Introduction to Periodic Geometry for applications in Materials Discovery

Vitaliy Kurlin’s group including
Olga Anosova, Phil Smith, Marco Mosca,
Matthew Bright, Daniel Widdowson, ... at
Materials Innovation Factory (MIF), Liverpool
Objects: all periodic crystals

Solid crystalline materials (periodic *crystals*) can have many types, all consist of elementary blocks (*motifs*) of atoms, ions or molecules in a *unit cell* periodically repeated in three directions.
A lattice is defined by a basis

Given any linear basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$, a lattice is $\Lambda = \left\{ \sum_{i=1}^{n} c_i v_i : c_i \in \mathbb{Z} \right\}$. Any lattice $\Lambda$ can be generated from infinitely many different bases.

Any lattice $\rho + \Lambda = \left\{ \rho + v : v \in \Lambda \right\}$ shifted by a vector (or obtained by rotation) is equivalent to the original lattice $\Lambda$ due to rigidity of crystals.
A periodic set = lattice + motif

Any linear basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ spans the parallelepiped $U = \left\{ \sum_{i=1}^{n} c_i v_i : c_i \in [0, 1) \right\}$ called a unit cell. A motif is a finite set of points $M \subset U$.

A periodic point set is the sum of a lattice and a motif: $S = M + \Lambda = \left\{ u + v : u \in M, v \in \Lambda \right\}$.
Our input: crystal = (cell, motif)

A crystal is usually given by a Crystallographic Information File (CIF) with parameters of a unit cell (coordinates of a basis $v_1, v_2, v_3$ or three edge-lengths and three angles) and fractional coordinates $(x, y, z)$ of atoms the cell basis.

We model a crystal as a periodic set of points at atomic centers, which are better defined than chemical bonds. Chemical types of atoms or bonds can be added but may not be needed.
Crystals up to isometries

Crystal structures are determined in a **rigid** form and can be studied up to *rigid motion*. An *isometry* of $\mathbb{R}^n$ is any map that preserves all distances between points. A *rigid motion* also preserves an orientation of $\mathbb{R}^n$, e.g. a reflection is not a rigid motion. Rigid motions in $\mathbb{R}^3$ are compositions of translations and rotations.
What is a crystal mathematically?

Since geometrically different periodic point sets (CIF) can represent equivalent crystals, define:

a **crystal** is an **isometry class** of infinitely many periodic point sets equivalent up to isometry.

These four lattices and infinitely many others form a **single isometry class** of square lattices.
Isometry invariants of crystals

An invariant (number, vector, matrix, ...) must take the same value on all isometric crystals.

crystal input = cell+motif, invariant: isometric crystals $\rightarrow$ one value

If a non-invariant takes two different values on two crystals, then no conclusion can be made.

Question: how about non-invariant big data?
Answer: use invariants.
No easy reduction to finite sets

From an infinite periodic crystal, a finite portion is selected, say in an extended cell or a ball of cut-off radius, missing points near boundaries.

Isometric crystals $\not\rightarrow$ non-isometric finite sets.
A partial match of molecules

Experimental T2-δ is overlaid with closest simulated version. RMSD irregularly grows if the COMPACK algorithm matches more molecules.

<table>
<thead>
<tr>
<th># matched molecules</th>
<th>5 of 5</th>
<th>8 of 10</th>
<th>10 of 15</th>
<th>11 of 20</th>
<th>16 of 25</th>
<th>18 of 30</th>
<th>21 of 35</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD, 1Å = 10^{-10}m</td>
<td>0.603</td>
<td>0.681</td>
<td>0.812</td>
<td>0.825</td>
<td>0.99</td>
<td>1.027</td>
<td>1.079</td>
</tr>
</tbody>
</table>
If the orthogonal projection of $v_2$ to the line of $v_1$, is outside $[-0.5|v_1|, 0.5|v_1|]$, add multiples of $v_1$ to make $v_2$ shorter. The cell $U$ is reduced to $U'$ or $U''$. 

One can exclude one of the projected endpoints $\pm 0.5|\vec{v}_1|$. This choice cuts the space of lattices. Any choice makes close lattices look distant similarly to measuring angles, for example angles 0, 359 are distant in the range $[0, 360)$. 

Unique Niggli’s reduced cell
Discontinuity of most invariants

Even if a cell is reduced (Niggli’s cell), any such reduction is discontinuous under perturbations.

All discrete invariants such as symmetry groups are discontinuous. Pseudo-symmetry depends on a threshold.

The primitive cell volume is discontinuous, the density is constant. Why is continuity important? All atoms vibrate, real measurements are noisy.
CSP: Crystal Structure Prediction

Aim: discover crystals with target properties, predict stable crystals obtained by embedding several molecules or atoms into a unit cell.

The stability is determined by an energy, which depends on atomic positions and charges.
Embarrassment of over-prediction coined by Prof Sally Price FRS (UCL) says that the state-of-the-art CSP software outputs too many approximate local minima of the energy: random start, perturb atoms, compute again, ...

The plot in Nature 2017 shows 5679 predicted crystals of T2 molecules, only 5 were synthesised. Each crystal is represented by (density, energy), insufficient to completely map a crystal space.
Isometry classification problem

We need a complete and continuous isometry \textit{invariant} of crystals $I : \{\text{crystals}\} \rightarrow \{\text{numbers}\}$

1) \textit{invariance} : if any crystals $S, Q$ are isometric, then $I(S) = I(Q)$, so $I$ should be well-defined on isometry classes of crystals, no cell parameters

2) \textit{completeness} : if $I(S) = I(Q)$, then $S, Q$ are isometric, i.e. $I$ distinguishes all crystals $S \neq Q$;

3) \textit{continuity} : the invariant $I$ slightly changes under perturbations to quantify a similarity, else we miss nearly identical crystals at local minima.
More classification requirements

4) *Computability*: a polynomial time in a motif size (the number $m$ of atoms in a unit cell).

Past: energy-vs-density plot of 5679 predicted T2 crystals took 12 weeks on a supercomputer.

5) *Inverse design*: a complete invariant should allow us to reconstruct a full 3D crystal so that we can choose a new invariant value and discover new crystals with extraordinary properties, e.g. an efficient solid-state battery.
Metric axioms and metric problem

A metric $d \geq 0$ on isometry classes of crystals:

1. $d(S, Q) = 0$ if and only if $S, Q$ are isometric;
2. symmetry: $d(S, Q) = d(Q, S)$;
3. △ inequality: $d(S, T) \leq d(S, Q) + d(Q, T)$.

The first axiom may fail for a non-complete invariant $I$ if $I(S) = I(Q)$ for non-isometric $S, Q$, then any distance between $I(S), I(Q)$ is 0.

The metric problem solves the classification: $S, Q$ are isometric if and only if $d(S, Q) = 0$. 
New continuous invariants


will be presented by Phil Smith at 13:30 today.

They extend the single-value density to functions depending on a variable radius.

+ continuous, + complete for generic crystals,
- slow (cubic time, hours on 5679 T2 crystals)
Distance-based invariants

Average Minimum Distances (AMD); D.Widdowson, M.M.Mosca, A.Pulido, V.Kurlin, A.I.Cooper; arXiv:2009.02488

will be presented by Marco Mosca at 14 today.

+ simple, + continuous, + fast (near linear time, seconds on 5679 T2 crystals), - incomplete?

Pointwise Distance Distributions (PDD) will be presented by Dan Widdowson at 14:30 today.
More complete invariants

Session MS25-5 tomorrow at 13:00-15:00.

Olga Anosova: complete and continuous isosets for any periodic point sets in any dimension.


Matt Bright: easily computable continuous metrics on lattices in dimensions 2 and 3.

Teresa Heiss: a persistence-based fingerprint.
First application: data reduction

For all crystal invariants, we have continuous distances satisfying the metric axioms.

Hence any dataset from Crystal Structure Prediction can be compressed by removing nearly identical crystals, hence saving time for simulating target physical properties.

The Cambridge Structural Database (CSD) of 1M+ crystals can be split into a smaller number of really different (not nearly identical) crystals.
Visualizations of crystal datasets
A metric map of a crystal space

Now: only peaks of height = -energy, no locations.

currently available CSP landscapes consist of only isolated dots without any metric information

transition paths between minima join the dots by distances

energy unknown barriers deep minima complete invariants map the crystal space
Summary: invariants of crystals

Periodic point sets model all crystalline solids. Continuous invariants give reliable distances between crystals having different symmetries. Complete isosets define continuous coordinates on the space of isometry classes of all crystals.

MACSMIN on 15-17 September 2021, MIF.

New collaborations are welcome: contact us.