Machine Learning for Materials Science

Assignment-1

Ms. Sulagna Ghosh

Senior Research Fellow,
VECC Kolkata, a CI of HBNI
Registration ID: **MLC3419**

Mr. Shuvankar Das

Senior Research Fellow,
SINP Kolkata, a CI of HBNI
Registration ID: **MLC3329**

Defining the problem (base paper: https://doi.org/10.1039/D3CP02431H)

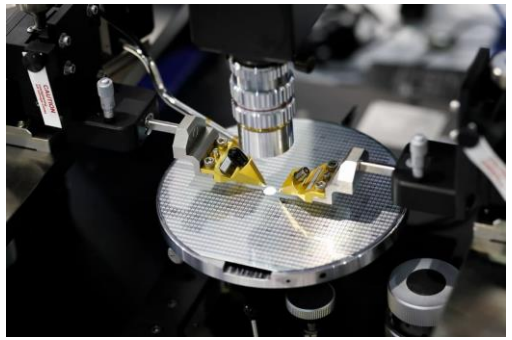
By low concentration doping the properties of a semiconductor can be modified

Application in:

- Electronic
- Optoelectronic
- Spintronic
- And many other devices

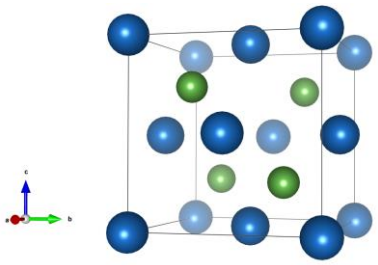
Intrinsic semiconductor
(GaAs)

Doping a semiconductor
($GaAs_xN_{1-x}$)



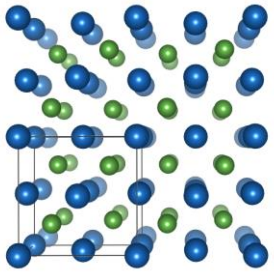
Defining the problem (base paper: https://doi.org/10.1039/D3CP02431H)

Calculation of low concentration doping is a problem using Supercell method in DFT



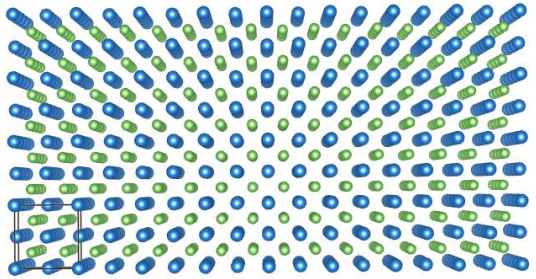
GaAs -*zinc blende*
(8 atoms)

For ~1.6% doping



$GaAs_xN_{1-x}$(x=0.0156)
(64 atoms)
(2x2x2) supercell

For ~0.1% doping



$GaAs_xN_{1-x}$(x=0.000976)
(1024 atoms)
(8x4x4) supercell

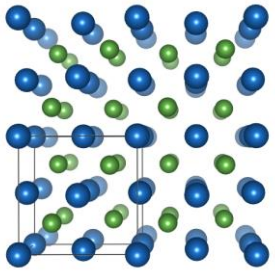
Low-Level Doping (< 10^{15} cm^{-3})
Medium Doping (10^{15} - 10^{17} cm^{-3})
High Doping (10^{18} - 10^{19} cm^{-3})

To perform Low-Level Doping calculation ($10^{15}/10^{23}=10^{-8}$ or 0.000001% doping)
We required ~125000 supercells

Huge/Unmanageable computational cost

Defining the problem

By changing the doping concentration (x) we can tune the properties of semiconductor



- Thousands of possible materials
- Experimental synthesis is expensive
- DFT calculations are time consuming

That's why the machine learning (ML) needed

Problem statement: "Band gap prediction of pristine and transition metal doped semiconductors using ML"

Workflow and Timeline of Deliverables

1. Creating the dataset (equal contribution)
   - Chemical composition of the host and dopant, doping concentration
   - Radius, charge state, electronegativity of the atoms
   - Lattice parameter, symmetry of the host, distance between the dopants
   - Band gap, DOS of the host
   - Formation energy, stability of the pristine and doped structure etc…
2. Training data and test data splitting (equal contribution)
3. Literature survey and selection of algorithm for the ML model (hoping to start by mid-term) (equal contribution)
4. Prediction of band gap
5. Use of the trained model for unknown materials. If possible, then extending the model to study about the induced magnetism (expected by end-term) (equal contribution)

Source

- Y. Tang, H. Chen, J. Wang and X. Neu, Phys. Chem. Chem. Phys. **25** (2023) 18086-18094. https://doi.org/10.1039/D3CP02431H
- Z. Wang, Y. Han, J. Cai, S. Wu, Jinjin, Energy Storage Materials **45** (2022) 1201-1211. https://doi.org/10.1016/j.ensm.2021.11.020
- Class-notes
- Wikipedia and internet