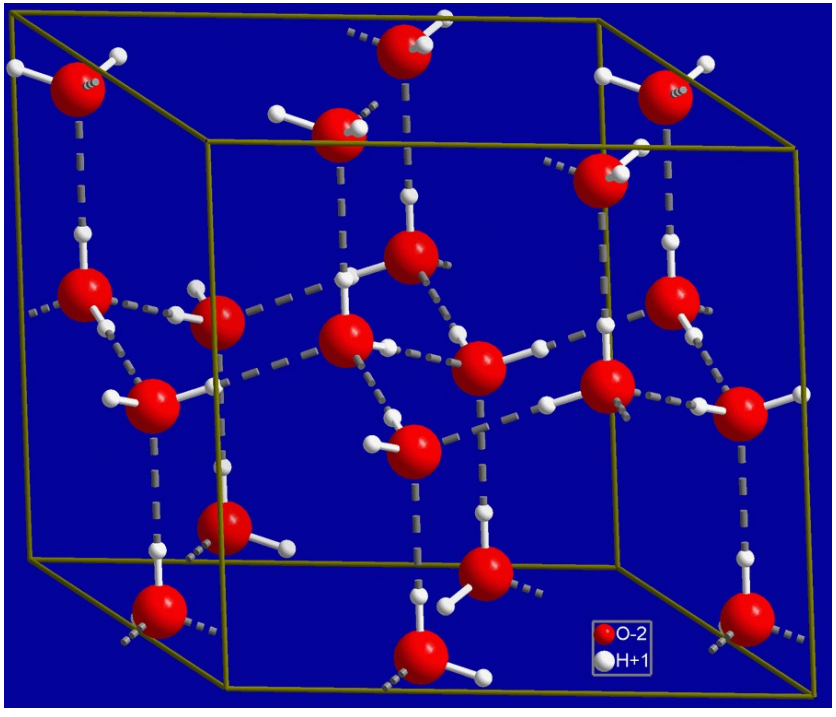
The discontinuity challenge of cell-based representations of crystals

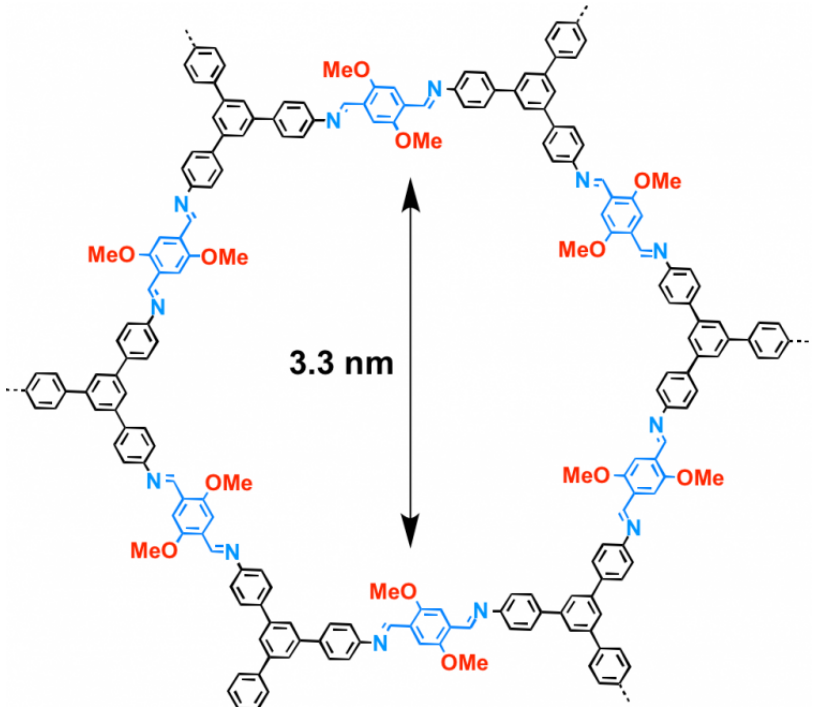
Vitaliy Kurlin's Data Science group: Dan Widdowson, Yury Elkin, Olga Anosova, ...

Materials Innovation Factory, Liverpool, UK

All types of periodic crystals

We study solid crystalline materials at the atomic level. What is a crystal on the left?



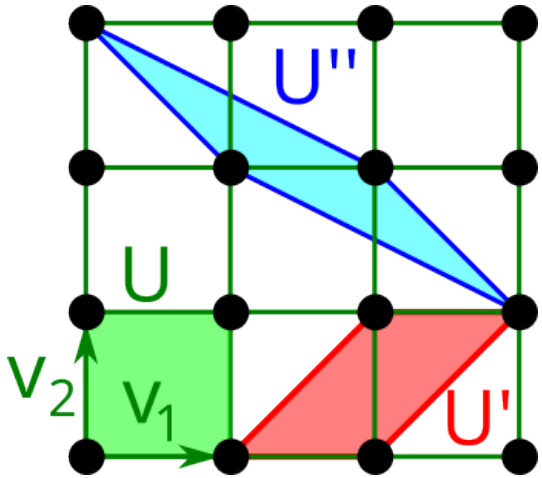


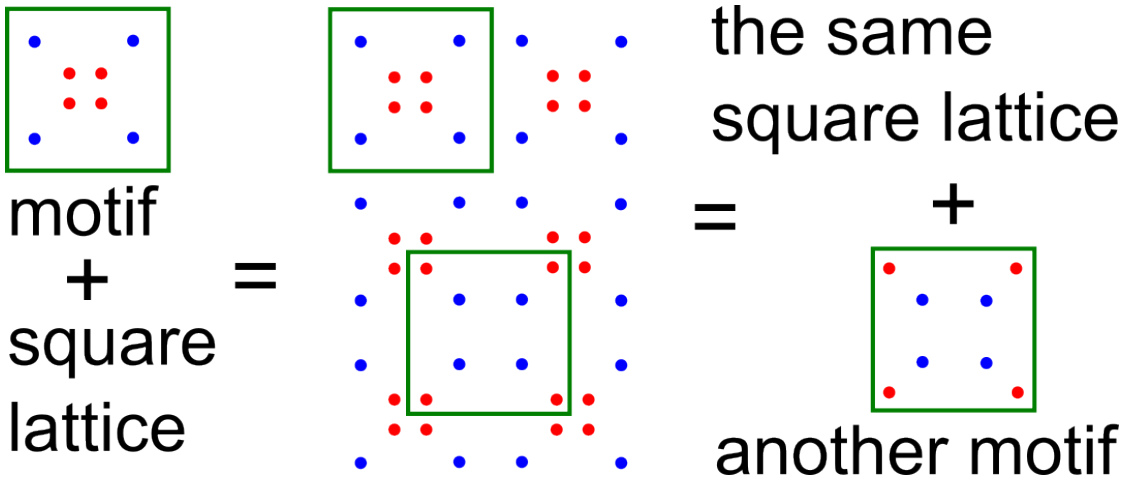
Q: What is a *crystal structure*? What structures are the *same*? If different, *how much different*?

A periodic point set (crystal)

Any basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ defines the *unit cell* U and generates the lattice $\Lambda = \{\sum_{i=1}^{n} c_i v_i \ : \ c_i \in \mathbb{Z}\}$.

For any finite *motif* $M \subset U$, the *periodic point set* is the sum $S = \Lambda + M = \{v + p \mid v \in \Lambda, p \in M\}$.





Different pairs (basis, motif) give equivalent sets.

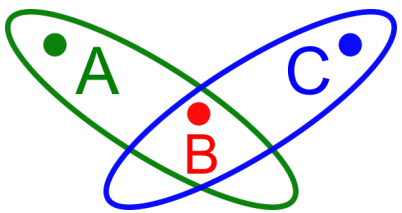
Three axioms of an equivalence

A relation $A \sim B$ between any data objects is called an *equivalence* if the three axioms hold:

(1) *reflexivity*: any object $A \sim A$;

(2) *symmetry*: if $A \sim B$ then $B \sim A$;

(3) *transitivity*: if $A \sim B$ and $B \sim C$, then $A \sim C$.



The transitivity axiom guarantees that all objects are in disjoint classes. Any justified classification needs an equivalence.

Equality is an equivalence: $0.5 = 50\% = \frac{1}{2} = 2 \div 4$

Different equivalence relations

Chemical: crystals $A \sim B$ if A, B have the same composition. Ok, but diamond and graphite with vastly different properties are in the same class.

By property: compounds $A \sim B$ if A, B have the same property. Ok, but molecules that share one property can differ by other properties.

3D space groups: crystals $A \sim B$ if A, B have isomorphic space groups. Fedorov (1891): 219 or 230 classes. Then NaCl, MgO, TiC, LaN, NaI, RbF, SrS, ... have the same group (225, $Fm\bar{3}m$).

What is the strongest relation?

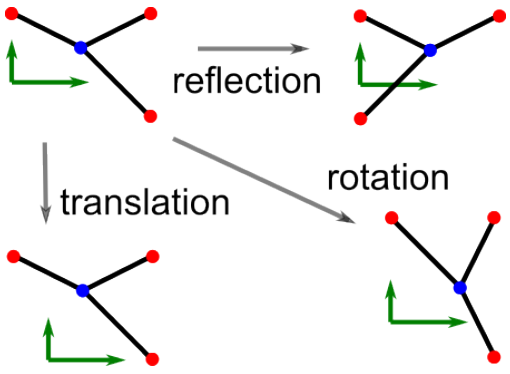
P. Sacchi et al. **Same or different - that is the question**: identification of crystal forms. CrystEngComm, 22(43), 7170-7185 (2020).

IUCr ACTIVITIES

CHANGE TO THE DEFINITION OF "CRYSTAL" IN THE IUCr ONLINE DICTIONARY OF CRYSTALLOGRAPHY

Definitions are not final without equivalence.

Definition of a crystal structure



Since crystal structures are determined in a *rigid form*, the strongest relation in practice is **rigid motion** = translations + rotations in $\mathbb{R}^3$.

Slightly weaker: **isometry** = rigid motion + reflections = any map preserving distances.

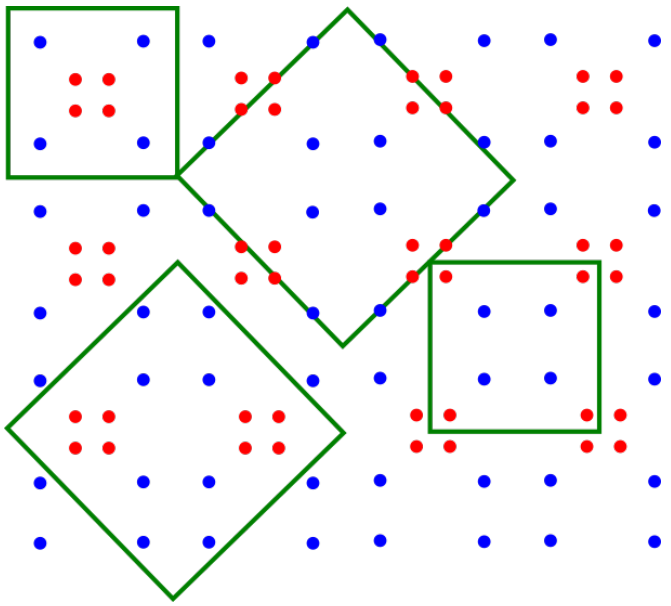
New definition by [Anosova et al, IUCrJ 2024]:

a **crystal structure** is a *rigid class* of crystals

$$= \frac{\text{infinitely many periodic crystals (CIFs) in } \mathbb{R}^3}{\textit{equivalence} \text{ under } \textit{rigid motion} \text{ (or isometry)}}$$

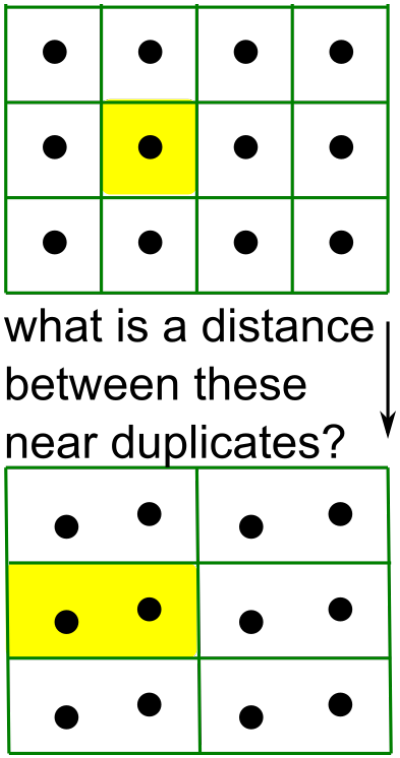
What crystals should be the same?

IUCr online dictionary: “crystals are said to be *isostructural* if they have the same structure ... CaCO3, NaNO3, FeBO3 are isostructural”.



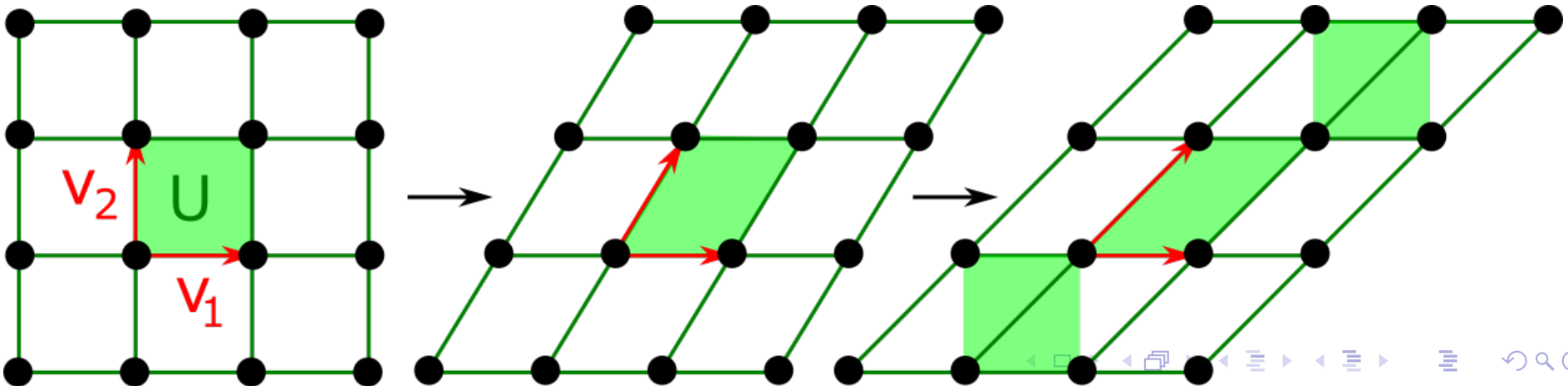
All conventional representations in the International Tables of Crystallography are *correct in theory* but are **no longer practical** because all data is noisy, and tiny displacements of atoms need very different (standard) settings.

Discontinuity of conventional cells

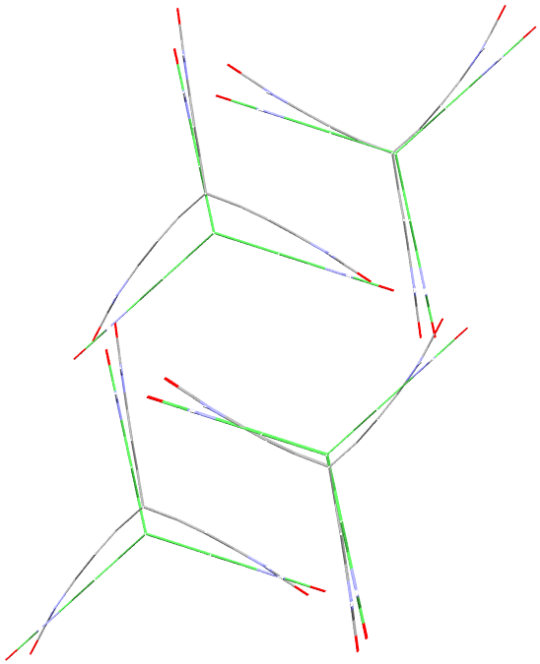


Any *reduced or conventional* cell is discontinuous under almost any noise. *Sorites paradox*: any equivalence up to a threshold $\varepsilon > 0$ implies that all objects become equivalent via a long enough sequence of ε-perturbations $S_1 \sim \cdots \sim S_k$. No noise can be ignored!

Here is a closed loop (moving v_2 to $v_2 + v_1$) in the continuous space of all 2D lattices.



Crystals live in a continuous space



All crystals consist of discretely located atoms, which have *continuous* real-valued coordinates in $\mathbb{R}^3$.

A small perturbation produces a slightly different crystal not rigidly equivalent to the original structure.

If we restrict comparisons only to a fixed space group, we cut the continuous space into disjoint pieces (230 in 3D), so many near-duplicates fall on different side of boundaries, which is tragic!

Descriptors vs invariants

An **invariant** is a map $I : \{\text{periodic crystals}\} \to$ simpler objects (numbers, vectors, matrices ...), preserved under an equivalence: all isometries.

Crystals can be distinguished by non-constant *invariants*, not by non-invariant descriptors:

if $S \simeq Q$ are *isometric*, then $I(S) = I(Q)$; or

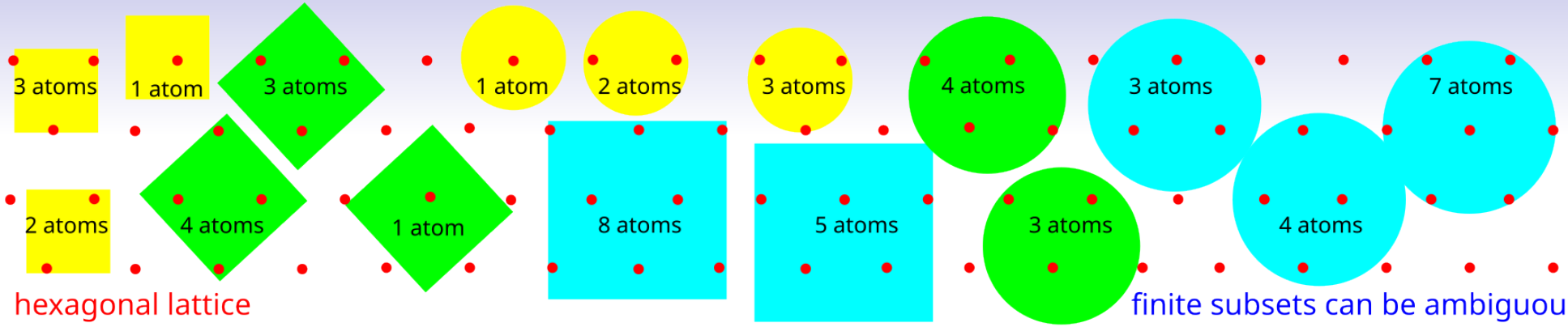
if $I(S) \neq I(Q)$, then $S \not\simeq Q$ are not isometric.

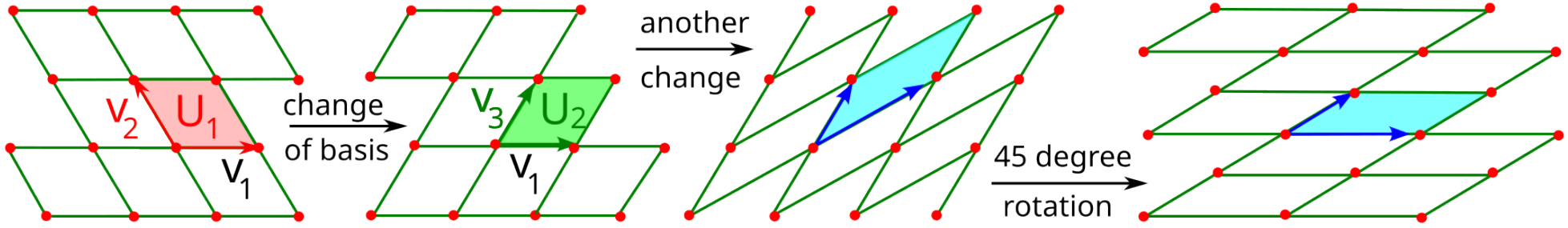
non-invariants
atomic coordinates in a cell basis, *cannot* distinguish crystals

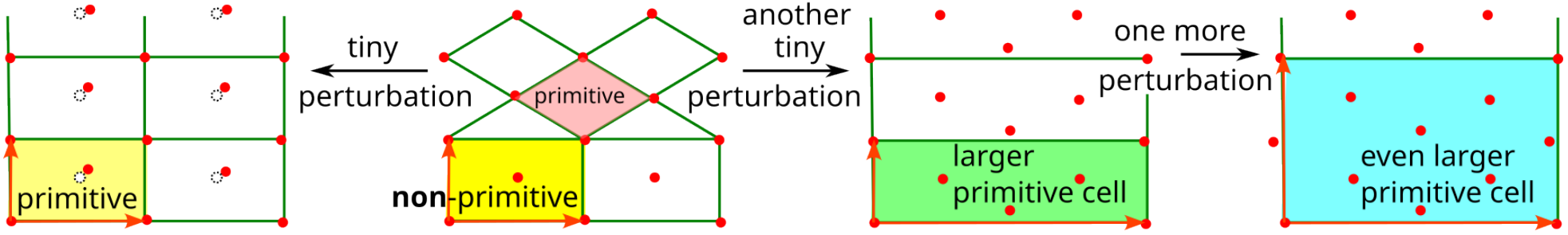
invariants can distinguish some, possibly not all crystals: density, symmetry

complete invariants, e.g. conventional representations distinguish all *in theory*

continuous **fast & reconstructable**







any periodic crystal ■ has infinitely many **near-duplicates with larger motifs**

■ the *smallest subspace* of crystals with m atoms in a fixed cell

a *larger subspace* of crystals with 2m atoms in a primitive cell

an even larger subspace of crystals with 3m atoms in a primitive cell

infinitely many layers in the continuous space of periodic crystals

Spaces and invariants of crystals

equivalences of **periodic crystals** have different *spaces of classes*

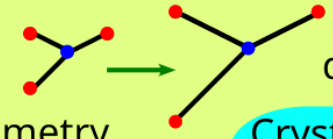
discrete spaces

by crystallographic symmetry groups → only **230 classes**

by chemical composition or structure types → still **finitely many** isolated classes

● ● ● **huge gap**

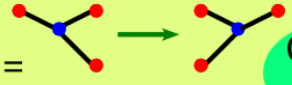
continuous spaces



all infinite dimensional

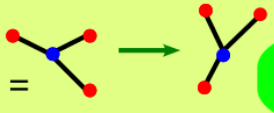
dilation = isometry + uniform scaling → Crystal Dilation Space (**CDS**)

smaller



isometry = rigid motion + reflection → Crystal Isometry Space (**CIS**)

larger



rigid motion = rotation + translation → Space of Rigid Crystals (**CRS**)

new definition: a **crystal structure** is a **class** of all **periodic sets** of atomic centers equivalent under **rigid motion**

crystal *descriptors* : how justified are they?

invariant under rigid motion?

no → *non-invariants* are ambiguous and always allow false negatives

yes → non-constant invariants *unambiguously distinguish* between some crystals

continuous?

no → can miss *near-duplicates*: space groups (1891), Niggli reduced cells (1926), structure standardizations (1987)

yes → *continuous* invariants (more practically useful) quantify the similarity

generically complete?

no → *incomplete* or *no distance:* composition, physical density, pair distribution function (PDF) Delone local theory (1976)

yes → *strong*: uniquely identify almost any periodic set

fully complete?

no → *gen.complete* densities (2021), *ultra-fast* Pointwise Distance Distribution (PDD, 2022)

yes → *strongest DNA-style*: distinguish between all periodic structures

exact distance metrics?

no → *complete* isosets (2021), approximate metrics (2025)

yes → invertible **crystal rigid code** (CRC, 2025)

Geo-mapping problem: crystals

Find a (complete, bi-continuous, and poly-time) *geocode* I for periodic sets of *unordered points*.

Invariance: if crystals $S \simeq Q$ are isometric, then $I(S) = I(Q)$, so I should be well-defined on isometry classes or I has *no false negatives*.
Completeness: if $I(S) = I(Q)$, then $S \simeq Q$ are isometric, hence I has *no false positives*.

Continuity: find a *metric* d and a constant λ such that if any point of S is perturbed within its ε-neighborhood, then $I(S)$ changes by max $\lambda\varepsilon$.

Harder practical requirements

Reconstruction (inverse design): any $S \subset \mathbb{R}^n$ can be reconstructed from its invariant $I(S)$.

Computability: I, d, and reconstruction of S from $I(S)$ can be obtained in polynomial time in the motif size (number of atoms in a unit cell), hence *no infinite/exponential size* invariants.

If all conditions hold, I is *universal* for all types of periodic crystals, independent of symmetry.

If I is simple enough, I defines geographic-style coordinates on the space of all periodic crystals.