A continuous map of the Cambridge Structural Database in meaningful coordinates

Daniel E Widdowson, Vitaliy A Kurlin

Materials Innovation Factory, University of Liverpool, United Kingdom, vitaliy.kurlin@gmail.com

Keywords: Crystal Isometry Principle, periodic crystals, data standard, materials genome, structural invariants, crystal descriptors

Since crystal structures are determined in a rigid form, their most practical equivalence is rigid motion, which is a composition of translations and rotations. If we also allow mirror reflections, we get a general isometry maintaining all inter-point distances.

Crystal structures can be distinguished up to isometry only by invariants that are preserved under all isometric transformations. Since atomic coordinates are not preserved even under translation, there are not isometry invariants. The parameters of Niggli’s reduced is invariant but is discontinuous [1] under almost any tiny displacement of atoms, which can make a primitive cell larger.

The physical density is a continuous invariant but is too weak to reliably distinguish many crystals. A strongest invariant is called complete and can be considered a materials genome or a DNA-style code that uniquely identifies any periodic crystal in practice.

The recent AMD (Average Minimum Distances) [2] and PDD (Pointwise Distance Distributions) [3] form an infinite hierarchy of isometry invariants that theoretically distinguish all periodic structures of atomic centers (without any chemical information) up to isometry in all general (non-singular) configurations. These invariants are so fast that more than 200 billion pairwise comparisons of all (more than 660 thousand) periodic crystals in the Cambridge Structural Database (CSD) [1] were completed within two days on a modest desktop and detected five pairs of unexpected duplicates that have all geometric data (even structure factors) identical to the last decimal place, but one atom was replaced with another one, for example, Cd with Mn in the pair HIFCAB vs JEPLIA.

Figure 1. The heat map of all periodic crystals in the Cambridge Structural Database by using density and new invariant AMD1.

The resulting Crystal Isometry Principle (CRISP) says that all periodic structures (irrespective of chemical composition) live in a common Crystal Isometry Space parametrized by complete invariant coordinates. The first maps for 2D lattices [4-5], Figure 1 visualizes the whole CSD in two meaningful coordinates, which can be replaced by more sophisticated ones in an interactive way.