

# Naturally parallelizable Monte Carlo algorithm for calculation of energy density of states

Jason D. Howard<sup>1</sup>

<sup>1</sup>*Materials Science Division, Argonne National Lab, Lemont, IL, 60439, USA*

(Dated: January 22, 2019)

In this work an algorithm is proposed that calculates the energy density of states, this algorithm combines random sets with Wang and Landau importance sampling and is naturally parallelizable. The algorithm is referred to as B<sub>L</sub>ENDER, which is an acronym for B<sub>L</sub>end Each New Density Each Round and an adjective for how it was created and functions. The algorithm was developed for the purpose of working towards the goal of using first principles simulations, such as density functional theory, to calculate the partition function of disordered sub lattices in crystal materials. In this work B<sub>L</sub>ENDER is tested with the 2d Ising model as an analogous system to the disordered cubic phase of the solid state lithium ion electrolyte Li<sub>2</sub>OHCl. This was done in the spirit of a “Fermi” problem to make a prediction of the wall time and core hours necessary for B<sub>L</sub>ENDER to evaluate the partition function of a 3×3×3 supercell of disordered cubic Li<sub>2</sub>OHCl. This work also demonstrates that, during the simulation to calculate the energy density of states, arithmetic averages of order parameters can be made for each energy level. With these averages and the energy density of states the ensemble average of the order parameter can be determined at any temperature.

For crystalline materials with disordered sub-lattices such as the Li-ion solid state electrolyte Li<sub>2</sub>OHCl[1–5] it is desirable to calculate from first principles methods(such as density functional theory[6]) the density of energy states  $G(E_j)$ . Here the energy density of states is meant to correspond to the energies of the distinct lattice configurations. With the energy density of states the partition function,

$$Z = \sum_i^{\Omega} e^{\frac{-e_i}{k_B T}} = \sum_j^{\Pi} G(E_j) e^{\frac{-E_j}{k_B T}}, \quad (1)$$

can be determined and from it many important thermodynamics properties such as the free energy, entropy, specific heat, and ensemble averages calculated. In Eq 1,  $\Omega$  corresponds to the number of possible configurations and energies in the set  $\{\Sigma_i, e_i\}_{\Omega}$ ,  $\Pi$  to number of possible distinct energies  $E_j$ ,  $k_B$  is Boltzman’s constant, and  $T$  is the temperature. One method to solve this problem could be temperature dependent simulations involving the Metropolis algorithm and histogram re-weighting techniques[7, 8]. Another algorithm called the Wang and Landau algorithm[9] has been developed which is temperature independent. An issue with these algorithms in use with first principles methods such as density functional theory is the large number of iterations needed which would require a prohibitively long wall time at the current performance power of computers. In this paper a method is proposed that combines the use of random sets along with the importance sampling method of the Wang and Landau algorithm that is meant to work towards the goal of highly parallel importance sampling algorithms that mesh well with high performance computing architectures. The algorithm developed in this work is referred to as the B<sub>L</sub>ENDER (B<sub>L</sub>end Each New Density Each Round) algorithm. The name B<sub>L</sub>ENDER functions as an

adjective as well as an acronym. This comes from how it blends the ideas of a random set and the Wang and Landau method, and also due to the nature of the algorithm iteratively blending histograms to produce a converged density of states. The Wang and Landau method does have parallel versions, including restricting random walkers to specific energy ranges or allowing the walkers to explore the entire space while periodically communicating with each other [10–12]. The B<sub>L</sub>ENDER algorithm is natural to parallelize as it is based on a set of random walkers that each can explore the entire energy range.

In a previous study[13] the notion of the master set  $\{\Sigma_i, e_i\}_{\Omega}$  of the  $\Omega$  total configurations and energies was put forth. The master set refers to the actual configurations of the defined system. In practice the calculation of the energy density of states  $G(E_j)$  may be possible through randomly sampling the configuration space. In a previous work it was shown that a properly scaled histogram of a uniformly sampled random set  $\{\Sigma_s, e_s\}_{\mathcal{S}}$  converges to the exact density of states  $G(E_j)$  as the number of samples  $\mathcal{S}$  goes to infinity. The problem with this method is that if  $\Omega$  is large, which it is for many problems, then the computational effort to achieve convergence is not feasible. This work tackles this issue and produces an algorithm that is highly parallel in terms of the calculation of the energies but also incorporates importance sampling such as in the Wang and Landau method.

The B<sub>L</sub>ENDER algorithm proposed in this work is given as follows. It is noted that the following algorithm is in terms of producing a relative density of states  $H(E_j)^I$ , where  $I$  is the iteration number.

1.  $H(E_j)^I, \{\Sigma_s, e_s\}_S^I$
2.  $\{\Sigma_s, e_s\}_S^I \rightarrow \{\Sigma'_s, e'_s\}_S^I$
3.  $H(E_j)^{II} = H(E_j)^I + \mathcal{H}(E_j, \{e'_s\}_S^I)$
4.  $\Sigma_s^I, e_s^I \rightarrow \Sigma_s^{I+1}, e_s^{I+1} \quad P = \min(1, H(e_s)^{II}/H(e'_s)^{II})$   
 $\text{else } \Sigma_s^I, e_s^I \rightarrow \Sigma_s^{I+1}, e_s^{I+1}$
5.  $H(E_j)^{I+1} = H(E_j)^I + C_o \mathcal{H}(E_j, \{e_s\}_S^{I+1}) \frac{H(E_j)^{II}}{\sum_j H(E_j)^{II}}$
6.  $N = \sum_j H(E_j)^{I+1}$

*if*  $H(E_j)^{I+1} \frac{\Omega}{N} < 1$ ,  $H(E_j)^{I+1} = \frac{N}{\Omega}$

(2)

Where  $H(E_j)^0 \equiv \mathcal{H}(E_j, \{e_s\}_S^0)$  with  $\mathcal{H}(E_j, \{e_s\}_S)$  being a histogram function that counts the number of energies  $E_j$  in the set  $\{e_s\}_S$ . In this work  $\{\Sigma_s, e_s\}_S^0$  is a randomly(uniformly) drawn set from the configuration space  $\{\Sigma_i, e_i\}_\Omega$ . In the second step a random change is applied to each element of the sampled set  $\{\Sigma_s, e_s\}_S^I$  to produced a “perturbed” set  $\{\Sigma'_s, e'_s\}_S^I$ , for the Ising model this could be randomly flipping a spin. In the third step a histogram of the “perturbed” set is added to the current estimate of the density of states  $H(E_j)^I$  to produce an intermediary density of states  $H(E_j)^{II}$ . In the fourth step a random number is drawn between zero and one for every sampled configuration, if this number is less then the ratio of the intermediary density of states  $H(e_s)^{II}/H(e'_s)^{II}$  then the perturbed configuration and energy  $\Sigma_s^I, e_s^I$  goes to  $\Sigma_s^{I+1}, e_s^{I+1}$ , else the unperturbed configuration and energy  $\Sigma_s^I, e_s^I$  goes to  $\Sigma_s^{I+1}, e_s^{I+1}$ . In the fifth step a histogram of the updated  $\{e_s\}_S^{I+1}$  energies is made and added (blended) in to the current density of states  $H(E_j)^I$  by multiplying by a constant  $C_o$ (which affects the convergence properties) and the relative probability of each energy  $E_j$  in the intermediary density of states  $H(E_j)^{II}$ . In this work it was found  $C_o = \Omega$  was computationally efficient. The sixth step corrects for a systematic error of a energy having a density of state less than one, in principle for a discrete system a given energy should have a density of state of a least one. After the algorithm is deemed to be complete it is necessary to re-normalize the iterated relative density of states  $H(E_j)^f$  at the final iteration  $I = f$  as follows,

1.  $A = \sum_j H(E_j)^f$
2.  $G(E_j) \approx H(E_j)^f \frac{\Omega}{A}$ ,

(3)

to produce the properly normalized estimated value of  $G(E_j)$ .

In this work the algorithm discussed is tested using the 2d square zero field Ising model with lattice dimen-

sion of even number[14–16]. The configurations  $\Sigma_i$  and energies  $e_i$  of the 2d Ising model are inherently defined by the lattice site spin variables and coupling constant  $J$ . In a previous study[13] uniform sampling was used with first principles simulations to approximate the partition function for a  $2 \times 2 \times 2$  supercell of disordered cubic  $\text{Li}_2\text{OHCl}$ . In going to just a  $3 \times 3 \times 3$  supercell the value of  $\Omega$  would jump from  $\sim 1e7$  to  $\sim 1e27$ . So computing the partition function for the  $3 \times 3 \times 3$  system of disordered cubic  $\text{Li}_2\text{OHCl}$  is completely intractable from uniform sampling. In this work a  $10 \times 10$  2d Ising model with  $\Omega = 1.3e30$  is used as an analogous system to predict the computational effort needed for the  $\text{B}_L\text{ENDER}$  algorithm to compute the partition function of a  $3 \times 3 \times 3$  supercell of disordered cubic  $\text{Li}_2\text{OHCl}$ . The first test is a test to show the convergence of the algorithm in terms of the number of samples  $\mathcal{S}$  and the number of iterations of the algorithm  $I$ . To test the accuracy of the simulations the results will be compared to the exact result solved by Beale [17]. The accuracy of the simulation will be determined by the average of errors given as,

$$\begin{aligned} \mathcal{E}(I, o) &= \langle |\epsilon(E_j, I, o)| \rangle_j \\ &= \frac{1}{\Pi} \sum_{j=1}^{\Pi} \frac{|\ln(G_{ex}(E_j)) - \ln(G_{cal}(E_j, I, o))|}{|\ln(G_{ex}(E_j))|} \end{aligned} \quad (4)$$

Where  $G_{ex}(E_j)$  is the exact density of states,  $G_{cal}(E_j, I, o)$  is the calculated density of states at iteration number  $I$  from initial conditions and trajectory  $o$ , and  $|\epsilon(E_j, I, o)|$  is the absolute value of the fractional error for a specific energy level. To test the algorithm on the  $10 \times 10$  Ising model, 144 individual calculations were performed at  $\mathcal{S} = 1, 10, 100, 1e3, 1e4$ , and  $1e5$ . The number of iterations to find all the energies and the number of iterations to reach  $\mathcal{E} = 1\%$  were recorded and averaged for each value of  $\mathcal{S}$ . These results are shown in Fig 1 (a) and (b). The plots show that the decrease in the number of iterations is not linear in including more samples  $\mathcal{S}$  in the set, in the sense that there is an exponential like decay in the reduction in number of iterations for an increase in samples  $\mathcal{S}$ . Something of interest to understand is some of the basic convergence properties of the error of the simulation. To test the convergence of the algorithm in terms of iterations, 36 simulations were done for  $\mathcal{S} = 100$  for  $1e9$  iterations. The average of the errors  $\langle \mathcal{E}(I, o) \rangle_o$  is shown in Fig 2 as a  $\log_{10} \log_{10}$  plot. The results show that the error appears to decrease approximately one order of magnitude for every 1.7 orders of magnitude increase in iterations. To further understand the errors, relative errors plots of  $\langle |\epsilon(E_j, I, o)| \rangle_o$  for the 144 simulations with  $\mathcal{S} = 100$  taken to  $\mathcal{E} = 1\%$  and the 36 simulations with  $\mathcal{S} = 100$  taken to  $I = 1e9$ , are shown in Fig 3 in black and red respectively. The results show that the relative error is larger at the less probable energy levels. The general shape of the errors seems to remain unchanged and the whole curve simply

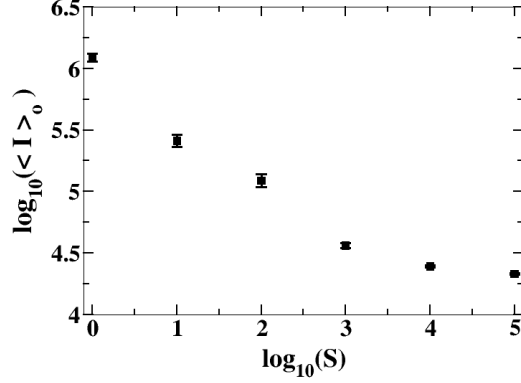
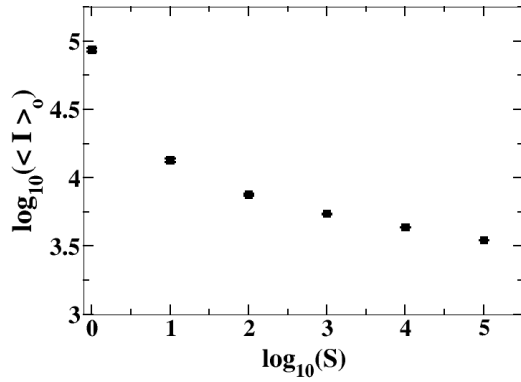
(a) average iterations to reach  $\mathcal{E} = 1\%$  vs  $\mathcal{S}$ (b) average iterations to find all energies vs  $\mathcal{S}$ 

FIG. 1. A  $\log_{10} \log_{10}$  plot of the average iterations  $I$  to reach  $\mathcal{E}(I, o) = 1\%$  vs the set size  $\mathcal{S}$  (a) and the average number of iterations  $I$  to find all the energies vs the set size  $\mathcal{S}$  (b). Error bars represent a linear propagation of the standard deviation of mean for  $\langle I \rangle_o$  in  $\log_{10}(\langle I \rangle_o)$ .

shifts when comparing the less converged results in the black to the more converged results red.

Now from Fig 1 (a) we could take  $\mathcal{S} = 100$  as an approximate optimal value to attempt a calculation of the density of states of disordered cubic  $\text{Li}_2\text{OHCl}$ . In willing to accept  $\mathcal{E} = 10\%$  there is an approximate  $10\times$  reduction in the number of iterations. So for an accuracy of  $\mathcal{E} = 10\%$  that gives an approximate number of iterations  $\sim 1e4$ . Using the VASP code[18–21] with the projector augmented wave formalism[22, 23] and the GGA-PBE exchange and correlation functional[24, 25], the calculation of a structural relaxation to obtain the energy for a given lattice configuration is on the order of 2 hours for a 36 core broadwell node. So to compute the partition function of the disordered cubic lattice of a  $3 \times 3 \times 3$  super-cell model of  $\text{Li}_2\text{OHCl}$  with 100, 36-processor broadwell nodes, at the level of static lattice internal energies (no phonon free energies) calculated with GGA-PBE would require approximately 70 million core hours and a wall

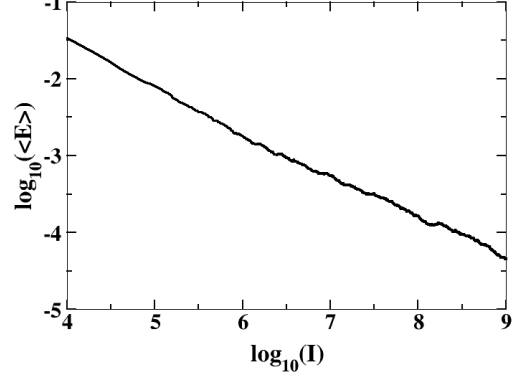


FIG. 2.  $\log_{10} \log_{10}$  plot of  $\langle \mathcal{E}(I, o) \rangle_o$  vs  $I$ . The iterations start at  $1e4$  because that is approximately the number of iterations when all the energies are found for a given simulation.

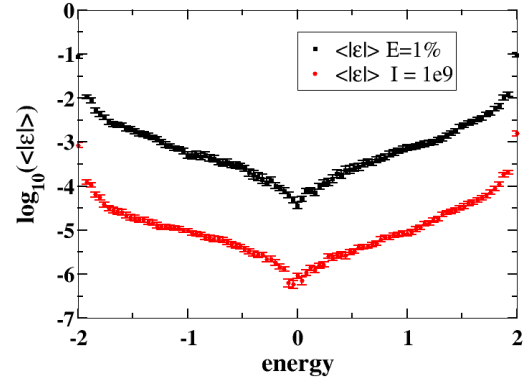


FIG. 3. plots of the  $\log_{10}(\langle |\epsilon(E_j, I, o)| \rangle_o)$  vs energy per lattice site for 144 simulations taken to  $\mathcal{E} = 1\%$  shown in black and 36 simulations taken to  $I = 1e9$  shown in red. Both simulations use  $\mathcal{S} = 100$ . Error bars are a linear propagation of the standard deviation of the mean for  $\langle |\epsilon(E_j, I, o)| \rangle_o$  in the equation  $\log_{10}(\langle |\epsilon(E_j, I, o)| \rangle_o)$ .

time of 800 days. So at the current moment it is still too large of a computational effort to make utilizing this algorithm on a  $3 \times 3 \times 3$  model of the  $\text{Li}_2\text{OHCl}$  system feasible but if Moore’s law continues within a decade it may be a viable option. It must be stated that this estimation on the computational effort required for the  $\text{Li}_2\text{OHCl}$  system is presented in the spirit of a “Fermi” problem and is only meant for a coarse estimate and that a more detailed analogous model would be required to improve the prediction. A more detailed analogous model might include using a 3d Ising model with fixed concentration of up and down spins and including longer range interactions. The benefit of the square 2d Ising model in this work is the availability of exact solutions to test the proficiency of the algorithm. Aside from the difficulties in

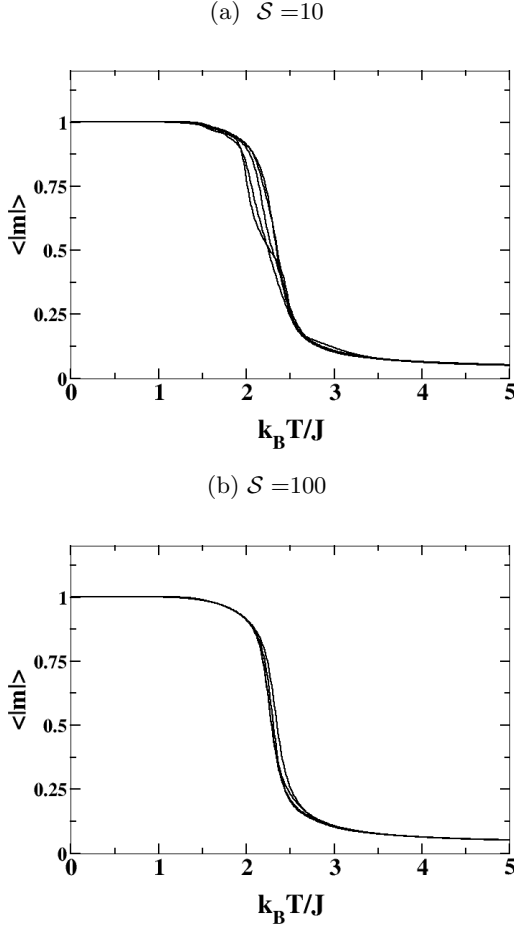


FIG. 4. The ensemble averaged absolute value of the magnetization per lattice site calculated for a  $26 \times 26$  ferromagnetic Ising model using  $S=10$  and  $S=100$  until  $\mathcal{E}(I, o) = 1\%$  shown in (a) and (b) respectively. Each plot displays five separate calculations at each value of  $S$ .

large supercell first principles simulations the algorithm could be used with first principles on smaller systems, with model Hamiltonians, or systems defined by classical potentials.

Once the energy density of states is obtained many important thermodynamic properties can be determined but it may be considered a draw back that ensemble averages of a general order parameter  $a_i \equiv a(\Sigma_i)$  are not accessible. It is shown in this work that this is possible to achieve by considering the following equation,

$$\langle a \rangle = \sum_{i=1}^{\Omega} a_i \frac{e^{-\frac{e_i}{K_b T}}}{Z} = \sum_{j=1}^{\Pi} \langle a_i \rangle_j \frac{G(E_j) e^{-\frac{E_j}{K_b T}}}{Z} . \quad (5)$$

Where  $\langle a_i \rangle_j$  is the arithmetic average of  $a_i$  over all configurations with energy  $E_j$ . This equation (Eq 5) follows from that given the ordered set of configurations

$\{\Sigma_i\}_{\Omega}$  with energies  $e_1 \leq e_i \leq e_{\Omega}$  and corresponding order parameters  $a_1 \dots a_i \dots a_{\Omega}$ , the order parameters can be partitioned into groups based on like energy. With this notion and defining  $K(E_j) \equiv \sum_{l=0}^{j-1} G(E_l)$  (note that for convenience:  $G(E_0) \equiv 0$ ), Eq 5 can be written as,

$$\langle a \rangle = \sum_{j=1}^{\Pi} \sum_{k=1}^{G(E_j)} a_{K+k} \frac{e^{-\frac{E_j}{K_b T}}}{Z} = \sum_{j=1}^{\Pi} \frac{e^{-\frac{E_j}{K_b T}}}{Z} \sum_{k=1}^{G(E_j)} a_{K+k} . \quad (6)$$

From Eq 6 it is straight forward to rewrite the sum over  $a_{K+k}$  in Eq 6 as,

$$\sum_{k=1}^{G(E_j)} a_{K+k} = G(E_j) \langle a_i \rangle_j , \quad (7)$$

that being the number of states with energy  $E_j$  times the average value of the order parameter  $a_i$  over those states. If an average of the values visited during the simulation is calculated then it is possible to calculate ensemble averaged order parameters over the entire temperature range. To test this the B<sub>L</sub>ENDER algorithm was applied to a  $26 \times 26$  ferromagnetic Ising model with zero field with  $S = 10$  and  $100$  to calculate the ensemble average of the absolute value of the magnetization per lattice site  $\langle |m| \rangle$ . In Fig 4 (a) and (b) the results of five separate simulations are shown for  $S = 10$  and  $100$ . The results show that the known phase transition behavior of the ferromagnetic Ising model is well represented which indicates that this is a viable method for calculating ensemble averaged order parameters and that a larger value of  $S$  appears to reduce the variability between separate calculations. This calculation is also evidence of B<sub>L</sub>ENDER being applicable to larger systems, although this is still small compared to the  $256 \times 256$  model tested by Wang and Landau. The Wang and Landau algorithm is in fact formulated in a way that updates the density of states by a multiplicative factor which allows for it to be formulated as a sum of logs, this is how they were able to use their algorithm for a system with such a large total number of states. For the B<sub>L</sub>ENDER algorithm there is a downside that it is formulated by updating the density of states through addition and if using a large constant  $C_o$  this can mean adding large numbers that over flow the real numbers for a typical programming language. In principle this can be overcome by using an arbitrary precision data type.

In summary the presented algorithm in this work has shown its effectiveness in accurately determining the density of states for the 2d square zero field Ising model. The simulations suggest that within a decade using the B<sub>L</sub>ENDER algorithm it will be feasible on a state of the art super computer to calculate the partition function of

a  $3 \times 3 \times 3$  135 atom model of disordered cubic  $\text{Li}_2\text{OHCl}$  at the static lattice internal energy level using density functional to a high level of accuracy. At the present time the algorithm should be suitable for first principles methods with smaller system sizes, with model lattice Hamiltonians, or with models defined through classical potentials. The algorithm shows promise for first principles methods in particular because the calculation of the energies can be done as individual job submissions to compute nodes managed by a script running on a head node. This work also demonstrated that if during the simulation to determine the energy density of states, arithmetic averages of order parameters are made for each energy level, then calculation of order parameters at any temperature is feasible.

I would like to thank the Laboratory Computing Resource Center (LCRC) faculty of Argonne National Lab for their support and maintenance of the computing resources that made this project possible.

- 
- [1] Z. D. Hood, H. Wang, A. S. Pandian, J. K. Keum, and C. Liang.  $\text{Li}_2\text{OHCl}$  crystalline electrolyte for stable metallic lithium anodes. *Journal of the American Chemical Society*, 138(6):1768–1771, 2016.
  - [2] Y. Li, W. Zhou, S. Xin, S. Li, X. Zhu, J. and L. Z. Cui, Q. Jia, J. Zhou, Y. Zhao, and J. B. Goodenough. Fluorine-doped antiperovskite electrolyte for all-solid-state lithium-ion batteries. *Angewandte Chemie International Edition*, 55(34):9965–9968, 2016.
  - [3] G. Schwing, A. Hönnerscheid, L. v. Wüllen, and M. Jansen. High lithium ionic conductivity in the lithium halide hydrates  $\text{Li}_{3-n}(\text{OH}_n)\text{Cl}$  ( $0.83 \leq n \leq 2$ ) and  $\text{Li}_{3-n}(\text{OH}_n)\text{Br}$  ( $1 \leq n \leq 2$ ) at ambient temperatures. *ChemPhysChem*, 4(4):343–348, 2003.
  - [4] J. D. Howard, Z D. Hood, and N. A. W. Holzwarth. Fundamental aspects of the structural and electrolyte properties of  $\text{Li}_2\text{OHCl}$  from simulations and experiment. *Phys. Rev. Materials*, 1:075406, Dec 2017.
  - [5] A.-Y. Song, Y. Xiao, K. Turcheniuk, P. Upadhyay, A. Ramanujapuram, J. Benson, A. Magasinski, M. Olguin, L. Meda, O. Borodin, and G. Yushin. Protons enhance conductivities in lithium halide hydroxide/lithium oxyhalide solid electrolytes by forming rotating hydroxy groups. *Advanced Energy Materials*, page 1700971, 2018.
  - [6] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133–A1138, 1965.
  - [7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
  - [8] D. P. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Physics*. Cambridge University Press, fourth edition edition, 2015.
  - [9] F. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86:2050–2053, Mar 2001.
  - [10] J. Yin and D. P. Landau. Massively parallel wanglandau sampling on multiple gpus. *Computer Physics Communications*, 183(8):1568 – 1573, 2012.
  - [11] L. Zhan. A parallel implementation of the wanglandau algorithm. *Computer Physics Communications*, 179(5):339 – 344, 2008.
  - [12] T. Vogel, Y. W. Li, T. Wüst, and D. Landau. Generic, hierarchical framework for massively parallel wang-landau sampling. 110:210603, 05 2013.
  - [13] J. D. Howard and N. A. W. Holzwarth. First-principles estimation of partition functions representing disordered lattices such as the cubic phases of  $\text{Li}_2\text{OHCl}$  and  $\text{Li}_2\text{OHBr}$ . *Phys. Rev. B*, 99:014109, Jan 2019.
  - [14] R. J. Baxter. *Exactly Solved Models in Statistical Mechanics*. Elsevier, 2016. ISBN 978-0-12-083180-7.
  - [15] L. Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65:117–149, 1944.
  - [16] B. M. McCoy and T. T. Wu. *The Two-Dimensional Ising Model*. Harvard University Press, 1973.
  - [17] P. D. Beale. Exact distribution of energies in the two-dimensional ising model. *Phys. Rev. Lett.*, 76:78–81, Jan 1996.
  - [18] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558–561, Jan 1993.
  - [19] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B*, 49:14251–14269, May 1994.
  - [20] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15 – 50, 1996.
  - [21] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct 1996.
  - [22] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, Dec 1994.
  - [23] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, Jan 1999.
  - [24] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.
  - [25] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple [phys. rev. lett. 77, 3865 (1996)]. *Phys. Rev. Lett.*, 78:1396–1396, Feb 1997.