
Simulation Methods 1: Exploring Materials Database

Contents

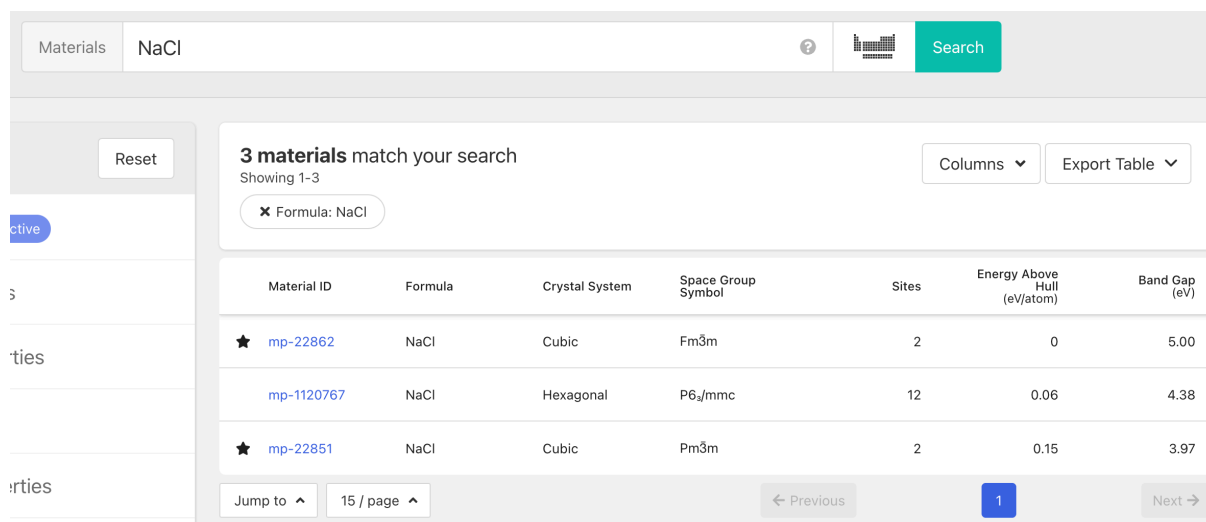
1. The Materials Project	2
1.1. Make a query through web	2
1.1.1. Energy above hull and thermodynamic stability	2
1.1.2. Crystal structure and symmetry	3
1.1.3. Wyckoff Positions	5
1.1.4. Band structure and band gap	7
2. Case study: CaTiO_3	8
3. Access database through API	9
3.1. Pull data from the material project	9
3.2. Band structure visualization with pymatgen	9
3.3. Matminer: An open source toolkit for materials data mining	9
4. Appendix	11
4.1. All 230 space groups	11
4.2. Colab notebook	13
Bibliography	14

1. The Materials Project

[The Materials Project](https://next-gen.materialsproject.org/materials) is a multi-institution, multi-national effort to compute the properties of all inorganic materials and provide the data and associated analysis algorithms for every materials researcher free of charge.

1.1. Make a query through web

One can make a query through the web interface: <https://next-gen.materialsproject.org/materials>



The screenshot shows the Materials Project web interface. At the top, there is a search bar with 'Materials' and 'NaCl' entered, and a 'Search' button. Below the search bar, a 'Reset' button is visible. The main content area displays '3 materials match your search' and 'Showing 1-3'. A filter 'Formula: NaCl' is applied. The results are shown in a table with columns: Material ID, Formula, Crystal System, Space Group Symbol, Sites, Energy Above Hull (eV/atom), and Band Gap (eV). The table lists three materials: mp-22862 (Cubic, Fm3m, 2 sites, 0 eV/atom, 5.00 eV band gap), mp-1120767 (Hexagonal, P6₃/mmc, 12 sites, 0.06 eV/atom, 4.38 eV band gap), and mp-22851 (Cubic, Pm3m, 2 sites, 0.15 eV/atom, 3.97 eV band gap). The first and third materials are marked with a star (★). At the bottom, there is a pagination bar showing 'Jump to', '15 / page', and navigation buttons for 'Previous', '1', and 'Next'.

Material ID	Formula	Crystal System	Space Group Symbol	Sites	Energy Above Hull (eV/atom)	Band Gap (eV)
★ mp-22862	NaCl	Cubic	Fm3m	2	0	5.00
mp-1120767	NaCl	Hexagonal	P6 ₃ /mmc	12	0.06	4.38
★ mp-22851	NaCl	Cubic	Pm3m	2	0.15	3.97

Figure 1: The Materials Project

- The star ★ indicates that the material is experimentally observed.
- Material ID: can be useful for accessing the database through the API.
- Crystal System: The crystal system of the material.
- Spacegroup: The space group of the material.
- Sites: The number of atoms in the unit cell.
- Energy Above Hull: The energy above the hull is the minimum energy required to remove an electron from the material.
- Band Gap: The band gap is the energy difference between the valence band and the conduction band.

1.1.1. Energy above hull and thermodynamic stability

The formation energy of a material defines the energy required to dissociate that material into its respective elements. It indicates the thermodynamic stability (room temperature ≈ 0.025 eV) against the elements it consists of. Usually, the density functional theory (DFT) based first-principles calculations are performed to calculate the formation energy. Formation energy indicates only whether the material is stable with respect to the constituent elements. However, the main enemy of the material is other competing phases rather than the constituent elements. The energy above the hull is a measure of the thermodynamic stability of the material against other competing phases, which measures how likely the material is to decompose to other phases. The larger the energy above the hull, the more likely the material is to decompose to other phases.

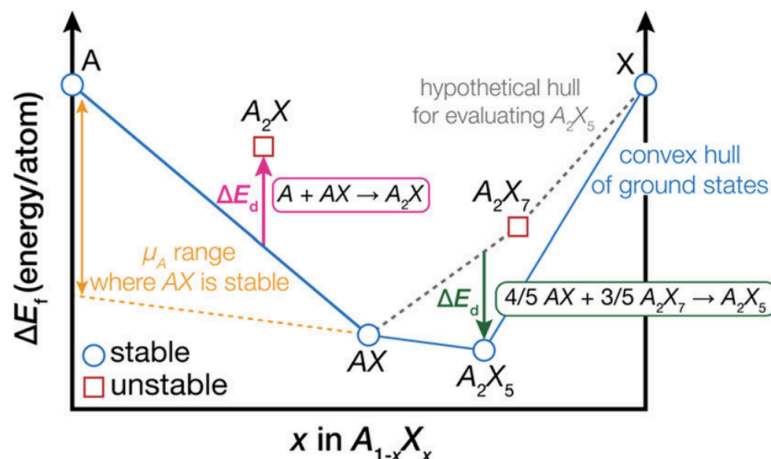


Figure 2: Energy above the convex hull. (Image source: [1])

As illustrated in Figure 2, for a formula of $A_{1-x}B_x$, multiple compounds can be formed. By plotting the formation energy as a function of x , we can find the convex hull. The energy above the hull is the difference between the formation energy and the convex hull. If a material is thermodynamically stable, the energy above the hull must be equal to zero. For example, in the plot, both A and AX are thermodynamically stable. Then the compound A_2X is thermodynamically unstable if the energy above the hull ΔE_d is larger than zero, because by decomposing A_2X into A and AX, the energy gain is larger than the energy cost by ΔE_d . On the other hand, it does not mean that A_2X cannot be synthesized. To decompose A_2X into A and AX, energy barrier must be overcome, which may be a very slow process.

1.1.2. Crystal structure and symmetry

The space group description of the salt crystal structure of NaCl (mp-22862) is:

- Crystal System: Cubic
- Lattice System: Cubic
- Hall Number: -F 4 2 3
- International Number: 225
- Symbol: $Fm\bar{3}m$
- Point Group: $m\bar{3}m$

Both the crystal system and lattice system are cubic. By checking the following table, we know the defining symmetry of cubic crystal system is the 4 three fold axes of rotation.

Crystal system	Symmetry	Point groups	Lattice system	Space groups
Triclinic		2	Triclinic	2
Monoclinic	1 twofold axis of rotation or 1 mirror plane	3	Monoclinic	13
Orthorhombic	3 twofold axes of rotation or 1 twofold axis of rotation and 2 mirror planes	3	Orthorhombic	59
Tetragonal	1 fourfold axis of rotation	7	Tetragonal	68
Trigonal	1 threefold axis of rotation	5	Rhombohedral	7
			Hexagonal	18
Hexagonal	1 sixfold axis of rotation	7		27
Cubic	4 threefold axes of rotation	5	Cubic	36
7	Total	32	7	230

Table 1: Crystal system, lattice system and space groups

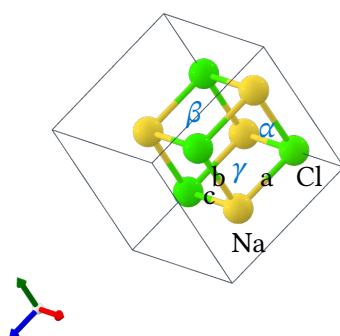
The international number 225, Hall number -F 4 2 3, and Symbol $Fm\bar{3}m$ are different ways to represent the same space group. By checking its space group $Fm\bar{3}m$ (Source: <https://www.globalsino.com/EM/page3025.html>), the 48 symmetry group operations can be resolved as follows:

x, y, z	$-x, -y, z$	$-x, y, -z$	$x, -y, -z$	z, x, y
$z, -x, -y$	$-z, -x, y$	$-z, x, -y$	y, z, x	$-y, z, -x$
$y, -z, -x$	$-y, -z, x$	$y, x, -z$	$-y, -x, -z$	$y, -x, z$
$-y, x, z$	$x, z, -y$	$-x, z, y$	$-x, -z, -y$	$x, -z, y$
$z, y, -x$	$z, -y, x$	$-z, y, x$	$-z, -y, -x$	$-x, -y, -z$
$x, y, -z$	$x, -y, z$	$-x, y, z$	$-z, -x, -y$	$z, x, -y$
$z, -x, y$	$-z, x, y$	$-y, -z, -x$	$y, z, -x$	$y, -z, x$
$-y, z, x$	$-y, x, -z$	$y, -x, -z$	y, x, z	$-y, -x, z$
$x, z, -y$	$-x, z, y$	$-x, -z, -y$	$x, -z, y$	$z, y, -x$
$z, -y, x$	$-z, y, x$	$-z, -y, -x$	$-x, -y, -z$	$x, y, -z$
$x, -y, z$	$-x, y, z$	$-z, -x, -y$	$z, x, -y$	$z, -x, y$
$-z, x, y$	$-y, -z, -x$	$y, z, -x$	$y, -z, x$	$-y, z, x$
$-y, x, -z$	$y, -x, -z$	y, x, z	$-y, -x, z$	$x, z, -y$
$-x, z, y$	$-x, -z, -y$	$x, -z, y$	$z, y, -x$	$z, -y, x$
$-z, y, x$	$-z, -y, -x$	$-y, \frac{1}{2} + x, \frac{1}{2} + z$	$x, \frac{1}{2} + z, \frac{1}{2} - y$	$-x, \frac{1}{2} + z, \frac{1}{2} + y$
$-x, \frac{1}{2} - z, \frac{1}{2} - y$	$x, \frac{1}{2} - z, \frac{1}{2} + y$	$z, \frac{1}{2} + y, \frac{1}{2} - x$	$z, \frac{1}{2} - y, \frac{1}{2} + x$	$-z, \frac{1}{2} + y, \frac{1}{2} + x$
$-z, \frac{1}{2} - y, \frac{1}{2} - x$	$-x, \frac{1}{2} - y, \frac{1}{2} - z$	$x, \frac{1}{2} - y, \frac{1}{2} - z$	$x, \frac{1}{2} + y, \frac{1}{2} + z$	$-z, -x, -y$
$z, x, -y$	$z, -x, y$	$-z, x, y$		

Each term in the table represents a symmetry operation. For example, the entry $-x, z, y$ means the atom at x, y, z is transformed to $-x, z, y$. A crystal possesses this symmetry means that if we can find an atom at x, y, z , then we can also find an atom at $-x, z, y$.

Together with the Wyckoff positions, we can visualize the crystal structure.

Wyckoff site	Element	x	y	z
4a	Na	0	0	0
4b	Cl	0	0	1/2



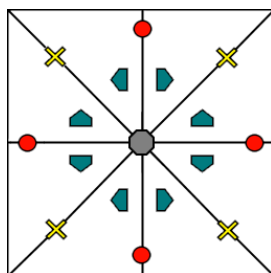
Lattice Parameters	Values
a	5.59 Å
b	5.59 Å
c	5.59 Å
α	90.00 °
β	90.00 °
γ	90.00 °
Volume	174.50 Å ³

Table 2: The crystal structure of NaCl. The lattice vectors are a, b, c , and the angles are $\alpha = \angle(b, c)$, $\beta = \angle(a, c)$, $\gamma = \angle(a, b)$. (Image source: The Materials Project)

1.1.3. Wyckoff Positions

- Ref: <https://pyxtal.readthedocs.io/en/latest/Background.html>

To complete defining a crystal structure, a final piece is required, which is the Wyckoff positions of different species of atoms. Because symmetry operations can be thought of as making copies of parts of an object, we can usually only describe part of a structure, and let symmetry generate the rest. This small part of the structure used to generate the rest is called the asymmetric unit. However, not all points in the asymmetric unit are generated the same. If an atom lies within certain regions - planes, lines, or points - then the atom may not be "copied" as many times as other atoms within the asymmetric unit. A familiar example is in the creation of a paper snowflake. We start with a hexagon, then fold it into a single triangle 6 sheets thick. Then, if we cut out a mark somewhere in the middle of the triangle, the mark is copied 6-fold. However, if we instead cut out a mark along the triangle's edge, or at the tip, the marks will only have 3 or 1 copies:



These different regions are called Wyckoff positions, and the number of copies is called the multiplicity of the Wyckoff position. So, if an atom lies in a Wyckoff position with multiplicity greater than 1, then

that Wyckoff position actually corresponds to multiple atoms. However, thanks to symmetry, we can refer to all of the copies (for that particular atom) as a single Wyckoff position. This makes describing a crystal much easier, since we no longer need to specify the exact location of most of the atoms. Instead, we need only list the space group, the lattice, and the location and type of one atom from each Wyckoff position. Just keep in mind that in this format, a single atomic entry may correspond to multiple atoms in the unit cell.

The largest Wyckoff position, which makes a copy for every symmetry operation, is called the general Wyckoff position, or just the *general position*. In the snowflake example, this was the large inner region of the triangle. In general, the general position will consist of every location which does not lie along some special symmetry axis, plane, or point. For this reason, the other Wyckoff positions are called the *special Wyckoff positions*.

The number and type of Wyckoff positions are different for every space group; a list of them can be found using the Bilbao utility WYCKPOS ([link](#)). In the utility, (TODO: point to the resource)

Wyckoff positions are described using the x, y, z notation, where each operation shows how the original (x, y, z) point is transformed/copied. In other words, if we choose a single set of coordinates, then plugging these coordinates into the Wyckoff position will generate the remaining coordinates. As an example, consider the general position of space group P222 (#16), which consists of the points (x, y, z) , $(-x, -y, z)$, $(-x, y, -z)$, and $(x, -y, -z)$. If we choose a random point, say $(0.321, 0.457, 0.892)$, we can determine the remaining points:

$$\begin{aligned}(x, y, z) &\rightarrow (0.321, 0.457, 0.892) \\ (-x, -y, z) &\rightarrow (0.679, 0.543, 0.892) \\ (-x, y, -z) &\rightarrow (0.679, 0.457, 0.108) \\ (x, -y, -z) &\rightarrow (0.321, 0.543, 0.108)\end{aligned}$$

Here a negative value is equal to 1 minus that value ($-0.321 = 1 - 0.321 = 0.679$).

To denote Wyckoff positions, a combination of number and letter is used. The number gives the multiplicity of the Wyckoff position, while the letter differentiates between positions with the same multiplicity. The letter 'a' is always given to the smallest Wyckoff position (usually located at the origin or z axis), and the letter increases for positions with higher multiplicity. So, for example, the space group I4mm (#107) has 5 different Wyckoff positions: 2a, 4b, 8c, 8d, and 16e. Here, 16e is the general position, since it has the largest multiplicity and last letter alphabetically.

Note that for space groups with non-simple lattices (those which begin with a letter other than 'P'), the Wyckoff positions also contain fractional translations. Take for example the space group I4mm (#107). The Bilbao entry can be found at the [url](#). Each listed Wyckoff position coordinate has a copy which is translated by $(0.5, 0.5, 0.5)$. It is inconvenient to list each of these translated copies for every Wyckoff position, so instead a note is placed at the top. This is why Wyckoff position 16e has only 8 points listed. In this case, to generate the full crystal, one could apply the 8 operations listed, then make a copy of the resulting structure by translating it by the vector $(0.5, 0.5, 0.5)$. Note that in space groups beginning with letters other than P, the smallest Wyckoff position will never have a multiplicity of 1.

In addition to the generating operations, the site symmetry of each Wyckoff position is listed. The site symmetry is just the point group which leaves the Wyckoff position invariant. So, if a Wyckoff position consists of an axis, then the site symmetry might be a rotation about that axis. The general position

always has site symmetry 1, since it corresponds to choosing any arbitrary structure or location can be made symmetrical by copying it and applying all of the operations in the space group.

Finally, since crystals are infinitely periodic, a Wyckoff position refers not only to the atoms inside a unit cell, but every periodic copy of those atoms in the other unit cells. Thus, the Wyckoff position (x, y, z) is the same as the position $(x + 1, y + 1, z)$, and so on. This is usually a minor detail, but it must be taken into account for certain computational tasks.

1.1.4. Band structure and band gap

Quantum many-body dynamics is generally intractable, making materials simulation a challenging task. However, in many cases, the electrons are not so strongly correlated, and we can treat them as single particles. The rest part, including the nuclei and the other electrons, is treated as a potential $V(\mathbf{r})$. We are quite familiar with solving the Schrödinger equation for a single particle, since we already encountered it in the context of the atomic structure. In the atomic case, the potential is the **spherical symmetric** Coulomb potential, and the single-particle wavefunction is the atomic orbital labeled by the quantum numbers n, l, m . In the solid case, the potential is the **translational invariant potential**, and the single-particle wavefunction is the Bloch wavefunction labeled by the crystal momentum (or wave vector) \mathbf{k} .

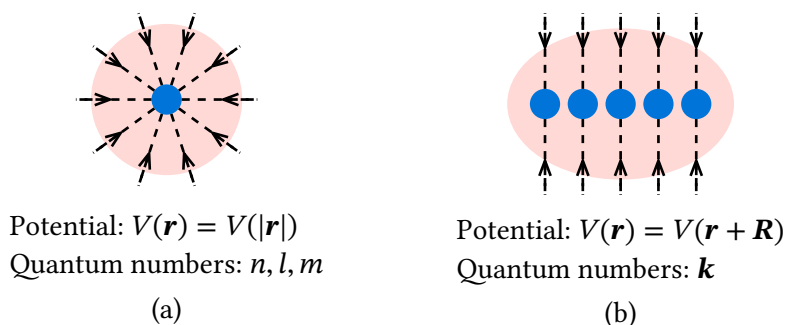


Figure 4: (a) Electron cloud (red) around an spherical symmetry (b) Electron cloud (red) around a translational invariant lattice potential, where \mathbf{R} is a lattice vector.

\mathbf{k} has discrete values, just like n, l, m have discrete values. They live in the **dual lattice** of the crystal, i.e. have a periodic structure. The unit cell of the dual lattice is called the **Brillouin zone**. Unsurprisingly, the Brillouin zone has the same point group symmetry as its real space lattice. By walking through high symmetry points (e.g. Γ, X, W, K, L below), we get the band structure. Although the band looks continuous, it is actually a series of many discrete energy levels. By counting the number of energy levels at each energy value, we get the density of states.

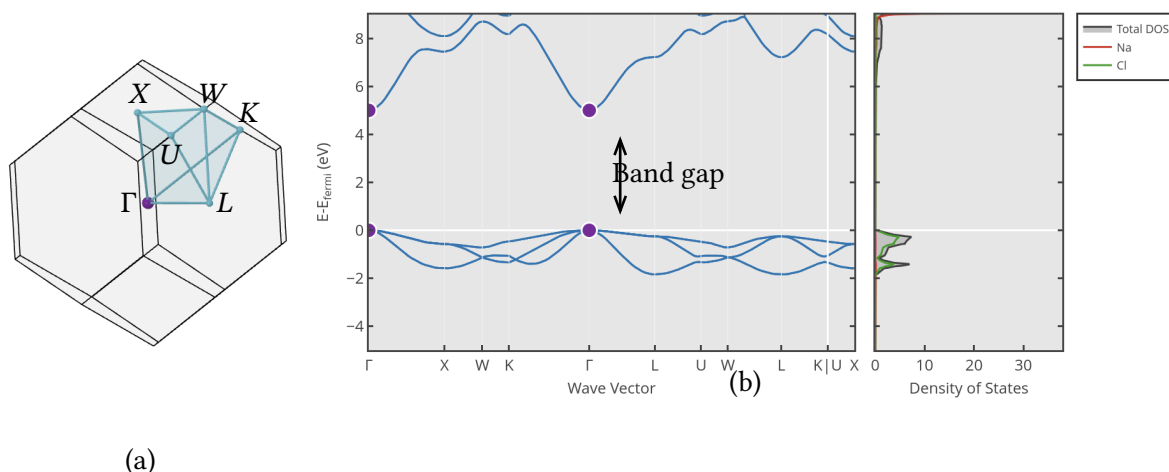
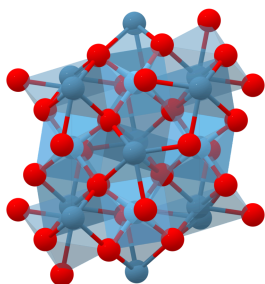


Figure 5: The electronic properties of NaCl. (a) The Brillouin zone. (b) The band structure and density of states.

2. Case study: CaTiO_3

Calcium titanate (CaTiO_3) is Orthorhombic Perovskite structured. The material ID is mp-4019: <https://next-gen.materialsproject.org/materials/mp-4019>



The crystal structure of CaTiO_3 .

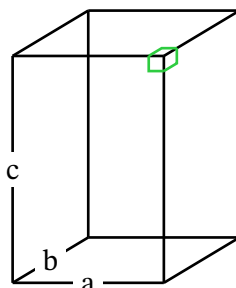
	Property	Value
Meta data	Material ID	mp-4019
	Experimentally Observed	Yes
Energy	Energy Above Hull	0.000 eV/atom
	Band Gap	2.31 eV
	Predicted Formation Energy	-3.453 eV/atom
Magnetic property	Magnetic Ordering	Non-magnetic
	Total Magnetization	0.00 $\mu\text{B}/\text{f.u.}$
Physical properties	Number of Atoms	20
	Density	4.03 $\text{g}\cdot\text{cm}^{-3}$
	Dimensionality	3D
	Possible Oxidation States	Ca^{2+} , Ti^{4+} , O^{2-}

The space group description of CaTiO_3 is:

- Crystal System: Orthorhombic
- Lattice System: Orthorhombic
- Hall Number: -P 2ac 2n
- International Number: 62
- Symbol: Pnma
- Point Group: mmm

The international number, Hall number, and Symbol are different ways to represent the same space group. With the following information, we can visualize the crystal structure.

Lattice Parameters	Values
a	5.37Å
b	5.46Å
c	7.64Å
α	90.00°
β	90.00°
γ	90.00°



Wyckoff site	Element	x	y	z
4b	Ti	0	1/2	1/2
4c	Ca	0.509067	0.542398	3/4
4c	O	0.076781	0.481328	3/4
8d	O	0.790294	0.789614	0.959841

Please refer to [Crystal morphology and symmetry](#) for more details.

3. Access database through API

3.1. Pull data from the material project

1. Go to website: <https://next-gen.materialsproject.org/>, Click “Login or Register”.
2. Click “API” in the top right corner, or just check [this link](#). Move to the section “API Key” and you will see your API key.
3. To use the Materials Project API, you need to install the mp-api package:

```
pip install mp-api
```


Please make sure you have a python version ≥ 3.11 (helpful links: [Miniconda](#)).
4. Then you can use the following Python code to retrieve a list of materials:

3.2. Band structure visualization with pymatgen

[Pymatgen \(Python Materials Genomics\)](#) is a robust, open-source Python library for materials analysis. It is based on the Materials Project database, and its has already been installed when you install the mp-api package.

3.3. Matminer: An open source toolkit for materials data mining

Four commonly used materials databases in matminer[2] are:

1. [Citrination](#)[3].
2. [The Materials Project \(MP\)](#)[4]
3. [The Materials Data Facility \(MDF\)](#)

4. [The Materials Platform for Data Science \(MPDS\)](#)
5. [MongoDB](#)

4. Appendix

4.1. All 230 space groups

Source: <http://pd.chem.ucl.ac.uk/pdnn/symm3/allsgp.htm> and <http://img.chem.ucl.ac.uk/sgp/large/sgp.htm>

Crystal System	Laue Class	Crystal Class	Lattice Centering	230 3D Space Groups
Triclinic	-1	1	P	P1
		-1		P-1
Monoclinic	2/m	2	P	P2, P2 ₁
			C	C2
		m	P	Pm, Pc
			C	Cm, Cc
		2/m	P	P2/m, P2 ₁ m, P2/c, P2 ₁ c
			C	C2/m, C2/c
Orthorhombic	mmm	222	P	P222, P222 ₁ , P2 ₁ 2 ₁ 2, P2 ₁ 2 ₁ 2 ₁
			C	C222, C222 ₁
			F	F222
			I	I222, I2 ₁ 2 ₁ 2 ₁
		mm2	P	Pmm2, Pmc2 ₁ , Pcc2, Pma2, Pca2 ₁ , Pnc2, Pmn2 ₁ , Pba2, Pna2 ₁ , Pnn2
			C or A	Cmm2, Cmc2 ₁ , Ccc2, Amm2, Abm2, Ama2, Aba2
			F	Fmm2, Fdd2
			I	Imm2, Iba2, Ima2
		mmm	P	Pmmm, Pnnn, Pccm, Pban, Pmma, Pnna, Pmna, Pcca, Pbam, Pccn, Pbcm, Pnnm, Pmmn, Pbcn, Pbca, Pnma
			C	Cmmm, Cmcm, Cmca, Cccm, Cmma, Ccca
			F	Fmmm, Fddd
			I	Immm, Ibam, Ibcm, Imma
Tetragonal	4/m	4	P	P4, P4 ₁ , P4 ₂ , P4 ₃
			I	I4, I4 ₁
		-4	P	P-4
			I	I-4
		4/m	P	P4/m, P4 ₂ /m, P4/n, P4 ₂ /n
			I	I4/m, I4 ₁ /a
	4/mmm	422	P	P422, P42 ₁ 2, P4 ₁ 22, P4 ₁ 2 ₁ 2, P4 ₂ 22, P4 ₂ 2 ₁ 2, P4 ₃ 22, P4 ₃ 2 ₁ 2
			I	I422, I42 ₁ 2

Crystal System	Laue Class	Crystal Class	Lattice Centering	230 3D Space Groups
		4mm	P	P4mm, P4bm, P4 ₂ cm, P4 ₂ nm, P4cc, P4nc, P4 ₂ mc, P4 ₂ bc
			I	I4mm, I4cm, I4 ₁ md, I4 ₁ cd
		-42m	P	P-42m, P-42c, P-42 ₁ m, P-42 ₁ c
			I	I-42m, I-42d
		-4m2	P	P-4m2, P-4c2, P-4b2, P-4n2
			I	I-4m2, I-4c2
		4/mmm	P	P4/mmm, P4/mcc, P4/nbm, P4/nnc, P4/mbm, P4/mnc, P4/nmm, P4/ncc, P4 ₂ /mmc, P4 ₂ /mcm, P4 ₂ /nbc, P4 ₂ /nnm, P4 ₂ /mbc, P4 ₂ /mcm, P4 ₂ /nmc, P4 ₂ /ncm
			I	I4/mmm, I4/mcm, I4 ₁ /amd, I4 ₁ /acd
Trigonal	-3	3	P	P3, P3 ₁ , P3 ₂
			R	R3
		-3	P	P-3
			R	R-3
	-3m	312	P	P312, P3 ₁ 12, P3 ₂ 12
				P321, P3 ₁ 21, P3 ₂ 21
		321	R	R32
		31m	P	P31m, P31c
				P3m1, P3c1
		3m1	R	R3m, R3c
		-31m	P	P-31m, P-31c
				P-3m1, P-3c1
		-3m1	R	R-3m, R-3c
Hexagonal	6/m	6	P	P6, P6 ₁ , P6 ₂ , P6 ₃ , P6 ₄ , P6 ₅
		-6		P-6
		6/m		P6/m, P6 ₃ /m
	6/mmm	622		P622, P6122, P6222, P6322, P6422, P6522
		6mm		P6mm, P6cc, P6 ₃ cm, P6 ₃ mc
		-6m2		P-6m2, P-6c2
		-62m		P-62m, P62c
		6/mmm		P6/mmm, P6/mcc, P6 ₃ /mcm, P6 ₃ /mmc
Cubic	m-3	23	P	P23, P2 ₁ 3
			F	F23
			I	I23, I2 ₁ 3

Crystal System	Laue Class	Crystal Class	Lattice Centering	230 3D Space Groups
		m-3	P	Pm-3, Pn-3, Pa-3
			F	Fm-3, Fd-3
			I	Im-3, Ia-3
	m-3m	432	P	P432, P4 ₂ 32, P4 ₃ 32, P4 ₁ 32
			F	F432, F4 ₁ 32
			I	I432, I4 ₁ 32
		-43m	P	P-43m, P-43n
			F	F-43m, F-43c
			I	I-43m, I-43d
		m-3m	P	Pm-3m, Pn-3n, Pm-3n, Pn-3m
			F	Fm-3m, Fm-3c, Fd-3m, Fd-3c
			I	Im-3m, Ia-3d

4.2. Colab notebook

https://colab.research.google.com/drive/1d_JOp2sTzBds4EAycePP-UDhapQQpIr9?usp=sharing

Bibliography

- [1] C. J. Bartel, “Review of computational approaches to predict the thermodynamic stability of inorganic solids,” *Journal of Materials Science*, vol. 57, no. 23, pp. 10475–10498, 2022.
- [2] L. Ward *et al.*, “Matminer: An Open Source Toolkit for Materials Data Mining,” *Computational Materials Science*, vol. 152, pp. 60–69, Sep. 2018, doi: [10.1016/j.commatsci.2018.05.018](https://doi.org/10.1016/j.commatsci.2018.05.018).
- [3] K. Michel and B. Meredig, “Beyond bulk single crystals: a data format for all materials structure–property–processing relationships,” *Mrs Bulletin*, vol. 41, pp. 617–623, 2016.
- [4] A. Jain *et al.*, “Commentary: The Materials Project: A materials genome approach to accelerating materials innovation,” *APL materials*, vol. 1, no. 1, 2013.