

AI in Crystallography: Mathematics exposes Artificial Illusions in material design

Tatiana Kurlina, University College London, tatiana.kurlina.25@ucl.ac.uk

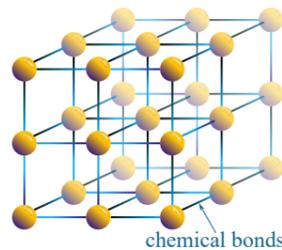
Introduction

Crystallography is the study of solid crystalline structures; one of the goals is to design new useful materials. AI has been generating a large amount of simulated crystals. It is important to analyse new creations to see which are the same and which are different.

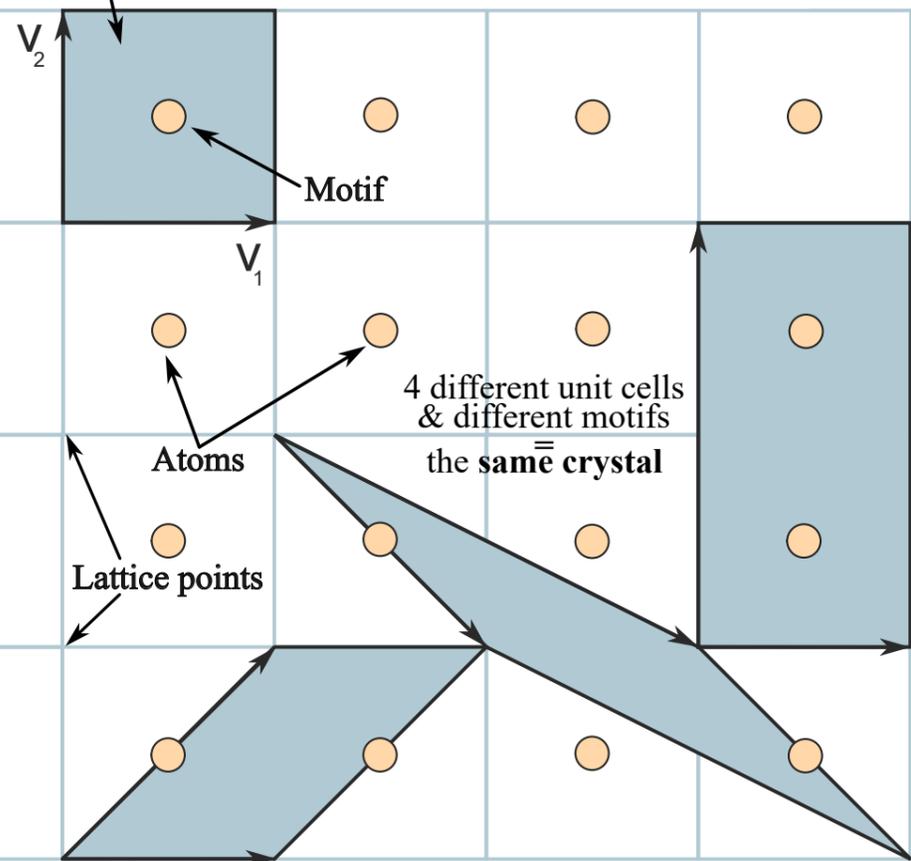
What is a crystal?

A crystal is a periodic arrangement of atoms. To mathematically construct a crystal, pick:

- 1) independent vectors $\{v_1, \dots, v_n\}$ of \mathbb{R}^n
- 2) a motif (a set of points: atoms)



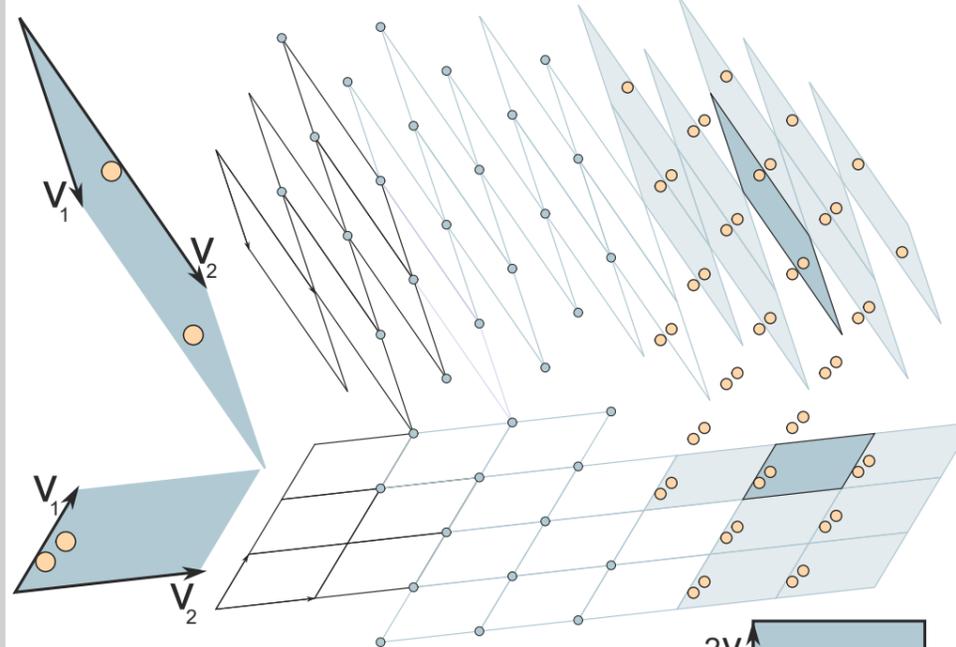
Unit Cell: The parallelepiped $U(v_1, \dots, v_n)$ consisting of all linear combinations $\sum_{i=1}^n t_i v_i$ with real coefficients $t_1, \dots, t_n \in [0, 1)$.



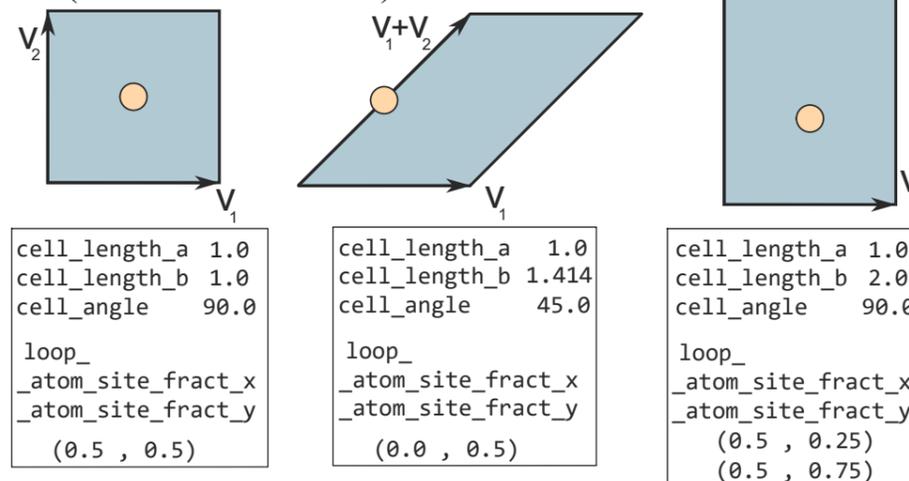
Motif: $M \subset U(v_1, \dots, v_n)$ is a finite set M of points.
Lattice: $\Lambda(v_1, \dots, v_n)$ consists of all linear combinations $\sum_{i=1}^n c_i v_i$ with integer coefficients $c_1, \dots, c_n \in \mathbb{Z}$.
Periodic Point Set: $S = M + \Lambda(v_1, \dots, v_n)$ is the set of points $p + v$ for all $p \in M$ and $v \in \Lambda$.

Which crystals are the same?

2 different unit cells + different motifs = the **same crystal**

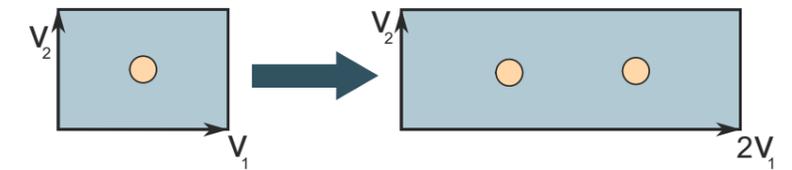


Databases store crystals in CIFs, Crystallographic Information Files containing the unit cell and motif (fractional coordinates)

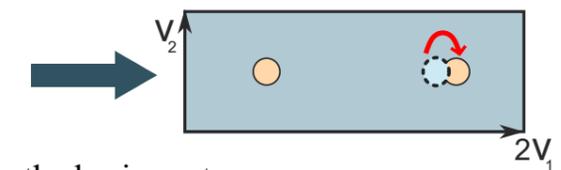


Designing your own Illusions

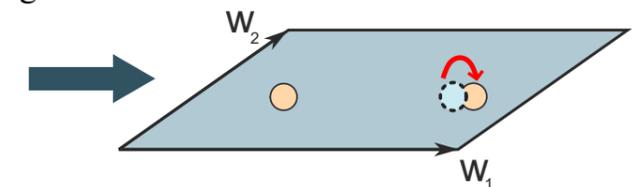
1) extend the unit cell



2) slightly shift some atoms in the motif



3) change the basis vectors



A simple Python app generating millions of new-looking crystals:

```
data_disguised
cell_length_a 1.117625288
cell_length_b 1.134880584
cell_length_c 9.67891891
cell_angle_alpha 19.47122063
cell_angle_beta 54.73561032
cell_angle_gamma 65.90515745
cell_volume 349.00260458

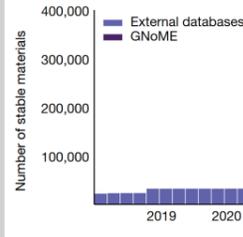
loop_
_atom_site_label
_atom_site_type_symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
Na+1 Na+ 0.99965316 0.0003467 0.99803752
Na+2 Na+ 0.29977722 0.50092328 0.4900313
Na+3 Na+ 0.99923756 0.49934218 0.00185047
Na+4 Na+ 0.25014822 0.9978209 0.50366005
Cl-1 Cl- 0.24966755 0.00138265 0.99773412
Cl-2 Cl- 0.9996186 0.50025784 0.50063548
Cl-3 Cl- 0.34923238 0.50000991 0.00012084
Cl-4 Cl- 0.00042397 0.99941199 0.49998203
Na+5 Na+ 0.50031141 0.000409 0.99937759
Na+6 Na+ 0.74916051 0.50053112 0.50012889
Na+7 Na+ 0.50025432 0.49917051 0.00115204
Na+8 Na+ 0.75014703 0.00129511 0.49773894
Cl-5 Cl- 0.75033765 0.00160787 0.99747864
Cl-6 Cl- 0.49993941 0.5006407 0.49900401
Cl-7 Cl- 0.76028128 0.50203556 0.99758844
```

A periodic structure is an *equivalence class* of periodic point sets $S \subset \mathbb{R}^n$ under rigid motion

which can be efficiently recognised by **complete invariants** I (such that $I(S) = I(Q)$ if and only if $S \cong Q$) with Lipschitz continuous metrics.

Artificial Illusions

Google's Deep Mind's AI tool GNoME claims to have discovered 2.2 million new crystals: "equivalent to nearly 800 years' worth of knowledge".



However, there is no guarantee that the the crystals they found are different or in fact real.

E.g. a group of 9 CIFs contain equal numbers but the chemical composition differs by only 1 or 2 atoms.

Conclusion

A rigorous geometric approach to crystal design suggests that the AI is often more Illusions than Intelligence. Equivalence classes, invariants and continuous metrics are required to compare crystals.

References

Anosova et al. The importance of definitions in crystallography. IUCrJ, v.11 (4), p.453-463, 2024, doi:10.1107/S2052252524004056
 Merchant et al. Scaling deep learning for materials discovery. Nature 624, 80-85 (2023) <https://doi.org/10.1038/s41586-023-06735-9>
<https://deepmind.google/blog/millions-of-new-materials-discovered-with-deep-learning/>