Workshop on Machine Learning Modalities for Materials Science

13.-17. May, Ljubljana, Slovenia

MAX PLANCK INSTITUTE
FOR SUSTAINABLE MATERIALS GMBH

Hands-on tutorial:

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

Sarath Menon and Jörg Neugebauer

Acknowledgments

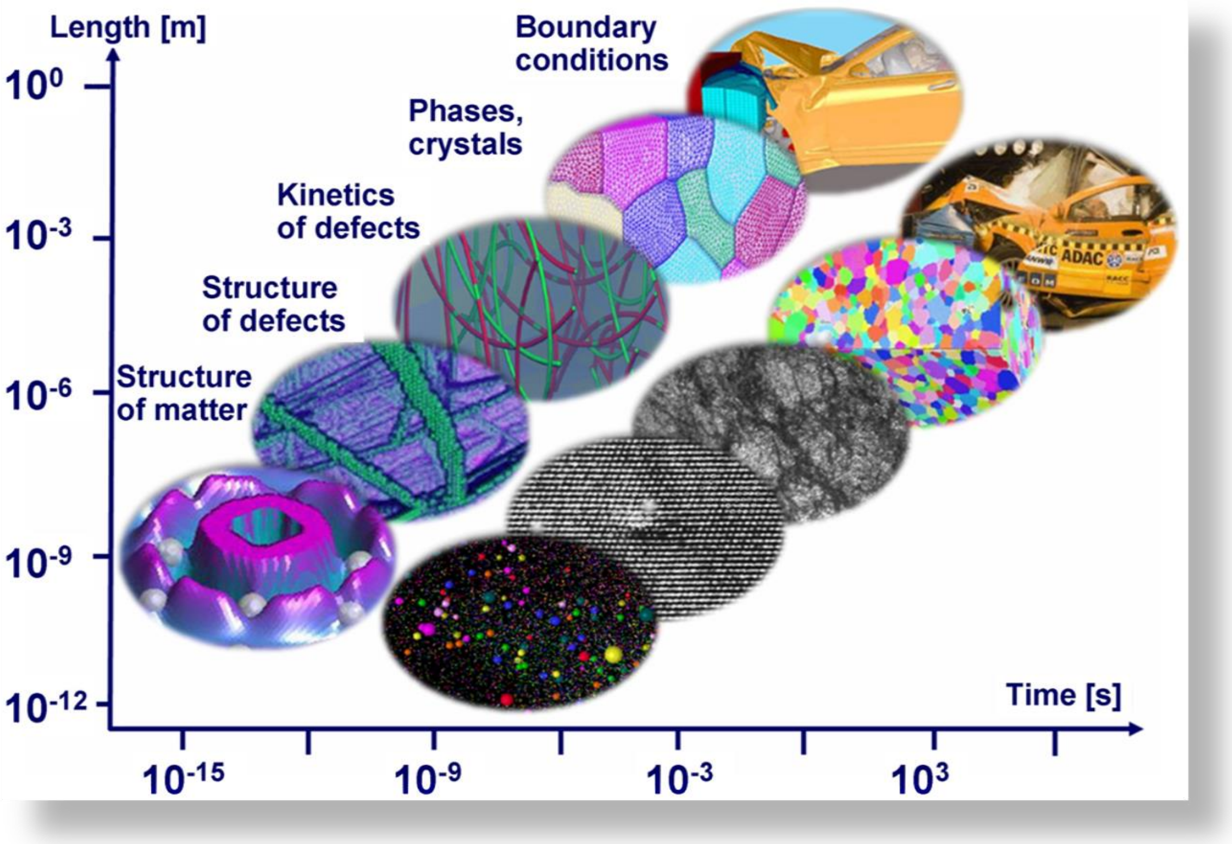
Marvin Poul[1], Jan Janssen[1], Minaam Qamar[3], Tilmann Hickel[1,2] and Ralf Drautz[3]

[1]Max-Planck-Institute for Sustainable Materials, Düsseldorf
[2]Bundesanstalt für Materialforschung und -prüfung, Berlin
[3]ICAMS, Ruhr-Universität Bochum

POTENTIALS

Multiscale challenge in materials modeling and experiment

Inherent multiscale character of materials



Multiscale & multiscale simulations often require complex simulation workflows and large-scale simulations

- **But: Tools have been independently developed and are not interoperable**

Key Challenge – Sampling Huge Configuration Spaces

$$Z(V,T,x,\alpha_{phase}) = \left\langle e^{-E^{QM}\left(\left\{\underbrace{\vec{R}_I}_{3N},\underbrace{Z_I}_{N},\underbrace{\sigma_I}_{3N},\underbrace{f_i}_{N},\ldots\right\}_V\right)/k_B T} \right\rangle_{V,T,x,\alpha_{phase}}$$

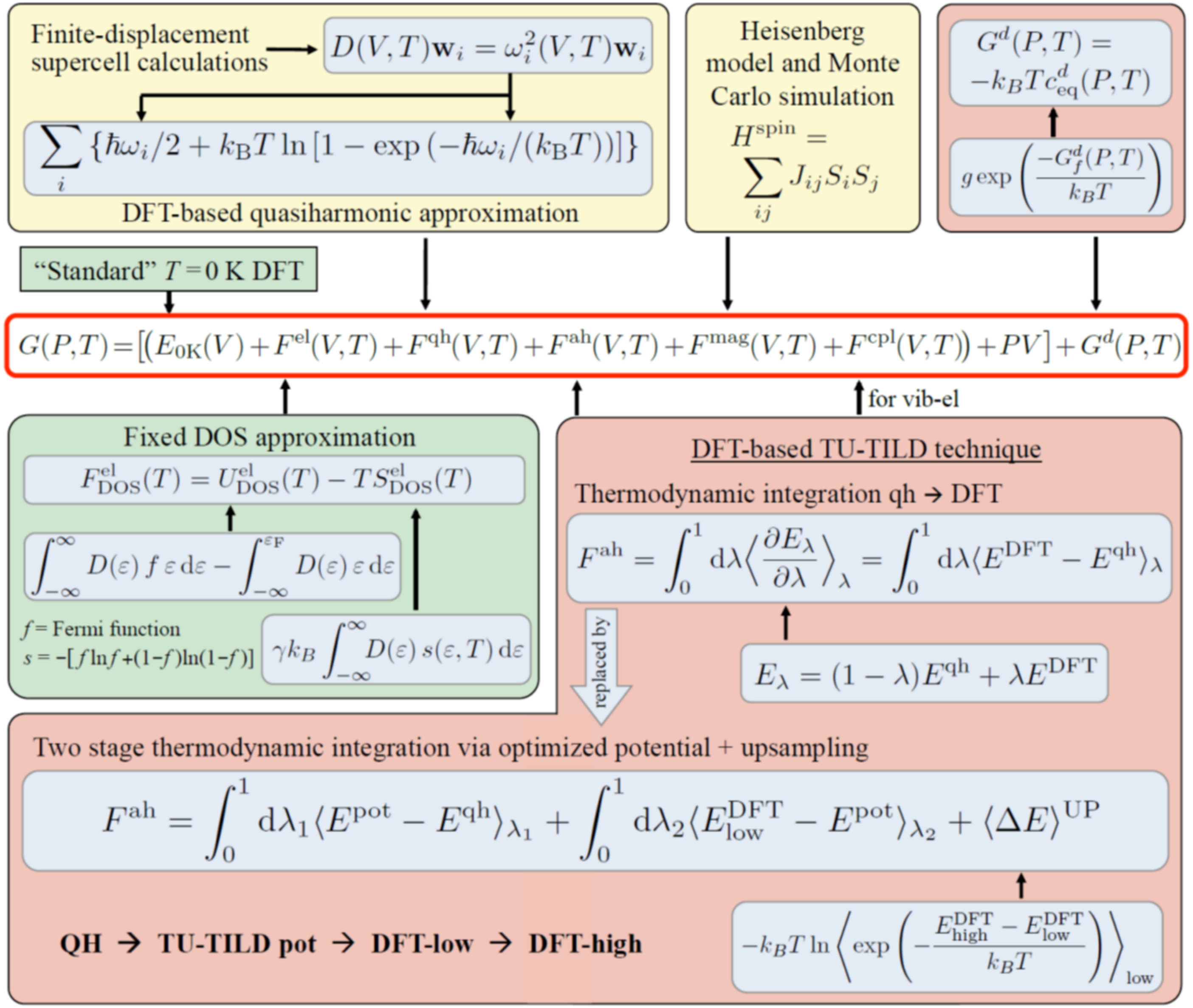
N ... Number of atoms

- Exploration of a huge (8N dimensional) configuration space!
  - In principle well-suited for exascale computing
  - But any brute force approach is unfeasible:
    - 10 data points on each coordinate
    - 100 atoms
    - 10^{800} configurations

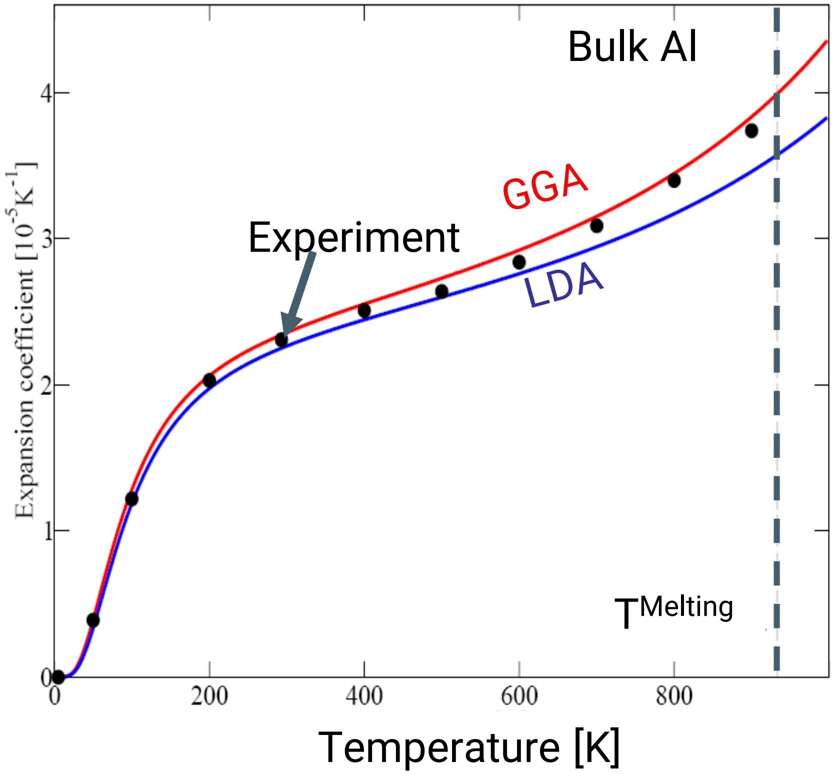
Dimensionality reduction is mandatory!
- Math + ML concepts

Practical examples of sampling/navigating in high-dimensional configuration spaces

Ab initio up to the melting temperature



Thermal expansion coefficient vs T



Workflows reduce computational effort from 10^7 down to 10^2 high-quality DFT calculations needed!

X. Zhang, T. Hickel, B. Grabowski, JN, Comp. Mat. Sci. (2018)

Ab initio up to the melting temperature

Harmonic Vibrations

phonopy

Spin MC

ALPS

Gibbs

"Standard" $T=0$ K DFT

$$G(P,T)=[(E_{0K}(V)+F^{el}(V,T)+F^{qh}(V,T)+F^{ah}(V,T)+F^{mag}(V,T)+F^{cpl}(V,T))+PV]+G^{d}(P,T)$$

for vib-el

Electronic DOS

Anharmonic Vibrations

phonopy

LAMMPS

High-Throughput Materials screening

scikit learn

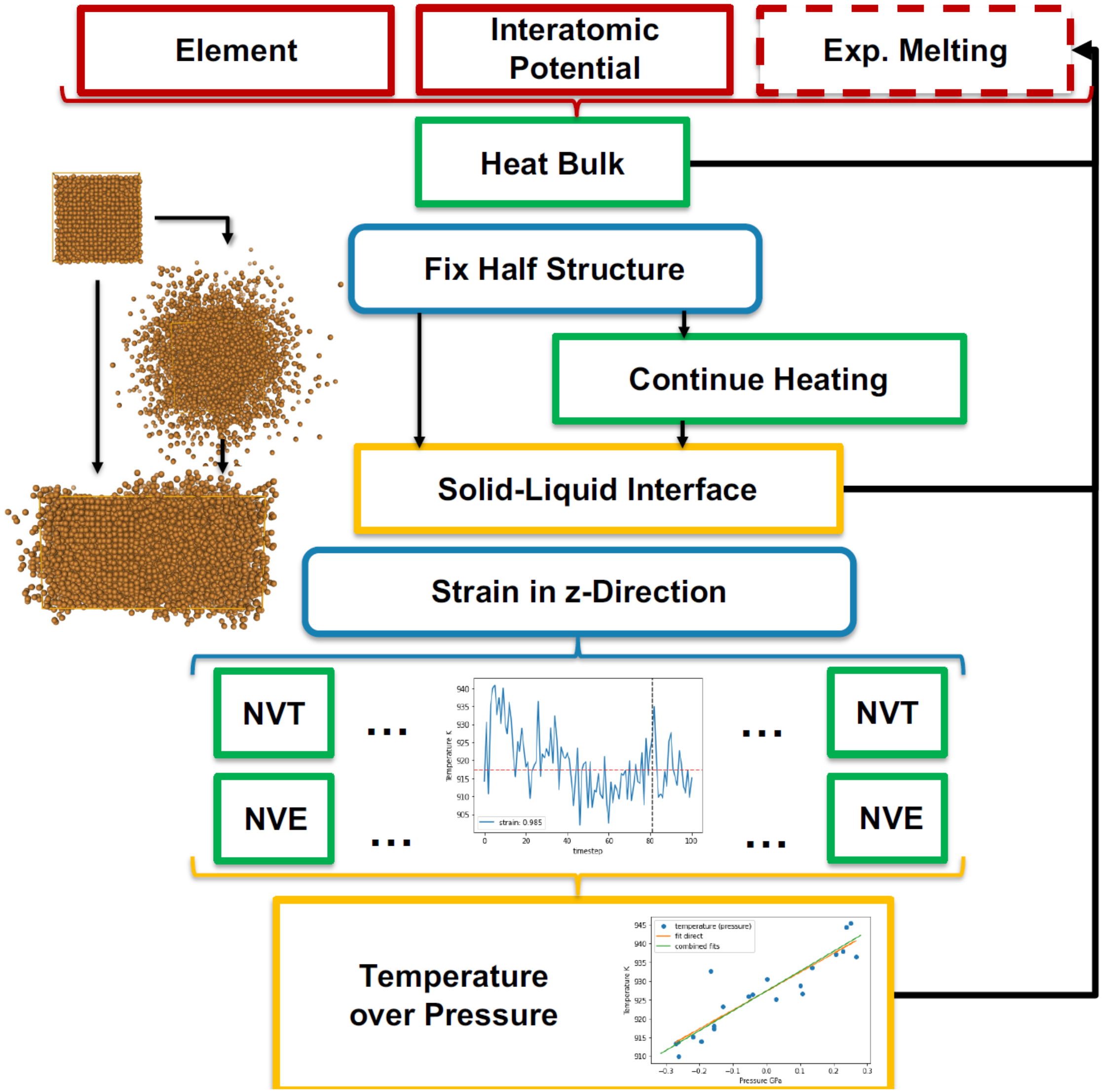
SciPy

NumPy

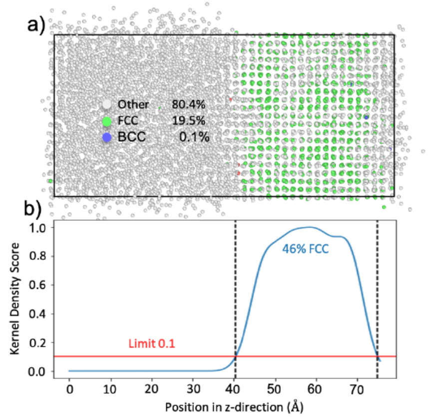
...

X. Zhang, T. Hickel, B. Grabowski, JN, Comp. Mat. Sci. (2018)

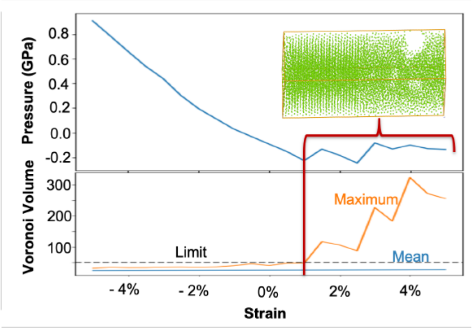
Fully autonomous algorithm to determine melting point



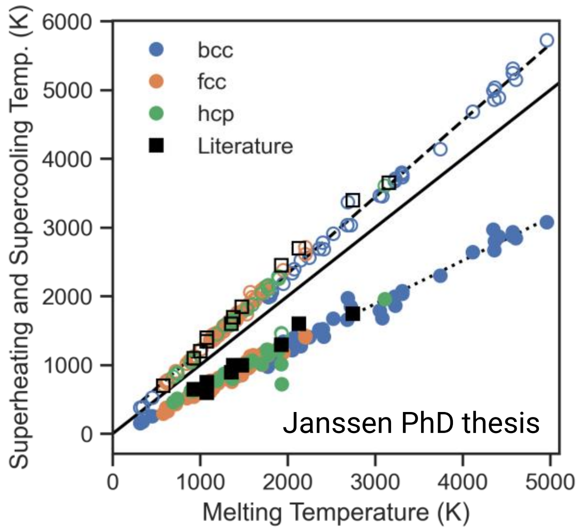
Solid Liquid Detector



Void Detector



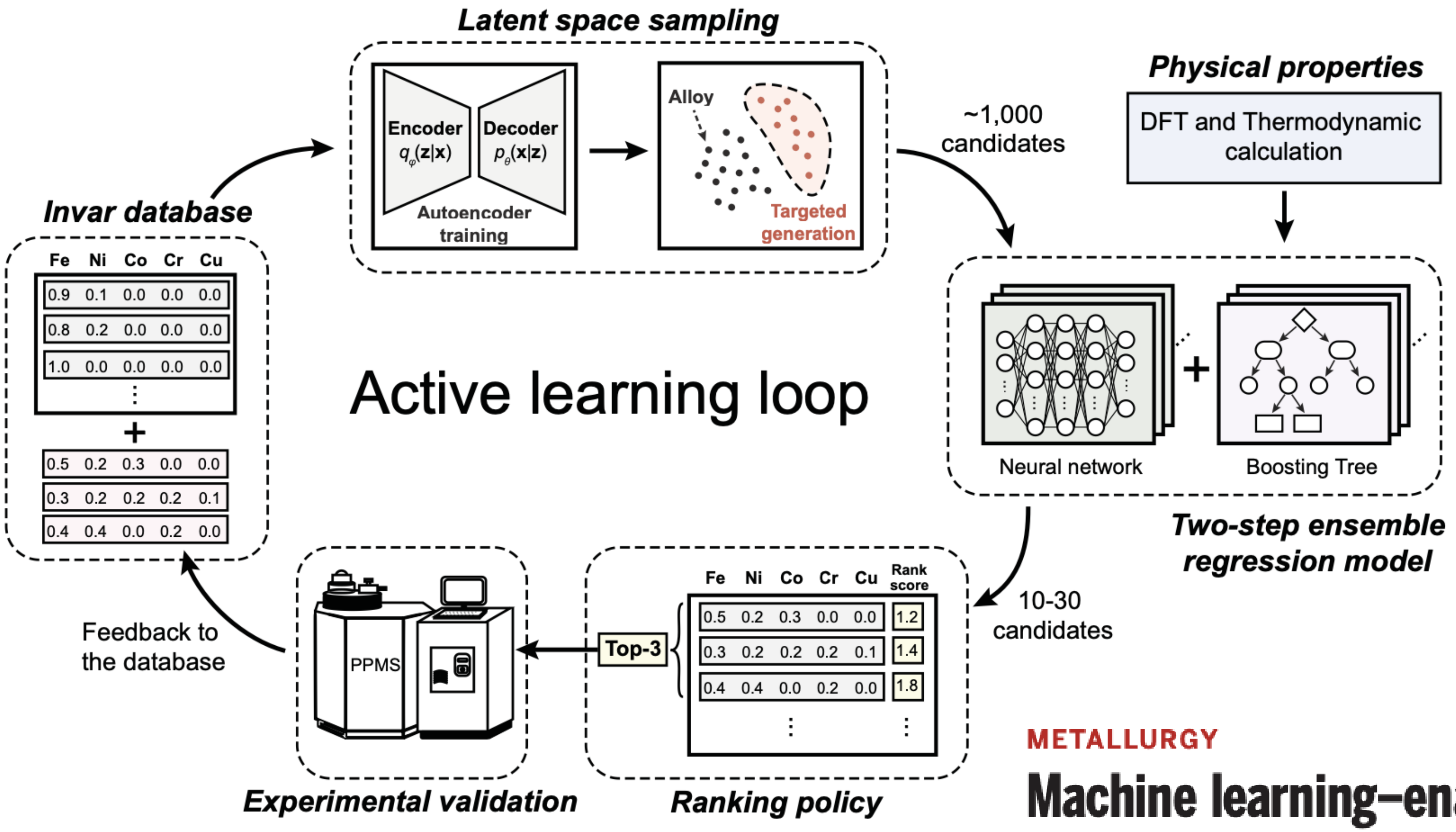
Chemical trends & descriptors



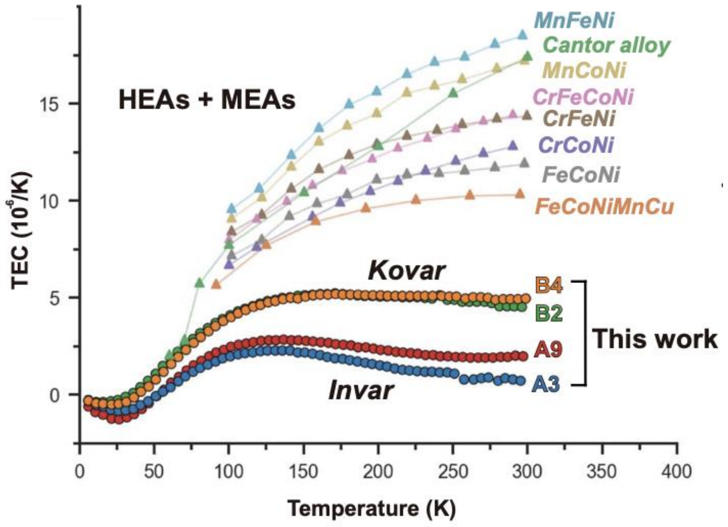
- Overheating/undercooling vs T_m
- > 300 emp. potentials

Zhu, Janssen, Ishibashi, Kormann, Grabowski, JN, Comp. Mat. Sci. 187, 12 (2021)

High-entropy alloy discovery



Experimental validation



Discovery of a new generation of invar alloys

METALLURGY

Rao *et al.*, *Science* **378**, 78–85 (2022)

Machine learning–enabled high-entropy alloy discovery

Ziyuan Rao[1], Po-Yen Tung[1,2], Ruiwen Xie[3], Ye Wei[1]*, Hongbin Zhang[3], Alberto Ferrari[4], T.P.C. Klaver[4], Fritz Körmann[1,4], Prithiv Thoudden Sukumar[1], Alisson Kwiatkowski da Silva[1], Yao Chen[1,5], Zhiming Li[1,6], Dirk Ponge[1], Jörg Neugebauer[1], Oliver Gutfleisch[1,3], Stefan Bauer[7], Dierk Raabe[1]*

Many of the recent breakthroughs in materials science simulations rely on complex simulation protocols (workflows)

past

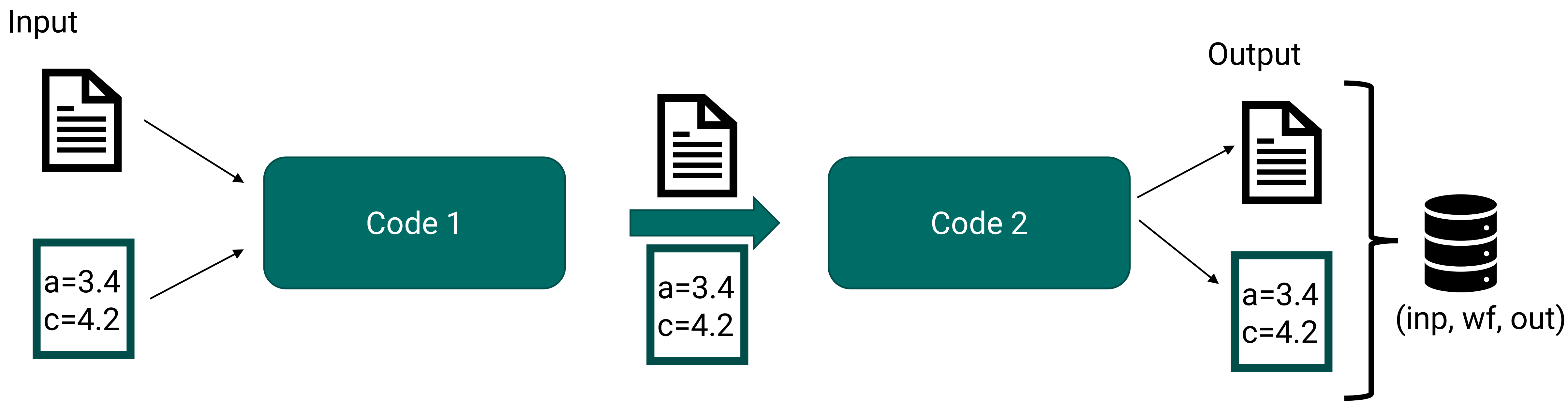
Monolithic program/code

paradigm change →

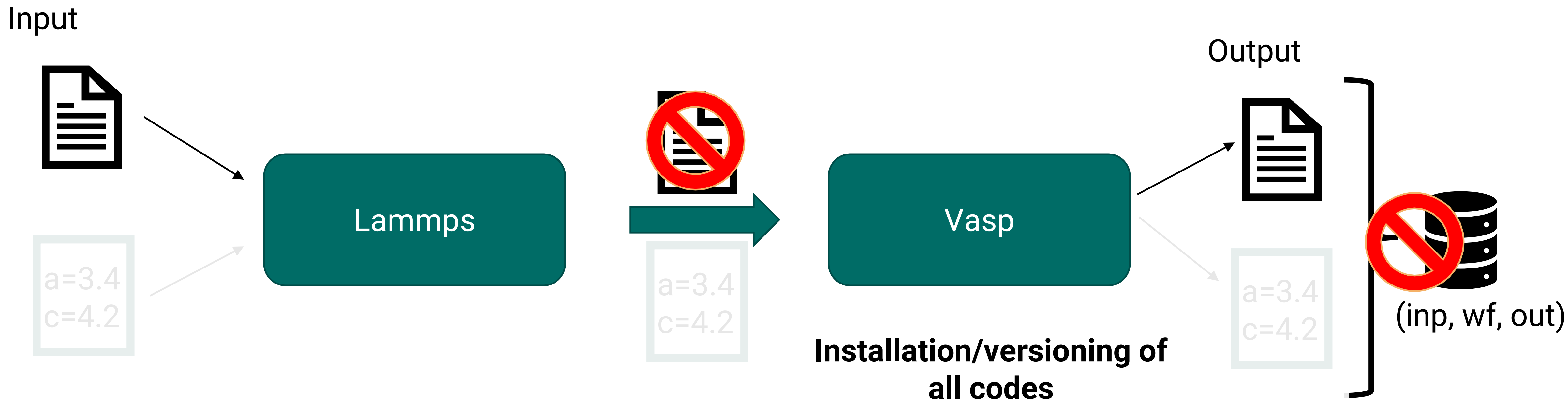
future

Large number of interacting codes and tasks

Workflows: Fundamentals



Workflows: Challenges



Code specific input

➢ Technical expertise for all codes in workflow is mandatory

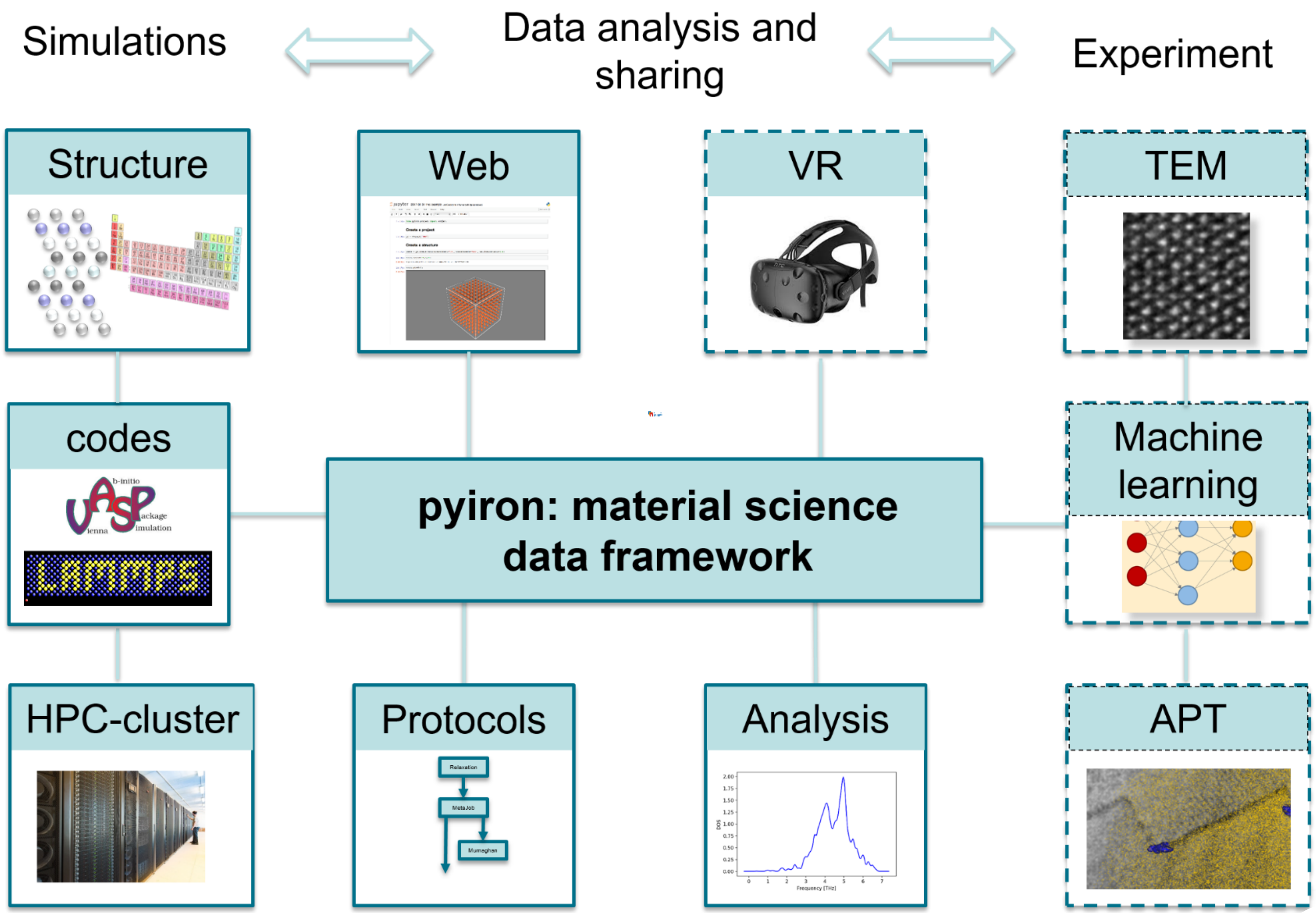
Lack of interoperability/ interactivity

Code specific output

Manual upload to repository

Needed: Tools for FAIR data and workflows

Pyiron: Python + Max Planck Institute for **Iron** research



Janssen, Surendralal, Hickel, Todorova, Lysogorskiy, Drautz, JN, Comp.Mat. Sci. **163**, 24 (2019)

Structural and Chemical Atomic Complexity

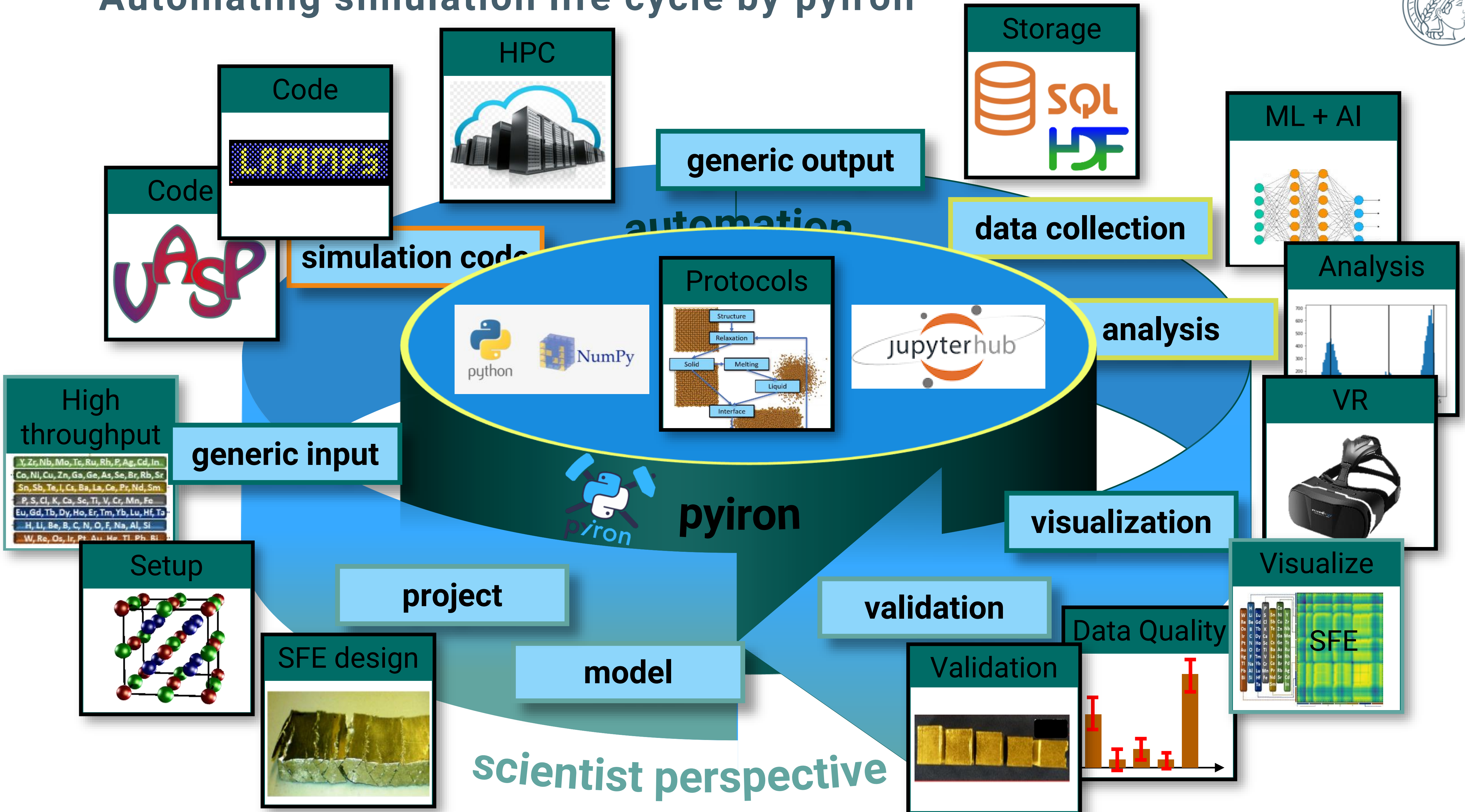
www.pyiron.org

Key pyiron features

User-friendliness:

- Fully integrated in Jupyter ecosystem (no need for console)
- Autocompletion (no need to memorize commands)
- Generic input/output (no need to learn code specific language)
- Upscaling (from interactive notebooks to high-throughput HPC workflows)
- Database centric (without having to ever know about it)
- IDE - Integrated Development Environment (all tools are accessed by same pyiron interface)

Automating simulation life cycle by pyiron



Tutorial

Concept:

- Cloud-Infrastructure and Jupyter Notebooks provided
- No installation, you just need a web browser

Sessions:

- Introduction into pyiron
- Generating (DFT) datasets and ML potential fitting (ACE)
- Validation and real-world applications
- FAIR data and workflows

Enjoy the tutorial!