CHEMISTRY OF MATERIALS

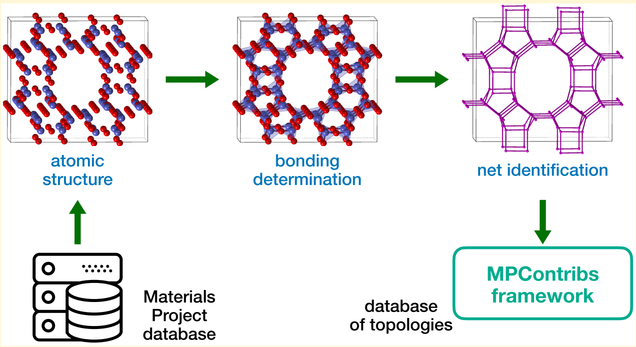
Large-Scale Characterization of Chemical Bonding and Topology in the Materials Project Database

Lionel Zoubritzky and François-Xavier Coudert*

Cite This: *Chem. Mater.* 2025, 37, 3137−3146 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Topology is key to the determination of many physical and chemical properties of materials, such as electrical, optical, and magnetic properties, as well as thermal and mechanical behavior. However, despite the growing number of databases of crystalline materials available, there has been very little systematic effort to date to analyze their topology. In this work, we have leveraged recent algorithmic advances in the analysis of chemical bonding and topology determination in order to perform high-throughput analysis of topology of materials on a large-scale database of existing and hypothetical materials, the Materials Project data set of more than 170,000 structures. Beyond the statistical analysis of the most frequent topologies and coordination environments, the publication of these topological data will allow researchers to search for materials by topology and chemical environment, paving the way to enhanced performance in materials screening for applications. We demonstrated two examples of the usefulness of topological considerations in such computational screening.



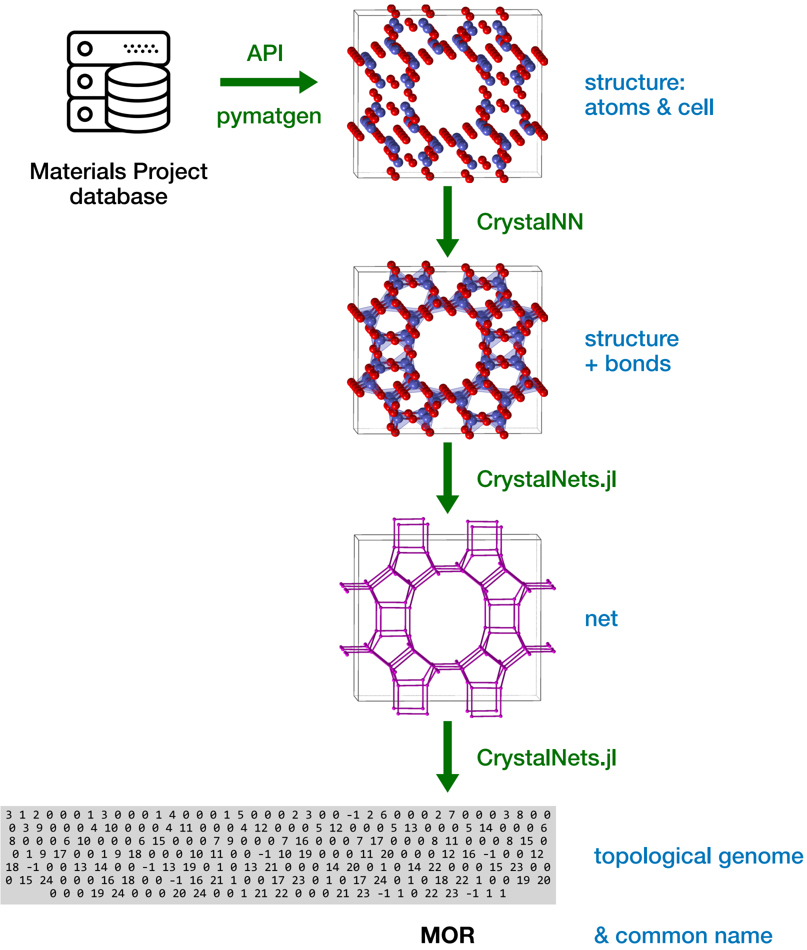
INTRODUCTION

The physical and chemical properties of materials at the macroscopic level naturally arise from the nature and microscopic arrangement of their atoms through the laws of statistical physics and the nature of the interactions between the atoms. In this, some geometrical quantities related to the local organization of matter are often considered as key by scientists because they play an important role in structure/property relationships: distances between nearest neighbors, bond angles, dihedral angles, etc. However, these very local descriptors of the environment of each atom cannot fully account for the macroscopic properties of the material. Longer-range information, related to the "packing" or "ordering" of atoms into space, is also important.[1−3] Features such as symmetry and topology play a critical role in determining the physical properties of materials, shaping their behavior across a range of fields from condensed matter physics to nanotechnology.[4,5] Topology is key to the determination of many physical and chemical properties, such as electrical, optical, and magnetic properties, as well as thermal and mechanical behavior, among others—in ways that go beyond simple chemical composition or crystallographic structure.[6,7]

This is particularly true for crystalline materials and framework materials built from building blocks (which can be individual atoms, molecular fragments, or larger structural elements) into ordered, periodic three-dimensional structures. A lot of effort has been extended to find topological characteristics that maximize the occurrence of specific desirable properties or suppress unwanted behavior. This concept is also at the core of the reticular chemistry approach,[8] in which well-defined molecular building blocks are assembled by strong bonds into crystalline extended frameworks, in a LEGO-like manner. There, the properties of the crystal derive from the properties of the individual building blocks as well as the topology of the three-dimensional net.

Recent years have seen the development and widespread use of an important number of databases of materials, whether experimental or hypothetical. Data-based approaches to the discovery of novel materials—or the identification of known materials with optimal performance for a specific application—are becoming a common research theme.[9,10] However, despite the many databases of crystalline or framework materials available, there has been very little systematic effort to date to analyze their topology. We think that this blind spot in previous research efforts is due to the combination of two factors. First, the determination of a material's topology

Received: November 15, 2024
Revised: April 4, 2025
Accepted: April 7, 2025
Published: April 17, 2025



```
3 1 2 0 0 0 1 3 0 0 0 1 4 0 0 0 1 5 0 0 0 2 3 0 0 -1 2 6 0 0 0 2 7 0 0 0 3 8 0 0
0 3 9 0 0 0 4 10 0 0 0 4 11 0 0 0 4 12 0 0 0 5 12 0 0 0 5 13 0 0 0 5 14 0 0 0 6
8 0 0 0 6 10 0 0 0 6 15 0 0 0 7 9 0 0 0 7 16 0 0 0 7 17 0 0 0 8 11 0 0 0 8 15 0
0 1 9 17 0 0 1 9 18 0 0 0 10 11 0 0 -1 10 19 0 0 0 11 20 0 0 0 12 16 -1 0 0 12
18 -1 0 0 13 14 0 0 -1 13 19 0 1 0 13 21 0 0 0 14 20 0 1 0 14 22 0 0 0 15 23 0 0
0 15 24 0 0 0 16 18 0 0 -1 16 21 1 0 0 17 23 0 1 0 17 24 0 1 0 18 22 1 0 0 19 20
0 0 0 19 24 0 0 0 20 24 0 0 1 21 22 0 0 0 21 23 -1 1 0 22 23 -1 1 1
```

Figure 1. Schematic representation of the workflow for topology determination, illustrated on the structure of the pure-silica mordenite zeolite, corresponding to Materials Project `mp-600063`. The topological genome for that net is a series of 241 integer values, and its three-letter code from the IZA-SC database is MOR.

requires the knowledge of chemical bonds in the system. While many heuristics exist for specific subclasses of systems, the determination of bonds is not a trivial problem, especially at large scale. Second, applying topology determination algorithms to a large number of structures (tens of thousands or more) requires the use of very fast, optimized software, which was not readily available until recently.[11]

In this article, we detail our methodological work to determine the topology of materials from databases at a large scale and analyze the resulting data. We demonstrate the usefulness of this approach by applying to more than 170,000 inorganic crystalline structures from the Materials Project database and publishing this new data set of chemical environment and topology information. This is the first systematic, high-throughput determination of material topology in a database of this size. We highlight how this can be statistically analyzed and propose two concrete examples of how topological considerations can be used in high-throughput computational screening for the identification of materials with desirable properties.

■ METHODS

We provide below a summary of the methodology followed in the present work. The associated code and data are available online at https://github.com/fxcoudert/topology_databases. A schematic representation of the workflow for topology determination is given in Figure 1.

Materials Project Database. We studied all inorganic structures available at the date of the study from the Materials Project database,[12,13] a part of the Materials Genome Initiative whose goal is to apply high-throughput computing to map the properties of "all known inorganic materials". The Materials Project API was queried using the `pymatgen` package (version 2025.2.18).[14]

The study was performed on database version `2025.02.12.post`, which was the latest version at the time of writing.[1] The query returned 170,470 crystalline inorganic structures, corresponding to known materials and many hypothetical ones, all optimized through Density Functional Theory (DFT) calculations of their atomic positions and cell parameters. The structures from the Materials Project are generally under the CC BY license (Creative Commons Attribution 4.0 International License), except the subset of structures from the GNoME database,[15] which are distributed under the terms of CC BY NC (Creative Commons Attribution Noncommercial 4.0 International License): this accounted for 15,483 structures.

Determination of Bonding. For all the structures retrieved from the Materials Project database, we performed calculations of bonding patterns through use of the CrystalNN method.[16] CrystalNN is a novel algorithm for determining near neighbors in crystals, introduced in 2021 and benchmarked against well-established algorithms on a set of reference crystal structures spanning elements, binaries, and ternary compounds. It has a stronger fundamental background than most of the previous algorithms, which typically rely on ad hoc heuristics. CrystalNN uses Voronoi decomposition[17] and solid angle weights to determine the probability of various coordination environments and select the one with highest probability. It is the same algorithm that is widely used throughout the Materials Project, including to display bonds on the web portal. CrystalNN is also the algorithm that is used for bond determination in the Materials Project's robocrystallographer, an open-source toolkit for analyzing crystal structures:[18] it generates textual descriptions including the local coordination and polyhedral type, polyhedral connectivity, octahedral tilt angles, and dimensionality.

We note here that while CrystalNN can be reparameterized for specific classes of systems, we used it here with its default parameters, tuned and recommended for inorganic materials. In order to improve the accuracy of the method, we made an exception for one specific class of system: intermetallic structures, defined as compounds that contain only metallic atoms. For those following the recommendations laid out in Pan et al.[16] (see section 8 of their Supporting Information), we used CrystalNN without electronegativity weighting (setting `x_diff_weight = 0`). Intermetallic structures account for 16.1% of the Materials Project database.

When applied to the 170,470 crystalline structures, CrystalNN successfully determined coordination environments in 170,455 structures (a failure rate of 0.0088%). It failed on 15 structures, where it could not find Voronoi neighbors for at least one atomic site: upon checking these structures, we therefore discarded them. For the remaining 170,455 structures, statistics about the coordination environments of each material were stored and later analyzed (see the Results section), and for each structure a CIF file was written including the bonding pattern—using the `geom_bond_atom_site_label`, `geom_bond_distance` and `geom_bond_site_symmetry` keywords from the CifCore dictionary.

There are several alternative methods to determine bonding between atoms. Some of the algorithms use heuristics refined over time, like those in Jmol[19] and OpenBabel,[20] while others are based on more mathematical foundations, such as the methods of Brunner,[21] Hoppe,[22] O'Keefe et al.,[23] and Isayev et al.[24] For a full review and comparison of these, we refer the reader to ref 16. Of course, the choice of bonding method impacts the choice of topology determination, but it is not entirely clear what is the "best" method at the moment—and it might depend on the category of systems studied. However, CrystalNN was found to perform well on a benchmark across multiple families of crystalline systems.[16]

One method that was developed at the same time as CrystalNN and is also implemented in `pymatgen` is the ChemEnv[25] algorithm, which has been field-tested by an analysis of the coordination environments in 8000 oxides[26] of the Inorganic Crystal Structure Database (ICSD).[27] ChemEnv features different possible strategies for the determination of the bonding and coordination environments of atoms in a crystal. However, even with the "simplest" strategy (`SimplestChemenvStrategy`), we found that the computational cost of ChemEnv analysis is larger than other methods, making it unsuitable to apply at the scale of hundreds of thousands of structures. To give an idea, the application of CrystalNN to the database took under four hours, with a timing of 81 ms per structure on average; ChemEnv analysis (performed on a random subset of 5000 structures) was 25 times slower, with 2.0 s per structure.

Determination of Topology. We then used the `CrystalNets.jl` software, previously developed in our group,[11] to retrieve the topology of all the structures. While CrystalNets internally has heuristics that can detect bonds when they are not provided, in this case, we strictly used the bonding information as provided by CrystalNN through CIF files. Given that we deal with inorganic structures, we used the `EachVertex` clustering type, meaning that we perform no clustering to identify secondary building units: each atom is its own vertex, and vertices with degree 2 or lower are iteratively collapsed into edges until all vertices have degree 3 or more. We also turn off the "split lone O vertex" option—a heuristic that is tuned for metal–organic frameworks but is unsuitable for oxygen-containing inorganic structures such as dense oxides.

For each crystal structure, CrystalNets then identified zero, one, or more periodic nets. If a structure does not have any net, it is called *aperiodic*: it means that it contains zero-dimensional clusters of atoms, unconnected to each other. On the other hand, a crystal structure may have multiple *interpenetrating nets*, a common case in porous framework materials—but rarer in dense inorganic crystals.

Then, each net in the structure is identified by its dimensionality (one, two or three-dimensional) and a unique identifier, its topological genome: it is a finite series of integers, such as `3 1 2 0 0 0 1 2 0 0 1 1 2 0 1 0 1 2 1 0 0` for the diamondoid (**dia**) net. This genome is provably unique for each net, depends only on the net itself and not its representation, and can be computed in polynomial time of the size of the net. In previous versions of `CrystalNets.jl`, some *unstable* nets[28] could not be assigned a unique genome: this happened when at least two vertices, in their equilibrium placement, have the same position—we refer the interested reader to ref 11. for more details on this. In the latest version, we introduced a new algorithm to uniquely identify unstable nets. However, this algorithm is factorial in complexity, while our algorithm for the general case is polynomial. Therefore, the algorithm is too slow to be used for some very large unstable nets: this accounted for 78 of 170,455 structures.

Working with topological genomes is suitable for computers but not for human chemists. Therefore, in a last task, the topological genome of each net is compared to the existing databases of named topologies: the Reticular Chemistry Structure Resource (RCSR),[29] the EPINET project,[30,31] and the zeolitic nets from the Structure Commission of the International Zeolite Association (IZA-SC).[32] The nets that have been attributed a name (often a three-letter code) by one of these entities can then be referred in a more human-readable fashion: **dia** for diamondoid net, SOD for sodalite zeolitic framework, etc. For nets that are not in any of these databases, we call them *unnamed*: they are uniquely identified by their genome but have not been attributed a name.

DFT Calculations. The single crystal elastic constants of selected structures from the Materials Project database were calculated in the density functional theory (DFT) approach with periodic unit cell and atomic basis sets, as implemented in the CRYSTAL17 code,[33] using the PBEsol exhange−correlation function[34] and POB double-ζ + polarization basis sets.[35] After the unit cell of the material is fully relaxed, optimizing both atomic positions and unit cell parameters, the elastic constants of these relaxed structures are then calculated by applying adimensional strains of ±1%.[36,37] The elastic stiffness matrices obtained were checked against Born's stability condition, namely that the symmetric stiffness tensor is positive-definite. The elastic tensors were then analyzed with the online ELATE application.[38]

Publication of the Resulting Data. All data produced in this work is published on our group repository at https://github.com/fxcoudert/topology_databases under the CC0 license, i.e., in the public domain. Additionally, all data was uploaded to the MPContribs framework under the `topology` project and available at https://next-gen.materialsproject.org/contribs/projects/topology. MPContribs is an extension of the Materials Project website that allows user-contributed data to be shown and analyzed alongside the core MP database.[39] This allows researchers to search for materials with specific topology or dimensionality, using the user-friendly Materials Project interface and its powerful in-browser visualization, as well as to link directly to other properties already present in the Materials Project.

RESULTS AND DISCUSSION

Diversity of the Topologies. Figure 2 presents the distribution of the dimensionality of nets identified in the Materials Project database—information that is available through our analysis but also stored in the MP itself and accessible through its `/materials/robocrys` API end point. We can see that 3.6% of the structures are aperiodic, i.e., correspond to clusters of atoms unconnected through space. 4.2% of the MP structures feature one-dimensional nets: these

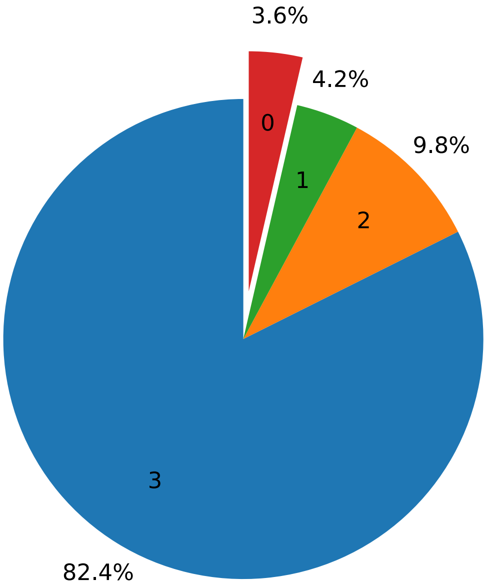


Figure 2. Distribution of dimensionality of nets identified in structures from the Materials Project database.

are typically linear chains of atoms, a net we call in this work the **rod** topology. Out of the 7631 one-dimensional nets identified, 6221 are **rod** nets; the remaining 1420 correspond to more complex one-dimensional nets (290 different types of 1D nets are represented). 2D nets represent 9.8% of the MP structures and will be discussed in more detail in their own section below. We focus in the rest of this section on three-dimensional nets, which represent the large majority of the MP database with 82.4% of structures.

The first thing we want to highlight from this analysis is the very large diversity of topologies encountered through the Materials Project database. In fact, the 149,204 materials with 3D nets are distributed into 40,246 different topologies, an average of ≈3.7 materials per topology. But the distribution is very unbalanced, with a fat-tailed nature: the most common topology (**pcu**) is encountered 8702 times, and the 20 most common topologies together account for 41,131 materials (27.6% of the database). On the other hand, the vast majority (30,101) of net topologies occur only once in the database. This huge diversity of the topologies encountered can be surprising or counterintuitive to solid-state chemists, who like to classify materials in well-known topologies (for example, for common minerals). However, the need for increasing topological diversity in databases of hypothetical materials has been highlighted before, for example, by Anderson et al.[40] in the case of metal−organic frameworks (MOFs). We believe that our present analysis reinforces this point: computational databases generated by enumeration of a limited number of identified topologies are inherently limited compared to the wide diversity observed here.

Another point that we can draw from this analysis is that the named nets, present in the RCSR, EPINET, and IZA-SC topology databases, only represent a small fraction of the many topologies identified. Indeed, of the 40,246 different topologies identified in the database, only 713 have common names (1.7%). The others are labeled *unnamed-n*, with integer values of *n* in decreasing order of frequency of their occurrence.

Most Frequent 3D Topologies. We present in Figure 3 the histogram of the occurrence of the most common 3D topologies in the Materials Project database, while Table 1 summarizes key properties of the 16 most common named 3D net topologies. It is not surprising that the most common topologies are very familiar to solid-state chemists, including various cubic-based structures (**pcu**, **bcu**, and **fcu**), hexagonal closest packing (**hcp**), and topologies from very common families of materials, such as perovskites, spinels, rutile, and diamond. Also unsurprisingly, the most common coordination environments include vertices with 4, 6, 8, and 12 neighbors, and less often, coordination numbers of 3, 10, and 16.

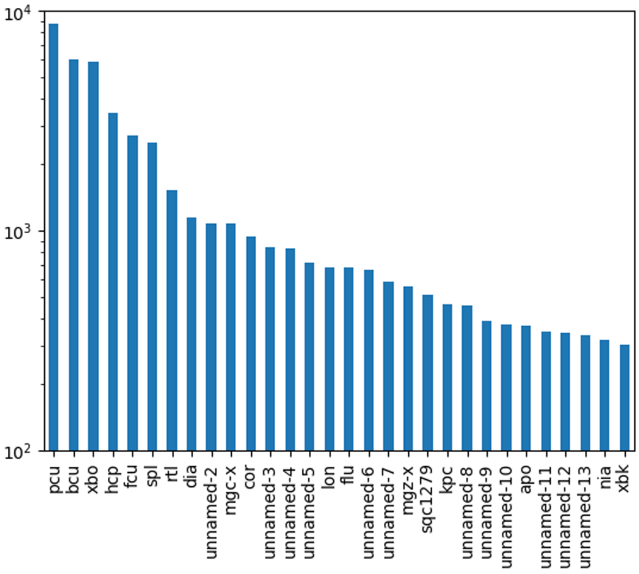


Figure 3. Most common three-dimensional nets in the Materials Project database (logarithmic *y* scale). Note that the first unnamed 3D net is "unnamed-2", because the most common unnamed net is a 2D net.

Table 1. Characteristics of the 16 Most Common Named 3D Net Topologies Encountered in the Materials Project Database

topology	common name	number of occurrences	type of vertices	coordination
pcu	primitive cubic lattice	8701	1	6
bcu	body-centered cubic	6044	1	8
xbo	perovskite	5882	3	6, 6 and 12
hcp	hexagonal closest packing	3443	1	12
fcu	face-centered cubic	2699	1	12
spl	spinel	2528	3	4, 4 and 6
rtl	rutile net	1538	2	3 and 6
dia	diamond	1143	1	4
mgc-x	$MgCu_2$, C15 Laves phase[41]	1074	2	16 and 12
cor	corundum	940	2	4 and 6
lon	lonsdaleite	682	1	4
flu	fluorite	678	2	4 and 8
mgz-x	$MgZn_2$, C14 Laves phase[41]	556	3	16 and 12
sqc1279		512	3	10 and 8
kpc	QOB net, K_2PtCl_6	461	3	12, 6 and 5
apo	α-PbO_2	369	2	3 and 6

Since we are not aware of any previous systematic study on topology in inorganic materials databases, we have no direct point of comparison for this list of topologies. What can, however, contrast the most frequent inorganic topologies with that of other materials that have been studied, namely, the supramolecular MOF frameworks. There, three recent studies have reported statistics of topologies: on the CoRE MOF database in ref 11, in the DigiMOF data set in ref 42, and in the ARC-MOF database in ref 43. All used the CrystalNets.jl[11] software, but on different sets of structures. Focusing on the most common 3D topologies, summarized in Table 2 in MOFs, we see that beyond relatively simple topologies like **pcu**, **bcu**, **rtl**, and **dia**, the top topologies of inorganics and MOFs differ for the most part. In fact, it appears clearly that MOFs, due to their supramolecular nature and large secondary building units, mostly feature low coordination numbers (3, 4, and 6 dominate), while dense inorganic materials often have topologies with higher connectivity.

Table 2. Characteristics of the 10 Most Common 3D Net Topologies Encountered in Three Different Metal-organic Framework (MOF) Databases[a]

CoRE MOF	DigiMOF	ARC-MOF
pcu (6)	**pcu** (6)	**pcu** (6)
dia (4)	**dia** (4)	**fsc** (4, 6)
ths (3)	**pts** (4)	**nbo** (4)
srs (3)	**rtl** (3, 6)	**pts** (4)
pts (4)	**cds** (4)	**sra** (4)
nbo (4)	**srs** (3)	**rna** (3, 6)
bpq (4, 6)	**sra** (4)	**dia** (4)
rtl (3, 6)	**bcu** (8)	**tfz-d** (3, 8)
lvt (4)	**ths** (3)	**cds** (4)
bcu (8)	**tfz-d** (3, 8)	**lvt** (4)

[a] Topology name, and coordination numbers between parentheses.

Focus on 2D Materials. As we have seen in Figure 2, materials with two-dimensional topologies represent 9.8% of the Materials Project database, corresponding to 17,577 crystalline structures. Also called "two-dimensional layered materials" (or simply "layered materials"), these materials consist of "sheets" without chemical bonds between them. They are interesting for two main reasons: (i) their layered nature gives rise to specific physical properties, marked by a very strong anisotropy; (ii) they can be considered precursors of truly two-dimensional materials (similarly to the relationship of graphite and graphene).

We plot in Figure 4 the distribution of the most common 2D topologies identified in the MP database. We can see four types of common 2D nets that are indeed very well-known:

- **kgd** is the Kagomé dual net, also called "rhombille tiling". It has one 6-coordinated vertex and one 3-coordinated vertex.
- **sql** is the square lattice, with only one type of vertex, which is 4-coordinated.
- **hcb** is the honeycomb net with one 3-coordinated vertex.
- **hxl** is the 6-coordinated hexagonal lattice, sometimes also called "triangular lattice".

It is also very interesting to note that the most common 2D net topology is unnamed; i.e., it is not featured in the RCSR database. It is, however, relatively simple and displayed in Figure 5 on the example of mp-7007, a layered $NbSe_2$ structure with this topology, which is known experimentally (it corresponds to ICSD entry 71,339).[44] This niobium(IV) selenide topology features 6-coordinated Nb and 3-coordinated Se atoms, with a hexagonal arrangement of niobium. It is interesting to note that this layered topology is very frequent in the Materials Project database, accounting for 4674 structures

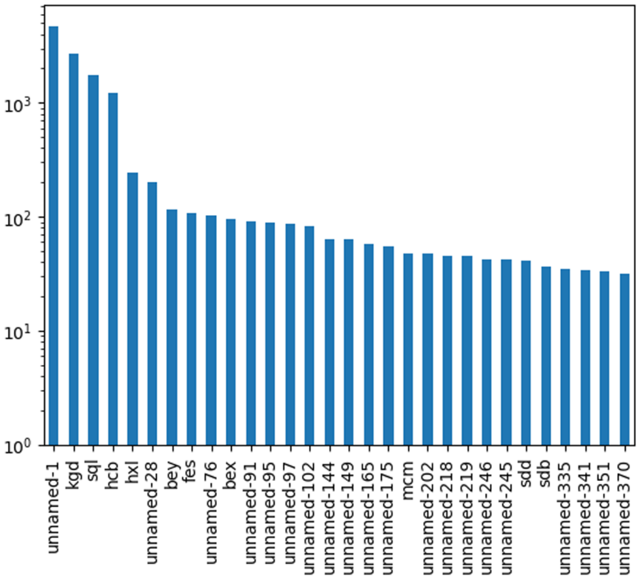


Figure 4. Most common two-dimensional nets in the Materials Project database (logarithmic y scale).

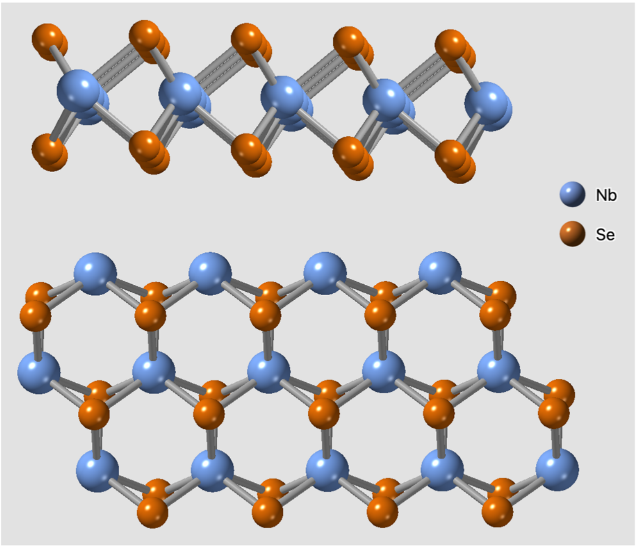


Figure 5. Illustration of the structure of material mp-7007, a layered $NbSe_2$ structure. Upper panel: view from the side; lower panel: view from the top.

The next net topologies in the list have a much lower occurrence, with less than 100 representatives in the database: the 3-coordinated **fes** net, and two **sql**-derived lattices mixing 3-fold and 4-fold coordination, **bex** and **bey**. They are illustrated in Figure 6. We note in passing that topologies with 5-fold coordination are very rare among 2D materials.

Statistics of Coordination Environments. As part of the data gathered through the use of the CrystalNN bond-detection algorithm, we store in our data set the coordination environment of all the materials. Meaning, for each material, we obtain a list of bonds present and the coordination number of each atom. This information is stored in JSON format like this:

```
{
  ’material_id’: ’mp-39105’,
  ’formula_pretty’: ’MgTa2N3’,
  ’nelements’: 3,
  ’theoretical’: True,
  ’is_stable’: None,
  ’crystal_system’: ’Hexagonal’,
  ’space_group’: 193,
  ’coordination’: [
      [’Mg’, 6],
      [’Ta’, 6],
      [’N’, 6]
    ],
  ’bonds’: [[’Mg’, ’N’], [’Ta’, ’N’]],
  ’topology’: [[1, 3, ’ste’]]
}
```

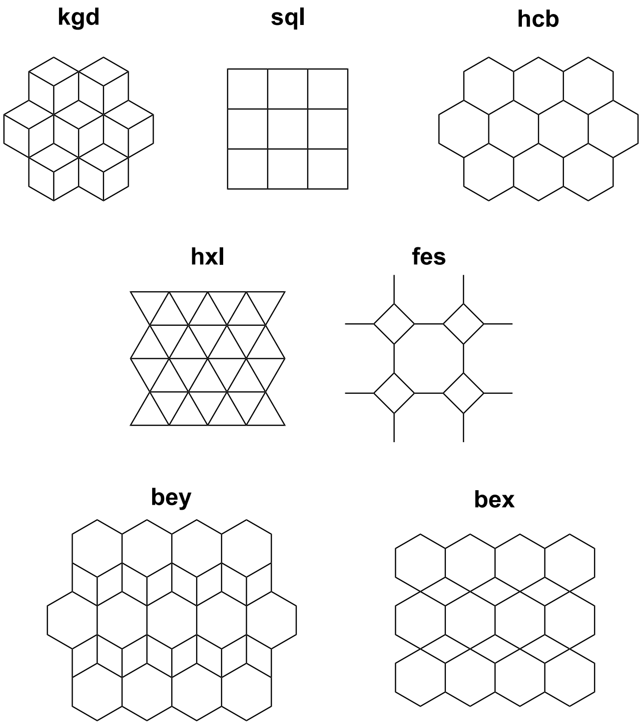


Figure 6. Illustration of some of the most common 2D nets in the Materials Project database.

In this case, we see that for material mp-39105, there are two types of bonds: Mg−N and Ta−N, and all atoms (Mg, Ta, and N) are systematically 6-coordinated. We also see that other information about the crystal system and space group, the chemical formula, and the topology is stored: here, the material has only one net, which has a three-dimensional **ste** topology.

This information on bonding and chemical environment can be used for searching through the Materials Project database in the future, answering questions such as “what materials contain Au−S or Ag−S bonds?”—there are 81 and 684, respectively. Or “what is the most common coordination number of Cu in compounds where it is bonded to F?”—it is 6, but 4- and 5-fold coordination also occur in significant numbers. Or “what are the most common coordination numbers for Ce?”—8 is the most common, followed by 6 and 12. Another possible use for this information is to draw statistics of coordination numbers for specific elements across the entire MP database, as shown in Figure 7.

Searching for Materials with Specific Topology or Coordination Environment. One of the goals of topological classification of materials in the MP database is the search or identification of materials with specific topologies. Computational high-throughput screening studies regularly use multi-objective optimization, i.e., the identification of top-performing materials for several physical or chemical properties, based on constraints on their chemistry (“the material should not contain Pb” or “should contain at least one S atom”), their geometry (“materials with two metal atoms with distance shorter than 2.5 Å”), or other properties (“presence of a cavity of diameter at least 4 Å”).[45] Before our present work, it was not possible to perform searches in the Materials Project database by topology. It is now possible to add criteria such as “structure with a five-coordinated S atom” or “structure with a diamond (**dia**) or diamond-derived topology (**dia**−*)”.

Because some physical properties are intrinsically tied to specific topologies, there is a clear interest in being able to filter and search material databases by topology. One very recent example is the search for classical spin liquids by looking at high symmetry nets and the impact of nearest neighbor interactions in these nets.[46] This lead the authors to identify, in addition to the well-known **crs** (pyrochlore) net, two other nets of interest that had not previously been investigated, **lcx** and **thp**.

We want to illustrate here briefly this approach on the MP database with two specific examples, tied to past work in our group. In the first case, we want to use topological search to identify materials with mechanical pliancy and negative linear compressibility, similar to the family of metal−organic frameworks MIL-53, which have a distinctive “wine-rack type” framework.[47] This framework has a net topology, **rna** or **bpq**, depending on the type of clustering used. Looking up materials with one of these two topologies in the MP database, we can identify 25 structures, including 16 experimental materials and 9 hypothetical structures. One family emerges in particular, of the type $ABF_{6-n}(H_2O)_{1+n}$ ($n = 0$ or 1), including for example materials like $NaPF_6H_2O$ (mp-767419) and $MgAlF_5(H_2O)_2$ (mp-24142). The structures are composed of BF_6 octahedra that share corners with AO_2F_4 octahedra, most often in the *Imma* space group (although some exceptions are in related *Imm*2 and *Pnna* space groups).

To confirm the potential of these structures for highly anisotropic mechanical properties, we performed DFT calculations on three matching structures, mp-23768 ($ZnAlH_4O_2F_5$), mp-24142 ($MgAlH_4O_2F_5$), and mp-725680 ($NaPOF_6$), in order to calculate the second-order elastic tensors of these materials and, therefore, their elastic mechanical properties. We find that all three materials exhibit anistropic mechanical properties, especially for their linear compressibility, with a factor of 4 between their stiffest and their softest directions. Of the three, $NaPOF_6$ is the only one that exhibits anomalous mechanical properties, with a very anistropic shear modulus and the existence of directions of negative Poisson’s ratio (a phenomenon also called *auxeticity*), as displayed in Figure 8. This is clearly linked to its “wine-rack type” structure, showing how screening of materials by topology can be integrated into multiscale high-throughput schemes for the identification of materials with targeted properties.

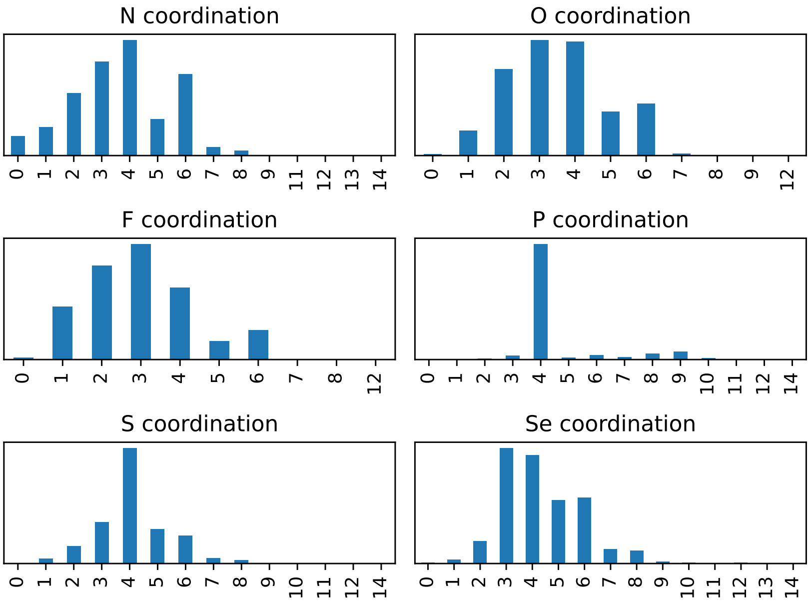


Figure 7. Distribution of coordination numbers for atoms N, O, F, P, S, and Se throughout all materials of the Materials Project database.

As a second example of identification of materials with specific topology, we focus here on the intriguing behavior of zirconium tungstate, ZrW_2O_8, which exhibits a number of counterintuitive mechanical responses to external stimuli: it has negative thermal expansion, very easy amorphization under hydrostatic pressure, and features negative hydration expansion, i.e., it contracts upon chemisorption of water molecules.[49] These unusual characteristics have been recently linked to its microscopic structure and in particular, the topology of this remarkable material (see Figure 9).[48] Looking then for structures with **fsn** topology in the Materials Project database, we identify 201 such materials, with 94 experimental structures and 107 hypothetical ones—including our original compound ZrW_2O_8 (a.k.a. mp-18778). Focusing on the experimental materials, 54 structures have the $Pa\bar{3}$ space group, 33 have the $P2_13$ space group, 5 have $Pca2_1$ symmetry, and two structures have $Pna2_1$ and $P1$ symmetry, respectively. If we further restrict our search for oxides (i.e., structures containing oxygen), we can identify four categories of candidate materials:

- only two tungstates: ZrW_2O_8 and HfW_2O_8 (expected due to the chemical similarity of zirconium and hafnium), in $P2_13$ space group;
- two vanadates: ZrV_2O_7 and HfV_2O_7 ($Pa\bar{3}$);
- several phosphates: MP_2O_7 for M = Si, W, Th, Mo, Ce, Ti, Sn, Zr (in the $Pa\bar{3}$ space group), as well as $InSb(P_2O_7)_2$ ($Pna2_1$) and UP_2O_7 ($P1$);
- some peroxides in the $Pa\bar{3}$ space group: NaO_2, ZnO_2, MgO_2, and CdO_2.

Further investigations would require in-depth screening of the properties of these materials, e.g., by DFT calculations and geometric determination of their void spaces, in order to identify the best candidates for thermal and mechanical properties similar to those of the parent compound ZrW_2O_8. We showed here how the availability of a materials database annotated with topological information can prove valuable in the exploration of chemical space of materials.

■ CONCLUSIONS AND PERSPECTIVES

From the viewpoint of a solid-state chemist, topology is a key feature of crystalline materials, as it influences many of their physical and chemical properties, including electrical, optical, and magnetic properties, as well as thermal and mechanical behavior. Crystalline topology has long been used to rationalize different categories of behavior, and researchers often look for materials with specific topologies in order to achieve the desired properties. So far, this has been a mostly manual process: despite the growing number (and size) of available databases of crystalline materials, these structures have not been annotated with topology information. Here, we have shown that it is now possible to achieve this goal at very large scale, by combining bonding detection algorithms (if bond information is not already present in the structures) and recent algorithmic advances in topology identification through the `CrystalNets.jl` library.

To demonstrate the capability of our workflow, we have applied this new workflow to the large-scale database of inorganic materials from the Materials Project, consisting of more than 170,000 structures—the first systematic determination of topology in a materials database of this size. This allowed us to report a statistical analysis of the most frequent topologies and coordination environments and publish this information online to allow researchers to search for materials by topology and chemical environment. This will pave the way to more efficient search in materials screening for applications, by allowing us to include more chemical insight and restrict the search space to specific topology, bonding patterns, and coordination environments. While we demonstrated two examples of the usefulness of topological considerations in such computational screening, we believe that this will be useful for many applications. It is one more tool in the toolbox of materials researchers; beyond its use in high-throughput screening, it will also be usable to experimentalists through the user-friendly MPContribs framework on the Materials Project website.

In the future, we intend to extend this approach to more material databases. Net detection and topology identification

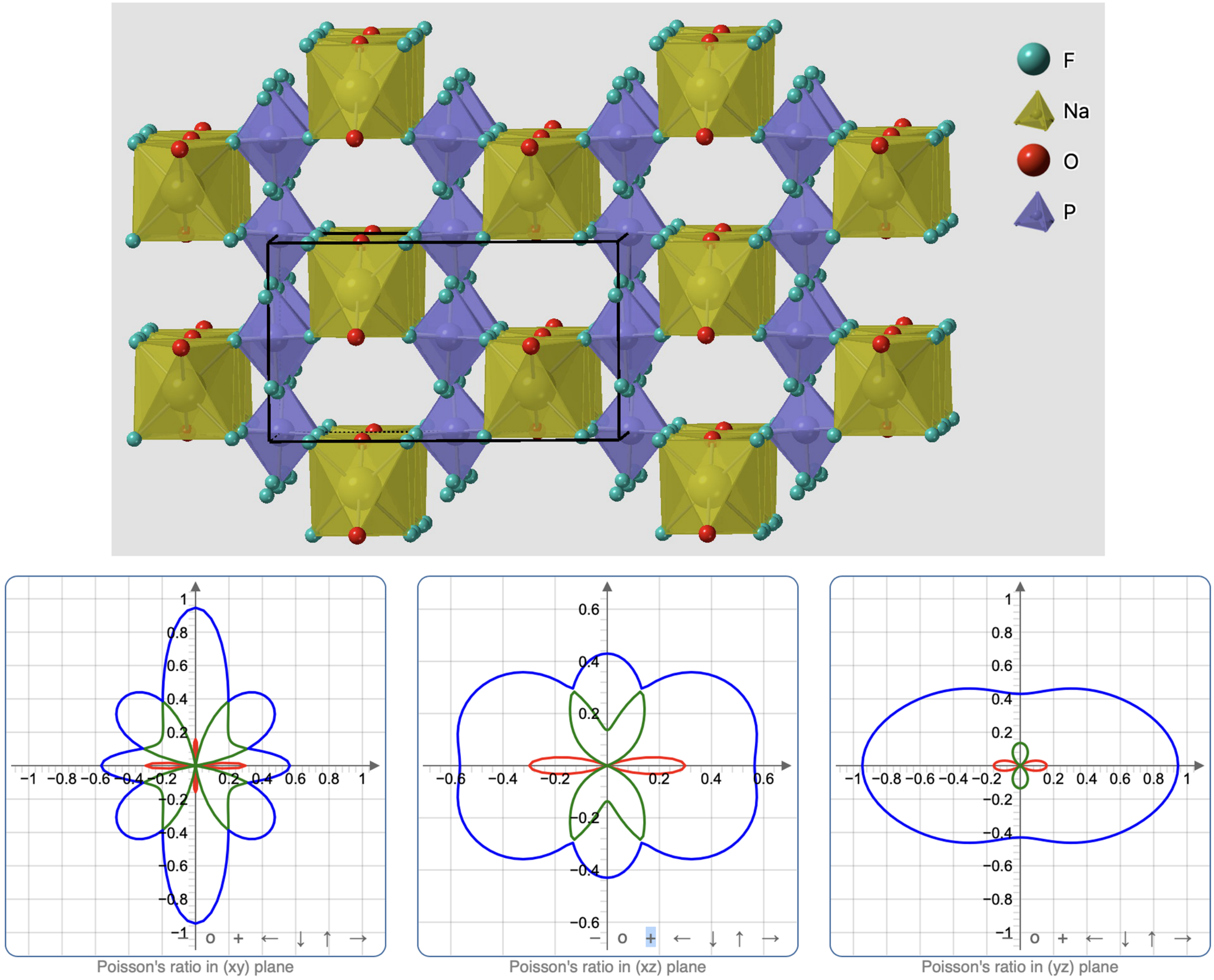


Figure 8. Top: structure of $NaPOF_6$ (mp-725680), a material with **bpq** topology. Bottom: 2D representations of the Poisson's ratio of this material, in the (*xy*), (*xz*), and (*yz*) planes. Blue: maximum value in each direction; green: minimum value in a given direction when it is positive; red: minimum value when it is negative, indicating a direction of auxeticity.

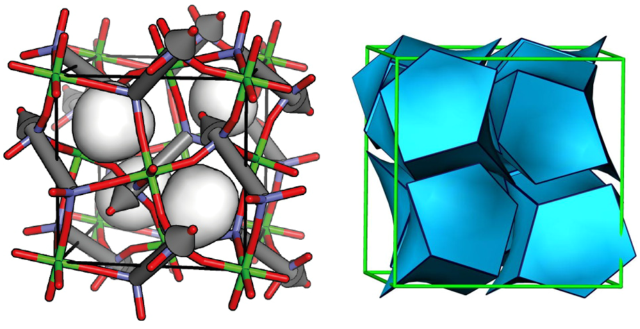


Figure 9. Left: crystal structure of ZrW_2O_8 (Zr in green, W in blue, and O in red), highlighting the four cavities of each unit cell as a white surface. Reproduced with permission from ref 48. Copyright 2018 American Physical Society. Right: tiling of the **fsn** topology of the material.

are easily automated, but the main limiting step at this moment in our methodology is detection of bonding. For inorganic structures, the CrystalNN algorithm has good accuracy,[16] but it is not well adapted and has important limitations for organic or organic-containing compounds or intermetallic structures. We have developed heuristics for subclasses of crystals, such as metal–organic frameworks, allowing us to screen MOF databases, such as CoRE MOF.[11] However, no truly general solution for accurate bond detection is available to the best of our knowledge. Further development in this direction is needed, and the database could also, in the future, include bonding information—as does the recently released MOSAEC-DB.[50] This will allow, in turn, to search for MOFs with specific properties linked to their topology, such as defect-tolerant MOFs with good mechanical stability[51] or acting as mechanical meta-materials.[52]

Another promising avenue of research is the use of more complex decomposition strategies to determine the topology in specific families of materials. For example, in ionic compounds, such as oxides, sulfides, or halides, one could choose a representation where only cations are included as vertices and bridging anions are transformed into edges.

AUTHOR INFORMATION

Corresponding Author

François-Xavier Coudert − *Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, Paris 75005, France;* orcid.org/0000-0001-5318-3910; Email: fx.coudert@chimieparistech.psl.eu

Author

Lionel Zoubritzky − *Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, Paris 75005, France; Air Liquide, Campus Innovation Paris, Les Loges-en-Josas 78350, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemmater.4c03172

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded under the France 2030 framework by Agence Nationale de la Recherche (project ANR-22-PEXD-0009 under PEPR DIADEM) and from ANRT CIFRE project 2021/1151 in association with Air Liquide. Access to HPC platforms was provided by a GENCI grant (A0170807069). We thank Frank Hoffmann for insightful discussions on 2D net topologies.

ADDITIONAL NOTE

[1] Information on the various Materials Project database versions is available at https://docs.materialsproject.org/changes/database-versions

REFERENCES

(1) Thomas, J. M. Topography and topology in solid-state chemistry. *Philos. Trans. R. Soc. London A* **1974**, *277*, 251−286.

(2) Zhao, D.; Timmons, D. J.; Yuan, D.; Zhou, H.-C. Tuning the Topology and Functionality of Metal−Organic Frameworks by Ligand Design. *Acc. Chem. Res.* **2011**, *44*, 123−133.

(3) Meekel, E. G.; Partridge, P.; Paraoan, R. A. I.; Levinsky, J. J. B.; Slater, B.; Hobday, C. L.; Goodwin, A. L. Enhanced elastic stability of a topologically disordered crystalline metal−organic framework. *Nat. Mater.* **2024**, *23*, 1245−1251.

(4) Tan, J. C.; Bennett, T. D.; Cheetham, A. K. Chemical structure, network topology, and porosity effects on the mechanical properties of Zeolitic Imidazolate Frameworks. *Proc. Nat. Acad. Sci.* **2010**, *107*, 9938−9943.

(5) Saurabh, S.; Gupta, A.; Chowdhury, R. Impact of parametric variation to achieve extreme mechanical metamaterials through topology optimization. *Compos. Struct.* **2023**, *326*, No. 117611.

(6) Cava, R. J.; Ji, H.; Fuccillo, M. K.; Gibson, Q. D.; Hor, Y. S. Crystal structure and chemistry of topological insulators. *J. Mater. Chem. C* **2013**, *1*, No. 3176.

(7) Bryant, M. J.; Maloney, A. G. P.; Sykes, R. A. Predicting mechanical properties of crystalline materials through topological analysis. *CrystEngComm* **2018**, *20*, 2698−2704.

(8) Jiang, H.; Alezi, D.; Eddaoudi, M. A reticular chemistry guide for the design of periodic solids. *Nature Rev. Mater.* **2021**, *6*, 466−487.

(9) Moosavi, S. M.; Jablonka, K. M.; Smit, B. The Role of Machine Learning in the Understanding and Design of Materials. *J. Am. Chem. Soc.* **2020**, *142*, 20273−20287.

(10) Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C. W.; Choudhary, A.; Agrawal, A.; Billinge, S. J. L.; Holm, E.; Ong, S. P.; Wolverton, C. Recent advances and applications of deep learning methods in materials science. *npj Comput. Mater.* **2022**, *8*, No. 59.

(11) Zoubritzky, L.; Coudert, F.-X. CrystalNets.jl: Identification of Crystal Topologies. *SciPost Chem.* **2022**, *1*, No. 005.

(12) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.

(13) Jain, A.; Montoya, J.; Dwaraknath, S.; Zimmermann, N. E. R.; Dagdelen, J.; Horton, M.; Huck, P.; Winston, D.; Cholia, S.; Ong, S. P.; Persson, K. *Handbook of Materials Modeling*; Springer International Publishing, 2018; pp 1−34.

(14) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.

(15) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80−85.

(16) Pan, H.; Ganose, A. M.; Horton, M.; Aykol, M.; Persson, K. A.; Zimmermann, N. E. R.; Jain, A. Benchmarking Coordination Number Prediction Algorithms on Inorganic Crystal Structures. *Inorg. Chem.* **2021**, *60*, 1590−1603.

(17) Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites. *J. Reine Angew. Math.* **1908**, *1908*, 97−102.

(18) Ganose, A. M.; Jain, A. Robocrystallographer: automated crystal structure text descriptions and analysis. *MRS Commun.* **2019**, *9*, 874−881.

(19) Hanson, R. M. Jmol − a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.* **2010**, *43*, 1250−1260.

(20) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, No. 33.

(21) Brunner, G. O. A definition of coordination and its relevance in the structure types AlB_2 and NiAs. *Acta Crystallogr., Sect. A* **1977**, *33*, 226−227.

(22) Hoppe, R. Effective coordination numbers (ECoN) and mean fictive ionic radii (MEFIR). *Z. Kristallogr. - Cryst. Mater.* **1979**, *150*, 23−52.

(23) O'Keefe, M.; Brese, N. E. Atom sizes and bond lengths in molecules and crystals. *J. Am. Chem. Soc.* **1991**, *113*, 3226−3229.

(24) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal fragment descriptors for predicting properties of inorganic crystals. *Nature Commun.* **2017**, *8*, No. 15679.

(25) Waroquiers, D.; George, J.; Horton, M.; Schenk, S.; Persson, K. A.; Rignanese, G.-M.; Gonze, X.; Hautier, G. ChemEnv: a fast and robust coordination environment identification tool. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2020**, *76*, 683−695.

(26) Waroquiers, D.; Gonze, X.; Rignanese, G.-M.; Welker-Nieuwoudt, C.; Rosowski, F.; Göbel, M.; Schenk, S.; Degelmann, P.; André, R.; Glaum, R.; Hautier, G. Statistical Analysis of Coordination Environments in Oxides. *Chem. Mater.* **2017**, *29*, 8346−8360.

(27) Zagorac, D.; Müller, H.; Ruehl, S.; Zagorac, J.; Rehme, S. Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *J. Appl. Crystallogr.* **2019**, *52*, 918−925.

(28) Delgado-Friedrichs, O.; Hyde, S. T.; Mun, S.-W.; O'Keeffe, M.; Proserpio, D. M. Nets with collisions (unstable nets) and crystal chemistry. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2013**, *69*, 535−542.

(29) O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Acc. Chem. Res.* **2008**, *41*, 1782−1789.

(30) Hyde, S.; Delgado Friedrichs, O.; Ramsden, S.; Robins, V. Towards enumeration of crystalline frameworks: the 2D hyperbolic approach. *Solid State Sci.* **2006**, *8*, 740−752.

(31) Ramsden, S. J.; Robins, V.; Hyde, S. T. Three-dimensional Euclidean nets from two-dimensional hyperbolic tilings: kaleidoscopic examples. *Acta Crystallogr., Sect. A:Found. Adv.* **2009**, *65*, 81−108.

(32) Baerlocher, C.; McCusker, L. B. Database of Zeolite Structures 2024 https://www.iza-structure.org/IZA-SC/ftc_table.php (accessed April 04, 2025).

(33) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. Quantum-mechanical condensed matter simulations with CRYSTAL. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1360.

(34) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100*, No. 136406.

(35) Peintinger, M. F.; Oliveira, D. V.; Bredow, T. Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations. *J. Comput. Chem.* **2013**, *34*, 451−459.

(36) Perger, W.; Criswell, J.; Civalleri, B.; Dovesi, R. Ab-initio calculation of elastic constants of crystalline systems with the CRYSTAL code. *Comput. Phys. Commun.* **2009**, *180*, 1753−1759.

(37) Erba, A.; Mahmoud, A.; Orlando, R.; Dovesi, R. Elastic properties of six silicate garnet end members from accurate ab initio simulations. *Phys. Chem. Miner.* **2014**, *41*, 151−160.

(38) Gaillac, R.; Pullumbi, P.; Coudert, F.-X. ELATE: an open-source online application for analysis and visualization of elastic tensors. *J. Phys.: Condens. Matter* **2016**, *28*, No. 275201.

(39) Huck, P.; Gunter, D.; Cholia, S.; Winston, D.; N'Diaye, A. T.; Persson, K. User applications driven by the community contribution framework MPContribs in the Materials Project. *Concurr. Comput. Pract. Exp.* **2016**, *28*, 1982−1993.

(40) Anderson, R.; Gómez-Gualdrón, D. A. Increasing topological diversity during computational "synthesis" of porous crystals: how and why. *CrystEngComm* **2019**, *21*, 1653−1665.

(41) Stein, F.; Leineweber, A. Laves phases: a review of their functional and structural applications and an improved fundamental understanding of stability and properties. *J. Mater. Sci.* **2021**, *56*, 5321−5427.

(42) Glasby, L. T.; Gubsch, K.; Bence, R.; Oktavian, R.; Isoko, K.; Moosavi, S. M.; Cordiner, J. L.; Cole, J. C.; Moghadam, P. Z. DigiMOF: A Database of Metal-Organic Framework Synthesis Information Generated via Text Mining. *Chem. Mater.* **2023**, *35*, 4510−4524.

(43) Burner, J.; Luo, J.; White, A.; Mirmiran, A.; Kwon, O.; Boyd, P. G.; Maley, S.; Gibaldi, M.; Simrod, S.; Ogden, V.; Woo, T. K. ARC-MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning. *Chem. Mater.* **2023**, *35*, 900−916.

(44) Wiegers, G.; Zhou, W. The misfit layer compound $(SnSe)_{1.16}NbSe_2$. *Mater. Res. Bull.* **1991**, *26*, 879−885.

(45) Ren, E.; Guilbaud, P.; Coudert, F.-X. High-throughput computational screening of nanoporous materials in targeted applications. *Digit. Discovery* **2022**, *1*, 355−374.

(46) Paddison, J. A. M.; Cliffe, M. J. Discovering Classical Spin Liquids by Topological Search of High Symmetry Nets. *ACS Cent. Sci.* **2024**, *10*, 1821−1828.

(47) Bousquet, D.; Coudert, F.-X.; Boutin, A. Free energy landscapes for the thermodynamic understanding of adsorption-induced deformations and structural transitions in porous materials. *J. Chem. Phys.* **2012**, *137*, No. 044118.

(48) Baise, M.; Maffettone, P. M.; Trousselet, F.; Funnell, N. P.; Coudert, F.-X.; Goodwin, A. L. Negative Hydration Expansion in ZrW_2O_8: Microscopic Mechanism, Spaghetti Dynamics, and Negative Thermal Expansion. *Phys. Rev. Lett.* **2018**, *120*, No. 265501.

(49) Banek, N. A.; Baiz, H. I.; Latigo, A.; Lind, C. Autohydration of Nanosized Cubic Zirconium Tungstate. *J. Am. Chem. Soc.* **2010**, *132*, 8278−8279.

(50) Gibaldi, M.; Kapeliukha, A.; White, A.; Luo, J.; Mayo, R. A.; Burner, J.; Woo, T. K. MOSAEC-DB: a comprehensive database of experimental metal−organic frameworks with verified chemical accuracy suitable for molecular simulations. *Chem. Sci.* **2025**, *16*, 4085−4100.

(51) Lee, S. J.; Doussot, C.; Baux, A.; Liu, L.; Jameson, G. B.; Richardson, C.; Pak, J. J.; Trousselet, F.; Coudert, F.-X.; Telfer, S. G. Multicomponent Metal−Organic Frameworks as Defect-Tolerant Materials. *Chem. Mater.* **2016**, *28*, 368−375.

(52) Coudert, F.-X.; Evans, J. D. Nanoscale metamaterials: Meta-MOFs and framework materials with anomalous behavior. *Coord. Chem. Rev.* **2019**, *388*, 48−62.