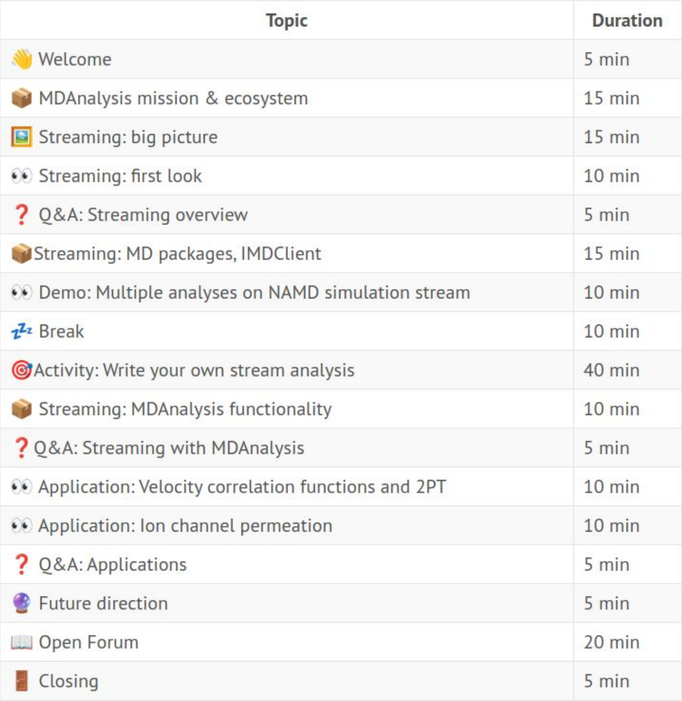
MDAnalysis Streaming Online Workshop

Welcome

- Welcome to the MDAnalysis streaming workshop!

- Collaboration with the **Heyden** and **Beckstein** labs at ASU

- Funded through NSF CSSI



Topic	Duration
👋 Welcome	5 min
📦 MDAnalysis mission & ecosystem	15 min
🖼️ Streaming: big picture	15 min
👀 Streaming: first look	10 min
❓ Q&A: Streaming overview	5 min
📦Streaming: MD packages, IMDClient	15 min
👀 Demo: Multiple analyses on NAMD simulation stream	10 min
💤 Break	10 min
🎯Activity: Write your own stream analysis	40 min
📦 Streaming: MDAnalysis functionality	10 min
❓Q&A: Streaming with MDAnalysis	5 min
👀 Application: Velocity correlation functions and 2PT	10 min
👀 Application: Ion channel permeation	10 min
❓ Q&A: Applications	5 min
🔮 Future direction	5 min
📖 Open Forum	20 min
🚪 Closing	5 min

NSF ASU NUMFOCUS [FISCALLY SPONSORED PROJECT]

Admin

- This meeting will be recorded

- When we do Q&A and interactive components
  - Raise your hand if you would like to talk
  - Wait until you are called on
  - Unmute yourself

- This event is administered under the MDAnalysis Code of Conduct
  - https://www.mdanalysis.org/pages/conduct

- If you are having trouble with an interactive component, ask in chat

- Workshop material will be made available under MDA org
  - https://github.com/mdanalysis/imd-workshop-2024

The MDAnalysis Project: An overview

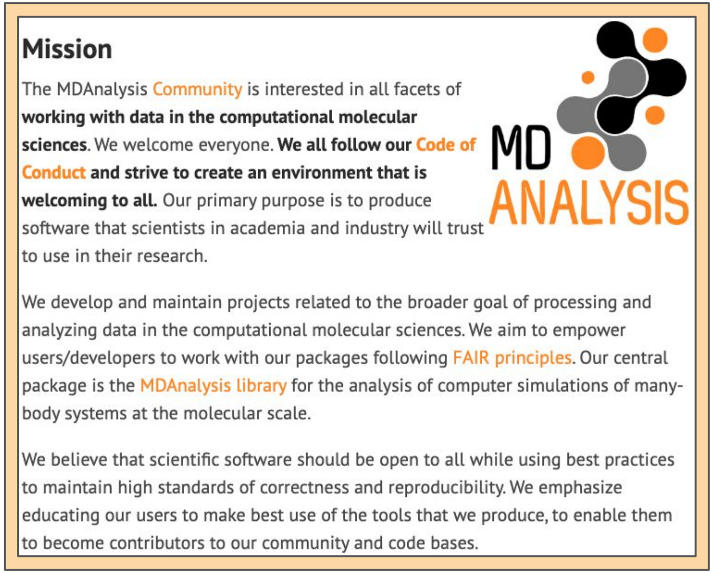
Hugo MacDermott-Opeskin, Oliver Beckstein, Jenna Swarthout Goddard

Originally prepared by, Oliver Beckstein, Jenna Swarthout Goddard for *CZI Open Science Meeting 2024*

MDANALYSIS

NUMFOCUS
[FISCALLY SPONSORED PROJECT]

The **MDAnalysis** Organization



Mission

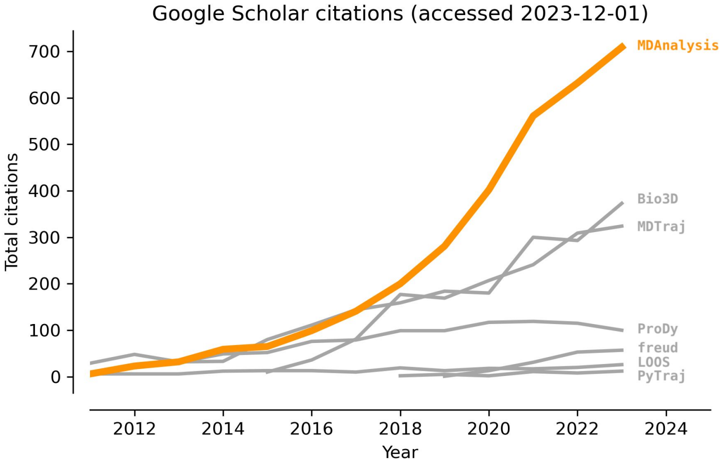
The MDAnalysis Community is interested in all facets of **working with data in the computational molecular sciences**. We welcome everyone. **We all follow our Code of Conduct and strive to create an environment that is welcoming to all.** Our primary purpose is to produce software that scientists in academia and industry will trust to use in their research.

We develop and maintain projects related to the broader goal of processing and analyzing data in the computational molecular sciences. We aim to empower users/developers to work with our packages following FAIR principles. Our central package is the MDAnalysis library for the analysis of computer simulations of many-body systems at the molecular scale.

We believe that scientific software should be open to all while using best practices to maintain high standards of correctness and reproducibility. We emphasize educating our users to make best use of the tools that we produce, to enable them to become contributors to our community and code bases.

- A Python library for analysis of simulation data - built *by* scientists, *for* scientists
- Community-led development
  - Majority non-funded work
  - More recently, contributors and work has been funded by grants (e.g., CZI EOSS program) and mentorship programs (e.g., GSoC, Outreachy)
    - More information in [About MDAnalysis · Funding]
- Mature, well-established library
  - Opportunity to focus on user experience

The **MDAnalysis** Organization



- Brief history of the project
  - Started by Naveen Michaud Agrawal ~2006 with Elizabeth Denning and Oliver Beckstein
  - Open source project since Jan 2008 on Google Code
  - NumFOCUS sponsored project since Jan 2020
  - Many users became developers and core developers (i.e., project leadership) over the years
- Today:
  - ≥ 205 contributors
  - Most used (thousands of users) Python library for analysis of molecular dynamics simulations (>30k downloads/month for new releases, >3700 citations of two MDAnalysis papers (2011, 2016))

MDAnalysis Project: www.mdanalysis.org

MDAnalysis library (`MDAnalysis`)

GitHub.org/MDAnalysis/mdanalysis

open source

community

forums (discord, GH discussions), X, bluesky, LinkedIn

other software (data processing in molecular sciences)

GitHub.org/MDAnalysis

education & outreach

workshops

Google Summer of Code

outreachy

Station1

eco-system

interoperability (RDKit, blender, …)

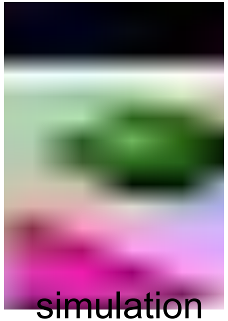
MDAKits

mdakits.mdanalysis.org

NUMFOCUS

OPEN CODE = BETTER SCIENCE

MDAnalysis library: heterogeneous data → uniform API

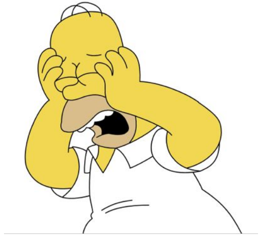


simulation trajectory

dcd, xtc, trr, ncdf, trj, pdb, pqr, gro, crd, dms, trz, mol2, xyz, config, history, gms, …

psf, tpr, prmtop, dms, mol2, hoomd xml, …

MD ANALYSIS



Oh nooooo!

NumPy

“accessible” structured data

`numpy.ndarray()`

python™

analysis algorithm

MD ANALYSIS

processed data

tables

images

graphs

Insights & publication!

doi:10.1016/j.jmb.2009.09.009 J. Mol. Biol. (2009) 394, 160–176

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open↔Closed Transitions

Oliver Beckstein[1,2*†], Elizabeth J. Denning[3†], Juan R. Perilla[3] and Thomas B. Woolf[2,3*]

[1]*Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK*

[2]*Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA*

[3]*Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA*

Received 17 June 2009; received in revised form 30 August 2009; accepted 4 September 2009
Available online 12 September 2009

Edited by M. Levitt

Adenylate kinase (AdK), a phosphotransferase enzyme, plays an important role in cellular energy homeostasis. It undergoes a large conformational change between an open and a closed state, even in the absence of substrate. We investigate the apo-AdK transition at the atomic level both with free-energy calculations and with our new dynamic importance sampling (DIMS) molecular dynamics method. DIMS is shown to sample biologically relevant conformations as verified by comparing an ensemble of hundreds of DIMS transitions to AdK crystal structure intermediates. The simulations reveal in atomic detail how hinge regions partially and intermittently unfold during the transition. Conserved salt bridges are seen to have important structural and dynamic roles; in particular, four ionic bonds that open in a sequential, zipper-like fashion and, thus, dominate the free-energy landscape of the transition are identified. Transitions between the closed and open conformations only have to overcome moderate free-energy barriers. Unexpectedly, the closed state and the open state encompass broad free-energy basins that contain conformations differing in domain hinge motions by up to 40°. The significance of these extended states is discussed in relation to recent experimental Förster resonance energy transfer measurements. Taken together, these results demonstrate how a small number of cooperative key interactions can shape the overall dynamics of an enzyme and suggest an "all-or-nothing" mechanism for the opening and closing of AdK. Our efficient DIMS molecular dynamics computer simulation approach can provide a detailed picture of a functionally important macromolecular transition and thus help to interpret and suggest experiments to probe the conformational landscape of dynamic proteins such as AdK.

© 2009 Elsevier Ltd. All rights reserved.

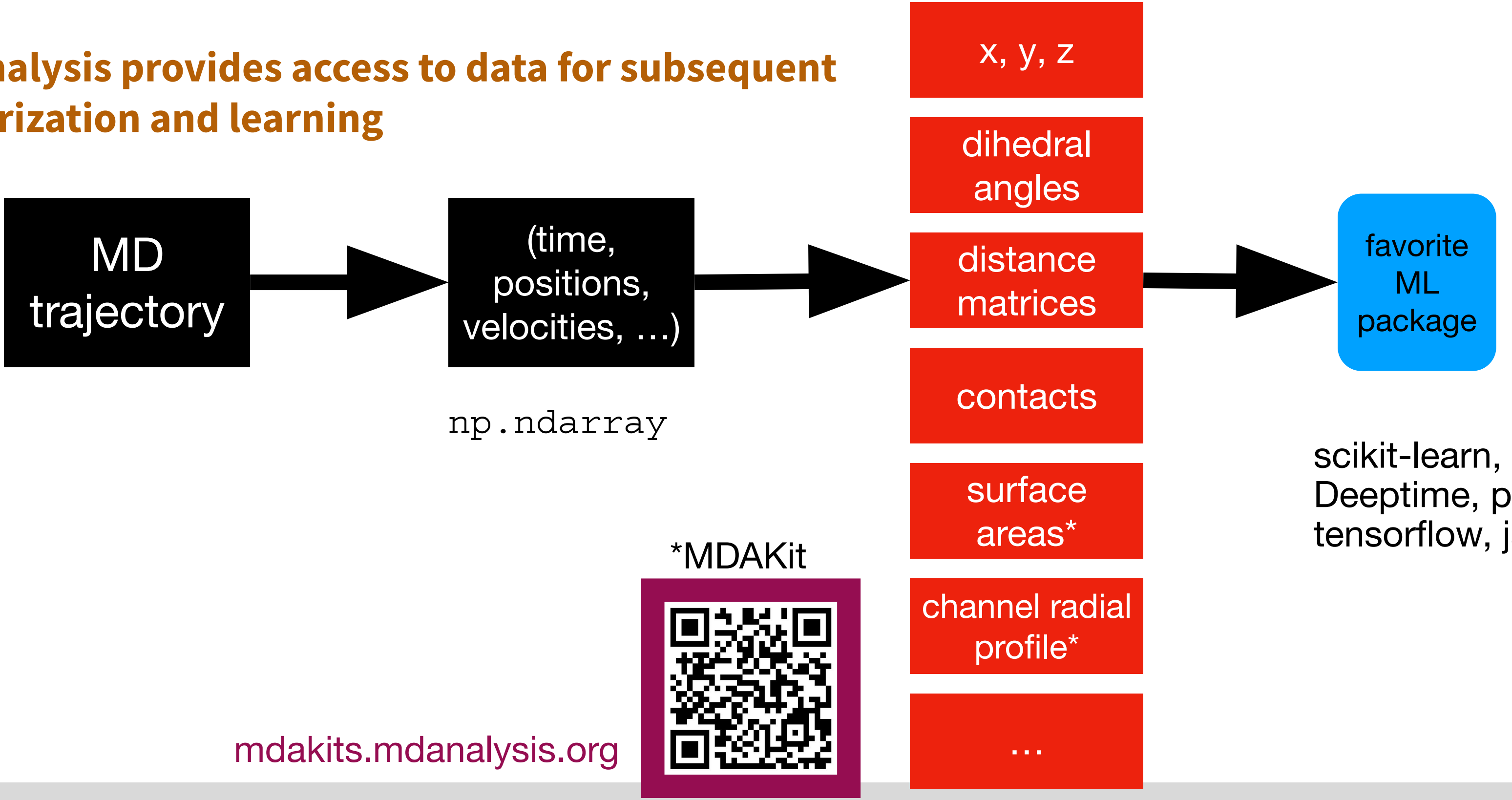
Keywords: adenylate kinase; importance sampling; molecular dynamics; transitions; free energy

> 40 file formats

1 analysis script

Example: MDAnalysis for ML

MDAnalysis provides access to data for subsequent featurization and learning



MD trajectory

(time, positions, velocities, …)

np.ndarray

x, y, z

dihedral angles

distance matrices

contacts

surface areas*

channel radial profile*

…

favorite ML package

scikit-learn, Deeptime, pytorch, tensorflow, jax

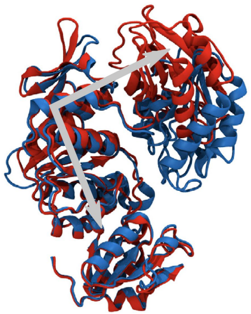
*MDAKit

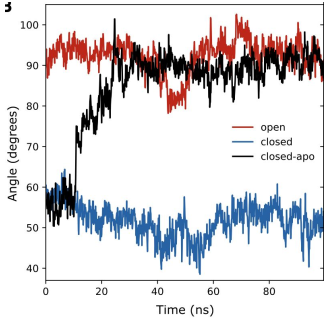
mdakits.mdanalysis.org

Example: MDAnalysis for ML: *MurD*

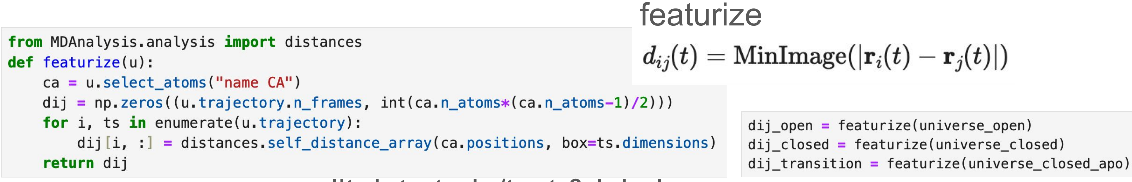
PCA and Random Forest classification of conformational states of MurD

MurD: closed → open transition





featurize

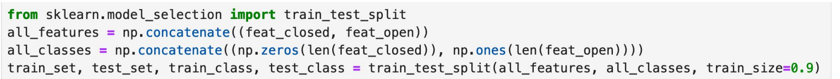


```
from MDAnalysis.analysis import distances
def featurize(u):
    ca = u.select_atoms("name CA")
    dij = np.zeros((u.trajectory.n_frames, int(ca.n_atoms*(ca.n_atoms-1)/2)))
    for i, ts in enumerate(u.trajectory):
        dij[i, :] = distances.self_distance_array(ca.positions, box=ts.dimensions)
    return dij
```

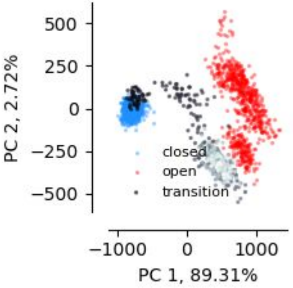
$$d_{ij}(t) = \mathrm{MinImage}(|\mathbf{r}_i(t) - \mathbf{r}_j(t)|)$$

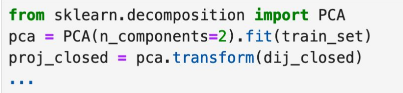
```
dij_open = featurize(universe_open)
dij_closed = featurize(universe_closed)
dij_transition = featurize(universe_closed_apo)
```

split data train/test & label



```
from sklearn.model_selection import train_test_split
all_features = np.concatenate((feat_closed, feat_open))
all_classes = np.concatenate((np.zeros(len(feat_closed)), np.ones(len(feat_open))))
train_set, test_set, train_class, test_class = train_test_split(all_features, all_classes, train_size=0.9)
```

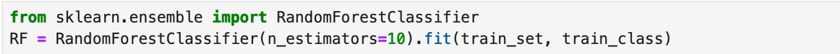




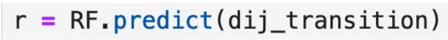
```
from sklearn.decomposition import PCA
pca = PCA(n_components=2).fit(train_set)
proj_closed = pca.transform(dij_closed)
...
```

principal component analysis (PCA)

Train RandomForest classifier

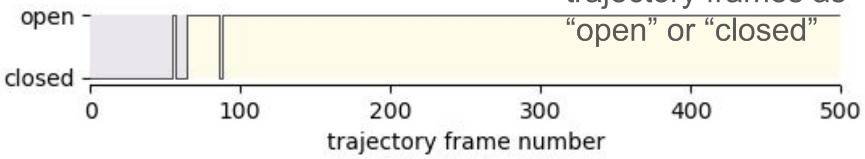


```
from sklearn.ensemble import RandomForestClassifier
RF = RandomForestClassifier(n_estimators=10).fit(train_set, train_class)
```



```
r = RF.predict(dij_transition)
```

Classify unknown trajectory frames as "open" or "closed"



Coupling Molecular Dynamics and Deep Learning to Mine Protein Conformational Space

Matteo T. Degiacomi[1],[2],*
[1]Department of Chemistry, Durham University, South Road, Durham DH1 3LE, UK
[2]Lead Contact
*Correspondence: matteo.t.degiacomi@durham.ac.uk

https://doi.org/10.1016/j.str.2019.03.018

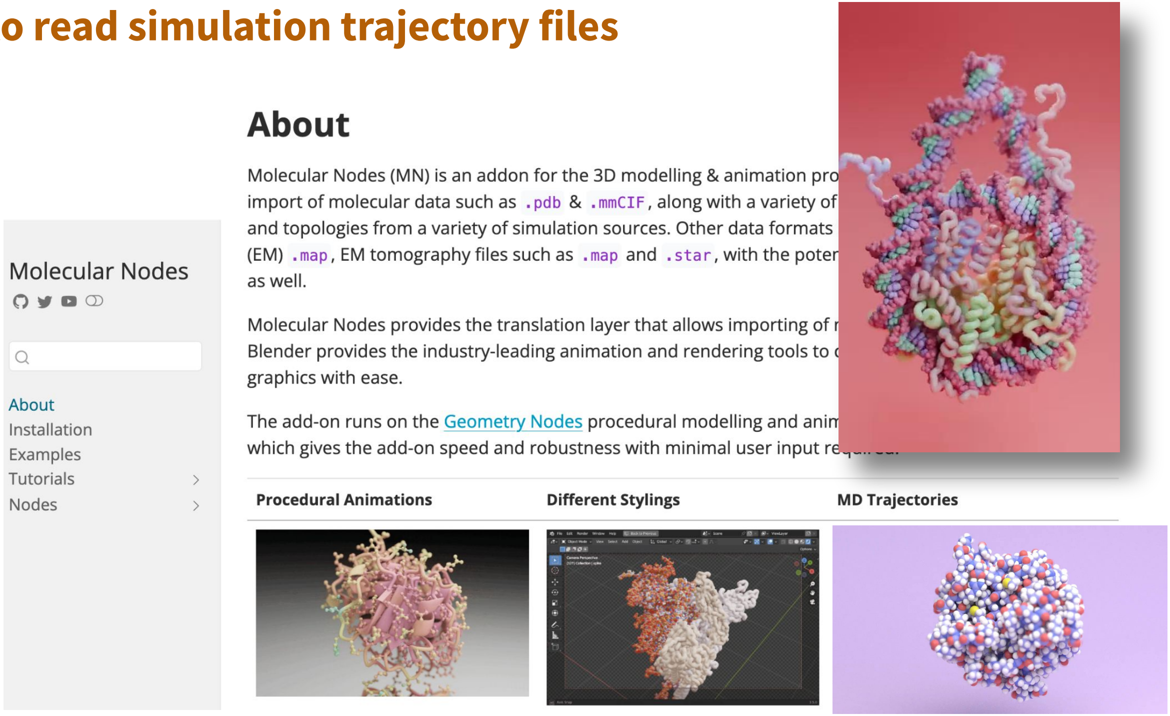
Structure *27*, 1034–1040, June 4, 2019

https://github.com/MDAnalysis/WorkshopMDMLEdinburgh2022/

Example: Render simulations with Blender/MolecularNodes

MolecularNodes uses MDAnalysis to read simulation trajectory files

- MolecularNodes renders molecular data in *blender*
- MDAnalysis enables reading of dynamic structural information: movies!
- Author: **Brady Johnston** `<brady.johnston@me.com>`



Molecular Nodes

About
Installation
Examples
Tutorials
Nodes

About

Molecular Nodes (MN) is an addon for the 3D modelling & animation pro
import of molecular data such as `.pdb` & `.mmCIF`, along with a variety of
and topologies from a variety of simulation sources. Other data formats
(EM) `.map`, EM tomography files such as `.map` and `.star`, with the poter
as well.

Molecular Nodes provides the translation layer that allows importing of r
Blender provides the industry-leading animation and rendering tools to c
graphics with ease.

The add-on runs on the Geometry Nodes procedural modelling and anim
which gives the add-on speed and robustness with minimal user input re

Procedural Animations | **Different Stylings** | **MD Trajectories**

https://bradyajohnston.github.io/MolecularNodes/
https://www.blender.org/

… and soon directly inside a notebook!

Roadmap: Towards **MDAnalysis 3.0** and Beyond

Extending “reach” and “detail”

1. **Streaming support** (cloud storage, running simulations): MDAnalysis as a universal data adapter for **machine learning workflows** on simulation data.
2. **Scientifically aware operations** (automatically infer chemistry using improved “guessers” and enable full cheminformatics with RDKit)

Improving user experience

3. **Improve performance**
    a. parallelization (multicore and high-performance computing via dask)
    b. redesigned precision model (control precision everywhere)
    c. replacement of Python code with cython/C++ code
4. **Improve interoperability** (support new file formats and “converters”=API-based plugins)

https://www.mdanalysis.org/2023/10/25/towards_3.0/

Contacts and Resources

GitHub

github.com/MDAnalysis

User Guide

userguide.mdanalysis.org

Join the conversation at:

https://mdanalysis.org

https://github.com/MDAnalysis/mdanalysis/discussions

MDAnalysis

@mdanalysis

@mdanalysis.bsky.social

mdanalysis

discord.gg/t3bx25zd7v

@mdanalysis3040

Acknowledgements

205 code contributors and countless community members

Naveen Michaud-Agrawal, Elizabeth J. Denning, **Oliver Beckstein**, Danny Parton, Philip Fowler, **Tyler Reddy**, Joseph Goose, **Jan Domanski**, Benjamin Hall, Paul Rigor, David Caplan, Christian Beckstein (logo), **Sébastien Buchoux**, Joshua L. Adelman, Lukas Grossar, Andy Somogyi, Lukas Stelzl, Jinju Lu, Joshua L. Phillips, Zhuyi Xue, Xavier Deupi, **Manuel Nuno Melo**, Robert McGibbon, **Richard J. Gowers**, Alejandro Bernardin, Lennard van der Feltz, Matthieu Chavent, Joe Jordan, Alex Nesterenko, Caio S. Souza, Sean L. Seyler, **David L. Dotson**, Carlos Yanez S., Kyle J. Huston, Isaac Virshup, **Max Linke**, Gorman Stock, **Jonathan Barnoud**, Hai Nguyen, Balasubramanian, Mattia F. Palermo, Utkarsh Saxena, Abhinav Gupta, John Detlefs, Eugen Hruska, Bart Bruininks, **Fiona B. Naughton**, Robert Delgado, Wouter Boomsma, **Matteo Tiberti**, Tone Bengtsen, Shantanu Srivastava, Pedro Reis, Ruggero Cortini, Zhiyi Wu, Kashish Punjani, Utkarsh Bansal, Shobhit Agarwal, Vedant Rathore, Akshay Gupta, Juan Eiros Zamora, Jon Kapla, Sang Young Noh, Andrew William King, Kathleen Clark, Dominik 'Rathann' Mierzejewski, Nestor Wendt, **Micaela Matta**, Jose Borreguero, Sören von Bülow, Nabarun Pal, Mateusz Bieniek, Paul Smith, Navya Khare, **Johannes Zeman**, Ayush Suhane, Davide Cruz, Shujie Fan, Andrew R. McCluskey, Henry Mull, **Irfan Alibay**, **Philip Loche**, Matthew W. Thompson, Ali Ehlen, Daniele Padula, Ninad Bhat, Fenil Suchak, Yibo Zhang, Luís Pedro Borges Araújo, Abhishek A. Kognole, **Rocco Meli**, **Lily Wang**, Matthijs Tadema, Joao Miguel Correia Teixeira, Charlie Cook, Yuanyu Chang, Guillaume Fraux, Ivan Hristov, Michael Quevillon, Hao Tian, **Hugo MacDermott-Opeskin**, Anshul Angaria, Shubham Sharma, **Yuxuan Zhuang**, Cédric Bouysset, Abhishek Shandilya, Morgan L. Nance, Faraaz Shah, Wiep van der Toorn, Siddharth Jain, Ameya Harmalkar, Shakul Pathak, Andrea Rizzi, William Glass, Marcello Sega, Edis Jakupovic, Nicholas Craven, Mieczyslaw Torchala, Ramon Crehuet, Haochuan Chen, Karthikeyan Singaravelan, Ian Aditya Kamath, Leonardo Barneschi, Henrik Jäger, Jan Stevens, Orion Cohen, Dimitrios Papageorgiou, Hannah Pollak, Estefania Barreto-Ojeda, Paarth Thadani, Henry Kobin, Kosuke Kudo, Sulay Shah, Alexander Yang, Filip T. Szczypiński, Marcelo C. R. Melo, Mark D. Driver, Kevin Boyd, Atharva Kulkarni, Yantong Cai, Bjarne Feddersen, Pratik Gupta, Alexander Gorfer, Aya M. Alaa, Kazi Shudipto Amin, Alia Lescoulie, Henok Ademtew, Uma D Kadam, Tamandeep Singh, Mingyi Xue, Meghan Osato, Anirvinya G, Rishabh Shukla, Manish Kumar, Aditi Tripathi, Sukeerti T, Kavya Bisht, Mark Verma, Marcelo D. Poleto, Ricky Sexton, Rafael R. Pappalardo, Tengyu Xie, Raymond Zhao, Haleema Khan, Jennifer A Clark, Jake Fennick, Utsav Khatu, Patricio Barletta, Mikhail Glagolev, Christian Pfaendner, Pratham Chauhan, Meet Brijwani, Vishal Parmar, Moritz Schaeffler, Xu Hong Chen, Domenico Marson, Ahmed Salah Ghoneim, Alexander Schlaich, Josh Vermaas, Xiaoxu Ruan, Egor Marin, Shaivi Malik, Daniel J. Evans, Mohit Kumar, Shubham Kumar, Zaheer Timol, Geongi Moon, Sumit Gupta, Heet Vekariya, Lawson Woods, Johannes Stöckelmaier, **Jenna M. Swarthout Goddard**, Aditya Keshari, Philipp Stärk, Kai Niklas Spauszus, Sampurna Mukherjee, Leon Wehrhan, Valerij Talagayev

Closing remarks

- Thank you all so much for coming!
- This is work in progress we welcome your feedback
- Get in touch on the MDAnalysis Discord, Github or other socials
- Or reach out to one of the PIs

https://github.com/mdanalysis/imd-workshop-2024

(post workshop section)

Join the conversation at:

https://mdanalysis.org

https://github.com/MDAnalysis/mdanalysis/discussions

MDAnalysis

@mdanalysis

@mdanalysis.bsky.social

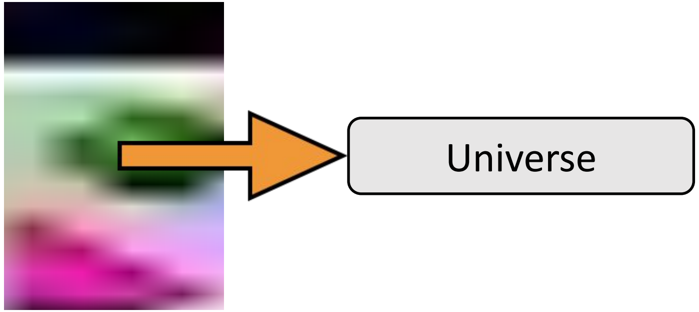
mdanalysis

discord.gg/t3bx25zd7v

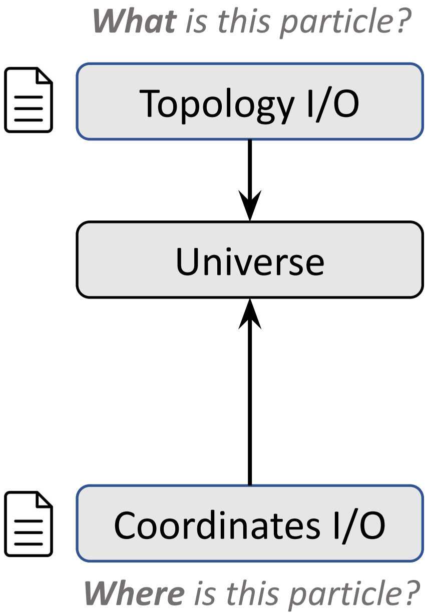
@mdanalysis3040

Archive

A Deep Dive into the **MDAnalysis** Library



A Deep Dive into the **MDAnalysis** Library



***What** is this particle?*

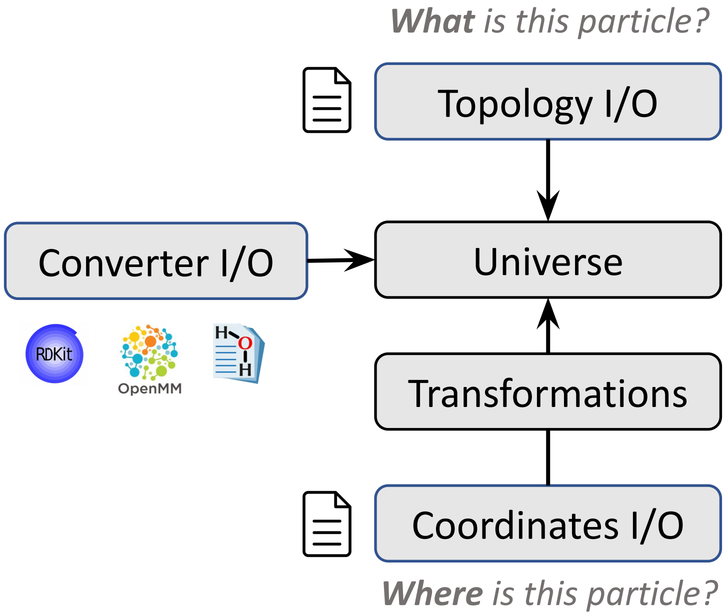
Topology I/O

Universe

Coordinates I/O

***Where** is this particle?*

A Deep Dive into the **MDAnalysis** Library



***What** is this particle?*

Topology I/O

Converter I/O

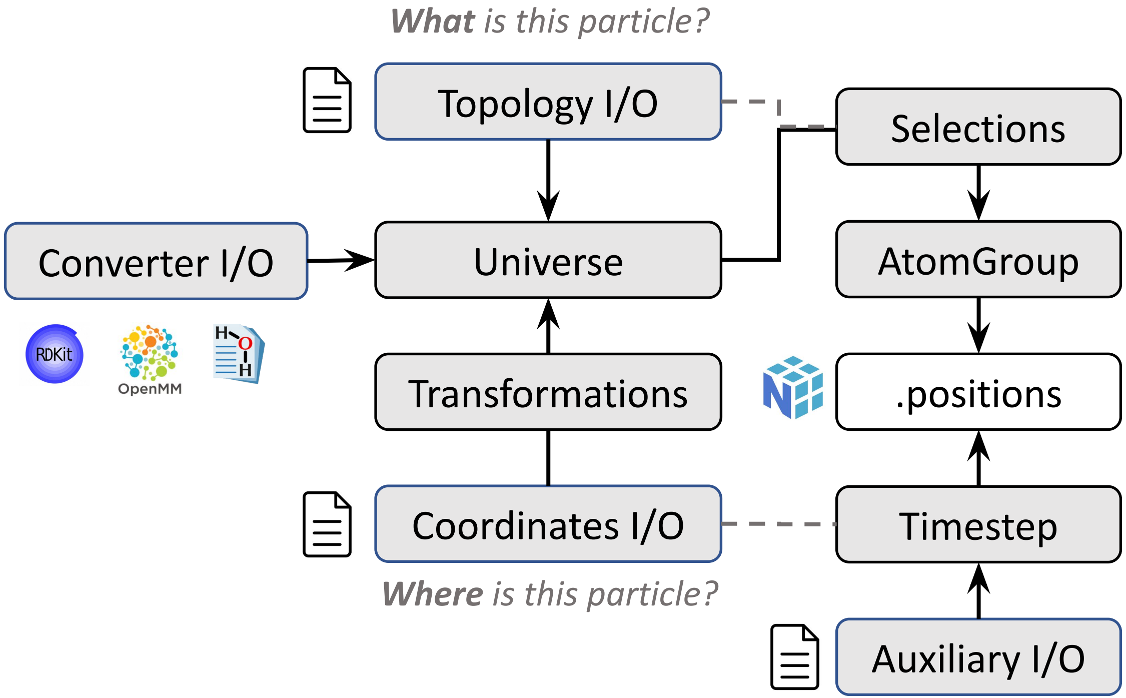
Universe

Transformations

Coordinates I/O

***Where** is this particle?*

A Deep Dive into the **MDAnalysis** Library



What *is this particle?*

Topology I/O

Selections

Converter I/O

Universe

AtomGroup

Transformations

.positions

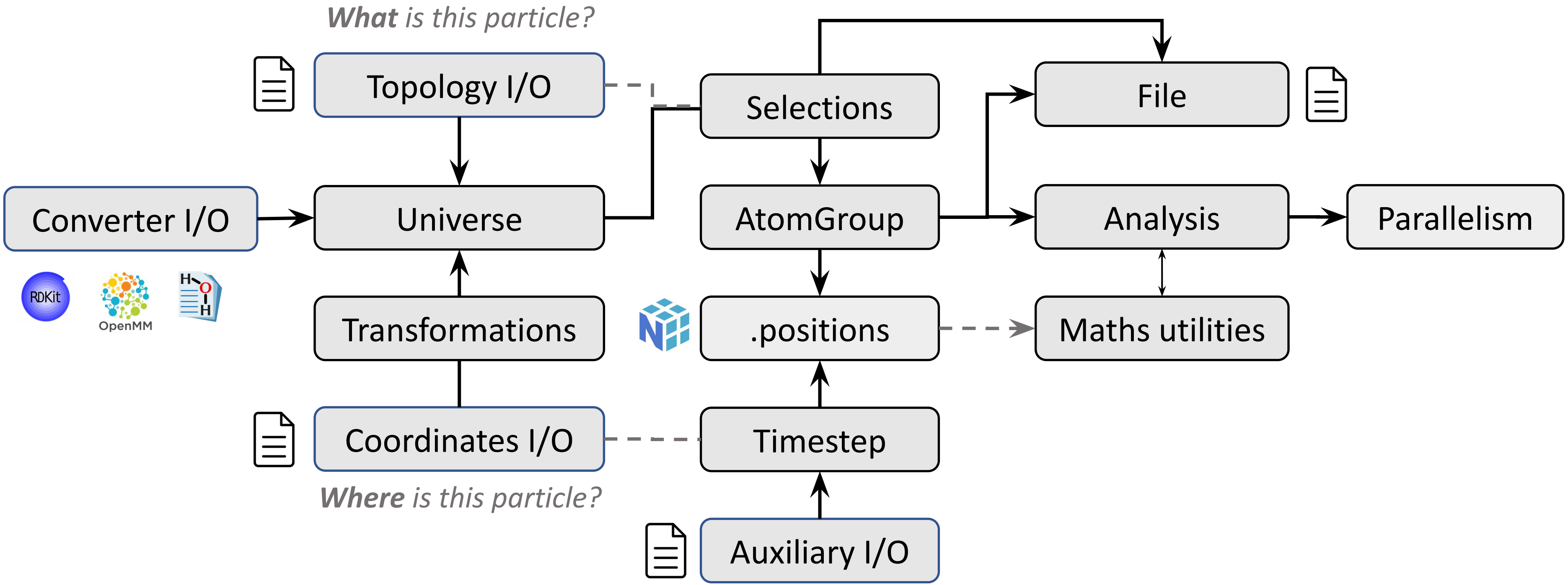
Coordinates I/O

Timestep

Where *is this particle?*

Auxiliary I/O

A Deep Dive into the **MDAnalysis** Library


What *is this particle?*
Topology I/O
Selections
File
Converter I/O
Universe
AtomGroup
Analysis
Parallelism
RDKit
OpenMM
Transformations
.positions
Maths utilities
Coordinates I/O
Timestep
Where *is this particle?*
Auxiliary I/O