

The Molecular Simulation Design Framework (MoSDeF): Capabilities and Applications

Peter T. Cummings
 Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN
 Vanderbilt University Multiscale Modeling and Simulation (MuMS) Center, Nashville, TN

mosdef.org

2019 International Workshop on Molecular Simulation
 Hangzhou, China
 October 12-13, 2019

SCHOOL OF ENGINEERING

MUMS

About Vanderbilt and Nashville

- Small, private university
 - Founded in 1873 with \$1M gift from Cornelius Vanderbilt
 - Small student body
 - 6750 undergraduates
 - 5960 graduate and professional degree students
 - 53% of student population are women
- Ranked (#15) in US News & World Report top 25 universities
 - Also ranked in top 25 in US Federal Government research funding
- Located in, and highly integrated with, Nashville
 - Music City, USA and the Athens of the South
 - ~1.6M population metropolitan area
 - Bon Appetit magazine: the "coolest, tastiest city in the south"
 - Athens of the South: 20 universities and colleges within a 60 km radius

SCHOOL OF ENGINEERING

MUMS

Motivation - Computational Materials Discovery

- Computational discovery of new materials is emerging as a new subfield of computational materials science developing out of the Materials Genome Initiative (MGI), e.g.:
 - The Materials Project
 - <https://materialsproject.org>
 - Properties of over 500,000 materials and screened 69,000 of these for Li-ion batteries.
 - Based on use of low-level electronic structure methods on crystalline (ordered) materials
 - Screening of metal-organic frameworks (MOFs)
 - 137,953 hypothetical MOF structures generated
 - Each MOF evaluated for methane storage at 35 bar and 298 K using grand canonical Monte Carlo
 - ~300 MOFs had higher methane-storage capacity at 35 bar than current world record
 - Crowdsourced Clean Energy Project database at Harvard
 - <http://www.molecularspace.org>
 - Data and analyses on 2.3 million candidate compounds for organic photovoltaics
 - Ab initio methods applied to single molecules, linking structural and electronic properties

Chemistry and conditions → Properties

SCHOOL OF ENGINEERING

MUMS

[1] Wilmer, et al., Large-scale screening of hypothetical metal-organic frameworks. *Nature Chemistry*, 4 (2011) 83-89. doi:10.1038/nchem.1192

Motivation - MGI for Soft Matter

- In soft matter, predominant physical behaviors occur at energy scales comparable with $k_B T$
 - Quantum aspects generally unimportant
 - Weak dispersion interactions frequently dominate
 - Contrast with hard materials
 - Energy scales large by comparison to $k_B T$
 - Relatively easy to predict properties since atoms/molecules are typically in crystalline lattice
 - It can be difficult to predict properties directly from atomic or molecular constituents
 - Soft matter often self-organizes into mesoscopic physical structures that are larger than microscopic scale yet much smaller than the macroscopic scale
 - Macroscopic properties are result of mesoscopic structure
 - Hard materials typically have no mesoscopic structuring
 - Many steps are needed to gather properties based on a given chemistry and condition

Chemistry and conditions → Properties

SCHOOL OF ENGINEERING

MUMS

MoSDeF

Molecular Simulation and Design Framework

- **Motivation** - develop a platform for performing MGI-style screening of soft-matter systems

- o **Goal:**

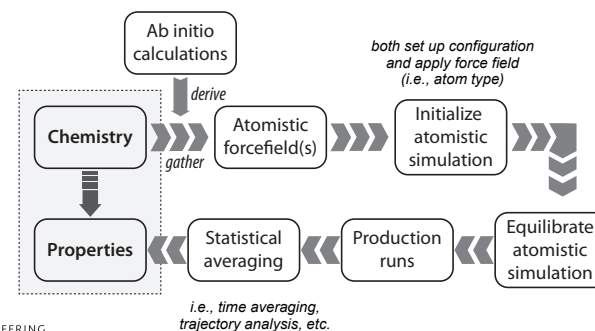


- o The procedures grouped into the arrow can be considered a “cottage industry” that is not always easy to reproduce
 - Only a few people have the knowledge to reproduce the specific simulation
 - Procedures are not always well documented/described
 - The properties may depend on the exact details of the procedures
 - Ad hoc tools/scripts are not always made available
 - Software may be highly specific and not extensible
 - Software may be poorly written and not be well validated

MoSDeF

Molecular Simulation and Design Framework

- **Motivation** - develop a platform for performing Materials Genome Initiative (MGI) style screening of soft-matter systems



MoSDeF

Molecular Simulation and Design Framework

- **Motivation** - develop a platform for performing Materials Genome Initiative (MGI) style screening of soft-matter systems

- o To do so, need to automate each of these steps and be able to combine them together

- **Model-Integrated Computing (MIC)**

- Developed by Vanderbilt researchers in the Institute for Software Integrated Systems (ISIS)

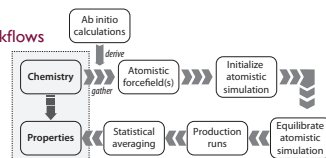
- Used in a wide variety of applications where scientific/engineering workflows require automation and integration to achieve complex goals

- Car manufacturing assembly line automation and control (GM)
- Health records management (NIH)
- Amphibious infantry fighting vehicle design (DARPA)

- MIC involves meta-language abstraction

- o MoSDeF is an outgrowth of three NSF grants joint with ISIS

- NSF CDI CBET 1028374, NSF SSI ACI 1047828, NSF SSE ACI 1535150
- MoSDeF also seeks to address **simulation reproducibility**



MoSDeF

Molecular Simulation and Design Framework

- **Why does MoSDeF promote simulation reproducibility and why is this important?**

- o In a science-skeptic world, the “reproducibility crisis” in scientific research has emerged as critical issue

- <http://www.nature.com/news/reproducibility-1.17552>

- o Reproducibility in computational science is becoming a forefront issue in the computational science community

- **Automation is a key device for reproducibility**

- Sandve, G. K. et al. (2013), Ten simple rules for reproducible computational research, *PLOS Comp. Bio.* (editorial), Vol. 9(10):1–4, doi: 10.1371/journal.pcbi.1003285

- **“Reproducible computational research, in which all details of computations—code and data—are made conveniently available to others, is a necessary response to [the credibility] crisis.”**

- Donoho, D. et al. (2009), Reproducible research in computational harmonic analysis, *Comp. Sci. Eng.* 11(1):8–18, doi: 10.1109/MCSE.2009.15

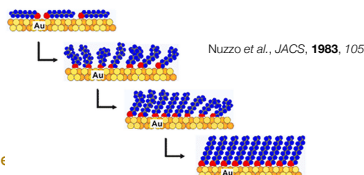
- **GUIs are the enemy of reproducibility**

- Lorena A. Barba, George Washington University; see her blog <http://lorenabarba.com/category/blog/> and “The hard road to reproducibility”, *Science*, 354(6308):142 (October 2016)
- » Reinhart and Rogoff 2000 Excel-based paper alleging a correlation between high government debt and low growth

- o Scriptability, open-source, complete data dissemination needed to reproduce published simulations are goals inherent to MoSDeF

Missing Functionality in Existing Tools

- ❑ Surface functionalization
 - Need ability to express patterns
 - Expose parameters that tune chemistry
- ❑ Flexible force field development and dissemination
 - Few force fields exist for all chemistries we care about
 - Not certain how accurate existing FFs will behave under high she
 - Collaborators and group members developing for specific systems
- ❑ Desire to use different simulation engines (engine agnostic)
 - LAMMPS, HOOMD, GROMACS all currently used in group
- ❑ MoSDeF is open source, built on open source components



IP[y]: IPython
Interactive Computing



MUMS

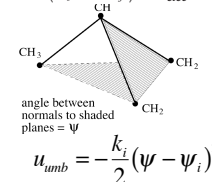
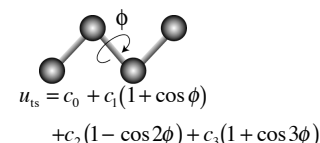
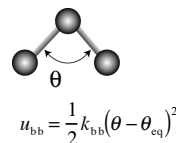
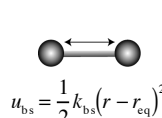
Aside on Force Fields

- ❑ In molecular dynamics simulation, we solve numerically some variant of Newton's equations

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$$

- ❑ Potential energy is described by a force field with as many as hundreds of parameters (e.g.)

$$U = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{\text{improper}} V_{imp} + \sum_{\text{LJ}} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}}$$



Atom-atom intramolecular interactions

MUMS

MoSDeF - Tools

- ❑ Three key tools being developed under the MoSDeF Suite
 - mBuild
 - A hierarchical, component-based toolkit for construction of complex molecular systems
 - Foyer
 - A general atom-typer for determining the proper force field parameters
 - metaMDS
 - Simulation configuration and execution
- ❑ MoSDeF also utilizes other open source software within the toolkit
 - InterMol, ParmEd, MDTraj, OpenMM
 - File format conversions, trajectory analysis, data structures
 - PACKMOL
 - Filling boxes and solvating structures
 - Python Scientific Stack: Numpy, Scipy, Matplotlib, Jupyter Notebooks, NGLView
 - Python libraries for numerical analysis, plotting, script encapsulation, visualization
 - OpenBabel, OpenMM
 - Energy minimization

mosdef.org

MUMS

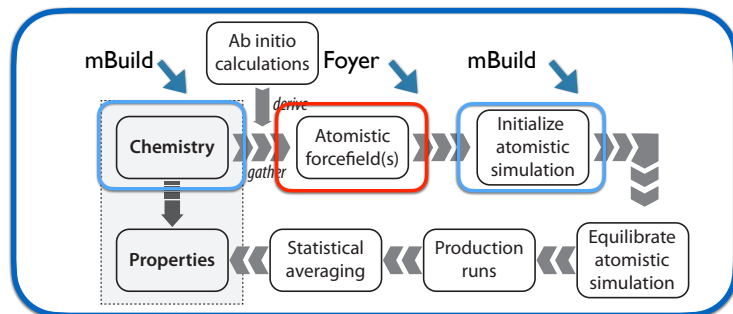
MoSDeF - Tools

- ❑ Three key tools being developed under the MoSDeF Suite
 - mBuild
 - A hierarchical, component based toolkit for construction of complex molecular systems
 - Foyer
 - A general atom-typer for determining the proper force field parameters
 - metaMDS **Signac**, **Signac-Flow** developed by Glotzer Group at U. Michigan
 - Simulation configuration and execution
- ❑ MoSDeF also utilizes other open source software within the toolkit
 - InterMol, ParmEd, MDTraj, OpenMM
 - File format conversions, trajectory analysis, data structures
 - PACKMOL
 - Filling boxes and solvating structures
 - Python Scientific Stack: Numpy, Scipy, Matplotlib, Jupyter Notebooks, NGLView
 - Python libraries for numerical analysis, plotting, script encapsulation, visualization
 - OpenBabel, OpenMM
 - Energy minimization

SCHOOL OF ENGINEERING

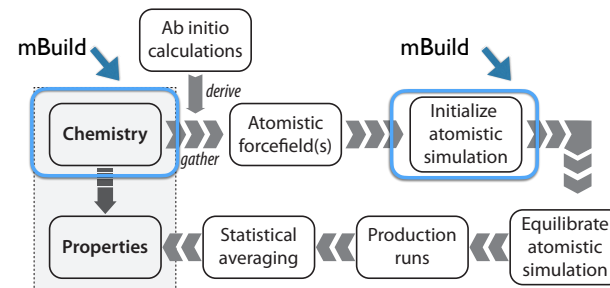
MUMS

Missing Functionality in Existing Tools



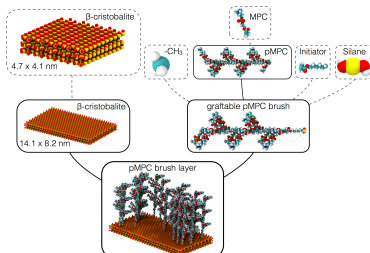
metaMDS/Signac, Signac-flow

Missing Functionality in Existing Tools



mBuild: a Hierarchical Molecular Builder

- ❑ Sketch or otherwise create simple components (dashed boxes)
- ❑ Combine operations on components into more complicated structures
 - Our applications (nanotribology, supercapacitors, skin lipid self-assembly) typically involve interfacial systems
- ❑ Everything is scriptable



Example: polymerization

```
import mbuild as mb
from mbuild.lib.moieties import CH2

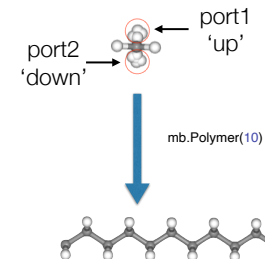
polymer = mb.Compound()

last_monomer = CH2()
polymer.add(last_monomer)

for _ in range(10):
    current_monomer = mb.clone(last_monomer)

    mb.force_overlap(move_this=current_monomer,
                    from_port=current_monomer['up'],
                    to_port=last_monomer['down'])

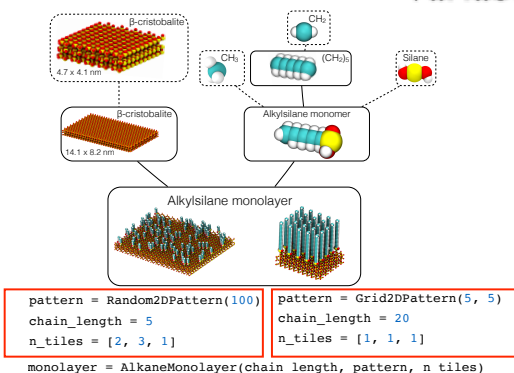
    polymer.add(current_monomer)
    last_monomer = current_monomer
```



- ❑ Above code with additional features is provided as mb.Polymer

Exposing Tunable Chemistry as Simple Variables

17



The idea is **abstraction**: users typically do not have to interact with low-level code constructing systems making it easier to perform screening

Systems are generated via well defined **recipes** that can easily be **shared** and **reproduced**

Using mBuild: CER Multilayer

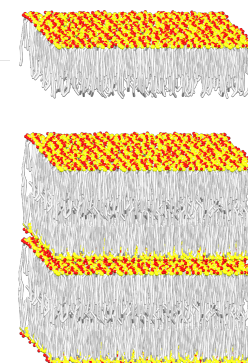
18

❑ Easily build a ceramide multilayer with just a few lines of python

- o Read in molecule from pdb file
- o Build leaflet, then stack leaflets
- o 18 lines!

```

1 import mbuild as mb
2 from numpy import pi
3
4 n_lipids = 100 # total number of lipids (must be a perfect square)
5 lipid_spacing = 0.4 # lipid spacing in nm
6 repeat_spacing = 7.75 # repeat spacing in nm
7
8 cer2 = mb.load('cer2.pdb')
9 cer2.spin(pi, around=[1, 0, 0])
10 grid = mb.pattern.Grid2DPattern(int(n_lipids**0.5), int(n_lipids**0.5))
11 grid.scale(int(n_lipids**0.5) * lipid_spacing)
12 stacked_bilayer = mb.Compound()
13 lipids = grid.apply(cer2) # put lipids on grid points
14 leaflet = mb.Compound(subcompounds=lipids)
15 for layer_num in range(4): # add 4 leaflets to the system
16     new_layer = mb.clone(leaflet)
17     if (layer_num + 1) % 2 == 1: # rotate every other layer 180deg
18         new_layer.spin(pi, around=[1, 0, 0])
19     new_layer.translate([0, 0, layer_num * repeat_spacing / 2])
20     stacked_bilayer.add(new_layer)
21 stacked_bilayer.save('stacked-bilayer.hoomdxml', overwrite=True)
    
```



Atom Typing and Applying Force Fields

20

❑ Need general purpose atom typer

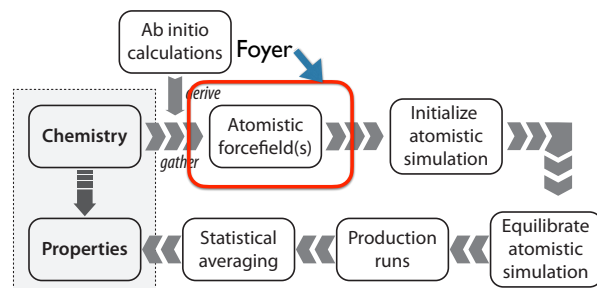
- o Arbitrary chemistry means we cannot rely on templates
- o Needs to be easy to develop new atom types and associated logic
- o Don't want to rely on rigid rule hierarchies for atom type definitions
 - With rigid rule hierarchies, order matters!
 - Ordered least to most general

❑ Removing gap between logic and parameters

- o Parameters exist in various files
- o Logic often encoded as series of if/else statements deep in source code
- o Why not combine the two?

❑ Similar concept being developed by Open Force Field group

- o <https://github.com/open-forcefield-group/>



Foyer: Applying and Disseminating force fields

❑ End user should only require

- o Force field file
- o Chemical topology

```
from foyer import Forcefield
import parmed as pmd
```

```
untyped_ethane = pmd.load_file('ethane.mol2', structure=True)
oplsaa = Forcefield(forcefield_files='oplsaa.xml')
ethane = oplsaa.apply(untyped_ethane)
```

```
# Save to any format supported by ParmEd
ethane.save('ethane.top')
ethane.save('ethane.gro')
```

Adding Atom Type Definitions to OpenMM XML

❑ OpenMM XML file convenient to extend

- o Existing infrastructure
- o Great support for custom functional forms

❑ Add machine/human readable SMARTS

- o Provides syntax for describing chemistry

Foyer Forcefield Format

Snippets from the OPLS force field

```
<ForceField>
<AtomTypes>
  <Type name="opls_148" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)H"
    desc="toluene CH3" overrides="opls_149,opls_135" doi="10.1021/ja9621760"/>
  <Type name="opls_149" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)*"
    desc="ethyl benzene CH2" overrides="opls_136" doi="10.1021/ja9621760"/>
  ...
</AtomTypes>
<HarmonicBondForce>
  <Bond class1="OW" class2="HW" length="0.09572" k="502080.0"/>
  ...
<NonbondedForce coulomb14scale="0.5" lj14scale="0.5">
  <Atom type="opls_001" charge="0.5" sigma="0.375" epsilon="0.43932"/>
  <Atom type="opls_002" charge="-0.5" sigma="0.296" epsilon="0.87864"/>
  ...
</NonbondedForce>
</ForceField>
```

Foyer Forcefield Format

SMARTS definition

```
def="[C;X4]([C;%opls_145])(H)(H)H"
```

```
<ForceField>
<AtomTypes>
  <Type name="opls_148" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)H"
    desc="toluene CH3" overrides="opls_149,opls_135" doi="10.1021/ja9621760"/>
  <Type name="opls_149" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)*"
    desc="ethyl benzene CH2" overrides="opls_136" doi="10.1021/ja9621760"/>
  ...
</AtomTypes>
<HarmonicBondForce>
  <Bond class1="OW" class2="HW" length="0.09572" k="502080.0"/>
  ...
<NonbondedForce coulomb14scale="0.5" lj14scale="0.5">
  <Atom type="opls_001" charge="0.5" sigma="0.375" epsilon="0.43932"/>
  <Atom type="opls_002" charge="-0.5" sigma="0.296" epsilon="0.87864"/>
  ...
</NonbondedForce>
</ForceField>
```


Foyer Forcefield Format

"Overrides" for assigning precedence

overrides="opls_149,opls_135"

```
<ForceField>
<AtomTypes>
  <Type name="opls_148" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)H"
    desc="toluene CH3" overrides="opls_149,opls_135" doi="10.1021/ja9621760"/>
  <Type name="opls_149" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)*"
    desc="ethyl benzene CH2" overrides="opls_136" doi="10.1021/ja9621760"/>
  ...
</AtomTypes>
<HarmonicBondForce>
  <Bond class1="OW" class2="HW" length="0.09572" k="502080.0"/>
  ...
<NonbondedForce coulomb14scale="0.5" lj14scale="0.5">
  <Atom type="opls_001" charge="0.5" sigma="0.375" epsilon="0.43932"/>
  <Atom type="opls_002" charge="-0.5" sigma="0.296" epsilon="0.87864"/>
  ...
</NonbondedForce>
</ForceField>
```

MUMS

Foyer Forcefield Format

DOIs for logging parameter sources

doi="10.1021/ja9621760",

```
<ForceField>
<AtomTypes>
  <Type name="opls_148" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)H"
    desc="toluene CH3" overrides="opls_149,opls_135" doi="10.1021/ja9621760"/>
  <Type name="opls_149" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)*"
    desc="ethyl benzene CH2" overrides="opls_136" doi="10.1021/ja9621760"/>
  ...
</AtomTypes>
<HarmonicBondForce>
  <Bond class1="OW" class2="HW" length="0.09572" k="502080.0"/>
  ...
<NonbondedForce coulomb14scale="0.5" lj14scale="0.5">
  <Atom type="opls_001" charge="0.5" sigma="0.375" epsilon="0.43932"/>
  <Atom type="opls_002" charge="-0.5" sigma="0.296" epsilon="0.87864"/>
  ...
</NonbondedForce>
</ForceField>
```

MUMS

Foyer Forcefield Format

DOIs for logging parameter sources

BibTeX file output

```
<ForceField>
<AtomTypes>
  <Type name="opls_148" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)H"
    desc="toluene CH3" overrides="opls_149,opls_135" doi="10.1021/ja9621760"/>
  <Type name="opls_149" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)*"
    desc="ethyl benzene CH2" overrides="opls_136" doi="10.1021/ja9621760"/>
  ...
</AtomTypes>
<HarmonicBondForce>
  <Bond class1="OW" class2="HW" length="0.09572" k="502080.0"/>
  ...
<NonbondedForce coulomb14scale="0.5" lj14scale="0.5">
  <Atom type="opls_001" charge="0.5" sigma="0.375" epsilon="0.43932"/>
  <Atom type="opls_002" charge="-0.5" sigma="0.296" epsilon="0.87864"/>
  ...
</NonbondedForce>
</ForceField>
```

MUMS

Foyer Forcefield Format

DOIs for logging parameter sources

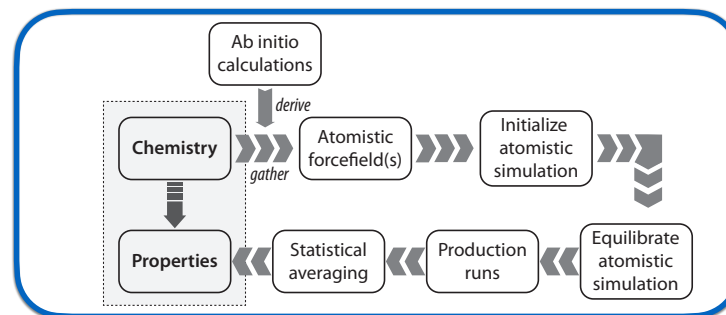
note = {Parameters for atom types: opls_761, opls_760, opls_764, opls_763}

```
<ForceField>
<AtomTypes>
  <Type name="opls_148" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)H"
    desc="toluene CH3" overrides="opls_149,opls_135" doi="10.1021/ja9621760"/>
  <Type name="opls_149" class="CT" element="C" mass="12.01100" def="[C;X4]([C;%opls_145])(H)(H)*"
    desc="ethyl benzene CH2" overrides="opls_136" doi="10.1021/ja9621760"/>
  ...
</AtomTypes>
<HarmonicBondForce>
  <Bond class1="OW" class2="HW" length="0.09572" k="502080.0"/>
  ...
<NonbondedForce coulomb14scale="0.5" lj14scale="0.5">
  <Atom type="opls_001" charge="0.5" sigma="0.375" epsilon="0.43932"/>
  <Atom type="opls_002" charge="-0.5" sigma="0.296" epsilon="0.87864"/>
  ...
</NonbondedForce>
</ForceField>
```

MUMS

Promoting Dissemination of Force Fields

- ❑ Template repo for building Foyer-compatible force field files
 - https://github.com/mosdef-hub/forcefield_template
 - Skeleton force field file
 - Skeleton testing setup
 - User needs to only add correctly typed .mol2 files
 - Tutorial for adding SMARTS definitions
 - Walk-through for adding more and more complex definitions
 - Testing with small example set of hydro carbons from alkanes through alkenes and benzene
 - How to run tests with py.test and enable TravisCI for your repo
- ❑ Force field development within our group
 - OPLS-AA compatible parameters for perfluoroethers
DOI: 10.5281/zenodo.56807
 - Coarse-grained models for skin lipids
 - Automating force field derivation for supercapacitor electrodes



metaMDS/Signac, Signac-flow

MetaMDS - Overview

- ❑ MetaMDS was designed with the goal of being able to perform screening simulations
 - It can submit large batches of jobs with various input parameters
- ❑ MetaMDS relies on 3 key inputs:
 - mBuild "recipe"
 - This tells metaMDS how to construct the system
 - Allows us to screen over not just thermodynamic properties, but over different structures/molecules/etc.
 - Simulation template
 - This is a script that encapsulate the underlying simulation routines/calls
 - List(s) of parameters to vary
 - These parameters are used to change thermodynamic states in the simulation script and change inputs to the mBuild recipe
 - These lists could also include different forcefields to evaluate

MetaMDS - Screening

- ❑ By using simulation templates, mBuild recipes, and simple parameter lists, we can create a streamlined, abstracted interface
 - Users only need a few lines of code to do large scale, complex screening
 - Individual elements can be easily tested before large scale deployment
 - All procedures are well documented and described

```

In [ ]: # Initialize a simulation instance with a template and some metadata
sim = mds.Simulation(name='monolayer', template=create_run_script, output_dir='output')

chain_lengths = [8, 12, 16, 20]
for length in chain_lengths:
    parameters = {'chain_length': length,
                  'n_molecules': 100,
                  'forcefield': 'OPLS-aa',
                  'build_func': build_monolayer}
    # Parameterize our simulation template
    sim.parameterize(**parameters)

In [ ]: # Run
sim.execute_all(hostname='rahman.vuse.vanderbilt.edu', username='ctk3b')

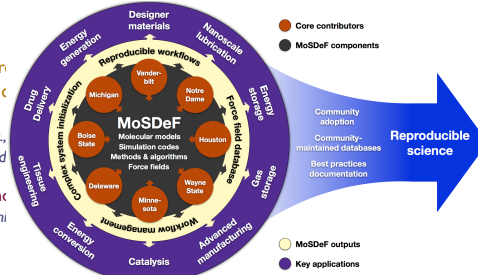
In [ ]: sim.sync_all()
  
```

User interface is Python notebook (c.f. Mathematica)

MoSDeF

Building a community

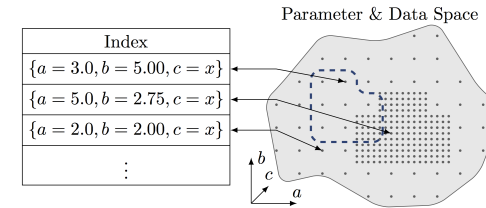
- o Hackathon in February 2017 with representation from
- o Goal of learning how to use MoSDeF and become a contributor
 - Add force fields
 - Currently support OPLS, TRAPPE, specific force fields (e.g., bio forcefield derivation)
 - Wish to add other standard forcefields (e.g., bio forcefield derivation)
 - Add simulation workflows for specific properties and
 - Currently support many standard workflows, coarse-grain
 - LONG wish list!
 - Add simulation packages
 - Currently support LAMMPS, GROMACS, HOOMD-Blue
 - Wish to add open-source AIMD codes, Monte Carlo (e.g., Cassandra), other MD codes
- o NSF project (2018-2021) to support next MoSDeF cycle
 - Michigan (Glotzer, HOOMD-blue, Signac), Notre Dame (Maginn, Cassandra, force fields), Wayne State (Potoff, GOMC, force fields), Minnesota (Siepmann, CP2K, TRAPPE force fields), Boise State (Jankowski, force fields, methods), Delaware (Jayaraman, polymer force fields, methods), Houston (Palmer, crystallization/melting methods)



MUMS

Putting it All Together: Signac Framework

- Recently adopted in place of metaMDS
- Define heterogeneous parameter spaces
- Provides link between data and metadata
 - o Every job and its data is tied to a point in parameter space
 - o Provides glue for seamless building, execution and analysis



Developed by Simon Adorf,
now in Glotzer group

MUMS

Putting It All Together

Our tribology operations

- o mBuild + save with foyer
- o Energy minimize
- o Equilibrate
- o Compress
- o Shear at range of normal loads

Customize compute environment

- o Defaults provided for common environments
- o Collection of environments for major computing centers
 - Oak Ridge Leadership Class Facility (OLCF)
 - National Energy Research Supercomputing Center (NERSC)
 - National Institute of Computational Science (part of XSEDE)

```
>>> project = MyProject()
>>> project.print_status(detailed=True, params={'a':1})
Status project 'test-project':
Total # of jobs: 10
Label      progress
-----
Initialized [#####] 100.00%
Processed   [#####] 66.67%
Detailed view:
job_id      5 next_job a labels
-----
10a1f78c38124440a188f931f480b U - - - - - 1 Initialized, processed
ba81a9f06a384dcf12c4a83ac9612f8a U 2 Initialized, processed
32764c28ef13b0aefeb0a76a158ac4e U 3 Initialized
```

Query state
point
statuses

Custom environment definition
(truncated for clarity)

```
class MyTorqueEnvironment(TorqueEnvironment):
    hostname_pattern = 'myoacluster.university.edu'
    cores_per_node = 16

    @classmethod
    def mpi_cmd(cls, cmd, np):
        return 'mpirun -np {np} {cmd}'.format(np=np, cmd=cmd)
```

MUMS

Intro to Nanotribology

Micro- and nano-electromechanical devices (MEMS and NEMS) used in a wide range of applications

- o Accelerometers
 - Airbag deployment
 - Consumer electronics
 - Apple iPhone, Nintendo Wii
- o Micromirror arrays
 - HD projectors (DLP TV)
- o Sensors



<http://mems.sandia.gov/gallery/images/gc8.jpg>

Typically applications avoid parts in contact

- o Surface forces such as adhesion, stiction, friction, and wear dominate

Lubrication schemes needed for devices to reach full potential



MUMS

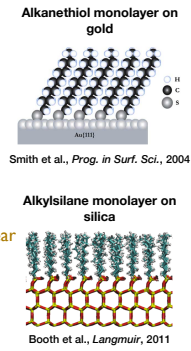
Monolayer Lubricants

Considerations for nanoscale lubrication systems

- Low surface energy
 - Minimize adhesion**
- Low friction coefficient (μ)
 - Minimize friction**
- Robustness/durability
 - Wear-resistance**

Alkylsilane monolayers

- Densely-packed chains reduce surface energy, protect underlying surface from wear
 - Low adhesion between surfaces even in the presence of water
- Friction coefficients vary widely by composition, density, and chain length
 - μ reduced to <50% of value for bare silica
- Durability concerns



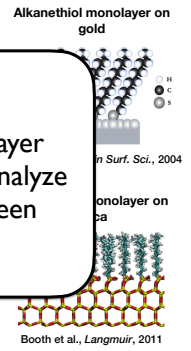
Monolayer Lubricants

Considerations for nanoscale lubrication systems

- Low surface energy
 - Minimize adhesion**
 - Low friction coefficient
 - Minimize friction**
 - Robustness/durability
 - Wear-resistance**
- ### Alkylsilane monolayers
- Densely-packed chains reduce surface energy, protect underlying surface from wear
 - Low adhesion between surfaces even in the presence of water
 - Friction coefficients vary widely by composition, density, and chain length
 - μ reduced to <50% of value for bare silica
 - Durability concerns

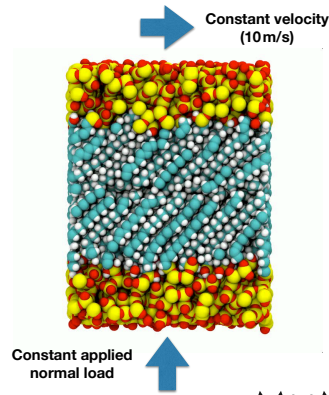
Objective

Perform MGI-style screening of monolayer lubricants using the MoSDeF toolkit to analyze structure-property relationships between monolayer chemistry and tribology



Studying Monolayers with MD

- ### Functionalize two opposing surfaces with monolayer films, bring into contact, and shear
- Atomically rough, amorphous silica surfaces mimic experimental conditions



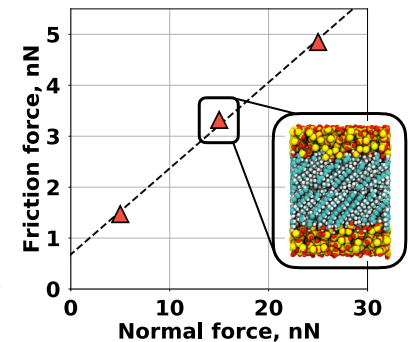
Studying Monolayers with MD

Derjaguin form of Amonton's Law of Friction

$$F_f = F_0 + \mu \cdot F_n$$

Friction force Adhesive force Coeff. of friction Normal force

- ### Functionalize two opposing surfaces with monolayer films, bring into contact, and shear
- Atomically rough, amorphous silica surfaces mimic experimental conditions
- ### Analyze tribology by using the Derjaguin form of Amonton's Law of Friction
- Perform simulations at several normal loads, friction vs. load relationship should be linear
 - Intercept represents adhesion, slope represents the coefficient of friction



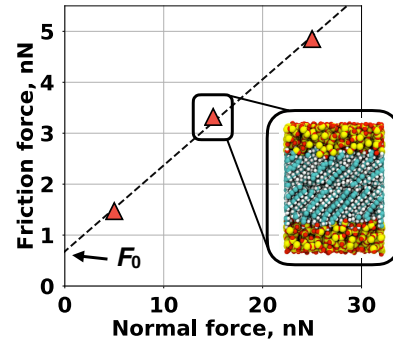
Studying Monolayers with MD

Derjaguin form of Amontons's Law of Friction

$$F_f = F_0 + \mu \cdot F_n$$

Friction force Adhesive force Coeff. of friction Normal force

- ❑ Functionalize two opposing surfaces with monolayer films, bring into contact, and shear
 - Atomically rough, amorphous silica surfaces mimic experimental conditions
- ❑ Analyze tribology by using the Derjaguin form of Amontons's Law of Friction
 - Perform simulations at several normal loads, friction vs. load relationship should be linear
 - Intercept represents adhesion, slope represents the coefficient of friction



MUMS

41

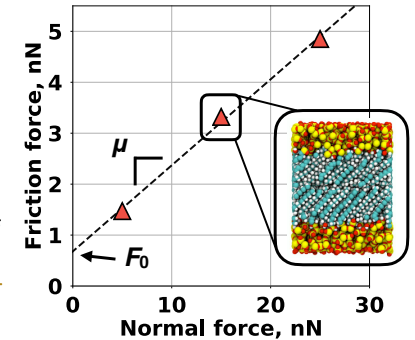
Studying Monolayers with MD

Derjaguin form of Amontons's Law of Friction

$$F_f = F_0 + \mu \cdot F_n$$

Friction force Adhesive force Coeff. of friction Normal force

- ❑ Functionalize two opposing surfaces with monolayer films, bring into contact, and shear
 - Atomically rough, amorphous silica surfaces mimic experimental conditions
- ❑ Analyze tribology by using the Derjaguin form of Amontons's Law of Friction
 - Perform simulations at several normal loads, friction vs. load relationship should be linear
 - Intercept represents adhesion, slope represents the coefficient of friction

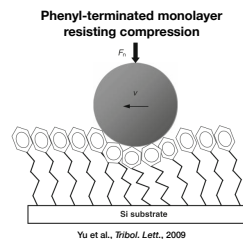


MUMS

42

Terminal Group Chemistry

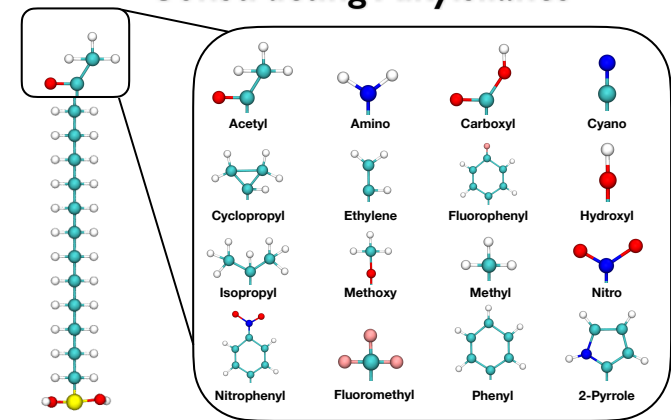
- ❑ Terminal groups form the boundary at the monolayer interface, thus having important effects on tribology
 - Methyl most common - low surface energy
- ❑ Higher friction forces in hydroxyl-terminated monolayers
 - Attributed to H-bonding across the monolayer interface
- ❑ Phenyl terminal groups can improve durability by resisting asperity penetration
- ❑ Only a few chemistries studied in the context of lubrication



MUMS

43

Constructing Alkylsilanes

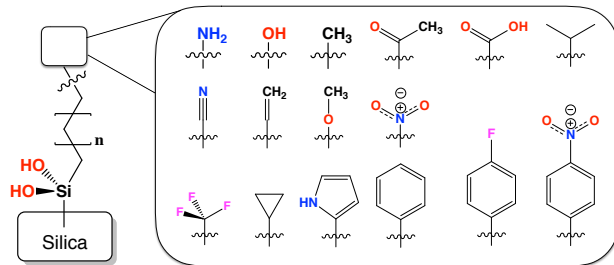


MUMS

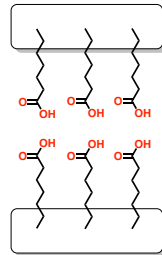
44

Constructing Alkylsilanes

Chemical parameter space
backbone chain length (n) and terminal group functionality

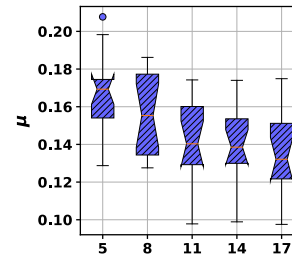


2D-depiction of two-monolayer system

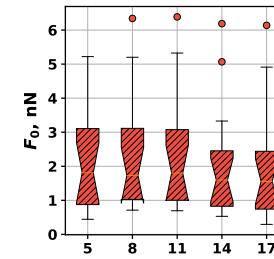


Chain length Effects

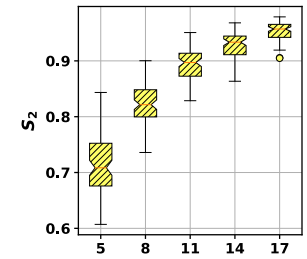
Decreased coefficient of friction
with increased chain length



Adhesion independent of
chain length



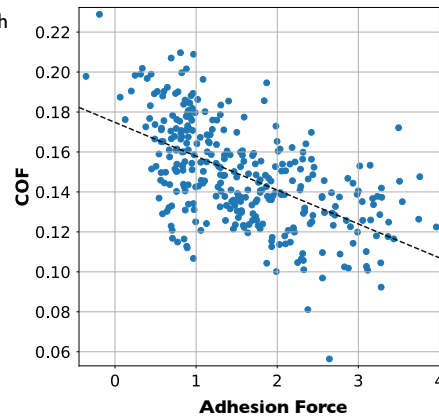
Orientational ordering of
monolayer chains increases
with chain length



Chain length, number of backbone carbons

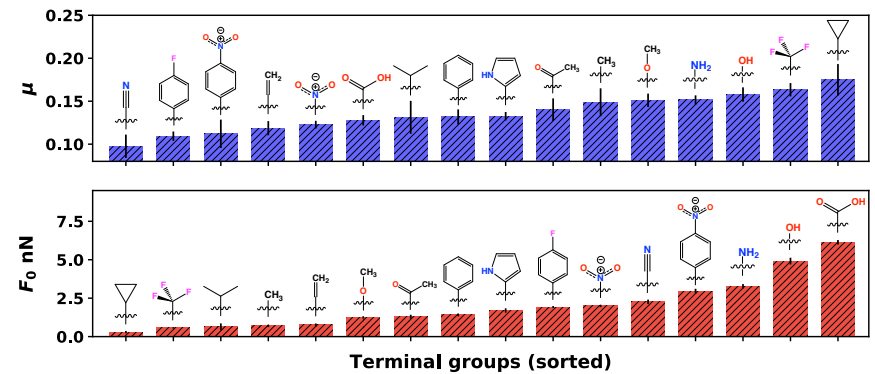
Coefficient of Friction vs Adhesion Force

□ One might expect th



oefficient of friction

Terminal Group Effects



Terminal groups (sorted)

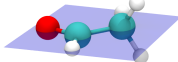
Structure-Property Relationships: Assigning Molecular Descriptors

53

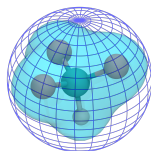
Fractional negative VDW surface area



Plane-of-best-fit



Asphericity

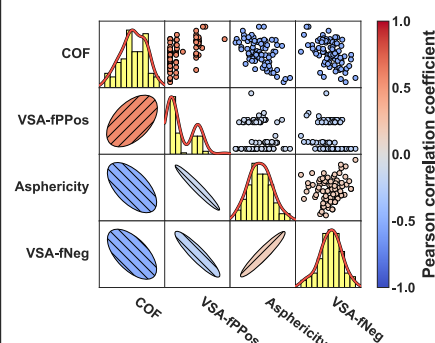


- Assign molecular descriptors to each terminal group
 - Provides numerical values to use for correlation analyses
- Describe characteristics such as molecular shape, size, and charge distribution
- >100 molecular descriptors assigned to each terminal group using the RDKit open-source cheminformatics package*

*<http://www.rdkit.org/>

Structure-Property Relationships: Coefficient of friction

54

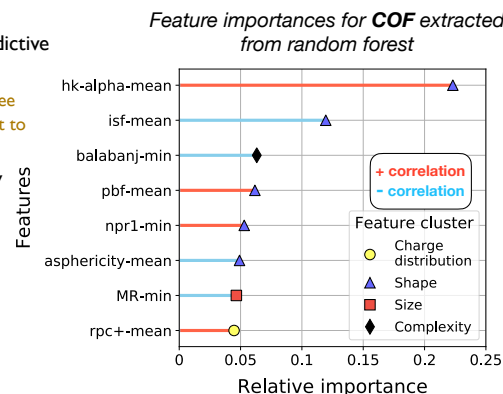


- Moderate correlations between descriptors of molecular shape, charge distribution and COF
 - **VSA-fPPos** (fractional positive polar van der Waals surface area)
 - **Asphericity**
 - **VSA-fNeg** (fractional negative van der Waals surface area)
- Reduced friction for groups that are non-spherical and have negative partial charges distributed onto larger atoms
- Similar analysis for adhesion force

Machine Learning: Feature Importance with Random Forest

55

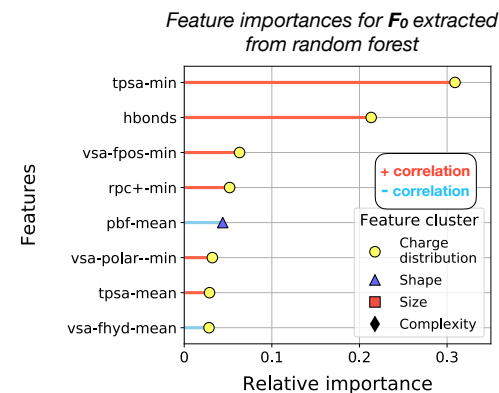
- Random forest is used to derive a predictive model for COF and F_0
 - Feature importances can be extracted to see what molecular properties contribute most to each tribological property
- COF appears to be influenced most by molecular shape
 - **hk-alpha** relates to hybridization
 - $sp < sp^2 < sp^3$
 - Positive correlation indicates linear, planar molecules will have lower COF



Feature Importance with Random Forest

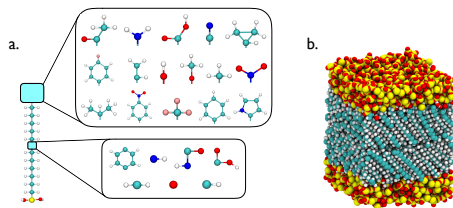
56

- Random forest is used to derive a predictive model for COF and F_0
 - Feature importances can be extracted to see what molecular properties contribute most to each tribological property
- F_0 is dominated by charge distribution
 - Low adhesion for terminal groups with a low polar surface area and w/o the ability to form inter-monolayer H-bonds



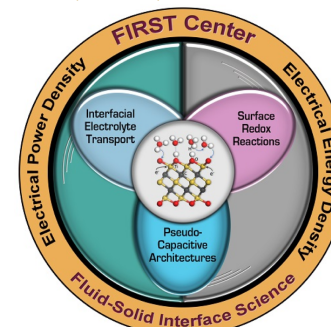
Next Steps

- ❑ Larger parameter set screening
 - Other factors which affect tribological behavior
 - Different tether chemistry
 - Multi-component monolayers
 - Mixed length, end-functionalization, tether chemistry
 - Asymmetric functionalization - different end groups on tethers
 - For N ways to produce a monolayer, $N(N+1)/2$ asymmetric functionalization scenarios
 - Identified ~45,000 variations to be screened
 - 2019 allocation (45M core hours) on Leadership Class supercomputer at Oak Ridge National Laboratory



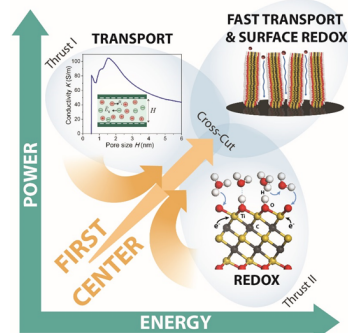
Diffusivity of Solvated Ionic Liquids through Molecular Dynamics Screening

- ❑ Fluid Interface Reaction Structure and Transport Center
 - DOE Energy Frontier Research Center (2009-2022)



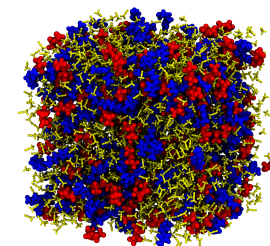
Advancing Energy Storage

- ❑ Current energy storage devices are limited
- ❑ Supercapacitors store energy through electric double layer (EDL)
 - High power density, low energy density
- ❑ Choice of electrolyte can enhance device performance
- ❑ Better understanding of electrolyte transport can lead towards more efficient supercapacitors



Ionic Liquids as Electrolytes

- ❑ Ionic liquids (ILs) exhibit different properties in comparison to conventional electrolytes
 - High electrochemical stability
 - Slow transport properties (diffusivity, viscosity, conductivity)
- ❑ Slow transport negatively affects power density of energy storage device
 - Solvate ionic liquids in organic solvents to mitigate slow transport
- ❑ Molecular dynamics (MD) simulations provide an atomic-level viewpoint of these systems and the ability to vary system parameters

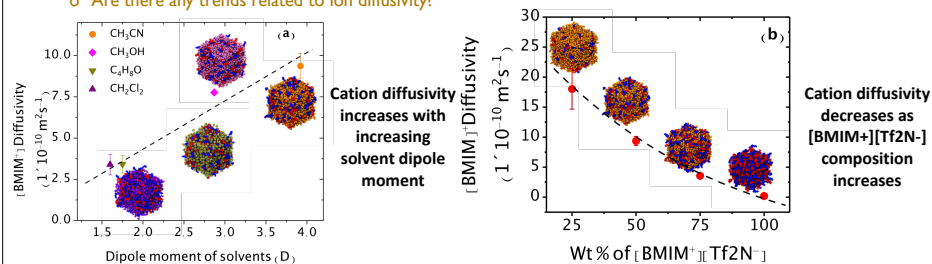


Snapshot of BMIM-Tf₂N-solvent system, Coloring: BMIM (Blue), Tf₂N (Red), Solvent (Yellow)

Solvation Effects on IL Transport

Previous Study:

- Quasi elastic Neutron Studying and MD of 1-Butyl-3-methylimidazolium bis(trifluoromethanesulfonyl)imide ([BMIM+]-[Tf2N-]) solvated in 5 unique solvents at 4 different concentrations
- Which solvent most increases ion diffusivity?
- Are there any trends related to ion diffusivity?



Cation diffusivity increases with increasing solvent dipole moment

Cation diffusivity decreases as [BMIM+][Tf2N-] composition increases



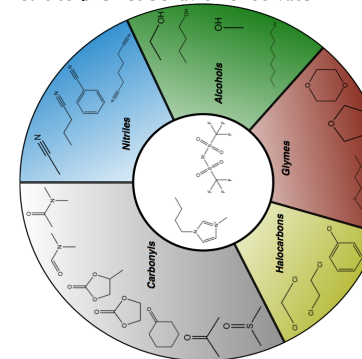
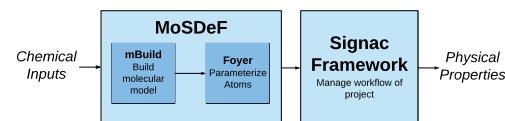
SCHOOL OF ENGINEERING

Osti, N. C.; Van Aken, K. L.; Thompson, M. W.; Tiet, F.; Jiang, D.; Cummings, P. T.; Gogotsi, Y.; Mamontov, E. *J. Phys. Chem. Lett.* **2016**, acs.jpcl.6b02587.

MUMS

Motivation for Screening

- Large chemical space and complex interactions make it difficult to predict behavior of solvated IL
- Greatly expand parameter space through screening:
 - 1 Ionic Liquid: [BMIM+]-[Tf2N-]
 - 5 solvents -> 22 solvents
 - 4 compositions -> 18 compositions
 - 396 unique systems

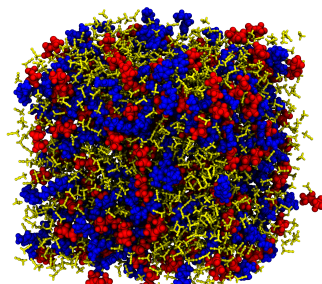


SCHOOL OF ENGINEERING

MUMS

Simulation Methods

- GROMACS 5.1.4 Simulation Engine
- OPLS-AA and CL&P Force Field Parameters
 - Automatic parametrization in Foyer
- Simulation Steps:
 - NVT Equilibration, NPT Equilibration, NPT Sample
- 30 ns sampling run with 1.5 fs time step
- T = 300 K, P = 1 bar



Snapshot of BMIM-Tf2N-solvent system, Coloring: BMIM (Blue), Tf2N (Red), Solvent (Yellow)

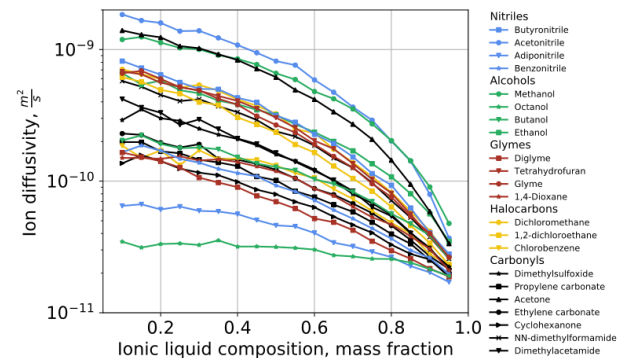


Adachi, H.; et al. *J. Chem. Phys.* **2015**, 143, 194701.
 Gogotsi, Y.; et al. *J. Phys. Chem. Lett.* **2016**, 7, 11225-11236.
 Gogotsi, Y.; et al. *J. Phys. Chem. Lett.* **2016**, 7, 11225-11236.

MUMS

Ion Diffusivity Decreases with Increasing IL Composition

- Highest ion diffusivity in [BMIM+]-[Tf2N-] with acetonitrile

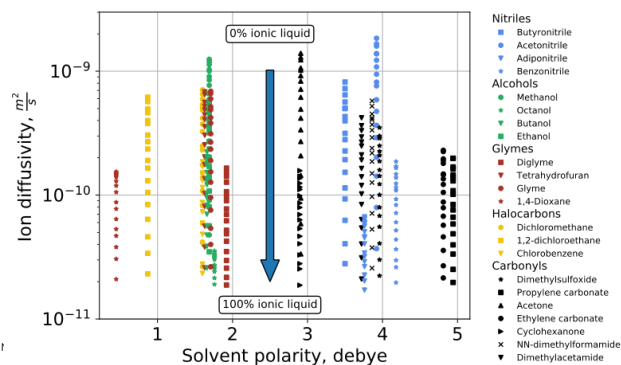


SCHOOL OF ENGINEERING

MUMS

No Correlation with Solvent Polarity

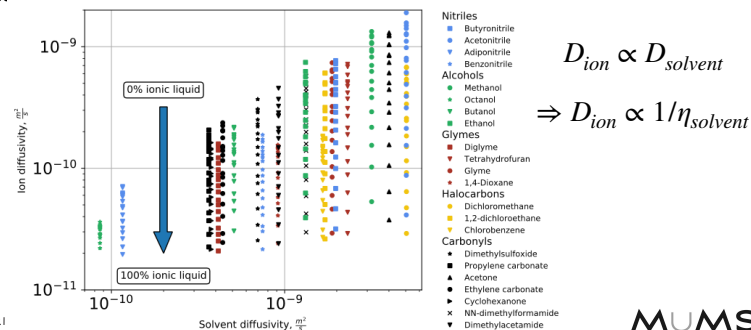
- Screening over a larger data set shows solvent polarity is not a good indicator of ion diffusivity



MUMS

Trend Between Ion and Solvent Diffusivity

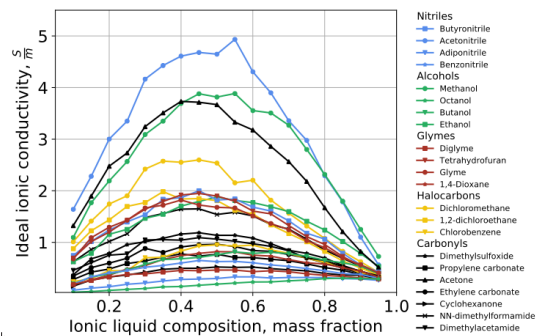
- Separate simulations of pure solvents to gather diffusivities
- Ionic liquids molecules more likely to exhibit increased molecular diffusivity in environment of molecules with faster motion



MUMS

Conductivity

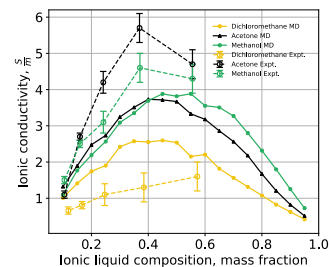
- Higher conductivity of electrolytes results in higher charging rates of supercapacitors
- Conductivity estimated through Nernst-Einstein Equation



MUMS

Conductivity

- Experimental conductivity measurements corroborate computational trends



M.W.Thompson, R. Matsumoto, R. L. Sacchi, N. C. Sanders, P.T. Cummings, "Scalable Screening of Soft Matter: Case Study of Mixtures of Ionic Liquids in Organic Solvents," J. Phys. Chem. B, **2019**, 123(6), 1340–1347. [10.1021/acs.jpcc.8b11527].

- And new QENS study: Osti, N. C.; Matsumoto, R.A.; Thompson, M.W.; Cummings, P.T.; Tyagi, M.; Mamontov, E. Microscopic Dynamics in an Ionic Liquid Augmented with Organic Solvents. J. Phys. Chem. C 2019, 123 (32), 19354–19361.

MUMS

Conclusions

- ❑ To perform MGI-style screening of soft materials requires automating many steps currently largely performed by hand
- ❑ MoSDeF is an environment for automating molecular simulations
 - mBuild, Foyer, metaMDS, Signac....
 - Everything is scriptable - important for screening
- ❑ Simulations performed within MoSDeF are completely reproducible
 - All steps are scripted, which can be distributed online through github
 - In line with identified best practices for reproducibility of computational research
- ❑ MoSDeF is being co-developed with other groups interested in transparency, reproducibility, usability, and extensibility (TRUE)

Conclusions

- ❑ Examples of materials screening enabled by MoSDeF and other tools (Signac flow)
- ❑ Nanoscale lubrication
 - Developed a model for lubrication performance based on molecular descriptors
 - Currently being extended from 820 distinct simulations to ~45,000
- ❑ Diffusivity of ionic liquids in organic solvents
 - Screened over 396 solvated IL systems
 - Acetonitrile-based systems exhibit fastest transport, alternatives exist
 - Maximum conductivity at intermediate ionic liquid composition
 - Correlation is consistent with Stokes-Einstein
 - Expanding parameter space through screening revealed true basis for trends in ion diffusivity and conductivity
- ❑ All resources available on github with version control
 - MoSDeF Hub - <https://github.com/mosdef-hub>
 - Foyer - <https://github.com/mosdef-hub/foyer> and mBuild - <https://github.com/mosdef-hub/mbuild>

Acknowledgments



❑ VU ChBE

- Clare McCabe
- Chris Iacovella
- Christoph Klein (now at Counsyl)
- Andrew Summers
- Justin Gilmer
- Matt Thompson
- Alex Yang
- Ray Matsumoto
- Jana Black
- Trevor Jones (ugrad; now at MIT for grad school)

❑ VU Institute for Software Integrated Systems

- Akos Ledezci, Gabor Karsai
- Janos Sallai, Peter Volbesi
- Umesh Timalsina



❑ Funding

- NSF ACI 1047828: Development of an Integrated Molecular Design Environment for Lubrication Systems (iMoDELS)
- NSF CDI CBET 1028374: Cyber-Enabled Design of Functional Nanomaterials
- NSF SSE ACI 1535150: Development of a Software Framework for Formalizing Forcefield Atom-Typing for Molecular Simulation
- NSF CSSI OAC 1835874: Collaborative Research: NSCI Framework: Software for Building a Community-Based Molecular Modeling Capability Around the Molecular Simulation Design Framework (MoSDeF)
- DOE: DE-AC05-00OR22725 Fluid Interface Reaction Structure and Transport Center

❑ Resources

- National Energy Research Supercomputing Center (NERSC), Lawrence Berkeley Laboratory
- Oak Ridge Leadership Class (OLC, Titan), Oak Ridge National Laboratory