

2023 INCITE Proposal Submission

Proposal

Title: Machine learning classical potentials for end-to-end design of organic electrolytes

Principal Investigator: Rafael Gomez-Bombarelli

Organization: MIT DMSE

Date/Time Generated: 6/17/2022 6:13:03 PM

Section 1: PI and Co-PI Information

Question #1

***Principal Investigator:** The PI is responsible for the project and managing any resources awarded to the project. If your project has multiple investigators, list the PI in this section and add any Co-PIs in the following section.*

Principal Investigator

First Name

Rafael

Last Name

Gomez-Bombarelli

Organization

MIT

Email

rafagb@mit.edu

Work Phone

617-253-5632

Address Line 1

77 Massachusetts Ave

Address Line 2

Bldg 13-5037

City

Cambridge

State

MA

Zip Code

02139

Question #2

Co-PI (s)

First Name

Yang

Last Name

Shao-Horn

Organization

MIT

Email

shaohorn@mit.edu

First Name

Bill

Last Name

Blaiszik

Organization

Materials Data Facility

Email

blaiszik@uchicago.edu

Question #3

***Institutional Contact:** For the PI's institution on the proposal, identify the agent who has the authority to review, negotiate, and sign the user agreement on behalf of that institution. The person who can commit an organization may be someone in the contracts or procurement department, legal, or if a university, the department head or Sponsored Research Office or Grants Department.*

Institutional Contact

Institutional Contact Name

Courtney Bensey

Institutional Contact Phone

617-324-7533

Institutional Contact Email

cbensey@mit.edu

Section 2: Project Information

Question #1

Select the category that best describes your project.

Research Category

Materials Science: Materials Discovery, Design, and Synthesis

Question #2

Please provide a project summary in two sentences that can be used to describe the impact of your project to the public (50 words maximum)

Project Summary

We will use machine learning to automate the parameterization of interatomic potentials for molecular dynamics simulations of solid polymer electrolytes. This will allow us to screen potential candidate systems while also providing a comprehensive single-reference dataset for determination of underlying design principles for novel electrolyte systems.

Section 3: Early Career Track

Question #1

Early Career

Starting in the INCITE 2022 year, INCITE is committing 10% of allocatable time to an [Early Career Track](#) in INCITE. The goal of the early career track is to encourage the next generation of high-performance computing researchers. Researchers within 10 years from earning their PhD (after December 31st 2012) may choose to apply. Projects will go through the regular INCITE Computational Readiness and Peer Review process, but the INCITE Management Committee will consider meritorious projects in the Early Career Track separately.

Who Can Apply: *Researchers less than 10 years out from their PhD that need LCF-level capabilities to advance their overall research plan and who have not been a previous INCITE PI.*

How to Apply:

In the regular application process, there will be a check-box to self-identify as early career.

- The required CV should make eligibility clear.*
- If awarded, how will this allocation fit into your overall research plan for the next 5 years?*

Projects will go through the regular INCITE review process. The INCITE Program is targeting at least 10% of allocatable time. When selecting the INCITE Career Track, PIs are not restricted to just competing in that track.

- What is the Early Career Track?*
 - The INCITE Program created the Early Career Track to encourage researchers establishing their research careers. INCITE will award at least 10% of allocatable time to meritorious projects.*
- Will this increase my chances of receiving an award?*
 - Potentially, this could increase chances of an award. Projects must still be deemed scientifically meritorious through the review process INCITE uses each year.*
- What do I need to do to be considered on the Early Career Track?*
 - In the application process, select 'Yes' at 'If you are within 10 years of your PhD, would*

you like to be considered in the Early Career Track?’ You will need to write a paragraph about how the INCITE proposal fits into your 5-year research and career goals.

- *What review criteria will be used for the Early Career Track?*
 - *The same criteria for computational readiness and scientific merit will be applied to projects in the Early Career Track as will be applied to projects in the traditional track. The difference will be manifest in awards decisions by the INCITE management committee.*
-

Early Career Track

If you are within 10 years of your PhD, would you like to be considered in the Early Career Track? Choosing this does not reduce your chances of receiving an award.

No

If ‘yes’, what year was your PhD? If ‘no’ enter N/A

N/A

If ‘yes’, how will this allocation fit into your overall research plan for the next 5 years? If ‘no’ enter N/A.

N/A

Section 4: INCITE Allocation Request & Other Project Funding/Computing Resources

Question #1

OLCF Summit (IBM / AC922) Resource Request - 2023

Node Hours

692500 GPU

Storage (TB)

100

Off-Line Storage (TB)

0

Question #2

OLCF Frontier (Cray Shasta) Resource Request – 2023

Question #3

OLCF Frontier (Cray Shasta) Resource Request – 2024

Question #4

OLCF Frontier (Cray Shasta) Resource Request – 2025

Question #5

ALCF Theta (Cray XC40) Resource Request - 2023

Node Hours

1000000 GPU

Storage (TB)

200

Off-Line Storage (TB)

0

Question #6

ALCF Polaris Resource Request - 2023

Question #7

ALCF Polaris Resource Request - 2024

Question #8

ALCF Polaris Resource Request - 2025

Question #9

ALCF Aurora (Intel Xe) Resource Request – 2023

Question #10

ALCF Aurora (Intel Xe) Resource Request – 2024

Question #11

ALCF Aurora (Intel Xe) Resource Request – 2025

Question #12

List any funding this project receives from other funding agencies.

Funding Sources

Funding Source

DOE AMO

Grant Number

DE-EE0009096

Question #13

List any other high-performance computing allocations being received in support of this project.

Other High Performance Computing Resource Allocations

Resource

Summit

Allocation Agency

NERSC/ERCAP

Allocation

10,000 CPU, 500 GPU

Allocation Year

2022

Section 5: Project Narrative and Supplemental Materials

Question #1

Using the templates provided here, please follow the [INCITE Proposal Preparation Instructions](#) to prepare your proposal. Elements needed include (1) Project Executive Summary, (2) Project Narrative, (3) Personnel Justification and Management Plan, (4) Milestone Table, (5) Publications Resulting from prior INCITE Awards (if appropriate), and (6) Biographical Sketches for the PI and all co-PI's. Concatenate all materials into a single PDF file. Prior to submission, it is strongly recommended that proposers review their proposals to ensure they comply with the proposal preparation instructions.

Concatenate all materials below into a single PDF file.

- 1. Project Executive Summary (One Page Max)**
- 2. Project Narrative (15 Pages Max)**
- 3. Personnel Justification and Management Plan (1 Page Max)**
- 4. Milestone Table**
- 5. Publications resulting from prior INCITE Awards (if appropriate)**
- 6. Biographical Sketches for the PI and all co-PI's.**

2023_INCITE_Final_submission_of_documents.pdf
The attachment is on the following page.

PROJECT EXECUTIVE SUMMARY

Title (80 characters max; strictly enforced): Machine learning classical potentials for design of organic electrolytes

PI and Co-PI(s): Rafael Gomez-Bombarelli (PI, MIT DMSE), Yang Shao-Horn (MIT DMSE) Ben Blaiszik (Materials Data Facility)

Applying Institution/Organization: Massachusetts Institute of Technology

Resource Name(s) and Number of Node Hours Requested: Summit-GPU: 0.6925M, Theta-GPU: 1M

Amount of Storage Requested: Summit: 100 TB, Theta: 200TB

Executive Summary/Abstract: The development of safe lithium batteries with high charge density to store renewable energy from intermittent sources requires solid electrolytes. Fast ion transport decoupled from matrix motions is needed to prevent concentration gradients that lead to dendrite formation. Molecular dynamics (MD) simulations are an important tool to screen potential electrolyte systems before pursuing costly experimental characterizations. However, exploration of novel, exotic electrolytes through MD is hindered by the sluggish development of interatomic potentials that are accurate, transferrable and scalable in length and time. To this end, we propose the development of two coupled autonomous and end-to-end computational frameworks to predict electrolyte conductivity and transference number for 1) a message-passing neural network-based interatomic potential for crystalline electrolyte materials to be routinely deployed to simulate hundreds of novel materials in the 1000-atom and nanosecond scale at density functional theory (DFT) accuracy; and 2) chemically and configurationally transferable classical potential parameters fit to DFT-computed forces and charges for amorphous polymer electrolyte systems to accelerate screening and elucidate underlying design principles for hundreds of novel polymers at the 10,000-atom, 100-nanosecond scale. The conductivities simulated herein will be integrated into a collaboration between MIT and PNNL through AMO award DE-EE0009096 to determine, parameterize, simulate, and characterize proposed electrolytes. All data and models developed in this allocation will be open-sourced through the Materials Data Facility.

MD simulations using neural network-based interatomic potentials are orders of magnitude faster than *ab initio* MD, but they require representative training data in order to remain stable for timescales on the order of ns, needed to determine the contribution of lithium-ion hopping to diffusion. The combinatorial space of possible aliovalent and isovalent substitutions requires significant computational resources to effectively explore the many possible inorganic argyrodite and perovskite crystalline systems. Through the use of an in-house automated screening platform for assessing the bulk conductivity of candidate inorganic crystalline electrolyte materials, using neural network potentials trained on data from first-principles calculations and active learning based on adversarial attacks on uncertainty at quantum chemical accuracies. Such a platform for obtaining ionic conductivity is material agnostic and could be applied to any inorganic or organic crystalline ion conductor (with Na^+ , Mg^{2+} , O^{2-} , etc). This could be used by the broader materials community since sensing and solid oxide fuel cells rely on similar phenomena.

Determination of conductivities for amorphous polymer systems require even longer MD simulations (~100s of ns for tens of thousands of atoms). Thus, classical interatomic potentials are required for reasonable simulation speeds. Traditional parameterization of these classical terms requires painstaking hand-tuning schemes for only a subset of possible atomic chemical environments. To this end, we have developed an autonomous framework to parameterize class 2 classical force fields learned from DFT to determine chemical environment-specific parameters that are learned automatically from data and transferable in chemical and configurational space. After showing learned potentials of prototypical solid polymer and liquid carbonate electrolytes are capable of predicting conductivities of similar magnitude to prior simulation literature and experimental measurements, this working group is poised to autonomously study the conductivity mechanisms across vast backbone and functionalization polymer chemistry space. Through ML-based end-to-end fitting of classical potentials, we will explore the accuracy-speed trade-off in data-driven potentials, and deploy the most effective simulations to understand and screen non-crystalline materials.

PROJECT NARRATIVE

1 SIGNIFICANCE OF RESEARCH

1.1 Solid-state electrolytes for next-generation energy storage.

The rapid development of high-charge density and high-power electrical energy storage is imperative to manage upcoming power demands of electric vehicles and intermittent renewable energy in a sustainable, cost effective manner. Lithium metal batteries have been proposed for high energy density applications to decarbonize the light- to heavy-duty vehicle sector, as well as for other applications with stringent energy density requirements¹. While these promising electrodes have been found to cycle reversibly in laboratory conditions, efforts to scale the technologies to commercial size have failed due to the formation of lithium metal dendrites that can cause performance and safety issues^{2,3}. High charge gradients from unequal ion transport in the electrolyte have been proposed as the cause of the dendrite formation, and thus there has been a strong push in the search for electrolyte systems that can optimally shuttle the cation between electrodes while slowing the anion or immobilizing it completely. Additionally, the conventional carbonate-based liquid electrolytes used in lithium-ion batteries degrade both chemically and electrochemically when in the presence of solid lithium electrodes. Thus, alternative electrolyte systems are being investigated to increase diffusion kinetics and chemical stability to ultimately develop long-lasting batteries for electric vehicles and intermittent renewable energy storage.

Crystalline (often ceramic) and amorphous (polymer) solid-state electrolytes have been proposed as non-flammable and electrochemically-stable alternatives to conventional liquid carbonate electrolytes. Solid-state electrolytes allow for leak-proof and thermally stable separators, significantly increasing the safety of resulting batteries when used in commercial transportation applications^{3,4}. Since the discovery of room temperature superionic conductors of Li⁺ such as Li_{1.3}Al_{0.3}Ti_{1.7}(PO₄)₃ (LATP) in 1990⁵, Li₇La₃Zr₂O₁₂ (LLZO) in 2007⁶, and Li₁₀Ge(PS₆)₂ (LGPS) in 2011⁷ with liquid-like diffusion of the lithium ion sublattice, there has been renewed interest in identifying novel crystalline electrolytes for batteries with similar properties but more Earth-abundant and distributed components. Additionally, crystalline electrolytes such as these and recently investigated plastic crystal-based electrolytes⁸ are of particular interest because they are also single-ion conductors – as in only the cation is mobile since the anions occupy fixed lattice points – removing the mechanism underlying the formation of concentration gradients that lead to cell performance loss. On the other hand, solid polymer electrolytes are of interest due to their ease of manufacturing and lower weight. However, the conventional solid polymer electrolyte system, polyethylene oxide (PEO), requires lithium to be introduced with a corresponding anion – often bis(trifluoromethane)sulfonimide (TFSI) – leading to electrode polarization and dendrite formation due to the anion being more mobile than the lithium cation due to polymer-lithium interactions².

While there has been significant progress in determining potential solid-state electrolyte systems in laboratory conditions, the determination of room temperature electrolytes that also exhibit electrochemical stability and favorable transport properties relies on the understanding of the ion diffusion mechanisms in each system. Following an unpublished work of collaborator Yang Shao-Horn (MIT), the primary mechanisms for ion conduction range from ion hopping to vehicular ion motion based on composition-dependent shear modulus and a characteristic volume for the diffusing species. Molecular dynamics simulations are of particular interest to study this problem as atomic-scale coordination environments and diffusion can be observed for physical time scales. Due to the end-to-end nature of our proposed computational workflows, we will be able to assess the relative solvent/framework and ion diffusion timescales and resulting mechanisms across immobile anion frameworks, through viscous organic crystal melts and entangled polymer solids, and potentially vehicular-motion driven liquid electrolytes with autonomously-derived, quantum-chemically accurate interatomic potentials for each system. Gaining access to the vast computational resources at LCF resources would allow the exploration across this bridging between mechanisms by expanding the capacity of this working team to study novel and potentially intermediate systems.

Historically, ab initio molecular dynamics (AIMD) has been used to evaluate the diffusivity and ionic conductivity of crystalline electrolytes, but this quantum-chemical calculation-based simulation requires months of compute time and is limited to nonphysical high temperatures in order to attain convergence in diffusivity statistics. However, the high-temperature regimes in which diffusivity is evaluated by AIMD often do not overlap with the near-room-temperature regime in which the diffusivity is measured experimentally, requiring extrapolation. Using neural network potentials, which predict the forces on each atom according to the local environment at each timestep, we have been able to access timescales on the order of 10s of ns with molecular dynamics using neural network potentials and achieved chemical accuracy (comparable to ab initio methods), allowing us to evaluate diffusivities for each composition of a crystalline material in a faster time than AIMD. This has allowed us to bridge the gap between the diffusivities at high temperature with AIMD and the diffusivities experimentally measured near room temperature (as seen in the figure below).

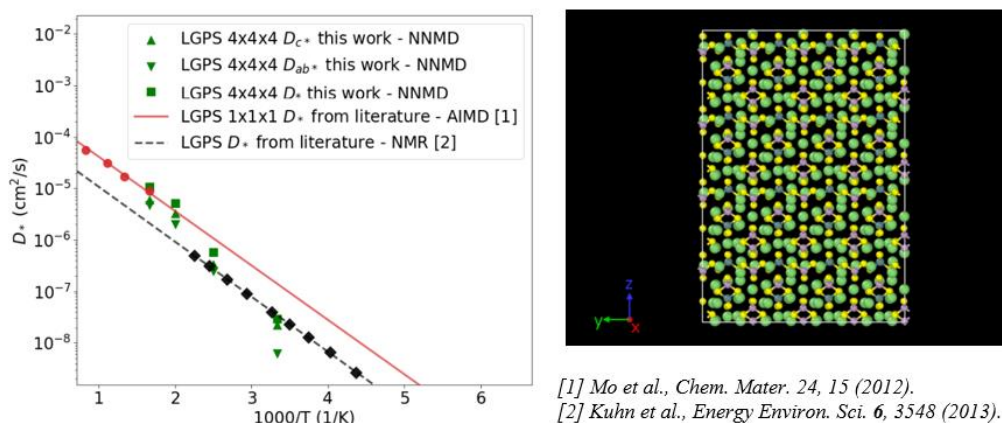


Figure 1. In-house neural network potential molecular dynamics simulation of crystalline LGPS and resulting diffusivities (green) across distinct temperature regimes inaccessible by ab initio molecular dynamics (red).

Meanwhile, polymer electrolyte studies have struggled to move beyond the seminal PEO-LiTFSI system because of its superior lithium salt solvation and low glass transition temperature^{9,10}. However, low lithium transference and resulting concentration and charge gradients plagued experimental researchers until all-atomistic molecular dynamics (MD) simulations revealed low or negative lithium transference was due to strong Lewis acidity interactions between PEO and the lithium ion once the salt was separated^{11,12}. Subsequent computational studies on promising polymer electrolyte systems have been hindered by unreliable and manually re-parameterized potentials that make comparing results across different studies difficult^{13,14}. Additionally, studies are typically limited to a few variations of known polymer backbones, such as glymes and carbonates^{15,16} due to the computational cost to run the simulations as well as the manual cost to readjust the interatomic potential partial charges.

Given a simulation scheme and initial positions of the atoms, the choice of interatomic potential used to compute the atom positions after each timestep dominates the behavior of a molecular dynamics (MD) simulation. These interatomic potentials have traditionally been determined in two main ways. The first is at each time step through computationally-expensive quantum chemical calculations in ab-initio molecular dynamics (AIMD) simulations. The second is in classical molecular dynamics simulations, which requires hand-tailored parameterization of analytical models for different types of interactions between atoms in classical molecular dynamics simulations based on costly quantum-chemical simulations such as density functional theory (DFT). In this project, we propose the use of in-house machine learning frameworks to accelerate the determination of interatomic potentials for the consistent and comprehensive study of novel ceramic and polymer electrolyte systems.

These frameworks could be used to screen potential electrolyte systems to later be passed to experimental collaborators on this project to validate and improve the machine learning predictions. More broadly, as these frameworks are developed independent of the choice of prescribed interatomic interactions, the effects of inclusion of difficult-to-parameterize interactions such as polarizability can be systematically investigated without the need to rely on previously-tailored legacy parameterizations or datasets.

We focus our efforts on exhausting the design space by determining interatomic potentials for previously unparameterized chemistries and thereby pushing performance targets of both crystalline and amorphous solid electrolytes forward in this project. The lead was recently awarded 10,000 CPU node hours and 500 GPU node hours through ERCAP Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0021172. The lead PI and co-PI have also received a 2021-2022 ALCC award on “Inverse design of multicomponent oxide catalysts with generative models and DFT,” however the current proposed project is unrelated.

1.2 Potential Impact.

The discovery of electrochemically stable electrolytes with high conductivity at room temperature could revolutionize the energy sector by significantly reducing the demand for fossil fuels in everyday transportation and by increasing the robustness of renewable energy-based power grids to intermittent generation and demand. However, optimal materials systems cannot be determined through conventional experimentation due to the combinatorial nature of potential electrolytes chemical space due to alloying and defect tuning in crystalline materials and backbone and functional group combinations for polymers. Also, simulations are burdened by the lack of reliable and chemically-relevant interatomic potentials needed to represent realistic diffusion mechanisms in these ionic systems.

Our proposed high-throughput, autonomous workflows aim to both determine better electrolytes as well as elucidate the interplay of electrolyte-ion interactions and resulting diffusion mechanisms. While many potential diffusion mechanisms for ion in both crystalline and amorphous systems have been proposed, all-atom simulations that seek to describe the mechanisms have relied on non-physical tricks such as artificially-high temperatures up to 1500K¹⁷ or reduced ion point charges to ± 0.7 ^{15,18} in order to attain the time scales of ion diffusion. This begs questioning the validity of resulting electrolyte-ion interactions.

Furthermore, due to significant monetary and computational costs of existing experimentation and simulations, the search for better electrolytes has focused on chemically-similar derivatives of known electrolyte systems, limiting the opportunity to uncover novel systems. In addition to predicting diffusivities, conductivities, and lithium transference of unknown candidate electrolyte systems, scientific insights into what physics dictate these mechanistic changes through the inclusion or removal of interaction models in our learned potentials.

Prior studies with classical interatomic potentials have found that class 1 interatomic potentials such as OPLS predict conductivities an order of magnitude below experiments due to significant coulombic overbinding from assuming point charges^{16,19}. Class 2 interatomic potentials such as COMPASS and PCFF+ predict conductivities half an order of magnitude below experimental results^{11,15}, while the use of polarizable force fields has been found to return much better agreement with experiments but have not yet been widely implemented due to a lack of reliable parameterization for the necessary chemistries²⁰. Thus, the implementation of polarizability into our end-to-end workflows can not only improve conductivity

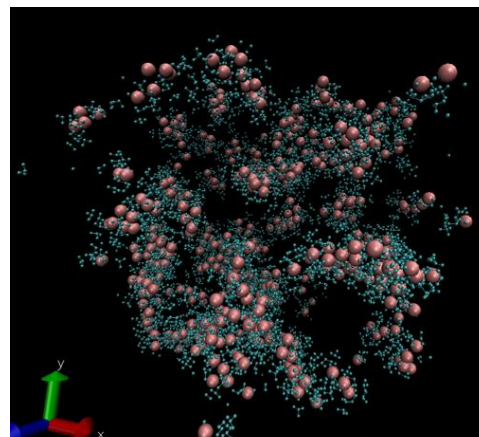


Figure 2. Over-bound ion cluster from in-group performed molecular dynamics simulation of amorphous PEO-LiTFSI using machine-learned OPLS interatomic potential. Pink=Li⁺, blue=TFSI⁻

predictions for novel electrolyte systems, but also provide physical parameters and insight to the broader community.

1.3 Innovativeness.

The chemical space of potential electrolytes is too large for traditional trial-and-error discovery. Synthesis and characterization of individual trials of alloying and featurization space is too slow, and simulations with quantum-chemically accurate interatomic potentials are either prohibitively expensive or do not have parameters describing the full breadth of local atomic environments.

In addition, experimental and computational efforts are currently only loosely coupled as computed conductivities are low-throughput and often orders of magnitude different than experimental measurements. In this proposal we will seek to further develop our lab and other's prior experience in high-throughput simulation by innovating on end-to-end integration.

Our approach will be to create a set of two fully autonomous machine-learning based workflows that will learn quantum chemically-accurate interatomic potentials to drive high throughput atomistic simulations using self-generated training data.

The existing, funded collaboration with world-class experimentalists including Yang Shao-Horn at MIT and Jun Liu and Jie Xiao at PNNL enables the computational conclusions of this work to reach out to the lab and to the broader scientific community through Co-PI Ben Blaiszik's work with the Materials Data Facility. Additionally, interfacing our predicted chemistries with computational collaborators Elsa Olivetti (MIT DMSE) and Tonio Buonassisi (MIT MechE) will further accelerate the discovery-to-implementation process through their work on predicting optimal synthesis and electrolytic cell parameters.

These experimental and computational efforts combine to form a fully end-to-end discovery pipeline from initial candidate chemistry screening through data mining and parallel autonomous high-throughput experimentation and simulation to full scale development and testing of electrochemical cells. This integrated approach balances physics-based models with big data machine learning techniques and autonomous experimentation to perform exhaustive optimization over crystalline and amorphous electrolyte systems. The proposed workflow is transferrable to other systems such as electrolytes for fuel cells or even for other materials where interatomic potentials are not well characterized such as redox-active polymers.

2 RESEARCH OBJECTIVES AND MILESTONES

We propose the use of two coupled, autonomous, and end-to-end machine learning frameworks to predict electrolyte conductivity and transference number using local atomic environment-aware interatomic potentials determined through 1) a message-passing neural network-based model for crystalline electrolyte materials to be routinely deployed to simulate hundreds of novel materials in the 100-atom and nanosecond scale at DFT accuracy 2) chemically and configurationally transferable classical potential parameters fit to DFT-computed forces and charges to be used for novel amorphous polymer systems at the 10,000-atom, 100-nanosecond scale.

These workflows will be used to screen crystalline chemically- and geometrically-disordered boron-containing argyrodites and oxides such as $\text{Li}_6\text{B}_7\text{S}_{13}\text{X}$, $\text{Li}_{6+x}\text{B}_{10}\text{S}_{18}\text{X}_x$ and $\text{Li}_{6+2x}\text{B}_{10}\text{S}_{18}\text{X}_x$ ($\text{X}=\text{Cl}, \text{Br}, \text{I}$) and organic plastic crystal electrolytes, as well as amorphous polymer systems such as polycarbonates and boron-containing Lewis basic pendant groups on a variety of polymer backbones.

2.1 Neural network (NN) potentials for ion hopping in organic crystalline materials

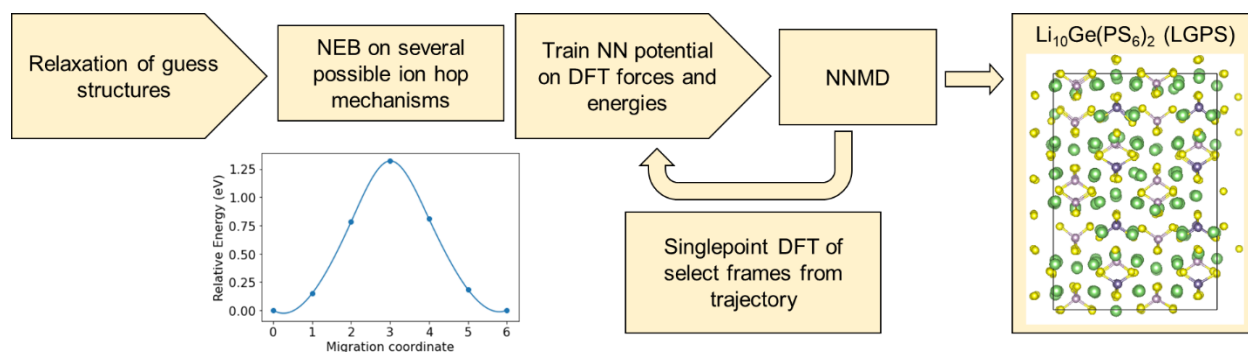


Fig. 3. Schematic of neural network potential workflow for crystalline electrolyte systems.

A workflow for efficiently assessing ionic conductivity in molecular crystalline systems (salts of solvate ionic liquids) was constructed. It consists of a training data generation step, where nudged elastic band (NEB) calculations using VASP are performed to sample off-equilibrium geometries near the transition state (ion hops), in conjunction with single-point VASP calculations of forces and energies for equilibrium geometries with random perturbations. The first-generation neural network potential is then trained on the forces and energies from these geometries and neural network molecular dynamics (NNMD) is performed. A message-passing neural network - the polarizable atom interaction neural network (PaiNN) - served as the underlying architecture for the potential. Charge is not explicitly included in the neural network architecture. With the initial NNMD trajectory, unique representative frames are strategically chosen to be evaluated by singlepoint DFT, sampling uniformly over phase space and filling in the missing regions of phase space when coarsely mapping the potential energy surface using NEB and random perturbations.

The underlying potential used to drive NNMD simulations is a message-passing neural network (MPNN), which initializes atomwise features and builds a graph within a local neighborhood of interacting atoms that repeatedly exchanges messages through updates of node features. The polarizable atom interaction neural network (PaiNN) is one example of a message-passing NN that is rotationally equivariant²¹ and is the NN architecture that will be used for the NN potential. While rotational invariance would require $O(N^2)$ scaling to store information about bond distances and angles that uniquely describes molecule conformations, rotationally equivariant vector representations store the same information that distinguishes unique molecule conformations and maintains $O(N)$ scaling. As a message-passing NN, if the cutoff is 5 Å, and there are 4 convolutional layers, due to the updates with message-passing the actual local environment is 20 Å. Intuitively, there is a tradeoff between accuracy and computational speed in considering the cutoff size and how large a local environment one is considering an atom/ion interacting with.

2.2 Autonomous parameterization of classical interatomic potentials for polymer electrolytes

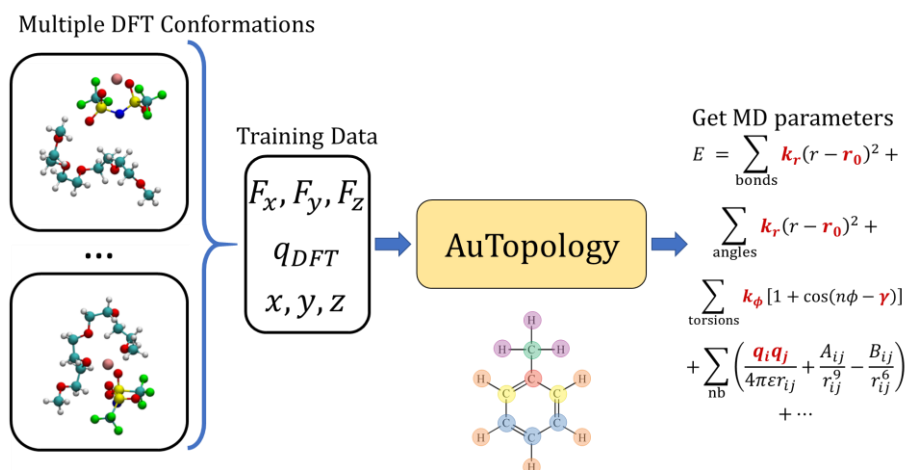


Fig. 4. Schematic of the AuTopology classical potential workflow for amorphous polymer systems, where the highlighted parameters are local chemical environment-specific to be learned.

Due to the 10-100s of ns time scale needed to observe polymer rearrangement and ion hopping in amorphous polymer electrolytes, AIMD and even NNMD are too computationally expensive for statistical sampling of configurational space. Thus, classical MD potentials such as the class 2 interatomic potentials which update the position of the atoms by calculating the net force on each atom through analytical relationships such as harmonic approximations of covalent bonds will be used to probe these systems instead. As an alternative to proprietary, and therefore non-augmentable, and limited-chemistry legacy classical parametrizations of interatomic potentials such as PCFF²² and COMPASS²³, we have developed an in-house machine learning framework called AuTopology which autonomously learns interatomic potential parameters based on local atomic environment using DFT forces and charges as training data. AuTopology allows for end-to-end parameterization of novel chemistries with only small amounts of DFT calculations and of different interatomic potential models through the inclusion of higher order potential terms such as polarizability.

AuTopology first featurizes each training conformer as a graph where each node represents an atom and each edge represents a bond. Atoms are each given a discrete type (represented as the colored circles in Fig. 4 by a one-hot encoding by their element and then further specialized by their 3-nearest graph neighbors to encode the local atomic environments. Bonds, angles, dihedrals, and impropers are subsequently labeled with a one-hot encoding according to the final atom labels contained within each topology. The interatomic potential parameters determined by these distinct topology types are then optimized through Adam optimizer stochastic gradient descent according to x, y, and z components of DFT forces, DFT partial charges, and integer ion charge.

In addition to discrete atom typing based on covalent bond graph connectivity, working group member Jurgis Ruza is in the process of abstracting the typing into a continuous, latent representation through the use of message passing neural networks such as SchNet²⁴. While the final parameters will be held constant after training similar to the discrete atom typing, passing the local atomic environments to a continuous latent space would allow AuTopology to learn the chemical similarities between training data environments and extrapolate to unseen chemistries, further improving the capabilities of the workflow towards the discovery of novel polymer electrolyte systems.

2.3 Autonomous, uncertainty-based training data for interatomic potential active learning

A large number of expensive DFT simulations are needed to appropriately sample the configurational space needed for training stable NN potentials or representative classical interatomic potential parameters. The exploration of such large compositional and configurational space is intractable with electronic

structure simulations only. We will thus use a two-stage approach to generate representative ensembles of training data for our workflows. The first stage is an initial set of configurations using either nudged elastic band or quantum chemical structure relaxation followed by an active learning loop where trained potentials are used to drive short molecular dynamics simulations to generate more training data from increasingly realistic potentials.

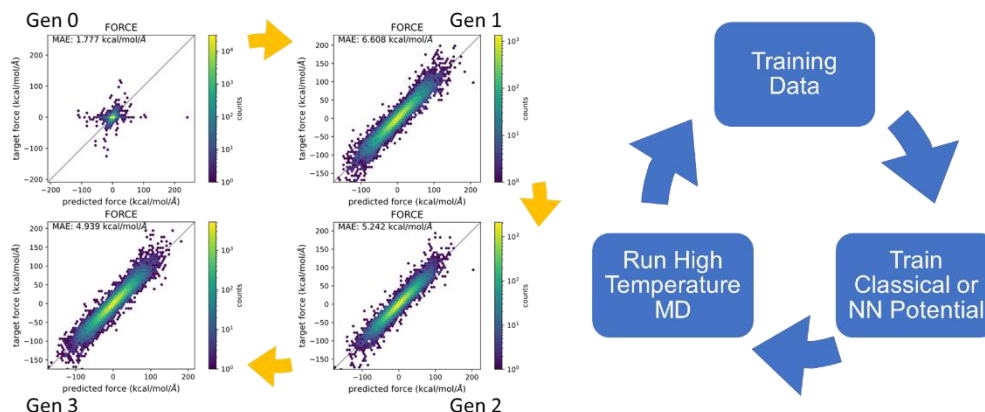


Fig. 5. Autonomous active learning loop which shows improved parity of learned classical potential forces relative to DFT training data for a broader range of conformers.

For crystalline systems, nudged elastic band (NEB)²⁵ will be used to create an initial set of training data that spans the trajectory of an ion hop through different pathways. NEB can be used to evaluate the enthalpic energy barrier of transition states, such as ion hopping in a crystal. In this project, NEB will be used to quickly gather DFT training data in the form of forces and energies for geometries corresponding to ion hopping events, which would otherwise require picoseconds of simulation time to gather using AIMD. The initial and final states for NEB are representative states before and after an ion hop, and the structures corresponding to these local equilibrium Li⁺ positions must already have been relaxed. For a concerted hop mechanism, the positions of two ions would change between the initial and final state. When there are a multitude of possible sites in the crystal framework's void space for the mobile species, tools that employ Voronoi decomposition²⁶ will be used to re-populate the interstitial sites with Li.

For amorphous polymer systems, quantum chemical geometry relaxation of representative ionic clusters will be used to generate an initial set of training data. An in-house clustering algorithm that places a coordination center [ex Li⁺] and coordinating atom [ex. sulfonate oxygen of TFSI- or ether oxygen in PEO] in a certain interaction distance with a Monte Carlo-based collision algorithm will be used to create physically meaningful solvent-ion clusters such as contact ion pairs and solvent-separated ion pairs^{11,16}.

These initial geometries will be used to train a “zeroth-generation” interatomic potential that will be used to run short, 100 picosecond simulations at high temperatures (400K) to explore accelerated ion hopping or internal molecular rotations. Frames will be selected for DFT single-point calculations based on an uncertainty quantification approach developed in-house by Schwalbe-Koda et al²⁷, and added to the training data to retrain a subsequent potential in an iterative manner as seen in Fig 6.

To determine the frames of maximum relevance for adding to training data, an ensemble of NN or classical potentials will be trained and used to identify trajectories that maximize uncertainty of the ensemble (adversarial loss) in physically-accessible phase space. First, non-physical frames from NNMD are filtered from the pool to sample from. These include: (1) Frames with atoms that collided (collision is determined based on cutoff radius from the pseudopotential or Lennard-Jones parameter for a given atom). (2) Frames with a molten framework or molecules separated by periodic boundary conditions. For crystalline systems, the Li⁺ sublattice is expected to be mobile, however the criteria for a molten framework is based on the root-mean-squared-displacement for framework atoms (in reference to the initial structure) with respect to a threshold value. Low-dimensional embedding of the similarity matrix between all filtered frames is performed to then do k-means (spectral) clustering, so as to ensure diversity in sampling. Then,

through Bayesian approaches, a decision strategy is put in place to avoid recalculating new data points if they resemble the training data. As a consequence, ground truth DFT data is only added for configurations that expand the breadth of the existing training data.

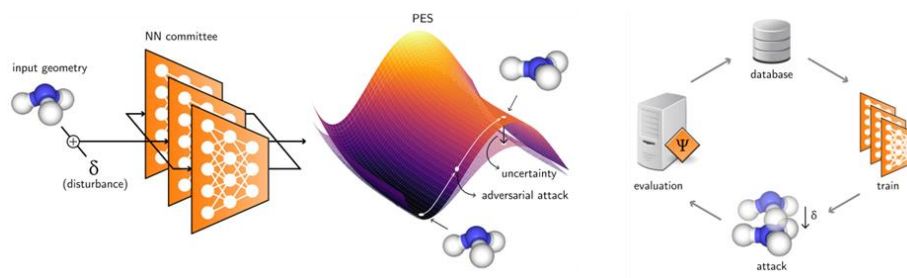


Fig 6. Uncertainty-based active learning procedure²⁷.

2.4 Determination of correlation-corrected ionic conductivity

Ionic conductivity is commonly described through the Nernst-Einstein formalization, where every charged species is assumed to be non-interacting with other species and the net conductivity is simply a function of the tracer diffusivity and charge carrier concentration of each ion. However, recent studies have found that accounting for both intra- and inter- species displacement correlations is imperative to compute conductivities similar to experimental results from molecular dynamics simulations^{28,29} in both single-ion and counter ion-containing electrolyte systems as seen in . Because of this, we will compute the Wheeler-Newman²⁹ based conductivity where Equation (1) describes the positional correlation function, which can describe interspecies correlation (ex. Li⁺ and TFSI⁻ correlation) or even intraspecies correlation (ex. one Li⁺ atom with the other Li⁺ atoms for concerted motion in single-ion conductors). The mass-transport coefficient, L_{ab} , in Equation (2) takes the place of the diffusivity from the Nernst-Einstein formalization and is similarly approximated from the time-averaged derivative of a second-order displacement (self-mean squared displacement or position correlation). Equation (3) describing the relationship between charge carrier concentration and transport parameter with system conductivity, where F is Faraday's constant and q_j and c_j are the charge number and concentration of species j , respectively.

$$PCF_{ab}(\tau) = \frac{V}{6k_B T} \langle [\vec{R}_a(\tau) - \vec{R}_a(0)] \cdot [\vec{R}_b(\tau) - \vec{R}_b(0)] \rangle \quad (1)$$

$$L_{ab} = \lim_{\tau \rightarrow \infty} \frac{\partial PCF_{ab}(\tau)}{\partial \tau} \quad (2)$$

$$\kappa_c = \sum_{ab} (F q_a c_a) (F q_b c_b) L_{ab} \quad (3)$$

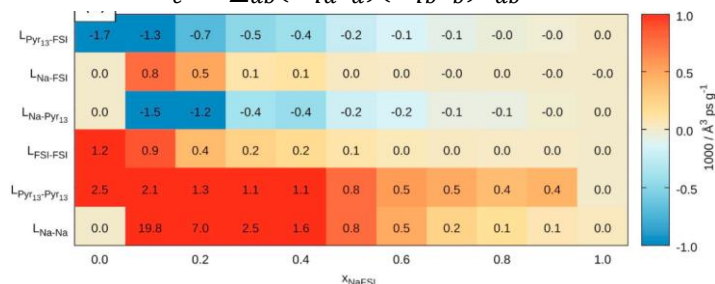


Fig. 7. Correlation-based mass-transport coefficient for a sodium ion electrolyte showing the importance of self- and cross-correlations across species at different ion concentrations²⁹.

This correlation-based conductivity analysis has been used in an unpublished report from this working group to show that potentials trained through AuTopology are able to recover conductivities closer to

experimental literature for a range of carbonate-based liquid electrolytes than traditional legacy parameterizations of classical potentials¹⁹.

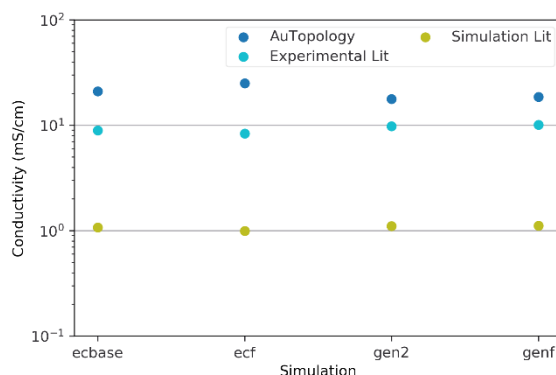


Fig 8. Conductivities of simulations using AuTopology force fields are in the same order of magnitude as experimental measurements while legacy parameterizations of classical force fields are an order of magnitude lower.

2.5 Data integration and data management plan

2.5.1 Database integration

Lead PI Gomez-Bombarelli houses a high throughput database that contains over 100 million in-house DFT simulations for molecules and solids. This software suite includes a PostgreSQL database schema, Django as an object relational manager and web framework, a Python-based codebase to automatically build simulation jobs from database objects, meta-scheduler programs that interface with supercomputer schedulers such as SLURM, automated batching of small jobs, and automated parsing to inject simulation outputs into the structured database. The toolset adds unique HT capabilities and integrates open-source libraries such as RDkit, ASE and Pymatgen for data processing, making it possible to generate a vast and comprehensive dataset of reactive silicate configurations, both in the gas phase and in vacuum for the first time to be used as training data for NNIP.

Both the NNMD and classical interatomic potential frameworks have successfully connected end-to-end with this database, querying the database for training data, running molecular dynamics simulations, and submitting new frames for DFT single-point calculations to be stored into the database for access for further training. **Data Access:** Access to all data collected will be made available to researchers through a programmatic interface (e.g, a REST API). We will utilize Python tools and scripts that enable rapid aggregation of data and metadata from the collected data stores and the search index. We expect in this project the primary way users will access data will be through scripts or Jupyter notebooks. **Data Publication:** To promote broader use of data from this project, many of the data artifacts will be published as appropriate. Data will be assigned a permanent unique identifier (e.g., DOI), will be described with the metadata collected previously, moved to persistent storage, and shared for free public access.

Separately, we will collect the models and code built in this project. Collaborative work on the code associated with models will be performed on Github. Each repository will contain the code needed to train the model, instructions on how to train the model, instructions on building the required computational environment (e.g., through a requirements.py file), and the trained model weights. After models are trained, they will be published to the DLHub service (with appropriate access control) to allow team members to invoke models on new data with only a few lines of code, and to receive³⁰ a DOI for permanent identification of the model and to enable citation and reference to the models in team publications.

This project will produce abundant data. The most copious data to be created are outputs of DFT calculations, including text output files and binary density and wavefunction files. We will use existing parsing programs, such as cclib and Atomic Simulation Environment to process these outputs and will

extend their capability in-house as needed. After processing, digests of the text-based human-readable outputs will be stored in a Postgresql database that uses django as an object-relational mapping layer. We will build on an existing schema that connects calculations with chemical graphs, crystalline structures, supramolecular clusters, job configuration files, and contains fully indexed and related record of job creation, completion and results. Our platform also includes an automated job meta-scheduler that interfaces supercomputer schedulers like SLURM and Torque to create and execute new submission files.

3 COMPUTATIONAL USAGE PLANS

*The **Milestone Table** is appended to the end of this document (after the References section).*

3.1 Use of Resources Requested and Parallel Performance

In order to determine the allocation request, speed tests for crystalline materials were conducted on the Cori (Intel Xeon Phi 7250) supercomputer and tests for neural network molecular dynamics (NNMD) have been performed on the Summit-GPU computing platform through the use of previous ALCC project allocations. Speed tests for classical molecular dynamics were conducted on Theta-GPU through the use of a director's discretionary funding.

GPU-driven neural network molecular dynamics (NNMD) have been enabled on the Summit-GPU computing platform by working group member Gavin Winter through the use of the Atomic Simulation Environment (ASE) and PyTorch Python packages. Access to large-scale GPU resources is required for running NNMD simulations on the order of nanoseconds with large systems containing many mobile ions and disorder. The underlying neural network potential is a message-passing neural network with a cutoff of 5 Å that evaluates forces and energies for a given set of atoms and atomic coordinates. Because the system size effects on periodic DFT simulations of crystalline electrolytes is only recently being explored, weak scaling relations were determined for the NNMD simulations with the results seen below.

Neural Network Molecular Dynamics (NNMD) with $\text{Li}_{10}\text{Ge}(\text{PS}_6)_2$

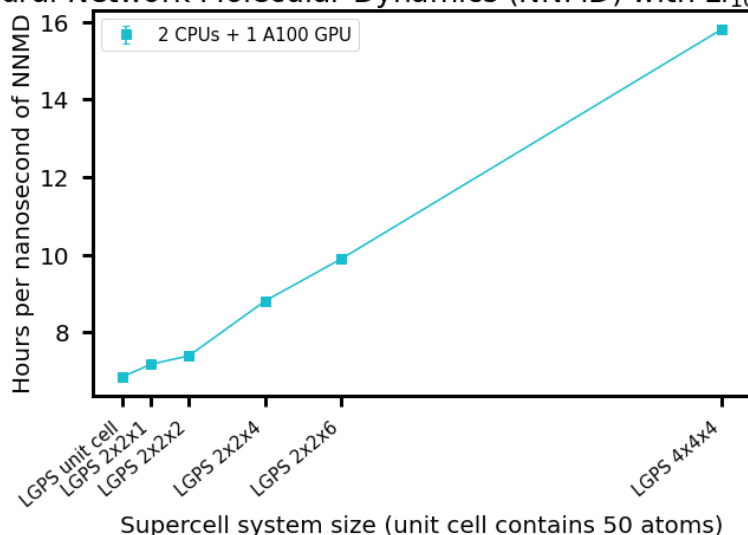


Figure 9. Weak scaling relation tests for crystalline neural network molecular dynamics simulations on Summit-GPU.

The ASE code for molecular dynamics has been adapted to (1) batch copies of trajectories to be run in parallel for improved diffusivity statistics or (2) parallelize a single trajectory of a large system over multiple GPUs. We observed on the Summit computing platform that the wall clock time is relatively insensitive to the number of CPU resources allocated for a job since most of the time-intensive operations for running MD are performed on GPU. More than 20,000 atoms can be fit on a single A100 GPU, but

scaling with system size is roughly linear. At least a 4x4x4 supercell is needed to avoid nearest-image effects, so 16 node-hour per simulated nanosecond are needed. In order to simulate 10 ns for 650 stoichiometries and 5 repetitions of 4x4x4 supercells, 32,500 ns of neural-network based MD will be conducted, requiring **0.52 million Summit-GPU node hours**. Due to the inability to split ASE Atoms objects between multiple GPUs, it is not currently possible to use multiple GPUs for running MD in a shorter wall time with more resources. While our current neural network architecture does not accommodate GPU parallelization (only uses 1 GPU for running MD), we are adapting to use a more scalable neural network architecture such as Allegro (NequIP)^{31,32}.

To generate the crystalline system training data with forces and energies from DFT, NNMD can be used to generate an initial trajectory from which unique representative frames are chosen to be evaluated by DFT, based on the model uncertainty obtained from the variance of an ensemble of related models. These single-point DFT calculations will be performed using this allocation. Vienna Ab Initio Simulation Package (VASP) will be used for electronic structure calculations based on density functional theory (DFT) using a plane-wave basis, and it is already compiled on Cori KNL and Perlmutter (GPU-enabled). Single-point density functional theory (DFT) calculations were performed to collect forces and energy for a randomly perturbed structure of the $\text{Li}_{20}\text{Ge}_2\text{P}_4\text{S}_{24}$ unit cell with 50 atoms (1x1x1), as well as on 2x1x1 and 2x2x1 supercells. This is representative of the main workload required for sampling structures from both NNMD trajectories as well as from random perturbations. The time to complete this single-point calculation is generally reflective of the time for a single timestep in Born-Oppenheimer Molecular Dynamics. We employed the generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) functional. Interactions between ion cores and valence electrons were treated using the projector augmented wave (PAW) method with an energy cutoff of 520eV, and a 2x2x1 k-point grid was used for all cells for direct comparison across system sizes.

For each system size, an allocation with 1, 2, and 4 nodes for each job were tested with both 4 cores per task and 16 cores per task. These calculations were performed on Cori KNL nodes (Intel Xeon Phi 7250), using VASP 5.4.4. The one exception is on the plot of wall clock times N1_n4_c32_G4 was included, which was run on one Perlmutter node (AMD EPYC 7763) with 4 GPUs (NVIDIA A100), using VASP 6.2.1 with GPU. Since it is not possible to directly compare between the two systems, the singlepoint DFT calculations on Perlmutter were not included when comparing the node-hours required for the other singlepoint DFT calculations.

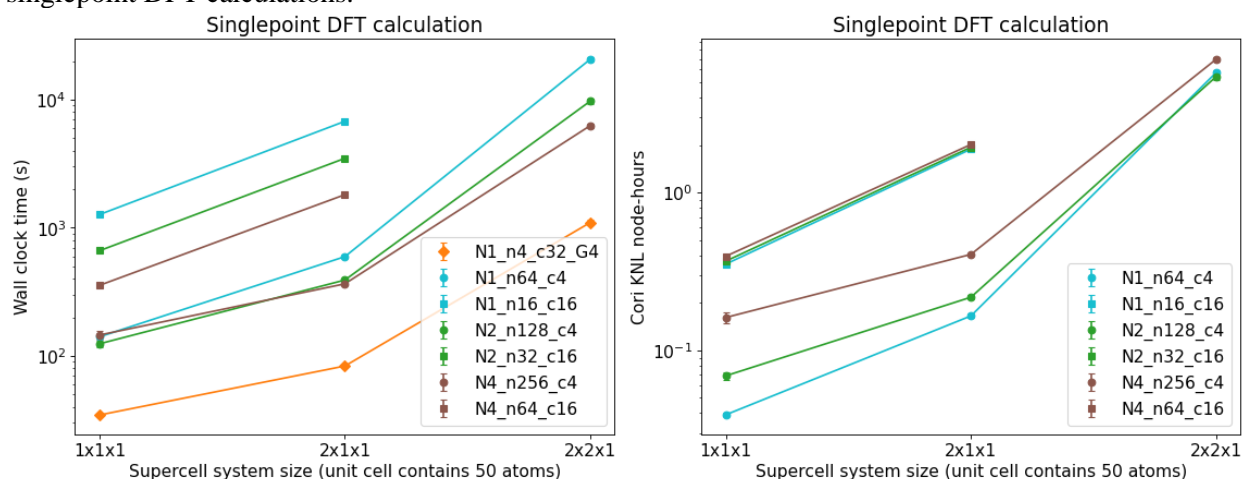


Fig. 10. Singlepoint DFT speed tests

As seen in Fig. 10, $\text{Li}_{20}\text{Ge}_2\text{P}_4\text{S}_{24}$ unit cell single-point calculations with 50 atoms (1x1x1) with 520eV energy cutoff and a 2x2x1 k-point grid take 0.1 node-hours to run on a single Cori KNL node which based on NERSC estimates is 4 times slower than Perlmutter, meaning 0.025 node-hours on all other clusters. These estimated times based on acceleration from Cori to Perlmutter also match prior speed tests conducted

in the Rafael Gomez-Bombarelli group for 2021-22 ALCC award titled “Inverse design of multicomponent oxide catalysts with generative models and DFT” on single-point calculations of 40 atom 1x1x1 perovskites on the Theta cluster, which took 0.02 node-hours. Speed tests on Cori for 2x2x1 supercells require a translated 0.05 node hours on non-Cori clusters for the same simulation parameters. Using this representative system to estimate the performance of training data collection for new stoichiometries, we will conduct 3,000 unit cell and 1000 supercell single-point calculations each for 300 stoichiometries (as many will be able to use the same training data), for a total of 1.2 million single-point DFT calculations for active learning training data generation, requiring **0.1725 million CPU node-hours split Summit (CPU)**. Notably, VASP with GPU on Perlmutter is significantly faster, by almost an order of magnitude, but since 4 GPUs are required, this may be reserved for singlepoint calculations with very large supercells. However, the bulk of training data for assessing bulk conductivity is done using the minimal representative structure, which in most cases can just be the unit cell.

Classical molecular dynamics simulations for the estimation of ionic diffusivities, solvation environments, and transference numbers consisting of 30k atoms will be conducted in LAMMPS with a real space cutoff of 12 Å with Particle Mesh Ewald summation beyond the cutoff. The working group has computational capabilities for training data generation and potential training through access to internal MIT clusters such as Supercloud and Engaging, and thus computational time for these will not be requested. LAMMPS was successfully installed and compiled for GPU acceleration on both Theta-GPU and Summit-GPU by working group member Pablo Leon. Speed tests were mainly conducted on Theta-GPU and these results will be described below. Systems of around 30,000 atoms are the standard for molecular dynamics simulations of polymer electrolyte systems, so strong scaling tests were conducted for a representative system of 33,160 containing 100-mer polyethylene oxide (PEO) chains with LiTFSI salt at an experimentally-relevant concentration (0.075 ethylene oxide monomer to salt molar ratio) with an AuTopology-trained class 2 interatomic potential. The results of this strong scaling test on Theta-GPU are described in Fig.11.

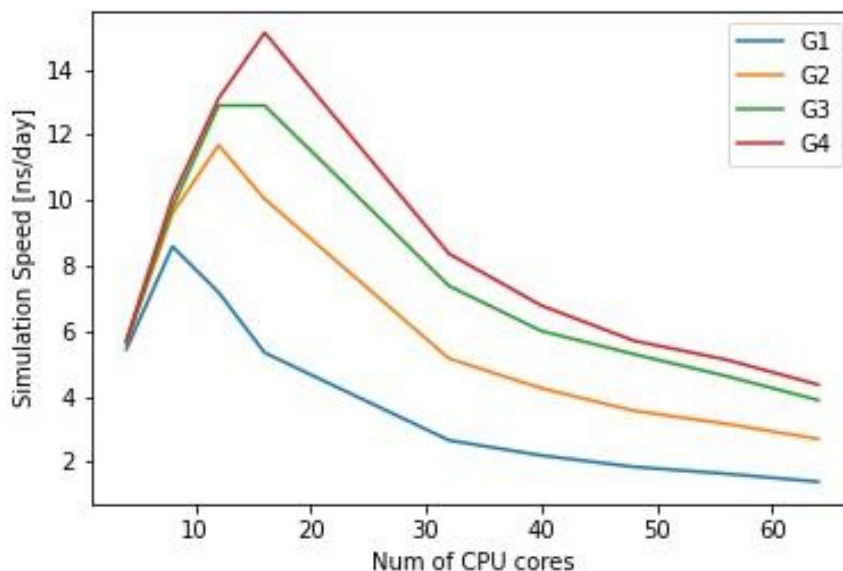


Fig. 11. GPU-accelerated classical MD speed tests on Theta-GPU

Although using 2 GPU with 12 CPU cores provides speeds of around 12 ns/day and is not the highest speeds seen in the speeding test, simulations of different chemistries and concentrations are embarrassingly parallel. Because Summit has 6 GPU and 44 CPU processors, three simulations run with 2 GPU and 12 CPU cores could be run in an embarrassingly parallel manner on a single node allowing a conversion to 36 ns/day or 0.67 node-hours/ns.

In order to explore the chemical space of possible polymer electrolytes, we request **1 million Theta-GPU node-hours** for 3 repetitions of 200 ns simulations for 500 polymers-salt systems at 5 different concentrations each. This is a total of 1.5 million ns of classical MD simulation. LAMMPS has well-documented restart commands that can be used to continue simulations past the maximum wall-clock restriction. These commands and necessary post-analysis python scripts to combine incomplete runs into a full trajectory have already been implemented widely by the proposing group to complete and analyze 200 ns runs of PEO-LiTFSI and liquid carbonates with lithium hexafluorophosphate (LiPF₆).

Having such a large range of polymer chemistries will allow us to exhaustively search the known backbone and functionalization space in a robust and cohesive manner, allowing for comparisons and determination of design principles in a manner not previously feasible in order to guide the exploration of novel polymer electrolyte systems. The proposed breadth of all atomistic simulation in collaboration with Co-PI Professor Yang Shao-Horn's novel high-throughput experimental machine for characterizing electrolyte conductivities provide the community with a fully comprehensive dataset of reliable and congruent experimental and simulation measurements.

All data will be stored long-term with the assistance of Co-PI Ben Blaiszek at ANL and through large archival and retrieval facilities like the Materials Data Facility. We expect the resource burnrate to have quartly peaks as the desired concentrations and alloying of new systems are determined and then set out for exploration. During those quarters, the simulations of the previous system will be analyzed using in-house python-based analysis tools.

3.2 Computational Approach

All calculations will be carried out using the Vienna Ab initio Simulation Package (VASP) for DFT, the Large-scale Atomistic/Molecular Massively Parallel Simulator (LAMMPS) for molecular dynamics, and PyTorch (<https://pytorch.org/>) for constructing, training, and employing deep learning models.

VASP and LAMMPS both handle MPI and mixed MPI-OpenMP parallelization internally, allowing for easy scaling of tasks to single node or multi node configurations. The single-point DFT calculations of active learning points for the deep neural network force field are embarrassingly parallel, as are the active learning loops and classical molecular dynamics simulations of distinct crystal stoichiometries and ion polymer electrolyte systems, respectively. Thus, the computational resources requested here will be used to sample the vast combinatorial space of crystalline framework elements and polymer backbones.

VASP is available for CPU-use only at and in a GPU-enabled build at NERSC and is widely applied within the proposing group as well as the broader community. The proposing group has licenses for VASP to enable use of allotted resources. LAMMPS is a free, open source software of equal renown in the molecular dynamics community as VASP to the quantum chemical community and has been used by the proposing group. LAMMPS is built with GPU capabilities at ANL and NERSC. GPU-accelerated LAMMPS builds have also been compiled by the working group on both Theta-GPU and Summit-GPU in addition to the Engaging and Supercloud clusters at MIT with combined MPI-OpenMP-Cuda acceleration in Intel Xen Silver 4216 and Xeon Gold 6248 CPUs with NVIDIA GeForce RTX 2080 and Volta V100 GPUs.

For the classical molecular dynamics simulations, the AuTopology parameter training automatically writes to LAMMPS-formatted input files, so only these single MB files need to be transferred to Summit to begin the simulation. The output of the simulation is a 20 GB trajectory file which can be converted to a Numpy array for conductivity and correlation analysis. The conversion will be performed and analyzed on Summit and then the final conductivity results will be combined into a pickled Pandas DataFrame object for transfer back to local computers for final visualization.

3.3 Developmental Work

VASP has well-optimized compiled models in both NERSC and ANL and the proposing team has extensive experience working with the software, including on Summit GPU through ERCAP Contract No. DE-AC02-05CH11231. Thus, production level results for the crystalline workflow are expected within the first few weeks. Additionally, 10 ns simulations of 4x4x4 supercell NNMD have already been run by working group member Gavin Winter through ERCAP0021172. The proposing team has significant experience compiling and running LAMMPS on a variety of platforms, including on MIT's (oversubscribed) Supercloud computing facility that contains the same NVIDIA V100 GPU as Summit, and LAMMPS has already been compiled and tested on Theta-GPU and Summit-GPU. Thus, classical molecular dynamics on Theta-GPU is also expected to begin within weeks. The meta-scheduling and database tasks are predicated on code that already exists and gets used daily by the 10+ members of the Gomez-Bombarelli group, whose database contains over 100 million in-house DFT simulations for molecules and solids.

Data acquisition

Data acquisition across all DFT calculations is embarrassingly parallel (linearly scaling). Similarly, data acquisition across different polymer chemistry and alloying are also embarrassingly parallel. Simulation of independent systems with DFT requires no communication between CPUs and can be made distributed and serial by centralizing results and proposing new steps from in the database on the fly. Our database-backed meta-scheduler and the small scale of each job allow us to perform the desired calculations with maximum throughput and make use of heterogeneous blocks in scheduling. Wall times for a full geometry optimization using VASP are expected to be within the allotted times. For LAMMPS, classical MD simulations will rely on the well-established restart functionality native to LAMMPS to reach 100s of ns simulations.

*The **Milestone Table** is appended to the end of this document (after the References section).*

4 REFERENCES

1. Zhao, Q., Stalin, S., Zhao, C. Z. & Archer, L. A. Designing solid-state electrolytes for safe, energy-dense batteries. *Nat. Rev. Mater.* 2020 53 **5**, 229–252 (2020).
2. Zhou, D., Shanmukaraj, D., Tkacheva, A., Armand, M. & Wang, G. Polymer Electrolytes for Lithium-Based Batteries: Advances and Prospects. *Chem* **5**, 2326–2352 (2019).
3. Karabelli, D., Birke, K. P. & Weeber, M. A Performance and Cost Overview of Selected Solid-State Electrolytes: Race between Polymer Electrolytes and Inorganic Sulfide Electrolytes. *Batter.* 2021, Vol. 7, Page 18 **7**, 18 (2021).
4. Zheng, F., Kotobuki, M., Song, S., Lai, M. O. & Lu, L. Review on solid electrolytes for all-solid-state lithium-ion batteries. *J. Power Sources* **389**, 198–213 (2018).
5. Aono, H., Sugimoto, E., Sadaoka, Y., Imanaka, N. & Adachi, G. ya. Ionic conductivity and sinterability of lithium titanium phosphate system. *Solid State Ionics* **40–41**, 38–42 (1990).
6. Murugan, R. *et al.* Fast Lithium Ion Conduction in Garnet-Type Li₇La₃Zr₂O₁₂. *Angew. Chemie Int. Ed.* **46**, 7778–7781 (2007).
7. Kamaya, N. *et al.* A lithium superionic conductor. *Nat. Mater.* 2011 109 **10**, 682–686 (2011).
8. Chandra, D., Day, C. S. & Barrett, C. S. Low- and high-temperature structures of neopentylglycol plastic crystal. *Powder Diffraction* **8**, 109–117 (1993).

9. Mindemark, J., Lacey, M. J., Bowden, T. & Brandell, D. Beyond PEO—Alternative host materials for Li⁺-conducting solid polymer electrolytes. *Prog. Polym. Sci.* **81**, 114–143 (2018).
10. An, Y. *et al.* Progress in Solid Polymer Electrolytes for Lithium-Ion Batteries and Beyond. *Small* **18**, 2103617 (2022).
11. Molinari, N., Mailoa, J. P. & Kozinsky, B. Effect of Salt Concentration on Ion Clustering and Transport in Polymer Solid Electrolytes: A Molecular Dynamics Study of PEO-LiTFSI. *Chem. Mater.* **30**, 6298–6306 (2018).
12. France-Lanord, A. & Grossman, J. C. Correlations from Ion Pairing and the Nernst-Einstein Equation. *Phys. Rev. Lett.* **122**, 136001 (2019).
13. Molinari, N. & Kozinsky, B. Chelation-Induced Reversal of Negative Cation Transference Number in Ionic Liquid Electrolytes. *J. Phys. Chem* **2020**, 3 (2022).
14. Savoie, B. M., Webb, M. A. & Miller, T. F. Enhancing Cation Diffusion and Suppressing Anion Diffusion via Lewis-Acidic Polymer Electrolytes. *J. Phys. Chem. Lett.* **8**, 641–646 (2017).
15. France-Lanord, A. *et al.* Effect of Chemical Variations in the Structure of Poly(ethylene oxide)-Based Polymers on Lithium Transport in Concentrated Electrolytes. *Chem. Mater.* **32**, 121–126 (2020).
16. Fong, K. D. *et al.* Ion Transport and the True Transference Number in Nonaqueous Polyelectrolyte Solutions for Lithium Ion Batteries. *ACS Cent. Sci.* **5**, 1250–1260 (2019).
17. Mo, Y., Ong, S. P. & Ceder, G. First principles study of the Li 10GeP 2S 12 lithium super ionic conductor material. *Chem. Mater.* **24**, 15–17 (2012).
18. Molinari, N., Mailoa, J. P. & Kozinsky, B. Effect of Salt Concentration on Ion Clustering and Transport in Polymer Solid Electrolytes: A Molecular Dynamics Study of PEO-LiTFSI. *Chem. Mater.* **30**, 6298–6306 (2018).
19. Hou, T., Kara Fong, ab D., Jingyang Wang ae, bc & Persson, K. A. The solvation structure, transport properties and reduction behavior of carbonate-based electrolytes of lithium-ion batteries †. (2021) doi:10.1039/d1sc04265c.
20. Bedrov, D. *et al.* Molecular Dynamics Simulations of Ionic Liquids and Electrolytes Using Polarizable Force Fields. *Chem. Rev.* **119**, 7940–7995 (2019).
21. Schütt, K. T., Schütt, S., Unke, O. T. & Gastegger, M. Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra. (2021).
22. Sun, H. Force field for computation of conformational energies, structures, and vibrational frequencies of aromatic polyesters. *J. Comput. Chem.* **15**, 752–768 (1994).
23. Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase Applications Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **102**, 7338–7364 (1998).
24. Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A. & Müller, K. R. SchNet - A deep

- learning architecture for molecules and materials. *J. Chem. Phys.* **148**, (2018).
25. Henkelman, G., Jóhannesson, G. & Jónsson, H. Chapter 10 Methods for Finding Saddle Points and Minimum Energy Paths. in *Theoretical methods in condensed phase chemistry* (2002).
 26. Rycroft, C. H. VORO++: A three-dimensional Voronoi cell library in C++. *Chaos An Interdiscip. J. Nonlinear Sci.* **19**, 041111 (2009).
 27. Schwalbe-Koda, D., Tan, A. R. & Gómez-Bombarelli, R. Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks. *Nat. Commun.* **2021 121** **12**, 1–12 (2021).
 28. He, X., Zhu, Y. & Mo, Y. Origin of fast ion diffusion in super-ionic conductors. *Nat. Commun.* **2017 81** **8**, 1–7 (2017).
 29. Molinari, N., Mailoa, J. P., Craig, N., Christensen, J. & Kozinsky, B. Transport anomalies emerging from strong correlation in ionic liquid electrolytes. *J. Power Sources* **428**, 27–36 (2019).
 30. Blaiszik, B. *et al.* A Data Ecosystem to Support Machine Learning in Materials Science. (2019).
 31. Musaelian, A. *et al.* Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. (2022).
 32. Batzner, S. *et al.* E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. (2021).

Proposal Title (exactly as it appears on submission): Machine learning classical potentials for design of organic electrolytes

Milestone:	Details (as appropriate):	Dates:
Active learning loop for neural network potential for boron-containing argyrodites and oxides	Resource: Summit GPU Node-hours: 86.25k Filesystem storage (TB and dates): 10 TB (temporary) Archival storage (TB and dates): N/A Software Application: VASP Tasks: single-point DFT calculations Dependencies: N/A	07/01/23-8/31/23
Neural Net Molecular Dynamics (NNMD) simulations on boron-containing argyrodites and oxides to determine optimal alloying and defect chemistries for ion hopping	Resource: Summit GPU Node-hours: 260k Filesystem storage (TB and dates): 100 TB (temporary) Archival storage (TB and dates): N/A Software Application: ASE, PyTorch Tasks: NNMD simulation Dependencies: Active learning loop for neural network potential for boron-containing argyrodites and oxides	09/01/23-12/31/23
Classical molecular dynamics simulations on functionalized polycarbonates and boron-containing polymer electrolytes	Resource: Theta GPU Node-hours: 500k Filesystem storage (TB and dates): 200 TB (temporary) Archival storage (TB and dates): N/A Software Application: LAMMPS Tasks: classical MD simulation Dependencies: Active learning loop for classical potential for functionalized polycarbonates	07/01/23-12/31/23
Active learning loop for neural network potential for organic plastic crystalline electrolytes	Resource: Summit GPU Node-hours: 86.25k Filesystem storage (TB and dates): 10 TB (temporary) Archival storage (TB and dates): N/A Software Application: VASP Tasks: single-point DFT calculations Dependencies: N/A	01/01/24-02/28/24
Neural Net Molecular Dynamics (NNMD) simulations on organic plastic crystalline electrolytes and defect chemistries for ion hopping	Resource: Summit GPU Node-hours: 260k Filesystem storage (TB and dates): 100 TB (temporary) Archival storage (TB and dates): N/A Software Application: ASE, PyTorch Tasks: NNMD simulation Dependencies: Active learning loop for neural network potential for organic plastic crystalline electrolytes	03/01/24-06/30/24
Classical molecular dynamics simulations for boron-containing polymer electrolytes	Resource: Theta GPU Node-hours: 500k Filesystem storage (TB and dates): 200 TB (temporary) Archival storage (TB and dates): N/A Software Application: LAMMPS Tasks: classical MD simulation Dependencies: Active learning loop for classical potential for boron-containing polymer electrolytes	01/01/23-06/30/23

PERSONNEL JUSTIFICATION AND MANAGEMENT PLAN

A multi-PI, multi-institution (MIT, Pacific Northwest National Lab) research program that combines elements of the proposed computation work with machine learning and high-throughput experimentation at the material and device scale has been awarded funding from the US Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office (AMO) Award Number AMO award DE-EE0009096. Financial support lasts 36 months with start date of 09/01/2020 and a renewal checkpoint that was successfully achieved on 02/28/2022.

In addition to the PIs identified in the proposal, this program provides broad support for personnel. Exclusively for tasks related to DFT and MD simulations and generative models, this includes 1.5 FTE appointments at the Gomez-Bombarelli group at MIT. Some of the personnel is already in place (such as 4rd year PhD student Pablo Leon, 3rd year PhD student Gavin Winter, and 2nd year PhD student Jurgis Ruza, all supervised by RGB at MIT DMSE). Co-PI Shao-Horn (MIT MechE) has extensive experimental and theoretical experience, including the development of a high throughput experimental machine that will be used in parallel to the computational screening to provide a replicable and comprehensive experimental dataset. Collaborators Jun Liu and Jie Xiao (PNNL) will perform experimental validation of the computational predictions by synthesizing and evaluating the electrolyte properties of the proposed materials while collaborator Elsa Olivetti (MIT DMSE) will perform natural language programming and text mining to determine appropriate experimental conditions and candidate material systems. PI Gomez-Bombarelli and Shao-Horn also have joint sponsored research program by Toyota Research Institute on design of polymer electrolytes, supporting 2.5 FTE on the computational and ML side until 2025.

We foresee very low project risk due to personnel turnover. The duration of the proposed ALCC program and the broader AMO support are significantly shorter than the turnover time for graduate student and co-PI appointments. In addition, the multi-FTE staffing offers opportunities for redundancy. Finally, because the proposed workflows are automated end-to-end, once the procedure is initialized, no further work is needed.

ADMINISTRATION DETAILS

As described in the research proposal, our approach involves autonomous potential training, molecular dynamics simulation, and analysis in an iterative pipeline that is agnostic to the user in charge of deploying them. Additionally, the program is designed to be guided from outside the HPC infrastructure by an autonomous database-backed engine housed in the lab of the Lead PI. This meta-scheduler has already been used to interface SLURM and Torque in a combination of hardware resources, including Harvard and MIT's research computing centers, NRSEC and Brookhaven National Laboratory, with peak performance occasionally passing 10,000 cores, 100+ GPU, and 500+ simultaneous calculations. Additionally, an unrelated framework to design multicomponent oxide catalysts has been successfully integrated with the Argonne's Theta supercomputer by other students in the Gomez-Bombarelli group through a separate ALCC allocation while neural network molecular dynamics simulations of crystalline materials has already been benchmarked on Summit-GPU through ERCAP0021172.

The co-PI's and collaborators will maintain ongoing conversations and triweekly meeting in the broader context of the AMO program and will contribute code and expertise towards extending and supervising the high-throughput simulation engine, regardless of individual allocations time. In principle, our complete program could run through a unique HPC user, with the co-PIs vision having been reflected in the HT settings.

Nevertheless, in order to facilitate human supervision accountability and division of labor, the distribution of efforts and resources is well partitioned. RGB is the main point of contact, supervises the code in the HT engine, and his team will fully design and deploy active learning loops and molecular dynamics simulations for crystalline systems (graduate student GAW) and amorphous systems (graduate student PAL, JR). All machine learning tasks will be performed by the RGB group using the infrastructures detailed in the Milestone Table. Lastly co-PI BB will coordinate data transfer and storage using existing tools such as the Materials Data Facility, and interface them with the RGB HT screening database.

PUBLICATIONS RESULTING FROM INCITE AWARDS

N/A

RAFAEL GOMEZ-BOMBARELLI

MIT Department of Materials Science and Engineering
77 Massachusetts Ave. Bldg 13-5037 Cambridge MA 02139
617 253 5632, rafagb@mit.edu

Professional Preparation

Universidad de Salamanca	Salamanca, Spain	Chemistry	B.S.+M.S., 2006
Universidad de Salamanca	Salamanca, Spain	Physical Chem.	Ph.D., 2011
Heriot-Watt University	Edinburgh, UK	Physics	Postdoc 2012-2014
Harvard University	Cambridge, MA	Chemistry	Postdoc 2014-2016

Appointments

Massachusetts Institute of Technology, Cambridge MA	
Assistant Professor, Materials Science and Engineering	2018-present
Kyulux North America, Boston MA	
Senior Scientist	2016-2018

Five Publications Most Relevant to This Proposal

1. Ang, S. J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Gomez-Bombarelli, R. Active learning accelerates ab initio molecular dynamics on pericyclic reactive energy surfaces. **2021** *Chem* 4 (6), 1189-1190
2. Schwalbe-Koda, D.; Kwon, S.; Paris, C.; Bello-Jurado, E.; Jensen, Z.; Olivetti, E.; Willhammar, T.; Corma, A.; Roman-Leshkov, Y.; Moliner, M.; Gomez-Bombarelli, R.; *A priori* control of zeolite phase competition and intergrowth with high-throughput simulations. **2021** *Science*. 374 308-315
3. Schwalbe-Koda, D*; Tan, AR*; Gomez-Bombarelli, R.; Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks **2021** *Nat. Commun.* 12, 5104
4. Qiao, B.; Mohapatra, S.; Lopez, J.; Leverick, G.; Tatara, R.; Shibuya, Y.; Jiang, Y.; France-Lanord, A.; Grossman, J. C.; Gómez-Bombarelli, R.; et al. Quantitative Mapping of Molecular Substituents to Macroscopic Properties Enables Predictive Design of Oligoethyleneglycol-Based Lithium Electrolytes. **2020**. *ACS Cent Sci* 6, 1115-1128
5. Wang, W.; Yang, T.; Harris, W.; Gomez-Bombarelli, R. Active Learning and Neural Network Potentials Accelerate Molecular Screening of Ether-Based Solvate Ionic Liquids. **2020**. *Chem Comm* 56, 8920-8923

Research Interests and Expertise

The RGB group at MIT DMSE carries out computational materials design using tools in the continuum between physics-based atomistic simulations and machine learning. High-throughput simulation pipelines use electronic structure and molecular dynamics simulations to screen novel materials and down select only the most promising ones for synthesis and laboratory testing. Machine-learning models are used as surrogate functions that replace expensive physics-based simulations, such as learning reactive interatomic potentials for organics and inorganics, or automatic building of coarse grained simulations. Novel statistical learning architectures and materials representations that capture physical principles are built. Descriptors extracted from simulations are used as inputs to train composition-structure-property models that guide materials discovery. Inverse design loops based on unsupervised generative models can automatically sample novel molecules and materials that maximize a desired property, without explicit enumeration of the search space. These tools are applied in collaboration with experimentalists in areas such as organic electronics (light emitting diodes, photodiodes), biomolecules like cell-penetrating peptides, multicomponent oxides catalysts for oxygen evolution, ion-conducting polymers or nanoporous silicates.

Synergistic Activities

1. **Journal peer reviewer:** *ACS Applied Mat and Interfaces*, *ACS Central Science*, *Advanced Materials*, *Advanced Energy Materials*, *Advanced Functional Materials*, *Bioch. Biophys. Acta*, *Chem*, *Chemical Science*, *Environmental Science and Technology*, *Joule*, *Journal of Chemical Information and Modeling*, *JACS*, *Journal of Chemical Theory and Computation*, *Journal of Cheminformatics*, *Journal of Materials C*, *Journal of Physical Chemistry*, *Journal of Physical Chemistry Letters*, *NPJ Computational Materials*, *Physical Chemistry Chemical Physics*, *PNAS*, *iScience*, *Scientific Reports*. **Conference peer reviewer:** Learning and Reasoning with Graph-Structured Data ICML 2019 Workshop; Graph Representation Learning Workshop at NeurIPS 2019. **Guest Associate Editor**, *Complexity*. Special issue *Complex Algorithms for Data-Driven Model Learning in Science and Engineering* (2018).
2. Consultant for Kyulux North America, Recursion Pharmaceuticals (2018)
3. Mentor at the International Mentor Program from the International Mentoring Foundation for the Advancement of Higher Education (2017)
4. LAUNCH Chemistry Big Think, Smarter Chemistry Challenge (2016-2017)
5. Founder of Calculario Inc. (2015)

Collaborators (past 5 years including name and current institution)

Thompson, Aidan	Sandia National Laboratory
Meredig, Bryce	Citrine Inc
Paris, Cecilia	Instituto de Tecnologia Quimica
Boronet, Mer	Instituto de Tecnologia Quimica
Moliner, Manuel	Instituto de Tecnologia Quimica
Varanashi, Kripa	MIT
Ang, Shi Jun	MIT
Barzilay, Regina	MIT
Buonassissi, Tonio	MIT
Coley, Connor	MIT
Corma, Avelino	Instituto de Tecnologia Quimica
Davenport, Timothy	UTRC
Green, William	MIT
Grossman, Jeffrey	MIT
Jaakkola, Tommi	MIT
Jensen, Klavs	MIT
Johnson, Jeremiah	MIT
Olivetti, Elsa	MIT
Pentelute, Bradly	MIT
Riley, Patrick	Google
Roman, Yuriy	MIT
Shao-Horn, Yang	MIT
Rimer, Jeff	University of Houston
Myers, Deborah	Argonne National Laboratory

Blaiszik, Ben	Argonne National Laboratory
Tabor, Daniel	Texas A&M
Sotos, Nancy	University of Illinois at Urbana–Champaign
Moore, Jeff	University of Illinois at Urbana–Champaign
Hartrampf, Nina	University of Zurich
Phillips, Scott	Boise State
Zhao, Xuanhe	MIT
Tuller, Harry L	MIT
Short, Michael P	MIT
Schuh, Christopher A	MIT
Ross, Caroline A	MIT
LeBeau, James	MIT
Hart, Anastasios J	MIT
Gedik, Nuh	MIT
Comin, Ricardo	MIT
Cappellaro, Paola	MIT
Beach, Geoffry S	MIT
Alamo, Jesus A	MIT
Cohen, Tal	MIT
Li, Ju	MIT
Tasan, Cem	MIT
Willard, Adam	MIT
Yildiz, Bilge	MIT
Ribbeck, Katharina	MIT
Smidt, Tess	LBNL
Das, Payel	IBM
Aspuru-Guzik, Alan	Toronto University
Aziz, Michael J	Harvard University
Dantus, Marcos	Michigan State University
Gordon, Roy	Harvard University
Harel, Elad	Northwestern University
Hirn, Matthew	Michigan State University
Hirzel, Timothy D	Kyulux North America
Kajino, Hiroshi	IBM
Piecuch, Piotr	Michigan State University
Adachi, Chiaya	Kyushu University
Adams, Ryan P	Princeton, Google
Beh, Eugene	Xerox
Chae, HyunSik	Samsung
Duvenaud, David	Toronto University
Gerhardt, Michale	Harvard University
Ha, Dong-Gwang	Samsung
Huh, Dal Ho	Samsung

Ihn, Soo-Ghang	Samsung
Jeon, Soon Ok	Amazon
Jung, Yongsik	Samsung
Kang, Hosuk	Samsung
Kim, Sunghan	Samsung
Lee, Namheon	Samsung
Lee, Sangyoon	Samsung
Lin, Kaixiang	Harvard University
Marshak, Michale	University of Colorado
Miyazaki, Hiroshi	Samsung
Numata, Masaki	Samsung
Sheberla, Denis	Harvard University
Sim, Myungsun	Samsung
Son, Young Mok	Samsung
Tong, Liuchuan	Harvard University
Wei, Jennifer	Google
Wong, Andrew	Harvard University
Wu, Tony	Toronto University
Sanchez-Lengeling, Benjamin	Google
Baldo, Marc	MIT
Chen, Qing	Hong Kong University
Einzinger, Markus	MIT
Fruchtman, Amir	Oxford University
Galvin, Cooper	Unknown
Gauger, Erik	Heriot-Watt University
Huang, Wenliang	MIT
Irish, Elinor	Southampton University
Kay, Christopher	University College London
Lovett, Brendon	Saint Andrews University
Maclaurin, Dougal	Google
MacRobert, Arthur	Heriot-Watt University
Morton, John	University College London
Noimark, Sacha	University College London
Parkin, I	University College London
Salvadori, Enrico	University College London
Valle, Alejandro	Unknown
Anderson, Harry	Cambridge University
Hong, Seongik	Cambridge University
Kelber, Julian	Strasbourg University
Markopoulos, Georgios	BASF
Panjwani, Naitik	Cambridge University
Pyzer-Knapp, Edward	IBM
Suh, Changwon	Nexight

Wu, Di	Cambridge University
Gagliardi, Laura	U Chicago
Ferguson, Andrew	U Chicago
Miller, Thomas	Caltech
Anandkumar, Anima	Caltech
Edelman, Alan	MIT
Rackauckas, Christopher	MIT
Marzouk, Youssef	MIT
Solomon, Justin	MIT
Van Voorhis, Troy	MIT

Yang Shao-Horn

Electrochemical Energy Laboratory
Department of Materials Science and Engineering
Department of Mechanical Engineering
MIT, Room 3-334, 77 Mass Ave, MA 02139

<http://web.mit.edu/eel>
shaohorn@mit.edu
Tel: (617) 253-2259
Fax: (617) 258-7018

Professional Preparation

Beijing University of Technology; Metallurgical and Materials Engineering; B.S.; Beijing, China 1992
Michigan Technological University; Metallurgical and Materials Engineering; Ph.D.; Houghton, MI 1998
ICMCB-CNRS; solid state chemistry; NSF International Research Fellow; Bordeaux, France 2000-2002

Appointments

W.M. Professor of Energy, MIT	7/2015-present
Professor, Department of Materials Science and Engineering, MIT	7/2013-present
Gail E. Kendall Professor in Mechanical Engineering, MIT	7/2011-6/2015
Associate Professor, Department of Mechanical Engineering, MIT	7/2007-present
Assistant Professor, Department of Mechanical Engineering, MIT	8/2005-6/2007
Atlantic Richfield Assistant Professor, Department of Mechanical Engineering, MIT	8/2002-6/2005
Staff Materials Scientist, Eveready Battery Company, Cleveland,	6/1998-9/2000

Five Publications Most Relevant to This Proposal

1. Hopkins, B.J., Y. Shao-Horn, and D. P. Hart, Suppressing Corrosion In Primary Aluminum–Air Batteries Via Oil Displacement, *Science*, **362**, 658-661 November 2018.
2. Hwang, J., R.R. Rao, L. Giordano, Y. Katayama, Y. Yu, and Y. Shao-Horn, Perovskites in Catalysis and Electrocatalysis, *Science*, **358**, 751-756 November 2017.
3. Sheberla, D., J. C. Bachman, J. S. Elias, C-J. Sun, Y. Shao-Horn, and M. Dinca, Conductive MOF Electrodes for Stable Supercapacitors with High Areal Capacitance, *Nature Materials*, **16**, 220-224 February 2017.
4. Grimaud, A., O. Diaz-morales, B.H. Han, W. T. Hong, Y.L. Lee, L. Giordano, K. A. Stoerzinger, M.T.M. Koper, Y. Shao-Horn, Activating lattice oxygen redox reactions in metal oxides to catalyze oxygen evolution, *Nature Chemistry*, **9**, 457-465 May 2017.
5. Bachman, J., S. Muy, Grimaud, A., H.H. Chang, N. Pour, S. Lux, O. Paschos, F. Maglia, S. Lupart, P. Lamp, L. Giordano and Y. Shao-Horn, Inorganic Solid-State Electrolytes for Lithium Batteries: Mechanisms and Properties Governing Ion Conduction, *Chemical Reviews*, **116**, 140-162 January 2016.

Research Interests and Expertise

Yang Shao-Horn studies materials for electrochemical and photoelectrochemical energy storage and conversion, which is centered on examining the influence of surface chemistry and electronic structures of thin films and nanomaterials on lithium storage and catalytic activity of small molecules of energy consequence, and applying fundamental understanding in reaction mechanisms to design new materials for lithium storage and electrocatalysis of oxygen reduction, water splitting, methanol oxidation and CO₂ reduction.

Synergistic activities:

- (a) Co-chaired the DoD workshop on Future Directions on Synthetic Biology for Energy and Power 2018; Co-Director for MIT Center on Energy Storage (2016-); MIT Energy Council (2018-); NSF MRSEC IRG Leader (2008-2013)
- (b) Member of National Academy of Engineering; the World's Most Influential Scientific Minds and *Highly Cited Researchers* (Thomson Reuters), Ideaslab of World Economic Forum in Davos 2017; advised 90+ students and postdoctoral associates at MIT, with 25+ in academia; received the Battery Division Research Award from the Electrochemical Society, the Research Award by the International Battery Association, the Faraday Medal from Royal Society of Chemistry, elected as a Fellow from the American Association for the Advancement of Science, the Royal Society of Chemistry, International Society of Electrochemistry, and the Electrochemical Society.

Collaborators

Abernathy, D. L.	ORNL
Adler, S. B.	U Washington
Allard, L.	ORNL
Almheiri, S.	Masdar Institute
Antonopoulos, B. K.	BMW Group
Ariando,	NUS
Bamgbopa, M.O.	Masdar Institute
Bansal, D.	ORNL
Belcher, Angela M.	MIT
Bulovic, Vladimir	MIT
Chiang, Y-M.	MIT
Chorkendorff, I.	DTU
Crabb, Emily J.	MIT
Christen, H.M.	ONRL
Cummins, Christopher C.	MIT
Delaire, O.	Duke/ORNL
Diaz-Morales, O.	Stockholm U.
Dokko, K.	YNU
Elias, J.	Harvard
Gallant, B.	MIT
Garg, A.	MIT
Gasteiger, H.	TUM, Germany
Grossman, Jeffrey C.	MIT
Jacobs, R.	U. Wisconsin-Madison
Jewett, M.	Northwestern U.
Kim, A.	IFM co.,Ltd
Kim, D.H.	IFM co.,Ltd
Kim, Jeehwan	MIT
Kim, T.Y.	SAIT
Kiong, C.W.	NUS
Koper, M.T.M.	Leiden U.
Lee, S.W.	Georgia Tech

Liu, Q.	LBNL
Mansour, A.	NSWC
Massimiliano, P.	Eni S.p.A
Meda, L.	Eni S.p.A.
Meng, S.	UCSD
Morgan, D.	U. Wisconsin-Madison
Mounfield III, W.P.	MIT
Nocera, D.G.	Harvard
Oldani, F.	Eni S.p.A.
Omampuliyur, R.	SMART
Østergaard, T.M.	U. Copenhagen
O'Sullivan, F.M.	MIT
Perego, D.	SMART
Phillips, K.	MIT
Risch, M.	Georg-August-Universität Göttingen
Rosowski, F.	BASF
Rossmeisl, Jan	U. Copenhagen
Ruettinger, W.	BASF
Sakti, A.	MIT
Sayed, S.Y.	Cairo U.
Schlögl, R.	Fritz Haber Institute, Max Planck Society
Spadavecchia, F.	Eni S.p.A.
Stephens, R.	Shell
Storey, B.	TRI Global
Suntivich, J.	Cornell
Thompson, Carl V.	MIT
Thu, Y.L.	MIT
van Gorp, J.J.	DSM
Varanasi, K.	MIT
Vegge, T.	DTU
Venkatesan, T.V.	NUSNNI
Voigt, Christopher	MIT
Wang, Evelyn	MIT
Wang, X.R.	NUS
Watanabe, Masayoshi	YNU
Weinberger, David S.	BASF
Willard, Adam P.	MIT
Xu, Z.	NTU
Yacoby, Y	Hebrew U.
Yao, K.P.C.	ANL

Ben Blaiszik
 University of Chicago – Globus
 Argonne National Laboratory - Data Science and Learning Division
 Argonne, IL 60439
 Phone: 630.252.1610
 Email: blaiszik@uchicago.edu

Professional Preparation

Elmhurst College	Elmhurst, IL	Physics/Math	B.S. 2003
University of Illinois at Urbana-Champaign	Urbana, IL	Theoretical and Applied Mechanics	M.S. 2005, Ph.D. 2009
University of Illinois at Urbana-Champaign	Urbana, IL	Materials Science	Oct 2009-2011
Argonne National Laboratory Director's Distinguished Postdoctoral Fellow	Lemont, IL	Center for Nanoscale Materials	2011-2013

Appointments

<i>Year(s)</i>	<i>Title</i>
2018-current	Research Scientist, Globus, University of Chicago / Argonne National Laboratory Data Science and Learning Division
2013-2018	Research Scientist, Globus, University of Chicago / Argonne National Laboratory
2012-2015	Founder of Inscites Research Inc., Chicago, IL
2011-2013	Director's Distinguished Postdoctoral Fellow, Argonne National Laboratory

Products

Peer-reviewed publications related to the proposed project

1. **B. Blaiszik**, L. Ward, M. Schwarting, J. Gaff, R. Chard, D. Pike, K. Chard, and I. Foster, "A data ecosystem to support machine learning in materials science," MRS Communications, pp. 1–9, Oct. 2019.
2. L. Ward, **B. Blaiszik**, I. Foster, R. S. Assary, B. Narayanan, and L. Curtiss, "Machine learning prediction of accurate atomization energies of organic molecules from low-fidelity quantum chemical calculations," MRS Communications, vol. 9, no. 03, pp. 891–899, Aug. 2019.
3. Abreha, Biruk, Snigdha Agarwal, Ian Foster, **Ben Blaiszik**, and Steven Lopez. "Virtual Excited State Reference for the Discovery of Electronic Materials Database (VERDE Materials DB): An open-access resource for ground and excited state properties of organic molecules." (2019). *The journal of physical chemistry letters* 10, no. 21 (2019): 6835–6841.
4. R. Chard, Z. Li, K. Chard, L. Ward, Y. Babuji, A. Woodard, S. Tuecke, **B. Blaiszik**, M.J. Franklin, and I. Foster. "DLHub: Model and Data Serving for Science." arXiv preprint arXiv:1811.11213 (2018) – accepted to IPDPS 2018.
5. L. Ward, M. Aykol, **B. Blaiszik**, I. Foster, B. Meredig, J. Saal, and S. Suram. "Strategies for accelerating the adoption of materials informatics." MRS Bulletin 43, no. 9 (2018): 683–689.

Research Interests

Ben is co-lead for the Materials Data Facility, DLHub, and various other data infrastructure projects. He also works closely with the Globus team, a project that is building platform as a service (PaaS) capabilities to simplify authentication, data movement, user groups, automation tasks and more for the research data community. Globus currently serves >120000 users across 1500 institutions in 80 countries. These projects are focused on building high-quality data services that specifically meet the growing data management challenges that scientific researchers face. He also co-leads a research group interested in solving problems in materials science and chemistry using machine learning techniques with specific interests in catalyst discovery and materials for next-generation energy storage applications.

Synergistic Activities

- Co-PI of the Materials Data Facility (MDF), a NIST-funded Materials Genome Initiative project to simplify the publication, indexing, subselecting, aggregation, and reuse of materials science data with a focus on large datasets (i.e., many TB or many millions of records). MDF holds over 45 TB of data, enables access to over 200 TB of data through its indexes, and has over 3.4M individual records (i.e., a set of results from an experiment or output files from a simulation) harvested from 116 diverse community sources.
- Co-PI for the Argonne DLHub project. DLHub will develop a flexible service to enable non-expert users to train and deploy machine learning models with data pulled from a variety of community data services, and train these models using ALCF resources. DLHub will also seek to catalog and make available best-in-class models from communities spanning materials science, chemistry, physics, genomics, and more.

Collaborators

Allison	John	University of Michigan Ann-Arbor
Ananthakrishnan	Rachana	University of Chicago
Antono	Erin	Citrine Informatics
Aykol	Muratahan	Toyota Research Institute
Bartolo	Laura	Northwestern University
Chard	Kyle	University of Chicago
Chard	Ryan	University of Chicago
Church	Carena P	Citrine Informatics
Davenport	Tim	UTRC
Ferris	Michael	University of Wisconsin-Madison
Franklin	Michael	University of Chicago
Gibbons	Breanna	Stanford University
Gomez-Bombarelli	Rafael	MIT
Hatrick-Simpers	Jason	National Institute of Standards and Technology
Hutchinson	Maxwell	Citrine Informatics
Ling	Julia	Citrine Informatics
McHenry	Kenton	University of Illinois at Urbana-Champaign
Mehta	Apurva	SLAC National Accelerator Laboratory

Meredig	Bryce	Citrine Informatics
Morgan	Dane	University of Wisconsin-Madison
Myers	Deborah	Argonne National Lab
Nahrstedt	Klara	University of Illinois at Urbana-Champaign
Niu	Changning	Questek Innovations
Olivetti	Elsa	MIT
Paradiso	Sean	Citrine Informatics
Perry	Michael	UTRC
Pruyne	James	University of Chicago
Saal	James E	QuesTek Innovations
Schleife	Andre	University of Illinois at Urbana-Champaign
Shao-Horn	Yang	MIT
Suram	Santosh	Toyota Research Institute
Trinkle	Dallas R	University of Illinois at Urbana-Champaign
Varanasi	Kripa	MIT
Voorhees	Peter	Northwestern University
Voyles	Paul	University of Wisconsin-Madison
Ward	Logan T	University of Chicago

Section 6: Software Applications and Packages

Question #1

Please list any software packages used by the project, and indicate if they are on open source or export controlled.

Application Packages

Package Name

VASP

Indicate whether Open Source or Export Controlled.

Export Controlled

Package Name

LAMMPS

Indicate whether Open Source or Export Controlled.

Open Source

Section 7: Wrap-Up Questions

Question #1

National Security Decision Directive (NSDD) 189 defines Fundamental Research as "basic and applied research in science and engineering, the results of which ordinarily are published and shared broadly within the scientific community, as distinguished from proprietary research and from industrial development, design, production, and product utilization, the results of which ordinarily are restricted for proprietary or national security reasons." Publicly Available Information is defined as information obtainable free of charge (other than minor shipping or copying fees) and without restriction, which is available via the internet, journal publications, textbooks, articles, newspapers, magazines, etc.

The INCITE program distinguishes between the generation of proprietary information (deemed a proprietary project) and the use of proprietary information as input. In the latter, the project may be considered as Fundamental Research or nonproprietary under the terms of the nonproprietary user agreement. Proprietary information, including computer codes and data, brought into the LCF for use by the project - but not for generation of new intellectual property, etc., using the facility resources - may be protected under a nonproprietary user agreement.

Proprietary Information

Are the proposed project and its intended outcome considered Fundamental Research or Publicly Available Information?

Yes

Will the proposed project use proprietary information, intellectual property, or licensing?

No

Will the proposed project generate proprietary information, intellectual property, or licensing as the result of the work being proposed?

If the response is Yes, please contact the INCITE manager, INCITE@doeleadershipcomputing.org, prior to submittal to discuss the INCITE policy on proprietary work.

No

Question #2

The following questions are provided to determine whether research associated with an INCITE proposal may be export controlled. Responding to these questions can facilitate - but not substitute for - any export control review required for this proposal.

PIs are responsible for knowing whether their project uses or generates sensitive or restricted information. Department of Energy systems contain only data related to scientific research and do not contain personally identifiable information. Therefore, you should answer "Yes" if your project uses or generates data that fall under the Privacy Act of 1974 U.S.C. 552a. Use of high-performance computing resources to store, manipulate, or remotely access any national security information is prohibited. This includes, but is not limited to, classified information, unclassified controlled nuclear information (UCNI); naval nuclear propulsion information (NNPI); and the design or development of nuclear, biological, or chemical weapons or of any weapons of mass destruction. For more information contact the Office of Domestic and International Energy Policy, Department of Energy, Washington DC 20585, 202-586-9211.

Export Control

Does this project use or generate sensitive or restricted information?

No

Does the proposed project involve any of the following areas?

i. Military, space craft, satellites, missiles, and associated hardware, software or technical data

ii. Nuclear reactors and components, nuclear material enrichment equipment, components (Trigger List) and associated hardware, software or technical data

iii. Encryption above 128 bit software (source and object code)

iv. Weapons of mass destruction or their precursors (nuclear, chemical and biological)

No

Does the proposed project involve International Traffic in Arms Regulations (ITAR)?

No

Question #3

The following questions deal with health data. PIs are responsible for knowing if their project uses any health data and if that data is protected. Note that certain health data may fall both within these questions as well as be considered sensitive as per question #2. Questions regarding these answers to these questions should be directed to the centers or program manager prior to submission.

Health Data

Will this project use health data?

No

Will this project use human health data?

No

Will this project use Protected Health Information (PHI)?

No

Question #4

The PI and designated Project Manager agree to the following:

Monitor Agreement

I certify that the information provided herein contains no proprietary or export control material and is correct to the best of my knowledge.

Yes

I agree to provide periodic updates of research accomplishments and to acknowledge INCITE and the LCF in publications resulting from an INCITE award.

Yes

I agree to monitor the usage associated with an INCITE award to ensure that usage is only for the project being described herein and that all U. S. Export Controls are complied with.

Yes

I understand that the INCITE program reserves the right to periodically redistribute allocations from underutilized projects.

Yes

Section 8: Outreach and Suggested Reviewers

Question #1

By what sources (colleagues, web sites, email notices, other) have you heard about the INCITE program? This information will help refine our outreach efforts.

Outreach

Question #2

Suggested Reviewers

Section 9: Testbed Resources

Question #1

The ALCF and OLCF have test bed resources for new technologies, details below. If you would like access to these resources to support the work in this proposal, please provide the information below.

(1 Page Limit)

The OLCF Quantum Computing User Program is designed to enable research by providing a broad spectrum of user access to the best available quantum computing systems, evaluate technology by monitoring the breadth and performance of early quantum computing applications, and Engage the quantum computing community and support the growth of the quantum information science ecosystems. More information can be found here: <https://www.olcf.ornl.gov/olcf-resources/compute-systems/quantum-computing-user-program/quantum-computing-user-support-documentation>.

The ALCF AI Testbed provides access to next-generation of AI-accelerator machines to enable evaluation of both hardware and workflows. Current hardware available includes Cerebras C-2, Graphcore MK1, Groq, Habana Gaudi, and SambaNova Dataflow. New hardware is regularly acquired as it becomes available. Up to date information can be found here: <https://www.alcf.anl.gov/alcf-ai-testbed>.

Describe the experiments you would be interested in performing, resources required, and their relationship to the current proposal. Please note, these are smaller experimental resources and a large amount of resources are not available. Instead, these resources are to explore the possibilities for these technologies might innovate future work. This request does not contribute to the 15-page proposal limit.

2023_INCITE_Publications_Testbed.pdf
The attachment is on the following page.

TESTBED RESOURCES

N/A