

Replica-exchange Wang–Landau sampling: pushing the limits of Monte Carlo simulations in materials sciences

Dilina Perera¹, Ying Wai Li², Markus Eisenbach², Thomas Vogel³, and David P. Landau¹

¹Center for Simulation Physics, The University of Georgia, Athens, GA 30602, USA

²National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

³Theoretical Division (T-1), Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Keywords: replica-exchange, Wang–Landau, Monte Carlo, Heisenberg model, bcc iron

Abstract

We describe the study of thermodynamics of materials using replica-exchange Wang–Landau (REWL) sampling, a generic framework for massively parallel implementations of the Wang–Landau Monte Carlo method. To evaluate the performance and scalability of the method, we investigate the magnetic phase transition in body-centered cubic (bcc) iron using the classical Heisenberg model parameterized with first principles calculations. We demonstrate that our framework leads to a significant speedup without compromising the accuracy and precision and facilitates the study of much larger systems than is possible with its serial counterpart.

Introduction

Computational methods have become indispensable for enhancing our understanding of material properties and making accurate predictions of their behavior. This can, in turn, lead to the development of improved materials with, e.g., increased strength and resistance to degradation. Even with recent advancements, first-principles based approaches to the electronic structure are restricted to the calculation of ground state properties of small systems and are incapable of predicting the system behavior at finite temperature. Instead, classical Monte Carlo (MC) simulations based on atomistic or coarse grained Hamiltonians have become a viable and powerful approach for probing the thermodynamic properties of large-scale, complex systems [1]. The Metropolis method [2] is the basis of many modern MC schemes where one probabilistically samples from the canonical (NVT) ensemble by generating a sequence of microstates according to Boltzmann weights. An inherent downside of Metropolis sampling, just as for NVT molecular dynamics simulations, is its tendency to get trapped in local free-energy minima that are frequently encountered in systems with complex free energy landscapes. To circumvent this pitfall, different methodologies have been introduced. One popular approach is to perform multiple Metropolis simulations at different temperatures and allow periodic conformational (replica) exchanges between them. This is widely known as parallel tempering or replica-exchange Monte Carlo [3, 4]. Another approach, strives to sample microstates from a flat probability distribution (the so-called multicanonical ensemble [5, 6]), where states have non-Boltzmann weights that are inversely proportional to the *a priori* unknown density of states. Such generalized-ensemble methods hope to escape free energy barriers and avoid long time scales which are encountered at low temperatures and near phase transitions. To this end, the Wang–Landau (WL) method [7–9] has emerged as a simple but powerful technique. The underlying idea behind Wang–Landau sampling is to perform a random walk in configuration space while iteratively adjusting the density of

states $g(E)$ and hence the simulation weights. This ultimately leads the accumulation of a uniform histogram in energy (E) space (or, for that matter, in any other reaction coordinate or collective variable). Eventually, the WL scheme delivers an estimator for the density of states so that the thermodynamic behavior of the system for the entire temperature range of interest can be extracted from a single simulation. The method has been successfully applied to a wide array of intricate problems in condensed matter and statistical physics including spin glasses, liquid crystals, polymers, protein folding etc. [10–13].

Multiple attempts have been made to improve the efficiency of the Wang–Landau method, either by finding more efficient trial moves, or by accelerating its convergence (see [14, 15] for examples). Nevertheless, in order to exploit the power of modern high performance computing systems and adapt the method to even larger-scale systems with increased complexity, an efficient parallelization scheme is essential. Recently, a generic framework was proposed, combining replica-exchange with Wang–Landau sampling into a massively parallel Monte Carlo simulation method [16–19]. The strategy is to split the energy range into a series of smaller, overlapping windows, which are sampled by independent random walkers with occasional replica-exchanges. Parallelism is then achieved mainly in a “divide-and-conquer” manner.

In this paper, we review this recently introduced parallelization scheme for WL simulations. To demonstrate its applicability to problems in material science, we investigate the ferromagnetic-paramagnetic phase transition in body centered cubic (bcc) iron via a classical spin model frequently used for describing magnetic materials, namely the Heisenberg ferromagnet. Furthermore, we will elaborate on the specific advantages of the parallel method and assess its scalability for large-scale simulations involving hundreds of thousands of spins, utilizing thousands of processors. This work serves as a precursor to the study of more realistic, and more complex, model Hamiltonians for bcc iron, where lattice vibrations contribute extra degrees of freedom and complicate the phase space along with the energy landscape [20]. The complexity of the problem restricts the study to small system sizes using the serial WL method or traditional parallel methods such as parallel tempering.

Model

In the Heisenberg model, atomic magnetic moments are represented as classical spin vectors of unit length. In the absence of magnetic anisotropies or external fields, the total energy for a system of interacting classical spins is given by the Hamiltonian

$$\mathcal{H} = - \sum_{i < j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j, \quad |\mathbf{S}_i| = 1 \quad \forall i, \quad (1)$$

where \mathbf{S}_i is the orientation of the i th spin, and J_{ij} is the exchange interaction between i th and j th spin. For the realistic modeling of magnetic properties of bcc iron, we choose numerical values of J_{ij} from the first principles based parameterization given in Ref. [21] which limits the interactions to the nearest and next nearest neighbors. The possible energies of the system lie within the range $[-N\epsilon_0, N\epsilon_0]$, where $\epsilon_0 \approx 0.14$ eV is the absolute value of the ground state energy per spin, and $N = 2L^3$ is the total number of spins. There are two spins in each unit cell and L is the lattice size, i.e., the number of unit cells in each direction. As our primary interest is the thermodynamic behavior near the critical temperature for the ferromagnetic–paramagnetic transition, we impose minimum and maximum energy cutoffs and sample configurations within a restricted global energy range, $[E_{\min}, E_{\max}]$. This reduction drastically reduces the computational cost of sampling rare configurations at extremely

low and high energies, which potentially becomes a bottleneck in serial WL simulations. The reduced energy range is still large enough though that the random walkers are able to visit all contributing microstates; therefore, this reduction does not introduce any systematic errors.

The Replica-Exchange framework for Wang–Landau sampling

Background in statistical mechanics

The density of states in energy, $g(E)$, measures the energy degeneracy of admissible states of a system, from which the partition function Z can be calculated:

$$Z(T) = \sum_{\mathbf{X}} e^{-E[\mathbf{X}]/k_B T} = \sum_E g(E) e^{-E/k_B T}, \quad (2)$$

where \mathbf{X} stands for a state or configuration that the system can reside in; k_B is the Boltzmann constant and T is the temperature. The first sum runs over all possible states of the system, whereas the second sum runs over all possible total energies and is calculable once $g(E)$ is known. While $g(E)$ is temperature independent and only depends on the definition of the Hamiltonian, Eq. (2) allows for the calculation of the temperature dependent Z via the corresponding Boltzmann factors. An important consequence is the possibility of calculating the thermodynamic quantities at *any* temperature with the sole knowledge of $g(E)$. For example, the average energy $\langle E \rangle$ and the heat capacity C_V can be calculated:

$$\langle E \rangle(T) = \frac{1}{Z} \sum_E g(E) E e^{-E/k_B T}, \quad C_V(T) = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2}. \quad (3)$$

The specific heat is defined as C_V/N . These thermodynamic observables provide a measure to identify and locate phase transitions and hence understand critical phenomena. For the Heisenberg ferromagnetic iron model we use in this study, we successfully capture the Curie temperature which corresponds to the ferromagnetic-paramagnetic transition.

The original Wang–Landau algorithm

The original Wang–Landau sampling [7–9] is classified as one of the generalized-ensemble MC methods where the final outcome of a simulation is an estimation of $g(E)$. Depending on the parameter space in which the random walk is performed, the density of states could as well be an estimate of the degeneracy of states as a function of other order parameters or reaction coordinates or even a higher-dimensional combination of such, see [22] for an example.

At the beginning of each Wang–Landau simulation the desired total energy range, $E \in [E_{\min}, E_{\max}]$, for which $g(E)$ should be obtained is determined or estimated. For a model with continuous energy domain an energy bin width δE is also chosen to control the resolution of $g(E)$. An initial guess of $\tilde{g}(E) = 1$ for all energies is used as a starting point, although other choices are suitable as well.¹ A histogram, $H(E)$, is introduced to keep track of the number of visits to each energy; all entries are set to zero at the beginning. The simulation starts with an arbitrary initial configuration, and new states B are generated by applying a Monte Carlo trial move (see below) to the current configuration A with energy E_A . The

¹Here, $\tilde{g}(E)$ denotes the instantaneous estimator for $g(E)$ which changes during the course of the simulation.

energy E_B of the trial state is then evaluated and the new state will be accepted according to the acceptance probability:

$$P(A \rightarrow B) = \min \left[1, \frac{\tilde{g}(E_A)}{\tilde{g}(E_B)} \right]. \quad (4)$$

If trial state B is accepted, $\tilde{g}(E)$ and $H(E)$ will be updated by $\tilde{g}(E_B) \rightarrow \tilde{g}(E_B) \times f$ and $H(E_B) \rightarrow H(E_B) + 1$. Otherwise, the update is done for the current state A . Here, f is the modification factor whose initial value is $f_0 = e^1$. Trial states are continuously generated and $\tilde{g}(E)$ and $H(E)$ are updated, until a “flat” histogram is achieved, i.e. all $H(E)$ entries are no less than a certain fraction p of its average value $\bar{H}(E)$ ($H(E) \geq p\bar{H}(E) \forall E$). At this point, the modification factor is reduced via $f \rightarrow \sqrt{f}$, the histogram is reset ($H(E) = 0$) but $\tilde{g}(E)$ is carried over to the next iteration. This procedure is repeated until f reaches a predefined, terminal value f_{\min} . Hence, $\tilde{g}(E)$ is modified by smaller and smaller amounts and eventually converges (asymptotically) to the true density of states $g(E)$.²

The parallel scheme

The recently introduced parallel implementation of Wang–Landau sampling [16, 17] incorporates the original, serial Wang–Landau algorithm into a replica-exchange framework. The global energy range $[E_{\min}, E_{\max}]$ is broken into a number (h) of smaller windows, each of which overlaps with its nearest neighbors on both sides with an overlapping ratio o (a schematic diagram is shown in the upper panel of Fig. 2). Multiple (m) Wang–Landau walkers can be employed in each of the windows to reduce statistical errors during the simulation [17], but each walker has its own $\tilde{g}(E)$ and $H(E)$ and all of them must proceed to the next iteration independently. Replica exchanges are proposed between walkers i and j from adjacent windows after a predefined number of MC sweeps (MCS) and are accepted according to the probability:

$$P_{\text{RE}} = \min \left[1, \frac{\tilde{g}_i(E_A)\tilde{g}_j(E_B)}{\tilde{g}_i(E_B)\tilde{g}_j(E_A)} \right], \quad (5)$$

where we have assumed that A (or B) is the current configuration of walker i (or j).

The simulation produces multiple, overlapping fragments of $g(E)$. They are joined at points where the slopes of $\ln g(E)$ (i.e., $d \ln g(E)/dE$, the inverse microcanonical temperature) best coincide. This practice reduces the introduction of artificial kinks in the combined $g(E)$ due to the joining process and eliminates resultant errors in thermodynamic quantities. See Refs. [18, 19] for illustrations.

Simulation details

To compare results for the parallel and serial WL schemes, we chose a relatively small system size ($L = 10$) and a global energy range ($-260 \text{ eV} \leq E \leq 120 \text{ eV}$) that are accessible by the serial method within a fairly reasonable time. For checking the convergence of $g(E)$, we used the parameter values from the original WL paper, an 80% flatness criterion and final modification factor of $\ln f_{\min} = 1 \times 10^{-8}$. For the trial updates we used the simplest MC trial move, i.e., randomly choose a spin and assign a arbitrary new direction.

We decided that eleven windows ($h = 11$) and replica exchanges between neighboring windows proposed every 500 MCS (one MCS is equivalent to N MC trial moves) are rea-

²Convergence proofs can be found in Ref. [14] and, for a directly related method [23], in Ref. [24].

sonable choices for the lattice size $L = 10$. Since replica exchanges can, by construction, only be accepted if both corresponding replica are in the energy overlap region between two windows, small overlaps lead to low acceptance rates whereas unnecessarily large overlap downgrades the performance. We found that a moderately large overlap of $o \approx 75\%$ can be a good choice for maintaining a balance between fast convergence and reasonable acceptance rates [17]. With this choice of the overlap, we observed acceptance rates in the range of 52 – 55% for replica exchanges.

Results and discussion: Accuracy and scalability

With those chosen values of $h=11$ and $o=75\%$, we performed parallel simulations for $L=10$ while employing a single walker per window ($m=1$) for simplicity. Fig. 1 shows the time series of a single replica as it performs smooth round-trips across the entire energy range (a) and through all the energy windows (b). Similar behavior was observed for all replicas, verifying that the random walkers are not restricted to certain regions of the energy space.

Fig. 2 (a) compares the density of states obtained from both parallel and serial runs. Data shown by filled dots were obtained from a single parallel simulation, the solid line represents the average of 10 independent serial runs, the standard deviation σ is shown in the inset. The absolute difference Δ between the results from the serial runs and the parallel run is also shown in the inset for comparison. Δ is of the same size as σ . Thus, the precision of the parallel WL results are on par with those from conventional, serial runs. To accommodate the increase in the global energy range for much larger systems, we increased the number of windows accordingly while keeping the energy window size ($\Delta E \approx 108.57$ eV) and the overlap ($o = 75\%$) constant. For example, for $L = 40$ (-16640 eV $\leq E \leq 7680$ eV), 893 windows were required. While a serial WL simulation would take years to estimate the density of states, the parallel scheme achieved this within two days. Fig. 2 (b) shows the specific heat derived from the estimated density of states for the system sizes $L = 10$ and $L = 40$. Despite the simplicity of our model, the peak positions match reasonably well with the experimental Curie temperature $T_C = 1043$ K [25]. For comparison, we also show the results obtained from the serial WL simulations for $L = 10$. The data from the serial and parallel runs are within the mutual error bars.

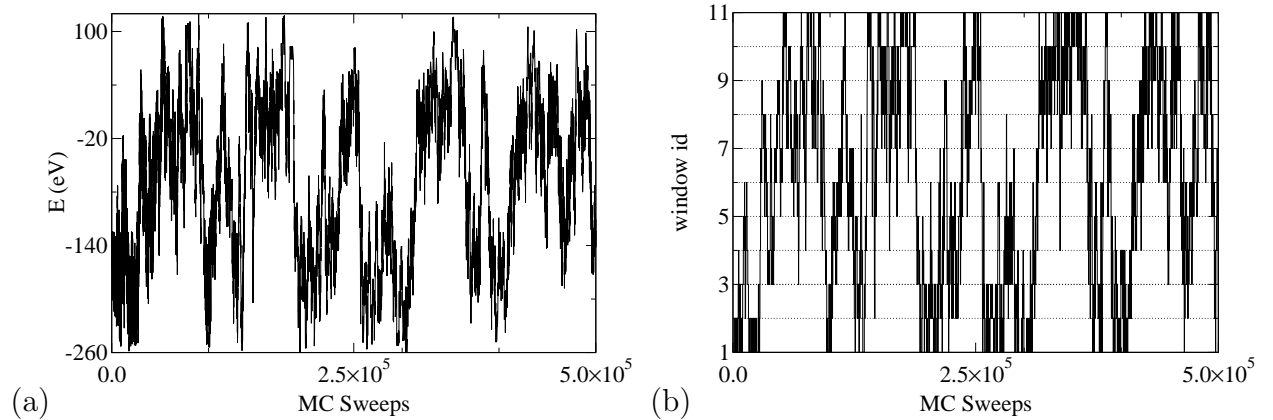


Figure 1: Path of a single replica through the entire energy space (a) and energy windows (b) during the first 5×10^5 MCS. Replica exchanges are attempted every 500 MCS, acceptance rates are in the range of 52 – 55%. The replica completes a round-trip approximately every 7×10^4 MCS.

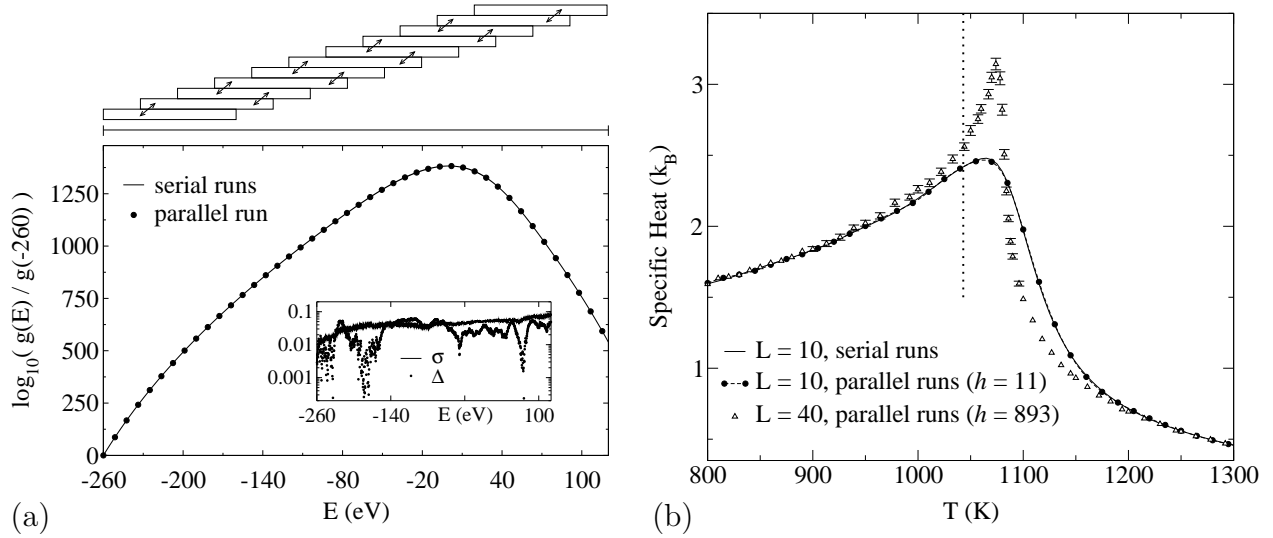


Figure 2: (a) Logarithm of the density of states for the lattice size $L = 10$. The solid line represents the average of 10 independent serial runs, filled dots show results from a single parallel run. The inset (note the logarithmic scale) illustrates the accuracy of the parallel method. Solid line in the inset shows the standard deviation σ of the serial results while the dots show the absolute numerical difference Δ between the serial and parallel results. (b) Specific heat curves as functions of temperature for $L = 10$ and $L = 40$. The dotted line marks the experimental Curie temperature, $T_C = 1043$ K. For $L = 10$, the results from the serial runs are shown for comparison.

The success of any parallel scheme depends not only on its accuracy and precision, but also on the scalability, i.e., the ability to effectively utilize an increasing number of processors. Ultimately, by increasing the number of computing cores one hopes to achieve two types of scaling behavior: to speed up the execution while fixing the problem size (strong scaling), or to increase the problem size without increasing execution time (weak scaling). Here, we will demonstrate the weak scaling behavior of our method using the Heisenberg model as a test case. For this analysis, we performed REWL simulations for lattice sizes $10 \leq L \leq 50$, (corresponding to spin numbers $2000 \leq N \leq 250000$ respectively), and measured the number of MC trial moves needed to complete the first Wang–Landau iteration. The results of this study are summarized in Fig. 3, where we compare the simulation time for serial runs (triangles) with the parallel performance (dots). For the serial runs, simulation time increases dramatically with system size whereas for the parallel runs it remains almost constant.

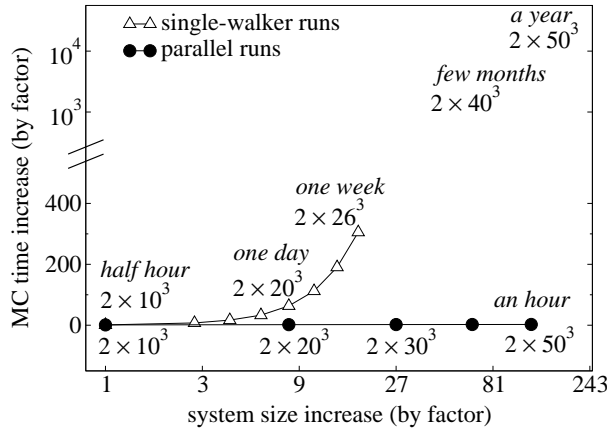


Figure 3: Relative increase in simulation time versus the system size increase. Filled dots are for parallel runs while the open triangles represent serial runs. Simulation time is determined by the time needed to complete the first WL iteration. The number of energy windows h were increased with the increasing system size while keeping the size of the windows and the overlap fixed.

Summary

The successful application of replica-exchange Wang–Landau (REWL) sampling to a Heisenberg bcc iron model in the study of magnetic phase transition proves its immense potential in obtaining finite temperature properties of materials via classical Monte Carlo simulations. The excellent agreement between REWL and the original, serial WL results, along with the nearly perfect weak scaling with the number of cores used, means that heretofore inaccessible large scale studies are now possible. With a model correctly parametrized by first principles calculations (using a relatively small system size), we can bridge different length scales between first principles calculations (typically in nanoscale) and model simulations (mesoscale or beyond) in computational physics/materials sciences. This will eventually enable the study of finite temperature properties of materials having the relevant length scales needed for real world applications.

Acknowledgements We wish to thank T. Wüst for stimulating discussions. This research was sponsored in part by the U.S. Department of Energy, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, “Center for Defect Physics”, an Energy Frontier Research Center. Y.W.L. and M.E. were partly sponsored by the Office of Advanced Scientific Computing Research of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Part of the work was performed at the Oak Ridge Leadership Computing Facility at ORNL, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. LA-UR-14-27450 assigned.

References

- [1] D.P. Landau and K. Binder, *A guide to Monte Carlo simulations in statistical physics*, 3rd edn. (New York, NY: Cambridge University Press, 2009).
- [2] N. Metropolis et al., “Equations of state calculations by fast computing machines,” *J. Chem. Phys.*, 21 (1953), 1087.
- [3] C.J. Geyer, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. E.M. Keramidas (Fairfax Station, VA: Interface Foundation, 1991), 156.
- [4] K. Hukushima and K. Nemoto, “Exchange Monte Carlo Method and Application to Spin Glass Simulations,” *Journal of the Physical Society of Japan*, 65 (1996), 1604.
- [5] B.A. Berg and T. Neuhaus, “Multicanonical algorithms for first order phase transitions,” *Phys. Lett. B*, 267 (1991), 249.
- [6] B.A. Berg and T. Neuhaus, “Multicanonical ensemble: A new approach to simulate first-order phase transitions,” *Phys. Rev. Lett.*, 68 (1992), 9.
- [7] F. Wang and D.P. Landau, “Efficient, Multiple-Range Random walk algorithm to calculate the density of states,” *Phys. Rev. Lett.*, 86 (2001), 2050.
- [8] F. Wang and D.P. Landau, “Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram,” *Phys. Rev. E*, 64 (2001), 056101.

- [9] D.P. Landau, S.-H. Tsai, and M. Exler, “A new approach to Monte Carlo simulations in statistical physics: Wang–Landau sampling,” *American J. Phys.*, 72 (2004), 1294.
- [10] S. Alder et al., “Dynamics of the Wang–Landau algorithm and complexity of rare events for the three-dimensional bimodal Ising spin glass,” *Journal of statistical mechanics: Theory and experiment*, 2004 (2004), P07008.
- [11] E.B. Kim, R. Faller, and J.J. de Pablo, “Potential of mean force between a spherical particle suspended in a nematic liquid crystal and a substrate,” *J. Chem. Phys.*, 117 (2002), 7781.
- [12] M.P. Taylor, W. Paul, and K. Binder, “Phase transitions of a single polymer chain: A Wang–Landau simulation study,” *J. Chem. Phys.*, 131 (2009), 114907.
- [13] N. Rathore and J.J. de Pablo, “Monte Carlo simulation of proteins through a random walk in energy space,” *J. Chem. Phys.*, 116 (2002), 7225.
- [14] C. Zhou and R. Bhatt, “Understanding and improving the Wang–Landau algorithm,” *Phys. Rev. E*, 72 (2005), 025701(R).
- [15] R. Belardinelli and V. Pereyra, “Fast algorithm to calculate density of states,” *Phys. Rev. E*, 75 (2007), 046701.
- [16] T. Vogel et al., “A generic, hierarchical framework for massively parallel Wang–Landau sampling,” *Phys. Rev. Lett.*, 110 (2013), 210603.
- [17] T. Vogel et al., “Scalable replica-exchange framework for Wang–Landau sampling,” *Phys. Rev. E*, 90 (2014), 023302.
- [18] T. Vogel et al., “Exploring new frontiers in statistical physics with a new, parallel Wang–Landau framework,” *J. Phys.: Conf. Series*, 487 (2014), 012001.
- [19] Y.W. Li et al., “A new paradigm for petascale Monte Carlo simulation: Replica exchange Wang-Landau sampling,” *J. Phys.: Conf. Series*, 510 (2014), 012012.
- [20] J. Yin et al., “Effect of lattice vibrations on magnetic phase transition in bcc iron,” *Phys. Rev. B*, 86 (2012), 214423.
- [21] P.-W. Ma, C.H. Woo, and S.L. Dudarev, “Large-scale simulation of the spin-lattice dynamics in ferromagnetic iron,” *Phys. Rev. B*, 78 (2008), 024434.
- [22] C. Zhou et al., “Wang–Landau Algorithm for Continuous Models and Joint Density of States,” *Phys. Rev. Lett.*, 96 (2006), 120201.
- [23] C. Junghans, D. Perez, and T. Vogel, “Molecular Dynamics in the Multicanonical Ensemble: Equivalence of Wang-Landau Sampling, Statistical Temperature Molecular Dynamics, and Metadynamics,” *J. Chem. Theory Comput. (JCTC)*, 10 (2014), 1843.
- [24] J. Dama, M. Parrinello, and G.A. Voth, “Well-Tempered Metadynamics Converges Asymptotically,” *Phys. Rev. Lett.*, 112 (2014), 240602.
- [25] E.P. Wohlfarth, *Ferromagnetic Materials* (Amsterdam, The Netherlands: North-Holland Publishing Company, 1980).