

# Atomistic Simulations of Phase Equilibria and Fluid Properties for Process Design: What is Possible Today and What is in Store for the Future?

Edward Maginn

Department of Chemical and Biomolecular Engineering  
University of Notre Dame, USA

[ed@nd.edu](mailto:ed@nd.edu)

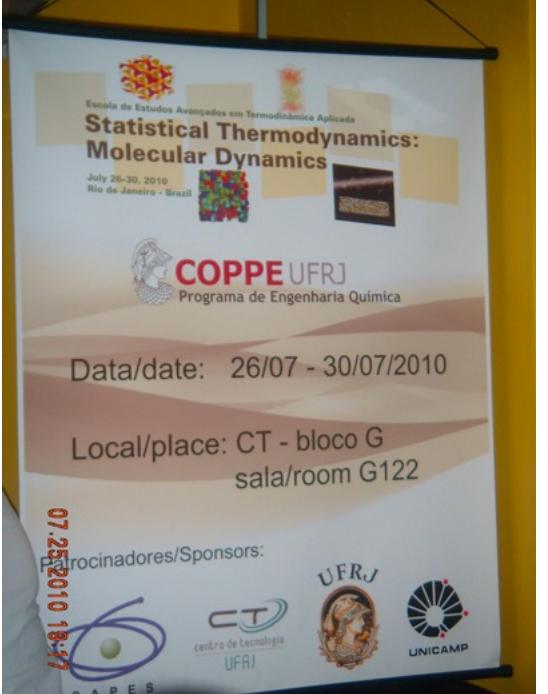


If you would like a copy of this presentation, email me [ed@nd.edu](mailto:ed@nd.edu) or download:  
<https://github.com/MaginnGroup/Public-Presentations>

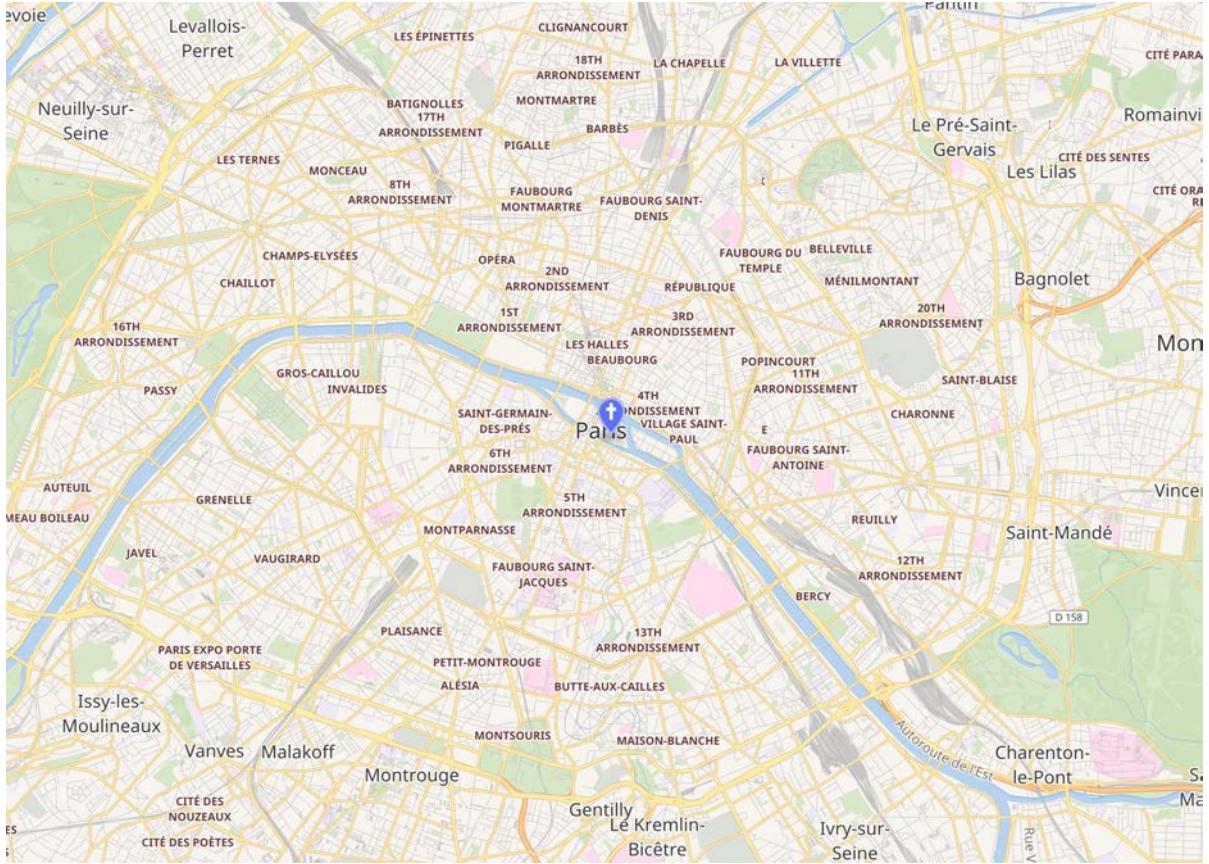
24 October, 2022



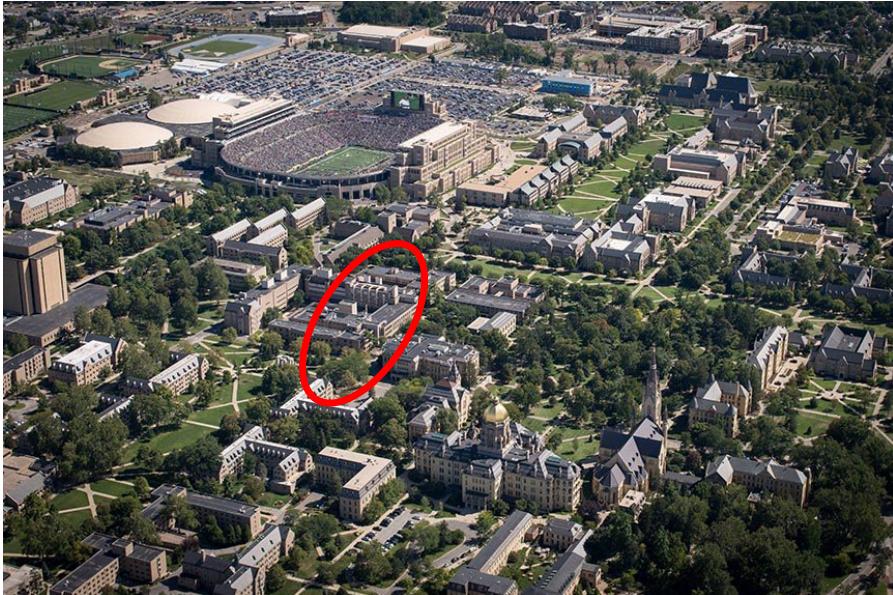
# July 2010



# Where is Notre Dame?



# Where is University of Notre Dame?



Notre Dame founded in 1842

- 8,700 undergraduate, 3,900 graduate
- 50 states, 71 nationalities

## Soft Matter & Nanomaterials



Phillip



Guo



Schaefer



Hill



Bruening

## Chemical Engineering (1905)

- ~210 UGs (3 classes)
- ~100 PhDs
- 24 tenure/tenure track and 4 teaching faculty



McGinn



Hicks



O'Brien



McCready



Myung



Zartman



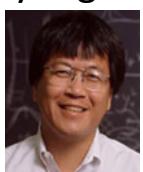
Bilgicer



Webber



Burks



Chang



Leighton



Bohn



Wang

## Energy & Sustainability

## Bioengineering & Medical Diagnostics

## Simulation & Data Science

## Teaching faculty

cbe.nd.edu



Maginn



Kantor



Schneider



Whitmer



Dowling



Colón



Kiziltepe



Vogel



Goodrich



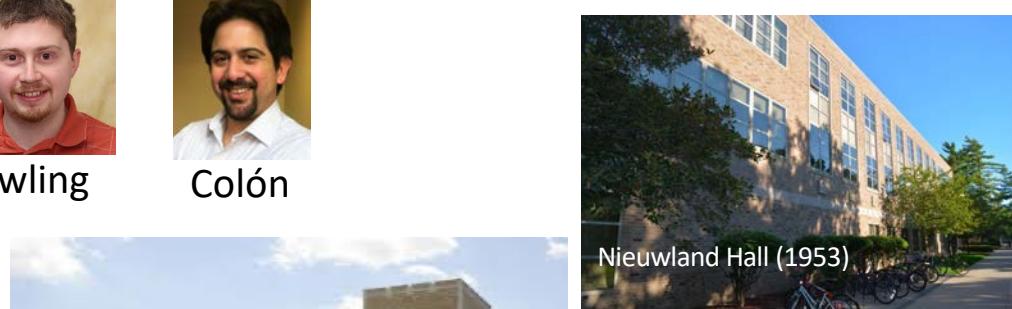
Ehren



Stinson-Remick Hall (2010)



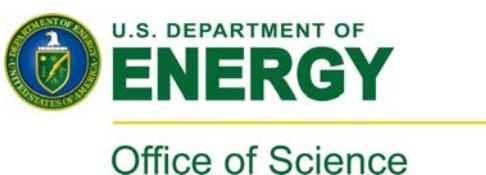
McCourtney Hall (2016)



Nieuwland Hall (1953)

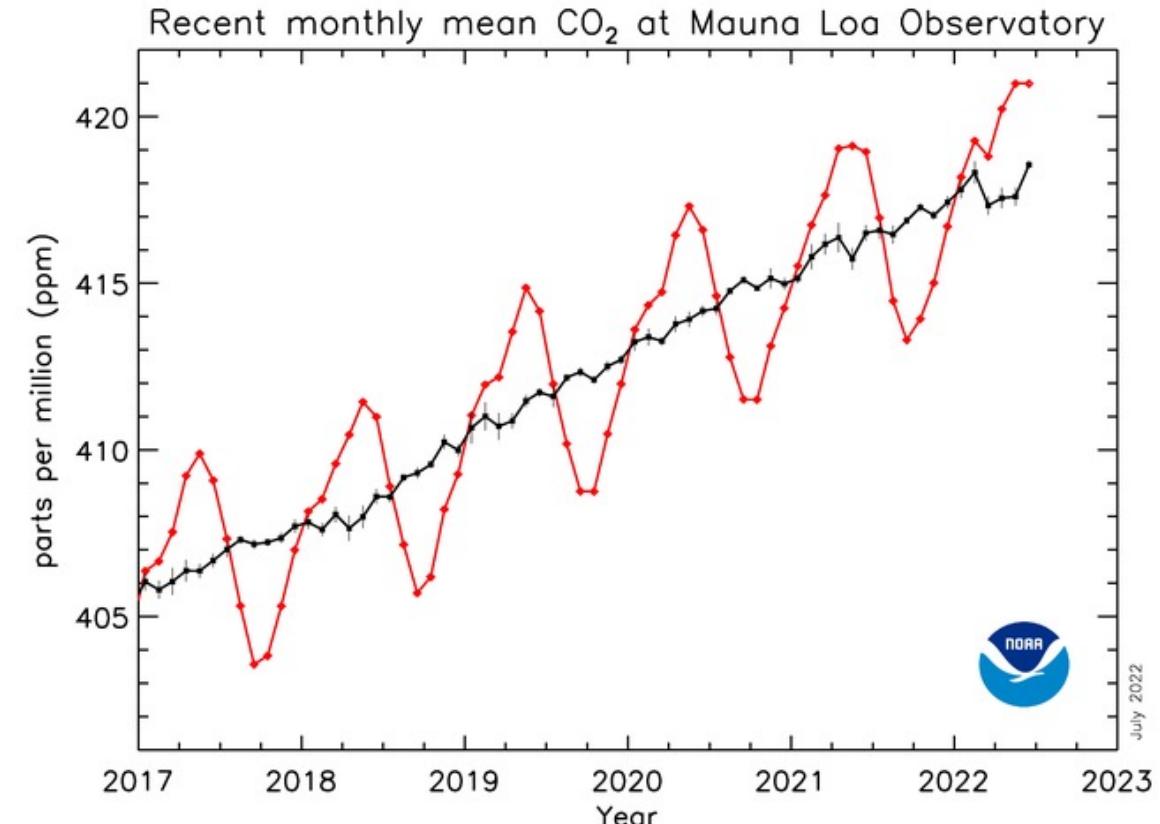
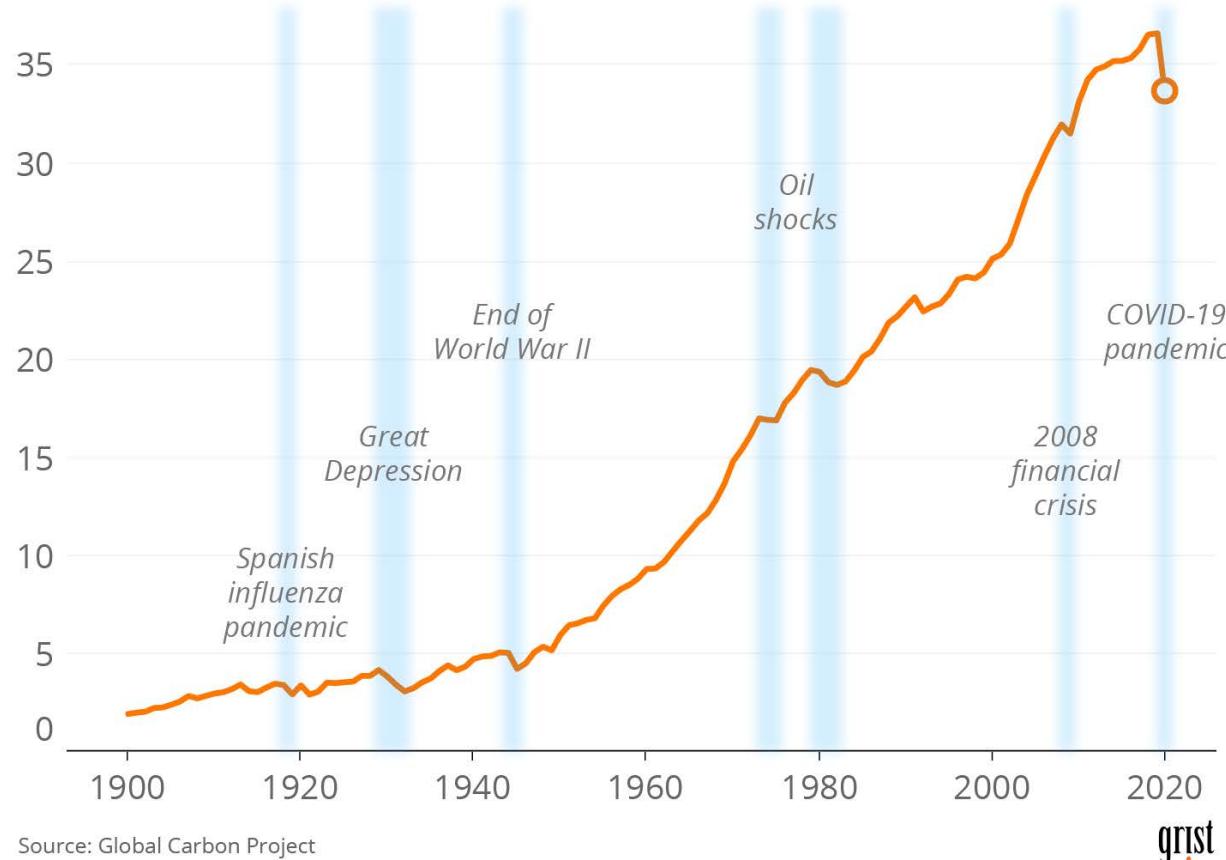
We are hiring faculty and always interested in PhD students and postdocs

# Thanks to my group and funding sources

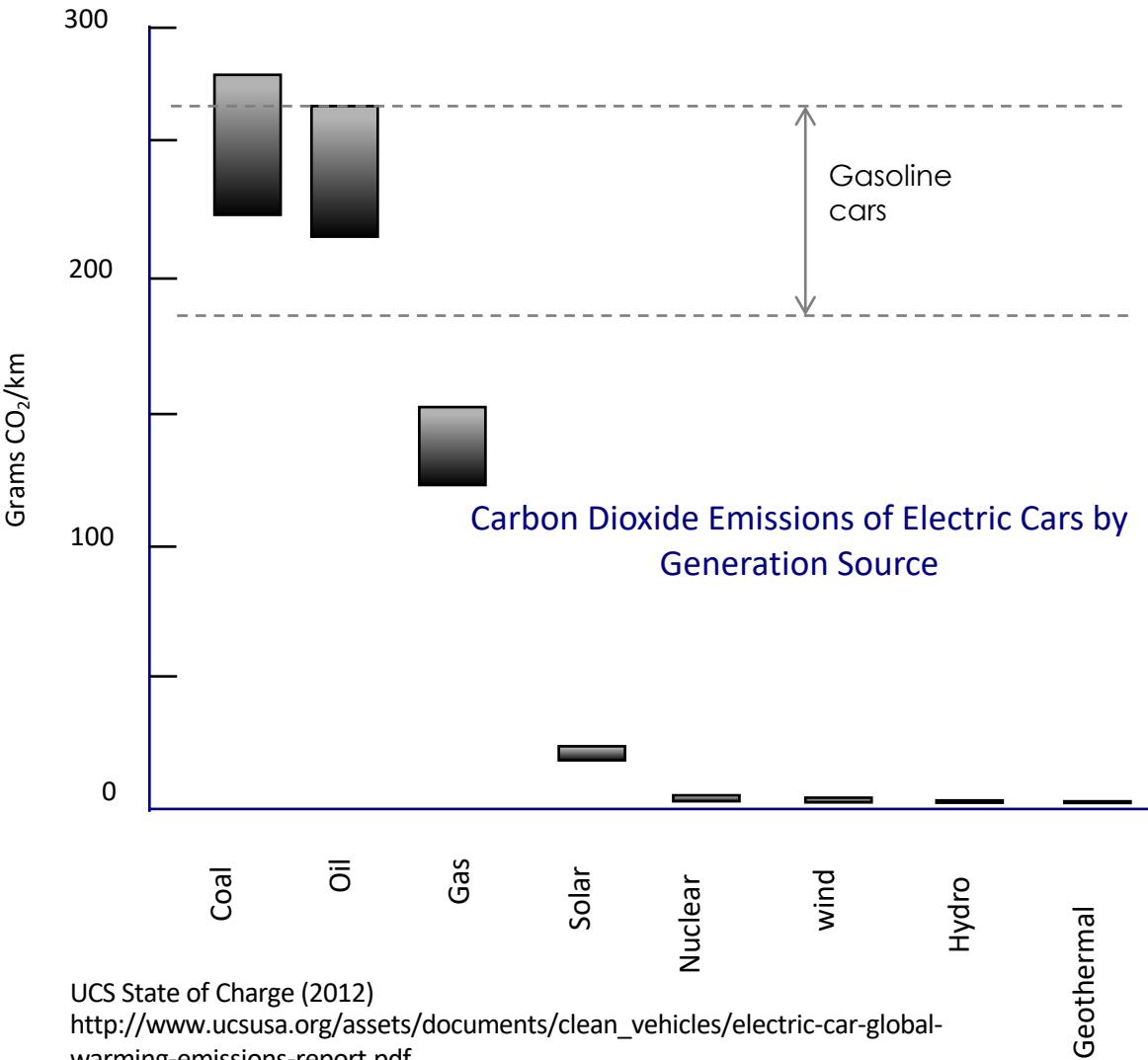


# Greenhouse gas emissions & sources

Annual global fossil emissions, billion metric tons of CO<sub>2</sub>



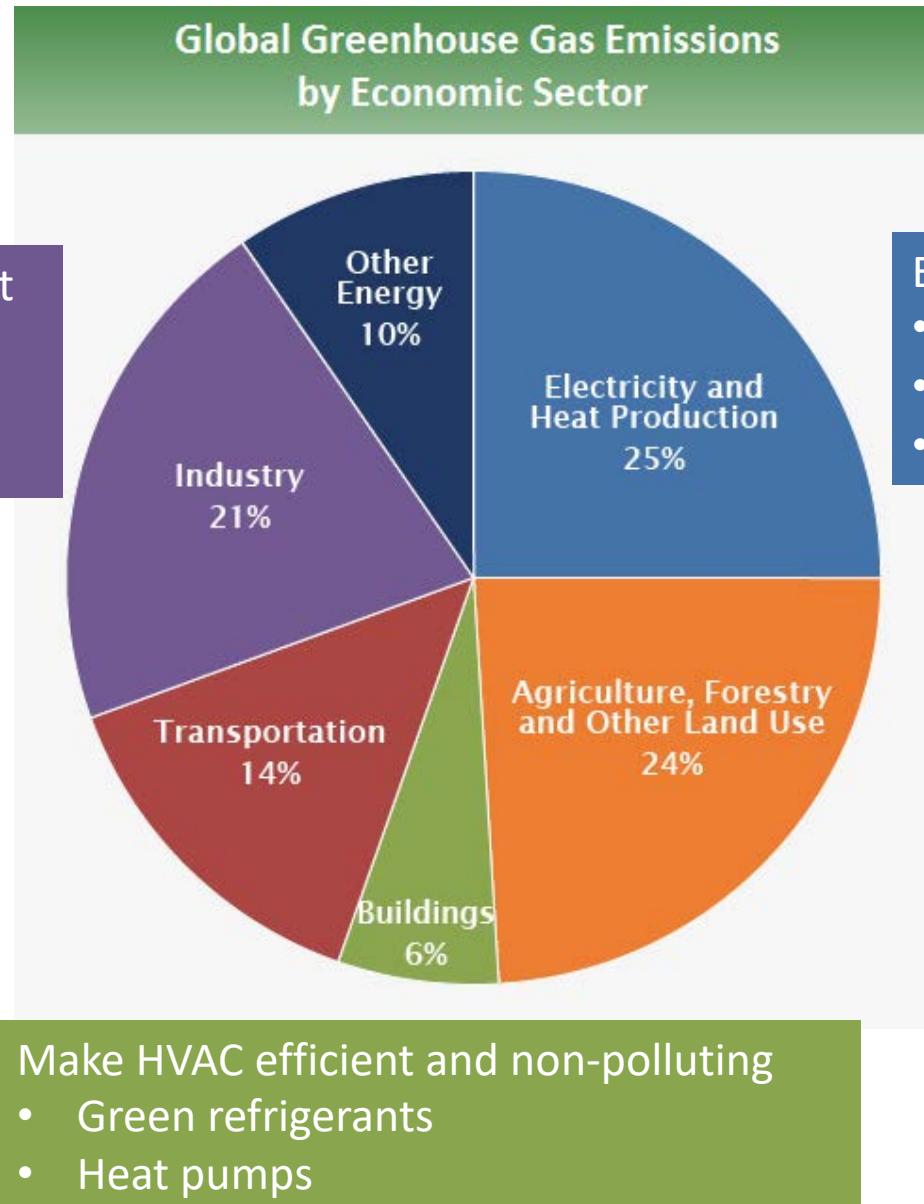
# Are electric cars the answer?



Thomas Edison and his electric car.  
Credit: *Scientific American*, January 14, 1911

- Depending on the power source, **some electric cars emit more CO<sub>2</sub> than gasoline cars!**
- Electricity must come from **nuclear or renewable sources**
- Energy storage and renewables are key!

# Equifase: Our role



Make process industries more efficient

- Separations
- Better catalysts
- Electrify thermal processes

Enable renewables and safe nuclear

- Energy storage
- Safer nuclear cycles
- Molten salts

Electrify transportation with nuclear/renewables

- Energy storage
- Critical element recovery
- New battery materials

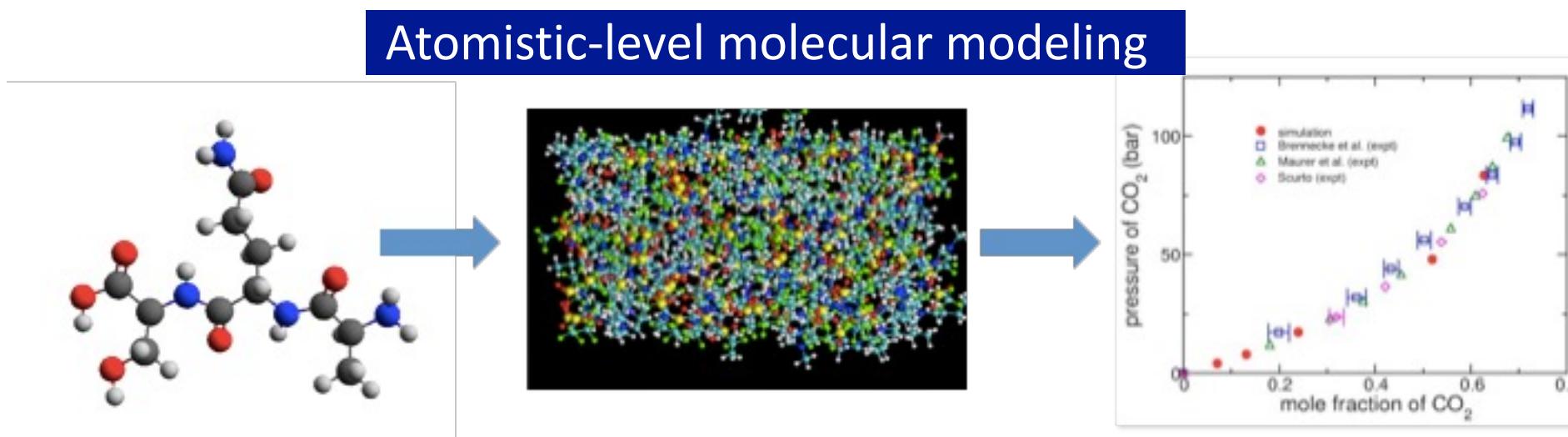
Provide water and fertilizer

- Green ammonia
- Water purification

# Goal of molecular simulations

- True *prediction* of phase equilibria and thermophysical properties

Draw - Simulate - Predict



# How it started...

Circa Equifase 1990 – Salvador de Bahía, Brazil



1990 workstation: 1 100 MHz processor; cost \$35K (\$75K today)



Cray supercomputer: 2-8 167 MHz processors; cost: \$5M (\$11.4M today)



```
IF (JRENE.GT.0) THEN
    IF (JRENE.EQ.1.) THEN ! Power-law energy spectrum
        RND=RNDM(5)
        ENLO=EK_LO
        ENHI=EK_HI
        AIND=AINDX
        CALL SPECT (RND, AIND, ENLO, ENHI, EKPAR)
        IF (KPART.EQ.1.) THEN
            PPART = EKPAR
        ELSE IF ((KPART.EQ.3).OR.(KPART.EQ.2)) THEN
            AM = 0.00051059906
            PPART=SQRT((EKPAR+AM)**2.-AM**2.)
        ELSE IF (KPART.EQ.14) THEN
            AM = 0.9382723
            PPART=SQRT((EKPAR+AM)**2.-AM**2.)
        ELSE
            PRINT *, 'YOU MUST INSERT THE PARTICLE MASS IN THE PROGRAM'
            STOP
        ENDIF
    ELSEIF (JRENE.EQ.2)THEN ! Spectrum for Background Positrons
        IF (KPART.NE.2) THEN
            PRINT *, 'THE CHOSEN SPECTRUM IS FOR POSITRONS!!!!'
            PRINT *, 'THE CHOSEN PARTICLES ARE NOT POSITRONS!!!!'
            STOP
        ENDIF
        .....
    ELSE
        .....
    ENDIF
ENDIF
```

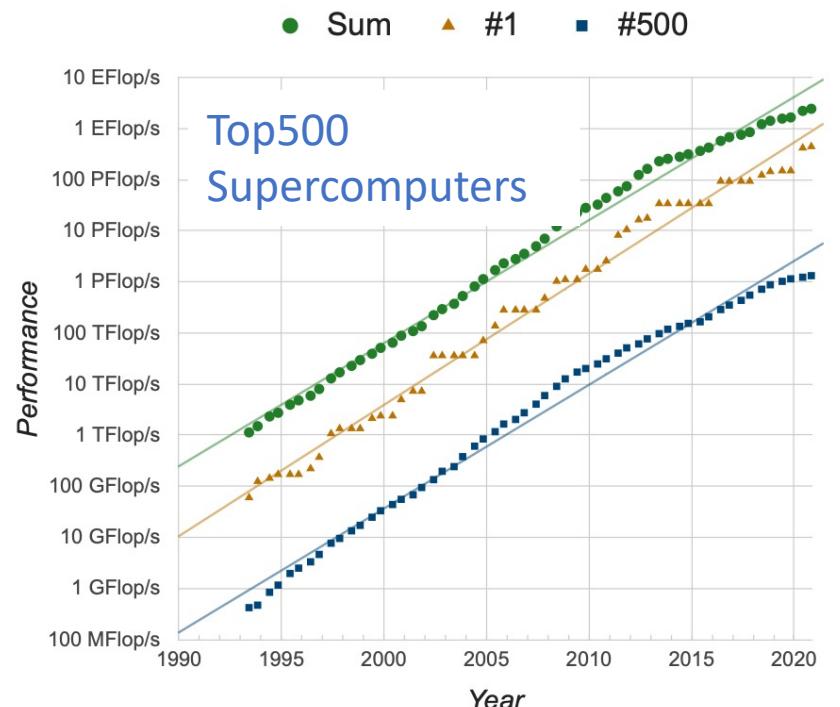
# How it is going...



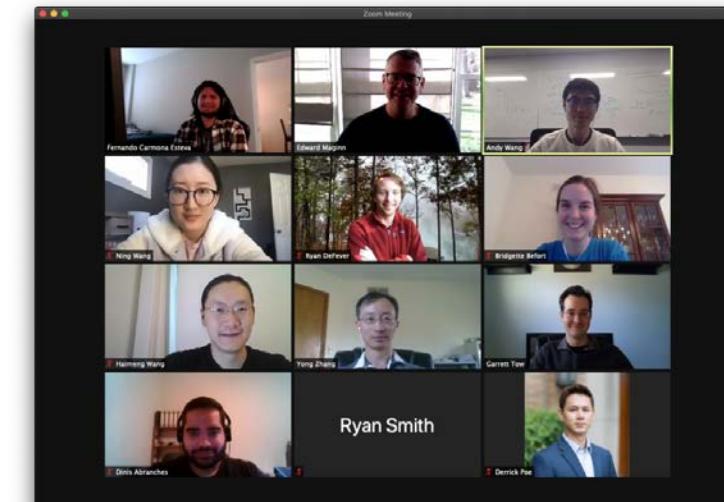
Frontier at Oak Ridge National Lab  
(2022, \$600M)  
1.1 exaflops ( $10^{18}$  calculations/s)  
9,472 CPUs and 37,888 GPUs



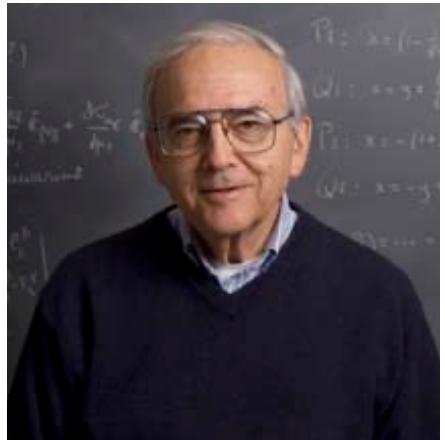
Macbook Pro; 8 3.5GHz cores ~\$2K



A screenshot of a GitHub repository page for the "Maginn Research Group". The repository has 3 followers and is associated with the University of Notre Dame. It contains three pinned projects: "Cassandra" (a Monte Carlo package for atomistic simulations in Fortran), "mosDef\_cassandra" (a MoSDef compatible wrapper for Cassandra Monte Carlo code in Python), and "PyLAT" (a Python library for lattice calculations). The main interface shows a search bar, pull requests, issues, marketplace, and explore sections.



# Method and algorithm advances



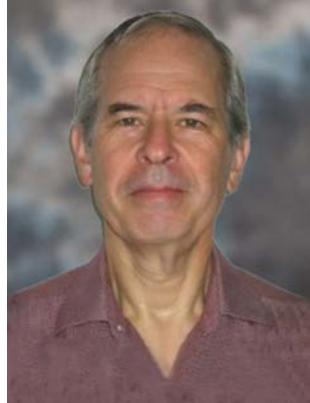
Widom insertion  
Ben Widom



Gibbs ensemble MC  
Thanos Panagiotopoulos



Solid free energies  
Daan Frenkel and Anthony J. C. Ladd



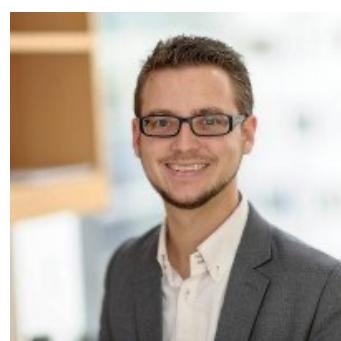
Great new method  
Your name here



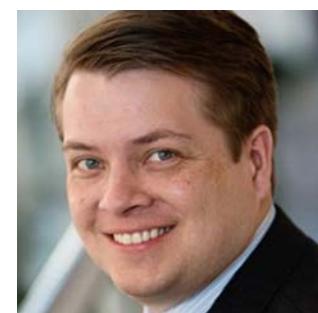
Overlapping distributions  
(BAR) Charles Bennett



Histogram reweighting  
Alan Ferrenberg and Robert Swendsen



MBAR  
John Chodera, Michael Shirts



*A few examples of advances in  
phase equilibria simulations*

# VLE simulations of alkanes

## A New Intermolecular Potential Model for the *n*-Alkane Homologous Series

Jeffrey R. Errington and Athanassios Z. Panagiotopoulos

[View Author Information](#) ▾

Cite this: *J. Phys. Chem. B* 1999, 103, 30, 6314–

6322

Publication Date: July 9, 1999 ▾

<https://doi.org/10.1021/jp990988n>

Copyright © 1999 American Chemical Society

Article Views

938

Altmetric

-

Citations

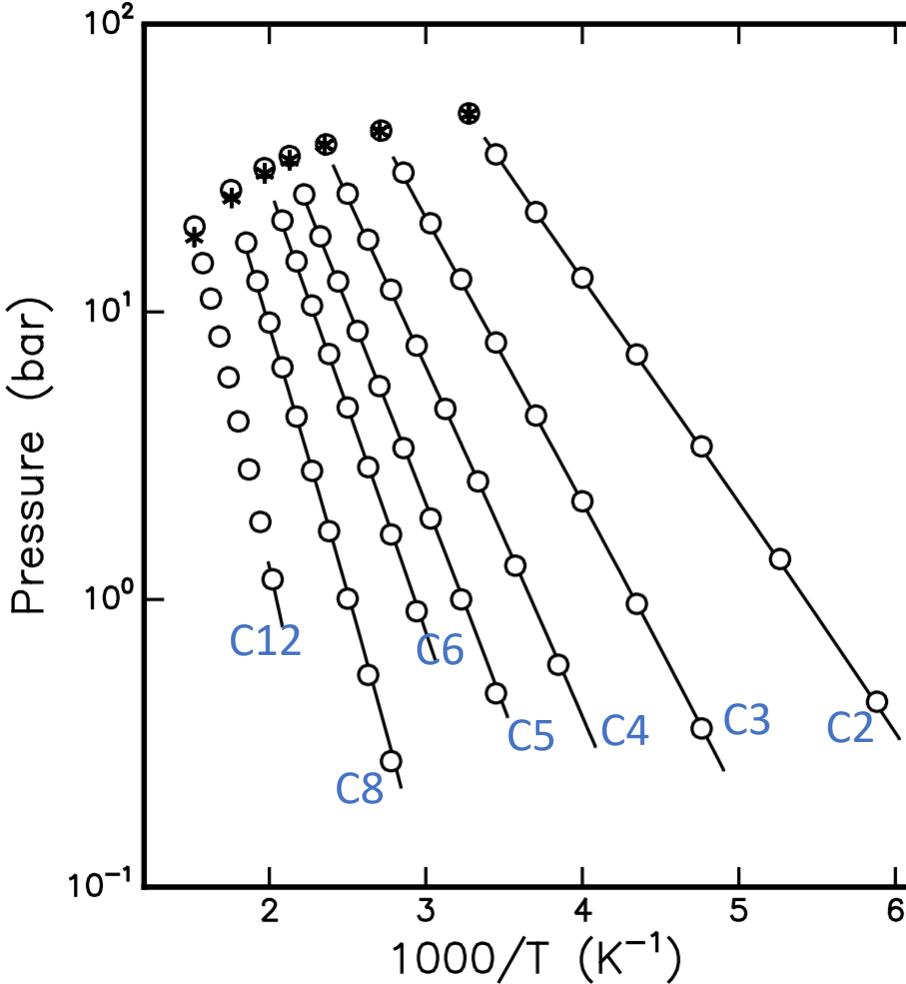
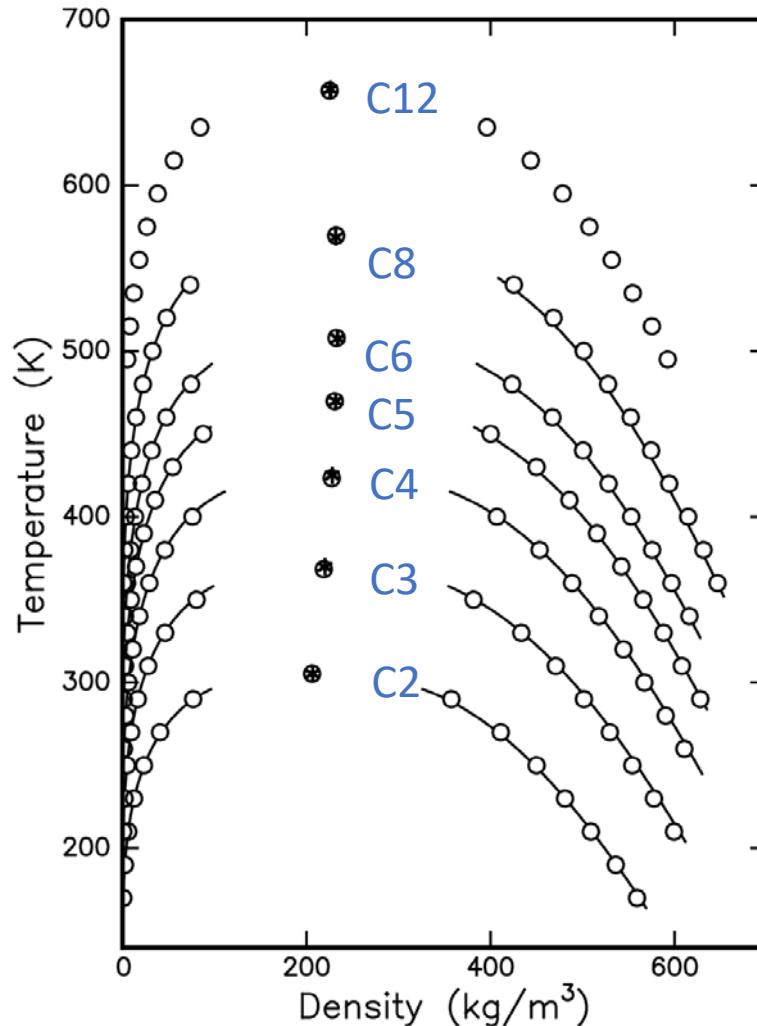
163

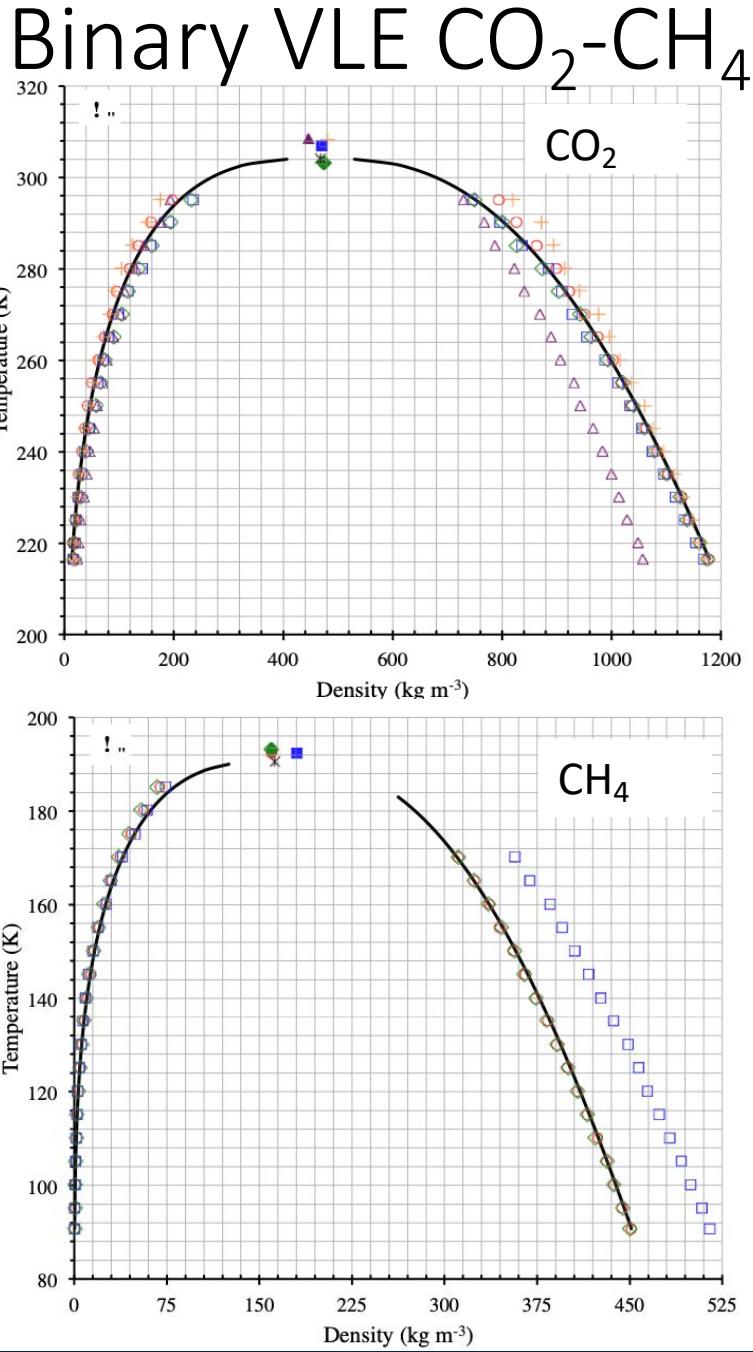
[LEARN ABOUT THESE METRICS](#)

Share Add to Export



GEMC of alkanes with **configurational biasing**

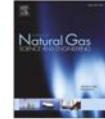




Mixed-gas VLE much harder to compute and model

SAFT  $\gamma$ -Mie potential\* does very well o

\*Avendaño et. al, *J. Phys. Chem. B* 2011, 115, 38, 11154–11169



Thermodynamic properties and fluid phase equilibrium of natural gas containing  $\text{CO}_2$  and  $\text{H}_2\text{O}$  at extreme pressures typically found in pre-salt reservoirs

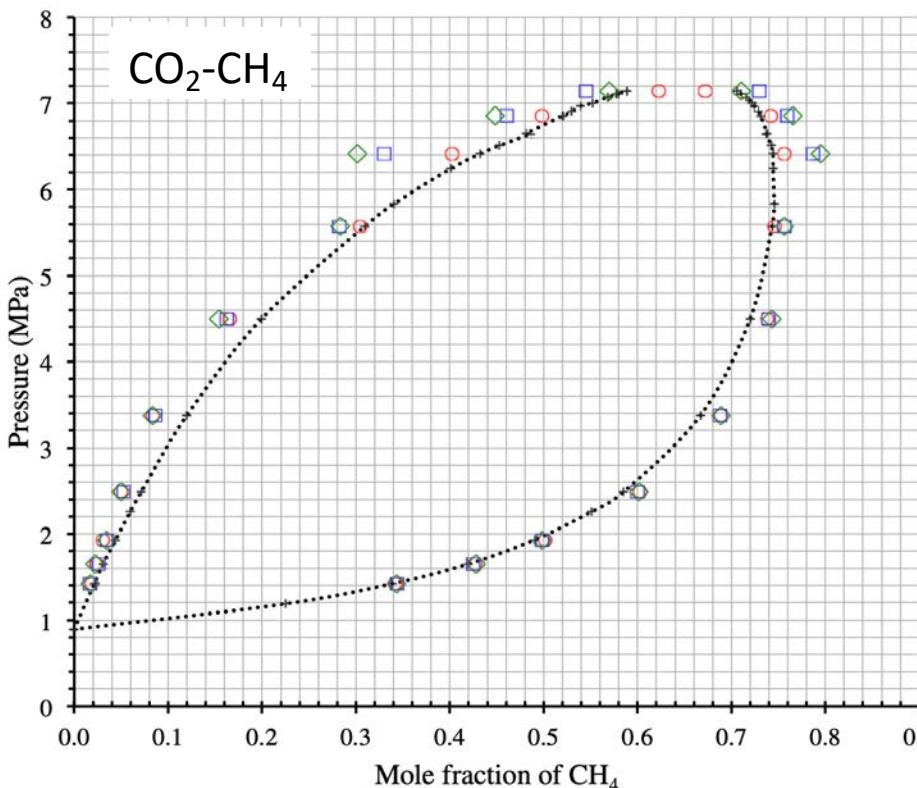
Cassiano G. Aimoli<sup>a,\*</sup>, Danilo P. de Carvalho<sup>b</sup>, Pedro A. Pessoa Filho<sup>b</sup>, Edward J. Maginn<sup>c</sup>, Charles R.A. Abreu<sup>d</sup>

<sup>a</sup> Petrobras, Petróleo Brasileiro S.A., Santos, SP, Brazil

<sup>b</sup> Chemical Engineering Department, Engineering School, University of São Paulo, São Paulo, SP, Brazil

<sup>c</sup> Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA

<sup>d</sup> School of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil



SAFT  $\gamma$ -Mie is a great example of combination of equation of state and molecular modeling

# Force field database/validated results

UNIVERSITY OF MINNESOTA  
Driven to Discover™

One Stop MyU For Students, Faculty, and Staff

The Siepmann Group  
Simulating Complex Chemical Systems

TraPPE Home Validation

About TraPPE

About the Validation Effort

The development of the TraPPE force field began in the mid 1990's with Ilja Siepmann's move to the University of Minnesota. The initial target was a united-atom model for linear alkanes (TraPPE 1). Since that time, TraPPE has been extended to include several different families, ranging from the widely used TraPPE-United Atom force field to newer extensions like TraPPE-Coarse Grain. During the same time span, simulation methods and computing power have progressively improved. We believe it is of interest to the TraPPE user community to validate the accuracy of early TraPPE models and to provide more precise simulation data using optimized simulation protocols, larger system sizes, and additional temperatures.

Search the Validation Database  
Select from the current list of validated models:  
(choose molecule)

2,2-dimethylpropane (UA)

Chemical Formula: C<sub>5</sub>H<sub>12</sub>

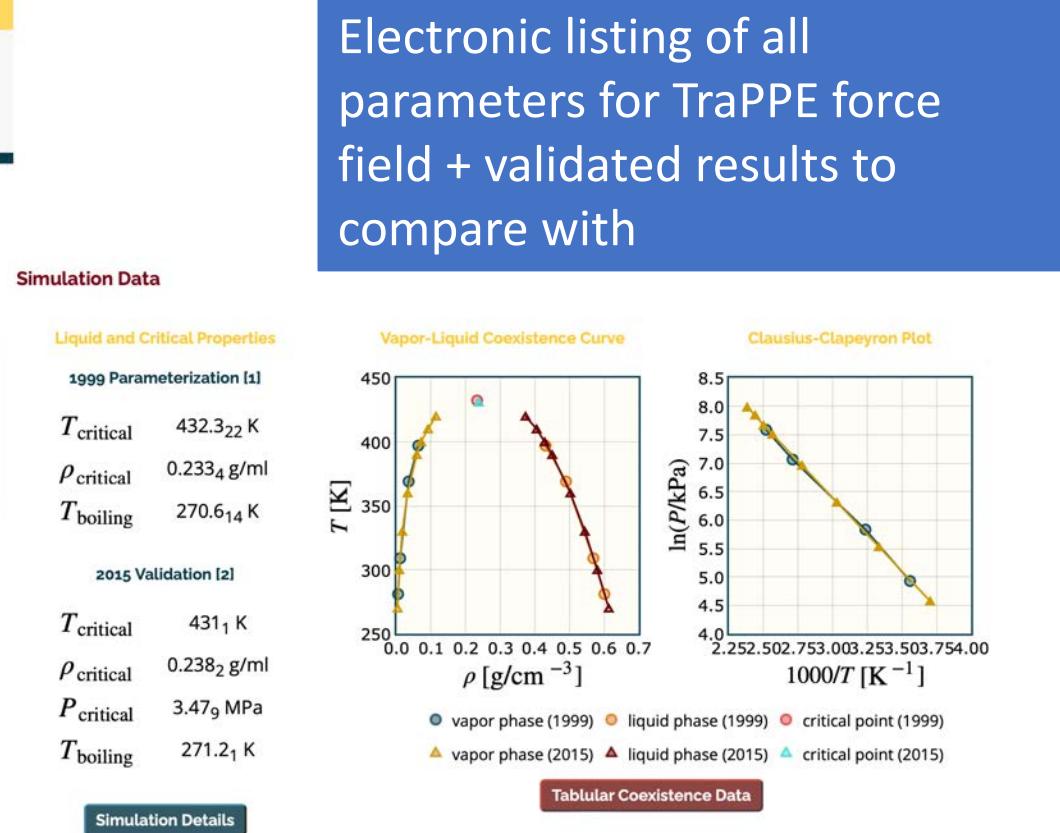
Molecular Weight: 72.15

Smiles String: CC(C)(C)C

InChI=1S/C5H12/c1-5(2,3)4/h1-4H3

This model was developed as part of the TraPPE parameterization process.

Parameters Simulation Data Downloads



## Citation List

When using the **2,2-dimethylpropane** force field or the parameterization data, please cite the following journal reference:

- [1] M.G. Martin, and J.I. Siepmann,  
["Novel configurational-bias Monte Carlo method for branched molecules. Transferable potentials for phase equilibria. 2. United-atom description of branched alkanes"](#)  
*J. Phys. Chem. B* **103**, 4508-4517 (1999).

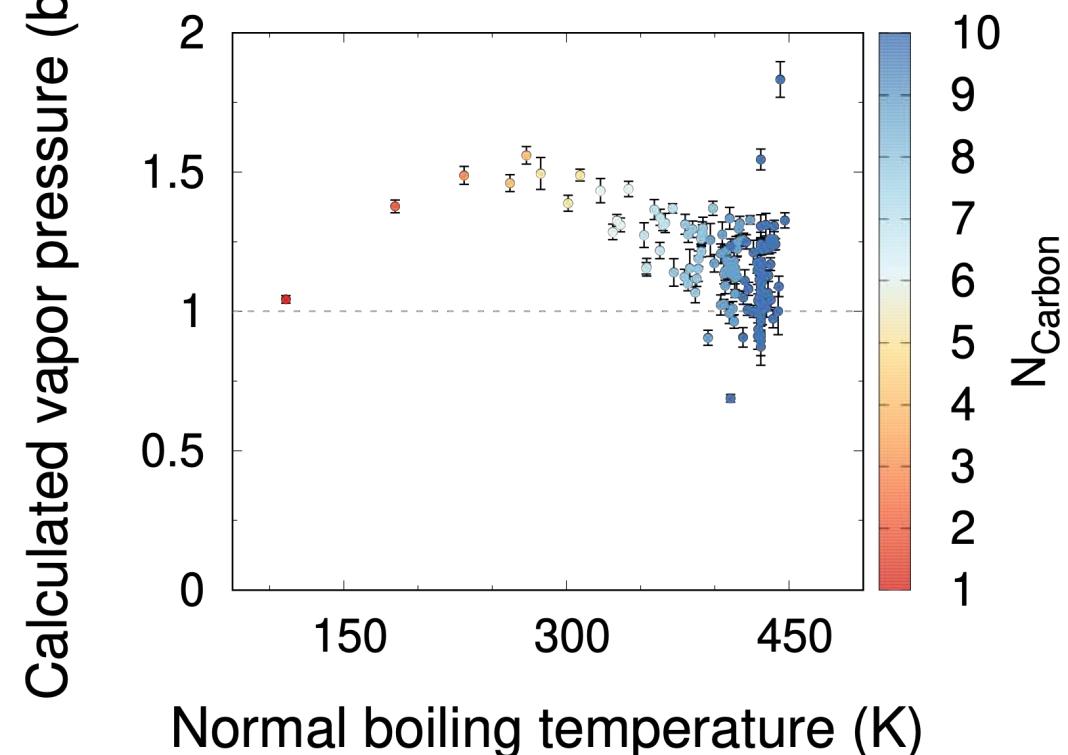
# High throughput testing of TraPPE

Run 150 independent GEMC simulations with unique force fields; job management done by signac

Linear and branched alkanes with 10 or less carbon atoms:

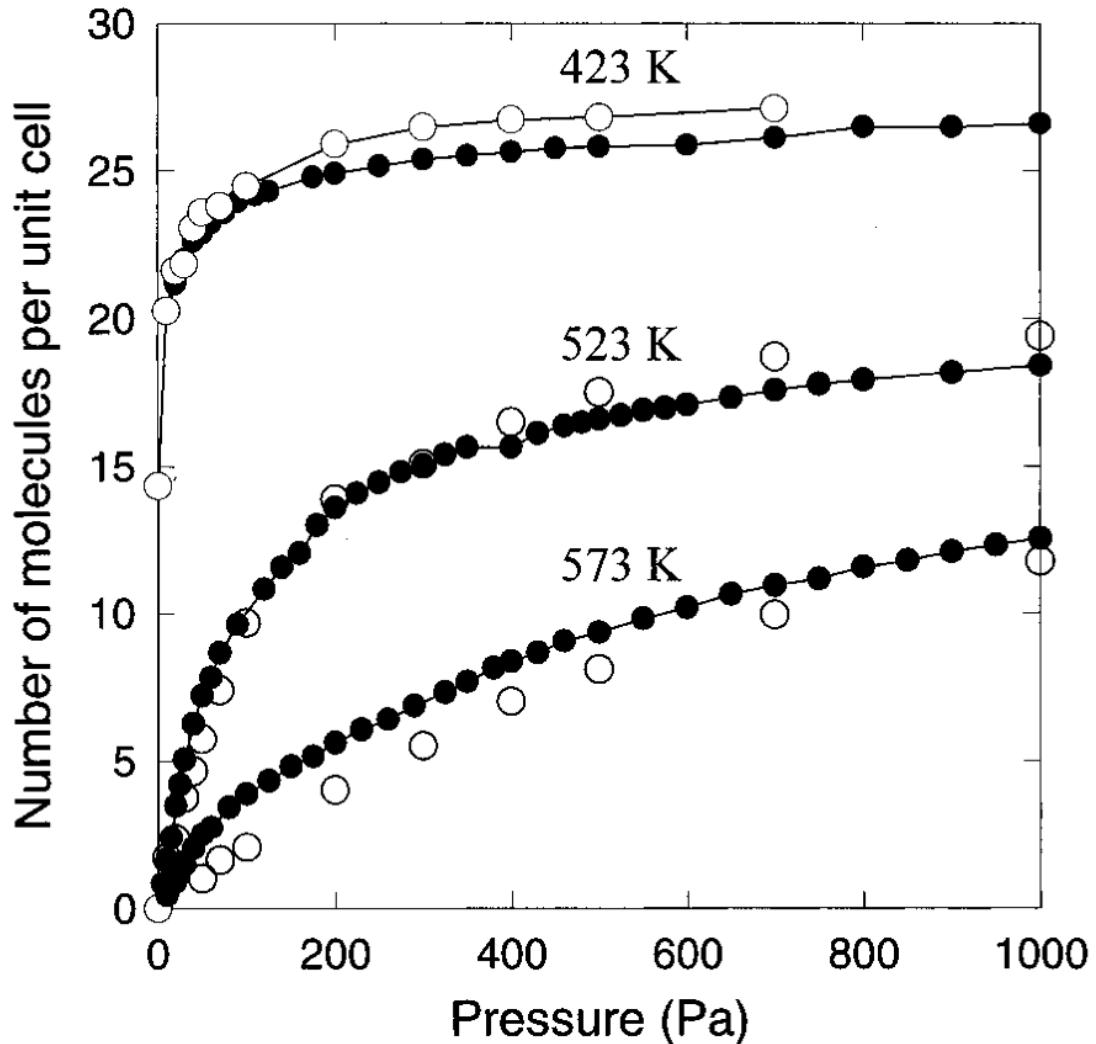
Number of Carbon	Number of isomers
1	1
2	1
3	1
4	2
5	3
6	5
7	9
8	18
9	35
10	75

Well-known problem that TraPPE overestimates vapor pressures



DeFever and Maginn, unpublished

# Adsorption – porous solids



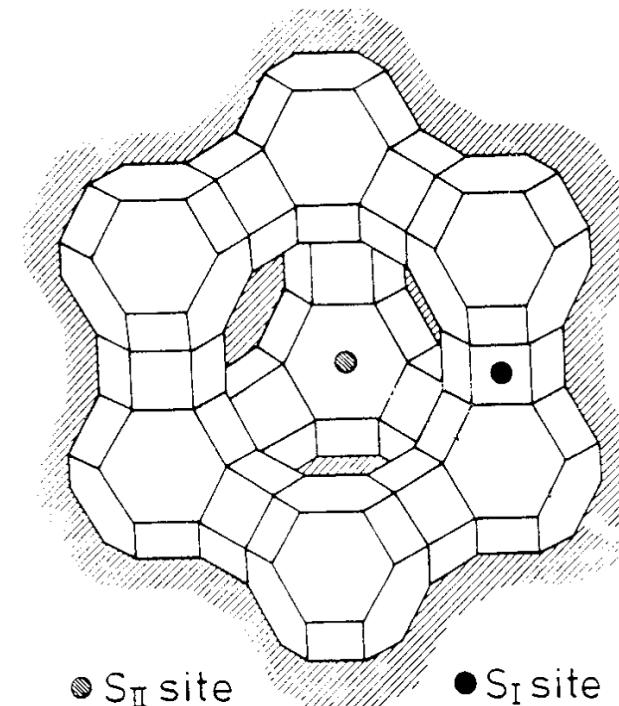
**Figure 15.** *m*-Xylene adsorption isotherms at 423, at 523, at 573 K. Comparison between simulation (open circle) and experiment (filled circle).<sup>64</sup>

## Grand Canonical Monte Carlo Simulations of Adsorption of Polar and Nonpolar Molecules in NaY Zeolite

Roland J-M Pellenq,<sup>†,‡</sup> Bernard Tavitian,<sup>§</sup> Didier Espinat,<sup>§</sup> and Alain H. Fuchs\*,<sup>†</sup>

Laboratoire de Chimie-Physique des Matériaux Amorphes, URA-CNRS 1104,  
Université Paris-Sud, 91405 Orsay cédex, France, and Département de Physico-Chimie  
Appliquée et Analyses, Institut Français du Pétrole, 1-4 avenue de Bois-Préau, BP 311,  
92506 Rueil Malmaison, cédex, France

### Faujasite – NaY zeolite



Pioneering study of adsorption  
in a zeolite with cations

# Alkane mixtures in silicalite with GCMC

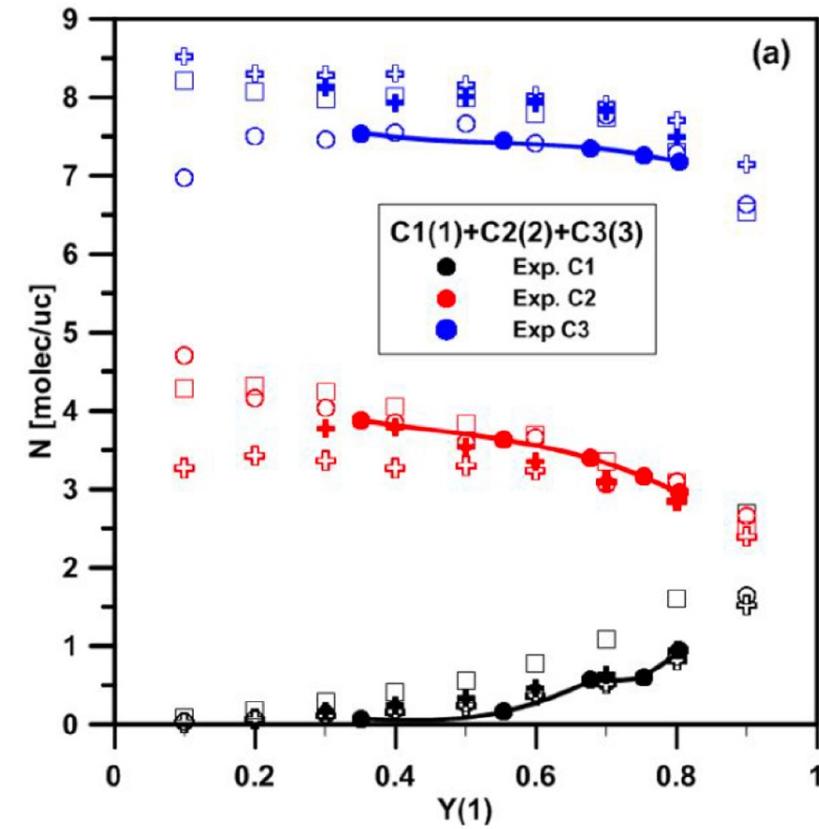
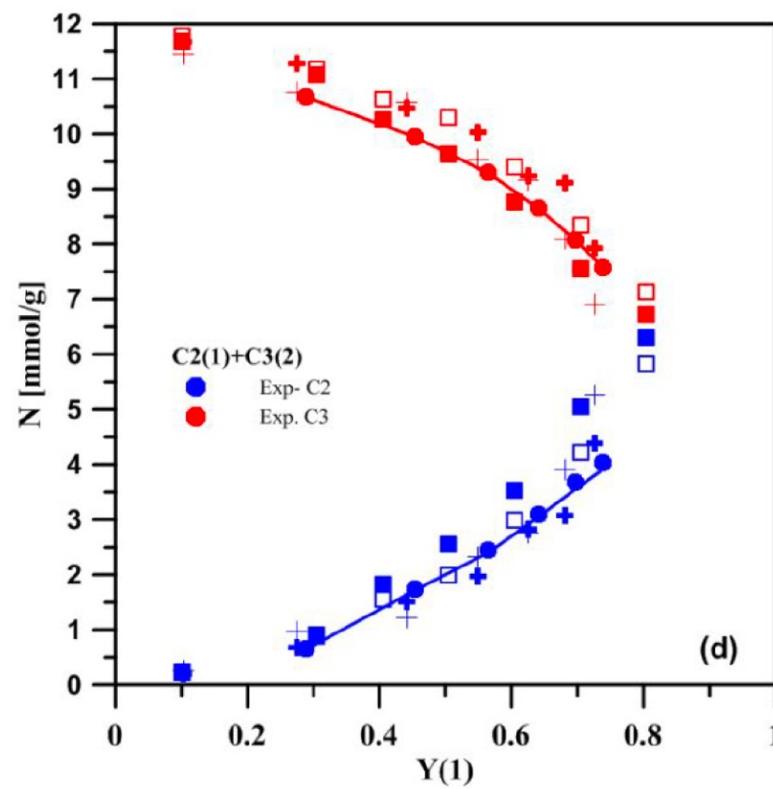
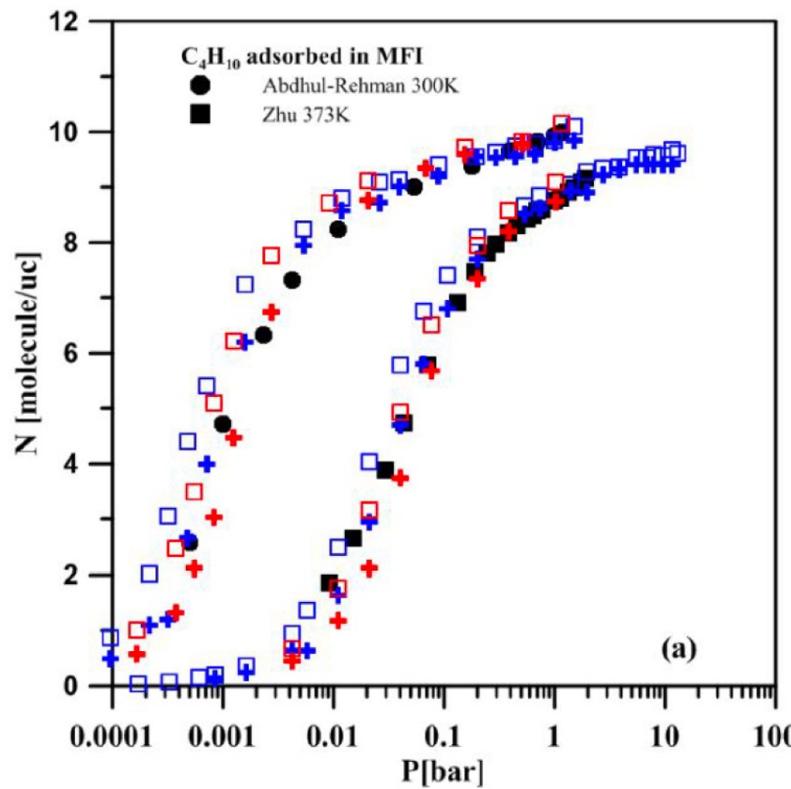
## Use of a New Size-Weighted Combining Rule to Predict Adsorption in Siliceous Zeolites

Published as part of the Journal of Chemical & Engineering Data Hans Hasse special issue.

L. L. Romaniello, M. G. M. V. Pereira, S. Arvelos, and E. J. Maginn\*

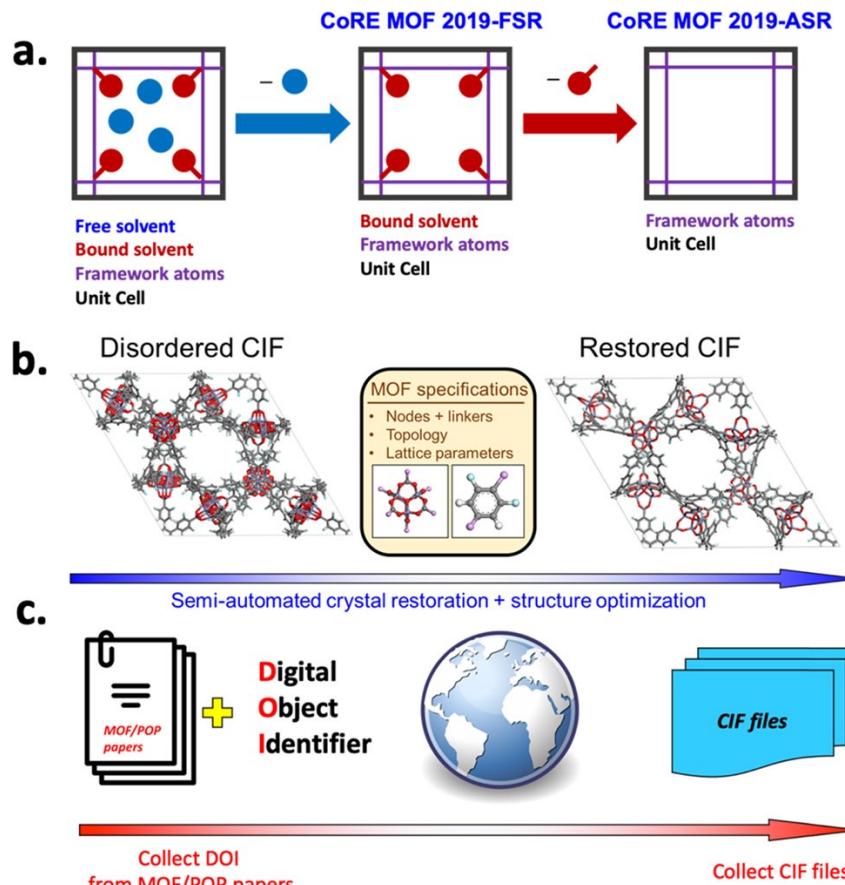
Cite This: *J. Chem. Eng. Data* 2020, 65, 1379–1395

Read Online



Very good agreement with experiment; effect of combining rule and details such as potential cutoff determined

# High throughput simulation of adsorption in MOFs



## Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019

Yongchul G. Chung\*, Emmanuel Haldoupis, Benjamin J. Bucior, Maciej Haranczyk, Seulchan Lee, Hongda Zhang, Konstantinos D. Vogiatis, Marija Milisavljevic, Sanliang Ling, Jeffrey S. Camp, Ben Slater, J. Ilja Siepmann\*, David S. Sholl\*, and Randall Q. Snurr\*

Cite this: *J. Chem. Eng. Data* 2019, 64, 12, 5985–5998

Publication Date: November 4, 2019

<https://doi.org/10.1021/acs.jcd.9b00835>

Copyright © 2019 American Chemical Society

RIGHTS & PERMISSIONS

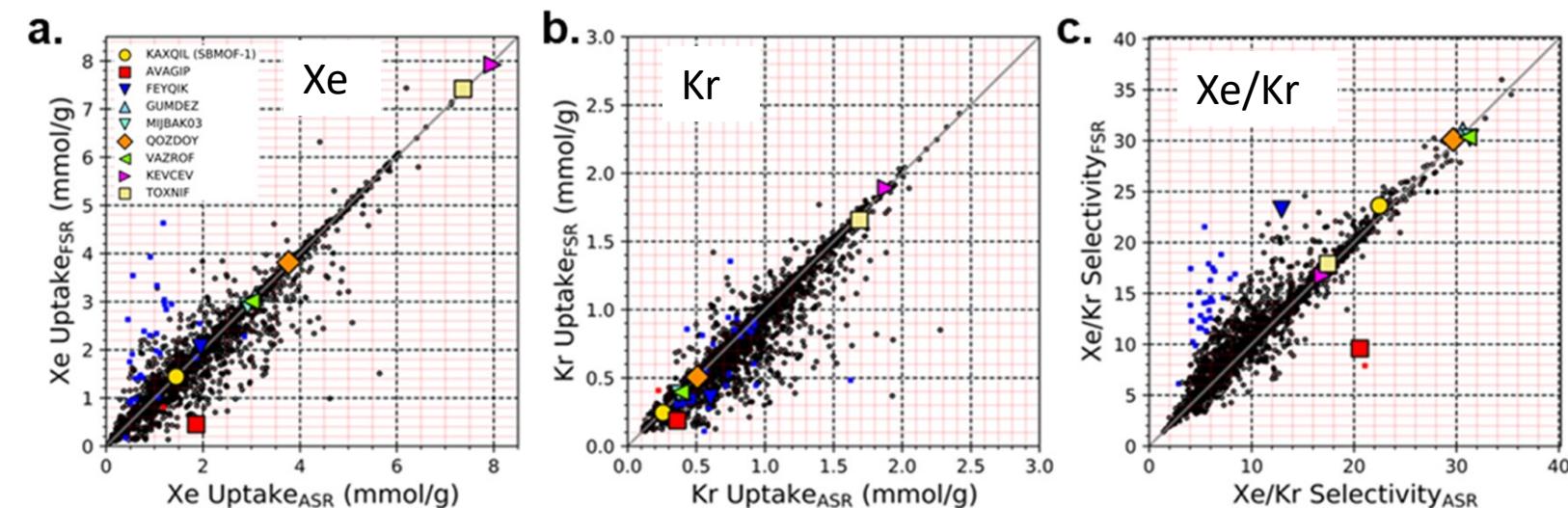
Article Views  
**12119**

Altmetric  
**20**

Citations  
**211**

LEARN ABOUT THESE METRICS

Share   Add to   Export



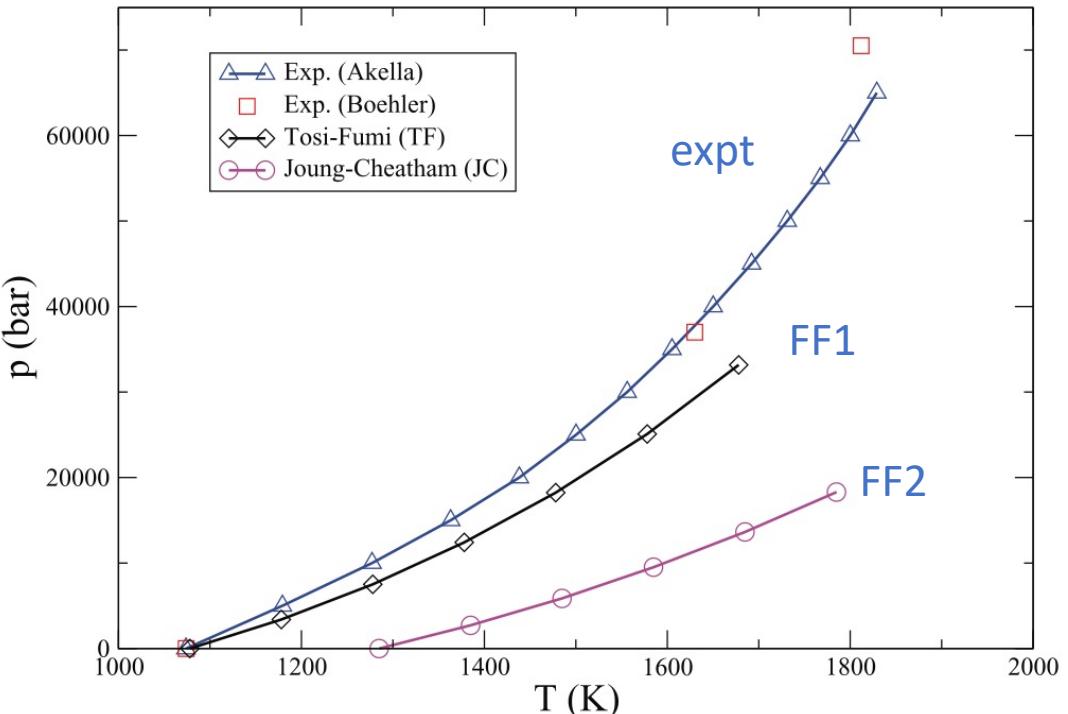
Free database of > 14000 MOF structures; high throughput GCMC simulations to screen materials

# Melting points

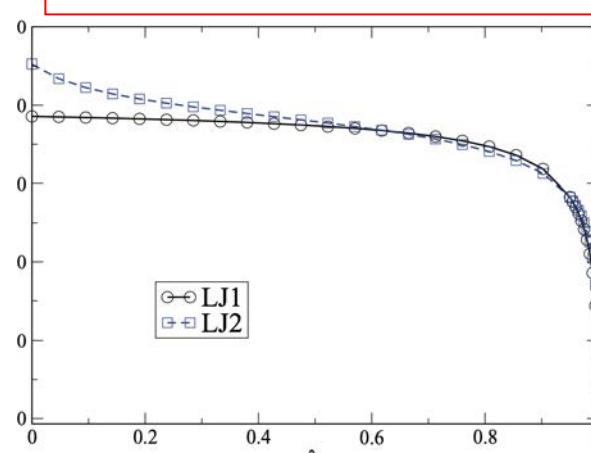
## Calculation of the melting point of alkali halides by means of computer simulations

J. L. Aragones, E. Sanz, C. Valeriani, and C. Vega<sup>a)</sup>

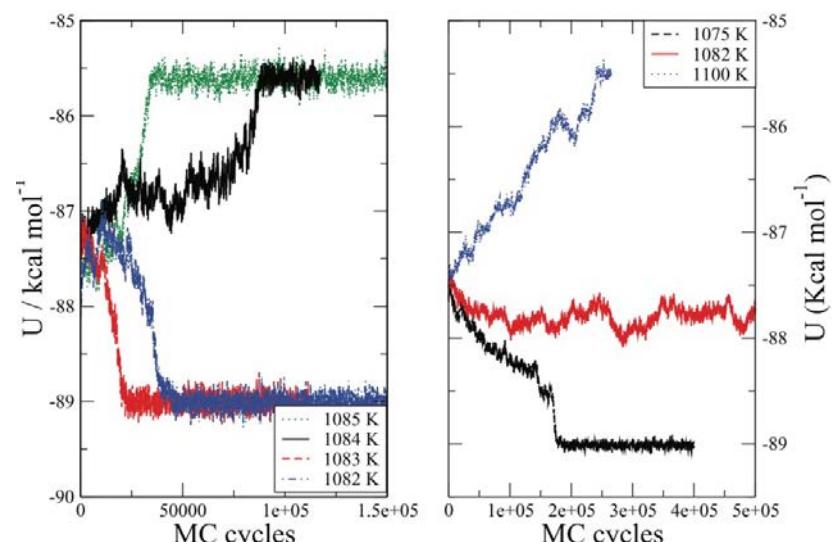
*Departamento de Química Física, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, Madrid 28040, Spain*



Free energy – Einstein  
Crystal method - NaCl

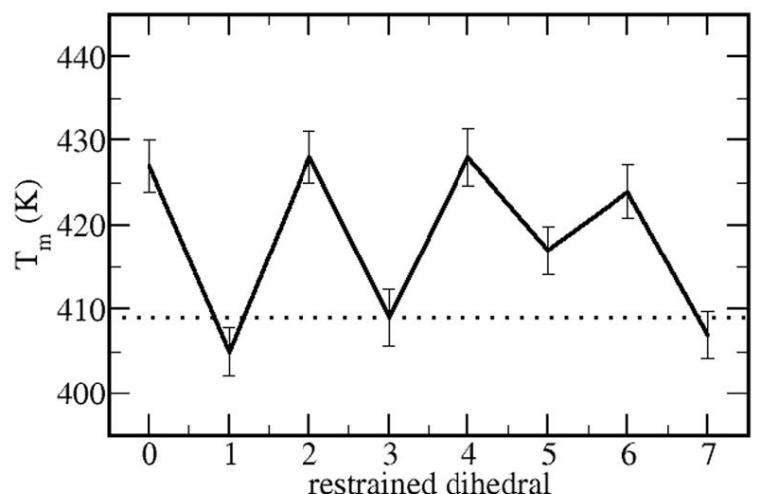
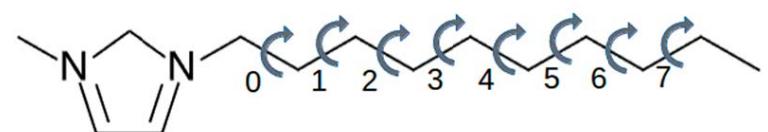
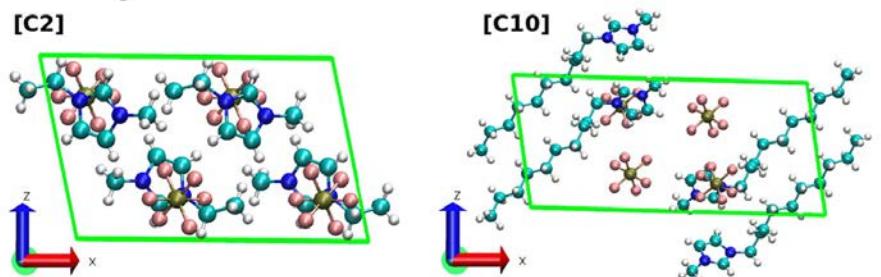
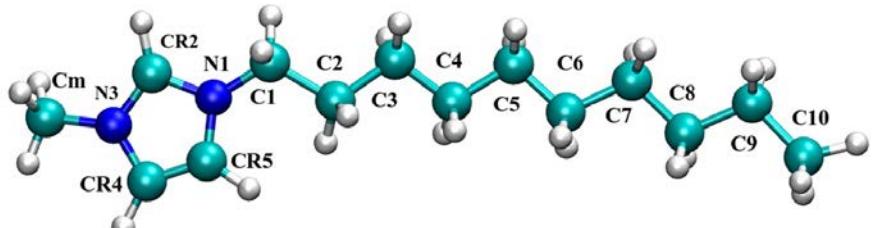


Direct coexistence - NaCl



Used three different methods to compute melting points; compared performance of two force fields

# Melting point – ionic liquids



## Effect of alkyl-group flexibility on the melting point of imidazolium-based ionic liquids

Cite as: J. Chem. Phys. 153, 044504 (2020); doi: 10.1063/5.0015992

Submitted: 31 May 2020 • Accepted: 3 July 2020 •

Published Online: 29 July 2020

Kalil Bernardino,<sup>1,a</sup> Yong Zhang,<sup>2,b</sup> Mauro C. C. Ribeiro,<sup>1,c</sup> and Edward J. Maginn<sup>2,d</sup>

### AFFILIATIONS

<sup>1</sup>Laboratório de Espectroscopia Molecular, Departamento de Química Fundamental, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo 05508-000, Brazil

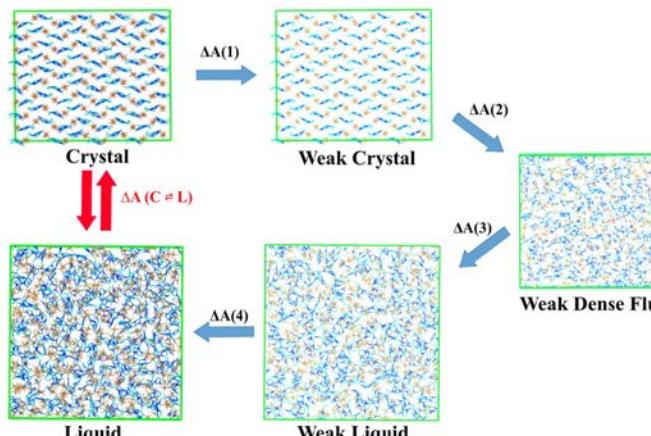
<sup>2</sup>Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA

### Revisiting the Pseudo-supercritical path Method: A Modified Formulation for the Alchemical Calculation of Solid-Liquid Coexistence

Gabriela B. Correa,<sup>\*†</sup> Yong Zhang,<sup>\*‡</sup> Frederico W. Tavares,<sup>\*†</sup> and Edward J. Maginn<sup>\*‡</sup>

<sup>†</sup>Chemical Engineering Program, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21941-909, Brazil

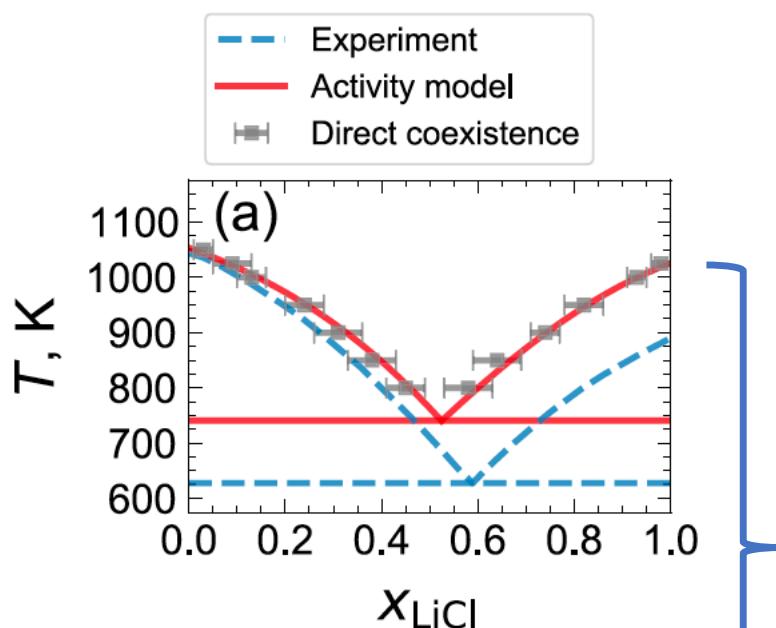
<sup>‡</sup>Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556



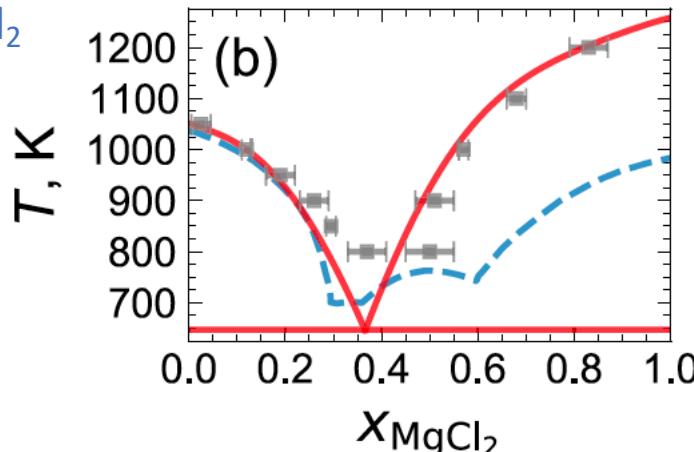
Demonstrated effect of alkyl chain flexibility and location on melting point

# Binary phase diagrams

KCl-LiCl



KCl-MgCl<sub>2</sub>



Pure MP of LiCl  
and MgCl<sub>2</sub> off

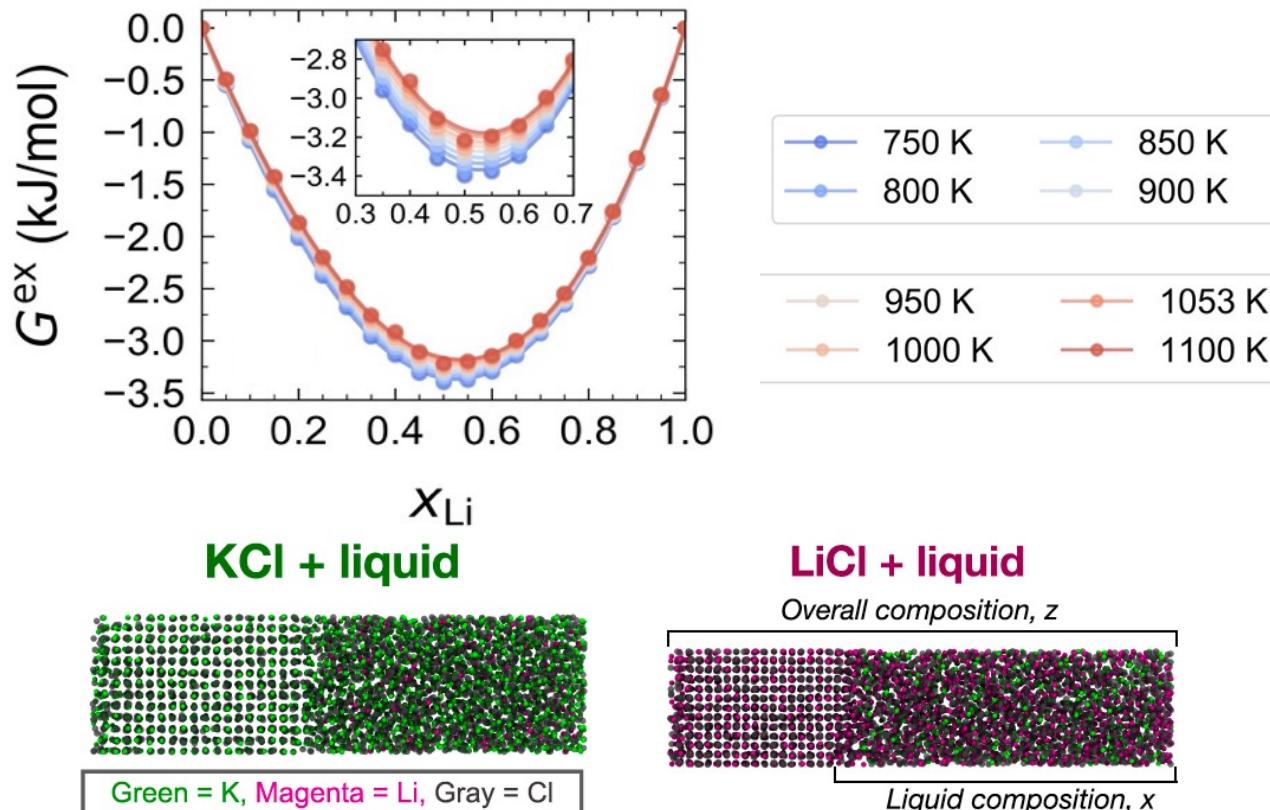
## Computing the Liquidus of Binary Monatomic Salt Mixtures with Direct Simulation and Alchemical Free Energy Methods

Published as part of *The Journal of Physical Chemistry* virtual special issue "125 Years of The Journal of Physical Chemistry".

Ryan S. DeFever and Edward J. Maginn\*

Cite This: *J. Phys. Chem. A* 2021, 125, 8498–8513

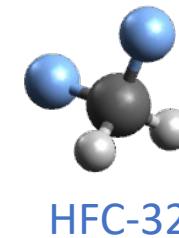
Read Online



Efficient method for computing SLE of mixtures;  
~ 10X faster than direct coexistence

# Gas solubility in ionic liquids – Osmotic ensemble MC

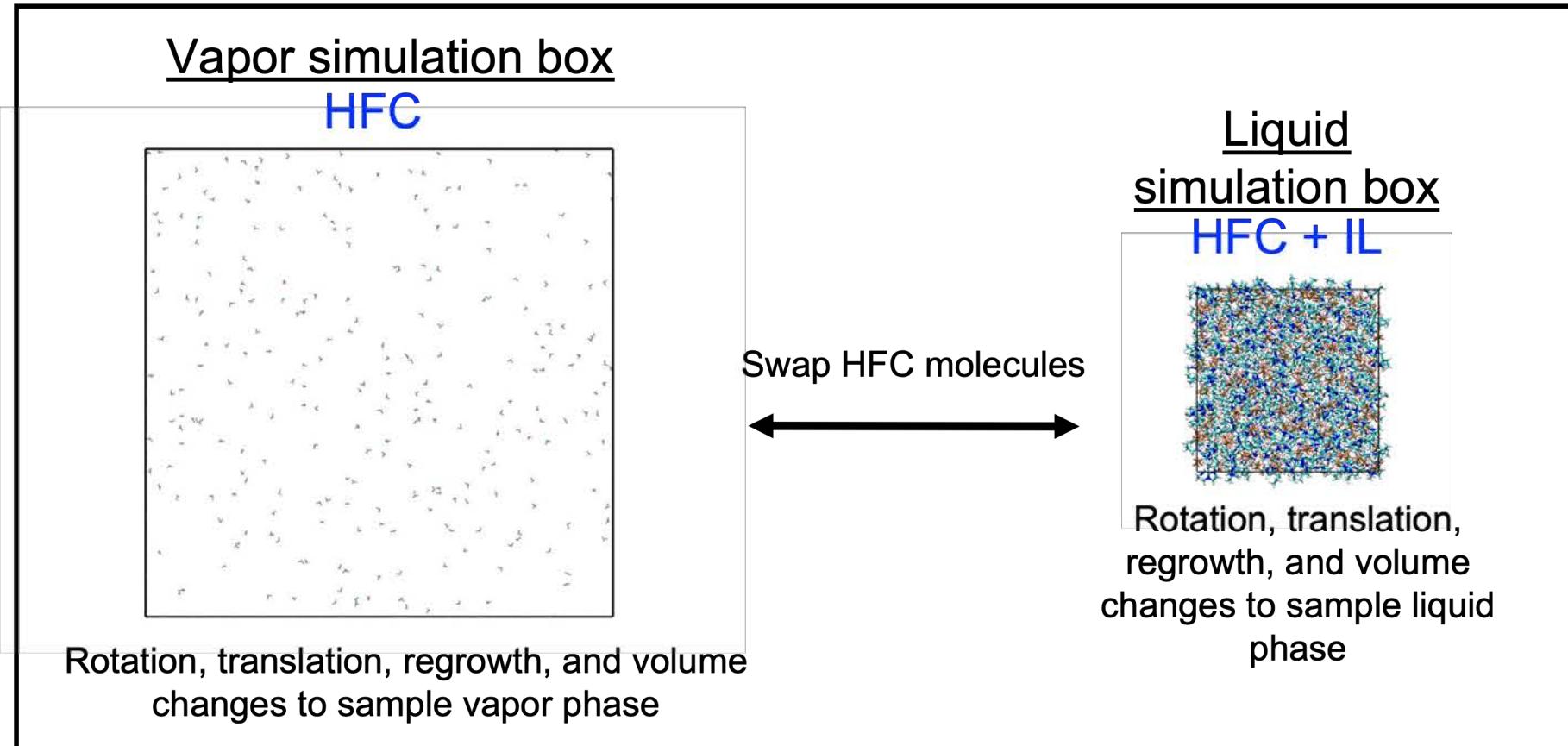
Hydrofluorocarbon (HFC)  
Ryan DeFever, Ning Wang  
(unpublished)



Gas

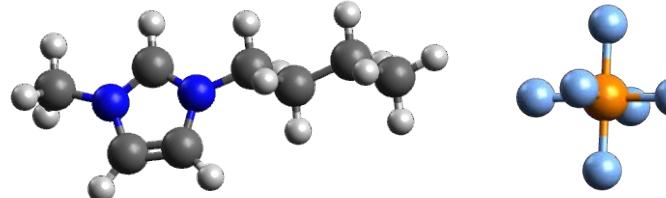
Liquid  
( $x_{\text{HFC}} = ?$ )

$$\begin{aligned}P^V &= P^L \\T^V &= T^L \\N_{\text{IL}}^L \\N_{\text{HFC}}^V + N_{\text{HFC}}^L \\ \mu_{\text{HFC}}^V &= \mu_{\text{HFC}}^L\end{aligned}$$



Markov Chain Monte Carlo simulation allows system to proceed to equilibrium. HFC solubility is determined from the composition of liquid simulation box after equilibrium is established.

# Monte Carlo convergence

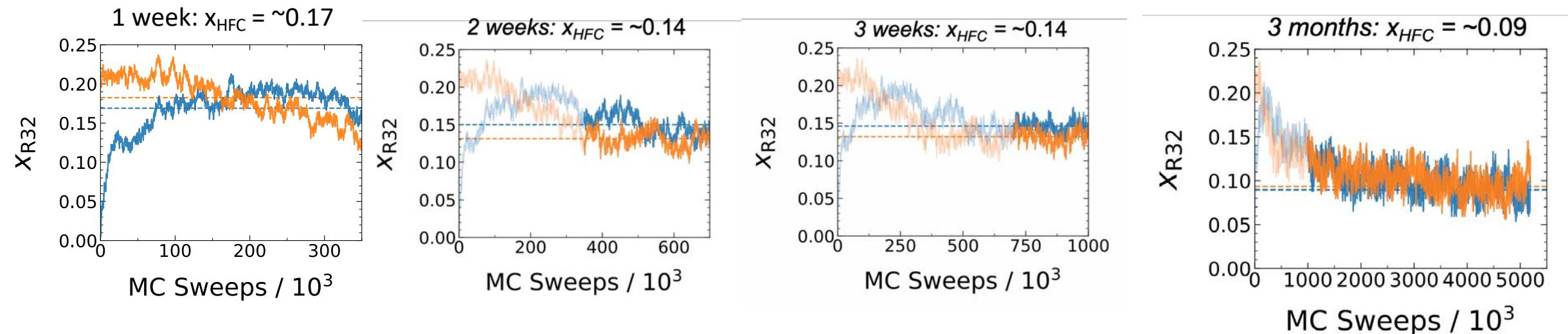


Gas

Liquid  
( $x_{HFC} = ?$ )

Orange = start with 20 mol% HFC  
Blue = start with 0 mol% HFC

Example results: HFC-32/[bmim][PF6], T = 298 K, P = 1.0 bar



The HFC-32 concentration in the ionic liquid equilibrates extremely slowly for these systems with osmotic ensemble Monte Carlo.

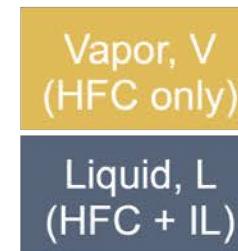
# Computing gas solubility faster with MD

At equilibrium  $\mu_i^V = \mu_i^L$ . This provides an alternative approach to computing the solubility.

$\mu_{\text{HFC}}^V$  = from equation of state (EOS) model

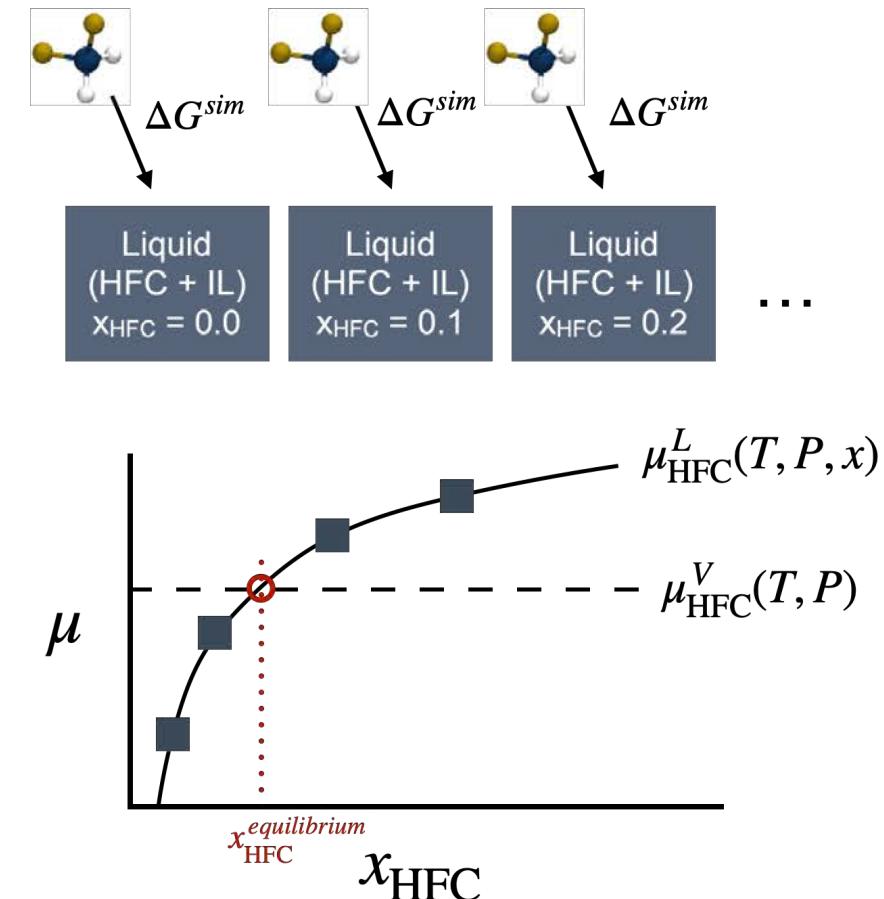
$$\mu_{\text{HFC}}^L \equiv G^L(N_{\text{HFC}} + 1, N_{\text{IL}}) - G^L(N_{\text{HFC}}, N_{\text{IL}})$$

$$\mu_{\text{HFC}}^L = \Delta G^{\text{sim}} - kT \ln \frac{V}{\Lambda_{\text{HFC}}^3 N_{\text{HFC}}}$$



## Assumptions:

- The vapor phase behavior of the molecular model is well described by an EOS.
- The liquid phase chemical potential is pressure independent to 1 MPa.



Instead of the “direct” MC method, we explicitly compute the chemical potential of the HFC in the liquid as a function of HFC concentration.

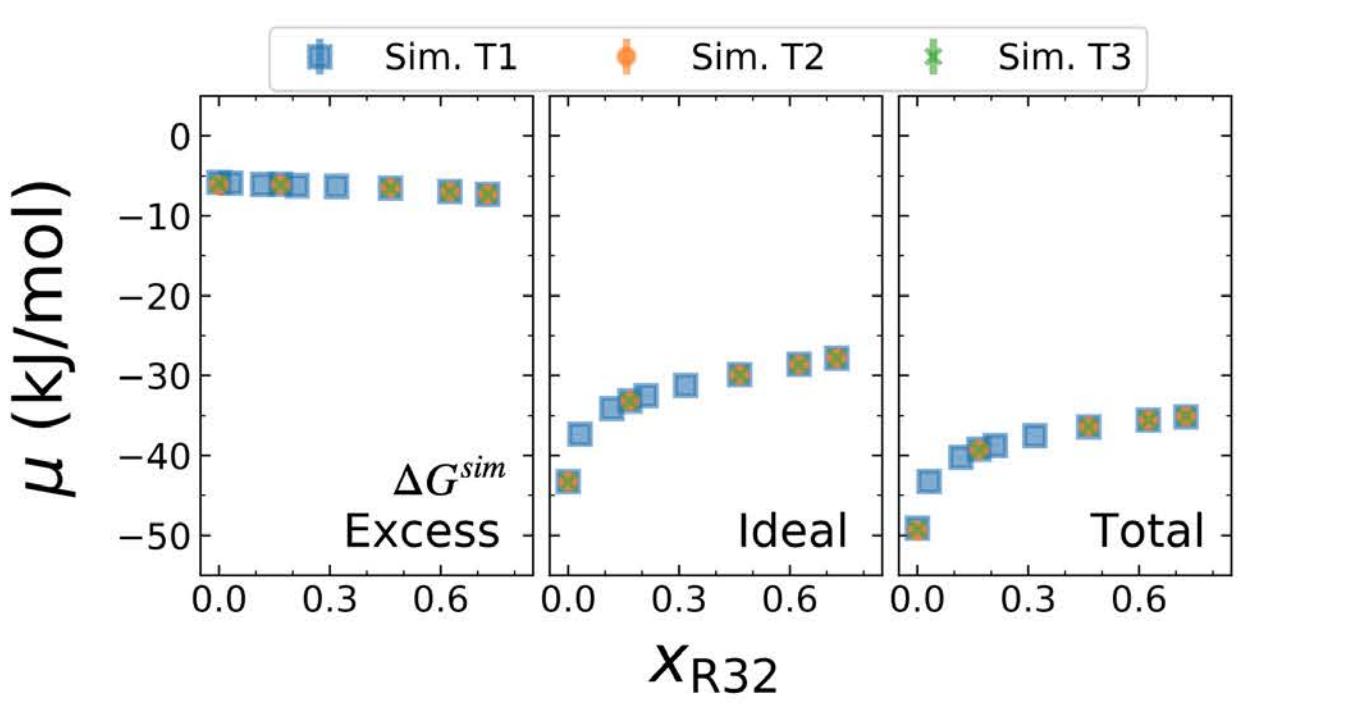
# How accurate is the method?

Test case: HFC-32/[bmim][PF<sub>6</sub>]

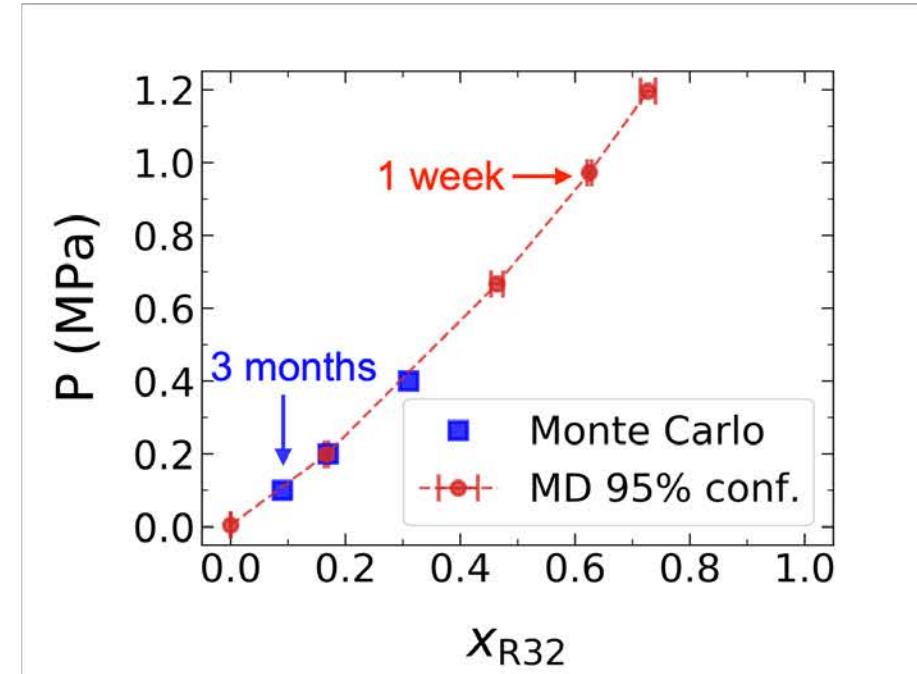
$$\mu_{\text{HFC}}^L = \mu_{\text{HFC}}^{L,\text{ex}} + \mu_{\text{HFC}}^{L,\text{id}}$$

$$\mu_{\text{HFC}}^L = \Delta G^{\text{sim}} - kT \ln \frac{V}{\Lambda_{\text{HFC}}^3 N_{\text{HFC}}}$$

Liquid chemical potential – three independent trials



Solubility prediction



Our method yields extremely accurate solubility prediction (uncertainty on the order of 1 mol%) with a much reduced wall clock time.

Lines to guide the eye

# Thermophysical and transport properties

# Collective properties much harder to compute than self-diffusivity

“Green-Kubo” correlation function

$$\gamma = \int_0^\infty dt \langle \dot{\xi}(t) \dot{\xi}(0) \rangle$$

$$D = \frac{1}{3} \int_0^\infty dt \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle$$

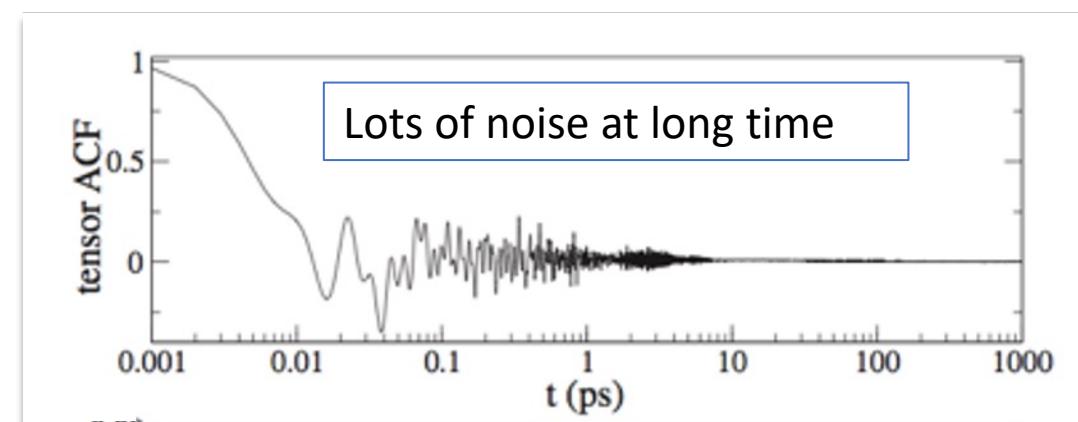
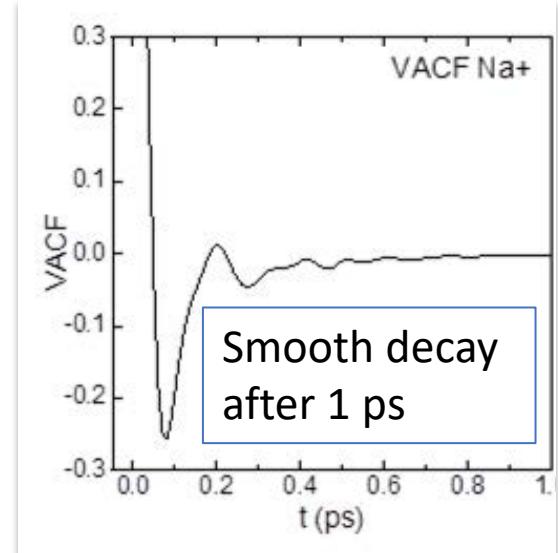
N independent averages, one for each molecule per frame

“Einstein” formulation

$$\gamma = \lim_{t \rightarrow \infty} \frac{\langle (\xi(t) - \xi(0))^2 \rangle}{2t} = \frac{1}{2} \lim_{t \rightarrow \infty} \frac{d}{dt} \langle (\xi(t) - \xi(0))^2 \rangle$$

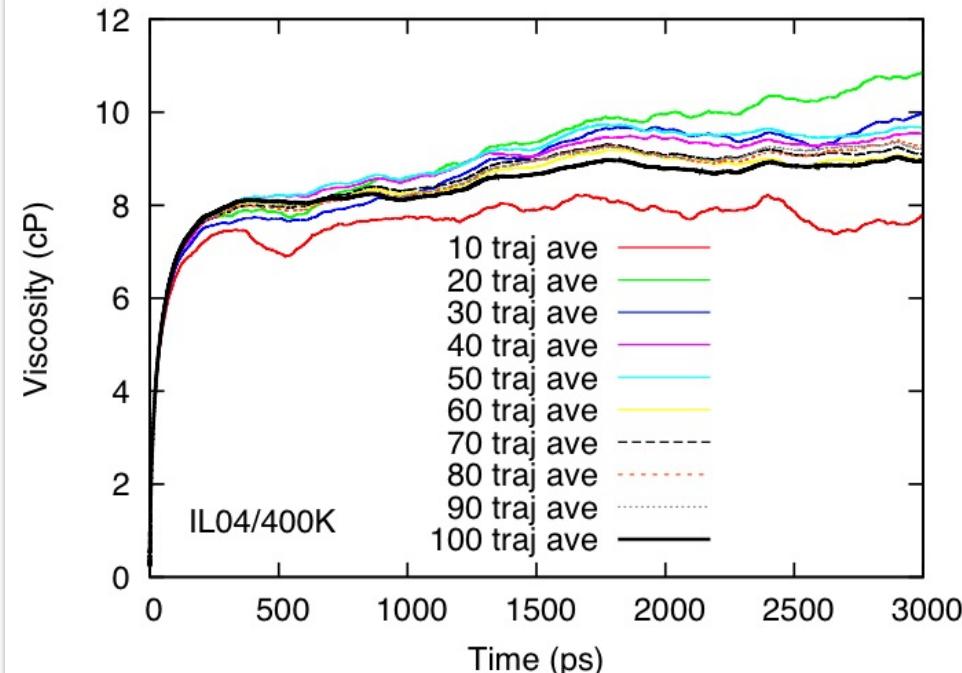
$$\eta = \frac{V}{k_B T} \int_0^\infty dt \langle P_{xy}(t) P_{xy}(0) \rangle$$

One value of stress tensor per frame



# “Time Decomposition” Method

Objective, automated viscosity calculation



Code that implements this:  
<https://github.com/MaginnGroup/PyLAT>

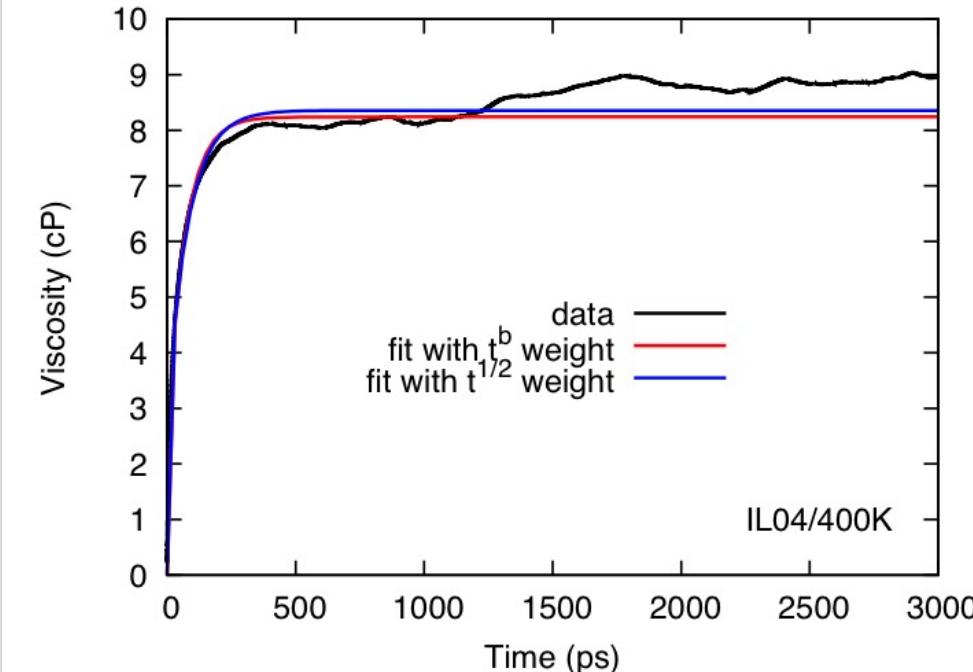
JCIM  
JOURNAL OF  
CHEMICAL INFORMATION  
AND MODELING

Cite This: *J. Chem. Inf. Model.*, 2019, 59, 1301–1305

Application Note  
pubs.acs.org/jcim

PyLAT: Python LAMMPS Analysis Tools

Michael T. Humbert,<sup>1</sup> Yong Zhang,<sup>1</sup> and Edward J. Maginn<sup>\*1,2</sup>



Fit time-dependent viscosity to a double exponential, with weight proportional to standard deviation

$$\eta(t) = A\alpha\tau_1(1 - e^{-t/\tau_1}) + A(1 - \alpha)\tau_2(1 - e^{-t/\tau_2})$$

Zhang, Otani, Maginn, *J. Chem. Theory Comput.* 2015, 11, 3537–3546

# $\text{CH}_4/\text{CO}_2$ at high T,P

Force field comparison and thermodynamic property calculation of supercritical  $\text{CO}_2$  and  $\text{CH}_4$  using molecular dynamics simulations

Cassiano G. Aimoli<sup>a,b,c</sup>, Edward J. Maginn<sup>c,\*</sup>, Charles R.A. Abreu<sup>b,d</sup>

<sup>a</sup> Santos Basin Exploration & Production Operations Unit, Petróleo Brasileiro S.A., Santos, SP 11015-001, Brazil

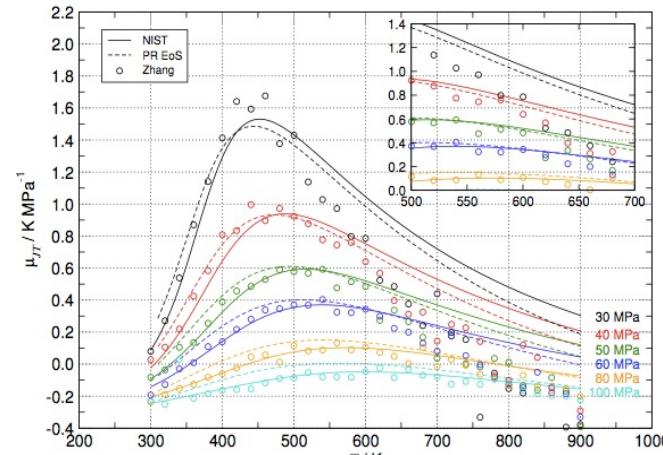
<sup>b</sup> School of Chemical Engineering, State University of Campinas, Campinas, SP 13083-970, Brazil

<sup>c</sup> Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

<sup>d</sup> School of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, RJ 21941-909, Brazil

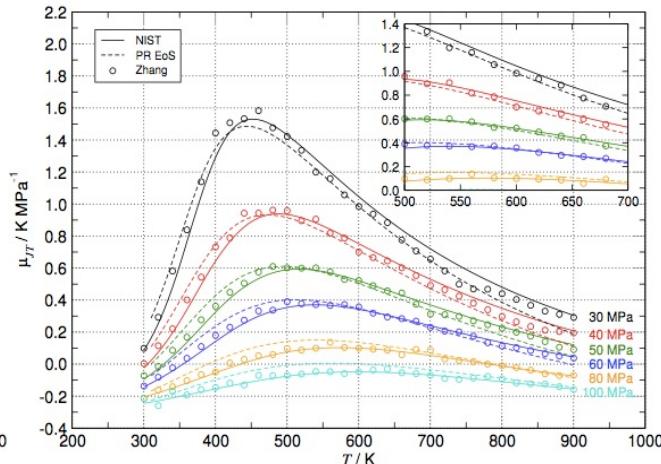
## Methane Joule-Thomson coefficient

STD

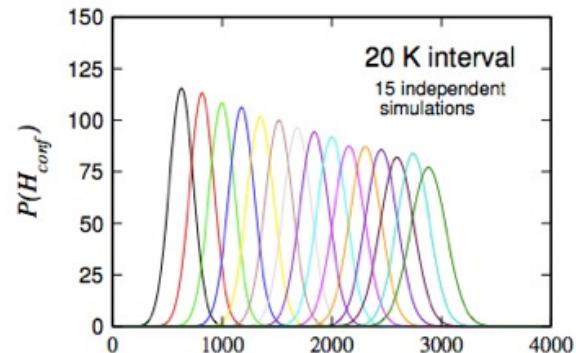


AAD = 197 %!!

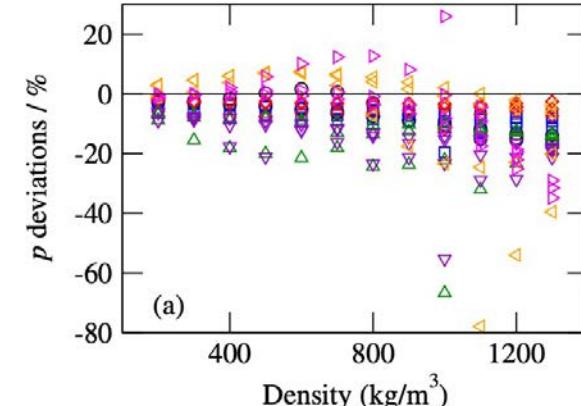
MBAR



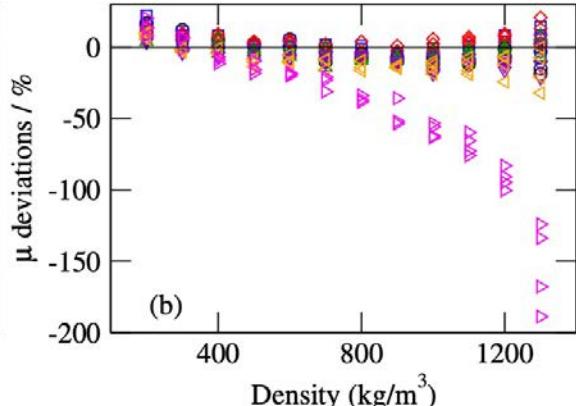
AAD = 7%



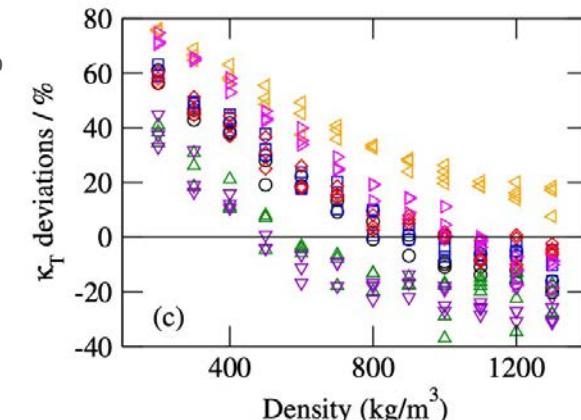
## $\text{CO}_2$ transport properties vs density



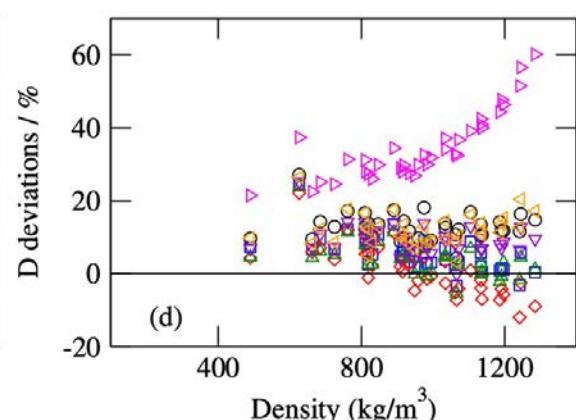
(a)



(b)



(c)



(d)

Large variations depending on force field

# Helpful resources



A LiveCoMS Best Practices Guide

## Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics [Article v1.0]

Edward J. Maginn<sup>1\*</sup>, Richard A. Messerly<sup>2\*</sup>, Daniel J. Carlson<sup>3</sup>, Daniel R. Roe<sup>4</sup>, J. Richard Elliott<sup>5</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, The University of Notre Dame;

<sup>2</sup>Thermodynamics Research Center, National Institute of Standards and Technology;

<sup>3</sup>Chemical Engineering Department, Brigham Young University; <sup>4</sup>Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health; <sup>5</sup>Department of Chemical and Biomolecular Engineering, The University of Akron

<https://github.com/ejmaginn/TransportCheckList>

## PyLAT: Python LAMMPS Analysis Tools

Michael T. Humbert,<sup>1,2</sup> Yong Zhang,<sup>1,2</sup> and Edward J. Maginn<sup>\*1,3</sup>

<https://github.com/MaginnGroup/PyLAT>



<https://github.com/MDAnalysis/mdanalysis>



1528

Biophysical Journal Volume 109 October 2015 1528–1532

## Computational Tools

### MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories

Robert T. McGibbon,<sup>1,\*</sup> Kyle A. Beauchamp,<sup>2</sup> Matthew P. Harrigan,<sup>1</sup> Christoph Klein,<sup>3</sup> Jason M. Swails,<sup>4</sup> Carlos X. Hernández,<sup>5</sup> Christian R. Schwantes,<sup>1</sup> Lee-Ping Wang,<sup>6</sup> Thomas J. Lane,<sup>7</sup> and Vijay S. Pande<sup>1,5</sup>

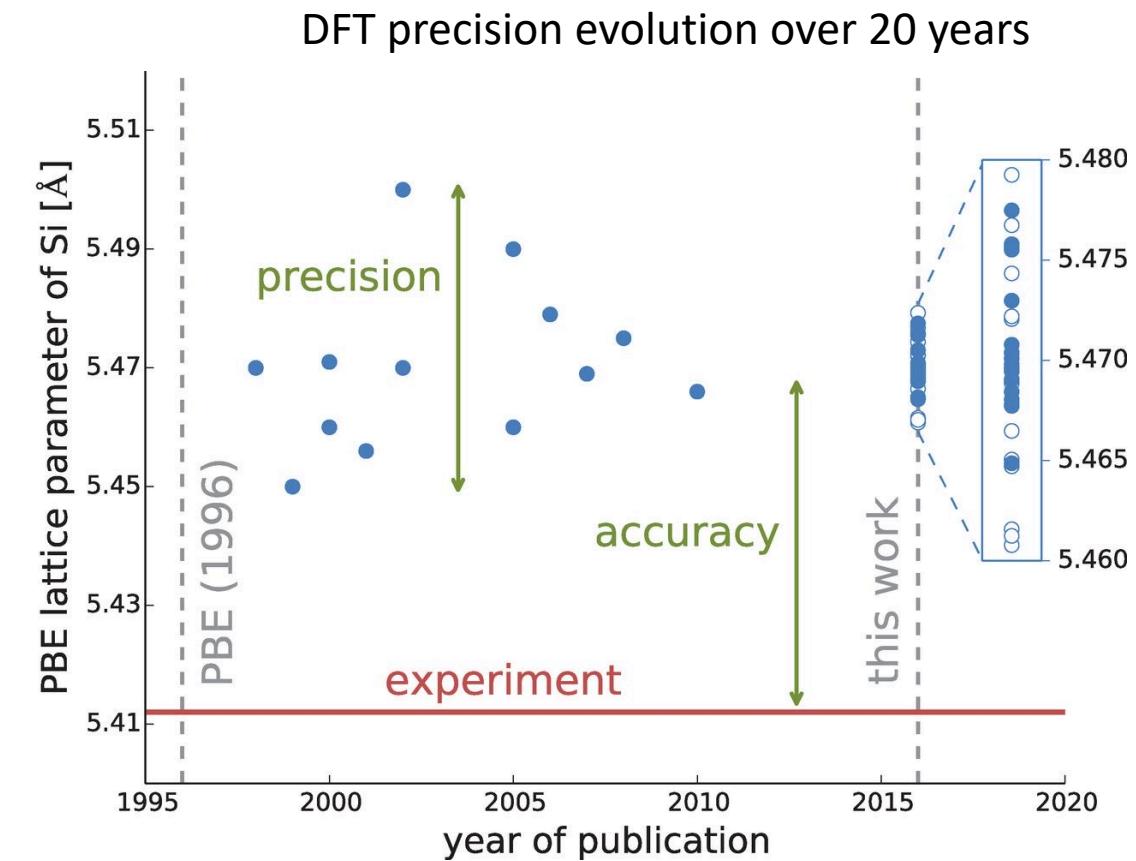
<sup>1</sup>Department of Chemistry, Stanford University, Stanford, California; <sup>2</sup>Computational Biology Program, Sloan-Kettering Institute, New York, New York; <sup>3</sup>Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee; <sup>4</sup>Department of Chemistry, Rutgers University, Piscataway, New Jersey; <sup>5</sup>Biophysics Program, Stanford University, Stanford, California; <sup>6</sup>Department of Chemistry, University of California, Davis, Davis, California; and <sup>7</sup>SLAC National Accelerator Laboratory, Menlo Park, California

# What is needed and what is coming?

- **TRUE** simulations: Transferable, Reproducible, Useable by others, Extensible
- Machine learning
- Integration of experiments, EOS modeling, molecular modeling: Molecule design

# Reproducibility is a real problem

- Emphasis on TRUE simulations:  
*Transparent, Reproducible, Usable by others, and Extensible*
  - If you get the “right answer” but nobody can see how you did it or get the same answer, what good is it?
- Reproducibility “crisis” in science
- Dramatic increase in computing power
  - Get the wrong answer faster!
- *Need improved tools to help users*



Kurt Lejaeghere et al. Science 26 March 2016;351, 6280, 1415

# Example: Computational Physics

Stodden V, Krafczyk MS, Bhaskar A. Enabling the Verification of Computational Results: An Empirical Evaluation of Computational Reproducibility. Paper presented at the Proceedings of the First International Workshop on Practical Reproducible Evaluation of Computer Systems. Tempe, AZ: New York: Association for Computing Machinery; 2018a, June 11. 10.1145/3214239.3214242.

An examination of the availability of artifacts for 307 articles published in the *Journal of Computational Physics*

- More than one-half (50.9%) of the articles were impossible to reproduce.
- About 6% of the articles (17) made artifacts available in the publication itself
- About 36% discussed the artifacts (e.g., mentioned code) in the article.
- Of the 298 authors who were emailed with a request for artifacts:
  - 37% did not reply
  - 48% replied but did not provide any artifacts
  - 15% supplied some artifacts.

## Round Robin Study: Molecular Simulation of Thermodynamic Properties from Models with Internal Degrees of Freedom

Michael Schappals,<sup>†</sup> Andreas Mecklenfeld,<sup>‡</sup> Leif Kröger,<sup>§</sup> Vitalie Botan,<sup>§</sup> Andreas Köster,<sup>||</sup> Simon Stephan,<sup>†</sup> Edder J. García,<sup>†</sup> Gabor Rutkai,<sup>||</sup> Gabriele Raabe,<sup>‡</sup> Peter Klein,<sup>⊥</sup> Kai Leonhard,<sup>§</sup> Colin W. Glass,<sup>#</sup> Johannes Lenhard,<sup>▽</sup> Jadran Vrabec,<sup>||</sup> and Hans Hasse<sup>\*,†</sup>

**Codes:** DL\_POLY, GROMACS, IMC, LAMMPS, ms2, NAMD, Tinker, Towhee

**Force fields:** OPLS, TraPPE, “OPLSAMBER”

**Systems:** ethane, propane, n-butane, i-butane

**Properties:** density and potential energy vs T, P

### Some interesting conclusions:

- Once some level of complexity is reached, not only statistical but also ***systematic errors become unavoidable in computer simulation***
- ...whether a ***source code is disclosed*** to the scrutiny of the scientific community or not is ***deeply related to their reliability***
- the user is “the most significant error source” and that “one does not become a theorist by buying chalk, experimentalist by buying a microscope, or a computational scientist by downloading software”

# MoSDeF: water in slit pore

Single script runs calculations

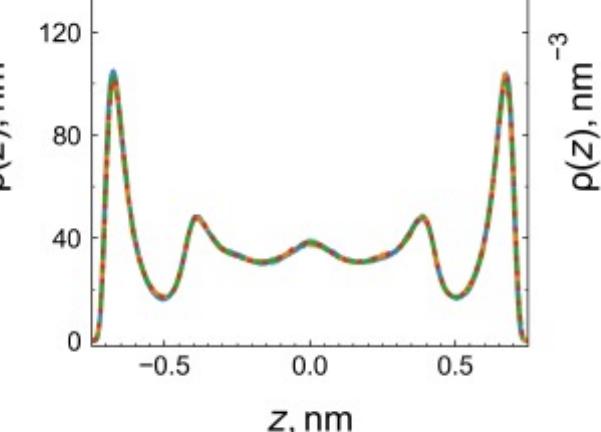
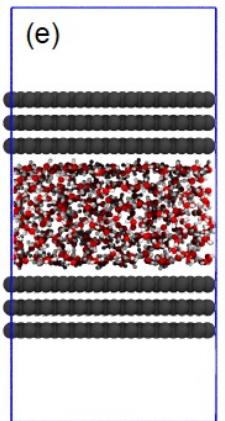
```

1  Desktop - rdefever@maginncr:~/repos/mosdef/cassandra/mosdef_cassandra/examples -- ssh -Y rdefever@maginncr.crc.nd.edu -- 120+40
2
3 import mbuild
4 import foyer
5 import mosdef_cassandra as mc
6
7 def run_nvt():
8
9     # Use mbuild to create molecules
10    methane = mbuild.load('C',smiles=True)
11
12    # Create an empty mbuild.Box
13    box = mbuild.Box(lengths=[3.,3.,3.])
14
15    # Load forcefields
16    oplsaa = foyer.forcefields.load_OPLSAA()
17
18    # Use foyer to apply forcefields
19    methane_ff = oplsaa.apply(methane)
20
21    # Create box and species list
22    box_list = [box]
23    species_list = [methane_ff]
24
25    # Use Cassandra to insert some initial number of species
26    nols_to_add = [[50]]
27
28    # Define the system object
29    system = mc.System(box_list,species_list,
30                        nols_to_add=nols_to_add)
31
32    # Get the move probabilities
33    moves = mc.Moves('nvt', species_list)
34
35    # Run a simulation with at 300 K with 10000 MC moves
36    mc.run(system,moves,300.0,'equilibration',10000)
37
38 if __name__ == "__main__":
39     run_nvt()

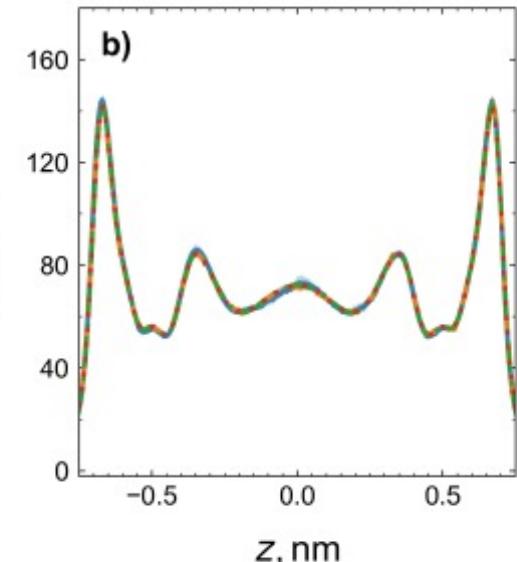
```

Keys:

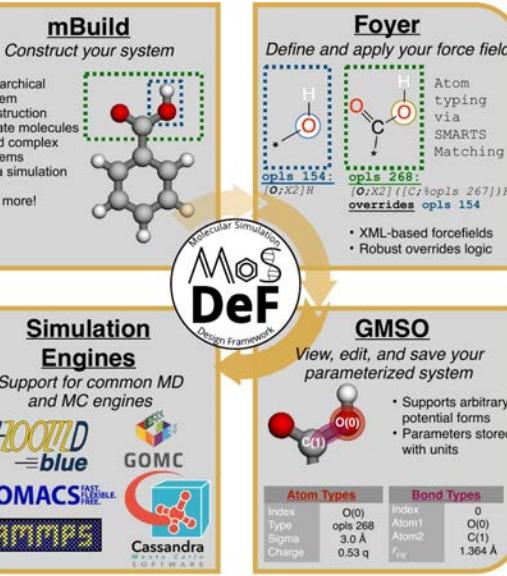
1. Open source codes
2. All files, parameters required are provided electronically
3. Analysis codes provided
4. Ideally, script all operations



Oxygen density profile



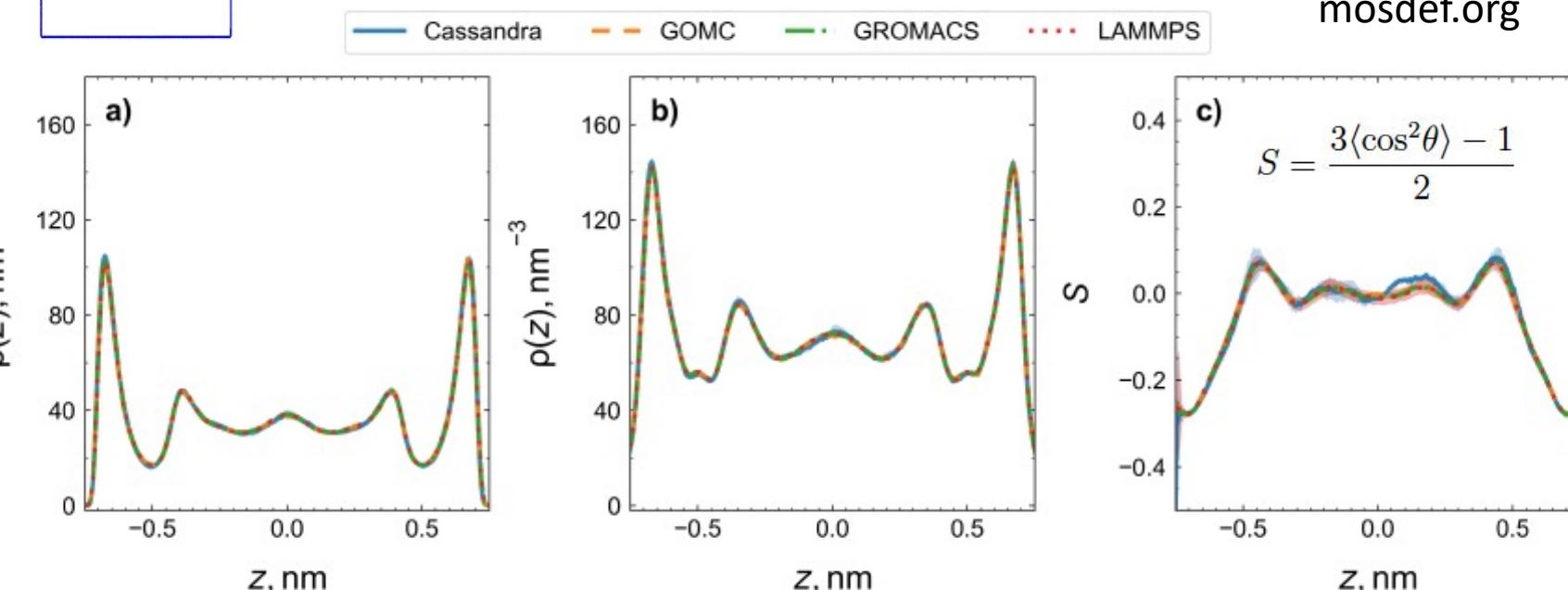
Hydrogen density profile



## Open-source molecular modeling software in chemical engineering focusing on the Molecular Simulation Design Framework

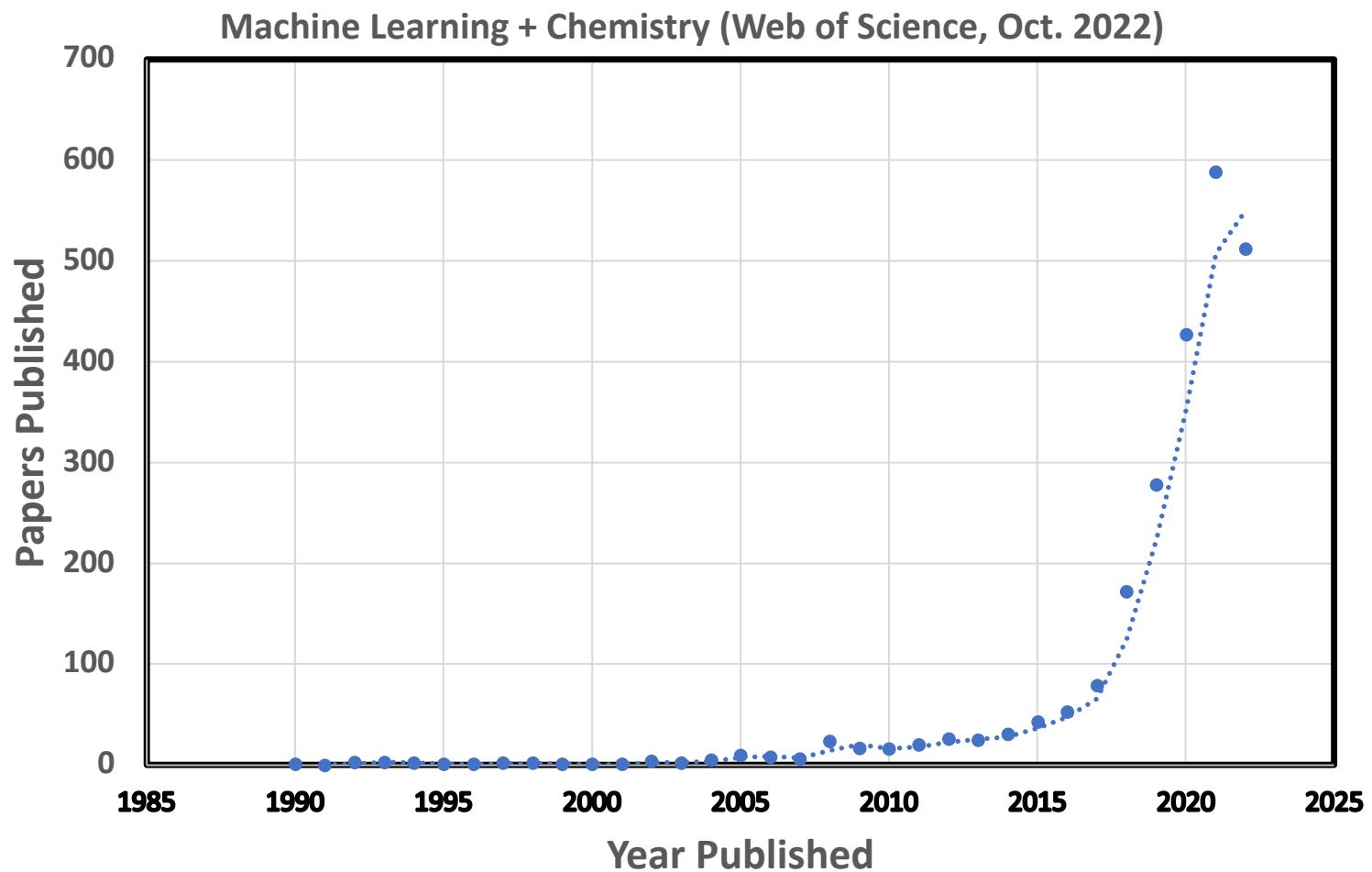
Peter T. Cummings Clare McCabe, Christopher R. Iacobella, Akos Ledeczi, Eric Jankowski, Arathi Jayaraman, Jeremy C. Palmer, Edward J. Maginn, Sharon C. Glotzer ... See all authors

First published: 23 January 2021 | <https://doi.org/10.1002/aic.17206> | Citations: 7



Order parameter

# Machine Learning



Rapid growth over last ~5 years; “Machine Learning” refers to many different things

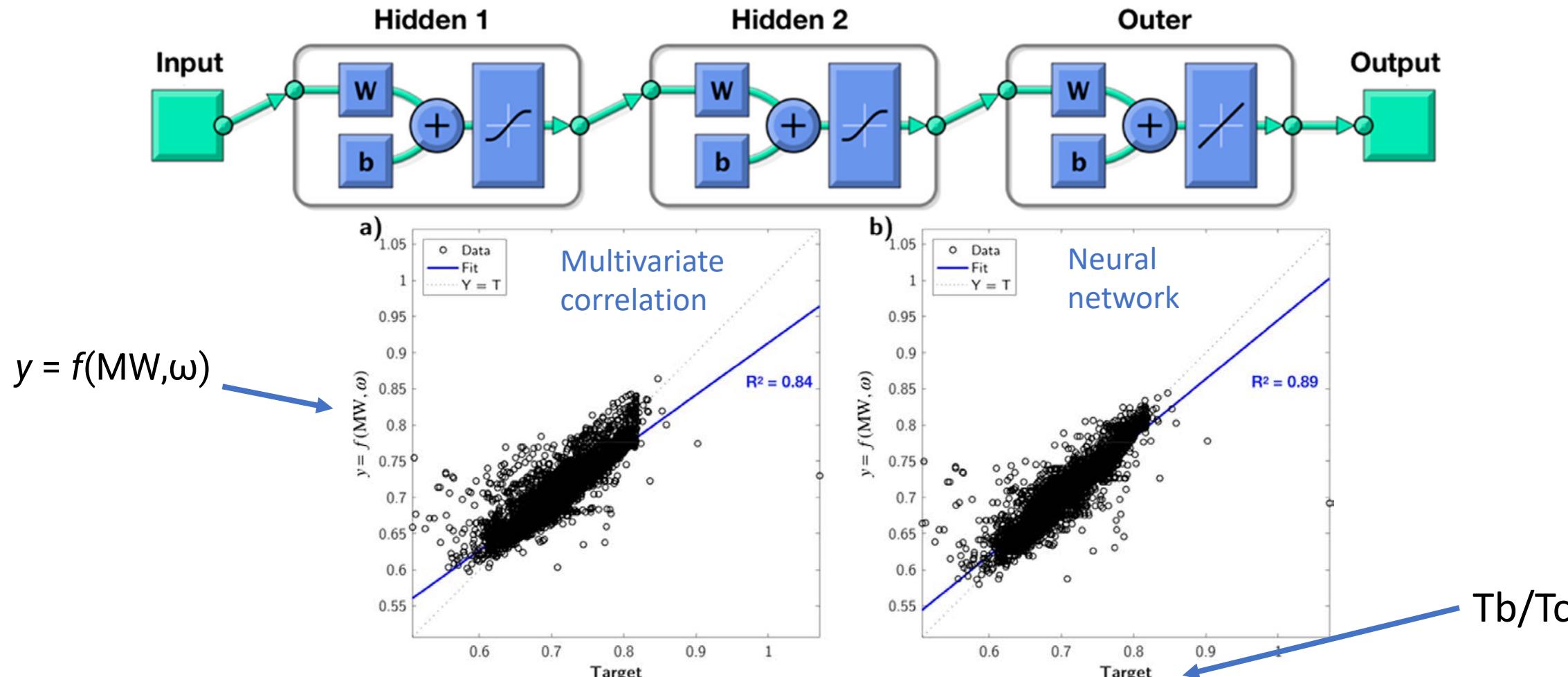
More from Venkat,  
Wednesday afternoon

# Machine learning: Properties

## Machine Learning for Fluid Property Correlations: Classroom Examples with MATLAB

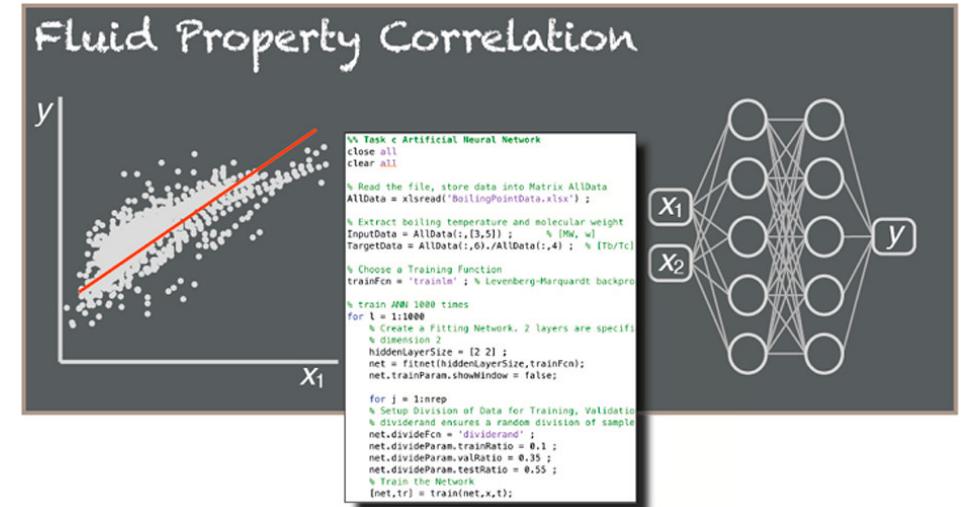
Lisa Joss<sup>①</sup> and Erich A. Müller<sup>\*②</sup>

Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom

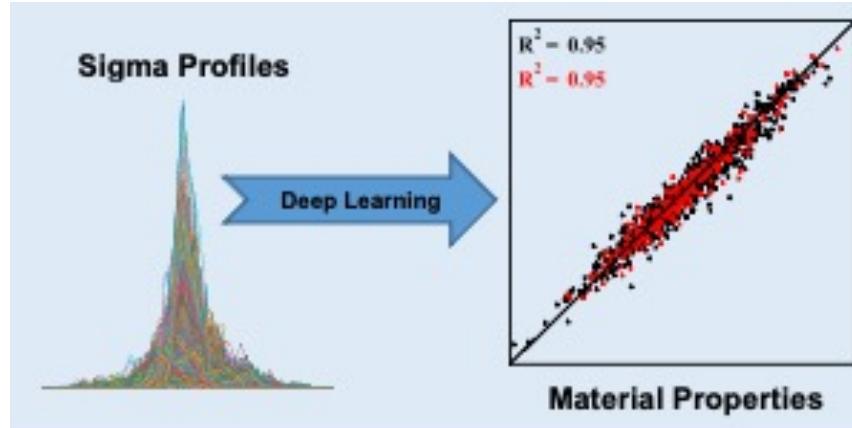


# Limitations

- Input layer to ANN (molecular “features”)
  - Often lack chemical information beyond connectivity, atom identity
  - Size depends on molecular size: Input to ANN as large as largest molecule
    - Huge amounts of data required
- Input features are abstract; lack physical meaning
  - Inverse design problem – how to use information for discovery?
- Limited predictions outside the training set
  - “Extrapolation” among similar molecules, not “discovery” of new molecules



# Convolution neural networks using sigma profiles for property prediction

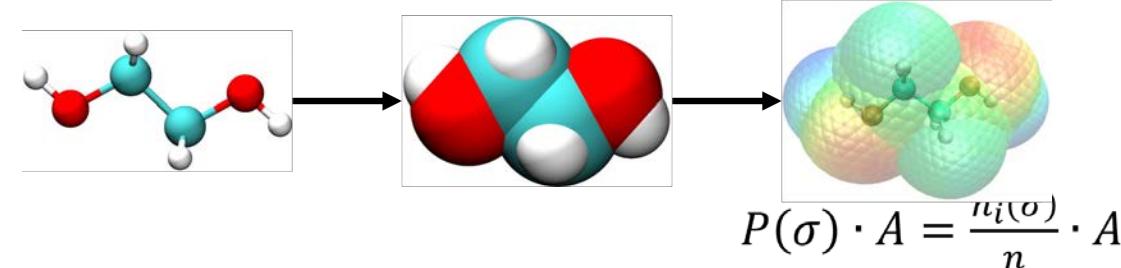


Cite this: DOI: 10.1039/d2cc01549h

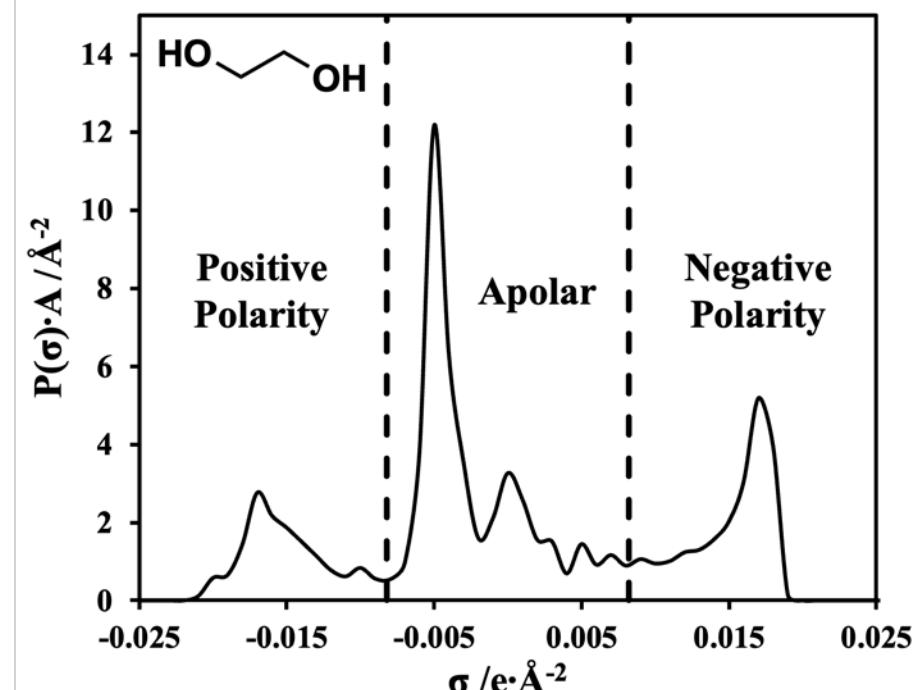
Received 17th March 2022,  
Accepted 5th April 2022

DOI: 10.1039/d2cc01549h

## Sigma profiles in deep learning: towards a universal molecular descriptor†

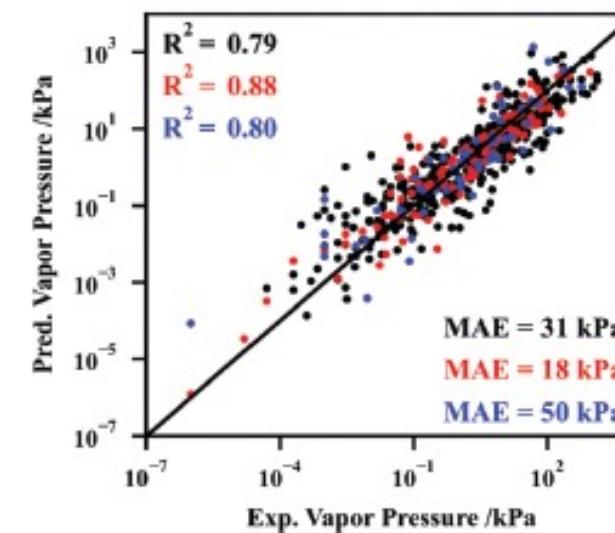
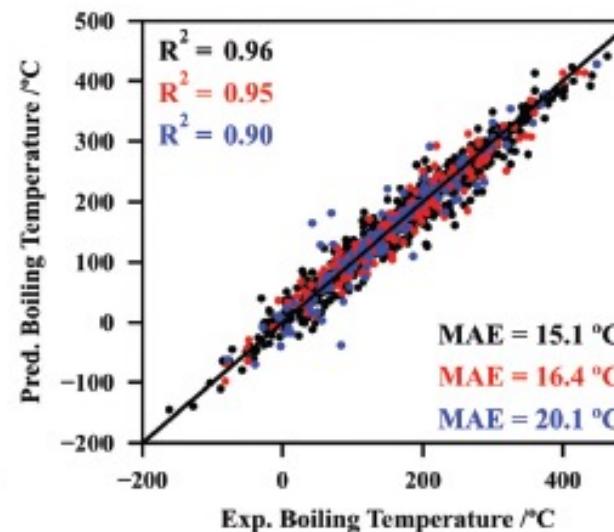
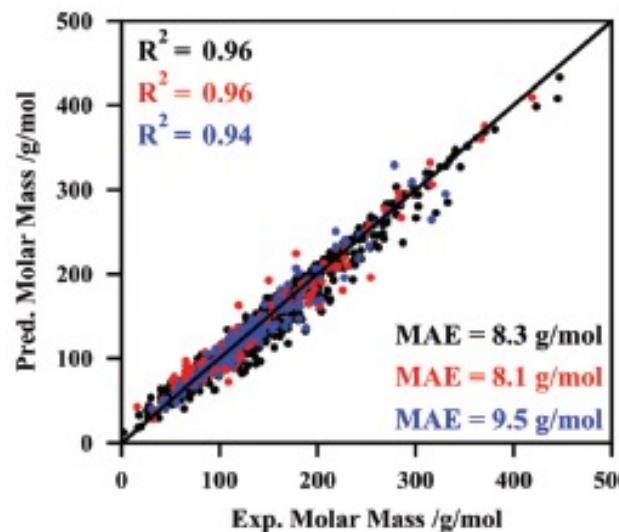
Dinis O. Abrançches,  Yong Zhang,  Edward J. Maginn  and Yamil J. Colón \*

- “Sigma profiles” are non-normalized surface charge density
- Molecular “fingerprint” that only contains surface charge
  - no atom identity, bonding, molecular weight, etc.
- Computed rapidly at various levels of theory (NWChem)
- Result: 51 data points *regardless of molecule size (fixed size CNN)*
- CNN only requires ~100 data points to train



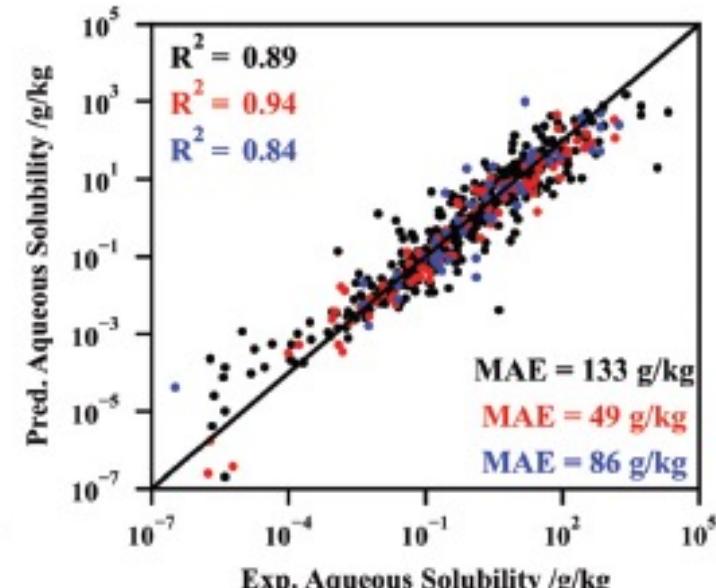
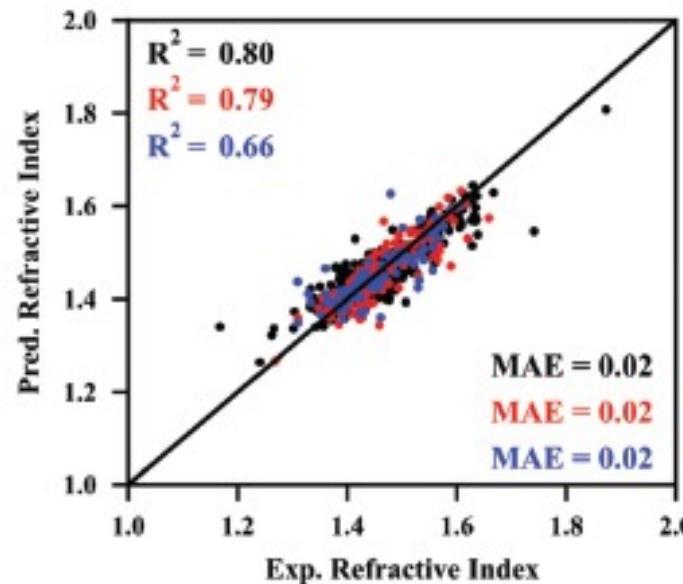
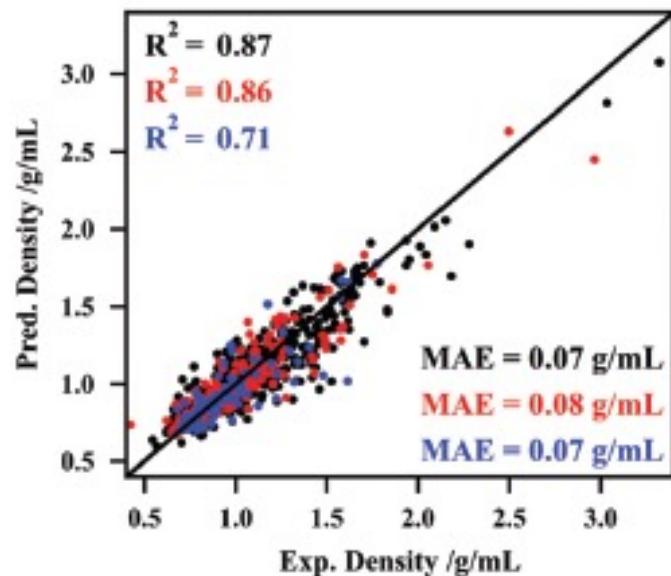
# “Predict” six basic properties

Property	Size	Range
Molar Mass	1432	2 – 447 g/mol
Normal Boiling Temperature	1208	-192 – 463 °C
Vapor Pressure (25 °C)	594	$10^{-6}$ – $10^3$ kPa
Density (20 °C)	711	0.6 – 3.3 g/mL
Refractive Index (20 °C)	834	1.26 – 1.87
Aqueous Solubility (25 °C)	327	$10^{-7}$ - $10^4$ g/kg



Training  
Testing  
Validation

# Very accurate correlation of range of properties



## Next steps:

- Train on IL and DES databases
- Computationally generate hypothetical structures, sigma profiles
- Inverse design: properties => sigma profile => desired molecule
- Hypothesis: more robust, predictive, and extensible than conventional ML/QSPR methods

[https://github.com/MaginnGroup/SP\\_ML\\_CC](https://github.com/MaginnGroup/SP_ML_CC)

# Machine learning *ab initio* potentials

## Machine learning potentials for complex aqueous systems made simple

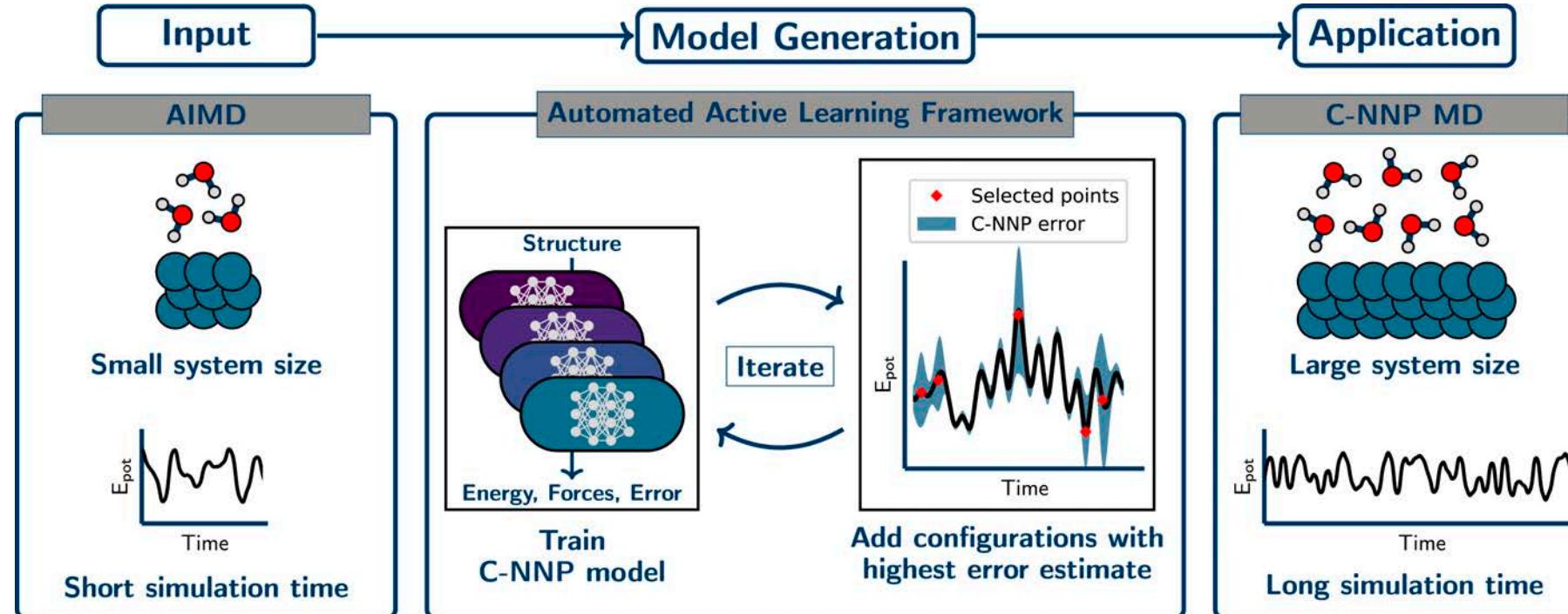
Christoph Schran<sup>a,b,c,d,1</sup> , Fabian L. Thiemann<sup>a,b,c,d,e,2</sup> , Patrick Rowe<sup>a,b,c,d,2</sup>, Erich A. Müller<sup>e</sup> , Ondrej Marsalek<sup>f</sup> , and Angelos Michaelides<sup>a,b,c,d,1</sup> 

<sup>a</sup>Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom; <sup>b</sup>Thomas Young Centre, University College London, London WC1E 6BT, United Kingdom; <sup>c</sup>London Centre for Nanotechnology, University College London, London WC1E 6BT, United Kingdom;

<sup>d</sup>Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom; <sup>e</sup>Department of Chemical Engineering, Sargent Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, United Kingdom; and <sup>f</sup>Charles University, Faculty of Mathematics and Physics, 121 16 Prague 2, Czech Republic

**PNAS**

Vol. 118 | No. 38



- How do you train?
- Are small boxes and short simulations “OK”?
- How would you know?

- Do computed properties “agree” with AIMD?
- Do results agree with experiment?
- Right level of AIMD used?

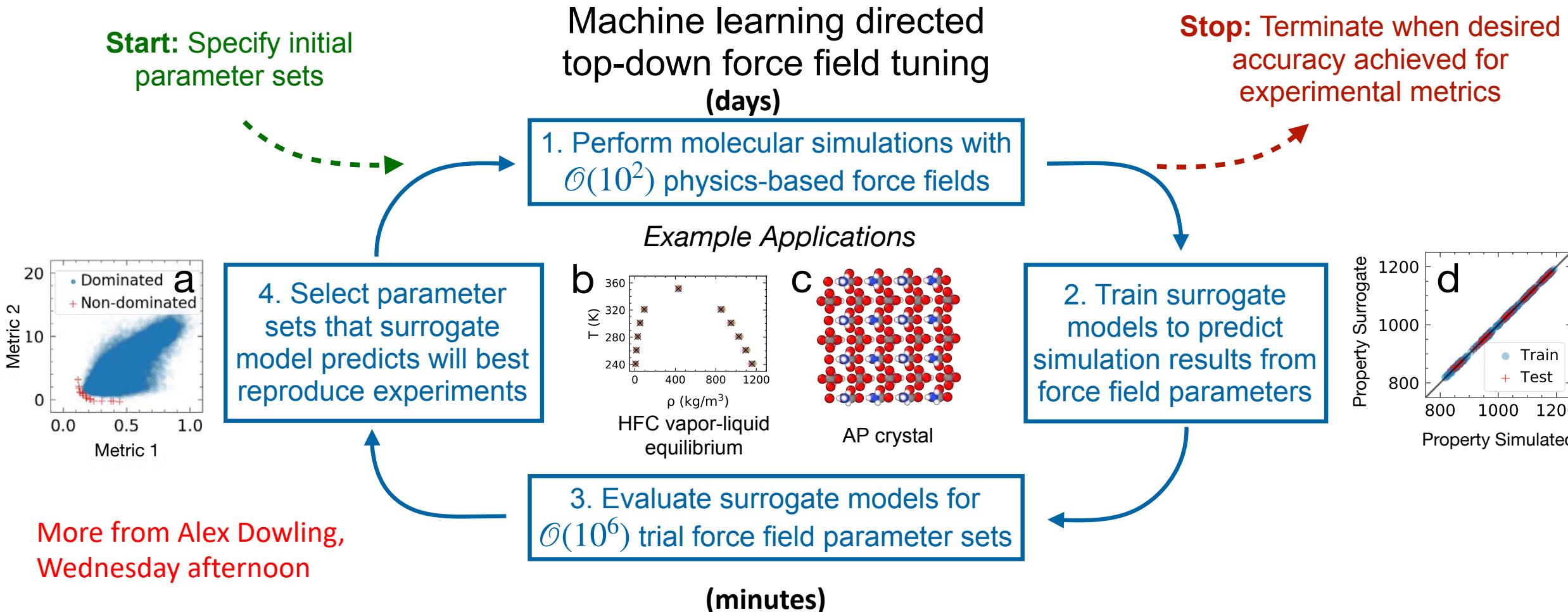
# Improving force fields – ML approach

## Machine Learning Directed Optimization of Classical Molecular Modeling Force Fields

Bridgette J. Befort,<sup>‡</sup> Ryan S. DeFever,<sup>‡</sup> Garrett M. Tow, Alexander W. Dowling, and Edward J. Maginn\*

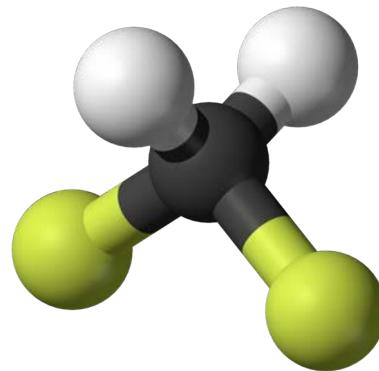
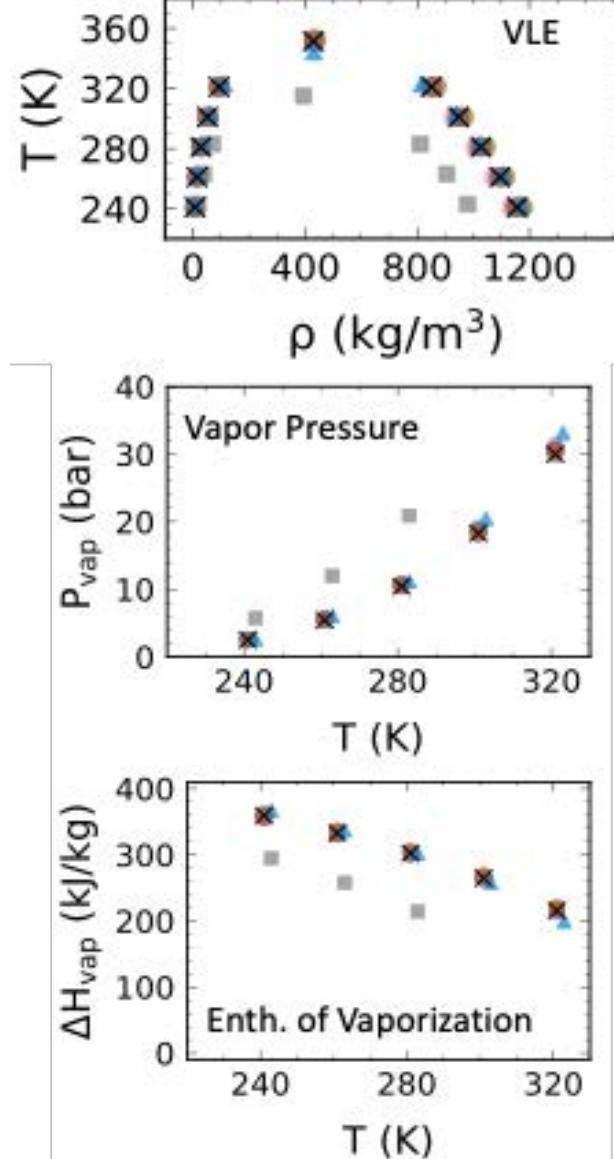
Cite This: *J. Chem. Inf. Model.* 2021, 61, 4400–4414

Read Online

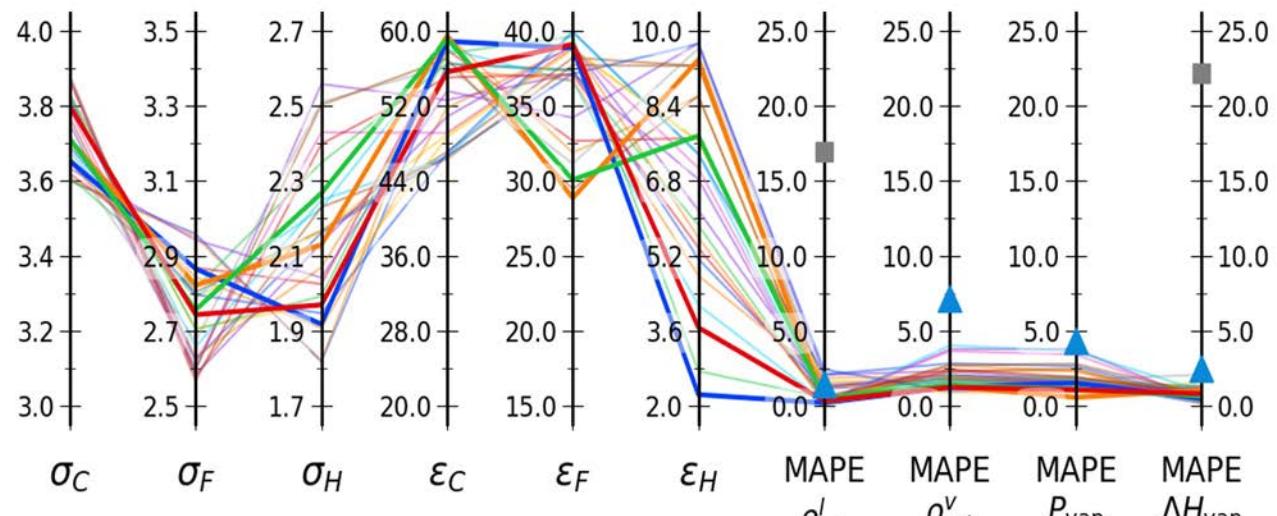


## HFC-32

### 26 Final Force Fields



$\sigma$  reported in units of Å;  $\varepsilon$  reported in units of K/k<sub>B</sub>.

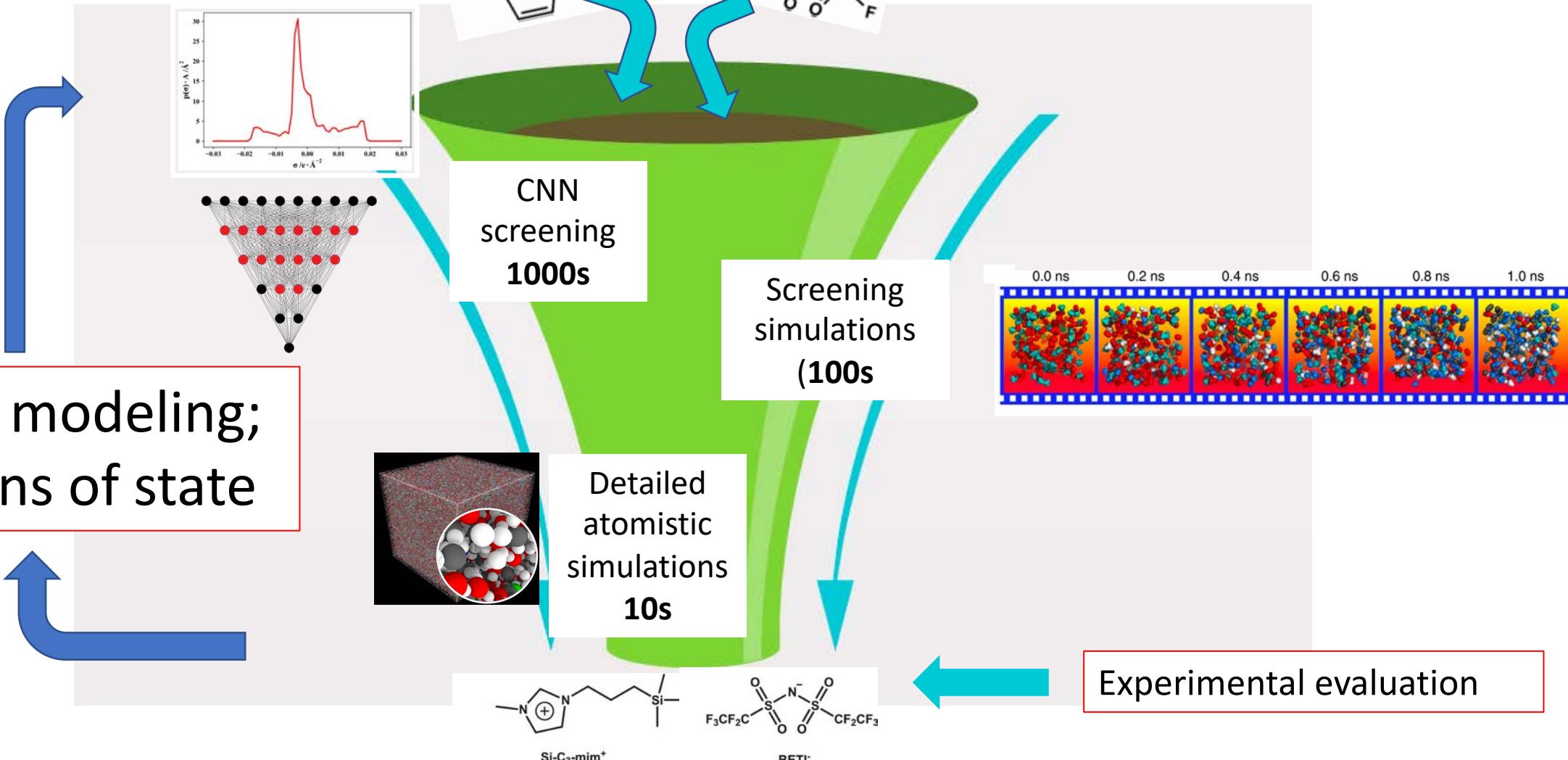


- Exhaustively search **all of parameter space** for "optimum" set of parameters
- Use of GP model enables this with **relatively few costly simulations**
- **No single set of force field parameters is "best"** - many do equally well for collection of properties
- Current: developing an optimized FF for all HFCs

<https://github.com/dowlinglab/hfcs-ffffit>

# What is the future?

## Generate hypothetical structures



# Key takeaways

- **Sophistication and quantity** of simulations have increased dramatically
  - Computing power (speed, cores, GPUs)
  - Algorithmic advances
  - Databases (experimental data, computational results, force fields)
  - Electronic publishing and collaboration tools
- **TRUE simulations** are needed
  - Reproducibility crisis requires open source, data sharing, full transparency
- **Machine learning** is having a major impact
  - Property correlations (predictions?)
  - Augmenting (replacing?) EOS and G<sup>E</sup> models
  - ML-based AIMD potentials and better physics-based force fields
- **Molecular design** is now within reach
  - Integration of molecular simulations, process simulations, thermodynamic modeling

Equifase research can truly save the planet!

# The stars of the show...



Cassiano Aimoli



Charles Abreu



Thanos  
Panagiotopoulos



Jeff Errington



Ilja Siepmann



Alain Fuchs



Roland Pellenq



from Saleem ur Rahman



Pedro Pessoa Filho



Danilo Carvalho



Lucienne Romanielo



Randy Snurr



David Sholl



Yongchul Chung



Carlos Vega

Notre Dame colleagues



Michael Humbert



Ryan DeFever



Alex Dowling



Eliseo Marin



Bridgette Befort



Ning Wang



Yong Zhang



Gabriela Correa



Garrett Tow



Dinis Abrantes



Yamil Colón



Kalil Bernadino



Mauro Ribeiro



Richard Messerly



Richard Elliott



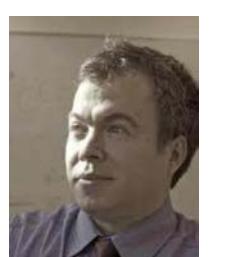
Daniel Roe



Peter Cummings



Sharon Glotzer



Erich Muller



Lisa Joss



Angelos Michaelides



Fabian Thiemann



Patrick Rowe



Ondrej Marsalek



Fred Tavares