Data-driven tolerance factor for chalcogenide perovskites and their suitability for photovoltaic applications

Diego A. Garzón[1, 2], Luisa Andrade[2], Sascha Sadewasser[1], José A. Marquéz[3]
[1]International Iberian Nanotechnology Laboratory, Braga, Portugal
[2]LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering and ALiCE – Associate Laboratory in Chemical Engineering, Faculty of Engineering, University of Porto, Porto, Portugal
[3]Department of Physics, Humboldt-Universität zu Berlin, Berlin, Germany

diego.garzon@inl.int

Abstract

Chalcogenide perovskites are a recent family of inorganic materials in which anions in the ABX_3 perovskite structure correspond to S^{2-}, Se^{2-} or Te^{2-}, and do not consider oxygen-based compounds. With less than a decade of development, these materials have acquired significant attention from the scientific community due to their potential as active material for energy conversion applications, especially in photovoltaic technology. Although theoretical studies predict that many chalcogenide perovskites have outstanding optoelectronic properties, only a few have been experimentally synthesized with the true corner-sharing perovskite structure. [1] Challenges arise due to the different possible edge-sharing crystal structures that can exist with this stoichiometric arrangement (Figure 1a-d).

For traditional perovskites, such as oxide or halide-based compounds a geometric factor based on the ionic radii is commonly used to predict phase stability. The Goldschmith tolerance factor, $t = (r_A + r_X)/ (\sqrt{2} (r_B + r_X))$ has traditionally been a rule of thumb for perovskite prediction, with stable structures typically falling within $0.8<t<1.0$. [2] However, this approach is less effective for predicting the stability of non-oxide perovskites, achieving an accuracy of less than 66%. Recently, other tolerance factors have been developed, incorporating terms related to electronegativity to account for the covalent nature of metal-chalcogen bonds. [3], [4] These newer tolerance factors improve accuracy to approximately 72% when compared with experimental data.

Despite these improvements, there remains significant room for enhancement. For oxide and halide perovskites, a new tolerance factor, τ, was developed using a machine learning approach called SISSO (sure independence screening and sparsifying operator). This tolerance factor achieves an accuracy of approximately 91% when validated against experimental results. Additionally, it facilitates the identification of new double perovskites and their probabilities of stability, providing a valuable guide for synthetic and computational efforts. [5]

Inspired by the SISSO method, [6] we constructed a new dataset from experimental reports of non-oxide perovskites, focusing particularly on sulfide and selenide compounds. This dataset contains 283 materials, with 27% corresponding to chalcogenides and the rest to halides. We then generated a set of features by applying linear combinations of elemental properties—such as electronegativity, ionic radius, and charge—using basic mathematical operations like subtraction, multiplication, division, exponentiation, square roots, and natural logarithms.

The features were ranked based on the overlap between stability ranges of stable and unstable materials, with the top 2,000 features exhibiting the smallest overlaps. Next, classification trees were trained using 5-fold cross-validation, and the features were ranked by mean accuracy. Notably, the classification trees were limited to a depth of one, ensuring a simple threshold-based decision-making process.

Through this process, we derived a new tolerance factor with greater accuracy than both t and τ for the same dataset. Using this factor, we identified a set of potential chalcogenide perovskites by fixing a stoichiometric ratio of 1:1:3 and ensuring charge-balanced compounds (Figure 1e). Solid solutions with multiple metals at the A or B positions were also considered. The new tolerance factor was then used to predict the stability of these unexplored chalcogenide perovskites.

The predicted stable chalcogenide perovskites were ranked based on the earth abundance of their constituents. Subsequently, state-of-the-art machine learning algorithms, such as MODNet [7] and CrabNet, [8] were applied to predict properties relevant to photovoltaic applications, particularly bandgaps. These algorithms were fine-tuned using experimental data from the stable perovskites' dataset, along with their corresponding experimental properties.

By considering stability, earth abundance, and photovoltaic properties, each material was ranked based on its suitability for next-generation solar cells. This approach requires no resource-intensive computational calculations, relying solely on stoichiometric formulae. Consequently, this pipeline provides a fast, reliable, and efficient method for guiding the experimental discovery of new chalcogenide perovskites.

In conclusion, the development of this data-driven tolerance factor represents a significant advance in the predictive design of chalcogenide perovskites, addressing limitations of earlier approaches fully based on density functional theory methods. By enhancing accuracy and integrating practical

considerations like earth abundance and photovoltaic properties, this pipeline accelerates the identification of stable perovskite structures. It offers a simple approach for guiding experimental efforts, enabling progress in the discovery of next-generation materials for sustainable energy technologies.

References

[1] K. V. Sopiha, C. Comparotto, J. A. Márquez, and J. J. S. Scragg, “Chalcogenide Perovskites: Tantalizing Prospects, Challenging Materials,” *Advanced Optical Materials*, vol. 10, no. 3, p. 2101704, 2022, doi: 10.1002/adom.202101704.

[2] V. M. Goldschmidt, “Die Gesetze der Krystallochemie,” *Naturwissenschaften*, vol. 14, no. 21, pp. 477–485, May 1926, doi: 10.1007/BF01507527.

[3] A. Jess, R. Yang, and C. J. Hages, “On the Phase Stability of Chalcogenide Perovskites,” *Chem. Mater.*, vol. 34, no. 15, pp. 6894–6901, Aug. 2022, doi: 10.1021/acs.chemmater.2c01289.

[4] J. W. Turnley, S. Agarwal, and R. Agrawal, “Rethinking tolerance factor analysis for chalcogenide perovskites,” *Mater. Horiz.*, vol. 11, no. 19, pp. 4802–4808, 2024, doi: 10.1039/D4MH00689E.

[5] C. J. Bartel *et al.*, “New tolerance factor to predict the stability of perovskite oxides and halides,” *Sci. Adv.*, vol. 5, no. 2, p. eaav0693, Feb. 2019, doi: 10.1126/sciadv.aav0693.

[6] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, and L. M. Ghiringhelli, “SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates,” *Phys. Rev. Mater.*, vol. 2, no. 8, p. 083802, Aug. 2018, doi: 10.1103/PhysRevMaterials.2.083802.

[7] P.-P. De Breuck, M. L. Evans, and G.-M. Rignanese, “Robust model benchmarking and bias-imbalance in data-driven materials science: a case study on MODNet,” *J. Phys.: Condens. Matter*, vol. 33, no. 40, p. 404002, Oct. 2021, doi: 10.1088/1361-648X/ac1280.

[8] A. Y.-T. Wang, M. S. Mahmoud, M. Czasny, and A. Gurlo, “CrabNet for Explainable Deep Learning in Materials Science: Bridging the Gap Between Academia and Industry,” *Integr Mater Manuf Innov*, vol. 11, no. 1, pp. 41–56, Mar. 2022, doi: 10.1007/s40192-021-00247-y.

Figures

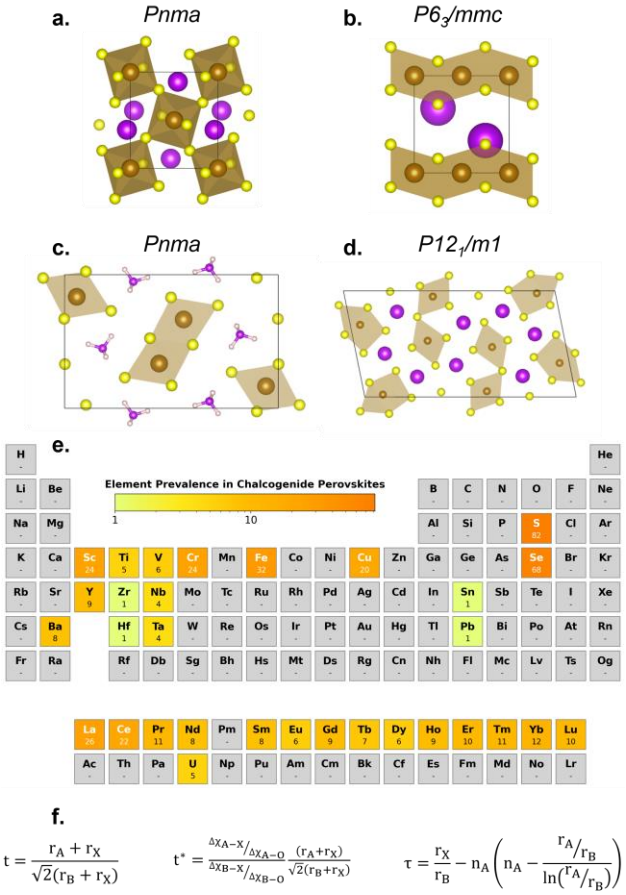

Figure 1. Typical crystal structures that ABX_3 compounds could have, where only the corner-sharing structure corresponds to the real perovskite (a), while the other crystals (b, c, d) are equivalent to edge-sharing structures and cannot be denominated perovskites. (e) Periodic table of the elements with a color code based on the element prevalence in predicted chalcogenide perovskites. (f) Different formulae corresponding to reported tolerance factors for perovskite materials.