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

Complete and return via email to brian_reich@ncsu.edu by 19 January 2018

Team Composition (2 people max.)

Name	Department	Institution	Email
Matthew Jones	Micron School of Materials Science and Engineering	Boise State University	MattyJones@boisestate.edu
Evan Miller	Micron School of Materials Science and Engineering	Boise State University	Evanmiller326@boisestate.edu

Project Title

Machine Learning for Structure-Performance Relationships in Organic Semiconducting Devices

Project Synopsis (approx. 100 words)

Organic semiconducting materials have the potential to provide an inexpensive and tunable alternative to conventional inorganic materials for use in the construction of electronic devices. The performance of these devices depends on the movement of charges through the fine intermolecular structure. Computational methods can predict these structures and subsequent electronic properties for the wide variety of candidate molecules, however, it is too computationally expensive to calculate the properties for the many combinations of molecules. To overcome this, we propose applying machine learning to predict electronic properties, thereby reducing computational bottlenecks and enabling a widespread investigation of the variables affecting device performance.

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

The rate at which a charge is able to move from one electronically active part of a molecule (chromophore) to another depends on the relative chromophore positions and orientations (transfer integral), as well as the energetic differences between the molecules (energy difference). Currently, we use computationally expensive quantum chemical calculations to identify how quickly charges can move between chromophores. We hypothesize that, after calibration, machine learning can be used to predict the transfer integrals and energy differences for the system, without performing these calculations. We would need a data-science expert to assist in implementing the machine learning techniques to convert the tens of millions of chromophore pair conformations we have into suitable inputs for a machine learning model, as well as help us calibrate the model to successfully predict unknown transfer integrals and energy differences for new chromophore conformations.

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

Data is generated on high-performance clusters located at Boise State University and clusters available through XSEDE (primarily XStream at Stanford). The code used to generate the data is open-sourced and freely-available to download at https://bitbucket.org/cmelab/morphct. The data themselves will be hosted online through the Albertsons Library at Boise State University (http://scholarworks.boisestate.edu/cme_lab/) in order to make the data freely available to the data science team.

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

<Please see attached>

The goal of the proposed work is to understand electron and hole transport in organic semiconductors, enabling the mitigation of global climate change through the production of high-efficiency, low cost solar panels. The challenge we address here is understanding how the chemistry and packing of photoactive molecules influences the ability of electrons and holes to move through the solar cell's active layer. We propose using machine learning techniques to replace expensive quantum chemical calculations needed in the prediction of charge mobility.

Organic semiconductors are becoming an increasing popular alternative to conventional inorganics in the construction of electronic devices, including thin-film transistors[1], light-emitting diodes[2] and photovoltaics[3] because of recent advances in synthetic chemistry and low-cost scalable manufacturing processes. Charge-carrier mobility describes the speed at which electrons and holes can move through the active layer of the device, and is a crucial factor in device efficiency[4]. The mobility often depends sensitively on the morphology of the active layer, which describes the relative positions and orientations of the component molecules. Therefore, in order to manufacture the most efficient devices, it is vital to optimize the morphology such that the charge-carrier mobility is maximized. The molecular morphology can be influenced by the choices of chemistries in the system, as well as the device processing conditions such as temperature, pressure, solvent choice, and annealing duration[5]. This massive phase space necessitates the use of computational methods (rather than manufacturing hundreds of millions of test devices in a wet lab) that are capable of spanning multiple length- and time-scales.

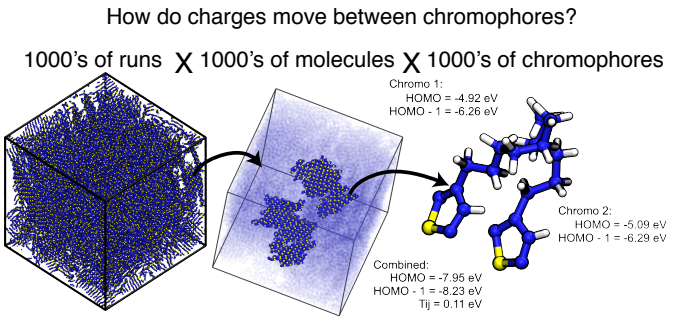

Figure 1: Predicting charge mobility for a single simulation snapshot requires quantum chemical calculations be performed on each pair of chromophores. Machine learning techniques represent a way to obtain transfer integrals between chromophore pairs, saving billions of unnecessary, relatively expensive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

Carriers move through the morphology via quantised tunnelling events - 'hops' - between electronically active functional groups on the molecules known as 'chromophores'. The rate at which a carrier hop can occur from chromophore i to chromophore j, k_{ij}, is given by the semi-classical Marcus expression[6]:

$$k_{ij} = \frac{|T_{ij}|^2}{\hbar}\sqrt{\frac{\pi}{\lambda k_B T}}\exp\left[-\frac{(\Delta E_{ij}+\lambda)^2}{4\lambda k_B T}\right], \quad (1)$$

where T_{ij} is the electronic transfer integral, ΔE_{ij} is the difference in energy between the initial and final hop sites, and the remaining parameters are material-specific, thermodynamic or fundamental constants. The speed at which a hop from one chromophore to a neighbour can occur is primarily governed by T_{ij}, which is a measure of the amount of molecular orbital overlap between the pair.

Current state-of-the-art predictions of mobility combine computational techniques: molecular dynamics simulations to obtain a candidate morphology, quantum chemical calculations to determine the transfer integrals and hop rates between chromophores, and kinetic Monte Carlo to simulate charge motion through the device (Figure 1)[7,8]. These simulations can take several days to run on supercomputers even with GPU acceleration hardware for just a single selection of component molecules and device processing techniques. Optimising the simulation pipeline will dramatically improve computational throughput of the screening process required to detect combinations of molecules and processing that will result in the most efficient devices. One area of opportunity we have identified is the calculation of transfer integrals via quantum chemical calculations: of the 10,000-10,000,000 chromophore pairs that make up a single simulation snapshot, many

pairs share the same local structure and therefore have the same transfer integrals. Performing quantum chemical calculations for each pair of chromophores therefore represents an inefficiency compared to a sufficiently accurate pattern recognition scheme that can look up transfer integrals based on their local structure.

We propose using Data Science techniques such as machine learning and neural networks to streamline the quantum chemical and Monte Carlo portions of the pipeline by predicting carrier transfer integrals between pairs of chromophores. Neural networks are an especially promising candidate due to their efficient parallelizability to be executed on GPUs, bringing the rest of the simulation pipeline in line with the molecular dynamics simulations[9]. We will train our model using the wealth of data already obtained from the pipeline and by providing key structural descriptors such as position and orientation of chromophore pairs in the system, and then measure its accuracy in predicting transfer integrals for preliminary data left out of the training set. We also aim to train a separate network to predict the relative deviations in energy levels between chromophores to obtain the ΔE_{ij} term in equation 1, effectively replacing both of the slowest components of the pipeline.

The current dataset to be used for training and testing consists of around 500 unique morphologies, covering 10 chemistries including polymers, fullerenes, block co-polymers and polycyclic aromatic hydrocarbons, each with at least 3 processing state points above and 3 state points below an order-disorder transition temperature for each chemistry. Each morphology contains, on average, 100,000 atoms resulting in 20,000 chromophore pairs per morphology. Each pair creates an output file in text format, with size $\sim$50 KB, describing the 3-dimensional positions of the constituent atoms as well as the scalar molecular orbital energies. In total, we have already generated electronic properties data for over 10,000,000 chromophore pairs, corresponding to several hundreds of GB of raw data. All data was generated using the open source MorphCT[8], HOOMD-Blue[10], and ORCA[11] software suites and will be made freely available through digital hosting provided by the Albertsons Library at Boise State University.

References

[1] A. Tsumura, H. Koezuka, and T. Ando. Macromolecular Electronic Device: Field-Effect Transistor with a Polythiophene Thin Film. *Applied Physics Letters*, 49(18):1210–1212, nov 1986.

[2] R. H. Friend, R. W. Gymer, A. B. Holmes, J. H. Burroughes, R. N. Marks, C. Taliani, D. D. C. Bradley, D. A. Dos Santos, J. L. Brédas, M. Lögdlund, and W. R. Salaneck. Electroluminescence in Conjugated Polymers. *Nature*, 397(6715):121–128, 1999.

[3] N. S. Sariciftci, L. Smilowitz, A. J. Heeger, and F. Wudl. Photoinduced Electron Transfer from a Conducting Polymer to Buckminsterfullerene. *Science*, 258(5087):1474–1476, 1992.

[4] H. Sirringhaus. 25th Anniversary Article: Organic Field-Effect Transistors: The Path Beyond Amorphous Silicon. *Advanced Materials*, 26(9):1319–1335, 2014.

[5] R. Noriega, A. Salleo, and A. J. Spakowitz. Chain Conformations Dictate Multiscale Charge Transport Phenomena in Disordered Semiconducting Polymers. *Proceedings of the National Academy of Sciences*, 110(41):16315–16320, 2013.

[6] R. A. Marcus. Chemical and Electrochemical Electron-Transfer Theory. *Annual Review of Physical Chemistry*, 15(1):155–196, 1964.

[7] M. L. Jones, D. M. Huang, B. Chakrabarti, and C. Groves. Relating Molecular Morphology to Charge Mobility in Semicrystalline Conjugated Polymers. *Journal of Physical Chemistry C*, 120(8):4240–4250, 2016.

[8] M. L. Jones and E. Jankowski. Computationally Connecting Organic Photovoltaic Performance to Atomistic Arrangements and Bulk Morphology. *Molecular Simulation*, 43(10-11):756–773, 2017.

[9] K.-S. Oh and K. Jung. GPU Implementation of Neural Networks. *Pattern Recognition*, 37(6):1311–1314, 2004.

[10] J. A. Anderson, C. D. Lorenz, and A. Travesset. General Purpose Molecular Dynamics Simulations Fully Implemented on Graphics Processing Units. *Journal of Computational Physics*, 227(10):5342–5359, 2008.

[11] F. Neese. The ORCA Program System. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78, 2012.