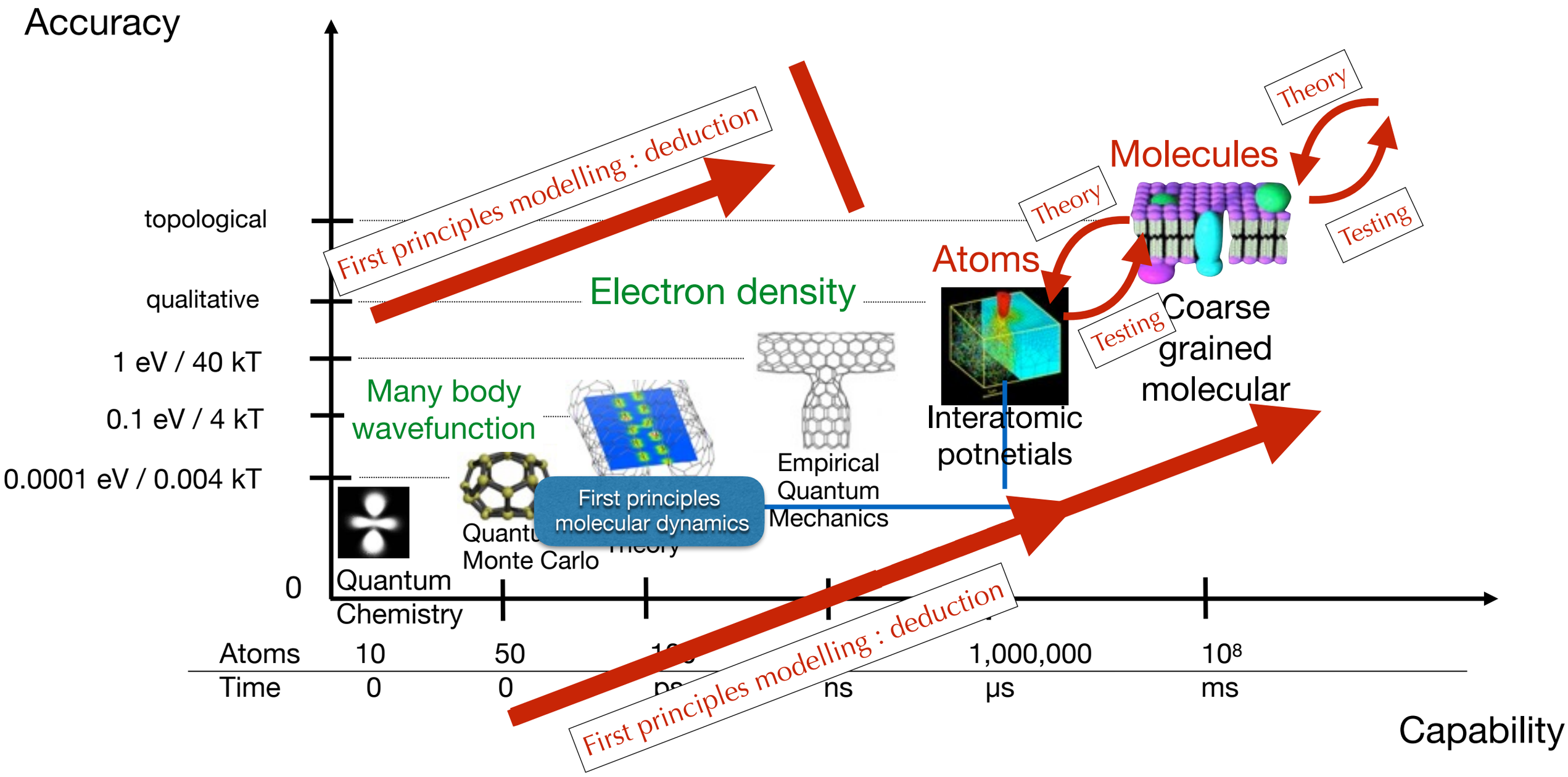
Interatomic potential fitting: illustration and workflows

Gábor Csányi

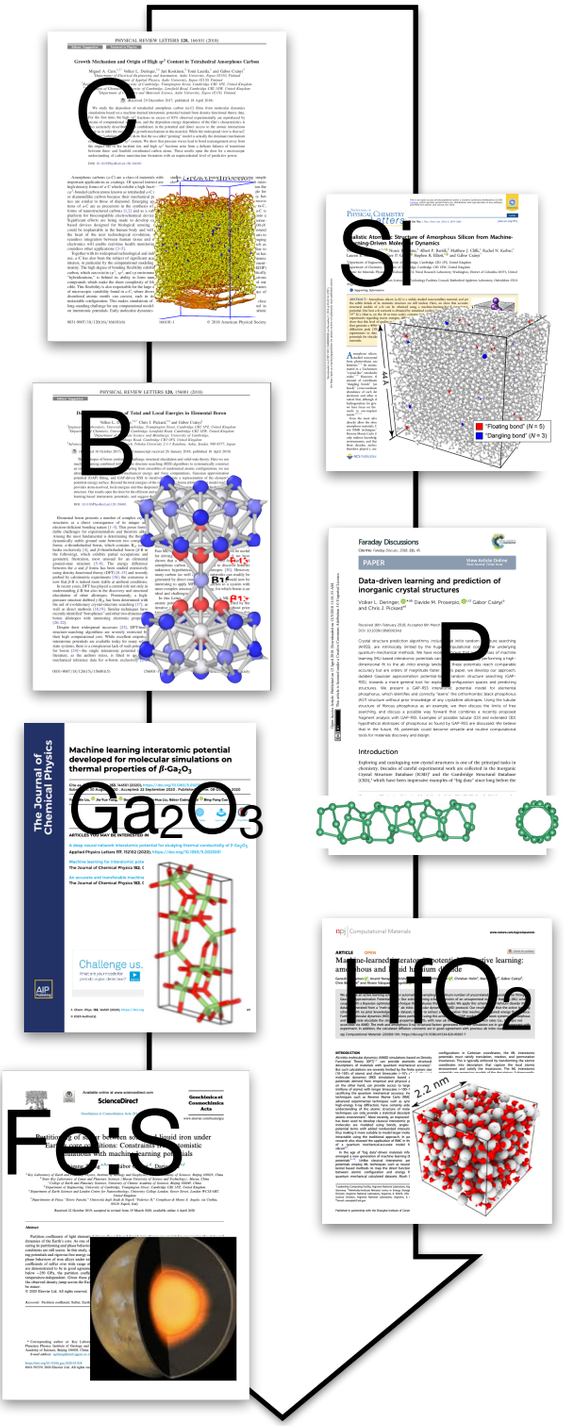
UNIVERSITY OF CAMBRIDGE

Multiple scales of modelling



What happened in the past ~15 years?

Applications



Function fitting (regression)

Analytical full PES fits

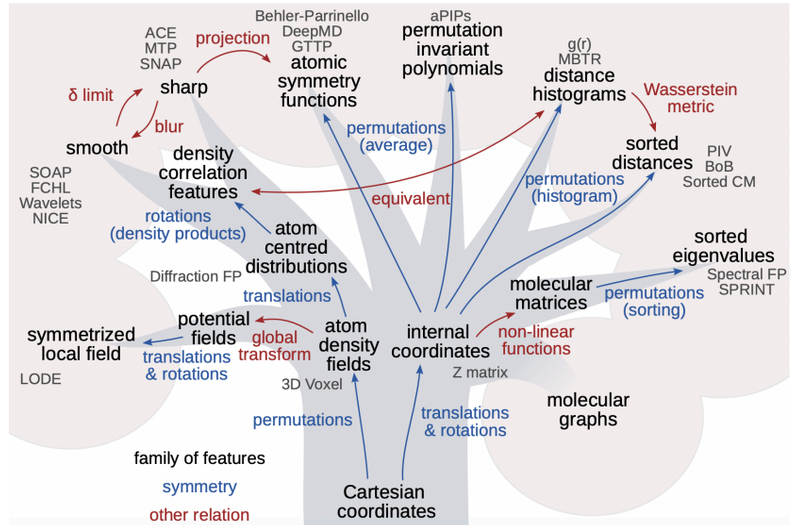
Empirical force fields

Kernel machines (e.g. Gaussian process)

Feed-forward neural networks

Symmetric polynomials

Local representations and descriptors



many body correlation-based representations

Body ordered force field terms

PIPs and molecular many body expansion

symmetric polynomials for atomic environments

Kernels and shallow NNs

Equivariant message passing neural networks

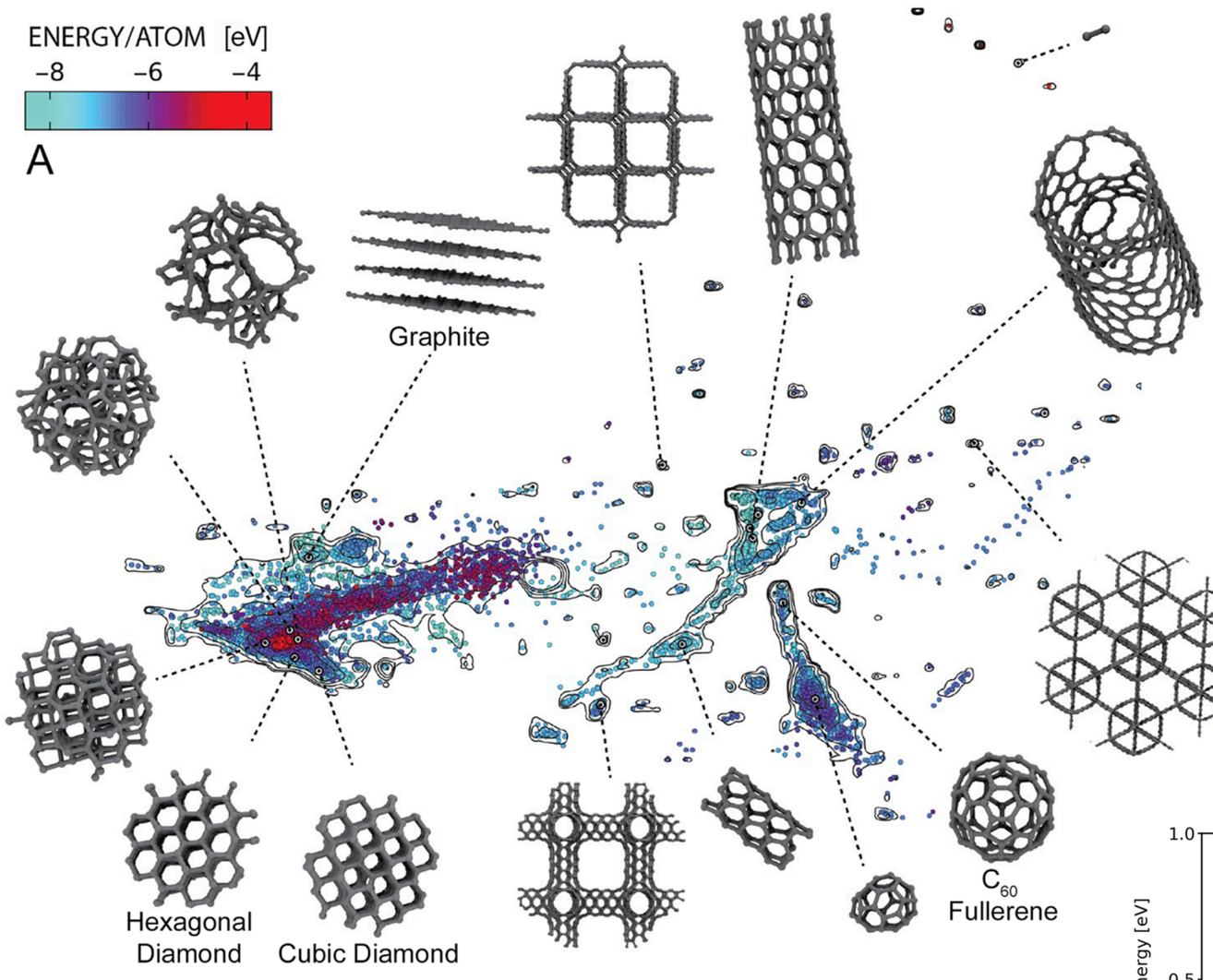
General carbon SOAP-GAP model (2020)

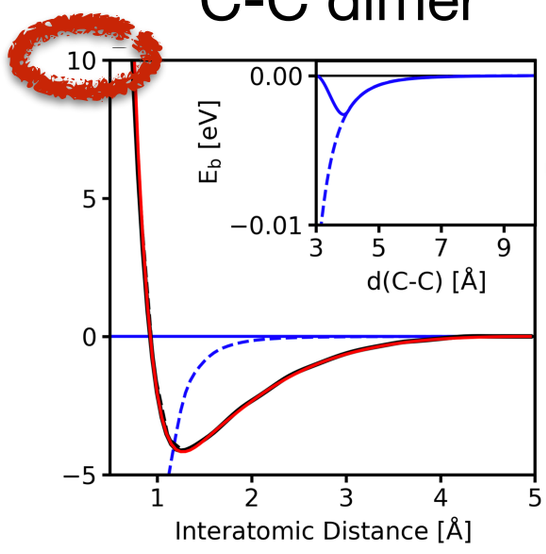
Full . Submitted: 17 February 2020 . Accepted: 22 June 2020 . Published Online: 15 July 2020

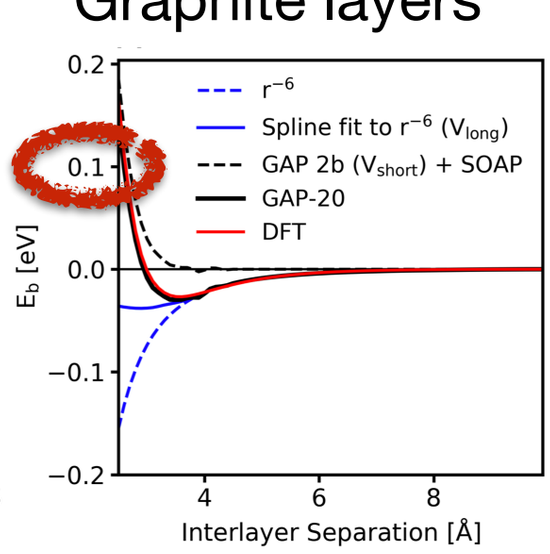
An accurate and transferable machine learning potential for carbon

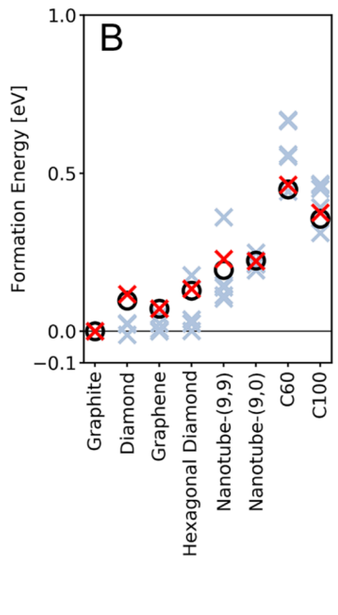
J. Chem. Phys. **153**, 034702 (2020); https://doi.org/10.1063/5.0005084

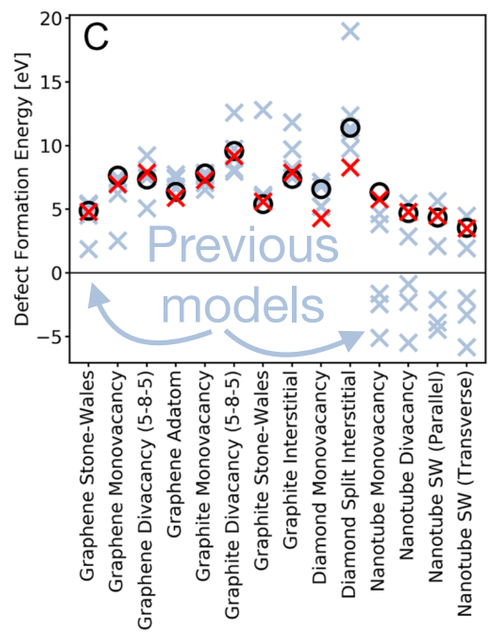
Patrick Rowe[1], Volker L. Deringer[2], Piero Gasparotto[1], Gábor Csányi[3], *and* Angelos Michaelides[1,a)]

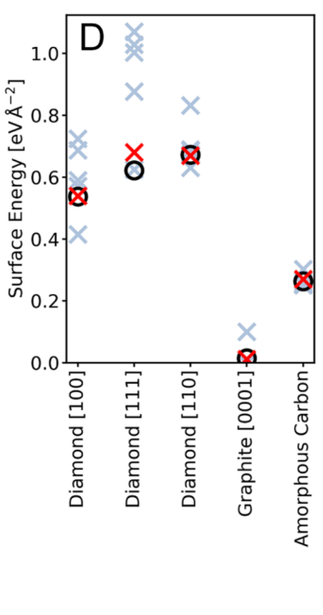












Sparse kernel model (SOAP kernel)
~ 9000 basis functions
~ 6000 oracle evaluations (~400k scalars)
> 10^6 speedup

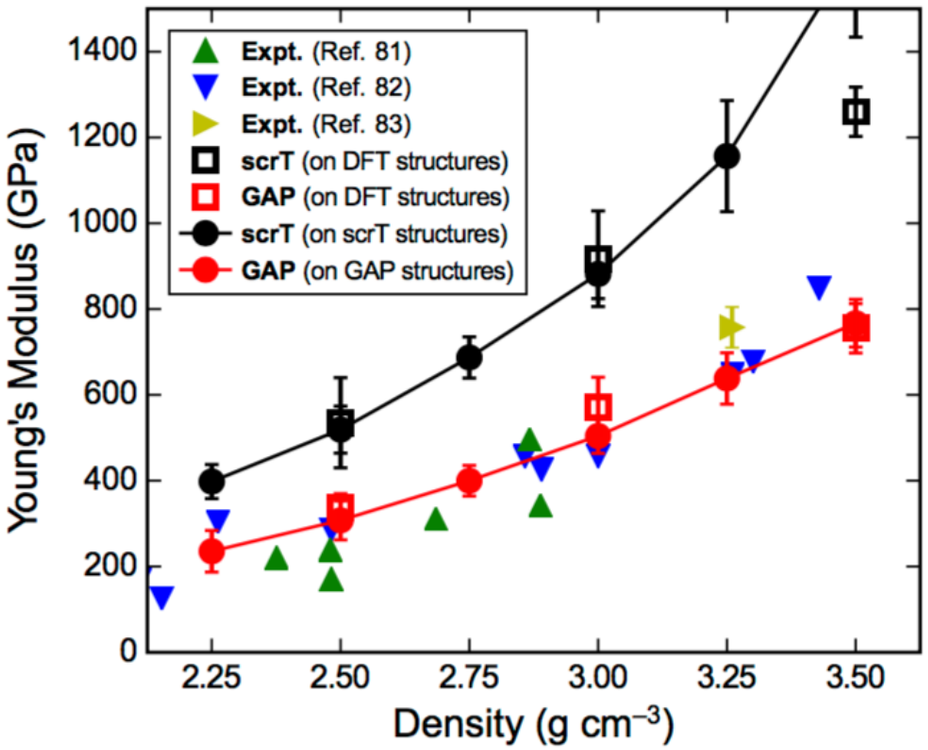
Target ○

Model ×

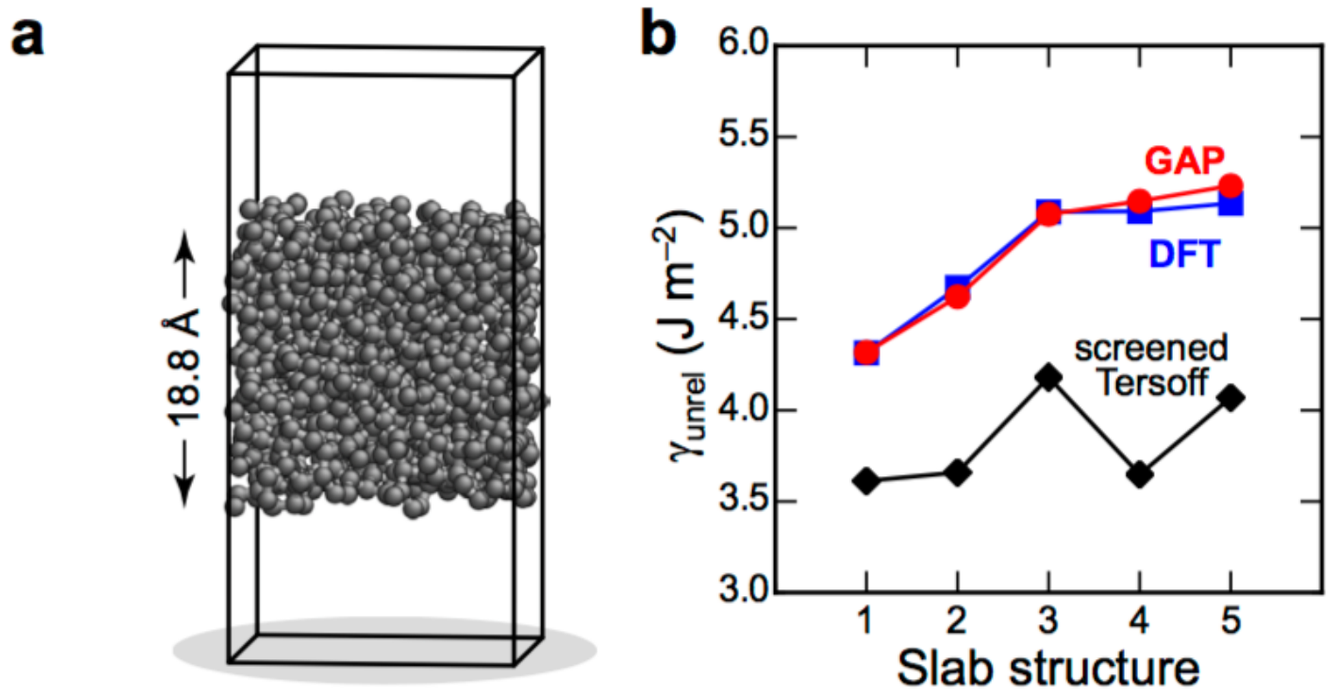
Further properties

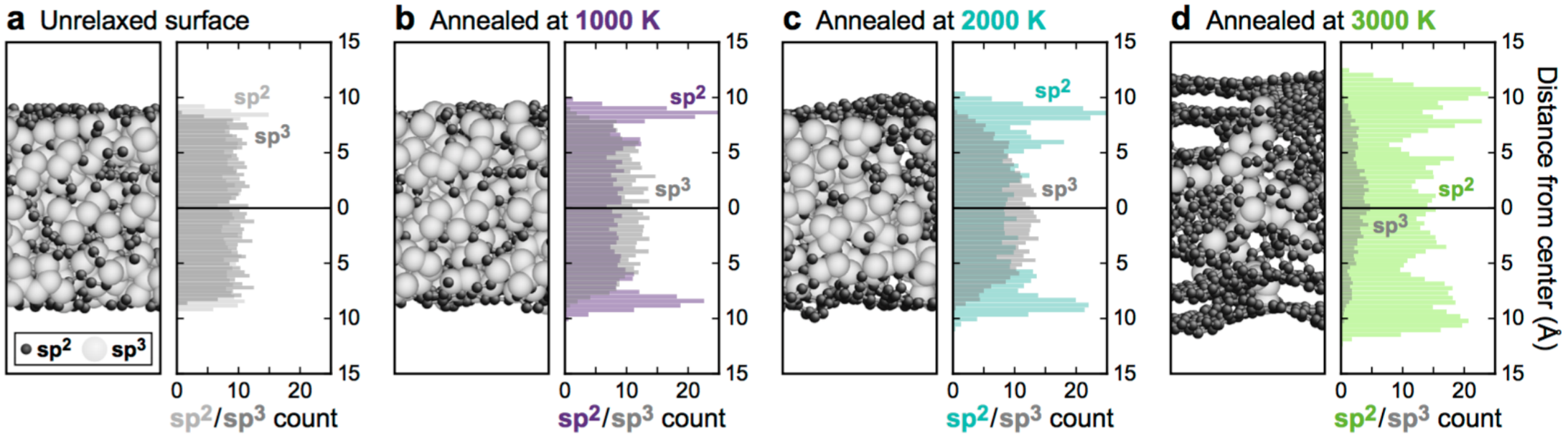
amorphous carbon

Young's modulus



Surface energies



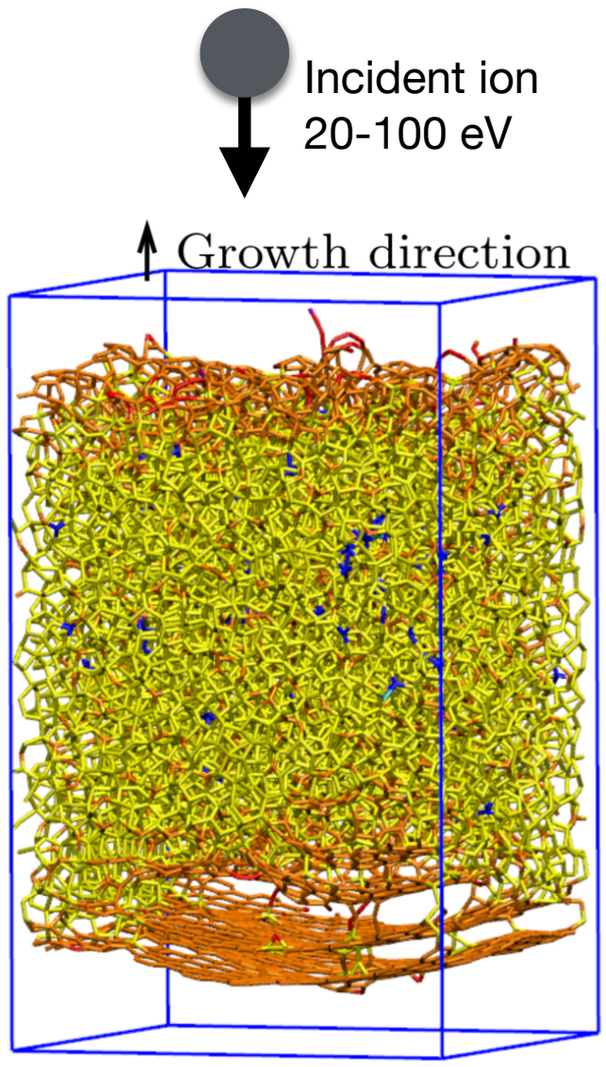


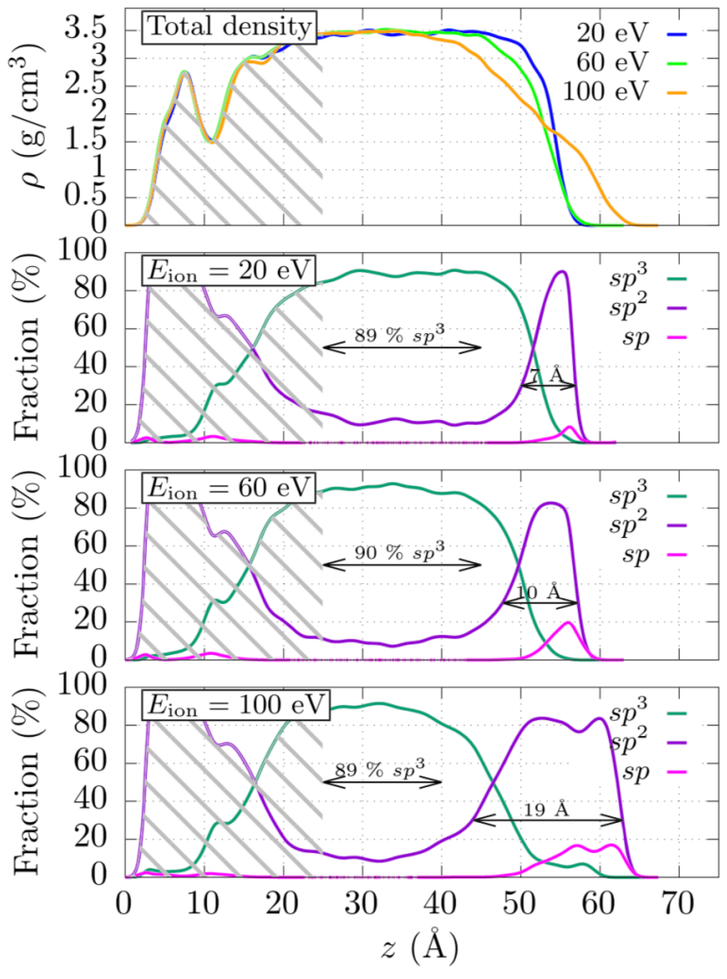
PHYSICAL REVIEW LETTERS **120,** 166101 (2018)

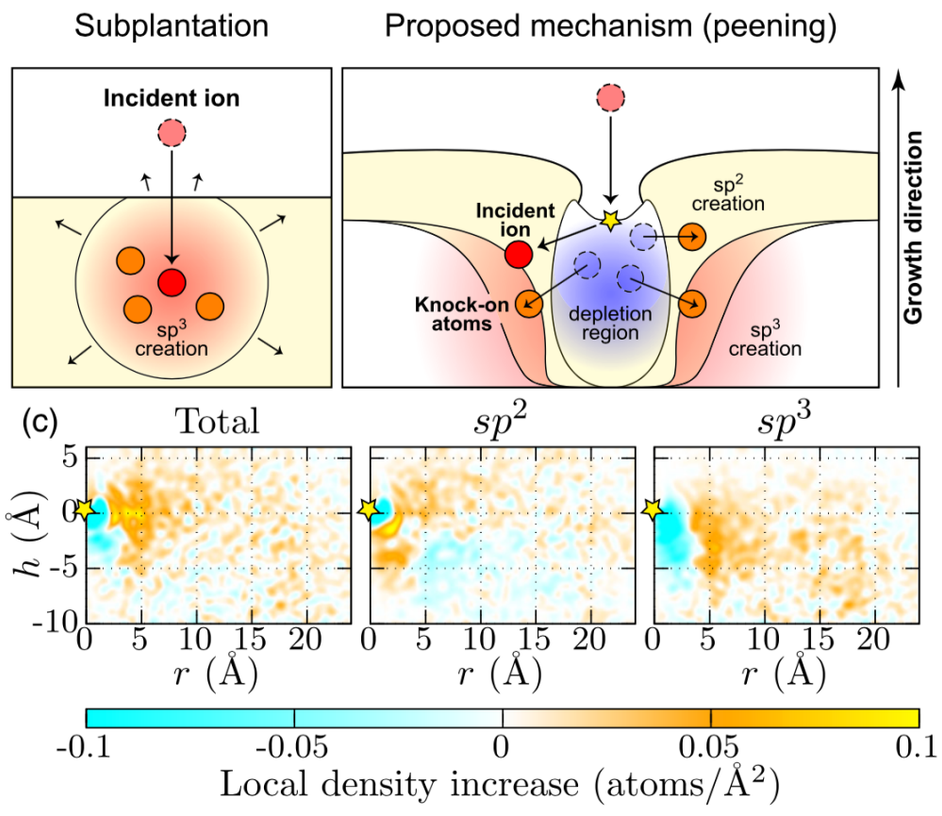
Editors' Suggestion | Featured in Physics

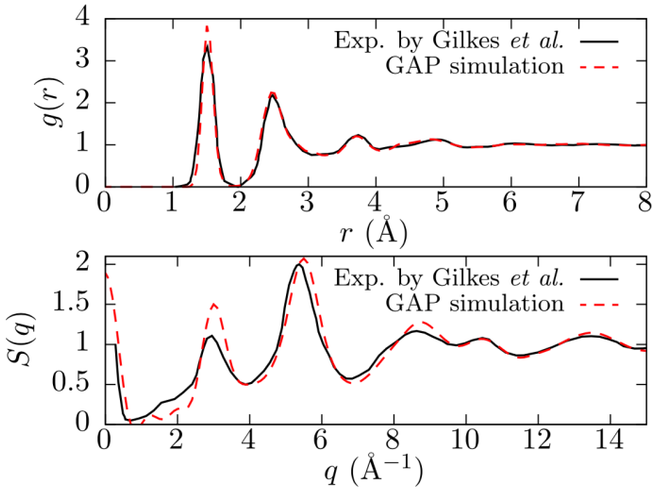
Growth Mechanism and Origin of High sp^3 Content in Tetrahedral Amorphous Carbon

Miguel A. Caro,[1,2,*] Volker L. Deringer,[3,4] Jari Koskinen,[5] Tomi Laurila,[1] and Gábor Csányi[3]

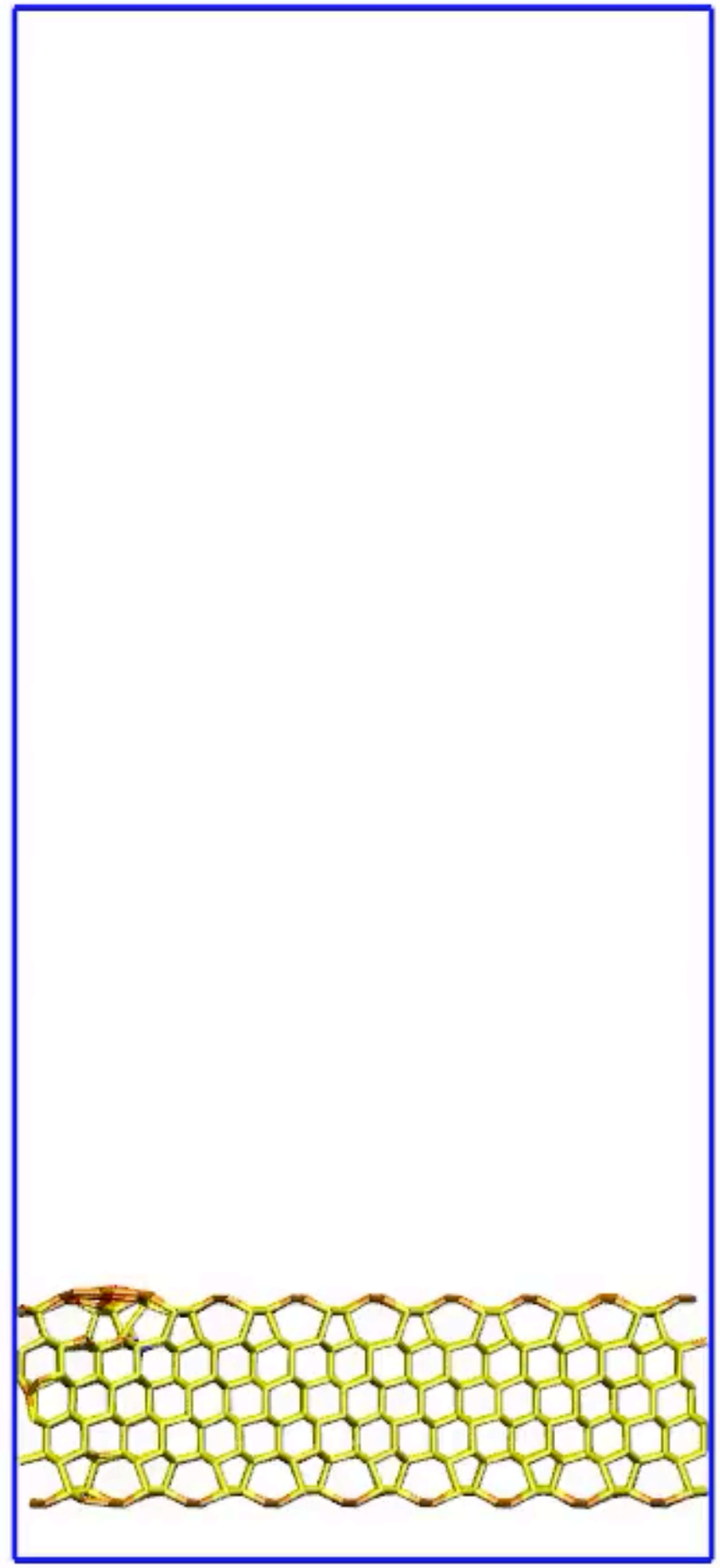


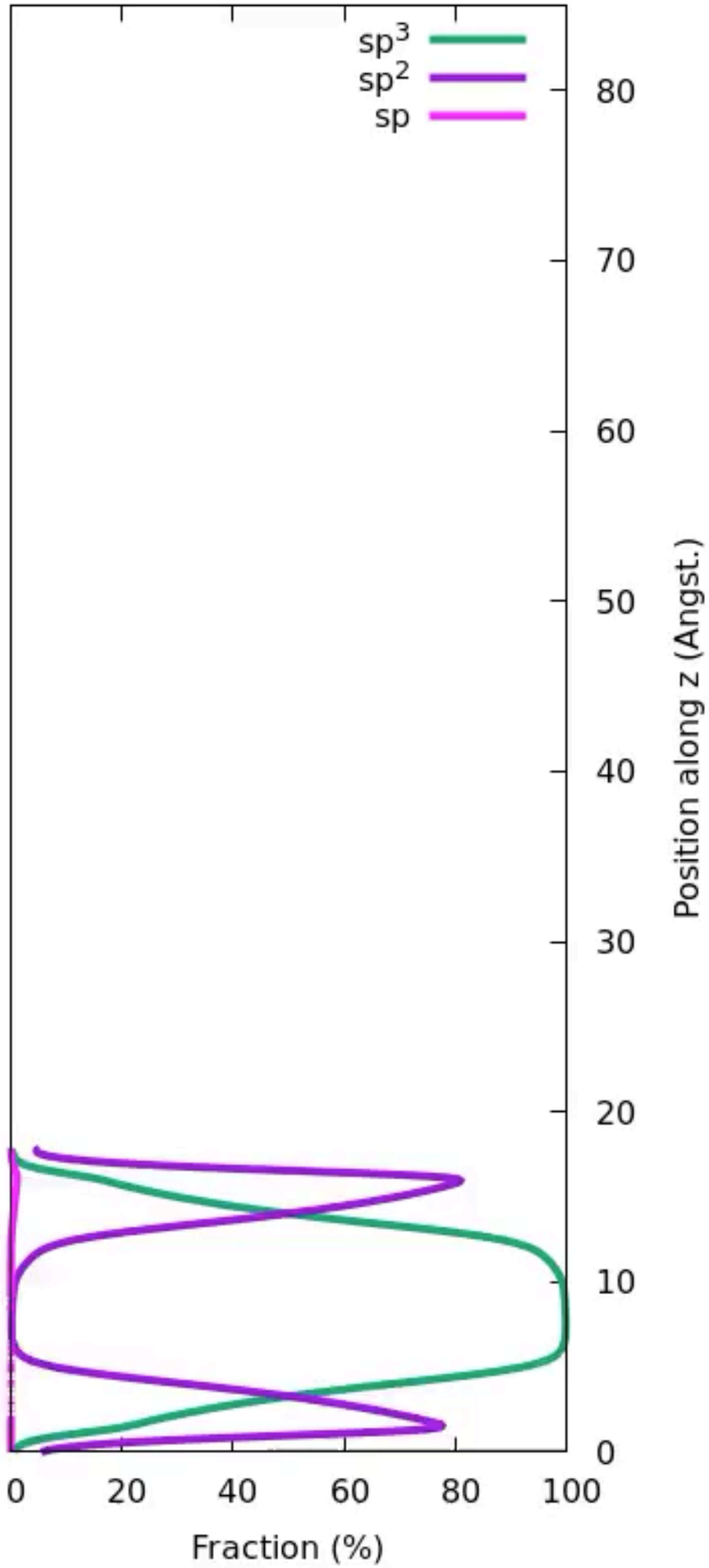






Speedup wrt. explicit electronic simulation: ~ 10^5
would have taken 30,000 years


60 eV 00001 impacts


sp3
sp2
sp
0
10
20
30
40
50
60
70
80
Position along z (Angst.)
0
20
40
60
80
100
Fraction (%)

Summary of silicon material properties

PHYSICAL REVIEW X

Highlights Recent Subjects Accepted Collections Authors Referees Search

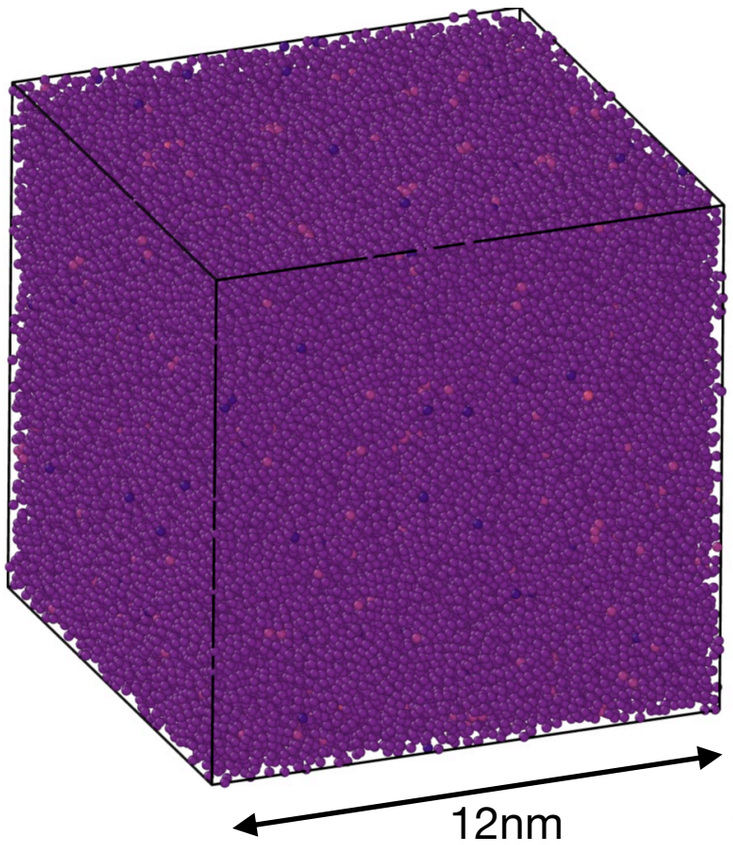
Open Access

Machine Learning a General-Purpose Interatomic Potential for Silicon

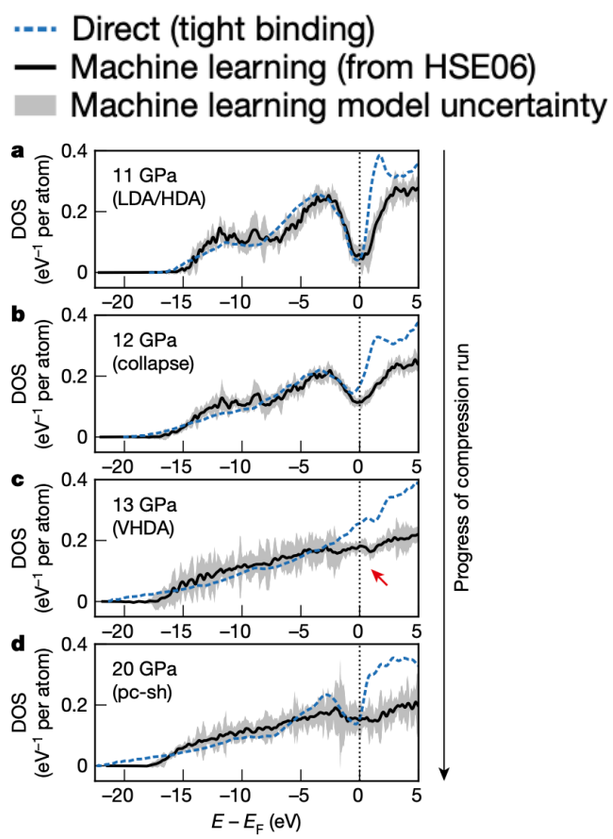
Albert P. Bartók, James Kermode, Noam Bernstein, and Gábor Csányi

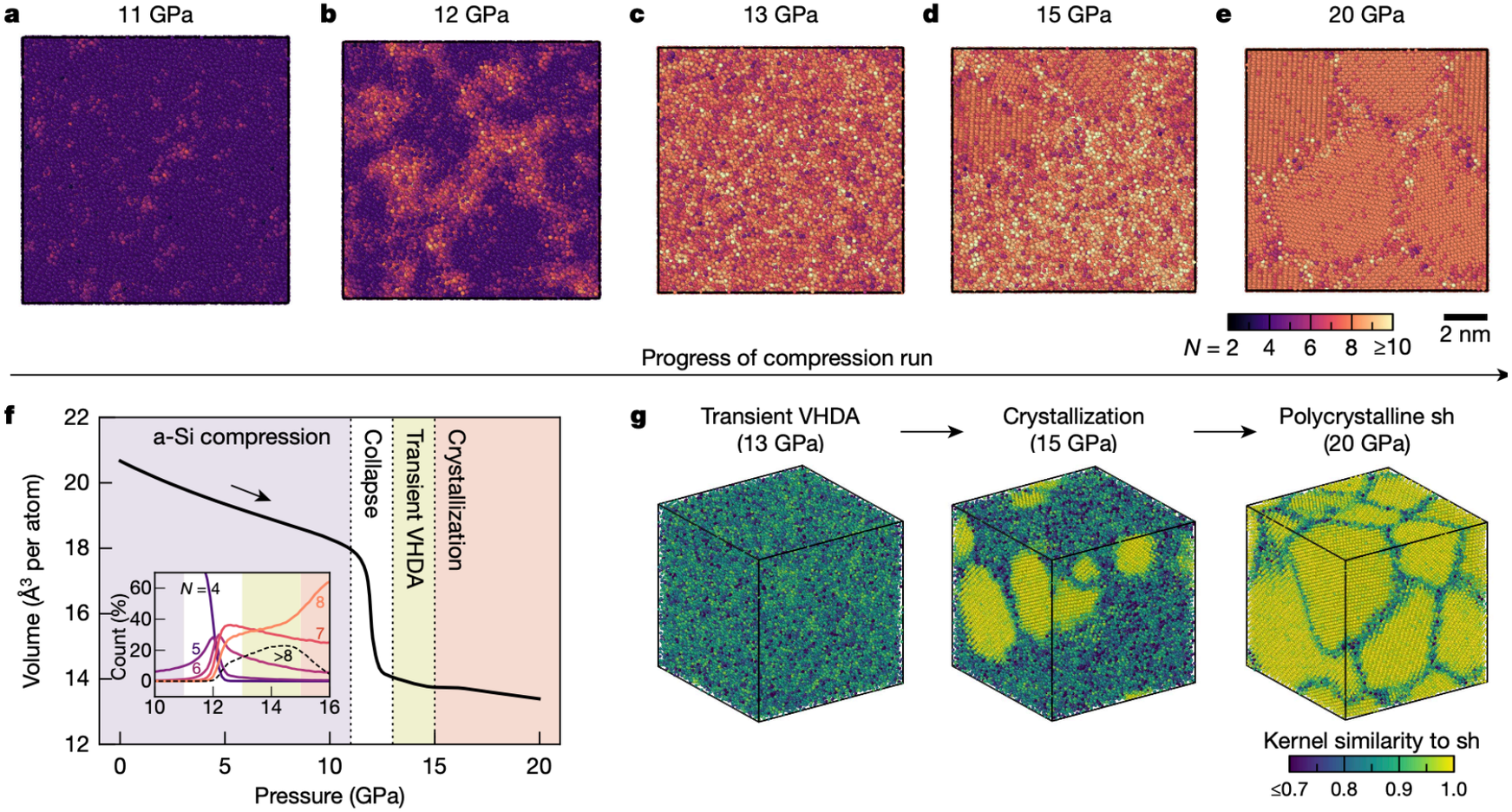
Phys. Rev. X **8**, 041048 – Published 14 December 2018

Amorphous Si science



100K atoms, 200 ps compression
Pressure-induced crystallisation
New very-high-density-amorphous phase





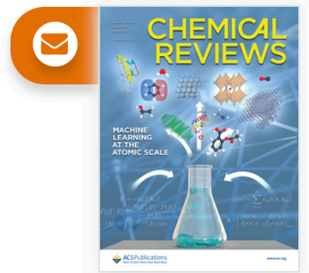
Pair of papers in Chemical Reviews

RETURN TO ISSUE | < PREV **REVIEW** NEXT >

Gaussian Process Regression for Materials and Molecules

Volker L. Deringer*, Albert P. Bartók*, Noam Bernstein, David M. Wilkins, Michele Ceriotti, and Gábor Csányi*

Cite this: *Chem. Rev.* 2021, 121, 16, 10073–10141
Publication Date: August 16, 2021
https://doi.org/10.1021/acs.chemrev.1c00022



Chemical Reviews

RETURN TO ISSUE | < PREV **REVIEW** NEXT >

Physics-Inspired Structural Representations for Molecules and Materials

Felix Musil, Andrea Grisafi, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti*

Cite this: *Chem. Rev.* 2021, 121, 16, 9759–9815
Publication Date: July 26, 2021
https://doi.org/10.1021/acs.chemrev.1c00021

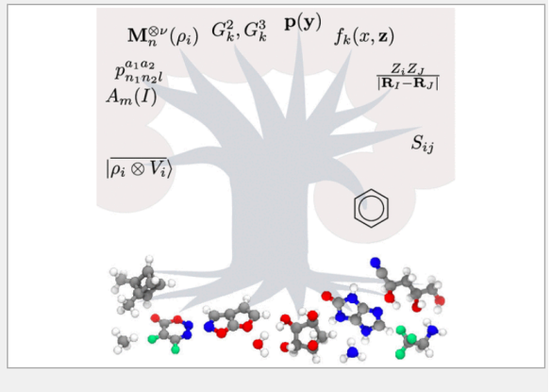
Chemical Reviews

PDF (22 MB)

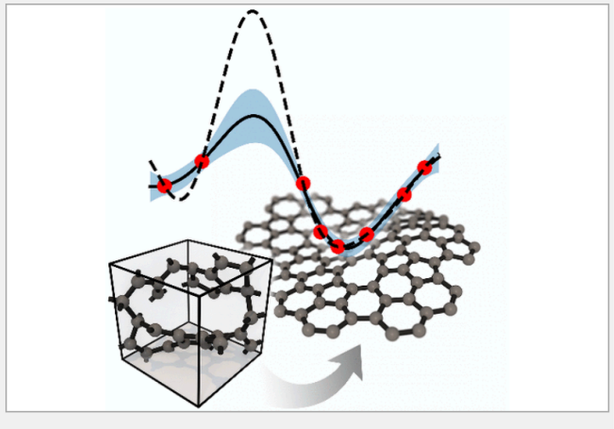
SUBJECTS: Group theory, Chemical structure, Molecular structure,

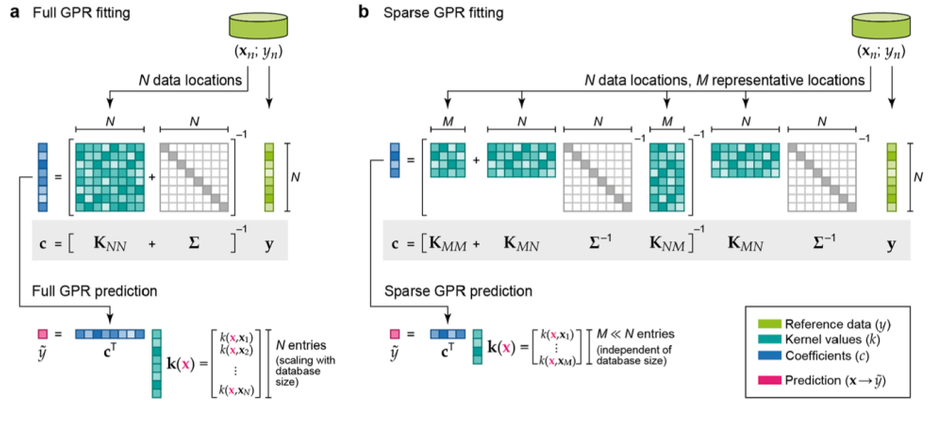
Abstract

The first step in the construction of a regression model or a data-driven analysis, aiming to predict or elucidate the relationship between the atomic-scale structure of matter and its properties, involves transforming the Cartesian coordinates of the atoms into a suitable *representation*. The development of atomic-scale representations has played, and continues to play, a central role in the success of machine-learning methods for chemistry and materials science. This review summarizes the current understanding of the nature and characteristics of the most commonly used structural and chemical descriptions of atomistic structures, highlighting the deep underlying connections between different frameworks and the ideas that lead to computationally efficient and universally applicable models. It emphasizes the link between properties, structures, their physical chemistry, and their mathematical description, provides examples of recent applications to a diverse set of chemical and materials science problems, and outlines the open questions and the most promising research directions in the field.



process regression (GPR) machine-
erials science and chemistry. The
ession of atomistic properties: in
omic potentials, or force fields, in
GAP) framework; beyond this, we
, vectorial, and tensorial quantities.
generation, representation, and
v a data-driven model may be
cussed. A survey of applications to
try and materials science illustrates
outlined for the development of the

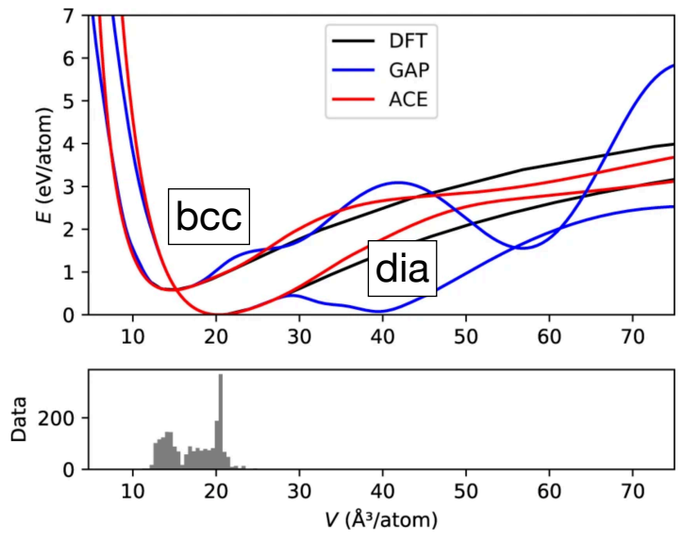




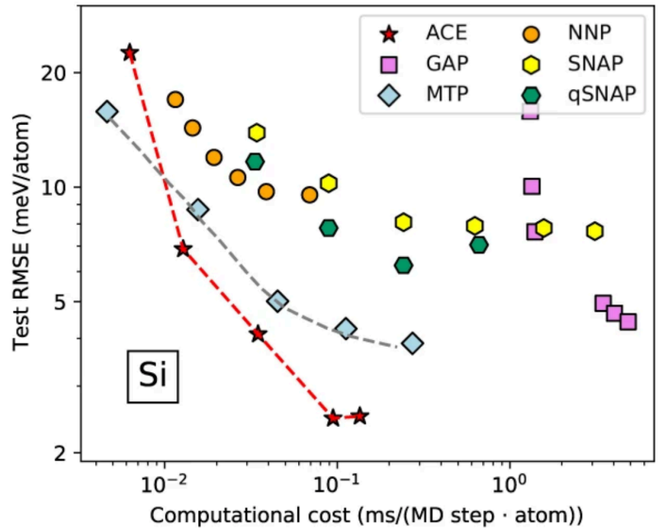
Latest developments: Atomic Cluster Expansion

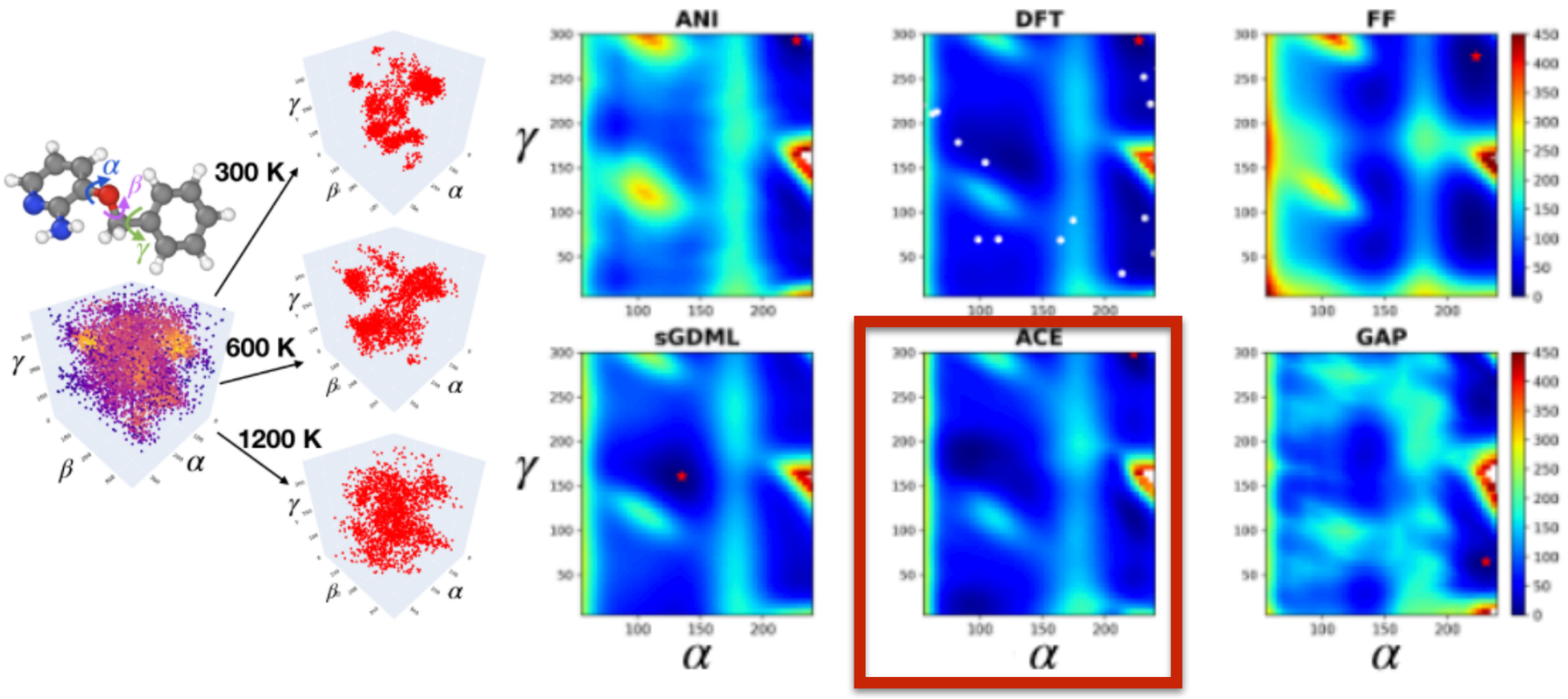
- ***Linear*** polynomial models
- Explicit high body order
- Generalisation of SOAP
- Smooth, extensive, transferable, fast
- Drautz (2019), Shapeev (2016, MTP)

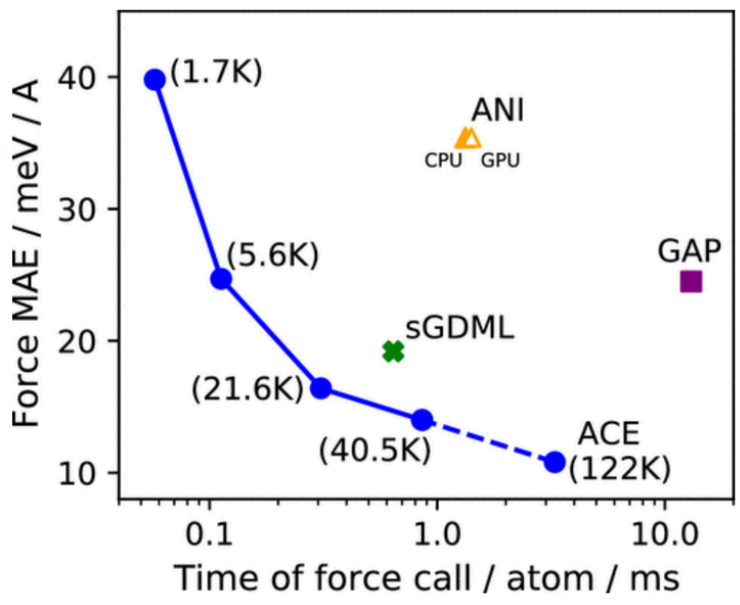
Extrapolation



Efficiency







Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon

Yury Lysogorskiy, Cas van der Oord, Anton Bochkarev, Sarath Menon, Matteo Rinaldi, Thomas Hammerschmidt, Matous Mrovec, Aidan Thompson, Gábor Csányi, Christoph Ortner & Ralf Drautz

npj Computational Materials **7**, Article number: 97 (2021) | Cite this article

Linear Atomic Cluster Expansion Force Fields for Organic Molecules: Beyond RMSE

Dávid Péter Kovács*, Cas van der Oord, Jiri Kucera, Alice E. A. Allen, Daniel J. Cole, Christoph Ortner, and Gábor Csányi

Cite this: *J. Chem. Theory Comput.* 2021, XXXX, XXX, XXX-XXX
Publication Date: November 4, 2021
https://doi.org/10.1021/acs.jctc.1c00647
© 2021 The Authors. Published by American Chemical

Article Views 932 | Altmetric 6 | Citations -
LEARN ABOUT THESE METRICS

Current modelling challenges

- Solved short range potential fitting problem, energy depends on near environment only
- Long range interactions: flexible electrostatics and charge transfer
  Need to model full self-consistency of DFT
- Higher length scales:

  "Coarse grained" models can be made similarly
  (But accuracy vs cost tradeoff is different!)

- Lower length scales:

  Full electronic wave functions have exploitable local structure, new representations can be used to escape exponential scaling

New workflows for atomistic modelling

1. “ML-enhanced *ab initio* MD”

   - Major DFT codes will have ML “interpolators” built-in:
     VASP already does, Castep coming soon

   - Allows extending time scales, complex chemistry, basically automatic

2. “General force field” for a given application

   - Assemble a range of configurations. For materials:
     pressure, temperature, composition, phases, defects etc

   - High throughput DFT, fit potential : iterative training

   - LAMMPS/OpenMM/Gromacs etc. for large scale production runs

     how do we know simulation is correct?

New *kinds* of modelling questions

- Forward problem:

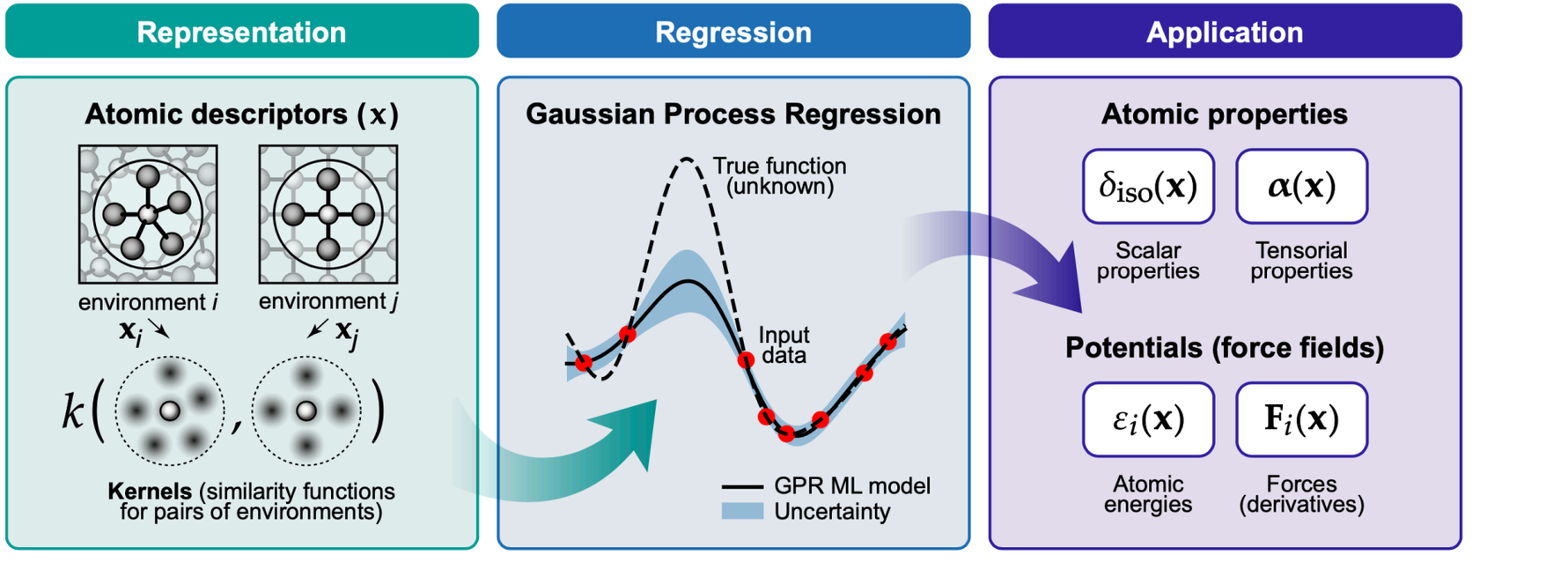
  Given a database of small structures, how does the accuracy of fit (controlled with weights) determine the accuracy of material properties?

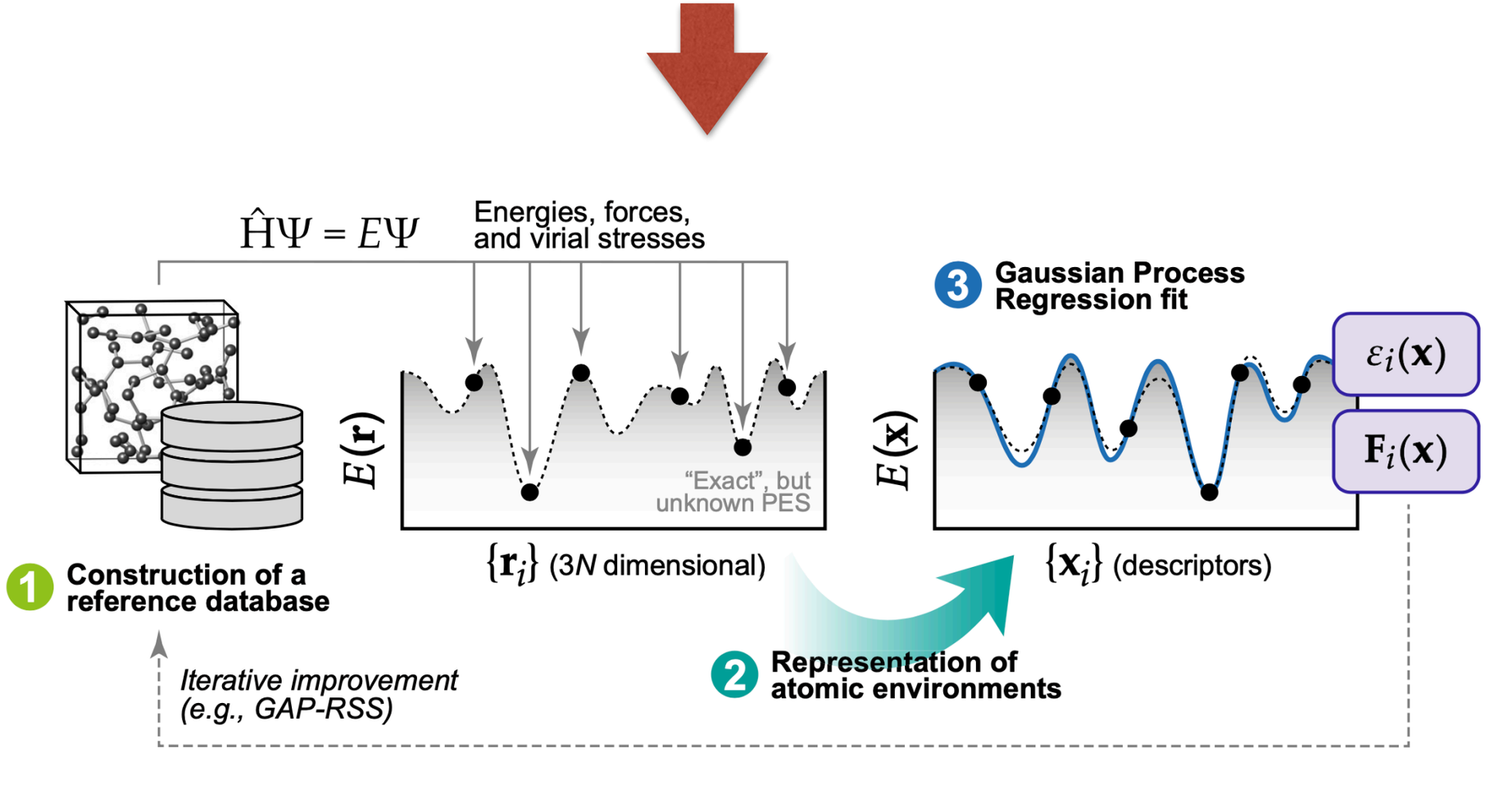
- Inverse problem:

  Given a material property of interest, what fitting database is required (and what accuracy of fit) to obtain a good prediction?

- These questions did not arise before because force fields were not systematic !

"First principles" force fields





“Semi-automated” potential fitting

PHYSICAL REVIEW LETTERS **120,** 156001 (2018)

Editors' Suggestion

Data-Driven Learning of Total and Local Energies in Elemental Boron

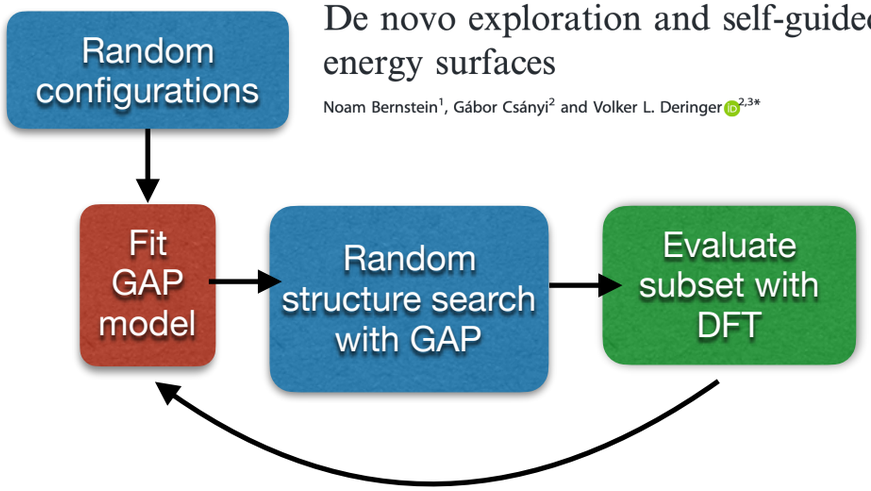
Volker L. Deringer,[1,2,*] Chris J. Pickard,[3,4] and Gábor Csányi[1]

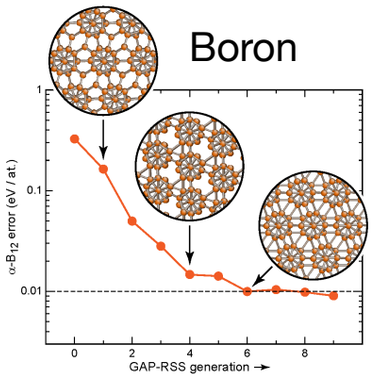
npj | Computational Materials

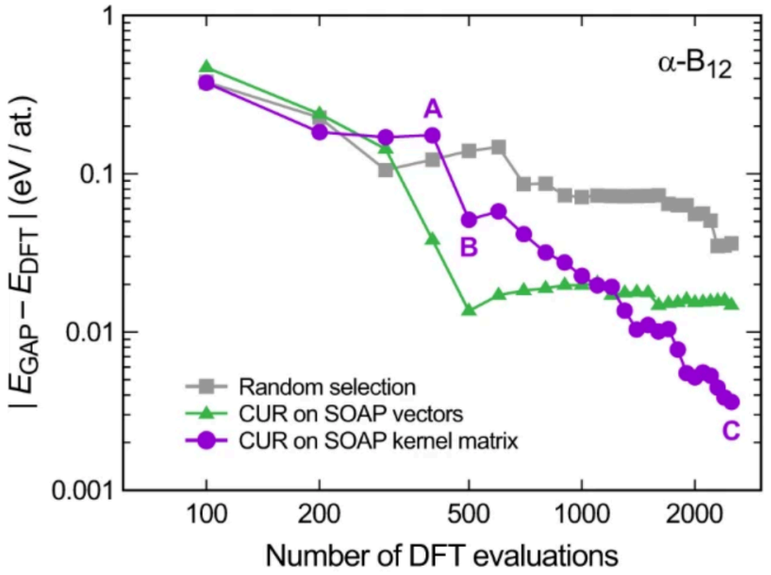
ARTICLE OPEN

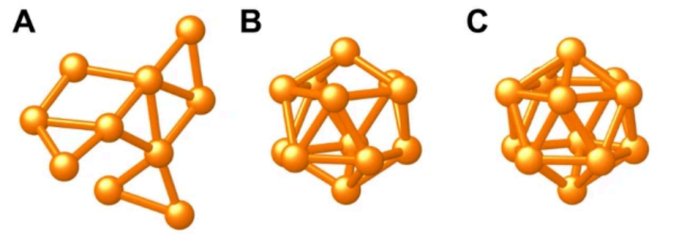
De novo exploration and self-guided learning of potential-energy surfaces

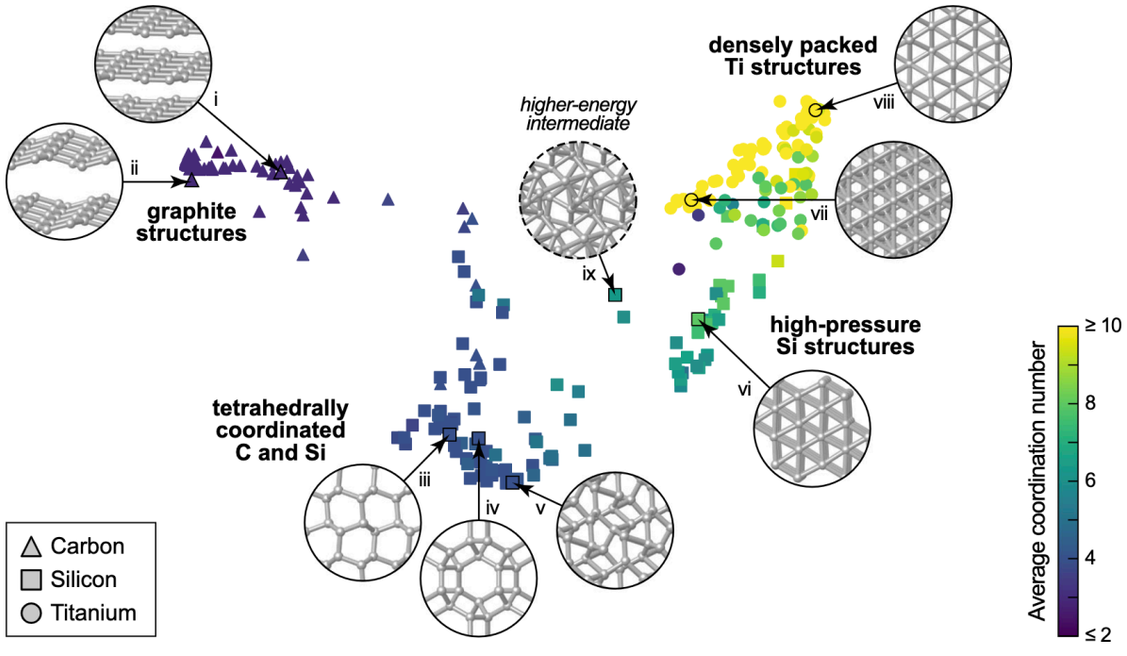
Noam Bernstein[1], Gábor Csányi[2] and Volker L. Deringer[2,3*]

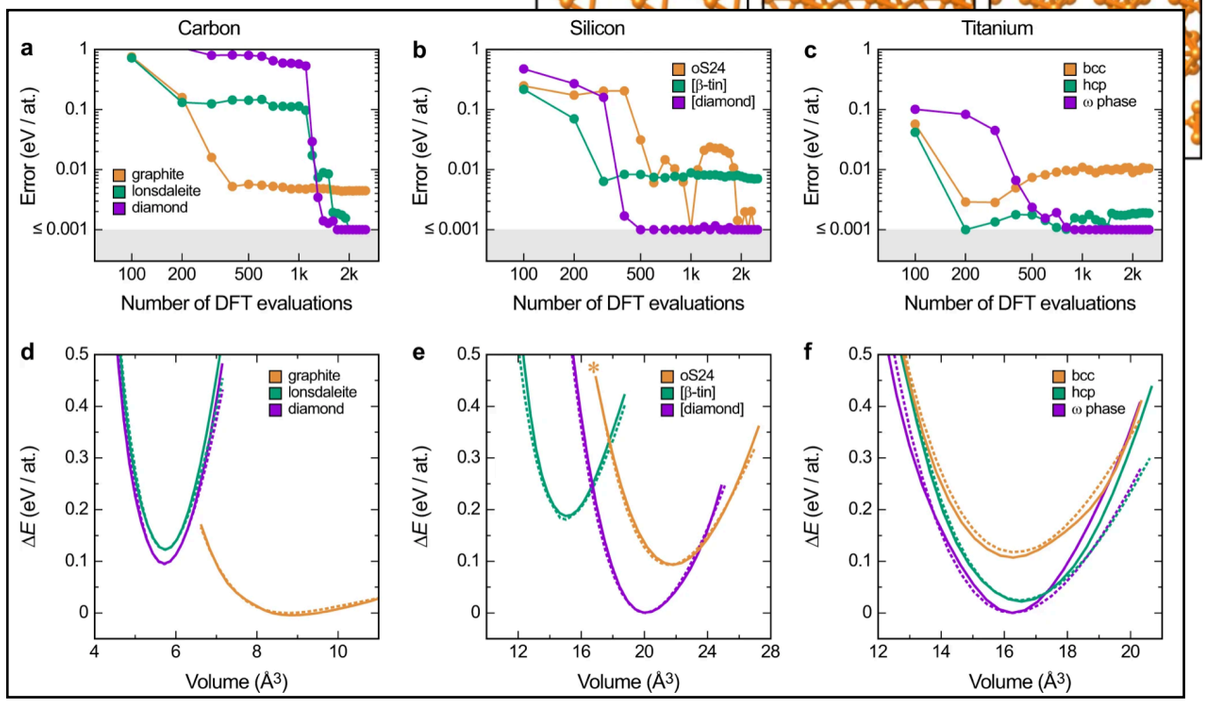




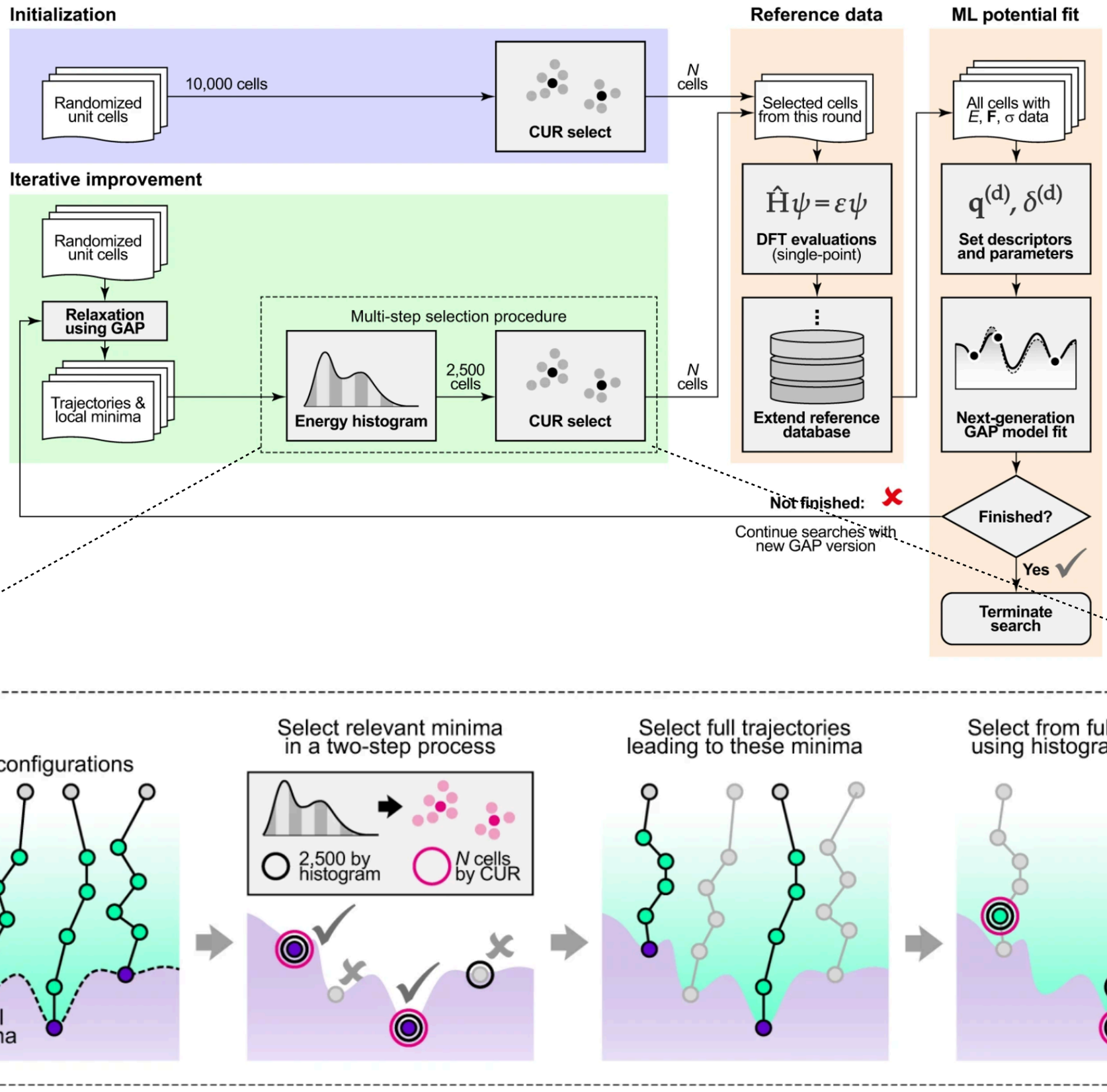






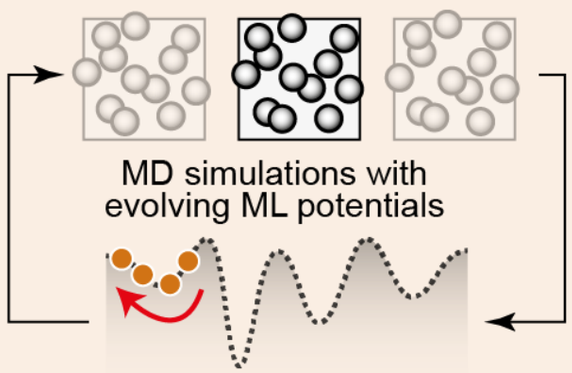


Random Structure Search workflow

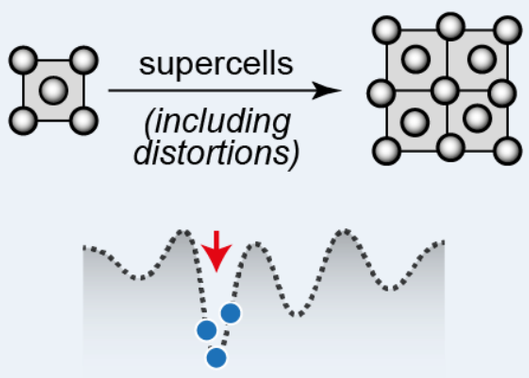


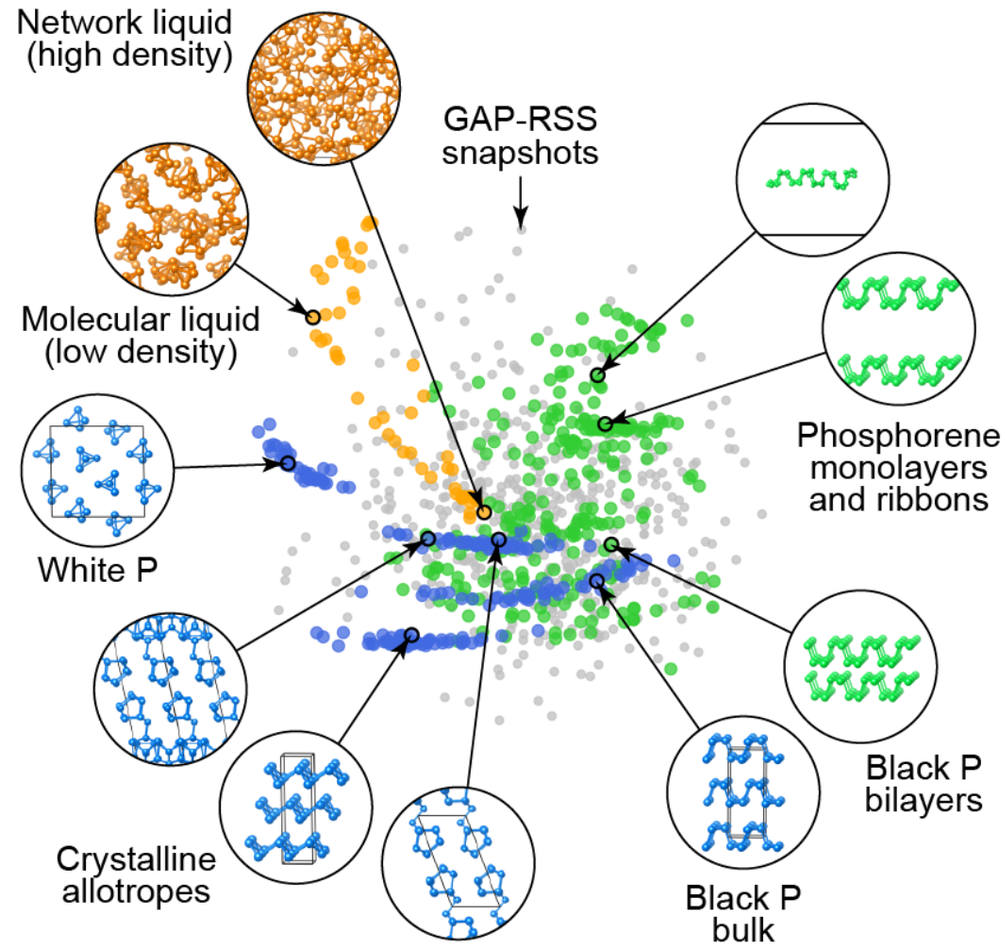
Building a general database: phosphorus

Iterative MD and fitting

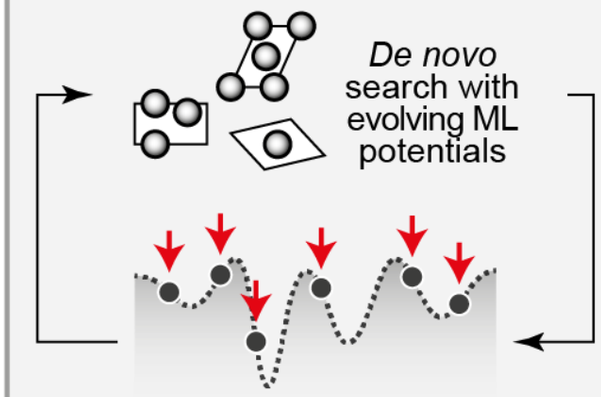


Crystallographic data

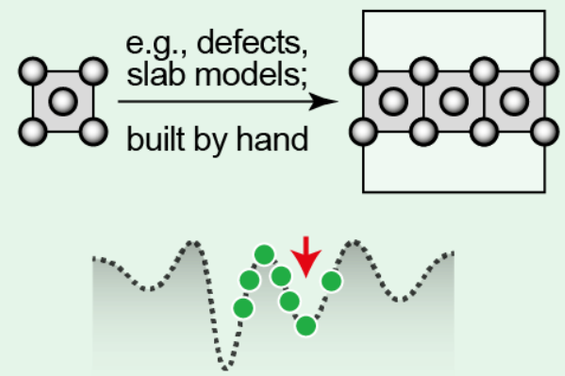


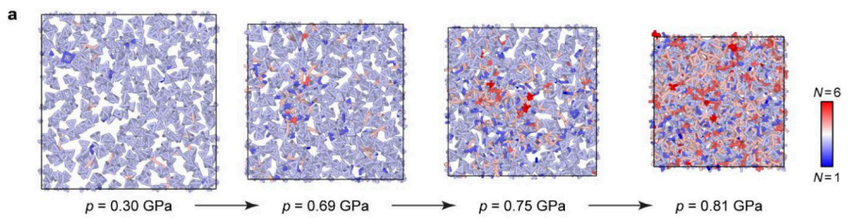


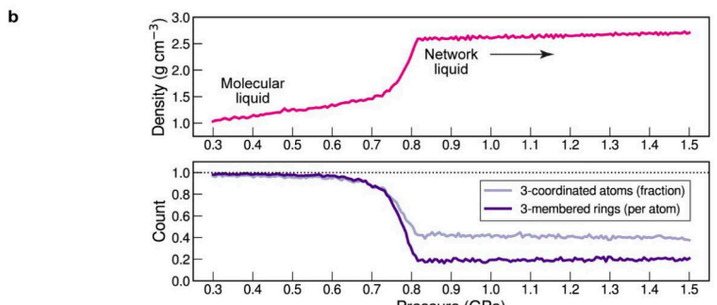
Random structure search (RSS)

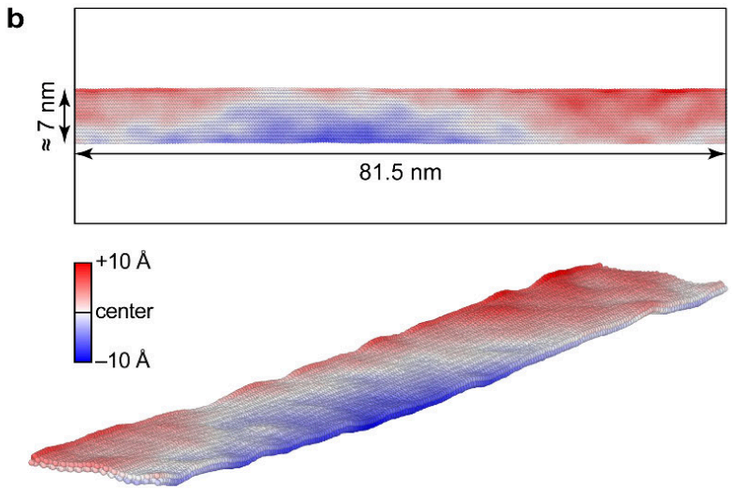


Manual construction

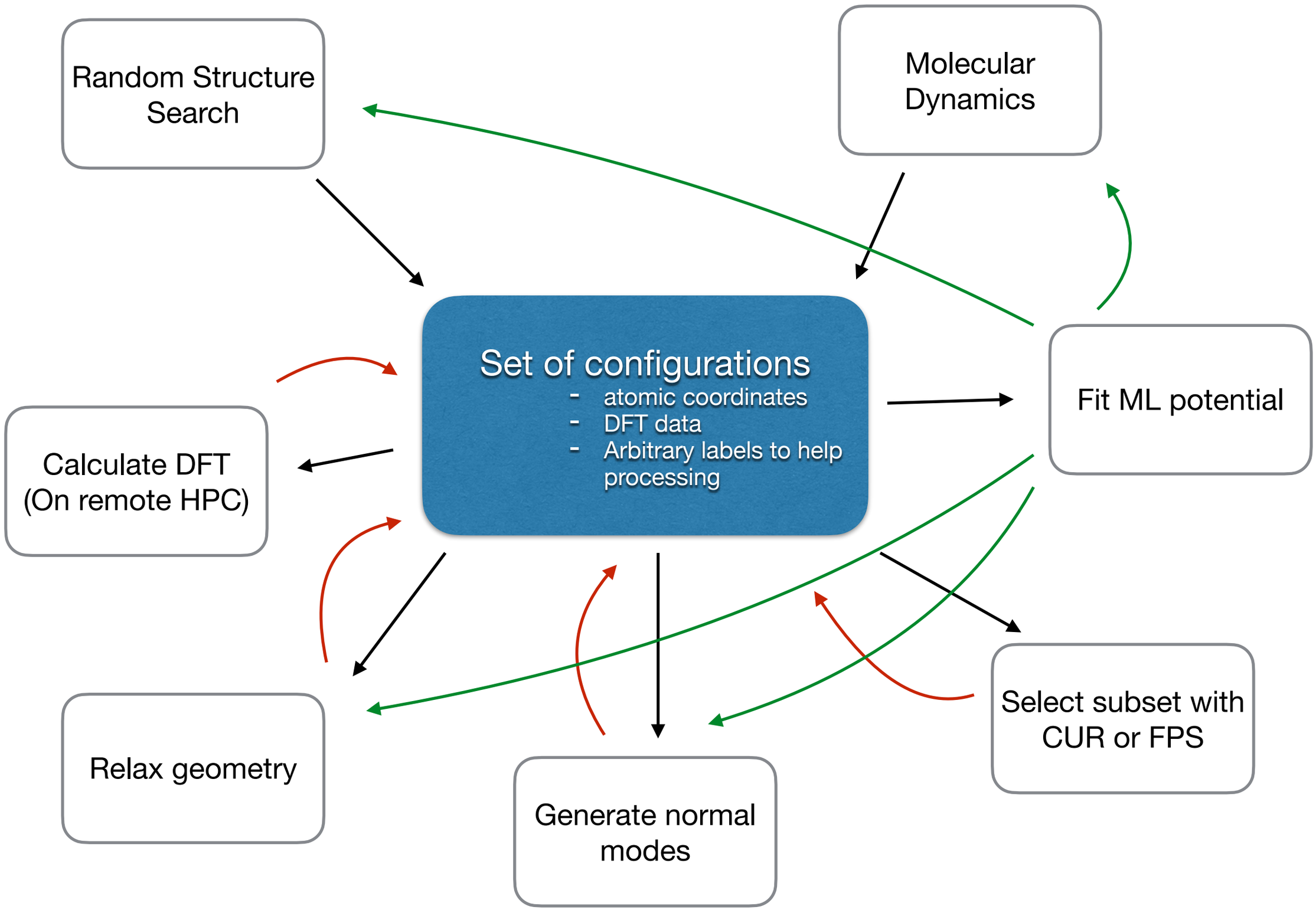








Example tasks



Workflow implementation: requirements

- Use ASE as glue: all calculators ready
- Modular: some steps still manual
- Heterogeneous computing environment
  - Laptop for development: frequently changing package stack
  - Local cluster with moderate resources: full control of stack
  - National HPC (large resource): no control over stack, only ssh

AiiDA
Automated Interactive Infrastructure and Database for Computational Science

✗ Heavy-weight, no ASE

FireWorks

✗ requires dbase on HPC