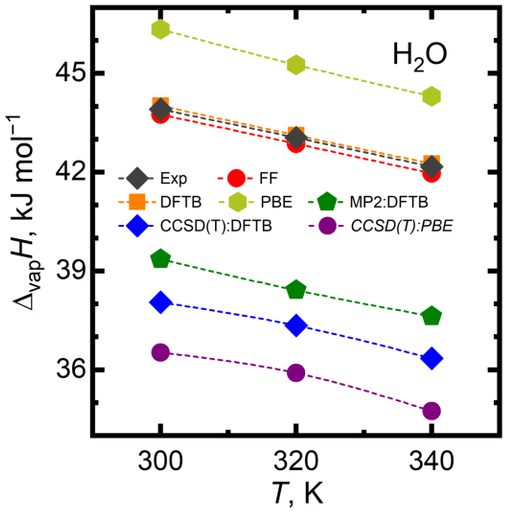


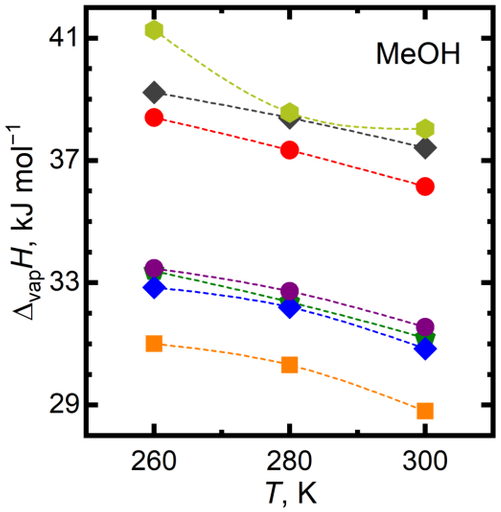
40 molecules
16 Å box

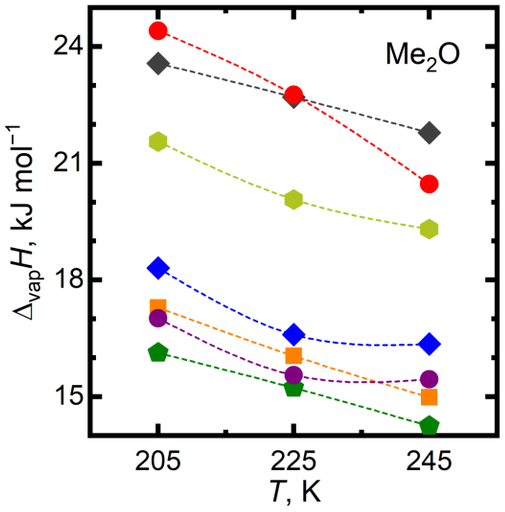
credit: Červinka, J. Chem. Theory Comput., 2025, 21, 9868.

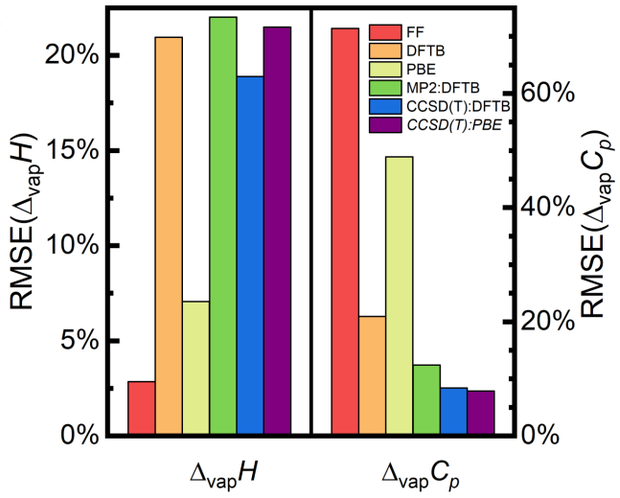
FrAMonC Results – Cohesion

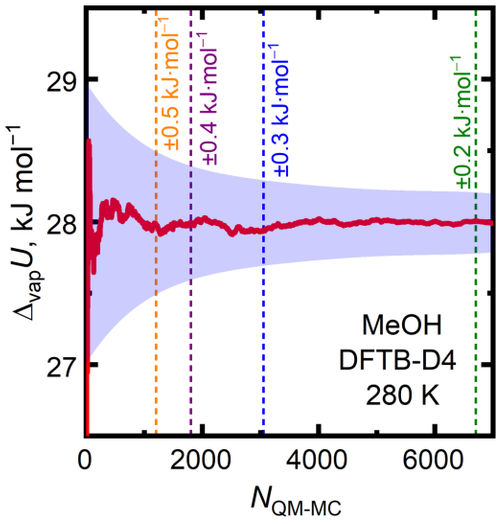
- Benchmarking vaporization enthalpy and heat capacity difference



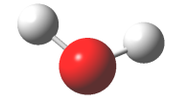




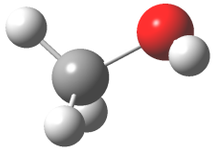




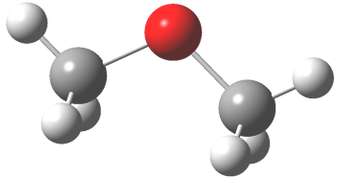
- Model **convergence** more **problematic**
- Chemical accuracy not yet reached
- **Thermal trends captured** very well



100 molecules
14 Å box



50 molecules
15 Å box

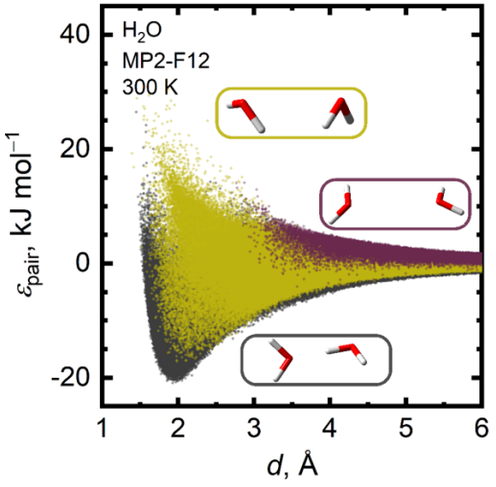


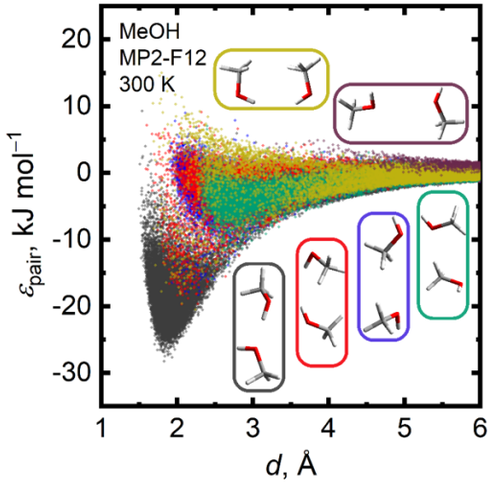
40 molecules
16 Å box

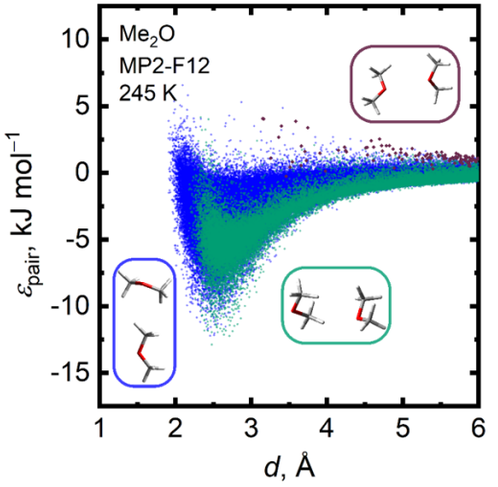
credit: Červinka, J. Chem. Theory Comput., 2025, 21, 9868.

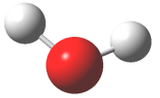
FrAMonC Results – Interaction Libraries

- Storing energies of monomers and dimers extracted from bulk liquid

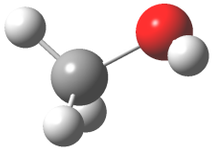




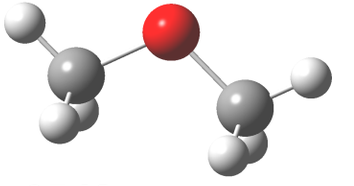




78 K monomers
6100 K dimers



50 K monomers
1660 K dimers



65 K monomers
780 K dimers

- Exploit the libraries for ML training of molecular interaction potential
  - High-tier model DLPNO-CCSD(T)/CBS in ORCA
  - Middle-tier model GFN2-xTB in DFTB+
  - Low-tier model DFTB3-D4 in DFTB+

ORCA

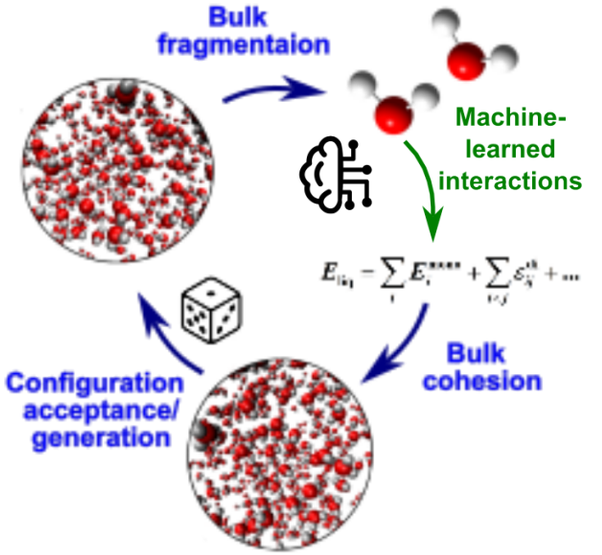
DFTB

credit: Červinka, J. Chem. Theory Comput., 2025, 21, 9868.

Computational methodology III

- Large numbers of particular **molecular interactions** to be sampled
  - **Ab initio** methods within short cut-offs or **unaffordable**

  ⇒ Use **machine-learning** molecular **potentials**

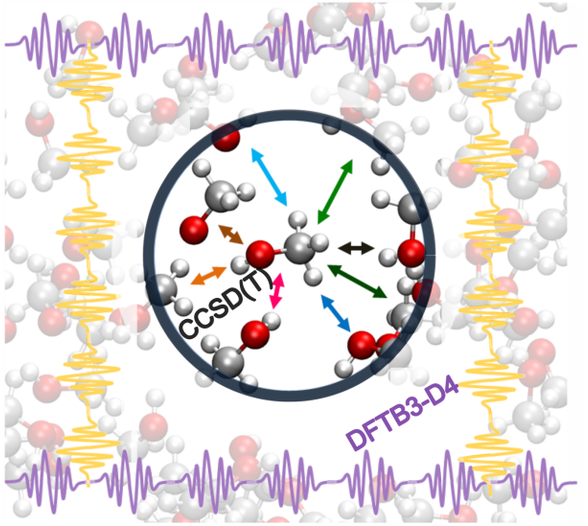


- Molecules are symmetric and their interactions may be so

  ⇒ Use **equivariant neural network** architecture methods for the **ML potential training**

- Targeting **sub-chemical accuracy** of the ML potential
  - Physical background vs. overfitting

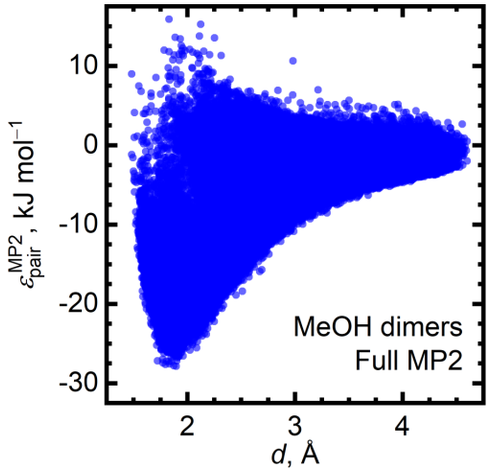
  ⇒ Use **Δ-ML scheme** correcting outputs of cheap QM methods

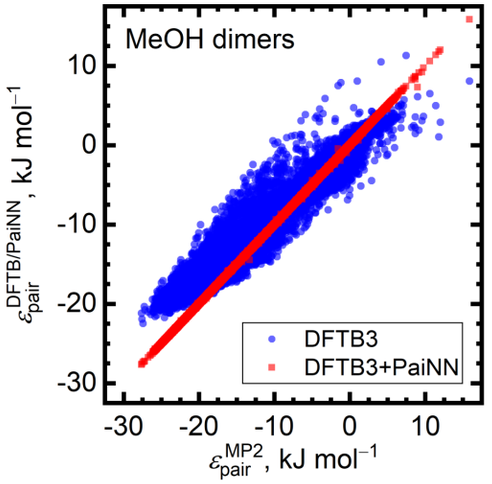


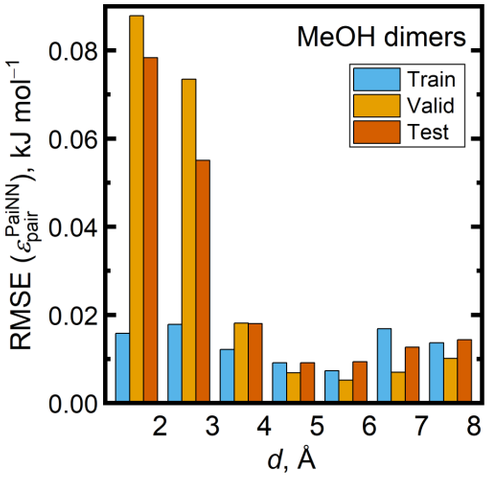
Exploitation of the **PaiNN architecture correcting** GFN2-xTB or DFTB3-D4 energies to **DLPNO-CCSD(T) quality**

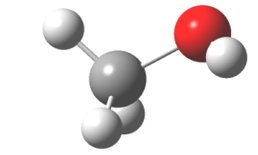
FrAMonC MLIP training

- Develop a potential **correcting DFTB3** pair interactions to the **MP2 quality**

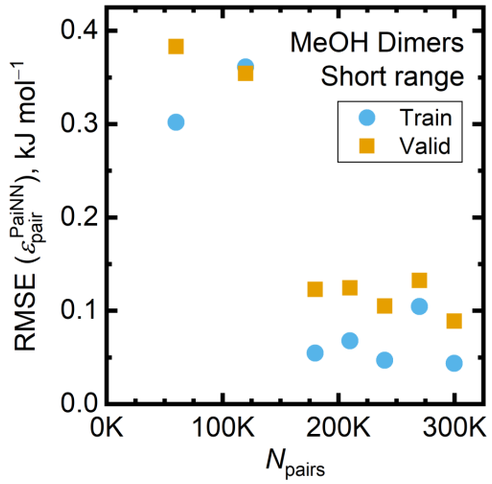


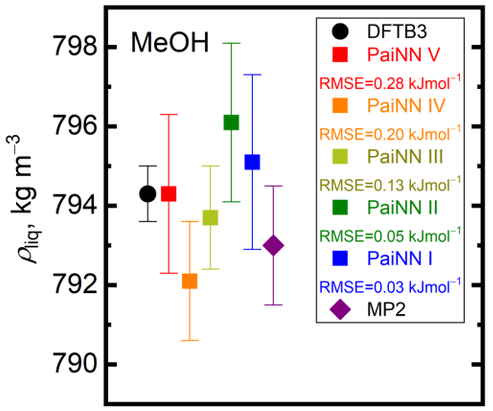






1005 K dimers

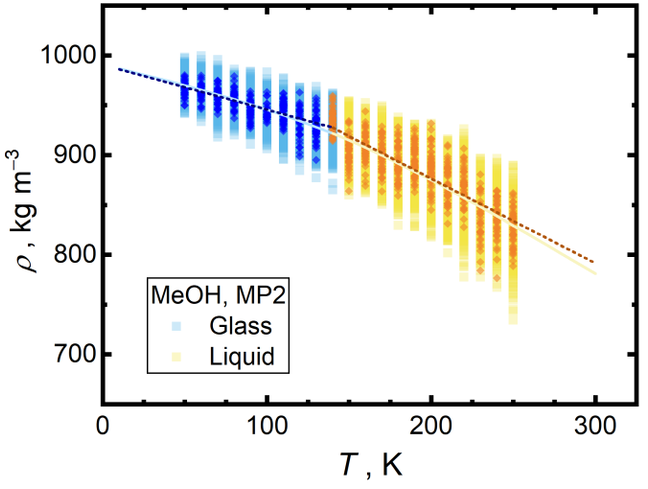


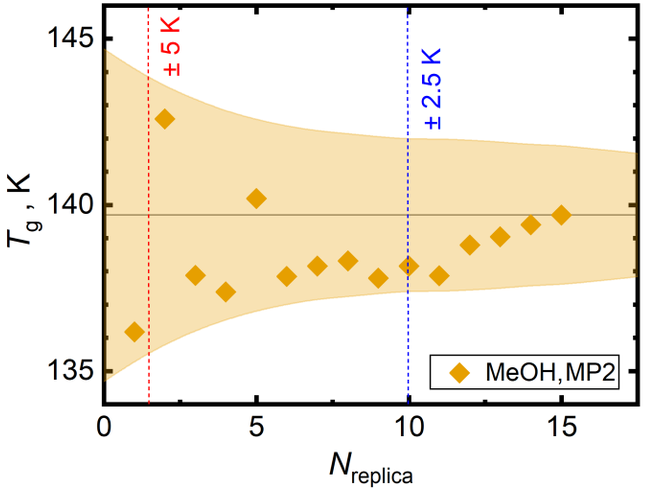


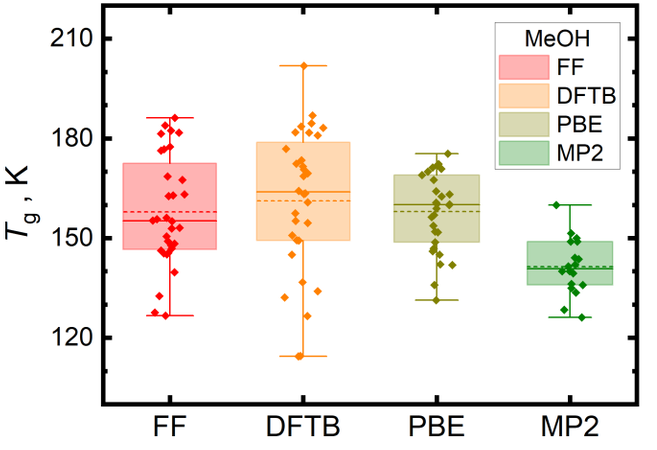
- PaiNN model **reproduces** MP2 data closely
- How **many training dimers** do we need?
- How does the **PaiNN uncertainty** propagate to **bulk simulations**?

FrAMonC Results – Vitrification

- **Glass-transition temperature** – to **dissect liquid** and amorphous **solid** (= glass)
- Challenge to derive material **dynamics from** stochastic and **static** Monte Carlo **simulations**
  - Analysis of volumetric data and **thermal expansivity**
  - Hyperbola fits to localize liquid and glassy branches
- Challenge to model properties of **amorphous molecular solids**
  - Potentially frozen in coincidental exotic states
  - Crucial to perform **multiple simulation replica**



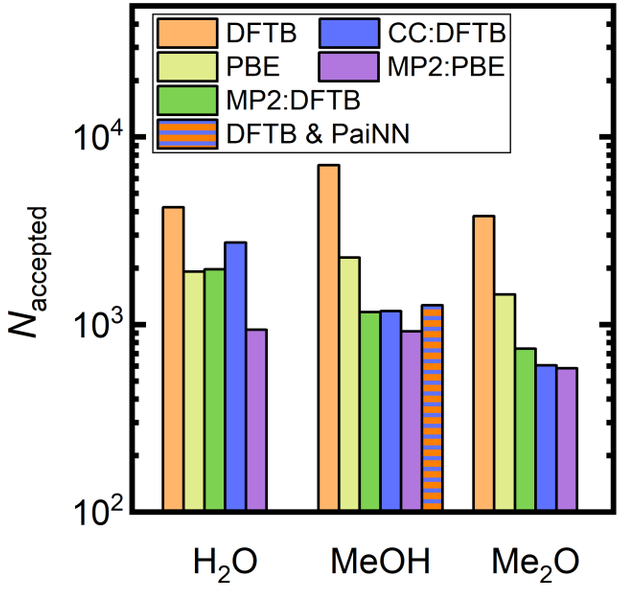


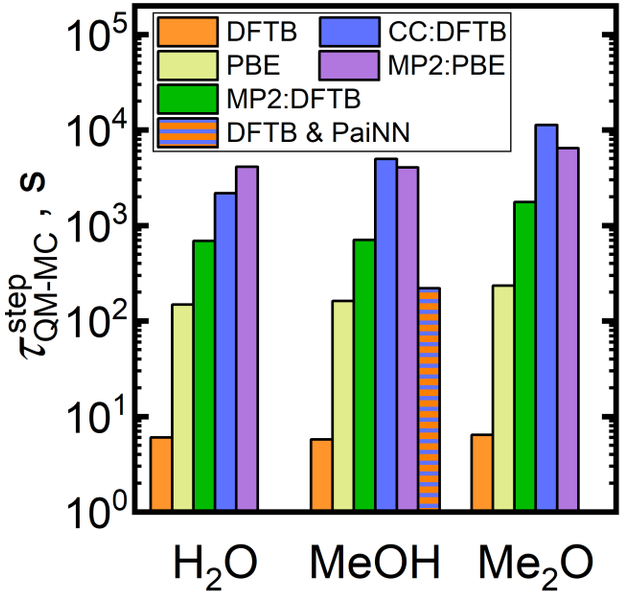


Considering Computational Costs

- Direct ab initio sampling is **expensive**, yet **feasible**
  - Hundreds to **thousands** of QM-accepted **configurations** required
  - QM-**acceptance rates** above **satisfactory** 75%
  - Perfect **parallelizability** thanks to the **many-body expansion** of bulk energy

- **Mimicking** the **ab initio** accuracy with ML potential **saves resources**
  - The most **expensive** ab initio calculations **by-passed**
  - **DLPNO-CCSD(T)** accuracy of bulk description at **sub-DFT costs**





1000 QM configs at 1 Karolina node		
Material	$\tau_{CCSD(T)}(h)$	$\tau_{PaiNN}(h)$
H_2O	680	*n/a*
MeOH	1750	60
Me_2O	4230	*n/a*

credit: Červinka, J. Chem. Theory Comput., 2025, 21, 9868.

Conclusions

- **Ab initio** simulations enabled for **thermodynamic** simulations of bulk **liquids** and **glasses**
  - Benchmarked for three simple (yet sometimes anomalous) liquids
  - Very **good performance** for densities, thermal expansion and heat capacities
  - Remaining enthalpic offsets for the overall bulk cohesion

- Further **development** using **machine-learned potentials**
  - Very **large libraries** of ab initio monomer energies and **pair interactions** created
  - **Neural networks** for a cheap ab initio mimicking potential
  - Extending the scope and **saving resources**

- More **ambitious simulation targets** in the future
  - Converge the **enthalpic accuracy** of bulk description
  - Focus on **solid – liquid** phase **transitions**

Acknowledgements

Coworkers

- @CervinkaGroup at UCT Prague
  - Zaheer Abbas
  - Petr Touš



Resources

- Czech Science Foundation JUNIOR-STAR Project 23-05476M
- COST Action CA22107 BEST-CSP
- IT4I OPEN-33-74 Project