MATDAT18: Materials and Data Science Hackathon
MATERIALS SCIENCE TEAM APPLICATION FORM

Complete and return via email to brian_reich@ncsu.edu by 15 January 2018

Team Composition (2 people max.)

Name	Department	Institution	Email
Jimmy-Xuan Shen	Physics	UC Santa Barbara	jshen@physics.ucsb.edu
Wennie Wang	Materials	UC Santa Barbara	wwwennie@ucsb.edu

Project Title

Rapid identification and characterization of defects in perovskite oxides

Project Synopsis (approx. 100 words)

The theory for identifying and characterizing defects and impurities has reached maturity, but remains a laborious process for charged defects. Simultaneously, charged defects play a significant role in the properties of semiconductor devices. We propose a collaborative effort to focus on identifying shallow versus deep defects in the perovskite oxides, which are promising material candidates for photovoltaic devices. Our goal is to create a framework for streamlining the identification and characterization of charged defects for electronic devices using descriptors beyond bulk materials properties. This shall be accomplished through mining of existing databases and building upon existing scalable workflows for density functional theory calculations.

Identified Data-Science Collaborative Need (approx. 100 words)

The project seeks to go beyond simple descriptors and take advantage of the orbital character data available in standard codes. We propose to be able to infer information about the defect formation energies from the orbital characteristics of the different states in the bulk band structure. This project is a key opportunity to combine the expertise and tools we have in semiconductor and defect physics with recent advancements in data science to tackle a highly specific, high impact materials problem requiring unique optimization.

Data Origin and Access (*data must be available and shareable with data science teams* – please address: data source/origin, access privileges, sharing privileges)

The comprehensive data for orbital characteristics of solid-state materials are not yet available. Existing, open access, databases like the MaterialsProject will provide us with the initial structures and basic properties for our calculations. The MaterialsProject database can be queried for all materials with the chemical formula ABO_3 (7965 distinct materials as of January 2018). For each material, DFT bulk calculations will be performed along with a series of defect calculations for the different charge states. The xml formatted VASP outputs can be stored on the NOMAD repository, which allows the simulated data to be requested and cited by a unique DOI. In order to share the results of the calculations with data science teams, the data will be stored in HDF5 format containing the orbital-projected wavefunction characters data of each band and along with the more traditionally available data such as atomic species and positions.

Project Description (approx. 1.5 pages, plus figures and references; please describe data size, form, dimensionality, uncertainties, number of examples, etc.)

Introduction and Background

Defects make the device. The basis of all electronic devices is the engineering and control of defects and impurities. The identification and characterization of deep and shallow defects is essential to device functionality [1-4]. However, this is both difficult and slow. Current methods for understanding point defects are tedious and often done on a case-by-case basis (see Fig. 1a), particularly for charged defects, which requires additional corrections due to spurious interactions between periodic images of the simulation cell [5].

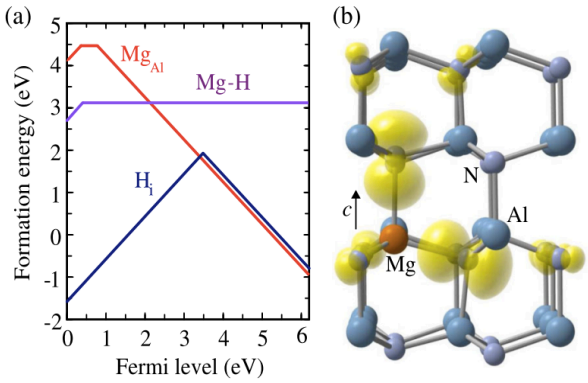

Fig. 1 (a) Example of formation energy diagram as a function of Fermi energy referenced to the valence band maximum for various substitutional and interstitial defects in AlN. The slope of each linear portion per curve represents a stable charge state. (b) Spatial charge distribution corresponding to Mg substituted on Al in the +1 charge state. Adapted from Ref. [2].

Many efforts have emerged to combine data science with materials science. These include the MaterialsProject [6], OQMD [7], AFLOW [8], NOMAD [9], and Atomate [10], which focus on various parts of the creation, collection, management, and analysis of computational materials data. This has lead to a vast exploration for materials screening and prediction [11-14].

Thus far, such efforts have built the foundation for computationally-driven materials science. Varley et al. [15] applied machine learning techniques and identified the branch-point energies of bulk structures as good descriptors for the formation energies and transition levels for II-VI and III-V compounds. Similar studies focusing on perovskites include predicting band gap [16] and predicting dielectric breakdown [17]. Baladandan et al. [18] studied formation energies and interactions of neutral oxygen vacancies, H interstitials, and substitutional dopants for fuel cell applications. However, these have mostly used descriptors based on bulk properties and chemical intuition, or focused only on neutral defects and impurities.

Proposed Project and Scope

Our aim for MATDAT18 is to create a data science framework for accurately identifying shallow versus deep dopants among possible candidate charged defects. We propose going beyond descriptors derived from bulk properties and incorporating orbital character resolved by

band and *k*-point using accurate and efficient first-principles calculations based on density functional theory. Orbital character is a promising descriptor as it can capture changes to atomic or electronic structure that occur locally (see Fig. 1b). Here, we shall study the perovskite compounds, and focus on material candidates for photovoltaic devices.

The first step will involve screening for stable compounds. As explained in the data origin/access section, we will use existing materials databases to select our candidates. Next, we will set up and execute a framework for calculating formation energies of charged defects, using the well-established method from Freysoldt [5]. Several tools, such as Atomate [10], the Atomic Simulation Environment [19] and pymatgen [6], can be adapted for our purposes for interfacing between *ab initio* codes and high-throughput calculations.

In order to cross-validate our chosen descriptor of orbital character, we will compare with explicit calculations. After calculating charge-state correction per charge state and defect, we can construct formation energy diagrams (see Fig. 1 for example). Based on formation energy diagrams, we identify charge-state transition levels, which reveals whether a defect is deep or shallow.

Data will involve collecting and processing multi-dimensional data in a combination of text and xml formats. Formation energies for the defects require only total energies of the defect-cell calculations. For the bulk calculations, the orbital projection data will be stored as a $K \times B \times N \times 10$ matrix, where K is the number of *k*-points in the calculation, B is the number of bands in the calculation, N is the number of atoms in the unit cell, and 10 represents the different *spd* orbitals at each atom onto which the wavefunction can be projected. For a particular materials system, files will at most be on the order of 130 MB.

Our main uncertainty is how well the band edges and band gaps are described with our calculations. Our goal is to achieve accuracies of better than state-of-the-art < 0.2 eV mean absolute error in charge state transition energies. We aim to integrate this framework into the relevant aforementioned efforts. We expect this effort to offer new insights into the relation between local physics and defect behavior as well as a pathway to integrate more complex materials phenomena into existing databases.

References

[1] Van de Walle, C. G., & Neugebauer, J. (2004). *Journal of Applied Physics*, 95(8), 3851.
[2] Lyons, J. L., Janotti, A., & Van de Walle, C. G. (2012). *Physical Review Letters*, 108(15), 156403.
[3] Baranowski, L. L., Zawadzki, P., Lany, S., Toberer, E. S., & Zakutayev, A. (2016). *Semiconductor Science and Technology, 31*(12), 123004.
[4] Smyth, D. M. (1985). *Annual Review of Materials Science, 15*(1), 329–357.

[5] Freysoldt, C., Grabowski, B., Hickel, T., Neugebauer, J., Kresse, G., Janotti, A., & Van de Walle, C. G. (2014). *Reviews of Modern Physics, 86*(1), 253–305.
[6] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., … Persson, K. A. (2013). *APL Materials*, *1*(1), 11002.
[7] Kirklin, S., Saal, J. E., Meredig, B., Thompson, A., Doak, J. W., Aykol, M., … Wolverton, C. (2015). *Nature Publishing Group, 1*.
[8] Calderon, C. E., Plata, J. J., Toher, C., Oses, C., Levy, O., Fornari, M., … Curtarolo, S. (2015). *Computational Materials Science, 108*, 233–238.
[9] Ghiringhelli, L. M., Carbogno, C., Levchenko, S., Mohamed, F., Huhs, G., Lüders, M., … Scheffler, M. (2017). *Npj Computational Materials*, *3*(1), 46.
[10] Mathew, K., Montoya, J. H., Faghaninia, A., Dwarakanath, S., Aykol, M., Tang, H., … Jain, A. (2017). *Computational Materials Science*, *139*, 140–152.
[11] Meredig, B., Agrawal, A., Kirklin, S., Saal, J. E., Doak, J. W., Thompson, A., … Wolverton, C. (2014). *Physical Review B*, *89*(9), 94104.
[12] Dagdelen, J., Montoya, J., de Jong, M., & Persson, K. (2017). *Nature Communications*, *8*(1), 323.
[13] Goldsmith, B. R., Boley, M., Vreeken, J., Scheffler, M., & Ghiringhelli, L. M. (2017). New Journal of Physics, 19(1), 13031.
[14] Seko, A., Maekawa, T., Tsuda, K., & Tanaka, I. (2014). *Physical Review B - Condensed Matter and Materials Physics*, *89*(5), 1–9.
[15] Varley, J. B., Samanta, A., & Lordi, V. (2017). *The Journal of Physical Chemistry Letters*, *8*(20), 5059–5063.
[16] Pilania, G., Mannodi-Kanakkithodi, A., Uberuaga, B. P., Ramprasad, R., Gubernatis, J. E., & Lookman, T. (2016). *Scientific Reports*, *6*(1), 19375.
[17] Kim, C., Pilania, G., & Ramprasad, R. (2016). *The Journal of Physical Chemistry C*, *120*(27), 14575–14580.
[18] Balachandran, J., Lin, L., Anchell, J. S., Bridges, C. A., & Ganesh, P. (2017). *The Journal of Physical Chemistry C*, *121*(48), 26637–26647.
[19] Hjorth Larsen, A., Jørgen Mortensen, J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., … Jacobsen, K. W. (2017). *Journal of Physics: Condensed Matter*, *29*(27), 273002.