AL4GAP: Active Learning Workflow for generating DFT-SCAN Accurate Machine-Learning Potentials for Combinatorial Molten Salt Mixtures

Jicheng Guo,[1] Vanessa Woo,[2] David A. Andersson,[3] Nathaniel Hoyt,[4] Mark Williamson,[4] Ian Foster,[5] Chris Benmore,[6] Nicholas E. Jackson,[7] and Ganesh Sivaraman[8]

[1])*Chemical and Fuel Cycle Technologies Division, Argonne National Laboratory, Lemont, IL 60439, USA*

[2])*School of Electrical, Computer, Energy Engineering, Arizona State University, Tempe, AZ 85287, USA*

[3])*Materials Science and Technology Division, Los Alamos National Laboratory P.O. Box 1663, Los Alamos, NM 87545, USA*

[4])*Chemical and Fuel Cycle Technologies Division, Argonne National Laboratory, Lemont, IL 60439, USA*

[5])*Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA*

[6])*X-ray Science Division, Argonne National Laboratory, Lemont, IL 60439, USA*

[7])*Department of Chemistry, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA*

[8])*Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA*

(*Electronic mail: gsivaraman@anl.gov)

(Dated: 4 April 2023)

Machine learning interatomic potentials have emerged as a powerful tool for bypassing the spatio-temporal limitations of *ab initio* simulations, but major challenges remain in their efficient parameterization. We present AL4GAP, an ensemble active learning software workflow for generating multi-composition Gaussian approximation potentials (GAP) for arbitrary molten salt mixtures. The workflow capabilities includes: (1) setting up user-defined combinatorial chemical spaces of charge neutral mixtures of alkaline/alkaline earth metals/heavy elements (e.g. Neodymium, Thorium) with halide ions, (2) configurational sampling using low-cost empirical parameterizations, (3) active learning for down-selecting configurational samples for single point density functional theory calculations at the level of SCAN exchange-correlation functional, and (4) Bayesian optimization for hyperparameter tuning of two-body and many-body GAP models. We apply the AL4GAP workflow to showcase high throughput generation of five independent GAP models for multi-composition binary-mixture melts, each of increasing complexity with respect to charge valency and electronic structure, namely: LiCl-KCl, NaCl-CaCl$_2$, KCl-NdCl$_3$, CaCl$_2$-NdCl$_3$ and KCl-ThCl$_4$. Our results indicate that GAP models, can accurately predict structure for diverse molten salt mixture with DFT-SCAN accuracy, capturing the intermediate range ordering characteristic of the multi-valent cationic melts.

Tremendous progress has been achieved in the past decade developing data driven surrogate models for learning molecular potential energy surfaces (PES)[1–3]. Data-driven models are functional forms with high dimensionality, typically containing between $\sim 10^4$ to $\sim 10^5$ parameters, that are machine-learned from from *ab initio* training datasets. The high-dimensionality of data-driven models allows for the capture of complex intermolecular potentials with *ab initio* accuracy that cannot be fit to low dimensional ($\sim 10^2$) empirical interatomic potentials[4]. Further, data-driven models can bypass the spatio-temporal limitations of density functional theory (DFT) based molecular dynamics (MD) simulations by dramatically reducing the associated computational cost of dynamical simulations. Many data-driven model architectures have been developed over the years, with some of the most relevant being Behler-Parrinello neural network, Gaussian approximation potential (GAP), Spectral Neighbor Analysis Potential (SNAP) , moment tensor potential (MTP), ANI, FCHL, SchNet, MBTR, DeepMD, linear atomic cluster expansion and NequIP[4–15]. Taken together, these classes of data-driven intermolecular potentials facilitate a new frontier for *ab initio* quality dynamical modeling at unprecedented spatio-temporal scales.

The bottleneck for developing data-driven intermolecular potentials concerns the computational inefficiencies associated with diverse training database generation and model hyperparameter tuning. Active learning (AL) has been proposed as an efficient data sampling heuristic and applied to select the most promising training samples from large unlabelled sample pools, which are typically generated from equilibrium MD simulations[16,17]. However, equilibrium MD based active learning does not appropriately sample meta-stable and out of equilibrium training sub-regions which are necessary for avoiding unphysical traps or numerical instability of the MD simulations[18,19]. To overcome such limitations, a number of recent studies have employed active learning in combination with enhanced sampling or the direct incorporation of experimentally measured meta-stable structures to improve the diversity of the training databases[20–23]. Even with a diverse training database, model hyperparameters still need to be carefully tuned to arrive at a model that can provide accurate and numerically stable MD simulations. To this end, data-driven model hyperparameter tuning schemes employing Bayesian optimization, particle swarm, and genetic algo-

rithms[17,24,25] have been introduced to the community.

Molten salts represent an iconic use-case for machine learning interatomic potentials in which conventional modeling paradigms (e.g. empirical potentials, *ab initio* dynamics) cannot obtain the necessary balance of accuracy (high polarizability salts) and accessible spatiotemporal scales ($\sim$ 10's nm, ns) necessary to estimate important physical properties. Molten salts have broad applications in concentrated solar power systems, liquid metal batteries, rare earth element (REE) production and molten salt reactors[26–30], but often molten salt mixtures are required to obtain targeted physical chemical properties. Specifically, to lower the working temperature of molten salts, salts are typically combined to create mixtures of eutectic composition with low melting points. The complexity of the molten salt mixtures in technological applications is further increased by the inclusion of multivalent ions and radioactive heavy elements. As modeling systems of this level of complexity requires the explicit inclusion of many-body effects such as polarization which require a fully quantum mechanical treatment, there are at present no chemically generalizable approaches for predicting the liquid phase structure of complex molten salt mixtures at the spatio-temporal scales relevant for industrial applications.

In this article, we present an active learning software workflow with data-efficient sampling and hyperparameter tuning for multi-composition Gaussian approximation potentials (GAP) in mixtures of molten salts. This workflow is a culmination of a number of best practices we have gained through our past publications[17,18,21,31,32]. We begin by providing a brief overview of the multi-composition active learning workflow for GAP, termed as AL4GAP, with an emphasis on the Python classes of AL4GAP. The details of the specific methodologies can be found in our prior publications. The AL4GAP workflow currently supports arbitrary molten mixtures spanning 11 cations (Li, Na, K, Rb, Cs, Mg, Ca, Sr, Ba and two heavy species, Nd and Th) and 4 anions (F, Cl, Br and I). We provide an overview of the software framework built around this active learning workflow for sampling arbitrary multi-component molten salt mixture melts. Each of the individual Python classes is subsequently detailed, with additional details provided in the GitHub repository (`https://github.com/pythonpanda2/AL4GAP_JCP`)[33]. In the results section, we showcase the utility of the AL4GAP for the high-throughput generation of five independent GAP models for multi-composition binary-mixture melts each of increasing complexity with respect to multiply charged ionic system, and electronic structure namely: LiCl-KCl, NaCl-CaCl$_2$, KCl-NdCl$_3$, CaCl$_2$-NdCl$_3$, and KCl-ThCl$_4$. The results section provides two different simulation scenarios for the developed potentials. The first scenario employs the molten NaCl-CaCl$_2$ as a model system to understand the liquid structure as functions of composition and temperature. For the second scenario, we discuss the chemical physics insights drawn from melt simulations for all five chemical systems with equal mixture composition at 1200K.

Listing 1: CSV formatted input file with three mixture compositions defined.

```
Ca,Nd,Cl,T,rho
0.0681,0.1989,0.7330,1003,3179.618
0.1577,0.1317,0.7106,1003,2852.936
0.2723,0.0458,0.6819,1093,2321.894
```

I. METHODOLOGY

A. AL4GAP workflow

The goal of the AL4GAP workflow is to accelerate the development of the GAP model that uses two-body squared exponential and many-body smooth overlap of atomic positions (SOAP) kernel chemical descriptors[34,35]. This workflow builds upon our previous AL approach that consists of an unsupervised clustering algorithm combined with Bayesian optimization for on-the-fly hyperparameter tuning of the GAP model[17,36]. A standalone step-by-step tutorial of the AL scheme is provided in the GitHub repository under the subheading "Simple AL4GAP Tutorial" ("Notebook/tutorial.ipynb") along with an accompanying YouTube video[37]. The AL4GAP workflow illustrated in Figure 1, which details the actions in the GitHub wrapper script named 'driver.py' that deploys compute resources and executes the workflow in ensemble mode[38]. The setup and prerequisites are listed in the GitHub repository. Each of the execution blocks from Figure 1 are further discussed below.

Composition space: The workflow takes the user-defined composition space as an input for setting up the sampling task. In the present study, all the composition space is defined by the user in a comma separated value (CSV) format. An example is illustrated for the molten CaCl$_2$-NdCl$_3$ melt system (Listing 1) also found as the 'density.csv' file in the GitHub repository, where a particular mixture composition is converted into its elemental composition. This example is provided to the user for a minimal trial run. The first three column headers correspond to abbreviated elemental symbols and the corresponding fractional composition. The second from last column with label 'T' corresponds to the target experimental measurement temperature in Kelvin. If the temperature is not known it can be left as a zero value as the actual sampling is performed at an elevated temperature. The final column with label 'rho' corresponds to the density in the units of kg. m^{-3}.

The csv formatted input file is imported as a pandas object and passed as input to the 'setup_inputs' Python method inside the driver.py script (invoked as 'from AL4GAP.setup_inputs import setup_inputs'). The 'setup_inputs' uses the information read from the csv file to create an atomic coordinate file ('opls.data') and a corresponding LAMMPS input/ force field file ('opls.in') for each of the composition rows, piping these to a separate directory.

Generation of LAMMPS input: The density and composition read from the .csv files are used by 'setup_inputs' to randomly pack the atoms into a LAMMPS readable coordinate file. The 'ffparam' argument to the 'setup_inputs'
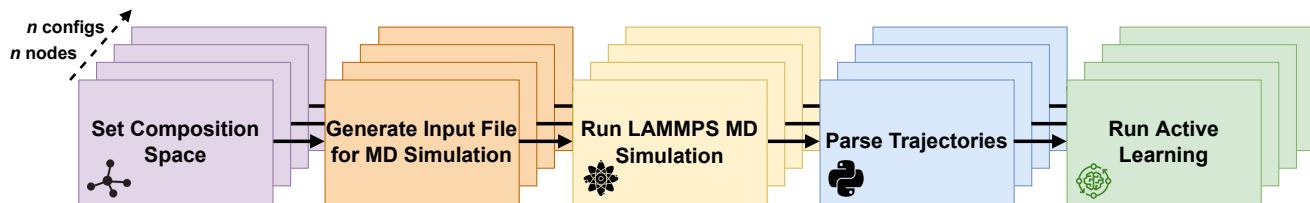
FIG. 1: AL4GAP Workflow for active leaning over a combinatorial composition map of molten salt mixtures. The compositions space is explored and sampled concurrently by providing dedicated compute resource for each composition.

method generates the necessary forcefield files needed for performing the LAMMPS simulation. The 'ffparam' can take two possible arguments namely 'TF' (i.e., rigid ion model) or 'OPLS' (Optimized Potentials for Liquid Simulations[39]) keywords. Our prior publications for molten LiCl and LiCl-KCl used the rigid ion based sampling which invoked the 'TF' keyword to perform RIM based sampling and the corresponding parametrization available in our GitHub only supports monovalent mixing of up to two salts[18,32]. The RIM model generator classes used in prior studies are found in the GitHub repository as 'moltensalt_tosifumi_gen.py' and 'tosi_fumi_params.py'. The 'OPLS' class was added to expand the number of chemical species in the composition space to support multiply charged cationic salt mixtures and is the new default argument. The 'setup_input' will internally invoke a Python class named 'moltensalt_gen' that contains all the coordinate generation and force field parameters for OPLS. A target "number of atom parameter ('set_target_number_of_atoms'), which is fixed at 64 atoms with a shortest distance cutoff of 2 Å('min_dist'), is used to arrive at the most plausible charge neutral atoms packing. The AL4GAP workflow through the OPLS sampler currently support 9 cations (Li, Na, K, Rb, Cs, Mg, Ca, Sr, and Ba), 4 anions (F, Cl, Br and I) and two heavy species (Nd and Th).

Run sampling with LAMMPS: After the input files are generated, the 'LAMMPS_ensemble' method embedded in the driver script is invoked to launch ensembles of MD simulations corresponding to the total number of compositions defined in the input. Each of the MD simulations (and subsequent steps) are allocated a dedicated computing resource to run sampling for compositions concurrently. Since there are three compositions used in the present example, three compute nodes are allocated with an additional compute node for database task. Further details are provided in driver submission bash script along with documentation listed in the GitHub. In our prior studies for LiCl-KCl, the RIM based MD sampling was performed at 2100 K. With the OPLS-based MD sampling used for all potential generation in this study, the temperature was further elevated to 5000K.

Parse MD trajectories: Following the completion of all MD simulations, the driver script invokes the 'AL_ensemble' method which parses the MD trajectories per composition and writes them in an extended xyz (exyz) format independent for each composition. Each of the parsed exyz files per composition contains $\sim$ 20,000 configurational samples.

Run active learning: Once the parsing of all the MD trajectories is completed, the 'AL_ensemble' method launches ensembles of active learning for each composition. The active learner acts on each of these compositions independently and parses them to find the most informative configurations (similar to the step-by-step tutorial listed earlier, except now each composition space is acted on by an independent instance of active learner concurrently).

B. DFT calculations

All training and validation "labels" produced by the AL4GAP workflows(Figure 1) are combined and single point DFT calculations are performed. Here we describe the generic calculation setting used in this study. Single point DFT calculations are performed on the configurations by using the SCAN exchange correlation (XC) functional[40], which shows superior performance compared to generalized gradient approximation (GGA) XC functionals[41,42]. DFT single point calculations are performed using the Vienna ab initio simulation package (VASP)[43]. The SCAN exchange-correlation functional and projector-augmented wave method are employed[40,44]. A large plane wave cutoff of 700 eV with an electronic convergence criterion of 10^{-7} eV is used. A Γ-centered $1 \times 1 \times 1$ k-mesh is used for reciprocal sampling. Spin polarization is applied to all the chemical mixtures involving heavy elements. The driver script can already handle many ensemble computational tasks and the DFT can very well be bundled along within that script. Here we do not choose to do so because we want to make the tutorial modular, and the users might have a different preferred choice for their reference electronic structure calculation code. An added benefit of this decoupling is that the sampling performed with the empirical forcefield is extremely cheap relative to expensive the DFT-MD runs and sanity checks can be done once the sampling is completed before proceeding with expensive single point DFT-SCAN calculations.

C. Hyperparameter optimization

Once the DFT-SCAN single point calculations for the AL4GAP training and validation sets are generated, the next step is to use Bayesian optimization (BO) to tune the a

GAP model hyperparameters employing the two body (2B) squared-exponential and many-body SOAP descriptors[17,45]. Here we use the DFT-SCAN dataset for molten KCl-ThCl$_4$ corresponding to the AL4GAP run for the composition space defined in Listing 6. The BO code and the datasets can be found in the Github folder 'HyperparameterOptimization/'. The BO script that utilizes the GPyOpt library[46] is implemented in the 'BayesOpt_SOAP.py'. This Python script performs BO to find the optimal values for the following hyperparamters: cutoff, scaling of kernel (2B, SOAP), number of representative sparse points (2B, SOAP), and number of angular and radial basis functions for SOAP. The best optimal hyperparameters found in the search range are written to the 'hyperparam_quip.json' file. More details can be found in the Github repository. The optimal BO hyperparameters are preserved through the iterative retraining discussed in the next subsection.

D. Metadynamics sampling:

An initial GAP model is fitted using the optimized hyperparameters for the AL4GAP-generated configurations across multiple-compositions (Fig. 1b), with "labels" computed using DFT-SCAN calculations. To circumvent the limitations of Boltzmann sampling, we showed in Guo et al. that metadynamics can be used to sample out-of-equilibrium training regions by utilizing the initial GAP model for LiCl-KCl as a model system[18]. Here we briefly reproduce the details in utilizing the well-tempered variation of Metadynamics sampling on the newly AL4GAP generated GAP model for an equal fraction of molten salt mixture (e.g., 50% molar fractions of LiCl and KCl) to perform configurational sampling near the melting point as a reference temperature, which potentially leads to improved coverage of the training space[47,48]. The unlike atom-atom ion pair (e.g., Li-Cl, K-Cl) coordination has been chosen intuitively as a collective variable (CV) to drive the exploration. The atom-atom pair coordination CV is parametrized with the values of the first minima in the partial PDFs. A Gaussian was deposited every 250 fs with a bias factor equal to 50. The potential is subject to retraining with an actively learned configuration drawn from this metadynamics sampling. A more detailed discussion is available in Guo et al.[18]

II. RESULTS

A. Application of AL4GAP to binary molten salt melts

We apply the AL4GAP framework to automatically generate GAP models for five independent binary molten salt melts. Here we carefully design the chemical systems and sampling space with close coordination with experimentalists[49]. These systems are LiCl-KCl, NaCl-CaCl$_2$, KCl-NdCl$_3$, CaCl$_2$-NdCl$_3$ and KCl-ThCl$_4$. The CSV formatted composition mapping input file used for AL4GAP for each of these

systems can be seen in the Appendix A Listing 2 to Listing 6. The composition space for LiCl-KCl (monovalent/monovalent cation) in Listing 2 corresponds to the input used in Guo et al.[18] The composition space for NaCl-CaCl$_2$ (monovalent/bivalent cations) corresponds to the eight composition rows, including two for pure salt melts, as shown in Listing 3. The composition space for KCl-NdCl$_3$ (monovalent/ trivalent cations) corresponds to the six composition rows, including two for pure salt melts, as shown in Listing 4. The composition space for CaCl$_2$-NdCl$_3$ (bivalent/ trivalent cations) corresponds to the five composition rows, including two for pure salt melts, as shown in Listing 5. Finally, the composition space for KCl-ThCl$_4$ (monovalent/ tetravalent cations) corresponds to the five composition rows, including two for pure salt melts, as shown in Listing 6.

B. Training Database, Model and Validation

The training databases have been summarized in Table I. As noted in subsection II A, there is a variation in the input compositions used for different molten salt mixtures. Consequently, the systems with larger input composition spaces exhibit a larger number of active learned training samples. The BO-optimized hyperparameters used to fit the final GAP models on these training databases are listed in Table II). The atomic forces validated for independent test samples drawn at each composition are visualized in Figure 2. The GAP models provide excellent force prediction accuracy with respect to DFT-SCAN across different molten salt mixture chemistries, with root mean squared errors (RMSE) ranging from 0.12 to 0.17 eV/Å. The larger distribution of atomic force in the LiCl-KCl system can be attributed to smaller chemical species having relatively higher diffusivity.

System	Training samples
LiCl-KCl	1127
NaCl-CaCl$_2$	801
CaCl$_2$-NdCl$_3$	762
KCl-NdCl$_3$	839
KCl-ThCl$_4$	632

TABLE I: GAP training database

C. GAP-MD simulation

The final GAP models are used to perform MD simulations with a system size over ∼1000 atoms. Detailed information on the simulated systems, their number of atoms, and densities estimated from GAP-MD for each composition and temperature are summarized in Appendix B Table V. We performed GAP MD using the LAMMPS software package compiled with the QUIP pair style[50,51]. Each simulation condition is initially thermalized at 1500K in the (NVT) ensemble[52,53], followed by volume relaxation in an isothermal-isobaric (NPT) ensemble with a pressure coupling of 1 bar[54–56]. The tem-

Parameter Name	LiCl-KCl		NaCl-CaCl$_2$		KCl-NdCl$_3$		CaCl$_2$-NdCl$_3$		KCl-ThCl$_4$	
	2B	SOAP	2B	SOAP	2B	SOAP	2B	SOAP	2B	SOAP
Cut off (Å)	5.92	5.92	5.97	5.97	6.074	6.074	5.594	5.594	6.11	6.11
Sparse points	65	1200	65	1100	65	1300	65	1300	55	1500
Delta (eV)	2.74	0.78	2.11	0.65	8.09	0.99	7.31	0.70	11.11	0.89
(l$_{max}$, n$_{max}$)	-	(4,8)	-	(4,8)	-	(6,9)	-	(4,9)	-	(4,9)

TABLE II: GAP model hyperparameter
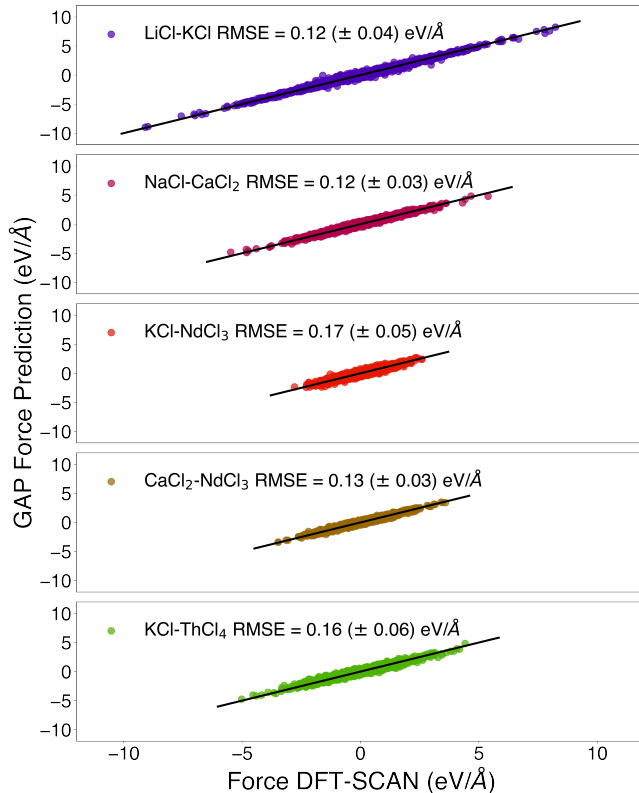
FIG. 2: Comparison of GAP models against exact DFT-SCAN-derived forces computed for diverse configurations drawn across independent compositions not explicitly included in the model training.

perature is decreased to the target temperature over 200 ps in the NPT ensemble. At the target temperature the NPT MD is continued for ~2 ns with a time-step of 0.5 fs and the last 1 ns is used for computing the structure. A representative set of simulated molten salt chemistries are shown in Figure 3. The results of these simulations are presented in the next two subsections.

D. Benchmark for molten NaCl-CaCl$_2$

From the MD simulation trajectory, we calculated the partial pair distribution function (PDF) of NaCl-CaCl$_2$ across multiple compositions from pure CaCl$_2$ (1200 K), NaCl-2CaCl$_2$ (903 K), NaCl-CaCl$_2$ (813 K), and 2NaCl-CaCl$_2$ (923

K) to pure NaCl (1148 K). The simulation temperature and compositions for mixtures were chosen to be consistent with the study of Igarashi et al.[58] The partial PDF functions are Fourier transformed to obtain partial structure factors and then weighted according to the Faber-Ziman formalism to obtain the total X-ray structure factor, which was then Fourier transformed to the total pair distribution function[59,60]. The total PDF is visualized in Figure 4. The first peak of the total PDF corresponds to the cation-anion interaction (i.e., Na-Cl and Ca-Cl) in the first coordination shell. As Na-Cl and Ca-Cl bond lengths are very close (i.e., 2.70 vs. 2.71 Å), the composition change causes minimal changes in the first peak position of the total PDF. The second peak of the total PDF is mainly from the Cl-Cl interaction, the length of which increases as the NaCl content increases as shown in Figure 4. This agrees with the x-ray diffraction (XRD) experimental results reported by Igarashi et al.[58] This change is further supported by the Cl-Cl partial PDF data (Figure 5).

The coordination number (CN) of each ion pair was calculated by integrating the partial PDF to its first minima and compared with previous studies in Table III[31,58,61]. We noticed that the Ca-Cl coordination number in pure CaCl$_2$ and the Na-Cl coordination number in pure NaCl are around 6.33 and 5.40 respectively, which are close to their crystalline state value of 6. Ca^{2+} is surrounded with more Cl$^-$ within the 1st coordination shell in the molten state than in the crystalline state, while Na$^+$ is surrounded by fewer. In the NaCl-CaCl$_2$ mixture, the coordination number for both Na-Cl and Ca-Cl decreases as the NaCl content increases.

The total structure factor data of the NaCl-CaCl$_2$ mixture at various compositions are shown in Figure 6a. A pre-peak at approximately 1.44 Å^{-1} appears in the CaCl$_2$ S(Q) plot which is not present in the NaCl S(Q) plot (zoomed in Figure 6b). A peak near 1.04 Å^{-1} appears in the structure factor when NaCl is introduced into CaCl$_2$. These two peaks are most likely the results of intermediate range order in CaCl$_2$ and CaCl$_2$-NaCl mixtures which are also observed in other AX$_2$ systems[64–68]. The peak at 1.04 Å^{-1} is probably attributable to the periodicity of Cl$^-$ in adjacent ionic networks (chains) as mentioned by Wu et al.[68] It is also worth pointing out that the simulations using the RIM model cannot accurately predict the structure of mixtures with multiply charged cations[69,70]. This means the ML-MD can accurately capture both the details in the short-range structure and the intermediate-range structure of complex molten salt mixtures.

We also explore the effect of temperature on the structure of molten NaCl-CaCl$_2$. The PDF plots shown in Figure 7a for 50-50 NaCl-CaCl$_2$ at 813.15K and 1200K indicate that the overall peak intensity reduces as temperature increases indi-
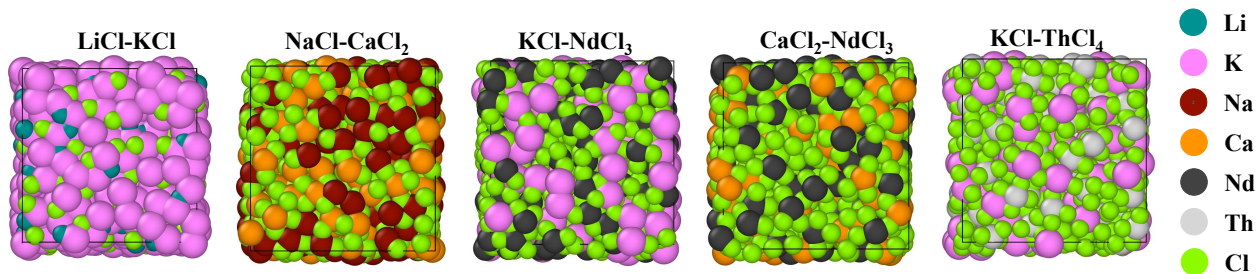
FIG. 3: MD simulation snapshots of the molten salt mixtures explored as a part of the combinatorial screening. The visualization is performed using OVITO[57].

Chemistry	Method	Temperature (K)	Na-Na		Na-Cl		Na-Ca		Ca-Ca		Ca-Cl		Cl-Cl	
			r (Å)	CN	r (Å)	CN	r (Å)	CN	r (Å)	CN	r (Å)	CN	r (Å)	CN
CaCl$_2$	GAP-MD	1200	-	-	-	-	-	-	4.52	10.75	2.72	6.33	3.68	16.31
	XRD[58]	1063	-	-	-	-	-	-	4.60	6.90	2.76	5.20	3.55	7.60
	DFT-MD[62]	1200	-	-	-	-	-	-	4.55	-	2.73	6.20	3.66	-
NaCl-2CaCl$_2$	GAP-MD	903	4.16	4.18	2.71	6.35	4.38	8.21	4.53	7.63	2.73	6.06	3.73	16.31
NaCl-CaCl$_2$	GAP-MD	813	4.16	6.26	2.71	6.22	4.46	6.79	4.50	5.96	2.73	6.06	3.79	16.41
	GAP-MD	1200	4.15	6.28	2.71	6.01	4.41	6.38	4.56	5.94	2.71	5.99	3.81	16.06
2NaCl-CaCl$_2$	GAP-MD	923	4.14	9.09	2.69	5.79	4.36	4.49	4.43	4.36	2.72	5.93	3.83	15.44
NaCl	GAP-MD[31]	1148	4.03	15.39	2.70	5.40	-	-	-	-	-	-	4.12	14.93
	NDIS+XRD[63]	1093	4.01	15.20	2.68	4.70	-	-	-	-	-	-	4.03	15.10
	PIM-MD	1100	4.05	15.40	2.66	5.20	-	-	-	-	-	-	4.07	15.30

TABLE III: Bond lengths and coordination numbers of ion pairs in the NaCl-CaCl$_2$ mixture with various compositions. Literature values are also reported when available in their original sources. All new GAP-MD simulations performed as a part of this study except for molten NaCl. Neutron diffraction with isotope substitution (NDIS). Polarizable ion model (PIM).
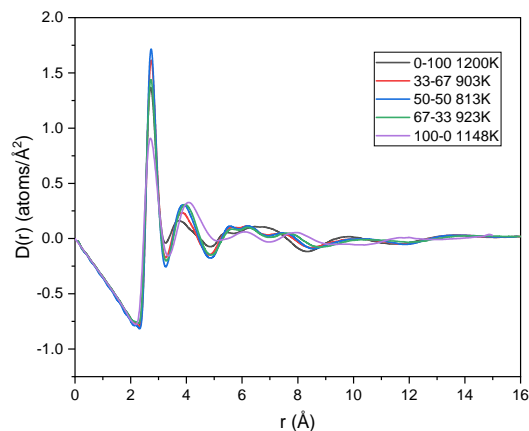

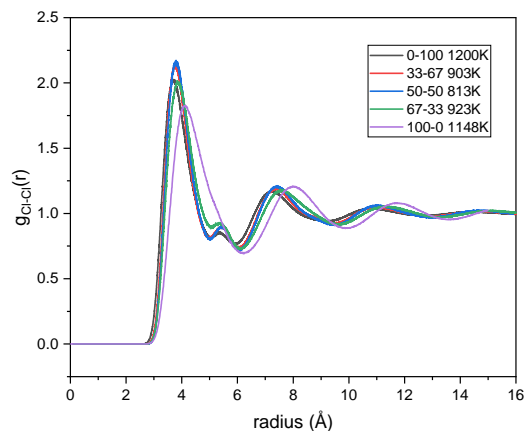FIG. 4: The pair distribution function plot for different compositions of molten NaCl-CaCl$_2$.


FIG. 5: Cl-Cl partial pair distribution functions for different compositions of molten NaCl-CaCl$_2$.

cating the atomic density decreases. The change also reflects in the coordination number (Table III) as the Na-Cl, Ca-Cl, and Cl-Cl coordination number decreases from 6.22, 6.06, and 16.41 to 6.01, 5.99, and 16.07, respectively. The structure factor (Figure 7b) shows that the pre-peaks at 1.0Å^{-1} and 1.4Å^{-1} increase while the first major peak at 2.48Å^{-1} reduces as temperature increases from 813K to 1200K. Similar behavior is observed in the molten MgCl$_2$-KCl system[64,68]. This suggests that at higher temperatures, shorter cation-anion-cation chains are favored which leads to more ions contributing to the inter-chain interaction and as a result, the enhanced pre-peak.
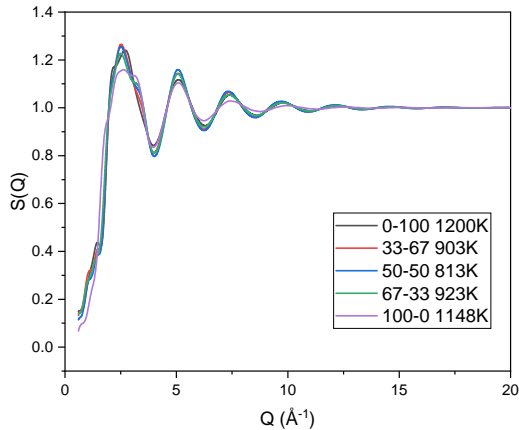
(a)
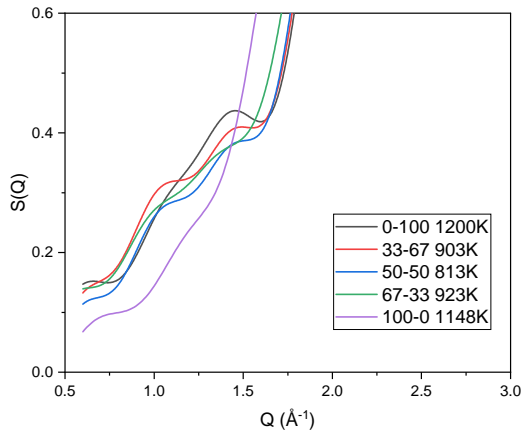
(b) Close up of Prepeaks in (a)

FIG. 6: Structure factor of NaCl-CaCl$_2$ mixture of different compositions.

E. GAP-MD simulation for comparing molten salts chemistry

Here, we also looked at the structure of multiple bi-valent, mono-valent, tri-valent, and quad-valent salt mixtures. In the total pair distribution function shown in Figure 8a, for light elements (i.e., Li, K, Ca, Na), the 1st peak has contributions from the cation-anion bond and the 2nd peak is mainly due to the contribution from the Cl-Cl bond. However, for mixtures with heavy elements, even though the Cl concentration is higher, according to the weight factor, the main contributions to the 2nd and 3rd peaks are from the heavy cation – heavy cation (e.g., Th-Th and Nd-Nd) correlations. The detailed structure information, such as bond length and coordination number for individual bonds, is derived from the partial PDF of these salt mixtures and listed in Table IV. The average Th-Cl coordination number (CN) in the KCl-ThCl$_4$ mixture was found to be 6.21, which is consistent with prior stud-

ies indicating that at high ThCl$_4$ concentrations, the primary species in ThCl$_4$-ACl (A = alkali metal) are 6-coordinated octahedra, including linked ThCl^{6-} and chained [Th$_n$Cl$_{4n+2}$]$^{2-}$ and [Th$_n$Cl$_{4n+2}$]$^{2+}$.[71,72] Similarly, the average Nd-Cl CN in the KCl-NdCl$_3$ mixture was 6.33, also in agreement with previous research suggesting that a distorted octahedral loose network structure predominates in NdCl$_3$-ACl mixtures with high NdCl$_3$ concentrations (> 25 mol%).[73] The structure factor of different salt mixtures is shown in Figure 8b. Salt mixtures with ThCl$_4$ and NdCl$_3$ show very intense pre-peaks at around ~ 0.8Å^{-1} to ~ 1.05Å^{-1} which indicates a strong intermediate range ordering.

III. CONCLUSION

We have described a software workflow, AL4GAP, for accelerating the development of machine learning interatomic potentials (ML-IP) via active leaning over a combinatorial composition space. The workflow provides an easy-to-use interface for setting up efficient sampling over arbitrary mixtures of molten salt chemistries. AL4GAP employs low cost empirical forcefields for sampling, and thereby effectively bypassing the need for expensive DFT-MD for training dataset generation. We also provide an easy-to-use interface for the Bayesian optimization of ML-IP model hyperparameters.

We showcase the power of AL4GAP in enabling facile generation of GAP models at DFT-SCAN level for five different three-component molten salt mixtures. Our results indicate that GAP models can accurately capture the structural characteristics of complex mixtures of molten salts. Our workflow accelerates the development of DFT-SCAN accurate multi-composition ML-IPs, which can be used for high accuracy property prediction across various compositions and simulation conditions opening up the possibility for providing rapid feedback/guidance to experiments in regimes challenged by high corrosion and radiation.

(a) Pair distribution function

(b) Structure factor

FIG. 7: NaCl-CaCl2 with 50-50 mol% composition simulated at 2 different temperatures.

(a) Pair distribution function

(b) Structure factor
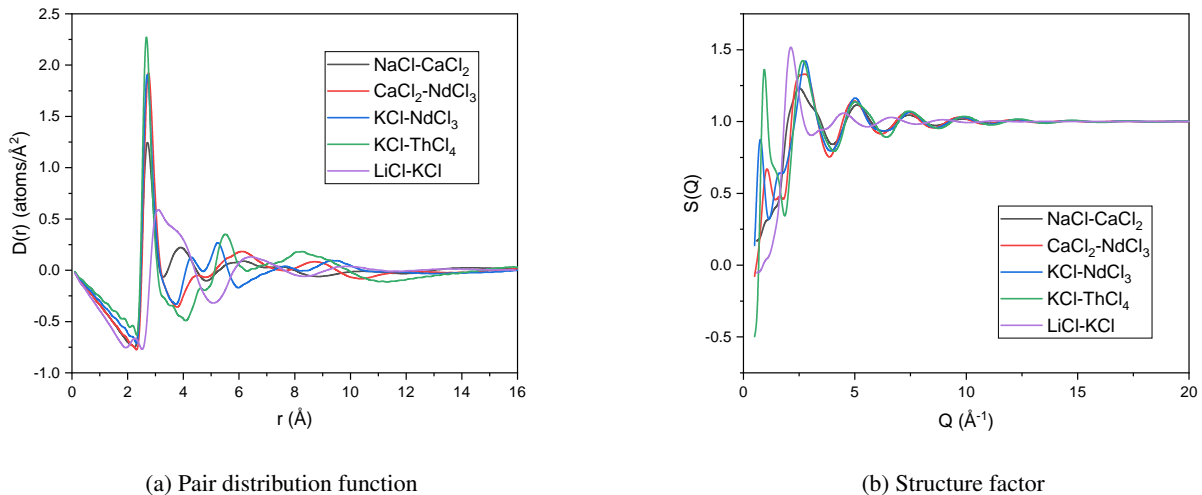
FIG. 8: Salt mixtures at 50-50 mol% composition simulated at 2 different temperatures.

IV. DATA AND CODE AVAILABILITY

The AL4GAP workflow is provided under an MIT license at Ref.[33](`https://github.com/pythonpanda2/AL4GAP_JCP`). The GAP model, training data, and the MD trajectories have been deposited in the Ref.[74].

Appendix A: Composition space for five molten salt mixture chemistry

Chemistry	r (Å)	CN	r (Å)	CN	r (Å)	CN	r (Å)	CN	r (Å)	CN	r (Å)	CN
KCl-ThCl$_4$	K-K		K-Cl		K-Th		Th-Th		Th-Cl		Cl-Cl	
	5.16	4.79	3.20	8.12	5.26	6.28	4.62/5.42	2.02	2.67	6.21	3.70	13.58
KCl-NdCl$_3$	K-K		K-Cl		K-Nd		Nd-Nd		Nd-Cl		Cl-Cl	
	5.25	5.29	3.07	6.19	5.04	6.31	4.21/5.16	2.96	2.69	6.33	3.53	14.46
CaCl$_2$-NdCl$_3$	Ca-Ca		Ca-Cl		Ca-Nd		Nd-Nd		Nd-Cl		Cl-Cl	
	4.86	4.02	2.71	6.32	4.55	4.09	4.52	2.92	2.74	6.78	3.58	17.15
LiCl-KCl	Li-Li		Li-Cl		Li-K		K-K		K-Cl		Cl-Cl	
	3.28	4.80	2.29	4.05	3.90	7.96	4.62	10.35	3.02	6.70	3.93	10.07
NaCl-CaCl$_2$	Na-Na		Na-Cl		Na-Ca		Ca-Ca		Ca-Cl		Cl-Cl	
	4.15	6.28	2.71	6.01	4.41	6.38	4.56	5.94	2.71	5.99	3.81	16.06

TABLE IV: Bond lengths and coordination numbers of ion pairs in salt mixtures at 1200K with 50-50 mol% compositions.

Mixture chemistry	Mixture composition	Number of atoms	Temperature (K)	Density from GAP-MD (g.cm^{-3})
LiCl-KCl	50-50	1024	1200	1.587 ($\pm$ 0.015)
NaCl-CaCl$_2$	33-67	1024	903.15	1.805 ($\pm$ 0.017)
	50-50	1040	813.15	1.799 ($\pm$ 0.013)
		1040	1200	1.690 ($\pm$ 0.015)
	67-33	1064	923.15	1.661 ($\pm$ 0.012)
KCl-NdCl$_3$	50-50	1008	1200	2.354 ($\pm$ 0.028)
CaCl$_2$-NdCl$_3$	50-50	1008	1200	2.506 ($\pm$ 0.023)
KCl-ThCl$_4$	50-50	1008	1200	2.551 ($\pm$ 0.021)

TABLE V: GAP MD simulation set up. The mixture compositions expressed in mol%. A 50-50 composition of LiCl:KCl with 256 anion-cation pairs for each salt would be equivalent to 1024 atom system.

Listing 2: CSV formatted input file for LiCl-KCl

```
Li ,K, Cl ,T, rho
0.45 ,0.05 ,0.5 ,820 ,1552
0.4 ,0.1 ,0.5 ,760 ,1598
0.35 ,0.15 ,0.5 ,740 ,1622
0.3335 ,0.1665 ,0.5 ,720 ,1646
0.25 ,0.25 ,0.5 ,780 ,1627
0.21 ,0.29 ,0.5 ,860 ,1595
0.165 ,0.335 ,0.5 ,900 ,1583
0.1 ,0.4 ,0.5 ,980 ,1548
0.05 ,0.45 ,0.5 ,1020 ,1534
0 ,0.5 ,0.5 ,1060 ,1520
```

Listing 3: CSV formatted input file for NaCl-CaCl$_2$

```
Na , Ca , Cl ,T, rho
0.500000 ,0.000000 ,0.500000 ,1090.0000 ,1554.5
0.425000 ,0.050000 ,0.525000 ,1090.0000 ,1635.0
0.337500 ,0.108333 ,0.554167 ,1090.0000 ,1742.4
0.245500 ,0.169667 ,0.584833 ,1060.0000 ,1854.8
0.179500 ,0.213667 ,0.606833 ,1080.0000 ,1912.9
0.112500 ,0.258333 ,0.629167 ,1080.0000 ,1971.9
0.099000 ,0.267333 ,0.633667 ,1040.0000 ,2002.9
0.000000 ,0.333333 ,0.666667 ,1070.0000 ,2073.2
```

Appendix B: GAP-MD simulations of molten salt mixtures

Listing 4: CSV formatted input file for KCl-NdCl$_3$

```
K, Nd , Cl ,T, rho
0 ,.25 ,.75 ,1090 ,3250.3474
.03705692803 ,.231471536 ,.731471536 ,997 ,3216.625
.08927519151 ,.2053624042 ,.7053624042 ,1045 ,2954.175
.202247191 ,.1488764045 ,.6488764045 ,1083 ,2920.545
.3680555556 ,.06597222222 ,.5659722222 ,1093 ,1872.805
.50 ,0 ,.50 ,1122 ,1493.702
```

Listing 5: CSV formatted input file for CaCl$_2$-NdCl$_3$

```
Ca , Nd , Cl ,T, rho
0.00 ,.2500 ,.7500 ,1270 ,3082.9222
.0681 ,.1989 ,.7330 ,1003 ,3179.618
.1577 ,.1317 ,.7106 ,1003 ,2852.936
.2723 ,.0458 ,.6819 ,1093 ,2321.894
.3333 ,0.00 ,.6667 ,1098 ,2122.448
```

[1] J. Behler and G. Csányi, "Machine learning potentials for extended systems: a perspective," The European Physical Journal B **94**, 1–11 (2021).

[2] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, "Gaussian process regression for materials and molecules," Chemical Reviews **121**, 10073–10141 (2021).

[3] M. Ceriotti, "Beyond potentials: Integrated machine learning models for materials," MRS Bulletin , 1–9 (2022).

[4] D. P. Kovács, C. v. d. Oord, J. Kucera, A. E. Allen, D. J. Cole, C. Ortner, and G. Csányi, "Linear atomic cluster expansion force fields for organic molecules: beyond rmse," Journal of chemical theory and computation **17**, 7696–7711 (2021).

[5] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Physical review letters **98**, 146401 (2007).

[6] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the elec-

Listing 6: CSV formatted input file for KCl-ThCl$_4$

```
K,Th,Cl,T,rho
0,.20,.80,1075,3317
.0771,.1692,.7538,883,3188.55
.1416,.1433,.715,748,3171.64
.3494,.0602,.5903,925,2111
.50,0,.50,1075,1506.375
```

trons," Physical review letters **104**, 136403 (2010).

[7] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, "Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials," Journal of Computational Physics **285**, 316–330 (2015).

[8] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, "The mlip package: moment tensor potentials with mpi and active learning," Machine Learning: Science and Technology **2**, 025002 (2020).

[9] J. S. Smith, O. Isayev, and A. E. Roitberg, "Ani-1: an extensible neural network potential with dft accuracy at force field computational cost," Chemical science **8**, 3192–3203 (2017).

[10] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "Schnet–a deep learning architecture for molecules and materials," The Journal of Chemical Physics **148**, 241722 (2018).

[11] A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. Anatole von Lilienfeld, "Fchl revisited: Faster and more accurate quantum machine learning," The Journal of chemical physics **152**, 044107 (2020).

[12] H. Huo and M. Rupp, "Unified representation of molecules and crystals for machine learning," Machine Learning: Science and Technology **3**, 045017 (2022).

[13] R. Lot, F. Pellegrini, Y. Shaidu, and E. Küçükbenli, "Panna: Properties from artificial neural network architectures," Computer Physics Communications **256**, 107402 (2020).

[14] A. M. Cooper, J. Kästner, A. Urban, and N. Artrith, "Efficient training of ann potentials by including atomic forces via taylor expansion and application to water and a transition-metal oxide," npj Computational Materials **6**, 1–14 (2020).

[15] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," Nature communications **13**, 1–11 (2022).

[16] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, "On-the-fly active learning of interpretable bayesian force fields for atomistic rare events," npj Computational Materials **6**, 1–11 (2020).

[17] G. Sivaraman, A. N. Krishnamoorthy, M. Baur, C. Holm, M. Stan, G. Csányi, C. Benmore, and Á. Vázquez-Mayagoitia, "Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide," npj Computational Materials **6**, 1–8 (2020).

[18] J. Guo, L. Ward, Y. Babuji, N. Hoyt, M. Williamson, I. Foster, N. Jackson, C. Benmore, and G. Sivaraman, "Composition-transferable machine learning potential for licl-kcl molten salts validated by high-energy x-ray diffraction," Phys. Rev. B **106**, 014209 (2022).

[19] D. Montes de Oca Zapiain, M. A. Wood, N. Lubbers, C. Z. Pereyra, A. P. Thompson, and D. Perez, "Training data selection for accuracy and transferability of interatomic potentials," npj Computational Materials **8**, 1–9 (2022).

[20] L. Bonati and M. Parrinello, "Silicon liquid structure and crystal nucleation from ab initio deep metadynamics," Physical review letters **121**, 265701 (2018).

[21] G. Sivaraman, G. Csanyi, A. Vazquez-Mayagoitia, I. T. Foster, S. K. Wilke, R. Weber, and C. J. Benmore, "A combined machine learning and high-energy x-ray diffraction approach to understanding liquid and amorphous metal oxides," Journal of the Physical Society of Japan **91**, 091009 (2022).

[22] N. Bernstein, G. Csányi, and V. L. Deringer, "De novo exploration and self-guided learning of potential-energy surfaces," npj Computational Materials **5**, 1–9 (2019).

[23] C. van der Oord, M. Sachs, D. Kovacs, C. Ortner, and G. Csanyi, "Hyperactive learning for data-driven interatomic potentials," (2022).

[24] G. C. Sosso *et al.*, "Soap_gas," `https://github.com/gcsosso/SOAP_GAS` (2022).

[25] S. K. Natarajan and M. A. Caro, "Particle swarm based hyper-parameter optimization for machine learned interatomic potentials," arXiv preprint arXiv:2101.00049 (2020).

[26] J. Dupont, "From molten salts to ionic liquids: a "nano" journey," Accounts of Chemical Research **44**, 1223–1231 (2011).

[27] H. Kim, D. A. Boysen, J. M. Newhouse, B. L. Spatocco, B. Chung, P. J. Burke, D. J. Bradwell, K. Jiang, A. A. Tomaszowska, K. Wang, *et al.*, "Liquid metal batteries: past, present, and future," Chemical Reviews **113**, 2075–2099 (2013).

[28] H. Gougar, D. Petti, P. Demkowicz, W. Windes, G. Strydom, J. Kinsey, J. Ortensi, M. Plummer, W. Skerjanc, R. Williamson, *et al.*, "The us department of energy's high temperature reactor research and development program–progress as of 2019," Nuclear Engineering and Design **358**, 110397 (2020).

[29] J. Guo, N. Hoyt, and M. Williamson, "Multielectrode array sensors to enable long-duration corrosion monitoring and control of concentrating solar power systems," Journal of Electroanalytical Chemistry **884**, 115064 (2021).

[30] M. Mehos, C. Turchi, J. Vidal, M. Wagner, Z. Ma, C. Ho, W. Kolb, C. Andraka, and A. Kruizenga, "Concentrating solar power gen3 demonstration roadmap," Tech. Rep. (National Renewable Energy Lab, Golden, CO, United States, 2017).

[31] S. Tovey, A. Narayanan Krishnamoorthy, G. Sivaraman, J. Guo, C. Benmore, A. Heuer, and C. Holm, "Dft accurate interatomic potential for molten nacl from machine learning," The Journal of Physical Chemistry C **124**, 25760–25768 (2020).

[32] G. Sivaraman, J. Guo, L. Ward, N. Hoyt, M. Williamson, I. Foster, C. Benmore, and N. Jackson, "Automated development of molten salt machine learning potentials: application to licl," The Journal of Physical Chemistry Letters **12**, 4278–4285 (2021).

[33] G. Sivaraman *et al.*, "Active learning workflow for gaussian approximation potential (al4gap)," `https://github.com/pythonpanda2/AL4GAP_JCP` (2023).

[34] A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," Physical Review B **87**, 184115 (2013).

[35] V. L. Deringer and G. Csányi, "Machine learning based interatomic potential for amorphous carbon," Physical Review B **95**, 094203 (2017).

[36] G. Sivaraman, L. Gallington, A. N. Krishnamoorthy, M. Stan, G. Csányi, Á. Vázquez-Mayagoitia, and C. J. Benmore, "Experimentally driven automated machine-learned interatomic potential for a refractory oxide," Physical Review Letters **126**, 156002 (2021).

[37] G. Sivaraman, "Ml-ip 2021, a psi-k tutorial workshop : From atomistic to coarse grained," `https://youtu.be/yDDMNh2-fbk`.

[38] S. Partee, M. Ellis, A. Rigazzi, A. E. Shao, S. Bachman, G. Marques, and B. Robbins, "Using machine learning at scale in numerical simulations with smartsim: An application to ocean climate modeling," Journal of Computational Science **62**, 101707 (2022).

[39] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids," Journal of the American Chemical Society **118**, 11225–11236 (1996).

[40] J. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly constrained and appropriately normed semilocal density functional," Physical review letters **115**, 036402 (2015).

[41] J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, *et al.*, "Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional," Nature chemistry **8**, 831–836 (2016).

[42] D. Tisi, L. Zhang, R. Bertossa, H. Wang, R. Car, and S. Baroni, "Heat transport in liquid water from first-principles and deep neural network simulations," Physical Review B **104**, 224202 (2021).

[43] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," Physical review B **54**, 11169 (1996).

[44] P. E. Blöchl, "Projector augmented-wave method," Physical review B **50**, 17953 (1994).

[45] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of bayesian optimization," Proceedings of the IEEE **104**, 148–175 (2015).

[46] J. González and Z. Dai, "GPyOpt: A Bayesian optimization framework in Python," (2016).

[47] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: a smoothly converging and tunable free-energy method," Physical review letters **100**, 020603 (2008).

[48] A. Laio and M. Parrinello, "Escaping free-energy minima," Proceedings of the National Academy of Sciences **99**, 12562–12566 (2002).

[49] G. J. Janz, *Molten salts handbook* (Elsevier, 2013).

[50] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," Journal of computational physics **117**, 1–19 (1995).

[51] G. Csányi, S. Winfield, J. R. Kermode, A. De Vita, A. Comisso, N. Bernstein, and M. C. Payne, "Expressive programming for computational physics in fortran 95+," IoP Comput. Phys. Newsletter , Spring 2007 (2007).

[52] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," The Journal of chemical physics **81**, 511–519 (1984).

[53] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," Physical review A **31**, 1695 (1985).

[54] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," Journal of Applied physics **52**, 7182–7190 (1981).

[55] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," The Journal of chemical physics **101**, 4177–4189 (1994).

[56] W. Shinoda, M. Shiga, and M. Mikami, "Rapid estimation of elastic constants by molecular dynamics simulation under constant stress," Physical Review B **69**, 134103 (2004).

[57] A. Stukowski, "Visualization and analysis of atomistic simulation data with ovito–the open visualization tool," Modelling and simulation in materials science and engineering **18**, 015012 (2009).

[58] K. Igarashi, T. Nijima, and J. Mochinaga, "Structure of molten cacl$_2$-nacl mixture," Proceedings of the first international symposium on molten salt chemistry and technology : Kyoto, Japan , 469 (April 20-22, 1983).

[59] T. Faber and J. Ziman, "A theory of the electrical properties of liquid metals: III. The resistivity of binary alloys," Philosophical Magazine **11**, 153–173 (1965).

[60] D. A. Keen, "A comparison of various commonly used correlation functions for describing total scattering," Journal of Applied Crystallography **34**, 172–177 (2001).

[61] X. Wei, D. Chen, S. Liu, W. Wang, J. Ding, and J. Lu, "Structure and thermophysical properties of molten calcium-containing multi-component chlorides by using specific bmh potential parameters," Energies **15**, 8878 (2022).

[62] M. Bu, W. Liang, G. Lu, and J. Yu, "Static and dynamic ionic structure of molten cacl 2 via first-principles molecular dynamics simulations," Ionics **27**, 771–779 (2021).

[63] A. Zeidler, P. S. Salmon, T. Usuki, S. Kohara, H. E. Fischer, and M. Wilson, "Structure of molten nacl and the decay of the pair-correlations," The Journal of Chemical Physics **157**, 094504 (2022).

[64] S. Day and R. McGreevy, "Structure factors of molten caci2 and mgci2 at low q," Physics and Chemistry of Liquids an International Journal **15**, 129–136 (1985).

[65] R. McGreevy and L. Pusztai, "The structure of molten salts," Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences **430**, 241–261 (1990).

[66] P. S. Salmon, "The structure of molten and glassy 2: 1 binary systems: an approach using the bhatia—thornton formalism," Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences **437**, 591–606 (1992).

[67] A. Zeidler, P. Chirawatkul, P. S. Salmon, T. Usuki, S. Kohara, H. E. Fischer, and W. S. Howells, "Structure of the network glass-former zncl2: From the boiling point to the glass," Journal of Non-Crystalline Solids **407**, 235–245 (2015).

[68] F. Wu, S. Sharma, S. Roy, P. Halstenberg, L. C. Gallington, S. M. Mahurin, S. Dai, V. S. Bryantsev, A. S. Ivanov, and C. J. Margulis, "Temperature dependence of short and intermediate range order in molten mgcl2 and its mixture with kcl," The Journal of Physical Chemistry B **124**, 2892–2899 (2020).

[69] M. Wilson and P. Madden, ""prepeaks"and "first sharp diffraction peaks"in computer simulations of strong and fragile ionic liquids," Physical review letters **72**, 3033 (1994).

[70] M. Salanne and P. A. Madden, "Polarization effects in ionic solids and melts," Molecular Physics **109**, 2299–2315 (2011).

[71] A.-L. Rollet and M. Salanne, "Studies of the local structures of molten metal halides," Annual Reports Section" C"(Physical Chemistry) **107**, 88–123 (2011).

[72] G. áM Photiadis *et al.*, "Co-ordination of thorium (iv) in molten alkali-metal chlorides and the structure of liquid and glassy thorium (iv) chloride," Journal of the Chemical Society, Dalton Transactions , 3541–3548 (1999).

[73] G. Photiadis, B. Brresen, and G. Papatheodorou, "Vibrational modes and structures of lanthanide halide–alkali halide binary melts lnbr 3–kbr (ln= la, nd, gd) and ndcl 3–acl (a= li, na, k, cs)," Journal of the Chemical Society, Faraday Transactions **94**, 2605–2613 (1998).

[74] J. Guo, V. Woo, D. Andersson, N. Hoyt, M. Williamson, I. Foster, C. Benmore, N. E. Jackson, and G. Sivaraman, "Metadynamics enhanced training Datasets, DFT-SCAN accurate GAP Model and MD trajectories for "AL4GAP: Active Learning Workflow for generating DFT-SCAN Accurate Machine-Learning Potentials for Combinatorial Molten Salt Mixtures"," (2023), 10.6084/m9.figshare.22534981.v1.

[75] J. Jerden, "Molten salt thermophysical properties database development: 2019 update," Tech. Rep. ANL/CFCT-19/6 154904; TRN: US2000034 (Argonne National Lab, Argonne, IL, United States, 2019).

[76] J. W. Mcmurray, K. Johnson, C. Agca, B. R. Betzler, D. J. Kropaczek, T. M. Besmann, D. Andersson, and N. Ezell, "Roadmap for thermal property measurements of molten salt reactor systems," Tech. Rep. (Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2021).