

Bayesian Machine Learning for Chemical Model Building - Research Overview

H. Deng,¹ H. O. Scheiber,¹ R. Krems,¹ M. Thachuk,¹ and G. N. Patey¹

*Department of Chemistry, University of British Columbia, Vancouver,
British Columbia, Canada V6T 1Z1*

(Dated: 17 April 2020)

I. WHAT WE HAVE ACHIEVED SO FAR

The research began with the development of two optimization algorithms. The first algorithm, written by Hayden, takes as input an initial guess geometry for an infinite LiX crystal and a model pairwise interaction of Coulomb-Lennard-Jones (CLJ, e.g. Jeung-Cheatham model¹) or Coulomb-Buckingham (e.g. Tosi-Fumi^{2,3}) type. It outputs the nearest local minimum of energy configuration, with its full geometry and 0K energetics. The algorithm allows for flexibility in the number of constraints. There are three levels of constraint:

1. The most constrained (and fastest) optimization constrains the fractional coordinates of atoms, the space group, and the full symmetry of the ideal/experimental unit cell. For example, this level of constraint applied to rocksalt optimization requires optimization of a single lattice parameter, a with the rest of the geometry being related by constraints.
2. At this level of constraint fractional coordinates of atoms within the unit cell are allowed to change, but the cell symmetry and space group remain fixed.
3. The fractional coordinates and cell symmetry are unconstrained. The unit cell is placed into the P1 space group and all atoms in the unit cell are allowed to optimize freely.

One should note that at all levels of constraint, the number of atoms in the unit cell remains fixed. The algorithm itself is a straightforward gradient descent algorithm, which is described in detail in the appendix. Note that the calculation of energies is performed via GROMACS, but GROMACS does not have an in-build optimization algorithm for lattice vectors so it was necessary to write one. We found that it was sufficient to use the most constrained optimization without significant deviation from less constrained options, as this greatly sped up calculation times.

The second algorithm, written by Hancock, is a Bayesian machine learning algorithm which is able to efficiently optimize input parameters in high-dimension spaces to multiple targets simultaneously. This second algorithm uses the first algorithm as a “black-box”, placing inputs and extracting outputs without regard to the internal workings. Hancock has developed his algorithm to search through the space of Lennard-Jones (LJ) parameters of a CLJ pairwise force field, which acts as an *effective potential* to use in classical lithium halide chemical simulations. The targets of his search are variable, but were always experimentally derived or calculated from high-accuracy DFT simulations of lithium halide crystals using CRYSTAL17⁴ software.

The full details of this algorithm can be found in Hancock’s thesis. In the case of the CLJ type (JC-like) potential,

$$u_{ij}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (1)$$

the algorithm takes as input a range for 6 (or 4 when using combining rules) model parameters, a LiX salt, and experimentally derived targets such as rocksalt lattice parameters and lattice energies. The model parameters to optimize are ϵ_{ij} and σ_{ij} . When using combining rules, only ϵ_{ii} and σ_{ii} are input. The remaining parameters are constructed from the Lorentz-Berthelot combining rules^{5,6}

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad (2)$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}. \quad (3)$$

The radius of cross-term LJ well is the arithmetic mean of the two radii, and the well depth is the geometric mean.

Hancock’s algorithm searches through the range of model parameters in an initial training phase, then begins optimization of the model parameters towards the given targets. The

algorithm uses Bayesian statistics to constantly update its internal knowledge of the best possible parameters. In this way, the algorithm is made extremely efficient for searching high-dimensional parameter spaces.

After a period of time spent exploring the effect of using various targets for Hancock’s algorithm, we settled on a series of LJ parameters (for the CLJ pairwise model) which are optimized as much as possible to the following targets:

1. The energy difference between rocksalt and wurtzite crystal structures is $15 - 5 \text{ kJ mol}^{-1}$, with rocksalt as the overall lowest energy crystal structure.
2. The experimental rocksalt lattice parameters.
3. The experimental (or DFT-calculated for LiF) wurtzite lattice parameters.

It was found that one *cannot* optimize to these targets and the *absolute* experimental lattice energy of the LiX rocksalt crystal structure simultaneously. It seems that only two of three types of targets can be solved simultaneously.

- The lattice parameters and relative energies crystal structures can be simultaneously solved for, leaving large errors in the absolute lattice energy.
- The lattice parameters and absolute rocksalt lattice energies can be simultaneously solved for, but wurtzite becomes the lowest energy crystal structure.
- The relative energies of the crystal structures and the absolute rocksalt lattice energy can be solved for, but lattice parameters become too large.

A few other observations are notable. The most useful properties to solve for when studying LiX nucleation (one possible end goal of this research) are relative crystal structure stability and (rocksalt) lattice parameters. When these targets are chosen for model parameter optimization, a trend in absolute rocksalt lattice energies appears. One observes

that the smaller the halide of the LiX salt, the larger the error in the absolute lattice energy becomes when the other targets are optimized. Furthermore, decreasing the relative difference in energy target between rocksalt and wurtzite structures from 15 to 0 kJ mol⁻¹ systematically reduces the error in the absolute rocksalt lattice energy. Something in the model itself appears to prevent optimization to all physical properties of LiX salts simultaneously. It is not currently known why this deficiency exists.

It was suggested that damping of the model’s $1/r^6$ dispersion term at close range could tip results in favour of rocksalt, since wurtzite crystals have a shorter nearest-neighbour bond length than rocksalt for a given salt and model. A fixed-parameter Becke-Johnson “rational damping”^{7,8} factor was introduced, transforming the $1/r^6$ term into

$$\frac{C_{6,ij}}{R_{0,ij}^6 + r_{ij}^6} \quad (4)$$

where $C_{6,ij} = 4\epsilon_{ij}\sigma^6$ and $R_{0,ij}$ defines the radius where damping begins to denominate the dispersion. Although $R_{0,ij}^6$ is in principle a parameter which could be further optimized, we instead derived its value in each case from the relationship

$$R_{0,ij} = \sqrt{\frac{C_{8,ij}}{C_{6,ij}}}. \quad (5)$$

where $C_{6,ij}$ is obtained directly from the input parameters, and $C_{8,ij}$ is derived from a recurrence relationship discussed in Ref. 8

$$C_{8,ij} = 3C_{6,ij}\sqrt{Q_i Q_j} \quad (6)$$

$$Q_i = s_{42}\sqrt{Z_i} \frac{\langle r^4 \rangle_i}{\langle r^6 \rangle_i}. \quad (7)$$

The empirical values of Q_i were taken directly from Grimme’s D3 paper.⁸ Application of

the Becke-Johnson rational damping scheme was found to improve the resulting rocksalt absolute lattice energies significantly when rocksalt-wurtzite energy differences and rock-salt lattice parameters were chosen as optimization targets. However, a large error in the absolute rocksalt lattice energies remains.

After much tinkering with Hancock’s algorithm, we found that when targetting relative rocksalt-wurtzite lattice energies and lattice parameters, the resulting optimized model parameters had the following characteristics:

- σ_{ij} model parameters are found to be clustered into small ranges. The parameter range is salt-dependent.
- ϵ_{ij} are found to be dispersed relatively widely.
- All model parameters yield very similar absolute rocksalt lattice energies, despite absolute lattice energy not being a target.

In other words, the location of the LJ well minimum (and corresponding repulsive wall) are much more important than the LJ well depth for the three properties of interest in this study. This begs the obvious question: is there another physical observable which we can use to narrow down the set of model parameters obtained so far? Or is the LJ well depth of little consequence for all physical properties of the lithium halides?

The first physical observable we decided to check is the melting point. Checking the melting point of all model parameter sets is computationally costly, so a selection of three optimized parameter sets for each LiX salt was chosen, representing a range of ϵ_{ij} values. To check the melting point of a model, there are several theoretical approaches which are discussed in detail in Ref. 9. So far, we have attempted to heat a perfect LiX crystal (consisting of 10000 atoms with periodic boundary conditions) slowly until melting is observed at T^+ , followed by a short equilibration period, and finally a slow cooling until nucleation occurs at T^- .

Using this technique, the melting point T_m can be approximated as

$$T_m \approx T^+ + T^- - \sqrt{T^+ T^-}. \quad (8)$$

II. FUTURE DIRECTION

At this point, there appears to be little need to perform further model parameter optimizations using Hancock’s machine learning algorithm. We have obtained a large set of viable LJ parameters for the pairwise CLJ LiX models. Performing machine learning on further targets such as the melting point becomes too computationally costly. A more traditional approach will be used moving forward.

A. Melting Points

High quality melting points should be obtained for a selection of models. It should be possible to obtain very precise measurements of melting points, using two phase solid-liquid coexistence simulations. A solid and liquid will only exist in equilibrium very near the true melting point of a given model.

We should therefore obtain high quality MD simulations for each LiX salt using three models of varying σ_{ij} . This should be performed in the NPT ensemble with $P = 1 \text{ atm}$ and compared with experimental melting points. The codes to initiate these calculations have already been written. One can obtain an initial geometry with both liquid and solid phases from preliminary melting/freezing cycle simulations already performed.

B. Dynamic Properties

In the future, other physical properties of the LiX system can potentially be studied in an attempt to improve narrow down our set of parameters. The following is a list of

experimental LiX properties from the literature.

- Thermal conductivities of molten LiF, LiCl, and LiBr have been measured as a function of temperature.¹⁰
- Thermal diffusivity of molten LiCl is available.¹¹
- Electrical conductivity of molten LiCl has been measured at 700 °C.¹²
- Electrical conductivity of molten LiBr and LiI are measured for a range of temperatures.¹³
- LiF, LiCl, LiBr, LiI compressibilities/bulk moduli are available under ambient conditions.^{14,15} However, we know the CLJ potential has the wrong shape of repulsive wall to model compressibilities correctly. We expect that compressibilities will be quite poor for all models.
- Molten LiCl, LiBr, and LiI radial distribution functions and neutron/X-ray scattering properties are available.¹⁶
- Molten LiX densities and surface tensions are available.¹⁷
- LiX solubilities in water are well known. The JC model¹ (CLJ type) gets solubilities off by an order of magnitude. It would be worth checking to see if any of these models improve upon the solubility. We would use the same water-ion parameters as in the JC model, only the Ion-Ion parameters would change.
- Molten LiCl and LiBr compressibilities are measured.¹⁸

It seems that the most relevant physical property to study is the LiX-water solubility. Exploring both melting points and solubilities, combined with everything else done so far, should be sufficient for the paper. For studies of melting points and solubilities, it's

important to quantify any errors that we can. These include errors due to finite size, finite time step, cutoffs, and random variability. MD simulations can be replicated to quantify random variability in the results.

REFERENCES

- ¹I. S. Joung and T. E. Cheatham, *J. Phys. Chem. B* **112**, 9020 (2008).
- ²M. P. Tosi and F. G. Fumi, *J. Phys. Chem. Solids* **25**, 45 (1964).
- ³F. Fumi and M. Tosi, *J. Phys. Chem. Solids* **25**, 31 (1964).
- ⁴R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civaleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1360 (2018), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1360>.
- ⁵H. Lorentz, *Ann. der Phys.* **248**, 127 (1881).
- ⁶D. Berthelot, *Compt. Rendus* **126**, 1703 (1898).
- ⁷A. D. Becke and E. R. Johnson, *J. Chem. Phys.* **127**, 124108 (2007), <https://doi.org/10.1063/1.2768530>.
- ⁸S. Grimme, S. Ehrlich, and L. Goerigk, *J. Comput. Chem.* **32**, 1456 (2011), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21759>.
- ⁹Y. Zhang and E. J. Maginn, *The J. chemical physics* **136**, 144116 (2012).
- ¹⁰M. Smirnov, V. Khokhlov, and E. Filatov, *Electrochimica acta* **32**, 1019 (1987).
- ¹¹Y. Nagasaka, N. Nakazawa, and A. Nagashima, *Int. journal thermophysics* **13**, 555 (1992).
- ¹²J. D. Edwards, C. S. Taylor, A. S. Russell, and L. F. Maranville, *J. The Electrochem. Soc.* **99**, 527 (1952).
- ¹³I. Yaffe and B. Van Artsdalen, *The J. Phys. Chem.* **60**, 1125 (1956).
- ¹⁴J. C. Slater, *Phys. Rev.* **23**, 488 (1924).
- ¹⁵S. Vaidya and G. Kennedy, *J. Phys. Chem. Solids* **32**, 951 (1971).
- ¹⁶H. Levy, P. Agron, M. Bredig, and M. Danford, *Ann. New York Acad. Sci.* **79**, 762 (1960).
- ¹⁷M. Smirnov and V. Stepanov, *Electrochimica Acta* **27**, 1551 (1982).

- ¹⁸J. O. Bockris and N. Richards, Proc. Royal Soc. London. Ser. A. Math. Phys. Sci. **241**, 44 (1957).
- ¹⁹L. Armijo, Pac. J. mathematics **16**, 1 (1966).
- ²⁰J. E. Dennis Jr and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Vol. 16 (Siam, 1996).
- ²¹E. Lindahl, B. Hess, M. Abraham, D. van der Spoel, and the GROMACS development team, “Gromacs reference manual version 2018,” www.gromacs.org (2018).

III. APPENDIX

A. Geometry Optimization of Lattice Parameters using Empirical Models

In order to generate fully optimized geometries with the TF, JC, and any DFT-D4 models, we designed and implemented a gradient descent optimization scheme written in MATLAB version R2019a. Initial input unit cell geometries were typically taken from the experiments where available, or from outputs of ab initio geometry optimization calculations with fixed ideal fractional coordinates. The first stage of the geometry optimization algorithm calculates three-point numerical derivatives with respect to all unit cell parameters (a , b , c , α , β , and γ). This is done via a triplet of single point energy calculations with GROMACS using settings tuned for high accuracy. Unit cell parameters which are constrained by the symmetry of the particular structure's space group were normally held fixed. Loosening of the symmetry constraint (i.e. placing the unit cells in space group P1) did not significantly change the resulting optimized structures or lattice energies. Numerical derivative step sizes h are initially chosen as

$$h_{\omega} = \max(10^{-4}\omega, 10^{-7}) \quad (9)$$

where ω is the differentiated parameter. After numerical differentiation of the lattice parameters, an appropriate step in the direction of the gradient is calculated. The step size is initialized as $\eta \vec{v}$ where

$$\vec{v} = -\vec{\nabla} E_L(\vec{x}). \quad (10)$$

η is a step-size parameter initially set to 1.0 but regularly updated. \vec{x} is a vector of the degrees of freedom (i.e. the differentiable lattice parameters). Each new step of the geometry

optimization is tested to satisfy the Armijo-Goldstein inequality^{19,20}

$$E(\vec{x} + \eta \vec{v}) \leq E(\vec{x}) - c\eta \|\vec{\nabla} E(\vec{x})\|^2, \quad (11)$$

where c is an arbitrary fixed parameter which we set to 0.5 for all calculations. If the above inequality is not satisfied at a particular optimization step, then η is repeatedly updated by

$$\eta_{new} = \lambda \eta \quad (12)$$

until the inequality is satisfied or 10 attempts are made. We found that $\lambda = 0.1$ is satisfactory for our geometry optimization calculations. If 10 attempts are made without satisfying the Armijo-Goldstein inequality, the gradient is discarded and a new gradient is calculated from numerical derivative step sizes 0.1 that of the current. Once the Armijo-Goldstein inequality is satisfied, $\vec{x} + \eta \vec{v}$ becomes the new central point with lattice energy $E(\vec{x} + \eta \vec{v})$. After each successful optimization step, η is scaled up by 1.5 for the next step.

During the second stage of our geometry optimization scheme, unit cell parameters are held fixed while the fractional coordinates of the ions are minimized by the built-in steepest-descent algorithm in GROMACS.²¹ This stage may be skipped, and instead the fractional coordinates of the asymmetric unit are minimized along with the unit cell parameters as previously described. However, the two stage process is faster. Within this second stage, minimization of fractional coordinates was considered converged once the maximum calculated force is smaller than $0.88 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$. After convergence of the second stage, the change between the initial energy (prior to stage 1) and final energy (following stage 2) was saved.

Three criteria are required for final convergence: the root-mean-squared of the (stage 1) gradient must be less than $0.59 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$ (the units change if cell angles or fractional coordinates are also optimized); the maximum absolute value of any component of

the gradient must be less than $0.88 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$ or $0.88 \text{ kJ mol}^{-1} \text{ }^{\circ}\text{ }^{-1}$; and the change in energy between optimization cycles must be less than $10^{-4} \text{ kJ mol}^{-1}$. If these three criteria were not met, the process was iterated at the new geometry. If the change in energy between optimization cycles was less than $10^{-4} \text{ kJ mol}^{-1}$, but the gradient criteria were not satisfied, then η was reset to 1.0 before the next iteration.

In practice, we found that optimization with respect only to symmetry-allowed lattice parameters was sufficient to reach near the unconstrained energy minimum, within 2 kJ mol^{-1} . By optimizing only these degrees of freedom, the geometry optimization convergence rate was accelerated greatly. By using a gradient descent algorithm, we ensured that the structure with the “nearest” local minimum of energy (without an intervening energy barrier) was always found.