Data Analysis of physicochemical properties based on molecular structure of ABO3 perovskites

Jesús Camilo, Díaz Olivella

Coderhouse
Data Science Course, comission 42370
Medellín, Colombia
E-mail: jesusc.diaz@udea.edu.co

27th april, 2023

Contents

I Introduction

Renewable energies are those that have a replenishment cycle faster than consumption. Technologies based on taking advantage of sunlight present a particular interest because the sun provides an abundant source. While most materials are capable of absorbing solar energy, it is a challenge to maximize the efficiency with which they do so while also preventing it from dissipating into the environment, thus resulting in it being unusable.[1] Many materials have been proposed to harness solar energy, among them are perovskites, which are a cheaper alternative to other traditional semiconductors.[2] Due to the low cost of perovskites and their competent current efficiency, they become a high cost/benefit material and therefore a very good candidate for the massive implementation of photovoltaic devices.[3] Perovskite solid-state solar cells (PSCs) are exposed to high temperatures, constant sunlight, and humidity. These conditions are ideal for them to suffer corrosion, which greatly limits their efficiency in devices.[4] Environmental and economic interests have promoted the realization of many investigations to make perovskites more stable. A characteristic found that is of paramount importance is the geometric arrangement and the type of chemical element that the atoms defining the solids of the perovskite family adopt for their stability under the mentioned conditions.[5]

II First delivery

1 Description of the data following the SMART methodology

2 Specific:

The database contains information on the stability of different perovskites according to the combination of 73 different elements with different spatial arrangements, for a total of 1930 rows of measurements.

3 Measurable:

The database contains columns with physicochemical variables that measure stability (resistance to reacting with other substances or decomposing with itself), thus making it possible to analyze how the relationships of atom position and type of chemical element influence their stability variables. Therefore, it will be very interesting to measure which class of atoms and in what position gives the greatest resistance to the first ionization energy for all perovskites (a factor that influences chemical reactivity, which is desired to be avoided).

4 Achievable:

We aim to study data that has already been collected, so no further experiments are required. The hypothesis is defined solely based on properties obtainable with the provided information.

5 Relevant:

It is an active field of study that presents great economic interests, as it belongs to the renewable energy sector. The analysis could be used subsequently for product development.

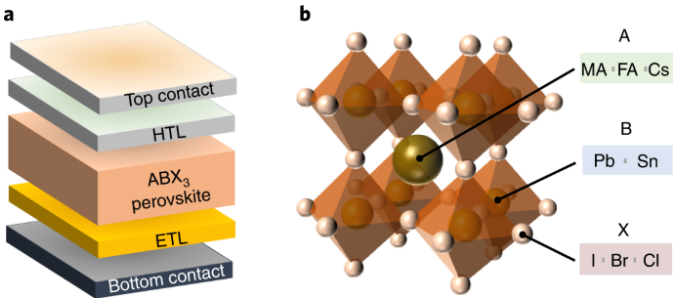

Figure 1. The figure shows: a) A perovskite solar panel diagram. b) A crystallographic arrangement of a perovskite of type ABX3, which will be studied in the project. The database contains many perovskites with this form, but with different ratios and different atoms in different positions, generating the 1930 rows of different arrangements. The image source is cited as: https://doi.org/10.1038/s41560-022-01096-5.

6 Time:

During the course, appropriate tools will be used to analyze the provided information, as well as the guidance of the tutors at Coderhouse. The necessary chemical knowledge to understand the variables in the columns has already been achieved, and adequate guidance is also available from a teaching professor at the University of Antioquia. This makes it possible to achieve the objectives within the course timeline.

7 Hypothesis

There is a relationship between positions and the types of atoms in perovskites that is related to physicochemical properties.

III Second Delivery

1 Description

Two databases will be used, one with structural properties and the other with physicochemical properties. In order for the bases to have a clear relationship, a filter was made where only the structures that are repeated in both tables were left. It will be systematically analyzed how atomic positions and chemical elements influence the different stability variables, for which all perovskites are classified in similar geometries with atoms. Thus, it will be possible to make correlation graphs with possible physical-chemical causalities. In the end, it will be known which perovskites present better stability properties based on their crystallographic arrangements.

2 Objective of the project

The objective of the project is to relate the geometric positions of the perovskite atoms with their physicochemical properties, seeking to find trends in why some are more stable than others. This will help to develop useful scientific knowledge for the implementation of perovskites as solar energy cells.

3 Scope

The project will cover a detailed analysis of structural (geometric) parameters with their relationship of physicochemical properties, thus delivering different graphs of the highest correlations between microscopic (structure) and macroscopic (thermodynamic properties) structure.

4 End User

The end user will be the scientific and engineering community, in this case it is expected that the users with the greatest benefit will be those who work in the area of materials science with development of photovoltaic cells.

5 Application level

The project is expected to have an operational scope, allowing the creation of new product designs by establishing conceptual tools that allow decision-making in the development and research sector. It could also reach a tactical level, by emphasizing the importance of basic science as a creator of new knowledge, which, although it may not seem useful at first, is an investment towards innovation and development.

6 List of tables

Four tables will be used, which were processed from two much larger tables with 1900 and 4000 rounds, the idea of the processing was to unify that both tables had a relationship between them, for this purpose SQL and Python were used for processing. In the end it was found optimal to divide the tables into 4.

1. A_site table 1: Contains the physicochemical information of the chemical elements that are located in site A, its PK is column 'A' which is a chemical element, this table has a cardinality of N: 1 with the table 3, its connection is then that the table 3 has many compounds (unions of A and B in different proportions).
    - PK:A
2. B_site table 2: It is analogous to A_site in that they are the chemical elements for position B, their PK being column 'B', but the physicochemical properties of interest are quite different being chemical elements of quite different groups, being A_site mostly metals and B_site mostly halogens. It can then be said that having few similar columns it was not advantageous to join it to A_site.
    - PK: B
3. Crystal_structure table 3: It is the central table, it relates the other tables, it contains most of the relevant geometric parameters, angles, distances and volume relationships necessary to assemble the assembly of the figure **??**. The PK is 'Compound' which is a reduced stoichiometric ratio, that is, they are all of the ABO3 form, unlike the table 4 where we find materials that have the same reduced form, but may have different geometric arrangements. , can be thought of as lego blocks, the table Crystal_structure contains the lego blocks while Material_properties contains the figures assembled with the lego blocks. Its relation to A and B is from FK, this table having its own columns A and B.
    - PK: Compound
    - FK:A
    - FK: B
4. Material_properties table 4: contains the physicochemical information of the material properties, that is, the properties of the macroscopic system, its PK is 'Material Composition' and its FK that connects it with Crystal_structure is 'compound'.

- PK: Material Composition
- FK: Compound

A Entity Relationship Diagram (EDR)

To facilitate the description given above, the EDR is appended at the end of the document in figure 2.

IV Third delivery

1 Data relationship and Cardinality

The tables are using a model that have some resemblance about the way we study mater by size, we say that atoms form the crystal lattice (the minimum arrangement of atoms needed to understand a crystal) and if we made and arrangement of one lattice together to other lattice and expand it to a physics infinity the crystal lattice form the structure as we know as crystals. In this sense the tables represent Atoms, lattice and crystals. Table 1 and table 2 are atoms in the sites A and B, these sites are connected to 3 (the lattice) which is the lattice, that's why table 3 is the table that work as the bridge between 4 and the Atoms A and B. These are showed in the figure 2.

2 Date tables

There were not used in these work, my databases don't have any important relationship that use time.

3 Calculated Measures

Many of these calculations were done in python previously, for these reason a measure table is not presented. Only the Goldschmidt tolerance factor was calculated in Power BI. The python Jupyter notebook would be found in the folder of the project.

- Formation energy is calculated by the formula

$$E_{\text{form}} = E_{\text{tot}} - \sum_{x} E_{\text{tot}(x)}$$

  the total energy is the energy of all the system, and the i subscript is all the species in his most common form in nature.

- All average AB columns like density_avg column in table 4 were calculated by and average formula:

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i$$

- The cohesive energy is calculated by the formula:

$$E_{\text{form}}^{\text{A}} = E_{\text{tot}}^{\text{A}} - E_{\text{tot}}^{\text{A}}/n_{\text{A}} - E_{\text{tot}}^{\text{As}}/n_{\text{B}}$$

  where n_A and n_{n_B} are the number of atoms in the bulk unit cell.

- The Goldschmidt tolerance factor is calculated by the formula:

$$t' = \frac{r_A + r_X}{\sqrt{2}\,(r_B + r_X)} \times \frac{\Delta(X-A)}{\Delta(O-A)}$$

  with rA, rB and rX being the ionic radii of A and B cations and X anions respectively. $\delta X(X-A)$ is the electronegativity difference between X and A, and $\delta X(O-A)$ is the electronegativity differences between O and A.[6]

4 Explaining the process

The idea of pibx presentation is to show the relationship about the properties of the electronic structure (at molecular level) with macro physicochemical properties at the crystal level. The pibx project is divided in four pages as was asked in the instructions.

A Title page

The idea of the title page is a quick inside in the project, a giant graph is presented in the center, showing the energy above hull which is the most important value of the stability. The formation energy is also present, this parameter is indirectly related to the stability but hot related with the structure change at molecular level. My intention in this graph is to show the extension of the project, showing the total number of perovskites in the database and the total elements that conform sites A and sites B.

B Microscopic properties

In this page, I wanted to exploit the amazing relationship capabilities of the tables in Power Bi and trying to accomplish a little immersive experience, presenting the microscopic data of all the compounds with a deformation tree, I also have a version with buttons instead of tree, this approach was more intuitive but was not as stunning and conceptually clear.

C Macroscopic properties

In this page all is about the compounds and important properties. In these are shown four average properties that are calculated, the first ones are a classic Ionization potential and Electron affinity, this show the possible chemical reactivity, if both are high is less probable to react then is more stable. The second graph show transport properties, these are not directly related to stability but are really useful to know two important characteristics of the perovskites, the first important characteristic is thermalconductivity. For photovoltaic applications, it is generally better to have low thermal conductivity perovskites. This is because in a solar cell, the aim is to convert as much of the incident solar radiation as possible into electrical energy. Any heat generated during this process is essentially wasted energy that is not contributing to the output of the cell. Electron conductivity is another important property of perovskites for photovoltaic applications, as it determines how effectively electrons can move through the material and contribute to the output of the solar cell.

In general, high electron conductivity is desirable for perovskites used in photovoltaic devices, as this enables more efficient charge transport and reduces losses due to resistance.

Therefore, in order to achieve high efficiency in perovskite solar cells, it is important to:

1. Optimize the balance between absorbing as much solar radiation as possible and minimizing any heat loss.

2. optimizing electron conductivity but balancing with band gap and hole defect formation.

Therefore, achieving high efficiency in perovskite solar cells requires a careful optimization of multiple properties, including electron conductivity, in order to maximize the overall performance of the device.

D Micro and Macro relationship

This page is all about the conclusions and aims of the project, it shows how the most important stability parameter change which is energy above hull when we are alternating along the chemical element in site A and site B (avg change, this is a many-to-one relationship). There are also showed the same properties mentioned in the Macroscopic properties section, but instead of showing by compound there is also show the avg per element, in these sense here I try to achieve the connection between the page 2 and page 3.

V Final Delivery

1 New navigation button and and new filter implementation

For the last delivery minor but very useful aditions were added, the idea was to be able to quickly navigate between pages.The second addition was 3 more filters, this filters made more the presentation more interactive:

- First filter is a position A atom filter for the Macroscopic properties, adding this filter we are able to see the behavior in a specif family of perovskites.
- The tree in the Macroscopic properties was not clear that was a filter, adding a explicit title who says to click in the bars would improve the user to understand the function.
- In the Micro-Macro relationship and ideal target slider were added to be able to bin the data intro desired categories, for this reason a new column were created using an equal weight of the stability variables, the variables were normalized so the perfect value of each variable would be cero.

Acknowledgment

The author thanks the Professor Jorge Mario Osorio Guillen and the PhD Juan Camilo Alvarez Quiceno, for the advice provided for the definition of the problem. We thank the tutor Flor de coderhouse for the recommendation of the glossary to facilitate reading. I also thanks Iara for the assessor in all the course.

References

[1] P. Zhao, B. J. Kim, and H. S. Jung, "Passivation in perovskite solar cells: A review," *Materials Today Energy*, vol. 7, pp. 267–286, 3 2018.

[2] N. G. Park, "Perovskite solar cells: An emerging photovoltaic technology," *Materials Today*, vol. 18, pp. 65–72, 3 2015.

[3] D. Arvizu, P. B. Singapore, L. F. Cabeza, K. T. Hollands, A. J.-W. Italy, M. Kondo, C. Konseibo, V. Meleshko, W. Stein, Y. Tamaura, H. Xu, and R. Zilles, "3 direct solar energy coordinating lead authors: Lead authors," 2010.

[4] Y. Yu, R. Liu, M. Xu, and H. Yu, "Diammonium-induced dion-jacobson 2d/3d wide-bandgap perovskite solar cells with enhanced efficiency and stability," *EcoMat*, 2022.

[5] L. Zhang, G. Qi, Y. Zhang, H. Wu, X. Xu, G. Zhou, H. Zhu, X. Li, G. Wu, and H. Chen, "Intermolecular interaction assisted fabrication of dion-jacobson perovskite film with promoted photovoltaic property," *Chemical Engineering Journal*, vol. 451, 1 2022.

[6] J. A. Brehm, J. W. Bennett, M. R. Schoenberg, I. Grinberg, and A. M. Rappe, "The structural diversity of ABS3 compounds with d electronic configuration for the B-cation," *The Journal of Chemical Physics*, vol. 140, no. 22, 06 2014, 224703. [Online]. Available: https://doi.org/10.1063/1.4879659

A Glossary

- Formation energy: The formation energy of a material is the difference between the Gibbs free energy of the material minus the energy of the reactants, the more negative the formation energy, the more stable it is.
- Energy above convex hull: It is a very important parameter of the stability of a material, it is a line created by the linear combination of the different phases, if a material is above this energy value, then it is called meta stable or not stable.
- Heat of vaporization: It is the energy needed to evaporate, the more energy it requires it could become more stable.
- Shannon radii: It is the distance between the positions A-O and B-O, remember that X for these data are only Oxygen.
- Ionization Energy: Energy required to remove an electron from an atom, is the difference in energy between the neutral atom and the cation of the atom.
- Electron Affinity: Energy that an atom would have when gaining an electron. It is the difference between the energy of the neutral atom and the anion.
- Thermal conductivity: It refers to the amount of energy that a material is capable of conducting in a given time.
- Atomic Volume A-B: It is the volume occupied by the averages of the individual volumes of A and B.
- Electrical conductivity: It is the property of electron transport in a certain interval of time.
- Mendeleev number: is the average of the numbers of electrons.
- Density avg: is the average density between atoms A and B.
- host Asite0 Heat of Vaporization: Is the heat of vaporization of the pure element in the site A.
- Asite BCCvolume pa weighted avg: Average volume of the atoms in the A site in a BCC crystal.
- Asite Ionic Radius (angstroms) max: The max ionic radio of the A site atom.
- Asite BCCenergy pa max and min: refers to a maximum and minimum energy value per atom in the A-site of a body-centered cubic (BCC) crystal structure.
- Asite Atomic Radius: The radious of the atom in A site.
- Asite Atomic Volume: Is the atomic volume of atom A.

- Asite BCCefflatcnt range: Is the distance between A sites in the pure BCC crystal structure.
- host Asite0 OrbitalD: Metal atoms has D orbital in his last valance shell.
- host Asite0 NsValence: refers to the valence state of the A-site cation in a host material. The A-site refers to the position of the larger cation in the crystal structure, while the host material could refer to the bulk material in which the A-site cation is located.
- Ionization potential: Is the energy required to make the transition from one electron to the continuum.
- Density: Is the avg density calculated as thinking of the atoms as hard spheres (Volumne = Atomic Volume / Atomic Mass)
- The energy above hull (EAH) represents the energy difference between the total energy of a material and the energy of the most stable material (or phase) with the same composition. The energy of the most stable material is defined as the energy at which the material has the lowest Gibbs free energy, which takes into account both the internal energy of the material and its entropy.
- Thermal and electric conductivity: is the capacity of the material to transport heat and electrons, respectively.
- V(X) is the valence number of A or B.
- r(AXII)(A): Ionic radius of cation A.
- r(AVI)(A): Ionic radius of cation B.
- En(X): Avg electronegativity of atoms.
- l(X-O): The average distance between atom the atom X and the Oxygen.
- δENR: The difference of formation energy between the species and the structure.
- Lowest distortion: Classification of the lattice.

B Tablas

C Figuras:

	mean	max	min	std	Type	Key
A		Y	Ba		object	PK
host_Asite0_Heat of Vaporization	246.08	399.57	127.60	98.73	float64	
Asite_BCCvolume_pa_weighted_avg	37.15	63.59	22.88	10.79	float64	
Asite_BCCvolume_padiff_weighted_avg	0.03	0.59	-1.32	0.59	float64	
Asite_At. Radius (angstroms)_weighted_avg	2.46	2.76	1.72	0.29	float64	
Asite_n_wsthird_weighted_avg	1.12	1.22	0.81	0.16	float64	
Asite_Ionic Radius (angstroms)_max	1.10	1.42	0.72	0.18	float64	
Asite_BCCenergy_pa_max	-3.79	-1.51	-6.34	1.58	float64	
Asite_Atomic Radius (Å)_max	1.87	2.22	1.60	0.17	float64	
Asite_At. Radius (angstroms)_max	2.46	2.76	1.72	0.29	float64	
Asite_Atomic Volume (cm^3/mol)_max	22.69	39.24	13.97	7.05	float64	
Asite_shannon_radii_min	1.23	1.61	0.89	0.19	float64	
Asite_BCCenergy_pa_min	-3.79	-1.51	-6.34	1.58	float64	
Asite_BCCenergydiff_min	0.10	0.15	0.00	0.06	float64	
Asite_GSmagmom_min	0.00	0.00	0.00	0.00	int64	
Asite_At. Radius (angstroms)_min	2.46	2.76	1.72	0.29	float64	
Asite_shannon_radii_range	0.00	0.00	0.00	0.00	float64	
Asite_BCCefflatcnt_range	0.00	0.00	0.00	0.00	float64	
Asite_IsBoron_weighted_avg	0.00	0.00	0.00	0.00	int64	
Asite_IsHalogen_weighted_avg	0.00	0.00	0.00	0.00	int64	
Asite_IsPnictide_weighted_avg	0.00	0.00	0.00	0.00	float64	
Asite_IsRareEarth_weighted_avg	0.58	1.00	0.00	0.51	float64	
Asite_NfUnfilled_weighted_avg	3.58	11.00	0.00	4.32	float64	
Asite_NfValence_weighted_avg	3.42	11.00	0.00	4.17	float64	
host_Asite0_IsBCC	0.08	1.00	0.00	0.29	int64	
host_Asite0_IsCubic	0.00	0.00	0.00	0.00	int64	
host_Asite0_IsAlkali	0.00	0.00	0.00	0.00	int64	
host_Asite0_OrbitalD	0.17	1.00	0.00	0.39	int64	
host_Asite0_NsValence	2.00	2.00	2.00	0.00	int64	
Asite_IsAlkali_max	0.33	1.00	0.00	0.49	int64	

Table 1. A_site table description, included some statistical descriptors, object type is a str in python or a char in msSQL

	mean	max	min	std	Type	Key
B		Zr	Al		object	PK
host_Bsite0_at. wt.	98.09	195.08	26.98	53.91	float64	
host_Bsite0_Ionization Energy (kJ/mol)	718.74	906.40	577.60	89.06	float64	
Bsite_ BP (K)_weighted_avg	3802.92	5870.00	1180.00	1152.07	float64	
Bsite_At. Radius (angstroms)_weighted_avg	1.89	2.27	1.53	0.19	float64	
Bsite_Second Ionization Potential (V)_weighted_avg	13.91	20.51	0.00	6.44	float64	
Bsite_electrical conductivity_weighted_avg	12.06	60.70	0.00	13.22	float64	
Bsite_ BP (K)_max	3802.92	5870.00	1180.00	1152.07	int64	
Bsite_At. Radius (angstroms)_max	1.89	2.27	1.53	0.19	float64	
Bsite_First Ionization Potential (V)_max	7.43	9.39	5.99	0.94	float64	
Bsite_Third Ionization Potential (V)_max	23.93	39.72	0.00	12.65	float64	
Bsite_MendeleevNumber_min	52.50	74.00	11.00	14.73	int64	
Bsite_n_wsthird_min	1.62	1.85	1.21	0.20	float64	
Bsite_At. #_weighted_avg	41.92	78.00	13.00	20.43	float64	
Bsite_Period_weighted_avg	4.73	6.00	3.00	0.87	float64	
Bsite_IsMetal_weighted_avg	1.00	1.00	1.00	0.00	float64	
Bsite_NdUnfilled_weighted_avg	4.19	9.00	0.00	2.99	float64	
Bsite_NpUnfilled_weighted_avg	0.38	5.00	0.00	1.36	float64	
host_Bsite0_At. #	41.92	78.00	13.00	20.43	int64	
host_Bsite0_IsHexagonal	0.42	1.00	0.00	0.50	int64	
host_Bsite0_IsNoblegas	0.00	0.00	0.00	0.00	int64	
Bsite_IsMetal_max	1.00	1.00	1.00	0.00	int64	

Table 2. Description of the B_site table, included some statistical descriptors, object type is a string in python or a char in msSQL

	mean	max	min	std	Type	Key
Compound		YVO3	BaAlO3		object	PK
A		Y	Ba		object	FK
B		Zr	Al		object	FK
In literature	0.634	True	False	0.485	bool	
v(A)		3.000	-		object	
v(B)		4.000	-		object	
r(AXII)(Å)	1.374	1.610	0.890	0.208	float64	
r(AVI)(Å)	1.112	1.350	0.720	0.193	float64	
r(BVI)(Å)	0.622	0.900	0.480	0.069	float64	
EN(A)	1.032	1.310	0.890	0.131	float64	
EN(B)	1.747	2.280	1.220	0.233	float64	
l(A-O)(Å)	2.748	2.943	2.401	0.156	float64	
l(B-O)(Å)	1.971	2.644	0.000	0.272	float64	
δENR	-2.127	-1.685	-3.202	0.254	float64	
tG	0.971	1.132	0.814	0.083	float64	
θ		5.712	-		object	
μ	0.445	0.643	0.343	0.049	float64	
Lowest distortion		tetragonal	cubic		object	

Table 3. Description of the Crystal_sructure table, included some statistical descriptors, object type is a str in python or a char in msSQL

	mean	max	min	std	Type	Key
Compound		YVO3	BaAlO3		object	FK
Material Composition		Y8V8O24	Ba8Al8O24		object	PK
energy_above_hull (meV/atom)	115.1496	643.7298	0.0000	141.7363	float64	
formation_energy (eV/atom)	-1.9110	-0.7112	-3.2085	0.6504	float64	
shannon_radii_AB_avg	0.9929	1.2550	0.7620	0.1156	float64	
Density_AB_avg	6.3517	13.0950	3.0450	2.1150	float64	
BCCefflatcnt_AB_avg	7.4818	9.0509	6.3557	0.5878	float64	
BCCvolume_padiff_AB_avg	-0.0276	0.3838	-0.6600	0.2879	float64	
GSenergy_pa_AB_avg	-5.5428	-1.5818	-7.9852	1.3287	float64	
ICSDVolume_AB_avg	30.8148	47.8500	17.5500	7.3001	float64	
covalent radius_AB_avg	1.4851	1.8000	1.2000	0.1394	float64	
Ionization Energy (kJ/mol)_AB_avg	626.0380	748.5000	540.2500	41.3616	float64	
Electron Affinity (kJ/mol)_AB_avg	7.8092	80.5000	-103.0000	42.7367	float64	
Atomic Volume (cm^3/mol)_AB_avg	18.7091	29.5200	10.5350	4.7976	float64	
MendeleevNumber_AB_avg	32.4718	61.5000	10.5000	7.7335	float64	
First Ionization Potential (V)_AB_avg	6.4878	7.7580	5.5990	0.4320	float64	
thermal conductivity_AB_avg	52.4669	209.7000	5.6750	40.3340	float64	
at. wt._AB_diff	23.4719	55.1742	-28.8750	23.0686	float64	
specific heat capacity_AB_diff	-0.0700	0.2855	-0.3480	0.0942	float64	
electrical conductivity_AB_diff	-1.9768	15.4000	-28.9500	6.5615	float64	
BCCefflatcnt_AB_ratio	1.5078	1.8327	1.1358	0.1652	float64	
Ionization Energy (kJ/mol)_AB_ratio	0.7768	1.0643	0.5548	0.1015	float64	
Heat of Vaporization_AB_ratio	0.6974	1.8184	0.1902	0.4264	float64	

Table 4. Description of the Material_properties table, included some statistical descriptors, object type is a str in python or a char in msSQL

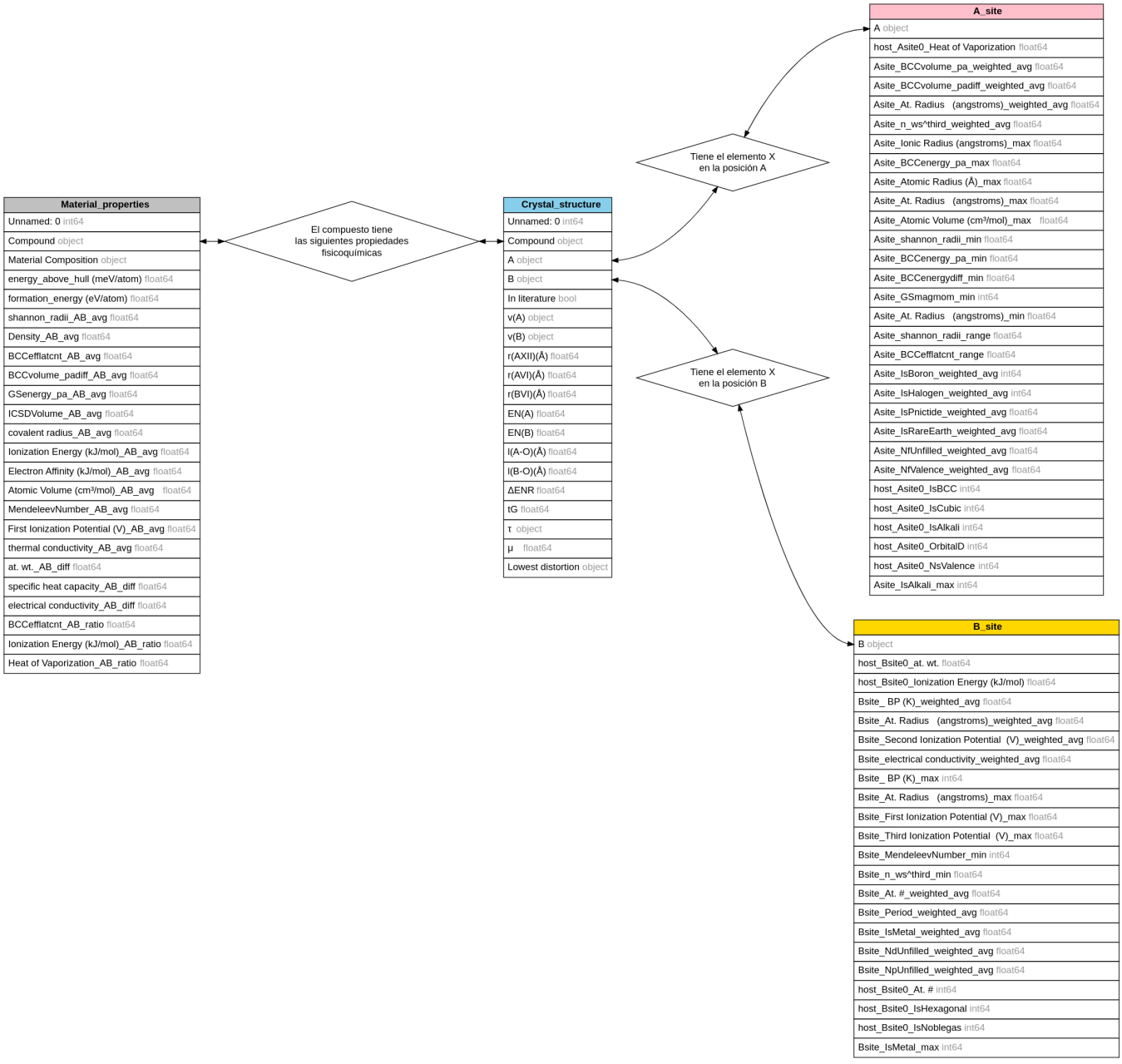

Figure 2. Entity-relationship diagram of the tables that will be used to analyze the stability and properties of interest of perovskite materials