Exploring Novel Materials Using Deep Learning Algorithm

Shu-Ting Pi*
Dept. of Physics, University of California, Davs, One Shields Ave., Davis, CA, 95616

Motivation

Density functional theory (DFT)[1] is a very powerful tool in condensed matter physics and material science. It calculates material properties from the first-principles (Kohn-Sham equation) and provides highly accurate description on their behaviors under various conditions such as magnetic fields, electric fields or high pressures. Several famous codes based on DFT, e.g. ABINIT, Quantum Espresso, WIEN2k, VASP, etc.[2], have been widely applied to explore novel materials and simulate advanced electronic devices.

Despite DFT has been proven successful in studying materials properties, there are still two major limitations: 1). Systems such as heterostructures or giant molecules are challenging for DFT calculation because they contain hundreds or more atoms per unitcell which makes the computation extremely heavy 2). DFT calculation usually fails in materials with strong correlations, which are commonly found in systems with atoms that include 5d, 4f or 5f valence electrons, because DFT presumes Coulomb interaction among electrons to be weak so that single particle treatment remains valid [1].

A conventional approach to solve the above issues is Wannier functions[3]. Wannier functions are a set of highly localized functions constructed from the electronic structures calculated via DFT. When using these functions to replace the original Kohn-Shame orbitals in our interested materials, the form of interactions among electrons will become short-range. As a result, the whole system can be described by a much simplified model called “tight-binding” (TB) which assumes all the interactions are very short-range and only neighboring atoms can talk to each other. Therefore, the first issue is automatically solved because the dimensions of the original problem is greately reduced in TB model. As for the second issue, it can also be solved by introducing extra interaction terms in the TB model, e.g. dynamical mean-field theory (DMFT) method[4], constraint random-phase-approximation (cRPA) [5], or multipolar exchange modesl[6,7], etc., to capture the effects of correlations that are neglected in usual DFT calculations.

Although Wannier functions is an excellent auxiliary to DFT, its generation is never easy. Currently, only the package *WANNIER90* can interface to most DFT codes[8]. It is slow and do not support parallel calculation. Besides, the generation of Wannier functions usually requires a lot of experience. One has to fine tune the parameters until the output TB Hamiltonian can really fit the electronic structures obtained via DFT. Therefore, it still requires many manual operations. Another serious issue is the losing of symmetry in Wannier functions. Symmetry is a critical property of materials. For example, crystal symmetry can determine the degeneracy of the electronic structures, atomic symmetry can determine the main features of electron around Fermi level, spin symmetry can determine the helicity of an electronic state. Nevertheless, due to some theoretical limitations, they are generally not included in the Wannier functions which makes it difficult to analyze orbital or topological related properties.

A possible way to build a model Hamiltoinan that contains full symmetries is simplified LCAO method (it is more commonly named Slaster-Koster method)[9]. It parameterizes a Hamiltonian using symmetry-protected variables such as σ-bond, π-bond, δ-bond, φ-bond, etc. Therefore, if one can

represent a Hamiltonian using those symmetry-protected variables, a TB model with well-defined symmetry is guaranteed and several material properties can be analyzed easily. However, a systemic approach to construct Slaster-Koster Hamiltonians from DFT outputs remains lacking. Therefore, an efficient algorithm to map the DFT results to a Slaster-Koster Hamiltonian is particularly desired. It can not only applied to explore the properties of large scale materials but also helps researchers to analyze symmetry-related characteristics.

Recently, conventional neural networks (CNN) [10] and recurrent neural network (RNN or LSTM) [11] have been proven a very powerful framework of image recognition [cite] and natural language processing. Several attempts have been proposed in the past few years such as the high accuracy CNN models in ImageNet and the RNN implementation in Google's voice recognition. nVidia's GPU calculation has also been proven the best platform in deep learning calculation. Our interested questions are:

1. Can we use a CNN or RNN to train a machine that automatically maps a calculated electronic structure from DFT to a Slaster-Koster Hamiltonian?
2. Can we build a transfer learning algorithm that fine-tune a pre-trained to newly discovered materials and speed up the already time- and resource-consuming material mining (search for novel materials with desired properties via performing massive calculations over all possible candidate materials) research [12]?
3. Can GPU calculation greatly increase the performance in material mining?

These are the problems I would like to address in this proposal.

Proposal

In this project, I will conduct three major topics:

- Build an effective algorithm to DFT outputs to Slaster-Koster Hamiltonian using CNN or RNN

A standard CNN is constructed by two parts, i.e. the convolutional parts and fully connected parts. The convolutional part uses several convolutional layers as "filter lens" to extract the features of the input image and uses several pooling layers to reduce the dimensions. Then it uses the reduced data as inputs to a fully connected neural network as usual multilayer perceptron. This framework has been proven useful in image recognition.

Actually, the data structure of an electronic band structure is very similar to an image. To explore the electronic structure of a given material, usually we will calculate the band structures over the Brillouin zone (BZ) using a grid mesh, e.g. (10 x 10 x 10). If we only want to use, say, 8 bands around the Fermi level, we will have 10 x 10 x 10 x 8 eigenvalues. We can directly perform a 3D convolution over the 10x10x10 "cubic-image" with 8 channels or we can simply arrange the data as a 10 x 100 image with 8 channels (here, we call it generalized colors, like RGB corresponds to 3 channels) to perform 2D convolution. For each 10 x 10 x 10 (or 10 x 100) image, it reflects the energy profile of the electronic structure on an extended plane and use them as inputs to CNN or treat the band structures as sequence data as an input to RNN. Therefore, we can first construct a Slaster-Koster Hamiltonian that corresponds to our interested material and generate several dozens of thousands band structures based on randomly assigned values of the Slaster-Koster parameters as input data to train a machine as shown in Fig.1. Then use it to predict future Slaster-Koster parameters based on the band structures obtained by DFT. Actually we have tested this approach on some simple materials such as NaCl and the output Slaster-Koster parameters can indeed reproduce the band structure calculated via DFT. However, tests on more complicated materials remain needed especially for systems with serious band entanglements.

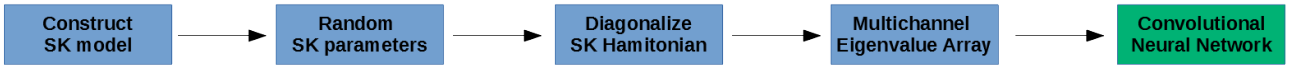

Fig 1: data flow chart of proposed training algorithm

- Build transfer learning algorithm to speed up material mining

Despite the above method is a possible approach to generate Slaster-Koster Hamiltonian, there are still some issues unsolved. Since different classes of materials will have various crystal structures and various orbitals, their model Hamiltonians would look very different and we have to retrain a machine for every class of materials. Although collecting new training data are theoretical possible (because they are generated via diagonalizing a bunch of Slaster-Koster Hamiltonian with randomly assigned values), it is still a heavy loading task. Taking a 20 x 20 x 20 grid mesh with 32 bands as an example, effectively we will have an image that contains 256,000 pixels and for each materials we should train a machine by at least 50,000-100,000 samples to get reliable predictions. Therefore, transfer learning seems a possible way to reduce the loading.

Transfer learning [13] uses the common part of a model trained by a large data set to reduce the required number of training data in a new but somewhat relevant model. This approach has been widely applied in image and speech recognition. CNN models like vgg16, vgg19, Xception, ResNet50, etc. are all famous base-models for transfer learning. In image recognition, the first few layers in a neural network correspond to the extraction of concrete features such as edge, lines, corner, etc., which are universal among most images. Therefore, if one transfers the first few layers from a well-trained image model to another image data set, the minimal number of training set would be much lesser (see Fig.2).

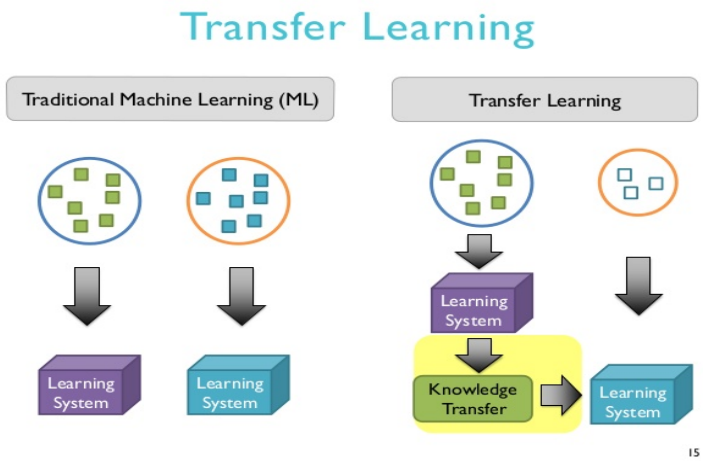

Fig 2: illustraion of transfer learning

Similar idea can be applied to our project. As mentioned above, a band structure can be considered as an image with several generalized colors on an extend plane. It essentially reflects the energy profiles within the 3D Brillouin zone. Therefore, there are also some universal features such as saddles, peaks,

plateau, band crossing, band degeneracy, etc., among all systems. If we train a model using a much larger data set from the randomly generate Slaster-Koster data and transfer these features to new models that apply to another class of materials, it could much speed up the research of “material mining”.

- Build a multitask learning so similar materials can share the same model

In the real world, there are countless number of materials. Although it is impossible to build an universal model to describe all materials, it is still possible to train a model that describes similar materials. Double perovskite materials are a paradigm. Double perovskite materials have chemcial formula: ABB’O6. Usually A-site are alkaline earth elements such as Ca, Sr, Ba, etc. B and B’ sites are usually atoms with p- or d-valence electrons. They combine together form a structure with cubic symmetry. These materials are known for high efficiency solar cells[14]. Therefore, if we can speed up the material mining research in this field, our energy industry can get benefits a lot. However, in the language of Slaster-Koster Hamiltonian, when B and B’ are p-orbital atoms (e.g. Ge, Bi, Se, Te, etc.) or d-orbitals (e.g. Fe, Co, Ni, Cu, etc.), their model would be very different since d-orbitals have different symmetry and different Slaster-Koster parameters. On the other hand, they also share some common features. For example, they both have cubic symmetry and oxygen sites with very high electron negativity that trap electrons as many as possible. Therefore, when training a neural network, it is straightforward to design a multitask learning [15]. L. Kaiser et. al., arXiv:1706.05137 (2017)

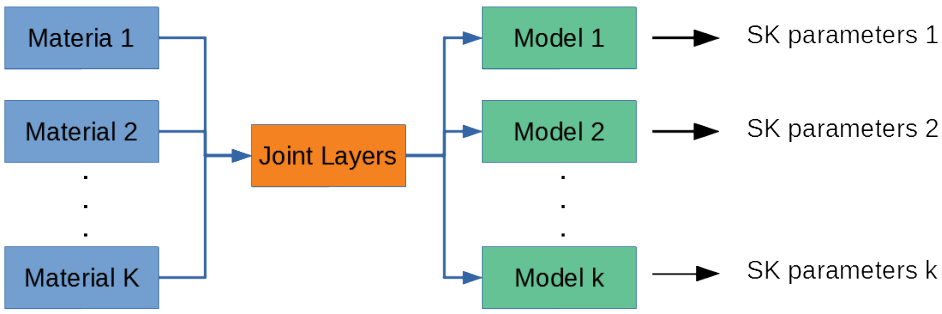

Fig 3: illustration of multitask training

Conclusion

In this proposal, I plan to study three topics:

1. Build an effective Slaster-Koster Hamiltonian mapping algorithm using CNN or RNN
2. Build transfer learning algorithm to speed up material mining
3. Build a multitask learning so similar materials can share the same model

While the first project can provide us a new and powerful tool to build a symmetry-protected model Hamiltonian that can apply to advanced material properties analysis and large scale material calculations, the later two projects can further speed up the research of material mining.

We have performed some initial attempts of the above proposals using Google’s tensorflow and Keras in combined with nVidia’s entry level GPU GeFroce 840m. Even though it’s just an low-end GPU, we

still get 408% improvement on efficiency than a intel's high end CPU i7-4710HQ in our CNN calculations. Apparently nVidia's GPU in combined with CUDA and cuDNN are much more appropriate tools in deep learning than any other choices. Since materials mining is a new and challenging field, therefore, GPU calculation would greatly benefit the project and have huge impact to the field of material science. We believe this research project is also able to explore the potential of nVidia's GPU in material science and deep learning.

In the future, we would like to extend the idea to more advanced property predictions such as directly predict complicated properties from some initial results from DFT without doing complicated postprocess calculations. I believe a powerful GPU calculation could even boost the field.

Reference

* email: spi@ucdavis.edu

[1]. J. Kohanoff, electronic structure calculations for solids and molecules, Cambridge Press, 2006
[2]. X. Gonze et. al., Comp. Phys. Comm. 205 106, 2016
[3]. N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. 84 1419 (2012)
[4]. G. Kotliar et. al., Rev. Mod. Phys. 78 865 (2006)
[5]. B. Amadon, T. Applencourt and F. Bruneval, Phys. Rev. B 89, 125110 (2014).
[6]. ST Pi, R Nanguneri, S Savrasov, Phys. Rev. Lett. 112 077203 (2014)
[7]. ST Pi, R Nanguneri, S Savrasov, Phys. Rev. B 90 045148 (2014)
[8]. AA Mostofi, JR Yates, G Pizzi, YS Lee, I Souza, D Vanderbilt, N Marzari, Comput. Phys. Commun. 185, 2309 (2014)
[9]. J. C. Slater and G. F. Koster, Phys. Rev. 94, 1498 (1954)
[10]. A. Krizhevsky, I. Sutskever, G. E. Hinton, NIPS Proceedings, 1097 (2012)
[11]. S. Hochreiter, J. Schmidhuber, Neural Computation. 9 1735 (1997)
[12]. L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, npj Comp. Mater. 2 16028 (2016)
[13]. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, arXiv 1411.1792 (2014)
[14]. A. Mei et. al., Science 345 295 (2014)
[15]. L. Kaiser et. al., arXiv:1706.05137 (2017)