Band Gap Prediction in h-BN Using MBJ Data and Physically Informed Descriptors

Buccino, Michael Scheele, Kendra Kim-Reyes, Greis J.

Department of Physics and Astronomy, SUNY New Paltz

Data-Driven Physics

DFT and ML are vital for testing material properties in pharmaceuticals, clean energy, and engineering industries.

- The Kohn-Sham equations are the basis to Density Functional Theory (DFT):

$$\left[-\frac{1}{2}\nabla^2 + V_n(r)\right]\phi_i(r) = \epsilon_i\phi_i(r)$$

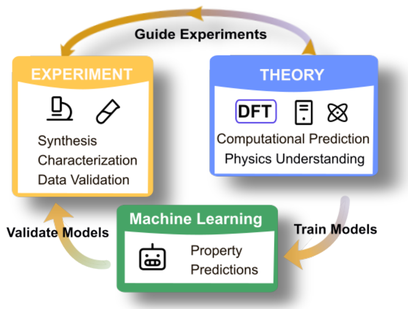
$$n(r) = \sum_i |\phi_i(r)|^2$$

$$V_H(r) = \int \frac{n(r')}{|r-r'|}\,dr'$$

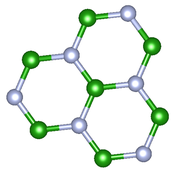
$$\left[-\frac{1}{2}\nabla^2 + V_n(r) + V_H(r)\right]\phi_i^{\text{new}}(r) = \epsilon_i^{\text{new}}\phi_i^{\text{new}}(r)$$

Iterate

- Physics from DFT, accuracy from Machine Learning (ML): Replace costly band gap corrections with learned corrections.



- The modified Becke–Johnson (mBJ) method band gaps closely approximate experimental values (~ 0.1–$0.3\,\text{eV}$ accuracy) and serve as the prediction target.
- Case Study: Hexagonal Boron Nitride (hBN) — monolayer

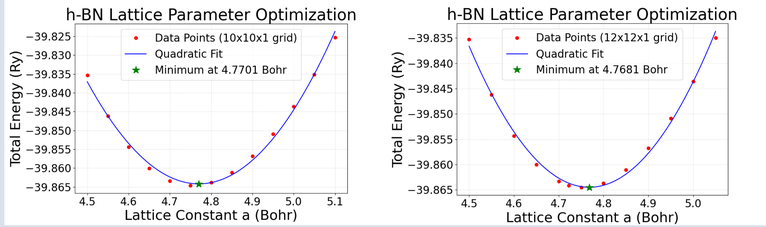


- Honeycomb lattice with alternating boron and nitrogen atoms.
- Ultrawide band gap insulator (about $6\,\text{eV}$).
- Insulating substrate and encapsulation layer for two-dimensional devices.

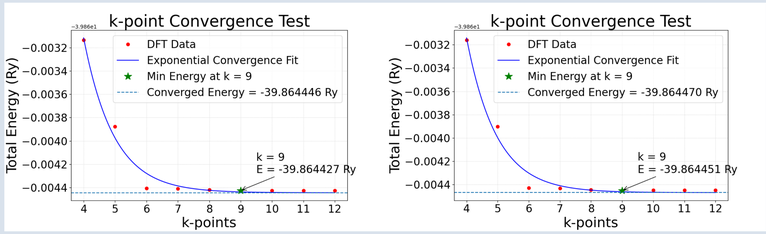
DFT Foundation

The graphs below illustrate the results of the calculations used to extract computational parameters for h-BN with k-point grids of 10x10x1 and 12x12x1 respectively.

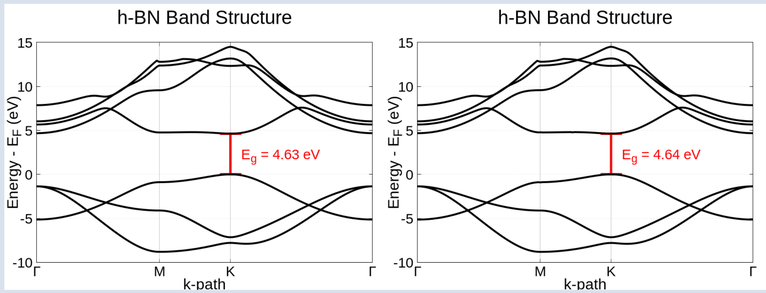
- Lattice parameter optimization from total energy minimization.



- Energy convergence with respect to k-point sampling.



- Standard PAW–DFT Band Structure of Monolayer h-BN



- The optimized lattice constant is $a = 4.7$ Bohr (≈ 2.49 Å), in excellent agreement with the accepted value for monolayer hBN (≈ 2.50 Å).
- Accepted/experimental-scale value is $\approx 6.08\text{eV}$, while our standard DFT (PAW) result is 4.6eV (typical underestimation by baseline DFT).

ML Correction

The workflow for obtaining and comparing MBJ data

Download data → 2D Materials (JARVIS)

~ 1100 materials → mBJ filtering → 246 materials

Train/Test Split → Training Set (80%) → Test Set (20%)

All non-hBN Hold-out non-hBN + hBN

Features (X) → Physics features

2 features:
- ◇ Standard DFT gap
- ◇ Electronic dielectric

Prediction: ~ 5.23 eV

Composition descriptors

22 chemical features:
- ◇ Z: atomic number
- ◇ X: electronegativity
- ◇ R: atomic radius
- ◇ IE: first ionization

Prediction: ~ 6.0 eV

Prepare Data → Convert to numeric

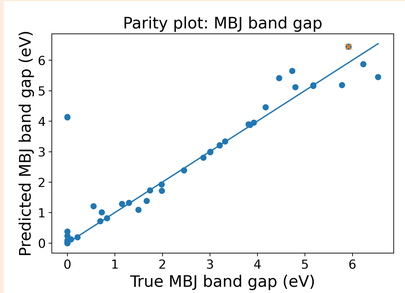
Mean Imputation

Build X_{train}, Y_{train}

Train + Evaluate → Random Forest Regressor

Metrics: MAE, R^2

Predict E_g^{MBJ}



- Overall model error (MAE): 0.336 eV
- Overall fit (R^2): 0.938
- Held-out monolayer hBN test (not used in training): The model predicts E_g^{MBJ} close to the reference value.
- JVASP-76649: $E_g^{\text{MBJ}} = 5.917\,\text{eV}$, Predicted $E_g^{\text{MBJ}} = 6.049\,\text{eV}$.

Analysis of Results

Density Functional Theory

- The standard DFT band gap is underestimated ($\sim 4.6\,\text{eV}$ vs. $\sim 6.08\,\text{eV}$), consistent with the expected band-gap error of conventional DFT.
- Energy vs. lattice parameter optimization yields an equilibrium lattice constant of ~ 4.7 Bohr, in very good agreement with the accepted experimental value for monolayer hBN.

Machine Learning

- The model achieved MAE = 0.336 eV and $R^2 = 0.938$ using 246 monolayer materials with MBJ band-gap references.
- Including composition descriptors improved the predicted hBN band gap from 5.23 eV (DFT-only features) to 6.0 eV, much closer to the MBJ value of 5.917 eV.

Overall

- Standard DFT captures the essential physics of monolayer hBN, but systematically underestimates its band gap.
- Machine learning uses DFT-based features and composition descriptors to learn this correction, giving near-mBJ band gap predictions at much lower computational cost.
- Future work will expand the dataset and explore more advanced machine learning models to further improve band gap predictions for 2D materials.

Acknowledgments

- This work was supported by the **Academic Year Undergraduate Research Experience (AYURE)** program at **SUNY New Paltz**.
- We gratefully acknowledge the **Research, Scholarship, and Creative Activities (RSCA) Office** for funding through the **Student Opportunity Grant**.

Selected References

- Cassabois *et al.*, *Nat. Photon.* **10**, 262–266 (2016). • Ward *et al.*, *npj Comput. Mater.* **2**, 16028 (2016). • Wang *et al.*, *Chem. Mater.* **32**, 4954–4965 (2020). • Schmidt *et al.*, *npj Comput. Mater.* **5**, 83 (2019). • Lee *et al.*, *Phys. Rev. B* **93**, 115104 (2016).

kimreyesg@newpaltz.edu