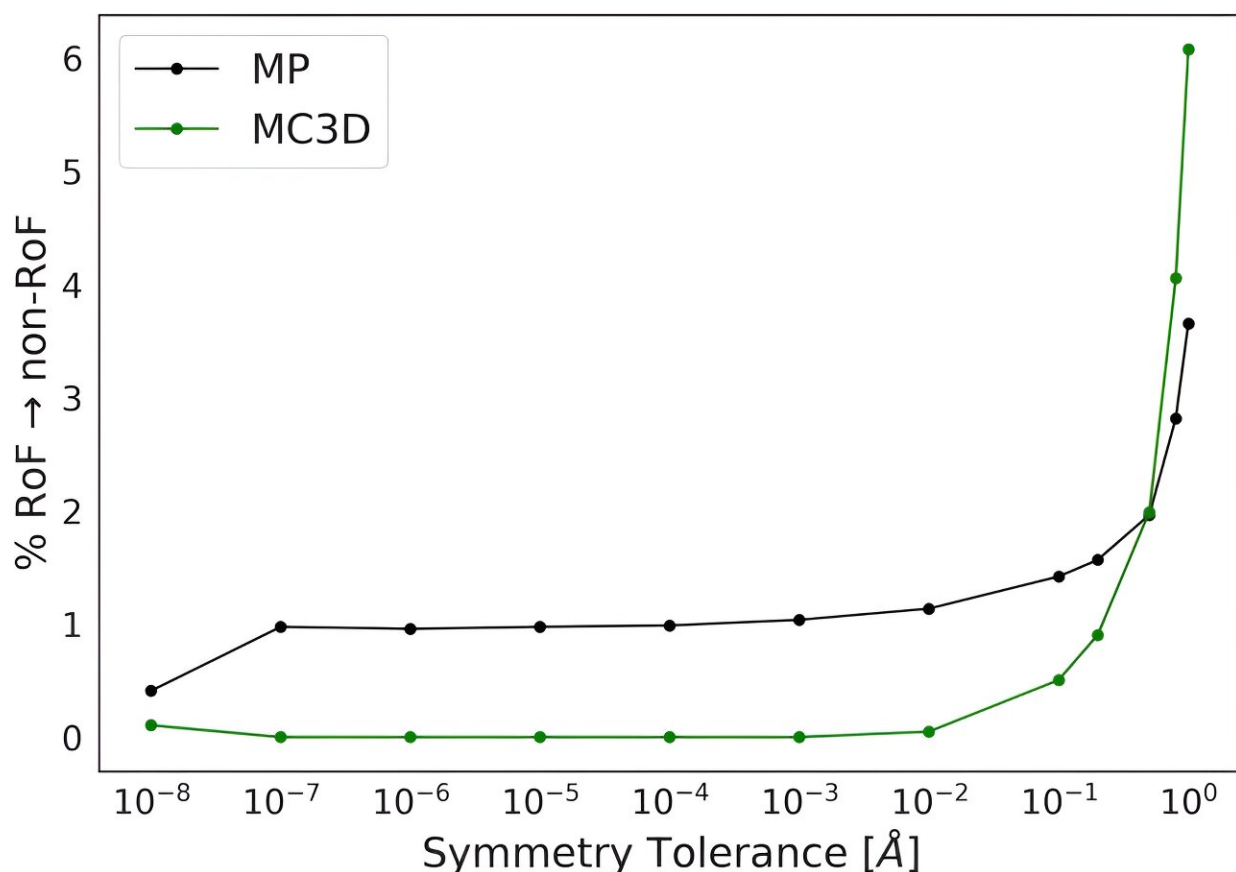# Materials follow the 'Rule of Four,' but scientists don't know why yet

April 18 2024

Percentage of RoF structures that become labeled non-RoF as a function of the symmetry tolerance parameter used for reduction to the primitive cell. The black and green lines correspond to structures in the MP and MC3D-source datasets, respectively. At typical symmetrization parameters, there is little to no change in the number of RoF structures (roughly 1% of RoF structures go to non-RoF). At larger symmetrization parameters (≈1 Å), this increases to roughly 6% based upon the large deviations allowed in considering sites as symmetrically

Scientists are normally happy to find regularities and correlations in their data—but only if they can explain them. Otherwise, they worry that those patterns might just be revealing some flaw in the data itself, so-called experimental artifacts.

That's what scientists in Nicola Marzari's group at the Swiss Federal Institute for Technology in Lausanne (EPFL) worried about when they noticed an unexpected pattern in two widely used databases of electronic structures, the Materials Project (MP) database and the Materials Cloud 3-dimensional crystal structures 'source' database (MC3Dsource).

The two collections include over 80,000 electronic structures of experimental as well as predicted materials, and in principle, all types of structures should be equally represented. However, scientists noticed that around 60 percent of structures in both databases have primitive unit cells (the smallest possible cell in a crystal structure) made out of a multiple of 4 atoms. The scientists named this recurrence the "Rule of Four" and started looking for an explanation.

"A first intuitive reason could come from the fact that when a conventional unit cell (a larger cell than the primitive one, representing the full symmetry of the crystal) is transformed into a primitive cell, the number of atoms is typically reduced by four times," says Elena Gazzarini, a former INSPIRE Potentials fellow in the Laboratory of Theory and Simulation of Materials (THEOS) at EPFL and now at CERN in Geneva.

"The first question we asked was whether the software used to

'primitivize' the unit cell had done it correctly, and the answer was yes."

From a chemical point of view, another possible suspect was the coordination number of silicon (the number of atoms that can bind to its atom), which is four. "We could expect to find that all the materials following this Rule of Four included silicon," says Gazzarini. "But again, they did not."

The formation energies of the compounds could not explain the Rule of Four either. "The materials that are most abundant in nature should be the most energetically favored, which means the most stable ones, those with negative formation energy," says Gazzarini. "But what we saw with classic computational methods was that there was no correlation between the Rule of Four and negative formation energies."

Because the materials space covered by the two databases is huge, going from small units to very large cells with dozens of different chemical species, there was still a chance that a more refined analysis looking for a correlation between formation energies and chemical properties may provide an explanation.

So, the team involved Rose Cernosky, a machine-learning expert at the University of Wisconsin, who developed an algorithm to group structures according to their atomic properties and look at formation energies within classes of materials sharing some chemical similarities. But again, this method did not provide a way to distinguish the Rule-of-Four compliant materials from the non-compliant ones.

Similarly, the abundance of multiple of fours does not even correlate with highly symmetric structures but rather with low symmetries and loosely packed arrangements.

In the end, the resulting article in *npj Computational Materials* is the rare

example of a scientific paper describing a negative result: the researchers could only describe the phenomenon and rule out several possible causes, without finding one.

But negative results can be just as important as positive ones for scientific advancement, because they point to difficult problems—which is why scientists often complain that journals should publish more such studies.

The failure to find a compelling explanation did not prevent the group from predicting, through a Random Forest algorithm, with an accuracy of 87% whether a given compound will follow the Rule of Four or not. "This is interesting because the algorithm uses only local rather than global symmetry descriptors, which suggests that there may be small chemical groups in the cells (still to be found) that may explain the Rule of Four," says Gazzarini.

More information: Elena Gazzarrini et al, The rule of four: anomalous distributions in the stoichiometries of inorganic compounds, *npj Computational Materials* (2024). DOI: 10.1038/s41524-024-01248-z

Provided by National Centre of Competence in Research (NCCR) MARVEL