### Materials modeling: Building bonds across atoms, code and people

### Michael F. Herbst<sup>1</sup> and Rachel C. Kurchin<sup>a</sup>

 Mathematics for Materials Modelling (matmat.org), EPFL
 a. Accelerated Computation of Materials for Energy (acme-group-cmu.github.io), Carnegie Mellon University

#### 10 July 2024 https://michael-herbst.com/slides/juliacon2024







### What is Materials Science and Engineering?

MSE is about understanding (and leveraging) relationships between key attributes of materials:



### MSE example: Cheese! (processing)



<sup>1</sup>Photo from

https://www.dairyfoods.com/articles/96116-moving-towards-carbon-neutral-cheesemaking

<sup>2</sup>Photo from https://www.cheeseprofessor.com/blog/brooklyn-cheese-caves

### MSE example: Cheese! (structure)



<sup>&</sup>lt;sup>3</sup>Image from DOI: 10.1088/2515-7639/aba2b7

<sup>&</sup>lt;sup>4</sup>Image from DOI: 10.22443/rms.inf.1.55

### MSE example: Cheese! (properties)



<sup>5</sup>Photo from https://thescranline.com/ultimate-grilled-cheese/

<sup>6</sup>Art from Sudowoodo on iStock

### MSE example: Cheese! (performance)



<sup>&</sup>lt;sup>7</sup>Photo from https://www.thetimes.com/life-style/health-fitness/article/ are-you-a-cheese-addict-time-to-quit-ph3thqvcd

<sup>&</sup>lt;sup>8</sup>Did I chose this photo because the source URL was funny? Mayyybe...

### What is Materials Science and Engineering?

What do we mean by a material?  $\rightarrow$  basically, anything solid or solid-ish

This includes many materials classes such as metals, ceramics, polymers, and composites, which can be atomically ordered (crystalline) or not, as well as exist in a variety of geometries (bulk, films, nanostructures...)



### What is Materials Science and Engineering?

MSE relies/builds on concepts from physics, chemistry, and engineering fields (mechanical, chemical, electrical, ...) – the field is highly interdisciplinary!

New/better materials are also crucial to the innovations that will enable society to address the grand challenges of the 21st century, including fighting the climate crisis, improving global health, and enabling next-generation technologies!



- Modeling materials through computer solutions of the governing equations for physical models that describe them
- "Computational experiments" can offer valuable complementarity to lab experiments and pencil/paper theory
- Computational MSE offers tremendous value for acceleration of materials design/discovery/optimization

### Computational materials science



Computational materials science: Electronic structure





**Key math/numerics ideas:** nonlinear complex-valued eigenvalue problems, fixed-point methods, preconditioners, constrained non-convex optimization, ...

### Computational materials science: Molecular dynamics



**Key math/numerics ideas:** massive systems of coupled ODE's (with symplectic integration), control schemes, constrained optimization, high-dimensional statistics and parameter inference, dynamical systems, ...

### Computational materials science: Mesoscale modeling



**Key math/numerics ideas:** partial differential equations, adaptive time-stepping, Monte Carlo, ...

### Computational materials science: Continuum modeling



**Key math/numerics ideas:** partial differential equations, finite elements, meshing, ...

### Towards materials design



Computation has a key enabling role to play in materials discovery/design due to slow speed and high monetary/energetic cost of experiment (1 fume hood  $\simeq 2-3$  average households<sup>9</sup>)

<sup>&</sup>lt;sup>9</sup>D. Wesolowski et. al. Int. J. Sustain. High. Edu. 11, 217 (2010).

### High-throughput materials screening









### High-throughput materials screening

• We can fully automate this !



### High-throughput materials screening



16 / 42

### Computational materials discovery



- Goal: Only promising candidates made in the lab
- Systematic simulations on  $\simeq 10^4 10^6 \ {\rm compounds}$ 
  - Noteworthy share of world's supercomputing resources
- This approach (and computational MSE more broadly) is used extensively in industrial contexts as well as academia!

# AIRBUS **D** • BASF **BOSCH** CITRINE Google GORE-TEX® Meta Microsoft



What is materials science and engineering?





State and future of JuliaMolSim

### Computing complex properties



Melting point

(lonic) transport

- Dynamics important for many properties: Involved workflows
- Example: Aluminium crystal (low-end of the scale)

### Aluminium dynamics: One time step



• Our video: 1000 time steps

### Aluminium dynamics: Many time steps



- Our video: 1000 time steps
- Runtime for one  $|\Psi^{(2)}|$ : hours

### Aluminium dynamics: Many time steps



- Our video: 1000 time steps
- Runtime for one  $\mathbf{ML}$ : < 1 second





- Large-scale data generation & regression/training
- Many balances to strike:
  - Expressiveness of model
  - Accuracy of model
  - Data requirements for constructing sensible model
  - Cost of data generation
  - Cost of training procedure
  - Cost of evaluation of ML model
  - . . .
- $\Rightarrow$  No one size fits all
- $\Rightarrow$  Flexibility in simulation (e.g. ML model) is key

### Flexibility in practice ...

- Text-based interfaces are still standard (to lesser and lesser extent)
- Vastly different syntax & features
  - Units, conventions, implicit assumptions, ...
- Sometimes surprising changes between versions ....



CIF structure file

QuantumESPRESSO

LAMMPS

Some standard input files for our aluminium case

### Solution: Invoke the power of abstraction



• JuliaMolSim: AbstractSystem (AtomsBase) and AtomsCalculator

- We are not the first to suggest such interfaces, examples:
  - Atomistic Simulation Environment (ASE)
  - Workflow managers like FireWorks PELOW & Aiii





- Important foundational work (and inspiration for us), but:
  - Often wrappers around text IO
  - New calculator is non-trivial effort
  - Two (or more) language problem
- $\Rightarrow$  Not in line with opportunities & strengths of julia
  - Hard to exploit unique features in existing **julia** materials tools
  - Let's see some flagship examples ...

### Our aluminium simulation: Driven by julia tools<sup>10</sup>



- Structures passed as AbstractSystem
- AtomsCalculator interface currently being finalised

<sup>10</sup>https://github.com/mfherbst/juliamolsim-demo

### Solution: Invoke the power of abstraction



- JuliaMolSim: AbstractSystem (AtomsBase) and AtomsCalculator
- We are not the first to suggest such interfaces, examples:
  - Atomistic Simulation Environment (ASE)
  - Workflow managers like
     Fire Works





- Important foundational work (and inspiration for us), but:
  - Often wrappers around text IO
  - New calculator is non-trivial effort
  - Two (or more) language problem
- $\Rightarrow$  Not in line with opportunities & strengths of julia
  - Hard to exploit unique features in existing julia materials tools
  - Let's see some flagship examples ...

# Inverse design of MI





- Goal: Fit parameters of  $\left| \begin{array}{c} \overset{\mbox{\tiny \ext{\tiny ML}}}{\mathbf{ML}} \right|$
- Traditionally: Parameters fitted by "proxy quantities"
- Here end-to-end: Use AD through the simulation
- $\Rightarrow$  Need parameters exposed in interface !

# Inverse design of MI





- Goal: Fit parameters of  $|_{ML}^{\ll}$
- Traditionally: Parameters fitted by "proxy quantities"
- Here end-to-end: Use AD through the simulation
- $\Rightarrow$  Need parameters exposed in interface !

# Inverse design of MI



#### Chemical Science





View Article Online View Journal | View Issue



Cite this: Chem. Sci., 2024, 15, 4897

All publication charges for this article have been paid for by the Royal Society of Chemistry

#### Differentiable simulation to develop molecular dynamics force fields for disordered proteins†

#### Joe G. Greener 💿

dynamics on disordered proteins, Here I improve the a9958-dig force field and the GBNed/2 implicit solvent model to better describe disordered proteins. Differentiable molecular simulations with 5 ns trajectories are used to jointly optimise 0.08 parameters to better mach explicit solvent trajectories. Simulations with the improved force field better reproduce the radius of gravition and secondary structure content seen in experiments, whilst showing sliptity degraded performance on folded proteins and protein complexes. The force field better reproduce the radius of a small molecule binding sludy and improves agreement with experiment for the aggregation of anyloid peptides. CB99dms, which can be used in OpenMML is available at https://github.com/greener-group/CB99dms. This work is the first to show that gradents can be obtained directly from nanoscend length differentiable simulations of biomolecules and highlights the effectiveness of this approach to training whole force fields to match disered properties.

Implicit solvent force fields are computationally efficient but can be unsuitable for running molecular

Received 3rd October 2023 Accepted 8th February 2024

DOI: 10.1039/d3sc05230c rsc.li/chemical-science

• Goal: Fit parameters of ML



- Traditionally: Parameters fitted by "proxy quantities"
- Here end-to-end: Use AD through the simulation
- ⇒ Need parameters exposed in interface !

# Active learning for $|_{\mathbb{N}}$





www.nature.com/npjcompumats

# ARTICLE OPEN (Construction) OPEN (Construction

Cas van der Oord<sup>1127</sup>, Matthias Sachs<sup>2</sup>, Dávid Péter Kovács<sup>1</sup>, Christoph Ortner<sup>3</sup> and Gábor Csányi <sup>[6]</sup>

Data-driven interatomic potentials have emerged as a powerful tool for approximating ab initio potential energy surfaces. The most time-consuming step in creating these interatomic potentials is typically the generation of a suitable training database. To aid this process hyperactive learning (HAL), an accelerated active learning scheme, is presented as a method for rapid automated training database assembly. HAL adds a biasing term to a physically motivated sampler (e.g. molecular dynamics) driving anotic structures towards uncertainting in turn generating unsene or valuable training configurations. The proposed HAL framework is used to develop atomic cluster expansion (ACE) interatomic potentials for the AISI10 alloy and polyethylene glycol (PEG) polymer starting from roughly a dozen initial configurations. The HAL generated ACE potentials are shown to be able to determine macroscopic properties, such as melting temperature and density, with close to experimental accuracy.

npj Computational Materials (2023)9:168; https://doi.org/10.1038/s41524-023-01104-6

• Active learning: Train  $\left| {{ { { { { { { { { ML} } } } } } } } } \right|$  while running dynamics

 $\Rightarrow$  If prediction not confident for | : Run extra |





• Breaks traditional workflows !

### Robust algorithms and error control for





- First-principle models challenging & lack of mathematical analysis
- Numerics features considerable trial and error with parameters
- Recent work: Parameter-free and self-adapting algorithms<sup>1,2</sup>



Silicon band structure with error bars

- Numerical error estimation or model uncertainties hardly established
- First works<sup>3,4</sup>: Error in bands, densities and forces
- Need close maths / physics collaboration
- **DFTK** is major research tool here

<sup>&</sup>lt;sup>1</sup>MFH, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).

<sup>&</sup>lt;sup>2</sup>MFH, A. Levitt. J. Comput. Phys. 459, 111127 (2022).

<sup>&</sup>lt;sup>3</sup>MFH, A. Levitt, E. Cancès. Faraday Discus. 223, 227 (2020).

<sup>&</sup>lt;sup>4</sup>E. Cancès, G. Dusson, G. Kemlin et. al. SIAM J. Sci. Comp., 44, B1312 (2022).





Letters in Mathematical Physics (2023) 113:21 https://doi.org/10.1007/s11005-023-01645-3



#### Numerical stability and efficiency of response property calculations in density functional theory

Eric Cancès<sup>1,2</sup> · Michael F. Herbst<sup>3</sup> · Gaspard Kemlin<sup>1,2</sup> · Antoine Levitt<sup>1,2</sup> · Benjamin Stamm<sup>4</sup>



$$\begin{split} \rho_*(\theta) &= \mathrm{argmin}_{\rho} E(\rho,\theta) \\ \mathrm{sensitivities} &= \frac{\mathrm{d} \, \nabla E(\rho_*(\theta),\theta)}{\mathrm{d} \theta} \end{split}$$

Generic and efficient solvers for AD gradients (linear response)
 ⇒ A first for plane-wave DFT !

- Towards routine sensitivity analysis, inverse problems
  - But needs care: Talk by Niklas Schmitz (Th 15:40, Function (4.1))
- $\Rightarrow$  To fully exploit: Interface supporting error propagation

# Efficient AD for $|_{\Psi}$





- Generic and efficient solvers for AD gradients (linear response)
   ⇒ A first for plane-wave DFT !
- Towards routine sensitivity analysis, inverse problems
  - But needs care: Talk by Niklas Schmitz (Th 15:40, Function (4.1))
- $\Rightarrow$  To fully exploit: Interface supporting error propagation

### What we want: A shopping list for modern research thrusts

- Improve composability of flagship julia materials tools
  - Error control and robust algorithms
  - New ML models and learning approaches
  - Algorithmic differentiation
- Integration across ecosystems:
  - Vast portfolio of relevant methods (e.g. in LAMMPS, ASE, QE)
  - ⇒ Need two-way communication (for testing & adoption!)
- Dynamic and heterogeneous modelling pipelines:
  - Interdependent steps (active learning, on-the-fly data generation)
  - Heterogenous hardware (ML on GPU, DFT on CPU)
  - $\Rightarrow$  Dynamic job scheduling is crucial
- Support cross-disciplinary research:
  - Speed of research problem
  - $\bullet\,$  Not all fundamental maths / CS ideas work in practical setting
  - ⇒ Need flexibility inside hot loops (ML kernels, algorithms, ...) & HPC performance (for testing & adoption!)
  - $\Rightarrow$  People compose if software composes



What is materials science and engineering?





State and future of JuliaMolSim

### julia materials ecosystem: What you can do today

- Interact with AbstractSystems:
  - AtomsBuilder.jl: Build structures (surfaces, defects, ...)
  - AtomsIO.jl: Read/write structure files
- Some calculators (not all AtomsCalculators ... yet):
  - DFTK.jl: Density-functional theory (1D & 1000 electrons) (Talk on Friday, 12:00, Function 4.1)
  - InteratomicPotentials.jl: Simple parametrised materials potentials
  - acesuit: ACE-based machine learning potentials
  - Various active learning tools (e.g. HAL, PotentialLearning.jl, Cairn.jl

- see talk on Friday, 16:20, Method 1.5)

- Dynamics and post-processing:
  - Molly.jl: Differentiable, GPU-enabled molecular dynamics (Talk on Friday, 11:00, REPL 2)
  - Wannier.jl: Localisation of electronic structures
- External ecosystem integrations:
  - ASEconvert, LAMMPS.jl, Spglib.jl ...

### Other cool things we did not have time to talk about



- <sup>1</sup>L. Martínez Comp. Phys. Comm. 279, 108452 (2022)
- <sup>2</sup>RCK and V. Viswanathan J. Chem. Phys. 153, 134706 (2020)
- <sup>3</sup>RCK, D. Gandhi, and V. Viswanathan J. Phys. Chem. Lett. 14, 35 (2023)
- <sup>4</sup>M. Babar, Z. Zhu, RCK, et. al. J. Am. Chem. Soc. 146, 23 (2024)
- <sup>5</sup>J. Gardner, O. Douglas-Gallardo, W. Stark, et. al. J. Chem. Phys. **156**, 174801 (2022)
- <sup>6</sup>J. Gardner, D. Corken, S. Janke, et. al. J. Chem. Phys. 158, 064101 (2023)

Gas

Particles

# Current state of 🕞 Julia MolSim

- Package development driven by research questions
  - Ecosystem is co-evolving with problems tackled
  - ⇒ Packages still lack breadth, but became more mature recently
    - julia tools have been key ingredient to push state of the art
  - ⇒ Community recognition (3 of JuliaMolSim now junior faculty)
- Interoperability and collaboration have been side projects
  - JuliaMolSim started around 2019 by Christoph Ortner, then AtomsBase (JuliaCon 2021), AtomsCalculator (end 2023)
  - Past year: Increased cross-group collaboration & community-centred packages (next slides)
  - $\Rightarrow$  Integration is early days
  - $\Rightarrow$  Potential is there, need funding to improve integration

## JuliaMolSim: Some ongoing projects

- Increase internal interoperability and collaboration:
  - Redesign of AtomsBase & AtomsCalculator
  - GeometryOptimization.jl: Based on above, generic algorithms to find minimal-energy material structure (standard time-consuming modelling task)
  - NeighbourLists.jl: Fast, multi-hardware implementation (crucial ingredient to molecular dynamics)
- Cross-ecosystem integration
  - **OFTK** + **AiiDA**: **julia** backend within high-throughput DFT workflows
  - AtomsCalculator + ASE: Python-based interface standard
  - AtomsCalculator + . Socket protocol for dynamics
  - $\Rightarrow$  Goal: Improve testing and adoption !

# JuliaMolSim: Gripes and issues

- Obstacles to julia adoption
  - Overly ambitious promises by the community
  - Too rudimentary binary support (dynamic libraries)
  - Gripes with python  $\rightarrow$  julia calls
- Obstacles to 🚱 Julia MolSim adoption
  - Limited portfolio of available physical models
  - Insufficient performance (often low-hanging fruits!)
  - Lack of documentation
    - Integration between 🛞 Julia MolSim packages
    - Across-ecosystem integration
    - Examples
- Lack of important utility packages
  - Structure visualisation, plotting, printing & standard analysis
- $\Rightarrow$  If you are interested to help, reach out !

### What we want: A shopping list for modern research thrusts

# Improve composability of flagship julia materials tools Error control and robust algorithms

- New ML models and learning approaches
- Algorithmic differentiation

- Integration across ecosystems:
   Vast portfolio of relevant methods (e.g. in LAMMPS, ASE, QE)

#### Dynamic and heterogeneous modelling pipelines:

- Interdependent steps (active learning, on-the-fly data generation)
- Heterogenous hardware (ML on GPU, DFT on CPU)

# Support cross-disciplinary research: • Speed of research problem

- Not all fundamental maths / CS ideas work in practical setting

### How to get involved

Try things out! Integrate your stuff! File issues! Please don't hesitate to reach out and tell us what works and what doesn't.



Connect with us at JuliaMolSim.org and join our Zulip server to chat! We aim to have ~monthly calls soon.

#### Here at JuliaCon:

- Find us after the poster session: socialise over drinks!
- Join us at the hackathon: we have a few project ideas

### Advertisement break

### Open PostDoc in the EPFL Mt Mat group



Topic: Efficient inverse materials design

- Interdisciplinary environment
- Bayesian optimisation
- AD & gradient approaches
- See https://matmat.org/jobs/

MolSSI workshop (R. C. Kurchin):

"Julia for computational molecular and materials science"

- 20–23 October 2024 at Carnegie Mellon University, Pittsburgh
- Targets: People within and outside julia & JuliaMolSim communities
- $\Rightarrow$  Stay tuned on 🚱 Julia MolSim Zulip for registration details !

Psi-k workshop (M. F. Herbst, A. Levitt, J. Haegeman): "Julia for numerical problems in quantum and solid-state physics"

- 26–28 November 2024 at EPFL, CECAM-HQ, Lausanne
- Targets: Linear algebra, physics and julia communities
- ⇒ https://www.cecam.org/workshop-details/1355 (Deadline: 20th Sep) 39/42

### ${\sf Conclusion}\ /\ {\sf Outlook}$

- Computational materials science
  - Established in academia & industry
  - High-throughput simulations: A different kind of HPC



- Importance of dynamic & heterogeneous workflows
  - Novel tools and paradigms (inverse design, active learning)
  - Composability is key  $\Rightarrow$  julia tools can be at the forefront
- 🚯 Julia MolSim ecosystem
  - Key simulation functionality available
  - Selected state-of-the-art features
  - Consensus building important, but (often) tiresome
  - Software integration (across ecosystems) is real work !
- Challenges looking forward
  - Expanding the user base beyond package developers
  - Reducing ecosystem (and julia) entrance barrier
  - More performance, documentation & core utilities

### Acknowledgements

AtomsBase/AtomsCalculators

- Christoph Ortner (UBC)
- Teemu Järvinen (UBC)
- Joe Greener (Cambridge)
- Spencer Wyant (MIT)

JuliaMolSim community

- Mt Mat group
- ACME group
- MIT CESMIX team



### Questions?



#### Rachel Kurchin

0	rkurchin
$\checkmark$	rkurchin@cmu.edu
Ø	<pre>mse.engineering.cmu.edu/ acmegroup</pre>



- https://juliamolsim.org/
- https://michael-herbst.com/slides/juliacon2024