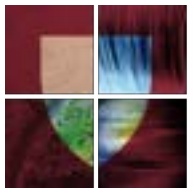
Defect Structure Searching & Automation

Seán R. Kavanagh, Environmental Fellow @ Harvard



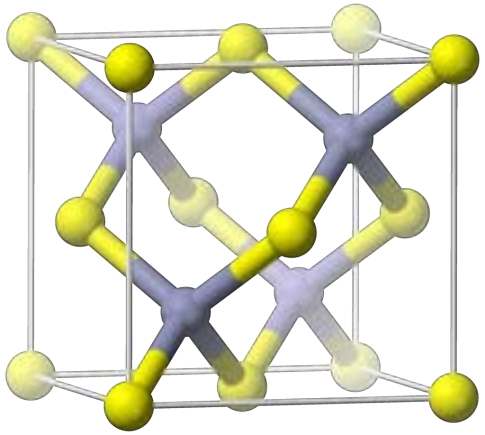
HARVARD UNIVERSITY
CENTER FOR THE ENVIRONMENT
A Center of the Salata Institute

Harvard John A. Paulson
School of Engineering
and Applied Sciences

CdTe: Motivation

Most commercially-successful thin-film PV solar cell technology

- Ideal band gap of ~1.5 eV
- Low carrier masses
- Strong absorption
- Relatively simple crystal structure (zinc blende)
- Well-studied, relatively well-optimized manufacturing
- Stable
- Low energy payback time (2x better than Si)
- Low degradation rate
- Good temperature coefficient
- …



First Solar®

Efficiency record holders, $20 bn market cap

CdTe Solar Farm in California (550 MW)

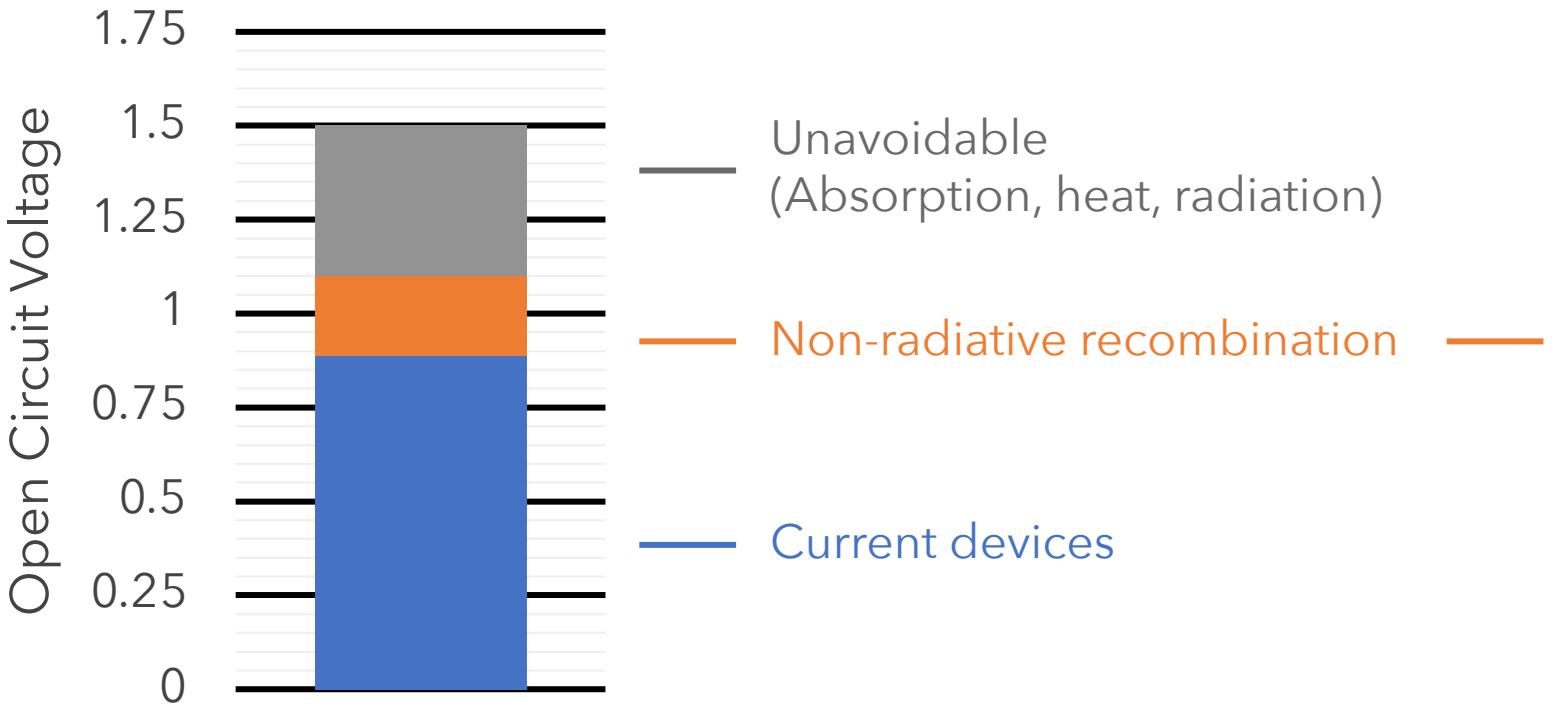
Scarpulla *et al. Sol. Energy Mater Sol. Cells* 2023

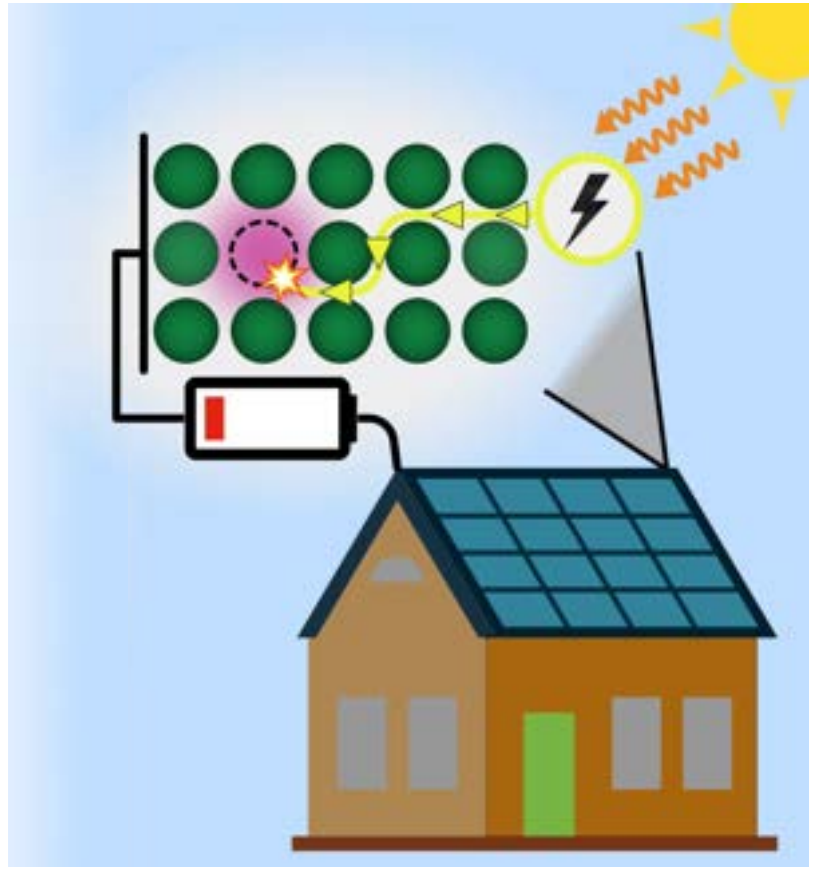
CdTe: Motivation

Most commercially-successful thin-film PV solar cell technology

But PV efficiency has stagnated at 22.1% (out of a potential ~32%), ***limited by defects***

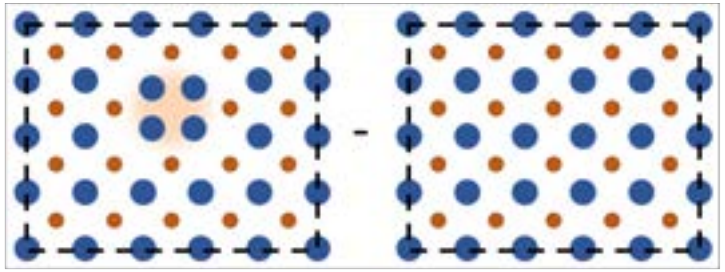
- Limited *p*-type doping concentrations
- Defect-mediated electron-hole recombination



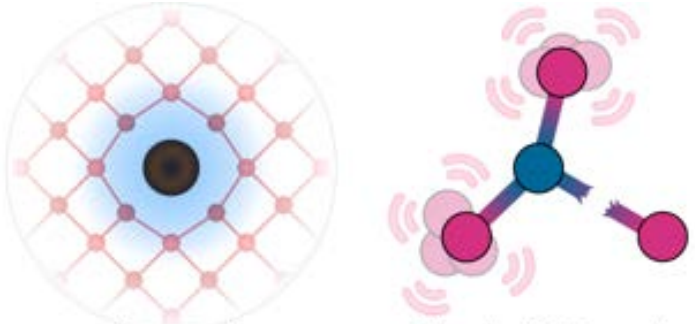


CdTe: Defects

Defect supercell Bulk supercell

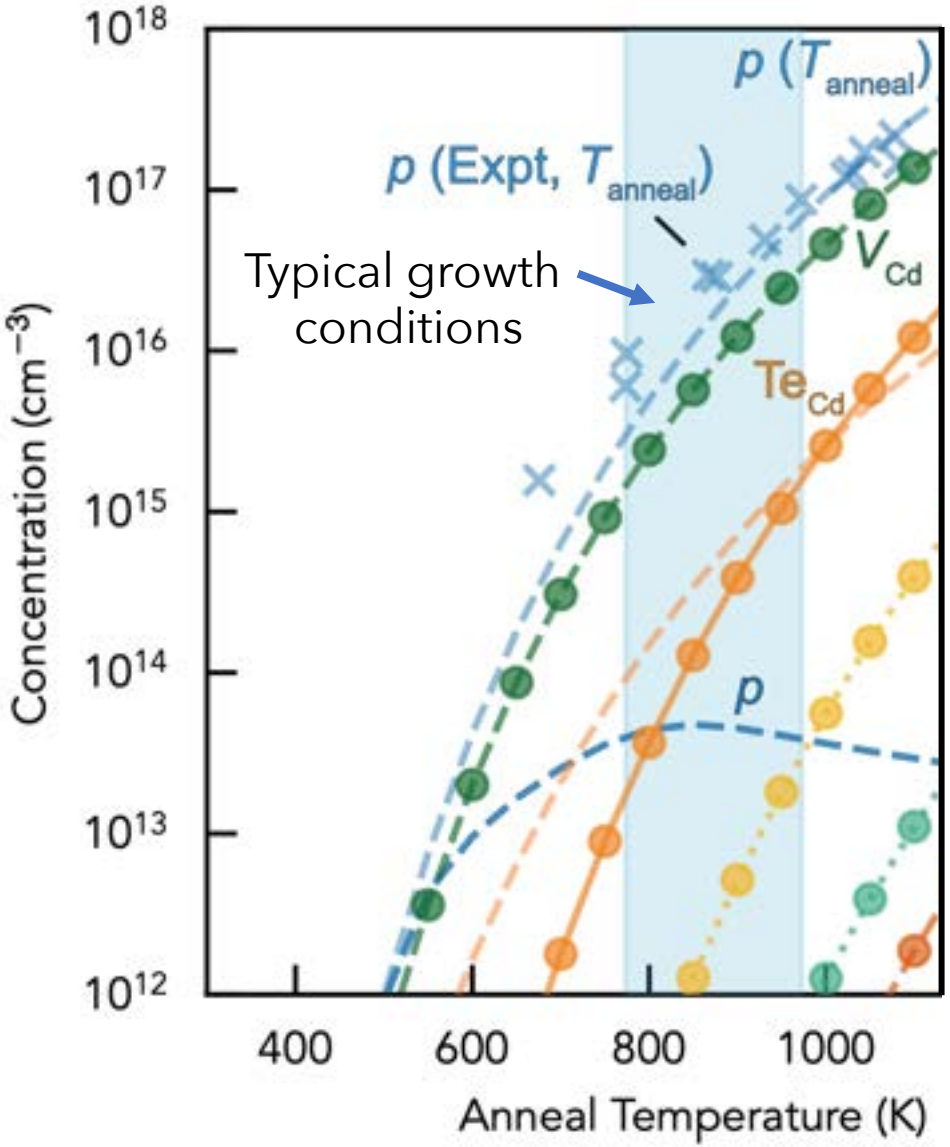


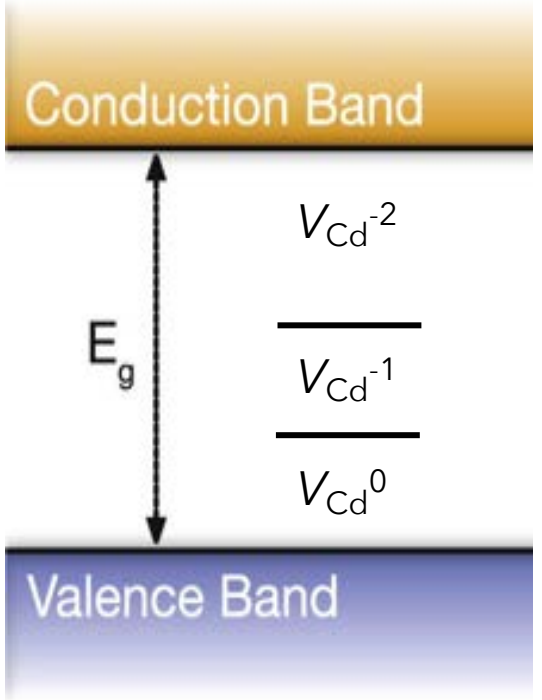
(+ corrections)



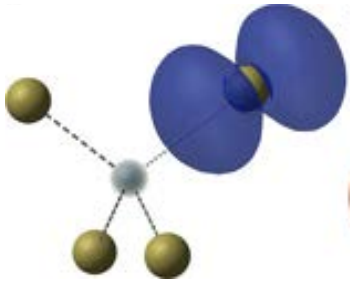
(early versions)

DFT: HSE(34.5% exchange), with spin-orbit coupling

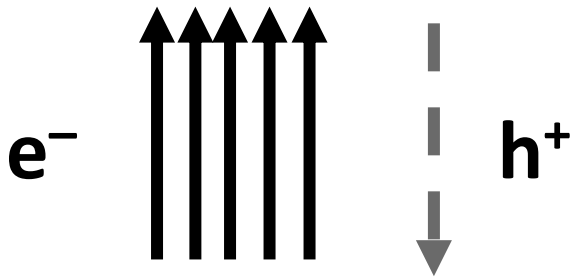




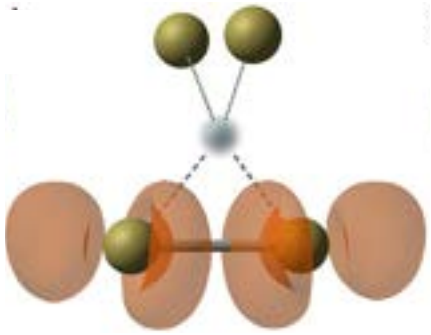
Modelling Electron-Hole Recombination

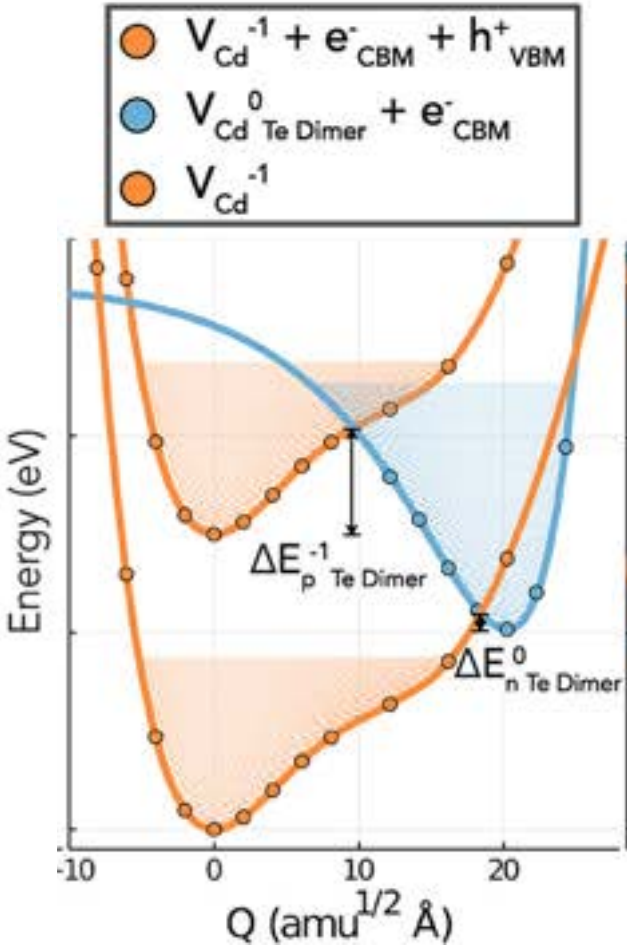


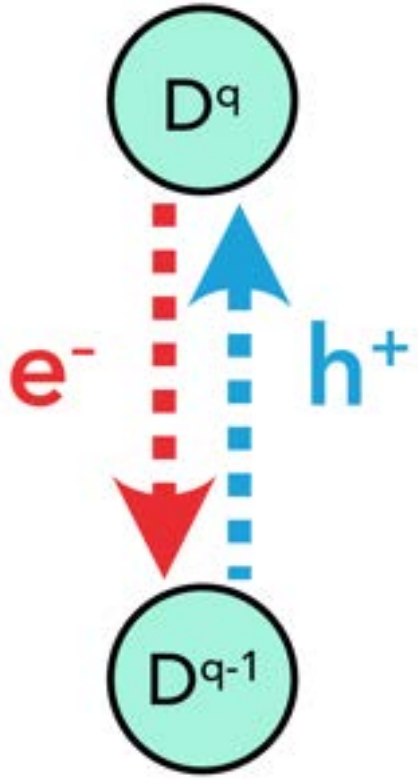
V_{Cd}^{-1}



V_{Cd}^{0}

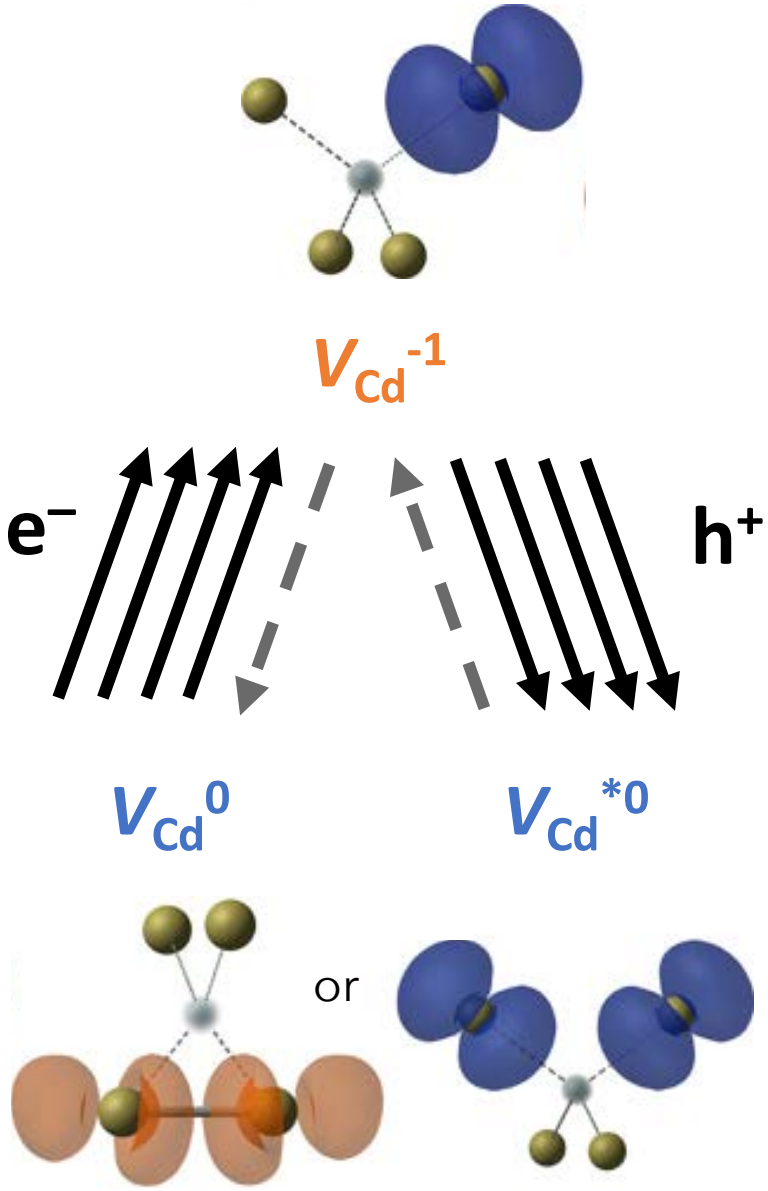


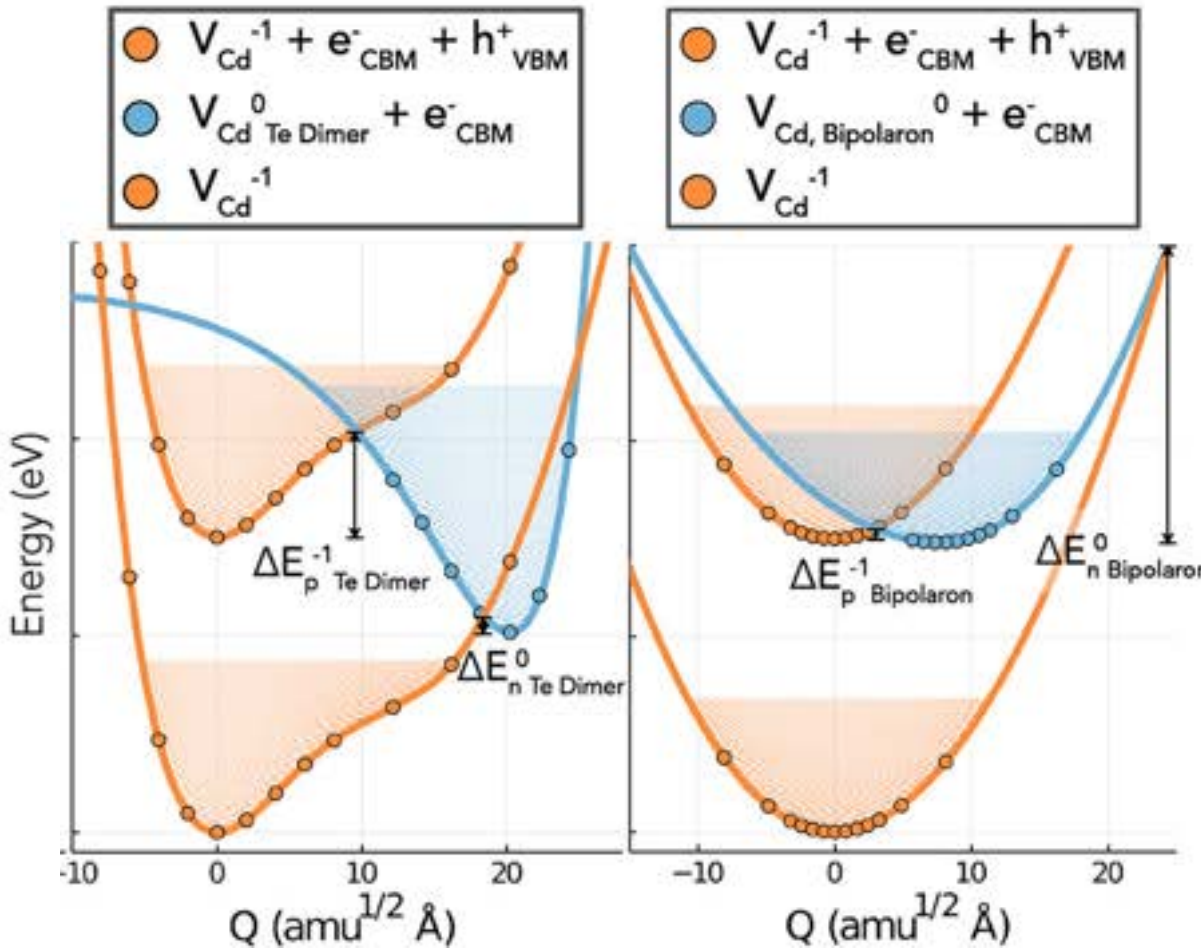




Kavanagh et al. *ACS Energy Lett* **2021**

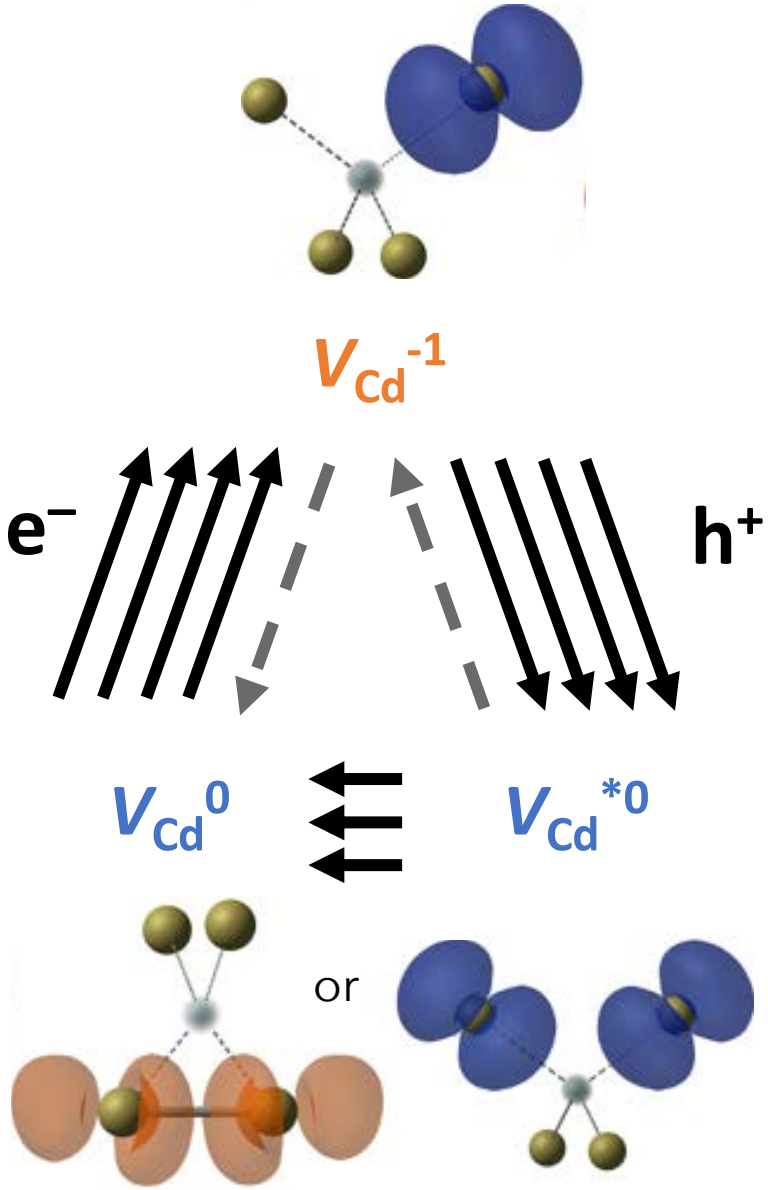
Modelling Electron-Hole Recombination

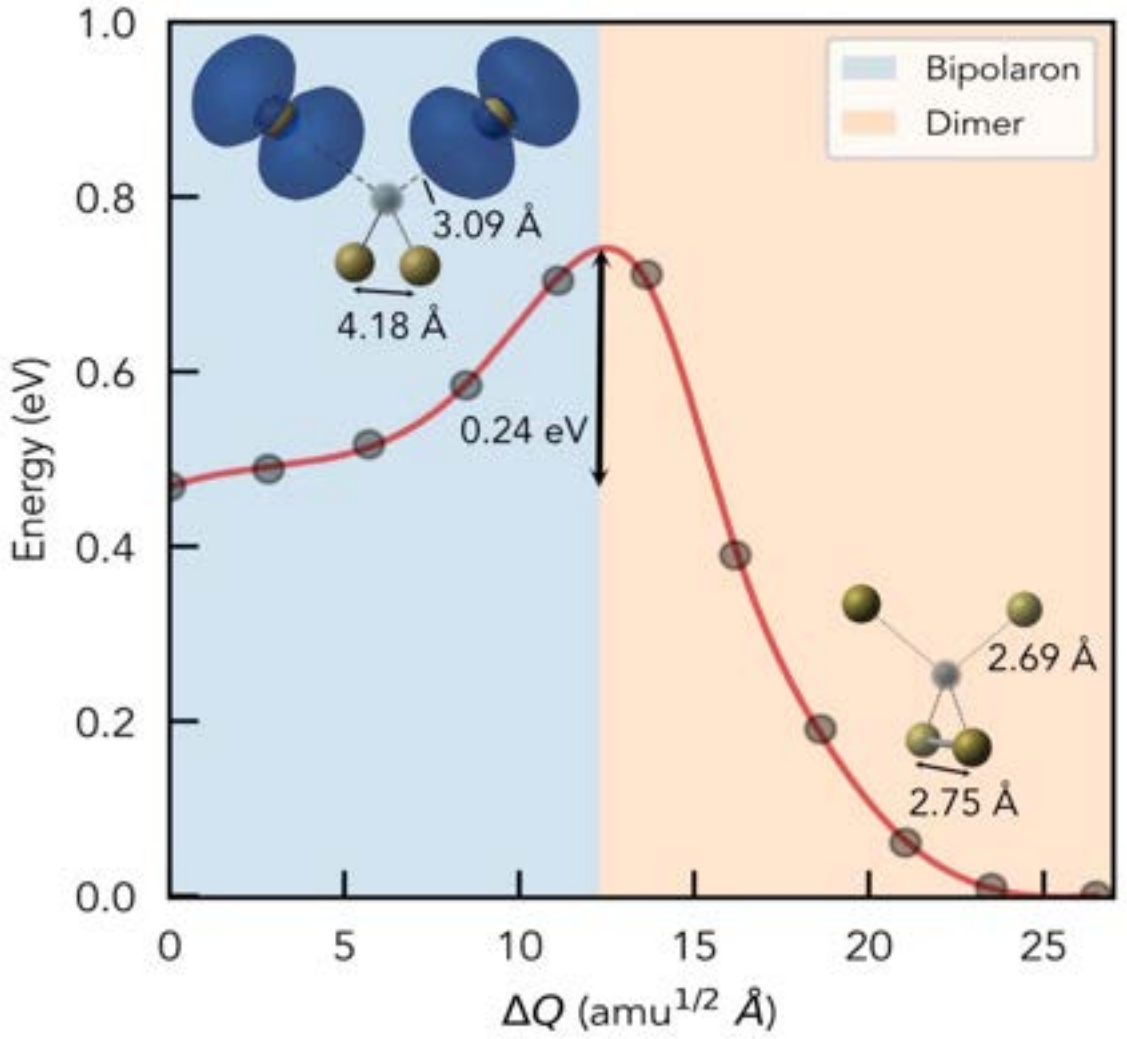




Kavanagh et al. *ACS Energy Lett* **2021**

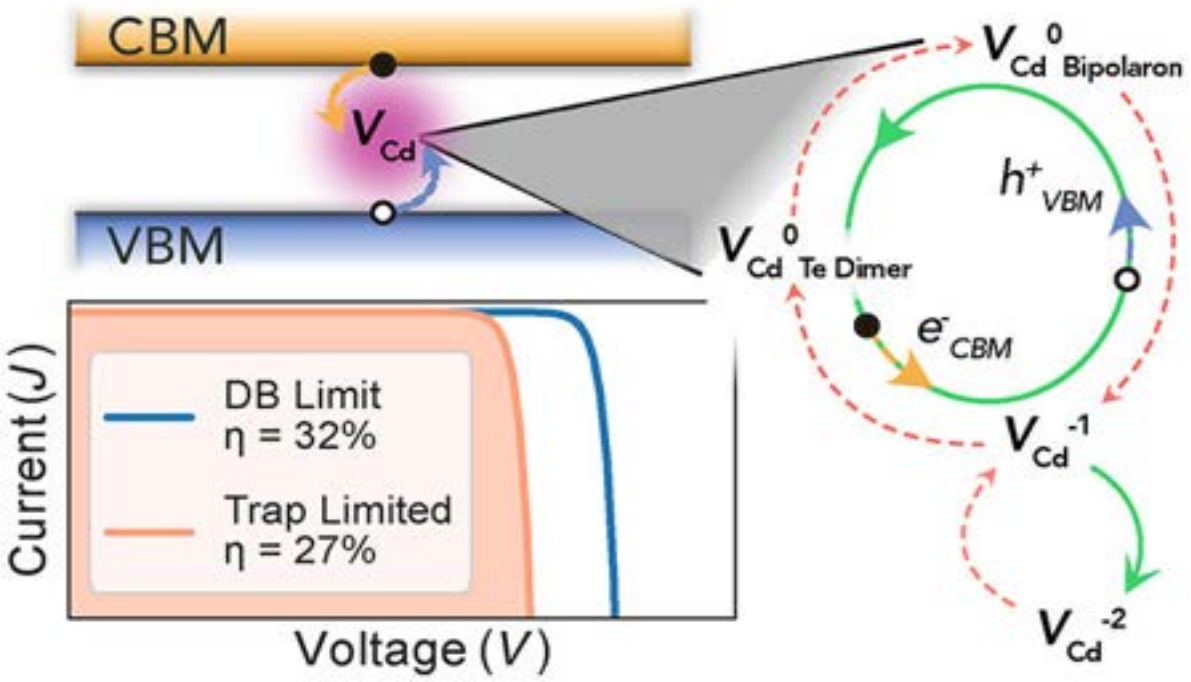
Modelling Electron-Hole Recombination



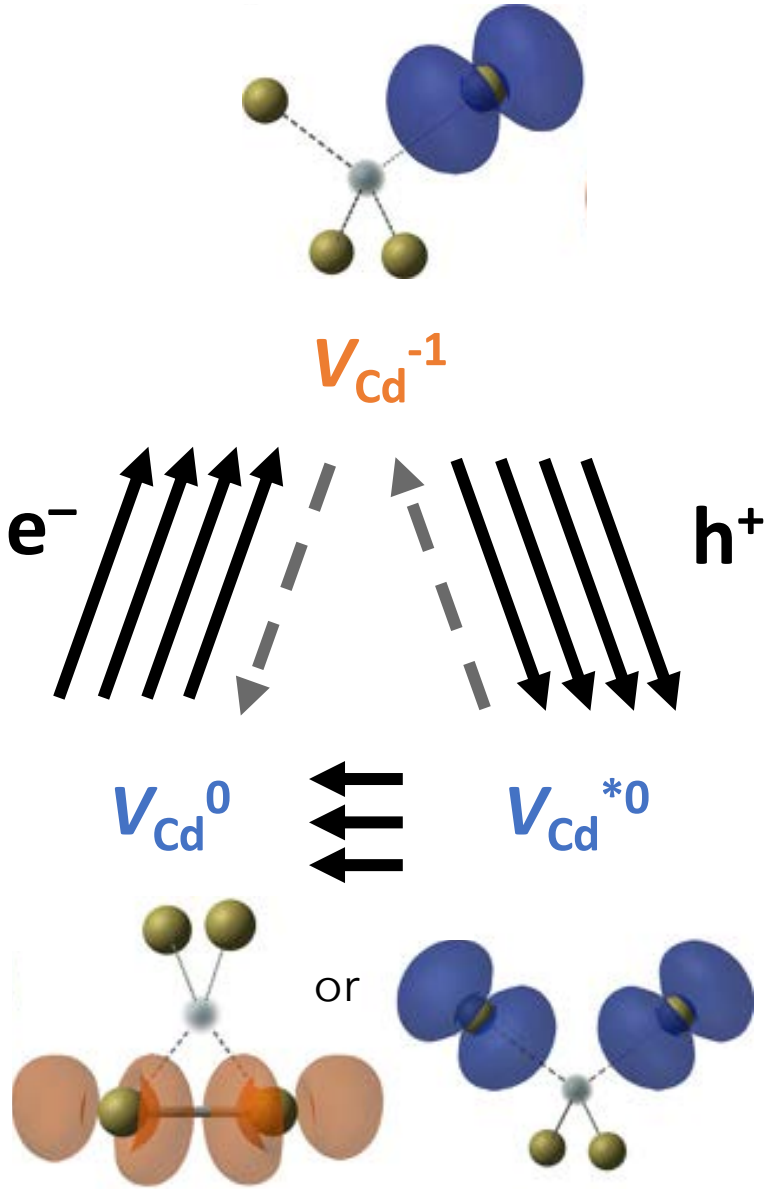


Kavanagh et al. *ACS Energy Lett* **2021**

Modelling Electron-Hole Recombination



ACS Energy Lett 2021



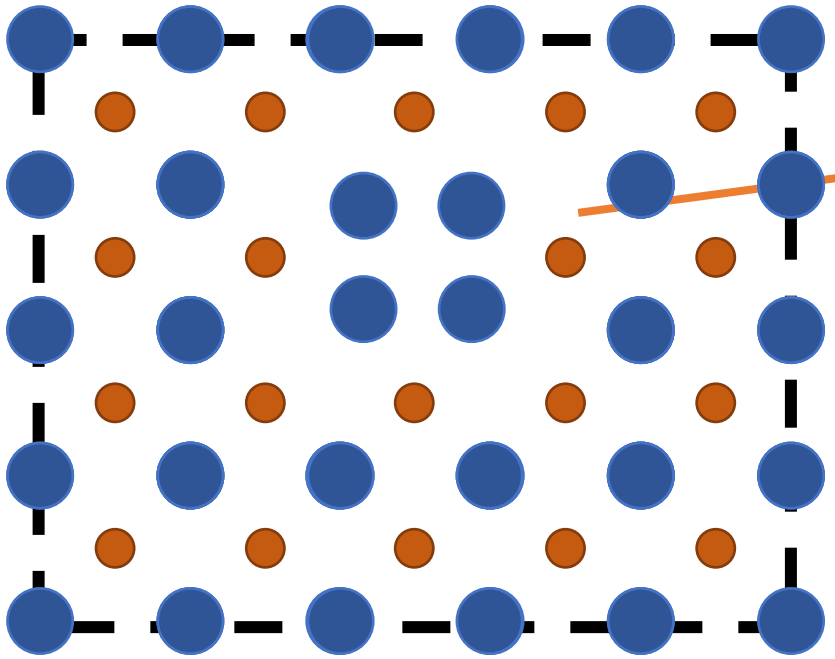
V_{Cd} is a crucial defect to avoid in CdTe!

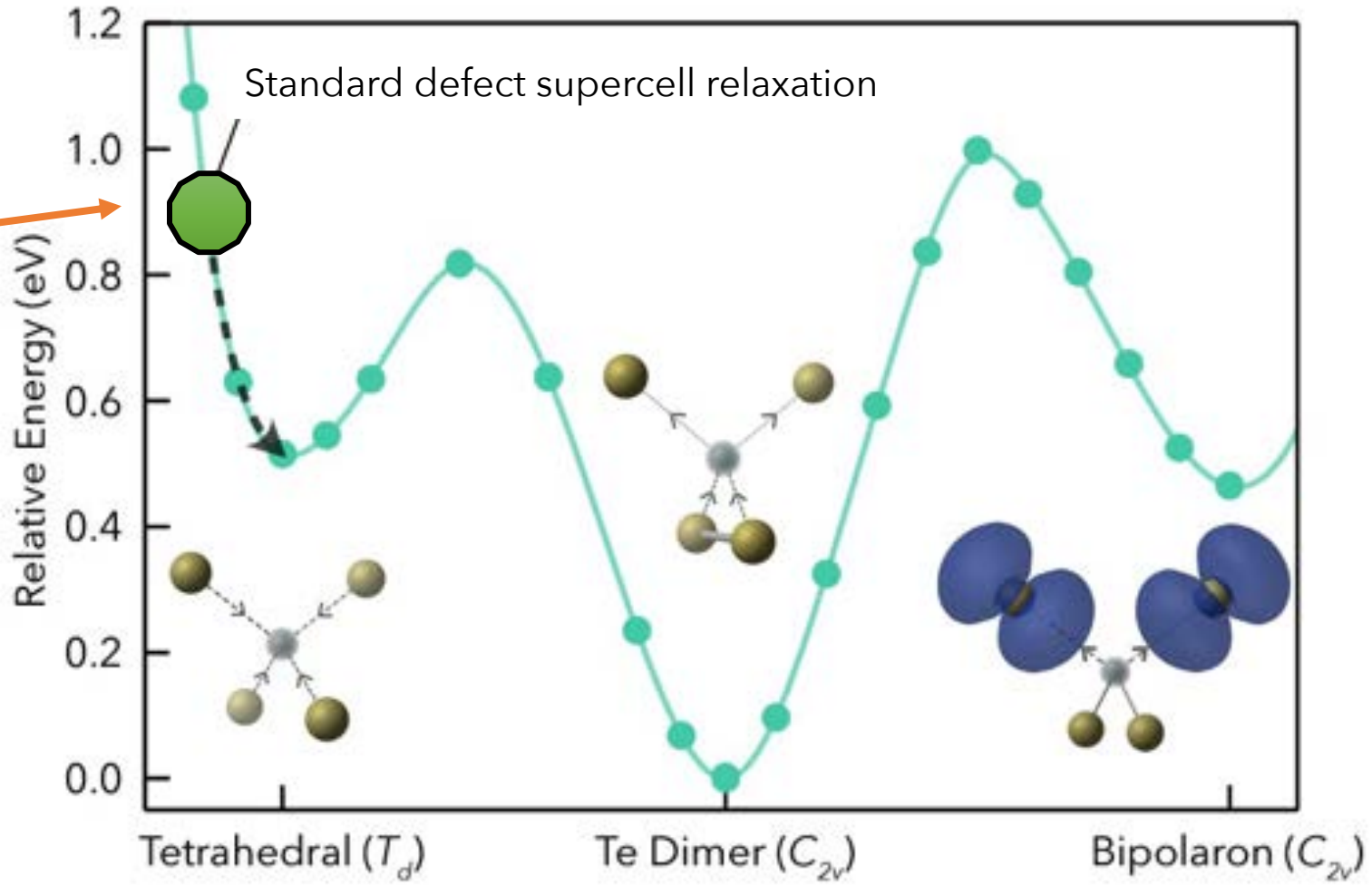
Explains importance of Cl treatment (partially passivates V_{Cd})

Suggests Cd-rich growth conditions are optimal -> Matching current record devices[1]

1. Li et al. *Nature Energy* 2021

Exploring the Energy Landscape

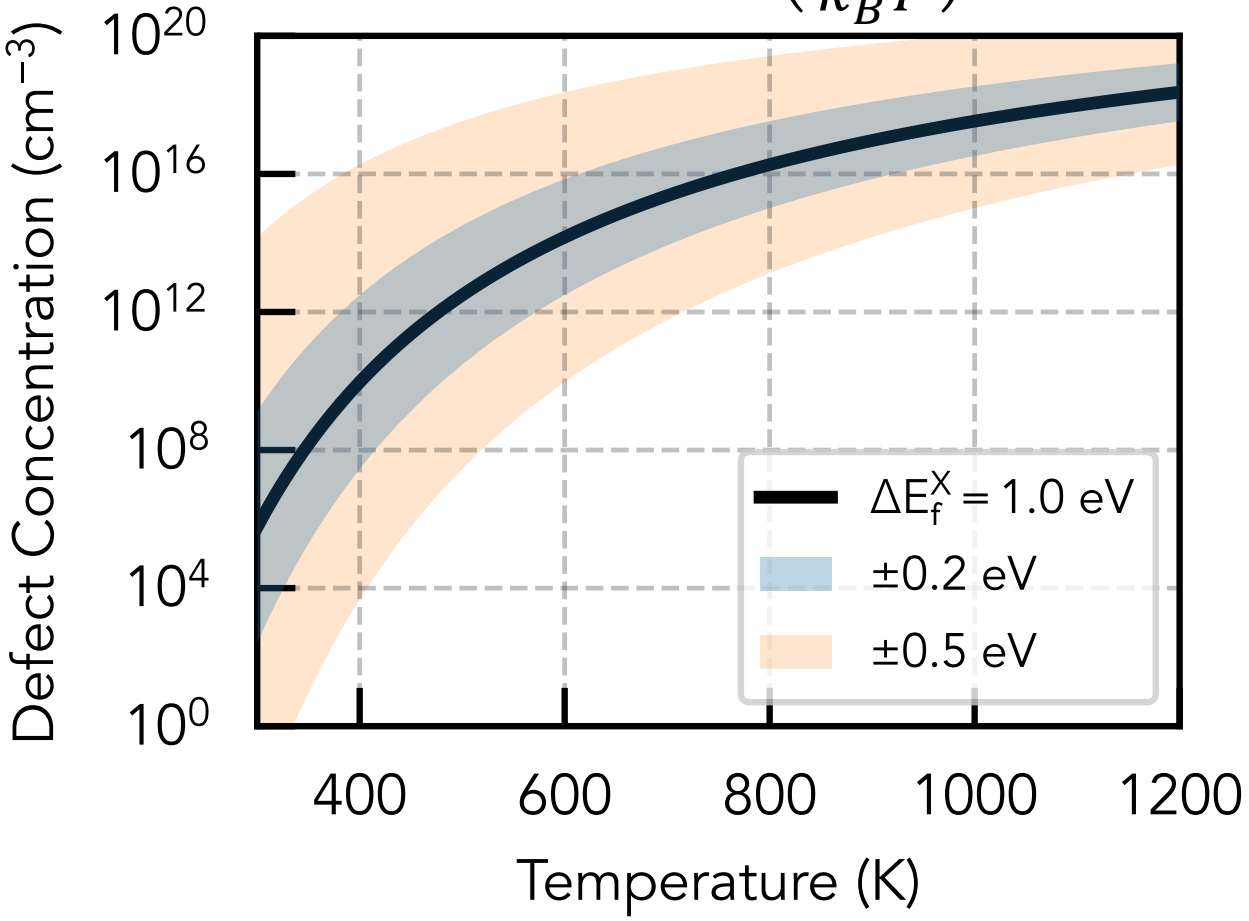




Mosquera-Lois & Kavanagh, *Matter* 2021

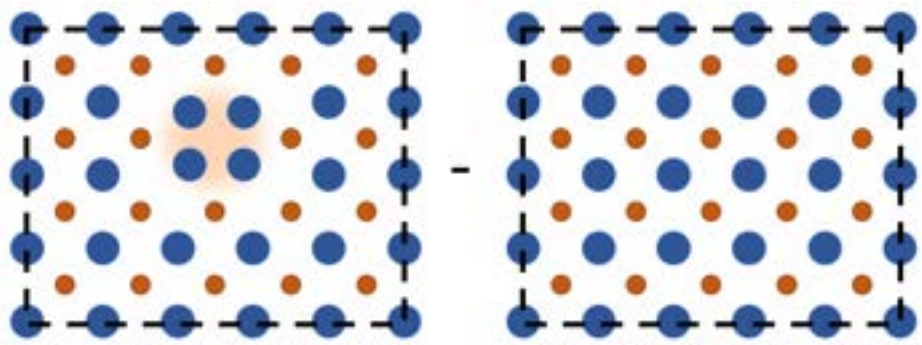
Defect Structure Searching

$$c_X = g\exp\left(\frac{-\Delta H}{k_B T}\right)$$

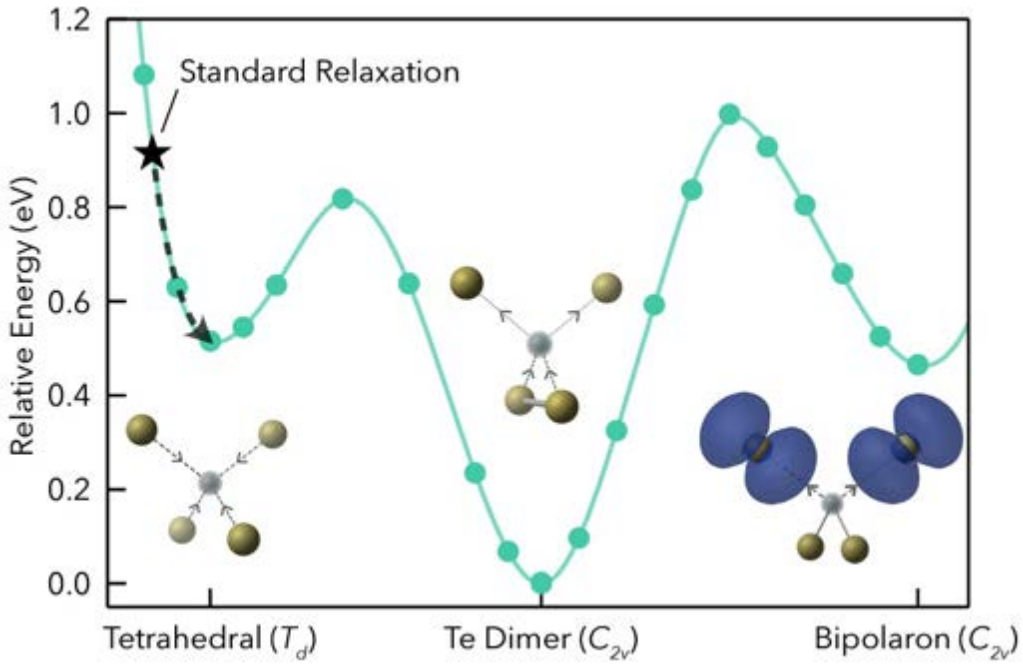


Defect supercell

Bulk supercell



$= \Delta H_d$



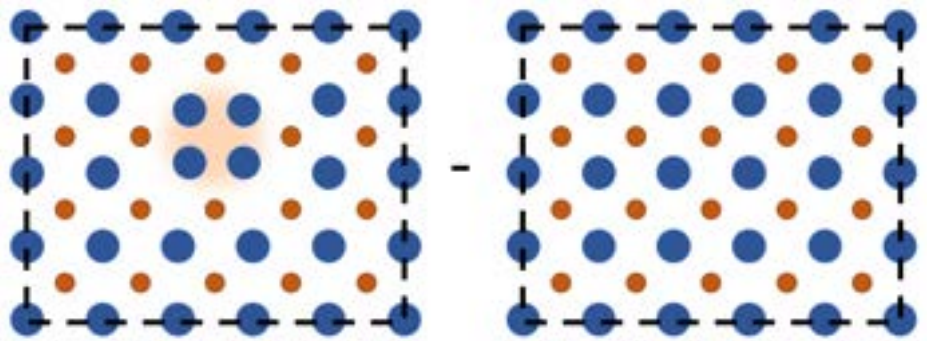
Mosquera-Lois, Kavanagh et al. *Chem Soc Rev* 2023

Defect Structure Searching

Structure determines:

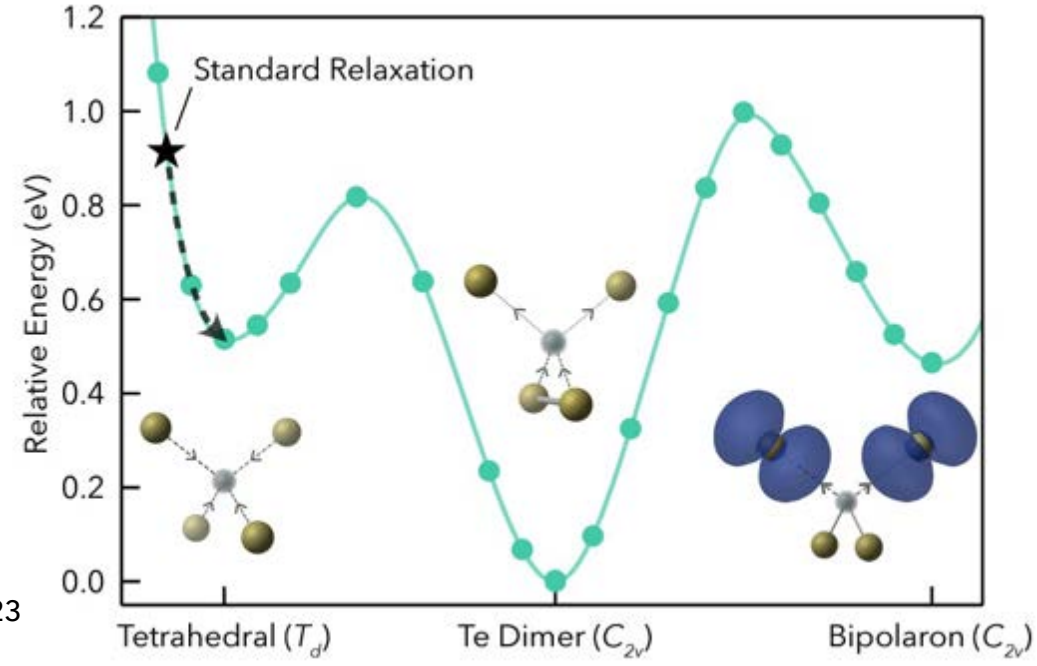
- Formation energies
  - ➢ Concentrations
  - ➢ Doping
  - ➢ Thermodynamics…
- Localisation (deep vs shallow)
- Recombination activity
- Migration
- Catalytic activity
- …

Defect supercell Bulk supercell



$= \Delta H_d$

Chadi & Chang *Phys Rev B* 1989
Lany & Zunger *Phys Rev Lett* 2004
Du & Zhang *Phys Rev B* 2005
Morris, Pickard, Needs *Phys Rev B* 2008
Varley et al *J. Phys.: Condens. Matter* 2011
Kehoe, Scanlon, Watson, *Chem Mater* 2011
Krajewska, Kavanagh et al. *Chem Sci* 2021
Osterbacka, Wiktor *J Phys Chem C* 2022
Kononov et al. *J. Phys.: Condens. Matter* 2023
…

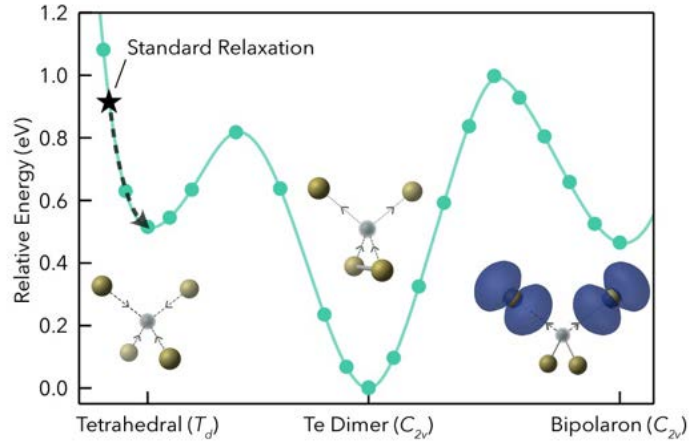


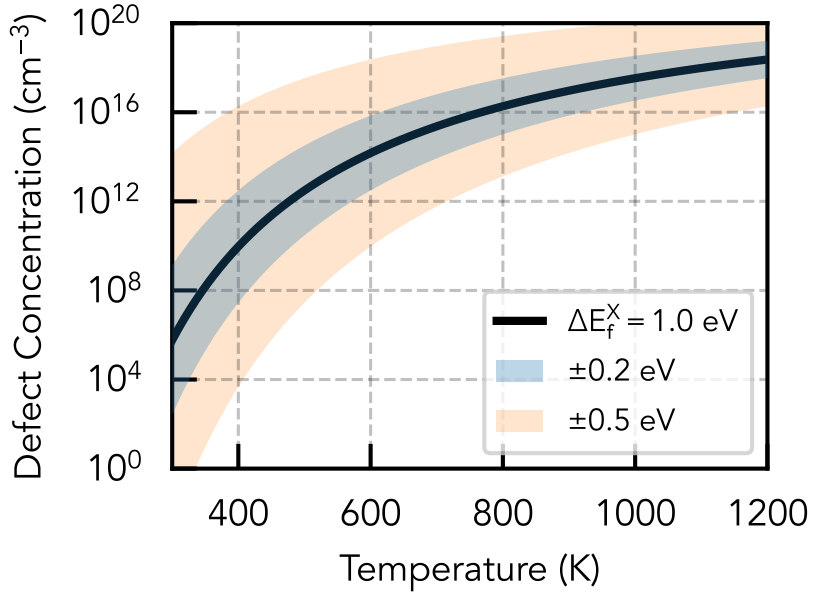
Does it matter?

Structure ➡️ Energy ➡️ Properties (recombination, catalytic activity, diffusion, conductivity…)

Further Examples:

- Gallium vacancies, migration and compensation in Ga_2O_3[1]
- Catalytic activity (divalent metal dopants in CeO_2)[2]
- Persistent Photoconductivity in Si, GaAs DX centres[3,4]
- Hydrogen Complexes in Silicon[5]
- Defect absorption / bandgap lowering (Sn-doped $Cs_3Bi_2Br_9$)[6]
- Oxide polarons (in $BiVO_4$)[7]
- Colour centres and deep anion vacancies in II-VI compounds[8]
- …





1. Varley et al *J. Phys.: Condens. Matter* **2011**
2. Kehoe, Scanlon, Watson, *Chem Mater* **2011**
3. Du & Zhang *Phys Rev B* **2005**
4. Chadi & Chang *Phys Rev B* **1989**
5. Morris, Pickard, Needs *Phys Rev B* **2008**
6. Krajewska, Kavanagh et al. *Chem Sci* **2021**
7. Osterbacka, Ambrosio, Wiktor *J Phys Chem C* **2022**
8. Lany & Zunger *Phys Rev Lett* **2004**

ShakeNBreak

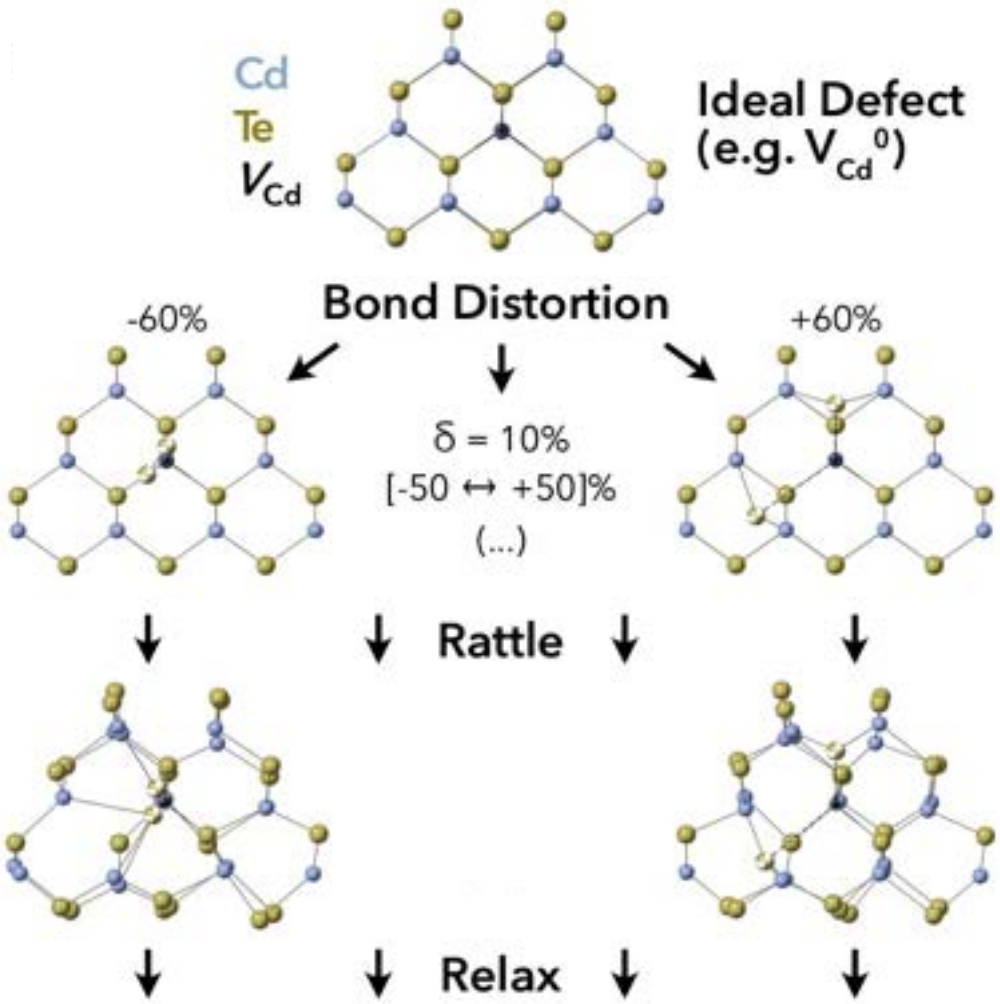
Idea: Leverage the localised "molecule-in-a-solid" behaviour of point defects:

- Chemically-guided neighbour bond distortions:
  No. distorted bonds = Δ{Valence Electrons}

- Stretch/compress neighbour bonds (±50% range)
  ➡️ Distortion mesh of trial structures

- 'Rattle': Add small random displacements to break symmetry and aid location of global minimum

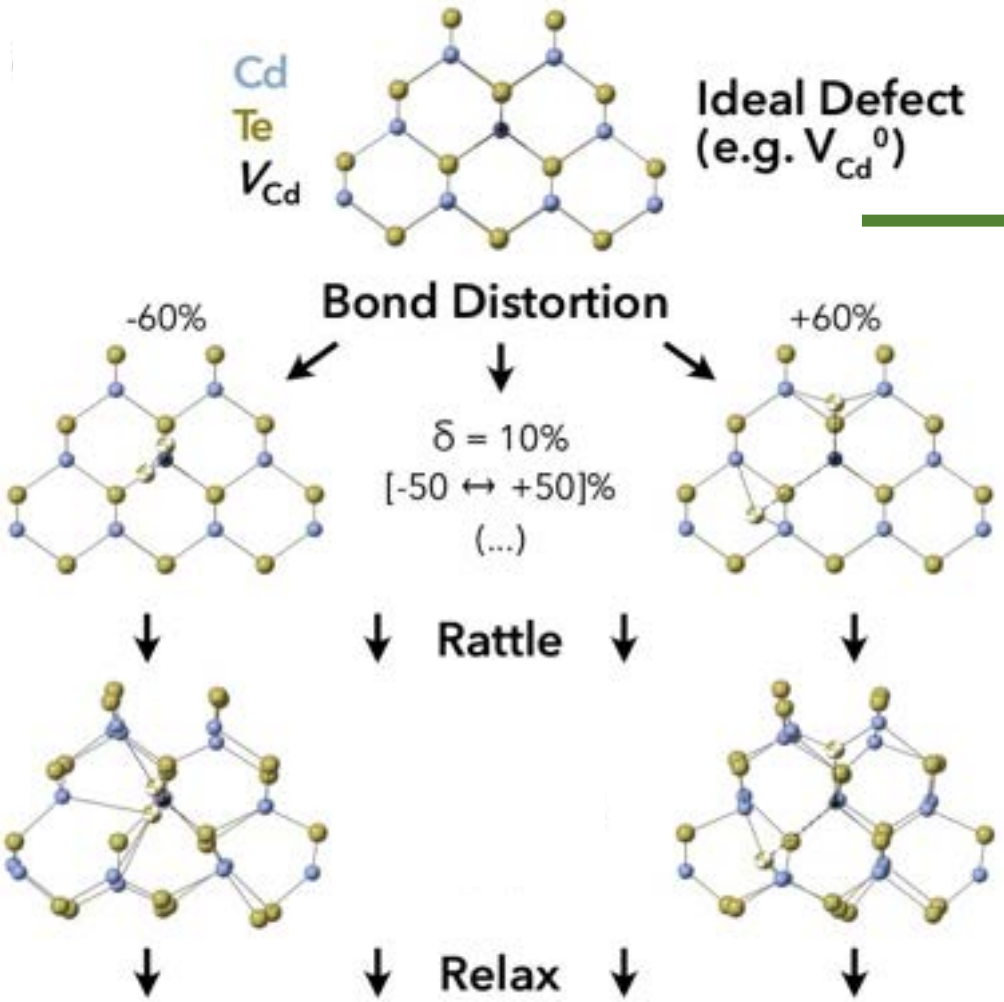
- -> Relax



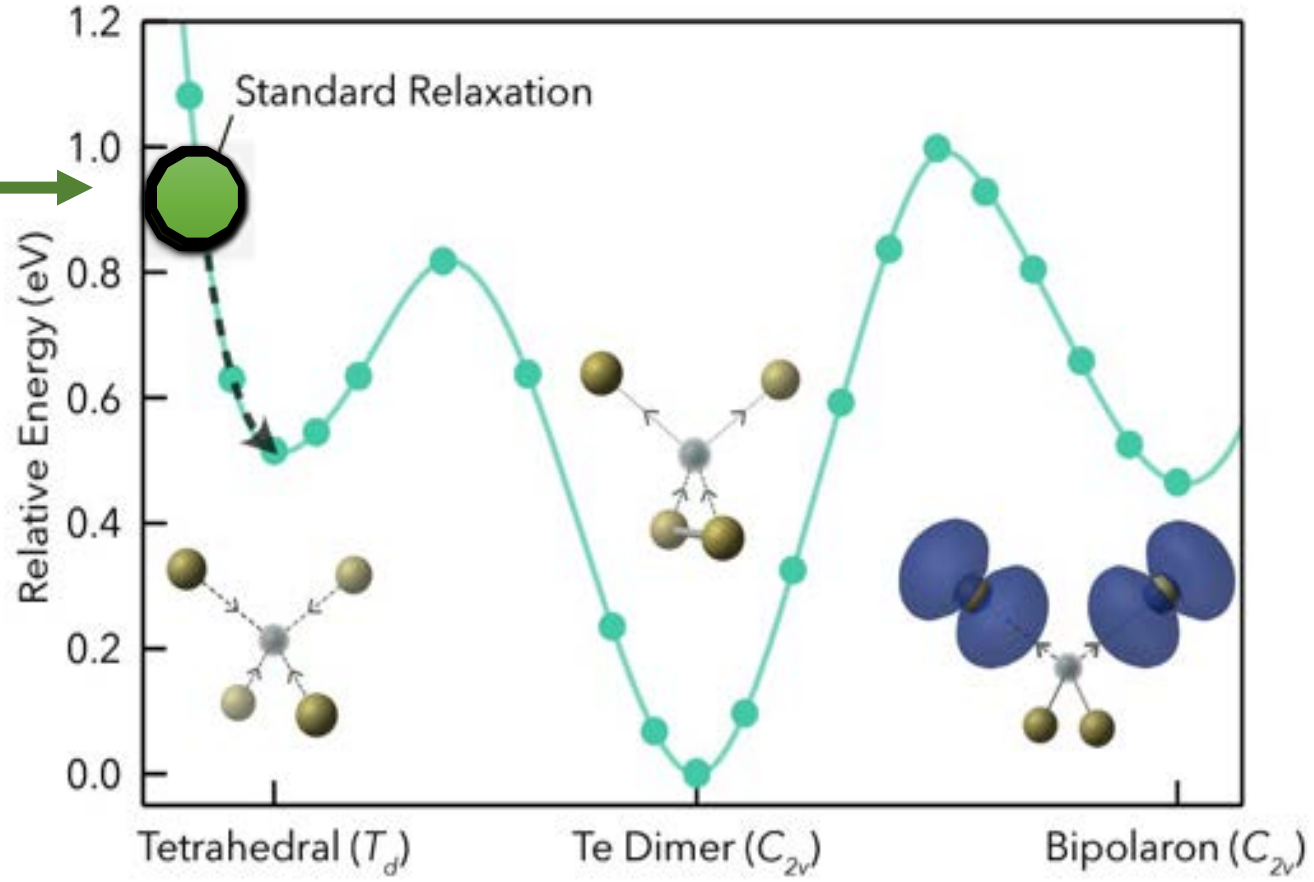
Mosquera-Lois et al. *npj Comput Mater* 2023
Mosquera-Lois et al. *J. Open Source Software* 2022

Pham & Deskins JCTC 2020

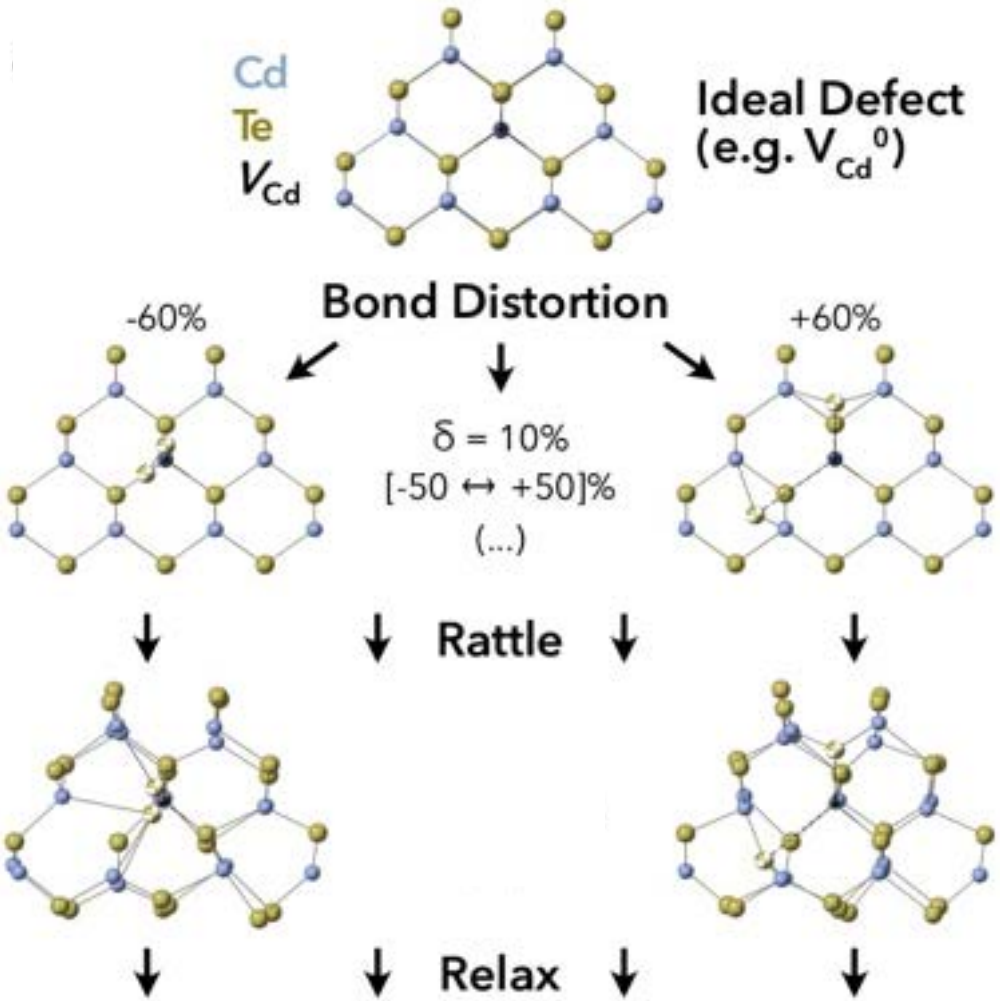
ShakeNBreak



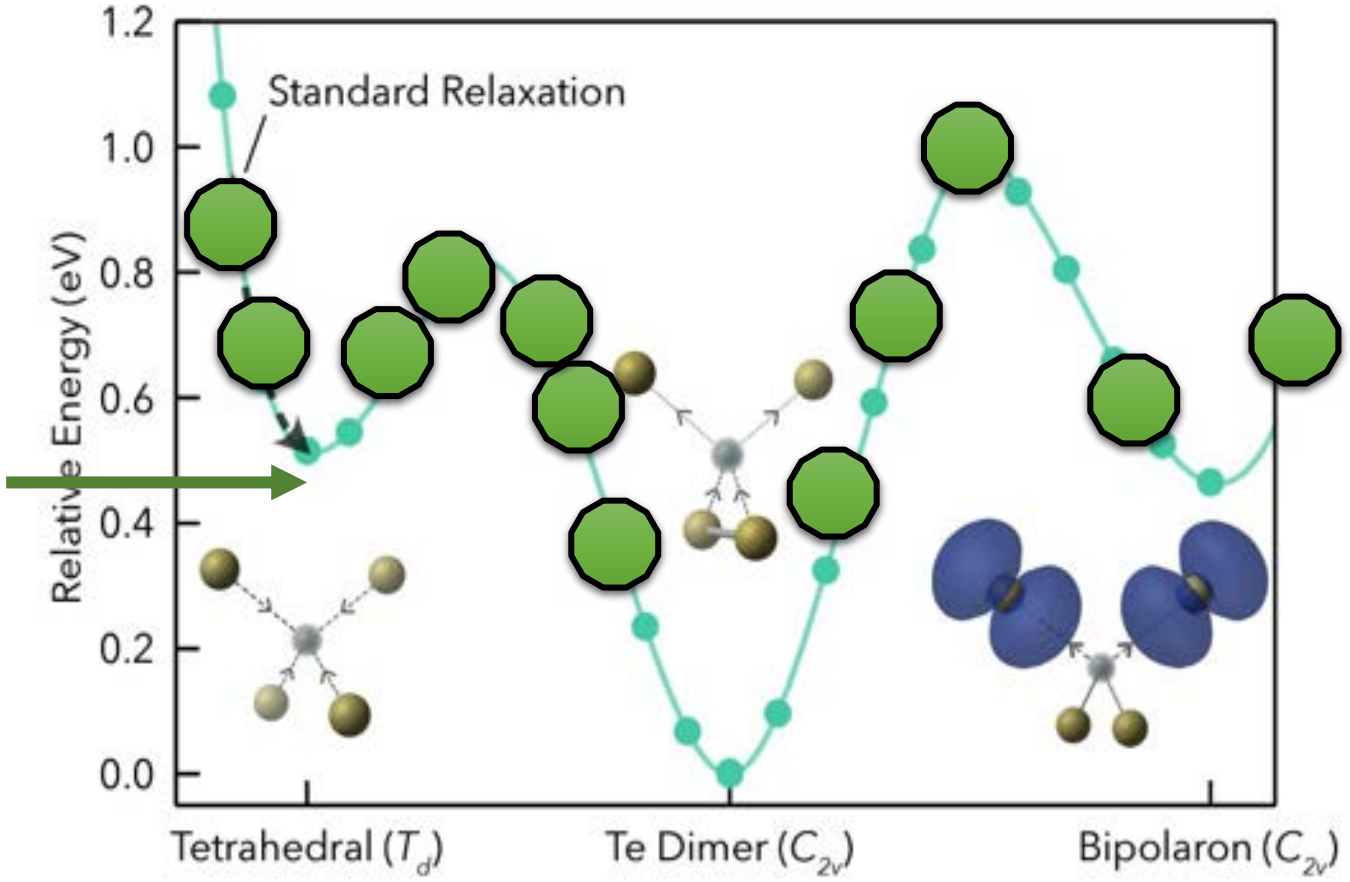
11 relaxations with **Γ**-only sampling



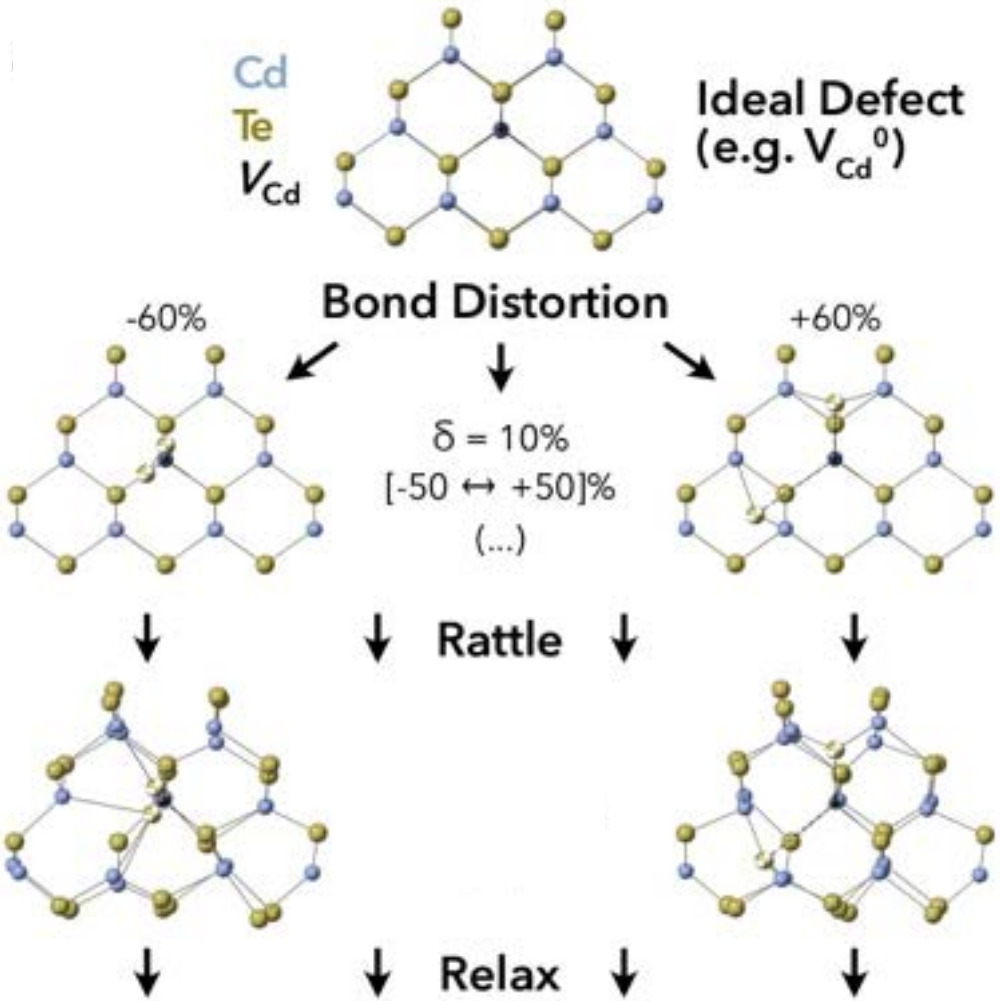
ShakeNBreak



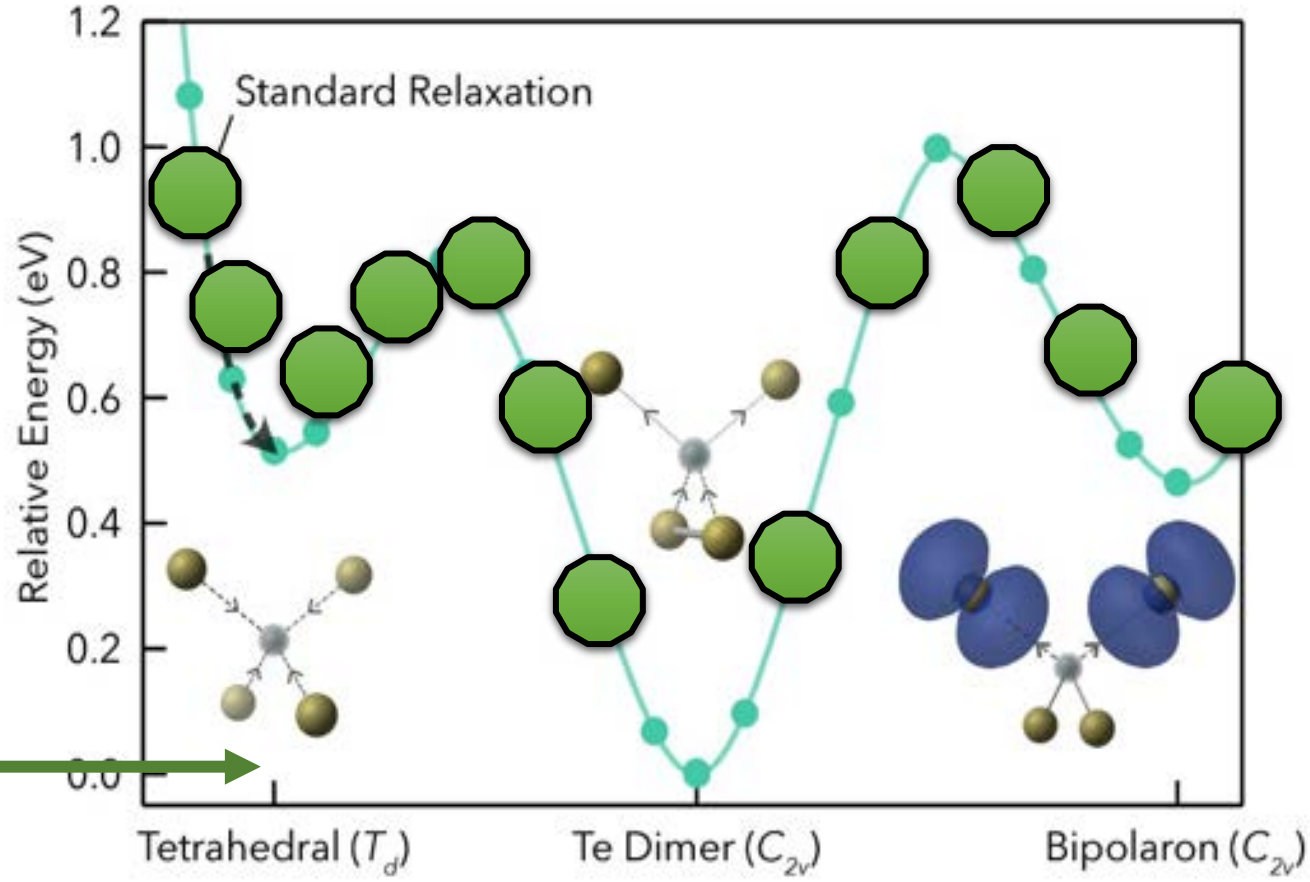
11 relaxations with **Γ**-only sampling



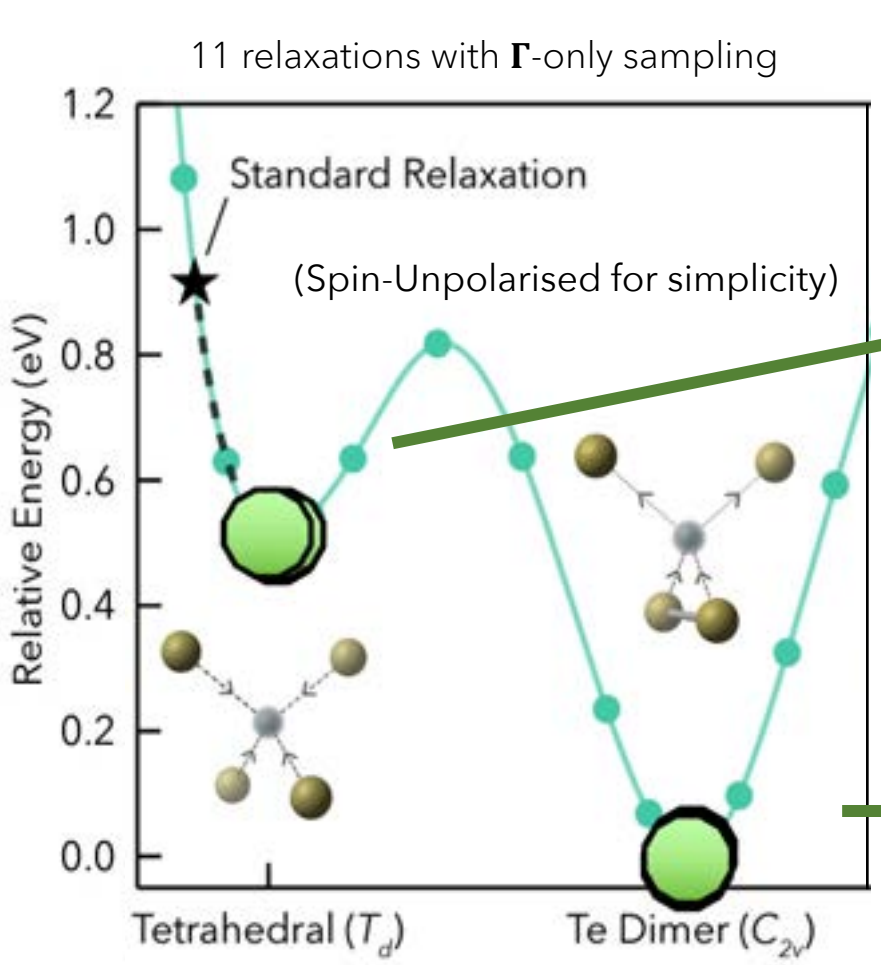
ShakeNBreak

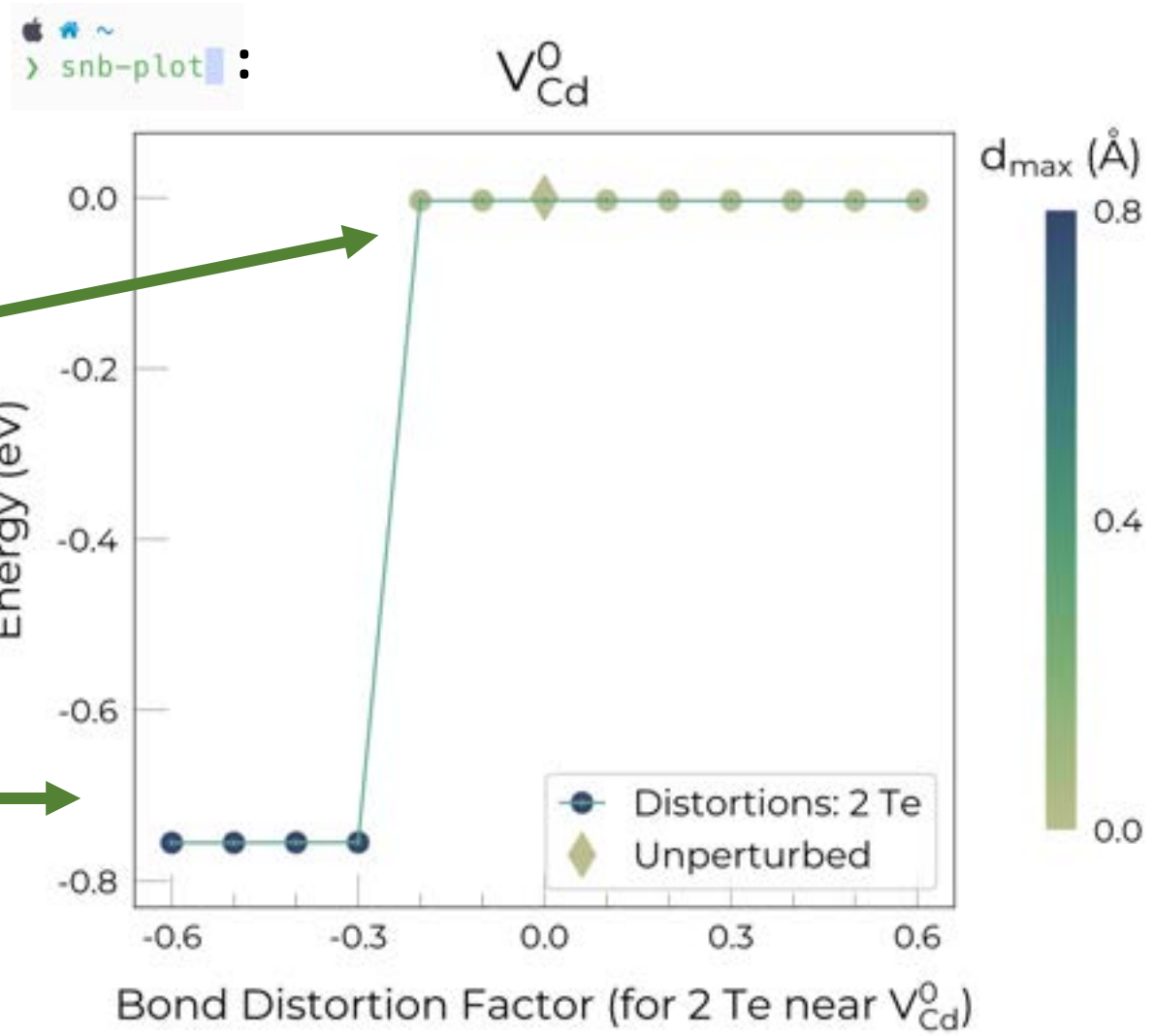


11 relaxations with **Γ**-only sampling



ShakeNBreak





ShakeNBreak

Successfully reproduces all previously-reported cases (Benchmarks: Si, CdTe, GaAs, CeO_2, ZnO…)

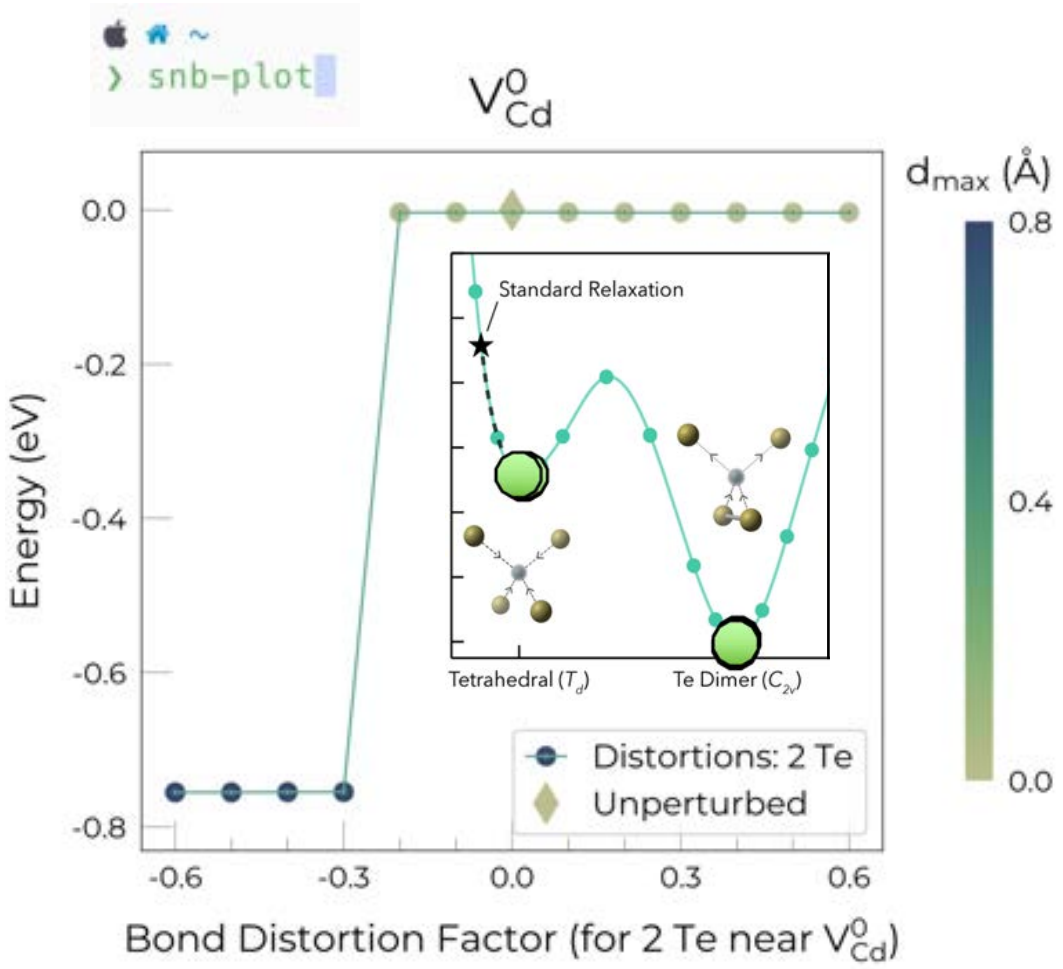
Energy-lowering reconstructions identified in a diverse range of materials & defects – found for defects in <u>every</u> material studied ($Sb_2(S/Se)_3$, In_2O_3, TiO_2, Se, $SrTiO_3$ …)

Can locate low-energy metastable structures

➡️ Important for diffusion (transition states) and carrier recombination

Efficient (<10% computational cost of full defect study)

Automated *user-friendly* (Python API or CLI), trivially parallel…



Mosquera-Lois et al. *npj Comput Mater* 2023
Mosquera-Lois et al. *J. Open Source Software* 2022

ShakeNBreak: Key Examples

Battery degradation processes in Ni-rich cathodes

[Murdock et al. *Adv Mater* 2024, Squires et al. *ACS Energy Lett* 2024, Cen et al. *JMCA* 2023…]

Oxygen vacancy reconstructions in $SrTiO_3$, key to photocatalyst performance

[Ogawa et al. *In Submission*]

Un-stable *polarons* in $CuSbSe_2$

[Lohan et al. *Nature Comms* 2025]

Charge compensation in transparent conducting oxides

[Li et al. *Chem Mater* 2024, Cai et al. *APL* 2025…]

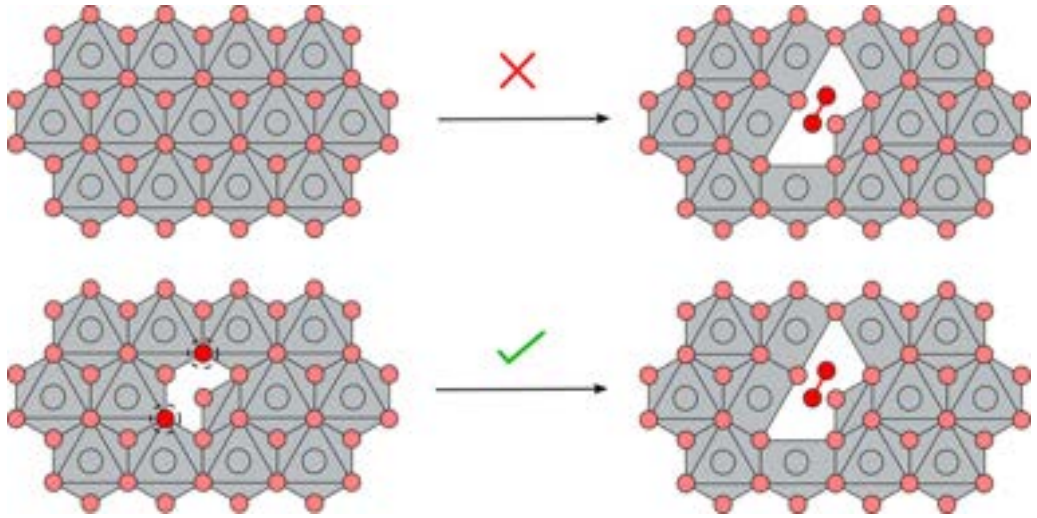
Extreme charge compensation, defect metastability & impact on electron-hole recombination in antimony chalcogenide solar cells ($Sb_2(S/Se)_3$)

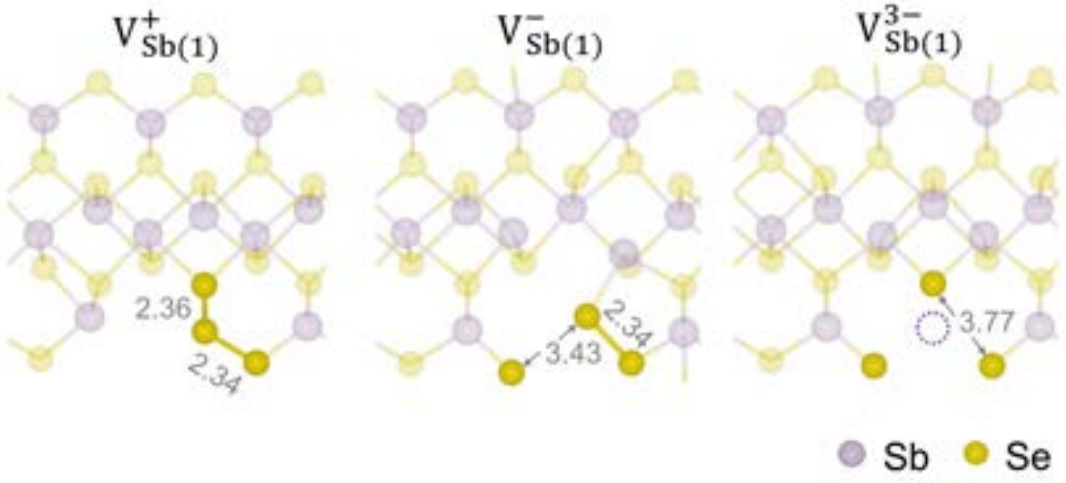
[Wang et al. *PCCP* 2022, *ACS Energy Lett* 2022, *Phys Rev B* 2023, *Joule* 2024, *ACS Energy Lett* 2024…]

Quantum defects in CaO

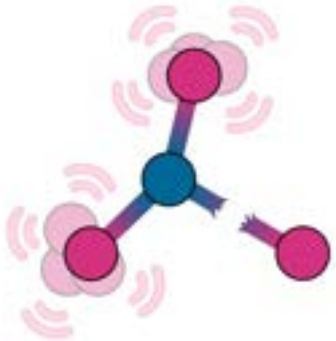
[Yuan & Hautier *APL* 2024]

…





Structural Reconstructions at Defects



Search ⌘ + K

Welcome to ShakeNBreak!

Usage

- Installation
- Python API
- Tutorials
- Miscellaneous Tips & Tricks

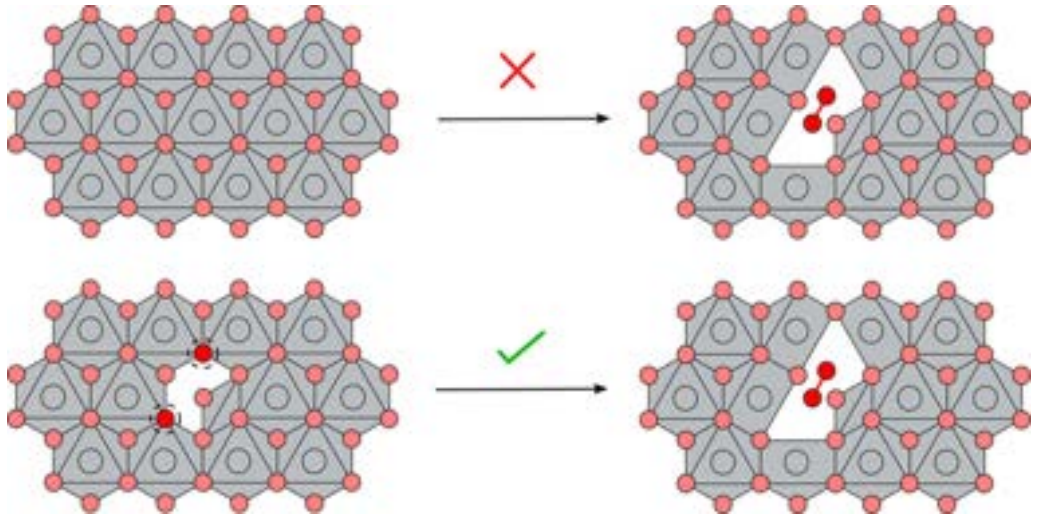
Information

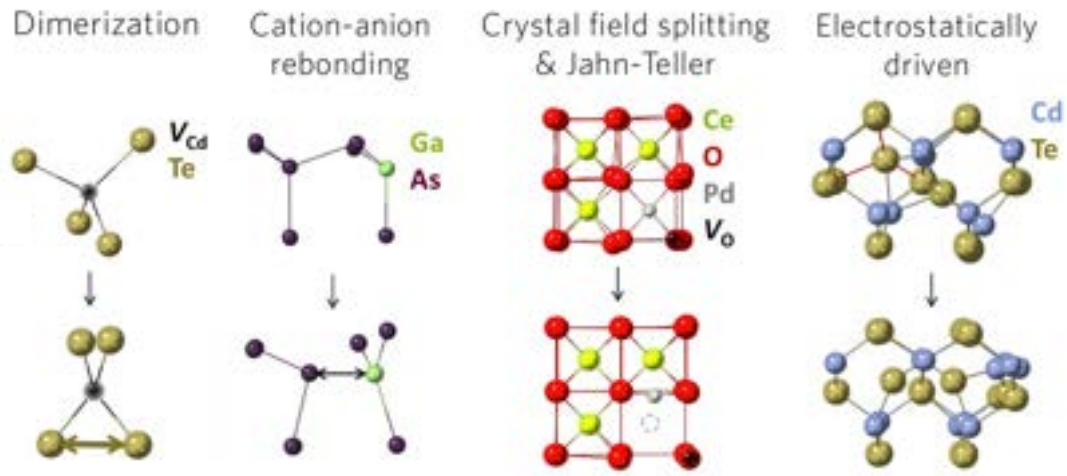
- Code Compatibility
- Contributing
- Change Log
- ShakeNBreak on GitHub

Studies using ShakeNBreak

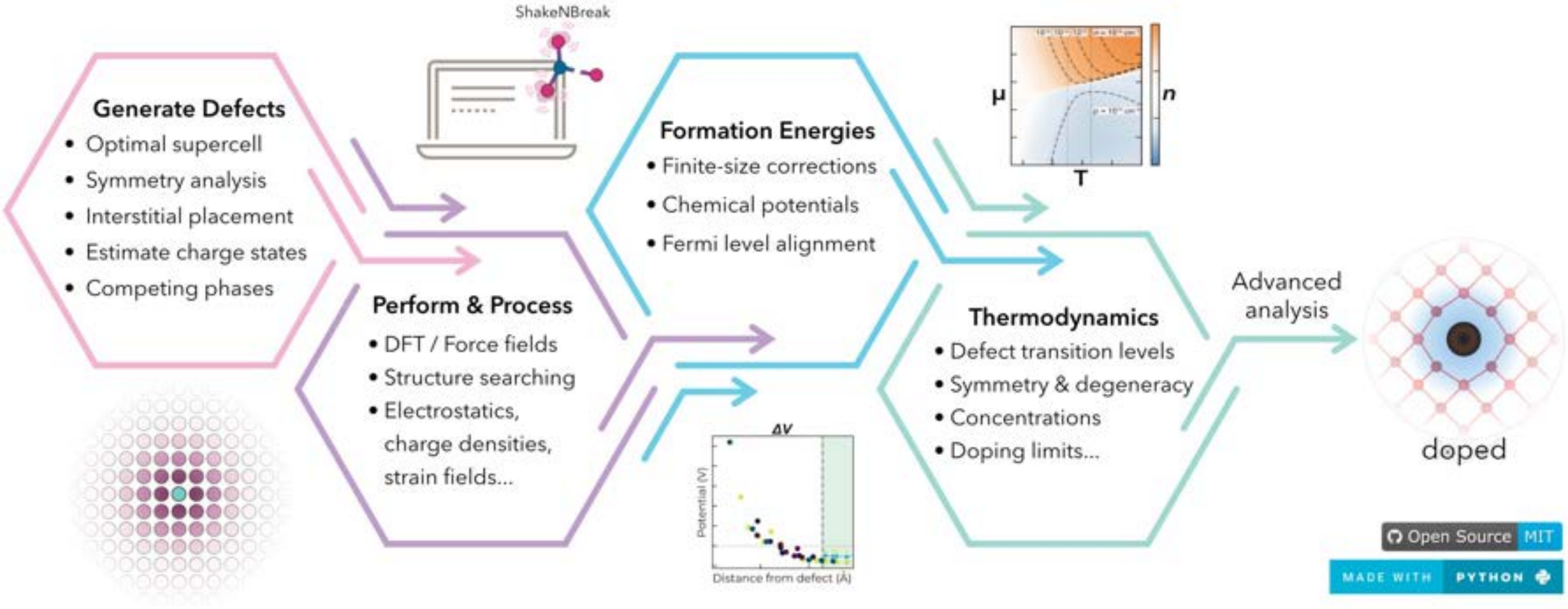
- Y. Fu & H. Lohan et al. **Factors Enabling Delocalized Charge-Carriers in Pnictogen-Based Solar Absorbers: In-depth Investigation into $CuSbSe_2$** Nature Communications 2025
- S. R. Kavanagh **Identifying Split Vacancies with Foundation Models and Electrostatics** arXiv 2025
- S. R. Kavanagh et al. **Intrinsic point defect tolerance in selenium for indoor and tandem photovoltaics** ChemRxiv 2025
- J. Hu et al. **Enabling ionic transport in Li_3AlP_2 the roles of defects and disorder** Journal of Materials Chemistry A 2025
- Z. Cai & C. Ma **Origin of oxygen partial pressure-dependent conductivity in $SrTiO_3$** Applied Physics Letters 2025
- W. D. Neilson et al. **Oxygen Potential, Uranium Diffusion, and Defect Chemistry in $UO_{2\pm x}$: A Density Functional Theory Study** Journal of Physical Chemistry C 2024
- X. Wang et al. **Sulfur vacancies limit the open-circuit voltage of Sb_2S_3 solar cells** ACS Energy Letters 2024
- Z. Yuan & G. Hautier **First-principles study of defects and doping limits in CaO** Applied Physics Letters 2024
- B. E. Murdock et al. **Li-Site Defects Induce Formation of Li-Rich Impurity Phases: Implications for Charge Distribution and Performance of $LiNi_{0.5-x}M_xMn_{1.5}O_4$ Cathodes (M = Fe and Mg; x = 0.05–0.2)** Advanced Materials 2024
- A. G. Squires et al. **Oxygen dimerization as a defect-driven process in bulk $LiNiO_2$** ACS Energy Letters 2024
- X. Wang et al. **Upper efficiency limit of Sb_2Se_3 solar cells** Joule 2024
- I. Mosquera-Lois et al. **Machine-learning structural reconstructions for accelerated point defect calculations** npj Computational Materials 2024
- S. R. Kavanagh et al. **doped: Python toolkit for robust and repeatable charged defect supercell calculations** Journal of Open Source Software 2024
- K. Li et al. **Computational Prediction of an Antimony-based n-type Transparent Conducting Oxide: F-doped Sb_2O_5** Chemistry of Materials 2024
- S. Hachmioune et al. **Exploring the Thermoelectric Potential of MgB_4: Electronic Band Structure, Transport Properties, and Defect Chemistry** Chemistry of Materials 2024
- X. Wang et al. **Four-electron negative-U vacancy defects in antimony selenide** Physical Review B 2023
- Y. Kumagai et al. **Alkali Mono-Pnictides: A New Class of Photovoltaic Materials by Element Mutation** PRX Energy 2023

...





doped: Calculating Defects in Solids

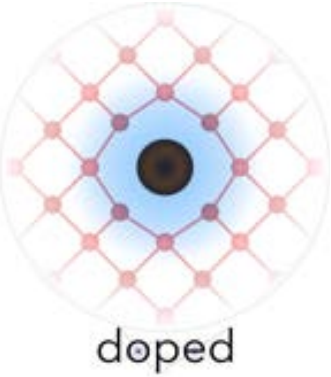


🔗 doped.readthedocs.io

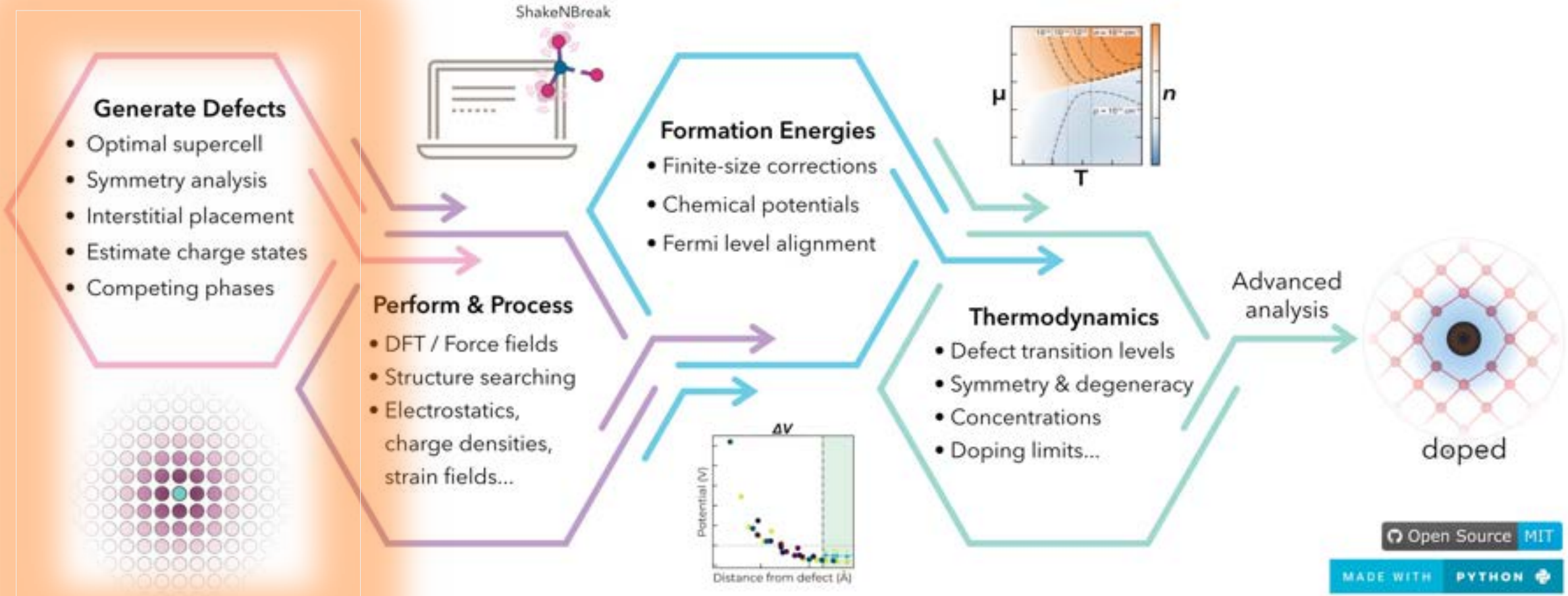
Kavanagh et al. *J Open Source Softw* 2024

doped: Design Philosophy

- Wide-spanning functionality
- Reasonable defaults, but with full control/flexibility for the user
- User-friendly, extensive tutorials/documentation, automated compatibility/sanity checks
- Aid and encourage reproducibility
- Computational efficiency through intelligent algorithms
- Publication-ready outputs



doped: Calculating Defects in Solids



🔗 doped.readthedocs.io

Kavanagh et al. *J Open Source Softw* 2024

doped: Defect Generation

```
from pymatgen.core.structure import Structure
from doped.generation import DefectsGenerator
```

```
# Load our relaxed bulk (host) structure:
relaxed_primitive_CdTe = Structure.from_file("CdTe/relaxed_primitive_POSCAR")

# generate defects:
defect_gen = DefectsGenerator(relaxed_primitive_CdTe)
```

```
Generating DefectEntry objects: 100.0%|██████████| [00:05,  19.58it/s]
```

Vacancies	Guessed Charges	Conv. Cell Coords	Wyckoff
v_Cd	[-2,-1,0,+1]	[0.000,0.000,0.000]	4a
v_Te	[-1,0,+1,+2]	[0.250,0.250,0.250]	4c

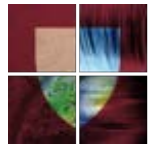
Substitutions	Guessed Charges	Conv. Cell Coords	Wyckoff
Cd_Te	[0,+1,+2,+3,+4]	[0.250,0.250,0.250]	4c
Te_Cd	[-4,-3,-2,-1,0,+1,+2]	[0.000,0.000,0.000]	4a

Interstitials	Guessed Charges	Conv. Cell Coords	Wyckoff
Cd_i_C3v	[0,+1,+2]	[0.625,0.625,0.625]	16e
Cd_i_Td_Cd2.83	[0,+1,+2]	[0.750,0.750,0.750]	4d
Cd_i_Td_Te2.83	[0,+1,+2]	[0.500,0.500,0.500]	4b
Te_i_C3v	[-2,-1,0,+1,+2,+3,+4]	[0.625,0.625,0.625]	16e
Te_i_Td_Cd2.83	[-2,-1,0,+1,+2,+3,+4]	[0.750,0.750,0.750]	4d
Te_i_Td_Te2.83	[-2,-1,0,+1,+2,+3,+4]	[0.500,0.500,0.500]	4b

- Candidate interstitial sites generated with Voronoi tessellation (most reliable approach)[1] – or user-specified.
- Auto-generated optimal supercell, or user-provided.
- Generates all intrinsic (and extrinsic if specified) point defects.
- Fully flexible

1. Kononov *et al* 2023 *J. Phys.: Condens. Matter* **35** 334002

doped: Defect Generation



- **Target Supercell**: You may want to use a specific supercell for your calculations, perhaps to reproduce or build on calculations from a previous study with a specific supercell.

```
# run this to see the function documentation:
DefectsGenerator?
```

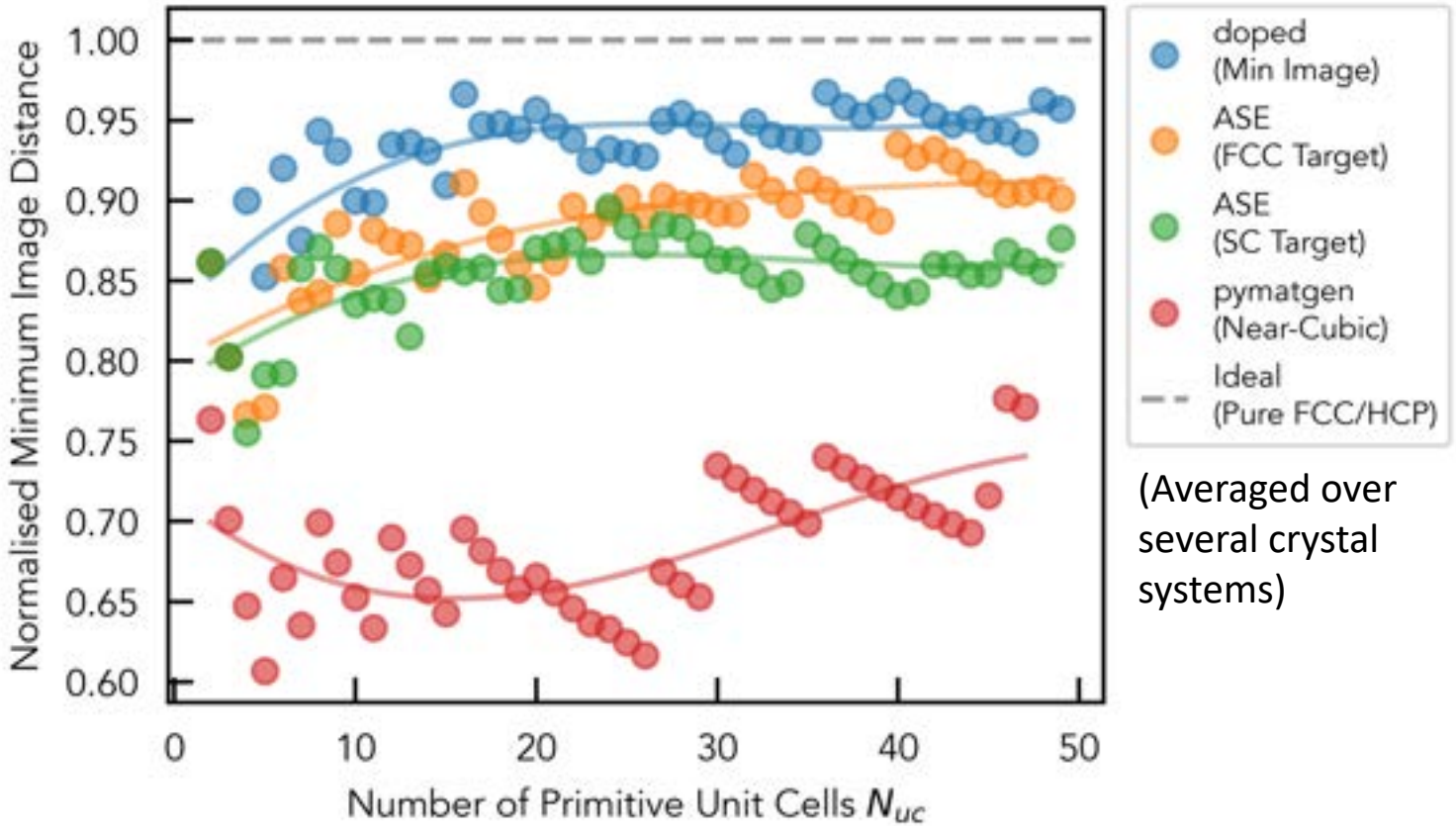
```
Init signature:
DefectsGenerator(
    structure: pymatgen.core.structure.Structure,
    extrinsic: Union[str, List, Dict, NoneType] = None,
    interstitial_coords: Optional[List] = None,
    generate_supercell: bool = True,
    charge_state_gen_kwargs: Optional[Dict] = None,
    supercell_gen_kwargs: Optional[Dict] = None,
    interstitial_gen_kwargs: Optional[Dict] = None,
    target_frac_coords: Optional[List] = None,
    processes: Optional[int] = None,
)
Docstring:      Class for generating doped DefectEntry objects.
Init docstring:
Generates doped DefectEntry objects for defects in the input host
structure. By default, generates all intrinsic defects, but extrinsic
defects (impurities) can also be created using the `extrinsic`
argument.

Interstitial sites are generated using Voronoi tessellation by default (found
to be the most reliable), which can be controlled using the
`interstitial_gen_kwargs` argument (passed as keyword arguments to the
`VoronoiInterstitialGenerator` class). Alternatively, a list of interstitial
sites can be manually specified using the `interstitial_coords` argument.
```

- Candidate interstitial sites generated with Voronoi tessellation (most reliable approach)[1] – or user-specified.
- Auto-generated optimal supercell, or user-provided.
- Generates all intrinsic (and extrinsic if specified) point defects.
- Fully flexible

1. Kononov *et al* 2023 *J. Phys.: Condens. Matter* **35** 334002

doped: Supercell Generation

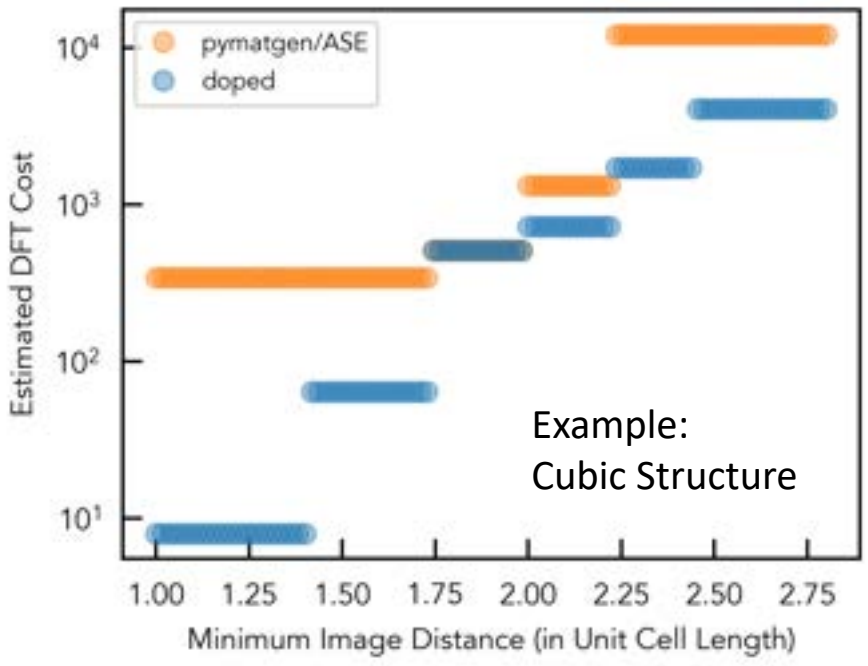
Efficient, direct optimization of image distance over all possible supercell transformations, and accounting for rotational invariances



(Averaged over several crystal systems)

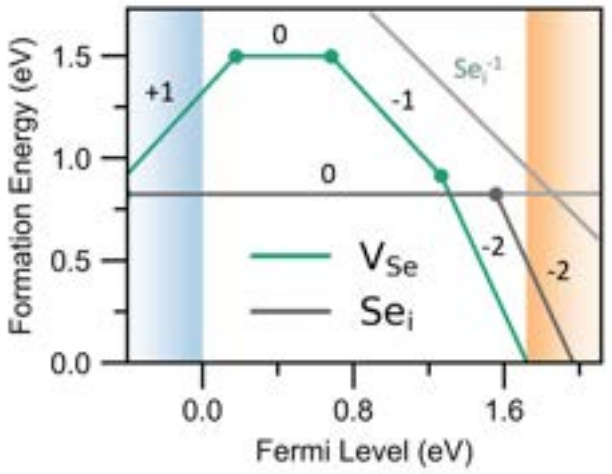
➡️ minimal DFT/computational cost

Mean improvements of ~10% compared to custom ASE scans
-> ~35% DFT cost reduction



doped: Charge State Estimation

- Defects can adopt various charge states in materials – but which ones will *actually be stable?*
- False positives = Charge states included, but not stable (*bad, but inevitable*)
- False negatives = Not included, but actually stable (*very bad*)




```python
from pymatgen.core.structure import Structure
from doped.generation import DefectsGenerator
```

```python
# Load our relaxed bulk (host) structure:
relaxed_primitive_CdTe = Structure.from_file("CdTe/relaxed_primitive_POSCAR")

# generate defects:
defect_gen = DefectsGenerator(relaxed_primitive_CdTe)
```

```
Generating DefectEntry objects: 100.0%|██████████| [00:05,  19.58it/s]
```

Vacancies	Guessed Charges	Conv. Cell Coords	Wyckoff
v_Cd	[-2,-1,0,+1]	[0.000,0.000,0.000]	4a
v_Te	[-1,0,+1,+2]	[0.250,0.250,0.250]	4c

Substitutions	Guessed Charges	Conv. Cell Coords	Wyckoff
Cd_Te	[0,+1,+2,+3,+4]	[0.250,0.250,0.250]	4c
Te_Cd	[-4,-3,-2,-1,0,+1,+2]	[0.000,0.000,0.000]	4a

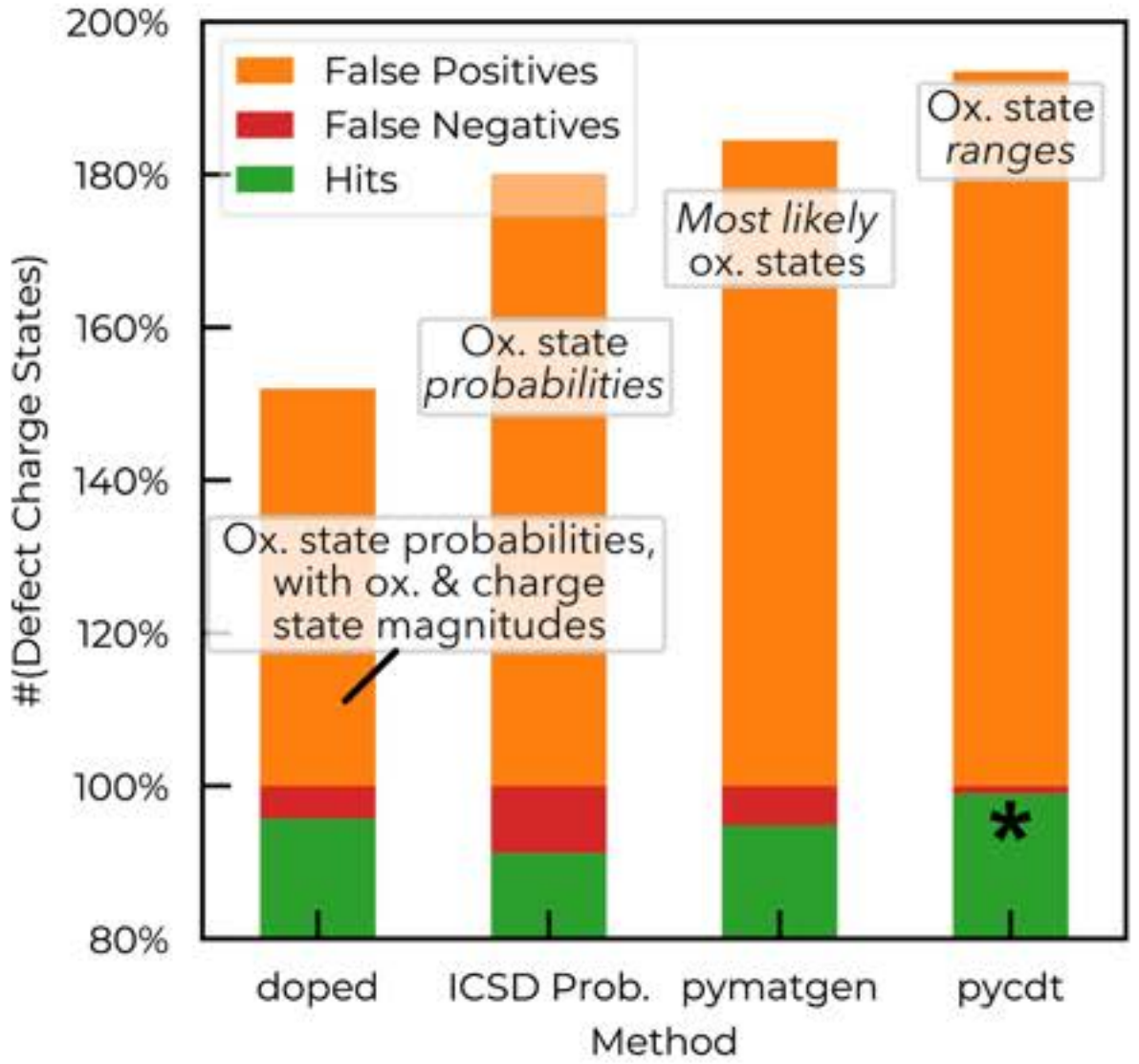
Interstitials	Guessed Charges	Conv. Cell Coords	Wyckoff
Cd_i_C3v	[0,+1,+2]	[0.625,0.625,0.625]	16e
Cd_i_Td_Cd2.83	[0,+1,+2]	[0.750,0.750,0.750]	4d
Cd_i_Td_Te2.83	[0,+1,+2]	[0.500,0.500,0.500]	4b
Te_i_C3v	[-2,-1,0,+1,+2,+3,+4]	[0.625,0.625,0.625]	16e
Te_i_Td_Cd2.83	[-2,-1,0,+1,+2,+3,+4]	[0.750,0.750,0.750]	4d
Te_i_Td_Te2.83	[-2,-1,0,+1,+2,+3,+4]	[0.500,0.500,0.500]	4b

doped: Charge State Estimation

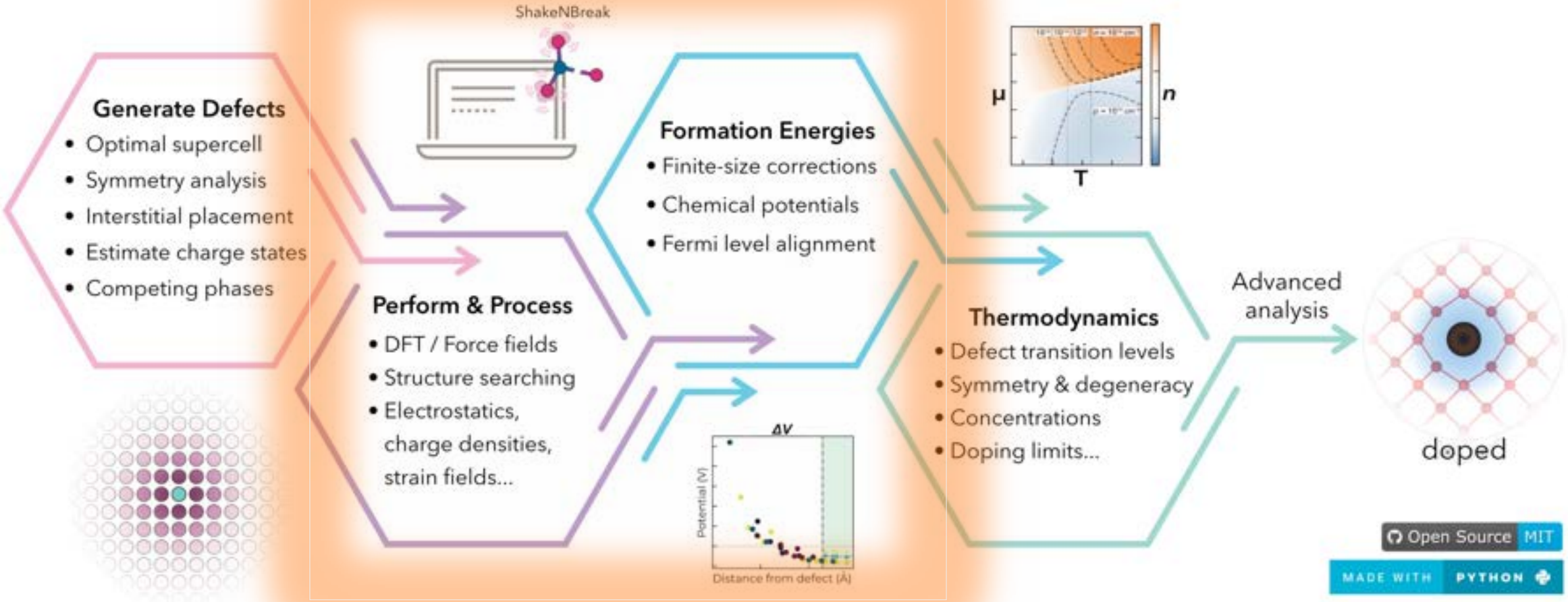
- Defects can adopt various charge states in materials – but which ones will *actually be stable?*
- False positives = Charge states included, but not stable (*bad, but inevitable*)
- False negatives = Not included, but actually stable (*very bad*)
- As always, fully tunable (probability threshold to control lean/completeness) & flexible

(Intermediate charge states which are metastable should still be calculated and are not considered false positives)

* False negatives underestimated for PyCDT due to bias in dataset



doped: Calculating Defects in Solids



🔗 doped.readthedocs.io

Kavanagh et al. *J Open Source Softw* 2024

doped: Finite-Size Corrections

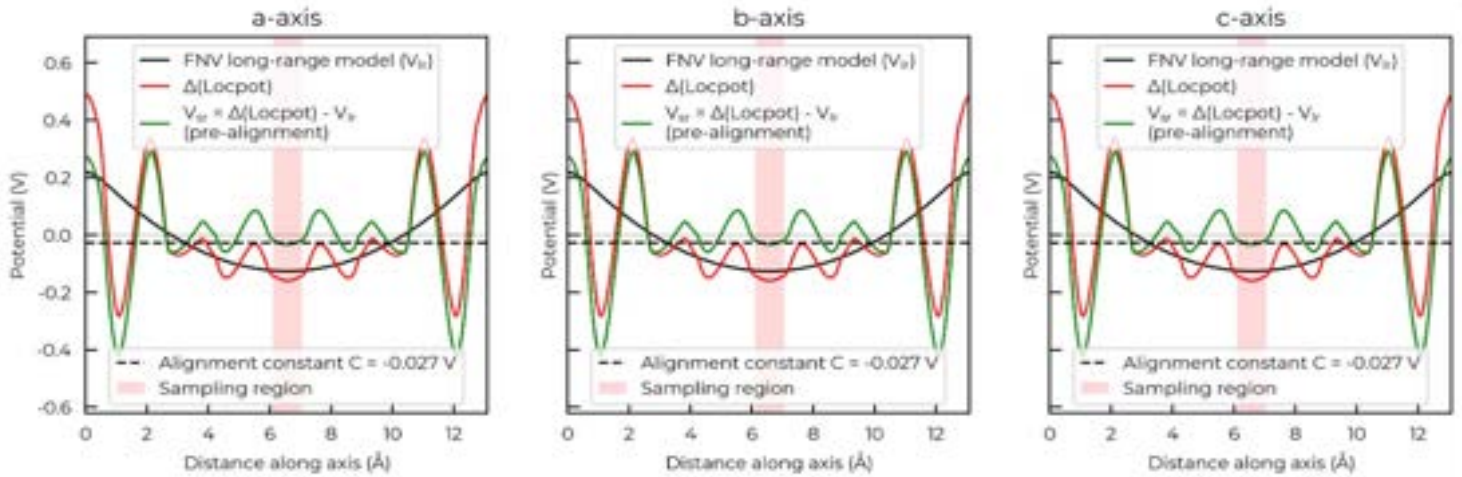
As always, automated but fully flexible/controllable (sampling region, excluded atoms – useful for low-dimensional systems…)

Automated estimation of error in charge correction energy, based on variance of site/electrostatic potential in sampling region

Freysoldt (FNV); isotropic

```
correction, plot = CdTe_defect_dict["v_Cd_-2"].get_freysoldt_correction(plot=True)
```

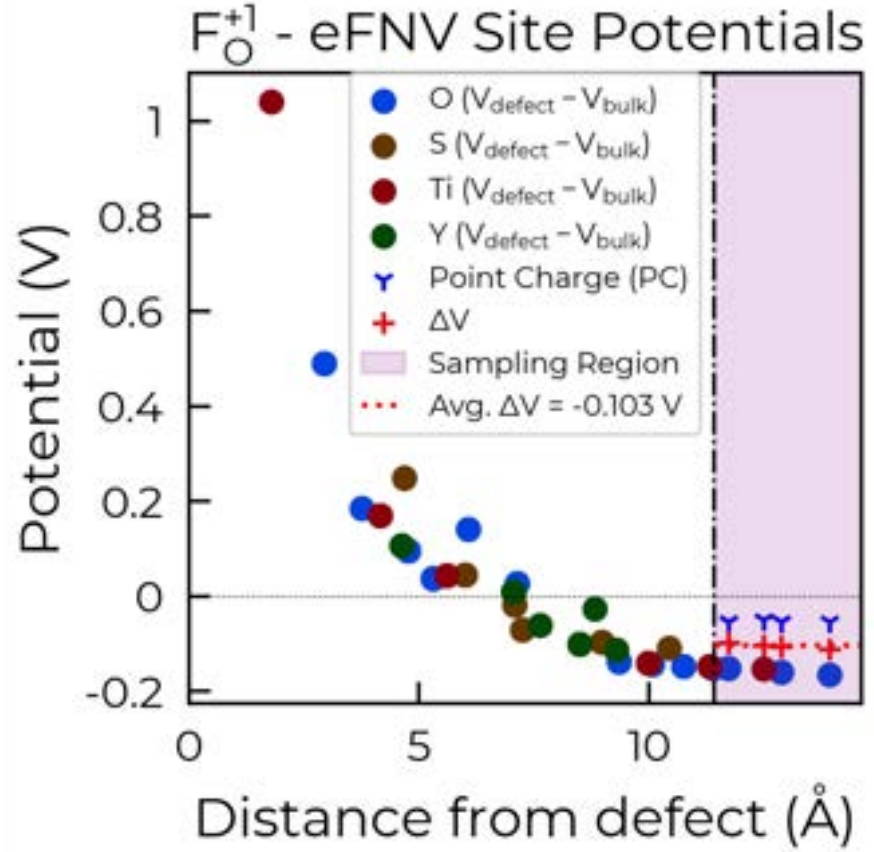
```
Calculated Freysoldt (FNV) correction is 0.738 eV
```



Kumagai-Oba (eFNV); anisotropic

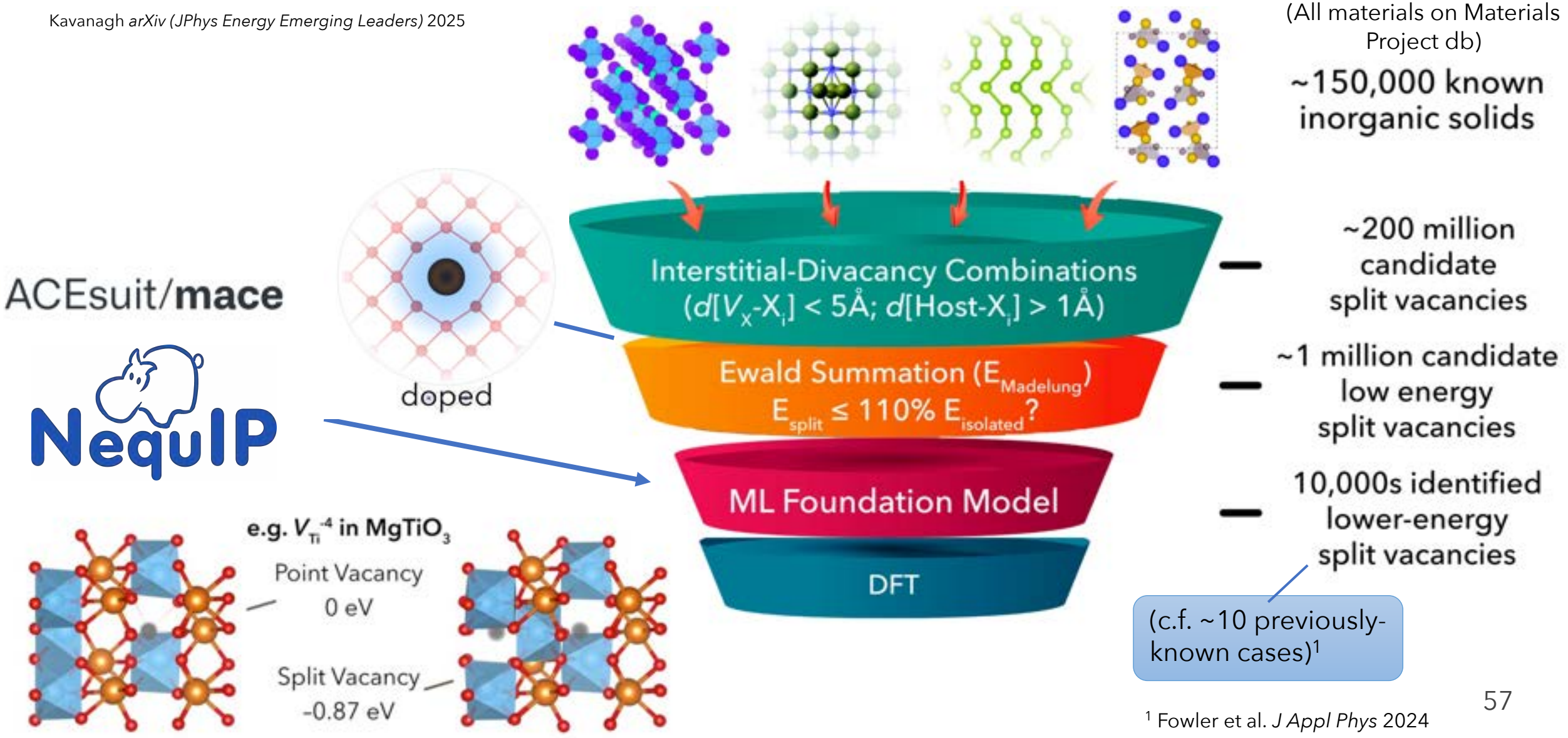
```
correction, plot = F_O_1_entry.get_kumagai_correction(plot=True)
```

```
Calculated Kumagai (eFNV) correction is 0.127 eV
```



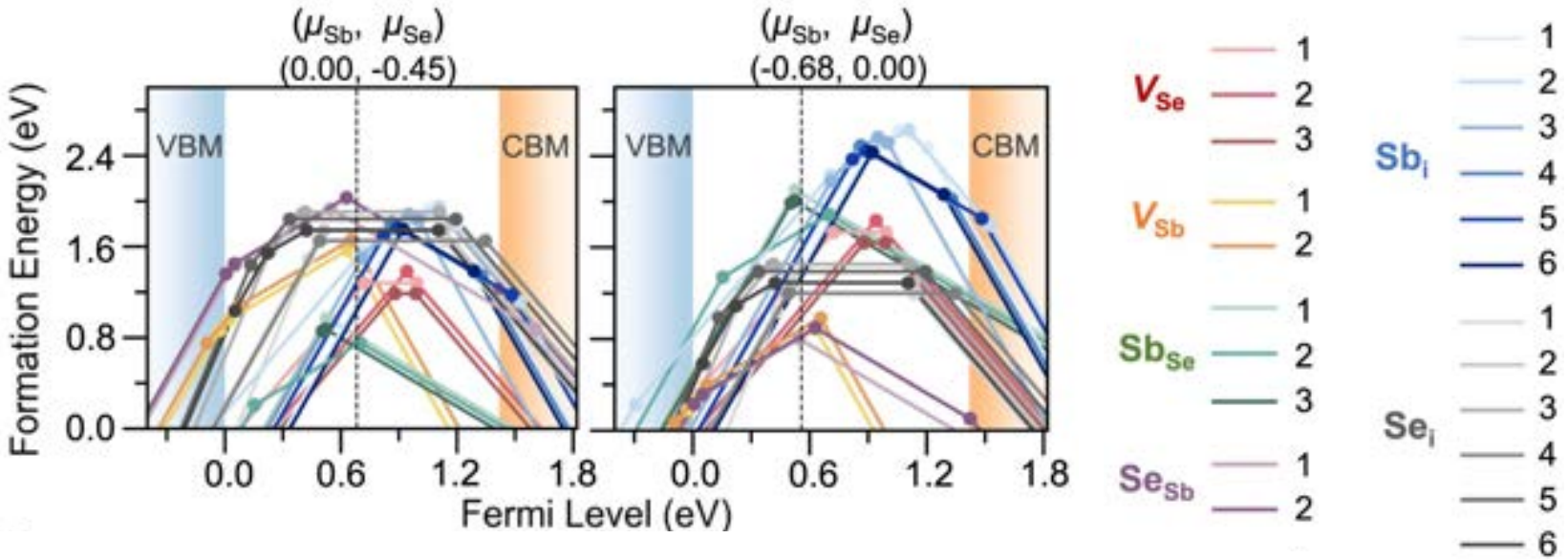
doped Case Studies: Split Vacancies

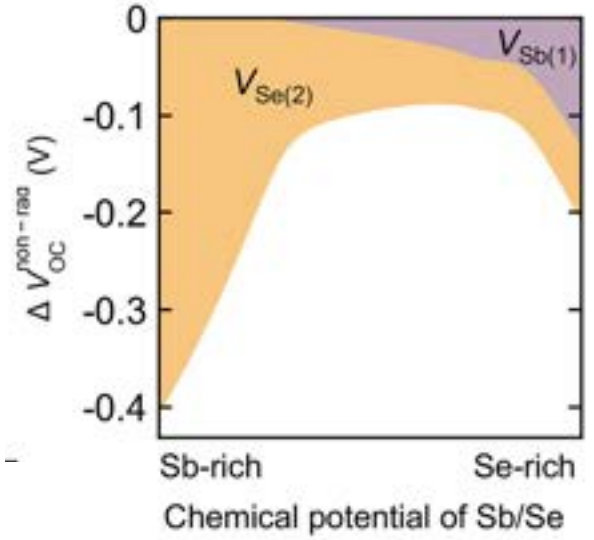
Kavanagh *arXiv (JPhys Energy Emerging Leaders)* 2025

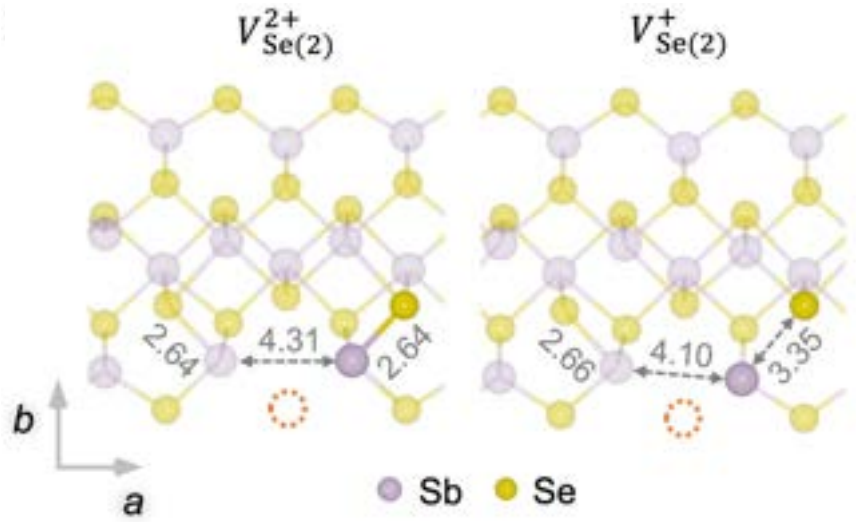


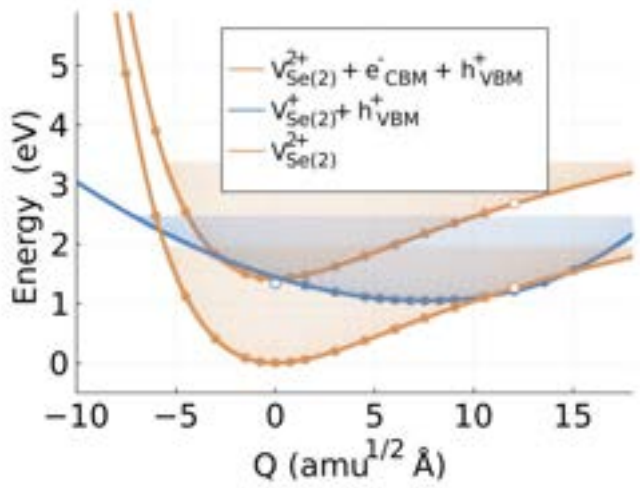
[1] Fowler et al. *J Appl Phys* 2024

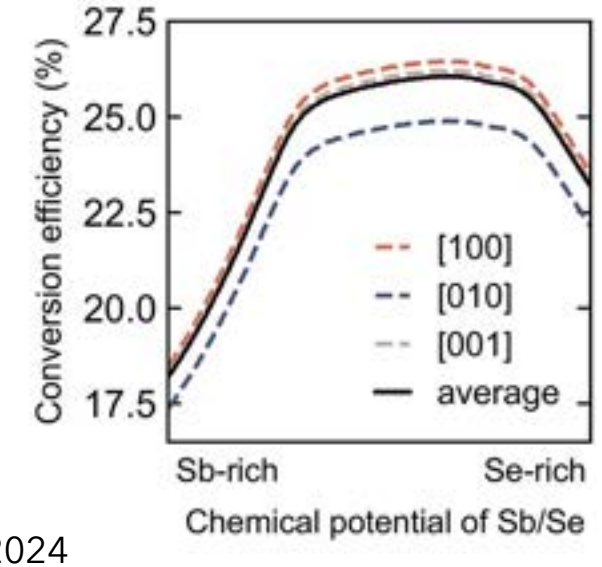
doped Case Studies: Defects in Sb_2Se_3





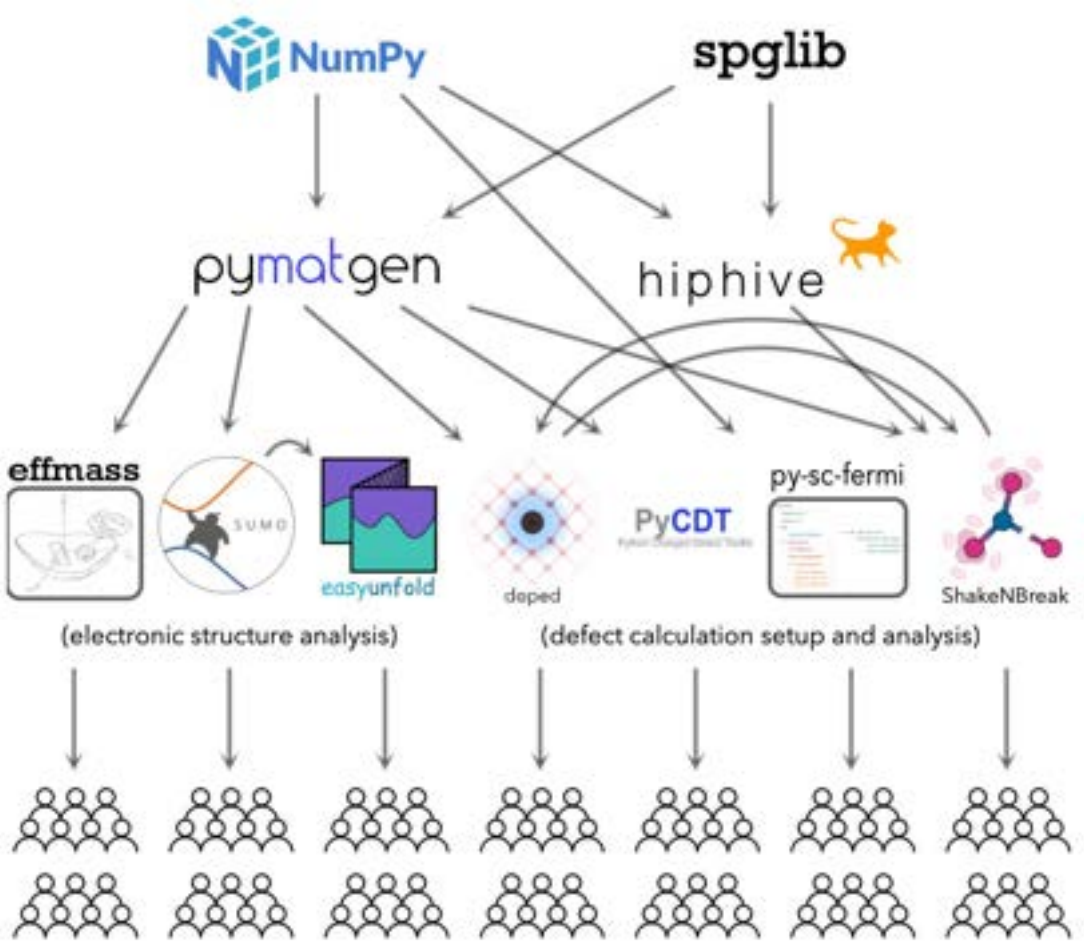






Wang et al. *Joule* 2024

Importance of (Good) Research Software



Good research software ➡️ exponential boosts in community research productivity

But mostly *indirect* impacts, with a major lack of funding for research software management positions

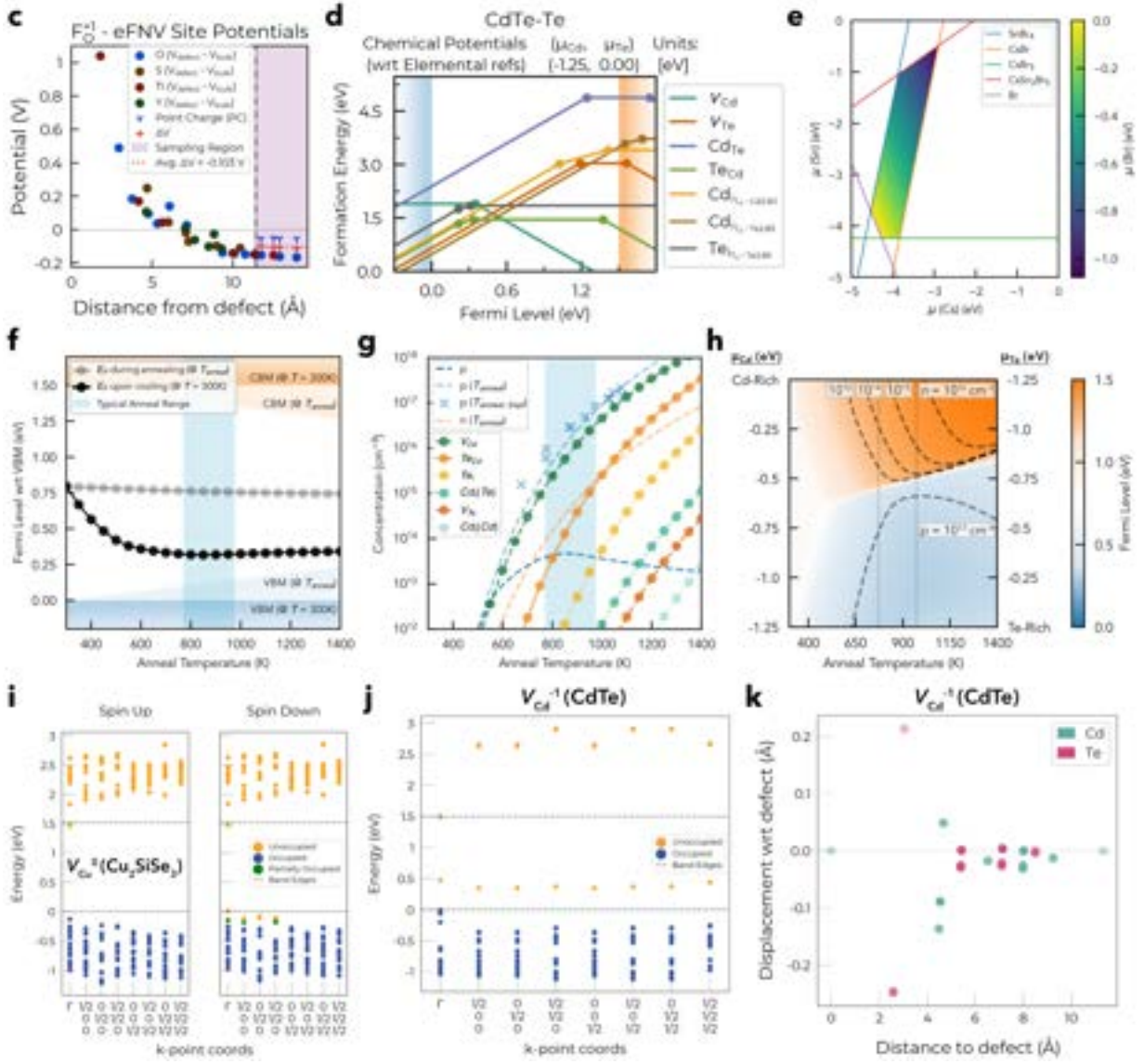
(especially when competing with Google etc.)

doped Case Studies: …

- Y. Fu & H. Lohan et al. **Factors Enabling Delocalized Charge-Carriers in Pnictogen-Based Solar Absorbers: In-depth Investigation into $CuSbSe_2$** Nature Communications 2025
- S. R. Kavanagh **Identifying Split Vacancies with Foundation Models and Electrostatics** arXiv 2025
- S. R. Kavanagh et al. **Intrinsic point defect tolerance in selenium for indoor and tandem photovoltaics** ChemRxiv 2025
- J. Hu et al. **Enabling ionic transport in Li_3AlP_2: the roles of defects and disorder** Journal of Materials Chemistry A 2025
- X. Jiang et al. **Carrier lifetime killer in 4H-SiC: carrier capture path via carbon vacancies** Journal of Materials Chemistry C 2025
- C. Kaewmeechai, J. Strand & A. Shluger **Structure and Migration Mechanisms of Oxygen Interstitial Defects in β-Ga_2O_3** Physica Status Solidi B 2025
- X. Wang et al. **Sulfur vacancies limit the open-circuit voltage of Sb_2S_3 solar cells** ACS Energy Letters 2024
- A. Zhang et al. **Optimizing the n-type carrier concentration of an $InVO_4$ photocatalyst by codoping with donors and intrinsic defects** Physical Review Applied 2024
- M-L. Wang et al. **Impact of sulfur doping on copper-substituted lead apatite** Physical Review B 2024
- S. Quadir et al. **Low-Temperature Synthesis of Stable $CaZn_2P_2$ Zintl Phosphide Thin Films as Candidate Top Absorbers** Advanced Energy Materials 2024
- M. Elgamil et al. **Controlling the Superconductivity of $Nb_2Pd_{1-x}S_5$ via Reversible Li Intercalation** Inorganic Chemistry 2024
- Z. Yuan & G. Hautier **First-principles study of defects and doping limits in CaO** Applied Physics Letters 2024
- B. E. Murdock et al. **Li-Site Defects Induce Formation of Li-Rich Impurity Phases: Implications for Charge Distribution and Performance of $LiNi_{0.5-x}M_xMn_{1.5}O_4$ Cathodes (M = Fe and Mg; x = 0.05–0.2)** Advanced Materials 2024
- A. G. Squires et al. **Oxygen dimerization as a defect-driven process in bulk $LiNiO_2$** ACS Energy Letters 2024
- X. Wang et al. **Upper efficiency limit of Sb_2Se_3 solar cells** Joule 2024
- I. Mosquera-Lois et al. **Machine-learning structural reconstructions for accelerated point defect calculations** npj Computational Materials 2024
- W. Dou et al. **Band Degeneracy and Anisotropy Enhances Thermoelectric Performance from $Sb_2Si_2Te_6$ to $Sc_2Si_2Te_6$** Journal of the American Chemical Society 2024
- K. Li et al. **Computational Prediction of an Antimony-based n-type Transparent Conducting Oxide: F-doped Sb_2O_5** Chemistry of Materials 2024
- S. Hachmioune et al. **Exploring the Thermoelectric Potential of MgB_2: Electronic Band Structure, Transport Properties, and Defect Chemistry** Chemistry of Materials 2024
- Y. Zeng et al. **Role of carbon in α-Al2O3:C crystals investigated with first-principles calculations and experiment** Ceramics International 2024
- X. Wang et al. **Four-electron negative-U vacancy defects in antimony selenide** Physical Review B 2023

…

Studies using doped



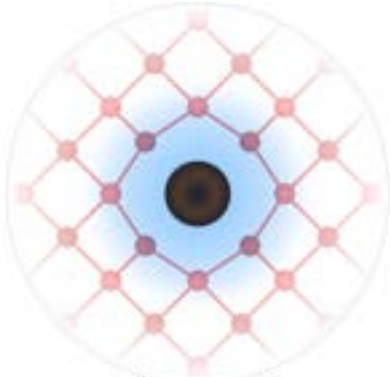
Acknowledgements

PhD (UCL & Imperial):

Profs David Scanlon & Aron Walsh

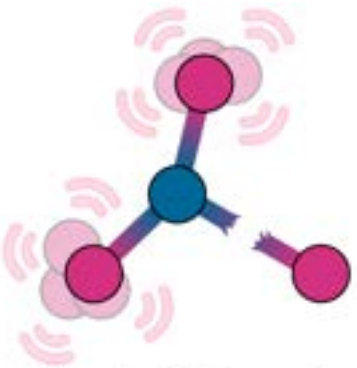
Fellowship (Harvard):

Prof. Boris Kozinsky



doped

doped.readthedocs.io



ShakeNBreak

shakenbreak.readthedocs.io

Thank you for listening!

skavanagh@seas.harvard.edu