Resolving Computational Challenges in Accelerating Electronic Structure Calculations using Machine Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

Recent advances in use of machine learned surrogates to accelerate electronic structure calculations provide exciting opportunities for materials modeling. While the new models are extremely effective, the training of such models require millions of samples for predicting the material properties for a configuration of atoms or snapshot in a single temperature, atomic density pair. This results in excessively high training costs when material properties for multiple snapshots at multiple temperatures and densities are needed. We present a novel atom-centered decomposition of local density of states for supervision, which reduces the number of samples for training and evaluation by orders of magnitude compared to past approaches. Combined with a new model for learning atomic environment descriptions end-to-end, our approach allows resolving downstream quantities such as band energy of melting point aluminum at a fraction of the cost of previous state of the art, with matching or greater accuracy. We further demonstrate that the new models generalize across multiple temperatures of Aluminum reducing computational costs even further. Finally, in order to extend the approach even further we devise an uncertainty metric to choose the next snapshot for training. We demonstrate the efficacy of this metric using liquid and solid aluminum snapshots.

1 Introduction

The ability to perform accurate materials modeling across different length and time scales holds promise in advancing key directions of material science research. Example applications include the discovery of new materials, or their behavior under extreme conditions. The primary challenge is being able to extend information from quantum mechanical calculation at microscale (nanometer, femtoseconds) to simulations operating at mesoscale and macroscale (centimeter, milliseconds).

Kohn-Sham density functional theory (DFT) has been the quantum mechanical method of choice for calculations fundamental to driving simulations at the microscale due to its accuracy and speed. Important outputs from DFT include the energy and forces of a system as a function of the atomic positions, which enable moving forward the dynamics of the simulation in time according to physical principles. However, effectiveness of DFT is limited to systems on the scale of hundreds of atoms, as its computational cost scales as the cube of the system size and becomes prohibitively expensive for larger systems. The fundamental bottleneck of DFT calculations is the Kohn-Sham differential equations, which has inspired recent efforts to use ML to approximate its solutions [Chandrasekaran et al., 2019, del Rio et al., 2020, Ellis et al., 2021]. One of the key quantities characterizing the

Submitted to NeurIPS 2021 AI for Science Workshop.

electronic structure is the electronic density of states (DOS), which describes the energy distribution of electrons of an atomic system.

Recent methods have had success using the local density of states (LDOS) as the supervised target [Chandrasekaran et al., 2019, Ellis et al., 2021], from which properties such as DOS and band energy can be computed inexpensively. While accurate, the LDOS is computationally expensive as it is defined over a 3D grid containing tens of thousands of points per atom, requiring that many predictions to resolve the properties of a snapshot. The size of the grid also needs to scale up with the size of the system in order to maintain accuracy, presenting a formidable scalability challenge. As these approaches have to be used to predict the properties at multiple temperatures and densities, the computational costs become infeasible. We solve these problems in three different ways.

First, we propose a new approach for atom-level supervision, *ADOS*, that reduces the total work for prediction by orders of magnitude by comparison with relatively same accuracy.

Additionally, existing ML approaches for resolving DOS have so far relied on hand-crafted descriptors to extract features (*fingerprints*) from local atomic environments, as the input to their ML model. While much progress has been made in the development of fingerprinting techniques, they share in common the constraint of being limited to fitting to fixed basis functions. This work proposes to instead use trainable neural descriptors for fingerprinting, specifically focusing on the Concentric Spherical Neural Network (CSNN) model [Fox et al., 2022] as extended to the DOS prediction problem. This allows the atomic environment fingerprinting to be adapted to the data and target problem, with the end goal of generalizing to greater types and complexities of environments within a single model. We experimentally evaluate our approach for accurately resolving the band energy (calculated from DOS) of aluminum at the melting point. Our overall approach is able to match and even surpass the accuracy of previous LDOS-based approach for aluminum [Ellis et al., 2021], at a fraction of the time. We demonstrate that the new fingerprinting can accurately predict the band energies for several temperature configurations not in the training data.

Finally, we examine the use of Monte Carlo dropout (or just dropout, for brevity) to assess the uncertainty of an ADOS model when predicting band energies of test atomic configurations. We show that for a model trained using solid aluminum data, the dropout uncertainty is generally higher for liquid configurations, as expected. This finding supports the use of Monte Carlo dropout to gauge the accuracy of ADOS models when predicting unlabeled data and for retraining ADOS models.

We believe our approach resolves the computational challenges in developing the surrogates for DFT calculations and also opens the door to resolving DOS for systems containing thousands of atoms or more that are beyond existing DFT capabilities.

2 Related Work

Molecular dynamics simulations depend on accurate determination of the energy of an atomistic system as a function of the atomic positions. Over at least the past decade, there has been an evolving body of work on using data to directly learn interatomic potentials (IAPs) that predict this energy. While different in their choice of method for the regression problem, these ML-based potentials share a need for *fingerprints*, or feature vector representations of localized atomic environments as input. Methods such as Bartók et al. [2010, 2013], Thompson et al. [2015], Huan et al. [2017] rely on hand-crafted descriptors for fingerprinting, while more recently some methods [Schütt et al., 2017, Lubbers et al., 2018] have used neural descriptors to learn the fingerprint end-to-end.

Recently there also have been efforts to use ML to approximate solutions to the fundamental bottleneck of DFT calculations, the Kohn-Sham differential equations [Chandrasekaran et al., 2019, del Rio et al., 2020, Ellis et al., 2021]. Solving these equations involves accurately resolving properties of the electronic structure, such as the electronic density of states (DOS). Existing ML approaches predict this quantity indirectly through spatially localized contributions, centered around 3D grid points [Chandrasekaran et al., 2019, Ellis et al., 2021] or atoms of the system [Schütt et al., 2014, Ben Mahmoud et al., 2020, del Rio et al., 2020]. In the former case, grid points correspond to supervised quantities from DFT calculation (LDOS), providing millions of training samples for a

single configuration of atoms. However, this leads to computationally intensive training and inference. Atom-centered contributions are significantly more cost effective for training and inference, but thus far do not have a well-defined formulation for localized supervision. As the only supervision is from the total DOS of the system, DFT calculations must be run for many more configurations of atoms in order to generate adequate training data, an expensive process. The proposed ADOS approach bridges this gap, providing local supervision while avoiding the cost of grid-centered LDOS.

The greatly reduced cost of the the ADOS approach opens up the possibility of assessing the uncertainty of model predictions. A low-cost means of computing uncertainty is of interest because it can warn us when model predictions are untrustworthy, and it also suggests how to augment a model's training set. We explored use of Monte Carlo dropout for this purpose. Monte Carlo dropout is a well-established technique to reduce overfitting [Hinton et al., 2012]. More recently, it has been suggested as a way to estimate the uncertainty of model predictions [Gal and Ghahramani, 2016].

3 Methods

This section covers key components of our overall machine learning approach for resolving the electronic density of states. Sec. 3.1 discusses the aluminum snapshots used in subsequent experiments. Sec. 3.2 presents a partition-of-unity approach for deriving atom-level supervision for the DOS, as targets for downstream ML. Sec. 3.3 gives an overview of the proposed neural fingerprinting model.

3.1 Dataset

The focus of our experiments is aluminum at ambient density (2.699g/cc) and over a range of temperatures between 0K and 1000K. The melting point of Al is 933K, and so this temperature range includes both solid and liquid aluminum. We used the electronic structure code *Quantum ESPRESSO* [Giannozzi et al., 2009, 2017, 2020] to generate LDOS for all atomic configurations. The configurations were generated from snapshots of DFT-MD trajectories of 256-atom supercells of aluminum. The entire dataset comprised approximately 50 such snapshots.

We considered four training sets. We used 933K liquid and solid snapshots to compare the ADOS-CSNN model with the LDOS-SNAP model. In the second set we included four low-temperature (100K, 200K) solid snapshots as well as four high-temperature (933K) solid snapshots. One low temperature and one high temperature snapshot were used for validation during training. We refer to this training set as *split-temperature*. The final training set we examined included only low temperature solid data (four 100K and 200K snapshots), and the third, only high temperature solids (four 933K snapshots). We refer to the Appendix (Sec. 6.1) and Ellis et al. [2021] for more details.

3.2 Atom-Decomposed Density of States

In order to reduce the number of predictions that are required in order to evaluate the DOS for a given system, we wish to replace the LDOS $D^L(r, E)$ evaluated at grid points r and energies E with an "Atom-Decomposed Density of States" (ADOS) $D_i^A(E)$ evaluated for atoms i and energies E. There are two requirements for the ADOS: (1) The DOS is given by a sum of the LDOS over grid points

$$D(E) = \sum_r D^L(r, E). \tag{1}$$

Summing the ADOS over atoms should produce the same DOS, i.e.,

$$\sum_i D_i^A(E) = D(E). \tag{2}$$

(2) If $D^L(r, E)$ can be accurately approximated as a function of the atomic positions in some local region around grid point r, then $D_i^A(E)$ can be accurately approximated by some function of the atomic positions in some local region around R_i, the position of atom i. Both of the above properties can be achieved if $D_i^A(E)$ is defined as a weighted sum of $D^L(r, E)$ over grid points, and the weighted sum is a local partition of unity. In particular, if $D(r, E)$ is the LDOS evaluated at grid

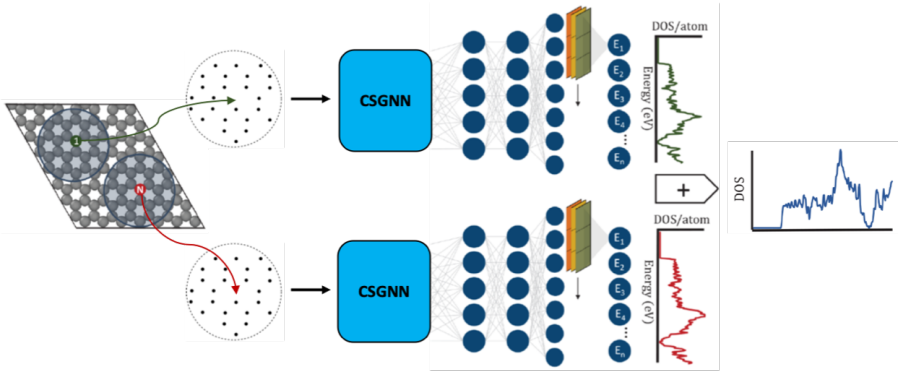


Figure 1: Atom-centered ML workflow: the local atomic environment of each atom, as positions, are input into the learned fingerprinting module (CSNN). Resulting outputs are mapped through additional neural layers to predict atom-level DOS. These are then summed to obtain the total predicted DOS for the system.

point r and energy E, we can define the ADOS associated with atom i as

$$D_i^A(E) = \sum_r w_i(r) D^L(r, E) \tag{3}$$

for some weighting functions $w_i(r)$. The set of weighting functions $w_i(r)$ is a partition of unity if

$$\sum_i w_i(r) = 1 \ \forall r. \tag{4}$$

Likewise, the partition of unity is local if every $w_i(r)$ decays sufficiently rapidly for large $\|r - R_i\|$. There are many way to define such a partition of unity, the approach that we have chosen is to define

$$w_i(r) = \frac{\exp\left[-\|r - R_i\|^2/2\sigma^2\right]}{\sum_j \exp\left[-\|r - R_j\|^2/2\sigma^2\right]}. \tag{5}$$

Given this definition, it is easy to verify that $w_i(r)$ is a partition of unity and that Requirement (1) above is satisfied. For atom positions R_i that are evenly distributed throughout space, $w_i(r)$ decays as a Gaussian tail for large $\|r - R_i\|$, and the partition of unity is local. For systems that involve large regions with no atoms, some of the weighting functions $w_i(r)$ can remain substantial throughout such regions. However, $D^L(r, E)$ is generally small in such regions, at least for energies E that are occupied by electrons, and thus, for practical purposes, we believe that Requirement (2) also holds for such systems.

When σ is much less than the distance between atoms, the partition of unity defined above closely approximates an approach in which the LDOS at each grid point is assigned to the nearest atom. In the opposite limit, which σ is comparable to the distance between atoms, the LDOS at each grid point is shared between several atoms. We have picked an intermediate value of $\sigma = 1.3$ Angstroms, compared to an average nearest neighbor distance of around 2.6 Angstroms. Thus, grid points near to an atom will mostly have their LDOS assigned to that atom, while grid points between atoms will have their LDOS shared between the nearby atoms.

Using the above approach, we calculated the ADOS from our previously evaluated LDOS in order to generate training data for a model that predicts the ADOS as a function of the local environment around each atom. This model can then be used to predict the ADOS directly while avoiding the computationally expensive evaluation of the LDOS.

3.3 Concentric Spherical Neural Network for Atomic Environments

A workflow of the overall ADOS ML approach is illustrated by Fig. 1. CSNN, the proposed model, operates on a concentric spherical spatial sampling of 3D space. Each individual sphere is discretized

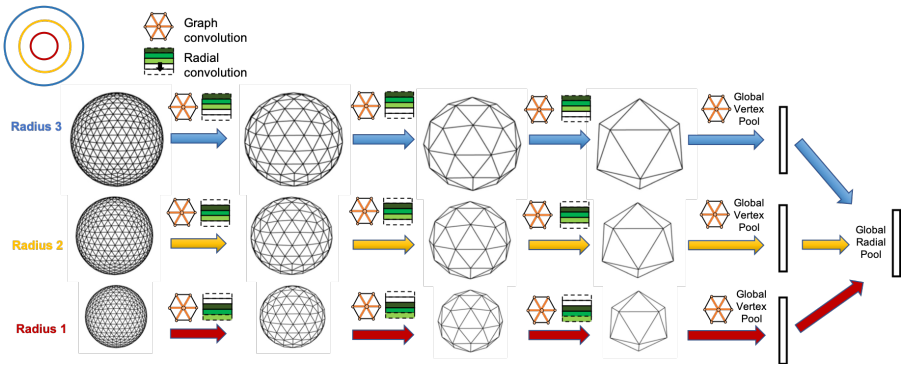


Figure 2: Example CSNN architecture with $R = 3$ concentric spheres. Graph convolutions are followed by radial convolutions at each density of spherical sampling. Graph convolution is applied within each sphere. 1D convolution is applied between co-radial vertices (3 in this example). Vertex pooling (not shown) and downsampling then coarsens the spherical sampling. Global pooling is applied at the end to obtain the final feature representation.

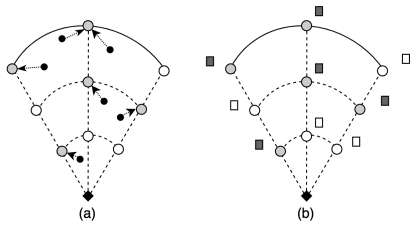


Figure 3: Shown is a 2D cross section of an atomic environment centered at a reference point (black diamond), for an example sector. (a) Each atom (black dot) in the environment a value of $\phi(r)$ to its nearest vertex in 3D space, where r is radial distance from the center and ϕ is a chosen distance mapping (such as the inverse function). (b) Values incident at any given vertex are summed, resulting in a scalar input feature per vertex.

by the icosahedral grid, resulting in a highly uniform sampling of spherical space. The grid is sub-divided recursively to create higher sampling resolution. The sampling is further extended radially, resulting in concentric spheres about a center, which is defined naturally as an atom for the ADOS problem. We refer to Fig. 2 for illustration of the concentric spherical grids. An atom's atomic environment is contained within the concentric spherical sampling, and mapped to an initial description over the sampling. Fig. 3 provides an illustration of this mapping.

Two types of convolutions are defined for representation learning over the concentric spherical grid: intra-sphere and inter-sphere convolutions. The former is implemented by graph convolutions [Kipf and Welling, 2017], with connectivity defined by each vertex's local neighborhood in the icosahedral discretization. Inter-sphere convolutions operate between co-radial vertices, orthogonally to intra-sphere convolutions. The combined use of the two convolution types permits extracting of features volumetrically over the concentric spherical sampling. Furthermore, the intra-sphere convolutions are by design rotationally equivariant to the icosahedral rotation group [Yang et al., 2020], and approximately equivariant to the general space of 3D rotations. We refer to Fox et al. [2022] for more detailed discussion of the concentric spherical convolutions. We combine the proposed convolutions into a hierarchical convolutional architecture, by also utilizing pooling and downsampling over the icosahedral grid. Fig. 2 illustrates an example CSNN architecture. Convolutions at different scales of spherical sampling enables learning representation of the input atomic environment analogously to 2D CNNs for images.

Method	Training Set	Total Training Samples	Test Set	Band Energy Max Error (meV/atom)	Band Energy Mean Error (meV/atom)
LDOS-SNAP Ellis et al. [2021]	6 liquid	4.8×10^7	3 liquid	21.3	17.1
	6 solid	4.8×10^7	3 solid	39.3	33.6
ADOS-CSNN	6 liquid	1.5×10^3	3 liquid	19.9	15.6
	6 solid	1.5×10^3	3 solid	5.3	3.3

Table 1: Band energy results, comparing the proposed ADOS-CSNN approach to prior LDOS-SNAP approach. Band energy error is calculated for the test set, and measured in terms of both max and mean absolute error.

4 Results

In this section we present main results of our atom-centered ML approach for electronic structure calculation, demonstrated for aluminum. Sec. 4.1 shows that the proposed ADOS permits faithful reconstruction of the original DOS, and therefore a sufficient target for atom-centered supervision. Sec. 4.2 presents band energy results using the proposed CSNN model for learned fingerprinting, combined with ADOS training. Finally, Sec. 4.2 demonstrates how the proposed ADOS approach leads to significant speedup over LDOS in practice for training and inference.

4.1 Reconstruction of DOS from ADOS

For the proposed ADOS to be useful, it must be possible to reconstruct the original DOS derived from LDOS. We experimentally verified that simple summation of the ADOS leads to nearly perfect reconstruction of the original DOS and the band energy derived from ADOS matches the original band energy. These results are shown in the Appendix (Section 6.2).

4.2 ML Model for Resolving Band Energy of Aluminum

For experiments, we consider a dataset of 20 total snapshots of aluminum at 933K, consisting of 10 liquid and 10 solid phase aluminum snapshots. For each phase, 6 snapshots are used for training, 1 snapshot for validation, and 3 for testing. Band energy is calculated from predicted DOS for each snapshot of the test set, and error from ground-truth is measured by meV per atom. We compare the proposed approach with LDOS-SNAP [Ellis et al., 2021]. Our approach uses atom-based ADOS for supervision, while LDOS-SNAP uses grid-based LDOS. Another key difference, orthogonal to the the form of supervision, is the method of fingerprinting. Whereas LDOS-SNAP used SNAP [Thompson et al., 2015] for fingerprinting, we use a neural fingerprinting approach, CSNN, to learn atomic environment descriptors end-to-end.

Table 1 presents results for the proposed model and comparisons. By using ADOS instead of LDOS, the total number of samples for prediction is reduced by a factor of 32,000 for training. This reduction also extends to inference, although not shown in table for brevity. This is a significant reduction, as the total number of samples directly reflects the total amount of actual work for the model, all else equal. Importantly, this reduction is achieved without any sacrifice to accuracy.

Compared to LDOS-SNAP , the ADOS-CSNN model reduces band energy error (mean absolute error) by 9% in the case of liquid phase aluminum, and by 90% the case of solid phase aluminum. For the liquid phase band energy, the ADOS-CSNN model achieves a slight improvement in accuracy over LDOS-SNAP. However, for the solid phase band energy, the ADOS-CSNN model achieves nearly 10x improvement, which represents a major advance in predictive power. We surmise that this large reduction in error is due to difference in the learning model–CSNN learns local environment descriptions end-to-end, which could prove beneficial when using a single model for hybrid dataset. However, this hypothesis remains to be investigated further.

Method	Training time (1 epoch)	Inference time (1 snapshot)
LDOS-SNAP	76 minutes	54 seconds
ADOS-CSNN	19 seconds	1 second

Table 2: Runtime comparison for training and inference, run on single V100 GPU. LDOS-SNAP takes grid-centered local descriptors as input to the neural model, but their generation time was not included in this comparison.

4.3 Runtime

In this section we explore actual runtime for training and inference of the ADOS-CSNN approach compared to the LDOS-SNAP approach. For training we consider the time for a single epoch (12 training snapshots), and for inference we consider the time to evaluate a single snapshot for its local DOS quantities. Both models are run on a single NVIDIA V100 GPU. Results are presented in Table 2. ADOS-CSNN provides a $240\times$ speedup in training per epoch and $54\times$ speedup in inference compared to LDOS-SNAP . While a very significant and practical improvement, the speedups fall short of the factor of reduction $(32,000)$ in the total amount of samples in switching from LDOS to ADOS. This is likely due to the difference in the neural models used in ADOS-CSNN vs. LDOS-SNAP . Additionally, while fingerprint generation is part of the neural model in the case of ADOS-CSNN , it is not in the case of LDOS-SNAP and was omitted from time comparison. The speedup of ADOS-CSNN should therefore be interpreted as a lower bound, especially in the case of inference, as the time to generate fingerprint for input cannot be ignored in practice.

4.4 Split-Temperature Model Predictions

Figure 4a shows the DFT-computed band energy (that is, the "truth") for all the snapshots in our dataset. A few features are noteworthy. First, for the solids, band energy decreases monotonically with temperature. Ideally, a model trained on snapshots at multiple temperatures will reproduce this trend. Second, liquids have substantially lower band energy than solids, even when they have the same temperature, as occurs at 933K.

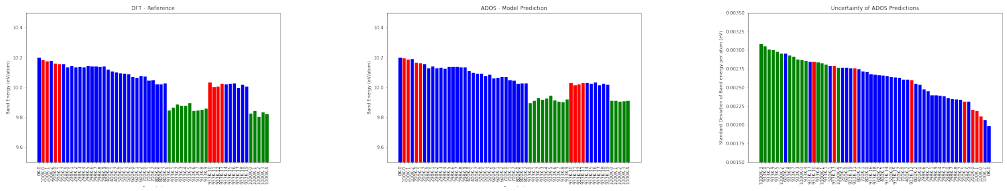
An ADOS model was trained on the split-temperature training set described in 3.1. The hyperparameters were selected using the procedure described in 6.4 and the experimental design approach is described in 6.5. This split-temperature model was used to predict ADOS (and band energy) for all snapshots. Inference was performed on the model without dropout. The predictions are shown in Figure 4b It can be seen that the model predictions qualitatively capture the two features we noted above—the temperature trend and the difference between solids and liquids—and that it is able to do so even though no liquids were present in the training set.

The parity plot in Figure 5a compares the DFT reference and ADOS model predictions directly. Points at higher energy are for lower temperatures, and vice versa. For all the solid snapshots, including those at intermediate temperatures, which were not included in the training set, very good agreement is obtained. The dashed lines on the plot are spaced at $\pm$ 10 meV/atom and predictions for solids are typically within this band. This result demonstrates that it is possible to generate machine learned models using the ADOS approach that exhibit at least some degree of transferability outside of the training set, which is an important property for practical use. Although predictions for the liquid snapshots are well outside the chemical accuracy window, they are not completely unphysical.

Figure 5 shows parity plots for models generated from the low-temperature and high-temperature training sets. These suggest that training on a single (or narrow range) of temperatures is inadequate to produce models that show the same level of transferability that the split-temperature model achieved.

4.4.1 Dropout Uncertainty

Monte Carlo dropout was used to compute the uncertainty of the split-temperature model for all snapshots. A dropout probability of 0.1 was used. A total of 512 inferences were run, and the standard

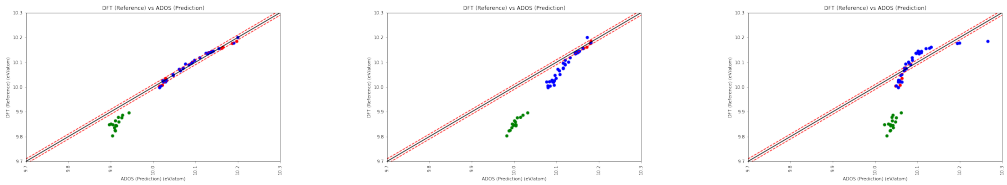


(a) DFT computed reference band energies.

(b) Band energies predicted by the split-temperature ADOS model.

(c) Standard deviation of dropout predictions.

Figure 4: Split Temperature Expected band energies, predicted band energies and uncertainty. The red bars are the solid training set snapshots. Blue are solid test snapshots, and green are liquid test snapshots. The snapshots have been sorted by decreasing uncertainty in (c).



(a) Split Temperature

(b) Low Temperature

(c) High Temperature

Figure 5: Parity plots for the low temperature and high temperature models. Red: solid training snapshot; Blue: solid test snapshot; Green: liquid test snapshot. The dash lines are $\pm$ 10 meV/atom above and below the center line.

deviation of the resulting band energies predictions was used. The results are shown in Figure 4c. Notably, the dropout uncertainty of many of the test snapshots are lower than that of snapshots that were present in the training set. This is perhaps not too surprising, considering how well the model predicts the band energies of all the solid snapshots. We also observe that liquid snapshots exhibit, for the most part, the greatest uncertainty as expected. There is no discernible step change between solid and liquid uncertainty, which may be surprising, considering the large difference between their band energies. We point to the success of the model in predicting the band energy of the liquids as a possible explanation. Uncertainty predicted using dropout provides a plausible way of selecting snapshots with which to augment an existing training set. It correctly indicates that test snapshots that intuitively are most different from the training set and that in fact have the greatest error should be included. We plan to validate these by augmenting the training set in the future.

5 Conclusion

In this work we present a machine learning approach for predicting key materials properties, such as the density of states and band energy, at a small fraction of the computational cost of existing LDOS approaches and without sacrificing accuracy. The first key piece of the proposed approach is to create atom-level supervision, ADOS, using a partition-of-unity approach. This reduces the total number of predictions required to resolve DOS compared to LDOS by orders of magnitude, for both training and inference. The second piece of our approach is to incorporate a neural model based on concentric spherical convolutions for learning atomic environment fingerprints end-to-end. We experimentally demonstrate that our overall approach allows resolving DOS and band energy many times faster than with LDOS-based approaches. In combination with our neural model for learned fingerprinting, we match and even outperform LDOS-based approaches in resolving band energy of melting point aluminum. The transferability of the split-temperature model to intermediate temperatures is an encouraging result that points to the broader applicability of the ADOS approach. In terms of future work, we believe that our atom-centered approach can be very feasibly extended to systems of size of $O(10^4)$ atoms, which is already well beyond the reach of DFT.

References

Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104:136403, Apr 2010. doi: 10.1103/PhysRevLett.104.136403. URL https://link.aps.org/doi/10.1103/PhysRevLett.104.136403.

Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Phys. Rev. B*, 87:184115, May 2013. doi: 10.1103/PhysRevB.87.184115. URL https://link.aps.org/doi/10.1103/PhysRevB.87.184115.

Chiheb Ben Mahmoud, Andrea Anelli, Gábor Csányi, and Michele Ceriotti. Learning the electronic density of states in condensed matter. *Physical Review B*, 102(23):235130, Dec 2020. ISSN 2469-9950, 2469-9969. doi: 10.1103/PhysRevB.102.235130.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statist. Sci.*, 10(3):273–304, 08 1995. doi: 10.1214/ss/1177009939. URL https://doi.org/10.1214/ss/1177009939.

Anand Chandrasekaran, Deepak Kamal, Rohit Batra, Chiho Kim, Lihua Chen, and Rampi Ramprasad. Solving the electronic structure problem with machine learning. *npj Computational Materials*, 5 (1):1–7, 2019.

Beatriz G. del Rio, Christopher Kuenneth, Huan Doan Tran, and Rampi Ramprasad. An efficient deep learning scheme to predict the electronic structure of materials and molecules: The example of graphene-derived allotropes. *The Journal of Physical Chemistry A*, 124(45):9496–9502, 2020. doi: 10.1021/acs.jpca.0c07458. URL https://doi.org/10.1021/acs.jpca.0c07458. PMID: 33138367.

J. A. Ellis, L. Fiedler, G. A. Popoola, N. A. Modine, J. A. Stephens, A. P. Thompson, A. Cangi, and S. Rajamanickam. Accelerating finite-temperature kohn-sham density functional theory with deep neural networks. *Phys. Rev. B*, 104:035120, Jul 2021. doi: 10.1103/PhysRevB.104.035120. URL https://link.aps.org/doi/10.1103/PhysRevB.104.035120.

James Fox, Bo Zhao, Beatriz Gonzalez del Rio, Sivasankaran Rajamanickam, Rampi Ramprasad, and Le Song. Concentric spherical neural network for 3d representation learning. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022. doi: 10.1109/IJCNN55064.2022.9892358.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 1050–1059. JMLR.org, 2016.

P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo, A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, E Küçükbenli, M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V Nguyen, A Otero de-la Roza, L Paulatto, S Poncé, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with quantum ESPRESSO. *Journal of Physics: Condensed Matter*, 29(46):465901, oct 2017.

P. Giannozzi, O. Baseggio, P. Bonfà, Davide Brunato, R. Car, I. Carnimeo, C. Cavazzoni, Stefano de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, Andrea Urru, and S. Baroni. Quantum espresso toward the exascale. *The Journal of chemical physics*, 152 15: 154105, 2020.

Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, sep 2009.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012. URL `http://arxiv.org/abs/1207.0580`. cite arxiv:1207.0580.

Tran Doan Huan, Rohit Batra, James Chapman, Sridevi Krishnan, Lihua Chen, and Rampi Ramprasad. A universal strategy for the creation of machine learning-based atomistic force fields. *npj Computational Materials*, 3(1):37, Dec 2017. ISSN 2057-3960. doi: 10.1038/s41524-017-0042-y.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. pages 448–456, 2015. URL `http://jmlr.org/proceedings/papers/v37/ioffe15.pdf`.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations*, 2017.

Nicholas Lubbers, Justin S. Smith, and Kipton Barros. Hierarchical modeling of molecular energies using a deep neural network. *The Journal of Chemical Physics*, 148(24):241715, Jun 2018. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.5011181.

Douglas Montgomery. *Design and Analysis of Experiments, 10th ed.* Wiley, 2019.

Thomas J Santner, Brian J Williams, and William I Notz. *The design and analysis of computer experiments*. Springer Series in Statistics, 2003.

K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. NIPS'17, page 992–1002, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Physical Review B*, 89(20):205118, May 2014. ISSN 1098-0121, 1550-235X. doi: 10.1103/PhysRevB.89.205118.

A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, and G.J. Tucker. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *Journal of Computational Physics*, 285:316–330, 2015. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2014.12.018. URL `https://www.sciencedirect.com/science/article/pii/S0021999114008353`.

Qin Yang, Chenglin Li, Wenrui Dai, Junni Zou, Guo-Jun Qi, and Hongkai Xiong. Rotation equivariant graph convolutional network for spherical image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4303–4312, 2020.

6 Appendix

6.1 Dataset

We expand on Section 3.1 and describe the data generation further here. The LDOS for all the Aluminum snapshots is calculated over a finite grid of evenly spaced energy values, with spacing of 0.1 eV ranging from -10 eV to 14.9 eV. The data for each grid point is then a vector length 250. The process used to generate the LDOS data is described in detail in Ref. Ellis et al. [2021], and we refer to it for more detailed discussion and justification of the procedures.

As described in Section 3.1, we considered four initial training sets (i) Solid and liquid snapshots at 933K; (ii) four low-temperature (100K, 200K) solid snapshots as well as four high-temperature (933K) solid snapshots (*split-temperature* data set); (iii) low temperature solid data (four 100K and 200K snapshots); and (iv) only high temperature solids (four 933K snapshots).

Data set (i) is used to show the efficacy of the ADOS-CSNN model and compare it to the LDOS-SNAP model. We selected temperatures at the two extremes of the range for the split-temperaature data because we wanted to understand how errors and uncertainty would grow for solid phase test data as we moved inward in temperature. That is, we wanted to discover how well a model trained on such a training set could "interpolate" at intermediate temperatures.

Additionally, by including only solids in the split-temperature training set, we set up an easy initial test case for the idea of using dropout for experimental design. A model trained only on solids ideally should exhibit high dropout uncertainty for liquid test data. This expectation is based partly on previous experience with grid-based models, which struggled to make cross-phase predictions, and on our physical understanding: liquid snapshots substantially differ from solids in terms of atomic positions and energies and hence contain many "out of distribution" inputs. Therefore, if uncertainty for liquid snapshots is not clearly higher than for solids, it would tend to argue against using dropout uncertainty for experimental design.

Results from models trained on (iii) and (iv) aided our interpretation of the models trained using (ii).

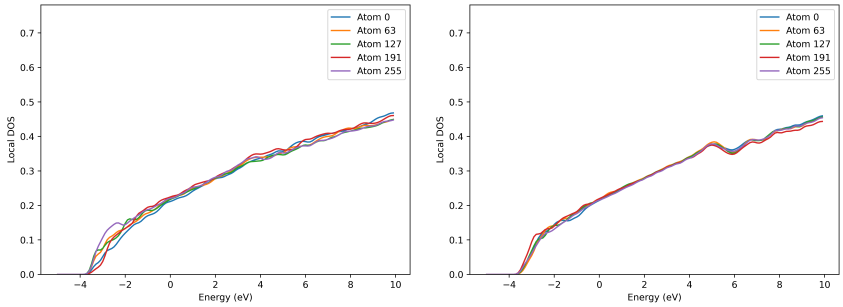
6.2 ADOS accuracy

Resulting ADOS curves are plotted for sampled atoms from liquid and solid snapshots in Fig. 6. Overall, the atom-centered DOS appears much more similar within each snapshot than between liquid and solid snapshots, with the solid snapshots showing prominent wiggles in the 5 to 8 eV range that are remnants of the Van Hove singularities that occur in a perfect crystal. Furthermore, the ADOS within each snapshot tends to reflect the profile of the DOS of their respective snapshots (see Fig. 7). These results are to be expected since both solid and liquid aluminum are generally homogeneous systems with each atom in a similar local environment. There are some fluctuations in the local environment, which are reflected in the variations between the ADOS for different atoms within the same phase. The local environment varies more for the liquid than for the solid, and correspondingly, the variation between the ADOS for different atoms is larger in the liquid. However, even in the liquid, these local fluctuations are not as significant as the difference between the solid and liquid phases. This shows that the atom-centered DOS profile is able to resolve the differences between liquid and solid phase aluminum, as well as fluctuations in the local environment of the atoms.

We further plot the DOS predicted by ADOS-CSNN to the reference DOS from DFT, and show these for example liquid and aluminum snapshots in Fig. 7. These plots confirm that the proposed approach is able to produce aluminum DOS closely matching DOS from quantum-mechanical calculation, and that the band energy accuracy is not resulting from some degeneracy.

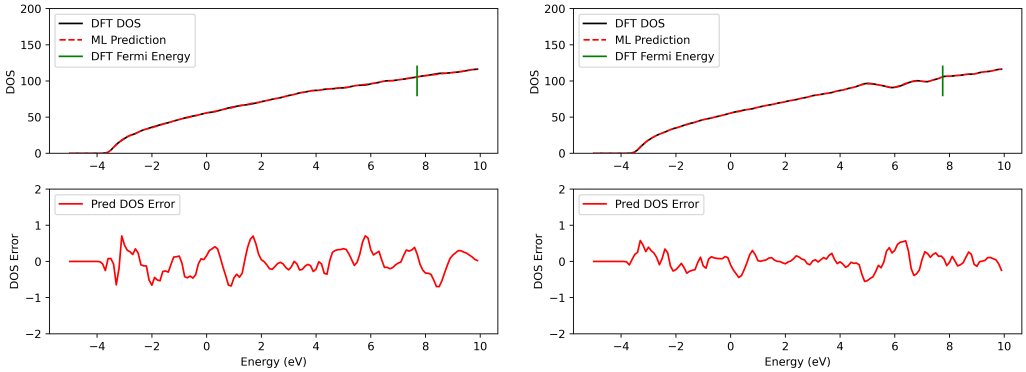
6.3 List of hyperparameters for the ADOS-CSNN model

We list the hyperparameter settings for the best-performing ADOS-CSNN model in Table 3. We also used batch normalization [Ioffe and Szegedy, 2015], which is not counted in the total number of layers. Finally, we also plot training and validation loss for the best-performing model in Fig. 8.



(a) ADOS for liquid aluminum, snapshot 9. (b) ADOS for solid aluminum, snapshot 19.

Figure 6: Atom-centered DOS values resulting from partition-of-unity, for liquid and solid aluminum snapshots at 933K. Shown are DOS from 5 sampled atoms of each snapshot.



(a) DOS of liquid aluminum for snapshot 8. (b) DOS of solid aluminum for snapshot 18.

Figure 7: Density of states for solid and liquid snapshots at 933K. Top row shows DOS curve predicted by ADOS compared to reference curve from DFT. X-axis is energy range from -5 to 10 eV. Units for y-axis is eV. Bottom row plots difference between predicted DOS and the reference DOS of respective snapshots.

Parameter	Value
Concentric spheres	16
Spherical resolution	642
Optimizer	Adam
Batch size	32
Learning rate	0.01
Activation	ReLU
Epochs	200
Layers	18
Total Weights	5.9×10^6

Table 3: List of parameter settings for ADOS-CSNN model used for experiments. Spherical resolution is number of vertices of icosahedral spherical sampling. Number of layers is trainable layers.

6.4 Hyperparameter Tuning

A full-factorial study of three hyperparameters was performed to identify the model parametrization that minimized the validation error for each of the three training sets. These hyperparameters were:

- **Learning rate**. This is the initial learning rate provided to the Adam optimizer, which was used for training. Permitted values were 0.01, 0.001, and 0.0001.

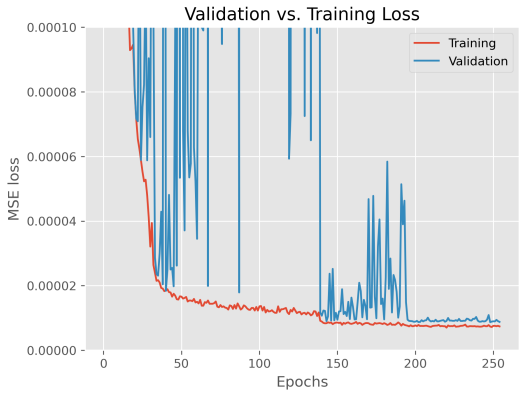


Figure 8: Training and validation loss curves for best-performing version of ADOS-CSNN used in experiments. Y-axis is mean-squared error loss for ADOS prediction, and x-axis is epoch number.

- **Number of hidden output layers**. In the ADOS model, the CSNN autoencoder provides features to a set of dense layers. Given a parameterization of the CSNN, the width of the layers is constant and determined by the output size of the CSNN. The numbers of layers considered were 3, 4, 6, 8, and 10.
- **The parameter factor**. The parameter factor scales the number of output channels in the convolutional layers . Permitted values were 5, 6, 8, 10, and 12.

Three replicates were performed for each hyperparameter combination. The training was permitted to run for a maximum of 600 epochs, which was sufficient for the learning rate convergence criterion (1e-5) to be met in nearly all cases.

Other hyperparameters matched those used in Table 3.

The optimal hyperparameters for the three training sets are shown in Table 4. The training history for the split-temperature model is shown in Figure 9. Because the low- and high- temperature training sets were not the primary focus of this work, their training histories are omitted for brevity.

Table 4: Optimal hyperparameters and resulting number of unknowns for the three training sets.

Training Set	Hidden Output Layers	Parameter Factor	Learning Rate	Unknowns
Split Temperature	6	10	0.1	6.22e7
Low Temperature	6	8	0.1	3.99e7
High Temperature	3	12	0.1	6.11e7

6.5 Experimental Design

Experimental Design refers to the selection of parameter settings at which to run physical or computational experiments [Santner et al., 2003, Montgomery, 2019]. The goal is to identify the next set of atomic configurations which should be run through the DFT calculations to generate more training data to improve the ADOS predictions. We did not pursue standard approaches such as Bayesian optimal design [Chaloner and Verdinelli, 1995] due to computational feasibility. We wanted an approach that would select configurations that are dissimilar to data that are already present and for which the current ADOS predictions are poor (there is little reason to add data for which we can already make accurate predictions). The approach also should be fast and based only on the input atomic configurations (we want to avoid performing costly DFT calculations on candidate data).

Dropout is a reasonable approach in this context: the test data for which model predictions are highly uncertain are good candidates for augmenting an existing training set, which can then be used to

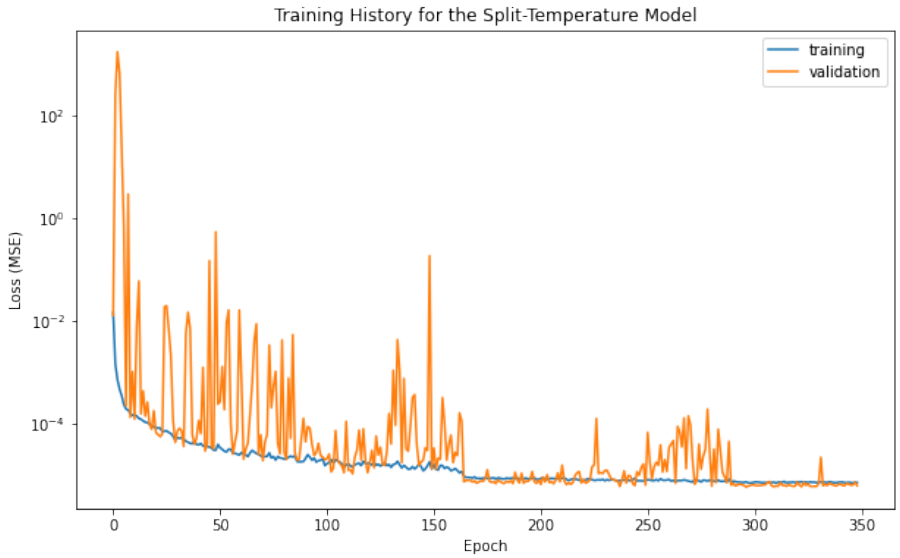


Figure 9: Training and validation errors during training for the split-temperature model.

update the model. Importantly, test data need not be labeled to use Monte Carlo dropout. In the context of our present problem, this implies that we can use dropout to select atomic configurations before incurring the computational expense of calculating their ADOS using DFT. For our experiments, dropout probability was set to 0.1. We applied droput both to intra- and inter-sphere convolution and in the hidden output layers of the ADOS model.