An isometry classification of periodic point sets *

Olga Anosova[1] and Vitaliy Kurlin[1][0000−0001−5328−5351]

University of Liverpool, Liverpool L69 3BX, UK vitaliy.kurlin@gmail.com
http://kurlin.org

Abstract. We develop discrete geometry methods to resolve the data ambiguity challenge for periodic point sets to accelerate materials discovery. In any high-dimensional Euclidean space, a periodic point set is obtained from a finite set (motif) of points in a parallelepiped (unit cell) by periodic translations of the motif along basis vectors of the cell.

An important equivalence of periodic sets is a rigid motion or an isometry that preserves interpoint distances. This equivalence is motivated by solid crystals whose periodic structures are determined in a rigid form.

Crystals are still compared by descriptors that are either not isometry invariants or depend on manually chosen tolerances or cut-off parameters. All discrete invariants including symmetry groups can easily break down under atomic vibrations, which are always present in real crystals.

We introduce a complete isometry invariant for all periodic sets of points, which can additionally carry labels such as chemical elements. The main classification theorem says that any two periodic sets are isometric if and only if their proposed complete invariants (called isosets) are equal.

A potential equality between isosets can be checked by an algorithm, whose computational complexity is polynomial in the number of motif points. The key advantage of isosets is continuity under perturbations, which allows us to quantify similarities between any periodic point sets.

Keywords: lattice · periodic set · isometry invariant · classification

1 Introduction: motivations and problem statement

One well-known challenge in applications is the *curse of dimensionality* meaning that any dataset seems sparse in a high-dimensional space. This paper studies the more basic *ambiguity challenge* in data representations meaning that equivalent real-life object can often be represented in infinitely many different ways.

Data ambiguity makes any comparison unreliable. For example, humans should be not be compared or identified by the average color of their clothes, though such colors are easily accessible in photos. Justified comparisons should use only *invariant* features that are independent of an object representation.

* Supported by the EPSRC grant Application-driven Topological Data Analysis

Our objects are periodic point sets, which model all solid crystalline materials (crystals). Solid crystal structures are determined in a rigid form with well-defined atomic positions. Atoms form strong bonds only in molecules, while all inter-molecular bonds are much weaker and have universally agreed definitions. Points at atomic centers can be labeled by chemical elements or any other properties, e.g. radii. Later we explain how to easily incorporate labels into our invariants. We start with the most fundamental model of a periodic point set.

The simplest example is a *lattice* Λ, a discrete set of points that are integer linear combinations of any linear (not necessarily orthogonal) basis in $\mathbb{R}^n$, see Fig. 1. More generally, a *periodic point set* is obtained from a finite collection (*motif*) of points by periodic translations along all vectors of a lattice Λ.

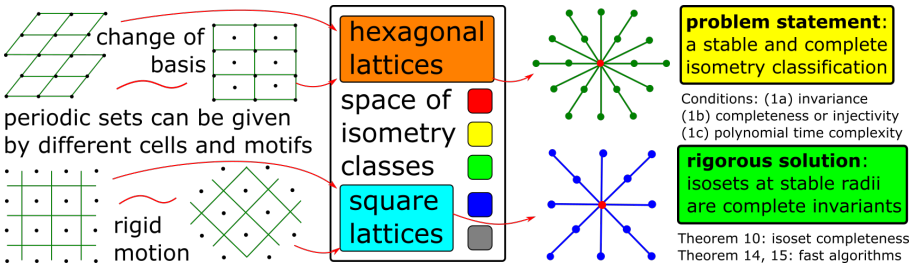

Fig. 1. **Left**: periodic sets represented by different cells are organized in isometry classes, which form a continuous space. **Right**: the new *isoset* resolves the ambiguity.

The same periodic point set S can be obtained from infinitely many different Minkowski sums $\Lambda + M$. For example, one can change a linear basis of Λ and get a new motif of points with different coordinates in the new basis.

The above ambiguity with respect to a basis is compounded by infinitely many rigid motions or isometries that preserve inter-point distances, hence produce equivalent crystal structures. Shifting all points by a fixed vector changes all point coordinates in a fixed basis, but not the isometry class of the set.

The curse of ambiguity for periodic sets can be resolved only by a complete isometry invariant as follows: two periodic sets given by any decompositions $\Lambda+M$ into a lattice and a motif should be isometric if and only if their complete invariants coincide. Such a complete invariant should have easily comparable values from which we could explicitly reconstruct an original crystal structure.

The final requirement for a complete invariant is its continuity under perturbations, which was largely ignored in the past despite all atoms vibrate above the absolute zero temperature. All discrete invariants including symmetry groups are discontinuous under perturbation of points. A similarity between crystals should be quantified in a continuous way to filter out nearly identical crystals obtained as approximations to local energy minima in Crystal Structure Prediction [17].

Problem 1 formalizes the above curse of ambiguity for crystal structures.

Problem 1 (complete isometry classification of periodic point sets)**.** Find a function I on the space of all periodic point sets in $\mathbb{R}^n$ such that

(1a) *invariance* : if any periodic sets S, Q are isometric, then $I(S) = I(Q)$;

(1b) *continuity* : $I(S)$ continuously changes under perturbations of points;

(1c) *computability*: $I(S) = I(Q)$ is checked in a polynomial time in a motif size;

(1d) *completeness* : if $I(S) = I(Q)$, then the periodic sets S, Q are isometric. ■

The main contribution is the new invariant *isoset* in Definition 9 whose completeness is proved in Theorem 10. Conditions (1cd) are proved in the recent work [3] introducing the new research area of Periodic Geometry and Topology.

2 A review of the relevant work on periodic crystals

Despite any lattice can be defined by infinitely many primitive cells, there is a unique Niggli's reduced cell, which can be theoretically used for comparing periodic sets [11, section 9.2]. Niggli's and other reduced cells are discontinuous under perturbations in the sense that a reduced cell of a perturbed lattice can have a basis that substantially differs from that of a non-perturbed lattice [2].

Continuity condition (1b) fails not only for Niggli's reduced cell, but also for all discretely-valued invariants including symmetry groups. The 230 crystallographic groups in $\mathbb{R}^3$ cut the continuous space of isometry classes into disjoint pieces. This stratification shows many nearly identical crystals as distant.

The first step towards a complete isometry classification of crystals has recently been done in [14] by introducing two proper distances between arbitrary lattices that satisfy the metric axioms and are also continuous under perturbations. Also [14, section 3] reviews many past tools to compare crystals.

The world's largest Cambridge Structural Database (CSD) has more than 1M crystals. Each crystal is represented by one of infinitely many choices of a unit cell and a motif M in the form of a Crystallographic Information File (CIF). The CSD is a super-long list of CIFs with limited search tools, mainly by chemical compositions, and without any organization by geometric similarity.

Quantifying crystal similarities is even more important for Crystal Structure Prediction (CSP). A typical CSP software starts from a given chemical composition and outputs thousands of predicted crystals as approximations to local minima of a complicated energy function. Any iterative optimization produces many approximations to the same local minimum. These nearly identical crystals are currently impossible to automatically identify in a reliable way [17].

Crystals are often compared by the Radial Distribution Function (RDF) that measures the probability of finding one atom at a distance of r from a reference atom, which is computed up to a manually chosen cut-off radius.

The new concept of a stable radius in Definition 8 gives exact conditions for a required radius depending on a complexity of a periodic set. The crystals indistinguishable by their RDF or diffraction patterns are known as homometric [15]. The most recent survey [16, Fig. S4] has highlighted pairs of finite atomic arrangements that cannot be distinguished by any known crystal descriptors.

On a positive side, the mathematical approach in [10] has solved the already non-trivial 1-dimensional case for sets whose points have only integer (or rational) coordinates. Briefly, any given points $c_0, \dots, c_{m-1}$ on the unit circle $S^1 \subset \mathbb{C}$ are converted into the Fourier coefficients $d(k) = \sum\limits_{j=0}^{m-1} c_j \exp \frac{2\pi ijk}{m}$, $k = 0, \dots, m-1$. Then all point sets in the unit circle can be distinguished up to circular rotations by the n-th order invariants up to $n = 6$, which are all products of the form $d(k_1) \cdots d(k_n)$ with $k_1 + \dots + k_n \equiv 0 \pmod m$.

The more recent advances in Problem 1 are Density Functions [9] and Average Minimum Distances [18]. The k-density function $\psi_k[S]$ of a periodic point set $S \subset \mathbb{R}^n$ measures the fractional area of the region within a unit cell U covered by exactly k closed balls with centers $a \in S$ and a radius $t \geq 0$. The density functions satisfy conditions (1abc) and completeness (1d) in general position.

However, the density functions do not distinguish the following 1-dimensional sets $S_{15} = \{0, 1, 3, 4, 5, 7, 9, 10, 12\}+15\mathbb{Z}$ and $Q_{15} = \{0, 1, 3, 4, 6, 8, 9, 12, 14\}+15\mathbb{Z}$ with period 15, see [3, Example 11]. The sets S_{15}, Q_{15} were introduced at the beginning of section 5 in [9] as $U \pm V + 15\mathbb{Z}$ for $U = \{0, 4, 9\}$ and $V = \{0, 1, 3\}$.

The above sets S_{15}, Q_{15} are distinguished by the faster Average Minimum Distances (AMD), see [3, Example 6]. For any integer $k \geq 1$, $\mathrm{AMD}_k(S)$ is the distance from a point $p \in S$ to its k-th nearest neighbor, averaged over all points p in a motif of S. For $k \to +\infty$, $\mathrm{AMD}_k(S)$ behaves as $\sqrt[n]{k}$, see [18, Theorem 14].

3 Necessary concepts from computational geometry

In the Euclidean space $\mathbb{R}^n$, any point $p \in \mathbb{R}^n$ is represented by the vector $\boldsymbol{p}$ from the origin of $\mathbb{R}^n$ to p. The *Euclidean* distance between points $p, q \in \mathbb{R}^n$ is denoted by $|pq| = |\boldsymbol{p} - \boldsymbol{q}|$. For a standard orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, the integer lattice $\mathbb{Z}^n \subset \mathbb{R}^n$ consists of all points with integer coordinates.

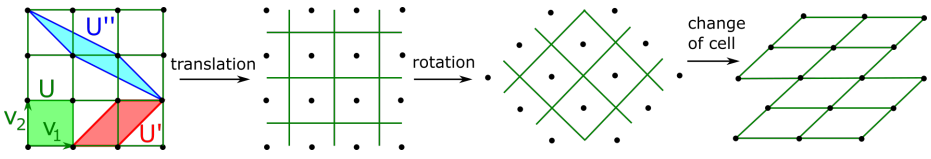

Fig. 2. Left: three primitive cells U, U', U'' of the square lattice S. Other pictures show different periodic sets $\Lambda + M$, which are all isometric to the square lattice S.

Definition 2 (a lattice Λ, a unit cell U, a motif M, a periodic set $S = \Lambda + M$)**.** For any linear basis $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ in $\mathbb{R}^n$, a *lattice* is $\Lambda = \{\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i : \lambda_i \in \mathbb{Z}\}$. The *unit cell* $U(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n) = \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i \ : \ \lambda_i \in [0,1) \right\}$ is the parallelepiped spanned by the basis. A motif M is any finite set of points $p_1, \dots, p_m \in U$. A *periodic point set* is the Minkowski sum $S = \Lambda + M = \{\boldsymbol{u} + \boldsymbol{v} : \boldsymbol{u} \in \Lambda, \boldsymbol{v} \in M\}$. A unit cell U of a periodic set $S = \Lambda + M$ is *primitive* if any vector $\boldsymbol{v}$ that translates S to itself is an integer linear combination of the basis of the cell U, i.e. $\boldsymbol{v} \in \Lambda$. ■

A primitive unit cell U of any lattice has a motif of one point (the origin). If U is defined as the closed parallelepiped in $\mathbb{R}^n$, hence includes 2^n vertices, one could count every vertex with weight 2^{-n} so that the sum is 1. All closed unit cells in Fig. 2 are primitive, because four corners are counted as one point in U.

The first picture in Fig. 3 shows a small perturbation of a square lattice. The new periodic set has a twice larger primitive unit cell with two points in a motif instead of one. All invariants based on a fixed primitive unit cell such as Niggli's reduced cell [11, section 9.2] fail continuity condition (1b) in Problem 1.

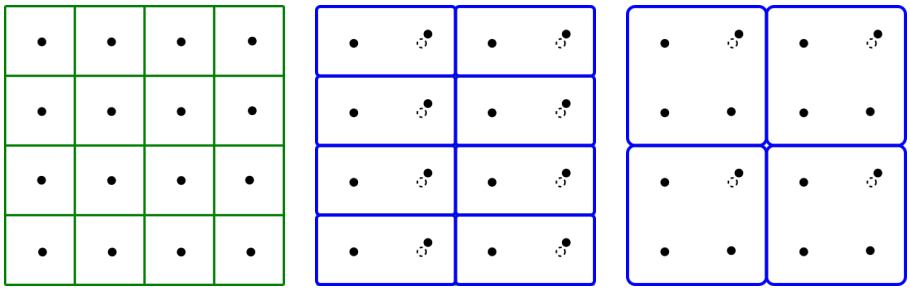

Fig. 3. A continuous invariant should take close values on these nearly identical periodic sets, though their symmetry groups and primitive cells substantially differ.

The auxiliary concepts in Definitions 3, 4, 5 follow Dolbilin's papers [7], [5].

Definition 3 (bridge distance $\beta(S)$)**.** For a periodic point set $S \subset \mathbb{R}^n$, the *bridge* distance is a minimum $\beta(S) > 0$ such that any two points $a, b \in S$ can be connected by a finite sequence $a_0 = a, a_1, \dots, a_m = b$ such that any two successive points a_i, a_{i+1} are close, i.e. the Euclidean distance $|\boldsymbol{a}_{i-1} - \boldsymbol{a}_i| \le \beta(S)$ for $i = 1, \dots, m$. Fig. 4 shows periodic sets with different bridge distances. ■

Definition 4 (m-regularity of a periodic set)**.** For any point a in a periodic set $S \subset \mathbb{R}^n$, the *global cluster* $C(S, a)$ is the infinite set of vectors $\boldsymbol{b} - \boldsymbol{a}$ for all points $b \in S$. Points $a, b \in S$ are called *isometrically equivalent* if there is an isometry $f : C(S, a) \to C(S, b)$ such that $f(a) = b$. A periodic set $S \subset \mathbb{R}^n$ is called *regular* if all points $a, b \in S$ are isometrically equivalent. A periodic set S is *m-regular* if all global clusters of S form exactly $m \ge 1$ isometry classes. ■

For any point $a \in S$, its global cluster is a view of S from the position of a, e.g. how we view all astronomical stars in the universe S from our planet Earth. Any lattice is 1-regular, because all its global clusters are related by translations.

Though the global clusters $C(S,a)$ and $C(S,b)$ at any different points $a,b \in S$ seem to contain the same set S, they can be different even modulo translations.

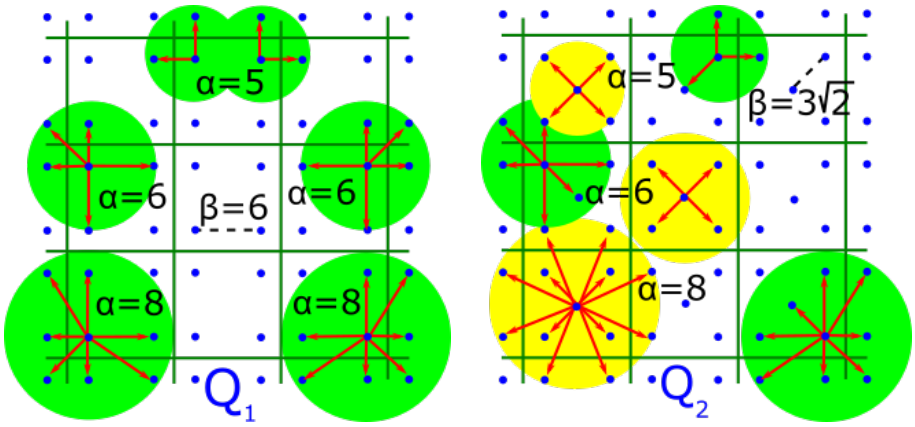

Fig. 4. Left: the periodic point set Q_1 has the four points $(\pm 2, \pm 2)$ in the square unit cell $[0,10]^2$, so Q_1 isn't a lattice, but is 1-regular by Definition 4, also $\beta(Q_1) = 6$. All local α-clusters are isometric, shown by red arrows for radii $\alpha = 5, 6, 8$, see Definition 5. **Right**: the periodic point set Q_2 has the extra point $(5,5)$ in the center of $[0,10]^2$ and is 2-regular with $\beta(Q_2) = 3\sqrt{2}$. Local clusters have two isometry types.

The first picture in Fig. 4 shows the 1-regular set $Q_1 \subset \mathbb{R}^2$, where all points have isometric global clusters related by translations and rotations through $\frac{\pi}{2}, \pi, \frac{3\pi}{2}$, so Q_1 is not a lattice. The global clusters are infinite, hence distinguishing them up to isometry is not easier than distinguishing the original sets. However, m-regularity can be checked in terms of local clusters defined below.

Definition 5 (local α-clusters $C(S,a;\alpha)$ and symmetry groups $\mathrm{Sym}(S,a;\alpha)$)**.** For a point a in a crystal $S \subset \mathbb{R}^n$ and any radius $\alpha \geq 0$, the *local cluster* $C(S,a;\alpha)$ is the set of vectors $\boldsymbol{b}-\boldsymbol{a}$ of lengths $|\boldsymbol{b}-\boldsymbol{a}| \leq \alpha$ for $b \in S$. An isometry $f \in \mathrm{Iso}(\mathbb{R}^n)$ between clusters should match their centers. The *symmetry* group $\mathrm{Sym}(S,a;\alpha)$ consists of *self-isometries* of $C(S,a;\alpha)$ that fix the center a. ■

If $\alpha > 0$ is smaller than the minimum distance between any points, then every cluster $C(S,a;\alpha)$ is the single-point set $\{a\}$ and its symmetry group $\mathrm{O}(\mathbb{R}^n)$ consists of all isometries fixing the center a. When the radius α is increasing, the α-clusters $C(S,a;\alpha)$ become larger and can have fewer self-isometries, so the symmetry group $\mathrm{Sym}(S,a;\alpha)$ becomes smaller and eventually stabilizes.

The 1-regular set Q_1 in Fig. 4 for any point $a \in Q_1$ has the symmetry group $\mathrm{Sym}(Q_1, a; \alpha) = \mathrm{O}(\mathbb{R}^2)$ for $\alpha \in [0, 4)$. The group $\mathrm{Sym}(Q_1, a; \alpha)$ stabilizes as $\mathbb{Z}_2$ for $\alpha \geq 4$ as soon as the local α-cluster $C(Q_1, a; \alpha)$ includes one more point.

4 The isotree of isometry classes and a stable radius

This section introduces the isotree and a stable radius in Definitions 6 and 8 by comparing local clusters at radii $\alpha - \beta$ and β, where β is the bridge distance.

Any isometry $A \to B$ between local clusters should map the center of A to the center of B. The *isotree* in Definition 6 is inspired by a dendrogram of hierarchical clustering, though points are partitioned according to isometry classes of local α-clusters at different radii α, not by a distance threshold.

Definition 6 (isotree $\mathrm{IT}(S)$ of α-partitions)**.** Fix a periodic set $S \subset \mathbb{R}^n$ and $\alpha \geq 0$. Points $a, b \in S$ are called *α-equivalent* if their α-clusters $C(S, a; \alpha)$ and $C(S, b; \alpha)$ are isometric. The *α-equivalence class* $[C(S, a; \alpha)]$ consists of all α-clusters isometric to $C(S, a; \alpha)$. The *α-partition* $P(S; \alpha)$ is the splitting of S into α-equivalence classes of points. The number of α-equivalence classes of α-clusters is the *cluster count* $|P(S; \alpha)|$. When the radius α is increasing, the α-partition can be refined by subdividing α-equivalence classes of points of S into subclasses. If we represent each α-equivalence class by an abstract point, the resulting points form the *isotree* $\mathrm{IT}(S)$ of all α-partitions, see Fig. 5, 6. ■

The α-equivalence and isoset in Definition 9 can be refined by labels of points such as chemical elements. Theorem 10 will remain valid for labelled points. Recall that isometries include reflections, however an orientation sign can be easily added to α-clusters, hence we focus on the basic case of all isometries.

When a radius α is increasing, α-clusters $C(S, a; \alpha)$ include more points, hence are less likely to be isometric, so $|P(S; \alpha)|$ is a non-increasing function of α. Fig. 5, 6 show α-clusters and isotrees of non-isometric 1D periodic sets S, Q [15, p. 197, Fig. 2], which have identical 1D analogs of diffraction patterns.

Any α-equivalence class from $P(S; \alpha)$ may split into two or more classes, which will not merge at any larger radius α'. Lemma 7 justifies that the isotree $\mathrm{IT}(S)$ can be visualized as a merge tree of α-equivalence classes of clusters.

Lemma 7 (isotree properties)**.** The isotree $\mathrm{IT}(S)$ has the following properties:

(7a) for $\alpha = 0$, the α-partition $P(S; 0)$ consists of one class;

(7b) if $\alpha < \alpha'$, then $\mathrm{Sym}(S, a; \alpha') \subseteq \mathrm{Sym}(S, a; \alpha)$ for $a \in S$;

(7c) if $\alpha < \alpha'$, the α'-partition $P(S; \alpha')$ *refines* $P(S; \alpha)$, i.e. any set from the α'-partition $P(S; \alpha')$ is included into a set from the α-partition $P(S; \alpha)$. ■

Proof. (7a) If $\alpha \geq 0$ is smaller than the minimum distance r between point of S, every cluster $C(S, a; \alpha)$ is the single-point set $\{a\}$. All these single-point clusters are isometric to each other. So $|P(S; \alpha)| = 1$ for all small radii $\alpha < r$.

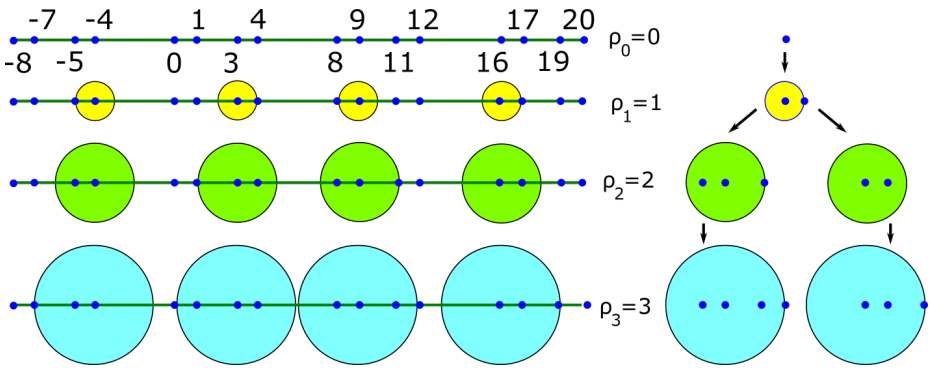

Fig. 5. Left: $S = \{0, 1, 3, 4\} + 8\mathbb{Z}$ has $t = 4$ and is 2-regular by Definition 4. **Right**: Local clusters with radii $\alpha = 0, 1, 2, 3$ represent vertices of the isotree IT(S) in Definition 6. All α-clusters are isometric for $\alpha < 2$, form two isometry classes for $\alpha \geq 2$.

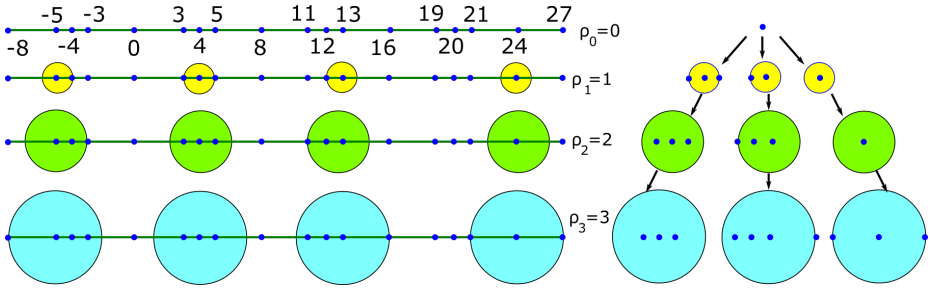

Fig. 6. Left: $Q = \{0, 3, 4, 5\} + 8\mathbb{Z}$ has $t = 3$ and is 3-regular by Definition 4. **Right**: Local clusters with radii $\alpha = 0, 1, 2, 3$ represent vertices of the isotree IT(Q) in Definition 6. All α-clusters are isometric for $\alpha < 1$, form three isometry classes for $\alpha \geq 1$.

(7b) For any point $a \in S$, the inclusion of clusters $C(S, a; \alpha) \subseteq C(S, a; \alpha')$ implies that any self-isometry of the larger cluster $C(S, a; \alpha')$ can be restricted to a self-isometry of the smaller cluster $C(S, a; \alpha)$. So $\mathrm{Sym}(S, a; \alpha') \subseteq \mathrm{Sym}(S, a; \alpha)$.

(7c) If points $a, b \in S$ are α'-equivalent at the larger radius α', i.e. the clusters $C(S, a; \alpha')$ and $C(S, b; \alpha')$ are isometric, then a, b are α-equivalent at the smaller radius α. Hence any α'-equivalence class is a subset of an α-equivalence class. □

Property (7c) can be illustrated by the examples in Fig. 5 and 6. For $\alpha = 1$, all points of the periodic set $S = \{0, 1, 3, 4\} + 8\mathbb{Z}$ are in the same α-equivalence class with 1-cluster $\{0, 1\}$. For $\alpha' = 2$, S splits in two α'-equivalence classes: one containing the points from $0 + \mathbb{Z}$ and $4 + \mathbb{Z}$ with the 2-clusters $\{0, 1\}$ and another one containing $1 + \mathbb{Z}$ and $3 + \mathbb{Z}$ with 2-clusters $\{-1, 0, 2\}$.

If a point set S is periodic, the α-partitions of S stabilize in the sense below.

Definition 8 (a stable radius)**.** Let a periodic point set $S \subset \mathbb{R}^n$ and β be an upper bound of its bridge distance $\beta(S)$ from Definition 3. A radius $\alpha \geq \beta$ is called *stable* if both conditions below hold:

(8a) the α-partition $P(S;\alpha)$ coincides with the $(\alpha-\beta)$-partition $P(S;\alpha-\beta)$;

(8b) the symmetry groups stabilize: $\mathrm{Sym}(S,a;\alpha) = \mathrm{Sym}(S,a;\alpha-\beta)$ for all points $a \in S$, which is enough to check for points only from a finite motif of S. ■

A minimum radius α satisfying the above conditions for the bridge distance $\beta(S)$ from Definition 3 can be called *the minimum stable radius* and denoted by $\alpha(S)$. Upper bounds of $\alpha(S)$ and $\beta(S)$ will be enough for all results below.

Due to Lemma (7bc), conditions (8ab) imply that the α'-partitions $P(S;\alpha')$ and the symmetry groups $\mathrm{Sym}(S,a;\alpha')$ remain the same for all $\alpha' \in [\alpha-\beta,\alpha]$.

Condition (8b) doesn't follow from condition (8a) due to the following example. Let Λ be the 2D lattice with the basis $(1,0)$ and $(0,\beta)$ for $\beta > 1$. Then β is the bridge distance of Λ. Condition (8a) is satisfied for any $\alpha \geq 0$, because all points of any lattice are equivalent up to translations. However, condition (8b) fails for any $\alpha < \beta + 1$. Indeed, the α-cluster of the origin $(0,0)$ contains five points $(0,0),(\pm 1,0),(0,\pm\beta)$, whose symmetries are generated by the two reflections in the axes x, y, but the $(\alpha-\beta)$-cluster of the origin consists of only $(0,0)$ and has the symmetry group $\mathrm{O}(2)$.

Condition (8b) might imply condition (8a), but in practice it makes sense to verify (8b) only after checking much simpler condition (8a). Both conditions are essentially used in the proofs of Isometry Classification Theorem 10.

For the set $S = \{0,1,3,4\} + 8\mathbb{Z}$ in Fig. 5 with the bridge distance $\beta(S) = 4$, any $\alpha \geq 6$ is a stable radius, because the partition $P(S;\alpha-4)$ splits S into the same two classes for any $\alpha \geq 6$. For the periodic set $Q = \{0,3,4,5\} + 8\mathbb{Z}$ in Fig. 6 with the bridge distance $\beta(Q) = 3$, any $\alpha \geq 4$ is a stable radius.

Any periodic set $S \subset \mathbb{R}^n$ with m motif points has at most m α-equivalence classes, because any point of S can be translated to a motif point. Hence it suffices to check condition (8a) about α-partitions only for the m motif points. Condition (8b) can be practically checked by testing if the inclusion $\mathrm{Sym}(S,a;\alpha') \subset \mathrm{Sym}(S,a;\alpha)$ from (7b) is surjective, which is needed only for one representative cluster from at most m isometry classes (exactly m is S is m-regular).

A stable radius in [5] was defined by using the notations ρ and $\rho+t$. This pair changed to $\alpha-\beta$ and α, because subsequent Theorem 10 is more conveniently stated for the larger radius α. Any 1-regular set in $\mathbb{R}^3$ with a bridge distance β has a stable radius $\alpha = 7\beta$ or $\rho = 6t$ in the past notations of [7].

5 Isosets completely classify periodic sets up to isometry

A criterion of m-regular sets [8, Theorem 1.3] has inspired us to introduce the new invariant isoset in Definition 9, whose completeness (injectivity) in the isometry classification of periodic sets will be proved in main Theorem 10.

Definition 9 (isoset $I(S;\alpha)$ of a periodic point set S at a radius α)**.** Let a periodic point set $S \subset \mathbb{R}^n$ have a motif M of m points. Split all points $a \in M$ into α-equivalence classes. Then each α-equivalence class consisting of (say) k points in M can be associated with the *isometry class* of $\sigma = [C(S,a;\alpha)]$ of an α-cluster centered at one of these k points $a \in M$. The *weight* of the class σ is defined as $w = k/m$. Then the *isoset* $I(S;\alpha)$ is defined as the unordered set of all isometry classes with weights $(\sigma; w)$ over all points $a \in M$. ■

All points a of a lattice $\Lambda \subset \mathbb{R}^n$ are α-equivalent for any $\alpha \geq 0$, because all α-clusters $C(\Lambda, a; \alpha)$ are isometrically equivalent to each other by translations. Hence the isoset $I(\Lambda;\alpha)$ is one isometry class of weight 1 for any α.

All isometry classes $\sigma \in I(S;\alpha)$ are in a 1-1 correspondence with all α-equivalence classes in the α-partition $P(S;\alpha)$ from Definition 6. So $I(S;\alpha)$ without weights is a set of points in the isotree $\mathrm{IT}(S)$ at the radius α. The size of the isoset $I(S;\alpha)$ equals the cluster count $|P(S;\alpha)|$. Formally, $I(S;\alpha)$ depends on α, because α-clusters grow in α. To distinguish periodic point sets S, Q up to isometry, we will compare their isosets at a common stable radius α.

An equality $\sigma = \xi$ between isometry classes of clusters means that there is an isometry f from a cluster $C(S,a;\alpha)$ representing σ to a cluster $C(Q,b;\alpha)$ representing ξ such that $f(a) = b$, i.e. f respects the centers of the clusters.

The set $S = \{0,1,3,4\} + 8\mathbb{Z}$ in Fig. 5 has the isoset $I(S;6)$ of two isometry classes of 6-clusters represented by $\{-4,-3,-1,0,1,4,5\}$ and $\{-3,-2,0,1,5,6\}$ centered at 0. The set $Q = \{0,3,4,5\} + 8\mathbb{Z}$ in Fig. 6 has the isoset $I(Q;4)$ of three isometry classes of 4-clusters represented by $\{-4,-3,0,3,4\}$, $\{-3,0,1,2\}$, $\{-4,-1,0,1,4\}$. To conclude that S, Q are not isometric, Theorem 10 will require us to compare their isosets at a common stable radius $\alpha \geq 6$. In the above case it suffices to say that the stabilized cluster counts differ: $2 \neq 3$.

An equality $\sigma = \xi$ between isometry classes means that there is an isometry f from a cluster in σ to a cluster in ξ so that f respects the centers of the clusters. This equality is checked in time $O(k^{n-2}\log k)$ for any dimension $n \geq 3$ by [1, Theorem 1(a)], where k is the maximum number of points in the clusters.

Theorem 10 (complete isometry classification of periodic point sets)**.** For any periodic point sets $S, Q \subset \mathbb{R}^n$, let α be a common stable radius satisfying Definition 8 for an upper bound β of $\beta(S), \beta(Q)$. Then S, Q are isometric if and only if there is a bijection between their isosets respecting weights: $I(S;\alpha) = I(Q;\alpha)$ means that any isometry class $(\sigma; w) \in I(S;\alpha)$ of a weight w coincides with a class $(\xi; w) \in I(Q;\alpha)$ of the same weight w and vice versa. ■

Theoretically a complete invariant of S should include isosets $I(S;\alpha)$ for all sufficiently large radii α. However, when comparing two sets S, Q up to isometry, it suffices to build their isosets only at a common stable radius α.

The α-equivalence and isoset in Definition 9 can be refined by labels of points such as chemical elements, which keeps Theorem 10 valid for labeled points.

Recall that isometries include reflections, however an orientation sign can be easily added to α-clusters, hence we focus on the basic case of all isometries.

The proposed complete invariant for classification Problem 1 is the function $S \mapsto I(S;\alpha)$ from any periodic point set S to its isoset at a stable radius α, which doesn't need to be minimal. All points a of a lattice $\Lambda \subset \mathbb{R}^n$ are α-equivalent to each other for $\alpha \geq 0$, because all α-clusters $C(\Lambda, a; \alpha)$ are related by translations, hence the isoset $I(\Lambda;\alpha)$ of any lattice is a single isometry class for any α.

Lemmas 11 and 12 help to extend an isometry between local clusters to full periodic sets to prove the complete isometry classification in Theorem 10.

Lemma 11 (local extension)**.** Let periodic sets $S, Q \subset \mathbb{R}^n$ have bridge distances at most β and a common stable radius α such that α-clusters $C(S,a;\alpha)$ and $C(Q,b;\alpha)$ are isometric for some $a \in S$, $b \in Q$. Then any isometry $f : C(S,a;\alpha-\beta) \to C(Q,b;\alpha-\beta)$ extends to an isometry $C(S,a;\alpha) \to C(Q,b;\alpha)$. ■

Proof. Let $g : C(S,a;\alpha) \to C(Q,b;\alpha)$ be any isometry, which may not coincide with f on the $(\alpha-\beta)$-subcluster $C(S,a;\alpha-\beta)$. The composition $f^{-1} \circ g$ isometrically maps $C(S,a;\alpha-\beta)$ to itself. Hence $f^{-1} \circ g = h \in \mathrm{Sym}(S,a;\alpha-\beta)$ is a self-isometry. Since the symmetry groups stabilize by condition (8b), the isometry h maps the larger cluster $C(S,a;\alpha)$ to itself. Then the initial isometry f extends to the isometry $g \circ h^{-1} : C(S,a;\alpha) \to C(Q,b;\alpha)$ as required. □

Lemma 12 (global extension)**.** For any periodic point sets $S, Q \subset \mathbb{R}^n$, let α be a common stable radius satisfying Definition 8 for an upper bound β of both $\beta(S), \beta(Q)$. Assume that $I(S;\alpha) = I(Q;\alpha)$. Fix a point $a \in S$. Then any local isometry $f : C(S,a;\alpha) \to C(Q,f(a);\alpha)$ extends to a global isometry $S \to Q$. ■

Proof. We shall prove that the image $f(b)$ of any point $a' \in S$ belongs to Q, hence $f(S) \subset Q$. Swapping the roles of S and Q will prove that $f^{-1}(Q) \subset S$, i.e. f is a global isometry $S \to Q$. By Definition 3 the above points $a, a' \in S$ are connected by a sequence of points $a = a_0, a_1, \dots, a_m = a' \in S$ such that $|\boldsymbol{a}_{i-1} - \boldsymbol{a}_i| \leq \beta$, $i = 1, \dots, m$, where β is an upper bound of both $\beta(S), \beta(Q)$.

The cluster $C(S,a;\alpha)$ is the intersection $S \cap B(a;\alpha)$. The ball $B(a;\alpha)$ contains the smaller ball $B(a_1;\alpha-\beta)$ around the closely located center a_1. Indeed, since $|\boldsymbol{a} - \boldsymbol{a}_1| \leq \beta$, the triangle inequality for the Euclidean distance implies that any $c \in B(a_1;\alpha)$ with $|\boldsymbol{a}_1 - \boldsymbol{c}| \leq \alpha - \beta$ satisfies $|\boldsymbol{a} - \boldsymbol{c}| \leq |\boldsymbol{a} - \boldsymbol{a}_1| + |\boldsymbol{a}_1 - \boldsymbol{c}| \leq \alpha$.

Due to $I(S;\alpha) = I(Q;\alpha)$ the isometry class of $C(S,a_1;\alpha)$ coincides with an isometry class of $C(Q,b;\alpha)$ for some $b \in Q$, i.e. $C(S,a_1;\alpha)$ is isometric to $C(Q,b;\alpha)$. Then the clusters $C(S,a_1;\alpha-\beta)$ and $C(Q,b;\alpha-\beta)$ are isometric.

By condition (8a), the splitting of Q into α-equivalence classes coincides with the splitting into $(\alpha-\beta)$-equivalence classes. Take the $(\alpha-\beta)$-equivalence class $[C(Q,b;\alpha-\beta)]$ containing b. This class includes the point $f(a_1) \in Q$, because f restricts to the isometry $f : C(S,a_1;\alpha-\beta) \to C(Q,f(a_1);\alpha-\beta)$ and $C(S,a_1;\alpha-\beta)$ was shown to be isometric to $C(Q,b;\alpha-\beta)$.

The α-equivalence class $[C(Q,b;\alpha)]$ includes both b and $f(a_1)$. The isometry class $[C(Q,b;\alpha)] = [C(S,a_1;\alpha)]$ can be represented by the cluster $C(Q,f(a_1);\alpha)$, which is now proved to be isometric to $C(S,a_1;\alpha)$.

We apply Lemma 11 for f restricted to $C(S,a_1;\alpha-\beta) \to C(Q,f(a_1),\alpha-\beta)$ and conclude that f extends to an isometry $C(S,a_1;\alpha) \to C(Q,f(a_1);\alpha)$.

Continue applying Lemma 11 to the clusters around the next center a_2 and so on until we conclude that the initial isometry f maps the α-cluster centered at $a_m = a' \in S$ to an isometric cluster within Q, so $f(a') \in Q$ as required. □

Lemma 13 (all stable radii of a periodic set)**.** If α is a stable radius of a periodic point set $S \subset \mathbb{R}^n$, then so is any larger radius $\alpha' > \alpha$. Then all stable radii form the interval $[\alpha(S), +\infty)$, where $\alpha(S)$ is the minimum stable radius of S. ■

Proof. Due to Lemma (7bc), conditions (8ab) imply that the α'-partition $P(S;\alpha')$ and the symmetry groups $\mathrm{Sym}(S,a;\alpha')$ remain the same for all $\alpha' \in [\alpha-\beta,\alpha]$. We need to show that they remain the same for any larger $\alpha' > \alpha$.

Below we will apply Lemma 12 for the same set $S = Q$ and $\beta = \beta(S)$. Let points $a, b \in S$ be α-equivalent, i.e. there is an isometry $f : C(S,a;\alpha) \to C(S,b;\alpha)$. By Lemma 12 the local isometry f extends to a global self-isometry $S \to S$ such that $f(a) = b$. Then all larger α'-clusters of a, b are isometric, i.e. a, b are α'-equivalent and $P(S;\alpha) = P(S,\alpha')$. Similarly, any self-isometry of $C(S,a;\alpha)$ extends to a global self-isometry, i.e. the symmetry group $\mathrm{Sym}(S,a;\alpha')$ for any $\alpha' > \alpha$ is isomorphic to $\mathrm{Sym}(S,a;\alpha')$. □

Proof of Theorem 10. The part *only if* $\Rightarrow$ follows by restricting any given global isometry $f : S \to Q$ between the infinite sets of points to the local α-clusters $C(S,a;\alpha) \to C(Q,f(a);\alpha)$ for any point a in a motif M of S.

Hence the isometry class $[C(S,a;\alpha)]$ is considered equivalent to the class $[C(Q,f(a);\alpha)]$, which can be represented by the α-cluster $C(Q,b;\alpha)$ centered at a point b in a motif of Q. Since f is a bijection and the point $a \in M$ was arbitrary, we get a bijection between isometry classes with weights in $I(S;\alpha) = I(Q;\alpha)$.

The part *if* $\Leftarrow$. Fix a point $a \in S$. The α-cluster $C(S,a;\alpha)$ represents a class with a weight $(\sigma,w) \in I(S;\alpha)$. Due to $I(S;\alpha) = I(Q;\alpha)$, there is an isometry $f : C(S,a;\alpha) \to C(Q,f(a);\alpha)$ to a cluster from an equal class $(\sigma,w) \in I(Q;\alpha)$. By Lemma 12 the local isometry f extends to a global isometry $S \to Q$. □

6 A discussion of further properties of isosets

This paper has resolved the ambiguity challenge for crystal representations, which is common for many data objects [4]. Crystal descriptors [13] are often based on ambiguous unit cells or computed up to a manual cut-off radii. Representations of 2-periodic textiles [6] should be similarly studied up to periodic isotopies [3, section 10] without fixing a unit cell. Definition 8 gives conditions for a stable radius so that larger clusters will not bring any new information.

The recent survey of atomic structure representations [16] confirmed that there was no complete invariant that distinguishes all crystals up to isometry.

Theorem 10 provides a complete invariant for the first time. The follow-up paper [3] discusses computations and continuity of the new invariant isoset. Isosets consisting of different numbers of isometry classes will be compared by the Earth Mover's Distance [12]. We thank all reviewers for their time and suggestions.

References

1. Alt, H., Mehlhorn, K., Wagener, H., Welzl, E.: Congruence, similarity, and symmetries of geometric objects. Discrete & Comp. Geometry **3**, 237–256 (1988)
2. Andrews, L., Bernstein, H., Pelletier, G.: A perturbation stable cell comparison technique. Acta Crystallographica A **36**(2), 248–252 (1980)
3. Anosova, O., Kurlin, V.: Introduction to periodic geometry and topology. arXiv:2103.02749 (2021)
4. Anosova, O., Kurlin, V.: Key data challenges: what can be wrong with AI? preprint (2021), `http://kurlin.org/research-papers.php#year2021`
5. Bouniaev, M., Dolbilin, N.: Regular and multi-regular t-bonded systems. J. Information Processing **25**, 735–740 (2017)
6. Bright, M., Kurlin, V.: Encoding and topological computation on textile structures. Computers & Graphics **90**, 51–61 (2020)
7. Dolbilin, N., Bouniaev, M.: Regular t-bonded systems in R^3. European Journal of Combinatorics **80**, 89–101 (2019)
8. Dolbilin, N., Lagarias, J., Senechal, M.: Multiregular point systems. Discrete & Computational Geometry **20**(4), 477–498 (1998)
9. Edelsbrunner, H., Heiss, T., Kurlin, V., Smith, P., Wintraecken, M.: The density fingerprint of a periodic point set. In: Proceedings of SoCG (2021)
10. Grünbaum, F., Moore, C.: The use of higher-order invariants in the determination of generalized patterson cyclotomic sets. Acta Cryst. A **51**, 310–323 (1995)
11. Hahn, T., Shmueli, U., Arthur, J.: Intern. tables for crystallography, vol. 1 (1983)
12. Hargreaves, C.J., Dyer, M.S., Gaultois, M.W., Kurlin, V.A., Rosseinsky, M.J.: The earth mover's distance as a metric for the space of inorganic compositions. Chemistry of Materials (2020)
13. Himanen, Land Jäger, M., Morooka, E., Canova, F., Ranawat, Y., Gao, D., Rinke, P., Foster, A.: Dscribe: Library of descriptors for machine learning in materials science. Computer Physics Communications **247**, 106949 (2020)
14. Mosca, M., Kurlin, V.: Voronoi-based similarity distances between arbitrary crystal lattices. Crystal Research and Technology **55**(5), 1900197 (2020)
15. Patterson, A.: Ambiguities in the x-ray analysis of crystal structures. Physical Review **65**, 195 (1944)
16. Pozdnyakov, S., Willatt, M., Bartók, A., Ortner, C., Csányi, G., Ceriotti, M.: Incompleteness of atomic structure representations. Phys. Rev. Lett. **125**, 166001 (2020), `arXiv:2001.11696`
17. Pulido, A., Chen, L., Kaczorowski, T., Holden, D., Little, M., Chong, S., Slater, B., McMahon, D., Bonillo, B., Stackhouse, C., Stephenson, A., Kane, C., Clowes, R., Hasell, T., Cooper, A., Day, G.: Functional materials discovery using energy–structure–function maps. Nature **543**, 657–664 (2017)
18. Widdowson, D., Mosca, M., Pulido, A., Kurlin, V., Cooper, A.: Average minimum distances of periodic point sets. arXiv:2009.02488 (2020)