MAX PLANCK INSTITUTE
FOR SUSTAINABLE MATERIALS

Maximizing High-Throughput Discovery and Machine Learning Efficiency Through Computational Workflows

Sarath Menon, Jörg Neugebauer

Acknowledgments: Liam Huber[1], Marvin Poul[2], Jan Janssen[2], Tilmann Hickel[2,3] and Ralf Drautz[4]

[1] Gray Haven Solutions, Victoria, British Columbia, Canada
[2] Max-Planck-Institute for Sustainable Materials, Düsseldorf
[3] Bundesanstalt für Materialforschung und -prüfung, Berlin
[4] ICAMS, Ruhr-Universität Bochum

14.05.2024 | ML4MS workshop

Hi! I am Sarath Menon

Doctorate in Mechanical Engineering,

Ruhr-Universität Bochum

Computational Materials Science,

Atomistic Modelling

Research Associate,

Max Planck Institute for Sustainable Materials

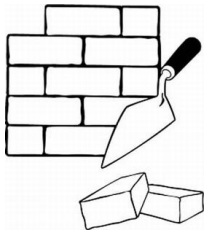
Workflows and Software Development,

@ NFDI-MatWerk

s.menon@mpie.de

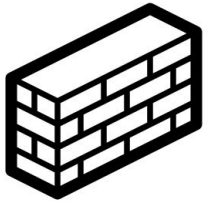
github.com/srmnitc

Validation of interatomic potentials



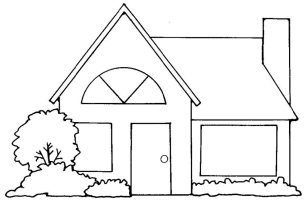
Level I

- Raw, unstructured data
- Energy, forces, stresses



Level II

- Basic material properties
- Computationally inexpensive
- Directly accessible by DFT
- Equilibrium energy/volume, bulk modulus, elastic constants

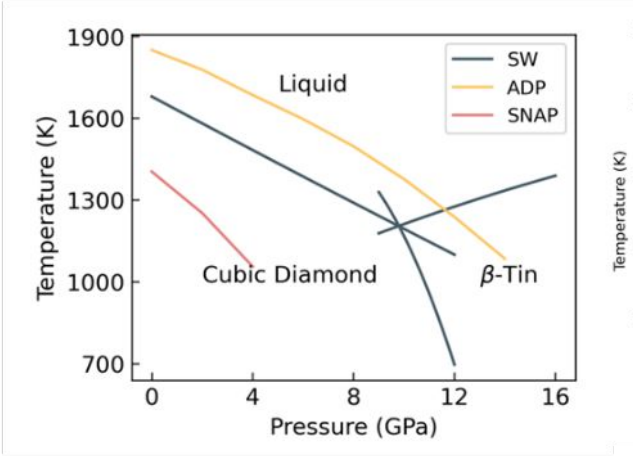


Level III

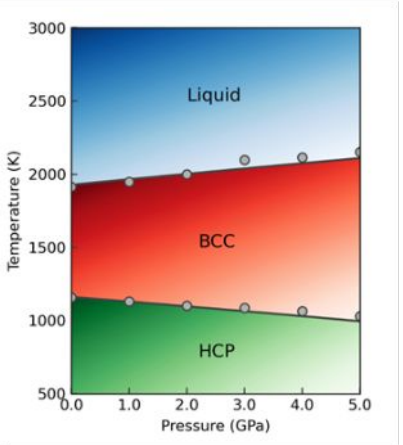
- Validation by application
- Expensive calculations
- specific heat, melting temperature, phase diagram

Tutorial 3

Calculated phase diagrams

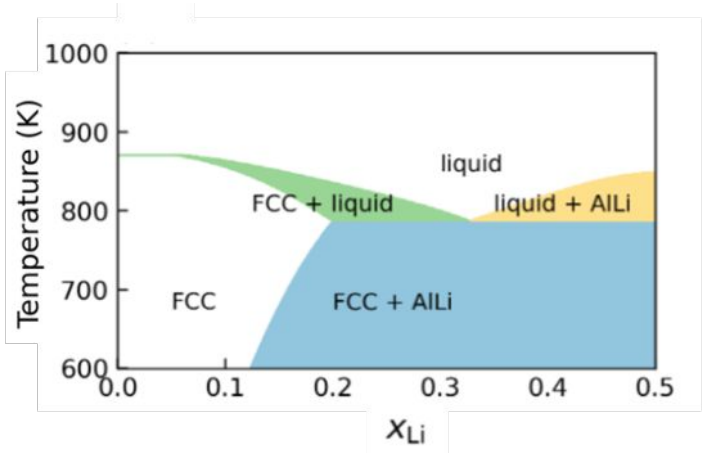


Silicon, SW



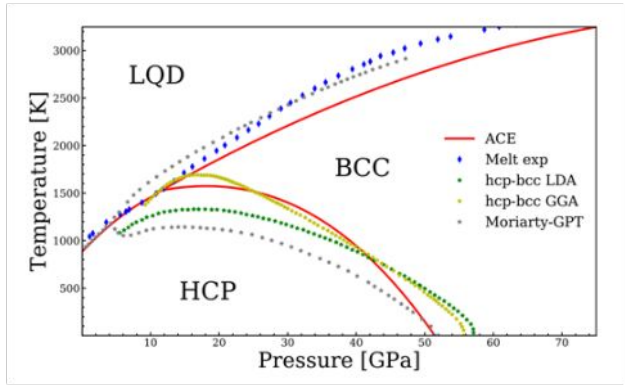
Titanium, EAM

Menon *et al.*, Phys. Rev. Mater. 5 (2021)



Aluminium-Lithium, NNP

Menon *et al.*, Arxiv 2403.05724. (2024)

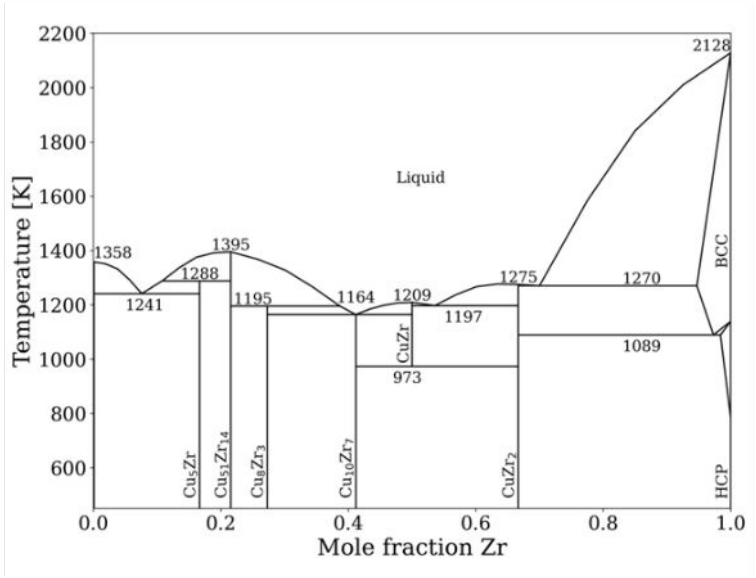


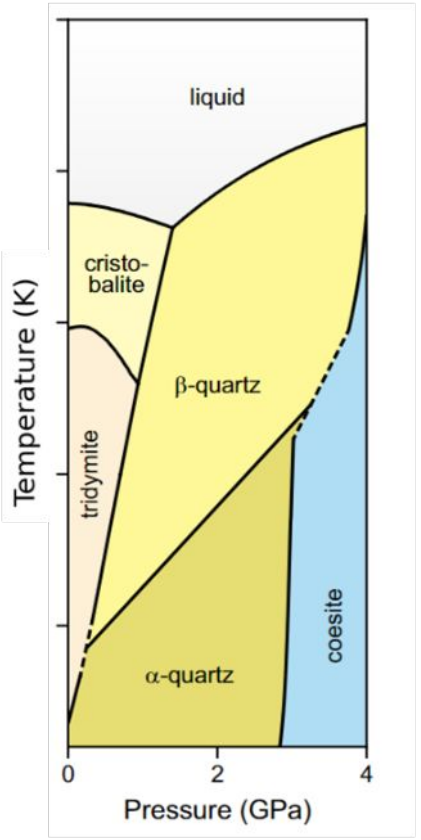
Magnesium, ACE

Ibrahim *et al.*, Arxiv. 2305.03577 (2023)

Copper-Zirconium, ACE

Leimeroth *et al.*, Acta. Mater. (2023)

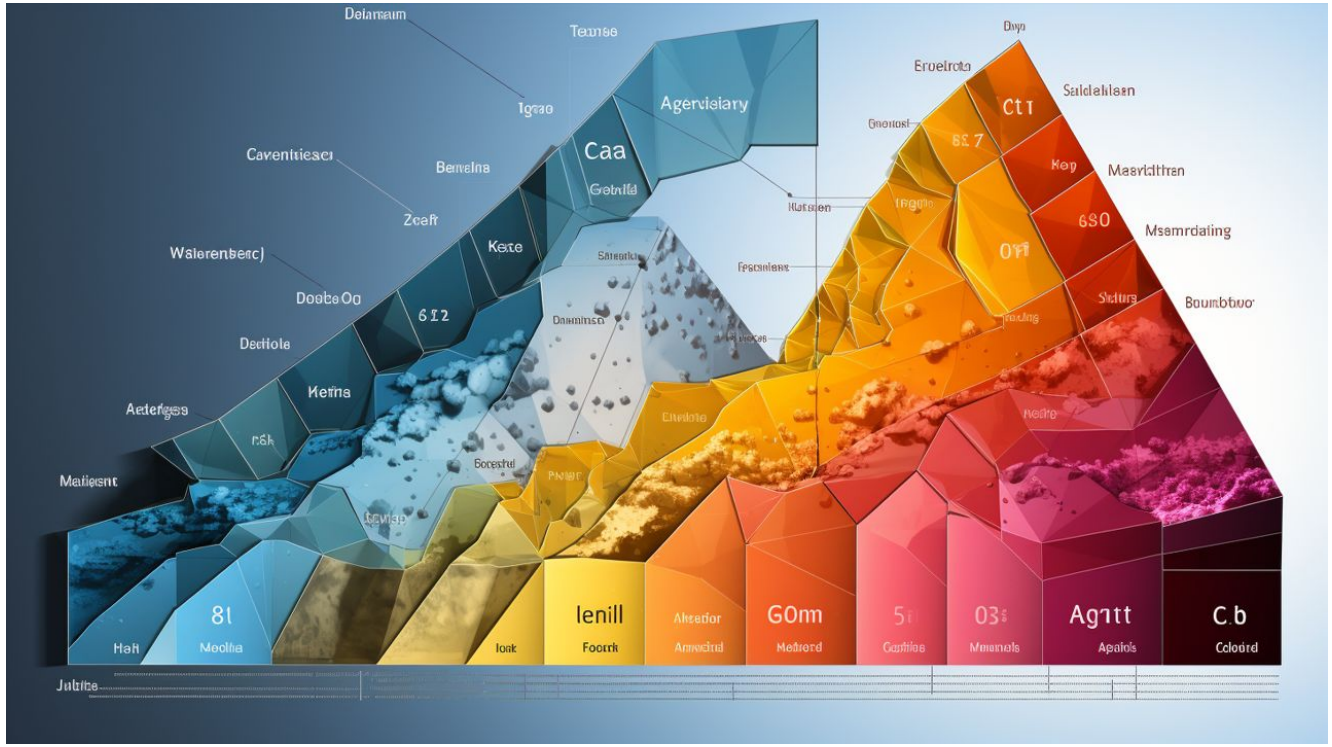




Silicon-Oxide, ACE

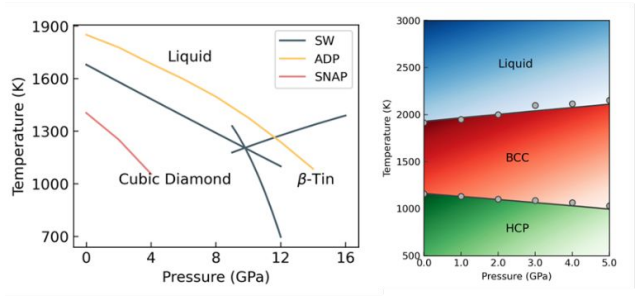
Erhard *et al.*, Nat. Commun. (2024)

Or just use AI



AI-Li phase diagram, as rendered by Midjourney, a generative AI program

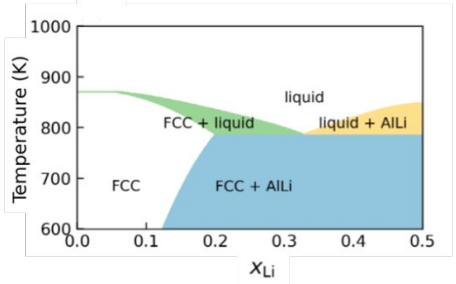
Calculated phase diagrams



Silicon, SW

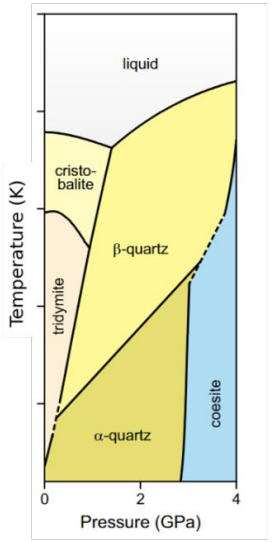
Titanium, EAM

Menon *et al.*, Phys. Rev. Mater. 5 (2021)



Aluminium-Lithium, NNP

Menon *et al.*, Arxiv 2403.05724. (2024)

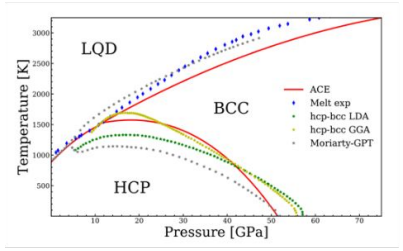


Silicon-Oxide, ACE

Erhard *et al.*, Nat. Commun. (2024)

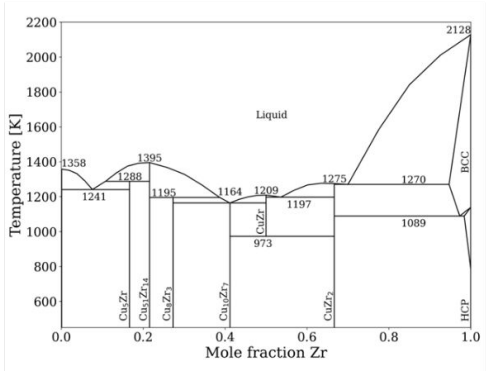
Copper-Zirconium, ACE

Leimeroth *et al.*, Acta. Mater. (2023)



Magnesium, ACE

Ibrahim *et al.*, Arxiv. 2305.03577 (2023)



A larger number of calculations are needed…
How do we ensure that they are reproducible?

Published might not mean reproducible

[HTML] A general-purpose machine learning framework for predicting properties of inorganic materials

L Ward, A Agrawal, A Choudhary, C Wolverton

npj Computational Materials, 2016 • nature.com

Abstract

A very active area of materials research is to devise methods that use machine learning to automatically extract predictive models from existing materials data. While prior examples have demonstrated successful models for some applications, many more applications exist where machine learning can make a strong impact. To enable faster development of machine-learning-based models for such applications, we have created a framework capable of being applied to a broad range of materials data. Our method works by using a

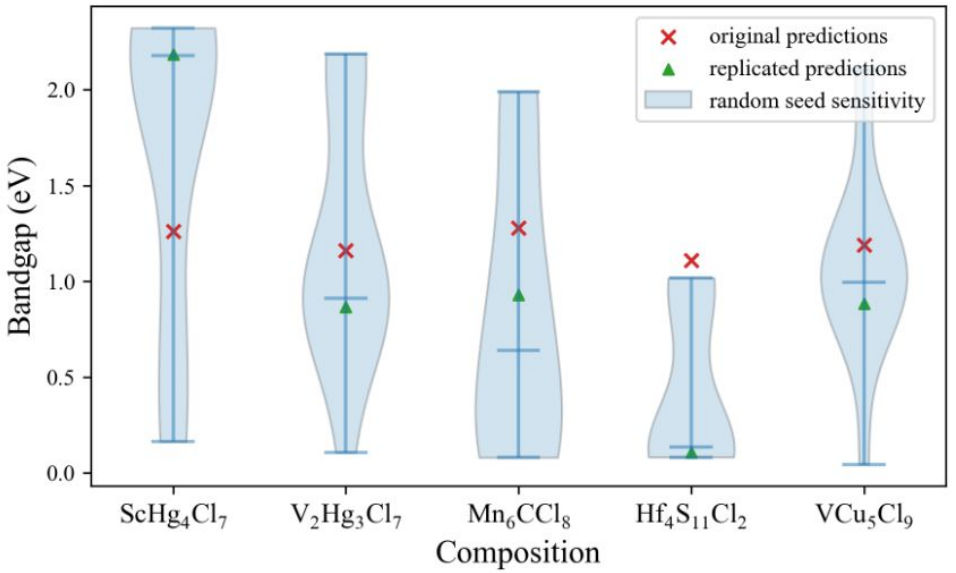
SHOW MORE

☆ Save 99 Cite Cited by 995 Related articles All 13 versions

Reproducibility in Computational Materials Science: Lessons from'A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials'

D Persaud, L Ward, J Hattrick-Simpers

arXiv preprint arXiv:2310.07044, 2023 • arxiv.org

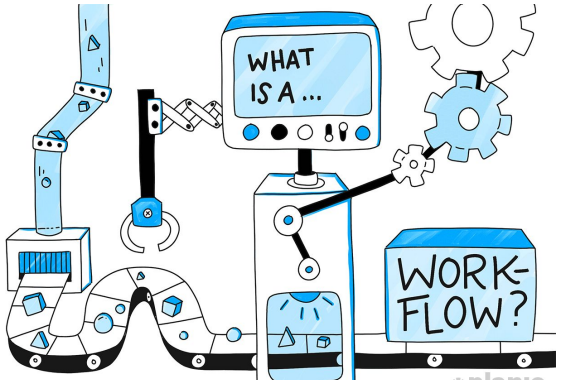


Major challenges:

- Sequential code organization: which steps in what order?
- Reporting computational dependencies and versioning

Towards FAIR workflows

- Complex multi-step methods
- Used for data collection, data preparation, analytics, predictive modelling, and simulation
- Computational workflows are enablers of automated data processing.



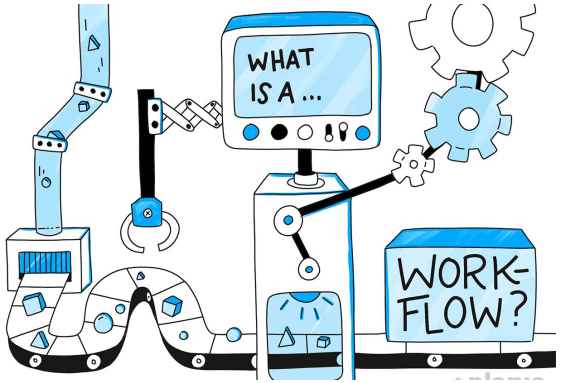
https://plan.io/blog/what-is-a-workflow/

FAIR workflows = FAIR data + FAIR software + Unique requirements

Goble *et al*; FAIR Computational Workflows. *Data Intelligence* 2020; 2 (1-2): 108–121. doi: https://doi.org/10.1162/dint_a_00033

Towards FAIR workflows

- Complex multi-step methods
- Used for data collection, data preparation, analytics, predictive modelling, and simulation
- Computational workflows are enablers of automated data processing.



https://plan.io/blog/what-is-a-workflow/

FAIR workflows = FAIR data + FAIR software + Unique requirements

FAI**R**
Reuse, Rerun, Replicate, Repurpose, Recycle, Reproduce

Goble *et al*; FAIR Computational Workflows. *Data Intelligence* 2020; 2 (1-2): 108–121. doi: https://doi.org/10.1162/dint_a_00033

Tutorial 4