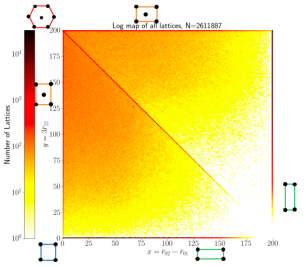
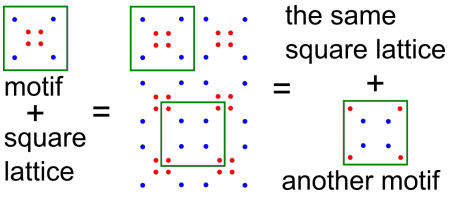
CRISP : the Crystal Isometry Principle

Andy Cooper and Vitaliy Kurlin's Data Science group
Materials Innovation Factory, University of Liverpool



What is a *periodic crystal S*?

S is often defined by periodically translating a motif of atoms along basis vectors of a unit cell. Infinitely many pairs (basis, motif) give identical structures up to *rigid motion* (compositions of translations and rotations) or *isometry* (with reflections).



Crystal structures are determined in a rigid form. Hence the most practical *equivalence* is rigid motion or isometry.

A **new** definition of a *crystal*

A periodic crystal is not a single periodic structure but an *isometry class* of infinitely many periodic point sets that are all isometric to each other [1]

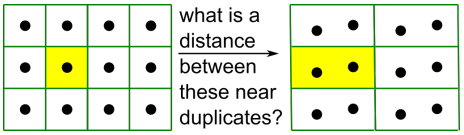
$$\textbf{Crystal}=\frac{\text{infinitely many structures}}{\text{equivalence up to isometry}}$$

The weaker equivalence by symmetry (the same space-group type) gives 219 classes (230 if mirror images are distinguished), insufficient to distinguish 1.1M+ crystals in the Cambridge Structural Database (CSD).

What is a crystal *invariant*?

An *invariant* is a property preserved by equivalence (isometry of crystals), eg symmetry, primitive cell volume.

Non-invariants (parameters of non-reduced cells, atomic coordinates) cannot reliably distinguish crystals as they can differ for the same crystal.



Can reduced cells help? No!

A reduced (Niggli's) cell is invariant but discontinuous under atomic displacement. Under tiny perturbations, the symmetry breaks down, the primitive cell volume doubles, see above.

Mapping crystals problem

Find a complete invariant I for all crystals with a continuous metric:

invariance : $S \cong Q$ are isometric $\Rightarrow I(S) = I(Q)$, so *no false negatives*;

completeness : $I(S) = I(Q) \Rightarrow S \cong Q$ are isometric, hence *no false positives*;

metric : $d(S,Q)$ satisfies the axioms 1) $d(S,Q) = 0 \Leftrightarrow S \cong Q$ are isometric, 2) $d(S,Q) = d(Q,S)$, 3) $\triangle$ inequality $d(S,Q) \le d(S,T) + d(T,Q)$.

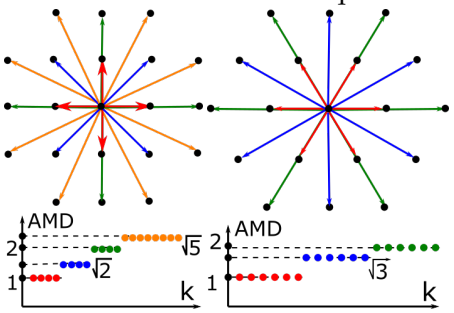
continuity : the metric d continuously changes under perturbations of S;

inverse design : a crystal S can be reconstructed from an invariant value.

There was no complete and continuous invariant I even for lattices [2].

AMD invariants of crystals

For any p_i (one of m motif points) in a unit cell of a crystal S, let d_{ik} be the distance to its k-th closest neighbour in the infinite S. The Average Minimum Distance [1] is $\mathrm{AMD}_k = \frac{1}{m}\sum_{i=1}^{m} d_{ik}$. The square and hexagonal lattices have these AMD sequences:

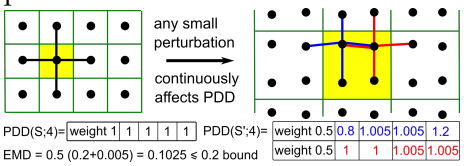


Stronger PDD$(S;k)$: Pointwise Distance Distribution

For any motif point p_i, put its distances $d_{i1} \le \cdots \le d_{ik}$ into a row of the $m \times k$ matrix. If j of m rows are identical, collapse them into one row of *weight* j/m. The matrix $\mathrm{PDD}(S;k)$ is independent of a crystal representation. Increasing k adds more columns without changing the initial columns.

PDD: continuous invariants

If atoms are perturbed up to ε, then $\mathrm{PDD}(S;k)$ changes up to 2ε in Earth Mover's Distance, which can compare PDD matrices of different sizes.



PDD(S;4)=						PDD(S';4)=					
	weight	1	1	1	1		weight	0.5	0.8	1.005	1.005	1.2
							weight	0.5	1	1	1.005	1.005

EMD = 0.5 (0.2+0.005) = 0.1025 ⩽ 0.2 bound

PDD: generically complete

Under tiny perturbation, any crystal becomes *generic*, e.g. no repeated distances except due to periodicity.

Any such crystal is uniquely reconstructed from lattice invariants [2] and $\mathrm{PDD}(S;k)$ for a large enough k.

200B+ pairwise comparisons of AMD, PDD for all 660K+ periodic crystals (no disorder, full 3D structure) in the CSD over two days on a modest desktop detected five pairs of suspicious entries with identical geometry and one atom replacement [1, section 7]:

HIFCAB and JEPLIA (Cd ↔ Mn),

LALNET and POCPAA (Cd ↔ Ni),

AFIBOH and NENCUF (Cd ↔ Zn),

COLYEI and POCLOK (Eu ↔ Sm),

DTBIPT and DTHBPD10 (Pt ↔ Pd).

Five journals are investigating the integrity of the underlying publications.

Crystal Isometry Principle

Map: periodic crystals → periodic point sets is *injective* modulo isometry. Any periodic crystal is determined by geometry of its atomic centres (without chemical labels) because replacing one atom with a different one should perturb distances to atom neighbors.

All known and undiscovered crystals live in the common *Crystal Isometry Space* parameterised by invariants.

[1] D.Widdowson et al. Average Minimum Distances of periodic sets. MATCH Comm. Math. Computer Chemistry, v.87(3), p.529-559, 2022.

[2] V.Kurlin. Mathematics of 2D lattices. arxiv:2201.05150.

Periodic Geometry for applications in Crystallography and Materials Design

The vision is to resolve the long-standing challenge of targeted design for solid crystalline materials by mapping the space of all periodic crystals using rigorous mathematical methods.

The key obstacle is the ambiguity of conventional crystal representations by primitive or reduced cells that are discontinuous under atomic displacements. Without continuously quantifying the similarity of periodic crystals, the brute-force Crystal Structure Prediction produces millions of nearly identical approximations to numerous local energy minima, see Fig. 1.

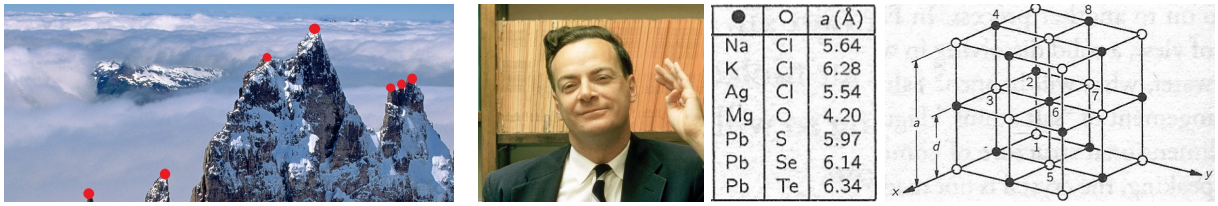


Figure 1: **Left**: energy landscapes show crystals as isolated peaks of height= −energy. To see beyond the 'fog', we need a map parameterised by invariant coordinates with a continuous metric. **Right**: thanks to R. Feynman's hint, all periodic crystals live in the common *Crystal Isometry Space* (CRISP) parameterized by distance-based invariants without atomic types.

The mapping problem is non-trivial already for 2D lattices whose continuous space up to rigid motion with uniform scaling is the square with identified sides or a punctured 2D sphere. Among 2.6M+ 2D lattices in 870K+ crystals (with full lattice data) from the Cambridge Structural Database, about 45% are oblique without extra symmetry and continuously fill the square apart from the bottom right corner representing too long and narrow cells, see Fig. 2.

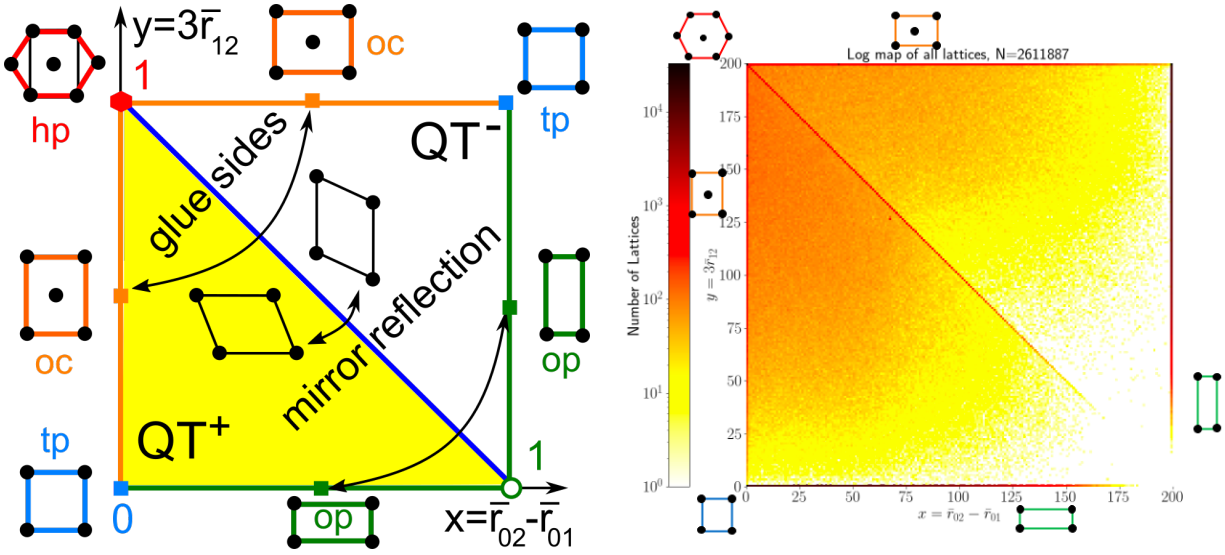


Figure 2: **Left**: any 2D lattice up to rigid motion and uniform scaling has a unique obtuse superbase with $v_0 + v_1 + v_2 = 0$, *root products* $r_{ij} = \sqrt{-v_i \cdot v_j}$, and complete invariant (x, y) in the unit square. **Right**: the log-scale density map of all 2D lattices in CSD crystals.

Extensions to 3D lattices and periodic crystals are progressing fast. If you would like to visualise your crystal datasets, we are open to collaboration, e-mail vitaliy.kurlin@gmail.com. All latest papers are linked at http://kurlin.org/research-papers.php#Geometric-Data-Science.