First principles calculation of the configurational energy density of states for LLTO with new Wang and Landau algorithm variant

Jason D. Howard[1]

[1]*Materials Science Division, Argonne National Lab, Lemont, IL, 60439, USA*[*]
(Dated: October 19, 2020)

In this work a variant of the Wang and Landau algorithm for calculation of the configurational energy density of states is proposed. The algorithm was developed for the purpose of using first principles simulations, such as density functional theory, to calculate the partition function of disordered sublattices in crystal materials. The expensive calculations of first principles methods make a parallel algorithm necessary for a practical computation of the configurational energy density of states within a supercell approximation of a solid state material. The algorithm developed in this work is tested with the 2d Ising model to bench mark the algorithm and to help provide insight for implementation to a materials science application. Tests with the 2d Ising model revealed that the algorithm has good performance compared to the original Wang and Landau algorithm and the 1/t algorithm, in particular the short iteration performance. A proof of convergence is presented within an adiabatic assumption, the anlaysis is able to correctly predict time dependant properties of the algorithm. The algorithm was then applied to the lithium and lanthanum sublattice of the solid state lithium ion conductor $Li_{0.5}La_{0.5}TiO_3$. This was done to help understand the disordered nature of the lithium and lanthanum. The results find, overall, that the algorithm performs very well for the 2d Ising model and that the results for $Li_{0.5}La_{0.5}TiO_3$ are consistent with experiment while providing additional insight into the lithium and lanthanum ordering in the material. The primary result is that the lithium and lanthanum become more mixed between layers along the c-axis for increasing temperature.

I. INTRODUCTION

For crystalline materials with disordered sublattices such as the lithium ion solid state electrolyte LLTO it is desirable to calculate from first principles methods(such as density functional theory[?]) the configurational energy density states $G(E_j)$, which will simply be referred to as the density of states. Here the density of states refers to the energies of the distinct lattice configurations. With the density of states the partition function,

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

can be determined and from it many important thermodynamic properties such as the free energy, entropy, specific heat, and ensemble averages calculated. In Eq. (??), Ω corresponds to the total number of possible configurations(Σ_i) and energies(e_i) referred to by the set $\{\Sigma_i, e_i\}_\Omega$, Π is the number of possible distinct energies E_j that appear in $\{\Sigma_i, e_i\}_\Omega$, k_B is Boltzmann's constant, and T is the temperature. If Ω is small enough one of the simplest ways to estimate $G_r(E_j)$ could be direct random sampling of the configuration space[?], which does have the advantage of being compeletly parrallel. For most problems Ω will be too large for direct random sampling of the configuration space to be practical, making an importance sampling algorithm necessary. A very well

known importance sampling method to solve this problem could be temperature dependent simulations involving the Metropolis algorithm and sampling with probability proportional to $exp(\frac{-e_i}{k_B T})$ along with histogram re-weighting techniques[? ? ?]. Another more advanced method is the multi-canonical method proposed by Berg et al. [? ?]. A variant of multicanonical sampling that samples the density of states directly known as entropic sampling developed by Lee [?] could also be used. The multicanonical method requires a careful choice of simulation paramaters while the entropic sampling requires a good estimate of the density of states to be effective. Another algorithm called the Wang and Landau algorithm [? ?] has been developed which is temperature independent, is based on a random walk in energy space with probability inversely proportional to the current estimate of the density of states, and builds up the density of states as the algorithm progresses. An issue with these algorithms(if using a single walker) 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 an algorithm is proposed that combines the use of random sets along with the importance sampling method of the Wang and Landau algorithm. The algorithm also used the principle of the Wang and Landau algorithm to build up an estimate of the density of states as the algorithm progresses. The proposed algorithm is meant to work towards the goal of a highly parallel importance sampling algorithm that directly calculates the density of states, meshes well with mid level high performance computing architectures(such as Argonne's BEBOP), and has a min-

---

[*] jdhoward@anl.gov

imum of paramaters for implementation. The algorithm developed in this work is referred to as the B$_L$ENDER (B$_L$end Each New Density Each Round) algorithm. The Wang and Landau method does have parallel versions, including restricting random walkers to specific energy ranges, allowing the walkers to explore the entire space while periodically communicating with each other, and methods based on a replica exchange frame work [? ? ? ?]. The B$_L$ENDER algorithm is characterized by allowing the walkers to explore the entire energy range and communication with each other through an update to the density of states at each iteration. In principle many of the different forms of Wang and Landau sampling currently used are based around the concept of sampling until a flat histogram of the visited energies is reached followed by a reduction in a modification factor to the density of states and that the calculated density of states is multiplied by this factor every time an energy level is visited. One issue with these types of Wang and Landau simulations is that, being based on a flat histogram of the visited energies, the energy range must be specified *apriori*. Another issue is that the original Wang and Landau formulation for the reduction in the modification factor to the density of states has been shown be non-convergent[? ? ? ?]. There have been advancements made in understanding how to reduce the modification factor by Belardinelli et al. [?] whom developed the 1/t algorithm which is proven to be convergent, this result was verified by the work of Zhou et al. [?]. An issue with the 1/t algorithm as presented by Belardinelli et al. is that it requires a non-trivial preconditioning using what is referred to as the N-fold way to be most effective. The ordinary 1/t algorithm relies on the original Wang and Landau algorithm for this preconditioning. The N-fold way requires a careful analysis of the method to perturb the systems affect on the energy which for local models like the Ising model is possible but for first principles simulations this is not feasible. The novel aspects of the B$_L$ENDER algorithm include, a continuous adaptation of the modification factor to the density of states using the current sum of the density of states as a regulator, using the number of configurations as a parameter in the modification factor, and using a histogram of the currently visited energies as a parameter in the modification factor. The algorithm in this work is believed to be convergent based on a mathematical analysis within an adiabatic assumption. The algorithm is also natural to parallelize as it is based on a set of random walkers. The algorithm was developed for ease of use in the application to disordered sublattices of crystal systems.

In this work the formulated algorithm is benched marked with the 2d Ising model as a standard means of testing performance. The tests allow for a comparison to exact results and to previous benchmarks of other algorithms. The tests with the 2d Ising model also allow for insight in how to implement the algorithm to a materials science problem. The main goal in this work was to calculate with first principles methods the density of states of the lithium ion conductor $Li_{0.5}La_{0.5}TiO_3$. There have been reports of the Wang and Landau algorithm used with linear scaling density functional theory methods to calculate magnetic properties of materials and order to disorder properties of alloys[? ?]. There has also been a report of first principles calculations with replica exchange canonical Metropolis sampling for investigating ion disorder in solids by Kasamatsu et al. [?]. $Li_{0.5}La_{0.5}TiO_3$ is part of a family of possible stoichiometries $Li_{3x}La_{2/3-x}TiO_3$ of interest as solid state lithium ion conductors[? ? ? ? ? ?]. For all of the possible stoichiometries there is a tendency towards ordering of the lithium and lanthanum into lithium rich layers and lanthanum rich layers. The primary calculation of this work is that of the temperature dependent order parameter related to the lanthanum rich layer in $Li_{0.5}La_{0.5}TiO_3$. This calculation both serves to benchmark the application of the algorithm to a materials science problem with experimental knowns and to provide further insight into the physics of the material.

The rest of the article is organized as follows; section ?? explains the B$_L$ENDER algorithm, section ?? covers the testing of the algorithm with the 2d Ising model including a comparison with the original Wang and Landau algorithm and the 1/t algorithm, section ?? covers an adiabatic analysis of the algorithm's convergence and the time dependence of the modification factor, section ?? covers the application of the algorithm to LLTO along with the computational details of the first principles methods, and section ?? is the conclusions.

In this work the software package XMGRACE[?] was used in the generation of plots and VESTA[?] for the generation of structural images of LLTO. The software package MATLAB was utilized in generating movies provided in the supplemental materials.

II. ALGORITHM

The B$_L$ENDER algorithm proposed in this work is given as follows for $\mathcal{S}$ random walkers. It is noted that the following algorithm is in terms of producing a relative density of states $G_r(E_j)^I$, where I is the iteration number.

1. $G_r(E_j)^I, \ \{\Sigma_s, e_s\}_{\mathcal{S}}^I$

2. $\{\Sigma_s, e_s\}_{\mathcal{S}}^I \rightarrow \{\Sigma_s^{'}, e_s^{'}\}_{\mathcal{S}}^I$

3. $\Sigma_s^{'I}, e_s^{'I} \rightarrow \Sigma_s^{I+1}, e_s^{I+1} \ \ P = \ min[1, G_r(e_s)^I/G_r(e_s^{'})^I]$
   $else \ \Sigma_s^I, e_s^I \rightarrow \Sigma_s^{I+1}, e_s^{I+1}$

4. $G_r(E_j)^{I+1} =$

$$G_r(E_j)^I + \frac{C_o \mathcal{H}^I}{(A^I)^{\frac{1}{N}}} G_r(E_j)^I =$$

$$G_r(E_j)^I (1 + \frac{C_o \mathcal{H}^I}{(A^I)^{\frac{1}{N}}})$$

$$\text{(2)}$$

Where $\mathcal{H}^I \equiv \mathcal{H}(E_j, \{e_s\}_{\mathcal{S}}^{I+1})$, that being an histogram function that counts the number of the currently visited energies E_j in the set $\{e_s\}_{\mathcal{S}}^{I+1}$. The "instantaneous" histogram $\mathcal{H}^I$ is related to the total histogram of visited energies $H(E_j)^I$ by $H(E_j)^I = \sum_{i=0}^{I} \mathcal{H}^i$. The value $A^I \equiv \sum_j G_r(E_j)^I$. In this work $\{\Sigma_s, e_s\}_{\mathcal{S}}^0$ is a randomly drawn set of $\mathcal{S}$ walkers from the configuration space $\{\Sigma_i, e_i\}_{\Omega}$. Also in this work the initial guess of the density of states was set to $G_r(E_j)^0 = 1 + \frac{C_o}{\mathcal{S}} \mathcal{H}(E_j, e_s{}^0)$, but in principle there could be other ways to intitialize the initial guess to the density of states. In the second step a random change is applied to each element of the sampled set $\{\Sigma_s, e_s\}_{\mathcal{S}}^I$ to produced a "perturbed" set $\{\Sigma_s', e_s'\}_{\mathcal{S}}^I$, for the Ising model this could be randomly flipping a spin. In the third step a random number is drawn between zero and one for every sampled configuration, if this number is less then the ratio of the current density of states of the unperturbed to perturbed energies $G_r(e_s)^I/G_r(e_s')^I$ 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 Σ_s^I, e_s^I goes to $\Sigma_s^{I+1}, e_s^{I+1}$. This step (third) is derived from the Wang and Landau method of sampling with probability inversely proportional to the density of states. In the fourth step a histogram of the updated $\{e_s\}_{\mathcal{S}}^{I+1}$ energies is made and added (blended) into the current density of states $G_r(E_j)^I$ by multiplying by a constant C_o(which affects the convergence properties) and $G_r(E_j)^I$ divided by the sum of the density of states to the $1/N$ power. The $1/N$ power is introduced as tuning parameter to affect the convergence properties and was discovered through empirical testing with the 2d Ising model. A better understanding of the algorithm and evidence for its convergence is found in Section **??** with a mathematical analysis within an adiabatic assumption. The mathematical analysis in Section **??** provides evidence that $1/N$ and C_o are simply computational parameters that will effect the initial dynamics of the simulation but not whether the algorithm will converge in the long iteration limit. The fourth step is also shown in terms of multiplication which is discussed later. In this work it was found $C_o = \Omega^{\frac{1}{N}}$ was computationally efficient. After the algorithm is deemed to be complete it is necessary to renormalize the iterated relative density of states $G_r(E_j)^f$ at the final iteration $I = f$ as follows,

$$
\begin{aligned}
&1.\ A^f = \sum_j G_r(E_j)^f \\
&2.\ G_{calc}(E_j) = G_r(E_j)^f \frac{\Omega}{A^f} ,
\end{aligned}
\tag{3}
$$

to produce the properly normalized estimated value of $G(E_j)$. In principle, $G_r(E_j)$ can also be renormalized based on information of the number of configurations in a given bin. For example if the ground state is known to have a given degeneracy then the entire density of states can be normalized such that the ground state bin has the correct degeneracy.

An important discussion point of this algorithm (Eq. **??**) is the update of the relative density of states(step four) being presented as addition and multiplication. In typical Wang and Landau sampling the update of the density of states is preformed by multiplication of the density of states by a factor greater than one every time an energy level is visited combined with a periodic reduction of the multiplication factor towards one when a histogram of visited energies reaches a predetermined flatness criteria, the histogram of the visited energies is then reset to zero. In the multiplication form of step four of this algorithm (Eq. **??**) it is seen that the dependence on one over the sum of the density of states serves to naturally reduce the multiplication factor towards one as the simulation progresses. The multiplication form is also useful when Ω is large and the sum of the density of states is larger than a typical floating point number. In this case the log of the density of states can be stored and the update performed through addition of logs. Taking $G_r^M \equiv max[G_r(E_j)]$ the log of A^I can be written as,

$$
\begin{aligned}
\log[A^I] &= \log[G_r^M \frac{A^I}{G_r^M}] = \\
&\log[G_r^M] + \log[\sum_j e^{\ln[G_r(E_j)] - \ln[G_r^M]}] .
\end{aligned}
\tag{4}
$$

With G_r^{LS} from Eq. **??** the log update form of step four of the algorithm (Eq. **??**) can be written as the following,

$$
\begin{aligned}
&\log[G_r(E_j)^I (1 + \frac{C_o \mathcal{H}^I}{(A^I)^{\frac{1}{N}}})] = \\
&\log[G_r(E_j)^I] + \log[1 + \mathcal{H}^I e^{\ln[C_o] - \frac{1}{N} \ln[A^I]}] .
\end{aligned}
\tag{5}
$$

In this form the algorithm can be implemented even when Ω is large. To implement the ratio of the density of states in step two of the algorithm,

$$
e^{\ln[G_r(e_s)^I] - \ln[G_r(e_s')^I]} ,
\tag{6}
$$

can be used.

III. BENCH MARK WITH 2D ISING MODEL

In this work the algorithm discussed is tested using the 2d square zero field Ising model with lattice dimension of even number[**? ? ?**] and periodic boundary conditions. The configurations Σ_i and energies e_i of the 2d Ising model are inherently defined by the lattice site spin variables $\sigma_{k,l}$ which take the values ± 1 and coupling constant J. Explicitly the energy e_i for a given configuration Σ_i of a n×n Ising lattice is given by,

$$
e_i = -J \sum_{k,l=1}^{n} \sigma_{k,l}^i (\sigma_{k+1,l}^i + \sigma_{k,l+1}^i) .
\tag{7}
$$

A. Performance and properties of the B_LENDER algorithm

The first test is the effectiveness of the algorithm in calculating the density of states of the 2d Ising model. To test the accuracy of the simulations the results will be compared to the exact result solved by Beale [?]. The accuracy of the simulation will be determined by the error defined as,

$$\mathcal{E}(I) = < |\epsilon(E_j, I)| >_j$$
$$= \frac{1}{\Pi} \sum_{j=1}^{\Pi} \frac{|\ln[G_{ex}(E_j)] - \ln[G_{calc}(E_j, I)]|}{|\ln[G_{ex}(E_j)]|} . \quad (8)$$

Where $G_{ex}(E_j)$ is the exact density of states, $G_{calc}(E_j, I)$ is the calculated density of states from Eq. ?? at iteration number I, and $|\epsilon(E_j, I)|$ is the absolute value of the fractional error for a specific energy level. The perturbed configurations in this work were generated by randomly flipping one spin on the Ising lattice

The first test of the algorithm is with the 32×32 Ising model. In a materials science problem with first principles calculations the system size is not expected to be anywhere near the size of the 32×32 Ising model so these results are included to show the algorithm has potential for larger system sizes. While the ideal value of $1/N$ is not known prior to the calculation it was found in this work that a value of $1/N = 0.1$ was computationally efficient for the 32×32 Ising model. In Fig. ??(a) the value of the average error calculated with Eq. ?? is shown up to 1e7 iterations for $\mathcal{S} = 1$, 10, 100, 1000, and 1e4. The data in Fig. ??(a) is averaged over 36 individual simulations for each value of $\mathcal{S}$. At $I = 1e7$ The results show linear scaling from $\mathcal{S} = 1$ to $\mathcal{S} = 10$ and then another order of magnitude improvement from $\mathcal{S} = 10$ to $\mathcal{S} = 1000$, and at $\mathcal{S} = 1e4$ there is marginal improvement at $I = 1e7$ but worse performance for $I < 1e6$. The periodic fluctuations in the average error are also noted in going to larger $\mathcal{S}$, these flucations are due to the calculated density of states(renormalized) oscillating about the exact solution and that for larger $\mathcal{S}$ these flucations are more in sync with each from one calculation to the next leading to a smooth average. For large $\mathcal{S}$ movies of convergence show that the wings of the density of states get more stretched before the walkers have completely scanned the energy range which leads to calculated solution experiencing large oscilations about the exact solutions on the wings. Movies of the convergence of $G_r(E_j)^I \frac{\Omega}{A^I}$ compared to the exact value $G(E_j)$ along with movies of $\ln[G_r(E_j)^I \frac{\Omega}{A^I}] - \ln[G(E_j)]$ are found in the supplementary movies SM1a,b through SM5a,b for $\mathcal{S} = 1$, 10, 100, 1000 and 1e4 respectively.

The next test is with the 10×10 Ising model. In Fig. ??(b) are the results for the average error of 36 independent simulations of a 10×10 Ising model for the different number walkers $\mathcal{S} = 1$, 10, 100, 1000, and 1e4, and with $1/N = 1$. The results show that the scaling is quite good

as the number of walkers increases. This result is encouraging because the number of configurations that are currently and in the near future acessible with density functional thoery simulations is not expected to exceed the large number of $\approx 10^{30}$ configurations in the 10×10 Ising model. Although in the work of Kahn et al. [?] they did tackel a configuration space of size of 10^{74} using linear scaling density functional theory for a 250 atom supercell. Typical density functional theory calculations scale roughly as the cube of the number of atoms, so their calculation was only feasible using the linear scaling methods.
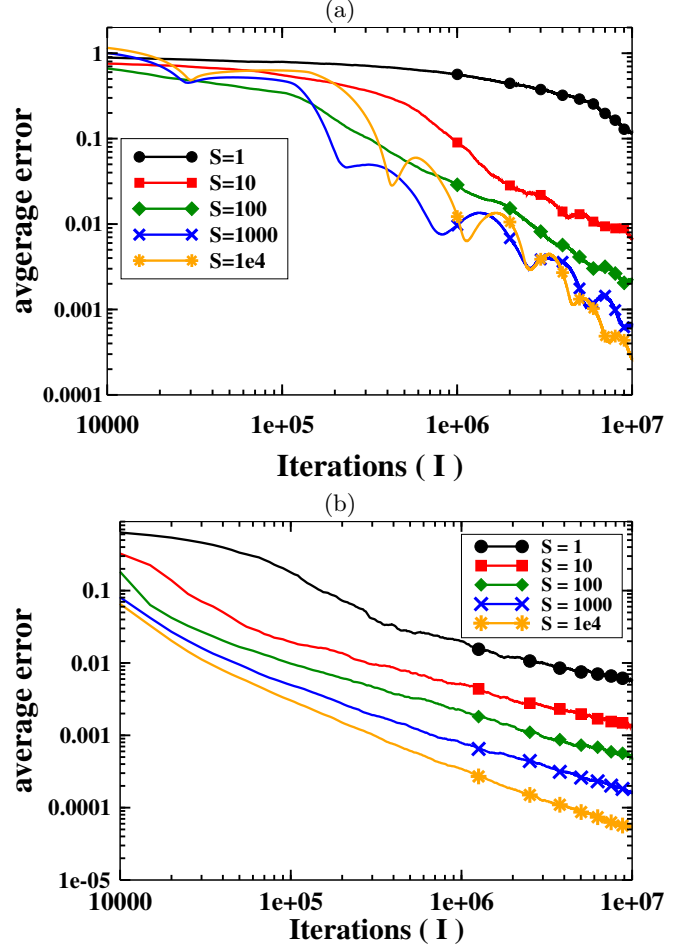

FIG. 1. (color online) Average error from 36 simulations calculated with Eq. ?? with $\mathcal{S} = 1$, 10, 100, 1000, and 1e4 for (a) the 32×32 Ising model with $1/N = 0.1$, (b) the 10×10 Ising model with $1/N = 1$.

Another aspect of the algorithm to consider is the dependence on the value of $1/N$ and of C_o. In Fig. ??(a) and (b) the dependence on $1/N$ is shown for the 32×32 and 10×10 Ising models, simulated to $I = 1e7$ and $I = 1e6$ respectively for $\mathcal{S} = 100$ (a) and $\mathcal{S} = 10$ (b). The results in Fig. ??(a) and (b) were averaged over 36 independent simulations. The results in Fig. ??(a) show that for the larger 32×32 model with $\mathcal{S} = 100$ the dependence

on $1/N$ is more pronounced and that the optimal value of $1/N$ is lower than for the 32×32 model with $\mathcal{S} = 10$ and the smaller 10×10 model for both $\mathcal{S} = 10$ and $\mathcal{S} = 100$. The more pronounced convergence dependence on $1/N$ for the larger 32×32 model at larger $\mathcal{S}$ does pose a problem if one was to implement the the algorithm for a new system where the density of states is not known beforehand because there is no current evidence to predict what the optimal parameter would be. The tests with the 10×10 model suggest that for a smaller system size that the convergence dependence on $1/N$ is less pronounced and that $1/N = 1$ is sufficient. The results presented here suggest that if one was to use the algorithm for a larger system size some method of predicting the optimal value of $1/N$ would be required. These tests have used a value of $C_o = \Omega^{1/N}$, this value was based on tests showing this to be a computationally efficient choice that can be predicted based on knowledge of the system. In Fig. **??**(a) and (b) is shown the error for the 32×32 and 10×10 Ising model vs $C_o/\Omega^{1/N}$ with, $1/N = 0.1$ and $1/N = 1$ respectively, $I = 1e7$ and $I = 1e6$ respectively, and with $\mathcal{S} = 100$ (a) and $\mathcal{S} = 10$ (b). The results in Fig. **??**(a) and (b) were averaged over 36 independent runs. The results in Fig. **??**(a) and (b) show that the errors are relatively insensitive to the value of C_o within several orders of magnitude of $\Omega^{1/N}$ and that the main feature is a sudden increase in error going below some lower bound of C_o and a plateau of the error for C_o above this lower bound.

B. Comparison with original Wang and Landau and the 1/t algorithms

In Wang and Landau's original work they simulated 32×32 and 50×50 2d Ising models for comparison to the exact density of states. The comparison in this section will use the 32×32 Ising model. To implement the original Wang and Landau algorithm the same parameters from their original paper will be used[**?**]. In the orginal Wang and Landau algorithm a histogram $H(E_j)$ is kept of the visited energies, it is perodically reset to zero when a predetermined flatness criteria is made. Also in the orginal Wang and Landau algoritm a modification factor f is defined such that everytime a energy level is visited(accepted) then $G_r(E_j)^{I+1} = G_r(E_j)^I f$. In Wang and Landau's original work they used a flatness criteria of 80% and a reduction schedule of $f^{I+1} = \sqrt{f^I}$. The flatness criteria was described such that all energies E_j had been visited and that $min[H(E_j)]/avg[H(E_j)] > 0.8$. When the flatness criteria is meet the reduction schedule $f^{I+1} = \sqrt{f^I}$ is then implemented. The value of $f^0 = e^1$ was used. In the 1/t algorithm as described by Belardinelli et al. [**?**] the system is first simulated with the Wang and Landau algorithm and when $\ln[f^I] < 1/t$ the time dependence of f^I switches to $\ln[f^I] = 1/t$, where $t = I/\Pi$ the Monte-Carlo time (Note: the inequality $\ln[f^I]$ is only checked when $t > 1$). Initial configurations
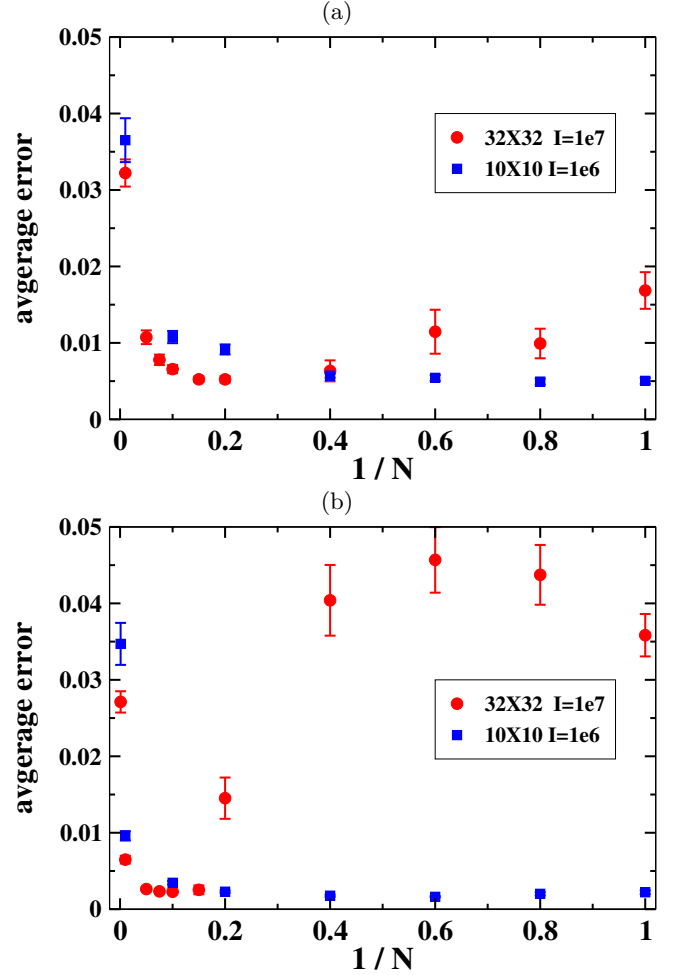

FIG. 2. (color online) Average error calculated from Eq. **??** from 36 simulations for (a) $\mathcal{S} = 10$ and (b) $\mathcal{S} = 100$ vs the value of $1/N$ for the 32×32 (red circles) and 10×10 (blue squares) Ising models simulated to 1e7 and 1e6 iterations respectively. Errorbars show the standard deviation of the mean.

were generated randomly and $G_r(E_j)^0$ was set to 1 when implementing the original Wang and Landau and 1/t algorithms. In this manner the B_LENDER algorithm with $1/N = 0.1$ and $\mathcal{S} = 1$ was compared to the perfomance of the original Wang and Landau algorithm. The results are shown in Fig. **??**(a), which are averaged over 36 independant simulations. The comparison shows a striking difference in the error, the B_LENDER algorithm tends to decrease over the entire time span while the original Wang and Landau implementation rises to a large peak early in the simulation then abrutly drops. At large number of iterations the error is nearly the same for both.

To compare potential parrallel performance of the orginal Wang and Landau and 1/t algorithms in a scheme similar to the B_LENDER algorithm the update,

$$G_r(E_j)^{I+1} = G_r(E_j)^I f^{\mathcal{H}^I} , \qquad (9)$$

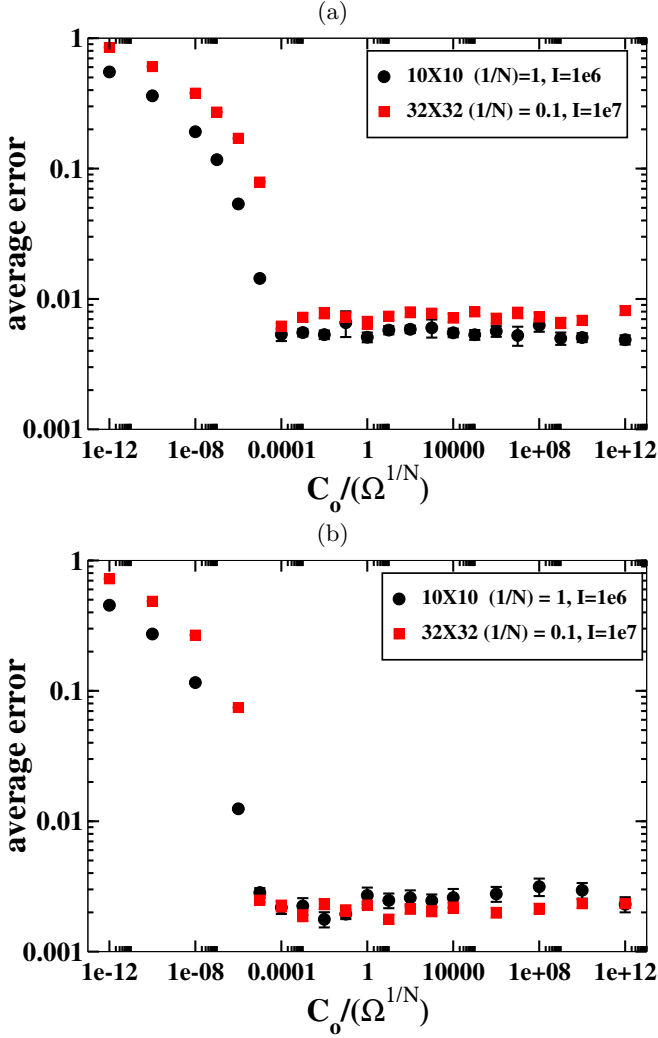is adopted. For the 1/t algorithm in the parrallel scheme

(a)

(b)

FIG. 3. (color online) Average error from 36 simulations calculated from Eq. **??** for the 10×10(black circles) and 32×32(red squares) Ising model simulated to 1e6 and 1e7 iterations and with $1/N = 1$ and $1/N = 0.1$ respectively for (a) $\mathcal{S} = 10$ and (b) $\mathcal{S} = 100$ vs the value of $C_o/\Omega^{1/N}$. Error bars show the standard deviation of the mean.

the Monte-Carlo time is calculated as $t = \mathcal{S}I/\Pi$. Where $\{e_s\}_{\mathcal{S}}^{I+1}$ and $\mathcal{H}^I$ are generated and function the same was an in Eq. **??**. In this manner a simulation was carried out with 100 walkers ($\mathcal{S} = 100$) for the Wang and Landau algorithm, with the same flatness criteria and reduction schedule as above, and for the B_LENDER algorithm to 1e7 iterations. The results are shown in Fig **??**(b), which are averaged over 36 independant simulations. The results show that both parallel implementations have significant improvement for equivalent number of iterations but that the original Wang and Landau still suffers the problem of rising to a large peak and then dropping suddenly. Both algorithms are shown to have similar errors at a large number of iterations until at some point the original Wang and Landau algorithm saturates and the B_LENDER algorithm conintues to converge.

One aspect of the tests so far in this work is that they have used a system size only as large as the 32×32 Ising model. Which in terms of number of configurations is very large for system sizes possible with density functional theory simulations but is small compared to many statistical models commonly studied in phyiscs. In fact Wang and Landau orginally tested their algorithm with the 256×256 Ising model[**?**], although they did this by placing walkers in different energy windows which required either building the larger model from smaller Ising lattices where the energy was known or doing a preliminary energy scan. In this work it was desriable to perform a benchmark with an entirely exploratory calculation of the large 256×256 Ising system. So starting from randomly generated configurations with $\mathcal{S} = 1000$ the B_LENDER algorithm was compared to the orginal Wang and Landau algorithm. It was found that utilizing $1/N = 0.01$ that it was possible to simulate this model quite efficiently with the B_LENDER algorithm as compared to the orginal Wang and Landau using the same parameters as in this section. These results can be considered as a preliminary test of the B_LENDER algorithm for large system size and are found in the supplemental materials, which includes movies of the convergence of the density of states and free energy.

A final comment to make on this section is that the comparisons made between the three algorithms only span a small part of the space of parameters available for each algorithm. The B_LENDER algorithm includes the choice of C_o and $1/N$. The Wang and Landau algorithm requires a choice of the definition of flatness along with a threshold factor for this definition, the value of b in $f^{I+1} = (f^I)^{1/b}$, and the initial value of f. The 1/t algorithm requires the same choices as the Wang and Landau algorithm along with the choice of m and p in m/t^p. The results in this section are meant to compare approximately ideal parameters for the algorithms.

IV. ADIABATIC ANALYSIS OF THE ALGORITHM

To better understand the nature of the algorithm we will consider the adiabatic properties of the histogram $\mathcal{H}^I \equiv \mathcal{H}(E_j, \{e_s\}_{\mathcal{S}}^{I+1})$. Considering that during a simulation the configurations are generated with probability proportional to $G(E_j)$ and accepted inversely proportional to $G_r(E_j)^I$ the probability distribution, which we will write as $\Phi(E_j)^I$, of $\mathcal{H}^I$ will follow the proportionality,

$$\Phi(E_j)^I \propto \frac{G(E_j)}{G_r(E_j)^I} . \qquad (10)$$

For short we will write $G_r(E_j)^I \rightarrow G_r^I$, $G(E_j) \rightarrow G$, and $\Phi(E_j)^I \rightarrow \Phi^I$. Considering that the sum over $\mathcal{H}^I$ is constrained to $\mathcal{S}$ if we normalize Φ^I to $\mathcal{S}$ we get

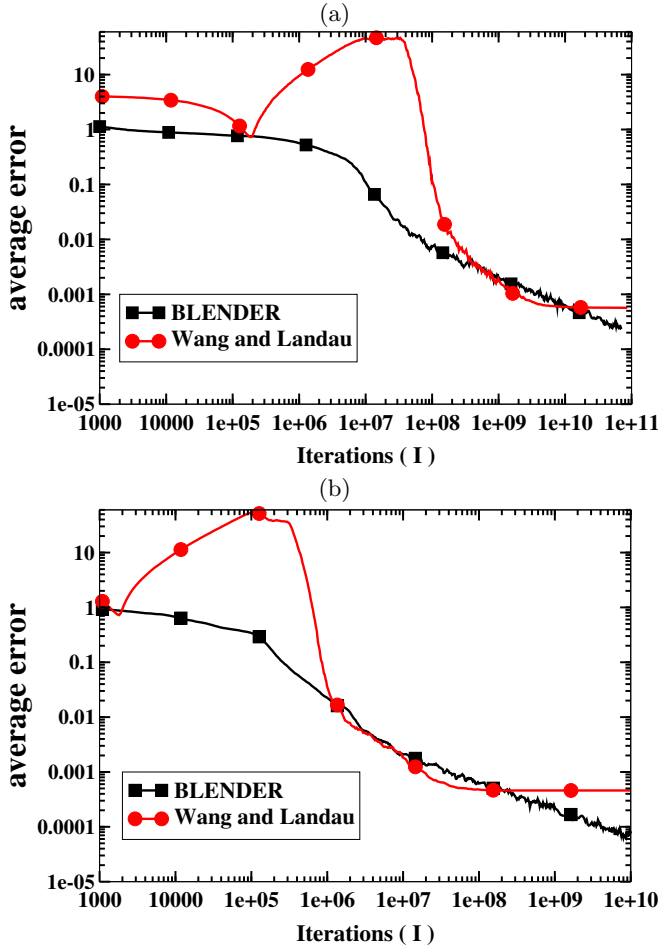$$\Phi^I = \mathcal{S}(\sum_j \frac{G}{G_r^I})^{-1} \frac{G}{G_r^I} . \qquad (11)$$

FIG. 4. (color online) Average error from 36 simulations for the orginal Wang and Landau algorithm in red with circles and the B$_L$ENDER algorithm in black with squares for (a) $\mathcal{S} = 1$ simulated to 1e9 iterations and (b) $\mathcal{S} = 100$ simulated to 1e7 iterations.

Now an adiabatic analysis will be considered by inserting this expression for the adiabatic distribution of the histogram $\mathcal{H}^I$ into the density of states update in step four of the algorithm in Eq. **??**. Doing this gives,

$$G_r^{I+1} = G_r^I + C_o \mathcal{S} (\sum_j \frac{G}{G_r^I})^{-1} \frac{G}{(A^I)^{1/N}} \ , \qquad (12)$$

which will be the basis for the adiabatic analysis of the algorithm. One point to clarify is what is meant by adiabatic. A rigorous definition of adiabaticity for the algorithm can be defined that for any initial $G_r(E_j)^0$, $\{\Sigma_s, e_s\}_S^0$, and ϵ an I and n can be found such that,

$$\langle \sum_j |\frac{1}{n+1} \sum_{i=I}^{I+n} \mathcal{H}^i - \Phi^I|\rangle_o < \epsilon \ , \qquad (13)$$

where $\langle\rangle_o$ means average over trajectories(same intital conditions at I but different subsequent random numbers) of simulation. This is a statement that the algorithm will progressively scan the energy range more and

more thoroughly before a significant change to the density of states is made. From here the analysis will be continued assuming the algorithm is adiabatic.

To continue the analysis we will first determine if and under what circumstances the algorithm will converge assuming adiabacitiy. The first step is to note that,

$$\mathcal{S}C_o(\sum_j \frac{G}{G_r^I})^{-1} \frac{1}{(A^I)^{1/N}} \ , \qquad (14)$$

is an iteration dependent constant(same for each bin j in the density of states), we will call this constant B^I. In this manner we will look at the progression of changes to the relative density of states,

$$G_r^{I+1} = G_r^I + GB^I$$
$$G_r^{I+2} = G_r^I + GB^I + GB^{I+1}$$
$$...$$
$$...$$
$$G_r^{I+n} = G_r^I + G \sum_{i=0}^{n-1} B^{I+i} \qquad (15)$$

For short $\sum_{i=0}^{n-1} B^{I+i}$ will be referred to as W^n The condition for convergence can be defined such that for two bins of the density of states l and k,

$$\lim_{n \to \infty} \frac{G_r^I(E_k) + G(E_k)W^n}{G_r^I(E_l) + G(E_l)W^n} \to \frac{G(E_k)}{G(E_l)} \ . \qquad (16)$$

For this to occur W^n must increase unbounded as $n \to \infty$ and not limit to zero or a constant. To make further progress we will first rewrite B^I as ,

$$\mathcal{S}C_o(\sum_j \frac{G}{G_r^I})^{-1} \frac{1}{(A^I)^{1/N}} =$$
$$\mathcal{S}C_o(\sum_j \frac{G}{G_r^I})^{-1} \frac{1}{A^I(A^I)^{1/N-1}} = \qquad (17)$$
$$\mathcal{S}C_o(\sum_j \frac{G}{G_r^I})^{-1} \frac{(A^I)^x}{A^I} \ ,$$

where $x = 1 - 1/N$. A sufficient condition to show convergence would be to show that the terms B^{I+i} do not go to zero as $\lim_{n\to\infty}$. In this context the lower bound of B^I will be studied. Considering that $\frac{min(G_r^I)}{\Omega} \leq (\sum_j \frac{G}{G_r^I})^{-1}$ and that $\frac{min(G_r^I)}{\Pi max(G_r^I)} \leq \frac{min(G_r^I)}{A^I}$ we can bound B^I as,

$$\frac{\mathcal{S}C_o}{\Pi\Omega} \frac{min(G_r^I)}{max(G_r^I)}(A^I)^x \leq B^I \leq \frac{\mathcal{S}C_o}{\Pi\Omega} \frac{max(G_r^I)}{min(G_r^I)}(A^I)^x \quad (18)$$

For the case of $0 \leq x < 1$ convergence is clear because $\frac{min(G_r^I)}{max(G_r^I)}$ can not limit to zero and $(A^I)^x \geq 1$. The term $\frac{min(G_r^I)}{max(G_r^I)}$ can not limit to zero for the adiabatic analysis because considering if the minimum is at the bin for E_k

and the maximum at the bin for E_l then ratio is of the form,

$$\frac{G_r^0(E_k) + G(E_k)W^I}{G_r^0(E_l) + G(E_l)W^I} \; . \tag{19}$$

For the case of $x < 0$ convergence is not so clear although in theory the algorithm may still be convergent based on contradiction. In Eq. **??** the upper bound on B^I tells us that for $x < 0$ if the algorithm does converge the B^I will tend towards zero since A^I must grow unbounded if the algorithm converges. This means that if the algorithm does converge for $x < 0$ that the sum of the B^I terms must increase unbounded although they tend towards zero. On the other hand if the algorithm does not converge, i.e. the sum of the B^I terms tends towards a constant, then A^I will not grow unbounded. This is a contradiction because if A^I does not grow unbounded then the B^I terms can't tend towards zero and the sum of them should grow unbounded and the algorithm should converge. This suggests that in theory the algorithm is convergent for $x < 0$ but it is dependent on the sum of terms that tend towards zero, which is essentially predicting slow convergence. In practice tests with the 8×8 Ising model reasonable convergence could only be achieved for "small" negative values of x.

So now considering the algorithm does converge within the adiabatic assumption we can derive the time dependence of $F_{ab}(E_j)^I = \ln[f_{ab}(E_j)^I]$ where $f_{ab}(E_j)^I = (1 + \frac{G}{G_r^I}B^I)$ within this analysis. Here the label ab is used to distinguish the values of $F(E_j)^I$ and $f(E_j)^I$ calculated within the adiabatic assumption, this point is discussed further later in this section. First rewriting $\frac{G}{G_r^I}B^I$ as $\frac{C_o \mathcal{H}^I}{(A^I)^{1/N}}$ it is clear that $\frac{G}{G_r^I}B^I$ becomes arbitrarily small for $I \to \infty$. With $\ln(1+a) \approx a$ for small a we now have for large I,

$$F_{ab}(E_j)^I = \ln[f_{ab}(E_j)^I] \approx \frac{GB^I}{G_r^I} \; . \tag{20}$$

Also because the algorithm is converging G_r^I is approaching $W^I G$ and $\sum_j G_r^I = A^I$ is approaching $W^I \Omega$ such that $(\sum_j \frac{G}{G_r^I})^{-1}\frac{1}{A^I} \approx \frac{W^I}{\Pi \Omega}$. With this we write in the large iteration limit $B^I \approx \frac{C_o \mathcal{S}}{\Pi \Omega}(A^I)^x$ and $G_r^I \approx G\frac{C_o \mathcal{S}}{\Pi \Omega}\sum_{i=0}^{I-1}(A^i)^x$ such that,

$$F_{ab}^I \approx \frac{(A^I)^x}{\sum_{i=0}^{I-1}(A^i)^x} \; . \tag{21}$$

The E_j as now been suppressed in Eq. **??** because in the convergent limit F_{ab}^I is uniform across bins j. In this form it is clear that the time dependence for the case of $x = 0$ is of $1/I$ form. For $x! = 0$ the time dependence is not as clear analytically but it can be simulated by using a boot strapping procedure. Assuming we start with $G_r^0 = cG$

so that $A^0 = c\Omega$ then,

$$\begin{aligned}
A^0 &= c\Omega \\
A^1 &= A^0 + \frac{C_o \mathcal{S}}{\Pi}(A^0)^x \\
A^2 &= A^1 + \frac{C_o \mathcal{S}}{\Pi}(A^1)^x \\
&\cdots\cdots \\
A^I &= A^{I-1} + \frac{C_o \mathcal{S}}{\Pi}(A^{I-1})^x
\end{aligned} \tag{22}$$

In this way a numerical simulation can be done to predict the value of F_{ab}^I in Eq. **??**. This was done using a starting value of $c = 1$ for $\mathcal{S} = 1$ and 100, and $1/N = 0.1$ and 0.5 to compare to results from an explicit simulation of the 32×32 Ising model. An important point to make is that the f_{ab}^I calculated from the adiabatic analysis in the convergent limit is effectively an average value per bin, this is because within the adiabatic analysis when the distribution Φ^I of $\mathcal{H}^I$ is inserted in the density of states updated the inherently discrete walkers are continously spread over all the bins. To compare to the explicit B_LENDER simulation we need to also calculate the average of $f(E_j)^I$ which is defined as,

$$\begin{aligned}
< f(E_j)^I >_j &\equiv (1 + \frac{1}{\Pi}\sum_j \frac{\mathcal{H}^I C_o}{(A^I)^{1/N}}) \\
&= (1 + \frac{\mathcal{S}C_o}{\Pi(A^I)^{1/N}}) \; ,
\end{aligned} \tag{23}$$

is calculated. With this then the value,

$$F_{avg}^I \equiv \ln[< f(E_j)^I >_j] \; , \tag{24}$$

is calculated for comparision to F_{ab}^I. The theoretical adiabatic results and the results using the B_LENDER algorithm are shown in Fig. **??**. The agreement is seen to be very good and both results have a m/I long time dependence with $m = 10$ and 2 for $1/N = 0.1$ and 0.5 respectively. The results in Fig. **??** also show that the long time behavior of F_{avg}^I is independant of $\mathcal{S}$. In fact the long time behavior of F_{avg}^I is predicted to equivalent for all $\mathcal{S}$, C_o, and model studied. It is notable that the predicted value of m is equal to the value of N in $1/N$, this was confirmed for many other simulations using Eq. **??**. So in general the prediction can be made that for the B_LENDER algorithm the long time behavior of F_{avg}^I is predicted to take the form N/I regardless of the choice of C_o and $\mathcal{S}$.

V. APPLICATION TO LLTO

The purpose of developing the B_LENDER algorithm was to develop an algorithm suitable for the needs of solid state density functional theory calculations of disordered crystal sublattices. Due to the long run time
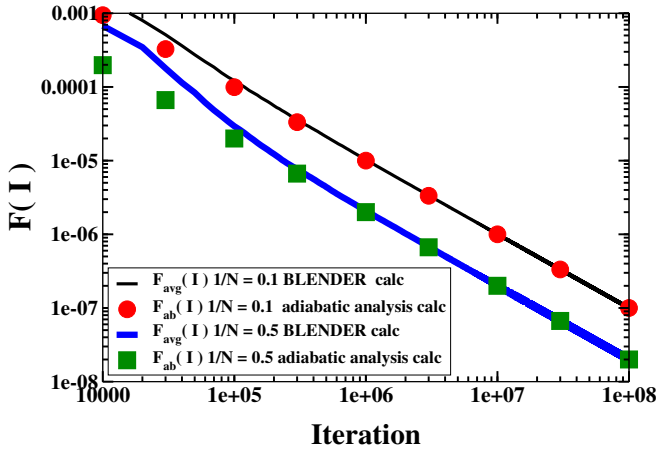
FIG. 5. F_{ab}^I calculated with the boot strapping method Eq. **??** and with an explicit simulation for the 32×32 Ising model using F_{avg}^I as defined by Eq. **??** for the B_LENDER algorithm with $\mathcal{S}$=1 and 100, $1/N = 0.1$ and 0.5.

of density functional theory calculations the parallel nature of the B_LENDER algorithm allows for calculations of each energy to be done as independent job submissions to a computer cluster. The results can then be processed by a script running on the head node and the implementation of the algorithm is relatively simple. In this work the B_LENDER algorithm is applied to the lithium and lanthanum sublattice of the solid state lithium ion electrolyte $Li_{0.5}La_{0.5}TiO_3$. The goal of this study was to both, perform a calculation with the B_LENDER algorithm of a real material system that is fairly well understood, and also to learn something new in the process. Specifically the desired knowledge to be gained is a better understanding of the lithium and lanthanum sublattice disordering.

A.   Background on LLTO

LLTO is a complex material comprised of a variety of stoichiometries and phases but in this work the study is restricted to the reported tetragonal P4/mmm phase of the stoichiometry $Li_{0.5}La_{0.5}TiO_3$[**? ?**]. A unit cell of this structure is shown in Fig. **??**. The lattice parameters for this unit cell were taken from the experimental results from Ibarra et al.[**?**]; 3.8688(4) Å for a- and b-axes, and 7.7463(2) Å for c-axis. This unit cell is representative of an ordered form of $Li_{0.5}La_{0.5}TiO_3$ where the lithium and lanthanum are separated into separate layers on the high symmetry A-sites. Where the A-site refers to the general perovskite formula unit ABX_3. The structure in Fig. **??** is actually structurally unstable and the energy can be lowered by lattice distortions which manifests as tilts in the titanium oxygen octahedra and the lithium distorting off of the high symmetry A-sites. The instability of the structure in Fig. **??** is evidenced by the imaginary phonon modes calculated by Moriwake et al. [**?**].

The physics of interest in this study is to understand the disordering of lithium and lanthanum between layers. It is reported for this phase that the lanthanum are mostly mixed between layers when the samples are slow cooled during synthesis and if quenched from high temperature the lanthanum ordering is reported to be completely mixed between layers [**?**]. In this work the B_LENDER algorithm is used to evaluate the density of states associated with local minimum corresponding to the lithium and lanthanum ordering and associated lattice distortions.
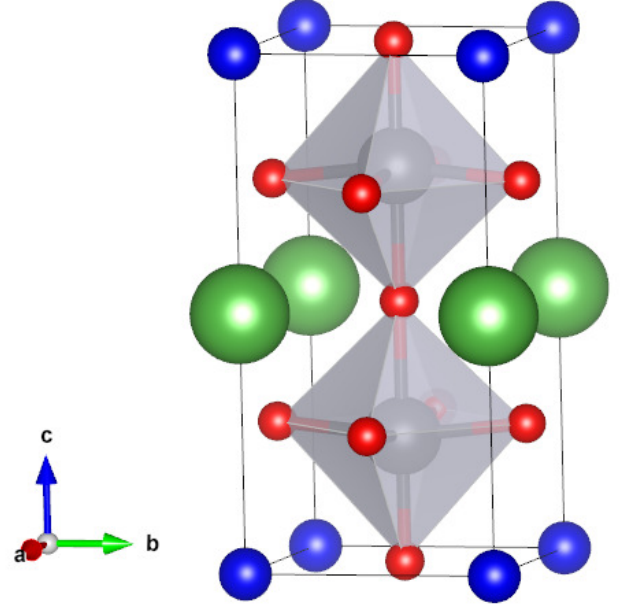

FIG. 6.   (color online) 10 atom unit cell of P4/mmm $Li_{0.5}La_{0.5}TiO_3$. Where dark blue spheres are lithium, green spheres are lanthanum, red spheres are oxygen, and grey spheres inside of octahedra are titanium.

B.   Computational Details

In this work a 3×3×1 90 atom supercell with periodic boundary conditions of the unit cell depicted in Fig. **??** was used as an approximation to bulk $Li_{0.5}La_{0.5}TiO_3$. While not an ideal size as it is restrictive of the possible lattice configurations and to the types of domains of octahedral tilting that can form it is the largest supercell practical for performing the configurational Monte Carlo in this work.

An important aspect of completing this study is a scheme for producing the initial and perturbed configurations in the iterative process of the B_LENDER algorithm. The scheme used in this study was to first generate a set of lithium and lanthanum randomly placed on the high symmetry A-sites where occupancy is restricted

to one, then a small amount of noise on the order of ± 0.2 $\mathring{A}$ was added to each lithium and lanthanum coordinate. These configurations were then relaxed to a local minimum which formed the first set of configurations in the iterative process. Then the perturbed configurations were formed by swapping a random lithium and lanthanum atom and placing them back on the high symmetry A-site along with a new amount of random noise, these configurations where then relaxed to a local minimum. The random noise off the A-site served to assist searching the distorted lattice configuration space.

The method used in the calculation of the total energies of the lattice configurations of LLTO in this work was density functional theory using the Vienna *Ab-initio* Simulation Package (VASP) [? ? ? ?] within the projector augmented wave formalism (PAW)[?]. The local density approximation(LDA) was used for the exchange and correlation functional[?]. The valence electron configurations for the PAW data sets were; $5p^6 5d^1 6s^2$ for La, $2s^1$ for Li, $3p^6 3d^2 3s^2$ for Ti, and $2s^2 2p^4$ for O. The calculations also took advantage of the "soft" option for La and O. The total energy cut off for expansion of the plane waves was 250 eV. Self consistent cycles were converged with a energy difference of $< 1e\text{-}5$ eV and relaxation of atomic coordinates was terminated when the difference in total energy between ionic relaxation steps was $< 1e\text{-}4$ eV. Electronic occupations used Guassian smearing with a width of 0.05 eV. A $1\times1\times2$ gamma centered k-point mesh was used for the $3\times3\times1$ supercells of the LiLaTiO6 unit cell. These cutoffs and parameters were chosen to maximize computational efficiency while retaining enough accuracy to capture important physical properties of LLTO. Each calculation of an energy was completed with a 36 processor broadwell node with an average wall time of approximately 5 minutes per calculation. To test the accuracy of these methods 10 structures were calculated with fixed coordinates at these convergence criteria and more accurate PAW data sets and cutoffs. The more accurate PAW data sets included the valence electron configurations; $5s^2 5p^6 5d^1 6s^2$ for La, $1s^2 2s^1$ for Li, $3p^6 3d^2 3s^2$ for Ti, and $2s^2 2p^4$ for O. The cutoffs for the more accurate calculations were 500 eV for the plane wave basis, and $2\times2\times3$ gamma centered k-points. The average magnitude in relative energy between structures from this test was 0.15 eV.

An important point to make about the calculations in this research is that they included 0 K static lattice internal energies, did not include phonon free energies, and they did not take into account relaxation of lattice parameters[?]. Ideally for fixed lattice parameters we would evaluate the parition function[?],

$$Z = \sum_{i=1}^{\Omega} e^{-\frac{(u_i + f_i(T))}{k_B T}} \ . \tag{25}$$

Where T is the temperature, k_B is Boltzmann's constant, u_i is the static lattice internal energy for configuration i, and $f_i(T)$ is the temperature dependant phonon free energy for configuration i. A configuration is defined in this work as a local minimum of the Born Oppenheimer potential energy surface. This form of the parition function poses the problem that the density of states is now temperature dependant. Take $u_i + f_i(T) \equiv e_i(T)$, let W be the number of unique $e_i(T)$, and $F_j(T)$ be the unique $e_i(T)$ then,

$$Z = \sum_{j=1}^{W} G(F_j(T)) e^{-\frac{F_j(T)}{k_B T}} \ . \tag{26}$$

This is a problem because to employ the Monte Carlo methods discussed in this work they would have to be applied at different temperatures which defeats the original purpose. This problem could be addressed by taking U_j to be the unique u_i, Π to be the number of U_j, and then considering a different form of the parition function in Eq. **??**,

$$Z = \sum_{j=1}^{\Pi} < e^{-\frac{f_i(T)}{k_B T}} >_j G(U_j) e^{-\frac{U_j}{k_B T}} \ . \tag{27}$$

Here $< \exp(-f_i(T)/k_B T) >_j$ is the arithmetic average of $\exp(-f_i(T)/k_B T)$ over all configurations with static lattice internal energy U_j. In this form the static lattice density of states $G(U_j)$ could first be determined with the Monte Carlo methods described in this work and then the $< \exp(-f_i(T)/k_B T) >_j$ could then be approximated by randomly choosing some of the perturbed configurations for each U_j to evaluate the phonon density of states. With the randomly generated phonon densities of states the $< \exp(-f_i(T)/k_B T) >_j$ could be approximated for any temperature. Even in this form though the computational expense is beyond the scope of this work as phonon calculations require high cutoffs and even within the harmonic approximation are much more computationally expense than static lattice calculations[?]. So in this work the approximation,

$$Z \approx \sum_{j=1}^{\Pi} G(U_j) e^{-\frac{U_j}{k_B T}} \ , \tag{28}$$

is made. This is reasonable because the phonon free energies are expected to vary much less from configuration to configuration than the static lattice internal energies. A final comment is that relaxation of the lattice paramters could be accomplished by calculating the partition function on a grid of lattice constants, then a free energy surface could be interpolated and the lattice paramters minimizing the free energy determined as a function of temperature. This procedure is also beyond the scope of the current research.

The calculations were performed at the experimental lattice parameters 3.8688 $\mathring{A}$ for a- and b-axes, and 7.7463 $\mathring{A}$ for c-axis. The parameters for the B$_L$ENDER algorithm were $\mathcal{S} = 10$ and $1/N = 1$. The bin width used for determining $G_r(E_j)$ was chosen to be 0.05 eV. The

value of Ω was estimated as 100 times the combinatoric number of configurations of the lithium and lanthanum ordering onto the A-site given as,

$$\Omega \approx 100 \frac{18!}{9!9!} \ . \tag{29}$$

While an exact value of Ω is not needed for the algorithm to converge experience from the 2d Ising model suggests that being within several orders of magnitude is sufficient. Estimating that Ω is greater than the combinatoric calculation of the lithium and lanthanum in the A-site cages comes from the possibility of multiple distinct lattice distortions for each type of A-site cage configurations. An approximate upper bound on the number of distinct lattice distortions for each A-site cage configuration can be based on the experimental and theoretical known that lithium tend to occupy the six possible sites corresponding to local minimum near the oxygen windows connecting different A-site cages and that lanthanum tend to occupy the center of the cage [**? ? ? ?**]. If every lithium could occupy one of these six locations within the A-site cage irrespective of the ordering of the other lithium the number of distinct lattice distortions could be estimated as $6^9 \approx 1e7$. So the approximation $\Omega \lesssim 1e7 \frac{18!}{9!9!}$ can be made. It is physically reasonable that a significant fraction of these configurations will be unstable so that the estimate of Ω in Eq. **??** is likely to be within several orders of magnitude of Ω.

C. Results

Using the parameters and configurational enumeration scheme specified above a simulation was performed to 10,000 iterations for the $3{\times}3{\times}1$, 90 atom supercell. After 150 iterations the algorithm was restricted to look in the energy range less than 2.8 eV higher than the lowest energy found at that time. This was to improve computational efficiency by preventing the walkers from exploring an unnecessarily high energy range. While 10,000 iterations is not ideally converged, it was sufficient to gain further understanding of the material. In principle it would be desirable to use some type of stopping criteria to determine convergence of the simulation such as in the work of Caprica[**?**] whom tracked thermodynamic quantities such as the peak of specific heat to determine convergence. In this work convergence is limited by computational resources. It is expected that the qualitative aspects of the results are well accounted for despite the limited number of iterations.

The main focus of the results is the nature of the lithium and lanthanum sublattice ordering. To accomplish this the order parameter of interest is that of the occupancy of lanthanum in the lanthanum rich layer along the c-axis. In the work by Ibarra et al. [**?**] they refer to this order parameter as $La1$, the same convention will be used in this work. This order parameter, $La1$, is defined as the number of lanthanum in the lanthanum

rich layer divided by the total number that could occupy the layer. As an example the unit cell in Fig. **??** would have $La1 = 1$. It is important to note in this work the $3{\times}3{\times}1$ supercell restricts the configurations along the a- and b-axes from having alternate layering of lithium and lanthanum rich layers. Ideally the calculations would be done with at least a $4{\times}4{\times}1$ supercell but the computational effort is beyond the scope of this work. The results later will have to be interpreted taking this systematic supercell error into account.

To calculate the ensemble average of these order parameters first arithmetic averages of the order parameter at each energy level E_j are calculated from the perturbed configurations that occurred during the simulation. The arithmetic average of a general order parameter O over all configurations with energy E_j is denoted by $<O>_j$. Then with these the ensemble average is computed as,

$$\langle O \rangle = \sum_{j=1}^{\Pi} <O>_j \ G_r(E_j) \frac{e^{-\frac{E_j}{k_B T}}}{Z} \ . \tag{30}$$

Where $Z = \sum_{j=1}^{\Pi} G_r(E_j) exp(-\frac{E_j}{k_B T})$. It is noted that normalization of the relative density of states to the appropriate number of configurations is not necessary for the calculation of the ensemble average of an order parameter. If wanting to compare free energies$(-k_B T \ln(Z))$ between phases it would be necessary to normalize the density of states properly to obtain an accurate calculation of the free energy.

The first main result is a view of the convergence of $G_r(U_j)$ as a function of the iterations. Here we have switched to using U_j to highlight that the calculations only include 0 K static lattice internal energies as discussed before. In Fig. **??**, $G_r(U_j)$ is shown at $I = 500$, 2000, and 10,000 with the y-axis plotted on a log scale. The $G_r(U_j)$ shown in Fig. **??** are plotted such that the lowest energy of $G_r(U_j)$ found at the particular iteration shown is set to zero on the x-axis, the sharp cutoff at higher energy was the upper limit to energy range, and the plots are normalized by dividing through by the minimum of $G_r(U_j)$ at that iteration. The main characteristic of the results by 6,000 iterations is the presence of some low energy states seperated by gaps followed by a guassian like distribution in the density of states starting at 0.8 eV f. It is noted, as the iterations increase, that $G_r(U_j)$ for the spectrum of higher energy states becomes noticeably smoother. The lowest energy configuration is characterized as having $La1 = 1$, that being having alternate layers of lithium and lanthanum along the c-axis, in fact all of the structures found up to 0.6 eV have $La1 = 1$. The are not however equivalent to the unit cell shown in Fig. **??**, in that the structure has distinct lattice distortions due to the lithium sitting off of the high symettry A-sites on the oxygen windows sepearting A-site cages.

The next result is the arithmetic averages of the $La1$ order parameter, which are shown in Fig. **??**(a) along with the number of samples used to determine each value
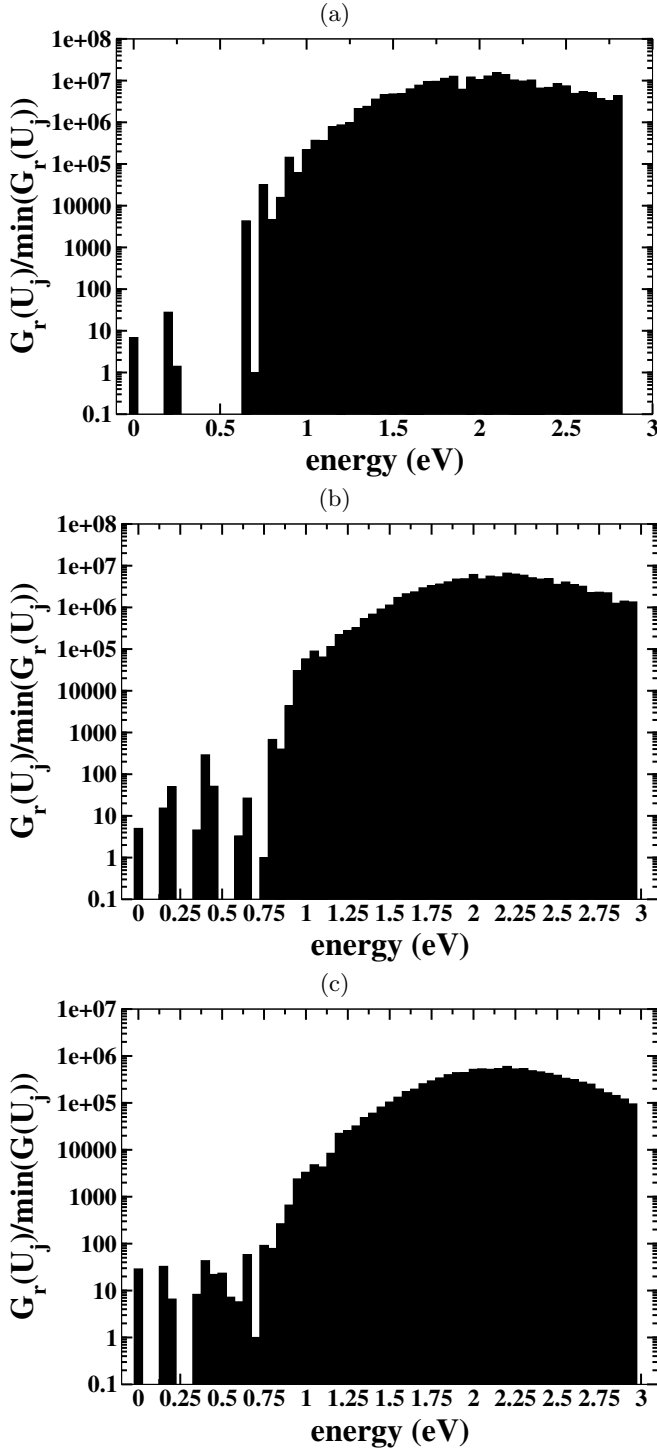
FIG. 7. Plots of $G_r(U_j)$ at (a) 500 iterations, (b) 2,000 iterations, and (c) 10,000 iterations. The plots are normalized by dividing through by the minimum value of $G_r(U_j)$ at that particular iteration. The plots are shown with a log scale on the y-axis.

in Fig. **??**(b). The results in Fig. **??**(a) show an overall tendency for more mixing of lithium and lanthanum between layers for higher energies. The ensemble aver-

age of $La1$ is shown in Fig. **??**, which shows a phase transition from completely segregated lithium and lanthanum between layers to mostly mixed between layers and increased mixing with increasing temperature. In Fig. **??** the value of the ensemble average of $La1$ shows a transition from a value of 1 to more mixed between layers starting at 1250K, this is lower than typical sintering temperatures of 1350K. The temperature is plotted to the arbitrarily high temperature of 4000K to show how the ensemble average of $La1$ decreases for increasing temperature for this fixed lattice simulation. For the 3×3×1 model the minimum possible value of $La1$ is $5/9 = 0.56$. These results are in qualitative agreement with the $La1 = 0.53$ reported by Ibarra et al. [**?**] and it is expected that that larger supercell models would increase the available configurational entropy and result in a reduced value of $La1$ in the disordered phase. The qualitative behavior refers to that the prediction that $La1$ becomes more mixed between layers as the temperature increases but because of the many factors affecting the accuracy of the simulation it is not expected that the temperature of the onset of this mixing to be a highly accurate prediction. These factors include the methodology(the local density approximation), the number of valence electrons, convergence criteria, k-points, size of supercell, and iterations performed. At a minimum to gain a better estimate of the transition temperature a calculation of a 4×4×1 supercell along with relaxation of the lattice parameters would need to be performed to arbitrarily large iteration number.

Some indication of convergence of the 10,000 iterations comes from inspecting the flatness of the histogram $H(U_j)$ of visited energies during the simulation. Why this is in an indicator of convergence can be understood by Eq. **??** which indicates that the adiabatic distribution of $\mathcal{H}^I$ is flat for bins that are converged realtive to each other. If $\mathcal{H}^I$ is flat then the total histogram $H(U_i)^I = \sum_{i=0}^{I} \mathcal{H}^i$ should form a flat histogram. A histogram of the visited energies during the simulation is shown in Fig. **??**. The results show that the histogram is qualitatively flat for $\gtrsim 0.8$ eV. This result along with the counts for calculating the $La1$ order parameter shown in Fig. **??**(b) suggest that the results are most well converged for $\gtrsim 0.8$ eV. While the first principles methods used can be considered coarse grained in terms of PAW data sets, total energy cutoffs, and k-points as observed from testing with more accurate methods, the trend seen in figure Fig. **??**(a) spans an energy range much greater than the expected relative error in energies between structures. It must be said that the ensemble average of $La1$ is highly dependent on the low energy structures as per the exponential nature of the partition function. In this regard the observed phase transition in Fig. **??** can not be expected to be an accurate prediction of a transition temperature. The most important result of Fig. **??** is the high temperature region above the phase transition showing a tendency for greater mixing of lithium and lanthanum for increasing temperature.
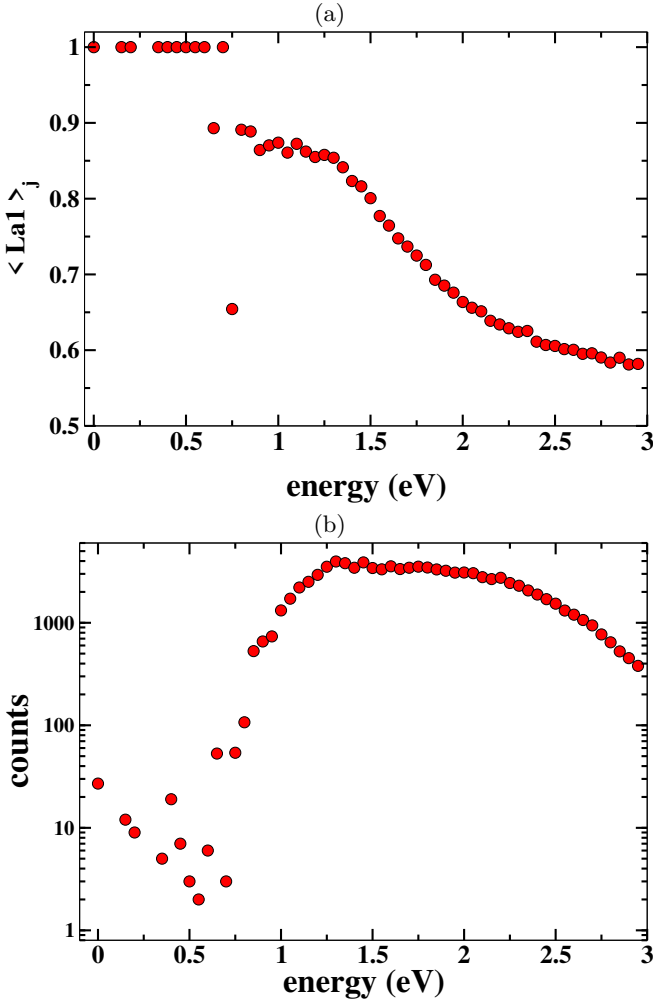
(a)

(b)

FIG. 8. (a) Arithmetic averages of the $La1$ order parameter as a function of energy. (b) Number of counts to determine each value of the $La1$ order parameter for each energy. For the bins with total count of 1 in (b) the standard deviation of the mean plotted in (a) is simply plotted as zero.

To gain some further insight into the structures found during the simulation the lowest energy structure and lowest energy structure with $La1$!= 1 are shown in Fig. **??**(a) and (b). In Fig. **??**(a) it is seen that the lowest energy structure has the lithium and lanthanum completely segregated between layers along the c-axis. In Fig. **??**(b) the lowest energy structure with $La1$!= 1 has a fully occupied layer of lithium along the a-axis along with two partially occupied layers, this structure is 0.75 eV in energy above the lowest energy structure. Structures similar to that in Fig. **??**(b) are in part responsible for the drop in the $La1$ order parameter in Fig. **??**(a) starting at 0.6 eV . Due to the pseudo cubic nature of the system it is expected that there would be nearly energetically identical structures consisting of completely segregated lithium and lanthanum layers along the a and b axes. This feature of ambiguity of the orientation of how the lithium and lanthanum rich layers can form are
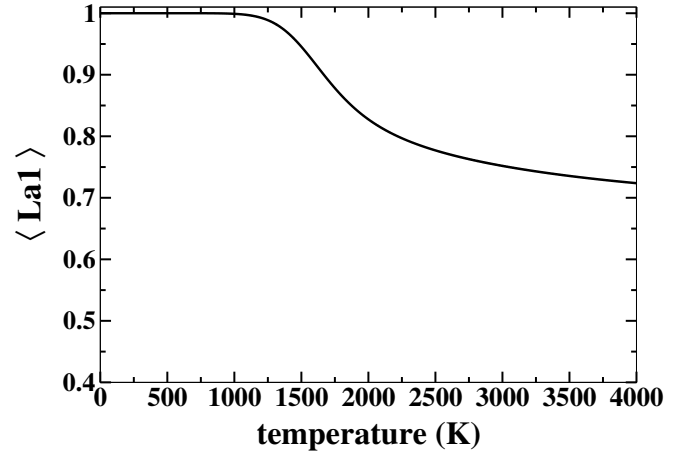

FIG. 9. Ensemble average of the $La1$ order parameter calculated with Eq. **??** as function of temperature.

a likely driving force for the numerous domain boundaries observed in experiments for other stoichiometries synthesized by annealing from high temperature[**? ?**]. It is a prediction of this work that if images were taken of the $Li_0.5La_0.5TiO_3$ stociometry they would also reveal numerous domains charecterized by segregation of the lithium and lanthanum into different layers. Due to the $3\times3\times1$ supercell used in this study these structures were not possible to fully realize in the simulation but as evidenced by Fig. **??**(b) segragation of the lithium and lanthanum into separate layers is still energetically favorable. A common feature of both Fig. **??**(a) and (b) is the lithium sitting off of the high symmetry A-site near the oxygen windows separating A-site cages. This feature has been previously reported experimentally and theoretically in the literature[**? ? ?**].

Another results of interest for this calculation is the value of the modification factor to the density of states as determined by Eq. . In Fig. **??** the modification factor calculated with Eq. is shown vs the iterations. As per the predictions previously made in Sec 4 the a linear fit of the modification factor on a log log plot should approach log(N) - log(I) which for this case would be 0 - log(I). A log10 log10 fit to Fig. in the range $I =$5000 to 10000 gives a log10(2.32) - 1.08log(I). So the calculation does appear to be approximately following the predictions. It would take significantly more iterations(5-10$\times$) to determine if the calculation continues to approach to the exactly predicted value.

A final comment of this calculation of the energy density of states of LLTO with first principles methods is how it highlights the great difficulty one faces in using this method for evaluation of the partition function and ensemble average order paramaters. Due to the N^3 scaling of the density functional theory calculations with number of electrons and the factorial scaling of the number of configurations the ability to reach larger system sizes is daunting. It is expected that without significant improvement in methods, algorithms, and

computing power it will remain challenging for over a decade to reach a resonable calculation of the partition function for even just the $4{\times}4{\times}1$ $Li_{0.5}La_{0.5}TiO_3$ system including lattice relaxation using density functional theory methods without the use of a very large amount of computing resources. A quick "Fermi" problem back of the envelope calculation based on the scaling in computing power for the methods used in this work along with the increase in number of configurations we may expect that the computing power to achieve similar convergence as for the $3{\times}3{\times}1$ for the $4{\times}4{\times}1$ system would take between 10times to 1e5times the computing resources. The lower bound represents equal number of iterations for the $4{\times}4{\times}1$ system and the upper represents the number of iterations required scaling directly with the number of configurations. So if Moore's law continues and computing power doubles every 2 years it is predicted that it will take greater than 6 years and less than approximately 3decades before the $4{\times}4{\times}1$ system could be computed with a similar wall time as the $3{\times}3{\times}1$ system in this work using the same methodology.

VI. CONCLUSIONS

This work has presented a parallel variant of the Wang and Landau algorithm referred to as B_LENDER($B_L$end Each New Density Each Round). The algorithm was developed purposely for use with disordered crystal sublattices and is naturally parallel. It's design makes it facile to implement on a mid level high performance computer such as Argonne's BEBOP where jobs for a structural energy calculation can be independently submitted to compute nodes and managed by a script running on a head node. It was trialed using the 2d Ising model and showed good performance for the $10{\times}10$ Ising model with a minimal number of implementation parameters. Results for the $32{\times}32$ Ising model suggest the algorithm could have applicability to larger system sizes provided a tuning parameter is chosen appropriately to maximize performance, currently it is not known how to choose this parameter before hand without numerical testing. Comparing performance with the the original Wang and Landau and $1/t$ algorithms showed that the B_LENDER algorithm has superior short time performance. Long time behavior of the B_LENDER algorithm and the $1/t$ algorithm are found to approximately equivalent in that they both appear to be convergent, while the orginal Wang and Landau algorithm suffers from a saturation in the error. Convergence of the algorithm is established within an adiabatic assumption and this analysis was able to correctly derive the time dependence of the modification factor for the algorithm for the $32{\times}32$ Ising model. The long time dependence of the natural log of the time dependence of the average modification factor per bin for the B_LENDER algorithm is predicted to take
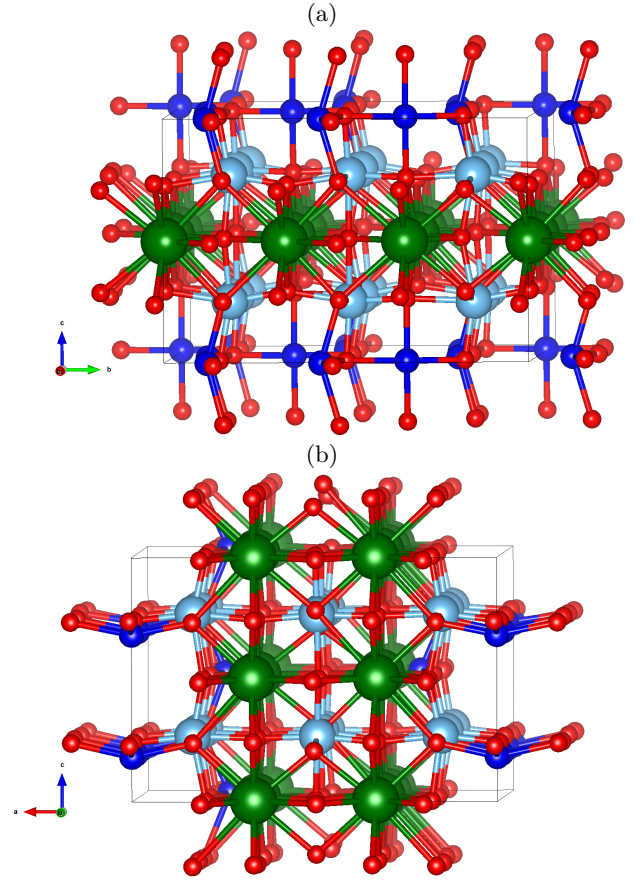
FIG. 10. (color online) (a) Lowest energy structure found during simulation. (b) Lowest energy structure found during the simulation that has $La1 ! = 1$. Where dark blue spheres are lithium, green spheres are lanthanum, red spheres are oxygen, and light blue spheres are titanium.
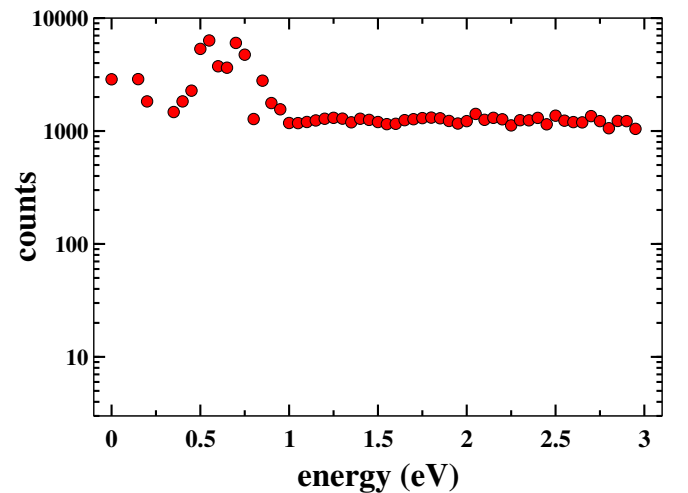
FIG. 11. Plot of the histogram of the visited energies over the 10,000 iterations of the simulation. Visited here means the accepted energies as per step 3 of Eq. **??**.

a N/I form. Knowledge gained from testing with the 2d Ising model allowed for an informed implementation to the real material science problem of studying the lithium and lanthanum sublattice disorder of $Li_{0.5}La_{0.5}TiO_3$ using density functional theory methods. The simulations of the disordered lithium ion conductor LLTO were in qualitative agreement with experiment and provided further insight into the disordered nature of the material. It was found that lower energy structures favored segregating lithium and lanthanum into separate layers and that structures with lithium and lanthanum mixed between layers were on average higher in energy than more segregated structures. Thermodynamic analysis of the order parameter related to lithium and lanthanum intermixing between layers showed a phase transition between completely segregated to mostly mixed tending to more mixed at higher temperatures. Overall the results show that the algorithm performed well in simulating both the 2d Ising model and with first principles calculations of a real material system. The long iteration behavior of the modification factor on a log log plot is found to approximately follow the prediction made from the adiabatic analysis to follow the log(N) - log(I) behavior. Overall the results of this work present a novel algorithm for calculation of the energy density of states that in principle if one knows the number of configurations of the system only requires the choice of one compuational parameter along with a choice in number of walkers. The application of the algorithm to a materials science problem highlights its utility along with the difficult task at hand with computing the energy density of states of a disordered crystal with first principles methods. In part the calculation of the LLTO system in this work serves as a benchmark of what size systems are achievable with the current state of the art in computing resources.
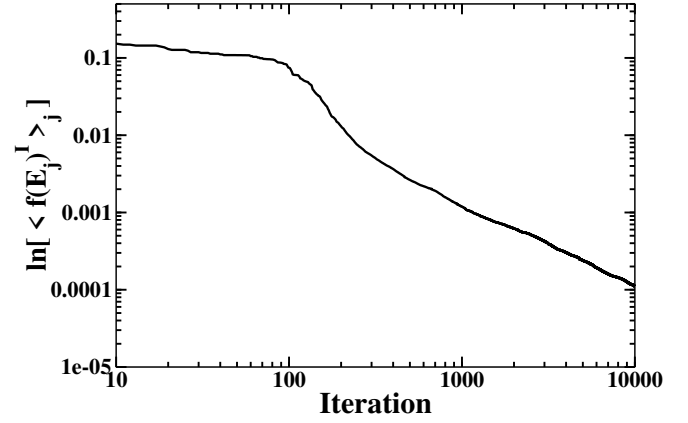

FIG. 12. Plot of the natural log of the energy bin average of the modification factor to the density of states calculated with Eq. **??** vs the iteration number.

ACKNOWLEDGMENTS