Continuous invariant-based asymmetries of periodic crystals quantify deviations from higher symmetry

Surya Majumder,[a] Daniel Widdowson,[b] Yury Elkin,[a] Olga Anosova,[a,c] Andrew I. Cooper,[b] Graeme M. Day[d] and Vitaliy A. Kurlin [a,b,c*1]

[a]*Computer Science department, University of Liverpool, Liverpool, L69 3BX, UK*, [b]*Materials Innovation Factory, University of Liverpool, Liverpool, L7 3NY, UK*, [c]*National Institute for Theory and Mathematics in Biology, Chicago, US, and* [d]*School of Chemistry and Chemical Engineering, University of Southampton, Southampton, SO17 1NX, UK. E-mail: vkurlin@liv.ac.uk*

Abstract

Ideal symmetry is known to break down under almost any noise. One measure of asymmetry in a periodic crystal is the relative multiplicity Z' of geometrically non-equivalent units. However, Z' discontinuously changes under almost any displacement of atoms, which can arbitrarily scale up a primitive cell. This discontinuity was recently resolved by a hierarchy of invariant descriptors that continuously change under all small perturbations. We introduce a Continuous Invariant-based Asymmetry (CIA) to quantify (in physically meaningful Angstroms) the deviation of a periodic crystal from a higher symmetry form. Our experiments on several Crystal Structure Prediction datasets show that about a half of simulated crystals have high values of CIA, while all experimental structures in these datasets have CIA $= 0$. On another hand, many crystals with high values Z' in the Cambridge Structural Database (CSD) turned out to be close to more symmetric forms with $Z' \leq 1$ due to low values of CIAs.

[1]

1. Introduction: motivations for a new continuous asymmetry of crystals

Many periodic crystals are highly symmetric, because a globally stable structure is usually formed by a few energetically favourable interactions, bonds, molecules, or formula units, which are repeated in $\mathbb{R}^3$ by symmetries (Lax, 2001). Though we were motivated by molecular crystals, our invariant-based approach to asymmetry extends to all non-molecular crystals and periodic sets in any Euclidean space $\mathbb{R}^n$.

While molecular crystals can contain many molecules in primitive unit cells, they are often obtained from a smaller number of molecules by *symmetry operations* preserving the whole crystal in $\mathbb{R}^3$ (Chapuis, 2024). For a non-molecular crystal, the chemical analogue of a molecule is a *formula unit* that is an electronically neutral group of atoms or ions, embedded in $\mathbb{R}^3$ and representing their relative numbers in a given compound, reduced to the smallest integer numbers. For example, table salt has the empirical formula NaCl with a formula unit consisting of two ions Na^+ and Cl^-. However, this pair of ions can be chosen in many geometrically different ways, because ionic bonds in NaCl do not define a bounded object, such as a molecule. Hence, non-molecular crystals should be unambiguously split into disjoint *geometric blocks*, for example, single ions, or metal blocks and organic linkers in a metal organic framework.

In this paper, a *crystal* S means a periodic crystal, while Z can be non-integer for disordered or aperiodic crystals (Senechal, 1996). The *multiplicity* Z usually denotes the number of formula units in a primitive unit cell. The *relative multiplicity* Z *prime* was often defined as $Z(S)$ divided by the number of independent general positions (Steed & Steed, 2015), which makes sense, if S consists of chemically equivalent molecules. For crystals with chemically different molecules (called co-crystals), (Van Eijck & Kroon, 2000) used another notation Z'' for the number of crystallographically non-equivalent molecules. To cover non-molecular crystals, we define Z' below for any periodic point set $S \subset \mathbb{R}^n$ with a given splitting into geometric blocks.

Definition 1 (relative multiplicity Z')**.** An *asymmetric unit* is a minimal, closed, and simply connected subset A of a unit cell of $S \subset \mathbb{R}^b$, whose images under all symmetry operations of S tile the whole space $\mathbb{R}^n$. Let $S \cap A$ split into geometric blocks $B_1, \dots, B_G$, which should be chemically different molecules, ions, or other disjoint blocks for crystals in $\mathbb{R}^3$. Let B_i have n_i symmetry operations (including the identity) that preserve both S and B_i, $i = 1, \dots, G$. The *relative multiplicity* is $Z'(S) = \sum\limits_{i=1}^{G} \dfrac{1}{n_i}$.

Geometric blocks B_i, B_j of $S \subset \mathbb{R}^n$ are called *rigidly equivalent* if there is a rigid motion of $\mathbb{R}^n$ that maps S to S and B_i to B_j. If all molecules of a crystal S are rigidly equivalent, an asymmetric unit A of S contains one molecule B, so $Z'(S) = \dfrac{1}{n} = \dfrac{Z}{N}$, where n is the number of symmetry operations preserving both S and B, while $N = nZ$ is the number of symmetry operations preserving S and the motif $S \cap U$, which can permute molecules within a primitive cell U of S. If $S \cap A$ consists of two non-rigidly equivalent molecules in 2-fold positions, then $Z'(S) = \dfrac{1}{2} + \dfrac{1}{2} = 1$. The crystal NaCl has ions Na^+ and Cl^- with point groups of order 48, so $Z'(\mathrm{NaCl}) = \dfrac{1}{48} + \dfrac{1}{48} = \dfrac{1}{24}$.

In about 90% of entries in the Cambridge Structural Database (CSD), an asymmetric unit includes only one molecule, so $Z' \leq 1$ (Anderson *et al.*, 2006). However, the CSD has many crystals with high Z' (Brock, 2016), e.g. OGUROZ has $Z' = 56$.

Crystal Structure Prediction (CSP) often starts with simulating $Z' = 1$ crystals for the most frequent space groups, but a final energy relaxation can produce structures with Z' values up to 36 (Pulido *et al.*, 2017). More importantly, almost any displacement of atoms or a whole rigid molecule discontinuously changes the size of a primitive (or reduced) cell and hence arbitrarily scales up Z'. Fig. 1 shows nearly identical structures with $Z' = 1, 2, 3$ and similarly applies to any periodic crystal.

Ignoring any noise up to a small threshold ε only shifts the problem from 0 to another number without guarantees of a continuous change. This *sorites* paradox (when a heap of sand stops being a heap while grains are removed one by one) has been known since

ancient times (sor, 2024). Its rigorous solution requires an *invariant* that is preserved by any rigid transformation and continuously changes under perturbations of atoms.

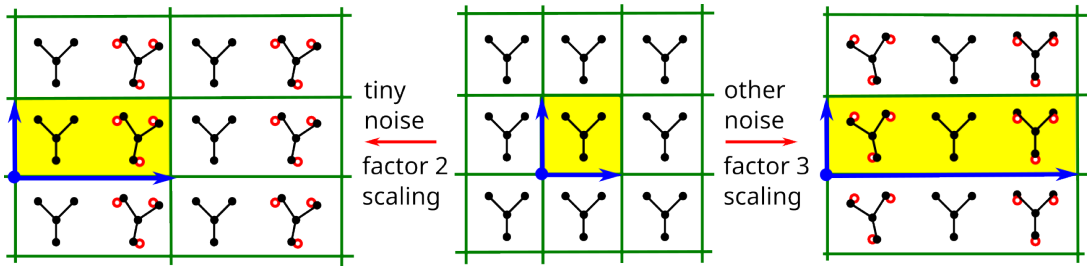

Fig. 1. Almost any noise arbitrarily scales up a primitive yellow cell and discontinuously changes the relative multiplicity Z' of molecules, which are represented by black Y graphs whose terminal vertices have initial positions shown by red circles.

While a full hierarchy of such invariants for periodic crystals from the fastest to complete is being finalised by (Anosova & Kurlin, 2025; Widdowson & Kurlin, 2025*b*), our continuous asymmetry will be based on the fast invariant PDD (Pointwise Distance Distribution), which distinguishes all non-duplicate crystals in the world's largest databases within two hours on a modest desktop (Widdowson & Kurlin, 2022).

2. Generically complete and continuous isometry invariants of crystals

This section recalls isometry invariants, which will be used to define a continuous invariant-based asymmetry in section 3. Any linear basis $\boldsymbol{v_1},\dots,\boldsymbol{v_n}$ of $\mathbb{R}^n$ generates the *lattice* $\Lambda=\{c_1\boldsymbol{v_1}+c_2\boldsymbol{v_2}+\dots+c_n\boldsymbol{v_n}\mid c_1,\dots,c_n\in\mathbb{Z}\}\subset\mathbb{R}^n$ and the *unit cell* $U=\{x_1\boldsymbol{v_1}+x_2\boldsymbol{v_2}+\dots+x_n\boldsymbol{v_n}\mid 0\le x_1,\dots,x_n<1\}\subset\mathbb{R}^n$.

For any finite motif $M\subset U$ of atoms (considered zero-sized points) in the unit cell, a periodic crystal is defined as the infinite set $S=\Lambda+M=\{\boldsymbol{v}+p\mid \boldsymbol{v}\in\Lambda,p\in M\}$ or a finite union $\cup_{p\in M}(\Lambda+p)$ of shifted lattices with origins at all points of M.

The definition above is widely used for representing crystals in Crystallographic

Information Files (CIFs), but is highly ambiguous in the sense that infinitely many pairs (basis, motif) represent the same crystal S. This ambiguity motivated us to distinguish between a crystal S and its *structure*, defined as the equivalence class of all periodic sets of atoms that are represented by different CIFs but can be exactly matched with each other by rigid motion, see Definition 6 in (Anosova *et al.*, 2024).

Any canonical (standard or conventional) choice of a cell for a periodic crystal is discontinuous under almost any noise, as in Fig. 1, which was experimentally demonstrated already in 1965, see p.80 in (Lawton & Jacobson, 1965). The new definition of a *crystal structure* as a rigid class (consisting of all crystals that can be exactly matched under rigid motion) has become practical due to the hierarchy of invariants that uniquely identify any crystal structure independent of its initial representation.

Definition 2 introduces the invariant PDD for any periodic set of points in $\mathbb{R}^n$, which can be all atomic centres of a crystal in $\mathbb{R}^3$, or other points defined by a crystal, for example, atoms of one specific type, or molecular centres, which form a periodic set.

Definition 2 (Pointwise Distance Distribution PDD)**.** Let $S \subset \mathbb{R}^n$ be a periodic point set with a motif $M = \{p_1, p_2, \dots, p_m\}$. Fix an integer $k \geq 1$. For every point $p_i \in M$, let $d_1(p) \leq \dots \leq d_k(p)$ be the distances from p to its k nearest neighbours within the full infinite set S not restricted to any cell. The matrix $D(S;k)$ has m rows consisting of the distances $d_1(p_i), \dots, d_k(p_i)$ for $i = 1, \dots, m$. If any $l \geq 1$ rows are identical to each other, we collapse them into a single row and assign the weight $\dfrac{l}{m}$ to this row. The resulting matrix of k columns and a maximum of m rows with weights, in the extra (say, 0-th) column, is called the Pointwise Distance Distribution PDD$(S;k)$.

The columns of the matrix PDD$(S;k)$ are ordered because each row consists of increasing values of distances to neighbours but without their indices. So PDD$(S;k)$ importantly differs from the matrix of pairwise distances between m points in the motif M, also because neighbours are not restricted to any (extended) cell of S.

Since many crystals consist of indistinguishable atoms, we consider all points of S unordered. Then $\mathrm{PDD}(S;k)$ has unordered rows and can be interpreted as a discrete distribution of rows (or unordered points in $\mathbb{R}^k$) with probabilities equal to the weights assigned to rows. The Pair Distribution Function is obtained from a single collection of all interatomic distances (usually normalised by frequencies and then smoothed) and hence is naturally weaker than $\mathrm{PDD}(S;k)$, which splits distances per point and avoids losing information under smoothing, see the discussion at the end of section 3 in (Widdowson & Kurlin, 2022). This probabilistic interpretation allows one to compare PDDs by many distance metrics on discrete distributions. We usually use the simplest metric called Earth Mover's Distance (EMD), which was adapted for comparing chemical compositions (Hargreaves *et al.*, 2020). Theorem 4.2 in (Widdowson & Kurlin, 2025*c*) proved that $\mathrm{PDD}(S;k)$ continuously changes in EMD under perturbations, including those that arbitrarily scale up a minimal cell as in Fig. 1.

The most important result about the PDD is its generic completeness: Theorem 5.8 in (Widdowson & Kurlin, 2025*c*) proved that $\mathrm{PDD}(S;k)$ with a lattice of S and the number m of points in a motif of S suffice to reconstruct any generic periodic point set $S \subset \mathbb{R}^n$, uniquely under isometry, for a large enough k with an explicit upper bound. In other words, $\mathrm{PDD}(S;k)$ with a few extra invariants provably distinguishes all crystals, possibly except singular examples that form a subspace of measure 0 within the continuous space of all periodic crystals. In practice, $\mathrm{PDD}(S;k)$ distinguished all non-duplicate crystals in the world's major databases within two hours on a modest desktop, see Table 3 in (Widdowson & Kurlin, 2025*c*). Theorem 3.7 in (Widdowson & Kurlin, 2025*c*) proved that, as $k \to +\infty$, the distances in each row of $\mathrm{PDD}(S;k)$ asymptotically approach $\mathrm{PPC}(S)\sqrt[n]{k}$, where the Point Packing Coefficient $\mathrm{PPC}(S)$ is inversely proportional to the point density, as defined below. This fact motivated us to subtract this asymptotic curve from $\mathrm{PDD}(S;k)$ to neutralise the influence of density.

Definition 3 (invariants PPC(S) and PDA($S;k$))**.** Let $S \subset \mathbb{R}^n$ be a periodic set with m points in a unit cell U of S. The *Point Packing Coefficient* is $\mathrm{PPC}(S) = \sqrt[n]{\dfrac{\mathrm{vol}(U)}{mV_n}}$, where vol($U$) is the volume of U, and V_n is the volume of the unit ball in $\mathbb{R}^n$, e.g. $V_3 = \frac{4}{3}\pi$. The *Pointwise Deviation from Asymptotic* is the matrix PDA($S;k$) obtained from PDD($S;k$) by subtracting PPC(S) $\sqrt[n]{j}$ from every distance in columns $j = 1, \ldots, k$.

Another advantage of PDA($S;k$) vs original PDD($S;k$) is the experimental convergence to 0 of the k-th values from the last column of PDA($S;k$) as $k \to +\infty$, see Fig. 4 in (Widdowson & Kurlin, 2025a). This convergence to 0 was formally proved for any cubic lattice $\mathbb{Z}^n$ in Example SM3.1 from (Widdowson & Kurlin, 2025c). Then there is no need to substantially increase the number k of neighbours, because more distant neighbours bring smaller contributions. We consider k not as a parameter that seriously affects PDA($S;k$), but as a degree of approximation like the number of decimal places on a calculator. The vector ADA($S;k$) of column averages in PDA($S;k$) for $k = 100$ atomic neighbours was sufficient to distinguish all non-duplicate crystals in the CSD (Widdowson & Kurlin, 2024). The components of this *Average Deviation from Asymptotic* vector ADA($S;k$) can be used as analytic coordinates on geographic-style maps of any materials database. Such maps were first developed for 2D lattices by (Bright *et al.*, 2023b; Bright *et al.*, 2023a; Kurlin, 2024).

3. A continuous invariant-based asymmetry (CIA) of periodic crystals

The discontinuity of Z' from Definition 1 under almost any perturbation has been known for 30+ years. The quote "two fairly unsymmetrical objects can be combined into a less unsymmetrical structural dimer" from (Wilson, 1993) means that a crystal with $Z' = 2$ can be geometrically close to a more symmetric crystal with $Z' = 1$.

This section first defines the Earth Mover's Distance (EMD) between geometric blocks within a periodic point set $S \subset \mathbb{R}^n$ by using the isometry invariant PDA($S;k$)

from Definition 3. The continuous invariant-based asymmetry of S will be defined through EMDs between all blocks in an asymmetric unit of S. The EMD needs a ground distance between vectors $\boldsymbol{b} = (b_1, \dots, b_k)$ and $\boldsymbol{c} = (c_1, \dots, c_k)$ in $\mathbb{R}^k$, such as rows of PDA$(S;k)$. The simplest choices are the *Chebyshev* distance $d_\infty(\boldsymbol{b},\boldsymbol{c}) = \max\limits_{1\le i\le k} |b_i - c_i|$ and the *Root Mean Square* (RMS) $d(\boldsymbol{b},\boldsymbol{c}) = \sqrt{\dfrac{1}{k}\sum\limits_{i=1}^{k}(b_i - c_i)^2}$.

These distances respect the continuity under perturbations as follows. If any b_i, c_i are perturbed up to ε, then $|b_i - c_i| \le 2\varepsilon$ for $i = 1, \dots, k$, and both $d_\infty(\boldsymbol{b},\boldsymbol{c}) \le 2\varepsilon$, $d(\boldsymbol{b},\boldsymbol{c}) \le 2\varepsilon$. We usually write d without a subscript for brevity. If d_∞ is used in computations, all relevant distances and asymmetry will have the subscript ∞.

For any periodic set in $\mathbb{R}^n$, Definition 4 introduces a distance between geometric blocks B, C (considered as finite sets of points), which are molecules, ions, or other well-defined disjoint subsets for crystals in $\mathbb{R}^3$. This distance measures how the positions of B, C differ within a common periodic set S containing both B, C. If B, C can be exactly matched by a rigid motion of $\mathbb{R}^n$ preserving S, then this distance is 0. In all real examples, any deviation from symmetry should be positive because of noise.

Though the EMD makes sense for distributions of different sizes, our experiments on crystals will use the EMD only for geometric blocks that are chemically identical molecules. More generally, we assume that every point in a periodic set $S \subset \mathbb{R}^n$ has a categorical label, which is an analogue of an atomic type, such as Na^+ and Cl^-.

Briefly, the EMD optimally splits and transports objects from one distribution to another by minimising the overall cost based on a ground distance between objects. If we need to guarantee matching of points only with the same label (atomic type for crystals), the ground distance can be adjusted by taking the maximum of d_∞ or $d = \text{RMS}$ with a discrete metric that is infinite between points of different labels.

Definition 4 (Earth Mover's Distance EMD between geometric blocks)**.** Let $S \subset \mathbb{R}^n$ be a periodic set of labeled points with an asymmetric unit A. Let $B, C \subset S \cap A$ be

geometric blocks (finite sets) that have $m(B), m(C)$ points of weights $\frac{1}{m(B)}, \frac{1}{m(C)}$, respectively. For $i = 1, \dots, m(B)$ and $j = 1, \dots, m(C)$, let $R_i(B), R_j(C)$ be the rows of i-th and j-th points in B, C, respectively. The distance below is independent of point ordering. The *Earth Mover's Distance* $\mathrm{EMD}(B, C) = \sum\limits_{i=1}^{m(B)} \sum\limits_{j=1}^{m(C)} f_{ij} d(R_i(B), R_j(C))$ is minimised over variable parameters $f_{ij} \in [0, 1]$ subject to $\sum\limits_{j=1}^{m} f_{ij} = \frac{1}{m(B)}$ and $\sum\limits_{j=1}^{m} f_{ij} = \frac{1}{m(C)}$ for all $i = 1, \dots, m(B)$ and $j = 1, \dots, m(C)$, respectively.

The distance $\mathrm{EMD}(B, C)$ measures the minimum perturbation of the rows of the geometric blocks B, C in $\mathrm{PDA}(S; k)$ to match (distance-based invariants of) B and C within the ambient periodic set S. This perturbation matching B and C reduces the number of geometrically non-equivalent blocks and hence makes S more symmetric.

If an asymmetric unit A of S has only one geometric block B, then S has no asymmetry because all blocks in S are images of B under symmetry operations of S. If A has only two blocks B, C, then $\mathrm{EMD}(B, C)$ is considered the asymmetry of S. In more general cases, Definition 5 introduces the continuous asymmetry below.

Definition 5 (Continuous Invariant-based Asymmetry $\mathrm{CIA}(S)$)**.** Let a periodic set $S \subset \mathbb{R}^n$ with labeled points have geometric blocks $B_1, \dots, B_G$ its asymmetric unit. Set $d_i = \max\limits_{j=1,\dots,G} \mathrm{EMD}(B_i, B_j)$ for $i = 1, \dots, G$. The *Continuous Invariant-based Asymmetry* is $\mathrm{CIA}(S) = \min\limits_{i=1,\dots,G} d_i$. The 'average' version is $\overline{\mathrm{CIA}}(S) = \frac{1}{G} \sum\limits_{i=1}^{G} d_i$.

The matrix of distances $\mathrm{EMD}(B_i, B_j)$ describes the relative positions of G blocks within an asymmetric unit of S in terms of their distances to atomic neighbours within the full S. For $i = 1, \dots, G$, the distance d_i measures how far B_i is from all other blocksThe standard (min-max) formula of $\mathrm{CIA}(S)$ means that the optimal i-th block B_i serves as a centre minimising its distance $\mathrm{EMD}(B_i, B_j)$ to the farthest block B_j, while $\overline{\mathrm{CIA}}(S)$ averages maximum deviations d_i from all blocks considered as centres. The default notation $\mathrm{CIA}(S)$ uses EMD based on the ground distance

d = RMS between rows of PDA$(S;k)$ with $k = 100$. For the Chebyshev distance d_∞, we keep the subscript ∞ in the notations EMD$_\infty$, CIA$_\infty$, and $\overline{\text{CIA}}_\infty$.

Lemma 6 (invariance of CIAs)**.** All CIAs in Definition 5 are invariant (remain unchanged) under any isometry and changes of a unit cell of a periodic point $S \subset \mathbb{R}^n$.

Lemma 6 and all further results below are proved in appendix B.

Lemma 7 (inequalities for CIAs)**.** In the notations of Definition 5, CIA $\le$ CIA$_\infty$, $\overline{\text{CIA}} \le \overline{\text{CIA}}_\infty$, and CIA $\le \overline{\text{CIA}} \le$ 2CIA hold for any periodic point set $S \subset \mathbb{R}^n$.

Since Definition 5 is based on the invariant PDA$(S;k)$, the full notation should be CIA(PDA$(S;k)$), where PDA$(S;k)$ can be replaced with another "pointwise" invariant, such as the higher-order PDA$^{(h)}$ (Widdowson & Kurlin, 2025*b*) or complete isoset (Anosova *et al.*, 2025). In this paper, we use only PDA$(S;100)$ and write CIA(S) for brevity. Theorem 8 justifies the continuity of the asymmetry CIA(S) under all small perturbations of points, including those that arbitrarily scale up an initial cell of S.

Theorem 8 (continuity of CIA under perturbations)**.** Let $S \subset \mathbb{R}^n$ be a periodic point set and $r(S)$ denote the minimum half-distance between any points of S. For any $0 < \varepsilon < r(S)$, let a periodic set $Q \subset \mathbb{R}^n$ be obtained by perturbing every point of S up to Euclidean distance ε. Then the CIAs based on the invariant PDA$(S;k)$ for any k in Definition 5 satisfy $|\text{CIA}(S) - \text{CIA}(Q)| \le 4\varepsilon$ and $|\overline{\text{CIA}}(S) - \overline{\text{CIA}}(Q)| \le 4\varepsilon$.

4. Fast detection of asymmetric crystals in large simulated CSP datasets

This section visualises several versions of CIA for 50+ thousand simulated crystals from four CSP datasets reported in (Pulido *et al.*, 2017). At that time, the synthesised crystals predicted by these CSPs substantially extended the small population of nanoporous crystals in the CSD. However, these predictions took more than 12 weeks on a supercomputer, also due to predictions of properties, such as gas capture.

In these cases, all experimental crystals have an asymmetric unit consisting of a single molecule, hence CIA = 0 for all versions, which confirms the symmetry principle saying that real crystals tend to be highly symmetric. All simulated crystals in the four CSP datasets are based on one of the four molecules in Fig. 2.

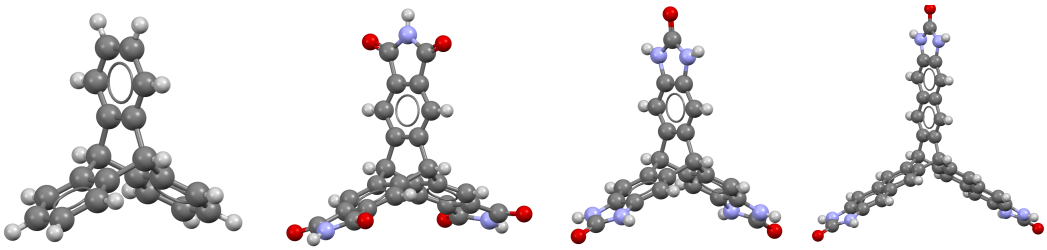

Fig. 2. T0, T1, T2, and T2E molecules in the four CSP datasets in this section.

Since each molecule has a rigid shape of three symmetric 'arms', its position in $\mathbb{R}^3$ is uniquely determined by 3 base points at the ends of these 'arms' that are most distant from the molecular centre. We selected the following 3 base points for each molecule. T0: mid-points defined by 3 pairs of the most distant carbons from the centre. T1: three nitrogens. T2 and T2E: three oxygens. All values of CIAs in this section were computed on periodic sets obtained by replacing each molecule with its three base points. The alternative option of considering all atoms is slower and unnecessary in these cases, because three base points per molecule suffice to completely reconstruct every crystal based on one of the molecules T0, T1, T2, and T2E in Fig. 2.

Fig. 3 has four histograms of the default CIA across four CSP datasets. In each histogram, the vertical y-axis shows the number of crystal structures on the log scale (as powers of 10) whose CIAs fall in a bin of size 0.01Å. The first vertical bin with CIA = 0 represents all crystals with CIA = 0. Since any CIA in Definition 5 is a min-max or an average of non-negative distances, all versions of CIAs vanish simultaneously.

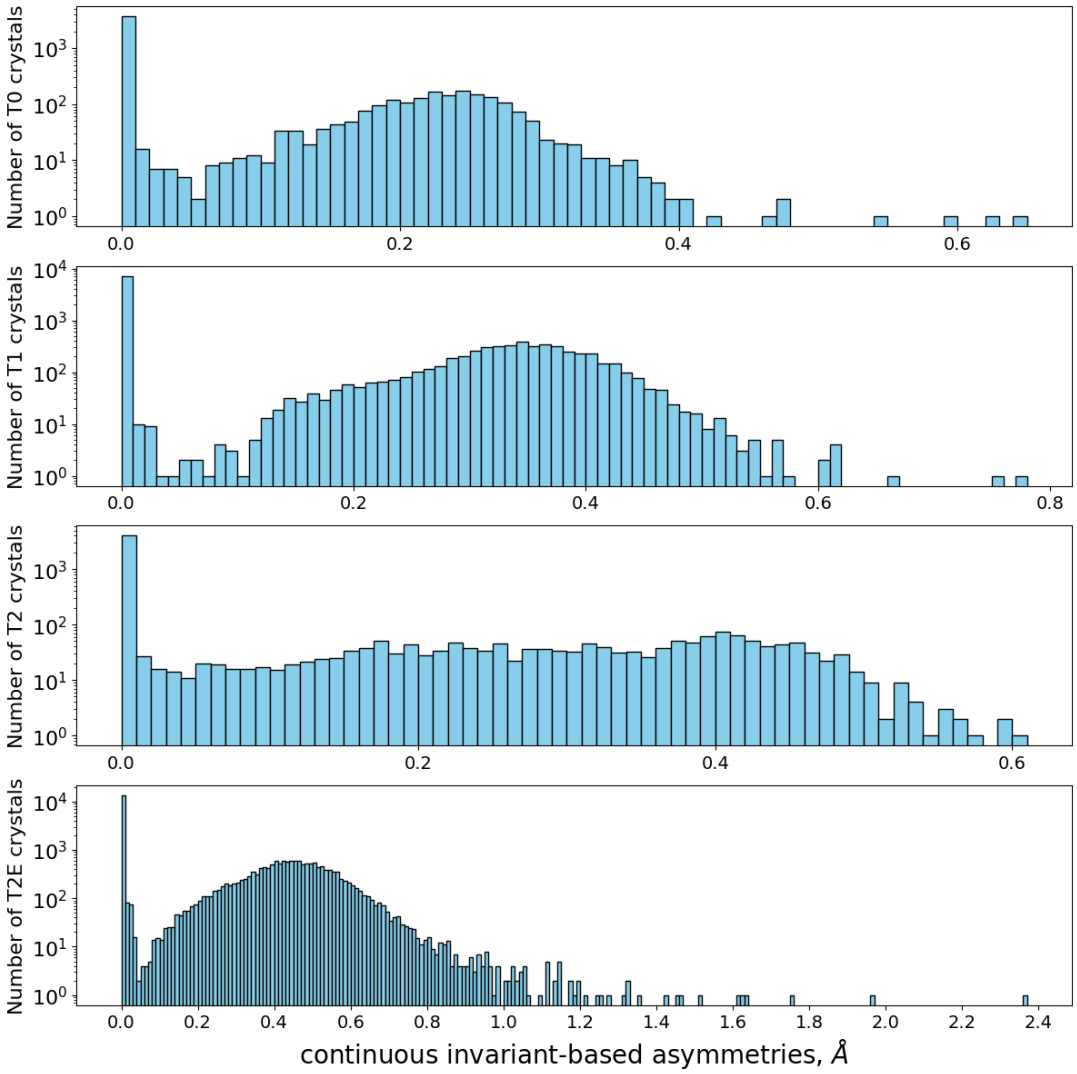

Fig. 3. The histograms of CIA for simulated crystals represented by 3 base points at 'ends' of molecules in Fig. 2. **Row 1**: T0. **Row 2**: T1. **Row 3**: T2. **Row 4**: T2E.

All structures in the four CSP datasets were generated with $Z' = 1$. The last stage of energy minimisation allowed this symmetry to be broken, which explains many cases of $Z' > 1$ in Table 1. If the generation stage included structures with $Z' \geq 2$, optimised crystals might have different distribution of CIAs than in Fig. 3.

In appendix A, Fig. 22 contains histograms of CIA_∞ based on the EMD with the ground metric d_∞ in Definition 4. The Chebyshev metric d_∞ captures the largest

deviations, while d = RMS averages over $k = 100$ adjusted inter-atomic distances, CIA_∞ has a larger range in comparison with CIA, see maximum values in Table 1.

Table 1. *Statistics of* CIA *values for the four CSP datasets from (Pulido* et al.*, 2017). The last rows contain Person correlations $r(x, y)$ between energy, density, and new* CIA*s.*

CSP datasets	T0 crystals	T1 crystals	T2 crystals	T2E crystals
# crystals: all	5645	12524	5679	29908
# crystals: CIA ≥ 0.001	2024	5363	1687	16491
percentage: CIA ≥ 0.001	35.8%	42.8%	29.7%	55.1%
maximum CIA, Å	0.642	0.779	0.605	2.364
$r(\text{energy}, \text{density})$	-0.909	-0.639	-0.377	-0.500
$r(\text{energy}, \text{CIA})$	-0.394	-0.202	$+0.022$	-0.026
$r(\text{density}, \text{CIA})$	$+0.317$	$+0.148$	$+0.040$	-0.021

CSP datasets are often visualised via energy-density plots, because density is a fast and continuous invariant. Moreover, density usually indicates stability, because stable crystals tend to be dense. Figures 4, 5, 6, 7 show these energy-density plots, where each crystal is represented by a point (density, energy), coloured according to its CIA. The colour bars on the right-hand side of the plots show the CIA range, with the bright red colour corresponding to high-symmetry structures with CIA = 0.

Table 1 highlights that large subsets (between 30% and 55%) of each CSP dataset have CIA > 0. Since all experimental crystals based on these molecules have CIA = 0, all non-symmetric crystals with CIA > 0 are likely non-ideal approximations to symmetric synthesised crystals. Indeed, if all non-red dots are removed from Figures 4, 5, 6, 7, the remaining red dots will still form roughly similar landscapes with all "minimal spikes" of density represented by only symmetric crystals with CIA = 0 in red.

The Pearson correlation $r(\text{energy}, \text{density})$ in Table 1 reflects the inverse dependence on density, because denser crystals tend to be more stable and have lower energies. This inverse correlation is the strongest with $r = -0.909$ for crystals based on the smaller molecule T0 and is still noticeable for crystals based on the larger molecules T1, T2, and T2E. The new asymmetries CIA and CIA_∞ are independent of density and energy due to their low correlations, especially for the T2 and T2E datasets.

All experimental crystals based on these molecules have CIA = 0, but their closest simulated analogues may not have the lowest energies as for the nanoporous T2-γ.

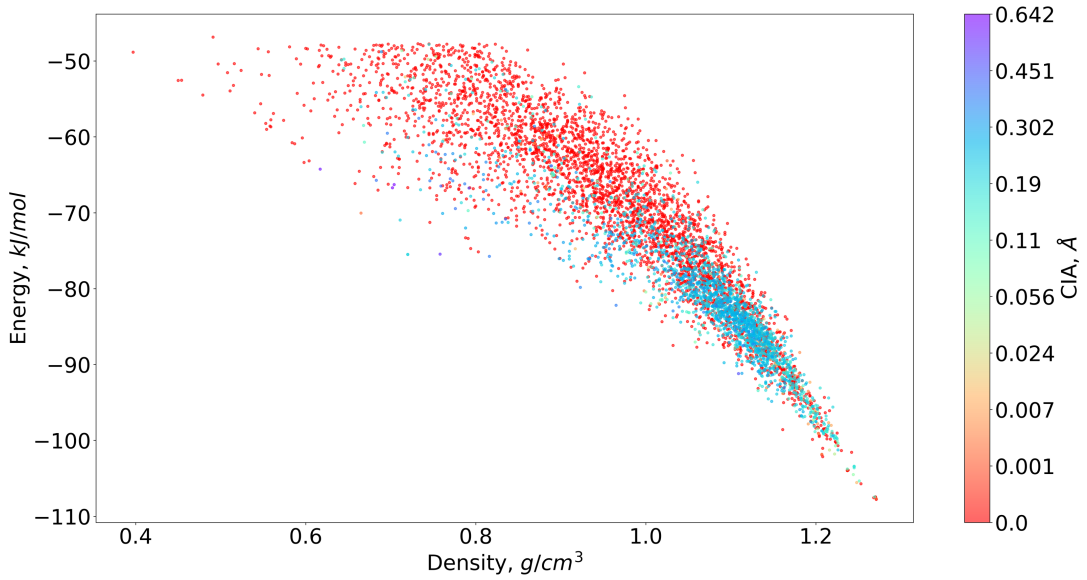

Fig. 4. Energy vs density for simulated T0 crystals, coloured by their CIA.

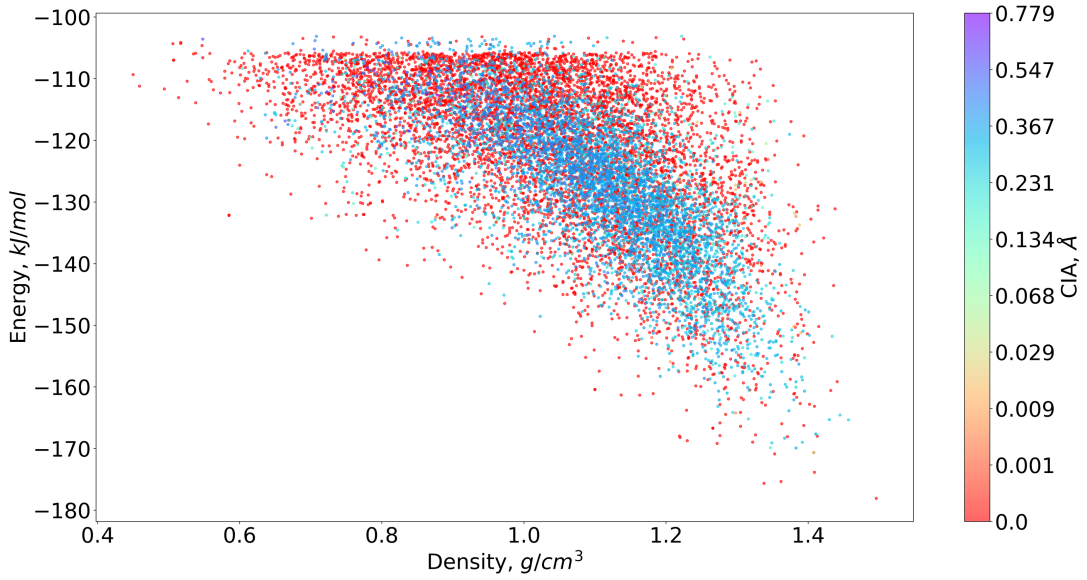

Fig. 5. Energy vs density for simulated T1 crystals, coloured by their CIA.

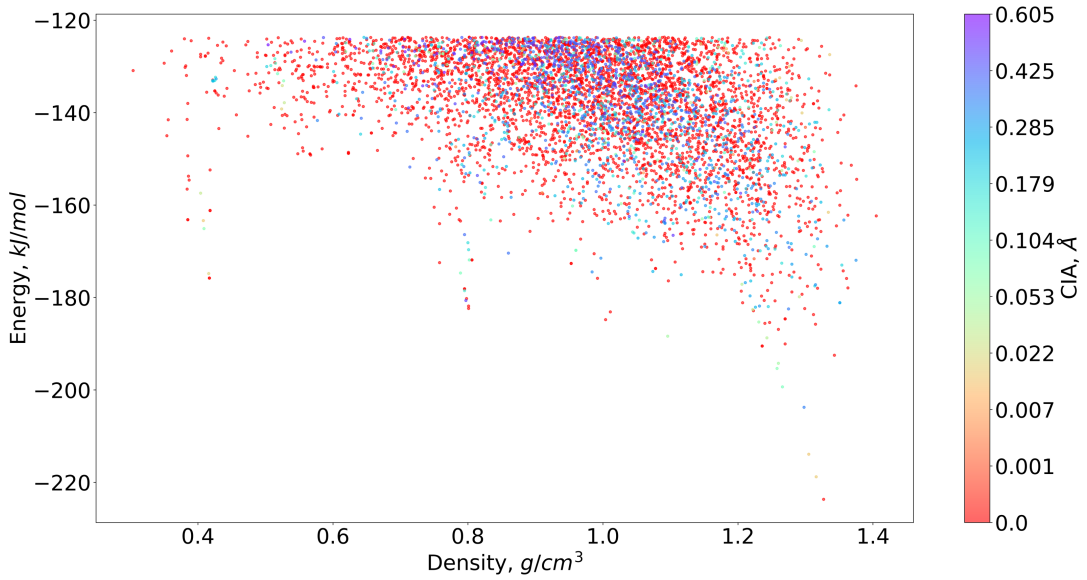

Fig. 6. Energy vs density for simulated T2 crystals, coloured by their CIA.

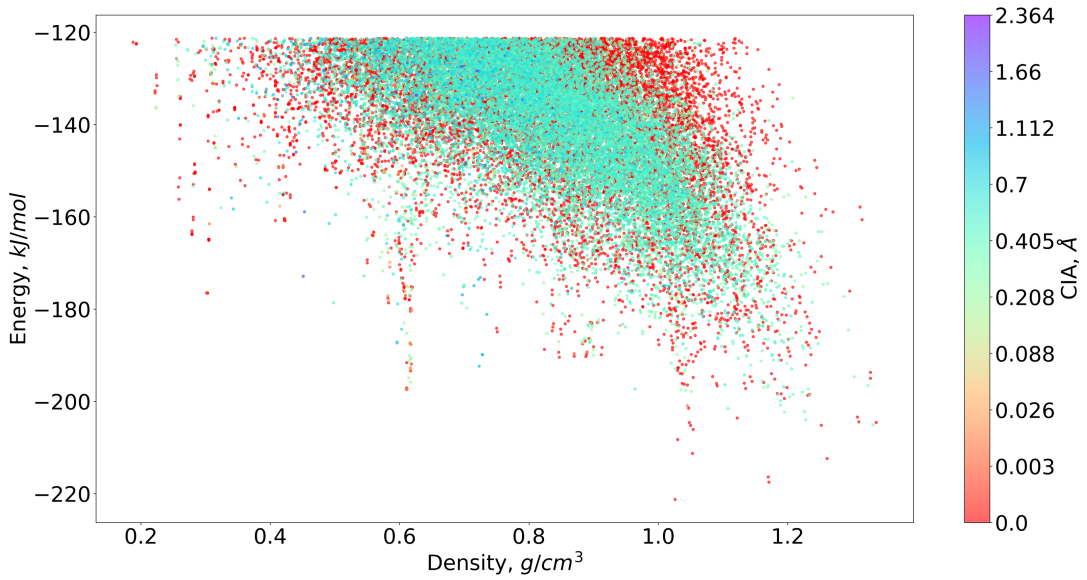

Fig. 7. Energy vs density for simulated T2E crystals, coloured by their CIA.

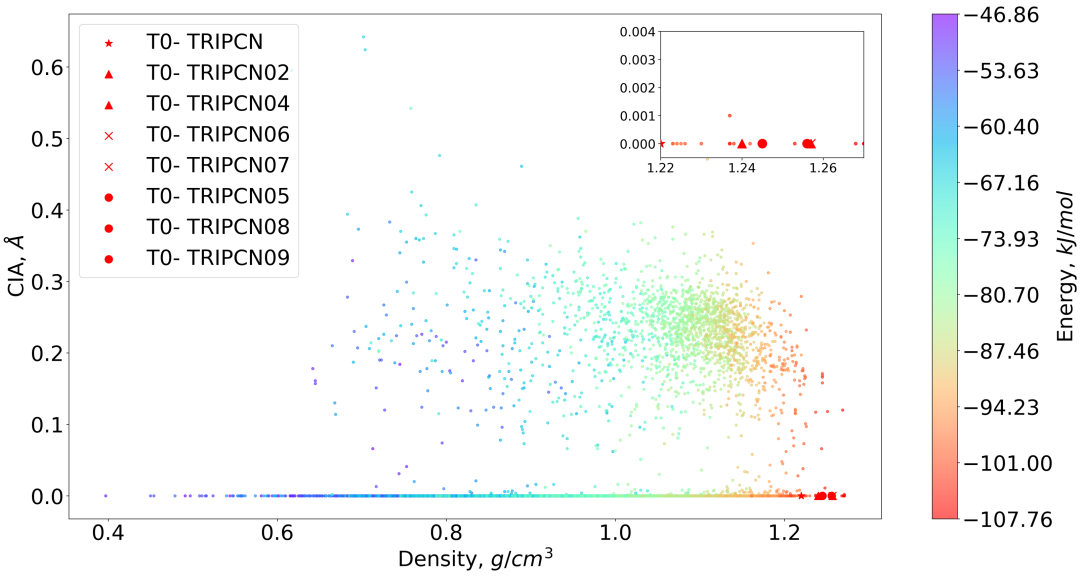

Fig. 8. CIA vs density for simulated and experimental T0 crystals in the CSD.

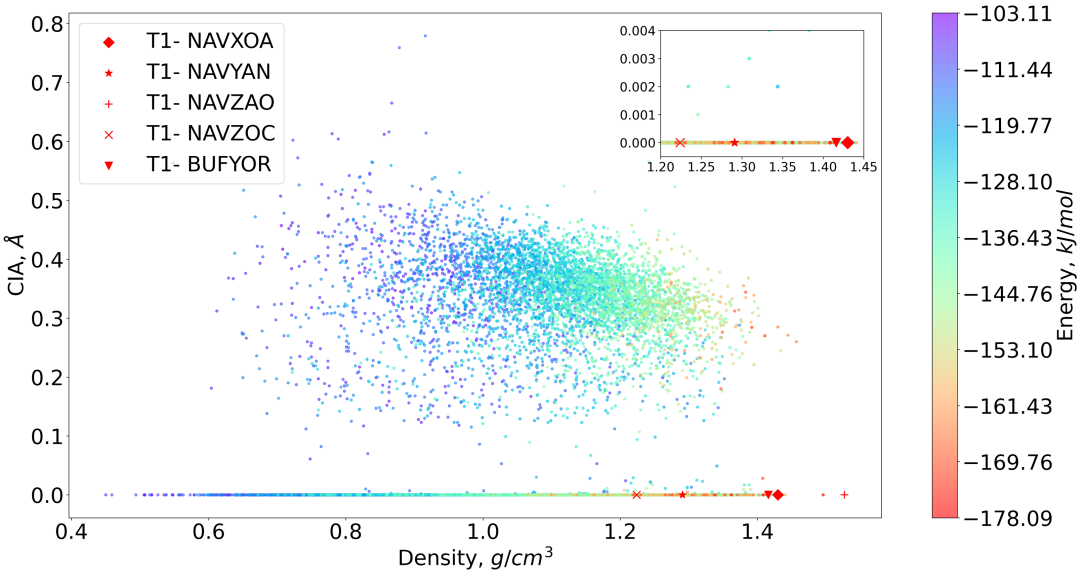

Fig. 9. CIA vs density for simulated and experimental T1 crystals in the CSD.

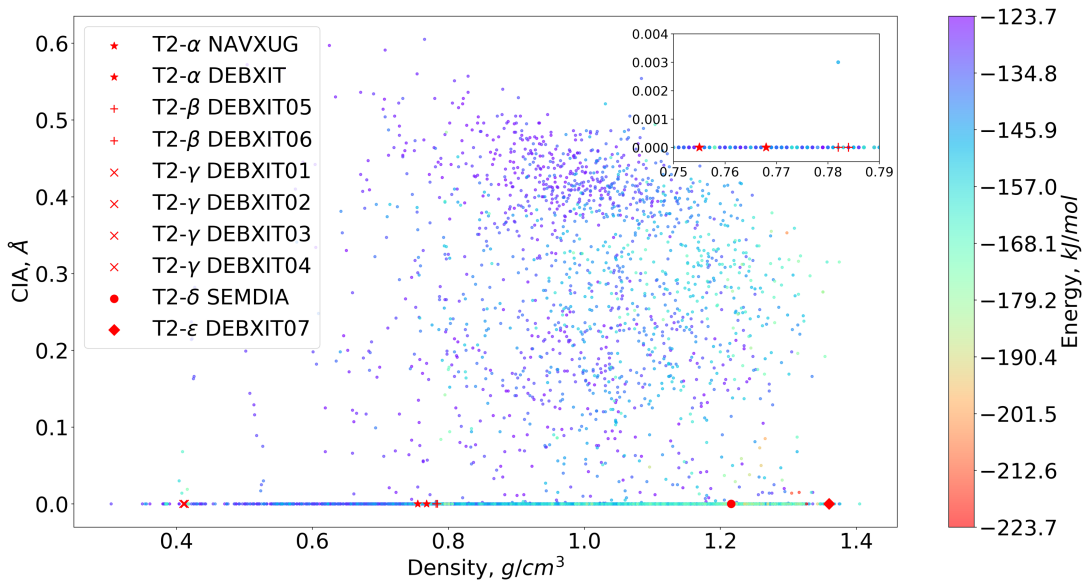

Fig. 10. CIA vs density for simulated and experimental T2 crystals in the CSD.

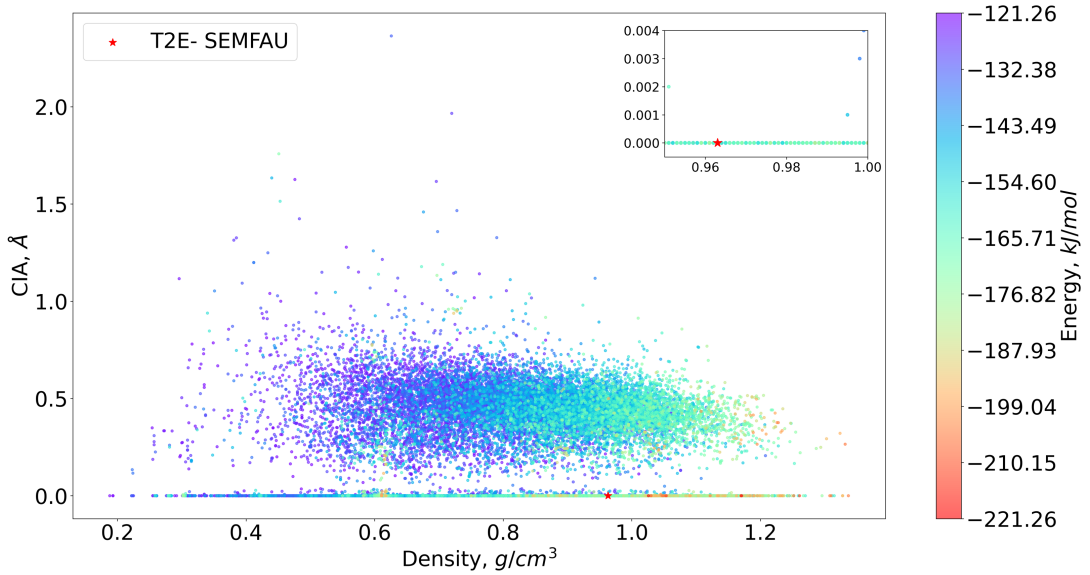

Fig. 11. CIA vs density for simulated and experimental T2E crystals in the CSD.

Figures 8, 9, 10, 11 show experimental crystals by red marks of various shapes in the coordinates (density, CIA), and indicate their apparent independence. In each figure,

the top right corner includes a zoomed-in image containing experimental crystals that are closest by density. Though many simulated crystals are symmetric with CIA = 0, all non-symmetric crystals form noisy clouds with variable energies. The visible gaps between these clouds and the horizontal axis CIA = 0 confirm a local version of the symmetry principle saying that a nearly symmetric structure likely converges to a higher symmetry version with CIA = 0.

Figure 12 shows the average running times vs the number Z of molecular components in a unit cell. This number Z goes up to 36 and coincides with G, because all finally optimised crystals are saved in the simplest translation group P1.

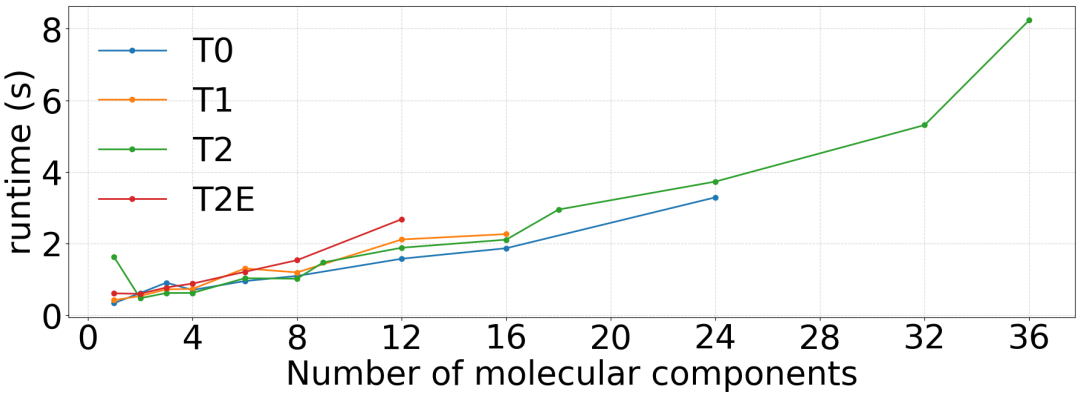

Fig. 12. Average running times (in seconds) of CIA on four CSP datasets vs the number G of molecules in asymmetric units, performed on a modest machine with CPU 13th Gen Intel(R) Core(TM) i7-1355U (1.70 GHz) and RAM 16GB.

5. Continuous asymmetries of all experimental crystals in the CSD

This section describes a large-scale analysis of asymmetries in the whole CSD. Each crystal is represented by a periodic set of all its atoms. We considered all periodic crystals with complete 3D geometry, no disorder, and based on a chemically unique molecule. Though Definition 5 can be applied to geometric blocks of different sizes, we postpone the more complicated case of co-crystals to future work.

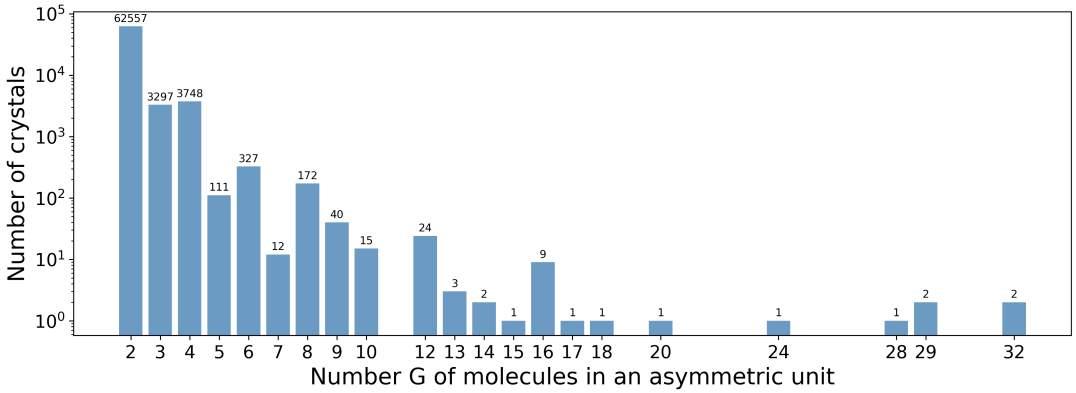

Fig. 13. The histogram of integer numbers G for all 69,196 periodic crystals in the CSD that have $G \geq 2$ chemically equivalent blocks in their asymmetric units.

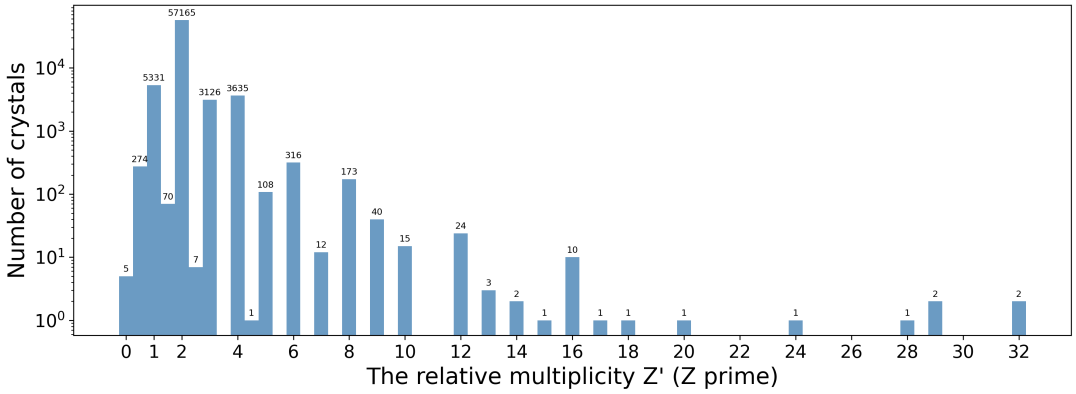

Fig. 14. The histogram of Z' with bin size 0.5 for all 69,196 periodic crystals in the CSD that have $G \geq 2$ chemically equivalent blocks in their asymmetric units.

The snapshot of the CSD on 12th November 2025 contained 1,394,755 entries, including 907223 crystals without disorder. Among them, 69,196 crystals have asymmetric units containing $G \geq 2$ molecules that all have the same composition, where G was computed by the CSD Python API as the number of components in the list crystal.asymmetric_unit_molecules. Some crystals with the highest Z' values from https://zprime.co.uk/database, such as OGUROZ ($Z' = 56$), TMESNH ($Z' = 32$),

IDOSID ($Z' = 24$), and VIFXEQ ($Z' = 24$), were excluded because of disorder.

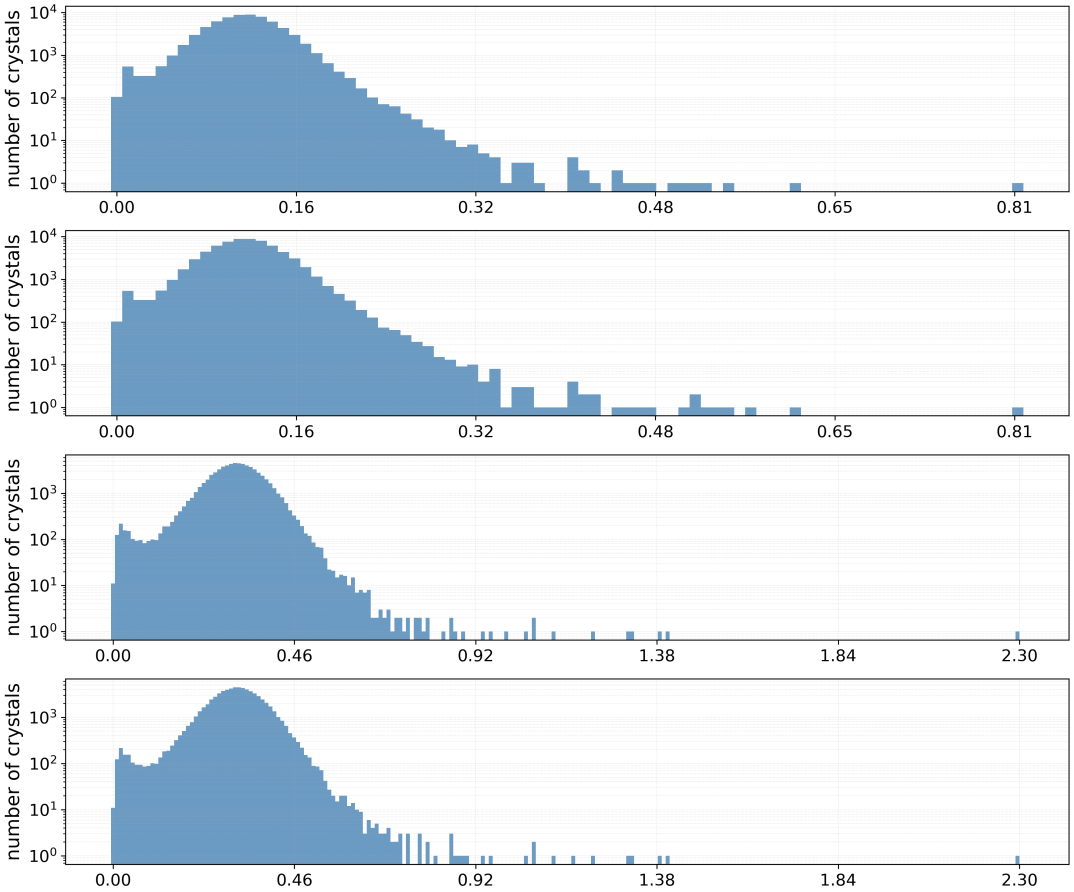

Fig. 15. The histograms of CIAs on the log scale with bin size 0.01Å for all 69,196 periodic crystals in the CSD that have $G \geq 2$ chemically equivalent molecules in asymmetric units. **Row 1**: CIA. **Row 2**: $\overline{\text{CIA}}$. **Row 3**: CIA_∞. **Row 4**: $\overline{\text{CIA}}_\infty$.

Figures 13, 14, and 15 show the histograms of G, Z', and four CIAs for this subset of the CSD, respectively, where Z' was computed as crystal.z_prime by the CSD Python API. The number Z[CIF] of molecules in the full motif was taken from lines "_cell_formula_units_Z" in CIFs from the CSD, which sometimes differs from Z[CSD], computed as the number of components in the list crystal.molecule. CSD Python API.

Table 2 shows all four versions of CIAs for the most extreme crystals in the CSD: five crystals with the lowest $Z' \leq 0.33$ and five crystals with the highest $Z' \geq 28$.

Table 2. CIA*s of the crystals with the lowest and largest relative multiplicities in the CSD. The numbers Z and G count molecules in a unit cell and an asymmetric unit, respectively.*

CSD id	Z[CIF]	G blocks	Z'[CSD]	CIA, Å	$\overline{\text{CIA}}$, Å	CIA_∞, Å	$\overline{\text{CIA}}_\infty$, Å
VESWEZ	2	2	0.083	0.204	0.204	0.580	0.580
ELIQIZ02	3	2	0.083	0.226	0.226	0.807	0.807
ZOKYEH01	16	2	0.167	0.086	0.086	0.241	0.241
RARTEK	16	2	0.17	0.125	0.125	0.332	0.332
ZAVJOV	2	2	0.33	0.090	0.090	0.241	0.241
QILJII01	112	28	28	0.168	0.185	0.397	0.426
LOFRAD	116	29	29	0.149	0.183	0.420	0.499
LOFRAD01	116	29	29	0.149	0.185	0.434	0.506
JIPTIL09	32	32	32	0.104	0.109	0.266	0.282
JIPTIL10	32	32	32	0.102	0.109	0.265	0.282

In Table 2, crystal VESWEZ has $G = 2$ components CN_2 in geometrically non-equivalent positions: in one CN_2, both nitrogens are linked to two carbons; in another CN_2, the two nitrogens are linked to 2 and 3 carbons, see Fig. 16. Crystal ELIQIZ02 has molecules C_6H_6 and C_2H_2, and its asymmetric unit consists of $G = 2$ geometrically different carbons: one from C_6H_6 and another from C_2H_2. Crystal ZOKYEH01 consists of a big molecule of C_{60} with extra tails, but its asymmetric unit was also split into $G = 2$ blocks $C_{10}O_2$, which apparently have non-isometric positions within the full crystal. Crystal RARTEK and ZAVJOV similarly consist of big molecules based on $G = 2$ blocks in asymmetric units, whose positions can not be matched by isometry preserving the whole crystal. The last three cases show that molecular crystals will benefit from quantifying asymmetry at the level of full molecules, because asymmetric units may not split into uniquely defined molecules or geometric blocks of atoms.

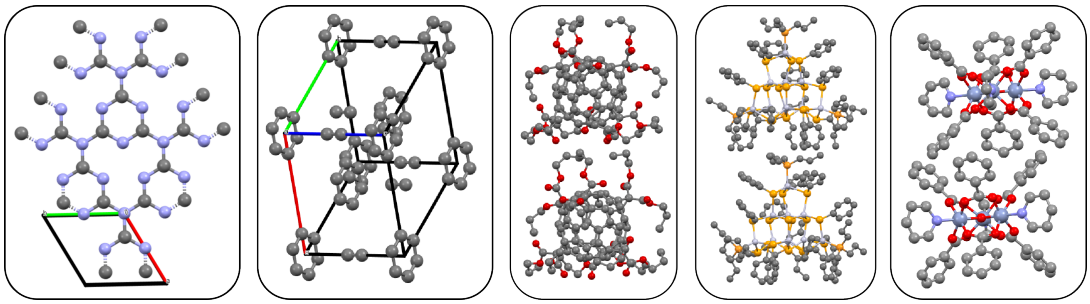

Fig. 16. The crystals with the lowest Z' from Table 2 shown without hydrogen atoms. **1st**: VESWEZ. **2nd**: ELIQIZ02. **3rd**: ZOKYEH01. **4th**: RARTEK. **5th**: ZAVJOV.

Table 3. *CIAs of the well-known polymorphs of artemisinin (QNGHSU01), pyridine (PYRDNA04), para-chlorophenol (α-form CLPHOL12 and β-form CLPHOL13), and the famous ROY molecule (R05 polymorph QAXMEH31 and R18 polymorph QAXMEH57).*

CSD id	Z[CIF]	G blocks	Z'[CSD]	CIA, Å	$\overline{\text{CIA}}$, Å	CIA_∞, Å	$\overline{\text{CIA}}_\infty$, Å
QNGHSU01	4	4	4	0.357	0.379	1.093	1.096
QAXMEH31	2	2	2	0.440	0.440	1.098	1.098
QAXMEH57	2	2	2	0.807	0.807	1.602	1.602
CLPHOL12	2	2	2	0.790	0.790	2.594	2.594
CLPHOL13	2	2	2	0.575	0.575	1.132	1.132
PYRDNA04	4	4	4	1.971	2.096	2.756	2.861

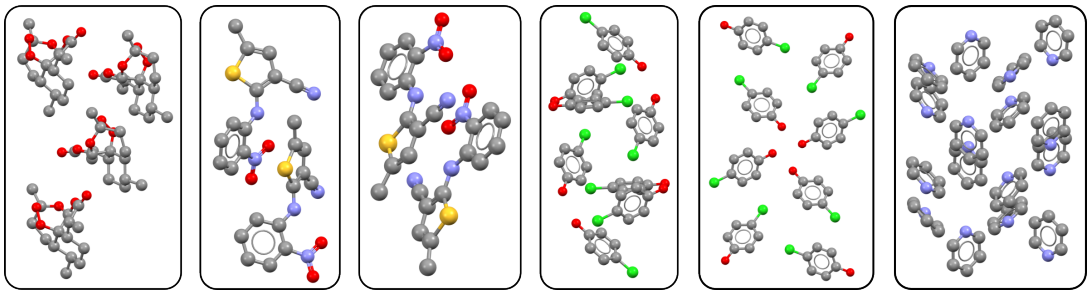

Fig. 17. Six famous polymorphs whose CIAs are listed in Table 3. From left to right: QNGHSU01, QAXMEH31, QAXMEH57, CLPHOL12, CLPHOL13, PYRDNA04.

Table 4 lists the 10 crystals from with the lowest values of CIAs. The first three crystals have CIA = 0 with 3 decimal places, so their $Z' \geq 2$ might be corrected.

Table 4. *Ten crystals with the lowest* CIA *among 69,196 periodic crystals in the CSD that have* $Z' \geq G \geq 2$ *chemically equivalent blocks in their asymmetric units.*

CSD id	Z[CIF]	G blocks	Z'[CSD]	CIA, Å	$\overline{\text{CIA}}$, Å	CIA_∞, Å	$\overline{\text{CIA}}_\infty$, Å
IYIWIY	8	8	8	0.000	0.000	0.000	0.000
GLYCIN81	2	2	2	0.000	0.000	0.000	0.000
YOSNEZ05	2	2	2	0.000	0.000	0.000	0.000
GIBVOG	2	2	2	0.000	0.000	0.001	0.001
GLYCIN82	3	3	3	0.001	0.001	0.002	0.002
KAVXUE	1	2	2	0.002	0.002	0.005	0.005
ADUWED	64	2	2	0.002	0.002	0.006	0.006
CINMAC13	2	2	2	0.002	0.002	0.010	0.010
XOTRAB	4	2	2	0.003	0.003	0.007	0.007
COTZES	6	2	2	0.003	0.003	0.009	0.009

The value CIA $= 0$ means that all molecules are geometrically equivalent, i.e. can be exactly matched by isometry that preserves the whole crystal. In this case, an asymmetric unit should contain only one molecule ($G = 1$), so $Z' \leq 1$ is expected.

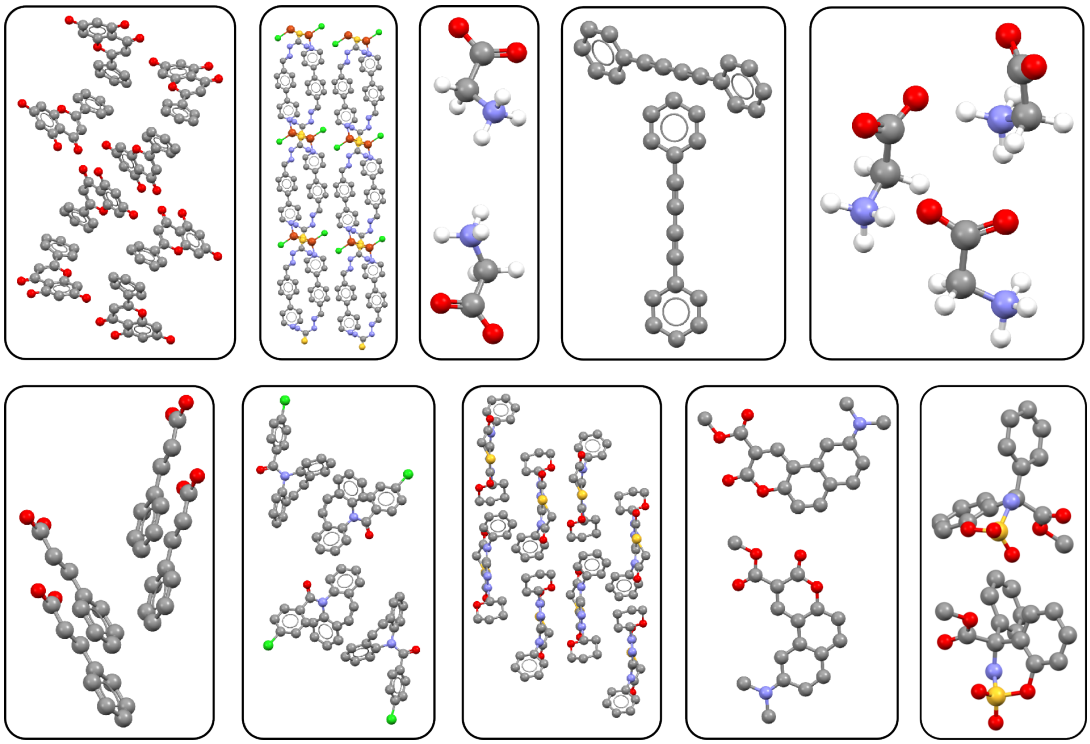

Fig. 18. Ten crystals (some shown without hydrogen atoms) from Table 4 with very low CIA ≥ 0. **Top** from left to right: IYIWIY, GIBVOG, GLYCIN81, YOSNEZ05, GLYCIN82. **Bottom**: CINMAC13, KAVXUE, ADUWED, XOTRAB, COTZES.

One explanation is a potentially wrong space group (Henling & Marsh, 2014). For example, IYIWIY has the space group P1, but looks more symmetric in the first picture of Fig. 18. Both structures IYIWIY and YOSNEZ05 were obtained from powder data, so their space groups might need re-checking. Since all CIAs continuously change under atomic perturbations by Theorem 8, there is no need to search for a higher symmetry group, which drops to the simplest group P1 under almost any noise anyway. The values of Z[CIF] can be corrected for all entries with $Z < G$ in Tables 4 and 5, because a unit cell should not have fewer molecules than in an asymmetric unit.

Table 5. *Almost symmetric crystals with high values* $Z' \geq 5$ *but low* CIA ≤ 0.021Å.

CSD id	Z[CIF]	G blocks	Z'[CSD]	CIA, Å	$\overline{\text{CIA}}$, Å	CIA_∞, Å	$\overline{\text{CIA}}_\infty$, Å
TEGBEP	1	6	6	0.010	0.011	0.030	0.032
HOGKAR	12	6	6	0.010	0.011	0.032	0.034
GINHIX	6	6	6	0.011	0.012	0.034	0.039
EVIWUE	12	6	6	0.012	0.014	0.051	0.059
LEMWOR	2	6	6	0.013	0.013	0.040	0.043
YIVHER	10	5	5	0.015	0.016	0.053	0.058
IFOFAN	10	5	5	0.020	0.020	0.070	0.077
EDUCAL	12	6	6	0.020	0.022	0.062	0.071
ROTSAY	18	9	9	0.021	0.023	0.060	0.067
CIDHAB	1	12	12	0.021	0.023	0.067	0.071

In conclusion, the relative multiplicity Z' discontinuously changes under almost any perturbation, the proposed CIA in Definition 5 is continuous by Theorem 8. For the CSP datasets in section 4, about a half of all 50K+ simulated crystals have CIA > 0, while all experimental crystals have CIA $= 0$, see Table 1. Moreover, these continuous and fast asymmetries are not correlated with density and energy. The large-scale experiments on the CSD show that many non-symmetric crystals with high Z' have low CIAs in Table 5 and hence are geometrically close to more symmetric forms. This work was supported by the Royal Society APEX fellowship "New geometric methods for mapping the space of periodic crystals" (APX/R1/231152) of the last author.

References

(2024). Continuum fallacy within the sorites paradox. https://en.wikipedia.org/wiki/Sorites_paradox#Continuum_fallacy.

Anderson, K. M., Afarinkia, K., Yu, H.-w., Goeta, A. E. & Steed, J. W. (2006). *Crystal growth & design*, **6**(9), 2109–2113.

Anosova, O. & Kurlin, V. (2025). *Geometric Data Science*. arXiv:2512.05040

Anosova, O., Kurlin, V. & Senechal, M. (2024). *IUCrJ*, **11**, 453–463.

Anosova, O., Widdowson, D. & Kurlin, V. (2025). *Pattern Recognition*, **171**(112108).

Bright, M. J., Cooper, A. I. & Kurlin, V. A. (2023*a*). *Chirality*, **35**, 920–936.

Bright, M. J., Cooper, A. I. & Kurlin, V. A. (2023*b*). *Acta Crystallographica Section A*, **79**(1), 1–13.

Brock, C. P. (2016). *Acta Cryst B*, **72**(6), 807–821.

Chapuis, G. (2024). Z and Z′ in the IUCr Online Dictionary of Crystallography. `https://dictionary.iucr.org/Z_and_Z'`.

Edelsbrunner, H., Heiss, T., Kurlin, V., Smith, P. & Wintraecken, M. (2021). In *Proceedings of Symposium on Computational Geometry*, vol. 189, pp. 32:1–32:16.

Hargreaves, C. J., Dyer, M. S., Gaultois, M. W., Kurlin, V. A. & Rosseinsky, M. J. (2020). *Chemistry of Materials*, **32**, 10610–10620.

Henling, L. M. & Marsh, R. E. (2014). *Acta Crystallographica Section C*, **70**(9), 834–836.

Kurlin, V. (2024). *Foundations of Computational Mathematics*, **24**, 805–863.

Lawton, S. L. & Jacobson, R. A. (1965). *The reduced cell and its crystallographic applications*. Tech. rep. Ames Lab., Iowa State Univ. of Science and Tech., US.

Lax, M. (2001). *Symmetry principles in solid state and molecular physics*. Courier Corporation.

Pulido, A. *et al.* (2017). *Nature*, **543**(7647), 657–664.

Senechal, M. (1996). *Quasicrystals and geometry*. CUP Archive.

Steed, K. M. & Steed, J. W. (2015). *Chemical Reviews*, **115**(8), 2895–2933.

Van Eijck, B. P. & Kroon, J. (2000). *Acta Cryst B*, **56**(3), 535–542.

Widdowson, D. & Kurlin, V. (2022). *Advances in Neural Information Processing Systems (NeurIPS)*, **35**, 24625–24638.

Widdowson, D. & Kurlin, V. (2024). *Crystal Growth and Design*, **24**, 5627–5636.

Widdowson, D. & Kurlin, V. (2025*a*). *Scientific Reports*, **15**, 27588.

Widdowson, D. & Kurlin, V. (2025*b*). *arXiv:2509.15088*.

Widdowson, D. & Kurlin, V. (2025*c*). *SIAM Journal on Applied Mathematics*.

Wilson, A. (1993). *Acta Cryst A*, **49**(6), 795–806.

Appendix A
Extra experimental results for simulated crystals

This appendix contains extra plots for other versions of CIAs.

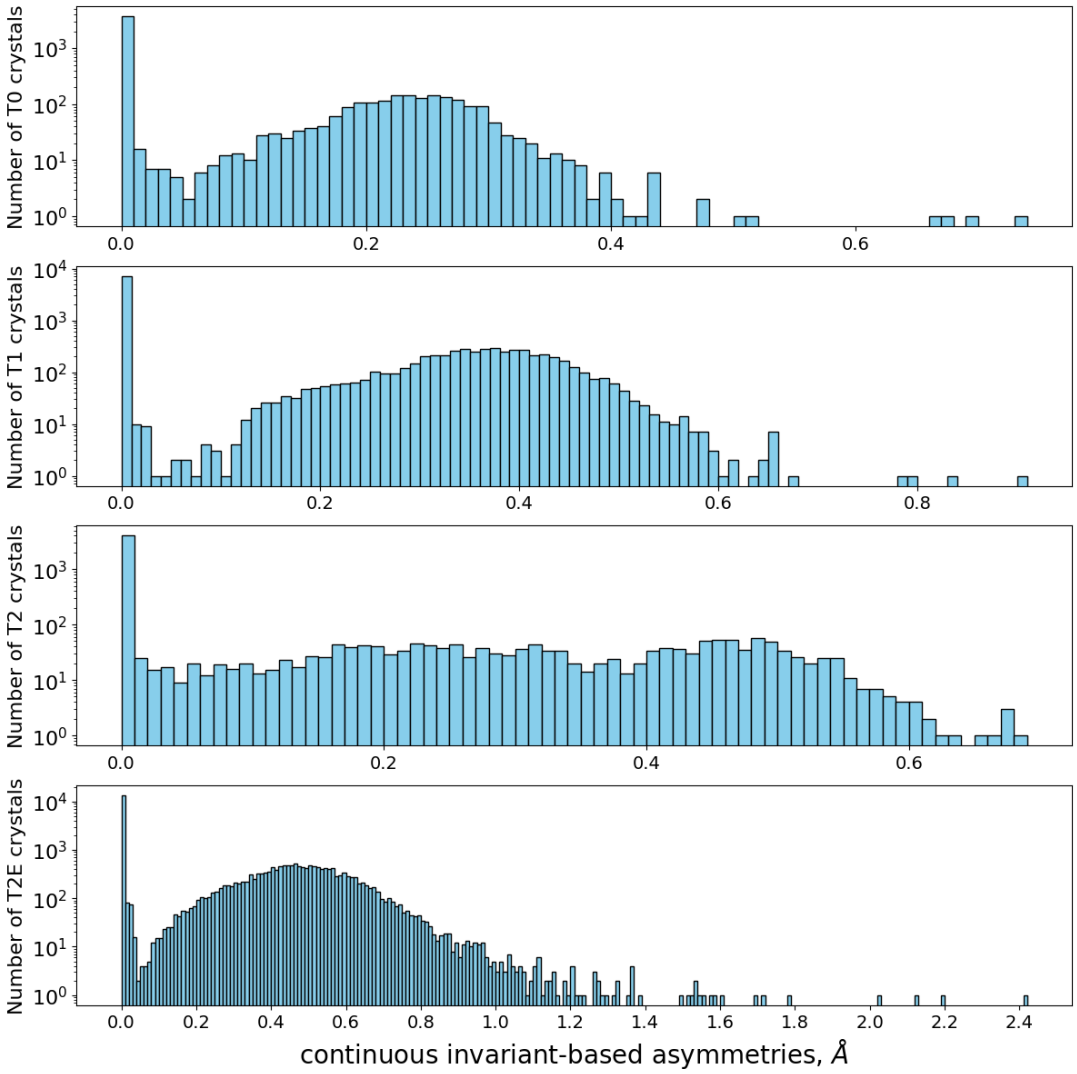

Fig. 19. The histograms of $\overline{\mathrm{CIA}}$ for simulated crystals represented by 3 base points at 'ends' of molecules in Fig. 2. **Row 1**: T0. **Row 2**: T1. **Row 3**: T2. **Row 4**: T2E.

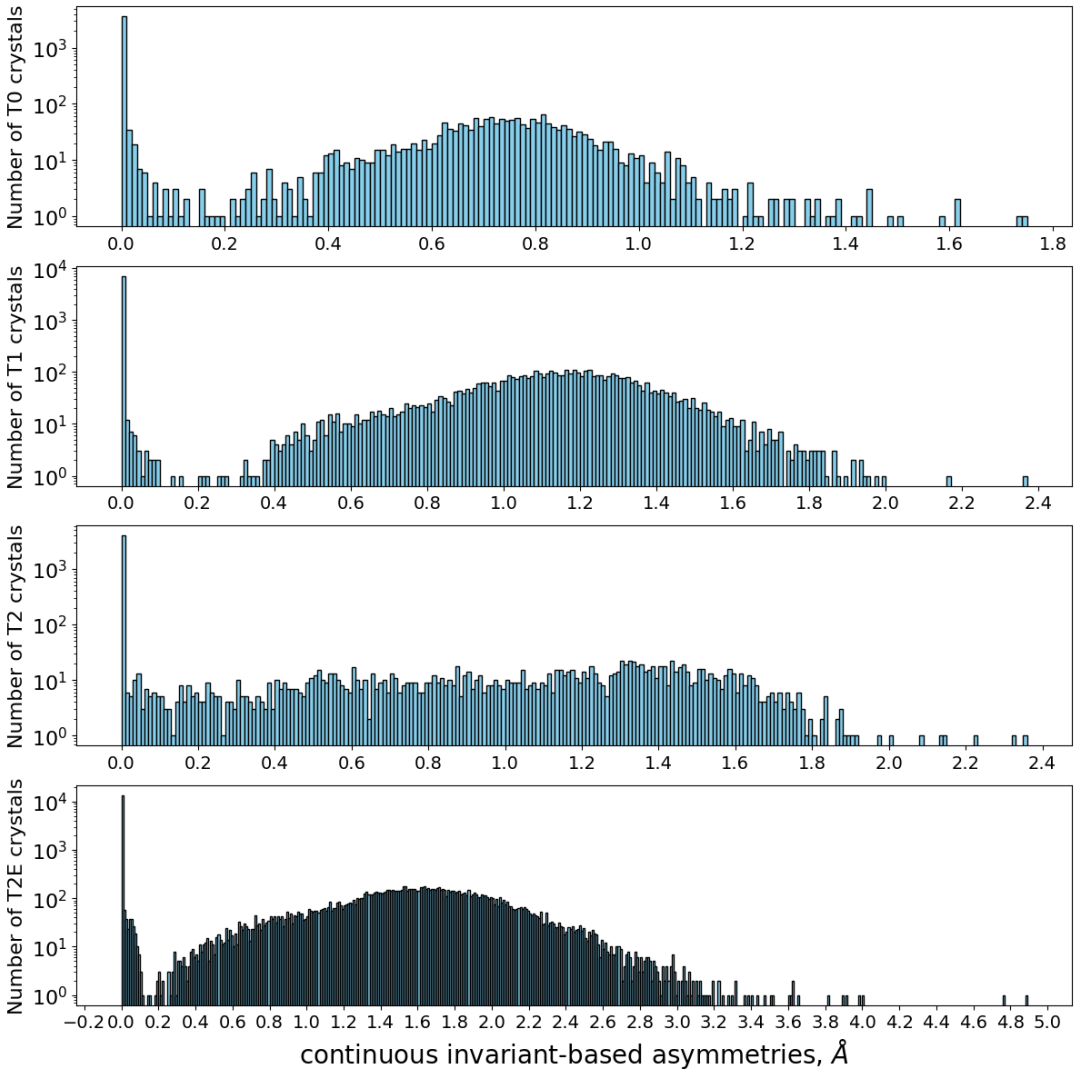

Fig. 20. The histograms of CIA_{∞} for simulated crystals represented by 3 base points at 'ends' of molecules in Fig. 2. **Row 1**: T0. **Row 2**: T1. **Row 3**: T2. **Row 4**: T2E.

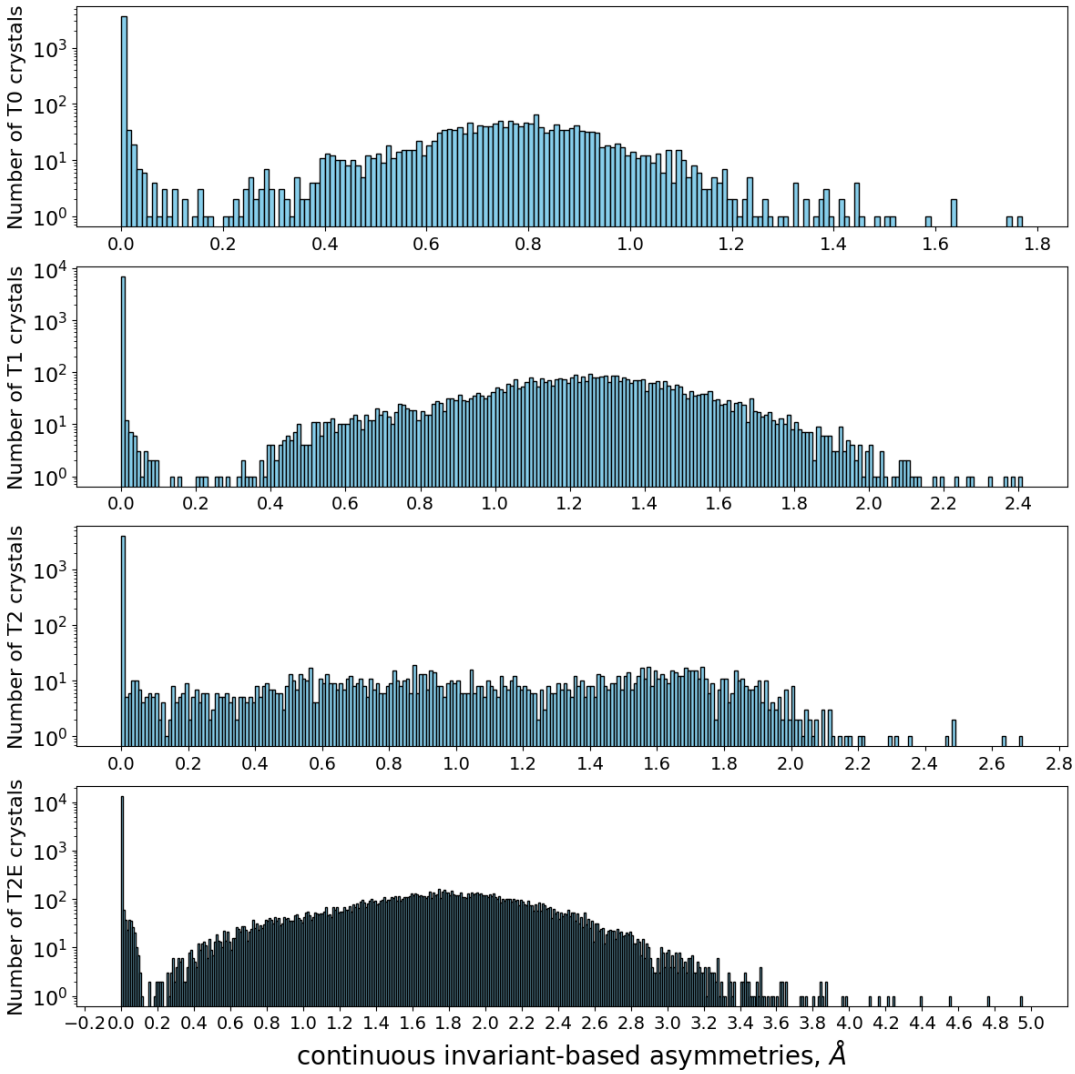

Fig. 21. The histograms of $\overline{\mathrm{CIA}}_\infty$ for simulated crystals represented by 3 base points at 'ends' of molecules in Fig. 2. **Row 1**: T0. **Row 2**: T1. **Row 3**: T2. **Row 4**: T2E.

Table 1. *Statistics of* $\overline{\mathrm{CIA}}, \mathrm{CIA}_\infty, \overline{\mathrm{CIA}}_\infty$ *for the four CSP datasets from (Pulido* et al.*, 2017). The last rows contain Person correlations* $r(x, y)$ *between energy, density, and new* CIA*s.*

CSP datasets	T0 crystals	T1 crystals	T2 crystals	T2E crystals
maximum CIA_∞, Å	1.748	0.902	2.352	4.882
$r(\text{energy}, \overline{\mathrm{CIA}})$	-0.393	-0.196	+0.035	-0.020
$r(\text{energy}, \mathrm{CIA}_\infty)$	−0.398	−0.196	+0.016	−0.019
$r(\text{energy}, \overline{\mathrm{CIA}}_\infty)$	-0.399	-0.186	+0.032	-0.014
$r(\text{density}, \overline{\mathrm{CIA}})$	+0.315	+0.144	+0.036	-0.021
$r(\text{density}, \mathrm{CIA}_\infty)$	+0.322	+0.133	+0.037	−0.033
$r(\text{density}, \overline{\mathrm{CIA}}_\infty)$	+0.323	+0.131	+0.032	-0.022

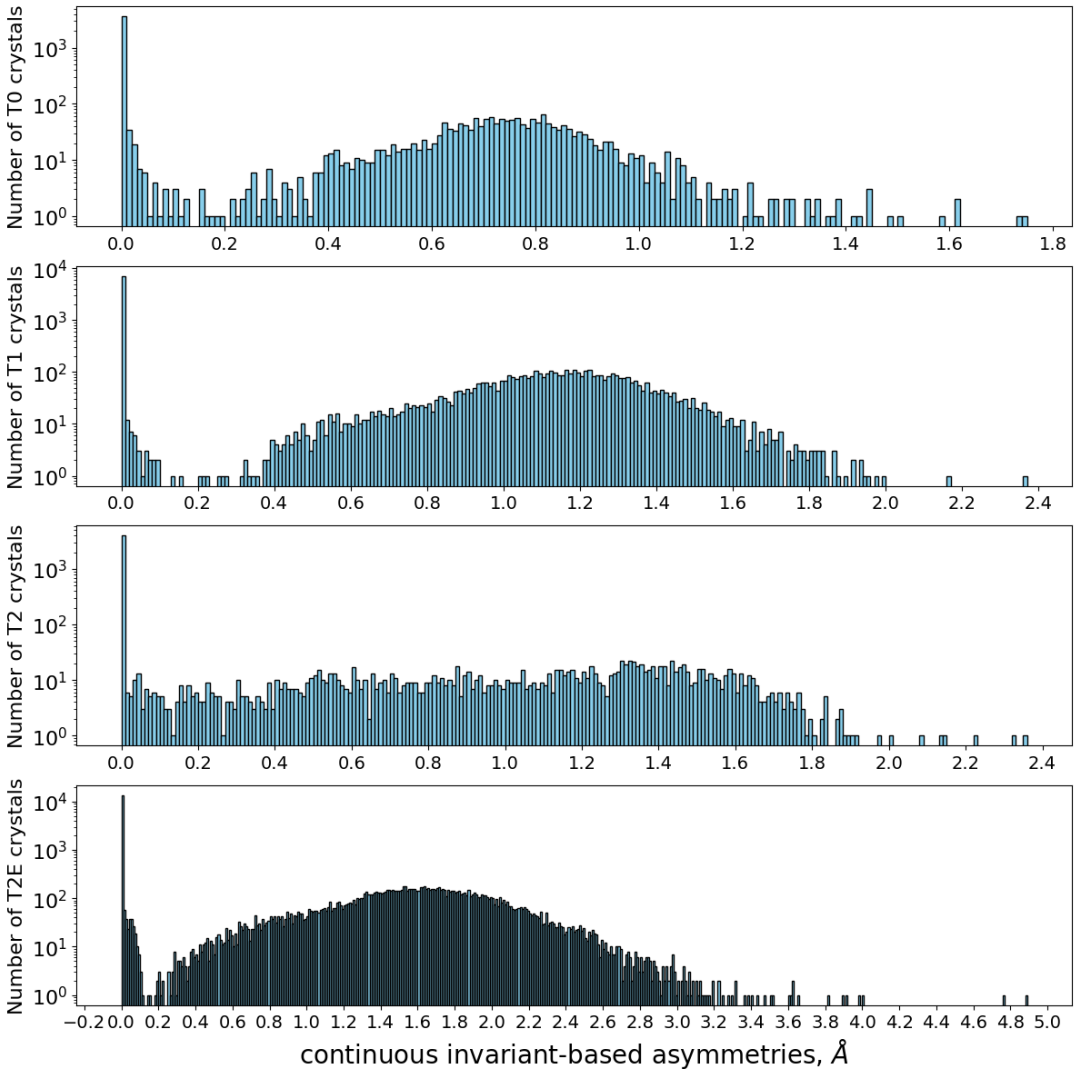

Fig. 22. The histograms of CIA_{∞} for simulated crystals represented by 3 base points at 'ends' of molecules in Fig. 2. **Row 1**: T0. **Row 2**: T1. **Row 3**: T2. **Row 4**: T2E.

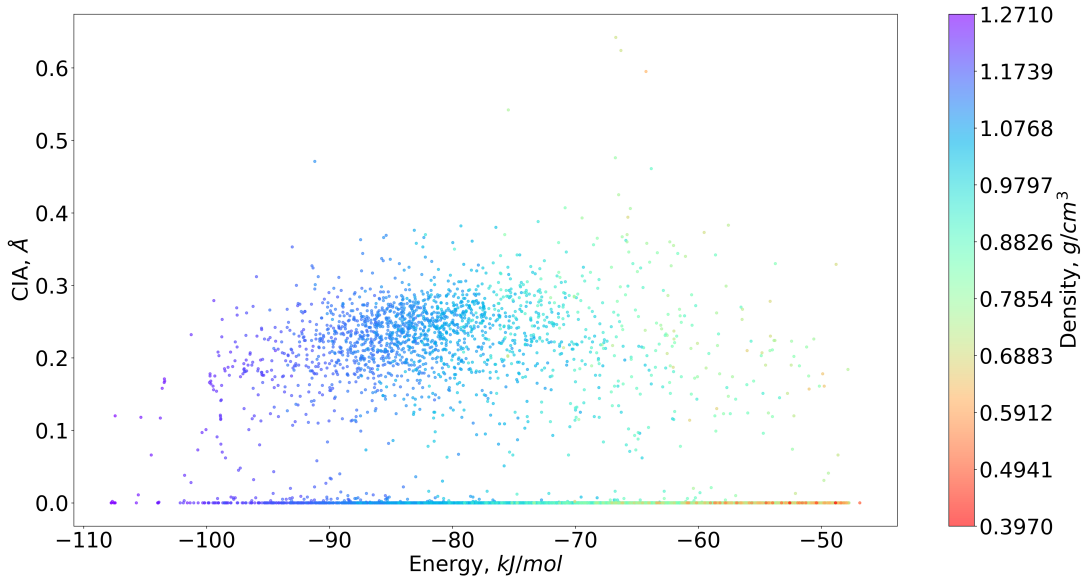

Fig. 23. CIA vs energy plot for simulated T0 crystals, coloured by their density

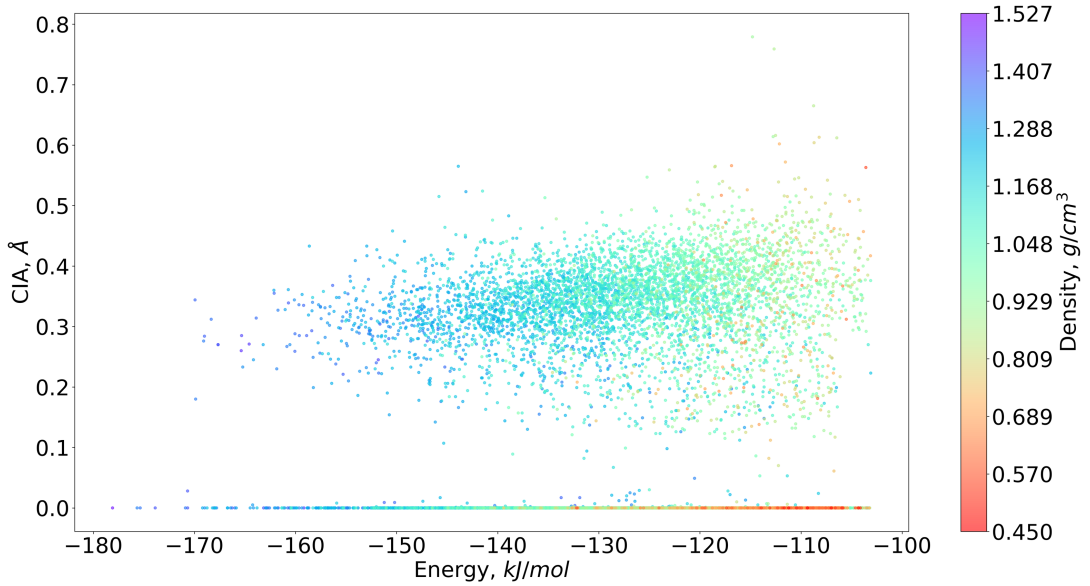

Fig. 24. CIA vs energy plot for simulated T1 crystals, coloured by their density

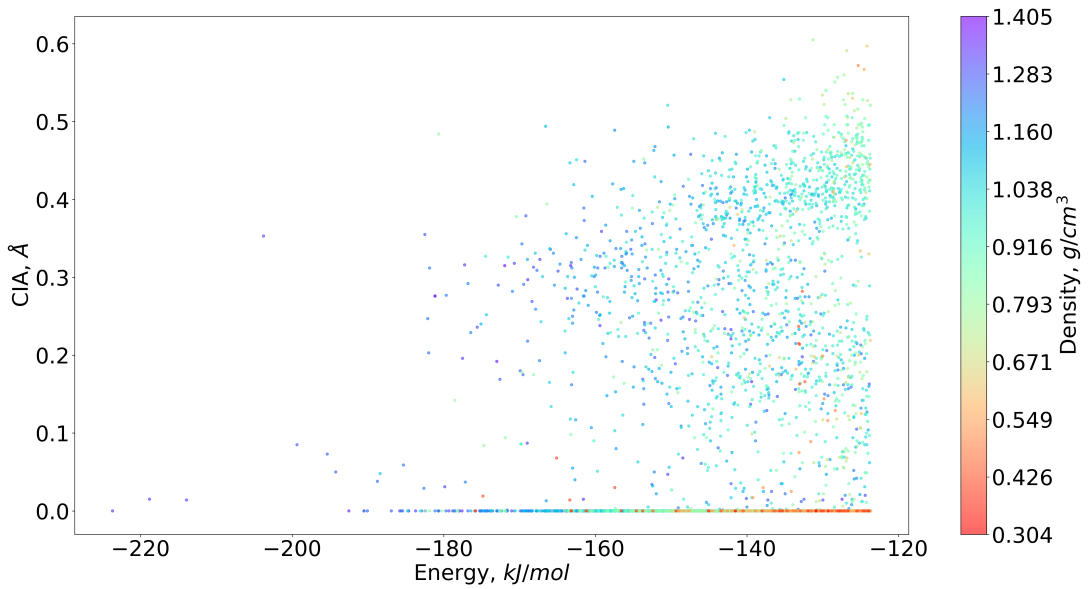

Fig. 25. CIA vs energy plot for simulated T2 crystals, coloured by their density

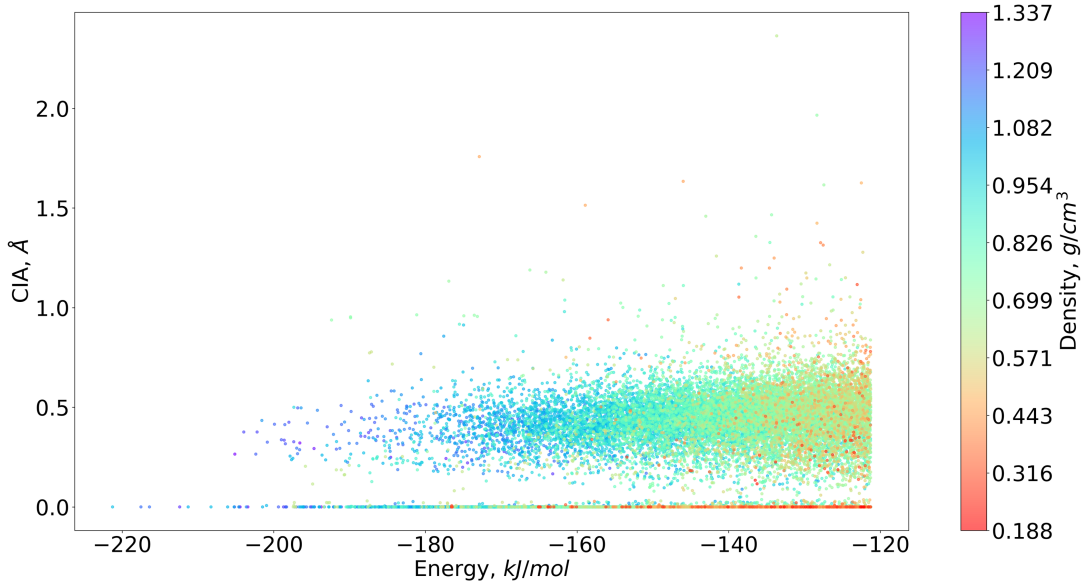

Fig. 26. CIA vs energy plot for simulated T2E crystals, coloured by their density

Figures 23, 24, 25, and 26 plot the CIA (Å) and CIA_{∞} (Å) vs lattice energy (kJ/mol) for T0, T1, T2, and T2E simulated crystals, respectively. Similar to Fig-

ures 8-11, the crystals are coloured with different density (g/cm^3) values. The colour bar for density is shown at the right side of each plot. Similar to the figures 8- 11, the colour gradient in these figures also indicate most of the crystals having lower energy, i.e., are stable, are often associated with higher density values compared to the crystals having higher energy which are mostly associated with lower density. While symmetric crystals exist across the full range of energies and densities, crystals with higher energies have larger asymmetries. High-density structures are observed with both zero and non-zero asymmetry, indicating that density alone does not determine symmetry. and higher asymmetry. -178.09 kJ/mol for T1, -123.7 to -223.7 kJ/mol for T2, and -121.26 to -221.26 kJ/mol for T2E predicted structures.

experimental structure SEMFAU. The CIA (Å) and CIA$_\infty$ (Å) is zero for the structure, indicating symmetry. Since the structure is nano-porous and has a low-density, the PPC of the structure is comparatively higher high average radius of balls 'packed' inside the unit cell. ADA$_1$ (Å) and ADA$_2$ (Å) representing how much AMD$_1$ and AMD$_2$ from the closest and second closest neighbour, respectively, deviate from the PPC, scaled for number of neighbours.

Appendix B
Detailed proofs of mathematical results

Proof of Lemma 6. We prove the invariance of CIA(S). The proof for other versions is almost identical. Since PDA($S;k$) consists of inter-point distances, which are invariant under any isometry, CIA(S) is also invariant. If a unit cell U of S is transformed to another cell U' by a matrix from SL($\mathbb{Z},n$), there is a 1-1 correspondence between all points in the original motif $S\cap U$ and the new motif $S\cap U'$ that respects all distances to neighbours. Then PDA($S;k$) and hence CIA(S) remain invariant. If a unit cell U

is scaled up by an integer factor c, the original motif $M = S \cap U$ transforms into the c-times larger motif M_c containing c isometric copies of M. The scaled-up asymmetric unit contains c times more blocks $B_1, \cdots, B_{cG}$, which can be considered as c copies of the original blocks. For each fixed $i = 1, \ldots, m$, the matrix of pairwise distances EMDs between cG blocks consists of $c \times c$ copies of the original matrix $G \times G$ of distances. The distances to the farthest units $d_{ij} = \max\limits_{j=1,\ldots,cG} \mathrm{EMD}(B_i, B_j)$ are obtained by concatenating c copies of the original vector $(d_{i1}, \ldots, d_{iG})$. Then the maximum and average values for each vector remain the same, so $\mathrm{CIA}(S)$ is invariant. □

Proof of Lemma 7. The inequalities $\mathrm{CIA}(S) \leq \mathrm{CIA}_\infty(S)$ and $\overline{\mathrm{CIA}}(S) \leq \overline{\mathrm{CIA}}_\infty(S)$ hold, because the RMS distance d is bounded from above by the Chebyshev distance d_∞. The inequality $\mathrm{CIA}(S) \leq \overline{\mathrm{CIA}}$ holds, because $\mathrm{CIA}(S) = \min\limits_{i=1,\ldots,G} d_i \leq \frac{1}{G}\sum\limits_{i=1}^{G} d_i = \overline{\mathrm{CIA}}(S)$. To prove the inequality $\overline{\mathrm{CIA}}(S) \leq 2\mathrm{CIA}(S)$, let B_i minimise $d_i = \max\limits_{j=1,\ldots,l} \mathrm{EMD}(B_i, B_j) = \mathrm{CIA}(S)$. For $j, k = 1, \ldots, G$, the triangle inequality

$$\mathrm{EMD}(B_k, B_j) \leq \mathrm{EMD}(B_k, B_i) + \mathrm{EMD}(B_i, B_j) \leq 2d_i = 2\mathrm{CIA}(S)$$

implies that $d_k = \max\limits_{j=1,\ldots,G} \mathrm{EMD}(B_k, B_j) \leq 2\mathrm{CIA}(S)$ for each $k = 1, \ldots, G$. Then $\overline{\mathrm{CIA}}(S) = \frac{1}{G}\sum\limits_{k=1}^{G} d_k \leq 2\mathrm{CIA}(S)$. □

Proof of Theorem 8. By Lemma 4.1 in (Edelsbrunner *et al.*, 2021), if periodic point sets $S, Q \subset \mathbb{R}^n$ are related by an ε-perturbation, then S, Q have a common lattice. Since CIA is invariant under changes of a unit cell by Lemma 6, we can assume that S, Q have the same number m of points in a common unit cell and equal Point Packing Coefficients $\mathrm{PPC}(S) = \mathrm{PPC}(Q)$ in Definition 3. By Lemma SM3.4 in (Widdowson & Kurlin, 2025*c*), all corresponding elements of $\mathrm{PDD}(S;k), \mathrm{PDD}(Q;k)$ differ by at most 2ε, which generalises the basic fact that perturbing any two points up to ε changes the distance between them up to 2ε by the triangle inequality. The same upper bound of 2ε holds for differences between all corresponding elements of $\mathrm{PDA}(S;k), \mathrm{PDA}(Q;k)$

in Definition 3 due to PPC(S) = PPC(Q). For both ground distances (Chebyshev and Root Mean Square) between rows of k distances, the upper bound of 2ε between corresponding distances $|b_i - c_i| \leq 2\varepsilon$, $i = 1 \ldots, k$, guarantees the same upper bound for $d_\infty = \max\limits_{i=1,\ldots,k} |b_i - c_i| \leq 2\varepsilon$ and $d = \sqrt{\frac{1}{k}\sum\limits_{i=1}^{k}(b_i - c_i)^2} \leq \sqrt{\frac{1}{k}\sum\limits_{i=1}^{k}(2\varepsilon)^2} = 2\varepsilon$.

Let $B_1, \ldots, B_G$ be all geometric blocks in an asymmetric unit of S. Denote by $C_1, \ldots, C_G$ the corresponding blocks in an asymmetric unit of Q so that each C_i is an ε-perturbation of B_i for $i = 1, \ldots, G$. By the argument above, all m_i corresponding points of B_i and C_i have 2ε-close rows in PDA($S; k$) and PDA($Q; k$), respectively, for $i = 1, \ldots, G$. Then $d(R_j(B_i), R_j(C_i)) \leq 2\varepsilon$ for $j = 1, \ldots, m_i$, where the ground distance d is Chebyshev or RMS. In the notations of Definition 4, if we set $f_{jj} = \dfrac{1}{m_i}$ for $j = 1, \ldots, m_i$, else 0, then EMD(B_i, C_i) $\leq 2\varepsilon$. The triangle inequality implies that

$$\mathrm{EMD}(B_i, B_j) \leq \mathrm{EMD}(B_i, C_i) + \mathrm{EMD}(C_i, C_j) + \mathrm{EMD}(C_j, B_j)| \leq \mathrm{EMD}(C_i, C_j) + 4\varepsilon.$$

Swapping the B-blocks and C-blocks, we similarly get EMD(C_i, C_j) $\leq$ EMD(B_i, B_j) + 4ε and $|\mathrm{EMD}(B_i, B_j) - \mathrm{EMD}(C_i, C_j)| \leq 4\varepsilon$, so the corresponding elements of the matrix of EMDs differ by at most 4ε. Then the maximum distances d_i in Definition 5 and hence the minima and averages over $j = 1, \ldots, G$ differ by at most 4ε. □

Synopsis

The new continuous invariant-based asymmetry quantifies a deviation of any periodic crystal from its closest higher symmetry neighbour where all molecules are geometrically equivalent.