

## Methods of measuring distances between chemical structures, shapes, and descriptors

Method	Goal and Key Inputs	Process or Definition	Limitations
Our tool, the <b>Cubic Deviation Metric (CDM)</b> <a href="https://arxiv.org/abs/2508.01177">arXiv: 2508.01177</a>	<b>Provide a numeric to quantify the deviation of unit cells from a given shape</b> , specifically a cube. Numerically anchor vaguely-defined terms such as “pseudocubic”. Volume normalized to allow comparison between unit cells of all sizes.	Polynomial derived from geometrical comparison to a perfect cube.	Unit cells of different shapes may produce identical or very similar values, making it impossible to rank all unit cells uniquely as more/less cubic. Can not compare cubic unit cells to each other. Appropriate use for supercells remains TBD.
<b>STRAIN method</b> <a href="https://cryst.ehu.es/cryst/strain.html">https://cryst.ehu.es/cryst/strain.html</a> <a href="#">Catti 1985</a>	<b>Provide a numeric to quantify the degree of lattice mismatch between two lattices</b> , particularly useful in thin film/epitaxial growth applications. Part of the Bilbao Crystallographic Server.	Lagrangian strain tensor calculated from unit cell parameters. Its eigenvalues can be used to calculate the numeric “degree of lattice distortion”.	Capable only of comparing two lattices to each other. No reference lattice exists and values are not normalized by unit cell volume, making it nonideal for comparison between different materials families. The numeric value is not commutative – it changes depending on which of the two structures is used as a reference.
<b>Voronoi-based similarity measures</b> <a href="#">Mosca &amp; Kurlin 2020</a>	<b>Provide a mathematically continuous metric to quantify the similarity of lattices</b> . Implementation of efficient algorithms enables filtering/clustering of large databases.	A distance measure based on Voronoi cells (defined as all points within a fixed $r$ distance from an atom in the unit cell) is calculated.	Voronoi areas/volumes change continuously but their shapes do not, making this method extremely susceptible to small perturbations in atomic positions, such as from experimental noise.
<b>Continuous Symmetry Measures (CSMs)</b> <a href="#">Zabrodsky, Peleg, &amp; Avnir 1992</a> <a href="#">Zabrodsky, Peleg, &amp; Avnir 1993</a> <a href="#">Pinsky, Dryzun, Casanova, &amp; Avnir 2008</a> <a href="#">Dryzun, Zait, &amp; Avnir 2011</a>	<b>Provide a numeric to quantify the deviation of molecular shapes from a given symmetry</b> (i.e. group or element of a group), notably <u>not</u> from a specific shape having said symmetry. There is a similar version for chirality (CCM <a href="#">Zabrodsky &amp; Avnir 1995</a> ).	Algorithm determines the distance (scaled to enforce a maximum value) each vertex must travel to produce the desired symmetry. The metric value is the average of these distances squared.	Most useful when the reference polyhedron is a Platonic solid, which is rare. The order of symmetry operations tested can result in multiple (but very similar) values for one specific shape and symmetry. For very large molecules

<u>Tuvi-Arad, Shalit, &amp; Alon 2024</u>			(i.e. complex shapes), only approximate algorithms exist.
<b>Continuous Shape Measures (CShMs)</b> <a href="#">Alvarez, Avnir, Llunell, &amp; Pinsky 2002</a> <a href="#">Casanova et al. 2004</a> <a href="#">Alvarez et. al. 2005</a>	<b>Provide a numeric to quantify the deviation of coordination polyhedra from a given shape.</b> One may also define a minimal distortion interconversion path between two polyhedra. This method can then be extended to quantifying deviation from some conversion path when discussing atomic rearrangements.	1. Find reference shape 2. Calculate the distances between the equivalent atomic positions in the real and reference shapes 3. Average squared distances 4. Optionally construct a “shape map” plot comparing value for multiple reference polyhedra and conversion paths between them.	It can be difficult to define the reference polyhedra, and often there is no great way to uniquely select one over another.
<b>Minimum Bounding Ellipsoid (MBE) approach</b> <a href="#">Cumby &amp; Attfield 2017</a>	<b>Provide a numeric to quantify the degree of distortion in coordination polyhedra, without reference to a specific shape</b> (useful for e.g. weird 7-coordinate systems). “Perfect” reference = a sphere and any compression, elongation, and/or angle changes yields different values.	Minimization algorithm to find three principal radii and a rotation matrix for the ellipsoid that contains all atoms in the polyhedron.	Any Platonic solid produces the minimum value of 0 = sphere, making this approach nonideal for comparing two Platonic solids (although this is a rare use case).
<b>CRYCOM</b> <a href="#">Dzyabchenko 1994</a> (see also <a href="#">Burzlaff &amp; Rothammel 1992</a> and <a href="#">Gelato &amp; Parthé 1987</a> )  (see also similar methods: CMPZ <a href="#">Hundt, Schönberg, &amp; Jansen 2005</a> and XtalCOMP <a href="#">Lonie &amp; Zurek 2012</a> )	<b>Identify similar crystal structures from information available in a CIF file</b> (aka the “formal parameters”). The degree and type of similarity is also of interest, and multiple measures of similarity are provided. This method incorporates earlier work matching Wyckoff positions only (STRUCTURE-TIDY <a href="#">Gelato &amp; Parthé 1987</a> ).	Algorithm comparing first structure fragments and then collections of fragments (whole structure) for low-symmetry structures.	The tool can’t relate structures on a “number line” by itself, but <a href="#">Burzlaff &amp; Rothammel</a> discuss this concept. Multiple measures of variation/distortion/similarity, with different applications requiring the user to look at different values. Works for structures of the same symmetry and same atom types.
<b>COMPACT</b> <a href="#">Chisholm &amp; Motherwell, 2004</a>	<b>Identify duplicate descriptions of crystal structures</b> or sections, to better optimize structure-prediction.	The nearest neighbors of a central atom (coordination sphere) are searched against possible	All of the nearest neighbors must be identified; the number to look for must be adjusted, and larger

<p>(see also similar method with noise tolerance, the Pointwise Distance Distribution <a href="#">Widdowson &amp; Kurlin 2025</a>)</p>	<p>Uses a generic search query based on molecular packing. Part of the Cambridge Structural Database.</p>	<p>duplicate sets. The RMS of the distances between each atom and its potential match is reported.</p>	<p>numbers are slower to compute. The mapping procedure is only designed to compare objects with the same symmetry and atoms. The approach is sensitive to small amounts of experimental noise.</p>
<p><b>COMPSTRU</b> <a href="https://cryst2.ehu.es/cryst/compstru">https://cryst2.ehu.es/cryst/compstru</a> <a href="#">de la Flor et al. 2016</a> (see also <a href="#">STRAIN</a> detailed above and the “structural descriptor” <math>\Delta</math> described in <a href="#">Bergerhoff et al. 1999</a>)</p>	<p><b>Quantify similarity of crystal structures/descriptions when the atoms are different</b> using only the information available in a CIF file. Certain similarity measures can be used to group structures into a series of classes. Part of the Bilbao Crystallographic Server.</p>	<ol style="list-style-type: none"> <li>1. Develop an equivalent description of structure #1 so its atoms are best aligned with those in #2</li> <li>2. Match the atoms in #1 and #2</li> <li>3. Report lattice distortion (see <a href="#">STRAIN</a>, above), atomic displacements from #1 to #2, and weighted mean differences of the atomic coordinates (see <a href="#">Bergerhoff et al. 1999</a>)</li> </ol>	<p>Designed to compare two structures at a time. Only works for structures of the same space group. Multiple measures are provided, with different values most useful in different applications. None of the three values reported are commutative – they change depending on which of the two structures is used as a reference. Values are also not continuous.</p>
<p><b>Geometric- and symmetry-adapted-cost metrics</b> <a href="https://prisms-center.github.io/CASMcode_docs/">https://prisms-center.github.io/CASMcode_docs/</a> <a href="#">Thomas, Natarajan, &amp; Van der Ven 2021</a>  (see also similar method: AFLOW-XtalFinder <a href="#">Hicks et al. 2021</a> or one implemented without unit cells <a href="#">Therrien, Graf, &amp; Stevanović 2020</a>)</p>	<p><b>Separately optimize the cost of both geometrical and symmetrical changes between structures.</b> As part of this process, methods of quantifying geometrical and symmetrical differences (between two structures at a time) are developed. Geometric here is a combination of volume-normalized stretching, strain, and atomic displacement. Different symmetries (spacegroups) are allowed. Part of the CASM software package.</p>	<ol style="list-style-type: none"> <li>1. Map lattice #2 onto #1, ignoring any symmetrical equivalents</li> <li>2. Reposition atoms in structure #2 to match locations in #1</li> <li>3. Calculate the lattice deformation, atomic displacement, and symmetry-breaking cost functions.</li> <li>4. Perform minimization to determine the “optimum” transformation between structures #1 and #2.</li> </ol>	<p>Compares only two structures at a time. Optimization efficiency dictates the speed of this algorithm.</p>
<p><b>Fingerprint functions and distance method</b> <a href="#">Oganov &amp; Valle 2009</a></p>	<p><b>Uniquely describe a [local] structure in a memory-efficient way and measure difference between</b></p>	<p>Normalized sum over all interatomic distances. Matrix form allows atomic ordering information</p>	<p>Although straightforward and memory-efficient from a computing standpoint, the fingerprint of a</p>

<p><u>Oguchi 2024</u></p> <p>(see also similar approaches: <a href="#">Zhu et al. 2016</a>; <a href="#">Yang, Dacek, &amp; Ceder 2014</a>; SOAP <a href="#">De et al. 2016</a>; Atomic Neighbor Densities <a href="#">Bartók, Kondor, &amp; Csányi 2013</a>; Radial Distribution Function method <a href="#">Willighagen et al. 2005</a>; CALYPSO <a href="#">Wang et al. 2012</a>; CCF <a href="#">Su et al. 2017</a>)</p>	<p><b>structures.</b> Closely related to a structure factor, with an analytical relation to the neutron scattering <math>S(Q)</math> and an approximate relation to the x-ray <math>S(Q)</math>. Approximating energy as pairwise interaction potentials, it can also be related to total energy. Can be used to predict phase transitions/new phases and enhance monitoring of structure-prediction programs.</p>	<p>to be preserved. When represented as vectors, a function may also be applied to measure the distance between two fingerprints (structures).</p>	<p>specific structure is not easy to analyze at a glance. Precise atomic positions must be known. Fingerprints do not establish a 1-to-1 correspondence between atoms, although for describing overall structures of periodic systems, this may be less of a concern. Some versions contain angular info (e.g. CALYPSO and CCF) or more atomic info (e.g. SOAP).</p>
<p><b>Tolerance factors:</b>  <a href="#">Goldschmidt 1926</a>  <a href="#">Mouta, Silva, &amp; Paschoal 2013</a>  <a href="#">Song, Zhou, &amp; Liu 2019</a>  <a href="#">Song &amp; Liu 2020</a>  <a href="#">Bassen et al. 2024</a>  <a href="#">Tschauner 2025</a></p>	<p><b>Predict material structure prior to synthesis.</b> Usually based on a sphere packing approximation using theoretical atomic radii and roles (cation vs. anion, etc.).</p>	<p>Polynomial derived from geometric understanding of sphere packing in a given unit cell shape.</p>	<p>Limited by accuracy of atomic radii and usually empirical. Often materials are exceptions rather than following the “rules” due to more complicated real physical behavior. Only works for solids.</p>

*This is not an exhaustive list; the reader is referred to the introductions of recent papers linked above, which often include a brief summary of pre-existing methods.*



= suitable for analysis by hand



= implemented for computing applications (algorithmic/optimization)

### **Key definitions:**

Lattice: a set of vectors or points which is periodic in space; unless speaking of an *atomic* lattice, this is usually a theoretical representation

Unit cell: a theoretical parallelepiped built on the basis vectors (crystallographic axes) of a lattice

Structure: the periodic arrangement of atoms in a real crystal; can be *represented* by a lattice/unit cell but is not the same thing

Coordination polyhedra: an approximation of the bonding environment around a central atom where bonds are formed only between the central atom and its nearest neighbors, and the locations of each neighbor are the vertices of a hypothetical polyhedron