The MAterials Simulation Toolkit for Defect and Diffusion Calculations

Tam Mayeshiba, Henry Wu, Zhewen Song, Ben Afflerbach, Professor Dane Morgan

University of Wisconsin-Madison

NIST/CHiMaD Data Workshop, Evanston, IL
May 2, 2016

http://pythonhosted.org/MAST
MAST’s role in the Materials Genome Initiative

Solid oxide fuel cell server, bloomenergy.com
TRISO fuel particle, US DOE
Reactor pressure vessel, US DOE

MGI image retouched from https://www.whitehouse.gov/mgi
What is MAST?

MAterials Simulation Toolkit

Motivated by materials science: defects and diffusion

Created for simulations (materials calculations)

Workflow manager and post-processor

\[ D = A \exp \left( \frac{-Q}{kT} \right) \]
Outline: Advantages of MAST

• Automatically manage workflows
  – MAST workflow management overview
  – Setting up a simple defect and diffusion workflow
  – Actual diffusion workflow schematic
  – Modify a workflow at any time
  – Manage workflows on various clusters
• Iterate easily over calculation system features
• Store and share workflow knowledge (VASP, diffusion, defects)
• Produce highly organized, consistently calculated data
  – Dilute solute diffusion
  – Scale examples of MAST-generated data
  – Storing MAST-generated data
MAST workflow management overview

- Crontab or user command
- Jobs on queue
- Workflow directories with input files and metadata
- Evaluate workflow logic
- New jobs to queue
- Finished workflows to archive
Setting up a simple defect and diffusion workflow

Directed Acyclic Graph (DAG) structure

How the DAG appears in the MAST input file:

- **perfect_A**
- **defect_B**
- **defect_C**
- **defect_B, defect_C**
- **transition_D**

MAST locates and creates defects

MAST pairs up defects and sets up the path
• 32 steps (not all steps are shown)
• Managing one case by hand is time-consuming and error prone
• Managing many workflows for many host-solute combinations by hand would be even worse => use MAST instead
Modify a workflow at any time

Missed a step? Add branches to workflows mid-completion and rerun

Example: Forgot the solute hop! Add a new branch with a defect calculation and a path calculation.
Manage workflows on various clusters

- Low headnode load
- Process-independent job monitor
  - Does not require a continuous process
  - Monitor gets information from text files: stable
  - Rebuilt each time: can be interrupted or fail without severe consequences
- Install and run without root access
Iterate easily over calculation system features

Same perovskite oxygen migration barrier workflow applied to different chemical systems, strain states, and dopant positions

Store and share workflow knowledge

Extensive out-of-the-box functionality

- Full diffusion coefficient workflow
  - Phonons
  - Multiple hops
  - Frequency models for diffusion

- Full defect formation energy workflow
  - Charged defects
  - Finite size scaling
  - Potential alignment

Save custom workflows for future use

- Input file saved with data
- Reproduce results
Produce highly-organized, consistently-calculated data

Example: Dilute solute diffusion

- **Al**: 7 hosts (5 FCC, 1 HCP, 1 BCC) > 230 host-solute combinations

- **Ni**: 0.1

Experiment Diffusion Coefficient [cm$^2$/s]

Theory Diffusion Coefficient [cm$^2$/s]

diffusiondata.materialshub.org; H. Wu et al., Scientific Data (2016, accepted).
Scale examples of MAST-generated data

- **Solute diffusion**: 20-30 steps per workflow x 230+ host-solute systems => over 6000 calculations

- **Perovskite oxygen migration**, with and without strain: 15-25 steps per workflow x 200+ workflows and 392 steps per workflow x 4 systems => over 5000 calculations

- **Surface exchange coefficient descriptor in perovskites**: 3 steps per workflow x 1850+ compositions => over 5500 calculations

MAST is good for handling complex workflows and for copying workflows over many systems.

[1] H. Wu et al., Scientific Data (2016, accepted)  
Storing MAST-generated data

Archived data
- NIST Data repository
  - http://hdl.handle.net/11256/102
  - http://hdl.handle.net/11256/76
- Figshare
  - https://dx.doi.org/10.6084/m9.figshare.1546772.v2
- Zenodo
  - https://dx.doi.org/10.5281/zenodo.48656

Interactive data
- Own dilute solute diffusion database website https://diffusiondata.materials hub.org
- Materials Project dataset: MPContribs (in progress)
- Citrine/Citrination?

Consolidation strategies?

Workflow management of calculations
3-30 steps per workflow

Data acquisition and processing
200-1850 workflows per data set

Data repository
5000-6000 calculations per data set, 5-30 GB
MAST was built to:

- Automatically manage workflows
- Iterate easily over calculation system features
  - Composition, strain, defects
- Store and share workflow knowledge
  - Included workflows emphasize diffusion and defect calculations
  - Almost all VASP calculations can be automated using MAST
- Produce highly organized, consistently calculated data

Github repository: [http://github.com/uw-cmg/MAST](http://github.com/uw-cmg/MAST)
Documentation: [http://www.pythonhosted.org/MAST](http://www.pythonhosted.org/MAST)
Python Package Index: [http://pypi.python.org/pypi/MAST](http://pypi.python.org/pypi/MAST)
### Acknowledgements

#### COMPUTATIONAL MATERIALS GROUP

**Faculty**
* Izabela Szlufarska  
* Dane Morgan  

**Postdocs**
* Georgios Bokas  
* Henry Wu  
* Mahmood Mamivand  
* Shipeng Shu  
* Yueh-Lin Lee  
* Guangfu Luo  
* Jia-Hong Ke  
* Ryan Jacobs  
* Wei Xie

**Graduate Students**
* Amy Kaczmarowski  
* Austin Way  
* Chaiyapat Tangpatjaroen  
* Franklin Hobbs  
* Huibin Ke  
* Jie Feng  
* Mehrdad Arjmand  
* Shuxiang Zhou  
* Xing Wang  
* Zhewen Song  
* Ao Li  
* Benjamin Afflerbach  
* Cheng Liu  
* Hao Jiang  
* Hyunseok Ko  
* Lei Zhao  
* Shenzhen Xu  
* Tam Mayeshiba  
* Yipeng Cao

**Visiting and Undergraduate Students**
* Aren Lorenson  
* Haotian Wu  
* Josh Perry  
* Liam Witteeman  
* Zachary Jensen  
* Benjamin Anderson  
* Jason Maldonis  
* Jui-Shen Chang  
* Tom Vandenberg

---

T. Mayeshiba: NSF GRFP DGE-0718123,  
UW-Madison GERS  
H. Wu, Z. Song, B. Afflerbach: NSF SI2-SSI 1148011
Calculations: UW-Madison Center for High Throughput Computing;  
University of Kentucky Lipscomb Center; NERSC DE-AC02-05CH11231