TYPE I: MACHINE LEARNING FOR STRUCTURE-PERFORMANCE RELATIONSHIPS IN ORGANIC SEMICONDUCTING DEVICES

PI: Eric Jankowski (Boise State University, ericjankowski@boisestate.edu)
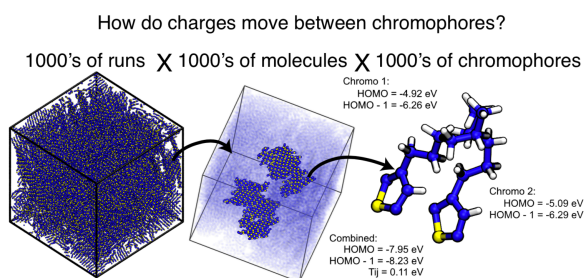**Team Members**: Matthew Jones (Boise State University, mattyjones@boisestate.edu), Evan Miller (Boise State University, evanmiller326@boisestate.edu), Bryan Stanfill (Pacific Northwest National Laboratory bryan.stanfill@pnnl.gov)

The goal of the proposed work is to transform the engineering of advanced electronic devices made from organic semiconducting molecules, enabling new solar panels, thermoelectrics, and video displays to be inexpensively deployed. The radically new approach is to replace expensive quantum chemical calculations with models trained on large data sets, significantly accelerating computer simulations that predict the electronic properties of soft material systems.

Preliminary successes enabled through the MATDAT18 hackathon underpin this work.

INTELLECTUAL MERIT

The challenge we address here is understanding how the chemistry and packing of photoactive molecules affects the motion of electrons and holes in organic active layers whose morphology has been predicted with molecular simulations. Ordinarily, this requires the execution of a large number of slow quantum chemical calculations to predict the electronic properties of the materials, demanding large amounts of time and computing power[1]. Our hypothesis is that machine learning techniques, trained on a small subset of quantum chemical data, can be used to predict the electronic capabilities of nanostructured morphologies. This will dramatically reduce the number of expensive calculations to be performed and facilitate high-throughput parameter-sweep studies of the vast phase space of candidate chemistries and processing methodologies, with the aim of developing a robust set of design-rules for manufacturing efficient devices.

Figure 1. Predicting charge mobility for a single simulation snapshot requires quantum chemical calculations be performed on each pair of chromophores. Machine learning techniques represent a way to obtain electronic transfer integrals between chromophore pairs, saving potentially billions of unnecessary, expensive chemical calculations per semiconductor study.

In summary, we propose performing the quantum chemical, soft matter, and machine learning computations to inform the ***fastest, most general model of organic charge transport properties in existence.*** Imagine being able to query the charge hopping rate between two arbitrary chromophores, retrieving an answer in microseconds rather than days, when hundreds of thousands of such calculations are needed to understand a single morphology. Furthermore, the fundamental understanding of machine learning model accuracy and transferability as a function of the training set size and quality developed herein has the potential to inform machine learning strategies for problems across the field of condensed matter.

PRELIMINARY WORK AND RESEARCH PLAN

At the two-and-a-half day MATDAT18 hackathon, we explored artificial neural networks, random-forest decision trees, support vector machines, simple linear regression, and semi-parametric penalized splines in predicting transfer integrals of poly(3-hexylthiophene) (P3HT) chromophores, each of which were trained on the same 135,000 quantum chemical calculations. A data science shortcoming exposed in this exercise is the fact that so many modern analytical techniques are geared towards the big data problem both in the feature space and dataset size. In this problem, there are a large number of data points to train on, but the feature set is relatively small (~10 features). Compounding this issue is the highly non-linear

B
BOISE STATE UNIVERSITY

relationship between the features and output, which makes it difficult to faithfully model with most machine learning techniques. Here, random-forest decision tree techniques proved to be both the most accurate and produced transparent feature importance metrics. The random forest fits resulted in a correlation coefficient of $R^2 = 0.978$, corresponding to an average absolute error of 30 meV, far below the expected 100 meV accuracy of even the most computationally intensive quantum chemical predictions (Figure 2). In short, *a machine learning model trained on transfer integrals in an afternoon from a single ordered simulation snapshot of P3HT is able to predict transfer integrals for both ordered and disordered morphologies with over 96% accuracy, in seconds rather than days*. This unexpectedly transferrable predictive capability for a simple random forest model provides compelling rationale for determining the how well such approaches can perform for arbitrary chemistries.
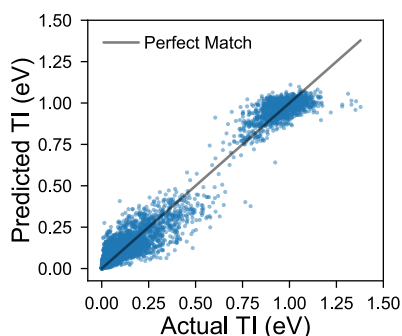
Figure 2. Prediction quality for the random forest model developed at MATDAT18. The average absolute error is 30 meV and $R^2=0.978$.

Beyond the high model accuracy after two days of dedicated work, two related observations further inform our hypothesis that a general transfer-integral model can be created with random forests. First is that when the P3HT model was applied to morphologies of a completely different macromolecule, the transfer integral predictions were better-than-random. This indicates that the center-of-mass position and orientation used by the random forest are (perhaps unsurprisingly) transferable metrics that determine the magnitude of transfer integrals between arbitrary chromophores. The second is that the 135,000 chromophore positions and orientations that were used in our preliminary work were not uniformly distributed: We realized that instead of drawing chromophore training data from simulation snapshots, we could more efficiently inform our machine learning techniques by iterating over specified distributions of relative position and orientation to maximize the information content of a training set.

In concert, our preliminary results suggest that by iterating over a minimal set of positions and orientations for a minimal (large, but "minimal") set of chromophore chemistries, an effectively exhaustive model of transfer integrals between arbitrary chromophores can be constructed. We propose performing the soft matter simulations, quantum chemical calculations, and data science investigations to measure the degree to which this can be accomplished.

The specific aims of this research are to:

1. *Generate transfer integral training data for 50 chromophore chemistries*—systematically determine the discretizations of distance- and rotation- space needed to capture transfer integral variance.

2. *Generate transfer integral training data for 1000 chromophore pairs*— systematically determine the discretizations of distance- and rotation- space needed to capture transfer integral variance for dissimilar chemistries

3. *Optimize machine learning models for predicting transfer integrals*—determine which molecular descriptors and training data characteristics are required to constrain average model error below 10 meV.

4. *Validate charge mobilities*—perform molecular dynamics simulations of 100 candidate chemistries available in the organic semiconductor literature, followed by kinetic Monte Carlo simulations, to measure the degree to which predicted transfer integrals and self-assembled morphology are predictive of zero-field charge mobility.

RESOURCES & BUDGET

The project team will continue to include PI Jankowski, postdoctoral researcher Matthew Jones, graduate student Evan Miller, and unfunded collaborator Bryan Stanfill. The proposed total budget of $250,000 over two years will support salary for PI Jankowski (1mo/yr), 60% of Postdoc Jones, and 100% of graduate student Miller, plus conference travel and publication costs in the second year.

All of the components of the proposed work requiring high performance computing resources will be performed at Boise State University. The NSF-funded Kestrel GPU cluster (Kestrel) (Award #1229709), PI Jankowski's cluster (Fry), and Boise State's research computing cluster (R2) will be employed for the proposed simulation research, a total of 753,360 GPU-hours/year. Kestrel provides 44 NVIDIA Kepler K20 GPU processors, Fry, provides 32 NVIDIA Pascal P100 GPUs, and R2 provides 10 P100 GPUs that are ideally suited for the proposed molecular dynamics simulations, quantum chemical calculations, and kinetic Monte Carlo simulations, and have been used to benchmark preliminary simulation results. The anticipated computational needs of the proposed work represent less than 5% of the computational resources available at Boise State.

BROADER IMPACTS

The ability to engineer low-cost solar panels has transformative potential to benefit society through the worldwide deployment of sustainable energy production. No other single advance will as broadly improve the health, security, and opportunity of global citizens[2–4].

The Jankowski lab is committed to maintaining diverse perspectives in an inclusive environment for STEM training, and *if funded, the proposed work will broaden PI Jankowski's existing mentoring and dissemination activities*. Eight out of fourteen undergraduate researchers mentored by PI Jankowski are female. PI Jankowski leads the *Inclusive Computing Initiative* at Boise State University, which has trained over 200 students in computational literacy through *Software Carpentry* workshops. PI Jankowski leads educational research into how student identity influences performance and retention in materials science, in particular through collaboration with *The Story Collider*, providing an international podcast audience for his thermodynamics students sharing stories of "True, personal stories about a time thermodynamics happened." These storytelling interventions show promise in improving student self-identification as materials scientists and increased public engagement with science and technology.

REFERENCES

1. Jones, M. L., Huang, D. M., Chakrabarti, B., and Groves, C. "**Relating Molecular Morphology to Charge Mobility in Semicrystalline Conjugated Polymers**" *Journal of Physical Chemistry C* **120**, 4240–4250 (2016)

2. Espinosa, N., Hösel, M., Angmo, D., and Krebs, F. C. "**Solar Cells with One-Day Energy Payback for the Factories of the Future**" *Energy & Environmental Science* **5**, 5117 (2012)

3. Mazzio, K. A. and Luscombe, C. K. "**The future of organic photovoltaics**" *Chemical Society Reviews* **44**, 78–90 (2015): Available at http://dx.doi.org/10.1039/C4CS00227J

4. Conti, J., Holtberg, P., Diefenderfer, J., LaRose, A., Turnure, J., and Westfall, L. "**International Energy Outlook 2016**" *U.S. Energy Information Administration*: Available at www.eia.gov/forecasts/ieo/pdf/0484(2016).pdf

B
BOISE STATE UNIVERSITY