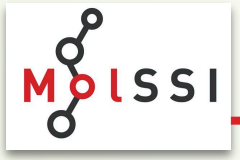
SEAMM

A Productivity Environment for Computational Materials Science

Eliseo Marin-Rimoldi, Jessica Nash, Paul Saxe
Molecular Sciences Software Institute, Virginia Tech
Blacksburg, VA 24060, USA



2020 Atomistic Simulations for Industrial Needs Workshop, NIST
Thursday 6 August 2020

What is the Issue?

- "**C**omputational **M**aterials **S**cience" and "**C**omputational **M**olecular **S**cience"
- Range of CMS Tools is Overwhelming!
  - Quantum Chemistry – codes like Gaussian, QChem, Gamess, Psi4, MOPAC*
  - Quantum Monte Carlo
  - Periodic Quantum – VASP, Quantum Espresso, FHI-aims, DFTB, Latte
  - Molecular Dynamics (using forcefield == atomistic potentials)
    - Biomolecular (forcefield) – AMBER, CHARMM, NAMD, GROMACS
    - Materials (potentials) – LAMMPS, HOOMD, GULP
  - Monte Carlo (forcefield) – Towhee, Cassandra, Gibbs
  - Helper codes: phonons, cluster expansion, GASP, K-Point grid, reaction kinetics, builders, …
  - Parameter databases: OpenKIM, Basis Set Exchange
- How to find, use and compare tools?

MolSSI

How many of these methods could help you solve your problem?

How many of these methods are you comfortable using?

Why Create SEAMM?

- There is no open-source environment covering CMS & CMS
- There are tools covering subdomains
  - ASE, Pymatgen, AiiDA, AMBER, CHARMM, RMG,…
  - SEAMM can leverage them to varying degrees

MolSSI

What is SEAMM?

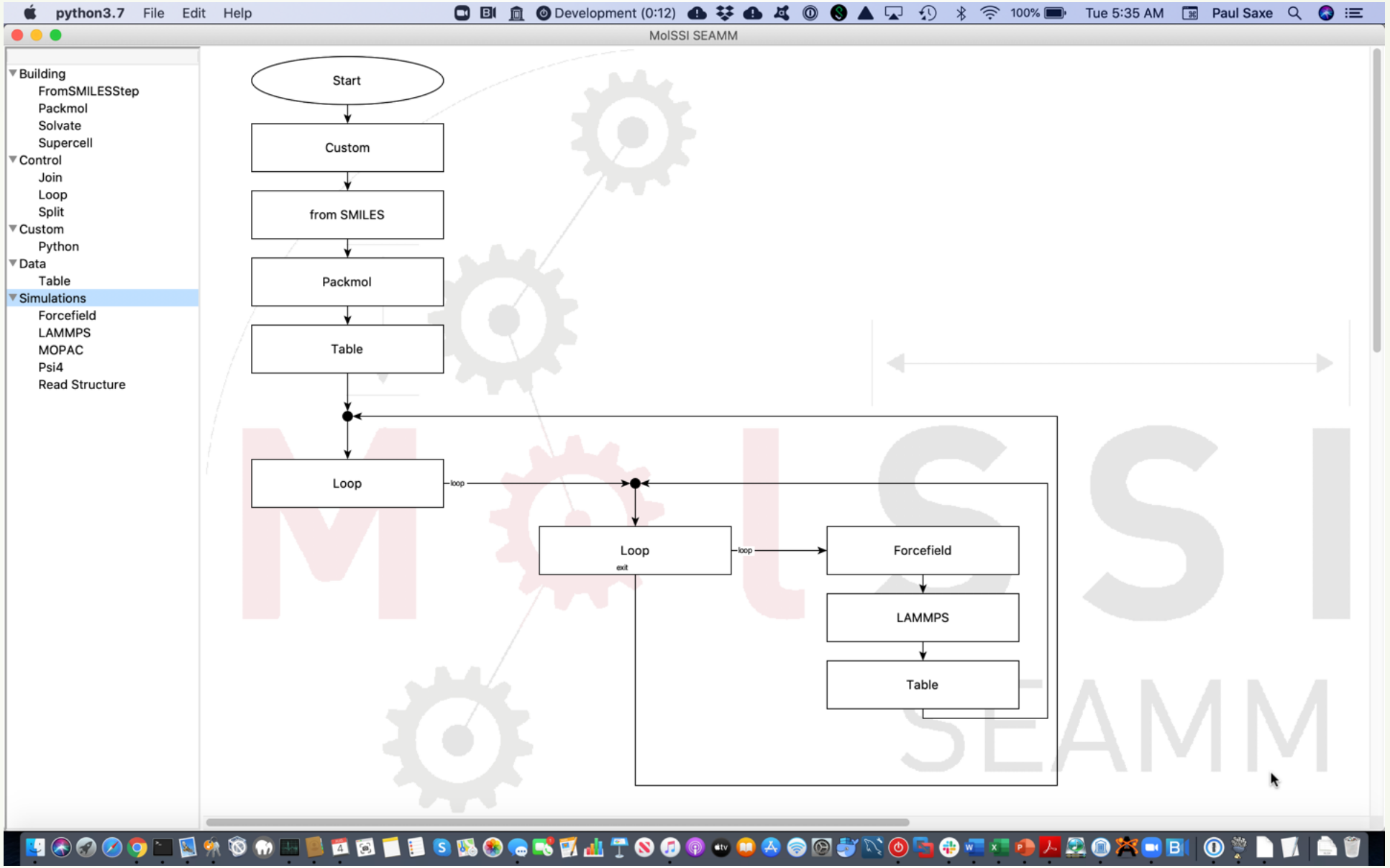
- A productivity environment for users
- A platform for developers

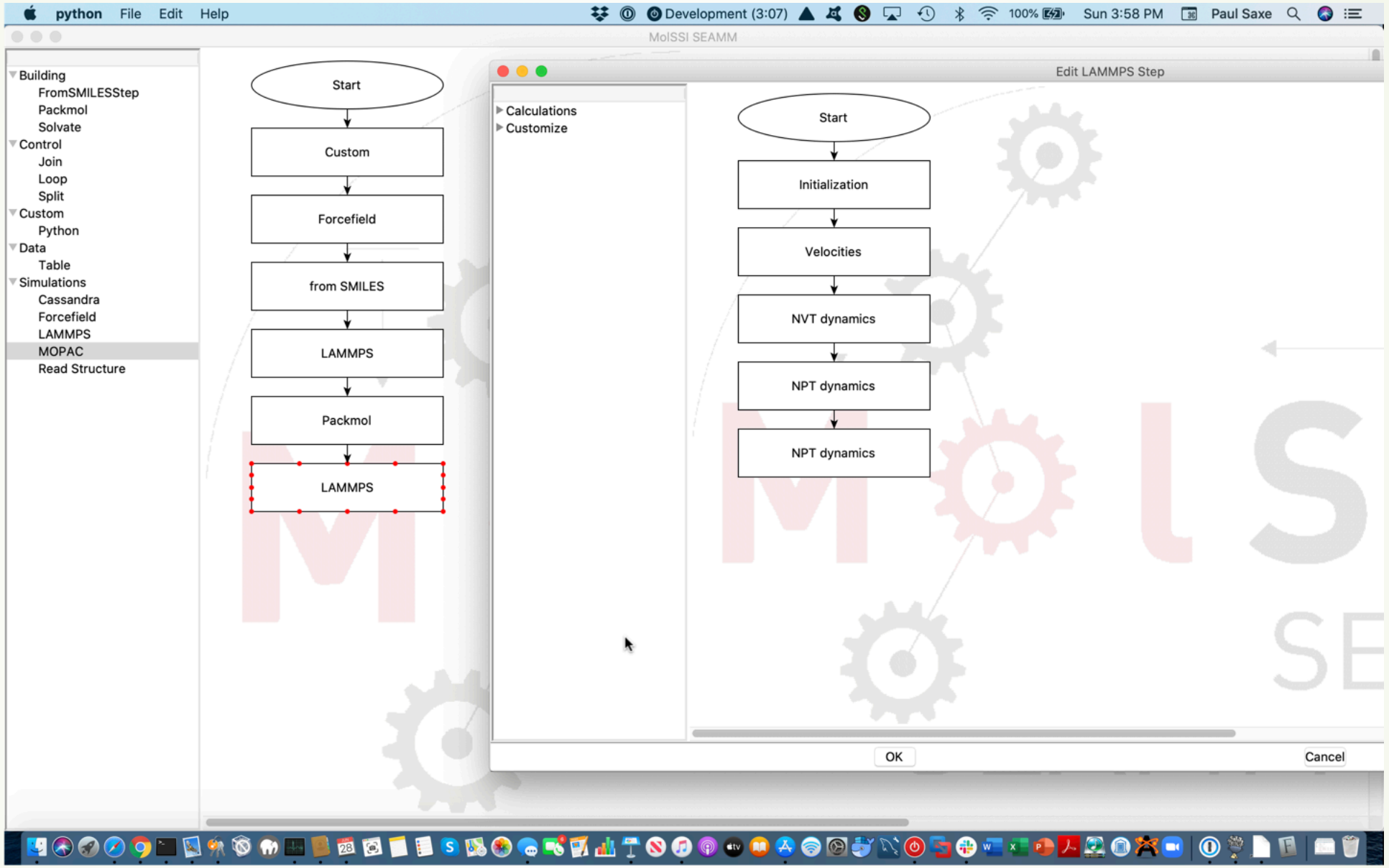
MolSSI

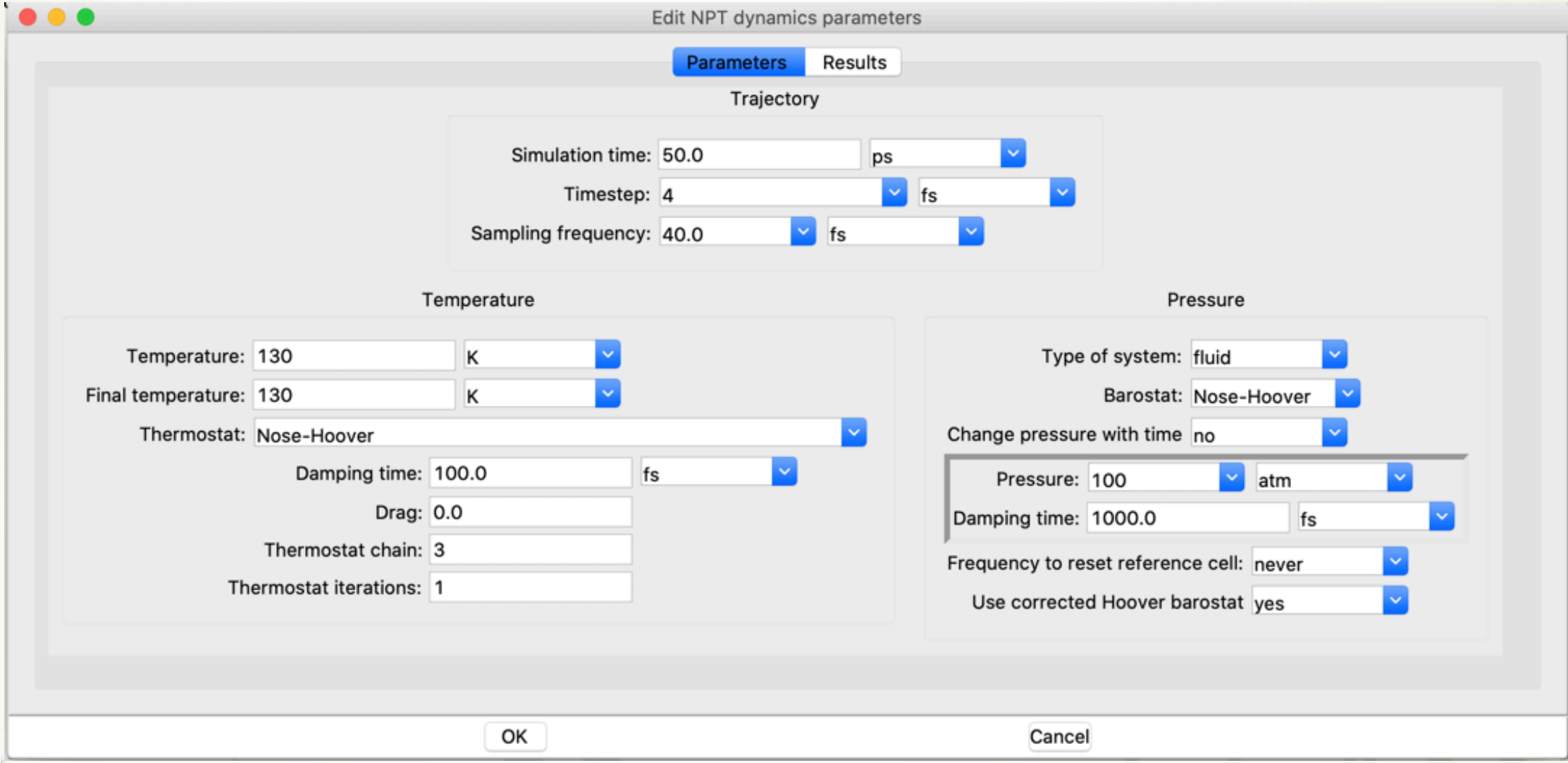
What is SEAMM, Technically?

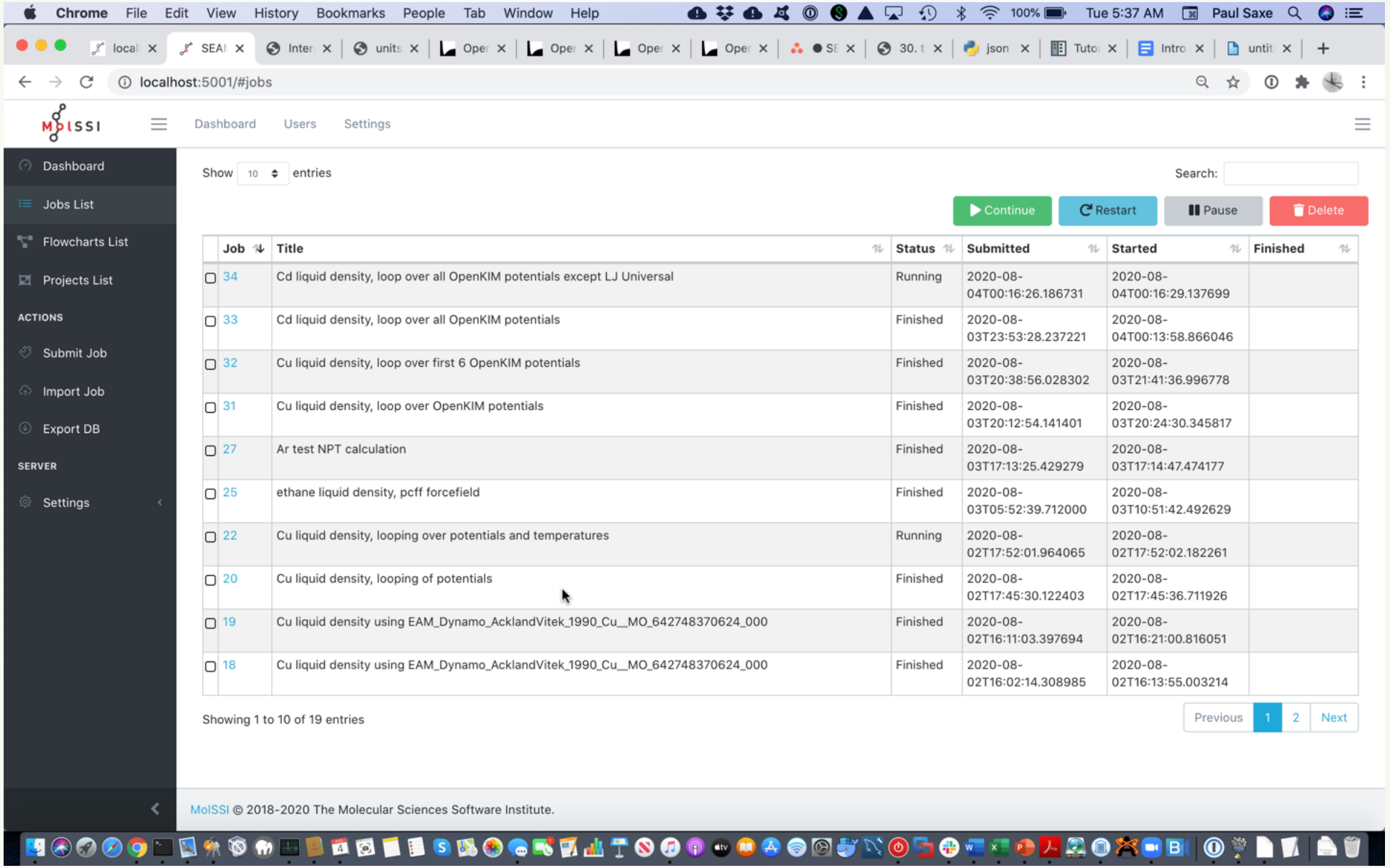
- A graphical frontend, which the user sees
- A plug-in system and extensive framework for CMS, which developers see
- A Dashboard, which is a portal to
  - A JobServer and management system for jobs
  - A Datastore containing all previous jobs and results

MolSSI

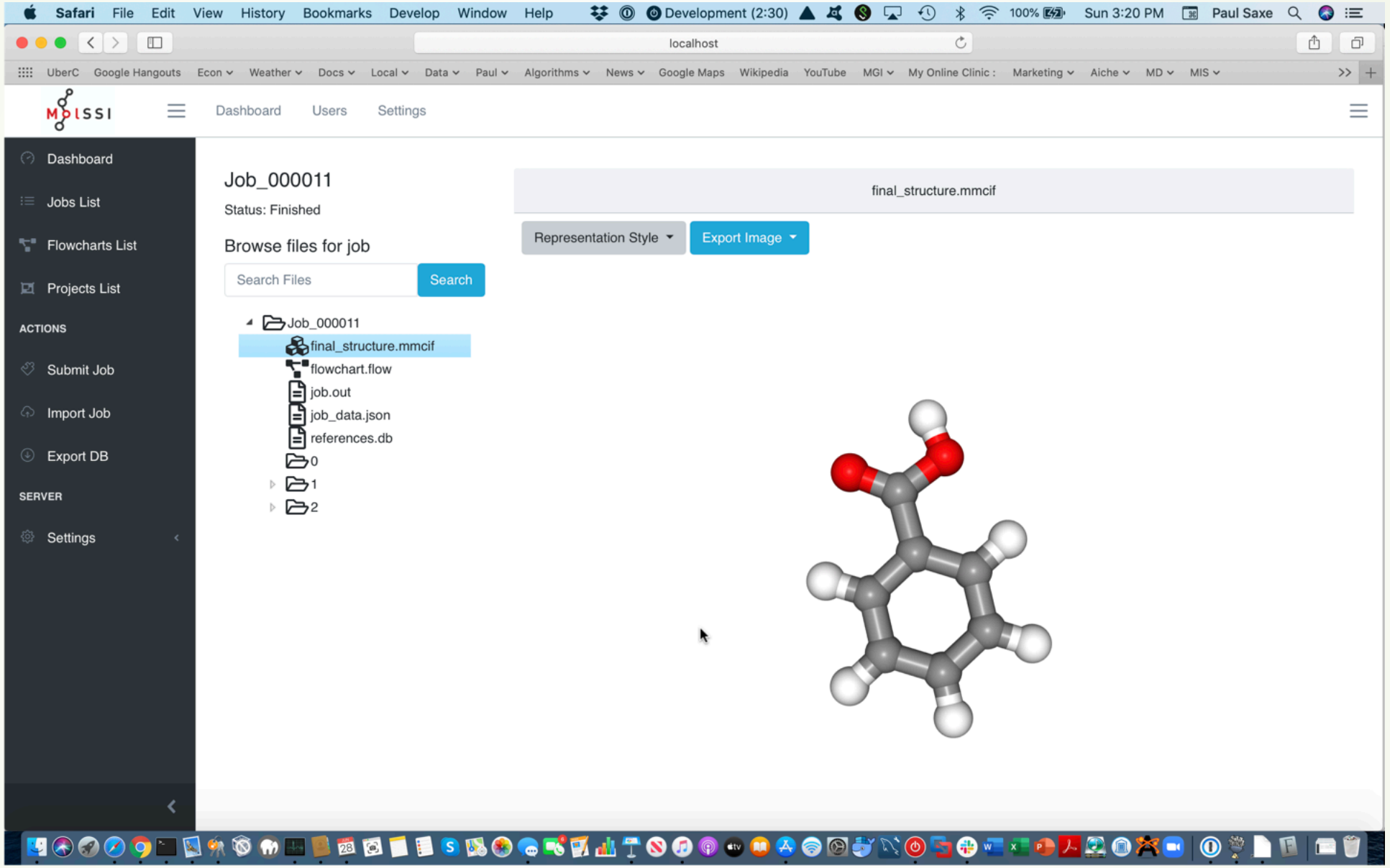

python3.7 File Edit Help
Development (0:12)
100%
Tue 5:35 AM
Paul Saxe
MolSSI SEAMM
Building
FromSMILESStep
Packmol
Solvate
Supercell
Control
Join
Loop
Split
Custom
Python
Data
Table
Simulations
Forcefield
LAMMPS
MOPAC
Psi4
Read Structure
Start
Custom
from SMILES
Packmol
Table
Loop
loop
Loop
exit
loop
Forcefield
LAMMPS
Table

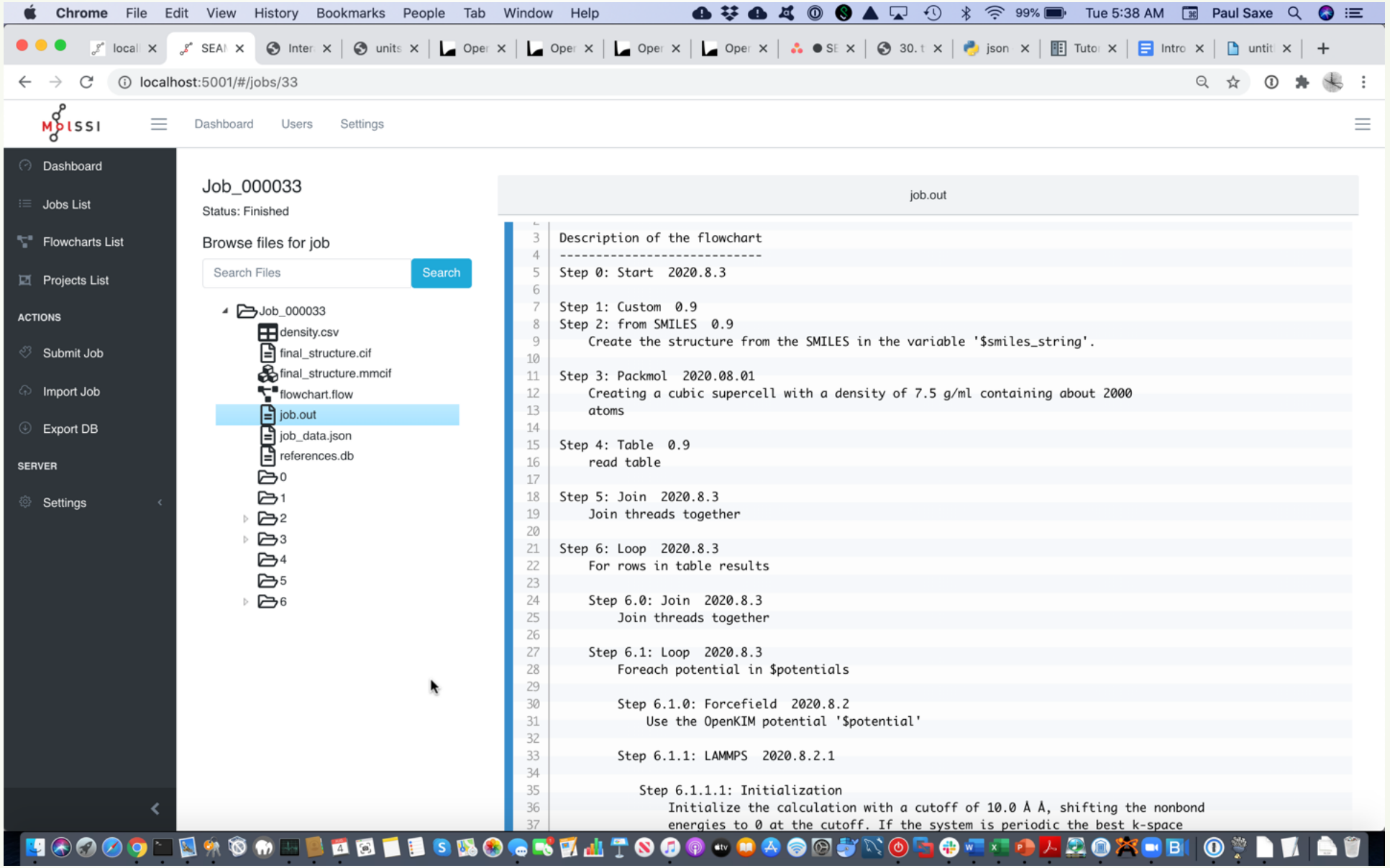

python File Edit Help
Development (3:07)
100%
Sun 3:58 PM
Paul Saxe
MolSSI SEAMM
Building
FromSMILESStep
Packmol
Solvate
Control
Join
Loop
Split
Custom
Python
Data
Table
Simulations
Cassandra
Forcefield
LAMMPS
MOPAC
Read Structure
Start
Custom
Forcefield
from SMILES
LAMMPS
Packmol
LAMMPS
Edit LAMMPS Step
Calculations
Customize
Start
Initialization
Velocities
NVT dynamics
NPT dynamics
NPT dynamics
OK
Cancel


Edit NPT dynamics parameters
Parameters
Results
Trajectory
Simulation time: 50.0 ps
Timestep: 4 fs
Sampling frequency: 40.0 fs
Temperature
Temperature: 130 K
Final temperature: 130 K
Thermostat: Nose-Hoover
Damping time: 100.0 fs
Drag: 0.0
Thermostat chain: 3
Thermostat iterations: 1
Pressure
Type of system: fluid
Barostat: Nose-Hoover
Change pressure with time no
Pressure: 100 atm
Damping time: 1000.0 fs
Frequency to reset reference cell: never
Use corrected Hoover barostat yes
OK
Cancel

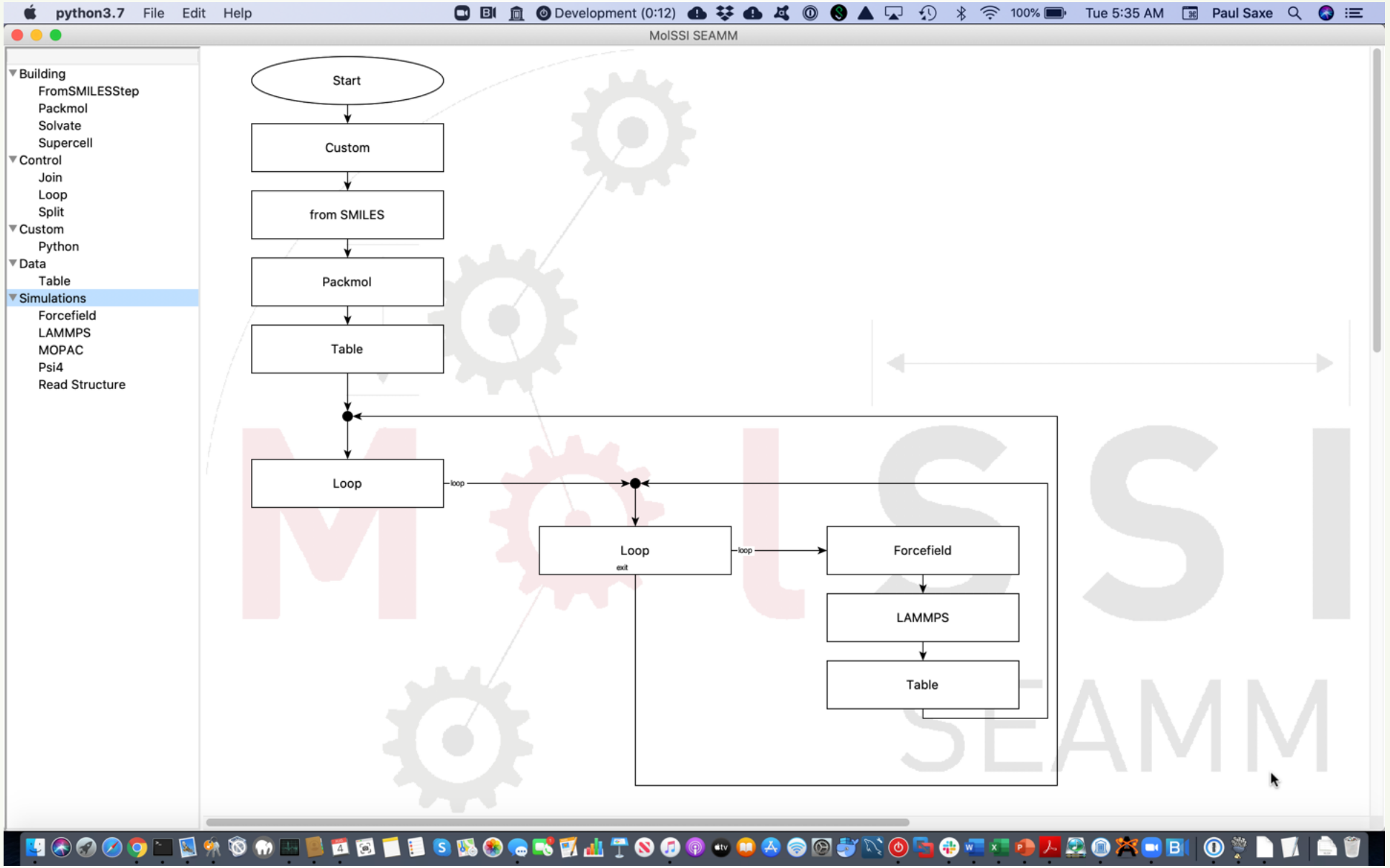

Chrome File Edit View History Bookmarks People Tab Window Help
Tue 5:37 AM
Paul Saxe
localhost:5001/#jobs
MolSSI
Dashboard
Users
Settings
Dashboard
Jobs List
Flowcharts List
Projects List
ACTIONS
Submit Job
Import Job
Export DB
SERVER
Settings
Show 10 entries
Search:
Continue
Restart
Pause
Delete
Job
Title
Status
Submitted
Started
Finished
34 Cd liquid density, loop over all OpenKIM potentials except LJ Universal Running 2020-08-04T00:16:26.186731 2020-08-04T00:16:29.137699
33 Cd liquid density, loop over all OpenKIM potentials Finished 2020-08-03T23:53:28.237221 2020-08-04T00:13:58.866046
32 Cu liquid density, loop over first 6 OpenKIM potentials Finished 2020-08-03T20:38:56.028302 2020-08-03T21:41:36.996778
31 Cu liquid density, loop over OpenKIM potentials Finished 2020-08-03T20:12:54.141401 2020-08-03T20:24:30.345817
27 Ar test NPT calculation Finished 2020-08-03T17:13:25.429279 2020-08-03T17:14:47.474177
25 ethane liquid density, pcff forcefield Finished 2020-08-03T05:52:39.712000 2020-08-03T10:51:42.492629
22 Cu liquid density, looping over potentials and temperatures Running 2020-08-02T17:52:01.964065 2020-08-02T17:52:02.182261
20 Cu liquid density, looping of potentials Finished 2020-08-02T17:45:30.122403 2020-08-02T17:45:36.711926
19 Cu liquid density using EAM_Dynamo_AcklandVitek_1990_Cu__MO_642748370624_000 Finished 2020-08-02T16:11:03.397694 2020-08-02T16:21:00.816051
18 Cu liquid density using EAM_Dynamo_AcklandVitek_1990_Cu__MO_642748370624_000 Finished 2020-08-02T16:02:14.308985 2020-08-02T16:13:55.003214
Showing 1 to 10 of 19 entries
Previous 1 2 Next
MolSSI © 2018-2020 The Molecular Sciences Software Institute.

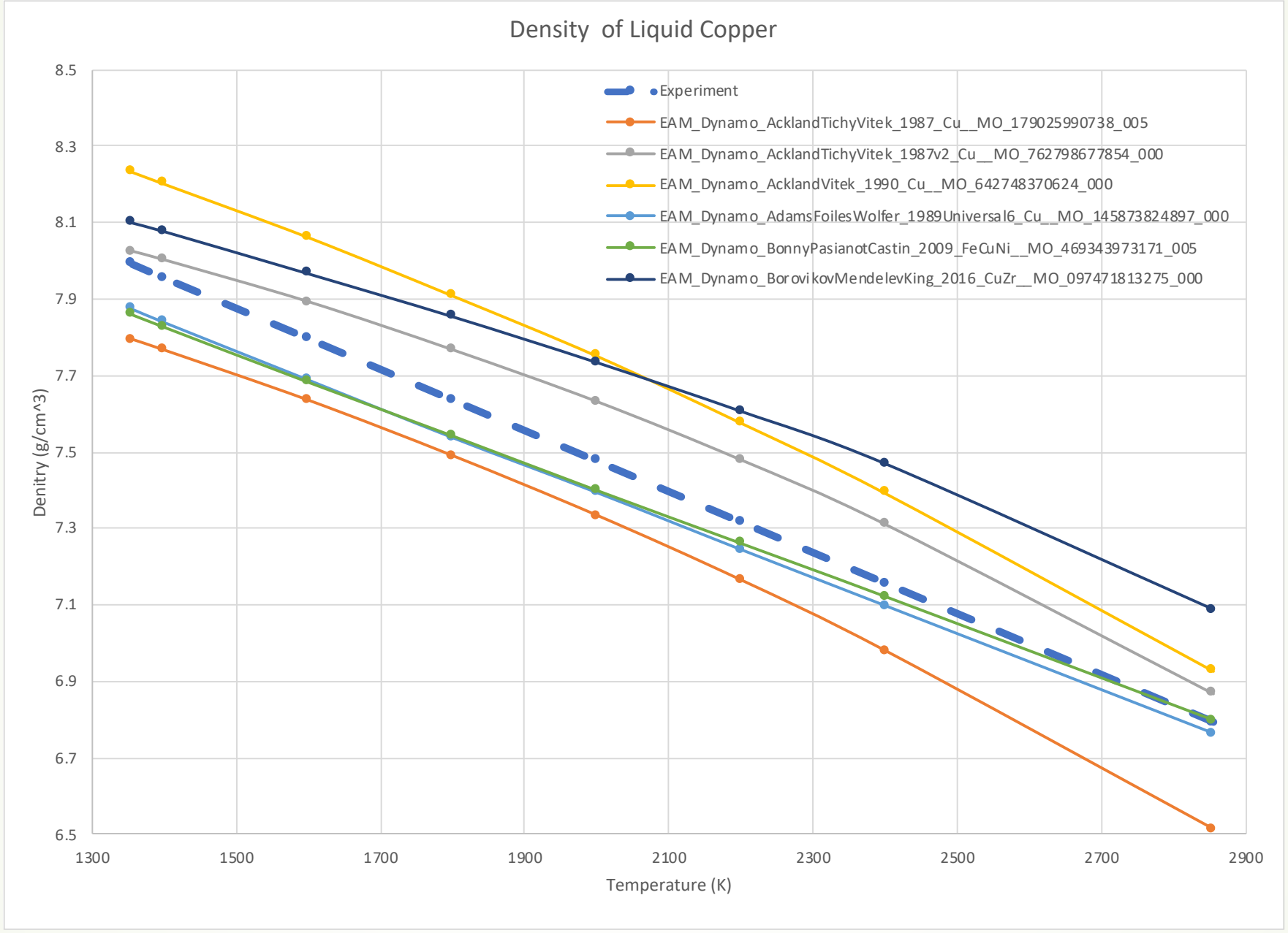
MolSSI


Safari File Edit View History Bookmarks Develop Window Help
Development (2:30)
100% Sun 3:20 PM Paul Saxe
localhost
UberC Google Hangouts Econ Weather Docs Local Data Paul Algorithms News Google Maps Wikipedia YouTube MGI My Online Clinic : Marketing Aiche MD MIS
MOLSSI
Dashboard Users Settings
Dashboard
Jobs List
Flowcharts List
Projects List
ACTIONS
Submit Job
Import Job
Export DB
SERVER
Settings
Job_000011
Status: Finished
Browse files for job
Search Files
Search
Job_000011
final_structure.mmcif
flowchart.flow
job.out
job_data.json
references.db
0
1
2
final_structure.mmcif
Representation Style
Export Image


Chrome File Edit View History Bookmarks People Tab Window Help
Tue 5:38 AM
Paul Saxe
localhost:5001/#/jobs/33
MOLSSI
Dashboard
Users
Settings
Dashboard
Jobs List
Flowcharts List
Projects List
ACTIONS
Submit Job
Import Job
Export DB
SERVER
Settings
Job_000033
Status: Finished
Browse files for job
Search Files
Search
Job_000033
density.csv
final_structure.cif
final_structure.mmcif
flowchart.flow
job.out
job_data.json
references.db
0
1
2
3
4
5
6
job.out
Description of the flowchart
Step 0: Start 2020.8.3
Step 1: Custom 0.9
Step 2: from SMILES 0.9
Create the structure from the SMILES in the variable '$smiles_string'.
Step 3: Packmol 2020.08.01
Creating a cubic supercell with a density of 7.5 g/ml containing about 2000
atoms
Step 4: Table 0.9
read table
Step 5: Join 2020.8.3
Join threads together
Step 6: Loop 2020.8.3
For rows in table results
Step 6.0: Join 2020.8.3
Join threads together
Step 6.1: Loop 2020.8.3
Foreach potential in $potentials
Step 6.1.0: Forcefield 2020.8.2
Use the OpenKIM potential '$potential'
Step 6.1.1: LAMMPS 2020.8.2.1
Step 6.1.1.1: Initialization
Initialize the calculation with a cutoff of 10.0 Å Å, shifting the nonbond
energies to 0 at the cutoff. If the system is periodic the best k-space

MOLSSI


python3.7 File Edit Help
Development (0:12)
100%
Tue 5:35 AM
Paul Saxe
MolSSI SEAMM
Building
FromSMILESStep
Packmol
Solvate
Supercell
Control
Join
Loop
Split
Custom
Python
Data
Table
Simulations
Forcefield
LAMMPS
MOPAC
Psi4
Read Structure
Start
Custom
from SMILES
Packmol
Table
Loop
loop
Loop
exit
loop
Forcefield
LAMMPS
Table


Density of Liquid Copper
Experiment
EAM_Dynamo_AcklandTichyVitek_1987_Cu__MO_179025990738_005
EAM_Dynamo_AcklandTichyVitek_1987v2_Cu__MO_762798677854_000
EAM_Dynamo_AcklandVitek_1990_Cu__MO_642748370624_000
EAM_Dynamo_AdamsFoilesWolfer_1989Universal6_Cu__MO_145873824897_000
EAM_Dynamo_BonnyPasianotCastin_2009_FeCuNi__MO_469343973171_005
EAM_Dynamo_BorovikovMendelevKing_2016_CuZr__MO_097471813275_000
Denitry (g/cm^3)
8.5
8.3
8.1
7.9
7.7
7.5
7.3
7.1
6.9
6.7
6.5
1300
1500
1700
1900
2100
2300
2500
2700
2900
Temperature (K)

Main Goals

- Productivity and Usability
  - For users
  - For developers
- Reproducibility and Replicability
- Recognizing and Crediting Authors

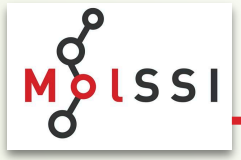
- To bootstrap a multi-sided platform:
  - Users benefit when there are more plug-ins
  - Plug-ins benefit from more users

    and
  - Plug-ins benefit from more plug-ins, because that brings more functionality and users

MolSSI

Productivity and Usability: Users

- Installation – begin at the beginning
  - Currently a few steps of manual installation
  - Moving to a single bootstrap install
    - Automatic installation and update of components
    - Know from a flowchart what is required
    - Will provide a database of information about plug-ins
  - Installation at computer centers is intrinsically more complicated
    - Leverage other tools and installations
    - Hopefully get the centers to help



Productivity and Usability: Users

- Graphical User Interface
  - Clearly easier for beginners
  - Flowcharts
    - Avoid having to set up the GUI repeatedly
    - Are editable and composable
    - Allow sharing and using experts' "best practices"
  - With care do not limit advanced users
- Job Management
- Job database stores all previous jobs
- List of appropriate citations based on actual calculations

MolSSI

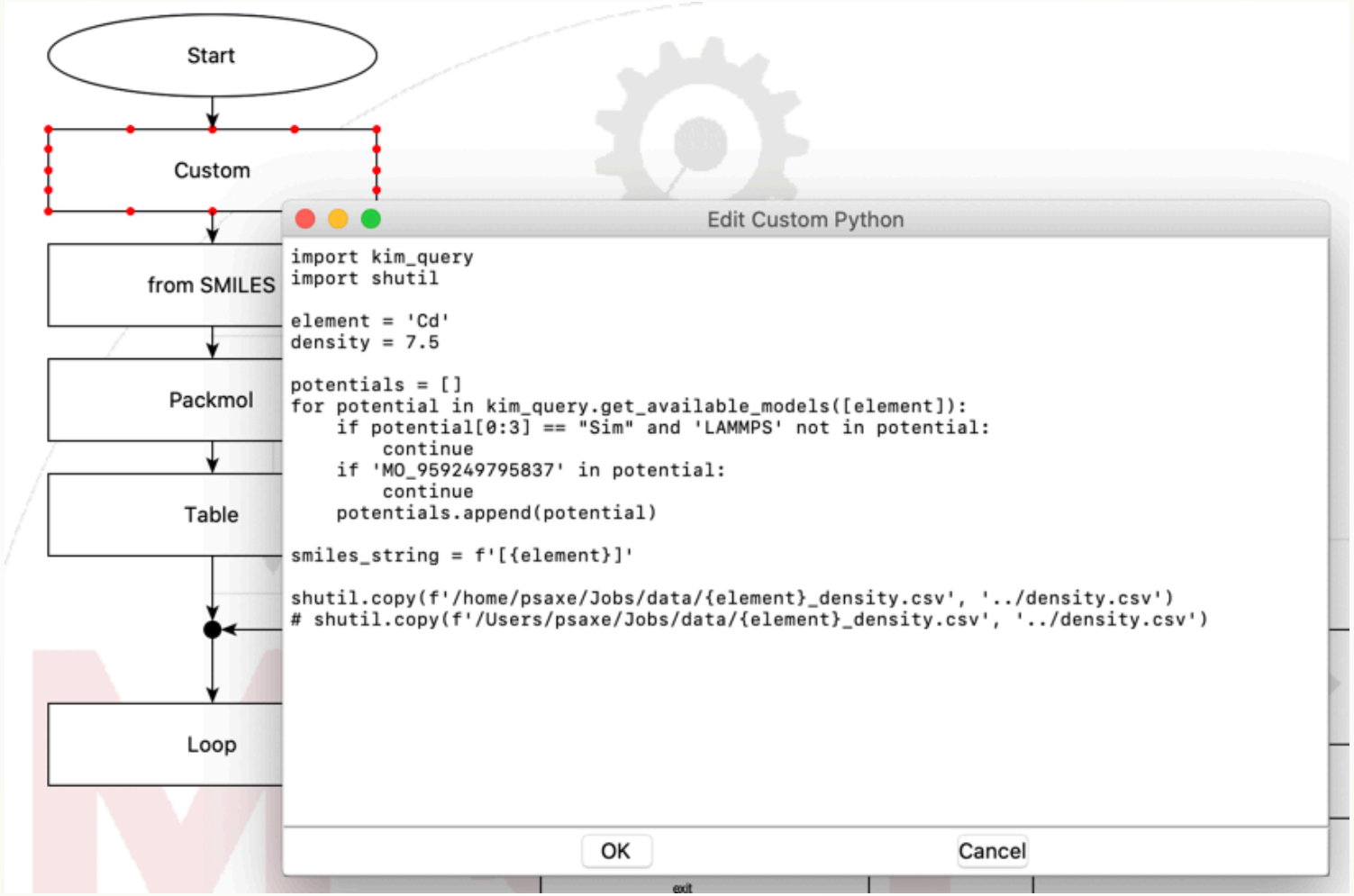
Productivity and Usability: Users

- Interactive Web Dashboard
  - Accessible from anywhere where you have permission to access host machine (i.e. ports are open, or can use ssh tunneling)
  - Project-based
  - Full permission handling (projects like folders, jobs like files) – in development
  - Allows online collaboration based on access and permissions

MolSSI

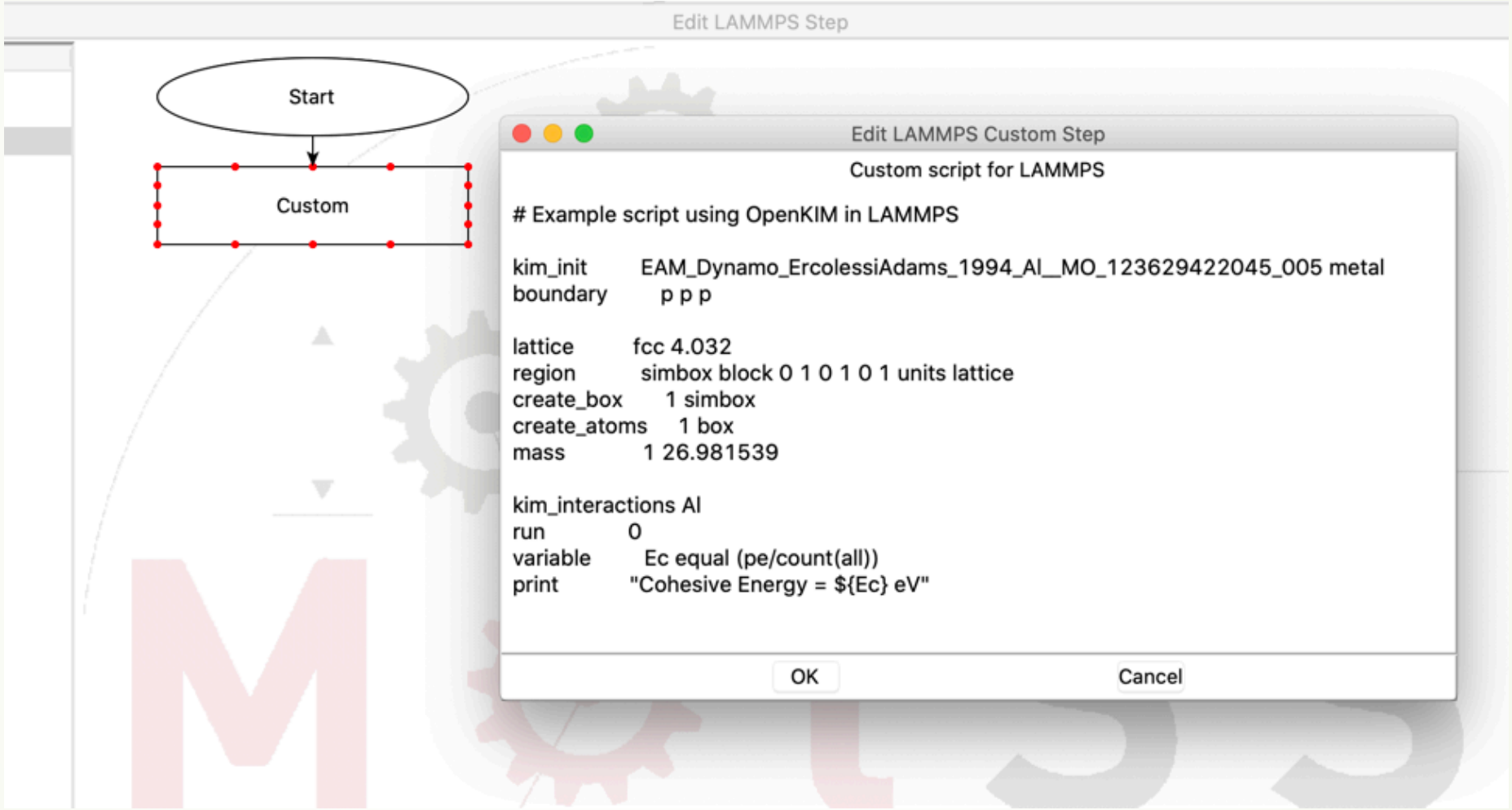
For Experts: Custom Steps and Overrides

Python



Custom Steps and Overrides

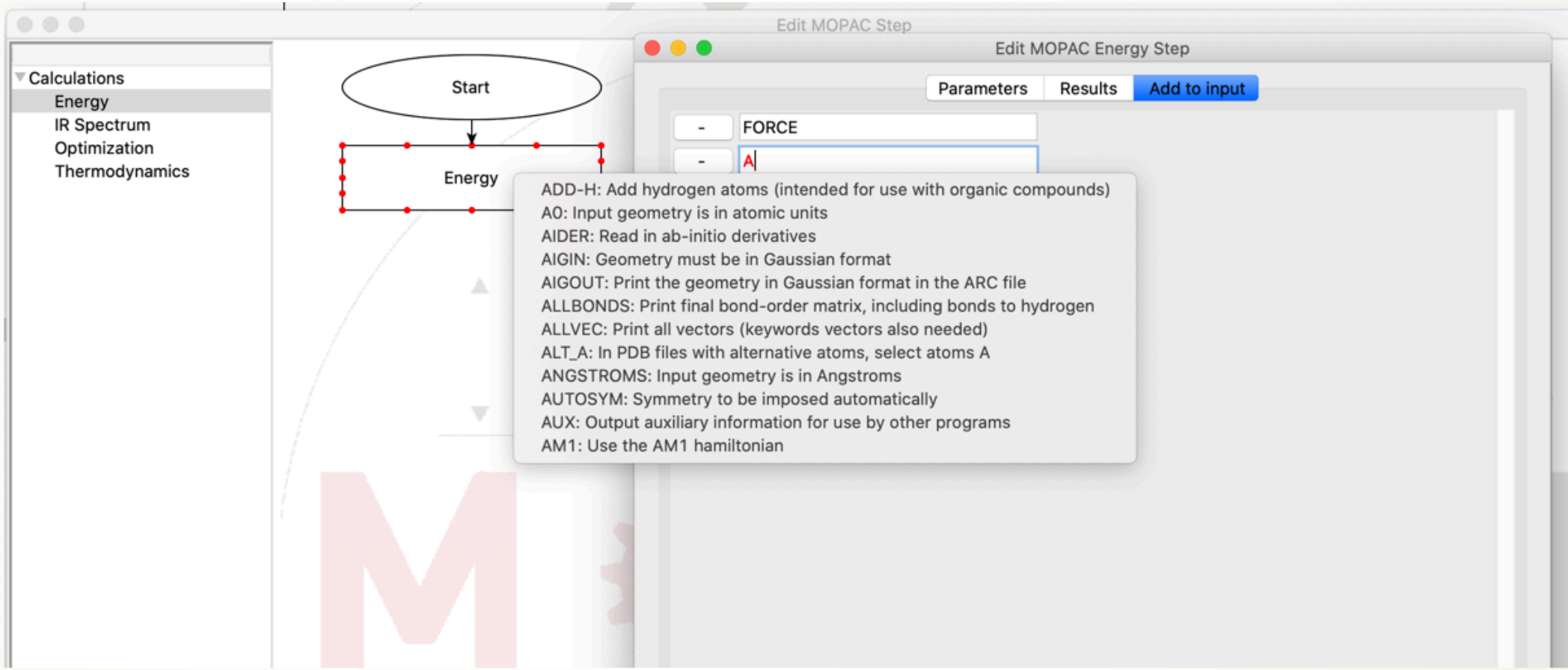
LAMMPS



Can be interspersed with other substeps, not just replacing everything as this example

Custom Steps and Overrides

MOPAC

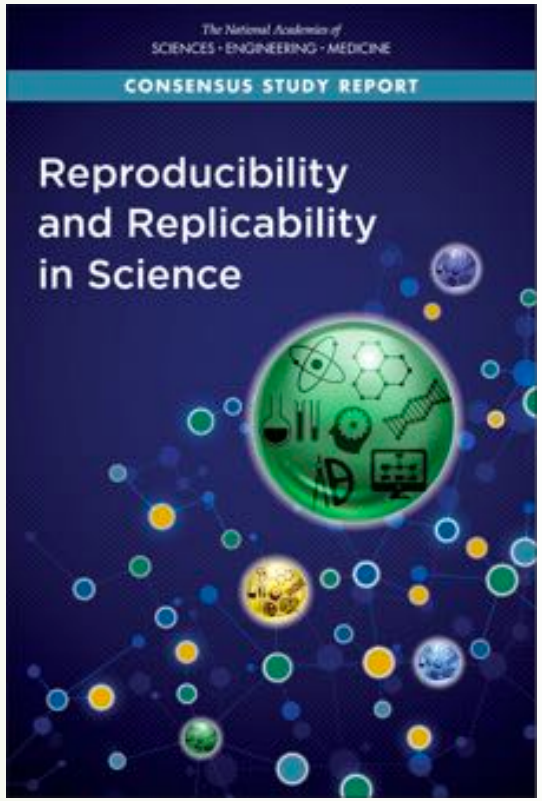


Productivity and Usability: Developers

- Cookie-cutters available to create template projects
- Extensive and growing infrastructure
  - For control options and defining results
  - For handling units and conversions
  - Specialized widgets including units, periodic table, etc.
  - General and flexible internal data structures for the molecule/crystal structure
  - Statistical analysis for MD and Monte-Carlo
  - Database for citations
  - Translators such as OpenBabel
  - Cheminformatics via RDKit
- Growing ecosystem of plug-ins

MolSSI

Reproducibility and Replicability



Reproducibility is obtaining consistent results using the same input data, computational steps, methods, and code, and conditions of analysis. This definition is synonymous with "computational reproducibility," and the terms are used interchangeably in this report.

Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data. Two studies may be considered to have replicated if they obtain consistent results given the level of uncertainty inherent in the system under study.

https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science

Reproducibility and Replicability

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

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

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

[†]Laboratory of Engineering Thermodynamics (LTD), University of Kaiserslautern, Kaiserslautern, Germany

[‡]Institut für Thermodynamik (ift), Technische Universität Braunschweig, Braunschweig, Germany

[§]Lehrstuhl für Technische Thermodynamik (LTT), RWTH Aachen University, Aachen, Germany

[∥]Thermodynamics and Energy Technology (ThEt), University of Paderborn, Paderborn, Germany

[⊥]Fraunhofer Institute for Industrial Mathematics (ITWM), Kaiserslautern, Germany

[#]High Performance Computing Center (HLRS), Stuttgart, Germany

[▽]Department of Philosophy, Bielefeld University, Bielefeld, Germany

MolSSI

http://pubs.acs.org/doi/10.1021/acs.jctc.7b00489

Reproducibility and Replicability

In the present round robin study, the following fundamental question is addressed: Will different user groups working with different simulation codes obtain coinciding results within the statistical uncertainty of their data?

A set of 24 simple simulation tasks is defined and solved by five user groups working with eight molecular simulation codes: DL_POLY, GROMACS, IMC, LAMMPS, ms2, NAMD, Tinker, and TOWHEE.

The results reveal the challenges of carrying out molecular simulations. Several iterations were needed to eliminate gross errors.

For most simulation tasks, the remaining deviations between the results of the different groups are acceptable from a practical standpoint, but they are **often outside of the statistical errors** of the individual simulation data. However, there are also cases where the deviations are unacceptable.

MOLSSI

1 Example for results after the first iteration

Figure 1 shows a typical example for results submitted to the central instance after the first iteration.

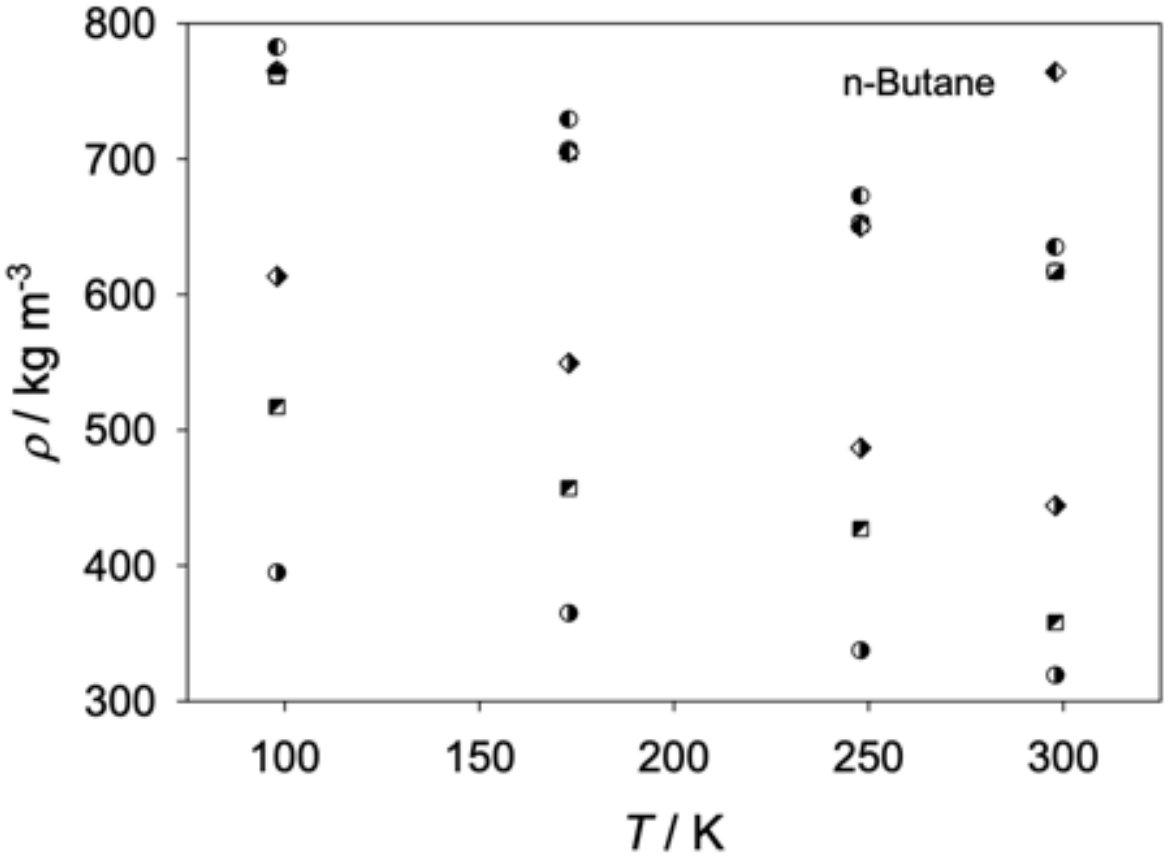


Figure 1: Example for results submitted to the central instance after the first iteration. The task was to simulate the density of n-butane at 41 MPa for the indicated temperatures with the OPLS force field. Different symbols stand for results from different user groups working with different simulation codes. There are seven different symbols, but due to overlap not all of them can be discerned. The extreme differences are in most cases due to input errors that were eliminated later, cf. Figure 3 from the main text.. A declaration of the symbols is

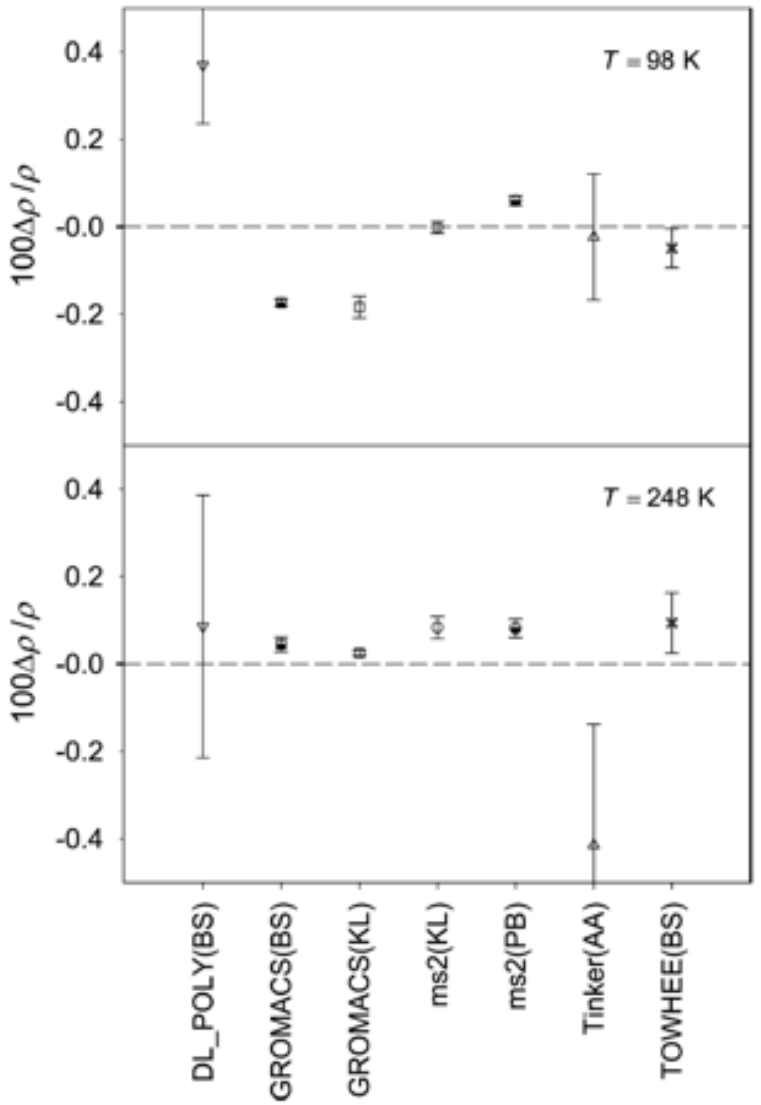


Figure 4. Statistical uncertainty of the data obtained for the density of *n*-butane at 41 MPa and 98 K (top) and 248 K (bottom) from the OPLS force field. Symbols: mean values with error bars determined from block averages of the production phase. The large error bars of the DL_POLY and Tinker data are due to the chosen barostat. Dashed line: arithmetic mean of all results.

Flowcharts Are Reproducible…

- …mostly! For a time they are absolutely reproducible, but it decays.
- Versions of plug-ins are known and stored
- More difficult to track e.g. LAMMPS versions used, hardware such as GPU's
- We can do a "good enough" but not perfect job

```
Description of the flowchart
----------------------------
Step 0: Start  2020.8.3

Step 1: Forcefield  2020.8.2
    Read the forcefield file 'pcff2018.frc' and use the default forcefield.

Step 2: from SMILES  0.9
    Create the structure from the SMILES '[Ar]'

Step 3: LAMMPS  2020.8.2.1+0.gd2a6e7b.dirty

    Step 3.1: Initialization
        Initialize the calculation with a cutoff of 10.0 Å Å, shifting the nonbond
        energies to 0 at the cutoff. If the system is periodic the best k-space
        acceleration method for the  molecular system will be chosen. The accuracy
        goal is 1.00e-05.

    Step 3.2: Energy
        Single-point energy calculation.


Step 4: Packmol  2020.08.01
    Creating a cubic supercell with a density of 1.1 g/ml containing about 1000
    atoms
```

development version

Flowcharts are Replicable

- Depends on care and thoughtfulness of plug-in designers
- Many "silly" errors like typos eliminated

- Editing and applying to other systems much less error prone

- Still challenging! But developing plug-ins gives an opportunity to rethink and do better

MolSSI

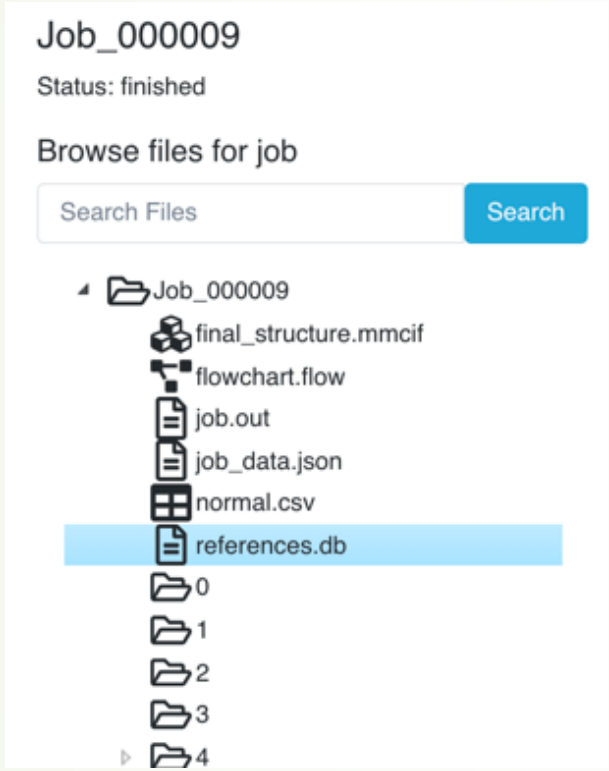
Recognizing and Crediting Authors

- Plug-ins use the Reference Handler to capture citations
  - At runtime, since that is only time all the details are known
    - Forcefields, basis sets, etc. depend on the system
  - Given an importance (1=most, 2=less ...)
  - Count of uses captured automatically.
- Citations printed and stored in a database with each job
- Working on developing a publishing tool
  - User marks all the jobs for e.g. a paper
  - System merges all the citations into a single list
  - Optionally prepares the flowcharts and critical files for archiving

MolSSI

https://github.com/molssi/reference_handler

Citations




```
Primary references:

(1) James J. P. Stewart; Optimization of parameters for semiempirical methods
    {VI}: more modifications to the {NDDO} approximations and re-optimization of
    parameters; Journal of Molecular Modeling; 2012; 19;
    10.1007/s00894-012-1667-x.
     (used 12 times)

(2) Jon Baker; An algorithm for the location of transition states; Journal of
    Computational Chemistry; 1986; 7; 10.1002/jcc.540070402.
     (used 12 times)
```

An Open-Source Platform for CMS

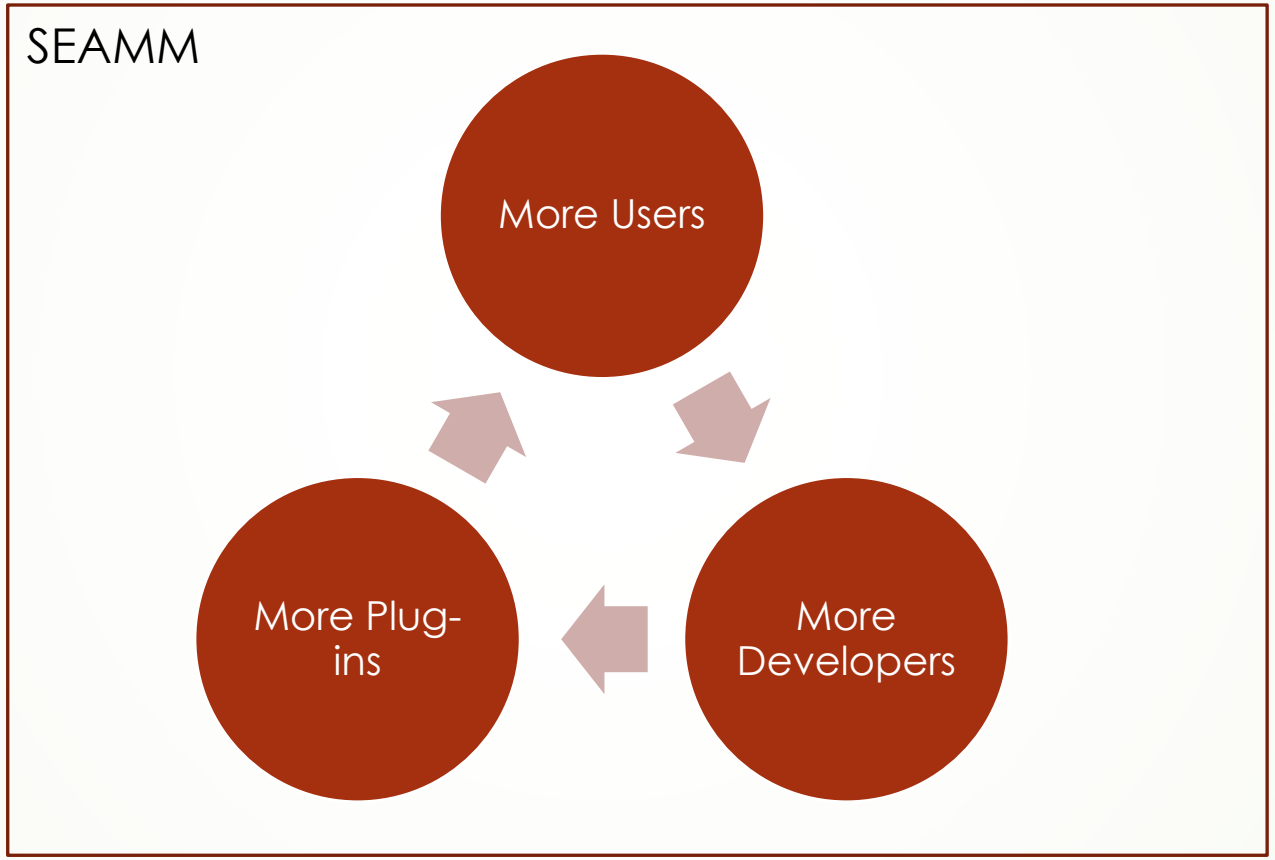
- As a user SEAMM will help you be more productive, reduce tedious work, focus on the science and engineering, solve more problems and, if you are in academia, publish more papers.
- As a developer, SEAMM will get your code into the hands of more users

  (and help them cite your work!)
- More plug-ins is good for everyone

MolSSI

What Does Success Look Like?

Funding From NSF ACI-1547580. Thanks!!!!



Demo tomorrow
2-3 PM EDT

https://molssi-seamm.github.io/index.html psaxe@vt.edu https://github.com/molssi-seamm