A machine learning surrogate for density functional theory based on the local density of states

PRESENTED BY

Normand A. Modine, *Sandia National Laboratories*

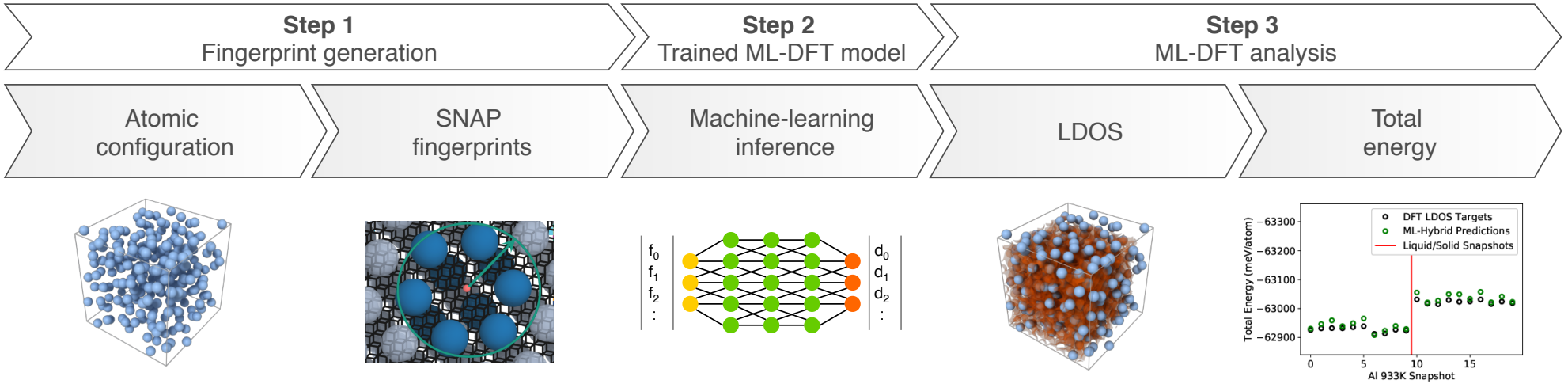
With Lenz Fiedler,[2,3] D. J. Vogel,[1] A. P. Thompson,[1] J. A. Ellis,[4] J. A. Stephens,[1] G. A. Popoola,[1] Attila Cangi,[2,3] and Siva Rajamanickam[1]

[1]*Sandia National Laboratories,*
[2]*Center for Advanced Systems Understanding,*
[3]*Helmholtz-Zentrum Dresden-Rossendorf,*
[4]*Oak Ridge National Laboratory*

The Center for Advanced Systems Understanding (CASUS) is financed by the German Federal Ministry of Education and Research (BMBF) and by the Saxon Ministry for Science, Art, and Tourism (SMWK) with tax funds on the basis of the budget approved by the Saxon State Parliament.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND No. _____

Workflow based on Machine Learning (ML) allows prediction of DFT results with O(N) cost



(a) Atomic configuration described by positions and types of atoms

(b) SNAP descriptors calculated on a grid of points encode the local structure at each point

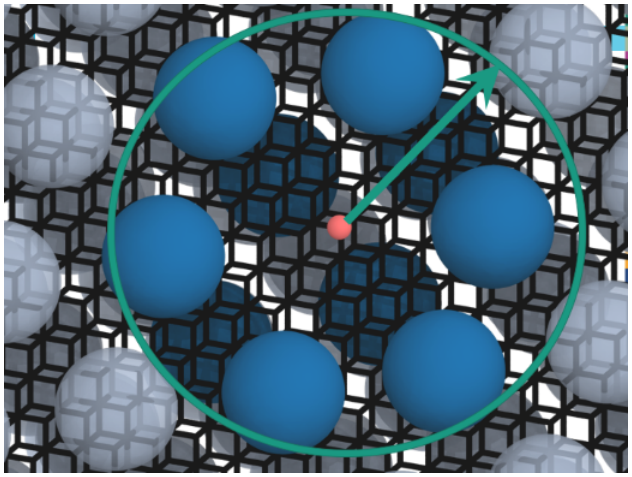
(c) A Feed-forward neural network runs independently at each grid point

(d) The Local Density of States (LDOS) is predicted at each grid point

(e) Densities, band energies, total energies, forces, etc. are calculated from the LDOS

- Relative to DFT, the only new approximation is the determination of the LDOS using a local ML model rather than by solving the Kohn-Sham equations
- Computations are independent at each point except for sums across points and calculation of the Hartree energy
- Once the ML model has been trained on the LDOS for cells where DFT calculations are practical, the model can be accurately and efficiently applied to much larger systems.

SNAP Descriptors Encode the Local Structure Near to Each Point in a Cartesian Grid



Spectral neighbor analysis potential (SNAP) descriptors (A. P. Thompson et al, J. Comput. Phys. **285**, 316 (2015)) are computed on a grid of points. Expand the "atom density"

$$\rho(\mathbf{r}) = \delta(\mathbf{0}) + \sum_{r_k < R_{\text{cut}}^{\nu_k}} f_c\left(r_k; R_{\text{cut}}^{\nu_k}\right) w_{\nu_k} \delta(\mathbf{r}_k),$$

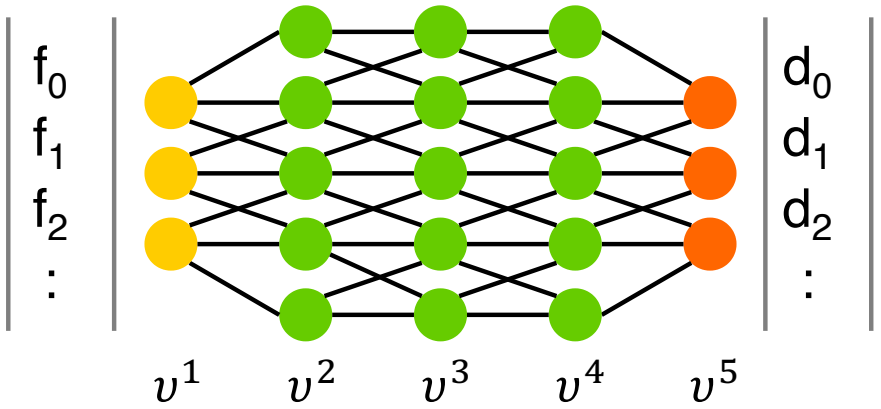
in hyperspherical harmonics $\mathbf{U}_j(\theta_0, \theta, \phi)$ where $\theta_0 = \theta_0^{\max} \dfrac{r_k}{R_{\text{cut}}^{\nu_k}}$. Expansion coefficients are

$$\mathbf{u}_j = \mathbf{U}_j(\mathbf{0}) + \sum_{r_k < R_{\text{cut}}^{\nu_k}} f_c\left(r_k; R_{\text{cut}}^{\nu_k}\right) w_{\nu_k} \mathbf{U}_j(\theta_0, \theta, \phi).$$

Then, descriptors are rotationally invariant tensor products of the expansion coefficients

$$B_{j_1 j_2 j} = \mathbf{u}_{j_1} \otimes_{j_1 j_2 j} \mathbf{u}_{j_2} \cdot (\mathbf{u}_j)^*$$

At Each Grid Point, a Feed-Forward Neural Network Predicts the LDOS from the Descriptors



A series of transformations on the input vector of descriptors yields the output vector of LDOS values for a grid of energies. The transformation at each layer is

$$v_{(n,1)}^{\ell+1} = \varphi\left(\mathbf{W}_{(n,m)}^{\ell} v_{(m,1)}^{\ell} + b_{(n,1)}^{\ell}\right),$$

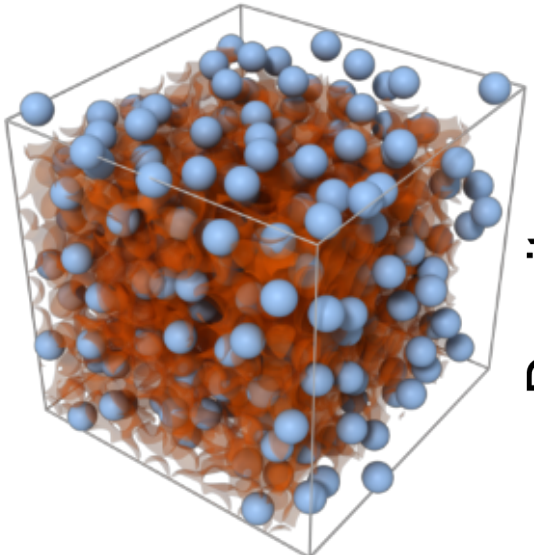
where φ is a non-linear activation function, and the weight matrices $\mathbf{W}_{(n,m)}^{\ell}$ and bias vectors $b_{(n,1)}^{\ell}$ are optimized to fit the training data using backpropagation gradients.

The predictions for a separate set of validation data are monitored to prevent over-fitting

As an outer loop, the hyperparameters used in training the model are optimized

To predict 250 LDOS values from 91 fingerprints at each point of a 200x200x200 grid, a typical network has 5 layers with a maximum width of 800-4000 and the LeakyReLU activation function

The Neural Network is Trained on LDOS Data Calculated with DFT for ~256 Atom Cells



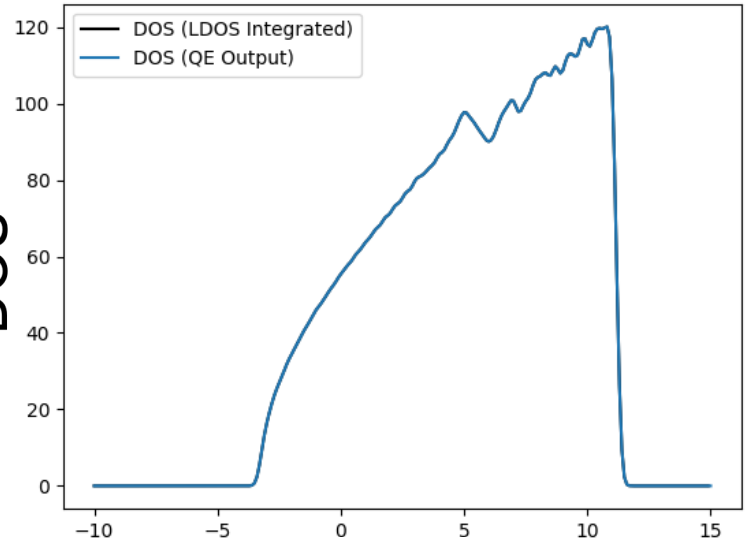
Position

Density

LDOS

Position and Energy

DOS



Energy

For training, the LDOS is calculated from the Kohn-Sham wavefunctions and energies as

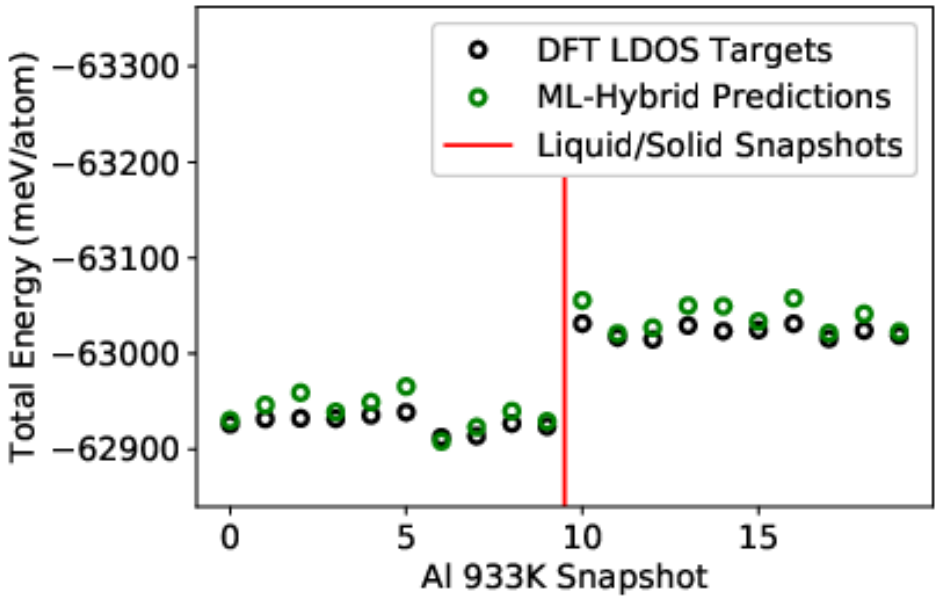
$$D(\epsilon, \boldsymbol{r}; \underline{\boldsymbol{R}}) = N_k^{-1} \sum_k^{N_k} \sum_j^{N_s} |\phi_{jk_k}(\boldsymbol{r}; \underline{\boldsymbol{R}})|^2 \, \tilde{\delta}(\epsilon - \epsilon_{jk_k})$$

where $\tilde{\delta}(x) = \exp(-x^2/\sigma^2)/\sqrt{\pi\sigma^2}$ represents a Dirac delta function. A fine grid of k-points is needed to get a smooth LDOS without excessive smearing. From the DFT or predicted LDOS, the electronic density and DOS are calculated as

$$n[D](\boldsymbol{r}; \underline{\boldsymbol{R}}) = \int d\epsilon \; f^{\beta}(\epsilon) D(\epsilon, r; \underline{\boldsymbol{R}}). \qquad \bar{D}[D](\epsilon; \underline{\boldsymbol{R}}) = \int d\boldsymbol{r} D(\epsilon, \boldsymbol{r}; \underline{\boldsymbol{R}}).$$

where we solve for the Fermi level in $f^{\beta}(\epsilon)$ that gives the correct total number of electrons.

Evaluating the DFT Total Energy from the LDOS



The DFT total free energy can be expressed as

$$A^{BO}[D](\underline{\boldsymbol{R}}) = E_b[D] - S_S[D]/\beta - U[D] + E_{XC}[D] - V_{XC}[D] + V^{ii}(\underline{\boldsymbol{R}}),$$

where the band energy and electronic entropy are determined from the DOS by

$$E_b[D] = \int d\epsilon \; f^{\beta}(\epsilon)\,\epsilon\,\bar{D}[D](\epsilon;\underline{\boldsymbol{R}})$$

$$S_S[D] = -\int d\epsilon \; \{f^{\beta}(\epsilon)\log[f^{\beta}(\epsilon)] + [1 - f^{\beta}(\epsilon)]\log[1 - f^{\beta}(\epsilon)]\}\bar{D}[D](\epsilon;\underline{\boldsymbol{R}})$$

The density determines the Hartree energy $U[n[D]]$, Exchange-Correlation energy $E_{XC}[n[D]]$, and

$$V_{XC}[D] = \int d\boldsymbol{r} \;\; n[D](\boldsymbol{r};\underline{\boldsymbol{R}}) \;\; \delta E_{XC}[D]/\delta n[D](\boldsymbol{r};\underline{\boldsymbol{R}})$$

Evaluating Forces (Mostly Implemented!)

Calculated with ML Backpropagation

In LAMMPS

In Quantum Espresso

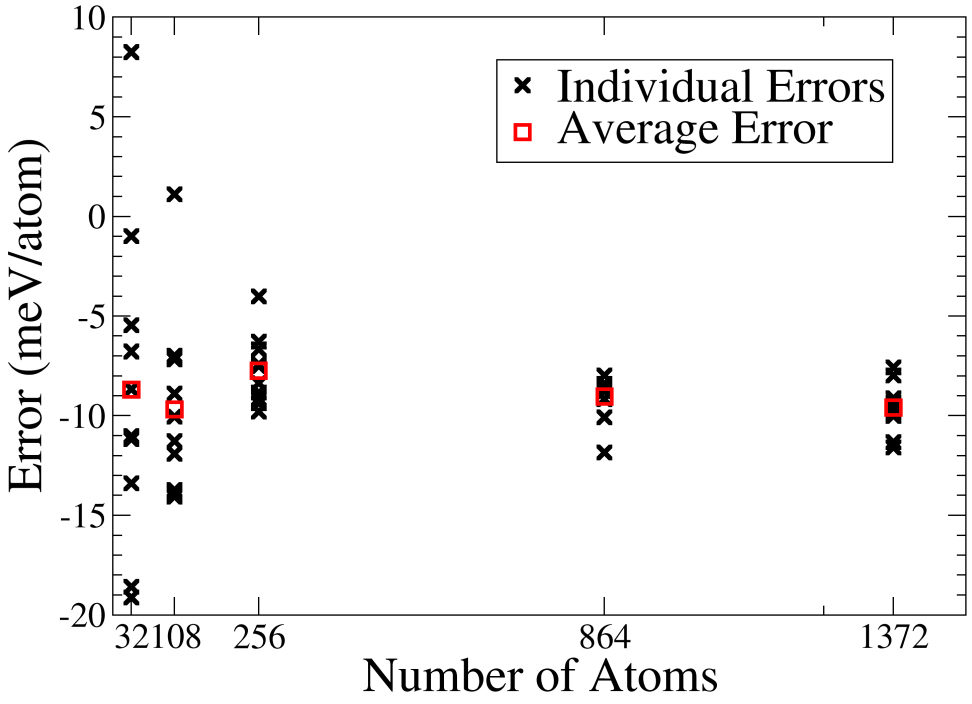
$$\mathbf{F}=\frac{dA^{BO}}{d\boldsymbol{R}}=\sum_{i,j,k}\left(\frac{dA^{BO}}{dD(\epsilon_i,\boldsymbol{r}_j)}\right)_{N_e}\frac{dD(\epsilon_i,\boldsymbol{r}_j)}{d\boldsymbol{B}_k(\boldsymbol{r}_j)}\frac{d\boldsymbol{B}_k(\boldsymbol{r}_j)}{d\boldsymbol{R}}+\frac{dV^{ii}(\boldsymbol{R})}{d\boldsymbol{R}}$$

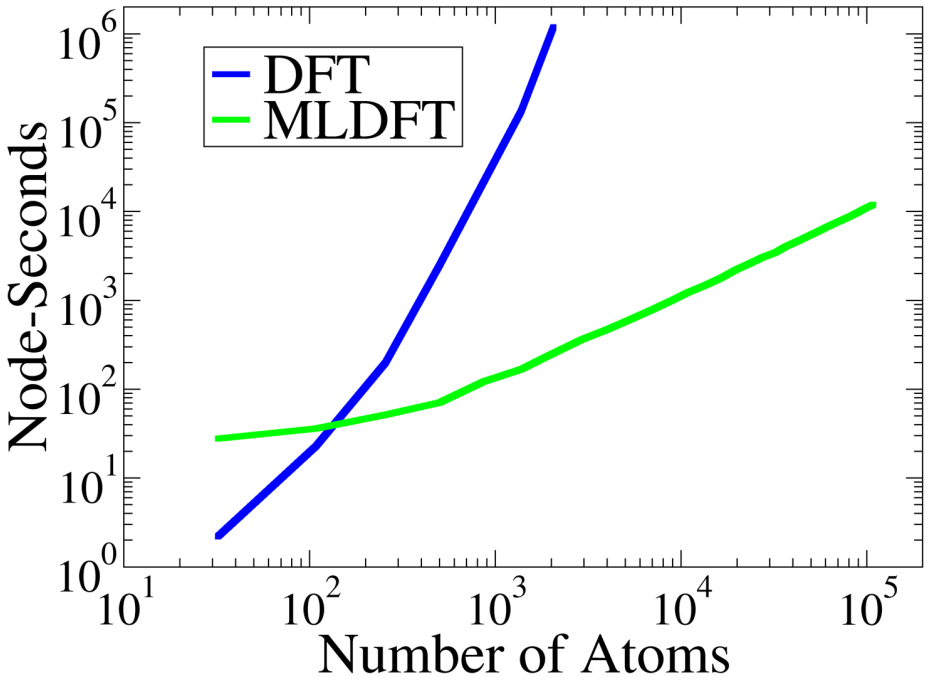
$$\left(\frac{dA^{BO}}{dD(\epsilon_i,\boldsymbol{r}_j)}\right)_{N_e}=\left(\frac{\partial A^{BO}}{\partial D(\epsilon_i,\boldsymbol{r}_j)}\right)_{n(\boldsymbol{r}_j),\mu}+\left(\frac{\partial A^{BO}}{\partial n(\boldsymbol{r}_j)}\right)_{D(\epsilon_i,\boldsymbol{r}_j),\mu}\left(\frac{\partial n(\boldsymbol{r}_j)}{\partial D(\epsilon_i,\boldsymbol{r}_j)}\right)_{\mu}$$
$$-\left(\frac{\partial A^{BO}}{\partial \mu}\right)_{D(\epsilon_i,\boldsymbol{r}_j)}\left(\frac{\partial N_e}{\partial D(\epsilon_i,\boldsymbol{r}_j)}\right)_{\mu}\Big/\left(\frac{\partial N_e}{\partial \mu}\right)_{D(\epsilon_i,\boldsymbol{r}_j)}$$

Most terms are done. Could use advice on using Quantum Espresso to evaluate:

$$\int d\boldsymbol{r}'\, n(\boldsymbol{r}')\frac{\delta^2 E_{xc}}{\delta n(\boldsymbol{r}')\delta n(\boldsymbol{r})}$$

Accurate and Efficient Results for Systems Sizes that are not Accessible with DFT





Model trained on 256 atoms systems give << 1 kcal/mol errors for 864 and 1372 atoms

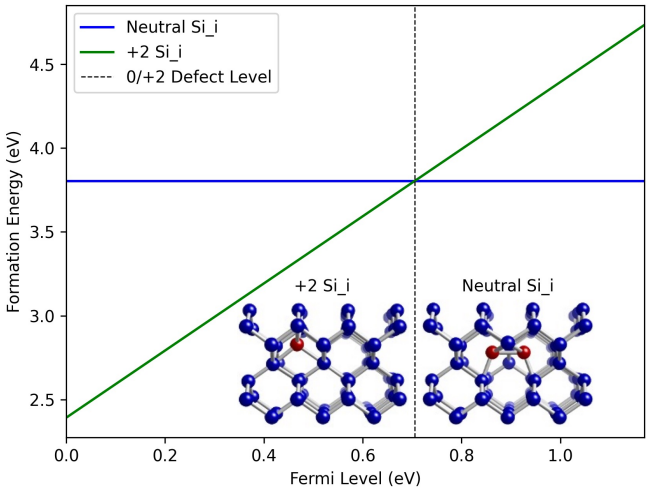
- Tests use 10 snapshots of room temperature thermalized aluminum
- A significant portion of errors are systematic – Probably do not matter in practice
- Errors somewhat larger for smaller systems --- Effects of periodicity are missing

ML approach scales linearly with system size, i.e., O(N), DFT scales as $O(N^3)$

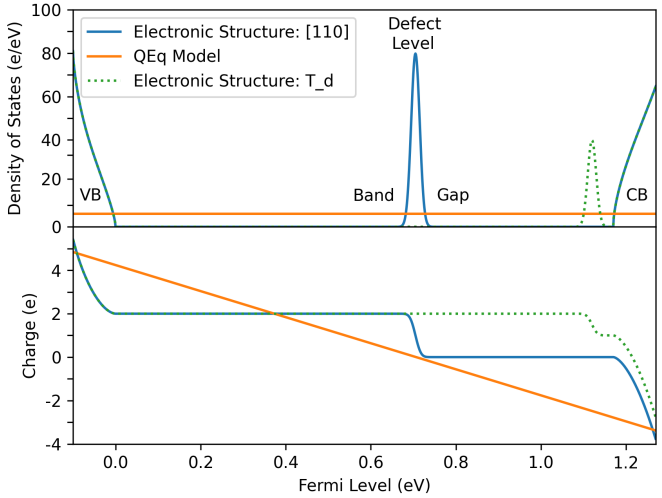
- ML approach feasible for $>10^6$ atoms, DFT becomes difficult for more than a thousand
- Crossover at ~100 atoms, or less for fully k-point converged DFT

Proposed Applications that are Very Challenging for Conventional Interatomic Potentials

Example 1: Charged defects in semiconductors or insulators



Modeling changes in defect charge state and structure with Fermi level requires capturing a complex, structure-dependent DOS



Example 2: The effects of electronic temperature on MD

- We get the dominant effects of electronic temperature automatically

Example 3: Metal-insulator transitions due to structural changes

- Properties are discontinuous when the band gap closes
- We explicitly treat the transition from metal to insulator

Getting More Information

Code is publicly available!

- https://github.com/mala-project
- Python code that calls LAMMPS and Quantum Espresso

Publication:

- Accelerating Finite-temperature Kohn-Sham Density Functional Theory with Deep Neural Networks, J. A. Ellis, L. Fiedler, G. A. Popoola, N. A. Modine, J. A. Stephens, A. P. Thompson, A. Cangi, S. Rajamanickam, Phys. Rev. B 104, 035120 (2021)

Our Team

Sandia Labs
- PI Siva Rajamanickam
- Normand Modine
- Jon Vogel
- Aidan Thompson
- Adam Stephens
- Laura Swiler
- Gabriel Popoola

CASUS (Germany)
- PI Attila Cangi
- Lenz Fiedler
- Daniel Kotik

ORNL
- Austin Ellis

Summer Students
- Max Win (UPenn)
- James Fox (Georgia Tech)
- Kyle Lennon (MIT)
- Eugene Uwiragiye (NCAT)