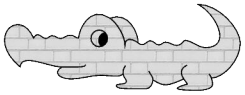


TCM

UNIVERSITY OF BIRMINGHAM

UNIVERSITY OF CAMBRIDGE

matador

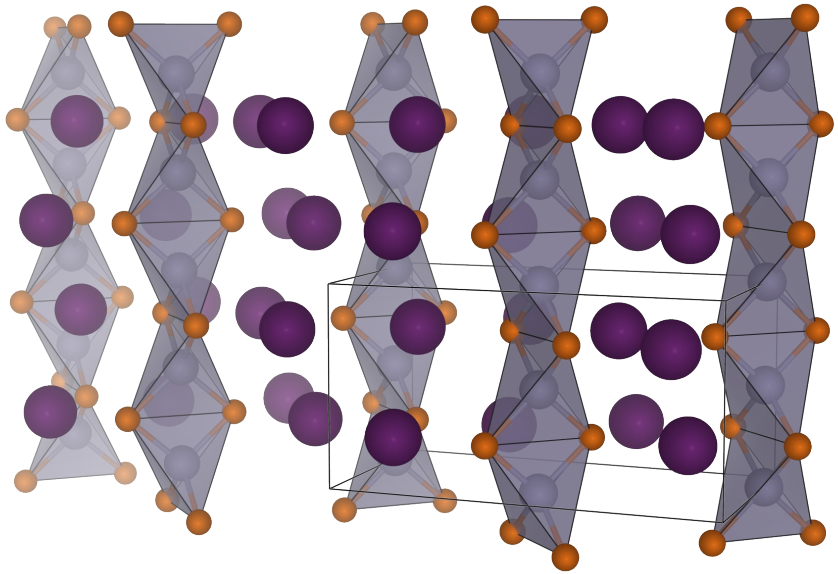
{Databases & Crystal Structure Prediction} in the Andrew Morris Group

Matthew L Evans
University of Cambridge

http://ml-evs.science

What flavour of crystal structure prediction?

- *Ab initio* random structure searching (AIRSS) provides an unbiased way of sampling configuration space.
- Supplemented by data mining, prototyping, evolutionary algorithms, etc.
- Focus on alloying/conversion anodes for batteries, where dense sampling of metastable phases is important.
- A more niche focus: encapsulated nanowires
- ~100 DFT relaxations per stoichiometry, ~100 stoichiometries per phase diagram.



Spec sheet

- ~850k geometry optimisations targeted towards alloying anodes for Li and beyond-Li batteries
- Native CASTEP 18, 17, 16, 9, 8… 7… 6
- Messy, nonuniform historical data
- Big phase diagrams: some as large as 100k calculations.
- Should fit CLI workflow of group, and be usable for new members
- Improved data mining capabilities
- Selectively open data, but either open to all or closed; no auth

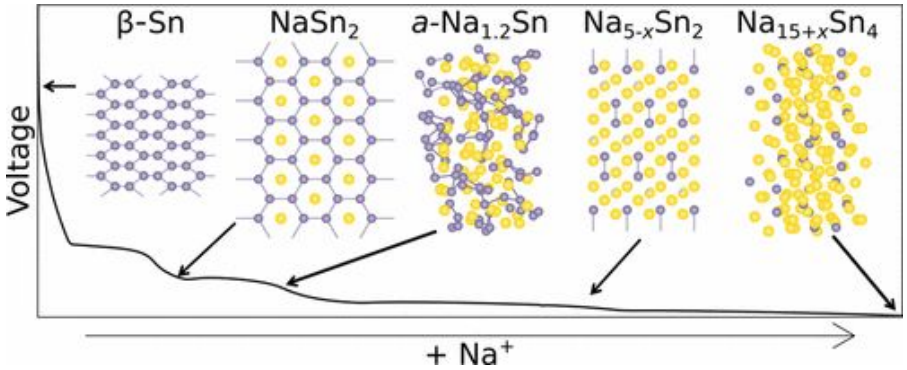


Figure 1: Na-Sn voltage curve adapted from Ref [1].

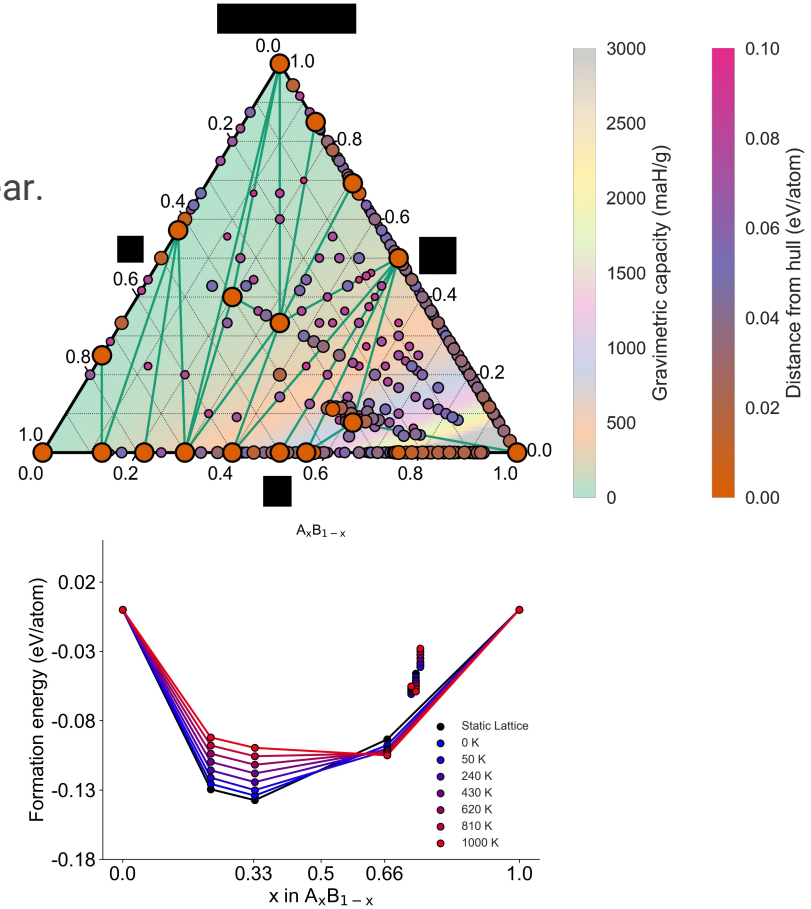
[1] P. K. Allan, et al. *J. Am. Chem. Soc.* **138** (2016)

https://matador-db.readthedocs.io

matador

- Open-source, Python 3.6, bus factor increasing by 1 per year.
- MongoDB backend.
- Fundamental sub-commands:
  - `import` data to local database, on-the-fly
  - `query` with powerful macros and DFT parameter matching
  - `hull` construct phase diagrams and make pretty plots, + ability to track changes over time (as searches progress)
  - `swap` leveraging OQMD, materials project and ICSD for structural prototypes
  - `run` high-throughput relaxations, electronic, vibrational calculations with CASTEP or QE.

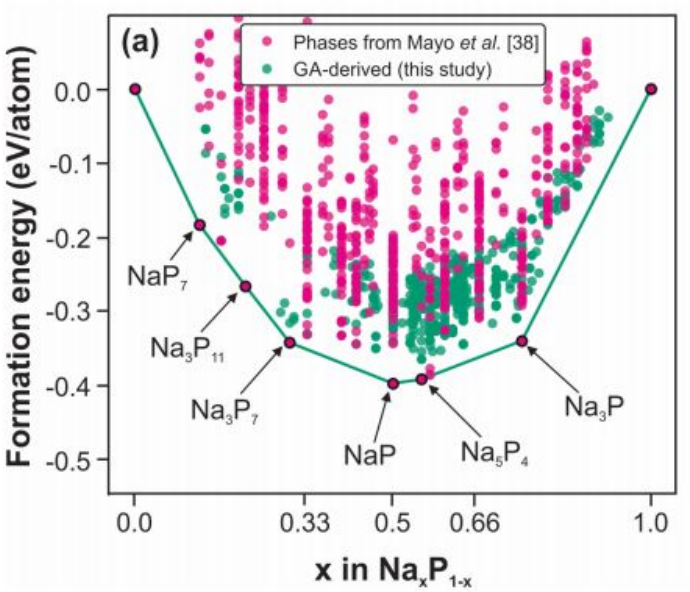
- **Somewhere between fully-automated luxury DFT and handle-turning monkey.**

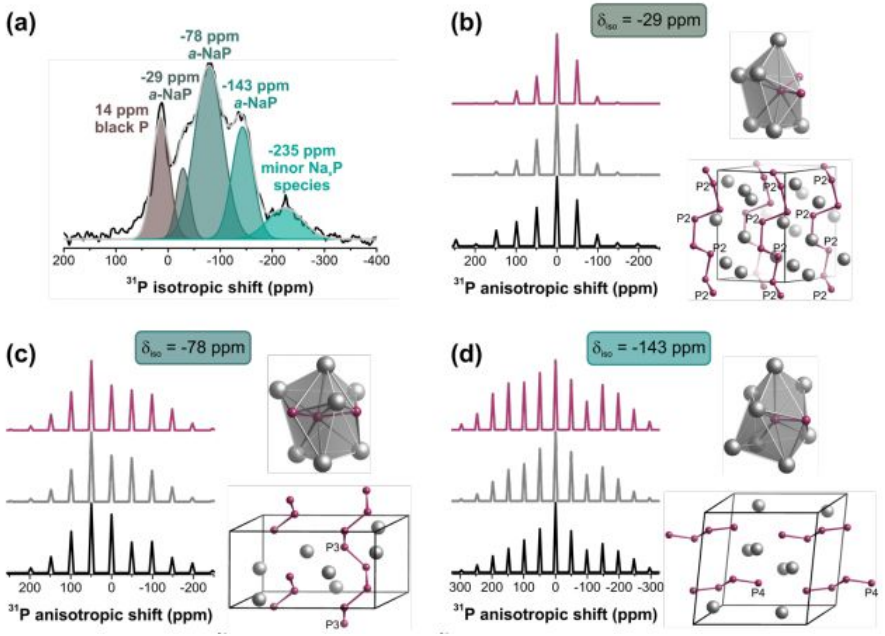


Python API for “Science”

- Structure analysis and descriptors:
  - Voronoi for local structure, others planned (SOAP, BP symmetry functions)
  - Crystal graphs for global structure
- Screening/fitting against experimental NMR, PDF, XRD, electrochemistry
- WIP: Finite-temperature phase diagrams and non-equilibrium electrochemistry
- WIP: Training of interatomic potentials
  - Overall ~850e3 geometry optimisations => ~50e6 “snapshots” of forces and energies
  - At least 10,000 per phase diagram; can learn on-the-fly
- Genetic algorithm running from database: separate open source code called `ilustrado`.

Case study: P helices in disordered NaP





[2] *(De)Sodiation via Helical Phosphorus Intermediates in High Capacity Anodes for Sodium-ion Batteries.* L. E. Marbella, M. L. Evans, M. Groh, J. Nelson, A. J. Morris, C. P. Grey, JACS (*In print*) (2018)

OPTiMaDe in matador 1.0

- Query selected OPTiMaDe databases from Python, using matador's own query structure.
  - Will implement two-way conversion between filter languages.
  - Will typically only query relaxations, preferably with quality tolerance.
- Host (at least) our own published data with OPTiMaDe REST-API (both internally and externally).
- **Provenance tracking**: we re-relax an OQMD prototype based on an ICSD phase?
  - Currently we would track both under `provenance: {oqmd_id: 1234, icsd_id: 5678}`,
  - List of DOIs enough? How many allowed per entry?
- **Discoverability**: we have many results on few systems
  - Access via e.g. http://matador.science/optimade or http://optimade.org/matador, should lead to same interface (c.f. Skyscanner comment)
  - Explorability? (e.g. wow, where else can I fly with this company?)
- **Calculation quality/type**: phase stability and interdependence of structures & calculations.
- **Fuzzier searches**: discussed in part yesterday,
  - `elements HAVE ALL ['[III]', '[V]'], n_elements IS 2`
  - Give me all A_2BO_3 ?

Thanks for listening!

EPSRC
Engineering and Physical Sciences Research Council

Andrew J. Morris
University of Birmingham

matador's (ab)users
In the AJM group

Computing:

- Tier-1: ARCHER
- Tier-2: CSD3, HPC Midlands +
- CFN Cluster Brookhaven National Labs

matador-db.readthedocs.io
www.andrewjmorris.org
ml-evs.science