Voronoi-based similarity distances between arbitrary crystal lattices

M. M. Mosca 1, Dr. V. Kurlin 2

Materials Innovation Factory and Computer Science department,

University of Liverpool, Liverpool L69 3BX, United Kingdom

E-mail: vitaliy.kurlin@liverpool.ac.uk

This paper develops a new continuous approach to a similarity between periodic lattices of ideal crystals. Quantifying a similarity between crystal structures is needed to substantially speed up the Crystal Structure Prediction, because the prediction of many target properties of crystal structures is computationally slow and is essentially repeated for many nearly identical simulated structures. The proposed distances between arbitrary periodic lattices of crystal structures are invariant under all rigid motions, satisfy the metric axioms and continuity under atomic perturbations. The above properties make these distances ideal tools for clustering and visualizing large datasets of crystal structures. All the conclusions are rigorously proved and justified by experiments on real and simulated crystal structures reported in the Nature 2017 paper "Functional materials discovery using energy–structure–function maps".

1. **Introduction: motivations for similarity distances from Crystal Structure Prediction**

Modern tools of the Crystal Structure Prediction (CSP) produce large datasets of thousands or even millions of simulated crystals based on the same chemical composition. [1] Many of these crystal structures are geometrically similar, because they were obtained as approximations to local minima of a complicated energy. The available similarity tests miss too many nearly identical structures that are ambiguously represented in different ways. That is why Prof Sally Price has summarized the state-of-the-art in CSP as `the embarrassment of over-prediction'.[1]

A dream CSP solution for the pharma industry would be a reliable method to output a short list of only few most stable polymorphs based on a given molecular input. [1] The drug design will be substantially sped up if one can enrich any dataset of simulated crystal structures by a justified distance that shows which structures are geometric neighbors, i.e. close to each other according to this distance, and how such neighborhoods are located relatively to each other. The astronomical analogy is to consider individual crystal structures as stars that form neighborhoods or galaxies. A distance information between visible stars (crystals) will allow one to visualize the whole universe. Such a geometric map of the crystal space[2] will enable a guided navigation in hot spots to further improve properties of known crystals; or a better search in unexplored regions that can contain exotic materials with extraordinary properties.

This paper proposes two new distances between arbitrary crystal lattices that are not restricted to the same crystal system or a Bravais type. All lattices can be continuously deformed into each other. Hence a similarity distance should be well-defined on the whole space of lattices. Section 2 defines necessary concepts and states the equivalence and distance problems for crystals and lattices. Section 3 discusses past approaches to similarities of crystal structures. Section 4 introduces three distances based on the Voronoi cell of an arbitrary lattice. Section 5 shows experimental results on the T2 dataset of simulated and real crystal structures that consists of nano-porous crystals structures that are all based on a single T2 molecule. [2]

2. **Rigorous definitions and problem statements for crystal equivalences and distances**

This section formalizes concepts of periodic crystals, lattices, equivalences and distances.

2.1. The comparison (equivalence) problem for periodic crystal structures

The general model of an ideal crystal is a periodic cloud of zero-sized points representing atoms. The periodicity is determined by a lattice whose nodes are abstract points, not atoms. A *lattice* in the Euclidean space R^3 is a set of points given by integer linear combinations $x\boldsymbol{u} + y\boldsymbol{v} + z\boldsymbol{w}$ of 3 basis vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$, where x, y, z are integer coefficients. If x, y, z are

restricted to the interval [0,1] in the real line, the points $x\boldsymbol{u} + y\boldsymbol{v} + z\boldsymbol{w}$ form a *primitive unit cell*, which is a parallelepiped or a non-rectangular box with parallel opposite sides. The same lattice can be generated by infinitely many bases, hence can have many different unit cells.

A *periodic crystal* is defined by a lattice L and a *motif*, which is a collection of molecules (for molecular crystals) or atoms or ions (in the case of a non-molecular crystals such as NaCl). The motif is periodically translated in the directions along the 3 vectors that define a unit cell of L. Because of many possible unit cells, it is not immediately obvious to decide if crystal structures (or lattices) represented by different unit cells are equivalent in the sense below.

2.2. The equivalence problem and geometric invariants for periodic crystals and lattices

Crystals are often represented by Crystallographic Information Files (CIFs), which contains edge-lengths and angles of a unit cell U and fractional coordinates of atoms in the basis of U, i.e. as numbers within the interval [0,1]. These coordinates are often given for an asymmetric unit that generates a full motif in U by applying symmetry operations specified in a CIF.

Crystals are solid materials, hence are invariant (unaffected) by rigid motions in R^3, which are compositions of rotations and translations. Hence any comparison of crystals should take into account infinitely many positions (of a crystal or its lattice) related by rigid motions in R^3.

Crystals structures (or lattices) are called *equivalent* (or isometric) R^3 if they can be obtained from each other by a rigid motion, which preserves distances between any points in R^3.

This equivalence is the minimal possible one to study crystals as solid materials. For example, if atom positions are perturbed, the perturbed crystal (or its lattice) can be geometrically different even if only slightly. So, the space of equivalence classes of lattices under rigid motions is infinite and continuous (or connected). Hence quantifying a similarity between perturbed crystals is an important problem, which is formalized in the next subsection.

The Bravais classification puts lattices into a much smaller number of classes (only 14 types in dimension 3), though lattices from different classes can be geometrically nearly identical.

Two randomly chosen lattices will share only the translation group of symmetries, hence we need other tools to check if given lattices are not equivalent. Such classification tools are called invariants. An *invariant* of lattices up to a certain equivalence relation, for example rigid motions, is a function that should take the same value on all equivalent lattices.

For example, the volume of a primitive unit cell is an invariant, because all primitive unit cells of a given lattice have the same volume. Edge-lengths and angles of a unit cell are not invariants, because there are infinitely many primitive cells that define the same lattices.

The equivalence problem for lattices is to design a robust algorithm that accepts two arbitrary lattices (without any extra parameters) and decides whether they are equivalent or not. Theoretically, such an algorithm can be based on Niggli's reduced cells[3] in subsection 3.3, and their instability under perturbations[4] motivates the harder distance problem below.

2.3. The distance problem for lattices of periodic crystal structures

This subsection states the metric axioms that are needed to successfully map the space of all crystal structures for any given composition of molecules, atoms or ions. If a distance function between crystal structures satisfies metric axioms below, the crystallography will be open to rigorous methods of metric geometry that will measure what portions of a crystal space are explored and what regions require more sampling in computer simulations.

Let R_+ denote the set of all non-negative real numbers. Let S be any set, e.g. S can be any collection of crystal structures or lattices. A *distance* (or a metric) on S is a function $d: S \times S \to R_+$, such that the following conditions (or metric axioms) hold:

(2.3a) for any $C, C' \in S$, the distance $d(C, C') = 0$ if and only if C, C' are equivalent (equal);

(2.3b) symmetry: $d(C, C') = d(C', C)$ for any $C, C' \in S$;

(2.3c) triangle inequality: $d(C, C') + d(C', C'') \geq d(C, C'')$ for any $C, C', C'' \in S$.

For $S = R^n$, one will use the Euclidean space $d(p, q) = \sqrt{(p_1 - q_1)^2 + \cdots + (p_n - q_n)^2}$ between points $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_n)$, which satisfies the axioms above.

For a set S of crystal structures or arbitrary lattices, it is a hard problem to define a distance function d satisfying the axioms above, because d should not depend on a way to represent crystal structures or lattices, hence should be independent of many potential unit cells.

Axiom (2.3a) avoids trivial examples when a distance d is constant, i.e. has the same value $d(C, C')$ for any non-equivalent crystal structure $C \neq C'$. Axiom (2.3b) says that a distance remains the same if endpoints are swapped. Axiom (2.3c) is motivated by the assumption that a shortest path from C to C'' should not be longer than a combination of shortest paths from C to C' and then from C' to C''. The metric axioms are claimed to be checked for the Euclidean distance between fingerprints proposed by Zhu et al. [5] Any such approach should justify that any non-equivalent crystal structures C, C' have different feature vectors. Else the distance between identical vectors of non-equivalent crystal structures C, C' is 0 and axiom (2.3a) fails.

The *distance problem* for crystal structures (or their lattices) is to find a distance function that satisfies metric axioms (2.3a), (2.3b), (2.3c) above and also the continuity condition below:

(2.3d) the distance $d(C, C')$ continuously changes under perturbations of crystal structures, e.g. if cell parameters or atomic positions are noisy; in particular, the range of d should be a continuous interval, possibly $[0, +\infty)$, but not only a finite collection of discrete values.

One more potentially useful property of similarity distance is invariance under scaling below:

(2.3e) the distance $d(C, C')$ should remain unchanged if both sets $C, C' \subset R^n$ are scaled by the same factor $s > 0$, i.e. $d(C, C') = d(sC, sC')$, where $sC = \{ s\, p \in R^n : \text{for any point } p \in C \}$.

3. Strengths and weaknesses of past approaches to a similarity of crystal structures

3.1. The COMPACK algorithm for the Cambridge Structural Database (CSD)

The widely used COMPACK algorithm[6] for identifying crystal structure similarity requires specified tolerances, e.g. 15% on distance constraints, relative to a reference structure S, and outputs a list of other structures (from a given dataset such as CSD) that are found to be close

to the reference S. The COMPACK output is a single set of similar crystal structures, though a continuous hierarchy ordered by distances to the reference S would be more informative. A numerical measure of similarity for two crystal structures offered by the Mercury software equals the root mean square deviation of atomic positions over finitely many (up to 15 by default) matched molecules or atoms. If this partial matching is extended to the full infinite crystal structures, the deviation of positions will infinitely grow, hence is defined only for finite portions, not for equivalence classes of periodic lattices considered up to rigid motions.

3.2. The COMPSTRU tool at the Bilbao Crystallographic Server (BCS)

Similarly to COMPACK, the recent COMPSTRU algorithm [7] measures a similarity between a given reference structure S and crystal structures whose lattice parameters should be close to those of S (by default 0.5A for distances and 5° for angles). This comparison is restricted to crystal structures that have the same space-group type. A slight perturbation of atomic positions of the reference S will produce a nearly identical crystal that is not comparable to the reference S, hence continuity condition (2.3d) is not satisfied. Many other approaches are based on closest parameters of crystal structures or reduced unit cells discussed below.

3.3. Comparison algorithms based on reduced cells of crystal lattices

Despite any lattice can be defined by infinitely many primitive unit cells, Niggli introduced a reduced cell, which is unique and can be theoretically used for comparing lattices. [3] Niggli's reduced cell is unstable 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. [4]

Figure 1 illustrates the 2-dimensional case of Niggli's reduction, where a narrow an original narrow unit cell is reduced by subtracting multiples of a horizontal basis vector $\boldsymbol{a}$ from another non-horizontal basis vector **b** until the projection of **b** to a is close to 0, i.e. fits the interval $[-0.5|\boldsymbol{a}|, 0.5|\boldsymbol{a}|]$. The endpoints $\pm 0.5|\boldsymbol{a}|$ correspond to two equivalent choices of $\boldsymbol{b}.$ Excluding one of these endpoints will give a unique basis $\boldsymbol{a}, \boldsymbol{b}$, but will make the reduction

discontinuous (unstable), because the two nearly identical vectors $\boldsymbol{b}$ whose projections are

equal to $0.5|\boldsymbol{a}| \pm \delta$ for any tiny $\delta > 0$ will be reduced to different vectors, hence fractional

coordinates of atoms in nearly identical crystal structures will have very different values.

Many software tools offer parameters that shift the perturbation problem to other bounds of

these parameters. The underlying reason of instability is similar to the choice of a range for

angles that can be measured within $[0, 360^o)$ or within $(-180^o, 180^o]$. The distance between

angles should be measured along a shortest round arc on a unit circle, not by breaking a circle

into an interval and taking the difference of angle values from this interval. Indeed, for any

choice of an interval, some angles that are close within a circle become distant in the interval.

Finally, crystal structures are compared (not successfully for Heusler structures [8]) by powder

diffraction patterns up to a cut off radius, which brings discontinuities similarly to other

parameters. To the best of our knowledge there was no distance on equivalence classes of

crystal structures that satisfies all metric axioms (2.3a), (2.3b), (2.3c) and continuity (2.3d).

4. New continuous Voronoi-based distances between arbitrary crystal lattices

This section starts from reminding auxiliary notions of the Hausdorff distance and Voronoi

cells that are needed to introduce two new distances on crystal lattices in subsection 4.3.

4.1. Offsets and the Hausdorff distance between arbitrary sets and crystal structures

For any subset $C \subset R^n$, its *r-offset* $N(C; r)$ consists of all points $p \in R^n$ that are at a

Euclidean distance at most r from C, i.e. $N(C; r) = \{p \in R^n \mid d(p, q) \le r$ for some $q \in C\}$.

 If C is one point, $N(C; r)$ is the ball with the center at C and radius r. If C is a set of points,

$N(C; r)$ is the union of balls that have the radius r and centers at all points of C. A crystal

consisting of the same atoms can be visualized as $N(C; r)$, where r is a bond atomic radius.

The *Hausdorff distance* d_H between any subsets $C, C' \subset R^n$ is the minimum $r \ge 0$ such that

r-offsets of C, C' cover each other, i.e. $C' \subset N(C; r)$ and $C \subset N(C'; r)$, or [10]

$$d_H(C, C') = min\{ r \ge 0 : C' \subset N(C; r) \text{ and } C \subset N(C'; r) \}. \tag{1}$$

For example, $d_H(C, C') = 0$ means that $C \subset C'$ and $C' \subset C$, hence $C = C'$ as needed in axiom (2.3a). This Hausdorff distance was previously used for comparing feature vectors of crystals [11] rather than for crystals. The metric axioms will be proved in the appendix for the extended Hausdorff distance between equivalence classes of lattices introduced in subsection 4.3.

4.2. The Voronoi cell of an arbitrary crystal lattice and its geometric stability

Fix an origin 0 in a lattice $L \subset R^n$. The *Voronoi cell* $V(L)$ is the set of all points $p \in R^n$ that are (not strictly) closer to 0 (in the usual Euclidean distance d) than to all other points of L, i.e. $V(L) = \{p \in R^n \mid d(p, 0) \leq d(p, q) \text{ for any } q \in L - 0\}$, see **Figure 2**. (2)

For a 2D lattice, the Voronoi cell is a rectangle or a centrally symmetric hexagon. [12] For a 3D lattice, a generic Voronoi cell is a truncated octahedron with 7 pairs of parallel faces. The Voronoi cell is centrally symmetric since any lattice is symmetric with respect to its origin. The combinatorial type of $V(L)$ can change under perturbations. But the geometric shape of $V(L)$ changes continuously by Reem's theorem below. A similar stability for Voronoi-based skeletons is known for point clouds. [13] The Voronoi cell of a lattice can be computed from a finite set $L_{[3]}$ of lattice nodes within a factor 3 extension of a suitably reduced [14] cell of L.

Geometric stability of Voronoi cells (simplified Reem's Theorem 5.1[15]) For any small $\epsilon > 0$, there is $r > 0$ such that any lattices L, L' with a small Hausdorff distance $d_H(L_{[3]}, L'_{[3]}) < r$ should have close Voronoi cells with a small Hausdorff distance $d_H(V(L), V(L')) < \epsilon$.

4.3. The rotational extension of the Hausdorff distance to equivalence classes of lattices

The geometric stability above holds for fixed lattices without equivalences up to rigid motions of R^3. For example, if a lattice L is shifted or rotated to a new position, then the new lattice L' remains equivalent to L, but the Hausdorff distance between non-identical Voronoi cells is positive: $d_H(V(L), V(L')) > 0$, which contradicts axiom (2.3a) of a metric. The key idea of the proposed extension from fixed lattices to their equivalence classes (up to rigid motions) is to minimize the distance over all possible rigid motions. Geometrically, the standard

8

Hausdorff distance measures how much one should enlarge each cell to fit into another. The extended distance d_H finds a minimal enlargement to fit one Voronoi cell into another over all rigid motions. Lemma 1 shows that a translation minimizing the Hausdorff distance makes the centers of Voronoi cells identical, which justifies the definition of the extended Hausdorff distance d_H as the minimum over only rotations about the origin, without any translations.

Lemma 1 (proved in Appendix). For any centrally symmetric polyhedra $P, P' \subset R^n$ and a translation T_v by a vector $v \in R^n$, the offset parameter $min\{ r : T_v(P) \subset N(P'; r)\}$ is minimal when T_v moves the center $c(P)$ of the polyhedron P to the center $c(P')$ of P'.

By Lemma 1 the extended Hausdorff distance can be minimized only over all rotations around the common center (say, the origin in R^n) of the Voronoi cells $V(L), V(L')$ as follows. Rigid motions that preserve the origin in R^n are defined by special orthogonal $n \times n$ matrices A such that the determinant of A is 1 and the inverse matrix A^{-1} equals the transpose matrix A^T. All these matrices form the group denoted by $SO(n)$. The group $SO(3)$ consists of rotations around axes that passes through the origin in R^3. Define the non-symmetric offset

$$offset(L, L') = min \{ r \geq 0 : R(V(L)) \subset N(V(L'); r) \}, \tag{3}$$

where the minimum is taken over all rotations $R \in SO(n)$. The *extended Hausdorff* distance for lattices is defined as the symmetric maximum of the two offset parameters:

$$d_H(L, L') = max\{ offset(L, L'), \ offset(L', L) \}, \text{ see **Figure 3**}. \tag{4}$$

Theorem 2 (proved in Appendix). The extended Hausdorff distance d_H is independent of a lattice representation and satisfies the axioms (2.3a), (2.3b), (2.3c) and condition (2.3d).

4.4. The scaling distance and its rotational extension to equivalence classes of lattices

The Hausdorff distance is additive in the sense that if lattices L, L' are scaled by the same factor $s > 0$, then $d_H(sL, sL') = sd_H(L, L')$. To get a distance invariant under scaling as in condition (2.3e), the new distance d_s is introduced first for arbitrary subsets $C, C' \subset R^n$.

Set $scale(C, C') = min\{ s > 0 : C \subset sC' \}$. Then $ln\ (max\{ scale(C, C'), scale(C', C) \})$

satisfies all metric axioms, which is proved below in a more general case for rigid motions.

For any lattices L, L' , their Voronoi cells are used to minimize two symmetric scales for

$V(L), V(L')$ over all rigid motions as in subsection 4.3. Lemma 3 below similarly to Lemma1

justifies that an optimal translation makes the centers of Voronoi cells identical.

Lemma 3 (proved in Appendix). For any centrally symmetric polyhedra $P, P' \subset R^n$ and the

translation T_v by a vector $v \in R^n$, the scale factor $min\{ s > 0 : T_v(P) \subset sP' \}$ is minimal

when T_v moves the center $c(P)$ of the polyhedron P to the center $c(P')$ of the polyhedron P'.

Now the scale factor s can be minimized only over all rotations $R \in SO(n)$ as follows:

$$scale(L, L') = min\ \{ s > 0 : R(V(L)) \subset sV(L') \} \}. \tag{5}$$

The dimension-less *scale-invariant distance* between equivalence classes of lattices is

$$d_s(L, L') = ln(\ max\ \{ scale(L, L'),\ scale(L', L) \}\ ), \text{see } \textbf{Figure 4}. \tag{6}$$

The logarithm above has the base e. Any other base changes d_s only by a constant factor.

The scale-invariant distance can help to recognize similar crystal structures based on scaled

motifs, for example when we `extend' arms of the T2 molecule by adding benzene rings.

Both new distances d_H and d_s are independent of the group symmetry of crystal structures.

Theorem 4 (proved in Appendix). The scale-invariant distance d_s is independent of a lattice

representation, satisfies the axioms (2.3a), (2.3b), (2.3c) and both conditions (2.3d), (2.3e).

5. Computations of new distances for lattices of simulated T2 crystal structures

5.1. A fast algorithm to approximate the new distances d_H and d_s on any lattices

Theorem 5 (proved in Appendix). For any polyhedra $P, P' \subset R^n$ symmetric with respect to 0,

$offset(P, P') = min\{ r > 0 : P \subset N(P'; r) \}$ and $scale(P, P') = min\{ s > 0 : P \subset sP' \}$ can

be computed in a linear time with respect to the numbers of vertices and faces of P, P'.

The Voronoi cell of any lattice in R^3 has at most 24 vertices and 14 faces, see Figure 1. Hence

Theorem 5 guarantees a constant upper bound not depending on lattices L, L'. The standard

way to optimize over all rotations $R \in SO(3)$ is to uniformly sample all rotations by using these parameters: a unit length vector $v \in R^3$ and an angle $\theta \in [0, 360^o)$ of the rotation around v. Take v in the upper hemisphere (to avoid opposite vectors giving the same axis) as

$$v = \left(\sqrt{1 - z^2} \cos \mu, \sqrt{1 - z^2} \sin \mu, z\right) \tag{7}$$

If the height parameter $z \in (0,1)$ has n samples, then the angles μ, θ have $[2\pi n]$ samples. In total, about $4\pi^2 n^3$ rotations from $SO(3)$ will be sampled. The experiments below use $n = 3$, hence more than 1000 sampled rotations. Any vector $u \in R^3$ is rotated by Rodrigues' formula

$$u \mapsto u \cos \theta + (v \times u) \sin \theta \, v(v \times u)(1 - \cos \theta). \tag{8}$$

5.2. Simulated and synthesized organic crystal structures based on the T2 molecule

The Nature paper by Pulido et al. [2] has demonstrated that functional organic materials can be discovered by simulating crystals build on a molecule with a desired function via costly optimizations of the energy and target properties such as gas adsorption. Only 5 crystals in **Figure 5** were synthesized, though predictions of properties were run for all crystals. The key bottleneck in this approach is the time-consuming prediction for nearly identical simulated crystals. For example, producing the CSP landscape energy-vs-density for 5688 crystals based on the T2 molecule has taken many weeks of the supercomputer time, see **Figure 6**.

The standard continuous similarity measures such as the energy and density are not enough to reliably quantify differences between crystals, because the same or almost identical energy and density cannot guarantee geometric similarity. The experiments on the T2 dataset of 5688 simulated crystal structures have found numerous pairs of crystals, e.g. with IDs (41,47), (68,71), (63,73), (71,83), (71,93), which have energy differences within 3 KJ/mol and also density differences within 0.01 g/cm^3. However, these crystal structures have extended Hausdorff distances $d_H \geq 15$ Angstroms and scale-invariant distances $d_s \geq 1.1$, i.e. with scale factors more than $e^{1.1} \approx 3$. Crystal lattices 41 and 47 have unit cell angles close to 90°, but very different unit cell sides: (53.3, 23.7, 7.3), (15.4, 12.9, 16.5). These differences in their

structure were found only now by using the new distances, not by the energy and density. Mercury has managed to match only 1 of attempted 50 T2 molecules in these structures, so the found deviation of positions is 0, because both crystals consist of the same T2 molecules.

5.3. Similarity matrices (grayscale maps) of the new distances on the T2 crystal lattices

For each of the distances d_H and d_s, all values (from a minimum to a maximum) were linearly scaled to the range [0,255] to visualize the distance matrix in grayscale. **Figure 6** shows the distances (rounded to integers) between 5 real T2 crystal structures.

Figure 7 shows the larger experiment on the first 100 crystals from the T2 dataset. Both distances d_H and d_s were computed for all 4950 unordered pairs of these 100 crystals, which took in total 6 hours on a modest laptop. The variability of intensities in both heatmaps justifies that d_H and d_s take many values not restricted to a small discrete set. The full 100×100 matrices for d_H and d_s are in the supplementary materials. The scale-invariant distance d_s shows more variability of colors in the heatmap in Figure 7, which confirms the usefulness of scale condition (2.3e), which led to the new scale-invariant distance d_s. The C++ code "Lattice Distances" for the new distances d_H and d_s will be available soon.

6. Conclusions: contributions to the state-of-the-art justified by proofs and experiments

- The new (extended Hausdorff d_H and scale-invariant d_s) distances are defined for equivalence classes of arbitrary lattices considered up to any rigid motions in R^n, hence are independent of choice of unit cells or coordinates in crystal representations.

- Both distances satisfy the metric axioms and the continuity under perturbations proved in Theorems 2 and 4, which allows one to quantity similarities in a continuous way. Such a quantification is the important step to produce continuous hierarchies of crystal structures and visualize patterns of clusters changing for a varied distance threshold.

- Experiments in section 5 on real and simulated T2 crystal structures [2] show that d_H and d_s better distinguish crystal lattices that have almost identical energy and density.

Received: ((will be filled in by the editorial staff))

Revised: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))

References

[1] A discussion with S.Price about her paper in *Faraday discussions*, **2018**, *211*, pp.9-30.

[2] A.Pulido, L.Chen, T.Kaczorowski, D.Holden, M.A.Little, S.Y.Chong, B.J.Slater, D.P.McMahon, B.Bonillo, C.J.Stackhouse, A.Stephenson, C.M.Kane, R.Clowes, T.Hasell, A.J.Cooper, G.M.Day, *Nature*, **2017** 543 (7647): 657.

[3] P.Niggli, Handbuch der Experimentalphysik, v.7, part 1, Akad. Verlagsgesellschaft, **1928**.

[4] L.C.Andrews, H.J.Bernstein, G.A.Pelletier, *Acta Crystallograph. A*, **1980**, 36(2), 248-252

[5] L.Zhu, M.Amsler, T.Fuhrer, B.Schaefer, S.Faraji, S.Rostami, S.A.Ghasemi, A.Sadeghi, M.Grauzinyte, C.Wolverton, S.Goedecker, *J. Chemical Physics,* **2016**, 144, 3, 034203.

[6] J.A.Chisholm, S.Motherwell. *J. Appl. Cryst,* **2005**, 38(1), 228-231.

[7] G.Flor, D.Orobengoa, E.Tasci, J.M.Perez-Mato, M.I.Aroyo, *J. Appl. Cryst*., **2016**, 49(2), 653-664.

[8] Oliynyk AO, Antono E, Sparks TD, Ghadbeigi L, Gaultois MW, Meredig B, Mar A. *Chemistry of Materials*. **2016**, ;28(20):7324-31.

[9] Blatov, I.A., Kitaeva, E.V., Shevchenko, A.P. and Blatov, V.A., *Acta Crystallogr. A*, **2019**, *75*(6).

[10] Kurlin V., *Int. Conference on Computer Analysis of Images and Patterns*, **2015**, 606-617.

[11] A.Collins, C.Wilson, C.Gilmore, *Cryst Eng Comm*, **2010**, 12, 801-809.

[12] Edelsbrunner, H., Iglesias-Ham, M. and Kurlin, V., *Canadian Conf. Comp. Geometry*, **2015**.

[13] Kurlin V., In *Proceedings of Computational Topology in Image Context*, **2014**, 594-601.

[14] N.Dolbilin, D.Huson, *Periodica Mathematica Hungarica*, **1997**, 34(1-2), 57-64.

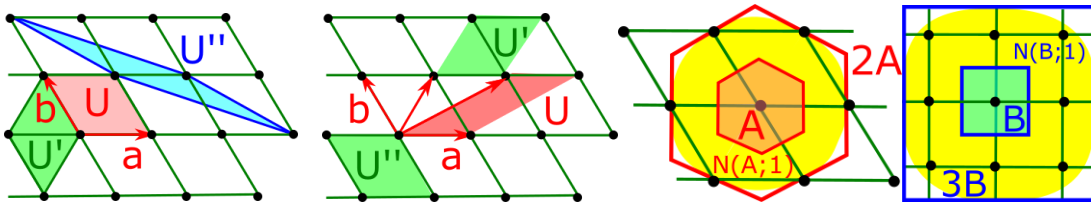[15] D.Reem, In *Proceedings of the 27th Symposium on Comp. Geometry,* **2011**, 254-263.

13

Figure 1. (**1st**: any lattice has infinitely many primitive cells, e.g. U, U', U''. **2nd**: Niggli's reduction of a vector v_2 relative to v_1 can lead to two cells U' and U'', a choice is unstable. **3rd**: the yellow offset $N(A; 1)$ of A is inside the scaled cell $2A$. **4th**: yellow offset $N(B; 1)$.)
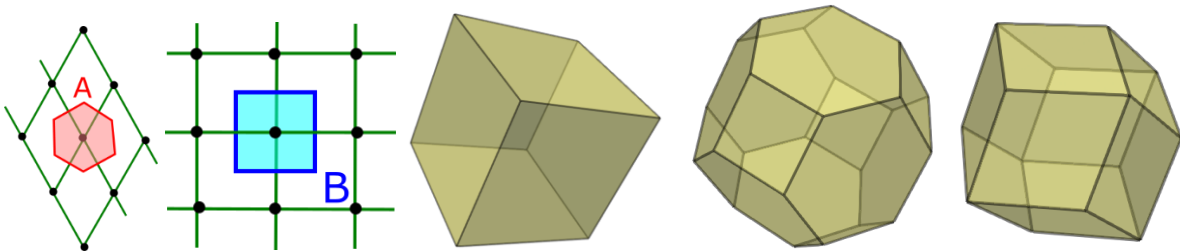

Figure 2. (The Voronoi cells of lattices: hexagonal, square, cubic, body-centered cubic (BCC), face-centered cubic (FCC), see definitions in subsection 4.2. the cubic lattice has the standard basis $(1,0,0), (0,1,0), (0,0,1)$; the body-centered cubic (BCC) lattice has the basis $(1,0,0), (0,1,0), (0.5,0.5,0.5)$; the FCC lattice has the basis $(1,0,0), (0.5,0.5,0), (0.5,0,0.5)$.)
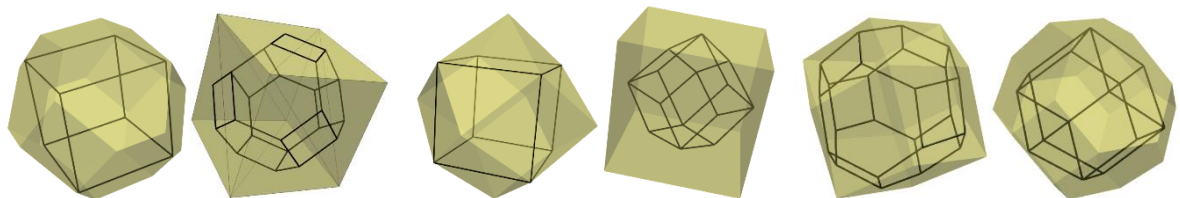
Figure 3. (Illustrations of extended Hausdorff distances: one Voronoi cell $V(L)$ is optimally rotated and inscribed into a minimal offset Voronoi cell $V(L')$ for pairs L, L' from the left to right: (cubic, BCC), (BCC, cubic), (cubic, FCC), (FCC, cubic), (BCC, FCC), (FCC, BCC).)
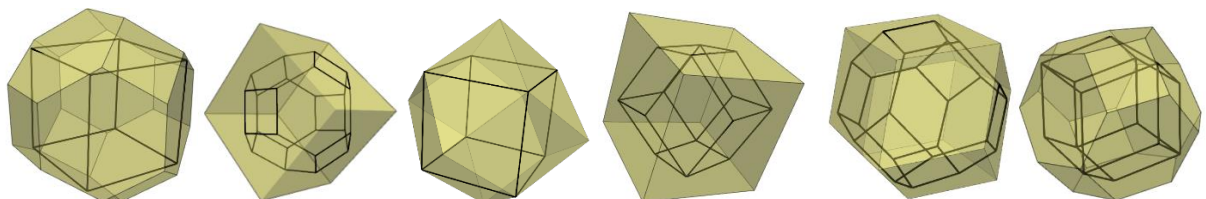

Figure 4. (Illustrations of scale-invariant distances $d_s(L, L')$ for pairs L, L' from the left to right: (cubic, BCC), (BCC, cubic), (cubic, FCC), (FCC, cubic), (BCC, FCC), (FCC, BCC).)
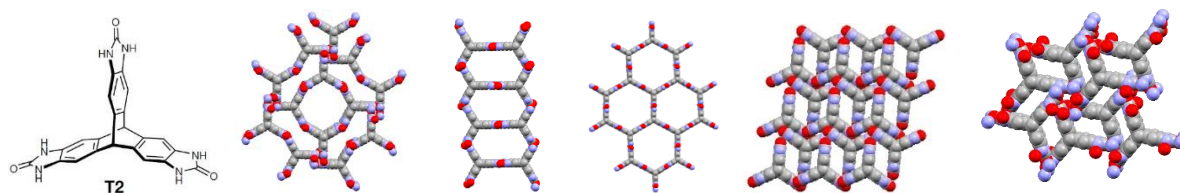
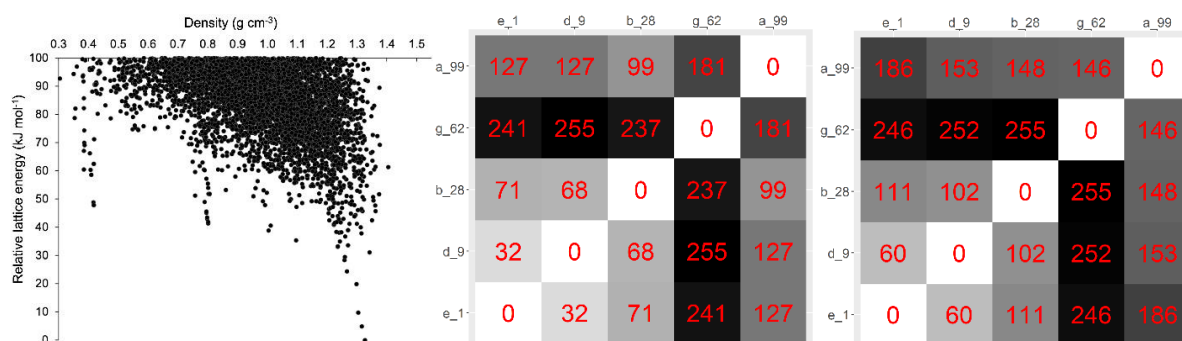**Figure 5**. (T2 molecule: triptycenetrisbenzimidazolon; crystals T2α,T2β,T2γ,T2δ,T2ε. [2])

Figure 6. (Left: energy-vs-density plot of simulated structures.[2] Middle: extended Hausdorff distances d_H, Right: scale-invariant distances d_s for real structures T2α(a_99), T2β(b_28), T2γ(g_62), T2δ(d_9), T2ϵ(e_1). All values are scaled to [0,255] and rounded to integers.)
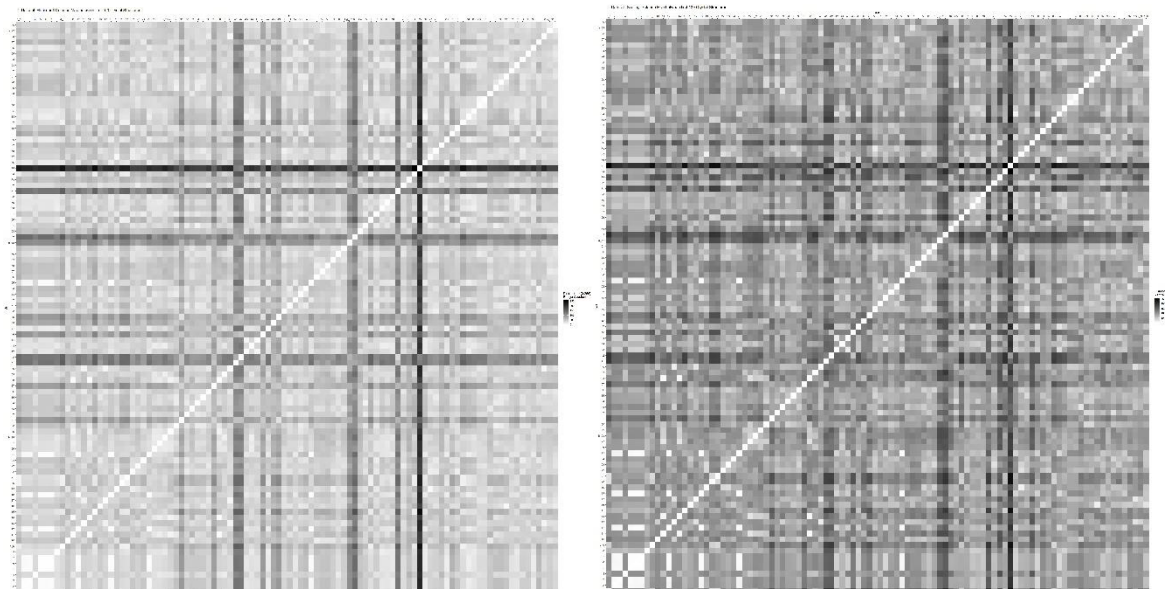

Figure 7. (Left: extended Hausdorff distances d_H, Right: scale-invariant distances d_s for first 100 of 5688 simulated crystal structures reported in the Nature paper. [2] All values are linearly scaled to [0,255] and displayed in a grayscale heatmap similarly to Figure 6.)