Learning from Machine Learning: The case of band-gap directness in semiconductors

Gustavo M. Dalpian[1]**,** Elton Ogoshi[2]**,** Mário Popolin Neto[3,4], Carlos Mera Acosta[2], Gabriel M. Nascimento[2], João N. B. Rodrigues[2,5], Osvaldo N. Oliveira Jr[4], Fernando V. Paulovich[6]

[1]University of São Paulo, São Paulo, SP, Brazil
[2]Federal University of ABC, Santo André, Brazil
[3]Federal Institute of São Paulo, Araraquara, Brazil
[4]University of São Paulo, São Carlos, Brazil
[5]Federal University of Pernambuco, Recife, Brazil
[6]Eindhoven University of Technology, Eidhoven, The Netherlands.

dalpian@usp.br

Machine learning methods are now integral to materials research thanks to the increasing availability of curated datasets containing structural and electronic data derived from quantum-mechanical calculations. Also, the accessibility and user-friendliness of Machine Learning algorithms and software have extended their utility to a broader scientific community. Machine Learning has been mostly used for predictive purposes. Common examples of such applications to Materials Science include predicting a material's crystal structure or band gap. An ultimate goal for Machine Learning, however, is to build models that can go beyond prediction, with scientific concepts being discovered via descriptive models. This is a formidable task, either because the feature space is too large, or because the physical mechanisms behind a specific phenomenon are unknown, barely understood or complex. Some cases of success have been reported, including some related to Materials Science. In this work, we use interpretable Machine Learning methods to analyze a fundamental property of semiconductors that lacks a detailed understanding: the reason why some semiconductors have a direct band gap while others have an indirect one. In semiconductors, the top of the valence band (representing occupied states) and the bottom of the conduction band (representing empty states) are separated by an energy gap. The top of the valence band and the bottom of the conduction band can be located at different momentum k-wavevectors (i.e., indirect band gap) or at the same k-wavevector (i.e., direct band gap). Determining the type of band gap is relevant for semiconductor applications. For instance, indirect band gap materials are usually not suitable for applications in optoelectronic devices because the absorption or emission of a photon requires an electron-lattice momentum exchanged, and such two-step process is less likely to occur. The most used material in the microelectronics industry, i.e. silicon, has an indirect band gap and thus is not suitable for optoelectronic devices such as light emitting diodes (LEDs). Transforming indirect band gap materials into direct band gap ones is still challenging. Common strategies for tuning indirect-direct band gaps include alloying, exploring strains, and quantum confinement. The directness of the band gap is not difficult to determine, either theoretically or experimentally, as discussed in fundamental physics textbooks. However, no unified theory exists to explain why one material has a direct or an indirect band gap. Semiconductors encompass a variety of materials and crystal structures, from simple diamond structures such as Si to perovskites, which contain at least three different types of atoms and many structural distortions that can change their properties. To the best of our knowledge, only a couple of works have focused on the explanation of the band gap directness in semiconductors. Yuan and collaborators [1] focused on Zincblende semiconductors, and have shown that materials with cations with occupied d-levels tend to have direct band gaps owing to the symmetry of the zincblende lattice. In the zincblende point group, the deep d-band interacts with the VB/CB at the Γ, L and X high- symmetry points in the reciprocal space according to wavefunction symmetry at this point which is dictated by their symmetry at the k-point and the band symmetry representation. The d-bands do not interact with the bottom of the conduction band at the Gamma point. This means that in the presence of occupied d-orbitals, the band repulsion at other k-points leads the CB higher in energy with respect to the Gamma point, which is kept fixed regardless the presence of d-orbitals due to the lack of interaction. This mechanism controls the direct/indirect gap for this specific crystal structure. In another paper, Choubisa and collaborators [2] proposed general rules to explain when a material has a direct or indirect band. The rules include the occupation of p orbitals, the position of the LUMO and the electronegativity of the constituent atoms. As we will show below, our rules are considerably different from these, providing new ways to design direct or indirect band gap materials. Herein, we conduct a broad analysis using data science and interpretable Machine Learning methods such as Decision Trees (DT) and Random Forests (RF). In particular, we perform a descriptive analysis using the VAX method [3], extracting Jumping Emerging Patterns (JEPs, descriptive logic rules) from Machine Learning models, and then look for causal relations or insights to explain why a material has a direct or indirect band gap. VAX is able to produce images such as the one shown in Figure 1, which are used in the interpretation analysis. Our ML models use a dataset extracted from the Materials Project (~10000 entries) and we've used atomic and structural features. Our findings indicate that the directness of the band gap depends on the symmetry of the different materials. No general pattern (rule) was encountered which would apply to all structural families of materials. However, if we divide the materials in smaller groups of compounds with similar symmetries (e.g, Zincblende, Wurtzite, Rock Salt, and Perovskite structures), VAX generates specific patterns that explain their bandgap directness. Specifically, we recovered the

known result for Zincblende structures, i.e., the existence of d orbitals. Furthermore, we have also found that relative energies of highest occupied states determine the direct–indirect bandgap transitions in other structures. We believe that these results demonstrate that explainable machine learning is promising to advance the understanding of physics problems.[4]

References

[1] Yuan LD, Deng HX, Li SS, Wei SH, Luo JW. Unified theory of direct or indirect band-gap nature of conventional semiconductors. Phys Rev B. 98 (2018).

[2] Choubisa H, Todorović P, Pina JM, Parmar DH, Li Z, Voznyy O, Tamblyn I, Sargent EH. Interpretable discovery of semiconductors with machine learning. NPJ Comput Mater. (2023).

[3] Neto MP, Paulovich FV. Multivariate data explanation by jumping emerging patterns visualization. IEEE Trans Visualizat Comput Graph. (2022).

[4] Ogoshi, E., Popolin-Neto, M., Acosta, C.M. et al. Learning from machine learning: the case of band-gap directness in semiconductors. Discov Mater 4, 6 (2024).
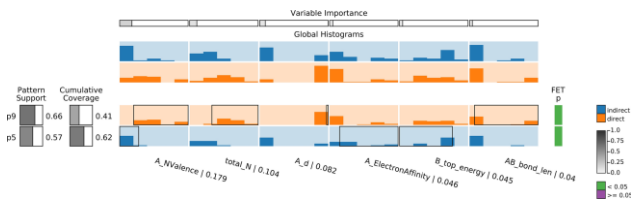
Figures

Figure 1. Representative diagram generated by the VAX method for the Zinc Blende materials. Lines indicate the patterns (rules) generated, whereas the columns represent the most important features. Blue boxes indicate indirect band gap materials, and orange indicates direct band gap materials. In this specific case, we observe that the number of valence electrons of the anion is important to separate the two groups. More information can be obtained in Ref. [4].