Thermodynamics of magnetic systems from first principles: WL-LSMS

M. Eisenbach,[1] C.-G. Zhou,[1*] D. M. Nicholson,[1] G. Brown,[2] T. C. Schulthess[3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA
[2]Department of Physics, Florida State University, Tallahassee, Florida 32306, USA
[3]Institute for Theoretical Physics and Swiss National Supercomputer Center, ETH Zürich, 8903 Zürich, Switzerland

Abstract. Density Functional calculations have proven to be a powerful tool to study the ground state of many materials. For finite temperatures the situation is less ideal and one is often forced to rely on models with parameters either fitted to zero temperature first principles calculations or experimental results. This approach is especially unsatisfactory in inhomogeneous systems, nano particles, or other systems where the model parameters could vary significantly from one site to another. Here we describe a possible solution to this problem by combining classical Monte Carlo calculations—the Wang-Landau method [2] in this case—with a first principles electronic structure calculation, specifically our locally self-consistent multiple scattering code (LSMS) [3]. The combined code shows superb scaling behavior on massively parallel computers. The code sustained 1.836 Petaflop/s on 223,232 cores of the Cray XT5 jaguar system at Oak Ridge.

1. Introduction

Density Functional based first principles electronic structure calculations for condensed matter systems have reached a high level of maturity over the last few decades and are now a standard tool for the study of ground state material properties [1]. While these methods have evolved to provide greater accuracy and deal with wider classes of materials, the field of finite temperature behavior has received less attention. The phase space usually is far too large to be dealt with directly. The usual methods of treating the thermodynamics of a physical system involve either the time evolution of an ensamble or the exploration of the most relevant parts of phase space by means of a Monte-Carlo method. Both these approaches require a large number of evaluations of the underlying Hamiltonian that describes the system ($> O(10^5)$), thus it is usually only feasible to treat severely simplified models that have to be designed to capture the essential physics, as opposed to a direct treatment of the Density-Functional Hamiltonian of the system.

To overcome this limitation we have developed the hybrid Wang-Landau/LSMS (WL-LSMS) code. This code combines revcent advances in computational statistical mechanics, namely the Wang-Landau method [2] with the LSMS first principles method that has already demonstrated superb scalability on massively parallel machines.

[*] Present address: Quantitative Research, J. P. Morgan Chase and Company, New York, New York 10017, USA.

2. Structure of WL-LSMS

The WL-LSMS code uses a multi level parallelization scheme. At the top level, the code parallelizes over concurrent random walkers, where we use a master-slave scheme, with a master that accumulates the density of states of the system, and the slaves that execute the random walks, each running its own instance of the LSMS method. The second parallelization level is the LSMS portion of the code. In typical production runs, the WL method would use a hundred to a few thousand concurrent walkers, and the LSMS portion would be parallelized over up to a few thousand processing cores. The method hence will scale to hundred thousand or millions of processing cores. The schematics of the parallelization structure are shown in Figure 1.

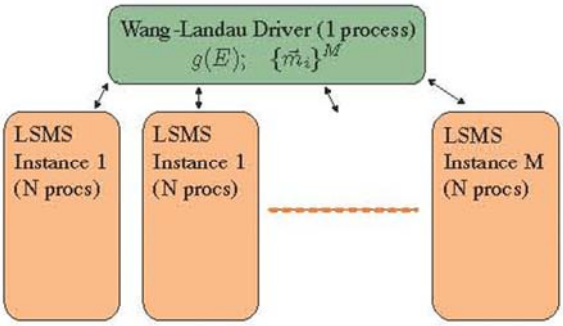

Figure 1. Parallelization strategy of the combined Wang-Landau/LSMS algorithm. The Wang-Landau process generates random spin configurations for M walkers and updates a single density of states $g(E)$. The energies for these N atom systems are calculated by independent LSMS processes (Figure 2). This results in two levels of communication, between the Wang-Landau driver and the LSMS instances, and the internal communication inside the individual LSMS instances spanning N processes each.

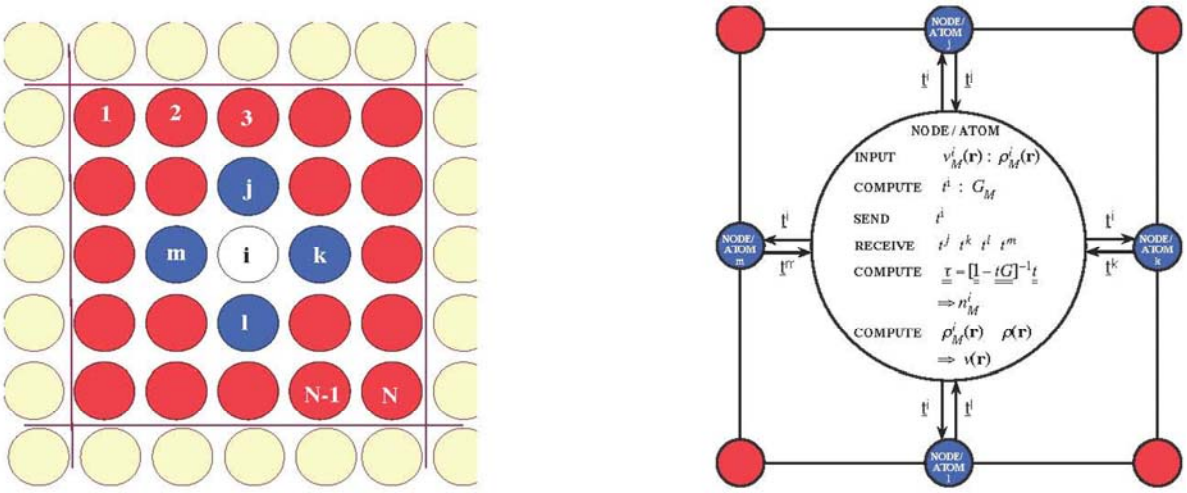

Figure 2. Schematic, Left: LIZ centered at processor/atom i; Right: message passing and computation.

3. The Wang-Landau Algorithm

All thermodynamic potentials can be derived from the partition function

$$Z(T) = \int e^{-E(\boldsymbol{X})/(k_B T)} d\boldsymbol{X} \tag{1}$$

where $E(\boldsymbol{X})$ is the internal energy of the system with the phase space described by the variable $\boldsymbol{X}$ in some high dimensional space consisting of all the microscopic degrees of freedom of the system (atomic positions, velocities, and/or magnetic moments). In importance sampling Monte Carlo simulations one performs a random walk through phase space that is biased in such a way that the walker spends most of the time where the integrand in equation (1) is largest, that is, where the energy $E(\boldsymbol{X})$ is small.

The partition function in equation (1) can be rewritten in the form

$$Z(T) = \int g(E) e^{-E/(k_B T)} dE \tag{2}$$

where the density of states is defined as

$$g(E) = \int \delta(E - E(X)) dX \tag{3}$$

and $\delta(E)$ is the Dirac δ-function.

Flat histogram methods, such as the Wang-Landau algorithm, use the density of states, $g(E)$, for importance sampling, thus accepting the new configuration with probability

$$\min[1, g(E_i)/g(E_{i+1})]. \tag{4}$$

The effect is to create an equal probability of visiting each energy level in the system.

The main obstacle of flat-histogram methods is that $g(E)$ is not known. Instead, an estimate of the density of states $\tilde{g}(E)$ must be constructed self-consistently as the Monte Carlo estimate is generated. The Wang-Landau algorithm accomplishes this as follows. It begins with a prior estimate of the density of states, $\tilde{g}_0(E)$, which might be just a constant. Assuming that a Monte Carlo move to a new configuration with Energy E_{i+1} is accepted according to the criterion of equation (4), the density of states is updated with

$$\ln[\tilde{g}(E_{i+1})] \leftarrow \ln[\tilde{g}(E_{i+1})] + \ln f \tag{5}$$

where f is the modification factor that is initially set to $\ln f = 1$. The modification factor is reduced in steps such that $\ln(f) \leftarrow \ln(f)/2$ and the density of states converges as $\ln(f) \to 0$.

Since the systems we set out to study here have continuous degrees of freedom, $g(E)$ is a function of a continuous variable as well. Hence, we employ the kernel update scheme described by Zhou et al. [8].

4. The LSMS Algorithm

For the energy evaluation, we employ the first principles framework of density functional theory (DFT) in in the local spin density approximation (LSDA). To solve the Kohn-Sham equations arising in this context, we use a real space implementation of the multiple scattering formalism. The details of this method for calculating the Green function and the total ground state energy $E[n(\vec{r}), \vec{m}(\vec{r})]$ are described elsewhere [3,4]. For the present discussion it is important to note that the computationally most intensive part is the calculation of the scattering path matrix τ for each atom in the system by inverting the multiple scattering matrix.

$$\tau = [I - tG_0]^{-1} t$$

The only part of τ that will be required in the subsequent calculation of site diagonal observables (i.e., magnetic moments, charge densities, and total energy) is a small (typically 32×32) diagonal block of this matrix. This will allow us to employ the algorithm described in the next section for maximum utilization of the on node floating point compute capabilities.

Most importantly for the application in the hybrid Wang-Landau LSMS method, our Locally Self-consistent Multiple Scattering (LSMS) method allows the possibility of non-collinear magnetism [5].

The orientation $\hat{e}^i$ of the magnetic moment for each site is determined by

$$\hat{e}^i = \int_{\Omega^i} d\vec{r}\, \vec{m}^i(\vec{r}) / |\int_{\Omega^i} d\vec{r}\, \vec{m}^i(\vec{r})|.$$

Since an arbitrary arrangement is not a DFT ground state we will have to deal with a constrained general state as presented by Stocks et al. [6,7]. In the constrained local moment (CLM) model the LSDA equations are solved subject to a constraint that ensures that the local magnetizations lie along the directions prescribed by $\{\vec{e}_i\}$. Thus this method enables the calculation of the energies of arbitrary orientational states as generated by the Wang-Landau algoritm.

5. Blockinversion in LSMS

The most time consuming part of the LSMS calculation is the inversion of the multiple scattering matrix. The amount of computational effort can be reduced by utilizing the fact that for each local interaction zone only the left upper block (τ_{00}) of the scattering path matrix τ is required. In this section we describe an algorithm that reduces the amount of work needed while providing excellent performance due to its reliance on dense matrix-matrix multiplications that are available in highly optimized form in vendor or third party provided implementations (i.e., ZGEMM in the BLAS library).

The method employed in LSMS to calculate the required block of the inverse relies on the well known expression for writing the invers of a matrix in term of inverses and products of subblocks:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} U & V \\ W & Y \end{pmatrix}$$

where

$$U = (A - BD^{-1}C)^{-1}$$

and similar expressions for V, W, and Y. This this method can be applied multiple times to the subblock U until the desired block τ_{00} of the scattering path matrix is obtained.

6. Performance

We analysed the performance of the code for systems consisting of both 250 and 1024 Fe atoms respectively. For these systems we study the scaling properties of the code as a function of number of walkers used in the WL simulation. Every individual LSMS calculation per walker can be distributed onto 250 or 1024 cores. On the Cray XT5 *jaguarpf* system at ORNL's National Center for Computational Sciences (NCCS), we can thus scale these calculations to up to 895 parallel WL walkers for 250 atoms on 223,752 cores and 218 walkers on 223,232 cores for 1024 Fe atoms. In these performance analysis runs, each walker executes 20 WL steps, which is far fewer than a real simulations.

In Figure 3 we show how the time to solution scales if we increase the number of WL walkers, and thus the total number of samples taken in the WL-LSMS simulation. The result shown thus represent a weak scaling plot, and the scaling behavior of the WL-LSMS method looks close to optimal. We find a similarly optimal strong scaling behavior, if we fix the number of samples taken for every run and increase the number of walkers. With the available size of machines today, we are still far from saturating the method in terms of scaling behavior.

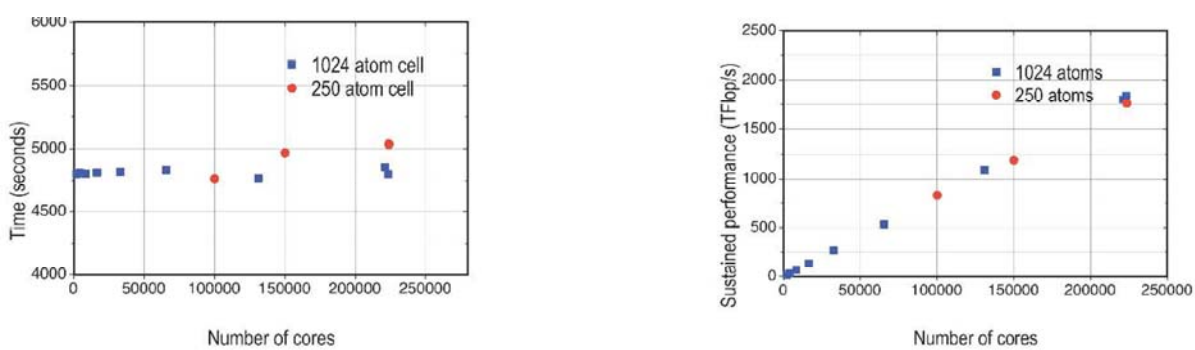

Figure 3. (left) The weak scaling behavior of the WL-LSMS code for 250 and 1024 atom systems with a varying number of Wang-Landau walkers. The times shown represent the total runtime of the code and include the startup costs of the calculations. This accounts for the jump in the runtime for the 250 atom systems as each Wang-Landau walker reads its initial input file. (right) The sustained performance of the Wang-Landau LSMS code on jaguarpf for systems of 250 iron atoms (blue squares) and 1024 atoms (red circles). The code reaches a performance of 1.755 Petaflop/s on 223,752 cores for 250 atoms and 1.835 Petaflop/s on 223,232 cores for a 1024 atom system.

The sustained floating point performance of the runs that correspond are shown also in Figure 3. In order to measure the executed floating point operations of the benchmark runs, we have instrumented the WL-LSMS code with PAPI calls. rFor the largest runs with 218 parallel Wang-Landau walkers of 1024 atoms each and 20 steps per walker, the measured sustained performances was 1.835 petaflop/s, which on the 223,232 AMD Opteron cores running at 2.6 GHz corresponds to a fraction of 79.0% of the theoretical peak performance.

7. Magnetic transition temperature for iron

Here we present a calculation of the magnetic transition temperature (Curie temperature) of iron. Utilizing the methods and algorithm described above we consider a periodically repeated cells of 250 iron atoms and converge the Wang-Landau density of states $g(E)$ for changes in the magnetization direction on the individual iron sites. For the underlying LSMS calculation of these iron cells, the atoms were placed on a body-centered cubic lattice with a lattice parameter of $5.42a_0$, corresponding to the experimental value, and the local interaction zone has a radius of $11.5a_0$. The self-consistently converged potential for the ferromagnetic ground state was used for all the individual frozen-potential energy calculations in the combined Wang-Landau/LSMS algorithm.

Using 400 Wang-Landau walkers the calculation converged in $590{,}000$ and required $4{,}885{,}720$ CPU hours on the jaguar Cray XT5 system. The resulting density of states is shown in Figure 4 and Figure 5. While the computational resources needed for a system with several hundred atoms are considerable, the remaining calculations to compute any desired temperature dependent thermodynamic properties is marginal.

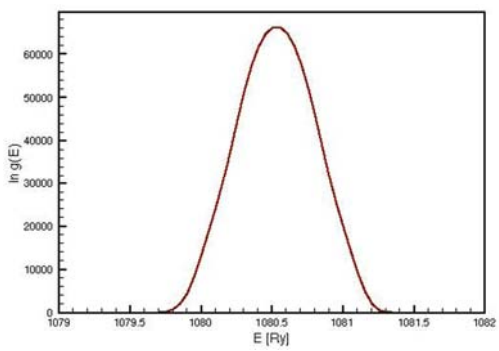

Figure 4. The unnormalized logarithmic Wang-Landau density of states $\ln g(E)$ for a periodic system of 250 iron atoms, respectively.

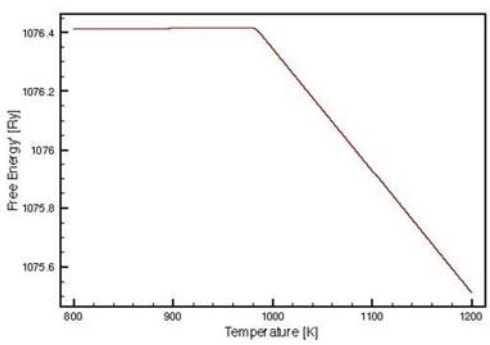

Figure 5. Free Energy F' for a system of 250 iron atoms.

With this density of states the partition function that describes the thermodynamics of the system can be calculated for any temperature.

These calculation yields a transition temperature of 980K. This result is in remarkably good agreement with the experimentally known Curie temperature of bulk iron of 1050K.

8. Summary

In the present paper we have demonstrated our approach for the first principles treatment of finite temperature behavior of magnetic systems. The combination of the most recent massively parallel supercomputing architectures and advances in both algorithmic developments and most importantly new computational methods have made this hybrid statistical mechanics/first principles method feasible.

The code presently is applicable to the evaluation of magnetic transition temperatures of transition metal alloys and has already reproduced the Curie temperature of bulk Iron [9].

Acknowledgments

We acknowledge use of resources of the National Center for Computational Sciences and the Center for Nanophase Materials Sciences, which are sponsored by the respective facilities divisions of the offices of Advanced Scientific Computing Research and Basic Energy Sciences (BES) of the U.S. Department of Energy. The present research was supported by the Laboratory Directed Research and Development fund of ORNL, as well as the division of scientific user facilities of BES. Computations were performed at the Oak Ridge Leadership Computing Facility under an INCITE allocation.

References

[1] Martin, R. M., *Electronic Structure: Basic Theory and Practical Methods*, Cambridge, 2004.

[2] F. Wang, D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).

[3] Yang Wang, G. M. Stocks, W. A. Shelton, D. M. C. Nicholson, Z. Szotek, and W. M. Temmerman, *Phys. Rev. Lett.* **75** 2867 (1995).

[4] M. Eisenbach, B. L. Györffy, G. M. Stocks, and B. Újfalussy, *Phys. Rev. B* **65**, 144424 (2002).

[5] G. M. Stocks, Y. Wang, D. M. C. Nicholson, W. A. Shelton, Z. Szotek, W. M. Temmerman, B. N. Harmon, V. P. Antropov, *Mater. Res. Soc. Symp. Proc.* **408**, 157 (1996).

[6] G. M. Stocks, B. Újfalussy, X. Wang, D.M.C. Nicholson, W. A. Shelton, Y. Wang, A. Canning, and B. L. Györffy, *Philos. Mag. B* **78**, 665 (1998).

[7] B. Újfalussy, X. Wang, D.M.C. Nicholson, W. A. Shelton, G. M. Stocks, Y. Wang, and B. L. Györffy, *J. Appl. Phys.* **85**, 4824 (1999).

[8] C.-G. Zhou, T. C. Schulthess, S. Torbrügge, D. P. Landau, *Phys. Rev. Lett.* **96**, 120201 (2006).

[9] M. Eisenbach, C.-G. Zhou, D. M. Nicholson, G. Brown, J. Larkin, T. C. Schulthess, SC'09: Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis, ACM, 2009.