8-2022

Direct Calculation of Configurational Entropy: Pair Correlation Functions and Disorder

Clifton C. Sluss
University of Tennessee, Knoxville, csluss@vols.utk.edu

Recommended Citation

To the Graduate Council:

I am submitting herewith a dissertation written by Clifton C. Sluss entitled "Direct Calculation of Configurational Entropy: Pair Correlation Functions and Disorder." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Materials Science and Engineering.

David J. Keffer, Major Professor

We have read this dissertation and recommend its acceptance:

Don M. Nicholson, Haikuan Xu, Orlando Rios

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Direct Calculation of Configurational Entropy: Pair Correlation Functions and Disorder

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Clifton Sluss

August 2022

ACKNOWLEDGEMENTS

ABSTRACT

Techniques such as classical molecular dynamics [MD] simulation provide ready access to the thermodynamic data of model material systems. However, the calculation of the Helmholtz and Gibbs free energies remains a difficult task due to the tedious nature of extracting accurate values of the excess entropy from MD simulation data. Thermodynamic integration, a common technique for the calculation of entropy requires numerous simulations across a range of temperatures. Alternative approaches to the direct calculation of entropy based on functionals of pair correlation functions [PCF] have been developed over the years. This work builds upon the functional approach tradition by extending the recently developed entropy pair functional theory [EPFT] to three new material systems. Direct calculations of entropy for the BCC iron and FCC copper (modeled with the modified embedded atom method [MEAM] potential) and the Diamond Cubic silicon system (modeled with the Tersoff potential) are compared against a target entropy as determined by thermodynamic integration. The sources of and correction to the high temperature error in several proposed functional approaches is explored in depth. Finally, a working code is provided to the community via Github to implement the extended EFPT to compute entropy using trajectory files generated from a single simulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Motivation

The calculation of entropy from classical molecular dynamics simulations is a time consuming and tedious process. The most common technique among practitioners, thermodynamic integration, requires a suite of simulations covering a range of temperatures sufficient to calculate entropy at a single temperature of interest. The entropy of a solid or liquid system consists of the contribution of entropy of the system as an ideal gas and what is referred to as the excess entropy. The entropy of the ideal gas of a system can be directly calculated from other easily obtainable thermodynamic values by way of the Sackur-Tetrode equation. It is the excess entropy the presents a challenge to the practitioner. The well-known equation for change in entropy is shown below.

$$\Delta S = \int_{T_0}^{T_1} \frac{C_V}{T} dT \tag{1.1}$$

This equation only provides the difference in entropy between two temperatures. In order to obtain the absolute value of entropy at T1, a ΔS must be calculated and referenced from a known absolute value of entropy at T0. In a case where no absolute entropy value is available for a system one must be obtained for the ideal gas and the thermodynamic integration must be performed across a range of temperatures bounded by the temperature of interest and a temperature sufficiently high to ensure that the assumptions of the ideal gas hold. For practical materials this temperature range will be quite large, on the order of 1 to 10 million K. For the modeler in classical molecular dynamics this will require a large set of time consuming simulations for a single entropy value.

Background

Classical Molecular Dynamics Simulations

Classical Molecular Dynamics (MD) simulation is a computer modeling technique based on classical mechanics that is useful for the modelling of materials at the atomic or molecular level. Classical MD simulations solve Newton's equations of motion to determine the positional configuration of a system of atoms or molecules. These positions are found for a series of incremental time steps defined by the user. Classical MD simulations rely on an interaction potential to determine the forces present between the particles in the system. All MD simulations in this proposal refer to classical MD simulations.

The advantage of classical MD simulation over quantum mechanical simulation techniques, such as ab initio MD (AIMD) lies in computational efficiency. Where AIMD simulations are generally limited to on the order of 10^3 atoms, MD simulations are routinely performed for 10^6 atoms. The disadvantage of classical MD is that the interaction potential is an input. If the interaction potential is not well suited (or well parameterized) to a particular property of interest, then the result will be unreliable, per the garbage-in/garbage-out principle.

In principle, the work presented here is equally applicable to either classical MD or AIMD simulation, since it relies on structural properties extracted from trajectories. However, in practice, we find that the characterization of structure required to reliably evaluate the entropy requires system sizes beyond that accessible to AIMD. Thus in this work, the exploration of entropy from atomic simulation is limited to classical MD.

The Interaction Potential

One of the simplest interaction potentials is the Lennard-Jones interaction potential shown in Figure 1-1. This interaction potential is the combination of a repulsive
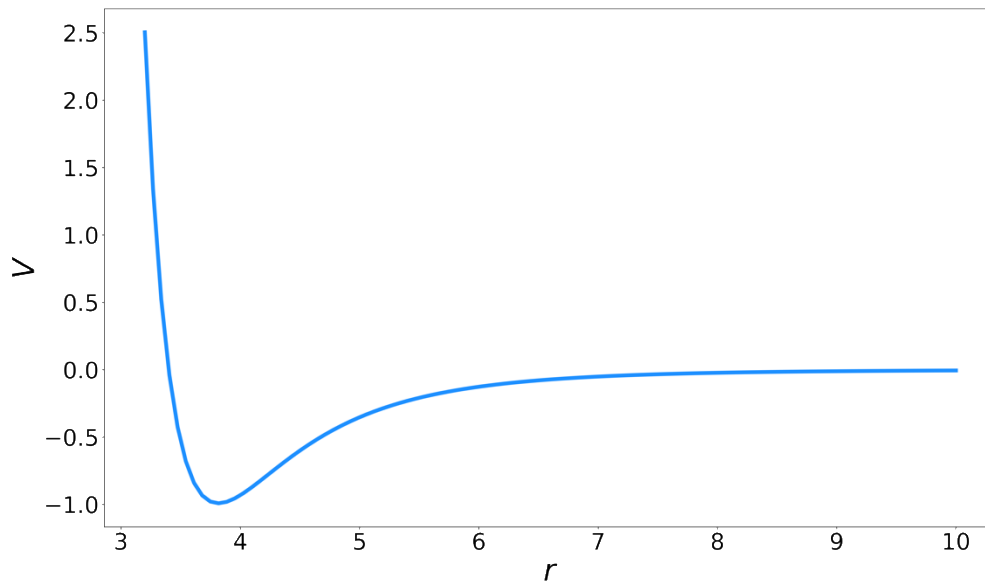
Figure 1-1. Lennard-Jones interaction potential.

component ($\sim 1/r^{12}$) and an attractive component, intended to capture the classically induced-dipole/induced dipole energy, $(1/r^6)$. When superimposed together these two components form the characteristic shape of the interaction potential seen in Figure 1-1. At small r the potential rises to infinity and at large r the potential approaches 0. The minimum potential energy forms an asymmetrical well that bounds the range of physically possible distances that could separate two interacting atoms. For later background discussions it is the infinite potential barrier at low r that is of particular interest. The manner in which the extremely repulsive component of the potential is modeled turns out to be of great importance to the calculation of entropy, though it is of little interest for many other thermodynamic and structural properties.

Entropy Pair Functional Theory

Boltzmann's famous entropy equation provides and expression for entropy (S) that is only dependent upon the number of microstates (W) in the system.

$$S = k_B lnW \tag{1.2}$$

At the beginning of the twentieth century Gibbs restated Boltzmann's equation in terms of the probability p_i of a microstate appearing in the system.

$$S = -k_B \sum_i p_i \, ln \, p_i \tag{1.3}$$

In 1948 Shannon[1], in the development of information theory, applied a more fundamental form (Boltzmann's H theory) of the entropy equation to information. Here H is entropy, $P(x_i)$ is the probability mass density function of each state x_i. As will be seen shortly, the discrete nature of encoding information necessitates the probability mass function instead of the more familiar probability density function.

$$H(X) = -\sum_i P(x_i) \log_b P(x_i) \tag{1.4}$$

It is helpful to mention Shannon's work because his approach with information can be reduced to a very simple case that provides some intuition for the somewhat abstract concept of entropy in general. For a fair coin toss where there are two possible states (heads or tails) each $P(x_i) = \frac{1}{2}$. For two possible states it makes sense to set b = 2 and so for the fair coin toss $H(X)$ evaluates to 1. 1 is the maximum value of H for the fair coin toss and it corresponds to the condition of the system when it is the least predictable. Any 'unfairness' of the coin will make the outcome of a toss more predictable and lower the entropy of the system. It is ultimately the same for a physical system even though the rules that govern the sets of possible states are vastly different.

Another tool at the disposal of the modern worker is the pair correlation function [PCF]. In general the PCF describes the correlation between two atoms. In three-dimensional space, the PCF is a function of six coordinates, the x, y and z positions of both atoms. In a homogeneous system, the position of the first particle is arbitrary, reducing the arguments of the PCF to 3, the difference in x, y and z positions. If we choose to avoid the angular dependence, as would be rigorously justified in an isotropic system, in the correlation and instead look only at the radial separation between particles, then the PCF is a function of only one variable, r. This limited PCF is referred to as the RDF [radial distribution function], $g(r)$. This RDF defined in Equation1.5 below provides the probability of finding an atom at a distance r from any atom in the system[2]. Where

$$\rho(\boldsymbol{r_1}, \boldsymbol{r_2}) = \rho g(\boldsymbol{r_1}, \boldsymbol{r_2}) \tag{1.5}$$

Equation 1.5 provides a valuable insight into the role of $g(r)$. Namely, when the atoms of a system are correlated, $g(r)$ provides a correction to the bulk density ρ to provide the r dependent density $\rho(r)$. For large separation between atoms, when there is no correlation, $\rho(r) = \rho$ (independent of r) and $g(r) = 1$.

$$\int_0^\infty \rho g(r) 4\pi r^2 dr = N - 1 \approx N \tag{1.6}$$

This equation states that the integral of an RDF $g(r)$ time the number density ρ (atoms

per unit volume) is equal to the number of atoms in the system minus 1. The minus 1 resulting from the exclusion of the central atom from which the beforementioned distances are calculated. For large N it is common to approximate the integral as being equal to N. In addition to the obvious characteristics present in the definition above, the RDF is defined such that for large r the value of $g(r) = 1$. From the perspective of any atom chosen as the central atom in a system there is a certain distance after which the odds of encountering an atom at that distance are unity.

In 1938 Kirkwood [3] developed an approximation for the calculation of the RDF that provided the ability to develop an expression of the probability density of a system from the RDF.

$$P(r_N) = \frac{\rho^N}{N!} g^N(r_N) \tag{1.7}$$

In a more detailed form below the limit to pair interactions is explicit in the form of g and in the product $N(N-1)/2$, the number of unique pairs in a system of N atoms.

$$P(r_N) = \frac{\rho^N}{N!} \Pi_{i,j}^{N(N-1)/2} g(r_i, r_j) \tag{1.8}$$

This expression paves the way to calculate entropy via the Gibb's equation directly from an RDF. Green and Wallace [4,5] proposed the following expression for excess entropy.

$$S_K^x[g] = -1 + \lim_{R \to \infty} \frac{1}{2} \{ -1 + \rho \int_0^\infty dr \bar{g}(r)(\ln \bar{g}(r) - (\bar{g}(r) - 1)) \tag{1.9}$$

Equation (1.9) is a functional of the RDF we will refer to as the Kirkwood entropy. The Kirkwood entropy is only accurate for a range of temperatures contained inside the liquid region. There are two well defined weaknesses of this functional, it fails to produce the correct entropy at high temperatures and it approaches negative infinity as the temperature approaches the melt temperature. Nicholson et al. proposed the existence of a "universal functional" of unknown form that removes these inadequacies. Preliminary

steps in addressing these shortcomings have been addressed by Nicholson et al. as the Entropy Pair Functional Theory[6].

For the high temperature correction it has been established that the error at high temperature approaches ½k$_B$/atom as the temperature approaches infinity. Nicholson et al. have identified that for very high temperature the restriction to $g(r) = 1$ at large r is no longer valid.

To understand the high temperature correction, it is informative to investigate the characteristics of the Johnson interaction potential that was used in the development of the Entropy Pair Functional Theory. Figure 1-2 shows a set of RDFs over a range of temperatures. These RDFs were generated from MD simulations that utilized the Johnson potential. In contrast to the Lennard-Jones potential discussed earlier, the Johnson potential has a finite potential barrier at low r.

In Figure 2-2 it is seen that for the lower temperatures there is a region in low r where the probability of finding a neighboring atom is zero. This region, created by the potential barrier found at low r in the interaction potential, is referred to as the excluded region. However, the RDF for 10^6 K demonstrates non-zero probability even at very low values of r. This characteristic is a direct result of the finite barrier in the Johnson potential and serves as a helpful visualization for the subsequent discussion of the high temperature correction.

We are now ready to return to the problem of the high temperature error in the Kirkwood entropy. Since the integral of the product of ρ and $g(r)$ must equal N-1 at any temperature, in order for there to be area under the RDF curve in the excluded region there must be a reduction of area under the remaining RDF curve to maintain the total area at N-1. This means that at large r the value of $g(r)$ must be something slightly less than 1. Baranyai[7] first proposed that the amount less than 1 must be 1/N, so that at large r at very high temperature $g(r) = 1 - 1/N$. Physically, this can be thought of as a penalty to the unity probability at large r. That penalty is the 1/N chance that a non-central atom would be found inside the excluded region. Because this transition from $g(r) = 1$ to $g(r) = 1 - 1/N$ is physical and potentially measurable from the RDF it forms the basis

for the high temperature correction. We reserve a detailed discussion of the high temperature correction for later chapters. At this time we simply point out that the EPFT utilizes an additive correction to the Kirkwood entropy as shown in Equation (1.9) below as $\frac{1}{2}\tilde{\phi}[g]$

$$S_K^x[g] = \frac{1}{2}\tilde{\phi}[g] - 1 + \lim_{R\to\infty}\frac{1}{2}\{-1 + \rho\int_0^\infty dr\bar{g}\,(r)(ln\bar{g}(r) - (\bar{g}(r) - 1)) \tag{1.10}$$

Nicholson et al. also provided a correction to the error as the liquid approaches the melt temperature. Starting from the knowledge that for a liquid and lower temperatures, at large r, $g(r) = 1$, and considering that in the lower temperature RDFs in Figure 1-2 that as more defined peaks emerge, $g(r)$ deviates from 1 over an increasing range of r. Nicholson et al. propose as an indicator of the approach to crystalline structure a variance based measure constructed as follows

$$h(r) = g_s(|\boldsymbol{r}|)\text{-}1 \tag{1.11}$$

$$G = 4\pi rh(r) \tag{1.12}$$

$$\kappa[\bar{g}] = \frac{\rho}{4\pi}\int dr G^2(r) \tag{1.13}$$

$$\gamma[g] = 1 + q_0\kappa[\bar{g}]^4 \tag{1.14}$$

With this addition, the final form of the EPFT is stated as follows,

$$S_K^x[g] = \frac{1}{2}\tilde{\phi}[g] - 1 + \lim_{R\to\infty}\frac{1}{2}\{-1 + \frac{\rho}{\gamma}\int_0^\infty dr\bar{g}\,(r)(ln\bar{g}(r) - (\bar{g}(r) - 1)) \tag{1.15}$$

The correction functionals $\tilde{\phi}[g]$, and γ are designed to act in two distinct regions of the liquid entropy as demonstrated in Figure 1-3.
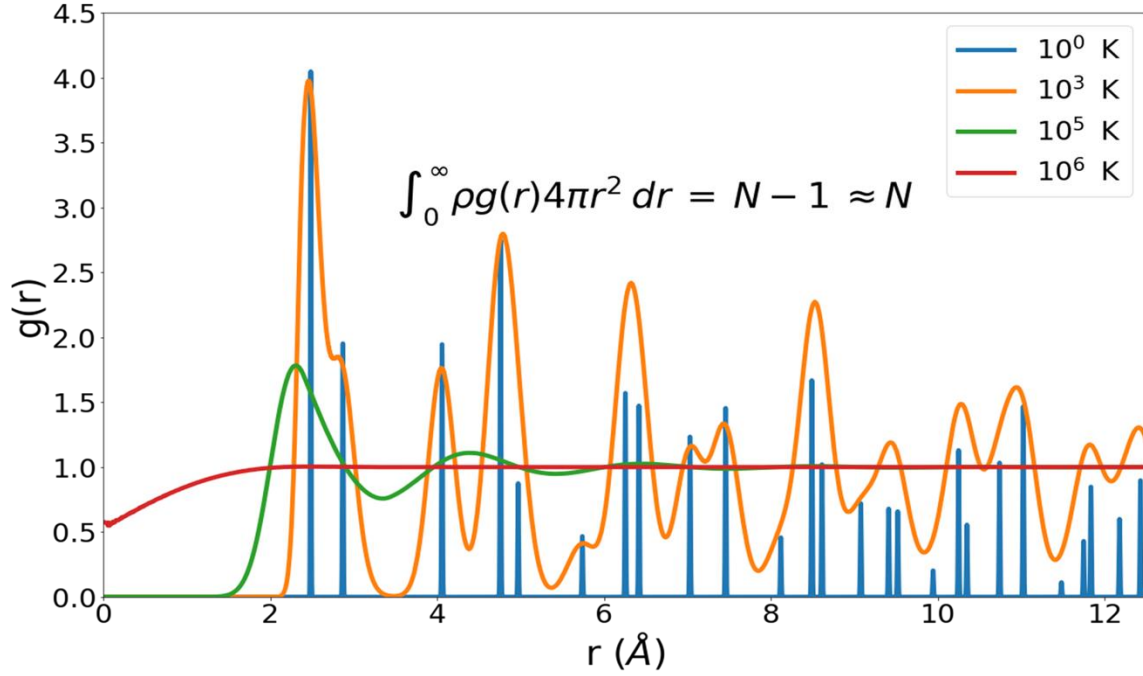
Figure 1-2. RDF for BCC iron modeled with the Johnson potential.
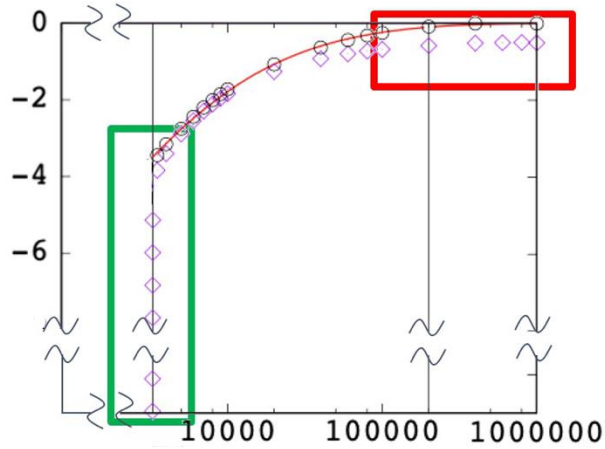
$$S_K^x[g] = \boxed{\frac{1}{2}\tilde{\phi}[g]} - 1 + \lim_{R \to \infty} \frac{1}{2}\left\{ -1 + \boxed{\frac{\rho}{\gamma}} \int_0^\infty dr\, \bar{g}(r)(ln\bar{g}(r) - (\bar{g}(r) - 1)) \right.$$

Figure 1-3. Liquid EPFT correction functionals regions of action.

9

For the entropy calculation of systems in the solid phase we are not limited to an approximation of the probability distribution based on the RDF. Instead, throughout the solid region the probability distribution of the atomic positions can be obtained directly from either the positional output of molecular dynamic simulations or through fitting techniques from the RDF. Nicholson explores a couple of different formulations which all share at a minimum the advantage that there are no fitting parameters used.

The first form of the entropy functional for the solid that we will be concerned with involves the application of the same extended expression of the Kirkwood approximation used above for the liquid. Working from the assumption that in the solid phase the relevant probability density functions can be modeled as Gaussian a reference entropy is first determined based on the variance of an atom around its ideal lattice site.

$$S_x^r[g] = -1 + \frac{3}{2} + \frac{3}{2} \ln \frac{\lambda_{00}^2}{\bar{\ell}^2} \tag{1.16}$$

Where λ_{00}^2 is the variance of an atom around its lattice position and

$$\ell = \frac{1}{\rho}, \quad \bar{\ell} = \frac{\ell}{\sqrt{\pi}}$$

S_x^r is said to be the reference entropy and is then refined by the addition of increasingly higher order nearest neighbor distances

$$S_x^{ex}[g(r)] = S_x^r + \frac{3}{2} \sum_i \ln \frac{\lambda_{0i}^2}{\lambda_\infty^2} + \frac{3}{2} \left(1 - \frac{\lambda_{0i}^2}{\lambda_\infty^2} \right) \tag{1.17}$$

Here the first subscript 0 of λ_{0i}^2 indicates the atom from which a neighbor is determined and the second subscript i is the order of nearest neighbor.

Nicholson et al. develop the second form of the solid entropy equation that we will examine from the work of Morris and Ho[8]. This form suggests a connection between the entropy in solids as derived from the extended expression of the Kirkwood approximation above and the entropy in harmonic solids. Nicholson et al. present two cases which again include a reference entropy, we will restrict ourselves to Case 1 for now.

$$S^r_{Ih}[g] = -1 + \frac{3}{2} + \frac{3}{2} ln \frac{\lambda^2_{00}}{\bar{\ell}^2} \tag{1.18}$$

This is easily recognizable as the reference entropy from the Kirkwood form above. The difference comes in with the addition of an off diagonal coupling term derived from the truncated correlation matrix. For Case 1 this term works out to be,

$$\epsilon_1 = -\frac{\frac{\lambda^2_{01}[\bar{g}]}{4} - \frac{\lambda^2_{00}[\bar{g}]}{2}}{\frac{\lambda^2_{00}[\bar{g}]}{2}} \tag{1.19}$$

And is included as a part of an additive term in the total entropy as shown below.

$$\tilde{S}^x_{Ih-TT} = S^r_{Ih}[g] + \frac{3}{2} ln \left(1 + \frac{1}{2} \left(\sqrt{1 - 4|\epsilon_1|^2} - 1 \right) \right) \tag{1.20}$$

These equations for liquid and solid entropy calculation along with the supporting information in this section should provide a sufficient foundation for the work discussed in the remaining sections of this proposal.

The Broader View

EPFT has been developed from a classical view point and is intended to calculate entropy from correlations between atoms determined only by the classical model. For this reason it is important to briefly examine the relationship between entropy as determined from the classical molecular dynamics simulations and the entropy resulting from the Debye model where quantum effects are taken into consideration.

To place the present work in the broader view we first preview the target entropy we will use to determine the quality of the EPFT results. Details of this target entropy will be provided in Chapter 2; here it is sufficient to note that, because EPFT calculates excess entropy and since thermodynamic integration provides a change in entropy ΔS, measured from some reference, our target entropy is developed to be 0 at the perfect gas limit (as

11

temperature approaches infinity). Figure 1-4 presents a pictorial development of total entropy as the sum of ideal and excess entropy.

We are now ready to compare the total entropy from the classical model to that obtained from the Debye model. Figure 1-5 provides this comparison. As expected the classical begins to break down as temperature approaches 0 K. The Debye model takes into account vibrations governed by quantum mechanics; the proper application of these quantum effects to a correction of the harmonic spectrum utilized in EPFT is a topic of ongoing interest.

Purpose

The purpose of this work is twofold. The first purpose is to implement the entropy pair functional as described above and then to extend and demonstrate its use for additional interaction potentials and crystal structures. This extension represents a first step toward a true universality of the EPFT. Because the high temperature correction is integral to achieving a universal EPFT, special attention is given to the development of this correction. The second purpose of this work is to provide a set of software tools that will allow other in the simulation community to accomplish similar work easily as a part of or in conjunction with simulations performed in LAMMPS.

Structure of the Dissertation

This dissertation is composed of three parts. The first part presents an exploration and extension of the Entropy Pair Functional Theory (EPFT) to three new material systems. This work is described in Chapter 2. The second part investigates the several proposed approaches for functionals to correct the high-temperature error present in the Kirkwood formulation. This work is described in Chapter 3. The third part is a working code to implement the EPFT to generate entropy from the trajectory files of a single simulation performed with the LAMMPS software. This code is available in a GitHub repository located at https://github.com/cliftonsluss/SFunk.

Excess   **Ideal**   **Total**
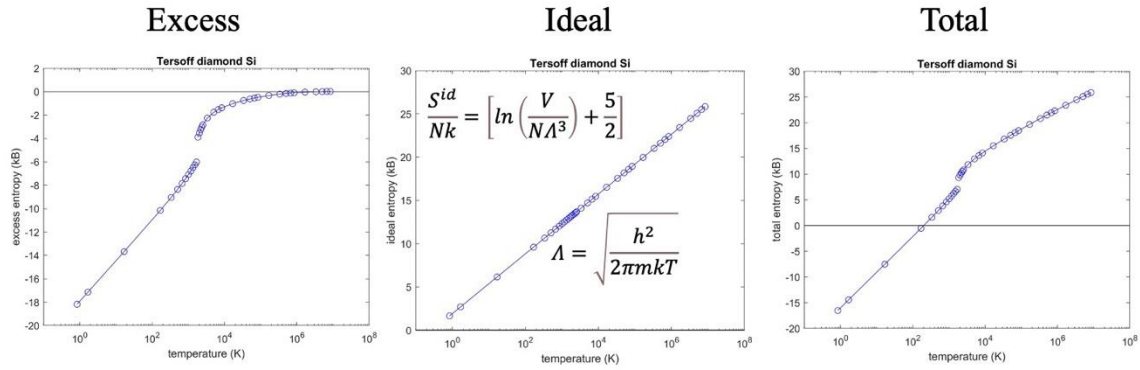
Figure 1-4. Total entropy from classical thermodynamic integration.
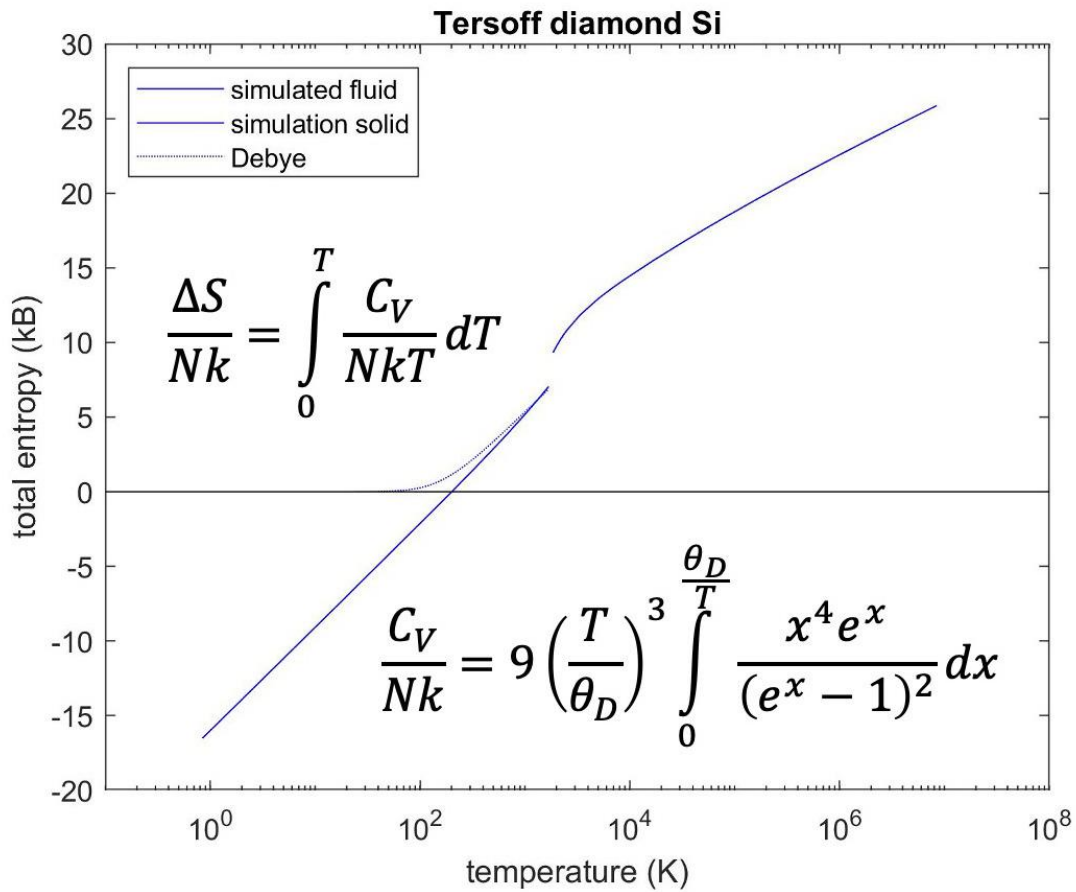
Figure 1-5. Comparison of classical entropy with entropy from Debye model.

13

REFERENCES

[1] C. E. Shannon, Bell System Technical Journal **27** (3), 379 (1948).

[2] Donald A. McQuarrie, *Statitical Mechanics*. (Harper & Row, New York, 1976).

[3] J. G. Kirkwood, J. Chem. Phys. **3** (5), 300 (1935).

[4] H.S. Green, *The Molecular Theory of Fluids*. (North-Holland, Amsterdam, 1952).

[5] D. C. Wallace, J. Chem. Phys. **87** (4), 2282 (1987).

[6] Donald M. Nicholson, Carrie Y. Gao, Marshall T. McDonnell, Clifton Sluss, and David J. Keffer, Entropy (2020).

[7] A. Baranyai and D. J. Evans, Physical Review A **40** (7), 3817 (1989).

[8] J. R. Morris and K. M. Ho, Phys. Rev. Lett. **74** (6), 940 (1995).

CHAPTER 2

EXPLORATION AND EXTENSION OF THE ENTROPY PAIR FUNCTIONAL THEORY

A version of this work was originally published by Clifton C. Sluss, Jace Pittman, Donald M. Nicholson, and David J. Keffer:

Clifton C. Sluss, Jace Pittman, Donald M. Nicholson, and David J. Keffer, "Exploration of Entropy Pair Functional Theory." Entropy 2022, 24, 603.

Abstract

Evaluation of the entropy from molecular dynamics [MD] simulation remains an outstanding challenge. The standard approach requires thermodynamic integration across a series of simulations. Recent work by Nicholson et al. demonstrated the ability to construct a functional that returns excess entropy, based on the pair correlation function [PCF]; it was capable of providing, with acceptable accuracy, the absolute excess entropy of iron simulated with a pair potential in both fluid and crystalline states. In this work, the general applicability of the Entropy Pair Functional Theory [EPFT] approach is explored by applying it to three many-body interaction potentials. These potentials are state of the art for large scale models for the three materials in this study: Fe modelled with a modified embedded atom method [MEAM] potential, Cu modelled with an MEAM and Si modelled with a Tersoff potential. We demonstrate the robust nature of EPFT in determining excess entropy for diverse systems with many-body interactions. These are steps toward a universal Entropy Pair Functional, EPF, that can be applied with

confidence to determine the entropy associated with sophisticated optimized potentials and first principles simulations of liquids, crystals, engineered structures, and defects.

Introduction

In material science, simulation is the third pillar of research, providing a complementary tool to experiment and theory. An attractive feature of simulation is that unambiguous access to all atomic coordinates is available. Density Functional Theory (DFT) simulation is the tool of choice for small systems and short time scales. Simulation with optimized classical potentials is the tool of choice for larger systems and longer time scales. Molecular dynamics (MD) has become a routine computational tool for investigating the structural, thermodynamic and transport properties of materials. MD simulations using optimized classical potentials for systems up to 10^5–10^6 atoms can be performed on modest compute clusters, while larger simulations are possible with access to supercomputing facilities. In terms of time scale, MD simulations from 1 to 10 ns are routine, while longer simulations are again possible given more extensive simulation resources. Routine simulations with DFT Hamiltonians are limited to 100 s of atoms for 100 s of ps.

As interest grows in materials with engineered disorder at the atomic scale, the ability to simulate systems with a sufficient number of atoms to capture the disorder further motivates interest in large scale MD simulations, where the use of first principles forces remains infeasible.[1,2] MD simulation of multicomponent materials with atomic-scale disorder, such as high entropy alloys (HEAs) or entropy stabilized oxides (ESOs) are limited by two issues. First, MD simulations require as input interaction potentials that describe how each type of atom interacts with each other type of atom. The robust and rapid determination of highly accurate interaction potentials for alloys or ceramics with arbitrary numbers of components is an area receiving great research interest.

The second challenge, the one on which this paper is focused, centers on the routine determination of entropy via MD simulation. The industry standard for MD simulation is the open-source simulation software, LAMMPS.[3] LAMMPS can generate instantaneous

17

values of many thermodynamic properties, including temperature, pressure, density, internal energy and enthalpy. Properties based on thermodynamic partial derivatives of the above properties, such as heat capacity or isothermal compressibility, can be obtained accurately with just a couple of simulations, using a centered finite difference approach. Mechanical properties, such as the elastic tensor or the bulk modulus, are also readily extracted. Structural properties, such as the radial distribution function (RDF) emerge from the straightforward post-processing of the trajectory file generated from an MD simulation. Algorithms for the determination of transport properties, such as diffusivities, shear viscosities or thermal conductivities, from both equilibrium and non-equilibrium MD simulations, abound. The reader is directed to the "examples" directory that accompanies the LAMMPS source code, which contains demonstration scripts for obtaining all of the properties listed above. The property that most resists straightforward determination in MD simulation is the entropy, and by extension the Helmholtz and Gibbs free energies.

Certainly, it is possible to calculate relative entropy differences through thermodynamic integration. However, this approach requires that a series of simulations be performed across the integration path. Examples of entropy differences that can be evaluated in this way are the entropy change due to a change in temperature at constant volume

$$\Delta S = \int_{T_1}^{T_2} \frac{C_v}{T} \, dT \tag{2.1}$$

or the entropy change due to a change in volume at constant temperature

$$\Delta S = \int_{V_1}^{V_2} \left(\frac{\partial p}{\partial T}\right)_V \, dV \tag{2.2}$$

which follows from the Maxwell relation

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V \tag{2.3}$$

The drawback to thermodynamic integration is the computational expense required to

perform the additional simulations. In some instances, there may also be a human-hour cost setting up addition configurations corresponding to each point along the integration path.

There remains interest in the determination of absolute entropy from a single simulation. To date, work has focused on determining the entropy as a functional of the PCF, which can be generated from a single simulation. The Gibbs formulation of Boltzmann entropy assumes the probability density of the atoms of a system in real space is known.[4]

$$S = -k_B \sum_i p_i ln\ p_i \qquad (2.4)$$

For the MD practitioner, computation of entropy from Equation (2.4) depends upon a method to calculate a probability distribution, p_i, from atomic position data obtained from simulations. Beginning with the Kirkwood approximation for the calculation of a discrete probability distribution, others have developed a probability density based on the RDF and a resulting functional for the direct calculation of entropy.[5-7] Kirkwood (K) entropy, S_K^x, provides a formulation of the absolute excess (x) entropy in the fluid state as a functional of the RDF.

$$S_K^x[g] = -1 + \lim_{R \to \infty} \frac{1}{2} \{ -1 + \rho \int_0^\infty dr g\ (r)(lng(\text{r}) - (g(r) - 1)) \qquad (2.5)$$

where $g(r)$ is the RDF. It is important to note that Equation (2.5) excludes all dependence on correlations higher in order than pair correlations. Kirkwood is one approximation to p_i; note that every approximation to the entropy can be related back to at least one approximation to p_i. For example, recent work by Haung and Widom[8] utilizes a Gaussian ansatz for p_i that is applicable to crystals. It is Gaussian in the sense that it is the exponential of a form quadratic in atomic displacements. Since each factor involves just two sites, they obtain entropy as a functional of the PCF. Their p_i is an approximate p_i that produces a harmonic oscillator entropy. Thus, entropy calculated in

quasi-harmonic approximation, based on first principles simulation of phonons[9], can be thought of as evaluating Equation (2.4) with an approximate p_i.

Contemporary approaches to the calculation of absolute excess entropy include improvements to Kirkwood entropy and the use of RDFs determined by first principles MD simulations[10-12] and machine learning techniques.[13] Recently, an approximation to the universal functional for the calculation of absolute excess entropy for pair potential Hamiltonians, from classical molecular dynamics simulations, has been developed.[14] This entropy pair functional theory (EPFT) has been demonstrated to provide reasonable agreement with excess entropy values produced by thermodynamic integration of MD simulation results for the Johnson iron (BCC) pair potential[15] across the entire temperature range, from crystalline solids at temperatures as low as 1 K up through the liquid state to a state approaching the perfect gas at 10^7 K. As it currently exists, the EPFT approach specifies a temperature independent functional of the PCF that returns the excess entropy. This single functional is constructed from subsidiary functionals that highlight specific traits of the PCF. The simplest of these functionals identifies the PCF as corresponding to a crystal or fluid. If the PCF corresponds to a crystal, the widths of the peaks of the PCF at lattice separation vectors can be quantified by the variance of separation vectors within each peak. If the PCF corresponds to a fluid, the PCF is isotropic and is equal to its spherical average RDF. We depend on several subsidiary functionals of the RDF, for example, $S_K^x[g]$ and the coordination number. Evaluating the excess entropy requires various special integrals involving the PCF at each temperature of interest. However, unlike thermodynamic integration, the EPFT holds the promise that an accurate approximation can be found to the universal functional, which would make numerous simulations outside the points of interest unnecessary.

Formally, the entropy of many-body potentials depends not just on the PCF, but also on the many-body correlations. However, it was proved that the error introduced by using the EPF is second order in many-body interactions.[14] Modern simulations are often first principles, or use many-body interactions based on a mix of first principles results and measured properties. The extent to which an EPF effectively models the entropy of

20

systems, governed by non-pairwise potentials, remains an open question. We apply EPFT to three many-body systems in order to explore the accuracy and universality of EPFT. EPFT must meet this challenge if it is to be adopted as an alternative to thermodynamic integration. The goal of this work is to investigate the universality of the EPFT approach with the explicit target of generating the absolute entropy from the pair correlation function (PCF) of a single simulation.

In this work we specifically explore the applicability of the EPFT to FCC copper and BCC iron systems simulated with the modified embedded atom method (MEAM) potential[16] and diamond cubic silicon system utilizing the Tersoff potential.[17] These systems and potentials were chosen out of a desire to take initial steps in demonstrating and expanding the universality of the EPFT. Copper and the MEAM potential provide the opportunity to test the EPFT with a new crystal structure (FCC), while iron provides a more direct comparison between MEAM and the Johnson potential of the original EPFT work. Silicon introduces a third crystal structure (diamond cubic) and the Tersoff potential tests the pair potential assumption of the underlying theory with the inclusion of large three--body angular terms.

Theory

Nicholson et al. have provided an extensive derivation of the EPFT approach and situated it within the historical framework of the Kirkwood superposition approximation for the fluid state and the harmonic oscillator approximation for the solid state.[14] For a full accounting, the interested reader is directed to that work. Here, we provide a summary of important points necessary to make this document self-sufficient.

The Entropy Pair Functional in[14] builds upon the Kirkwood entropy in the fluid state by introducing two new functionals that correct for two problem areas of the Kirkwood entropy.[5-7,18,19] First a corrective functional $\tilde{\phi}[g]$ is introduced to ensure that excess entropy approaches zero as the system approaches a perfect gas at high temperature. This high temperature limit is a subject that has previously been investigated with great interest.[5] The second functional $\gamma[g]$ provides a correction as liquid approaches

21

crystallization, where the Kirkwood entropy diverges. With the incorporation of these two corrections, the modified Kirkwood entropy in the fluid state becomes

$$\tilde{S}_K^x[g] = \frac{1}{2}\tilde{\phi}[g] - 1 + \lim_{R\to\infty}\frac{1}{2}\{-1 + \frac{\rho}{\gamma[g]}\int_0^\infty dr\, g(r)(lng(r) - (g(r) - 1)) \qquad (2.6)$$

A further examination of the corrective functionals $\tilde{\phi}[g]$ and $\gamma[g]$ will be presented in the methods section of this paper.

EPFT also extends the Kirkwood entropy into the crystalline phase down to arbitrarily low temperature. Several forms of the functional for the crystal have been proposed. In this work, we focus on three forms that utilize self and pair correlations only. The reference entropy of the crystalline state, $S_r^x[g]$, depends only on the variance of an atom around its lattice position, $1/2\lambda_{00}^2$,

$$S_r^x[g] = -1 + \frac{3}{2} + \frac{3}{2}ln\frac{\lambda_{00}^2}{\bar{\ell}^2} \qquad (2.7)$$

where $\ell = 1/\rho$ and $\bar{\ell} = \ell/\sqrt{\pi}$. S_x^r is said to be the reference entropy, which is then refined by the addition of terms that arise from neighboring atoms. The second form for solid entropy implemented here is also based on Equation (2.7), but utilizes the variance of first nearest neighbor distances, $1/2\lambda_{01}^2$, instead of the variance of atoms around their ideal lattice positions, $1/2\lambda_{00}^2$. Comparing Equation (2.8) to Equations (2.5) and (2.7) it can be seen that it represents a direct analog of the Kirkwood liquid excess entropy for solid excess entropy.

$$S_{II\,h}^x[g] = -1 + \frac{3}{2} + \frac{3}{2}ln\frac{\lambda_{01}^2}{\bar{\ell}^2} \qquad (2.8)$$

The third form for solid entropy we calculate was presented by Nicholson et al. as a connection between the Kirkwood entropy and the harmonic solid technique of Morris and Ho.[20]

$$\tilde{S}^x_{\text{I h}-\text{TT}} = S^r_x[g] + \frac{3}{2}\ln\left(1 + \frac{1}{2}\left(\sqrt{1 - 4|\epsilon_1|^2} - 1\right)\right) \qquad (2.9)$$

where S^r_x is defined in (2.3) and ϵ_1 provides an off-diagonal coupling term derived from the truncated correlation matrix

$$\epsilon_1 = -\frac{\dfrac{\lambda^2_{01}[\bar{g}]}{4} - \dfrac{\lambda^2_{00}[\bar{g}]}{2}}{\dfrac{\lambda^2_{00}[\bar{g}]}{2}} \qquad (2.10)$$

These three forms of solid entropy represent upper bounds. Due to this, in practice, the lower of the values produced should be considered the best estimate of entropy. Finally, it must be noted that any calculation of excess entropy is impacted by the accounting of additional degrees of freedom. While our treatment here is limited to the degrees of freedom of the atoms in a system, additional sources of entropy may be of great interest to other workers. For example, electrons contribute to the entropy on their own in several ways and, through their impact, on the degrees of freedom that describe the nuclei. This is particularly true for iron; in addition to electron-hole entropy there is entropy resulting from the formation of local moments.[21-23] For work such as this, it is important to note that EPFT applies with only minor changes when the scope of simulations is expanded to include other degrees of freedom, e.g., those associated with site occupation, as in alloys, or spin degrees of freedom, as in magnetic materials. EPFT is expanded by indexing $g_{\alpha,\beta}(r,r')$ where α and β refer to the atomic number and local atomic moment orientation at each nucleus.

Methods

Molecular Dynamics Simulations

A suite of classical MD simulations was performed for copper, iron and silicon. LAMMPS[3] was used to perform simulations of the three materials. The MEAM potential for Fe and for FCC Cu[16], and the Tersoff potential for Si[17], were taken from the literature.

For a given material, the density remained constant at all temperatures. The densities are reported in Table 2-1. A summary of the simulation size is included in Table 2-1 below. As can be seen in the table, the length of the cubic simulation volume in any dimension was in the order of 10^2 Å. This size was necessary to be able to calculate the RDF up to a maximum value of 50 Å.

For each material, a set of 40 simulations was performed for reduced temperatures in the range from 0.001 to 5000, where the temperatures were normalized by the melting temperature reported for each potential in the literature. (See Table 2-1) The Nose-Hoover thermostat was used to maintain the target temperature in the canonical ensemble (NVT). The size of the time step in each simulation was determined based on energy conservation in simulations in the microcanonical (NVE) ensemble, performed explicitly for this purpose. The time steps used ranged from ~13 fs at the lowest temperatures, nominally 1 K, to ~1.6×10^{-2} fs at the highest temperatures, nominally 10^6 K. In general, the total duration of the simulations was determined in order to ensure convergence of thermodynamic values. Simulations were run for a duration to produce convergence of thermodynamic values and sufficient variability for configurational statistics calculations. All simulations were at least 4.57 ps in length.

The thermodynamic data generated was used to determine the potential energy of the systems for each temperature. For each simulation, atomic coordinates were recorded every 100 timesteps and this trajectory data was used to calculate the pair correlation functions, and the statistical values used as input for the entropy functional. The system sizes were chosen to meet three criteria that we established. First, we wanted to demonstrate EPFT on systems sizes comparable to those typically used by MD simulators. Second, the system dimensions provide for RDF calculation out to 50 Å, which is commensurate with the range of correlations most often provided by experiment. Finally, by working with sufficiently large systems, complicating terms involving 1/N are avoided by operating at the thermodynamic limit. The simulations required an average of 20 h of wall time to complete when run on 2 nodes. This performance is very practical for the typical researcher compared to DFT-based

Table 2-1. Molecular Dynamics Simulations Parameters

System	T_{melt} (K)	Structure	Potential	N(Atoms)	Box Size (Å^3)	ρ (N/Å^3)
Cu	1347	FCC	MEAM	87,808	1,041,357.395	8.432×10^{-2}
Fe	1812	BCC	MEAM	93,312	1,081,182.881	8.631×10^{-2}
Si	1687	Diamond Cubic	Tersoff	54,872	1,100,297.642	4.987×10^{-2}

approaches, which require tens of thousands of memory-laden (>2 TB/core) nodes to simulate systems of comparable size.[24] Note that the computer resources needed to evaluate the entropy with EPFT by post-processing of the trajectories of any simulation is negligible. For comparison, calculation of the entropy through direct calculation of phonon frequencies is computationally more intense for systems of this size.

Target Entropy Development

Thermodynamic Integration

To validate the results from the EPFT, a target entropy was calculated across the whole temperature range of the simulations for each system to use as the standard. The standard entropy from thermodynamic integration and EPFT entropy are based on exactly the same Hamiltonian as rendered by LAMMPS. This is a better standard for comparison than are experimental entropies. The average potential energy was calculated for each temperature and an equation for $U(T)$ was fit to the data. See Appendix 2.A. The derivative $C_V = dU/dT$ and integral $\Delta S = \int_{T_0}^{T_1} \frac{C_V}{T} dT$ were determined analytically and curves for each were generated. This process was repeated for each material for both liquid and solid phases. Since the volumes of the simulations are constant across temperatures, the heat capacity is the constant-volume heat capacity. As the energy term employed is the potential energy, the heat capacity that is generated is the excess heat capacity, which does not include the kinetic contribution of the perfect gas. Similarly, the entropy arising from the excess heat capacity is strictly the excess entropy. Due to the discontinuity in the potential energy and entropy at the melt temperature, the liquid and solid target entropy curves must be developed separately and then reconnected via the calculation of the entropy of fusion.

Entropy of Fusion

The Gibbs Phase Rule states that the number of Degrees of Freedom (DOF) required to fully define a thermodynamic state of system composed of C components and φ phases is given by

$$DOF = C - \varphi + 2 \qquad (2.11)$$

26

In the case of a single component system and a two-phase, e.g., solid-liquid, equilibrium, there is only one degree of freedom. Often, this *DOF* is chosen as the temperature, though that is a choice made out of practical considerations, rather than a theoretical requirement. All other thermodynamic properties, including the pressure, chemical potential and density of the phases are defined once the temperature has been specified. Notably, the densities of the solid and liquid phases at equilibrium are not the same. As the target entropy developed for this work is along an isochor, the two phases present at the melt temperature are not in equilibrium. Thus, the discontinuity in entropy between the solid and liquid in this series of simulations does not correspond to the entropy of fusion of two phases at equilibrium. This motivated an approach to calculate the entropy difference between a liquid at a state defined by temperature and density (T, ρ), and a solid at the same temperature and density.

The entropy difference between a liquid at a thermodynamic state defined by arbitrary temperature, T_1, and arbitrary density, ρ_1. and a solid at the same temperature and density can be broken into three terms that describe a thermodynamic path. Since entropy is a state variable, this calculation is independent of path.

$$\Delta S_{tot} = \Delta S_L + \Delta S_{L \to S} + \Delta S_S \tag{2.12}$$

The thermodynamic path we invoke is as follows. In step 1, a liquid at (T_1, ρ_1) undergoes an isothermal compression (or expansion) to liquid at (T_1, ρ_2), denoted ΔS_L. In step 2, a liquid at (T_1, ρ_2) undergoes a phase change to solid at (T_1, ρ_3), with which it is in thermodynamic equilibrium, denoted $\Delta S_{L \to S}$. In step 3, a solid at (T_1, ρ_3) undergoes an isothermal expansion (or compression) to solid at (T_1, ρ_1), denoted ΔS_S.

We choose T_1 to be a temperature where coexistence of the liquid and solid is possible. We choose ρ_2 to correspond to the dependent liquid phase density at the equilibrium state uniquely defined by T_1. We choose ρ_3 to correspond to the dependent solid phase density at the equilibrium state uniquely defined by T_1. This path provides the entropy difference between a liquid at a thermodynamic state defined by arbitrary temperature, T_1, and arbitrary density, ρ_1. and a solid at the same temperature and density. Practically

speaking, we chose the temperature to correspond to the melting temperature at one atmosphere reported in the literature and reported in Table 2-1. In this case, the density of the coexisting liquid and solid were known and the entropy of fusion was reported in the literature.[25,26] The terms describing the entropy change due to a change in density were computed via thermodynamic integration using Equation (2.2). If the integral in Equation (2.2) is approximated with the trapezoidal rule, then, for the liquid and solid phases, Equation (2.12) becomes

$$\Delta S_L = \frac{1}{2}\left(\frac{1}{\rho_2} - \frac{1}{\rho_1}\right)\left(\left(\frac{\partial p}{\partial T}\right)_{\rho_2}^{L@T_1} + \left(\frac{\partial p}{\partial T}\right)_{\rho_1}^{L@T_1}\right) \qquad (2.13.a)$$

$$\Delta S_S = \frac{1}{2}\left(\frac{1}{\rho_1} - \frac{1}{\rho_3}\right)\left(\left(\frac{\partial p}{\partial T}\right)_{\rho_1}^{S@T_1} + \left(\frac{\partial p}{\partial T}\right)_{\rho_3}^{S@T_1}\right) \qquad (2.13.b)$$

The thermodynamic partial derivative $\left(\frac{\partial p}{\partial T}\right)_V$ evaluated under four conditions, as specified in (2.13.a) and (2.13.b) was evaluated using the centered finite difference formula

$$\left(\frac{\partial p}{\partial T}\right)_{\rho_1} \approx \frac{p(T + \delta T, \rho_1) - p(T - \delta T, \rho_1)}{2\delta T} \qquad (2.14)$$

where δT is a temperature offset, chosen to be sufficiently large to provide a reliable gradient, given the noise present in the pressure calculation. Each derivative requires two simulations.

Pair Correlation Functions

For the liquid entropy, the entropy functional takes, as input, radial distribution functions of the type defined below.[27] RDFs used as input to the entropy functional were calculated with bins of width 10^{-3} Å utilizing an in-house code.

$$\int_0^\infty \rho g(r) 4\pi r^2 dr = N - 1 \approx N \qquad (2.15)$$

We return now to a more detailed discussion of the corrective functionals $\tilde{\phi}$ and γ introduced in the theory section. We consider first the error introduced in the approach to the perfect gas and begin with some observations about the RDF in general. It is evident from Equation (2.15) that the integral of the RDF is a volume that contains all the atoms under consideration less the volume of the central atom, from which the nearest neighbor distances are measured. This volume, referred to as the excluded volume, is shown to decrease as the temperature of the system increases. Examination of Figure 2-1 reveals that as temperature increases, the region in r that corresponds to zero probability of finding a nearest neighbor decreases. The RDF is defined such that at large r $g(r) \approx 1$. This means that at sufficiently long distance there is unity probability of finding an atom in the next increment of volume.

High Temperature Liquid Correction

We reintroduce a functional $Q[g]$ that indicates the departure of the excluded volume from the origin; it is built upon the concept of the Wigner-Seitz cell. For the perfect crystal at 0 K each of the atomic cells in a system emerges as a Voronoi polyhedron (VP), centered on a single atom. The VP is defined to have faces that are perpendicular bisectors of the central atom and its neighbors. A corollary to this definition is that a point found inside the VP will have unity probability of being closer to the central atom than to any of the neighbors outside the VP. We adopt a probabilistic interpretation; at *T* = 0 (stationary atoms) *P(r)* = 1 for points inside the VP and 0 outside the VP; at finite temperature the VP changes over time but *P(r)* remains well defined.

The instantaneous VPs of the crystal evolves with time as atoms move at finite temperature. Eventually, the probability that a point displaced from atom *i* by *r* will be closer to atom *i* than to any other particle becomes spherical at melting. For this reason, a spherical probability was chosen as an approximate boundary for the measurement of the encroachment on the excluded volume as the temperature of the system approaches infinity.

At infinite temperature the probability that a point a distance *r* from an atom is closer to that atom than any other atom is
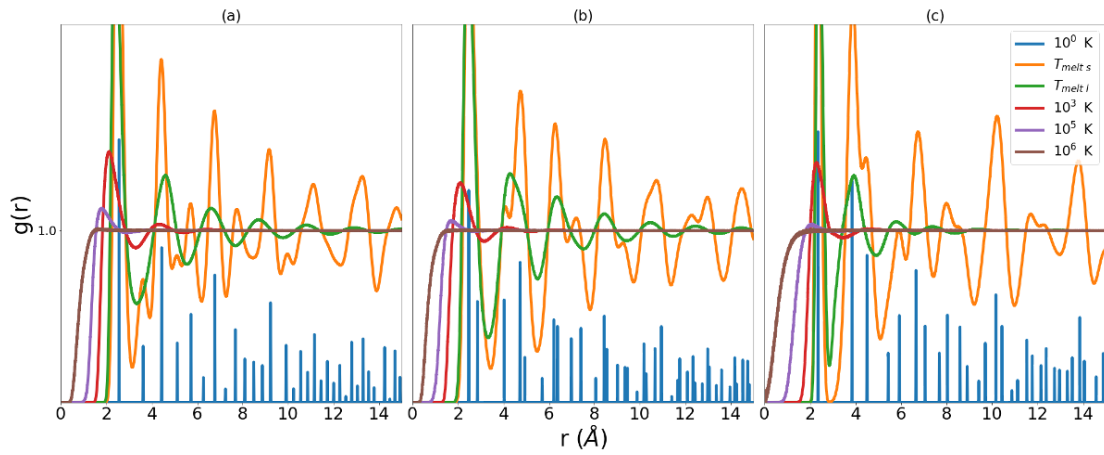
Figure 2-1. RDFs for MEAM copper (a), MEAM iron (b), and Tersoff silicon (c). Low temperature data has been scaled in order to highlight low r behavior at higher temperatures.

$$P^{id}(r) = \lim_{T \to \infty} \left(1 - r^3\big/Nr_s^3\right)^{(N-1)} = exp\left(-\left(\frac{r}{R_s}\right)^3\right) \qquad (2.16)$$

where R_s is the radius of a sphere of the atomic volume of the system, where the atomic volume is the inverse of the density. The derivation of this ideal gas probability is given in Appendix 2.B. The intrusion, at any temperature, of neighboring atoms into the infinite temperature VP is given by

$$I[g] = \int 4\pi r^2 \rho P^{id}(r)g(r)dr \qquad (2.17)$$

$$Q[g] = \max\left(0, \frac{I[g] - I_0[g]}{1 - I_0[g]}\right) \qquad (2.18)$$

As T is lowered toward T_m the peaks in $g(r)$ become increasingly narrow. In the limit that the peaks have zero width, the intrusion of the nearest neighbors becomes

$$I_0[g] = n[g]exp\left(-\left(\frac{R_p[g]}{R_s}\right)^3\right) \qquad (2.19)$$

where R_p is the radius of the first peak in $g(r)r^2$ and

$$n[g] = 2 \int_0^{R_p} 4\pi r^2 \rho g(r)dr \qquad (2.20)$$

$I_0[g]$ serves as a baseline for intrusion. Due to the fact that our results depended only weakly on $I_0[g]$, we made the simplifying, but not essential, choice that $I_0[g]$ is the minimum value of $I[g]$, $I_0[g] = I_{min}$. This means that, in this work, the entropy calculation for the liquid is a single additional calculation of I_{min} at the melting temperature. However, this calculation is less intensive than a full exploration of temperature required for thermodynamic integration. Consequently, the functional that characterizes the escape of the excluded volume from the origin is

31

$$Q[g] = \frac{I[g] - I_{min}}{1 - I_{min}} \qquad (2.21)$$

$Q[g]$ is used for the liquid phase only. It approaches zero at melting and $1 - 1/N$ in the perfect gas limit. The high temperature correction $\tilde{\phi}[g]$ appearing in (2.6) is a functional of the functional $Q[g]$.

$$\tilde{\phi}[g] = Q + q_1 Q(1 - Q) + q_2 Q^2(1 - Q) \qquad (2.22)$$

$\tilde{\phi}[g]$ possesses the same limits as $Q[g]$, namely zero, at the melting temperature, and approaching one (within $O(1/N)$) at very high temperature. The parameters q_1 and q_2 fit the functional $\tilde{\phi}[g]\phi$ to the target entropy. These parameters, which are the same for all materials, allow the functional to match the behavior of the entropy at intermediate temperatures.

Low Temperature Liquid Correction

In Figure 2-2 one PCF for each system is shown multiplied by density, $\rho g(r)$. This quantity is referred to as the neighbor density. The three materials are very different. They correspond to very different temperatures. However, their common attribute is that they correspond to essentially the same excess entropy. One of the challenges for the EPF is to take these very different functions and return the same value. Note that by plotting the neighbor density, we have emphasized that the density of Si is significantly different from that of Cu and Fe, and that the coordination number of Si is much smaller than those of Cu and Fe. Furthermore, the nearest neighbor peak positions almost line up even though the atomic radius in Si is much smaller. The three systems have different packing fractions, f_p. The packing fraction of Si is considerably smaller than that of Fe and Cu. Packing fraction, coordination number, intrusion, and λ_{01} are descriptors of $g(r)$ that depend only on the nearest neighbor peak; λ_{00} can also be considered very local. On the other hand, the integrals of $g(r)ln(g(r))$ and $(g(r) - 1)^2$ have contributions from all r; they emphasize the peaks, valleys, and their long-range persistence. This handful of physically motivated quantities provides a reasonable model of the entropy of the three
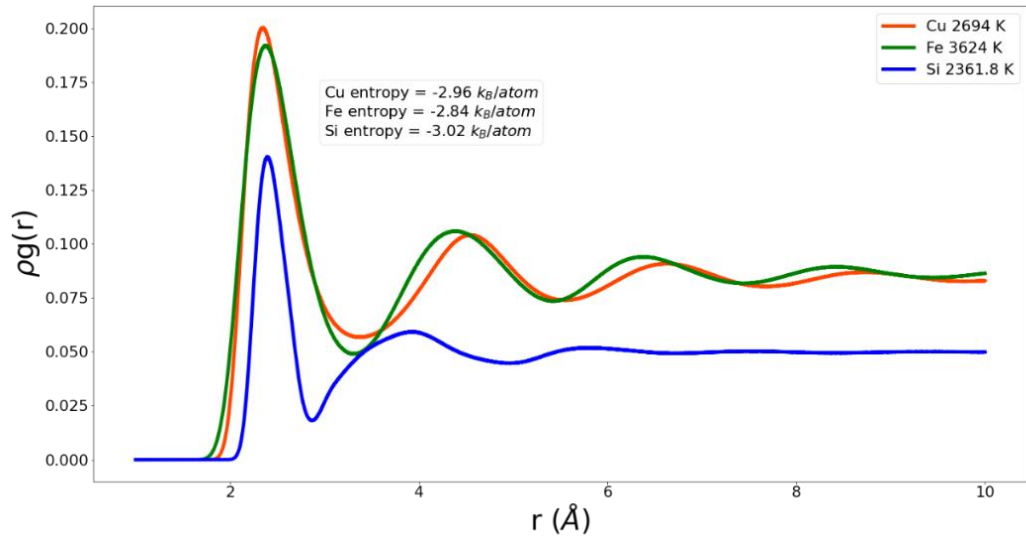
Figure 2-2. Neighbor density of copper, iron, and silicon at small r and similar excess entropy.

systems studied here. Figure 2-2 shows that there are significant differences in the behavior in, for example, the first valley. These differences could be further exploited in the EPF. However, at this stage in the development of an EPF we prefer to show reasonable agreement with a small number of descriptors and parameters. These will naturally build up as we, and others, extend the range of universality by modeling additional systems.

As the liquid approaches T_m, the atoms in the system begin to be distributed near separations found in their ideal lattice. This can be seen in the 'T melt l' data series in Figure 2-1. Turning our attention now to the error in S_K^x for the liquid in the region nearing crystallization, we observe, that as $g(r)$ takes on the characteristics of a set of more and more discrete distributions around the ideal lattice separations, the natural log term in Equation (2.5) produces larger and larger negative values of excess entropy. This results in a gross under-estimation of excess entropy as the liquid cools toward crystallization. Any corrective functional must be constructed with this trend of $g(r)$ in mind. In this case, Nicholson et al. proposed an indicator of the approach to crystalline structure and consequent correction measure constructed as follows

$$h(r) = g(|r|) - 1 \tag{2.23}$$

$$G = 4\pi r h(r) \tag{2.24}$$

$$\kappa[g] = \frac{\rho}{4\pi} \int dr G^2(r) \tag{2.25}$$

$$\gamma[g] = 1 + q_0 \kappa[g] \tag{2.26}$$

In liquid Fe near melting, a typical atom is surrounded by about 14 neighbors (6 BCC nearest neighbors and 8 BCC next-nearest neighbors) with very strict specifications of the distance to each of these neighbors. In liquid Si there are only four. Restrictions reduce the phase space available to the system and reduce its entropy. To ascertain an

34

appropriate level of restriction imposed by the neighbors, we can be guided by the basic fact that each atom is specified by only three coordinates, often {x, y, z}. If the structural environment of an atom is described by coordinates shared with neighbors the number of shared coordinates needed to maintain the correct total number of coordinates, 3 N, is 6 shared coordinates at each atom. For example, crystal entropy is well represented by harmonic models based on a linear chain where the three components of the two vectors to neighbors along the chain comprise the six shared coordinates. In fluids, the RDF gives information only about scalar separations. For a fluid with a coordination of six, the distances to the six neighbors provides a good accounting of restricting coordinates; in such a liquid we expect the corrections to Kirkwood to be small. For fluids, e.g., Fe, with coordination greater than six too many constraints are imposed by Kirkwood and for coordination less than 6, such as Si, it is anticipated to under-restrict the structure. Here we propose a form of $\gamma[g]$ [1] that reflects our understanding of the trends with respect to coordination that need to be incorporated when the peaks of $g(r)$ are narrow. For the differences in coordination number (C.N.) between different systems:

$$\frac{1}{\gamma[g]} = 1 + \frac{c_1 e^{-(c_2 \rho / \kappa)^2}}{(CN - 6)^3 f_p} \tag{2.27}$$

The correction $\gamma[g]$ was fit to the target entropy with the parameters c_1 and c_2 and then applied to the functional (2.5). As was the case for the parameters q_1 and q_2, c_1 and c_2 are optimized to the target entropy. It should be noted that while q_1, q_2, c_1, and c_2 all fit to the entropy obtained from thermodynamic integration, the three materials were fitted

---

[1] In practice , for systems of $5 < CN < 7$, $\gamma = 1$. For all other systems the form of γ in equation 2.27 should be used.

simultaneously with the goal of finding a universal set of these parameters that might serve a wide range of systems.

Results

Low and High Temperature Liquid Corrective Functionals

The high and low temperature liquid correction functionals $\tilde{\phi}$ (Equation (2.22)) and γ (Equation (2.27)) were developed from their constituent functionals Q (Equation (2.20)) and κ (Equation (2.25)) respectively. The results are shown in Figures 2-3, 2-4, 2-5, and 2–6. In Figure 2-3, the functional $Q[g]$ is plotted for Cu, Fe and Si as a function of reduced temperature. In each case the functional approaches zero near the melting temperature (reduced temperature of unity) and approaches one as the temperature increases. The fact that $Q[g]$ does not reach one, even at the highest temperatures simulated, indicates that 10^6 K is not sufficient to force these materials (as governed by the MEAM and Tersoff interaction potentials) to behave as a perfect gas. Certainly $Q[g]$ shares a similar qualitative shape for all three materials as a function of reduced temperature. To be clear, there are no fitting parameters in the functional $Q[g]$.

In Figure 2-4, the functional $\tilde{\phi}[g]$ is plotted for Cu, Fe and Si as a function of reduced temperature. In each case the functional approaches zero near the melting temperature (reduced temperature of unity) and approaches one as the temperature increases. We observe that the functionals for Cu and Si exceed unity at intermediate temperatures. Again, the parameters, q_1 and q_2, were optimized to fit the excess entropy functional (Equation (2.5)) to the target entropy obtained via thermodynamic integration. The behavior of $\tilde{\phi}[g]$ is therefore a consequence of this optimization procedure.

The fitting constants q_1, q_2, c_1, and c_2 given in Table 2-2 are universal to the three systems examined here. It is evident from Table 2-2 that the EPFT provides a fivefold improvement over results based on Kirkwood entropy alone. For the two MEAM systems the maximum error in Kirkwood entropy is predicably at the PG limit where it approaches ~0.5 k_B/atom. However, the maximum EPFT errors for these systems occur at

36
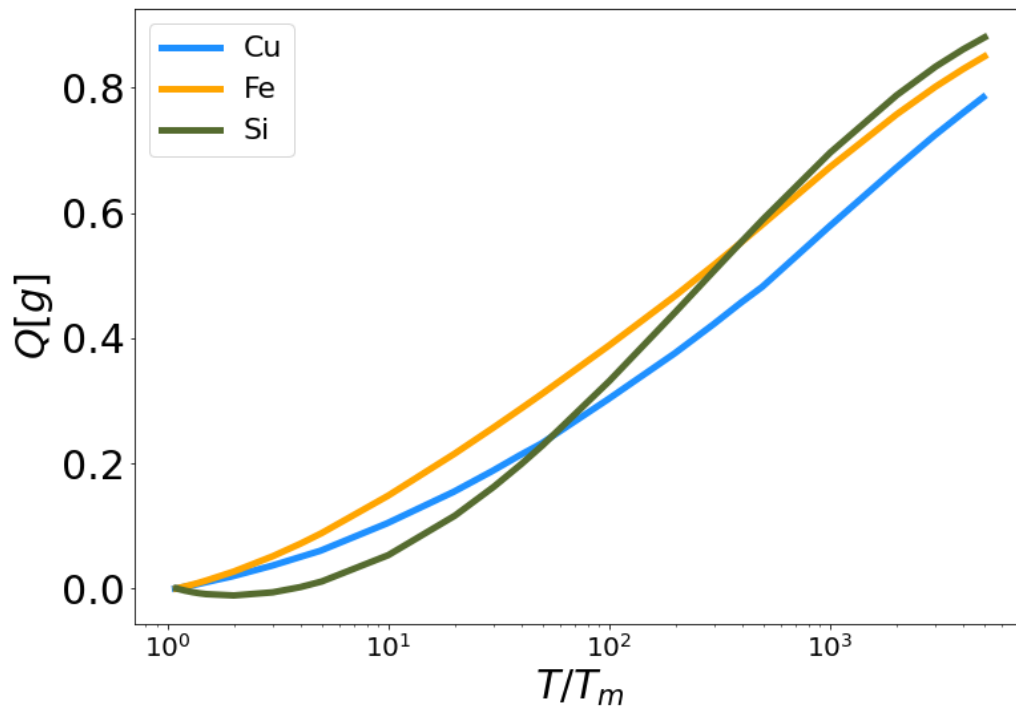
Figure 2-3. High temperature corrective functional Q vs. reduced temperature for MEAM copper and iron, and Tersoff silicon.
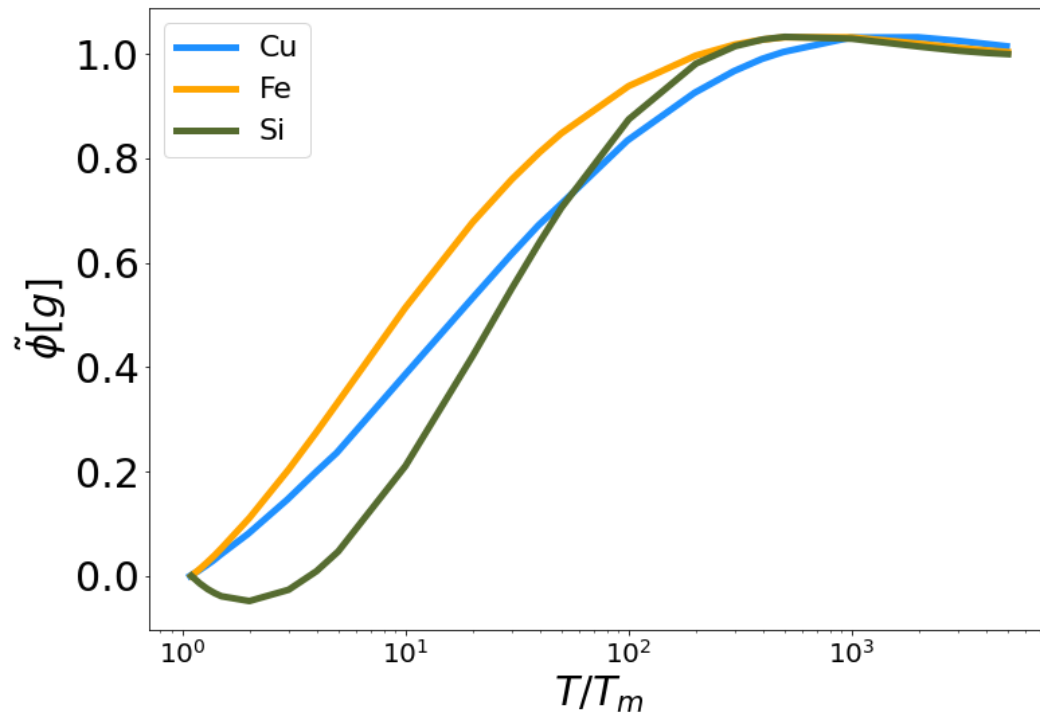
Figure 2-4. The fit high temperature corrective functional $\tilde{\phi}$ vs. reduced temperature for MEAM copper and iron, and Tersoff silicon.
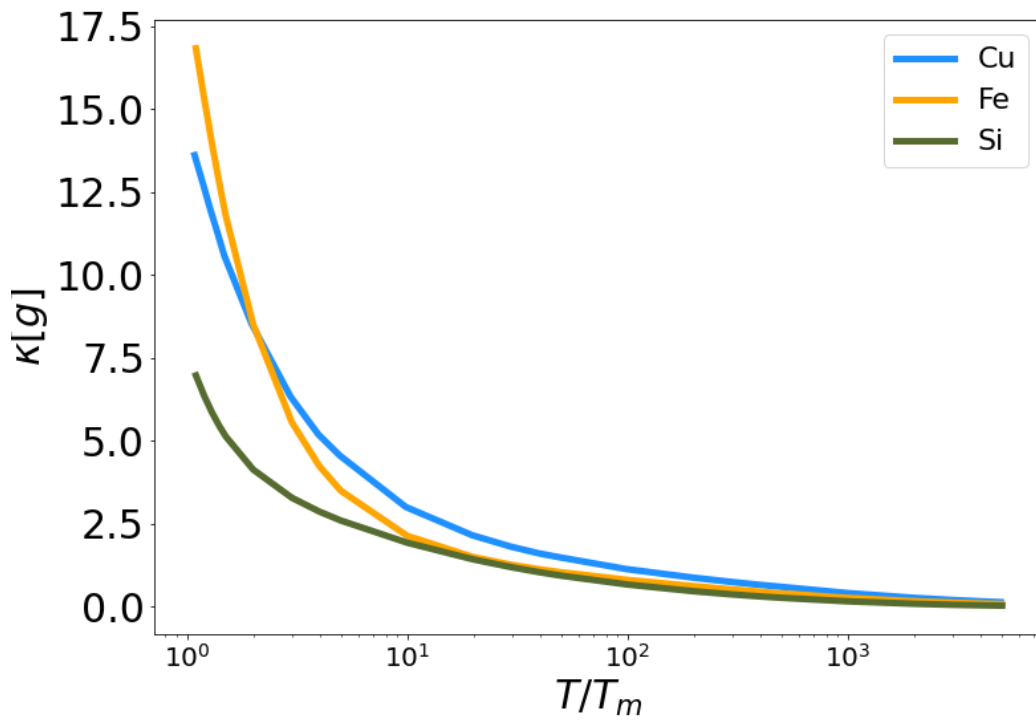
Figure 2-5. Low temperature corrective functional κ vs. reduced temperature for MEAM copper and iron, and Tersoff silicon.
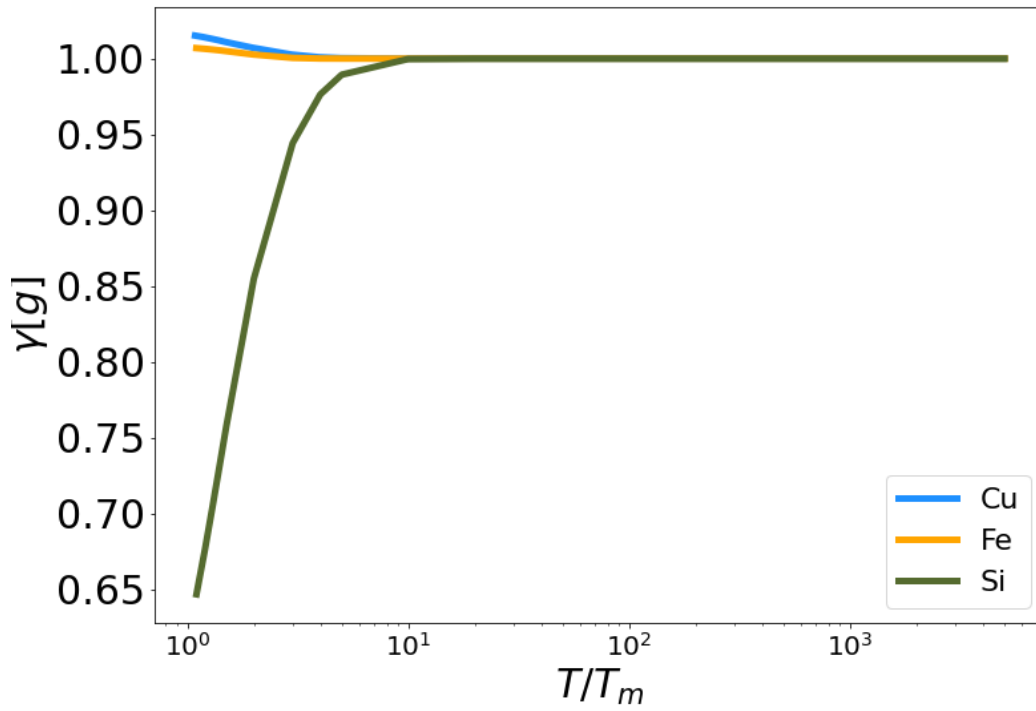
Figure 2-6. The fit low temperature corrective functional γ vs. reduced temperature for MEAM copper and iron, and Tersoff silicon.

Table 2-2. Summary of fit parameters and errors of EPFT and Kirkwood entropy functionals. Error is reported as the difference between the target entropy and the functional entropy. The fit parameters are unitless.

System	q_1	q_2	c_1	c_2	Kirk. Avg. Err (k_B/Atom)	EPFT Avg. Err (k_B/Atom)
Cu					0.32	0.06
Fe	3.24834	−2.406550	−320.1305	1.02966	0.41	0.08
Si					0.61	0.05

intermediate temperatures. At the PG limit the error in the EPFT for the MEAM systems approaches 0 k_B/atom.

In Figure 2-5, the functional $\kappa[g]$ is plotted for Cu, Fe and Si as a function of reduced temperature. This functional characterizes pre-melting structure in the liquid at temperatures close to the melt temperature, so it should deviate from zero as the temperature decreases. In each case the functional approaches zero at high temperature and becomes positive as the temperature approaches the melt temperature (reduced temperature of unity). Again, this functional, with no fitting parameter, is qualitatively similar for all three materials, when plotted with respect to reduced temperature.

In Figure 2-6, the functional $\gamma[g]$ is plotted for Cu, Fe and Si as a function of reduced temperature. The purpose of this functional is to influence the calculated entropy near the melting temperature, so it should deviate from unity only where there is pre-melting structure in the liquid at temperatures close to the melt temperature. In each case the functional approaches unity at high temperature. However, the functional increases for Si while decreasing to differing extents for Fe and Cu. Again, the parameters c_1 and c_2 were optimized to fit the excess entropy functional (Equation (2.5)) to the target entropy obtained via thermodynamic integration. The values of c_1 and c_2 are reported in Table 2-2. As with the parameters for the high temperature corrections, c_1 and c_2 are universal for the systems examined here.

Target Entropy

Figures 2-7, 2-8, and 2–9 show the solid and liquid target entropy developed as described in section III for each of the three systems we have investigated. For each composite figure, the left column describes the solid and the right column the liquid. The x axis is reported in absolute temperature from 1 K to the melting temperature (solid) and from the melting temperature to nearly ten million K (liquid). Each column contains three figures, the potential energy (top), the excess constant volume heat capacity (middle) and the excess entropy (bottom).

Collectively we observe several features of these thermodynamic properties, which qualitatively validate the simulations. The potential energies for all materials
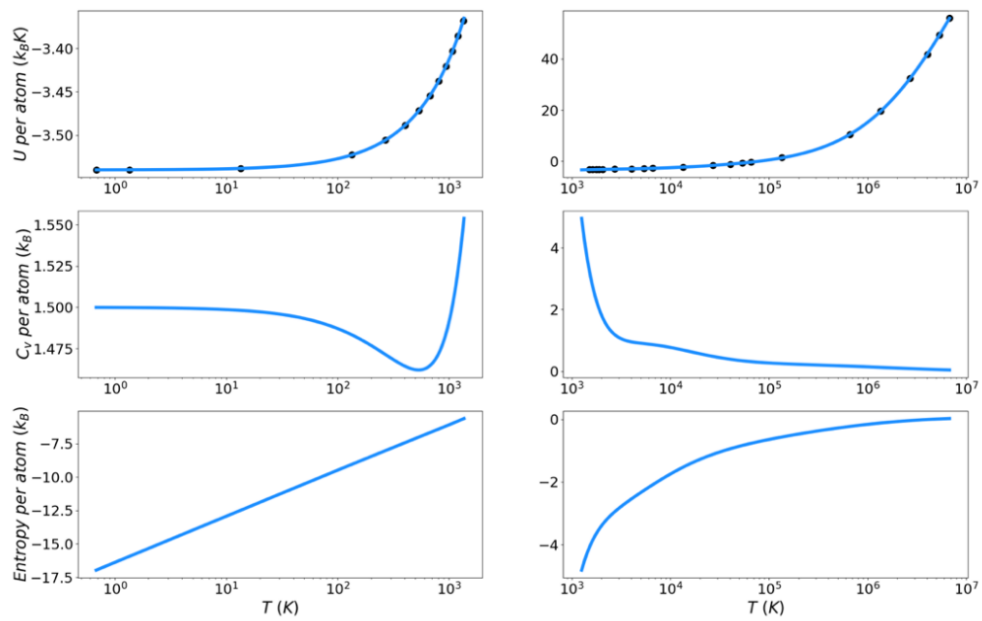
42

Figure 2-7. Thermodynamic integration development of solid and liquid entropy for copper.
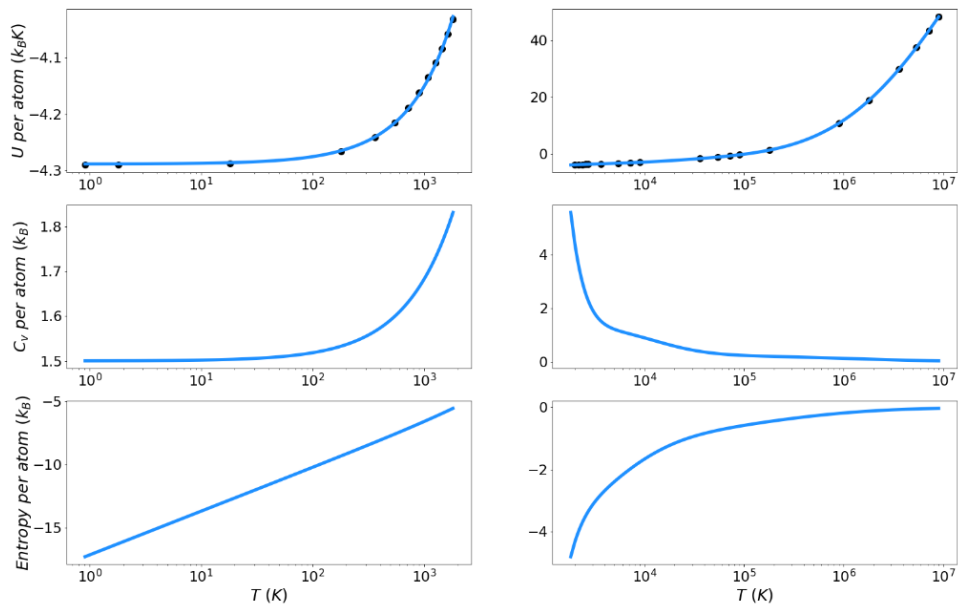
Figure 2-8. Thermodynamic integration development of solid and liquid entropy for iron.
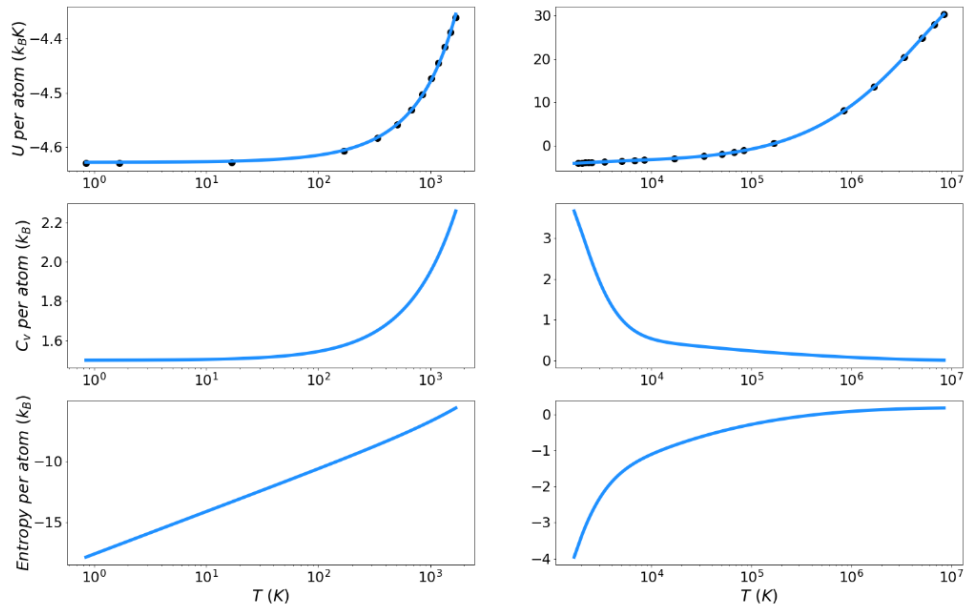
Figure 2-9. Thermodynamic integration development of solid and liquid entropy for silicon.

monotonically increase with increasing temperature. The excess heat capacities are always positive and thus the excess entropy monotonically increases with increasing temperature. The excess heat capacities further demonstrate three qualities deemed to represent the physical system accurately. First, the solid heat capacity approaches $3/2$ k_B on the approach to 0 Kelvin. Second, there is a sharp rise in heat capacity on the approach to the melt temperature in both the solid and liquid. Finally, the excess heat capacity approaches 0 as the temperature approaches infinity. These features provide confidence in the quality of the target entropies obtained through further thermodynamic integration. The only anomalous behavior from these target thermodynamic properties is that we observe, for some cases, unexpected fluctuations in the slope of the heat capacity immediately before (Cu) and after the melting temperature (Cu, Fe). When the scale of the y-axis is taken into account, these fluctuations are deemed to be minor. They are artifacts of fitting discrete points of the potential energy to an integrable function. In this work, the reference point for excess entropy is that it be zero at infinite temperature. To put the solid phase entropy on this same scale, the solid entropy must be shifted by a constant related to the entropy of fusion, as defined in Equation (2.13). It is important to remember that this constant is required only for the target entropy. It is not required by the entropy functional, which delivers an absolute excess entropy. These shift factors are reported in Table 2-3. The decomposition of the entropy shift into the three terms on the right-hand side of Equation (2.13) is also reported as fractions of the total.

Entropy Functionals

Excess entropy from the functionals has been plotted with the target entropy. These results for copper, iron, and silicon are included in Figures 2-10, 2-11, and 2-12 respectively. Comparable plots for Johnson Fe appear in in Figures 1–3 of Nicholson.[14] Collectively, in each plot, the reference entropy approaches zero in the high temperature limit. The excess entropy monotonically increases with increasing temperature. As a consequence of these two facts, the excess entropy is always negative.

Table 2-3. Excess entropy shift between solid and liquid phases and its decomposition.

System	ΔS_{tot} (J/mol/K)	ΔS_L (Fraction)	ΔS_s (Fraction)	$\Delta S_{L \to S}$ (Fraction)
Cu	−7.03	−0.930	0.540	1.390
Fe	−4.27	−1.177	0.743	1.434
Si	−13.82	−0.016	−0.009	1.025

Figure 2-10. Thermodynamic integration development of solid Excess entropy of copper, comparing target entropy from thermodynamic integration with values from solid and liquid functionals.

Figure 2-11. Thermodynamic integration development of solid Excess entropy of iron, comparing target entropy from thermodynamic integration with values from solid and liquid functionals.
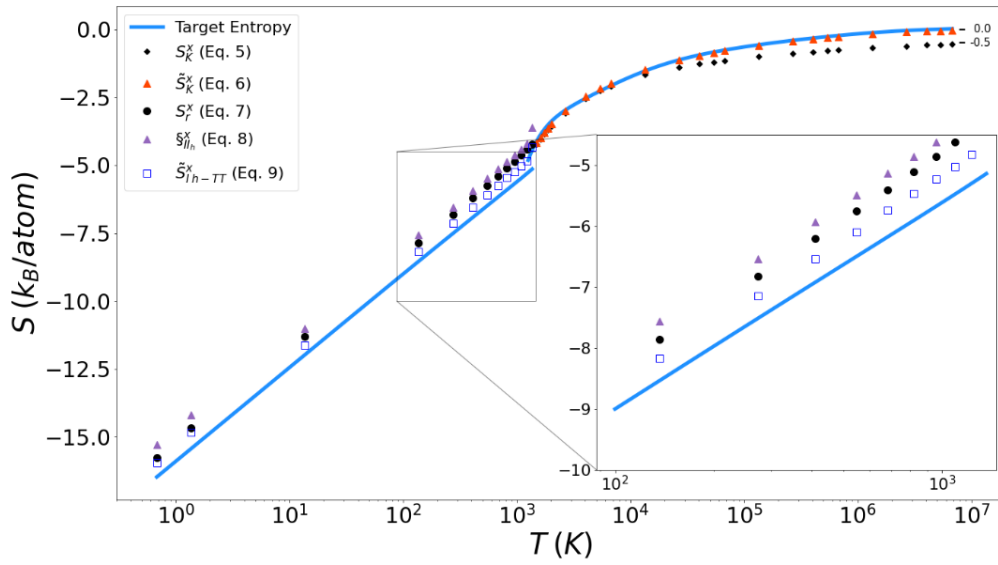
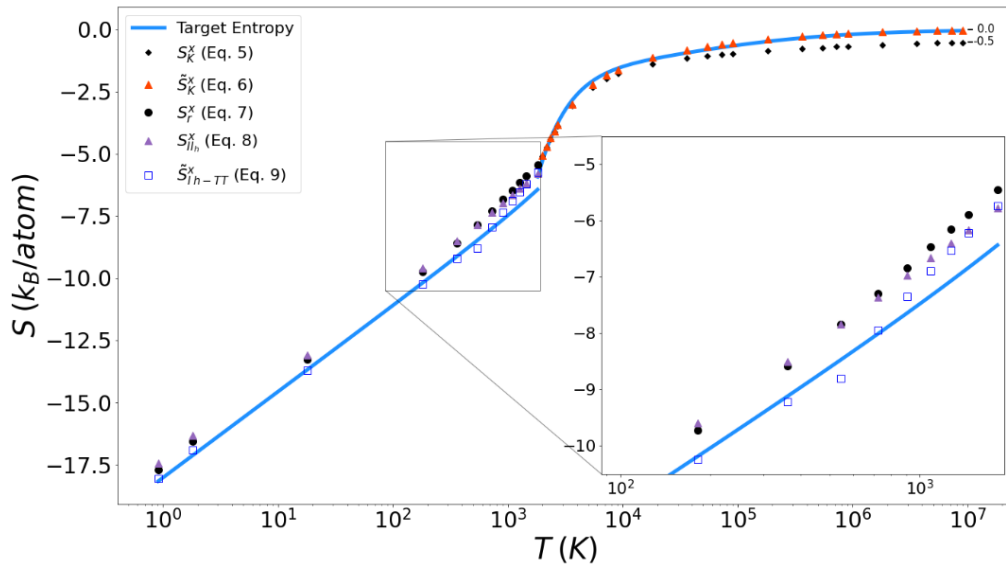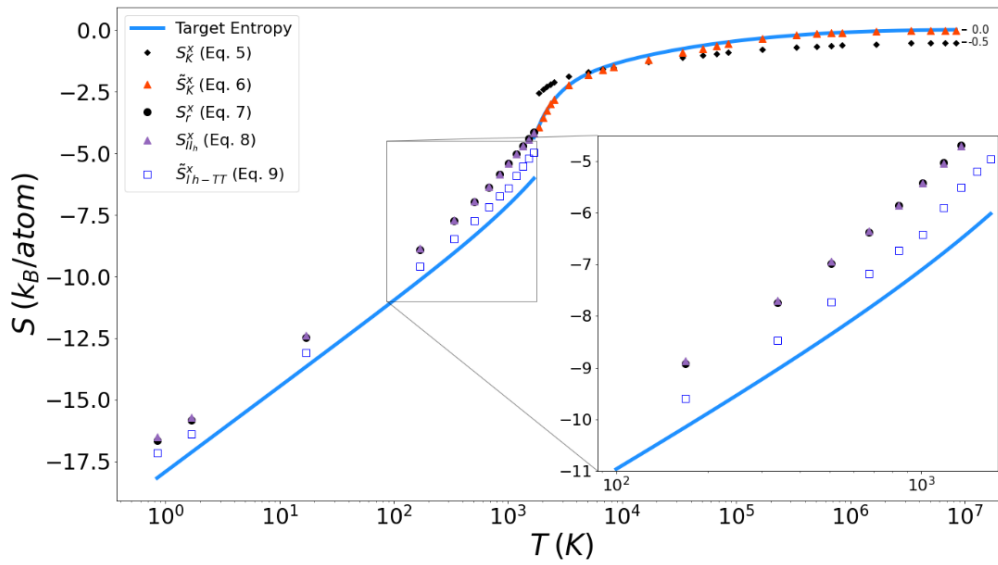**Figure 2-12. Thermodynamic integration development of solid Excess entropy of silicon, comparing target entropy from thermodynamic integration with values from solid and liquid functionals.**

On the liquid side, both the unmodified Kirkwood entropy (Equation (2.5)) and the modified Kirkwood entropy (Equation (2.6)), are plotted. In each case, the incorrect high temperature limit of the unmodified Kirkwood entropy is corrected by being shifted up a factor of ½ k_B. At intermediate temperatures, the presence of the $\tilde{\phi}[g]$ in the modified Kirkwood formulation significantly improves the ability of the functional to describe the simulated entropy. Near the melt temperature, the presence of the $\kappa[g]$, in the modified Kirkwood formulation, significantly improves the ability of the functional to describe the accelerated decrease in the simulated entropy.

In Figure 2-12 it can be seen that in general the Kirkwood entropy for Si varies from the target entropy differently than is the case for the Kirkwood entropy for Cu and Fe. For temperatures approaching T_m the Kirkwood entropy for Cu and Fe tend to undershoot the target while for Si it overshoots target entropy. There is also a subtle change in slope of the entropy for Si at intermediate temperatures in modified Kirkwood entropy, that is not evident in the Cu and Fe entropies.

On the solid side of the curve, we compare the target entropy with the three versions of the solid entropy functional explored in this work (Equations (2.7)–(2.9)). All models produce quantitatively similar results to the target entropy obtained from thermodynamic integration. The slope is well captured. The degree to which the intercept is captured varies. For Cu (Figure 2-10), Fe (Figure 2-11), and Si (Figure 2-12), Equation (2.9) gives the best fit. It is further observed that the trend in relative entropies from Equations (2.7)–(2.9) is not the same for all three materials.

Discussion

Table 2-2 shows that EPFT provides a significant reduction in error over the unmodified Kirkwood entropy. The error for the entropy of the EPFT compared to thermodynamic integration is reported in Table 2-2 and is below 0.1 k_B/atom for all three systems. In the case of silicon this represents an order of magnitude improvement over Kirkwood entropy.

The lower performance of Kirkwood entropy for the Tersoff silicon system most likely originates from the construction of the Tersoff potential, which takes into account multibody interactions through angle-dependent terms. As discussed, the Kirkwood entropy is ultimately a sub-functional of the EPFT and assumes pair potentials only. This, in fact, is the reason the multibody Tersoff potential was chosen, to test the compatibility of functionals accounting for only pair correlations with a multibody interaction potential. The new forms of $\tilde{\phi}[g]$ and $\gamma[g]$ presented take into account the effects of coordination number and density. These forms bring us closer to an accurate pair entropy functional and, therefore, overcome some of the limitations of Kirkwood entropy when applied to multibody potential systems.

It is well known that entropy should approach zero as temperature approaches absolute zero. The reader is reminded of two important restrictions we have imposed on this work. First, we are only addressing excess configurational entropy. We have developed the target entropy for this purpose from only the potential energy of the systems. This can be seen in the fact that the solid heat capacity approaches $3/2\,k_B$ as temperature approaches 0 Kelvin. By way of the equipartition theorem, this represents half of the $3\,k_B$ dictated by the law of Dulong Petit. Second, this work is based on classical MD simulations that do not take into account any quantum effects that begin to dominate as temperature approaches 0 K. A similar examination of the high temperature limit can be used to confirm the expected result of excess entropy approaching zero as the system approaches a configuration equivalent to that of a perfect gas.

Conclusions

The goal of this work was to explore the universality of a recently published Entropy Pair Functional Theory (EPFT) and its applicability to many-body interactions. The EPFT demonstrated that it could accurately describe materials obeying classical many-body (non-pairwise) potentials (MEAM and Tersoff) and we can expect that it would apply equally well to evaluation of the entropy to other many-body potentials, including first principles simulations. The practical value of EPFT is its potential to deliver excess

52

entropy from a single simulation. A suite of classical molecular dynamics simulations was performed for Cu (FCC, MEAM), Fe (BCC, MEAM) and Si (diamond cubic, Tersoff) over a temperature range from 1 K to over 10^6 K. Using thermodynamic integration, the excess entropy was calculated across this temperature range. The universality of EPFT was investigated by comparing the excess entropies of these materials from EPFT with the standard obtained from thermodynamic integration over the entire temperature range. The EPFT approach provides a significant improvement over Kirkwood entropy, yielding average errors of 0.06, 0.08, and 0.05 k_B/atom for Cu, Fe and Si, respectively.

As presented here, the EPFT approach to computing liquid phase excess entropies contains four parameters that are fit to simulation data obtained across a temperature range. These parameters are universal and can be made more robust as simulation data from additional systems is added to the data pool. Of note, the EPFT formalism for the solid phase contains no adjustable parameters and is universal to within the accuracy reported above. Utilization of EPFT to explore entropies of defective crystals and high entropy alloys is underway.

REFERENCES

[1] Z. H. Aitken, V. Sorkin, and Y. W. Zhang, Journal of Materials Research **34** (9), 1509 (2019).

[2] A. Ferrari, B. Dutta, K. Gubaev, Y. Ikeda, P. Srinivasan, B. Grabowski, and F. Kormann, J. Appl. Phys. **128** (15) (2020).

[3] S. Plimpton, J Comp Phys (117), 1 (1995).

[4] J. W. Gibbs, *Elementary Principles in Statistical Mechanics*. (Yale University Press, 1902).

[5] A. Baranyai and D. J. Evans, Physical Review A **40** (7), 3817 (1989).

[6] H. J. Raveche, J. Chem. Phys. **55** (5), 2244 (1971).

[7] D. C. Wallace, J. Chem. Phys. **87** (4), 2282 (1987).

[8] Y. Huang and M. Widom, **24** (Press) (2022).

[9] A. Togo and I. Tanaka, Scripta Materialia **108**, 1 (2015).

[10] M. C. Gao and M. Widom, Journal of Physical Chemistry B **122** (13), 3550 (2018).

[11] Y. Huang, M. Widom, and M. C. Gao, Physical Review Materials **6** (1) (2022).

[12] M. Widom and M. Gao, Entropy **21** (2) (2019).

[13] C. Desgranges and J. Delhommelle, Physical Review E **98** (6) (2018).

[14] D. M. Nicholson, C. Y. Gao, M. T. McDonnell, C. C. Sluss, and D. J. Keffer, Entropy **23** (2) (2021).

[15] R. A. Johnson, Physical Review a-General Physics **134** (5A), 1329 (1964).

[16] M. I. Baskes, Physical Review B **46** (5), 2727 (1992).

[17] J. Tersoff, Physical Review B **37** (12), 6991 (1988).

[18] J. G. Kirkwood, J. Chem. Phys. **3** (5), 300 (1935).

[19] R. E. Nettleton and M. S. Green, J. Chem. Phys. **29** (6), 1365 (1958).

[20] J. R. Morris and K. M. Ho, Phys. Rev. Lett. **74** (6), 940 (1995).

[21] P. W. Ma, S. L. Dudarev, and J. S. Wrobel, Physical Review B **96** (9) (2017).

[22] D. Perera, M. Eisenbach, D. M. Nicholson, G. M. Stocks, and D. P. Landau, Physical Review B **93** (6) (2016).

[23] J. Tranchida, S. J. Plimpton, P. Thibaudeau, and A. P. Thompson, Journal of Computational Physics **372**, 406 (2018).

[24] A. Erba, J. Baima, I. Bush, R. Orlando, and R. Dovesi, Journal of Chemical Theory and Computation **13** (10), 5019 (2017).

[25] S. A. Etesami and E. Asadi, Journal of Physics and Chemistry of Solids **112**, 61 (2018).

[26] G. P. P. Pun and Y. Mishin, Physical Review B **95** (22) (2017).

[27] Donald A. McQuarrie, *Statitical Mechanics*. (Harper & Row, New York, 1976).

APPENDICES

Appendix 2.A

Equations of the form A1 were used to fit potential energy where the order i is in-creased as necessary to obtain properly behaved heat capacity functions but not to exceed the number of potential energy data points.

$$U = \sum_i a_i ln(T)^i \tag{2.A1}$$

The constant volume heat capacity was determined as follows.

$$C_v \equiv \left(\frac{\partial U}{\partial T}\right)_v \tag{2.A2}$$

$$C_v = \sum_i \frac{i a_i ln(T)^{i-1}}{T} \tag{2.A3}$$

The target excess entropy, by which the EPFT was validated, was determined as follows.

$$S \equiv \int_{T_{ref}}^{T} \frac{C_v}{T'} dT' \tag{2.A4}$$

$$S = \sum_i \int_{T_{ref}}^{T} \frac{i a_i ln(T')^{i-1}}{T'^2} dT' = \sum_i i a_i \int_{T_{ref}}^{T} \frac{ln(T')^{i-1}}{T'^2} dT' \tag{2.A5}$$

Appendix 2.B

To determine the likelihood that a trial volume contains only one atom we determine the probability that $N - 1$ atoms in the system are outside of the trial volume. For a system of N atoms and trial volume $V = 4/3\pi r^3$ and a total system volume $V_s = 4/3\pi R_s^3$ the probability that any one of the N atoms is outside the trial volume is

$$p^{o1}(r) = 1 - \frac{4/3\pi r^3}{N4/3\pi R_s^3} \tag{2.B1}$$

The joint probability that all but one of the atoms in the system are outside the trial volume then becomes

$$p^o(r) = \left(1 - \frac{r^3}{NR_s^3}\right)^{N-1} \tag{2.B2}$$

The binomial theorem can be used to expand equations of this form. Setting $a = r^3/R_s^3$

$$p^o(r) = \lim_{N \to \infty} \left(1 - \frac{a}{N}\right)^{N-1} = 1 - \frac{a}{1!} + \frac{a}{2!} - \frac{a}{3!} \tag{2.B3}$$

This series is recognizable as the expansion of e^a and forms the basis for the high temperature correction $Q[r]$.

$$p^o(r) = \exp(-a) = \exp\left(-r^3 / R_s^3\right) \tag{2.B4}$$

CHAPTER 3

HIGH TEMPERATURE ERROR – SOURCES AND SOLUTIONS

Abstract

Direct entropy calculation for liquids are known to underestimate the excess entropy in the perfect gas [PG] limit. From a combinatorial perspective this error has been shown to originate from the absence of higher order correlation terms in the Kirkwood entropy. This work explores the connection between the combinatorial physical and configuration perspectives, ultimately providing a means to extract the higher order correlation information from the pair correlation function [PCF].

Introduction

Kirkwood Entropy

A functional of g(r) based on the Kirkwood approximation has been demonstrated as a means of direct calculation of the excess entropy for liquids[1-5]. The Kirkwood entropy functional is constructed as follows.

$$\tilde{S}_K^x[g] = \lim_{R \to \infty} \frac{1}{2}\{-1 + \rho \int_0^\infty dr\, g(r)(lng(\mathrm{r}) - (g(r) - 1)) \tag{3.1}$$

Where $g(r)$ is a pair correlation function defined as

$$\int_0^\infty \rho g(r) 4\pi r^2 dr = N - 1 \approx N \tag{3.2}$$

However, the so called Kirkwood entropy functional is known to possess a $\frac{1}{2}k_B/atom$[1] error at high temperatures and to decrease without bound upon approach to the melting temperature. This is evident in Figure 3-1 where the high temperature error can be seen to range from 0 at an intermediate temperature up to $1/2k_B$ at the perfect gas (PG) limit. Recently, Nicholson et al have presented the entropy pair functional theory (EPFT) demonstrating the direct calculation of entropy for $T = 0\,K$ to the PG limit[3]. In addition to extending the Kirkwood functional to solids, EPFT provides a functional form for liquid entropy that corrects the shortcomings of the Kirkwood entropy.
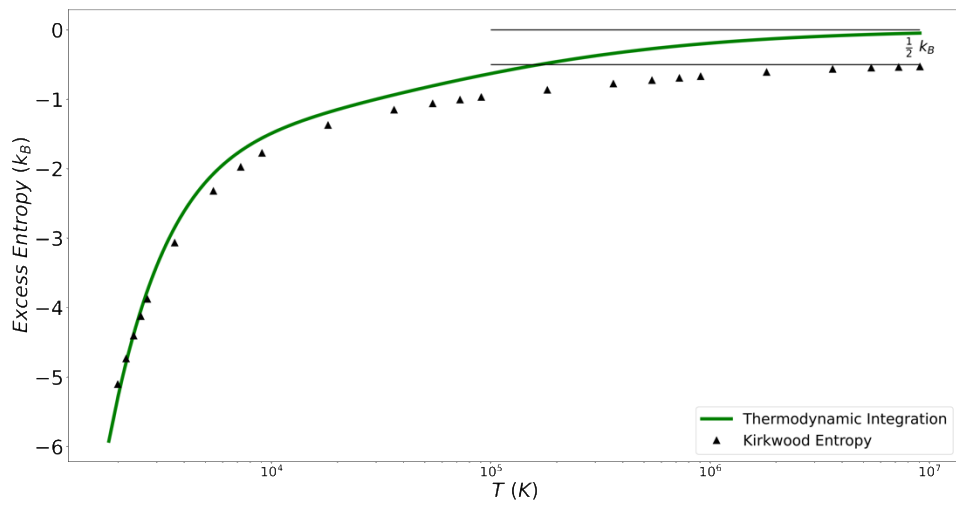
Figure 3-1. Excess entropy of MEAM iron from thermodynamic integration compared to calculation from the Kirkwood entropy.

The high temperature correction proposed by EPFT is in the form of a sub-functional $Q[g]$ that is fit with two parameters and included as $\tilde{\phi}[g]$.

$$\tilde{S}_K^x[g] = \frac{1}{2}\tilde{\phi}[g] - 1 + \lim_{R\to\infty}\frac{1}{2}\{-1 + \rho\int_0^\infty dr\, g(r)(lng(r) - (g(r) - 1)) \tag{3.3}$$

Where $\tilde{\phi}[g]$ is defined as

$$\tilde{\phi}[g] = Q + q_1 Q(1 - Q) + q_2 Q^2(1 - Q) \tag{3.4}$$

Parameters q_1 and q_2 in Equation 3.4 are fit to a target entropy developed by thermodynamic integration across the full range of temperature. In this work we examine the particularities of the functional $Q[g]$ specifically the form utilized by Sluss et al in their extension of the EPFT[6].

Background

Origins of the error in the PG limit

For systems of size N and correlations up to n = N, the Kirkwood entropy is defined as

$$\tilde{S}_K^x[g] = \frac{1}{2}\left\{-1 + \rho\int_0^\infty dr\, g(r)^{(2)}\left(lng(r)^{(2)} - \left(g(r)^{(2)} - 1\right)\right)\right\} \tag{3.5}$$
$$+ \frac{1}{6}\left\{-1 + \rho^2\int_0^\infty dr\, g(r)^{(3)}\left(ln\delta g(r)^{(3)}\right.\right.$$
$$\left.\left. - \left(g(r)^{(3)} + 3g(r)^{(2)}g(r)^{(2)} + 3g(r)^{(2)} - 1\right)\right)\right\}\ldots$$

The three terms in Equation (3.5) are referred to as the one, two, and three particle entropies. $\delta g(r)^{(n)}$ is the factor used to regain the complete n order correlation from the results of the superposition approximation based on pair correlations only[1,5,7]. Computation of the higher order terms is impractical and it is common to truncate the

terms higher than pair correlation. In this simplified case of n=2 Equation (3.5) reduces to Equation (3.2).

Baranyai et al provided a connection between the $1/2k_B$ error at the PG limit and the absence of higher order correlation terms in the Kirkwood entropy. As a review, $g(r)^{(n)}$ at the PG limit and at large r is

$$g(r)^{(n)} = \frac{N(N-1)\dots(N-n-1)}{N^n} \tag{3.6}$$

For pair correlations only, $n = 2$ and at large r

$$g(r)^{(2)} = \frac{N(N-1)}{N^2} = 1 - \frac{1}{N} \tag{3.7}$$

Equation (3.5) can be extended to include all n-particle terms up to $n = N$. By calculating these higher order contributions at the PG limit, Baranyai simply points out that at the PG limit the pair correlation terms in Equation (3.5) are equal to $1/2k_B$ and the sum of all higher order correlation terms account for an additional $1/2k_B$. It is clear then that the source of the $1/2k_B$ error in the Kirkwood entropy results from the omission of the contributions of the higher order $\delta g(\text{r})^{(n)}$.

One possible conclusion from the results above may be that $g(r)^{(2)}$ does not contain sufficient information for the calculation of entropy. In fact, for any system governed by pair interactions alone, all higher order correlations are wholly dependent upon the pair correlations. Nicholson et al have provided an ad absurdum proof of statement, demonstrating that for systems described by pair potential Hamiltonians pair correlations determine all higher order correlations. This means that the pair correlation $g(r)^{(2)}$ contains all of the configurational information needed for the calculation of the entropy and that the source of the $1/2k_B$ error at the PG limit is not attributable to a deficiency in the pair correlation $g(r)^{(2)}$ but only to the truncation of the higher order terms in the Kirkwood entropy. The remainder of the theory we will discuss here will deal specifically

with the pair correlation functions and as such we will refer to $g(r)$ without the order of correlation designator.

In addition to the mathematical origins above it is also necessary to develop some physical explanation for the high temperature error. There are two qualities of the $g(r)$ expression in Equation (3.2) that are important to remember at this time. First, $g(r)$ provides the probability of finding an atom at a distance r from a central atom. For this reason $g(r)$ is normalized to be equal to 1 at large r meaning that for any incremental increase in r at long distances from the central atom there is unity probability of finding another atom. In other words, at sufficiently large separation there is no correlation between two atoms in the system. Second, the right hand side of Equation (3.1) represents the total number of atoms in the volume $\frac{3}{4}\pi R^3$ less the central atom from which the separation distance r is measured. This is equivalent to observing that $g(r) = 0$ for r in the region around the central atom. The small r region where $g(r) = 0$ is referred to as the excluded volume and is ultimately defined by the governing interaction potential of the system.

Figure 3-2 provides an insight into the evolution of $g(r)$ over the temperature range in question. As temperature increases two phenomena are readily observable. First, as expected correlation between atoms decreases with temperature. Second, the excluded volume decreases with temperature. The first observation is directly in line with the definition of a perfect gas which assumes no interaction between particle and consequently there is no correlation. The second observation is a result of the energy of the system overcoming the repulsive component of the interaction potential as the temperature increases. Imperceptible in Figure 3-2, but of note, is that the value of $g(r)$ at large r has also been lowered at the PG limit ($10^6 \, K$) by $\frac{1}{N}$ as indicated in Equation (3.7).
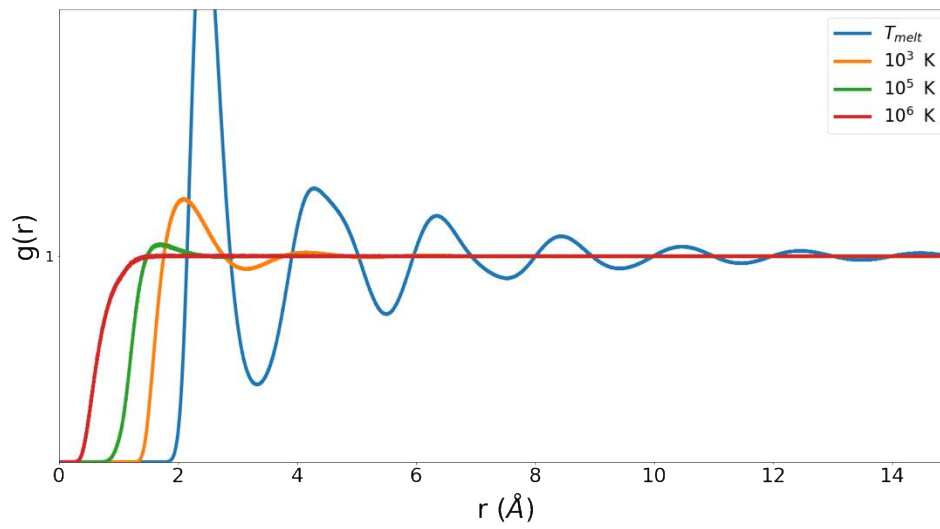
Figure 3-2. Radial distribution functions for MEAM iron across a full range of liquid temperatures.

High Temperature Correction

Direct Calculation from g(r)

Because the pair correlation $g(r)$ contains all of the configurational information needed to calculate entropy, The central challenge in the development of $Q[g]$ is to craft the functional in such a way as to utilize the relevant information present in $g(r)$. We know that a proper $Q[g]$ will increase monotonically from 0 at some intermediate temperature to 1 at the PG limit. We also require that $Q[g]$ is dependent only on $g(r)$ and that it is computationally accessible (is not of a form prone to errors related to machine precision). With all that in mind we begin the construction of the high temperature correction functional $Q[g]$ by ensuring the proper output at the low and high temperature limits. Namely, for the liquid just above T$_{\text{melt}}$, $Q[g] = 0$ and at T$_{\text{PG}}$, $Q[g] = 1$. We know from Equation (3.2) that for the liquid, $4\pi\rho \int_0^\infty r^2 g(r) dr = N - 1$. So we can say that

$$Q[g] = 4\pi\rho \int_0^R r^2 g(r) dr + x = 0 \tag{3.8}$$

Clearly, $x = 1 - N$. Because $4\pi\rho \int_0^\infty r^2 dr = N$, we can write $Q[g]$ as,

$$Q[g] = 1 - 4\pi\rho \int_0^R r^2 (1 - g(r)) dr \tag{3.9}$$

Next we insert Equation (3.7) into (3.9)

$$Q[g] = 1 - 4\pi\rho \int_0^R r^2 \left(1 - (1 - \frac{1}{N_{sim}})\right) dr = 1 - \frac{N_{sph}}{N_{sim}} \tag{3.10}$$

To ensure $Q[g]_{PG} = 1$ we then normalize $Q[g]$ with the right hand side of Equation (3.9).

$$Q[g] = \frac{1 - 4\pi\rho \int_0^R r^2 (1 - g(r)) dr}{1 - \dfrac{N_{sph}}{N_{sim}}}$$

(3.11)

N_{sph} is the number of atoms in the volume $V_g = \frac{4}{3}\pi R^3$ (the volume over which $g(r)$ is calculated). N_{sim} is the total number of atoms in the simulation cube. For this reason the denominator of equation 11 can be rewritten as f_{out}, the fraction of atoms in the simulation that are found outside the $g(r)$ sphere. A similar change of variable in the numerator of Equation (3.11) results in the exact form of $Q[g]$ presented by Nicholson et al.

$$Q[g] = \frac{1 - N_{diff}}{f_{out}}$$

(3.12)

Here N_{diff} represents the difference between the number of atoms inside V_g for an evenly distributed density and the number of atoms in V_g as determined by the distribution $g(r)$. This change of variables exposes the $Q[g]$ as a additional entropy functional. Ultimately, $g(r)$ is the correlation function that modifies the trivial distribution ρ (bulk density) to provide the r dependent density [7]. At the PG limit there is no correlation and $g(r)$ provides a unity modification of the bulk density. For this reason it would be expected that N_{diff} would approach 0 at the PG limit. Here we have made the case that at the PG limit $N_{diff} = f_{in}$ and that this is a direct result of the deficiencies of $g(r)$ that only include pair correlations. In this way $Q[g]$ accounts for the missing information described by Baranyai [1]. Furthermore, as $Q[g]$ is measuring the shift from r dependent density (order) to evenly distributed density (disorder), it is an entropy functional in its own right.

Unfortunately, Equation (3.12) demonstrates two key shortcomings. First, for lower temperatures $Q[g]$ is a function of the cutoff distance R_{cut} used in the calculation of N_{diff}. This is not particularly surprising when it is considered that the lower temperatures have densities of higher r dependence. The second issue with $Q[g]$ arises due to the assumption implicit in the inclusion of Equation (3.7). By saying that in the PG

limit that $g(r) = 1 - \frac{1}{N}$ we are assuming that the large r behavior of $g(r)$ is sufficient to describe the entire range of $g(r)$. For certain interaction potentials that have a finite repulsion energy at small r, this assumption is valid. However, for many common interaction potentials this assumption does not hold and consequently $Q[g]$ is not able to measure the system changes on the approach to the PG limit.

It is important to highlight the origin of the applicability of Equation (3.7) to finite interaction potentials and consequently why it fails for potentials that approach infinity at zero separation. In Figure 3-3 it can be seen that at the PG limit the Johnson potential provides for a non-zero chance for neighbors to be found at zero r. In contrast, the modified embedded atom method (MEAM) potential has a clearly defined region around the central atom where the chance of finding a neighbor atom is zero. We refer to this region as the excluded volume. For the Johnson potential we observe that the lack of a real excluded volume is advantageous to the applicability of Equation (3.7). We know that $\int_0^r g(r)dr$ is a volume fixed by the limits of integration. For this reason, any volume that is accounted for in the small r region must be absent at larger r. This provides a physical interpretation for the shift from $g(r) = 1$ to $g(r) = 1 - \frac{1}{N}$ for large r at the PG limit. Specifically, in this interpretation, the volume that would be reserved for the central atom is now being shared or distributed by all of the atoms in the system. From a computational perspective, because Equation (3.7) assumes a singular value for $g(r)$ across all r, the small r characteristic of the Johnson potential introduces less error that MEAM.

An Idealized Model of N$_{diff}$

It can be seen from the temperature evolution of $g(r)$ in Figure 3-2 that systems governed by the MEAM potential demonstrate some reduction in the size of the excluded region on the approach to the PG limit. This provides motivation to improve the method of calculation of N_{diff} for MEAM and similar systems. As a first step however, we direct our attention to eliminating the r dependence of $Q[g]$. This is accomplished by
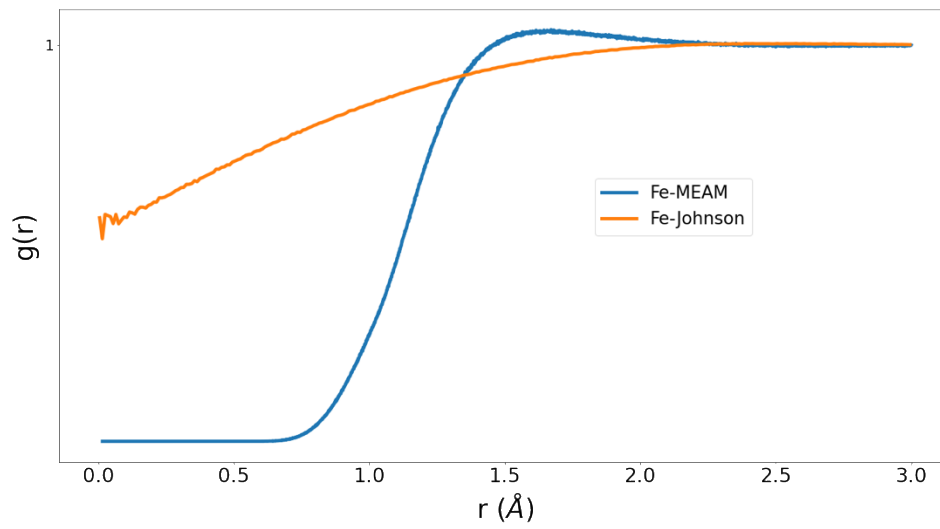
Figure 3-3. Comparison of small r behavior of systems governed by finite (Johnson) and infinite (MEAM) repulsive potentials.

calculating $g(r)$ over the entire simulation cube, for R in range $\left(0, \frac{\sqrt{3}}{2}\alpha\right)$ where α is the length of a simulation box side. $g(r)$ calculated in this manner will be referred to as the cumulative RDF or CRDF.

$N_{diff}[g,r]$ is the difference between the average number of atoms in a sphere and the observed number of atoms based on integration of the RDF. The sphere used in the definition of N_{diff} and $Q[g]$ is allowed to become larger than half the box size, which means the appropriate volume is no longer a sphere but the cube-sphere intersection (csi) volume, V_{csi}. Figure 3-4 provides a visual progression for the csi volume as the RDF sphere radius increases from half the box length up to half the box length times $\sqrt{3}/2$. For RDFs calculated over the csi volume, we redefine N_{diff} and $[g]$,

$$N_{diff}[g,r] = \rho \int_{V_{csi}(r)} 1 - g(r')dr' \tag{3.13}$$

$Q[g]$ is a functional of the RDF defined as

$$Q_{csi}[g] = \frac{1 - N_{diff}[g,r]}{1 - \dfrac{V_{csi}}{V_{cube}}} \tag{3.14}$$

It turns out that $N_{diff}[g,r]$ is a strong function of r, the upper limit of integration, making $Q[g]$ a function of r, which is undesirable from both practical and theoretical grounds. In Equation (3.14), we observe that when $r/\alpha \geq \sqrt{3}/2$, $V_{csi}(r) = V_{cube}$ so that $Q\left(\sqrt{3}/2\right)$ is formally undefined, 0/0. However, in practice, we have observed that the $Q\left(\sqrt{3}/2\right)$ converges to a finite number bounded between 0 and 1. Furthermore, the definition in Equation (3.14) yields $Q[g,r]$ as a function of r with substantial statistical noise. Figure 3-5 exemplifies the r dependance present in $Q[g]$.

To obtain $Q[g]$ independent of r and subject to less noise, we fit $N_{diff}[g,r]$ to a model. The model RDF of the liquid phase can be described in terms of two regions. In the first region, there is some excluded volume. In this region,
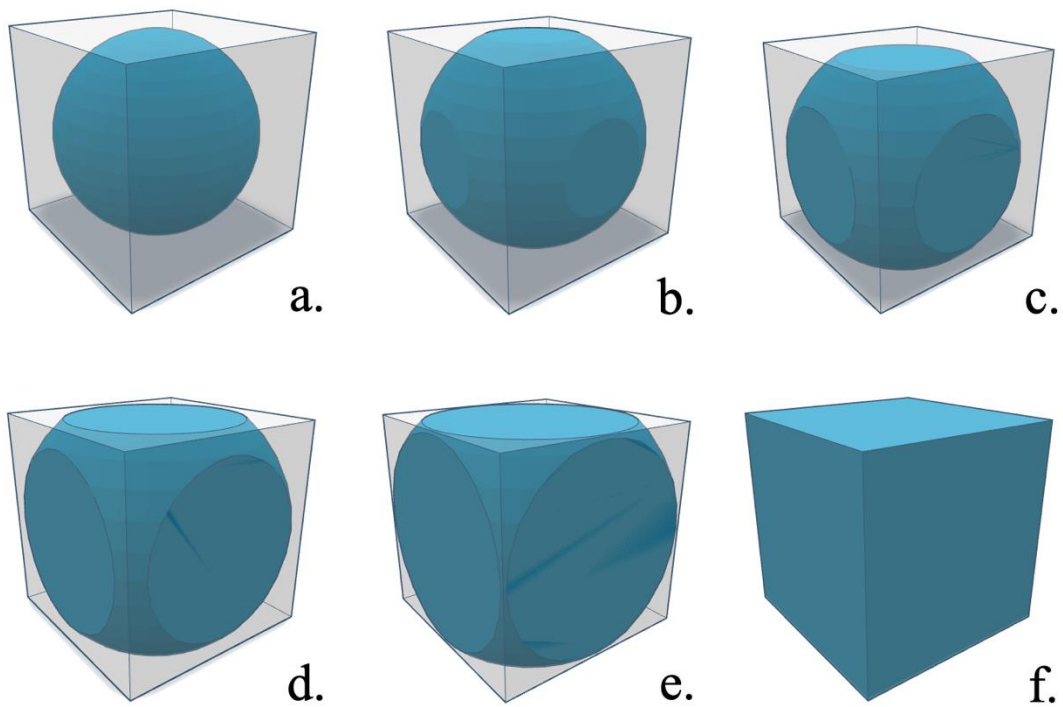
69

Figure 3-4. Evolution of the cube sphere intersection (csi) volume V_{csi}. For sphere radius r and half box length α, V_{csi} increases a. to f. for $r = \alpha$ (a.) to $r \geq \alpha\frac{\sqrt{3}}{2}$ (f.).
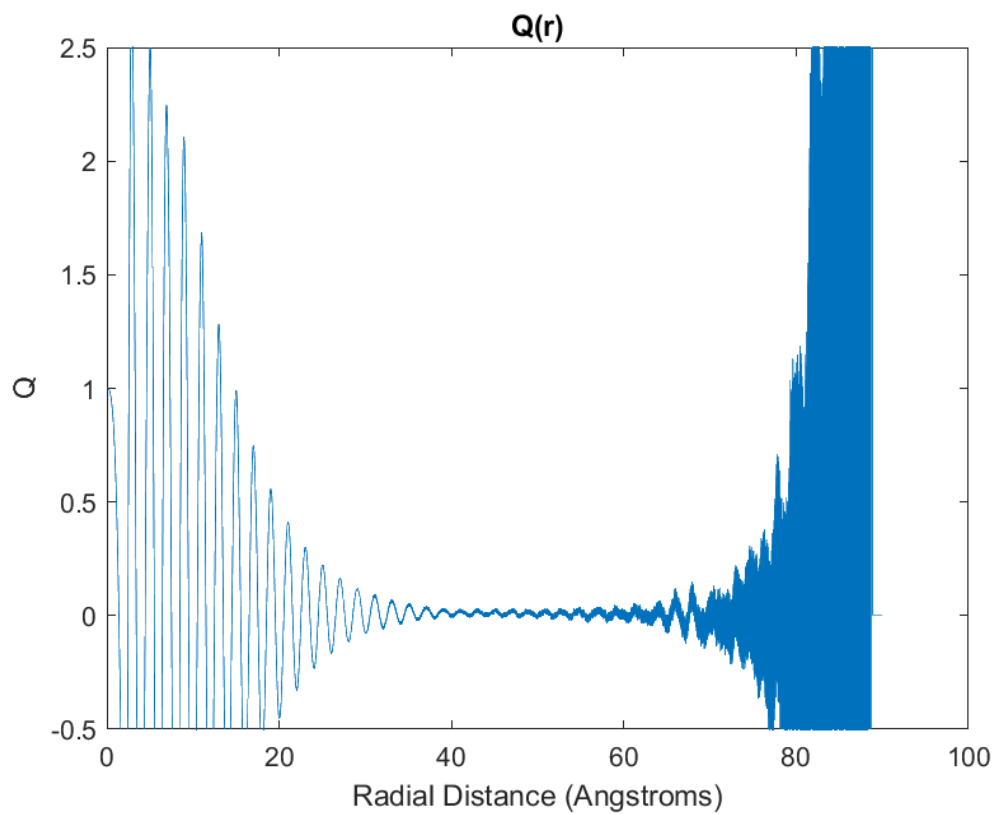
Figure 3-5. $Q[g]$ **for MEAM Fe developed from the CRDF demonstrating strong** r **dependance.**

$$g_{region\_1}(r) = 0 \ \text{for} \ 0 \leq r \leq r_{exclude} \tag{3.15}$$

where the upper limit of region 1, is given by $r_{exclude}$, which defines the excluded volume of the central atom,

$$V_{exclude} = \frac{4\pi}{3} r_{exclude}^3 \tag{3.16}$$

where the excluded volume can be related to the fraction of an excluded atom at the core,

$$V_{exclude} = \frac{f_{exclude}}{\rho} \tag{3.17}$$

where $0 \leq f_{exclude} \leq 1$. $r_{exclude}$ or equivalently $f_{exclude}$ is the only adjustable parameter in this model. Importantly, it will be shown below that for this model RDF, $Q[g]$ uniquely defined as

$$Q_{csi}[g] = 1 - f_{exclude} = 1 - \rho V_{exclude} = 1 - \rho \frac{4\pi}{3} r_{exclude}^3 \tag{3.18}$$

which has the desired property of being independent of r beyond the initial excluded volume. If we optimize the parameter, $f_{exclude}$, in this model to fit simulation data it will necessarily output Q.

In the second region, there is a plateau in the RDF, but the value of this plateau is set so that the finite size of the simulation box is realized. This manifests in the fact that N_{diff}. must go to unity at values of r which enclose the entire simulation box, $r/\alpha \geq \sqrt{3}/2$. This constraint dictates the value of the RDF to be

$$g_{region\_2}(r) = \frac{N_{cube} - 1 - N_{region\_1}}{\rho V_{region\_2}} \ \text{for} \ r_{exclude} \leq r \tag{3.19}$$

where N_{cube} is the number of atoms in the cubic simulation box, $N_{regions\_1}$ is the total number of atoms in region 1, as given by the cumulative RDF, r is the average density and V_{region_2} is the volume of region 2, that portion of the cubic simulation volume that extends beyond region 1.

$$V_{region\_2} = V_{cube} - V_{region\_1} \tag{3.20}$$

Also, the number of atoms in region 1, N_{region_1} is zero, since $g_{region\_1}(r) = 0$. In this model,

$$g_{region\_2}(r) = \frac{N_{cube} - 1}{\rho V_{region\_2}} \text{ for } r_{exclude} \leq r \tag{3.21}$$

There are no adjustable parameters in region 2.

From an analysis of a simulation, we have the RDF and cumulative RDF (CRDF). Both functions extend to the corner of the cubic unit cell, using the formula for the cube-sphere-intersection. We perform a single-variable non-linear optimization in the only parameter of our model RDF, namely, $f_{exclude}$, or equivalently $Q = 1 - f_{exclude}$, to match the simulation data. We tried both the Nelder and Mead's Downhill Simplex Method (the amoeba method) and the Polak-Ribiere Conjugate Gradient method, without alteration from <u>Numerical Recipes</u>.[8] In general, this is a very robust one-dimensional optimization. It turns out that $N_{diff}[g, r]$ has much less noise than $Q[g, r]$ (as determined in Equation (3.14)), as a so we optimize to $N_{diff}[g, r]$. Since Q is essentially a parameter in the RDF, the optimization returns Q. Figure 3-6 shows N_{diff} along with the single parameter ($f_{exclude}$) model used to produce stable values of Q for the high temperature correction.

Rethinking the Excluded Volume

In the first version of $Q[g]$ (Equation (3.12)) we developed a measurement of the r dependence of density from high dependence around T_{melt} to zero dependence at the PG limit. We also provided a tentative relationship between this transition and the shrinking of the excluded volume at high temperatures. The key insight provided by the model of N_{diff} and $Q_{csi}[g]$ (Equation (3.18)) is that the essential information for the high temperature correction is contained within the small r regions of $g(r)$ and its dependent functionals. We now turn out attention to finding more direct access to the excluded volume information contained in $g(r)$.
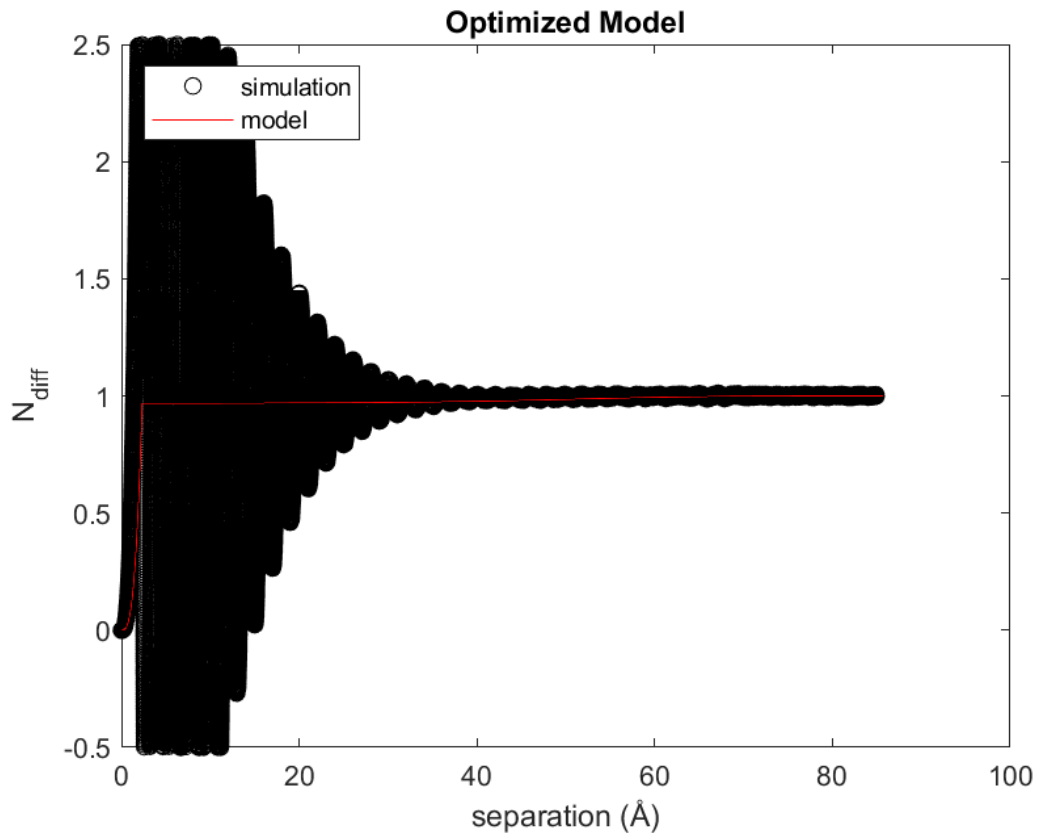
73

Figure 3-6. N_{diff} for MEAM Fe with single parameter model.

Interaction potentials by definition provide the relationship between energy and the atomic separation. It is clear that as the temperature of the system increases the repulsive energy barrier is overcome to some degree and atoms can be found closer together. This is one mechanism causing the shrinking of the excluded volume at high temperature. It is important to remember two things at this time. First, an important precept of EPFT is a functional of $g(r)$ not temperature. Second, the observation that $Q[g]$ is measuring the extent to which density is dependent upon r remains true. For these reasons we take the view that while the interaction potential provides a governing Hamiltonian for the system, it is the resulting configurational information that determines the entropy. What is needed then is a model of the excluded volume robust enough to account for a variety of atomic configurations found in nature yet measurable in $g(r)$. For this model we choose the polyhedron defined by the Voronoi tessellation of the entire simulation space. Voronoi polyhedra, VP, are commonly used in the study of material systems and when applied to a crystal lattice are known as Wigner-Seitz cell [9-14]. In the crystal, VP have faces equal in number to the coordination number of the system and defined by planes perpendicularly bisecting the distance between neighboring atoms. Because the model excluded volume must be measurable from $g(r)$ we translate the VP to a sphere with equal volume. Observing that the sum of the volume of every VP in a system is equal to the total volume of a the system, we consider the inverse of the bulk density, $\frac{1}{\rho}$, to be equivalent to the volume of the VP and define the radius, R_s, of the model excluded volume below.

$$R_s = \left(\frac{3}{4\pi\rho}\right)^{\frac{1}{3}} \tag{3.22}$$

The probability of finding any of the N atoms in a system outside of the model excluded volume is then,

$$p^{o1}(r) = 1 - \frac{4/3\pi r^3}{N 4/3 \pi R_s^3} \tag{3.23}$$

Here r is the separation between an atom and the centroid of the model excluded volume. The probability of finding all but one of the atoms outside of the trial volume is,

$$p^o(r) = \left(1 - \frac{r^3}{NR_s^3}\right)^{N-1} \tag{3.24}$$

Applying the binomial theorem to Equation (3.23) results in a more simplified form.

$$p^o(r) = \exp\left(-\left(\frac{r}{R_s}\right)^3\right) \tag{3.25}$$

We can then define the intrusion into the trial volume for any $g(r)$ as,

$$I[g] = \int 4\pi r^2 \rho P^{id}(r)g(r)dr \tag{3.26}$$

We define $Q[g]$ as,

$$Q[g] = \frac{I[g] - I_{min}}{1 - I_{min}} \tag{3.27}$$

I_{min}, is the minimum intrusion into the excluded volume and provides a reference for the full scale measurement $Q[g]$.

Results

Using MEAM iron as a test case, a comparison of the three forms of $Q[g]$ discussed was performed. The three forms of $Q[g]$ discussed here are presented across the liquid temperature range for MEAM iron in Figure 3-7. Equation (3.18) shows a vast improvement over the untenable results of Equation (3.12). The VP based approach, Equation (3.27), provides the least noise of the three techniques.

Discussion and Conclusions

We have provided an explanation for the high temperature error found in the Kirkwood entropy functional. This explanation connects the known combinatorial deficiencies at

76
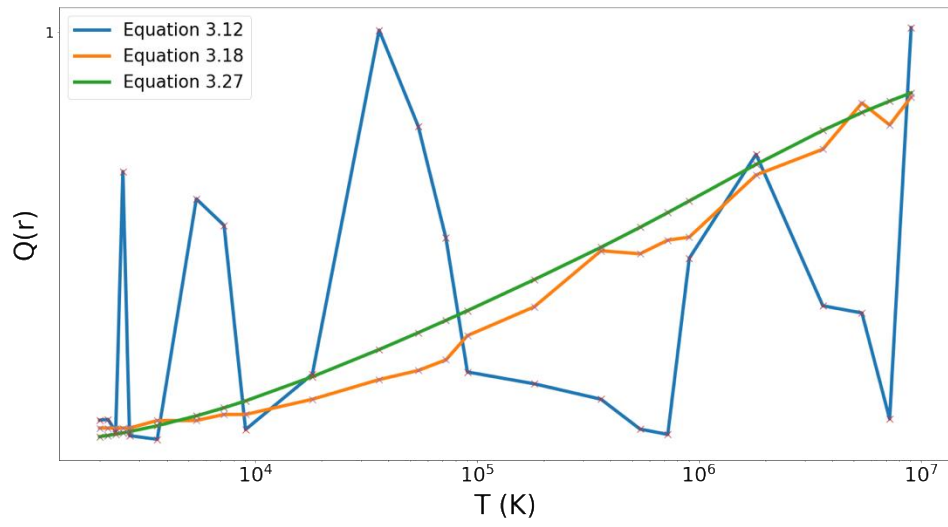
Figure 3-7. Comparison of three methods to calculate Q[g] for liquid MEAM iron across a full range of liquid temperatures.

the PG limit when only pair correlations are considered to the evolution of the r independence of density. Furthermore, we have demonstrated that a high temperature correction functional based on this explanation is a first class measure of disorder as well as providing a state of the art correction to PCF entropy functionals.

REFERENCES

[1]   A. Baranyai and D. J. Evans, Physical Review A **40** (7), 3817 (1989).

[2]   R. E. Nettleton and M. S. Green, J. Chem. Phys. **29** (6), 1365 (1958).

[3]   D. M. Nicholson, C. Y. Gao, M. T. McDonnell, C. C. Sluss, and D. J. Keffer, Entropy **23** (2) (2021).

[4]   H. J. Raveche, J. Chem. Phys. **55** (5), 2244 (1971).

[5]   D. C. Wallace, J. Chem. Phys. **87** (4), 2282 (1987).

[6]   Clifton C. Sluss, Jace Pittman, Donald M. Nicholson, and David J. Keffer, Entropy **24** (5), 603 (2022).

[7]   Donald A. McQuarrie, *Statitical Mechanics*. (Harper & Row, New York, 1976).

[8]   William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3 ed. (Cambridge University Press, 2007).

[9]   G. Lejeune Dirichlet, Journal für die reine und angewandte Mathematik (Crelles Journal) **1850** (40), 228 (1850).

[10] F. C. Frank and J. S. Kasper, Acta Crystallographica **11** (3), 184 (1958).

[11] Charles Kittel, *Introduction to solid state physics*, 8th ed. (Wiley, Hoboken, NJ, 2005).

[12] Steven H. Simon, *The Oxford Solid State Basics*, First edition. ed. (Oxford University Press, Oxford, 2013).

[13] G. Voronoi, Journal Fur Die Reine Und Angewandte Mathematik **134** (1/4), 198 (1908).

[14] E. Wigner and F. Seitz, Physical Review **43** (10), 0804 (1933).

CHAPTER 4

SUMMARY

Exploration of EPFT

We explored the general applicability of the Entropy Pair Functional Theory [EPFT]
approach by applying it to three many-body interaction potentials. These potentials are
state of the art for large scale models for the three materials in this study: Fe modelled
with a modified embedded atom method [MEAM] potential, Cu modelled with an
MEAM and Si modelled with a Tersoff potential. We demonstrated the robust nature of
EPFT in determining excess entropy for diverse systems with many-body interactions.
Finally, these steps toward a universal Entropy Pair Functional, EPF, can be applied with
confidence to determine the entropy associated with sophisticated optimized potentials
and first principles simulations of liquids, crystals, engineered structures, and defects.

High temperature correction

By exploring the origin of the high temperature error in the Kirkwood entropy we
demonstrated a means to correct the error. Specifically, by connecting the combinatorial
and physical configuration perspectives of the error a first class functional was
developed, providing a means to extract the higher order correlation information from the
pair correlation function [PCF].

Broader impact

Many of the problems facing the human race in the coming years depend upon materials
that do not yet exist or those that must be obtains from the reuse of the constituents of the
built environment. To expedite the development of these materials we must rely on the
computational tools that have emerged over the past half century. Free energy and
consequently entropy calculations are essential in any material discovery or engineering

processes. A direct, efficient means to calculate entropy will play an import role in this work.

Entropy Codes

The complete set of codes utilized in this work are provided by way of a GitHub repository located at https://github.com/cliftonsluss/SFunk. This code base includes C++ source code for the calculation of absolute liquid and solid entropy based on the EPFT methods used and described in this work. The program to calculate liquid entropy takes as input the radial distribution function for the modeled system, the system number density, and the four parameters p_1, p_2, c_1, and c_2 (default values for the four parameters are the four 'universal' parameters we have presented). The solid entropy code takes as input a trajectory file containing a set of frames of atomic positions and returns the three upper bounds to solid entropy we presented here. Some additional Python scripts are provided to the community to assist those who wish to develop a target entropy following the thermodynamic integration method we have used.

Future Work

Next steps with the EPFT should include the extension to additional interaction potentials and systems of higher complexity. The ubiquitous Lennard-Jones potential is an obvious next choice for EPFT extension. Logical examples of higher complexity systems, include the calculation of the entropy of defect formation (vacancy or interstitial) as well as multicomponent systems. As motivation we briefly present the current state of results for the Lennard-Jones argon system. We have developed the target entropy for a liquid argon system modeled with the Lennard-Jones potential. Utilizing the four parameters developed from the Cu, Fe, and Si systems we compare the EPFT and Kirkwood results for Lennard-Jones argon to target entropy. Figure 4-1 shows good qualitative agreement between the EPFT and target entropy. Of note is the error for Kirkwood entropy (Equation (2.5)) is larger in magnitude than the typical $\frac{1}{2} k_B$ at the PG limit. A summary of the absolute error is reported in Table 4-1. Error from the Cu, Fe, and Si systems is

repeated here for comparison. Lennard-Jones argon demonstrates marginally larger error but still provides a large improvement over the base Kirkwood functional. Future work with the Lennard-Jones systems should be used to further refine the universal parameters.
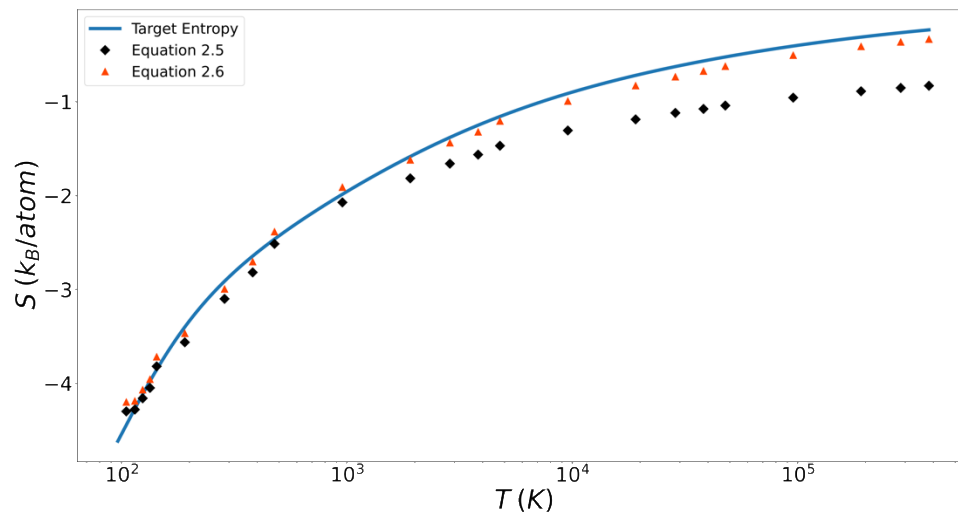
Figure 4-1. Liquid Lennard-Jones argon entropy.

System	q_1	q_2	c_1	c_2	Kirk. Avg. Err (k_B/Atom)	EPFT Avg. Err (k_B/Atom)
Cu					0.32	0.06
Fe					0.41	0.08
Si	3.24834	−2.406550	−320.1305	1.02966	0.61	0.05
Ar					0.30	0.09

Table 4-1. EPFT error for liquid systems utilizing universal parameters.

VITA

An East Tennessee native, Clifton Sluss explored several majors and worked for a time as a brick mason before finally receiving a Bachelor of Science degree in Mechanical Engineering from the University of Tennessee. After graduation he worked as an engineer at the Y-12 national security complex until changing disciplines and obtaining a professional engineering license in electrical power. Clifton spent 6 years designing protection and control systems for the electrical power grid before returning to Y-12 and ultimately to the University of Tennessee to pursue an advanced degree in material science. Clifton spends his free time working on and riding mountain bikes.