ACED: ACCELERATED COMPUTATIONAL ELECTROCHEMICAL SYSTEMS DISCOVERY

Rachel C. Kurchin, Lance Kavalsky, Xiaoyu Sun, Venkatasubramanian Viswanathan
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{rkurchin,seansun,venkatv}@andrew.cmu.edu

Eric Muckley, Vinay Hegde, James Saal & Bryce Meredig
Citrine Informatics
Redwood City, CA 94063, USA
{emuckley, vhegde, jsaal, bryce}@citrine.io

Dhairya Gandhi & Viral Shah
Julia Computing
Cambridge, MA 02139, USA
{dhairya, viral}@juliacomputing.com

Matthew Johnson, Alan Edelman, & Christopher Vincent Rackauckas
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
{mattsj, edelman, crackauc}@mit.edu

ABSTRACT

Large-scale electrification is vital to addressing the climate crisis, but many engineering challenges remain to fully electrifying both the chemical industry and transportation. In both of these areas, new electrochemical materials and systems will be critical, but developing these systems currently relies heavily on computationally expensive first-principles simulations as well as human-time-intensive experimental trial and error. We propose to develop an automated workflow that accelerates these computational steps by introducing both automated error handling in generating the first-principles training data as well as physics-informed machine learning surrogates to further reduce computational cost. It will also have the capacity to include automated experiments "in the loop" in order to dramatically accelerate the overall materials discovery pipeline.

1 ELECTROCHEMISTRY AND CLIMATE CHANGE

Electrification of virtually every energy-consuming sector is critical in the fight against climate change, as it will enable society to rely on carbon-free energy sources such as solar and wind. Increased performance and reduced cost of electrochemical technologies will be key to this electrification process. Many are already familiar with phenomena such as "range anxiety", as well as sticker shock for electric vehicles. Perhaps less familiar are the large swaths of the chemical industry that rely on extreme conditions (heat and pressure) produced by burning fossil fuels for large-scale synthesis of chemicals that are essential for fertilizers, steel, cement and other aspects of modern life that many of us take for granted.

To meet the technoeconomic targets posed by these challenges, novel materials and systems will need to be designed. In this project, we propose to develop a generalizable and automated workflow for discovering and developing these materials and systems. It will have the capability to build the necessary first-principles models and use their results to train machine learning surrogates that can be evaluated orders magnitude more quickly. The results of these models can be used in a series of

coarser models to bridge from the atomic to the device scales and evaluate performance potential in a real system, including the possibility for an "in-the-loop" automated experimental evaluation. Using sequential learning, we can then select the next candidate from a specified design space and proceed through the workflow as many times as necessary to meet target performance specifications.

2 Proposed Workflow

2.1 Overview

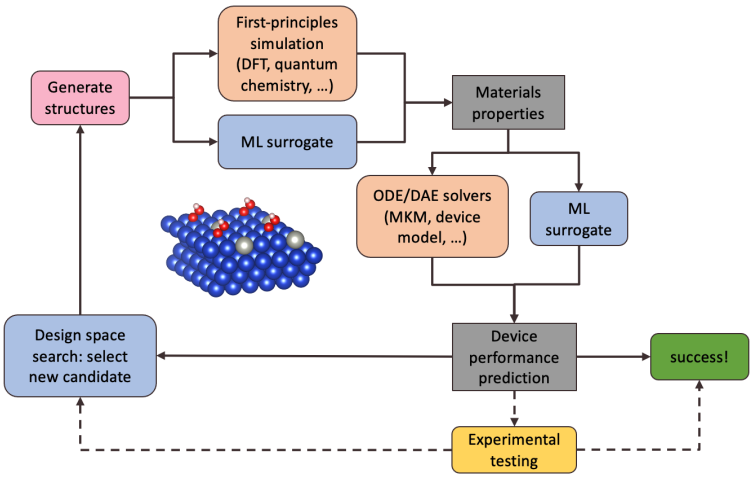

Figure 1: ML-aided materials discovery workflow. See Section 2.1 for a detailed description.

The workflow we envision is summarized at a high level in Figure 1. The first step (upper left) is to generate relevant atomic structures for the candidate molecules/materials. Initially, these will be fed into a first-principles simulation engine such as density functional theory (DFT) or a quantum chemistry code. These calculations result in predicted materials properties that serve as parameters in a larger-scale model (such as one describing chemical kinetics of catalysis or the operation of a battery), from which device performance can be predicted. If the workflow is purely computational, this result will determine if the search has succeeded or whether a new candidate needs to be identified. If an automated experiment is "in the loop," then this serves as a decision point for whether an experiment should be done. Assuming performance criteria have not yet been met, a new candidate material/molecule is selected using a sequential learning algorithm, and structures are generated to proceed through the loop again.

Once enough first-principles simulations have been run, the resulting data can be used to train an ML surrogate model to speed up candidate evaluation. A separate surrogate can be trained to accelerate the device modeling step.

In addition to building these ML surrogates, we are also placing emphasis on automating any step in this workflow that currently requires human intervention, with the ultimate goal of the entire materials discovery loop being autonomous. For example, running forward models (orange boxes in Figure 1) often requires, in addition to substantial computational resources, multiple steps of human decision-making, either to determine which specific calculations to do, or to resolve convergence errors arising from parameter choices. A significant portion of this project is developing an automated DFT workflow that can proceed entirely without this type of human input and hence eliminate associated delays.

The machine learning models for this work (with the exception of the design space search described below, for which a large body of code in Python already existed) take advantage of the Julia programming language's unique combination of ease of use and best-in-class performance, as well as the existence of a robust language-wide automatic differentiation system (Innes, 2019). In partic-

ular, we use the Flux machine learning library (Innes, 2018) and the DifferentialEquations.jl package (Rackauckas & Nie, 2017) extensively in this work.

While the strict data-dependencies of the workflow would traditionally require the all steps of the simulation process to be done sequentially, we are breaking the flow by training ML surrogates of the microkinetic models (MKM) during the first-principles simulations. This allows us to amortize the training time during the previous step of the process and thus receive the benefits of the ML augmentation while masking its cost. We plan to develop a surrogate of the whole MKM analysis process, i.e. a surrogate from the materials properties directly to the device performance prediction, so that the moment the DFT calculations are completed the neural network will bypass the stiff ill-conditioned kinetic simulation and directly predict the outcomes.

Search of the design space and candidate selection will be carried out using Citrine's Citrination cloud-based machine learning engine. Communication with Citrination will primarily be executed using the Citrination Python API (Citrine). Modeling of the design space will be performed using lolo (Citrine Informatics), Citrine's custom random forest regression algorithm which incorporates uncertainty estimates. Material candidates will be chosen based a strategy which balances greedy and exploratory selection strategies including maximum likelihood of improvement (MLI), maximum expected improvement (MEI), and maximum uncertainty (MU) (Ling et al., 2017). This processed will be repeated for each material candidate of interest to enable iterative sequential learning of material properties across the design space.

3 Case Studies

3.1 Phase I: Electrochemical Nitrogen Reduction

The nitrogen reduction reaction (NRR) is central to global food supply as it produces ammonia, a critical component in fertilizers. Roughly 80% of the nitrogen in an average human body today has been produced through the Haber-Bosch process (Howarth, 2008), the state-of-the-art industrial method for NRR. However, this process takes place at extreme pressure ($\sim$10MPa) and temperature (400-500 ^{o}C), conditions that require burning fossil fuels to achieve.

A promising alternative approach is electrochemical nitrogen reduction, where the activation energy currently provided by heat and pressure is instead supplied by electric voltage. This approach currently faces obstacles due to low activity and selectivity of catalysts. In this work, we will investigate two promising catalyst design spaces with potential to surmount these obstacles: single-atom catalysts and multi-principal-element alloys.

To substitute for DFT calculations to obtain binding energies for NRR intermediates with our candidate catalysts, we make use of atomic graph convolutional neural nets. This concept was originally popularized by Xie & Grossman (2018) for bulk crystals, and we are developing the AtomicGraphNets.jl package (currently available on GitHub, eventually in the Julia Package Registry) as a flexible implementation for crystals and molecules. In support of this, we are also developing ChemistryFeaturization.jl to provide a unified interface for building graphs from a variety of input structure files, and assigning feature matrices using data from several online databases.

3.2 Phase II: Novel Li Battery Electrolytes

Achieving higher specific energy (i.e. energy per unit mass) and power in batteries, as well as higher energy and power density (per volume), is critical to further expanding electric transportation (Sripad & Viswanathan, 2017) as well as to eventually electrify flight (Fredericks et al., 2018). A promising way to achieve these targets is by shifting to metallic lithium as the anode. Removing the typical graphite anode reduces both weight and volume and increases voltage and power capabilities, but introduces challenges due to the tendency of lithium to form dendrites during charging. These dendrites can reach across the separator and short the device, reducing cycle life, and in the worst cases, cause dangerous fires.

Critical to ameliorating these issues is the development of novel battery electrolytes that can block dendrites from growing. Typically, screening of candidate electrolyte molecules makes use of computationally intensive quantum chemistry simulations. However, there are several promising ML

approaches to building fast and accurate surrogates for such calculations. We are in the process of building the Julia language port of the popular DeepChem machine learning package (Ramsundar et al.). These property prediction methods will be used in conjunction with pseudo 2-dimensional porous electrode models for predicting the performance of lithium metal batteries.

4 Conclusion

Rapid materials discovery is critical in variety of climate change challenges, including and especially electrifying the chemical and transportation industries. Given the urgency of the climate challenge, we no longer have the luxury of time to go about materials discovery in the "traditional" paradigm. Machine learning approaches along with automation of both simulation and experimentation have great potential to dramatically accelerate the cycle of learning and help us to discover and develop the new materials and systems that will be essential in engineering a green future.

Acknowledgments

We gratefully acknowledge the ARPA-E DIFFERENTIATE program for financial support of this work.

References

Citrine. Citrination Python API. https://github.com/CitrineInformatics/python-citrination-client.

Citrine Informatics. lolo machine learning library. https://github.com/CitrineInformatics/lolo.

William L Fredericks, Shashank Sripad, Geoffrey C Bower, and Venkatasubramanian Viswanathan. Performance metrics required of next-generation batteries to electrify vertical takeoff and landing (VTOL) aircraft. *ACS Energy Letters*, 3(12):2989–2994, 2018.

Robert W Howarth. Coastal nitrogen pollution: a review of sources and trends globally and regionally. *Harmful algae*, 8(1):14–20, 2008.

Michael Innes. Don't unroll adjoint: differentiating SSA-form programs. *arXiv:1810.07951 [cs]*, March 2019. URL http://arxiv.org/abs/1810.07951. arXiv: 1810.07951.

Mike Innes. Flux: Elegant machine learning with Julia. *Journal of Open Source Software*, 3(25): 602, 2018. ISSN 2475-9066. doi: 10.21105/joss.00602.

Julia Ling, Maxwell Hutchinson, Erin Antono, Sean Paradiso, and Bryce Meredig. High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integrating Materials and Manufacturing Innovation*, 6(3):207–217, 2017.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia. *Journal of Open Research Software*, 5(1), 2017.

B Ramsundar, P Eastman, E Feinberg, J Gomes, K Leswing, A Pappu, M Wu, and V Pande. Deepchem: Democratizing deep-learning for drug discovery, quantum chemistry, materials science and biology.

Shashank Sripad and Venkatasubramanian Viswanathan. Performance metrics required of next-generation batteries to make a practical electric semi truck. *ACS Energy Letters*, 2(7):1669–1673, 2017.

Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters*, 120(14):145301, 2018. ISSN 10797114. doi: 10.1103/PhysRevLett.120.145301.