

Active Learning: Accelerating Discovery of Optimal Optical Materials through Synergistic Computational Approaches

Victor Trinquet¹, Matthew L. Evans¹,
Cameron Hargreaves¹, Pierre-Paul De Breuck¹,
Gian-Marco Rignanese¹

¹UCLouvain, Institut de la Matière Condensée et des
Nanosciences (IMCN), Chemin des Etoiles 8, Louvain-la-Neuve
1348, Belgium
victor.trinquet@uclouvain.be

Inorganic crystals combining both a high refractive index and a wide band gap are of great importance for many advanced optical applications. However, the observed inverse relationship between these two properties hampers a straightforward identification of promising materials. To go beyond a costly and time-consuming trial and error approach, an *ab initio* high-throughput screening was performed on more than 4000 semiconductors [1]. Yet, this selection approach is limited to the compounds that are considered initially. It may therefore inadvertently overlook potentially valuable areas within the materials space. To go one step further in this search, we here introduce an active learning framework powered by the synergistic use of an AI agent, open databases (via OPTIMADE [2]) and density-functional theory calculations. This framework strategically guides the exploration of crystal structures, emphasizing diversity within the chemical space and the exploitation of essential criteria for optical materials. Our findings demonstrate a substantial acceleration in comparison to conventional methods as emphasized in Figure 1. by the new data embracing the Pareto front. The framework operates autonomously, eliminating the need for human intervention. Furthermore, we anticipate its potential to generalize to other material properties in the future.

References

- [1] Naccarato, Francesco, et al., *Physical Review Materials* 3.4 (2019): 044602.
- [2] Andersen, Casper W., et al., *Scientific data* 8.1 (2021): 217.

Figures

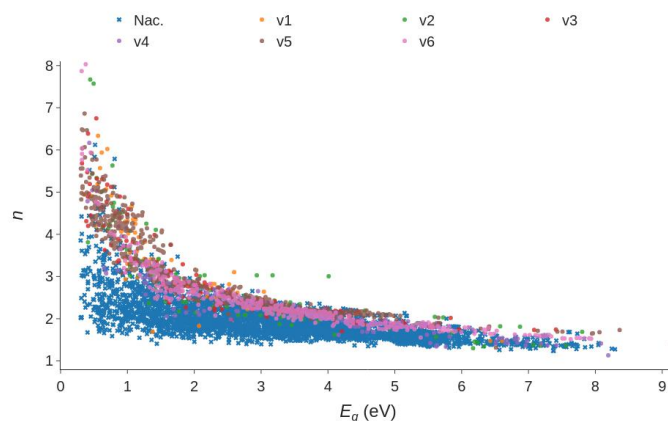


Figure 1. Distribution of the refractive index (n) vs. bandgap (E_g) in eV for the starting dataset labeled as Nac. (blue crosses) and the new data added iteratively at each active learning loop labeled as v1-v6 (dots).