Springer Theses
Recognizing Outstanding Ph.D. Research

Robert John Nicholas Baldock

Classical Statistical Mechanics with Nested Sampling

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Robert John Nicholas Baldock

Classical Statistical Mechanics with Nested Sampling

Doctoral Thesis accepted by
the University of Cambridge, UK

Springer

Author
Dr. Robert John Nicholas Baldock
Cavendish Laboratory, Department of Physics
University of Cambridge
Cambridge
UK

Supervisor
Prof. Mike Payne
Cavendish Laboratory, Department of Physics
University of Cambridge
Cambridge
UK

ISSN 2190-5053 ISSN 2190-5061 (electronic)
Springer Theses
ISBN 978-3-319-66768-3 ISBN 978-3-319-66769-0 (eBook)
https://doi.org/10.1007/978-3-319-66769-0

Library of Congress Control Number: 2017950271

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Supervisor's Foreword

Wikipedia tells us "In physics, a partition function describes the statistical properties of a system in thermodynamic equilibrium. Partition functions are functions of the thermodynamic state variables, such as the temperature and volume. Most of the aggregate thermodynamic variables of the system, such as the total energy, free energy, entropy, and pressure, can be expressed in terms of the partition function or its derivatives" [1]. Knowledge of the partition function, thus, gives us access to complete finite temperature phase diagram of any material. However, no one has previously been able to compute the absolute partition function as an explicit function of its parameters for any atomistic system, using realistic models of the interatomic interactions and, indeed, it has generally been believed that such a computation is intractable. The methods developed in this thesis allow the partition function to be computed for realistic atomistic models of materials and thus, for the first time, allow the complete phase diagram of any material to be determined. Although the work presented in the thesis uses empirical interatomic potentials, the methods are efficient enough that it will be possible to use quantum mechanical models in a few years' time either directly or, more likely, indirectly through the use of machine learning approaches such as the Gaussian approximation potentials [2].

The starting point for the work presented in the thesis is nested sampling, a novel inference algorithm developed by Prof. John Skilling [3], [4]. The work presented in this thesis introduces two significant advances that together allow the partition function of realistic materials to be calculated. The first advance is the use of molecular dynamics to explore the configuration space in place of the more conventionally applied random walks. The use of molecular dynamics significantly speeds up the exploration but, unfortunately, it biases the sampling of the partition function. However, the required corrections to remove this bias have been determined and implemented allowing the potential performance advantage of molecular dynamics to be realised. The second advance is to develop a constant pressure method for sampling configurations that is far more efficient than sampling at constant volume. Of course, the partition function has to be sampled over all volumes and pressures and so the two approaches are ultimately equivalent but the

performance advantage is crucial. Together, these advances allow finite temperature phase diagrams to be computed for any atomistic model and a number of proof of concept applications are presented in this thesis. It should also be emphasised that the methodology runs without user intervention and thus can be implemented in "high-throughput" approaches.

Many of us are aware of the increasing demands for materials' modellers to have greater impact on real-world problems. This clearly requires many advances in our field, but one crucial requirement is the ability to predict phase diagrams of materials at finite temperature. Hence, the work described in this thesis represents a significant step forward in our ability to confront these real-world challenges.

Cambridge, UK
May 2016

Prof. Mike Payne

References

1. Wikipedia, Partition Function (statistical mechanics)—Wikipedia, The Free Encyclopedia, (2016), https://en.wikipedia.org/w/index.php?title=Partition_function_(statistical_mechanics)
2. A.P. Bartók, M.C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys. rev. lett. **104**(13), 136403 (2010)
3. J. Skilling, Nested sampling, in AIP *Conference Proceedings*, vol. 735, (2004), p. 395
4. J. Skilling, Nested sampling for general Bayesian computation. Bayesian Anal.**1**, 833 (2006)

Acknowledgements

First, I would like to thank Mike Payne for accepting me into TCM and his research group, for his direction and encouragement throughout my PhD studies, and for his work facilitating my research. I would also like to thank Gábor Csányi for the many long and exciting discussions that steered and invigorated me throughout the past 3 years. I am also deeply indebted to my colleagues Lvia Bartók-Pártay and Albert Bartók-Pártay, who both contributed directly to the work, and were generous with their time, often discussing technical matters at length.

In TCM and in the Csányi group, I had the privilege to be surrounded by many brilliant colleagues, too many to mention by name, with which I had many interesting discussions, and enjoyed many convivial moments. In particular, I am grateful to my dear friends Greg Lever and Pascal Bugnion (my fellow Cornish boys), and to Matthew Lyle.

Penultimately, I would like to thank the many friends I have shared time with during the past few years. You are too many to name and you know who you are. Finally, Mum, Dad, and my family: thank you.

Contents

Chapter 1
Introduction

1.1 The Goal: Classical Statistical Mechanics as a Black Box Tool

Classical statistical mechanics tells us the phase of a material for any particular values of the thermodynamic parameters.[1] We might find that, at these parameter values, the system is either a gas, a liquid, or a solid. If we find the system to be a solid, classical statistical mechanics will also tell us what kind of solid phase it forms: it could be a disordered solid, a crystal, or something between the two. In particular, both theorists and materials scientists are interested in locating phase transitions: the thermodynamic parameter values at which the material changes from one phase to another. There are therefore two primary tasks that "*users*" of classical statistical mechanics are interested in accomplishing:

1. Determining the stable phase at thermodynamic parameter values of interest.
2. Locating phase transitions between regions of stability.

Both of these tasks amount to pen-and-paper mathematical calculations, which are typically intractable for realistic models of atomic interactions. The promise of statistical mechanics' predictive power, and the infeasibility of performing the mathematics in closed form, have led scientists around the world to develop computer algorithms for accomplishing the two tasks listed above. At their core, such algorithms must combine efficient exploration of phase space with fast and accurate calculations of the interatomic potential energy function. This thesis is concerned with the former of these two challenges: phase space exploration. Phase space exploration began between the 1950 s and 1960 s with the Monte Carlo [1] and molecular dynamics [2–4] algorithms: two algorithms which remain cornerstones of

[1] Typical thermodynamic parameters for materials include temperature, pressure, and chemical potential. Alternatively, we can replace any of these parameters with volume (pressure), entropy (temperature), and the number of particles (chemical potential), to match the experimental situation.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_1

Table 1.1 A short illustrative list of algorithms used for identifying stable phases, or calculating phase transitions

Informed-search algorithms	Blind-search algorithms
Gibbs ensemble Monte Carlo	Wang-Landau sampling
Thermodynamic integration	Nested sampling
Self-consistent phonon methods	Parallel tempering
Umbrella sampling	

computational science to this day. Readers new to the field of computational statistical mechanics are referred to two textbooks on the subject: [5, 6].

The state of the art algorithms that exist today fall into two broad categories: "Informed-search" algorithms, and "blind-search" algorithms. The former can be specialised for comparing the stability of phases of a particular type (crystal structures, for example) or for locating a particular type of phase transition. Such algorithms typically require a great deal of prior knowledge about the system: a short list of possible solid structures, for example, or that the transition occurs across a known range of values for some order parameter. The second group of algorithms explore all possible phases of the system in an unbiased way, and require no specialist system knowledge. The advantage of the "informed-search" algorithms is that, by exploiting a large degree of prior knowledge, they are able to perform the tasks 1 and 2, in a tiny fraction of the computer time required by the "blind-search" algorithms. The disadvantage of the "informed-search" algorithms is that, to calculate an entire phase diagram, showing all phases and their phase transitions, we would need to enlist a zoo of different algorithms, each requiring specific expertise and different pieces of system knowledge. The advantages of the blind algorithms are that they can reveal phases and collective phenomena that we had not considered, and that they are capable of calculating entire phase diagrams in a single framework, generically applicable to any material. A short list of algorithms of both kinds is given in Table 1.1.

This thesis is concerned with the development of a promising "blind-search" algorithm called nested sampling [7, 8]. Nested sampling takes as input the interatomic potential, and gives as output the integrated density of states for a material. From the integrated density of states we can write down the partition function as an explicit function of temperature. Nested sampling also returns a series of atomic configurations, from which one may compute ensemble averages of observables and free energy landscapes. By yielding the partition function together with these atomic configurations, nested sampling enables us to perform statistical mechanics "from first principles". The aim of this thesis is to make nested sampling fast enough that it might practically be used to calculate entire phase diagrams, and indeed the complete statistical description of a material. Achieving this goal would turn classical statistical mechanics into a black box tool that almost anyone could use. This should have a profound effect on the way statistical mechanics is used, both in science and in industry.

1.2 Nested Sampling in Bayesian Statistics and in Statistical Mechanics

Take a quick glance at the current literature for nested sampling and one will find articles on cosmology, high energy physics, acoustics, and image reconstruction, as well as a few articles on classical statistical mechanics. Nested sampling is a method for calculating the Bayesian "*evidence*", which takes the same role in Bayesian statistics, as the partition function in statistical mechanics: it is the normalising factor for probability distributions. In classical statistical mechanics, nested sampling is used to calculate the partition function.

The evidence is the denominator in Bayes' theorem, Eq. (2.18) on Page 15. Bayes' theorem is

$$\text{prob}\,(\theta|D) = \frac{\text{prob}\,(D|\theta) \times \text{prob}\,(\theta)}{\text{prob}\,(D)}$$

where D can represent some data that we have obtained, and θ can represent parameters of some model for the data. For example, D could represent measurements of the temperature on London Bridge, at noon on the 3rd of May, every year for the last twenty years. We might be confident that this data should be described by a Gaussian distribution, but unsure of the appropriate parameters $\theta = (\mu, \sigma)$, where μ represents the mean, and σ the standard deviation. In this case, prob (θ) is the extent to which we believe each value of the parameters to begin with, and we call prob (θ) the "prior". The "likelihood", prob $(D|\theta)$ is the probability of the data for these particular parameter values: $\text{prob}\,(D|\theta) = \left(2\,\pi\,\sigma^2\right)^{-10} \prod_{i=1}^{20} \exp\left(-\frac{[D_i-\mu]^2}{2\sigma^2}\right)$. Finally, the "evidence", $\text{prob}\,(D) = \int \text{d}\theta \text{prob}\,(D|\theta)\,\text{prob}\,(\theta)$ is the normalising factor. Once we've calculated each of these quantities, we obtain a new, "posterior" probability distribution over the parameters, prob $(\theta|D)$. This represents the (updated) confidence we should have in the parameter values, given the data. In this way, Bayesian statistics allow us to compare models or theories to experimental data, using Bayes' theorem. In particular, calculating the different evidences for several models, for example this Gaussian model and also a Cauchy distribution, allows one to compare the models. In the broadest sense, this is the way in which nested sampling is used in other parts of the literature.

We can consider statistical mechanics in the language of Bayesian statistics if we associate the "data" D with inverse temperature β, and then make associations for the parameters θ, such as those shown in Table 1.2.

If we associate $\theta \to (\mathbf{q}, \mathbf{p})$, then Bayes' theorem reads

$$\text{prob}\,((\mathbf{q}, \mathbf{p})\,|\beta) = \frac{1}{N!h^{3N}} \frac{\text{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})}}{Z\,(\beta)} \tag{1.1}$$

Here we have identified the following:

Table 1.2 Association of Bayesian statistics and statistical mechanics. Here we show two possible associations for the parameters, θ. Other equivalent associations exist, such as $D \to \beta$, and $\theta \to E$ (energy)

	Bayesian statistics		**Classical statistical mechanics**
	Data	$D \to \beta$	Inverse temperature
1.	Parameters	$\theta \to (\mathbf{q}, \mathbf{p})$	Position in phase-space
2.	Parameters	$\theta \to \Phi(E)$	Number of states with energy E or less

$$\text{prob}(\beta|(\mathbf{q},\mathbf{p})) \to \mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})} \tag{1.2}$$

$$\text{prob}(\mathbf{q},\mathbf{p}) \to \frac{1}{N!h^{3N}} \tag{1.3}$$

$$\text{prob}(D) \to Z(\beta) = \frac{1}{N!h^{3N}} \iint \mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}\mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})}. \tag{1.4}$$

Equation (1.2) is just the Boltzmann distribution, Eq. (1.3) corresponds to the principle of equal a priory probability (see Sect. 4.1.2), and Eq. (1.4) identifies the Bayesian evidence, prob (D), with the partition function. In fact, nested sampling can be most clearly understood by making the second (equivalent) association shown in Table 1.2. This association also identifies the Bayesian evidence with the partition function. Since we shall return to the nested sampling algorithm at length, we consider it sufficient only to highlight that association here.

1.3 Outline of the Thesis

Part I is a summary of statistical and thermal physics.

Part II introduces the nested sampling algorithm. We compare the approach of nested sampling to that of parallel tempering. We then explore how the method works, before developing a Markov chain Monte Carlo implementation of the algorithm for materials simulated at constant pressure. We demonstrate that this method is orders of magnitude more efficient than parallel tempering. The method is then applied to simple Lennard-Jonesium, a binary Lennard-Jonesium "alloy", and in Appendix 8.10, to an embedded atom model potential for aluminium. We calculate the pressure-temperature phase diagrams for these systems, and in each, nested sampling reveals phenomena we had not anticipated. In Chap. 9 we show how the results of nested sampling can be post-processed to calculate (P, V, T) equations of state. This method is applied to Lennard-Jonesium as an example. Finally, we compare two methods of parallelising the algorithm.

Part III explores methods of performing nested sampling using Hamiltonian Monte Carlo. In particular, we develop a method of performing nested sampling in

the total Hamiltonian. We compare the performance of this new method to our Markov chain Monte Carlo algorithm, and find that the Hamiltonian Monte Carlo algorithm decorrelates approximately 64 times more efficiently with 64 atoms in the simulation cell.

1.4 A History of Nested Sampling

The nested sampling algorithm was originally proposed by John Skilling [7, 8], who did his Ph.D. in astrophysics at the Department of Physics at the University of Cambridge, and later worked on Bayesian data processing at the Department of Applied Mathematics and Theoretical Physics, also at the University of Cambridge. The first Ph.D. work on nested sampling was performed by Iain Murray [9] who was supervised by David MacKay. Their work was fundamental, and considered Markov chain Monte Carlo inference methods in the broadest sense. Since then, and around that time, other people have also worked on the nested sampling method for general Bayesian computation, applying it to other fields. Of particular significance is the work of Farhan Feroz, Mike Hobson and others from the Astrophysics Group in the Department of Physics, at the University of Cambridge. They have developed several nested sampling algorithms, and in particular "MultiNest" [10], which is extremely effective for performing nested sampling on multimodal distributions, in relatively low numbers of dimensions (fewer than fifty). MultiNest has already found widespread utility in astrophysics, cosmology and high-energy physics.

Also in Cambridge, Gábor Csányi, a long-term friend of John Skilling, had the idea to explore the use of nested sampling for classical statistical mechanics. The first Ph.D. work done on using nested sampling for classical statistical mechanics was performed by Lívia Bartók-Pártay, who collaborated with Albert Pártay-Bartók, under the supervision of Gábor Csányi. Their paper on Lennard-Jones clusters [11] created something of a stir in the classical statistical mechanics community, and subsequently several papers inspired by their work have been published. In particular, work has been done on Lennard-Jones clusters [12, 13] and proteins [14]. Lívia, Albert and Gábor also applied nested sampling to the hard-sphere model [15] and in 2011, Brendon Brewer, Lívia and Gábor developed the general "diffusive nested sampling" algorithm [16] (sometimes also referred to as the "energy partitioning method"). The diffusive nested sampling algorithm has also found use in condensed matter, where Hainam Do and Richard Wheatley have [17] applied it to calculate the free energies of solid phases using a cage model [17] and separately to simulate evaporation [18].

References

1. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087 (1953)
2. E. Fermi, J. R. Pasta, S.M. Ulam, Studies of nonlinear problems. LASL Report LA-1940 (1955)
3. B.J. Alder, T. Wainwright, Studies in molecular dynamics. I. General method. J. Chem. Phys. **31**, 459 (1959)
4. A. Rahman, Correlations in the motion of atoms in liquid argon. Phys. Rev. **136**, A405 (1964)
5. D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Computational Science Series (Elsevier Science, 2001)
6. M. Tuckerman, Statistical Mechanics and Molecular Simulations (Oxford University Press, 2008)
7. J. Skilling, Nested sampling. AIP Conf. Proc. **735**, 395 (2004)
8. J. Skilling, Nested sampling for general Bayesian computation. Bayesian Anal. **1**, 833 (2006)
9. I. Murray, Advances in Markov chain Monte Carlo methods. Ph.D. thesis, Gatsby Computational Neuroscience Unit, University College London, 2007
10. F. Feroz, M. Hobson, M. Bridges, MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. Mon. Not. R. Astron. Soc. **398**, 1601 (2009)
11. L.B. Pártay, A.P. Bartók, G. Csányi, Efficient sampling of atomic congurational spaces. J. Phys. Chem. B **114**, 10502 (2010)
12. S. Martiniani, J.D. Stevenson, D.J. Wales, D. Frenkel, Superposition enhanced nested sampling. Phys. Rev. X **4**, 031034 (2014)
13. S.O. Nielsen, Nested sampling in the canonical ensemble: direct calculation of the partition function from NVT trajectories. J. Chem. Phys. **139**, 124104 (2013)
14. N.S. Burkoff, C. Várnai, S.A. Wells, D.L. Wild, Exploring the energy landscapes of protein folding simulations with Bayesian computation. Biophys. J. **102**, 878 (2012)
15. L.B. Pártay, A.P. Bartók, G. Csányi, Nested sampling for materials: The case of hard spheres. Phys. Rev. E **89**, 022302 (2014)
16. B.J. Brewer, L.B. Pártay, G. Csányi, Diffusive nested sampling. Stat. Comput. **21**, 649 (2011)
17. H. Do, R.J. Wheatley, Density of states partitioning method for calculating the free energy of solids. J. Chem. Theory Comput. **9**, 165 (2012)
18. H. Do, J.D. Hirst, R.J. Wheatley, Rapid calculation of partition functions and free energies of fluids. J. Chem. Phys. **135**, 174105 (2011)

Chapter 2
A Primer in Probability

This section provides a summary of probability theory, as necessary to understand this thesis. We consider both discrete and continuous random variables.

▷ A discrete random variable S takes on one of a set of possible values $\mathcal{A}_S = \{a_1, a_2, \ldots, a_I\}$ with probabilities $\mathcal{P}_S = \{p_1, p_2, \ldots, p_I\}$ such that $\text{prob}\,(S = a_i) = p_i$, $p_i \geq 0$ and $\sum_{a_i \in \mathcal{A}_S} \text{prob}\,(S = a_i) = 1$. The probability that S is found in W, a subset of $\mathcal{A}_S$ is

$$\text{prob}\,(S \in W) = \sum_{a_i \in W} \text{prob}\,(S = a_i) \tag{2.1}$$

For a continuous random variable X, we only assign probabilities to ranges of values for X. The probability that $a \leq X \leq b$ is

$$\text{prob}\,(a \leq X \leq b) = \int_a^b \varrho\,(x)\,\mathrm{d}x. \tag{2.2}$$

Here $\varrho\,(x)$ is termed the probability density function (pdf). The pdf is a non-negative, integrable function of x. The differential relation (2.3) is also true provided that the pdf is continuous at x.

$$\frac{\mathrm{d}}{\mathrm{d}x}\text{prob}\,(X \leq x) = \varrho\,(x) \tag{2.3}$$

As in the discrete case, the total probability is one.

$$\int_{-\infty}^{\infty} \varrho\,(x)\,\mathrm{d}x = 1 \tag{2.4}$$

▷ We have the pdf for a random variable X. We desire the pdf for another variable Y which is a single valued function of X: $Y = f\,(X)$. This transformation is called a "change of variables".

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_2

$$\varrho\,(Y = y) = \varrho\,(X = x) \times \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \tag{2.5}$$

$$= \varrho\,(X = x) \div \left|\frac{\mathrm{d}f}{\mathrm{d}x}\right| \tag{2.6}$$

▷ Consider ordered pairs $(S = a_i, T = b_i)$: realisations of the discrete random variables $S \in \mathcal{A}_S = \{a_1, a_2, \ldots, a_I\}$ and $T \in \mathcal{A}_T = \{b_1, b_2, \ldots, b_I\}$. We call the probability distribution over these pairs the joint probability of S and T, $\mathrm{prob}\,(S, T)$.
In a similar manner for two continuous random variables X and Y we define the joint probability distribution via a "multivariate" pdf

$$\mathrm{prob}\,(a \leq X \leq b, c \leq Y \leq d) = \int_{x=a}^{b} \int_{y=c}^{d} \varrho\,(X = x, Y = y)\,\mathrm{d}x\mathrm{d}y. \tag{2.7}$$

▷ Two random variables (S and T) or (X and Y) are independent if and only if

$$\mathrm{prob}\,(S, T) = \mathrm{prob}\,(S)\,\mathrm{prob}\,(T) \tag{2.8}$$

$$\varrho\,(X = x, Y = y) = \varrho\,(X = x)\,\varrho\,(Y = y) \tag{2.9}$$

▷ We can recover the probability distribution for S alone $\mathrm{prob}\,(S = a_i) = p_i$ from the joint distribution $\mathrm{prob}\,(S, T)$ by summing over all values of T: a process called marginalisation.

$$\mathrm{prob}\,(S = a_i) = \sum_{b_j \in \mathcal{A}_T} \mathrm{prob}\left(S = a_i, T = b_j\right) \tag{2.10}$$

Likewise, in the continuous case, we can recover the pdf for either variable by integrating the multivariate pdf.

$$\varrho\,(X = x) = \int_{-\infty}^{\infty} \varrho\,(X = x, Y = y)\,\mathrm{d}y \tag{2.11}$$

▷ The conditional probability $\mathrm{prob}\left(S = a_i | T = b_j\right)$ represents "the probability that $S = a_i$ *given* $T = b_j$". It is given by

$$\mathrm{prob}\left(S = a_i | T = b_j\right) \equiv \frac{\mathrm{prob}\left(S = a_i, T = b_j\right)}{\mathrm{prob}\left(T = b_j\right)} \text{ if } \mathrm{prob}\left(T = b_j\right) \neq 0. \tag{2.12}$$

(If $\mathrm{prob}\left(T = b_j\right) = 0$ then $\mathrm{prob}\left(S = a_i | T = b_j\right)$ is undefined.)
If U is the range $[a, b]$ and V the range $[c, d]$

$$\mathrm{prob}\,(X \in U | Y \in V) \equiv \frac{\mathrm{prob}\,(X \in U, Y \in V)}{\mathrm{prob}\,(Y \in V)} \text{ if } \mathrm{prob}\,(Y \in V) \neq 0. \tag{2.13}$$

(Again, if $\text{prob}\,(Y \in V) = 0$ then $\text{prob}\,(X \in U | Y \in V)$ is undefined.)

▷ From the definition of conditional probability we have the "product rule" (also called the "chain rule").

$$\begin{aligned}\text{prob}\left(S = a_i, T = b_j\right) &= \text{prob}\left(S = a_i | T = b_j\right) \times \text{prob}\left(T = b_j\right) \\ &= \text{prob}\left(T = b_j | S = a_i\right) \times \text{prob}\left(S = a_i\right)\end{aligned} \tag{2.14}$$

$$\begin{aligned}\varrho\,(X = x, Y = y) &= \varrho\,(X = x | Y = y) \times \varrho\,(Y = y) \\ &= \varrho\,(Y = y | X = x) \times \varrho\,(X = x)\end{aligned} \tag{2.15}$$

▷ The sum rule is obtained by rewriting marginalisation using the product rule.

$$\begin{aligned}\text{prob}\,(S = a_i) &= \sum_{b_j \in \mathcal{A}_T} \text{prob}\left(S = a_i, T = b_j\right) \\ &= \sum_{b_j \in \mathcal{A}_T} \text{prob}\left(S = a_i | T = b_j\right) \times \text{prob}\left(T = b_j\right)\end{aligned} \tag{2.16}$$

$$\begin{aligned}\varrho\,(X = x) &= \textstyle\int_{-\infty}^{\infty} \varrho\,(X = x, Y = y)\,\mathrm{d}y \\ &= \textstyle\int_{-\infty}^{\infty} \varrho\,(X = x | Y = y) \times \varrho\,(Y = y)\,\mathrm{d}y\end{aligned} \tag{2.17}$$

▷ We obtain Bayes' theorem by rearranging the product rule.

$$\text{prob}\,(S|T) = \frac{\text{prob}\,(T|S) \times \text{prob}\,(S)}{\text{prob}\,(T)} \tag{2.18}$$

$$\text{prob}\,(S|T) = \frac{\text{prob}\,(T|S) \times \text{prob}\,(S)}{\sum_S \text{prob}\,(T|S) \times \text{prob}\,(S)} \tag{2.19}$$

$$\text{prob}\,(X \in U | Y \in V) = \frac{\text{prob}\,(Y \in V | X \in U) \times \text{prob}\,(X \in U)}{\text{prob}\,(Y \in V)} \tag{2.20}$$

$$\text{prob}\,(X \in U | Y \in V) = \frac{\text{prob}\,(Y \in V | X \in U) \times \text{prob}\,(X \in U)}{\int_{y \in V} \int_{x=-\infty}^{\infty} \varrho\,(Y = y | X = x) \times \varrho\,(X = x)\,\mathrm{d}x\mathrm{d}y} \tag{2.21}$$

Part I
Statistical and Thermal Physics

Chapter 3
Introduction

3.1 Phase Space

Consider N particles with coordinates $\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N\}$ and momenta $\mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N\}$. The space spanned by the points $(\mathbf{q}, \mathbf{p})$ is called the phase space of the system. It has $6N$ dimensions, and each point $(\mathbf{q}, \mathbf{p})$ completely describes the state of those N particles. For this reason $(\mathbf{q}, \mathbf{p})$ is called a representative point. In addition to phase space, we also define configuration space as the $3N$ dimensional space spanned by the points $\mathbf{q}$.

3.2 Classical and Statistical Mechanics

The theory of classical mechanics describes the time evolution of representative points, and their paths are called trajectories. Given initial coordinates $(\mathbf{q}_0, \mathbf{p}_0)$, classical mechanics tells us the coordinates at any arbitrary time later. By contrast, statistical mechanics presents the probability density function in phase space, given limited information about the system's external conditions. For example, we might know that the system is in contact with a much larger system, which holds the temperature and pressure of our system constant, or that it is in complete isolation and therefore has constant energy. Gibbs introduced the idea of a statistical ensemble: infinitely many imagined copies of our system distributed through phase space according to the probability density function. Each of these copies evolves according to classical mechanics and therefore the time invariant probability density function of statistical mechanics is just a stationary distribution of that same dynamics. This creates an important link between the theories of classical and statistical mechanics.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_3

3.3 Hamilton's Equations of Motion

In classical mechanics, representative points move through phase space according to Hamilton's equations of motion. The central object in Hamilton's equations of motion is the Hamiltonian $\mathcal{H}(\mathbf{q}, \mathbf{p})$. The Hamiltonian is the sum of the potential and kinetic energies, $U(\mathbf{q})$ and $K(\mathbf{p})$

$$\mathcal{H}(\mathbf{q}, \mathbf{p}) = K(\mathbf{p}) + U(\mathbf{q}). \tag{3.1}$$

Both the potential and kinetic energies are bounded from below. The kinetic energy has the form

$$K(\mathbf{p}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i} \tag{3.2}$$

and is thus an even function

$$K(\mathbf{p}) = K(-\mathbf{p}). \tag{3.3}$$

The same may be said for the potential energy of closed systems.

Hamilton's equations of motion are

$$\frac{\partial \mathbf{q}}{\partial t} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}}, \quad \frac{\partial \mathbf{p}}{\partial t} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} \tag{3.4}$$

We return to Hamilton's equations of motion in Part III, where their properties are discussed in some detail.

3.4 Statistical Equilibrium

At an initial time t_0 we know the coordinates of a system to be $(\mathbf{q}_0, \mathbf{p}_0)$ with some error $(\Delta\mathbf{q}, \Delta\mathbf{p})$. A short time later Δt, we still have an idea of the system's coordinates, but with a larger uncertainty. A longer time later, after what's called the equilibration time, the probability density function (pdf) for the coordinates of our system is independent of the starting conditions. Once the pdf has relaxed to this state the system is said to be in statistical equilibrium. The subject of statistical mechanics is this pdf, which is by definition independent of the initial conditions, and depends instead on the external conditions of the system. For example, our system may be in complete isolation from all other systems, or alternatively it might be connected to a much larger system with which it can exchange energy.

3.5 Correlation Lengths and Statistically Independent Subsystems

Imagine a large volume of fluid. We divide the fluid in half, keeping the intensive variables, temperature, pressure and chemical potential, the same as before. In this case each half will have the same macroscopic properties as the original fluid. We repeat this operation a number of times. The fluid is made up of atoms and molecules so after many iterations something different must happen. The characteristic length scale at which the overall properties of the pieces begin to differ substantially from the original fluid is called the correlation length, ξ. This is the typical length scale over which movements of the atoms and molecules in the fluid are correlated.

Atoms separated by distances greater than the correlation length are effectively independent and we may therefore consider our original fluid to have been made up of many statistically independent subsystems of length scale greater than ξ. Statistical independence for two subsystems means that the pdf for the combined system $\varrho_{1,2}$ may be written as a product of the pdfs for the two subsystems ϱ_1 and ϱ_2.

$$\varrho_{1,2} = \varrho_1 \varrho_2 \tag{3.5}$$

3.6 Fluctuations

Consider some observable f that depends on the system coordinates $(\mathbf{q}, \mathbf{p})$. The mean value of the observable is given by

$$\overline{f} = \int f(\mathbf{q}, \mathbf{p})\, \varrho(\mathbf{q}, \mathbf{p})\, \mathrm{d}^N\mathbf{q}\, \mathrm{d}^N\mathbf{p}. \tag{3.6}$$

If we measure f we may not observe $\overline{f}$ exactly, but some value close to it. Let us represent the difference between the two as Δf. The difference Δf is traditionally called the "fluctuation" in this context.

$$\Delta f = f - \overline{f} \tag{3.7}$$

By definition, the mean value of the fluctuation is zero: $\overline{\Delta f} = 0$. The mean square fluctuation (the variance), and the root mean square fluctuation (the standard deviation) are given by

$$\mathrm{Var}(f) = \langle (\Delta f)^2 \rangle = \overline{f^2} - \left(\overline{f}\right)^2 \tag{3.8}$$

$$s = \langle (\Delta f)^2 \rangle^{\frac{1}{2}} = \left[\overline{f^2} - \left(\overline{f}\right)^2\right]^{\frac{1}{2}} \tag{3.9}$$

The ratio $\frac{s}{\overline{f}} = \frac{\langle(\Delta f)^2\rangle^{\frac{1}{2}}}{\overline{f}}$ is a measure of our uncertainty in f and is called "the relative fluctuation in f".

We shall now show that the relative fluctuations in extensive physical quantities decrease rapidly as the size of a system (the number of particles that make it up) increases. For extensive quantities we have $f = \sum f_i$. Therefore, if our total system is made up of N statistically independent subsystems, the average value $\overline{f}$ for the total system may generally be written as

$$\overline{f} = \sum_{i=1}^{N} \overline{f_i}. \tag{3.10}$$

The size of $\overline{f}$ is therefore approximately proportional to N.

Now consider the variance of f. In general

$$\langle(\Delta f)^2\rangle = \left\langle \left(\sum_i \Delta f_i \right)^2 \right\rangle. \tag{3.11}$$

Since the subsystems are statistically independent, mean values of products such as $\Delta f_i \Delta f_j$ vanish[1]

$$\overline{\Delta f_i \Delta f_j} = \overline{\Delta f_i}\ \overline{\Delta f_j} = 0 \ : \ i \neq j. \tag{3.12}$$

This simplifies the total variance (3.11) to a sum of N independent terms

$$\langle(\Delta f)^2\rangle = \sum_{i=1}^{N} \langle(\Delta f_i)^2\rangle. \tag{3.13}$$

The variance $\langle(\Delta f)^2\rangle$ is therefore also approximately proportional to N.

Hence we can say of the relative fluctuation that

$$\frac{\langle(\Delta f)^2\rangle^{\frac{1}{2}}}{\overline{f}} \propto \frac{1}{\sqrt{N}}. \tag{3.14}$$

[1] Here is a brief derivation of Eq. (3.12).

$$\begin{aligned} \Delta f_i \Delta f_j &= \left(f_i - \overline{f_i}\right)\left(f_j - \overline{f_j}\right) \\ &= f_i f_j - f_i \overline{f_j} - f_j \overline{f_i} + \overline{f_i}\ \overline{f_j} \\ \Rightarrow \overline{\Delta f_i \Delta f_j} &= \overline{f_i f_j} - \overline{f_i}\ \overline{f_j} \end{aligned}$$

Since f_i and f_j are independent

$$\overline{f_i f_j} = \overline{f_i}\ \overline{f_j}.$$

Therefore we have

$$\overline{\Delta f_i \Delta f_j} = 0 \ : \ i \neq j.$$

This is true for any extensive physical property. Since the number of independent subsystems is itself proportional to the number of atoms in the system, we could equally well replace N in (3.14) with the number of particles in our system.

In the limit of very large N, the relative fluctuation becomes negligible and the mean value is asymptotically equal to the value of maximum probability. This limit of large N is called the thermodynamic limit.

Statistical mechanics was developed to describe very large systems and traditionally makes predictions about the most probable value of a physical observable. If the ideas of statistical mechanics are to be applied to smaller systems, where probability distributions are broader and possibly even multimodal, then a single value may not characterise the pdf sufficiently well. In this case it is important to know the complete pdf for an observable.

Chapter 4
Phase Space Probability Distributions for Various External Conditions

4.1 Isolated Systems: Fixed N, E and V

This section describes the pdf in phase space $\varrho(\mathbf{q}, \mathbf{p})$ for a system with fixed total energy E and a fixed number of particles N which are constrained inside a volume V. The statistical ensemble corresponding to these conditions is named the micro-canonical ensemble.

Beyond restricting $\mathbf{q}$ to a region $D(V)$ and $(\mathbf{q}, \mathbf{p})$ to a surface of constant E, the information we have tells us nothing about $(\mathbf{q}, \mathbf{p})$. Therefore the only pdf that we can reasonably assign is a uniform distribution on that surface. This is a subjective probability argument, traditionally approached from the "principle of equal equilibrium probability" which is discussed at the end of this Sect. 4.1.

$$\varrho(\mathbf{q}, \mathbf{p}|E) = \frac{1}{N!h^{3N}} \frac{\delta(E - \mathcal{H}(\mathbf{q}, \mathbf{p}))}{\Omega(N, E, V)} \tag{4.1}$$

$$\Omega(N, E, V) = \frac{1}{N!h^{3N}} \int \mathrm{d}^N\mathbf{p} \int_{D(V)} \mathrm{d}^N\mathbf{q}\, \delta(E - \mathcal{H}(\mathbf{q}, \mathbf{p})) \tag{4.2}$$

The normalising constant $\Omega(N, E, V)$ is the number of microstates with energy E, also called the density of states. The factor $\frac{1}{N!h^{3N}}$ can be separated into two factors. The first, $\frac{1}{N!}$ accounts for the overcounting of configurations when the particles are indistinguishable. Although our notation distinguishes between particles i and j, nature may not. For the second, we consider that our system is a quantum system in the quasi-classical limit. In this case h^{3N} is the "volume" of phase space corresponding to a single quantum state for N particles. Thus $\frac{1}{h^{3N}}$ relates the continuous phase space volume element $\Delta\mathbf{q}\Delta\mathbf{p}$ to the corresponding number of quantum states $\frac{\Delta\mathbf{q}\Delta\mathbf{p}}{h^{3N}}$. This is how the density of states $\Omega(N, E, V)$ can be understood as the number of states with energy E.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_4

In addition to the density of states Ω, we also define the integrated density of states

$$\Phi(N, E, V) = \int_{-\infty}^{E} \Omega\left(N, E', V\right) \mathrm{d}E'. \tag{4.3}$$

This is the number of states with energy less than or equal to E, and is therefore a monotonic increasing function of E. I will often write $\Omega(E)$ and $\Phi(E)$ as a shorthand for $\Omega(N, E, V)$ and $\Phi(N, E, V)$. For real and simulated systems, which always have finite size, $\Omega(E)$ is a continuous function:

$$\Omega(E) = \frac{\mathrm{d}\Phi}{\mathrm{d}E}. \tag{4.4}$$

4.1.1 Distributions for U, K

We now introduce separate densities of states for the potential and kinetic energies $g(N, U, V)$ and $h(N, K)$.

$$g\left(N, U', V\right) = \frac{1}{N! h^{3N}} \int \mathrm{d}^N \mathbf{q}\, \delta\left(U\left[\mathbf{q}\right] - U'\right) \tag{4.5}$$

$$h\left(N, K'\right) = \int \mathrm{d}^N \mathbf{p}\, \delta\left(K\left[\mathbf{p}\right] - K'\right) \tag{4.6}$$

$$= \frac{\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} (2m)^{\frac{3N}{2}} K'^{\frac{3N}{2}-1} \tag{4.7}$$

The form of (4.7) applies to the standard quadratic form of the potential energy (3.2) assuming equal masses.[1] Shorthand notations for $g(N, U, V)$ and $h(N, K)$ are $g(U)$ and $h(K)$.

[1] Derivation:

$$|\mathbf{p}| \equiv p \tag{4.8}$$

$$\mathrm{prob}(p) = \frac{2\,\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} p^{3N-1} \tag{4.9}$$

$$\mathrm{prob}(K) = \mathrm{prob}(p) \times \left|\frac{\mathrm{d}p}{\mathrm{d}K}\right| \tag{4.10}$$

$$K = \frac{1}{2m} \sum_{i=1}^{3N} p_i^2 = \frac{p^2}{2m} \tag{4.11}$$

$$\frac{\mathrm{d}K}{\mathrm{d}p} = \frac{p}{m} \tag{4.12}$$

$$\mathrm{prob}(K) = \frac{\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} 2m p^{3N-2} \tag{4.13}$$

The multivariate pdf for the potential and kinetic energies is

$$\varrho\left(U, K|E\right) = \frac{1}{\Omega\left(E\right)} g\left(U\right) f\left(K\right) \delta\left(E - U - K\right) \tag{4.16}$$

Substituting (4.7) into the joint distribution (4.16) we obtain

$$\varrho\left(U, K|E\right) = \frac{\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} \left(2m\right)^{\frac{3N}{2}} \frac{1}{\Omega\left(E\right)} g\left(U\right) K^{\frac{3N}{2}-1} \delta\left(E - U - K\right). \tag{4.17}$$

The pdf for U is obtained by marginalisation of (4.17)

$$\begin{aligned}\varrho\left(U|E\right) &= \frac{\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} \left(2m\right)^{\frac{3N}{2}} \frac{1}{\Omega\left(E\right)} g\left(U\right) \int_0^\infty K^{\frac{3N}{2}-1} \delta\left(E - U - K\right) \mathrm{d}K \\ &= \frac{\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} \left(2m\right)^{\frac{3N}{2}} \frac{1}{\Omega\left(E\right)} g\left(U\right) \left(E - U\right)^{\frac{3N}{2}-1}\end{aligned} \tag{4.18}$$

or more simply

$$\varrho\left(U|E\right) \propto g\left(U\right) \left(E - U\right)^{\frac{3N}{2}-1}. \tag{4.19}$$

In addition to the density of states for U (4.5) we introduce the integrated density of states for U

$$\chi\left(N, U, V\right) = \int_{-\infty}^{U} g\left(N, U', V\right) \mathrm{d}U'. \tag{4.20}$$

In the same way as $\Phi\left(E\right)$ for the total energy, $\chi\left(N, U', V\right)$ represents the number of states with $U\left(\mathbf{q}\right) \leq U'$, and is a monotonic increasing function of U'. As for $\Phi\left(E\right)$, $\chi\left(U\right)$ is used as shorthand for $\chi\left(N, U, V\right)$. Furthermore, just as in (4.4),

$$g\left(U\right) = \frac{\mathrm{d}\chi}{\mathrm{d}U}. \tag{4.21}$$

Performing a change of variables from U to $\chi\left(U\right)$ we obtain

$$\varrho\left(\chi\left(U\right)|E\right) \propto g\left(U\right) \left(E - U\left(\chi\right)\right)^{\frac{3N}{2}-1} \div g\left(U\right) \tag{4.22}$$

$$\propto \left(E - U\left(\chi\right)\right)^{\frac{3N}{2}-1}. \tag{4.23}$$

(Footnote 1 continued)

$$= \frac{\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} 2m \left(2mK\right)^{\frac{3N}{2}-1} \tag{4.14}$$

$$= \frac{\pi^{\frac{3N}{2}}}{\Gamma\left(\frac{3N}{2}\right)} \left(2m\right)^{\frac{3N}{2}} K^{\frac{3N}{2}-1} \tag{4.15}$$

.

4.1.2 Principle of Equal Equilibrium Probability

In addition to energy conservation, a system may have conserved values for momentum (**P**) and angular momentum (**J**). The principle of equal equilibrium probability states that a very large number of repeat measurements of (**q**, **p**) will result in a uniform density of observations across the surface $E = E_0, \mathbf{P} = \mathbf{P}_0, \mathbf{J} = \mathbf{J}_0$. In practice, this locus may consist of two or more disconnected surfaces, with no mechanism for an isolated system to move between them. Therefore the principle of equal equilibrium probability does not hold in all circumstances. Nevertheless it is used as an axiom of traditional statistical mechanics.

4.2 Systems That Exchange Energy with the Universe: Fixed N, T, and V

A number of particles N are restricted to a volume V, as in an isolated system. Unlike the system in the microcanonical ensemble, these particles are in contact with a much larger system, such that energy may pass between the two. The term often applied for the larger system here is "a heat bath" or "the universe". We call the statistical ensemble corresponding to these conditions the "canonical ensemble".

Together the universe and our system form a combined system. This combined system has fixed total energy E and is described by the microcanonical ensemble. The small system has an energy ϵ and the universe has energy $E - \epsilon$. The density of states for the combined system is

$$\Omega_{1,2}(E, \epsilon) = \Omega_1(\epsilon)\,\Omega_2(E - \epsilon)\,. \tag{4.24}$$

Here system 1 refers to the small system and system 2 refers to the universe. This density of states $\Omega_{1,2}$ is a very large number, so we consider the logarithm

$$\log \Omega_{1,2}(E, \epsilon) = \log \Omega_1(\epsilon) + \log \Omega_2(E - \epsilon)\,. \tag{4.25}$$

As the combined system is in the microcanonical ensemble, we assume that all states with total energy E are equally probable. This most probable division of energy $(\epsilon, E - \epsilon)$ is the one that corresponds to the largest number of microstates in the combined system. We find this by setting $\frac{\mathrm{d}\log\Omega_{1,2}}{\mathrm{d}\epsilon} = 0$.

$$\frac{\mathrm{d}\log\Omega_{1,2}}{\mathrm{d}\epsilon} = \frac{\mathrm{d}\log\Omega_1(\epsilon)}{\mathrm{d}\epsilon} + \frac{\mathrm{d}\log\Omega_2(E-\epsilon)}{\mathrm{d}\epsilon} \tag{4.26}$$

$$= \frac{\mathrm{d}\log\Omega_1}{\mathrm{d}\epsilon} - \frac{\mathrm{d}\log\Omega_2}{\mathrm{d}\epsilon} \tag{4.27}$$

$$\frac{\mathrm{d}\log\Omega_{1,2}}{\mathrm{d}\epsilon} = 0 \tag{4.28}$$

$$\Rightarrow \left.\frac{\mathrm{d}\log\Omega_1}{\mathrm{d}E'}\right|_{E'=\epsilon} = \left.\frac{\mathrm{d}\log\Omega_2}{\mathrm{d}E'}\right|_{E'=E-\epsilon} \tag{4.29}$$

Another way of writing this for more than two systems is

$$\begin{aligned} &\frac{\mathrm{d}\log\Omega_1}{\mathrm{d}E_1} = \frac{\mathrm{d}\log\Omega_2}{\mathrm{d}E_2} = \frac{\mathrm{d}\log\Omega_3}{\mathrm{d}E_3} = \ldots \\ &E = E_1 + E_2 + E_3 + \ldots \end{aligned} \tag{4.30}$$

where E_i and $\log\Omega_i$ are the energy and log density of states of system i.

The differential $\frac{\mathrm{d}\log\Omega}{\mathrm{d}E}$ is the rate at which the (log) number of states at energy E changes with respect to E. The most probable division of energy balances the gain in the number of states accessible to one system with the loss in the number of accessible states in the other. Any deviation from this point would reduce the number of accessible states overall. We will return to $\frac{\mathrm{d}\log\Omega}{\mathrm{d}E}$ shortly.

We have derived the most probable energy of our small system, but what is the pdf? To derive this we first consider that our system is much smaller than the universe. Therefore we could assume that the narrow neighbourhood of energies close to ϵ contains only one quantum energy state. Conversely, quantum states in the universe are much closer together in energy, and the region close to $E-\epsilon$ contains, effectively, an almost continuous spectrum of energy states. In this approximation the combined density of states (4.24) is

$$\Omega_{1,2}(E,\epsilon) = \Omega_1(\epsilon)\,\Omega_2(E-\epsilon) \tag{4.31}$$

$$\simeq 1 \times \Omega_2(E-\epsilon) \tag{4.32}$$

$$\simeq \mathrm{e}^{\log\Omega_2(E-\epsilon)} \tag{4.33}$$

Since our system is so small compared to the universe $\frac{\epsilon}{E} \simeq 0$. The log density of states $\log\Omega_2(E-\epsilon)$ therefore changes slowly on the scale of ϵ, and we may sensibly make a Taylor expansion about $\epsilon = 0$ to first order in ϵ.

$$\mathrm{e}^{\{\log\Omega_2(E-\epsilon)\}} \simeq \exp\left\{\log\Omega_2(E) - \epsilon\left.\frac{\mathrm{d}\log\Omega_2}{\mathrm{d}E'}\right|_{E'=E} + O\left(\epsilon^2\right) + \ldots\right\} \tag{4.34}$$

Making a further Taylor expansion of $\left.\frac{\mathrm{d}\log\Omega_2}{\mathrm{d}E'}\right|_{E'=E}$ around $E' = E-\epsilon$ we obtain

$$\Omega_{1,2}(E,\epsilon) \simeq \exp\left\{\log\Omega_2(E) - \epsilon\left.\frac{\mathrm{d}\log\Omega_2}{\mathrm{d}E'}\right|_{E'=E-\epsilon} + O\left(\epsilon^2\right) + \ldots\right\} \tag{4.35}$$

$$\Rightarrow \Omega_{1,2}(E,\epsilon) \propto \exp\left\{-\epsilon\left.\frac{\mathrm{d}\log\Omega_1}{\mathrm{d}E'}\right|_{E'=\epsilon}\right\} \tag{4.36}$$

In (4.36) we have substituted from (4.29). Equation (4.36) is written entirely in terms of quantities that relate only to our system. We have therefore arrived at the pdf for our system. Since $\frac{\mathrm{d}\log\Omega_i}{\mathrm{d}E_i}$ is both the quantity common to all equilibrium systems able to exchange energy and also appears in the pdf for our system we give it a special symbol, β or $\frac{1}{k_\mathrm{B}T}$

$$\frac{\mathrm{d}\log\Omega}{\mathrm{d}E} \equiv \beta \equiv \frac{1}{k_\mathrm{B}T}. \tag{4.37}$$

This is the statistical definition of temperature T. β is called the "inverse temperature" and k_B is "Boltzmann's constant" which converts between units of temperature and energy. In light of this definition, Eq. (4.30) indicates that when two or more systems are able to exchange energy, once equilibrium is achieved their temperatures are the same.

Writing the pdf for the canonical ensemble in full we have

$$\varrho\left(\mathbf{q},\mathbf{p}\right) = \frac{1}{N!h^{3N}}\frac{\mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})}}{Z\left(N,\beta,V\right)} \tag{4.38}$$

$$Z\left(N,\beta,V\right) = \frac{1}{N!h^{3N}}\int \mathrm{d}^N\mathbf{p}\int_{D(V)}\mathrm{d}^N\mathbf{q}\,\mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})}. \tag{4.39}$$

The function $Z\left(N,\beta,V\right)$ is called the "partition function". Just like $\Omega\left(N,E,V\right)$ in the microcanonical ensemble, $Z\left(N,\beta,V\right)$ is a measure of the number of accessible microstates. Assuming equal masses and substituting for the kinetic energy (3.2) we obtain

$$Z\left(N,\beta,V\right) = \frac{1}{N!h^{3N}}\int \mathrm{d}\mathbf{p}_1\ldots\mathrm{d}\mathbf{p}_N\,\exp\left(-\beta\sum_i\frac{\mathbf{p}_i^2}{2m}\right)\int_{D(V)}\mathrm{d}^N\mathbf{q}\,\exp\left[-\beta U\left(\mathbf{q}\right)\right] \tag{4.40}$$

$$Z\left(N,\beta,V\right) = \frac{1}{N!h^{3N}}\left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}}\int_{D(V)}\mathrm{d}^N\mathbf{q}\,\exp\left[-\beta U\left(\mathbf{q}\right)\right] \tag{4.41}$$

The marginal pdf for the coordinates $\mathbf{q}$ alone is obtained by integrating out the momentum coordinates in (4.38) and is given by

$$\varrho\left(\mathbf{q}\right) = \frac{1}{N!h^{3N}}\left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}}\frac{\mathrm{e}^{-\beta U(\mathbf{q})}}{Z\left(N,\beta,V\right)} \tag{4.42}$$

The pdf for U is given by

$$\varrho\left(U\right) = \frac{1}{N!h^{3N}}\left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}}\frac{g\left(U\right)\mathrm{e}^{-\beta U}}{Z\left(N,\beta,V\right)}. \tag{4.43}$$

The partition function can therefore be written as

$$Z(N, \beta, V) = \frac{1}{N! h^{3N}} \left(\frac{2\pi m}{\beta} \right)^{\frac{3N}{2}} \int_{-\infty}^{\infty} \mathrm{d}U g(U) \mathrm{e}^{-\beta U}. \tag{4.44}$$

As in the microcanonical ensemble we now obtain the pdf for χ by a change of variables. Starting from (4.43) and recalling (4.21) we have

$$\varrho(\chi) = \frac{1}{N! h^{3N}} \left(\frac{2\pi m}{\beta} \right)^{\frac{3N}{2}} \frac{\mathrm{e}^{-\beta U(\chi)}}{Z(N, \beta, V)} \tag{4.45}$$

$$Z(N, \beta, V) = \frac{1}{N! h^{3N}} \left(\frac{2\pi m}{\beta} \right)^{\frac{3N}{2}} \int_{0}^{\infty} \mathrm{d}\chi \mathrm{e}^{-\beta U(\chi)} \tag{4.46}$$

The simple form of the partition function given by (4.46) forms the basis of much of this thesis.

4.3 Systems That Exchange Both Energy and Volume with the Universe: Fixed N, T, and P

We will first establish that the most probable combined state of two systems that can exchange energy and volume is that their pressures and temperatures are equal. The arguments we will see here are a straightforward generalisation of those presented in Sect. 4.2.

Consider once again that our system is in contact with a much larger system, the universe. The two systems are able to exchange energy and volume with one another, but the combined system is in the microcanonical ensemble. Therefore the total energy E and volume V are both fixed. Our system (labelled 1) has energy and volume (ϵ, ν) while the universe (labelled 2) has $(E - \epsilon, V - \nu)$. The number of microstates of the combined system is $\Omega_{1,2}(E, \epsilon, V, \nu)$, and its logarithm is

$$\log \Omega_{1,2}(E, \epsilon, V, \nu) = \log \Omega_1(\epsilon, \nu) + \log \Omega_2(E - \epsilon, V - \nu). \tag{4.47}$$

The total derivative of (4.47) is

$$\mathrm{d} \log \Omega_{1,2} = \frac{\partial \log \Omega_{1,2}}{\partial \epsilon} \mathrm{d}\epsilon + \frac{\partial \log \Omega_{1,2}}{\partial \nu} \mathrm{d}\nu \tag{4.48}$$

For $\log \Omega_{1,2}$ to be a maximum both $\frac{\partial \log \Omega_{1,2}}{\partial \epsilon}$ and $\frac{\partial \log \Omega_{1,2}}{\partial \nu}$ must vanish separately. Thus we obtain

$$\left. \frac{\partial \log \Omega_1}{\partial E'} \right|_{E'=\epsilon} = \left. \frac{\partial \log \Omega_2}{\partial E'} \right|_{E'=E-\epsilon} \tag{4.49}$$

$$\left. \frac{\partial \log \Omega_1}{\partial V'} \right|_{V'=\nu} = \left. \frac{\partial \log \Omega_2}{\partial V'} \right|_{V'=V-\nu} \tag{4.50}$$

Equation (4.49) tells us that, as in the canonical ensemble, the most probable division of energies is when the temperatures of the two systems are equal: $T_1 = T_2$.

The second equation (4.50) identifies a new quantity that is equal in the system and the universe. We define this to be βP:

$$\frac{\partial \log \Omega}{\partial V} \equiv \beta P. \tag{4.51}$$

This is the statistical definition of pressure P. Two systems that can exchange energy and volume have the same values for temperature and pressure in equilibrium.

The derivation of the combined density of states follows the same lines of reasoning as in the canonical ensemble. Therefore we simply state the result here.

$$\Omega_{1,2}(E, \epsilon, V, \nu) \propto \exp\left\{-\epsilon \frac{\partial \log \Omega_1}{\partial E'}\bigg|_{E'=\epsilon} - \nu \frac{\partial \log \Omega_1}{\partial V'}\big|_{V'=\nu} + O\left(\epsilon^2\right) + O\left(\nu^2\right)\right\} \tag{4.52}$$

$$\simeq \exp\{-\beta(\epsilon + P\nu)\} \tag{4.53}$$

This leads to the pdf and partition function

$$\varrho(\mathbf{q}, \mathbf{p}, V) = \frac{1}{N! h^{3N} V_0} \frac{\mathrm{e}^{-\beta(\mathcal{H}(\mathbf{q},\mathbf{p})+PV)}}{\Delta(N, \beta, P)} \tag{4.54}$$

$$\begin{aligned} \Delta(N, \beta, P) &= \frac{1}{N! h^{3N} V_0} \int_0^\infty \mathrm{d}V \int_{D(V)} \mathrm{d}^N\mathbf{q}\, \mathrm{d}^N\mathbf{p}\, \mathrm{e}^{-\beta[\mathcal{H}(\mathbf{q},\mathbf{p})+PV]} \\ &= \frac{1}{V_0} \int_0^\infty \mathrm{d}V Z(N, \beta, V) \mathrm{e}^{-\beta PV}. \end{aligned} \tag{4.55}$$

The constant V_0 has dimensions of volume, and is included to correct the dimensions. It does not feature in the predictions of statistical mechanics. Performing the momentum integration for the case of equal masses and a quadratic kinetic energy (3.2) we obtain the marginal pdf for $(\mathbf{q}, V)$ and a new expression for the partition function

$$\varrho(\mathbf{q}, V) = \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \frac{\mathrm{e}^{-\beta(U(\mathbf{q})+PV)}}{\Delta(N, \beta, P)} \tag{4.56}$$

$$\Delta(N, \beta, P) = \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \int_0^\infty \mathrm{d}V \int_{D(V)} \mathrm{d}^N\mathbf{q}\, \mathrm{e}^{-\beta[U(\mathbf{q})+PV]} \tag{4.57}$$

4.3.1 Microscopic Enthalpy

We introduce the microscopic enthalpy, in order to define a density of states and integrated density of states for systems that can exchange energy and volume with the universe. The aim is then to write the pdfs (4.54), (4.56) and partition functions (4.55), (4.57) in terms of the density of states and integrated density of states.

We define the microscopic enthalpy (a microscopic potential) as

$$H(\mathbf{q}, \mathbf{p}, P, V) \equiv E(\mathbf{q}, \mathbf{p}) + PV. \tag{4.58}$$

It is now simple to rewrite the pdf (4.54) and the partition function (4.55) for a system at fixed N, P, T as

$$\varrho(\mathbf{q}, \mathbf{p}, V) = \frac{1}{N! h^{3N} V_0} \frac{\mathrm{e}^{-\beta H(\mathbf{q}, \mathbf{p}, P, V)}}{\Delta(N, \beta, P)} \tag{4.59}$$

$$\Delta(N, \beta, P) = \frac{1}{N! h^{3N} V_0} \int_0^\infty \mathrm{d}V \int_{D(V)} \mathrm{d}^N \mathbf{q}\, \mathrm{d}^N \mathbf{p}\, \mathrm{e}^{-\beta H(\mathbf{q}, \mathbf{p}, P, V)} \tag{4.60}$$

Next we introduce the density of states for enthalpy H' at fixed N, P

$$\Omega\left(N, H', P\right) = \frac{1}{N! h^{3N} V_0} \int_0^\infty \mathrm{d}V \int \mathrm{d}^N \mathbf{p} \int_{D(V)} \mathrm{d}^N \mathbf{q}\, \delta\left(H' - H\left[\mathbf{q}, \mathbf{p}, P, V\right]\right). \tag{4.61}$$

Again, we define the integrated density of states, $\Phi(N, H, P)$

$$\Phi(N, H, P) = \int_{-\infty}^{H} \Omega\left(N, H', P\right) \mathrm{d}H'. \tag{4.62}$$

We also define the configurational enthalpy

$$Y(\mathbf{q}, V, P) = U(\mathbf{q}) + PV. \tag{4.63}$$

From Y we define the configurational density of states $g(N, Y, P)$ and configurational integrated density of states $\chi(N, Y, P)$ thus:

$$g(N, Y, P) = \frac{1}{N! h^{3N} V_0} \int_0^\infty \mathrm{d}V \int_{D(V)} \mathrm{d}^N \mathbf{q}\, \delta(Y - U(\mathbf{q}) - PV) \tag{4.64}$$

$$\chi(N, Y, P) = \int_{-\infty}^{Y} g\left(N, Y', P\right) \mathrm{d}Y' \tag{4.65}$$

For convenience, we will often write $\Omega(H)$, $\Phi(H)$ and $g(Y)$, $\chi(Y)$, without explicit mention of N and P.

As in the microcanonical ensemble, we have the differential relations (4.66) and (4.67)

$$\Omega(H) = \frac{\mathrm{d}\Phi}{\mathrm{d}H} \tag{4.66}$$

$$g(Y) = \frac{\mathrm{d}\chi}{\mathrm{d}Y}. \tag{4.67}$$

We obtain the marginal pdfs for H and Y from (4.54) and (4.56)

$$\varrho(H|N, P, \beta) = \frac{1}{N!h^{3N}V_0} \frac{\Omega(H)\,\mathrm{e}^{-\beta H}}{\Delta(N, \beta, P)} \tag{4.68}$$

$$\varrho(Y|N, P, \beta) = \frac{1}{N!h^{3N}V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \frac{g(Y)\,\mathrm{e}^{-\beta Y}}{\Delta(N, \beta, P)}. \tag{4.69}$$

We may also rewrite the partition function as

$$\Delta(N, P, \beta) = \frac{1}{N!h^{3N}V_0} \int_{-\infty}^{\infty} \mathrm{d}H\ \Omega(H)\,\mathrm{e}^{-\beta H} \tag{4.70}$$

$$= \frac{1}{N!h^{3N}V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \int_{-\infty}^{\infty} \mathrm{d}Y\ g(Y)\,\mathrm{e}^{-\beta Y} \tag{4.71}$$

Making changes of variables to Φ and χ we have the pdfs for Φ and χ and alternative expressions for the partition function

$$\varrho(\Phi|N, P, \beta) = \frac{1}{N!h^{3N}V_0} \frac{\mathrm{e}^{-\beta H(\Phi)}}{\Delta(N, \beta, P)} \tag{4.72}$$

$$\varrho(\chi|N, P, \beta) = \frac{1}{N!h^{3N}V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \frac{\mathrm{e}^{-\beta Y(\chi)}}{\Delta(N, \beta, P)}. \tag{4.73}$$

$$\Delta(N, P, \beta) = \frac{1}{N!h^{3N}V_0} \int_{0}^{\infty} \mathrm{d}\Phi\ \mathrm{e}^{-\beta H(\Phi)} \tag{4.74}$$

$$= \frac{1}{N!h^{3N}V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \int_{0}^{\infty} \mathrm{d}\chi\ \mathrm{e}^{-\beta Y(\chi)} \tag{4.75}$$

4.4 Systems That Exchange Energy and Particles with the Universe: Fixed V, T, and μ

Since the arguments here are the same as in Sects. 4.2 and 4.3 I simply state the results.

We define the chemical potential μ by

$$\mu\beta = -\frac{\partial \log \Omega}{\partial N}. \tag{4.76}$$

We consider a combined system with N particles and total energy E. Next we make a division of this system into a small system (labelled 1) that can exchange particles and energy with a larger system (labelled 2). We find finally that the maximum number of states in the combined system occurs when

$$\frac{\partial \log \Omega_1}{\partial E_1} = \frac{\partial \log \Omega_2}{\partial E_2} \ : \ E_1 + E_2 = E \tag{4.77}$$

$$\frac{\partial \log \Omega_1}{\partial N_1} = \frac{\partial \log \Omega_2}{\partial N_2} \ : \ N_1 + N_2 = N \tag{4.78}$$

This corresponds to equal values of $\beta = \frac{1}{k_B T}$ and μ in the system and the universe. Therefore, if we bring our system into contact with the universe such that the two can exchange energy and particles, then their most probable combined state is one where their temperatures and chemical potentials are the same.

Assuming that quantum states of our small system are clearly spaced, while those of the universe are far closer together, we find the combined density of states to be approximately

$$\Omega_{1,2}(E, \epsilon, N, n) \propto \exp\left\{-E_1 \frac{\partial \log \Omega_1}{\partial E} - N_1 \frac{\partial \log \Omega_1}{\partial N} + O\left(E_1^2\right) + O\left(N_1^2\right)\right\} \tag{4.79}$$

$$\simeq \exp\{-\beta(E_1 - \mu N_1)\}. \tag{4.80}$$

We therefore obtain the pdf and partition function

$$\varrho(\mathbf{q}, \mathbf{p}, N|\beta, V, \mu) = \frac{1}{N! h^{3N} \Xi} \mathrm{e}^{-\beta[\mathcal{H}(\mathbf{q},\mathbf{p}) - \mu N]} \tag{4.81}$$

$$\begin{aligned} \Xi(V, \beta, \mu) &= \sum_{N=0}^{\infty} \frac{1}{N! h^{3N}} \int \mathrm{d}^N \mathbf{q} \, \mathrm{d}^N \mathbf{p} \, \mathrm{e}^{-\beta[\mathcal{H}(\mathbf{q},\mathbf{p}) - \mu N]} \\ &= \sum_{N=0}^{\infty} Z(N, \beta, V) \, \mathrm{e}^{\beta \mu N}. \end{aligned} \tag{4.82}$$

4.5 Entropy

Consider a system with Ω discrete quantum states, with probabilities $\{\mathcal{P}_i\}$. Gibbs defined the entropy of the system to be

$$S = -k_{\mathrm{B}} \sum_{i=1}^{\Omega} \mathcal{P}_i \log \mathcal{P}_i . \tag{4.83}$$

In cases such as the microcanonical ensemble where all states have equal probability $\mathcal{P} = \frac{1}{\Omega}$ then (4.83) simplifies to

$$S = k_{\mathrm{B}} \log \Omega . \tag{4.84}$$

Equation (4.84) is called the "Boltzmann Entropy".

Chapter 5
Relating Probability Density Functions to the Behaviour of Systems

5.1 Phase Transitions

How do we decide whether a material should be a liquid or a gas? Let's say that we have a system at fixed pressure and temperature. The liquid and gas phases correspond to regions of extended phase space $(\mathbf{q}, \mathbf{p}, V)$: $D(L)$ and $D(G)$ respectively. We may calculate the probabilities for the liquid and gas phases by integrating the pdf (4.54) over these two regions

$$\text{prob}(L) = \int_{D(L)} \mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}\,\mathrm{d}V\,\varrho(\mathbf{q}, \mathbf{p}, V) \tag{5.1}$$

$$\text{prob}(G) = \int_{D(G)} \mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}\,\mathrm{d}V\,\varrho(\mathbf{q}, \mathbf{p}, V)\,. \tag{5.2}$$

Typically for a large number of particles the probabilities of the two phases will be very different. There is a single inverse temperature β at which the probabilities of the two phases are equal, and we call this point the "phase transition". Since the two phases have equal probabilities they coexist at the phase transition.

There are two broad classes of phase transition.

First order phase transitions involve discontinuous changes in the thermodynamic properties. They occur when the balance of probability shifts abruptly from one phase to another which has different physical characteristics. The correlation length ξ is generally finite at a first order phase transition. An example of a first order phase transition is evaporation of a liquid. Here the density and energy per particle change discontinuously when the system changes from liquid to gas.

Second-order phase transitions are continuous transitions. The correlation length becomes effectively infinite in a second order transition and fluctuations become

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_5

correlated over all distances. Approaching the transition from either side, the two phases become identical. An example is the critical point at the end of the evaporation line which separates the liquid and gas phases.

5.2 The Equation of State

Consider a very large system with effectively infinitely many particles N. The system is at fixed temperature and pressure far from any second order phase transition. Therefore the correlation length is finite and we may consider the system to be made up of very many statistically independent subsystems. According to equation (3.14) the relative fluctuations of extensive variables such as volume go to zero for infinite N. Therefore, when we fix (P, T) we in fact fix (P, T, V). Therefore there exists a unique equation of state

$$\phi\,(P, T, V) = 0. \tag{5.3}$$

This is true everywhere except at first order phase transitions where the system may make a discontinuous change in volume. At these points (a set of measure zero), the equation of state is multivalued. Additionally, at second order phase transitions the correlation length diverges. However, arbitrarily close to the phase transition the correlation length is finite, though large. Considering that the mean volume changes continuously across such second order phase transitions, we see that the equation of state is a well behaved, simple function, in the region of a second order phase transition. It should be noted that fluctuations of volume become large at a second order phase transition, and therefore the value reported by the equation of state there is nothing but the mean.

As a last remark, it is quite possible to specify a temperature at which a first order phase transition occurs together with a volume between those volumes corresponding to either state. In this case the system separates into two phases of density ρ_1 and ρ_2, the amount of each being controlled by these quantities, the total number of particles, and the volume of the system.

A typical equation of state for an atomic fluid is shown in Fig. 5.1. There are many points to note here, and we will return to the figure in detail in the next Sect. 5.3.

5.3 Phase Diagrams

Projecting the equation of state, Fig. 5.1 into any of the 2D planes $P - T$, $P - V$, $V - T$ we obtain a "phase diagram". The phase diagram shows which phase is stable at each combination of those variables. Typical $P - T$ and $P - V$ phase diagrams for this equation of state are shown in Fig. 5.2.

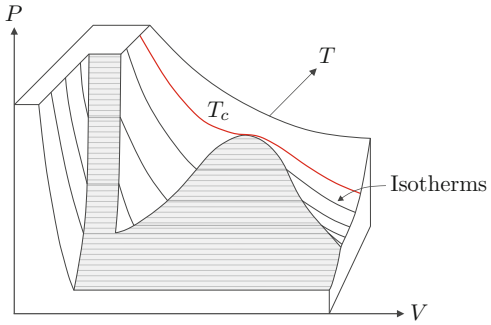

Fig. 5.1 A typical equation of state for an atomic or molecular fluid

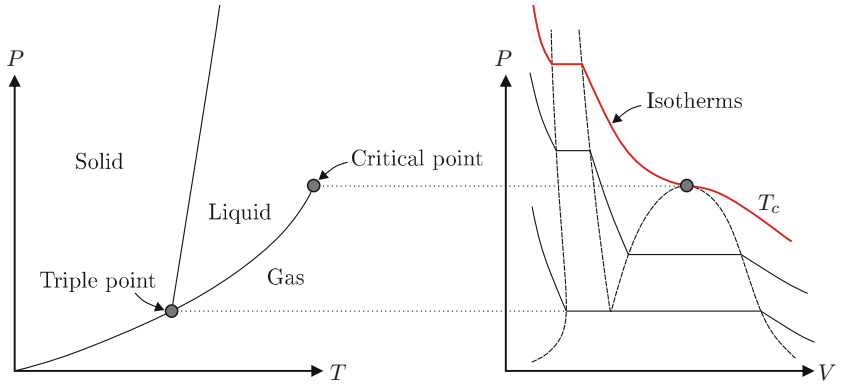

Fig. 5.2 Typical $P - T$ and $P - V$ phase diagrams, corresponding to the equation of state shown in Fig. 5.1

There are many things to note from these phase diagrams. First, transitions between solids, liquids and gases are first order phase transitions, where the volumes of the phases are not equal. From Sects. 4.2–4.4 the pressure, temperature and chemical potentials of the two phases must be equal at a phase transition. As a result, phase transitions correspond to points in the $P - T$ diagram, and horizontal isotherms in the $P - V$ diagram.

At each point on the first order phase boundaries the volume is multivalued. In particular, at the triple point in the $P - T$ diagram, the solid, liquid and gas phases have equal probability and the pdf for volume is the sum of three delta functions. At the top of the liquid - gas line there is a second order phase transition at the "critical point" (T_c, P_c), above which the volume becomes single valued. The state at pressures and temperatures greater than (T_c, P_c) is often referred to as the "supercritical phase", although a precise definition of this phase is contentious [1].

It is often stated that there is no distinction between liquid and gas above (T_c, P_c), but we will see in Sect. 5.5 that there is still structure within this region.

5.4 Thermodynamic Potentials

It is traditional to define thermodynamic potentials such as the free energy

$$F(\beta, V, N) = -\frac{1}{\beta} \log Z(\beta, V, N). \tag{5.4}$$

Recall that the partition function Z is just the sum of the unnormalised probabilities of all states; the normalising factor in the pdf (4.38). Consider two phases L and G that correspond to the configuration space regions $D(L)$ and $D(G)$. One may define free energies for each phase separately by evaluating only the contribution to Z made by the regions $D(L)$ and $D(G)$. Since the logarithm is a monotonic function, the phase with highest probability also has the lowest free energy. This is the language in which thermodynamics determines the relative stability of phases.

Clearly, it is only appropriate to compare the free energies of phases when we are interested in conditions of fixed β, V, N. We define the Gibbs free energy (5.5) in order to compare phases under conditions of fixed β, P, N.

$$G(\beta, P, N) = -\frac{1}{\beta} \log \Delta(\beta, P, N) \tag{5.5}$$

Additionally, we define the Grand Potential (5.6) for comparing the stability of phases at fixed β, V, μ.

$$\Phi(\beta, V, \mu) = -\frac{1}{\beta} \log \Xi(\beta, V, \mu) \tag{5.6}$$

In each case, the phase with the lowest value of the appropriate thermodynamic potential has the highest probability.

5.5 Signatures of Phase Transitions

Consider a system at fixed pressure and temperature. We have seen in Sects. 5.1 and 5.2 that extensive quantities, such as volume and enthalpy, take well defined values for each phase, and that these values can be very different for each phase, at a phase transition (this is the basis for a first order transition). For each phase, the pdf of each variable will be a unimodal, sharply peaked distribution. At a phase transition, the pdf of each variable will have two modes of equal weight, their means separated by far more than the widths of the individual peaks. Consider the variance of such a distribution. The variance is small for a sharp unimodal distribution, which occurs away from a phase transition. At the phase transition, where the pdf becomes suddenly bimodal, the variance peaks dramatically. In this way the variance of extensive quantities are clear indicators of a first order phase transition. Similarly for second order phase transitions, such variances also peak. This is because

the pdfs of extensive variables become extremely broad at a second order phase transition.

For a system at fixed β, P and N, the variance of enthalpy and volume are both sharply peaked at a phase transition. Rather than address these statistics directly, it is more useful to refer to them via closely related measurable quantities: the isobaric heat capacity C_P, and isothermal compressibility κ_T. The heat capacity and isothermal compressibility, together with many other similar functions are collectively referred to as "response functions" or "susceptibilities".

The heat capacity for a system at fixed T, P, N is defined as the rate at which the enthalpy changes with respect to temperature:

$$C_P = \frac{\partial H\left(T, P, N\right)}{\partial T} \tag{5.7}$$

$$= k_{\mathrm{B}}\beta^2 \frac{\partial^2 \log \Delta\left(\beta, P, N\right)}{\partial \beta^2} \tag{5.8}$$

$$= k_{\mathrm{B}}\beta^2 \mathrm{Var}\left(H\left(\mathbf{q}, \mathbf{p}, P, V\right)\right). \tag{5.9}$$

The isothermal compressibility is similarly defined as (minus) the rate of change of the volume with respect to pressure, normalised by the volume:

$$\kappa_T = -\frac{1}{V}\frac{\partial V\left(T, P, N\right)}{\partial P} \tag{5.10}$$

$$= \frac{1}{\beta V}\frac{\partial^2 \log \Delta\left(\beta, P, N\right)}{\partial P^2} \tag{5.11}$$

$$= \frac{\beta}{V}\mathrm{Var}\left(V\right). \tag{5.12}$$

As a final warning, it should be noted that although a peak in the heat capacity and isothermal compressibility is always present at a phase transition in a system at fixed N, P, T, the presence of a peak does not guarantee that a phase transition has occurred. Such peaks also occur along the "Widom lines" [1, 2] in the supercritical region, where no phase transitions occur.

The liquid and gas phases have different characteristic values for most quantities, and so the response functions for those quantities will all exhibit a peak at the evaporation transition, where the two phases coexist. In the supercritical region the clear distinction between the phases is lost. Nevertheless, each response function peaks where the variance of its corresponding observable is largest. Since there is no clear distinction between the liquid and gas phases, these maxima do not occur at the same values of (P, β), as they do along the evaporation transition. The lines of maxima for the response functions in the supercritical region are called the Widom lines [1]. Below the critical point the Widom lines correspond to the evaporation transition. Above the critical point the Widom lines diverge. The peaks for each response function become broader and smaller as we move further along the Widom line, into the supercritical region [1], until the point where it is impossible to resolve the peak.

5.6 Finite Size Simulations

This thesis is concerned with atomistic simulation. Here one often aims to calculate properties of an infinite system from simulations of a much smaller system. For simplicity, let us consider N particles constrained to a cell of side length L with periodic boundary conditions. When the correlation length ξ is finite, the infinite system may be considered as the combination of many statistically independent subsystems of size $L \gg \xi$, an idea already introduced in Sect. 3.5. In this case the pdf for any extensive quantity as calculated in a subsystem has the same form as the distribution in the infinite system, but with a larger relative fluctuation. This was described in Sect. 3.6, where it was shown that the relative fluctuation goes as $\frac{1}{\sqrt{N}}$. Therefore we may extrapolate to the infinite system limit from a finite calculation, so long as we ensure that $L \gg \xi$.

In practice, the correlation length can be very large, so that it is unfeasible to simulate a statistically independent subsystem. This is always the case at a second order phase transition where the correlation length becomes effectively infinite, $\xi \to \infty$, but this limitation also holds more generally. We might assume that it would be impossible to obtain meaningful results from a system with $L < \xi$. Considerable work has been done on the problem of extrapolating to larger system sizes. We will not be concerned with finite size corrections in this thesis, and instead refer the interested reader to the texts [3–6].

5.7 Periodic Boundary Conditions

Periodic boundary conditions introduce an unphysical periodicity into simulations of non-periodic systems, such as a gas. However, if our simulation cell is larger than the correlation length, then these periodic images are effectively independent atoms, "unaware" of each other, and results calculated for the subsystem are correct. Even in simulations that are smaller than the correlation length but "large enough", the effect of these correlations may be only slight. Periodic boundary conditions are an effective way of accounting for the surface interaction of subsystems. This ensures that quantities such as the energy and volume are extensive quantities, which is essential for deducing properties of the infinite system limit as described in Sect. 5.6. It should be noted that under periodic boundary conditions, there is nothing unique about the edges or corners of the cell, since these points can be mapped to any other part of the cell by translating all the atoms and applying the boundary conditions.

5.8 Fractional Coordinates and Fully Flexible Simulation Cells

This thesis is concerned with approximating infinite systems by using only a small, fixed number of particles, N. Different solid phases may have different crystal lattice structures, and there may be no way to arrange N particles in a fixed shape simulation cell to produce a particular lattice structure. Simulating N particles in, for example, a cubic simulation cell, may exclude certain phases from the results of our calculation. This is an example of a finite size effect, which can have terrible consequences. A second finite size effect arises from the impossibility of representing a crystal motif with an inappropriate number of nuclei. To pick a specific example, it is not possible to arrange 65 nuclei into a periodic hexagonal close-packed (HCP) lattice. Nor is it possible to construct a HCP lattice using periodic images of just one nucleus ($N = 1$); we require at least two.

Using a fully flexible simulation cell [7, 8] helps to alleviate the first of these system size effects, by allowing the simulation cell to take any shape. Sampling with a fully flexible simulation cell requires that we explore both atomic coordinates and shapes of the simulation cell. To this end it is useful to work in "fractional coordinates" $\mathbf{s}$, related to $\mathbf{q}$ and the "cell matrix" $\mathbf{h}$ by

$$\begin{aligned} \mathbf{s}_i &\in [0, 1)^3 \\ \mathbf{q}_i &= \mathbf{h}\mathbf{s}_i \end{aligned} \tag{5.13}$$

The cell matrix $\mathbf{h}$ is a real 3×3 matrix and need not be symmetrical. Columns of the cell matrix correspond to the cell vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. This is illustrated in Fig. 5.3.

The volume of the cell can be obtained from $\mathbf{h}$ by

$$V = \det \mathbf{h}. \tag{5.14}$$

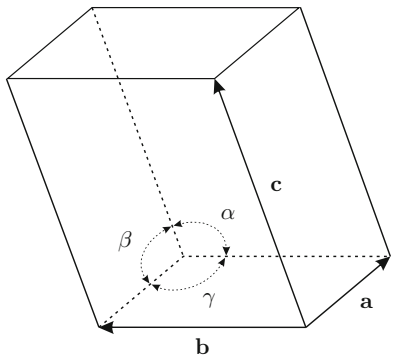

Fig. 5.3 A general simulation cell in the fully flexible simulation cell formulation. Rows of the cell matrix $\mathbf{h}$ correspond to the cell vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. Also shown are the conventional axial angles α, β and γ

Next we introduce the "unit cell matrix" $\mathbf{h}_0$ which corresponds to the cell matrix scaled to unit volume.

$$\mathbf{h}_0 = \mathbf{h} V^{-\frac{1}{3}} \tag{5.15}$$

$$\det \mathbf{h}_0 = 1 \tag{5.16}$$

We will find it useful to relate the coordinates $\mathbf{q}$ and $\mathbf{s}$ by $\mathbf{h}_0$ and V explicitly:

$$\mathbf{q}_i = V^{1/3} \mathbf{h}_0 \mathbf{s}_i . \tag{5.17}$$

One generalisation of the pdf and partition function for systems at fixed N, P, β to incorporate a fully flexible simulation cell was introduced by Martyna, Tobias and Klein [7]. In that picture $\mathbf{h}_0$ samples a uniform distribution over the surface $\det \mathbf{h}_0 = 1$.

$$H(\mathbf{s}, \mathbf{p}, P, V, \mathbf{h}_0) = E\left(V^{\frac{1}{3}} \mathbf{h}_0 \mathbf{s}, \mathbf{p}\right) + PV \tag{5.18}$$

$$\varrho(\mathbf{s}, \mathbf{p}, V, \mathbf{h}_0) = \frac{V^N}{N! h^{3N} V_0} \frac{\mathrm{e}^{-\beta H(\mathbf{s}, \mathbf{p}, P, V, \mathbf{h}_0)}}{\Delta(N, \beta, P)} \delta(\det \mathbf{h}_0 - 1) \tag{5.19}$$

$$\Delta(N, \beta, P) = \frac{1}{N! h^{3N} V_0} \int_0^\infty \mathrm{d}V \, V^N \int \mathrm{d}\mathbf{h}_0 \, \delta(\det \mathbf{h}_0 - 1) \int \mathrm{d}^N \mathbf{s} \, \mathrm{d}^N \mathbf{p} \, \mathrm{e}^{-\beta H(\mathbf{s}, \mathbf{p}, P, V, \mathbf{h}_0)} \tag{5.20}$$

There are three points to note when comparing (5.19) and (5.20) to the pdf and partition function for a fixed simulation cell (4.59) and (4.60). First, we are now expressing the pdf as a density over the fractional coordinates $\mathbf{s}$. Second, the transformation from $\mathbf{q}$ to $\mathbf{s}$ has introduced a factor of V^N. This is because $\mathrm{d}q_i = V^{\frac{1}{3}} \mathrm{d}s_i$ and there are $3N$ such terms in $\mathrm{d}^N \mathbf{q}$. Third, integration over elements of $\mathbf{h}_0$ in (5.20) is over all real values of the elements of $\mathbf{h}_0$, from $(-\infty, \infty)$. Therefore (5.20) corresponds to uniform integration over the surface where $\mathbf{h}_0$ has determinant 1.

Tuckerman gives a detailed derivation of the pressure tensor with partition function (5.20) in [9], page 230. That derivation makes it clear that the uniform distribution over $\mathbf{h}_0$ is required to obtain the correct pressure virial theorem and work virial. For this reason, I will use the fromulation of Martyna, Tobias and Klein [7] throughout this thesis.

References

1. V.V. Brazhkin, A.G. Lyapin, V.N. Ryzhov, K. Trachenko, Y.D. Fomin, E.N. Tsiok, Where is the supercritical fluid on the phase diagram? Phys. Uspekhi **55**, 1061 (2012)
2. V. Brazhkin, Y.D. Fomin, A. Lyapin, V. Ryzhov, E. Tsiok, Widom line for the liquid-gas transition in Lennard-Jones system. J. Phys. Chem. B **115**, 14112 (2011)

3. K. Binder, Finite size scaling analysis of Ising model block distribution functions. Z. Phys. B Condens. Matter **43**, 119 (1981)
4. K. Binder, D. Landau, Finite-size scaling at first-order phase transitions. Phys. Rev. B **30**, 1477 (1984)
5. J. Cardy, *Finite-Size Scaling* (Elsevier, 1988)
6. V. Privman, *Finite Size Scaling and Numerical Simulation of Statistical Systems* (World Scientific, 1990)
7. G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms. J. Chem. Phys. **101**, 4177 (1994)
8. M. Parrinello, A. Rahman, Crystal structure and pair potentials: a moleculardynamics study. Phys. Rev. Lett. **45**, 1196 (1980)
9. M. Tuckerman, *Statistical Mechanics and Molecular Simulations* (Oxford University Press, 2008)

Part II
Nested Sampling

Chapter 6
The Strategy of Nested Sampling

For the purposes of this discussion we will consider conditions of constant pressure. This is only so that we may illustrate points with real data calculated as described later in this thesis. The same discussion could be had for conditions of constant volume and the same conclusions would apply.

In Part I we saw how all observables in statistical mechanics follow from the partition function. Recall these forms of the partition function, both written in terms of the configurational enthalpy Y

$$\Delta(N, P, \beta) = \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \int_{-\infty}^{\infty} \mathrm{d}Y \; g(Y) \, \mathrm{e}^{-\beta Y} \qquad \text{(4.71 revisited)}$$

$$= \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \int_{0}^{\infty} \mathrm{d}\chi \; \mathrm{e}^{-\beta Y(\chi)}. \qquad \text{(4.75 revisited)}$$

Fundamentally, calculating the partition function, or any quantity derived from it, amounts to evaluating the functions $Y(\chi)$ and $g(Y)$, at all values of χ and Y that contribute appreciable weight to (4.71) and (4.75) for the external conditions (P, β). If our approach to calculating the functions $Y(\chi)$ and $g(Y)$ is to be one of statistical sampling, then this in turn requires that we characterise the pdfs for χ and Y (4.73) and (4.69) sufficiently well.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_6

$$\varrho\left(\chi|N,P,\beta\right)=\frac{1}{N!h^{3N}V_0}\left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}}\frac{\mathrm{e}^{-\beta Y(\chi)}}{\Delta\left(N,\beta,P\right)} \qquad \text{(4.73 revisited)}$$

$$\varrho\left(Y|N,P,\beta\right)=\frac{1}{N!h^{3N}V_0}\left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}}\frac{g\left(Y\right)\mathrm{e}^{-\beta Y}}{\Delta\left(N,\beta,P\right)} \qquad \text{(4.69 revisited)}$$

Figure 6.1 shows the function $Y\left(\log_{10}\chi\right)$ (black line). The three bimodal distributions (blue, green and red in order of increasing temperature) show pdfs for Y (4.69), close to the evaporation temperature[1] T_{e}^{*} of Lennard-Jonesium.[2] Configurational enthalpy is shown on the vertical axis, with the value of these pdfs shown on the top horizontal axis. The green curve is very close to the evaporation temperature.

Obtaining an accurate value for the relative weight of the two peaks is the goal of any sampler seeking to resolve a phase transition, since at a phase transition two phases are in thermal equilibrium and therefore have equal probability. Locating a phase transition therefore amounts to finding the thermodynamic conditions in which the two modes in Fig. 6.1 have equal weight. Thermal sampling schemes, which seek to sample according to the physical pdf, are usually based on some process that mimics the physical system. As in the physical process the low probability range for Y inhibits movement of the sampler between the two phases. Without frequent transitions between the phases it is very unlikely that the sampler will obtain an accurate value for the relative weight of the two peaks. Indeed, in their paper Pártay et al. [1] showed that the error for the heat capacity of an evaporating atomic cluster is much greater close to the transition peak when calculated using parallel tempering (a temperature scheduling algorithm). This is shown in Fig. 6.2.

Nested sampling [2, 3] takes a different approach: it attempts to resolve the value of Y at evenly spaced values of $\log\chi$. If one could achieve this aim, then it is easy to believe that this approach should be an efficient means of resolving the function $Y\left(\log\chi\right)$; the approach pays no regard to whether χ corresponds to a first order phase transition or not and $Y\left(\log\chi\right)$ is a smooth function. We will describe the methodology of nested sampling in detail during the chapters that follow. For now however, notice how well behaved the errors are at the peaks of the C_V curves shown in Fig. 6.2. This is the motivation for this thesis, in which we aim to resolve phase transitions such as that shown in Fig. 6.1 where χ can vary by hundreds of orders of magnitude.

[1] Reduced temperature is defined as $T^{*}\equiv\frac{k_{\mathrm{B}}T}{\epsilon}$.

[2] The results shown were calculated using a periodic system of 128 Lennard-Jonesium particles at reduced pressure $\log_{10}\left(P^{*}\equiv\frac{P\sigma^{3}}{\epsilon}\right)=-1.194$. The interatomic potential was exactly that described later in Sect. 8.8. These results were calculated using Nested Sampling, as described in Sect. 8.3.

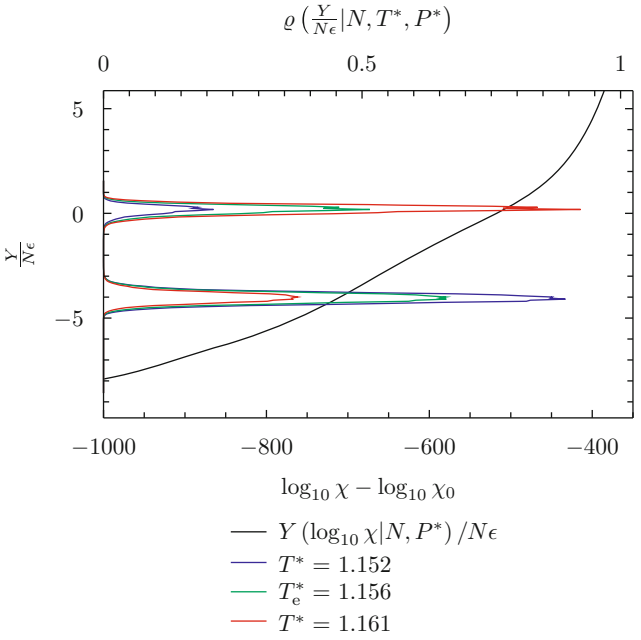

Fig. 6.1 The function $Y\left(\log_{10}\chi\right)$ (black line) is shown for 128 Lennard-Jonesium particles. χ_0 is the total volume of the configuration space (see Sects. 8.6 and 8.7). Notice the inflection in the black curve at $\log_{10}\chi - \log_{10}\chi_0 \simeq -650.0$: this corresponds to the entropy jump from the liquid to the gas phase. The three bimodal distributions (blue, green and red in order of increasing temperature) show pdfs for Y (4.69), close to the evaporation temperature at reduced pressure $\log_{10} P^* = -1.194$

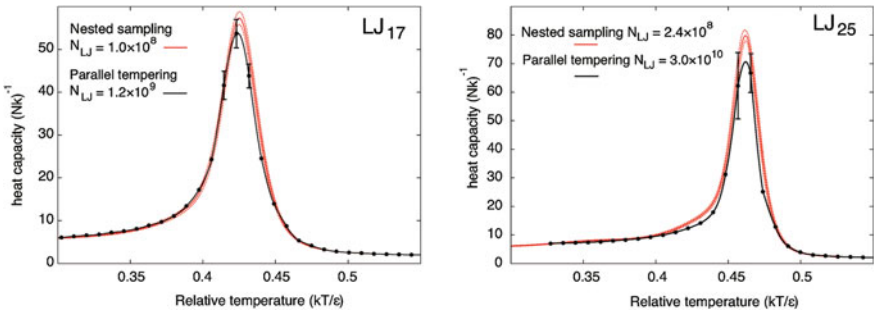

Fig. 6.2 Temperature scheduling algorithms (in this case parallel tempering) have largest error close to a phase transition. (Here, interpolation between temperatures in parallel tempering was performed using the method of Ferrenberg and Swendsen [4].) Nested sampling does not have the same bad behaviour. For LJ_{25}, nested sampling has resolved the transition with lower error bars, at a hundredth of the cost. This figure shows the "evaporation" transition of Lennard-Jonesium clusters LJ_{17} and LJ_{25} Reprinted (adapted) with permission from [1]. Copyright (2010) American Chemical Society

References

1. L.B. Pártay, A.P. Bartók, G. Csányi, Efficient sampling of atomic configurational spaces. J. Phys. Chem. B **114**, 10502 (2010), http://dx.doi.org/10.1021/jp1012973
2. J. Skilling, Nested sampling, in *AIP Conference Proceedings*, vol. 735, (2004), pp. 395
3. J. Skilling, 'Nested sampling for general Bayesian computation. Bayesian Anal. **1**, 833 (2006)
4. A.M. Ferrenberg, R.H. Swendsen, Optimized Monte Carlo data analysis. Phys. Rev. Lett. **63**, 1195 (1989)

Chapter 7
An Introduction to Nested Sampling

Throughout this introduction we assume that we are interested in the NVT ensemble with fixed simulation cell, described in Sect. 4.2.

As described in Chap. 6, we seek to calculate the inverse function to the integrated density of states $U(\chi)$, in order to calculate the partition function $Z(N, V, \beta)$ 4.46.

$$Z(N, \beta, V) = \frac{1}{N! h^{3N}} \left(\frac{2\pi m}{\beta} \right)^{\frac{3N}{2}} \int_0^\infty \mathrm{d}\chi \mathrm{e}^{-\beta U(\chi)} \qquad \text{(4.46 revisited)}$$

$$\chi(N, U, V) = \int_{-\infty}^{U} g\left(N, U', V\right) \mathrm{d}U' \qquad \text{(4.20 revisited)}$$

$$g\left(N, U', V\right) = \frac{1}{N! h^{3N}} \int \mathrm{d}^N \mathbf{q}\, \delta\left(U\left[\mathbf{q}\right] - U'\right) \qquad \text{(4.5 revisited)}$$

The aim of nested sampling [1, 2] is to calculate $U(\chi)$ at a geometric series of values for χ:

$$\chi \in \left\{ \chi_0 t, \chi_0 t^2, \ldots, \chi_0 t^{N_{\text{its}}} \right\}. \qquad (7.1)$$

For each of $\{\chi_i\}$ we calculate the respective energy $U(\chi_i)$.

The strategy for achieving this is to set each χ_i within some error, in a probabilistic way, then calculate the exact energy for the value of χ_i we have obtained. Convergence of the algorithm is achieved by reducing the error on $\{\chi_i\}$.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_7

7.1 The Nested Sampling Method

We begin with a graphic illustration of the central nested sampling procedure, without mathematical background. We follow this with a technical exposition of the nested sampling algorithm and its basis. Finally, we give an intuitive explanation of the algorithm as a microcanonical cooling algorithm.

7.1.1 A Graphical Introduction to the Nested Sampling Procedure

A graphical illustration of the central nested sampling procedure (or loop) is shown in Fig. 7.1.

As described in Fig. 7.1, at the start of each iteration we have $\mathcal{K}$ configurations drawn from a uniform distribution over the region $U(\mathbf{q}) \leq U_{\text{lim}}$, where U_{lim} depends on the iteration. At each iteration U_{lim} is reduced. This is the basis for the name "nested sampling".

It is slightly less obvious that samples drawn uniformly at random from the region $U(\mathbf{q}) \leq U_{\text{lim}}$ have values $\chi(U)$ (4.20) that are drawn uniformly at random from the interval $\chi \in [0, \chi(U_{\text{lim}})]$. We delay a rigorous explanation of this until the next section. However, me might consider dividing our (continuous) configuration space into small pieces of equal volume, and arranging those pieces along a line, sorted from high energy to low energy. Lengths of the line correspond to volumes of configuration space. That is, sections of the line correspond to intervals of χ. A configuration chosen at random from the space corresponds to a point chosen at random on the line. The iteration shown in Fig. 7.1 is represented again in this picture, in Fig. 7.2.

7.1.2 The Nested Sampling Algorithm in Detail

Each χ_i is found from the previous χ_{i-1}, where $\chi_{i-1} > \chi_i$ and $U(\chi_{i-1}) > U(\chi_i)$. Thus sampling is performed "inwards" towards lower and lower energies. This is the basis for the name "nested sampling".

At any given iteration i we have $\mathcal{K}$ configurations $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{\mathcal{K}}\}$ distributed uniformly throughout the region $U(\mathbf{q}) \leq U_{\text{lim}}$. Here $U_{\text{lim}} = U_{i-1}$, the configurational energy exactly corresponding to the approximate value we obtained for $\log \chi_{i-1}$. The pdf describing this situation is

$$\varrho(\mathbf{q}|U_{\text{lim}}) = \begin{cases} \frac{1}{\chi(U_{\text{lim}})}, & U(\mathbf{q}) \leq U_{\text{lim}} \\ 0, & \text{Elsewhere.} \end{cases} \tag{7.2}$$

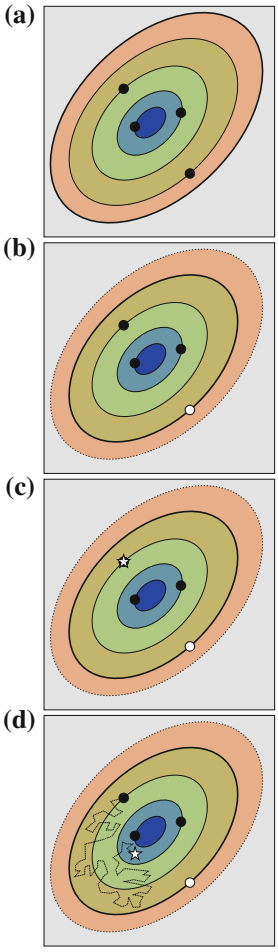

Fig. 7.1 A graphical representation of the nested sampling loop. The square box represents our configuration space, and *dots* represent configurations. The configurations have different energies, as represented by the energy contours. The (*central*) *blue* region has lowest energy, and the (*outer*) *grey* region has highest energy. **a** At the start of each iteration we always have $\mathcal{K}$ configurations placed uniformly at random throughout the region of configuration space with energy $U(\mathbf{q}) \leq U_{\text{lim}}$. The contour $U(\mathbf{q}) = U_{\text{lim}}$ is represented by the *thick line*. **b** Next, we identify the configuration with highest energy from the $\mathcal{K}$ configurations, indicated by the hollow *dot*. The value of U_{lim} is updated to the energy of that configuration. We save the new value of U_{lim} and the configuration with $U(\mathbf{q}) = U_{\text{lim}}$ as output. Finally, the configuration with $U(\mathbf{q}) = U_{\text{lim}}$ is removed. This leaves $\mathcal{K} - 1$ configurations in our sample set, all with $U(\mathbf{q}) \leq U_{\text{lim}}$. **c** We choose a configuration at random from the remaining $\mathcal{K} - 1$ configurations and copy it, producing a "clone". The clone is represented by the *white* star. **d** We perform a random walk on the clone through the region $U(\mathbf{q}) \leq U_{\text{lim}}$, until it has no memory of its starting configuration. The walk is constructed so that the pdf for the clone's final position is a uniform distribution in the region $U(\mathbf{q}) \leq U_{\text{lim}}$, and zero for $U(\mathbf{q}) > U_{\text{lim}}$. We now return to the start of the loop, and the loop is repeated using the most recent value of U_{lim}

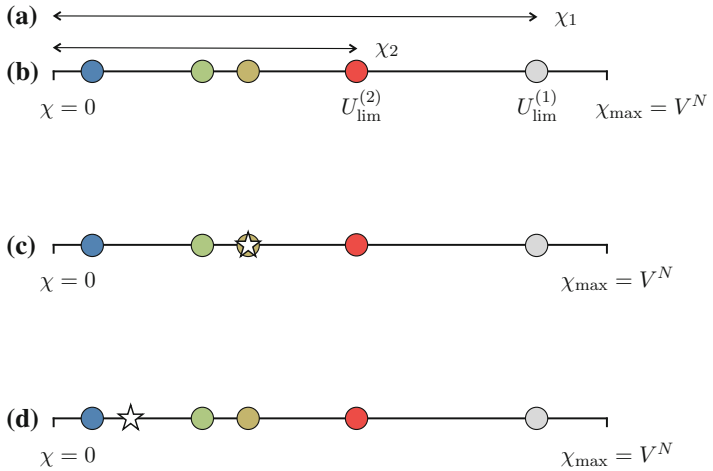

Fig. 7.2 We imagine dividing (continuous) configuration space into small pieces of equal size, and sorting those pieces along a line, from low energy (left) to high energy (right). Starting from the *left* end, the length of the line up to $U = U_{\rm lim}$ corresponds to the volume of configuration space with $U \leq U_{\rm lim}$. This is precisely $\chi(U_{\rm lim})$ (4.20). At the *left* of the line we have $\chi(U_0) = 0$, and at the *right* we have $\chi(U_{\rm max}) = V^N$ (see Sect. 7.1.4). Each infinitesimal length of the line corresponds to a tiny region of configuration space. Configurations chosen randomly from the space correspond to a point chosen randomly from the line. In this way, configurations chosen randomly from the configuration space with $U \leq U_{\rm lim}$ correspond to uniformly randomly distributed points along the line with $\chi \leq \chi(U_{\rm lim})$. The labels on the *left* hand side indicate the stages (**a**)–(**d**) in Fig. 7.1 to which each line corresponds. Here we include a *grey circle* to indicate the sample that occurred at $U_{\rm lim}^{(1)}$, in the previous iteration. On the first line we show the relationship between χ_1 and χ_2. It can be shown [2] that, with $\mathcal{K}$ points along a unit interval $\chi \in [0, \chi_1]$, the largest value of χ in our sample set is a beta distributed random number: $\text{prob}(\chi_2|\mathcal{K}, \chi_1) = \mathcal{K}\left(\frac{\chi_2}{\chi_1}\right)^{\mathcal{K}-1} \times \chi_1$. Therefore, χ_2 has mean value $\langle\chi_2\rangle = \frac{\mathcal{K}}{\mathcal{K}+1}\chi_1$, with error that decreases as $\mathcal{K}$ is made larger. These details are described in the next section. In stage (**c**), a random configuration is copied, producing a "clone", as indicated by the *white* star In stage (**d**), this clone is walked throughout the region $U(\chi) < U_{\rm lim}^{(2)}$, until it is completely independent of the its starting configuration

We call this distribution the "bounded uniform distribution". The pdf for U for a configuration drawn uniformly from configuration space with configurational energy $U(\mathbf{q}) \leq U_{\rm lim}$ is simply the density of states below $U_{\rm lim}$

$$\varrho(U|U_{\rm lim}) = \begin{cases} \frac{g(U)}{\chi_{\rm lim}}, & U \leq U_{\rm lim} \\ 0, & \text{Elsewhere} \end{cases} \tag{7.3}$$

Here $\chi_{\rm lim} \equiv \chi(U_{\rm lim})$. The pdfs for χ and $\log\chi$ are related to (7.3) by a change of variables.

$$\varrho\left(\chi|\chi_{\text{lim}}\right) = \begin{cases} \frac{1}{\chi_{\text{lim}}}, & \chi \leq \chi_{\text{lim}} \\ 0, & \text{Elsewhere} \end{cases} \tag{7.4}$$

$$\varrho\left(\log\chi|\log\chi_{\text{lim}}\right) = \begin{cases} \frac{e^{\log\chi}}{\chi_{\text{lim}}}, & \log\chi \leq \log\chi_{\text{lim}} \\ 0, & \text{Elsewhere.} \end{cases} \tag{7.5}$$

From (4.20), U and $\chi(U)$ are monotonically related. Therefore, if we sort our $\mathcal{K}$ configurations in order of descending U they are also sorted in order of descending χ. Picking the first (highest energy) configuration from our sorted list, the pdf for $t \equiv \frac{\chi_{\text{largest}}}{\chi_{\text{lim}}}$ is a beta distribution [2]

$$\varrho(t) = \mathcal{K}t^{\mathcal{K}-1}. \tag{7.6}$$

This pdf (7.6) has mean and standard deviation

$$t = \left(\frac{\mathcal{K}}{\mathcal{K}+1} \pm \frac{1}{\mathcal{K}+1}\sqrt{\frac{\mathcal{K}}{\mathcal{K}+2}}\right). \tag{7.7}$$

Picking the configuration with highest U, we know that this corresponds to a mean value of $\frac{\chi_i}{\chi_{i-1}} = \frac{\mathcal{K}}{\mathcal{K}+1}$ with standard deviation $\frac{1}{\mathcal{K}+1}\sqrt{\frac{\mathcal{K}}{\mathcal{K}+2}}$. We refer to the energy of the configuration with highest U in the i^{th} iteration as U_i. We store this configuration as output:

$$\chi_i \leftarrow \chi_{i-1} \times \frac{\mathcal{K}}{\mathcal{K}+1} \tag{7.8}$$

$$U(\chi_i) \leftarrow U_{\text{largest}} \tag{7.9}$$

Finally, having updated $U_{\text{lim}} \leftarrow U_i$ we generate a fresh configuration picked uniformly as in (7.2). The process is repeated to find U_{i+1}. For clarity, the algorithm is written again in Algorithm 7.1.

```
! This is iteration i
! U_lim = U_{i-1}
! We have K configurations distributed uniformly
! throughout the region U(q) ≤ U_lim according to (7.2).

Save sample with highest U = U_highest
  χ_i ← χ_{i-1} × K/(K+1)
  U_i ← U_highest
  U_lim ← U_i
call gen_new_sample      ! Generate new sample from (7.2)
```

Algorithm 7.1: The central loop of the nested sampling algorithm.

7.1.3 Nested Sampling as a Microcanonical Cooling Algorithm

An intuitive way to understand the nested sampling process is as a microcanonical cooling algorithm for calculating the density of states.[1] Random samples are drawn from configuration space below some upper energy limit $U \leq U_{\mathrm{lim}}$. The potential energy of such a sample is a continuous random variable distributed according to the density of states $g(U)$ below U_{lim}. The density of states is a rapidly increasing function of the potential energy, and therefore the overwhelming majority of our samples will have potential energies close to U_{lim}. As a result we resolve the density of states accurately in the interval just below U_{lim}, but our sampling at lower energies is noisy.

The nested sampling algorithm repeats this process in stages. In the upper panel of Fig. 7.3, we accurately resolve the density of states in the interval $\left[U_i, U_{i-1}\right]$. Nested sampling then generates new samples in the region $U \leq U_i$, before repeating the process in the interval $\left[U_{i+1}, U_i\right]$.

The really ingenious part of the nested sampling algorithm is that it chooses the appropriate energy interval automatically to minimise the uncertainty in χ_i. Recall from Sects. 7.1.1 and 7.1.2 that the values of χ for our samples are uniformly distributed along the interval $\chi \in [0, \chi(U_{\mathrm{lim}})]$. The number of samples in the interval $\chi \in [0, \chi(U)]$ is proportional to $\chi(U)$. Therefore, if we make the ratio $\frac{\chi(U)}{\chi(U_{\mathrm{lim}})}$ as close as possible to unity, we have a proportionally larger number of samples with which to estimate the same ratio. By adjusting the number of configurations $\mathcal{K}$ that nested sampling uses, we control the contraction ratio $\frac{\chi_i}{\chi_{i+1}}$, and the noise in our estimate of the density of states.

7.1.4 Initialisation of Nested Sampling

For the case of constant volume in a fixed cell, we can generate initial samples by generating uniformly random configurations $\mathbf{q} \in [0, V^{\frac{1}{3}})^{3N}$. In this case, the initial value for χ is

$$\chi_0 = V^N. \tag{7.10}$$

The integrated density of states and partition function are therefore absolute.

[1] The density of states may be recovered from the nested sampling output by numerical differentiation, $g(U) \simeq \frac{\Delta\chi_i}{\Delta U_i}$.

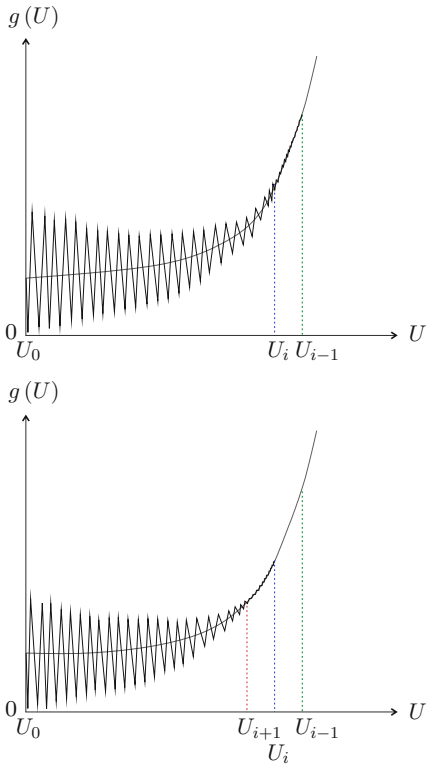

Fig. 7.3 Nested sampling may be understood as a microcanonical cooling algorithm. See Sect. 7.1.3 for discussion

7.2 Calculating the Partition Function and Other Quantities

The partition function may be approximated as follows. The full partition function is

$$Z(N, \beta, V) = A \times \int_0^\infty \mathrm{d}\chi \mathrm{e}^{-\beta U(\chi)} \qquad \text{(4.46 revisited)}$$

$$A = \frac{1}{N! h^{3N}} \left(\frac{2\pi m}{\beta} \right)^{\frac{3N}{2}}$$

The partition function can be approximated from the output of the nested sampling procedure as

$$Z(N, \beta, V) \simeq A * \sum_{i=1}^{N_{\text{its}}} \Delta\chi_i \, \mathrm{e}^{-\beta U_i} \tag{7.11}$$

$$\Delta\chi_i = \chi_0 (\zeta_{i-1} - \zeta_i) \tag{7.12}$$

Here ζ_i is the total phase space contraction after i iterations

$$\zeta_i = \prod_{n=1}^{i} t_n \tag{7.13}$$

$$\simeq \left(\frac{\mathcal{K}}{\mathcal{K}+1}\right)^i \tag{7.14}$$

We have made the simplest possible interpolation of the energy: that it is constant between successive values of χ_i. This estimate could be improved by using the trapezium rule, but the errors introduced by uncertainty in $\{\zeta_i\}$ tend to overwhelm small differences introduced by the interpolation scheme.

The partition function (7.11) is an explicit function of β and therefore the heat capacity at fixed volume, $C_V = k_\mathrm{B}\beta^2 \frac{\partial^2 \log Z(\beta)}{\partial \beta^2}$, can be calculated from (7.11) at any inverse temperature. As discussed in Sect. 5.5, the heat capacity is an important tool for identifying phase transitions.

Other quantities can also be calculated from the output of the nested sampling procedure. As an example, let us consider calculating the free energy of a region in configuration space. In this case one first decides which of the configurations saved as output belong to that region. Second, the contributions to (7.11) made by those configurations are summed. This process may be repeated for another region that does not overlap with the first. In this way it is straightforward to compare the probabilities of the two regions, or equivalently, their free energies (5.4). It is a simple task to recalculate this sum at a number of values of β. As a generalisation of this method, one may calculate entire free energy surfaces. A free energy surface is defined over particular collective variables, (ϕ, ψ): $F\left(\phi', \psi'\right) = -\frac{1}{\beta} \log\left(\frac{1}{N! h^{3N}} \int \mathrm{d}\mathbf{x} \mathrm{e}^{-\beta \mathcal{H}(\mathbf{x})} \delta\left[\phi(\mathbf{x}) - \phi'\right] \delta\left[\psi(\mathbf{x}) - \psi'\right]\right)$. The nested sampling approximation to this function is obtained by dividing the space $\{(\phi, \psi)\}$ into an array of bins, placing the nested sampling output configurations into their respective bins, and then computing directly the free energy of each bin. Computing the free energy of each bin is performed by evaluating the total partition function for configurations within that bin according to (7.11).

It is also possible to produce projections of the configuration space volume itself onto reduced coordinates, or individual potential energy minima [3, 4]. Here, the nested sampling output configurations are mapped onto the reduced coordinates or minima, and their configuration space volumes $\Delta\chi_i$ (7.12) are binned, as for a free energy surface calculation. Summing the configuration space volumes in each bin yields the volume of configuration space associated with each minimum, or small interval of the reduced coordinates.

7.3 Generating New Samples: Markov Chain Monte Carlo

New samples from (7.2) cannot be generated by rejection sampling, where random configurations are proposed, and only samples with $U \leq U_{\text{lim}}$ are accepted. This is because the volume of the allowed region in configuration space collapses exponentially as the calculation progresses. We can see this from Eq. (7.23). If we used rejection sampling, then after $100\mathcal{K}$ iterations the probability of accepting proposed configurations would be e^{-100}. Instead, new samples are generated by Markov chain Monte Carlo (MCMC), which is a two stage process. First we clone an existing sample from the region $U \leq U_{\text{lim}}$. Second, that sample is moved through configuration space by a random walk. This random walk must be long enough that the final configuration has no memory of its starting point. The random walk must also have the bounded uniform pdf (7.2) as its stationary distribution.

An example of one-dimensional random walk process that has the energy bounded uniform pdf (7.2) as its stationary distribution is a random walk of length L with the iterator described in Algorithm 7.2.

```
subroutine mcmc_step_1d          ! a single MCMC step
                                 ! The input configuration
                                 ! is q_in

step = 2*(ranf()-0.5)*step_size  ! random number in range
                                 ! [-step_size , +step_size]
trial = q_in + step              ! trial configuration
e_trial = U(trial)               ! U = potential energy
if (e_trial <= e_lim) then
  q_out = trial                  ! accept trial
else
  q_out = q_in                   ! reject trial
                                 ! take q_in: a null step
end if

end subroutine
```

Algorithm 7.2: A one-dimensional random walk iterator with the energy bounded uniform pdf (7.2) as its stationary distribution. `ranf()` is a random number uniform in [0,1].

We can derive Algorithm 7.2 by demanding that our integrator satisfy detailed balance, and asserting that its stationary distribution should be the energy bounded uniform distribution (7.2). Our first assertion, that the iterator should satisfy detailed balance between any two configurations o and n, can be written as

$$\mathcal{N}(o)\,\alpha(o \to n) \times \text{acc}(o \to n) = \mathcal{N}(n)\,\alpha(n \to o) \times \text{acc}(n \to o) \qquad (7.15)$$

We call Eq. (7.15) the detailed balance equation. Here $\alpha(x \to y)$ is the probability of proposing the trial configuration y when at the configuration x. The target stationary

distribution is $\mathcal{N}(x)$, which is zero if $U(x) > U_{\text{lim}}$ and equal to a constant otherwise. The final term acc $(x \to y)$ is the probability of accepting the move from x to y once we have proposed it. In Algorithm 7.2 we set the proposal distribution $\alpha(x \to y)$ to a symmetrical, uniform distribution over the range [-step_size, +step_size]. Therefore we have $\alpha(o \to n) = \alpha(n \to o)$. The configuration o is already inside the good region. Therefore, substituting the above into the detailed balance Eq. (7.15) and rearranging we have

$$\frac{\text{acc}(o \to n)}{\text{acc}(n \to o)} = \frac{\mathcal{N}(n)}{\mathcal{N}(o)} \tag{7.16}$$

An acceptance scheme that satisfies this is

$$\text{acc}(o \to n) = \begin{cases} 1, & U(n) \leq U_{\text{lim}} \\ 0, & \text{Elsewhere.} \end{cases} \tag{7.17}$$

This is exactly the scheme presented in Algorithm 7.2, and therefore that iterator has the required stationary distribution (7.2).

The complete MCMC process to generate a new sample in our 1D example is described in Algorithm 7.3.

```
subroutine gen_new_sample_1d     ! new 1D sample produced
                                 ! by cloning and MCMC

call copy_random_config          ! copies a random
                                 ! configuration from
                                 ! the allowed region

do i = 1, n_steps                ! walk length n_steps
  call mcmc_step_1d              ! calls algorithm 7.2
end do

end subroutine
```

Algorithm 7.3: The "clone and decorrelate" algorithm: generating a new sample in the region $U(x) \leq U_{\text{lim}}$ by first cloning, then decorrelating it using MCMC.

7.4 Multimodal Potentials

Suppose that the potential energy function has two modes, so that the region with $U(\mathbf{q}) \leq U_{\text{lim}}$ consists of two disconnected regions. The MCMC iterator described in Algorithm 7.2 cannot move between modes. As a result, if at any point we have zero samples in either mode, the "clone and decorrelate" Algorithm 7.3 will never find that mode again. Let the "gateway" energy at which the two modes become disconnected be U_g. Such a gateway energy contour is illustrated in Fig. 7.4.

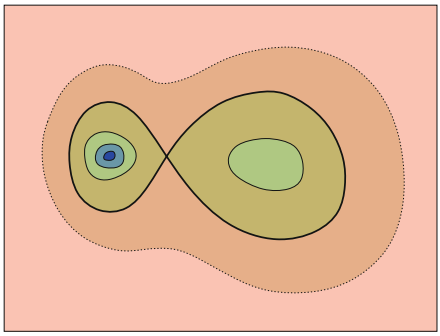

Fig. 7.4 Illustration of a bimodal potential energy function. The thick dark line corresponds to the "gateway" energy at which the two modes of the bounded uniform distribution (7.2) become disconnected. Once separated, no MCMC trajectory can move between the two modes. In this case it is possible for the (left) mode that contains the global minimum to become depopulated before the conclusion of the nested sampling calculation

At the gateway energy the configuration space "volumes" of the two modes are X_1 and X_2 where $X_1 + X_2 = \chi\left(U_g\right)$. Without loss of generality let $X_1 \leq X_2$. Let us also suppose that by bad luck, the smaller mode is the "dominant" mode, meaning it contains the global minimum of the potential energy $U(\mathbf{q})$. It is therefore of utmost importance that we have at least one sample inside the smaller, dominant mode. We are performing nested sampling using $\mathcal{K}$ configurations. For each configuration, the probability of being inside the dominant mode is

$$w = \frac{X_1}{X_1 + X_2}. \tag{7.18}$$

Therefore the probability of having one or more samples in the dominant mode is

$$\text{prob}\left(\text{success}|\mathcal{K}\right) = 1 - (1 - w)^{\mathcal{K}}. \tag{7.19}$$

We need $\mathcal{K} \geq \frac{1}{w}$ to have a reasonable chance of finding the global minimum.

Now consider the case where the bounded uniform distribution divides in this way, several times. These divisions are nested sufficiently that in each case the bounded uniform distribution may be thought as dividing from one mode into two. The probability of having at least one sample in the dominant mode after every one of these divisions is

$$\text{prob}\left(\text{success}|\mathcal{K}\right) = \prod_i \left[1 - (1 - w_i)^{\mathcal{K}}\right]. \tag{7.20}$$

If the smallest value of $\{w\}$ is $w_{\min}$, then we need roughly $\mathcal{K} \simeq \frac{1}{w_{\min}}$ configurations to achieve success: the smallest w dominates the rest.

Systems in which the dominant mode divides from the rest of the space at very high energy, with a small value of w constitute pathological problems for nested sam-

pling, which requires very large numbers of samples to achieve meaningful results under these circumstances. Some interesting work has been done on alleviating this problem [5, 6].

7.5 Error Propagation in Nested Sampling

After n nested sampling iterations, the volume of configuration space in the region $U \leq U_{\rm lim}$ has contracted from χ_0 to $\chi_0 \zeta_n$ where χ_0 is the total volume of the configuration space, and ζ_n is given by

$$\zeta_n = \prod_{i=1}^{n} t_i . \qquad \text{(7.13 revisited)}$$

The log of this is just

$$\log \zeta_n = \sum_{i=1}^{n} \log t_i . \qquad (7.21)$$

Making a change of variables (2.5) from the polynomial pdf for t (7.6), we find that $\log t$ is exponentially distributed

$$\varrho\,(\log t) = \mathcal{K} \mathrm{e}^{\mathcal{K} \log t} . \qquad (7.22)$$

This pdf (7.22) has mean and standard deviation

$$\log t = (-1 \pm 1) \,/\, \mathcal{K} . \qquad (7.23)$$

It follows that $\log \zeta_n$ has mean and standard deviation

$$\log \zeta_n = \left(-n \pm \sqrt{n}\right) / \mathcal{K} \qquad (7.24)$$

After $\mathcal{K}$ iterations, $\log \zeta_{\mathcal{K}}$ has mean value -1, with standard deviation $\frac{1}{\sqrt{\mathcal{K}}}$.

In (7.24) we have assumed for simplicity that t is a Gaussian distributed random variable, which it isn't. A better measure of the error introduced by the nested sampling procedure can be obtained by re-sampling $\{t_i\}$ from the pdf (7.6). If the quantity of interest is recomputed a dozen or so times using fresh $\{t_i\}$, then the mean value and standard deviation may be obtained. We call this error estimate the "internal" error for the nested sampling calculation. This error arises directly from the nested sampling procedure itself and is separate to errors produced by poor exploration of configuration space. Those "external" errors, due to improper decorrelation of samples, can result from using a walk length that is too short or not using enough live configurations $\mathcal{K}$ so that the global minimum is not discovered. In our experience, the total error is typically dominated by external errors.

7.6 Conclusion

The nested sampling procedure is an algorithm for calculating the inverse integrated density of states, $U(\chi)$. The central procedure of nested sampling has only two parameters: the walk length L and the number of samples to be used $\mathcal{K}$. Nested sampling produces a series of nested sample sets, where the energies of those samples are distributed according to the density of states. The rate of error accumulation in χ is set by the parameter $\mathcal{K}$, being proportional to $\frac{1}{\sqrt{\mathcal{K}}}$. The parameter L should be set large enough that the decorrelated sample is equally likely to have originated from each of the other samples.

Determining the location of a phase transition requires that we characterise the pdf for each phase, and calculate the relative probabilities of each phase, extremely accurately. At first order phase transitions, samplers that mimic physical processes, like the system they represent, spend long periods of time in one phase or the other, and transition between the two phases only rarely. We saw in Fig. 6.2 that such samplers lead to relatively large uncertainty at the peaks of the heat capacity. This is exactly the point that we must resolve when identifying the location of phase transitions.

In contrast, nested sampling explores a series of nested uniform distributions, constructed to have constant fractional overlap. As such, these distributions are computationally efficient to equilibrate between and the pathological behaviour of systems at first order phase transitions is avoided. This makes nested sampling an ideal tool for finding phase transitions.

Nested sampling's microcanonical cooling procedure ought to be a powerful tool for discovering "unknown unknown" phase transitions. In the rest of this thesis we develop nested sampling to perform this task for classical materials.

References

1. J. Skilling, Nested sampling. AIP Conf. Proc. **735**, 395 (2004)
2. J. Skilling, Nested sampling for general Bayesian computation. Bayesian Anal. **1**, 833 (2006)
3. L.B. Pártay, A.P. Bartók, G. Csányi, Efficient sampling of atomic configurational spaces. J. Phys. Chem. B **114**, 10502 (2010)
4. N.S. Burkoff, C. Várnai, S.A. Wells, D.L. Wild, Exploring the energy landscapes of protein folding simulations with Bayesian computation. Biophys. J. **102**, 878 (2012)
5. S. Martiniani, J.D. Stevenson, D.J. Wales, D. Frenkel, Superposition enhanced nested sampling. Phys. Rev. X **4**, 031034 (2014)
6. F. Feroz, M. Hobson, M. Bridges, MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. Mon. Not. R. Astron. Soc. **398**, 1601 (2009)

Chapter 8
Nested Sampling for Materials

8.1 Introduction

In Chap. 7 we identified nested sampling [1, 2] as a promising strategy for calculating the inverse integrated density of states $U(\chi)$ from first principles. In this chapter we develop nested sampling into a generic, automated tool for finding "unknown unknown" phase transitions.

We first apply nested sampling to Lennard-Jonesium, demonstrating that it gives the correct result for the pressure-temperature phase diagram. Second, we apply nested sampling to a binary Lennard-Jonesium "alloy". The alloy exhibits an order-disorder transition within the solid phase. Within the ordered phase we find two different structures, distinct in the arrangements of columns of A and B atoms within the crystal. A phase transition occurs between these two ordered phases at $\log_{10}(P^*) = -1.194$. Third, we apply nested sampling to the NPB embedded atom model potential for aluminium [3], and recover the crystal solid-solid phase transitions as predicted by this model.

8.2 Fixed Volume or Fixed Pressure?

In this section we address the question "in which ensemble should we perform nested sampling?" concluding that it is best to simulate conditions of constant pressure. MCMC simulations performed at fixed pressure require just a fraction of the computational expense as equivalent calculations performed at fixed volume. There are two reasons for this.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_8

First, allowing the system to change volume by dilating or contracting expedites the cooperative freeing of jammed atoms. In contrast, at fixed volume, atoms that have become jammed are only freed by the coincidental movement of all atoms to separate them. Consequently, MCMC simulations at fixed pressure explore configuration space far more rapidly than simulations at fixed volume.

The second reason arises from the thermodynamic behaviour of systems at a first-order phase transition. At a phase transition under constant volume conditions the two phases coexist and an interface forms between them. Such interfaces are large on the atomic scale [4], and the behaviour of atoms at an interface is not representative of the behaviour of atoms in the equilibrium phases. As a result the interface introduces a systematic error that is overcome only by simulating very large numbers of atoms.

Such interfaces also occur under constant pressure conditions in the infinite system size limit. The contribution to the Gibbs free energy from an interface is proportional to $\gamma N^{\frac{2}{3}}$, where γ is the interfacial tension. In contrast, the Gibbs free energies of each of the pure phases are extensive (proportional to N). Therefore the Gibbs free-energy cost per atom of the interface is negligible for thermodynamic systems. Conversely, for the relatively small system sizes amenable to density of states calculation methods such as nested sampling, the Gibbs free-energy cost per atom of the interface is appreciable, provided γ is not close to zero. Consequently, at a constant pressure phase transition between phases with identical atomic compositions, configurations containing an interface have negligible statistical weight in such simulations, and a discontinuous transition is observed from one equilibrium phase to the other. This enables the accurate simulation of phase transitions using much smaller numbers of atoms. We will therefore develop our schemes in the constant pressure ensemble.

8.3 Nested Sampling at Fixed Pressure

In this section we develop a method for performing nested sampling at fixed pressure. The result is then the integrated density of states at fixed pressure (4.65).

Our aim is to calculate the full phase diagram of the system, including all its solid phases. As described in Sect. 5.8, important phases can be completely excluded from simulations of a fixed number of particles in a cell of fixed shape. We use the fully flexible simulation cell formulation described in Sect. 5.8 to reduce this finite size effect.

Recall that the microscopic enthalpy, pdf and partition function with a fully flexible simulation cell are

$$H\left(\mathbf{s}, \mathbf{p}, P, V, \mathbf{h}_0\right) = E\left(V^{\frac{1}{3}}\mathbf{h}_0\mathbf{s}, \mathbf{p}\right) + PV \qquad \text{(5.18 revisited)}$$

$$\varrho\left(\mathbf{s},\mathbf{p},V,\mathbf{h}_0\right)=\frac{V^N}{N!h^{3N}V_0}\frac{\mathrm{e}^{-\beta H\left(\mathbf{s},\mathbf{p},P,V,\mathbf{h}_0\right)}}{\Delta\left(N,\beta,P\right)}\delta\left(\det\mathbf{h}_0-1\right) \tag{5.19 revisited}$$

$$\Delta\left(N,\beta,P\right)=\frac{1}{N!h^{3N}V_0}\int_0^\infty \mathrm{d}V\ V^N\int \mathrm{d}\mathbf{h}_0\,\delta\left(\det\mathbf{h}_0-1\right)\int \mathrm{d}^N\mathbf{s}\,\mathrm{d}^N\mathbf{p}\,\mathrm{e}^{-\beta H\left(\mathbf{s},\mathbf{p},P,V,\mathbf{h}_0\right)} \tag{5.20 revisited}$$

Performing the momentum integration explicitly and assuming a quadratic form for the kinetic energy and equal masses of all particles in the system we obtain the configurational enthalpy, marginal pdf and a new form for the partition function

$$Y\left(\mathbf{s},P,V,\mathbf{h}_0\right)=U\left(V^{\frac{1}{3}}\mathbf{h}_0\mathbf{s}\right)+PV \tag{8.1}$$

$$\varrho\left(\mathbf{s},V,\mathbf{h}_0\right)=\frac{V^N}{N!h^{3N}V_0}\left(\frac{2\,\pi\,m}{\beta}\right)^{\frac{3N}{2}}\frac{\mathrm{e}^{-\beta Y\left(\mathbf{s},P,V,\mathbf{h}_0\right)}}{\Delta\left(N,\beta,P\right)}\delta\left(\det\mathbf{h}_0-1\right) \tag{8.2}$$

$$\Delta\left(N,\beta,P\right)=\frac{1}{N!h^{3N}V_0}\left(\frac{2\,\pi\,m}{\beta}\right)^{\frac{3N}{2}}\int_0^\infty \mathrm{d}V\ V^N\int \mathrm{d}\mathbf{h}_0\int \mathrm{d}^N\mathbf{s}\,\mathrm{e}^{-\beta Y\left(\mathbf{s},P,V,\mathbf{h}_0\right)}\delta\left(\det\mathbf{h}_0-1\right). \tag{8.3}$$

The configurational density of states $g\left(Y',N,P\right)$ and integrated density of states $\chi\left(Y,N,P\right)$ are given by

$$g\left(Y',N,P\right)=\frac{1}{N!h^{3N}V_0}\int_0^\infty \mathrm{d}V\ V^N\int \mathrm{d}\mathbf{h}_0\,\delta\left(\det\mathbf{h}_0-1\right)\int \mathrm{d}^N\mathbf{s}\,\delta\left(Y\left(\mathbf{s},\mathbf{p},P,V,\mathbf{h}_0\right)-Y'\right) \tag{8.4}$$

$$\chi\left(Y,N,P\right)=\int_0^Y g\left(Y',N,P\right)\mathrm{d}Y' \tag{8.5}$$

Rewriting the partition function proceeds exactly as in Sect. 4.3.1 except that we must keep in mind that the definition of Y and χ are now (8.1) and (8.5) respectively. Thus we have

$$\Delta\left(N,P,\beta\right)=\frac{1}{N!h^{3N}V_0}\left(\frac{2\,\pi\,m}{\beta}\right)^{\frac{3N}{2}}\int_{-\infty}^\infty \mathrm{d}Y\ g\left(Y\right)\mathrm{e}^{-\beta Y} \tag{4.71 revisited}$$

$$\Delta(N, P, \beta) = \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \int_0^\infty \mathrm{d}\chi \, \mathrm{e}^{-\beta Y(\chi)}. \qquad \text{(4.75 revisited)}$$

As described in Chap. 7, the aim of nested sampling is to calculate the function $Y(\chi)$. The approach is to set a geometric series of values for χ (7.1) which we obtain in a probabilistic way with some error.

$$\chi \in \left\{\chi_0 t, \chi_0 t^2, \ldots, \chi_0 t^{N_{\text{its}}}\right\}. \qquad \text{(7.1 revisited)}$$

We calculate the exact values $Y(\chi_i)$ for the estimates of χ_i we have obtained. Convergence of the algorithm is possible by reducing the error on $\{\chi_i\}$. For fixed volume conditions with a fixed shape simulation cell, at each iteration we generate a fresh sample with $U \leq U(\chi_{i-1})$, distributed according to the density of states

$$\varrho(U|U(\chi_{i-1})) = \begin{cases} \frac{g(U)}{\chi_{i-1}}, & U \leq U(\chi_{i-1}) \\ 0, & \text{Elsewhere.} \end{cases} \qquad \text{(7.3 revisited)}$$

This is achieved by producing samples from configuration space with a uniform pdf throughout the region $U(\mathbf{q}) \leq U_{\text{lim}}$:

$$\varrho(\mathbf{q}|U(\chi_{i-1})) = \begin{cases} \frac{1}{\chi(U(\chi_{i-1}))}, & U(\mathbf{q}) \leq U(\chi_{i-1}) \\ 0, & \text{Elsewhere.} \end{cases} \qquad \text{(7.2 revisited)}$$

In the same way, for the fixed pressure ensemble with fully flexible simulation cell, we also generate a fresh sample at each iteration of the nested sampling procedure with $Y \leq Y(\chi_{i-1})$, distributed according to the density of states

$$\varrho(Y|Y(\chi_{i-1})) = \begin{cases} \frac{g(Y)}{\chi_{i-1}}, & Y \leq Y(\chi_{i-1}) \\ 0, & \text{Elsewhere.} \end{cases} \qquad (8.6)$$

This is achieved by generating samples according to the following pdf

$$\varrho\left(\mathbf{s}, \mathbf{h}_0, V|Y\left(\chi_{i-1}\right), P\right) = \begin{cases} \frac{V^N \delta(\det \mathbf{h}_0 - 1)}{\chi_{i-1}}, & \mathbf{s} \in [0,1)^{3N},\ Y(\mathbf{s}, \mathbf{h}_0, V|P) \leq Y\left(\chi_{i-1}\right) \\ 0, & \text{Elsewhere.} \end{cases} \qquad (8.7)$$

As in the fixed volume ensemble with a fixed shape simulation cell, we approximate the partition function (4.75) by

$$\Delta\left(N, P, \beta\right) \simeq \frac{1}{N! h^{3N} V_0} \left(\frac{2\,\pi\, m}{\beta}\right)^{\frac{3N}{2}} \sum_{i=1}^{N_{\text{its}}} \Delta\chi_i\, \mathrm{e}^{-\beta Y_i} \tag{8.8}$$

$$\Delta\chi_i = \chi_0\left(\zeta_{i-1} - \zeta_i\right) \tag{7.12 revisited}$$

Here, ζ_i is the factor by which the integrated density of states $\chi\left(Y_{\text{lim}}\right)$ is reduced from χ_0 at the end of iteration i.

$$\zeta_n = \prod_{i=1}^{n} t_i \tag{7.13 revisited}$$

$$\simeq \left(\frac{\mathcal{K}}{\mathcal{K}+1}\right)^n. \tag{7.14 revisited}$$

In Eq. (7.14) we have assumed zero error in integrated density of states $\chi\left(Y_{\text{lim}}\right)$ at every iteration. Errors in the integrated density of states were discussed in Sect. 7.5.

8.4 Minimum Cell Height Criterion

We now consider finite system size effects introduced by the fully flexible simulation cell formulation. The fully flexible cell formulation [5] described in Sect. 5.8 was introduced to remove the finite size effect whereby it may not be possible to arrange a fixed number of particles in a fixed shape cell into certain crystal structures. In unfortunate cases this can exclude thermodynamically relevant structures from the results of our calculation.

In this section we establish the following:

1. For a finite number of particles, there exist simulation cells **h** such that parallel faces of the cell are separated by only a few layers of atoms.
2. Moreover, such structures are the statistically most probable state for solids.
3. The resulting system of reduced dimension does not approximate the infinite fluid in three dimensions.
4. This problem can be solved by the introduction of a "minimum cell height" parameter.

For a Finite Number of Particles, There Exist Simulation Cells $\mathbf{h}$ *Such that Parallel Faces of the Cell are Separated by Only a Few Layers of Atoms*

This statement is very simple to prove. In the fully flexible cell formulation, the unit volume cell matrix $\mathbf{h}_0$ is allowed to become arbitrarily thin. Therefore, for any finite number of atoms, there exist cells $\mathbf{h}$ with parallel faces separated by as few as one layer of atoms.

Moreover, Such Structures are the Statistically Most Probable State for Solids

The overwhelming majority of cell shapes $\mathbf{h}_0$ are extremely thin, and we are able to show this very generally.

We begin by introducing a quantity called the "cell height", `ch`. The perpendicular distance between faces of the unit cell made by lattice vectors $\mathbf{h}^{(i)}$ and $\mathbf{h}^{(j)}$ is given by

$$d^{\perp}_{\mathbf{h}^{(k)}} = \frac{|\det \mathbf{h}|}{\left|\mathbf{h}^{(i)} \times \mathbf{h}^{(j)}\right|}. \tag{8.9}$$

The cell height is defined as the minimum value of $d^{\perp}_{\mathbf{h}^{(k)}}$, for the cell at normalised (unit) volume $\mathbf{h}_0$.

$$\mathtt{ch} = \min_{i=1,2,3} \left(d^{\perp}_{\mathbf{h}_0^{(i)}} \right) \tag{8.10}$$

The cell height measures how thin the cell has become. The integrated density of states for `ch` was calculated using nested sampling, and is shown in Fig. 8.1. Relative to the integrated density of states at $\mathtt{ch} = 0.4$, we find $\chi_{0.5} = 4.16 \times 10^{-3}$, $\chi_{0.7} = 2.79 \times 10^{-5}$, and $\chi_{0.9} = 1.77 \times 10^{-7}$. Evidently, the overwhelming majority of simulation cells are thin in at least one dimension.

The Resulting System of Reduced Dimension Does not Approximate the Infinite Fluid in Three Dimensions

Figure 8.2 shows the heat capacity of Lennard-Jonesium as calculated with a simulation cell containing 64 atoms at pressure $\log_{10} P^* = -1.194$. In each calculation we constrained the cell height to be greater than some minimum value, `mch`. The transition to a quasi-2D system as `mch` is reduced is clear. The location of the melting

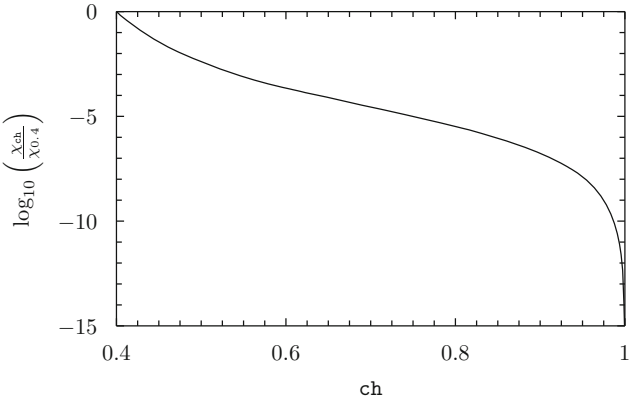

Fig. 8.1 Integrated density of states for cell height, shown for $\mathtt{ch} > 0.4$. This graph illustrates that there are many more simulation cells with small values of ch (8.10) than large values (close to 1). The cell height is the minimum distance between opposite faces of the unit volume (normalised) simulation cell. A small value indicates that the simulation cell is thin in at least one dimension. As fractions of the integrated density of states at $\mathtt{ch} = 0.4$, we find $\chi_{0.5} = 4.16 \times 10^{-3}$, $\chi_{0.7} = 2.79 \times 10^{-5}$, and $\chi_{0.9} = 1.77 \times 10^{-7}$. We conclude that the overwhelming majority of simulation cells are thin in at least one dimension

transition becomes independent of mch for $\mathtt{mch} \geq 0.65$, and the location of the evaporation transition, for $\mathtt{mch} \geq 0.35$. This is most likely because the atomic potential used a fixed radial cutoff ($r_c = 3\sigma$, see Sect. 8.8) and for any particular $\mathbf{h}_0$, at lower densities there will be fewer periodic self images of each atom within the cutoff. At low values of mch the system's behaviour is dominated by the fictitious periodicity imposed by the boundary conditions.

This Problem can be Solved by the Introduction of a "Minimum Cell Height" Parameter

It is clear from Fig. 8.2 that by imposing a suitable minimum cell height we can remove the unphysical behaviour from the full flexible cell formulation.

This condition,

$$\mathtt{ch}\,(\mathbf{h}_0) > \mathtt{mch} \tag{8.11}$$

leads us to write a new density of states

$$g\left(Y', N, P\right) = \frac{1}{N!h^{3N}V_0}\int_0^\infty \mathrm{d}V\; V^N \int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}} \mathrm{d}\mathbf{h}_0\,\delta\left(\det\mathbf{h}_0 - 1\right)\times \int \mathrm{d}^N\mathbf{s}\,\delta\left(Y\left(\mathbf{s}, \mathbf{p}, P, V, \mathbf{h}_0\right) - Y'\right). \tag{8.12}$$

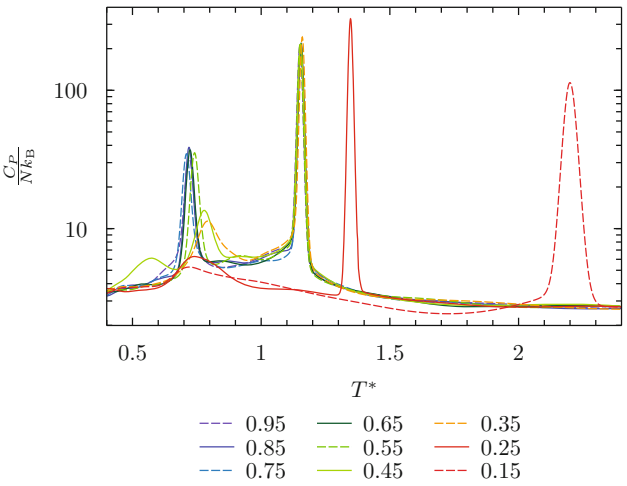

Fig. 8.2 Convergence of heat capacity with respect to minimum cell height, `mch`. The legend on the right shows the value of `mch` used in each simulation. A higher value of `mch` restricts the system to more cube-like cell shapes, while a low value allows the system to become essentially flat. Curves show the heat capacity of a periodic system of 64 Lennard-Jonesium particles, where the cell height (8.10) is constrained to be greater than `mch`. The system was simulated at a reduced pressure $\log_{10} P^* = -1.194$ using nested sampling, as in this section. The peak at lower temperature corresponds to melting, and the peak at higher temperature to evaporation. The transition to an effective 2D system can be seen clearly as `mch` is reduced. The location of the evaporation transition is converged for $\texttt{mch} \geq 0.35$, but melting requires a higher $\texttt{mch} \geq 0.65$. This makes intuitive sense, since the atomic potential used a fixed radial cutoff ($r_c = 3\sigma$, see Sect. 8.8) and for any particular $\mathbf{h}_0$, at lower densities there will be fewer periodic self images of each atom within the cutoff

As always in nested sampling, we seek to produce $\mathcal{K}$ samples with configurational enthalpies distributed according to (8.12). This can be achieved by sampling from configuration space according to

$$\varrho\left(\mathbf{s}, \mathbf{h}_0, V | Y\left(\chi_{i-1}\right), P\right) = \begin{cases} \frac{V^N \delta(\det \mathbf{h}_0 - 1)}{\chi_{i-1}}, & \mathbf{s} \in [0,1)^{3N} \text{ AND} \\ & \texttt{ch}\left(\mathbf{h}_0\right) > \texttt{mch} \text{ AND} \\ & Y\left(\mathbf{s}, \mathbf{h}_0, V | P\right) \leq Y\left(\chi_{i-1}\right), \\ 0, & \text{Elsewhere.} \end{cases} \tag{8.13}$$

The density of states (8.12) replaces (8.4) in the partition function (4.71).

We stated in Sect. 5.8 where the fully flexible cell formulation was introduced, that it is not in general valid to impose an arbitrary pdf over $\mathbf{h}_0$. It can be seen from Tuckerman's derivation of the pressure virial theorem, that this is not the case: [6], p. 230. We are not proposing to do that here. Rather, we have changed the limits on the integration for $\mathbf{h}_0$ in the partition function (8.13). The interested reader will see from Tuckerman's derivation that this leaves the pressure tensor unchanged.

8.5 MCMC Exploration in the Fully Flexible Cell Formulation at Fixed Pressure

In Sects. 8.3 and 8.4 we saw that for nested sampling at fixed pressure, at each iteration we require a new sample distributed as the pdf (8.13).

$$\varrho\left(\mathbf{s}, \mathbf{h}_0, V | Y\left(\chi_{i-1}\right), P\right) = \begin{cases} \frac{V^N \delta(\det \mathbf{h}_0 - 1)}{\chi_{i-1}}, & \mathbf{s} \in [0,1)^{3N} \text{ AND} \\ & \mathtt{ch}\left(\mathbf{h}_0\right) > \mathtt{mch} \text{ AND} \\ & Y\left(\mathbf{s}, \mathbf{h}_0, V | P\right) \leq Y\left(\chi_{i-1}\right), \\ 0, & \text{Elsewhere.} \end{cases} \tag{8.13 revisited}$$

This is achieved by the "clone and decorrelate" Algorithm 7.3. The MCMC walk is a random series of single atom steps, volume steps, lattice shear steps and lattice stretch steps. We describe these Monte Carlo (MC) steps below.

```
subroutine single_atom_mc
                              ! a single atom MC step
                              ! The input configuration
                              ! is s_in

at = ranf()*N + 1             ! random atom chosen from
                              ! {1, 2, ... , N}
dx = 2.0*step_l_s*(ranf() - 0.5)
dy = 2.0*step_l_s*(ranf() - 0.5)
dz = 2.0*step_l_s*(ranf() - 0.5)
                              ! dx, dy, dz each chosen
                              ! uniformly from
                              ! [-step_l_s, +step_l_s]

s_trial = s_in                ! trial s coordinates
s_trial(x, at) = s_trial(x, at) + dx
s_trial(y, at) = s_trial(y, at) + dy
s_trial(z, at) = s_trial(z, at) + dz

e_trial = Y(s_trial)          ! Y = configurational enthalpy
if (e_trial <= e_lim) then
  s_out = s_trial             ! accept trial
else
  s_out = s_in                ! trial rejected
                              ! s unchanged: a null step
end if

end subroutine
```

Algorithm 8.1 Single atom MC step in fractional coordinates **s**. Samples from the enthalpy bounded uniform pdf (8.13). `ranf()` is a random number uniform in [0,1].

Single atom steps. We attempt to displace one atom at a time, rather than all atoms at once. It is well established that this leads to more efficient exploration of configuration space [4]. The single atom MC step is shown in Algorithm 8.1.

Volume steps. Volumes have a weight V^N in our target pdf (8.13). Volumes are sampled according to this polynomial distribution by Algorithm 8.2, which implements a symmetric proposal distribution, $\alpha(o \to n) = \alpha(n \to o)$, together with the MC acceptance probability

$$\text{acc}(o \to n) = \min\left[1, \left(\frac{V_n}{V_o}\right)^N\right]. \tag{8.14}$$

We can see that applying this gives us the correct stationary distribution by comparing it with the detailed balance Eq. (7.15).

$$\mathcal{N}(o)\,\alpha(o \to n) \times \text{acc}(o \to n) = \mathcal{N}(n)\,\alpha(n \to o) \times \text{acc}(n \to o) \tag{7.15 revisited}$$

Since we are using a symmetric proposal distribution, $\alpha(o \to n) = \alpha(n \to o)$, (7.15) implies

$$\frac{\text{acc}(o \to n)}{\text{acc}(n \to o)} = \frac{\mathcal{N}(n)}{\mathcal{N}(o)}. \tag{7.16 revisited}$$

Therefore the MC acceptance probability (8.14) leads to

$$\mathcal{N}(V) \propto V^N. \tag{8.15}$$

For reasons explained in Sect. 8.6 we also impose a maximum value for the volume, $V < V_\text{u}$. That upper bound, V_u is chosen to easily contain a portion of the ideal gas phase.

```
subroutine volume_mc
                              ! volume MC step
                              ! The input volume is V_in

dV = 2.0*step_l_V*(ranf() - 0.5)
                              ! dV chosen uniformly from
                              ! [-step_l_V, +step_l_V]

V_trial = V_in + dV           ! trial V

if (V_trial > V_u) then       ! impose V < V_u
  V_out = V_in                ! trial rejected
  exit                        ! V unchanged: a null step
end if

p_accept = min(1.0, (V_trial/V_in)**N)
                              ! probability of accepting this
                              ! trial volume

if (ranf() < p_accept) then
  e_trial = Y(V_trial)        ! Y = configurational enthalpy
  if (e_trial <= e_lim) then
    V_out = V_trial           ! accept trial
  else
    V_out = V_in              ! trial rejected
                              ! V unchanged: a null step
  end if
else
  V_out = V_in                ! V unchanged: a null step
end if

end subroutine
```

Algorithm 8.2 Volume MC step. Samples generated according to the enthalpy bounded polynomial pdf (8.13). `ranf()` is a random number uniform in [0,1].

Lattice shape moves. We explore the surface $\det \mathbf{h}_0 = 1$, subject to the minimum cell height criterion discussed in Sect. 8.4, $\mathtt{ch}(\mathbf{h}_0) > \mathtt{mch}$. This minimum cell height criterion excludes quasi one and two-dimensional systems that would otherwise be an unavoidable artefact of simulating a finite size system using a fully flexible simulation cell. Our MC strategy employs both shear and stretch moves.

Shear moves. A simulation cell with side vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ was shown in Fig. 5.3. The volume of the cell is given by $|\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}|$: the area of the face in the $\mathbf{b} - \mathbf{c}$ plane, multiplied by the height of the cell perpendicular to that face. Adding vectors to $\mathbf{a}$ that lie in the $\mathbf{b} - \mathbf{c}$ plane does not change the height of the cell perpendicular to the $\mathbf{b} - \mathbf{c}$ face. Therefore, we may apply such shear deformations to the unit cell without changing its volume. Note that there is nothing special about the $\mathbf{a}$ vector, $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \mathbf{b} \cdot \mathbf{c} \times \mathbf{a} = \mathbf{c} \cdot \mathbf{a} \times \mathbf{b}$, so we can apply similar shearing MC moves to any of the lattice vectors.

Like the single atom MC steps, this MC scheme samples uniformly below the configurational enthalpy limit. We also enforce the minimum cell height condition introduced in Sect. 8.4.

The resulting MC strategy is best expressed in algorithmic form, which is given in Algorithm 8.3.

```
subroutine lattice_shear_mc
! lattice shear MC step
! The input unit cell matrix is h0_in(3,3)

h0_trial = h0_in                ! copy h0_in into trial array

do repeat = 1, 3                ! repeat three times

  h0_save = h0_trial            ! save h0 in case of rejection

  chng_vec = ranf()*3 + 1       ! randomly selected lattice
                                ! vector. integer: 1, 2, or 3

  basis = orthonormal_basis(h0_in,chng_vec)
  ! basis(3,2) is a pair of orthonormal vectors in the plane
  ! defined by the two edge vectors in h0_in other than chng_vec

  pair: do
    x = 2.0*(ranf() - 0.5)      ! generate (x,y) pair from
    y = 2.0*(ranf() - 0.5)      ! inside unit circle
    if (x**2 + y**2 < 1.0) exit pair
  end do pair
  x = x*step_l_sh               ! x,y updated to uniform pair in
  y = y*step_l_sh               ! a circle of radius step_l_sh

  h0_trial(:,chng_vec) = h0_trial(:,chng_vec)
                       + x*basis(:,1) + y*basis(:,2)
  ! change to cell vector chng_vec is circularly symmetric
  ! and in the plane defined by the other two edge vectors

  if ( ch(h0_trial) < mch ) h0_trial = h0_save
  ! trial rejected due to minimum cell height criterion.
  ! minimum cell height of trial cell is smaller than
  ! allowed value: ch(h0) < mch

end do
```

```
H_trial = H(h0_trial)      ! H = configurational enthalpy
if (H_trial <= H_lim) then
  h0_out = h0_trial        ! accept trial
else
  h0_out = h0_in           ! trial rejected
                           ! h0 unchanged: a null step
end if

end subroutine
```

Algorithm 8.3 Lattice shear MC step. Samples the surface $\det \mathbf{h}_0 = 1$ according to the enthalpy and cell height bounded uniform pdf (8.13). `ranf()` is a random number uniform in [0,1].

Stretch moves. In the stretch MC move we first generate a uniform random number u from the range [−step_l_st, step_l_st]. Second, we pick two cell vectors at random, and then stretch one by a factor e^{u}, and the other by e^{-u}. The proposal is reversible and we use the following acceptance probability

$$\text{acc}\,(o \to n) = \begin{cases} 1, & \texttt{ch}\,(\mathbf{h}_0) > \texttt{mch AND } Y\,(\mathbf{h}_0|P, V, \mathbf{s}) \le Y\,(\chi_{i-1}) \\ 0, & \text{Elsewhere.} \end{cases} \tag{8.16}$$

For clarity, the complete lattice stretch MC step is given in Algorithm 8.4.

```
subroutine lattice_stretch_mc
! lattice stretch MC step
! The input unit cell matrix is h0_in(3,3)

h0_trial = h0_in              ! copy h0_in into trial array

do repeat = 1, 3              ! repeat three times

  h0_save = h0_trial          ! save h0 in case of rejection

  vec1 = ranf()*3 + 1         ! (vec1,vec2) is a random
  vec2 = ranf()*2 + 1         ! pair of different lattice
  if (vec1 == vec2) then      ! vectors
    vec2 = 3
  end if

  ! u is a uniform random variable from
  ! [-step_l_st,step_l_st]
  u = 2.0*(ranf() - 0.5)*step_l_st

  ! multiply edge vectors vec1, vec2 by exp(u), exp(-u)
  h0_trial = h0_in
  h0_trial(:,vec1) = h0_trial(:,vec1)*exp(u)
  h0_trial(:,vec2) = h0_trial(:,vec1)*exp(-u)
```

```
  if ( ch(h0_trial) < mch ) h0_trial = h0_save
  ! trial rejected due to minimum cell height criterion.
  ! minimum cell height of proposed cell is smaller than
  ! allowed value: ch(h0) < mch

end do

H_trial = H(h0_trial)      ! H = configurational enthalpy
if (H_trial <= H_lim) then
  h0_out = h0_trial        ! accept trial
else
  h0_out = h0_in           ! trial rejected
                           ! h0 unchanged: a null step
end if

end subroutine
```

Algorithm 8.4 Lattice stretch MC step. This algorithm samples the surface $\det \mathbf{h}_0 = 1$ according to the enthalpy and cell height bounded uniform pdf (8.13). `ranf()` is a random number uniform in [0,1].

8.6 Initialising Nested Sampling at Fixed Pressure

The configuration space for particles in a simulation cell of variable volume, where the volume is allowed to vary without limit, is not a compact space. To generate the initial live samples for nested sampling we require a compact configuration space, and therefore we must impose a maximum allowed volume on our simulation cell. A sensible way to do this is to place the boundary well inside the region corresponding to the ideal gas. Having chosen a contour well inside the ideal gas, we can safely neglect the contribution made to statistical quantities by the excluded ideal gas configurations, provided $k_\mathrm{B}T \ll PV_u$. This is described in Sect. 8.7.

From the very start of a nested sampling calculation we require samples distributed as (8.13) with $Y_\mathrm{lim} = \infty$, and with the added constraint $V < V_\mathrm{u}$. The volume bound V_u is chosen to be so large that it is, effectively, independent of the system, comfortably within the ideal gas phase. An example value is $V_\mathrm{u} = N \times 10^7 \text{Å}^3$. The time for this part of the calculation scales as $\log\left(\frac{V_\mathrm{u}}{V_\mathrm{t}}\right)$ where V_t corresponds to the highest volume at which the gas undergoes a phase transition. The fractional coordinates $\mathbf{s}$ should be initialised as univariate random numbers in the interval $[0, 1)$. Lattice matrices $\mathbf{h}_0$ are generated by applying Algorithms 8.3 and 8.4 in a random sequence, starting from a cube. Step lengths of 1 can be used for both these algorithms at this point. Initial volumes may be generated by repeat application of Algorithm 8.2, starting from any allowed volume. It has been found that a good choice of starting volume is the mean of the distribution $\varrho(V) \propto V^N$: $\frac{N+1}{N+2}V_\mathrm{u}$. A good choice of step length for volume initialisation has been found to be $\frac{V_\mathrm{u}}{N}$.

8.7 Partition Function and Thermodynamic Variables

The partition function we seek to calculate is

$$\widetilde{\Delta}\,(N,\,\beta,\,P,\,\mathtt{mch}) = \frac{1}{N!h^{3N}V_0\int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}}\mathrm{d}\mathbf{h}_0\delta\,(\det\mathbf{h}_0-1)}\left(\frac{2\,\pi\,m}{\beta}\right)^{\frac{3N}{2}}\times \int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}}\mathrm{d}\mathbf{h}_0\delta\,(\det\mathbf{h}_0-1)\int_0^\infty \mathrm{d}V\;V^N\int_{(0,1)^{3N}}\mathrm{d}\mathbf{s}\;\mathrm{e}^{-\beta Y(\mathbf{s},P,V,\mathbf{h}_0)}. \tag{8.17}$$

The factor of $\frac{1}{\int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}}\mathrm{d}\mathbf{h}_0\delta(\det\mathbf{h}_0-1)}$ ensures that the partition function is independent of mch in the ideal gas limit. Above some sufficiently large volume V_u we approximate the system as an ideal gas, neglecting interatomic interactions. This corresponds to the condition $U\,(\mathbf{s},\,\mathbf{h}_0,\,V)\ll PV$. In this approximation the volume integral in (8.17) is the sum of two parts

$$\widetilde{\Delta}(N,\,P,\,\beta,\mathtt{mch})\approx\frac{1}{N!h^{3N}V_0}\left(\frac{2\,\pi\,m}{\beta}\right)^{\frac{3N}{2}}\left[\frac{\Delta_{\mathrm{NS}}(N,\,P,\,\beta,\,V_u,\mathtt{mch})}{\int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}}\mathrm{d}\mathbf{h}_0\delta\,(\det\mathbf{h}_0-1)}+\int_{V_u}^\infty \mathrm{d}V\,V^N e^{-\beta PV}\int_{(0,1)^{3N}}\mathrm{d}\mathbf{s}\right] \tag{8.18}$$

where

$$\Delta_{\mathrm{NS}}(N,\,P,\,\beta,\,V_u,\mathtt{mch}) = \int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}}\mathrm{d}\mathbf{h}_0\delta\,(\det\mathbf{h}_0-1)\times \int_0^{V_u}\mathrm{d}V\,V^N\int_{(0,1)^{3N}}\mathrm{d}\mathbf{s}\,e^{-\beta[U(\mathbf{s},\mathbf{h}_0,V)+PV]}. \tag{8.19}$$

We calculate Δ_{NS} using the nested sampling method described in this chapter. The nested sampling approximation for Δ_{NS} is

$$\Delta_{\mathrm{NS}}(N,\,P,\,\beta,\,V_u,\mathtt{mch})\approx\sum_{i=1}^{N_{\mathrm{its}}}(\widetilde{\chi}_{i-1}-\widetilde{\chi}_i)\;\mathrm{e}^{-\beta Y_i} \tag{8.20}$$

$$\approx\sum_{i=1}^{N_{\mathrm{its}}}\Delta\widetilde{\chi}_i\;\mathrm{e}^{-\beta Y_i} \tag{8.21}$$

where $\widetilde{\chi}_i\approx\widetilde{\chi}_0\left(\frac{\mathcal{K}}{\mathcal{K}+1}\right)^i$, $\widetilde{\chi}_0=\frac{V_u^{N+1}}{N+1}\int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}}\mathrm{d}\mathbf{h}_0\delta\,(\det\mathbf{h}_0-1)$, and $\Delta\widetilde{\chi}_i\approx\widetilde{\chi}_{i-1}-\widetilde{\chi}_i$.

Let us now consider the second integral in Eq. (8.18), which corresponds to the ideal gas. For the ideal gas $\int_{(0,1)^{3N}}\mathrm{d}\mathbf{s}=1$ and so we have

$$\begin{aligned}\int_{V_u}^{\infty} \mathrm{d}V V^N e^{-\beta PV} \int_{(0,1)^{3N}} \mathrm{d}\mathbf{s} &= \int_{V_u}^{\infty} \mathrm{d}V V^N e^{-\beta PV} \\ &= \frac{1}{(\beta P)^{N+1}} \Gamma(N+1, \beta P V_u),\end{aligned} \tag{8.22}$$

where $\Gamma(N+1, \beta p V_u)$ is the upper incomplete Gamma function. This form is closely related to the partition function for hard spheres [7, 8].

We observe that $\Gamma(N+1, \beta P V_u) \to 0$ in the limit $k_\mathrm{B}T/PV_u \to 0$. Therefore, provided $k_\mathrm{B}T \ll PV_u$, we have[1]

$$\widetilde{\Delta}(N, P, \beta, \mathtt{mch}) \approx \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \frac{\Delta_{\mathrm{NS}}(N, P, \beta, V_u, \mathtt{mch})}{\int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}} \mathrm{d}\mathbf{h}_0 \delta(\det \mathbf{h}_0 - 1)}. \tag{8.23}$$

Note that $\int_{\mathtt{ch}(\mathbf{h}_0)>\mathtt{mch}} \mathrm{d}\mathbf{h}_0 \delta(\det \mathbf{h}_0 - 1)$ cancels between the denominator of (8.23) and $\widetilde{\chi}_0$ in the approximation to Δ_{NS} (8.21). For convenience, we write (8.23) as

$$\widetilde{\Delta}(N, P, \beta, \mathtt{mch}) \approx \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \sum_{i=1}^{N_{\mathrm{its}}} \Delta\chi_i \, \mathrm{e}^{-\beta Y_i} \tag{8.24}$$

where $\chi_i \approx \chi_0 \left(\frac{\mathcal{K}}{\mathcal{K}+1}\right)^i$, $\chi_0 = \frac{V_u^{N+1}}{N+1}$, and $\Delta\chi_i \approx \chi_{i-1} - \chi_i$.

One can always assert the condition $k_\mathrm{B}T \ll PV_u$, and in practice it is easy to find values of V_u suitable for physically relevant conditions. As stated in Sect. 8.6, we found $V_0 = 10^7 N \text{Å}^3$ to be suitable for all conditions considered in this thesis.

From (8.24) we obtain the expected enthalpy

$$\begin{aligned}\langle H \rangle &= -\frac{\partial \ln \Delta(N, P, \beta)}{\partial \beta} \\ &= \left(\frac{3N}{2}\right) \frac{1}{\beta} + \langle Y \rangle\end{aligned} \tag{8.25}$$

and the heat capacity at constant pressure

$$C_P = -k_\mathrm{B}\beta^2 \frac{\partial \langle H \rangle}{\partial \beta} \tag{8.26}$$

$$= \frac{3Nk_\mathrm{B}}{2} + k_\mathrm{B}\beta^2 \left(\langle Y^2 \rangle - \langle Y \rangle^2\right) \tag{8.27}$$

where

[1] Although V_u appears in Eq. (8.23) as an argument of Δ_{NS} but not of $\widetilde{\Delta}$, there is no inconsistency because (8.23) is only valid in the limit $k_\mathrm{B}T/PV_u \to 0$ where the value of Δ_{NS} is independent of V_u.

$$\begin{aligned}\langle Y\rangle &= \frac{\sum_{i=1}^{N_{\text{its}}} \Delta\chi_i\, Y_i\, \mathrm{e}^{-\beta Y_i}}{\sum_{i=1}^{N_{\text{its}}} \Delta\chi_i\, \mathrm{e}^{-\beta Y_i}},\\ \langle Y^2\rangle &= \frac{\sum_{i=1}^{N_{\text{its}}} \Delta\chi_i\, Y_i^2\, \mathrm{e}^{-\beta Y_i}}{\sum_{i=1}^{N_{\text{its}}} \Delta\chi_i\, \mathrm{e}^{-\beta Y_i}}.\end{aligned} \tag{8.28}$$

This form (8.28) naturally does not depend on the contribution made by the low density configurations omitted from the NS calculation, or explicitly on the value of `mch`. We used Eq. (8.28) when calculating the heat capacities presented in this thesis.

8.8 Lennard-Jonesium

To validate the method described in this chapter, we applied it to the Lennard-Jonesium fluid. The Lennard-Jonesium potential is given by

$$\begin{aligned}U\left(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N\right) &= u_{\text{c}} + \sum_{i\geq j} u_{ij}\\ u_{ij} &= \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], & r_{ij} \leq r_{\text{c}}\\ 0, & \text{Elsewhere}\end{cases}\\ u_{\text{c}} &= \frac{8\,\pi\,\epsilon\sigma^3}{3}\frac{N^2}{V}\left[\frac{1}{3}\left(\frac{\sigma}{r_{\text{c}}}\right)^9 - \left(\frac{\sigma}{r_{\text{c}}}\right)^3\right]\end{aligned} \tag{8.29}$$

The term u_{c} corresponds to a mean field approximation, which is applied beyond the radial cutoff r_{c}. That is, we assume that particle mass is distributed uniformly beyond the radial cutoff. Throughout this section we use the standard definitions for reduced pressure and temperature: $P^* = \frac{P\sigma^3}{\epsilon}$, $T^* = \frac{k_{\text{B}}T}{\epsilon}$.

8.8.1 Simulation Details

The calculation was performed using 64 atoms, with $r_{\text{c}} = 3\sigma$, and a mean-field long-range correction, as in Eq. (8.29). A minimum cell height of `mch` $= 0.9$ was employed throughout. MCMC trajectories were a random sequence of atom, volume, lattice shear and lattice stretch MC moves. The MC move type was selected at random for each step, using a ratio 64:10:1:1 respectively.

Each MC step type had its own step length. These step lengths were updated every $\frac{\mathcal{K}}{2}$ iterations, which corresponds to a reduction of $\log\chi$ to approximately $\frac{1}{2}\log\chi$. Step lengths were adjusted to obtain an acceptance rate of 0.25 ± 0.05, using separate

trajectories not used in the rest of the calculation. The theoretically optimal value for sampling from the canonical distribution for a potential energy that is the sum of $3N + 1$ independent variables is 0.23 [9]. The acceptance rate 0.25 is purposefully chosen to be slightly larger than 0.23, since the acceptance rate of all MC proposals decreases as the space becomes constricted between step length updates.

All calculations were performed using $\mathcal{K} = 640$ live points. Fresh samples were decorrelated using a random walk of length L. For melting and evaporation, a walk length $\frac{L}{N} = 2.58 \times 10^3$ was more than sufficient, giving good agreement with the literature values above the triple point.

Equilibrating directly between the gas and the solid is difficult. The following strategy was used to extrapolate to very large walk lengths $\left(\frac{L}{N} \to \infty\right)$:

1. Repeat the calculation three times, using $\frac{L}{N} \in \left\{2.58 \times 10^2, 2.58 \times 10^{2.5}, 2.58 \times 10^3\right\}$, obtaining sublimation temperatures $\{T_0, T_1, T_2\}$.
2. Perform a least squares fit of the pairs $\left(\frac{1}{L_i}, T_i\right)$ using a straight line.
3. Take the temperature at the intercept, $T\left(\frac{1}{L} = 0\right)$.

An example of this interpolation is shown in Fig. 8.3.

In Sect. 5.5 we saw that the heat capacity is proportional to the variance of the enthalpy

$$C_P = k_\mathrm{B}\beta^2 \mathrm{Var}\left(H\left(\mathbf{q}, \mathbf{p}, P, V\right)\right). \qquad \text{(5.9 revisited)}$$

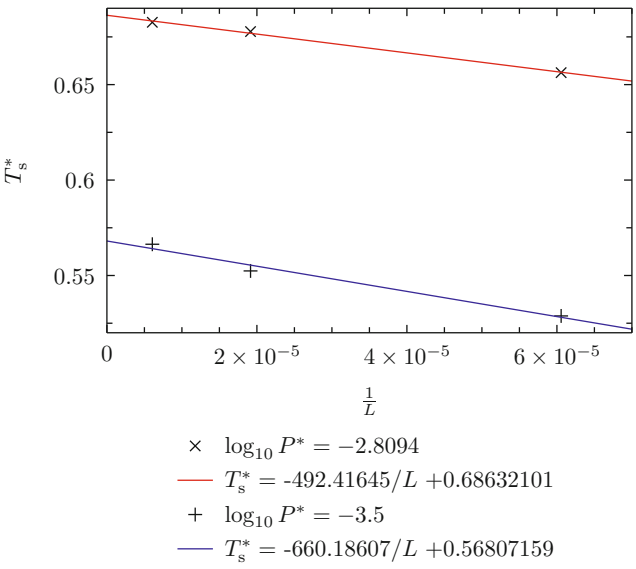

Fig. 8.3 Convergence of sublimation temperature T_s^* with respect to walk length L. The converged value is taken to be the intercept $T_s^*\left(\frac{1}{L} = 0\right)$. All calculations were performed with $N = 64$ particles in the simulation cell and $\mathcal{K} = 640$ live points

Sublimation, melting and evaporation are all first order phase transitions between phases with markedly different enthalpies. As a result, the variance of the enthalpy exhibits a peak at such phase transitions, where both phases have appreciable probability. At each (fixed) pressure we identified the temperature of these phase transitions by the corresponding peaks in the heat capacity. Above the critical point, where there is no first order phase transition, there is still a point where the variance of the enthalpy is largest and the heat capacity has a peak. The line of such maxima above the critical point is called a Widom line [10]. Therefore the critical point does not correspond to the end in the line of maxima in the heat capacity.

Along the evaporation line the density distribution is bimodal, with formally zero probability of densities between those corresponding to the two phases. At the critical point, this probability distribution is still bimodal, but intermediate densities, between the two modes, have a formerly non-zero probability, even in the thermodynamic limit [11]. At the critical point it is the mean value of the density distribution that is reported by the equation of state. To identify the point where the evaporation line ends and the Widom line begins we examined the distribution for volume at the temperature corresponding to the peak in the heat capacity.

8.8.2 Results

Literature values for the $P - T$ phase diagram of Lennard-Jonesium are shown in Fig. 8.4 and summarised in Table 8.1.

In Fig. 8.5 we present the $P - T$ phase diagram calculated using nested sampling. For comparison, a selection of the literature values given in Fig. 8.4 are also shown. Note that we also obtain the first part of the Widom line, corresponding to maxima of the heat capacity, C_P. At higher temperatures the Widom line is indiscernible from noise.

At pressures in the region of the triple point and below, equilibrium occurs between the solid and gas phases. Where this is the case, substantial differences in the phase transition temperatures are observed; the gas is stabilised at lower temperatures and higher pressures. In contrast to the nested sampling calculations, the literature values were almost all performed using a radial-cutoff equal to the radius of the largest sphere that will fit inside the simulation cell; a density dependent cutoff. For fixed cutoff values ($r_c = x\sigma$) it is well known that varying r_c can change the relative Gibbs free energies of the solid and liquid phases, altering the melting temperature by as much as 2% [13]. At a typical sublimation pressure, $\log_{10} P^* = -3.42$, with 500 particles in a cubic cell, the densities of the solid and gas phases [17] correspond to $r_c = 4.00\sigma$ in the solid phase, and $r_c = 46.8\sigma$ in the gas phase. Such a large difference will quite reasonably be expected to affect the sublimation temperature by between 3 and 9%. The differences observed are consistent with the hypothesis that the major source of error in the sublimation temperature is the cutoff scheme: the difference is positively correlated with the density difference between the solid

Table 8.1 Summary of literature for $P - T$ phase transitions in Lennard-Jonesium as shown in Fig. 8.4. Columns: phase transition; number of particles in simulation cell (N); radial cutoff (r_c); long-range correction scheme (LRC); shape of simulation cell; method used. For radial cutoff distance, all studies except Mastny and de Pablo [13] used the radius of the largest sphere that will fit inside the simulation cell (ROLS)

Reference	Transition	N	r_c	LRC	Cell shape	Method
Kofke [12]	Evaporation	Liquid/Gas: 256/256	ROLS	Mean-field	Cube	Thermodynamic integration along evaporation line
Mastny and de Pablo [13]	Melting	2048	6σ	Mean-field	Not specified	Thermodynamic integration
McNeil-Watson and Wilding [14]	Melting	108, 256, 500	ROLS	Mean-field	Cube	Phase switch Monte Carlo
Ahmed and Sadus [15]	Melting	Solid: 500 Fluid: 432	ROLS	Mean-field	Cube	Gibbs-Duhem integration
Agrawal and Kofke [16]	Melting, sublimation	Solid/Fluid: 500/432 Solid/Gas: 108/128	ROLS	Mean-field	Solid: Cube Liquid/Gas: Truncated Octahedron	Gibbs-Duhem integration
Barroso and Ferreira [17]	Melting, sublimation	500	ROLS	Solid/Fluid: Average of mean-field and FCC / Mean-field	Not specified	Thermodynamic integration
Apte and Kusaka [18]	Sublimation	500	ROLS	Mean-field	Cube	Thermodynamic integration

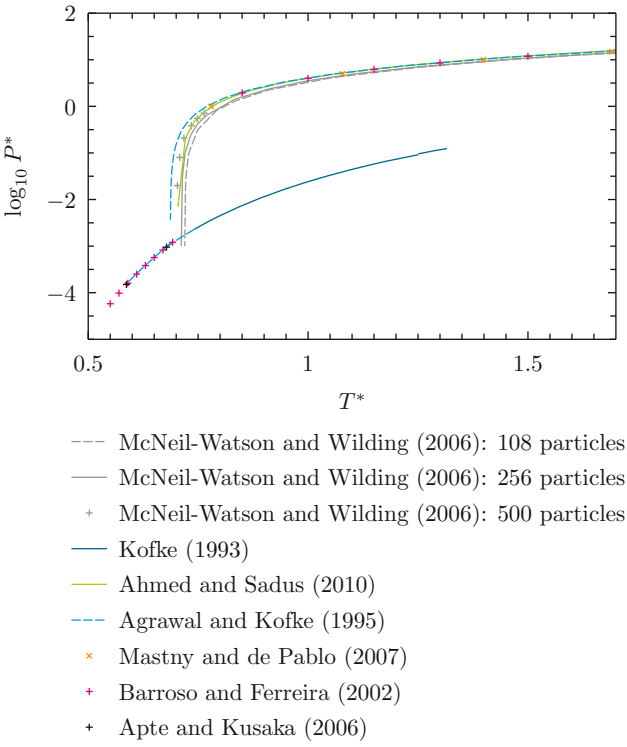

Fig. 8.4 Literature values for Lennard-Jonesium $P - T$ phase transitions. These calculations were performed with different radial cutoff schemes, long-range corrections and numbers of particles in the simulation cell. The most important differences are summarised in Table 8.1

and gas phases. That is, the difference in the sublimation temperature is negatively correlated with pressure.

At pressures above the triple point the nested sampling results lie within the spread of the literature values. This strongly suggests that finite size effects caused by only having $N = 64$ particles in the simulation cell do not dominate the error.

8.8.2.1 Identifying the Critical Point

We identify the critical point (P_c, T_c) by examining the volume distribution at maxima of the heat capacity $C_P(P, T)$. Along the evaporation curve, $T < T_c$, a first order phase transition occurs between the liquid and gas phases. Correspondingly, the volume distribution at this point has two separate modes of equal weight. At the critical point, intermediate densities, between those corresponding to the two phases, have formerly non-zero probability. Along the Widom line $T > T_c$ there is no first order phase transition and the volume distribution smoothly transitions from the bimodal distribution at the critical point, to a unimodal distribution.

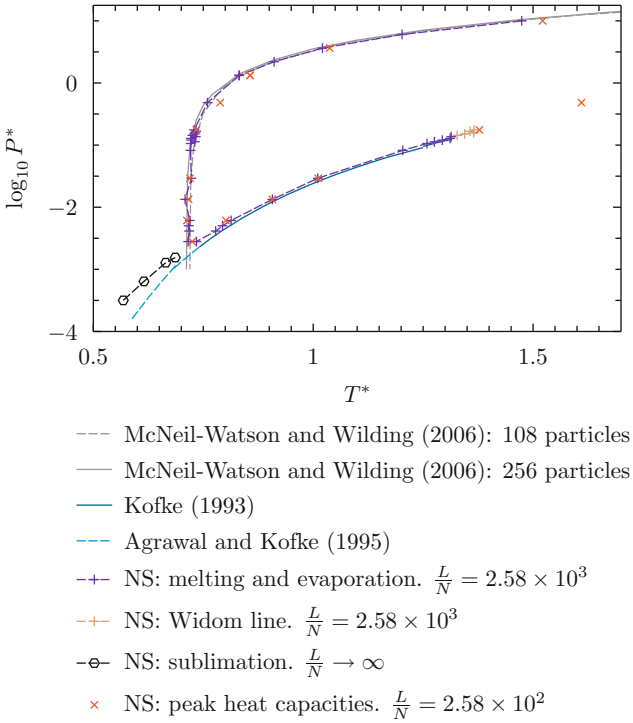

Fig. 8.5 $P - T$ phase diagram for Lennard-Jonesium from nested sampling (NS). These calculations were performed using 64 atoms, $r_c = 3\sigma$, and a mean-field long-range correction, as in Eq. (8.29). All calculations were performed using $\mathcal{K} = 640$ live points. The walk length (L) used to decorrelate fresh samples is given in the legend. A minimum cell height of $\texttt{mch} = 0.9$ was employed throughout. Convergence of sublimation peaks ($\frac{L}{N} \to \infty$) is described in the text. Good agreement with the literature is observed at pressures above the triple point. At pressures in the region of the triple point and below, where equilibrium occurs between the solid and gas phases, substantial differences are observed, stabilising the gas phase. As discussed in the text, this is probably the result of the differing r_c values. The literature values use $r_c \propto V^{\frac{1}{3}}$, which is very different in the solid and gas phases. In contrast, these nested sampling calculations were performed using $r_c = 3\sigma$. We find the critical point to be in the range $-0.86480 < \log_{10} P_c^* < -0.84286$, $1.3138 < T_c^* < 1.3273$. This is in agreement with literature values of the critical point [12, 19]. Note that we also obtain the first part of the Widom line, corresponding to a maximum of the heat capacity, C_P. At higher temperatures the Widom line is indiscernible from noise

Our simulations were performed with only 64 particles and we did not perform a finite-size scaling analysis. For such small system sizes the two modes of the volume distribution are rather broad, and the critical point is difficult to estimate, since the two modes meet, combining continuously as the temperature is raised along the evaporation line. Instead, we approximate the critical point by the following upper-bound, which is easier to estimate: we approximated the critical point as that point at which the minimum between the two maxima in the volume distribution vanishes.

We find the critical point to be in the range $-0.86480 < \log_{10} P_c^* < -0.84286$, $1.3138 < T_c^* < 1.3273$. This is consistent with the literature values of the critical point [12, 19].

The process of identifying the critical point is illustrated in Fig. 8.6.

8.8.2.2 The Solid Region

It is well known that solid Lennard-Jonesium forms close-packed face-centered cubic (FCC) and hexagonal close-packed (HCP) structures [20]. The HCP and FCC structures differ only in the stacking of adjacent layers. In a layer of close-packed spheres there are two equivalent sets of interstitial sites into which one can put another layer of close-packed spheres. In HCP the layers are arranged ABABAB, each second layer sitting directly above next-nearest layer. In FCC, layers are arranged ABCABC, where the third layer has been positioned into the other interstitial set.

At low temperatures and moderate pressures the solid forms a HCP lattice, with the FCC structure becoming stable at higher temperatures.[2] At high pressures, the FCC structure is stable at all temperatures [22]. The zero-temperature phase transition occurs at $\log_{10} P^* = 2.94$ [23]. This is far above the pressure range containing the typical phase diagram that we simulated. Furthermore, the stability of the HCP structure over FCC is only marginal; even at zero pressure the lattice energies of the HCP and FCC structures have a fractional difference of order 10^{-4} [23].

There are two consequences of these facts that are relevant to this study. First, it is very difficult to resolve the free energy difference of the two phases with a small number of particles. With a small number of particles, the enthalpy distribution is wider than the enthalpy difference between the two phases. Second, failure to resolve the correct solid phase at high temperatures has little or no effect on the melting curve, since the free energy difference between the solid and the liquid is far larger than the free energy difference between the solid phases.

We found that with $\mathcal{K} = 640$ live points, we could not reliably resolve HCP and FCC structures of the solid phase. With $\mathcal{K} = 6400$, we find that the HCP structure is stable at low temperatures, in agreement with the literature. At high temperatures, we do not observe the simple FCC structure. Rather, we observe a stacking defect phase with mixed ABA and ABC stacking of layers. This structure was identified by looking at the pdf for the bond order parameter Q_6 [24], which measures the angular arrangement of atoms in a system. The values of Q_6 for FCC and HCP are 0.57452 and 0.48476 respectively. The mixed phase takes discrete intermediate values for Q_6, corresponding to stacking faults.

The observation of a stacking fault phase is highly suggestive, and demands closer examination. Let us begin by constructing the following next-nearest-neighbour model of layers in a perfect crystal. We assume that each layer mainly interacts with those exactly adjacent to it and only weakly with the two next nearest layers

[2] Note that the behaviour of Lennard-Jonesium is different to that of a harmonic-crystal, where nearest neighbour atoms are connected by springs: at zero pressure the harmonic-crystal slightly favours the FCC lattice over the HCP lattice [21].

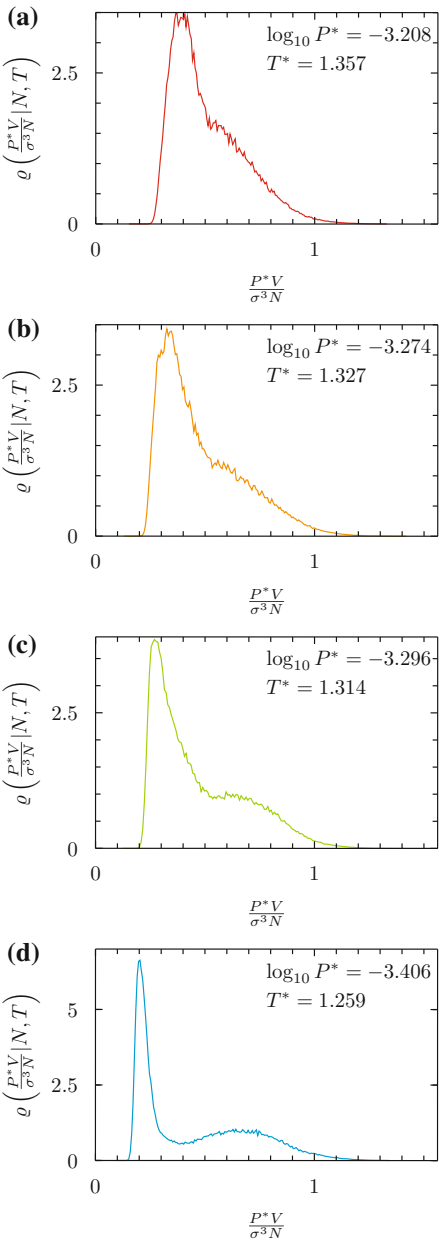

Fig. 8.6 These four volume distributions are taken at maxima of the heat capacity $C_P\,(P,T)$. They show Lennard-Jonesium at ascending pressures from lowest to highest. As described in the text, approached from along the evaporation curve, we approximate the critical point as that point at which the minimum between the two modes vanishes. We find the approximate critical point to be between Figures (**b**) and (**c**). That is, we find the approximate critical point to be in the region $-0.86480 < \log_{10} P_c^* < -0.84286$, $1.3138 < T_c^* < 1.3273$. This is in agreement with literature values of the critical point [12, 19]

either side. The energy of a layer is determined by whether it is in the centre of a stacking ABA or ABC. Relative to any particular layer, the stacking of each layer can be either A, B or C. Stackings ABA, BCB and ACA are related by symmetry oper-

ations and therefore have the same energy. Stackings BAB, CBC and CAC are also equivalent. The same is true for any combination ABC, BCA, CAB. Furthermore, stackings ABC and CBA are also related by a symmetry operation, and are equivalent. Therefore there are two states for the central layer, ABA and ABC, each with a multiplicity of 6. Since in Lennard-Jonesium HCP (ABA) has lower energy, we assign this state energy 0 and FCC (ABC) energy ϵ. Accounting for the multiplicity of the stacking faults, the partition function for a single layer is therefore

$$Z_1 = 6\left(1 + \exp\left(-\beta\epsilon\right)\right). \tag{8.30}$$

For M layers we have

$$Z_M = \frac{6^M}{M!}\left[1 + \exp\left(-\beta\epsilon\right)\right]^M. \tag{8.31}$$

For a (roughly cubic) system of N particles with density ρ we have $M = (\rho N)^{\frac{1}{3}}$, $\epsilon = \epsilon_0 (\rho N)^{\frac{2}{3}}$. Here $\frac{\epsilon_0}{2}$ is the energy per unit area and the surface area of each layer is $2(\rho N)^{\frac{2}{3}}$. Substituting for M and ϵ in (8.31), and calculating the expected number of ABC layers for our system of N particles, we obtain

$$\frac{\langle U\rangle}{\epsilon_0} = -\frac{1}{\epsilon_0}\frac{\partial \log Z_M}{\partial\beta} \tag{8.32}$$

$$= (\rho N)^{\frac{1}{3}}\,\frac{\mathrm{e}^{-\beta\epsilon_0(\rho N)^{\frac{2}{3}}}}{1 + \mathrm{e}^{-\beta\epsilon_0(\rho N)^{\frac{2}{3}}}}. \tag{8.33}$$

For sufficiently small temperatures (8.33) approaches zero, irrespective of the other parameters. Similarly, for sufficiently large N we expect no ABC layers, at all finite β and positive ρ. This supports the hypothesis that the mixed (defect phase) is an artefact of small system size.

Pronk and Frenkel [25] and Frenkel and Ladd [26] observed similar behaviour in small simulations of the hard sphere model, where the macroscopically stable FCC phase competes with the randomly stacked HCP (rHCP) phase. Indeed, in [25] the authors calculated the relative stability of the FCC and rHCP phases under fixed volume conditions, at a packing fraction $\eta = 0.5760$. By explicitly calculating the interfacial free energy between HCP and FCC phases at this density, they conclude that the rHCP phase is stable for crystals with fewer than a thousand particles per plane, while the FCC structure is stable for larger crystals.

8.9 A Binary Lennard-Jonesium Alloy

We investigated a binary Lennard-Jonesium alloy. The two species have equal atomic radii (equal σ) but different interaction strengths $\epsilon_{\mathrm{AA}}, \epsilon_{\mathrm{AB}}, \epsilon_{\mathrm{BB}}$. The potential is shown in Eq. (8.34).

$$
\begin{aligned}
U\left(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N\right) &= u_{\mathrm{c}} + \sum_{i \geq j} u_{ij} \\
u_{ij} &= \begin{cases} 4\epsilon_{ij}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], & r_{ij} \leq r_{\mathrm{c}} \\ 0, & \text{Elsewhere} \end{cases} \\
u_{\mathrm{c}} &= \frac{8\,\pi\,\sigma^3}{3V}\left[N_{\mathrm{A}}^2\epsilon_{\mathrm{AA}} + 2N_{\mathrm{A}}N_{\mathrm{B}}\epsilon_{\mathrm{AB}} + N_{\mathrm{B}}^2\epsilon_{\mathrm{BB}}\right]\left[\frac{1}{3}\left(\frac{\sigma}{r_{\mathrm{c}}}\right)^9 - \left(\frac{\sigma}{r_{\mathrm{c}}}\right)^3\right]
\end{aligned}
\tag{8.34}
$$

For the binary system we introduced an additional MC movement class: AB atom swaps. Here, two atoms are selected at random: one of type A, one of type B. An attempt is then made to exchange the coordinates of those two atoms, accepted with probability

$$
\mathrm{acc}\left(o \to n\right) = \begin{cases} 1, & Y\left(n\right) \leq Y_{\mathrm{lim}}, \\ 0, & \text{Elsewhere.} \end{cases} \tag{8.35}
$$

This MC move is given in Algorithm 8.5.

```
subroutine AB_atom_swap_mc
! AB atom swap MC step

i = choose(type=A)          ! random atom of type A
j = choose(type=B)          ! random atom of type B

save_i(x,y,z) = coordinates(x,y,z,i)

coordinates(x,y,z,i) = coordinates(x,y,z,j)
coordinates(x,y,z,j) = save_i(x,y,z)

e_trial = Y(coordinates)    ! Y = configurational enthalpy
if (e_trial > e_lim) then   ! trial rejected
  coordinates(x,y,z,j) = coordinates(x,y,z,i)
  coordinates(x,y,z,i) = save_i(x,y,z)
end if                      ! coordinates unchanged:
                            ! a null step

end subroutine
```

Algorithm 8.5 AB atom swap MC step for binary systems. Two atoms are chosen at random: one of type A, the other of type B. An attempt is made to exchange their coordinates. The swap is accepted if the new state has $Y \leq Y_{\mathrm{lim}}$.

8.9.1 Simulation Details

Our simulation used equal numbers of A and B particles: $N_A = 32$, $N_B = 32$. We used the following interaction strengths: $\epsilon_{AA} = \frac{1}{2}$, $\epsilon_{AB} = 1$, $\epsilon_{BB} = \frac{1}{2}$. The radial cutoff was fixed, $r_c = 3\sigma$.

MCMC trajectories were random sequences of atom, volume, lattice shear, lattice stretch, and AB swap MC moves. The MC move type was selected at random for each step, using a ratio 64:10:1:1:1 respectively. All MC proposals (except AB atom swaps) had a separate MC step length. These step lengths were updated exactly as described in Sect. 8.8.1, every $\frac{\mathcal{K}}{2}$ iterations, to obtain an acceptance rate of 0.25 ± 0.05.

Fresh samples were decorrelated from their clones using a random walk length $\frac{L}{N_A+N_B} = 2.58 \times 10^3$. Most pressures were simulated using $\mathcal{K} = 640$ live points, although some pressures were simulated using $\mathcal{K} = 5120$: those points are clearly marked on Fig. 8.7. A minimum cell height of $\texttt{mch} = 0.9$ was used for all calculations. We applied the same methodology as described in Sect. 8.8.

8.9.2 Results

Figure 8.7 shows the phase diagram for this binary Lennard-Jonesium system. An order-disorder transition occurs at $T^* = 2.53 \times 10^{-1} \pm 1.15 \times 10^{-2}$. The transition is between ordered and disordered solid phases of very nearly the same volume, and is therefore almost temperature independent. Throughout this section we use $P^* = \frac{P\sigma^3}{\epsilon_{AB}}$, $T^* = \frac{k_B T}{\epsilon_{AB}}$.

The radial distribution function of the system was calculated either side of the mixing transition, at $\log_{10} P^* = -2.398$, $T^* \in \{0.172, 0.388\}$. When calculating the radial distribution function, periodic images were ignored. Consequently the calculated radial distribution function goes to zero, rather than one, at large radii. The radial distribution functions, shown in Fig. 8.8, indicate that the phase transition corresponds to a loss of long range order with respect to positions of A and B atoms on sites of the same lattice.

As in single-species Lennard-Jonesium, small system size precluded the formation of pure structural phases. For this binary system we observed stacking defects, even at 0 K. Nevertheless the ordering of atoms on the close-packed structure was consistent. Two phases are formed: phases I and II. A phase transition occurs between phases I

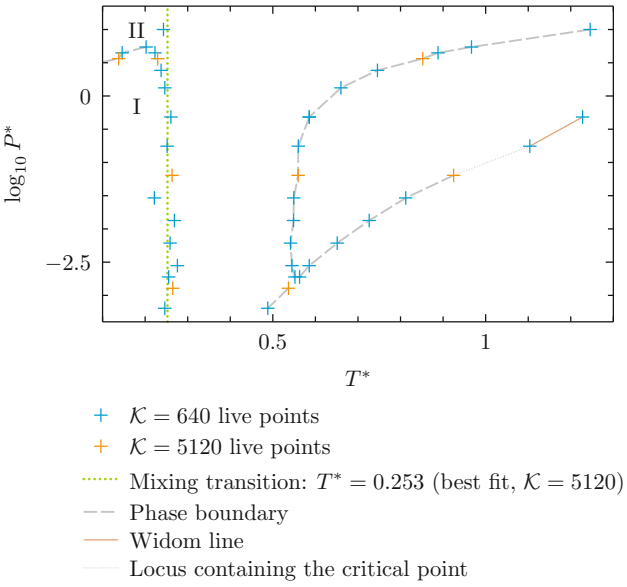

Fig. 8.7 Phase diagram of a binary Lennard-Jonesium alloy. $N_\mathrm{A} = 32$, $N_\mathrm{B} = 32$, $\epsilon_\mathrm{AA} = \frac{1}{2}$, $\epsilon_\mathrm{AB} = 1$, $\epsilon_\mathrm{BB} = \frac{1}{2}$, $\sigma_{ij} = \sigma \; \forall \, (i, j)$. All calculations used $\mathtt{mch} = 0.9$ and a random walk length $\frac{L}{N_\mathrm{A}+N_\mathrm{B}} = 2.58 \times 10^3$. Each pressure used either 640 or 5120 live points, as indicated in the legend. There is a solid-solid chemical ordering transition at $T^* = 2.53 \times 10^{-1} \pm 1.15 \times 10^{-2}$. This mixing transition corresponds to a loss of long range ordering of A and B atoms on sites of the same lattice. The site-ordered solid has two phases, I and II. Both of these phases are close-packed, with atoms arranged ABAB in one dimension. A phase transition occurs between phases I and II at zero temperature within the pressure range $0.386 < \log_{10} P^* < 0.561$. In phase I, atoms are arranged AABB in the other two dimensions, whereas in phase II the atoms are arranged into sheets of A and B atoms. This allows phase II to become slightly thinner in the dimension normal to the planes, which stabilises phase II at high pressures. The two phases are shown in Fig. 8.9

and II at zero temperature within the pressure range $0.386 < \log_{10} P^* < 0.561$. The regions occupied by each are shown on the phase diagram, Fig. 8.7. In both phases, atoms are arranged ABAB along one dimension and in columns of pure A or B atoms in another. In phase I, those columns are arranged in a pattern AABB, whereas in phase II the columns are arranged into planes of pure A or B atoms. Both phases are shown in Fig. 8.9. Our parameters favour adjacency of A and B atoms. The AABB pattern allows maximal adjacency of A and B atoms, lowering the potential energy of configurations corresponding to phase I. Conversely, arranging the atoms into planes reduces the lattice spacing between those planes in phase II and hence reduces the volume overall. This stabilises phase II at high pressures.

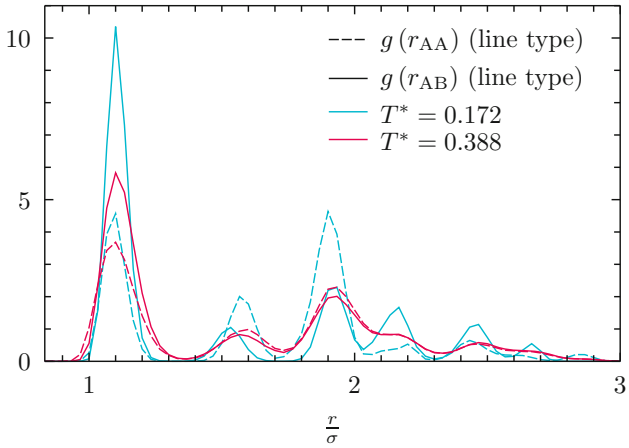

Fig. 8.8 Radial distribution functions of a binary Lennard-Jonesium system exhibiting a mixing transition. The potential energy function parameters for this binary Lennard-Jonesium alloy are described in the text. The mean radial distribution function was calculated from the nested sampling output either side of the mixing transition shown in Fig. 8.7, at $\log_{10} P^* = -2.398$, $T^* \in \{0.172, 0.388\}$. These radial distribution functions are centred on A-type atoms. The dotted line indicates the radial distribution function for A-A pairs, whilst the continuous line indicates the radial distribution function for A-B pairs. First, the radial distribution functions go nearly or exactly to zero between peaks, and the peaks are at the same positions for the two temperatures. This indicates that both phases correspond to solids, and that those solids have the same lattice structure. At temperatures above the transition (red lines), we can see that (apart from short range ordering) the A and B atoms are equivalent. Below the transition (blue lines) there is a clear difference between the two atom types. This indicates that the transition corresponds to a loss of long range order with respect to positions of A and B atoms on sites of the lattice

8.10 NPB Embedded Atom Model for Aluminium

We applied nested sampling to the NPB embedded atom model (EAM) potential of aluminium [3] to demonstrate how nested sampling can be applied to many body potentials that correspond reasonably well to physical systems. Figure 8.10 shows the phase diagram as calculated by nested sampling for this interatomic potential. The MCMC method described in Chap. 8 was used for the nested sampling calculations. Also shown are experimental results for the evaporation and melting transitions, as well as the critical point. Phase transitions were obtained from maxima of the heat capacity, and the critical point was obtained using the methodology described in Sect. 8.8.2.1. Most of these calculations used $\mathcal{K} = 804$ samples with a walk length $L = 3120$. More sampling was required to resolve the solid-solid transitions. For those calculations $\mathcal{K} = 6432$ samples were used, with a walk length $L = 15120$. A minimum cell height `mch` $= 0.65$ was used for all of the calculations.

Nested sampling obtained the same location for the evaporation line and the critical point as Gibbs ensemble Monte Carlo [27]. Maxima of the heat capacity along the

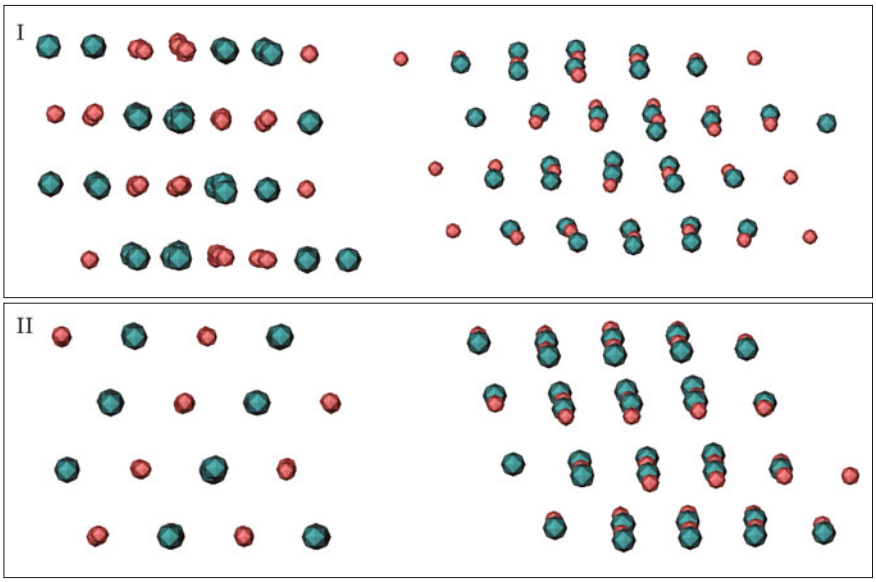

Fig. 8.9 Binary alloy phases I and II. In both phases, atoms are arranged ABAB in one dimension (right side), and into columns of pure A or B in another (left side). In phase I columns of pure A and B are arranged into a pattern AABB, whereas in phase II the columns form planes of pure A or B. Our parameters favour adjacency of A and B atoms. The AABB pattern allows maximal adjacency of A and B atoms, lowering the potential energy of configurations corresponding to phase I. Conversely, arranging the pure atoms into planes, which strongly attract one another, reduces the lattice spacing between those planes in phase II and hence reduces the volume overall. For this reason phase II is stable at high pressures

evaporation transition are joined by a red line. A section of the Widom line above the critical point was also resolved. Those points are not connected by the red line.

The nested sampling results agree well with the experimental values for the melting transition, and show reasonable agreement with empirical results for the critical point.

Empirically, aluminium exhibits FCC, HCP and BCC structures in the solid phase at 0 K. The FCC structure is stable at low pressures, and a transition to HCP occurs at 217 GPa [28]. A further transition from HCP to BCC occurs at between 360 and 565 GPa [29–31]. We can see the structure predicted by the NPB–EAM potential within the solid region at zero temperature is qualitatively incorrect. There is no pressure for which the HCP phase is stable. The NPB–EAM potential does identify the correct phases and in the correct order at finite temperatures, but at pressures a fraction of the experimental values.

We conclude that the NPB–EAM potential is successful in that it identifies reasonable values for the melting transition, and qualitatively correct structural phase

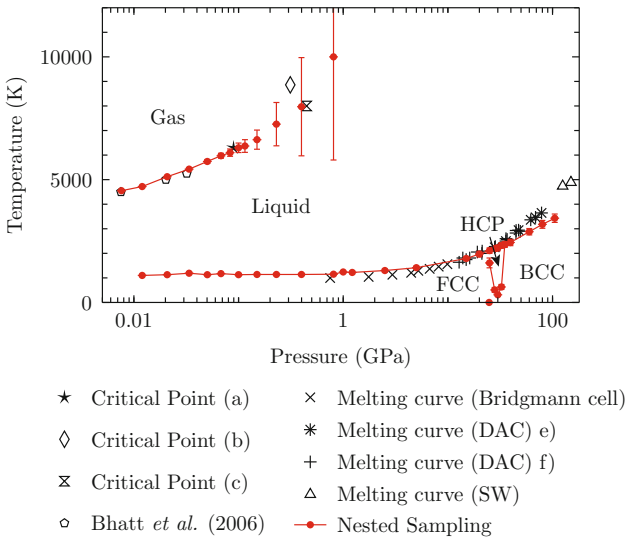

Fig. 8.10 Phase diagram for aluminium with the NPB–EAM potential. Red symbols show the nested sampling results, with error bars calculated as the width at half maximum of the peaks on the heat capacity curves. The estimates of the critical points from the literature were obtained as follows: **a** Gibbs ensemble Monte Carlo [32] simulation of NPB-EAM potential [27] **b** estimate from experiments [33], **c** estimate from experiments [34]. Black symbols show experimental melting points measured with Bridgmann cell experiments [35], with Diamond anvil cell measurements (DAC **e** [36] and DAC **f** [37]) and Shock wave experiments [38]. The evaporation temperatures from "Bhatt et al." correspond to the same paper as the critical point (**a**), and were also calculated by Gibbs ensemble Monte Carlo [32] with the NPB-EAM potential

transitions at finite temperature in the solid region. The presence of the Widom line and the solid-solid transitions were unexpected at the time we performed the calculation. Nested sampling resolved each of these virtually without guidance.

8.11 Comparison to Parallel Tempering

The calculations presented in this section were performed by Lívia Bartók-Pártay.

The performance of nested sampling was compared to that of parallel tempering. Parallel tempering is a state of the art temperature scheduling algorithm, that simultaneously equilibrates sampling at a series of temperatures. At each temperature, Metropolis Monte Carlo [39] is used to explore the isobaric isothermal distribution (4.54). Parallel tempering hugely accelerates exploration of those distributions at low temperatures by periodically attempting to swap current configurations between different temperatures. This allows the sampler to escape local free energy minima. These swaps are accepted according to the Metropolis-Hastings rule for the joint distribution

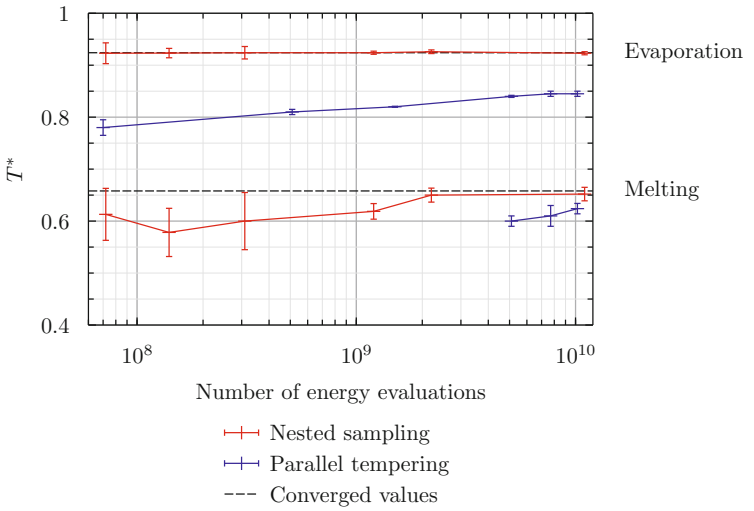

Fig. 8.11 Comparison of the convergence of nested sampling and parallel tempering. Details of these calculations are given in the text. Where no data is shown for the melting transition for parallel tempering, the solid phase had not yet been discovered. Evaporation is converged for nested sampling using fewer than 7.2×10^7 enthalpy evaluations, and the melting transition is converged using 2.2×10^9 enthalpy evaluations. Extrapolating using a line of best fit for the parallel tempering data as shown here, we predict that parallel tempering would require 3×10^{10} enthalpy evaluations to converge the melting transition, and 3×10^{12} to converge the evaporation transition. The converged values were obtained using nested sampling, with 5.7×10^{10} enthalpy evaluations. These calculations were performed by Lívia Bartók-Pártay

$$\mathrm{acc}\left[\left(\mathbf{R}_i, \beta_i\right), \left(\mathbf{R}_j, \beta_j\right) \rightarrow \left(\mathbf{R}_j, \beta_i\right), \left(\mathbf{R}_i, \beta_j\right)\right] = \min\left(1, \frac{\exp\left[-\beta_i Y\left(\mathbf{R}_j\right) - \beta_j Y\left(\mathbf{R}_i\right)\right]}{\exp\left[-\beta_i Y\left(\mathbf{R}_i\right) - \beta_j Y\left(\mathbf{R}_j\right)\right]}\right). \tag{8.36}$$

In this way, the swaps do not disturb the isobaric isothermal distributions.

We simulated a periodic system of $N = 64$ Lennard-Jonesium with the "truncated and shifted" potential described in Sect. 12.3.7 on page 136, at a reduced pressure of $\log_{10} P^* = -1.57$. The simulation was performed in the fully flexible cell formulation. For parallel tempering 128 temperatures were spaced evenly across the range $T^* \in [0.4, 1.4]$. For nested sampling, the cost of each calculation was set by changing both the walk length L used to decorrelate fresh samples, and the number of live configurations used, $\mathcal{K}$. The nested sampling calculations were performed as described in Sect. 8.8. Initial configurations for parallel tempering were taken from the gas phase. Figure 8.11 shows the results of these calculations.

We see from Fig. 8.11 that nested sampling requires fewer than 7.2×10^7 energy evaluations to converge the evaporation transition, and 2.2×10^9 to converge the melting transition. For parallel tempering, neither result has converged after 1.02×10^{10} energy evaluations. Extrapolating linearly from Fig. 8.11 we predict that parallel tempering would require 3×10^{10} energy evaluations to converge the melting

transition, and 3×10^{12} to converge the evaporation transition. We conclude that for 64 particles, nested sampling is at least an order of magnitude faster than parallel tempering for calculating the melting temperature, and four orders of magnitude faster for calculating the evaporation temperature.

Parallel tempering works most efficiently when the isobaric isothermal distributions at neighbouring temperatures have a high degree of overlap. When this is not the case, the acceptance rate for configuration swaps (8.36) becomes very low. This problem is typically pronounced at first order phase transitions of periodic simulations, unless an accurate estimate of the transition temperature is included as one of the temperatures used in parallel tempering. This problem of a reduced exchange acceptance rate in parallel tempering is more evident for the evaporation transition than the melting transition because the modes of the enthalpy distribution for the hot solid and the cold liquid have overlapping tails, whereas this is not the case for the liquid and gas phases. The behaviour of nested sampling is different to parallel tempering in this regard. From Fig. 8.11 we see that nested sampling is much more efficient at resolving the evaporation transition (which corresponds to a large change in χ) than the melting transition (which corresponds to a smaller change in χ, but for which the sampler must efficiently explore condensed phases). This suggests that the variation in χ across a phase transition is not as important a factor for nested sampling as the rate of decorrelation in the clone and decorrelate algorithm. Exploration of condensed phases using MCMC is evidently inefficient, even at constant pressure. This is the motivation for Part III of this thesis, where we explore the use of classical dynamics for decorrelation.

8.12 Summary

We have developed nested sampling into an automated tool for calculating phase diagrams from first principles. By inspecting generic physical quantities such as the heat capacity and volume distribution, we are able to identify phase transitions that we were not expecting. Examples of this include as the order-disorder transition in the binary Lennard-Jonesium alloy, the transition between phases I and II, and the Widom line in the Lennard-Jonesium, binary Lennard-Jonesium, and NPB-EAM aluminium systems.

In Sect. 8.4 we showed that the overwhelming majority of simulation cells in the fully flexible cell formulation are very thin. Such simulation cells correspond to systems of effectively reduced dimension, their physics being dominated by periodic images of each atom which are introduced by periodic boundary conditions. We showed how one can rigorously modify the fully flexible simulation cell formulation, introducing a minimum cell height criterion, $\mathtt{ch}\,(\mathbf{h}_0) \geq \mathtt{mch}$. This excludes simulation cells that are "too thin".

We tested nested sampling on simple Lennard-Jonesium, and verified that it gives the same results as found in the literature above the triple point. Below the triple point our results do not agree with the literature values. The literature values all use

a radial cutoff scheme in which the r_c is proportional to $\rho^{-\frac{1}{3}}$. In contrast, we use a fixed value for r_c. We identified this as the cause of the discrepancy.

The free energy difference between HCP and FCC structures in the solid region of Lennard-Jonesium is very small. The HCP and FCC structures are related by a series of stacking faults, with HCP and FCC as the limiting structures. We identify HCP as the stable phase at zero temperature for the pressures considered, but did not find a transition to FCC at higher temperatures. Instead we witnessed the gradual introduction of stacking faults as the temperature is increased. All calculations were performed using periodic simulations of only 64 particles. We argued, by introducing a simple model, that it is not possible to discern whether this is a real phenomenon, or a system size effect from these simulations, which used only 64 particles.

The heat capacity of Lennard-Jonesium was calculated at a pressure $\log_{10} P^* = -1.57$ using both nested sampling and parallel tempering. Sixty-four particles were used for the simulation. We found that nested sampling requires fewer than 7.2×10^7 energy evaluations to converge the evaporation transition, and 2.2×10^9 to converge the melting transition. After 1.02×10^9 energy evaluations we found that neither transition was converged with parallel tempering. Extrapolating from our results we predict that parallel tempering would require 3×10^{10} energy evaluations to converge the melting transition, and 3×10^{12} to converge the evaporation transition. We conclude that, to the best of our knowledge, the implementation of nested sampling developed here is the state of the art Monte Carlo algorithm for simulating materials, when we wish to uncover "unknown unknown" collective phenomena.

8.13 Further Work

We compared the performance of Nested Sampling to parallel tempering in Sect. 8.11. We noted that nested sampling is markedly slower at converging the melting transition than the evaporation transition. The reason for this is that the single atom MC steps used in the MCMC exploration are not cooperative. In a condensed phase, simple MCMC is not an efficient method by which to explore configuration space. In Part III we explore the use of classical mechanics for decorrelating configurations in the condensed phase.

The "clone and decorrelate" Algorithm 7.3 creates a "hole and a bump" in the uniform distribution at each iteration. Much of the computational expense in nested sampling is spent re-equilibrating the uniform distribution. In recent discussions, John Skilling described a method by which one can often clone from the highest energy configuration, without introducing errors for multimodal potentials. This ought to reduce the cost of the algorithm dramatically, and mitigate the problem of multimodal potentials explained in Sect. 7.4.

References

1. J. Skilling, Nested sampling. AIP Conf. Proc. **735**, 395 (2004)
2. J. Skilling, Nested sampling for general Bayesian computation. Bayesian Anal. **1**, 833 (2006)
3. A.W. Jasper, N.E. Schultz, D.G. Truhlar, Analytic potential energy functions for simulating aluminum nanoparticles. J. Phys. Chem. B **109**, 3915 (2005)
4. D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications, ser* (Elsevier Science, Computational science series, 2001)
5. G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms. J. Chem. Phys. **101**, 4177 (1994)
6. M. Tuckerman, *Statistical Mechanics and Molecular Simulations* (Oxford University Press, 2008)
7. W. Wood, Monte Carlo calculations for hard disks in the isothermal-isobaric ensemble. J. Chem. Phys. **48**, 415 (1968)
8. L.B. Pártay, A.P. Bartók, G. Csányi, Nested sampling for materials: the case of hard spheres. Phys. Rev. E **89**, 022302 (2014)
9. R.M. Neal, MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo* (CRC Press, New York, NY, 2011), p. 113
10. V.V. Brazhkin, A.G. Lyapin, V.N. Ryzhov, K. Trachenko, Y.D. Fomin, E.N. Tsiok, Where is the supercritical fluid on the phase diagram? Phys. Uspekhi **55**, 1061 (2012)
11. A. Bruce, N. Wilding, Scaling fields and universality of the liquid-gas critical point. Phys. Rev. Lett. **68**, 193 (1992)
12. D.A. Kofke, Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line. J. Chem. Phys. **98**, 4149 (1993)
13. E.A. Mastny, J.J. de Pablo, Melting line of the Lennard-Jones system, infinite size, and full potential. J. Chem. Phys. **127**, 104504 (2007)
14. G.C. McNeil-Watson, N.B. Wilding, Freezing line of the Lennard-Jones fluid: a phase switch Monte Carlo study. J. Chem. Phys. **124**, 064504 (2006)
15. A. Ahmed, R.J. Sadus, Effect of potential truncations and shifts on the solidliquid phase coexistence of Lennard-Jones fluids. J. Chem. Phys. **133**, 124515 (2010)
16. R. Agrawal, D.A. Kofke, Thermodynamic and structural properties of model systems at solid-fluid coexistence. Mol. Phys. **85**, 43 (1995)
17. M.A. Barroso, A.L. Ferreira, Solid-fluid coexistence of the Lennard-Jones system from absolute free energy calculations. J. Chem. Phys. **116**, 7145 (2002)
18. P.A. Apte, I. Kusaka, Direct calculation of solid-vapor coexistence points by thermodynamic integration: application to single component and binary systems. J. Chem. Phys. **124**, 184106 (2006)
19. B. Smit, Phase diagrams of Lennard-Jones fluids. J. Chem. Phys. **96**, 8639 (1992)
20. T. Kihara, S. Koba, Crystal structures and intermolecular forces of rare gases. J. Phys. Soc. Japan **7**, 348 (1952)
21. W.G. Hoover, Entropy for small classical crystals. J. Chem. Phys. **49**, 1981 (1968)
22. A.N. Jackson, A.D. Bruce, G.J. Ackland, Lattice-switch Monte Carlo method: application to soft potentials. Phys. Rev. E **65**, 036710 (2002)
23. F.H. Stillinger, Lattice sums and their phase diagram implications for the classical Lennard-Jones model. J. Chem. Phys. **115**, 5208 (2001)
24. P. Rein ten Wolde, M.J. Ruiz-Montero, D. Frenkel, Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. J. Chem. Phys. **104**, 9932 (1996)
25. S. Pronk, D. Frenkel, Can stacking faults in hard-sphere crystals anneal out spontaneously? J. Chem. Phys. **110**, 4589 (1999)
26. D. Frenkel, A.J.C. Ladd, New Monte Carlo method to compute the free energy of arbitrary solids, application to the fcc and hcp phases of hard spheres. J. Chem. Phys. **81**, 3188 (1984)
27. D. Bhatt, A.W. Jasper, N.E. Schultz, J.I. Siepmann, D.G. Truhlar, Critical properties of aluminum. J. Am. Chem. Soc. **128**, 4224 (2006)

28. Y. Akahama, M. Nishimura, K. Kinoshita, H. Kawamura, Y. Ohishi, Evidence of a fcc-hcp transition in aluminum at multimegabar pressure. Phys. Rev. Lett. **96**, 045505 (2006)
29. J.C. Boettger, S.B. Trickey, High-precision calculation of the equation of state and crystallographic phase stability for aluminum. Phys. Rev. B **53**, 3007 (1996)
30. G.V. Sin'ko, N.A. Smirnov, Ab initio calculations of elastic constants and thermodynamic properties of bcc, fcc, and hcp Al crystals under pressure. J. Phys. Condens. Matter **14**, 6989 (2002)
31. M.J. Tambe, N. Bonini, N. Marzari, Bulk aluminum at high pressure: a first-principles study. Phys. Rev. B **77**, 172102 (2008)
32. A.Z. Panagiotopoulos, Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. Mol. Phys. **61**, 813 (1987)
33. A.A. Likalter, Critical points of metals of three main groups and selected transition metals. Phys. A **311**, 137 (2002)
34. V. Fortov, I. Iakubov, *Non-Ideal Plasma* (Plenum Press, New York, 2000)
35. D. Errandonea, The melting curve of ten metals up to 12 GPa and 1600 K. J. Appl. Phys. **108**, 033517 (2010)
36. R. Boehler, M. Ross, Melting curve of aluminum in a diamond cell to 0.8 Mbar: implications for iron. Earth Planet. Sci. Lett. **153**, 223 (1997)
37. A. Hänström, P. Lazor, High pressure melting and equation of state of aluminium. J. Alloys Compd. **305**, 209 (2000)
38. J.W. Shaner, J.M. Brown, R.G. McQueen, Melting of metals above 100 GPa, *High Pressure in Science and Technology* (North-Holland, Amsterdam, 1984), p. 137
39. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087 (1953)

Chapter 9
Equations of State

9.1 Introduction

It was explained in Sect. 5.2 that for an atomic fluid containing very many nuclei, pressure temperature and volume are related by an equation of state

$$\phi\,(P, T, V) = 0. \qquad \text{(5.3 revisited)}$$

A typical equation of state was illustrated in Fig. 5.1 on page 34.

We arrive at this simple picture as follows. At any pressure and temperature, there exists a pdf for V: $\varrho\,(V|P, T)$. The relative fluctuations of volume are proportional to $1/\sqrt{N}$, according to Eq. (3.14). Therefore in the infinite system limit, the pdf for V must be a delta function, except at a phase transition, where it is the sum of two delta functions. Provided that our simulated system is larger than the correlation length, this delta function will have the same mean value as the pdf in our small system. Therefore, we can calculate $V\,(P, T)$ for the infinite system at any given pressure and temperature by identifying the mean of the dominant mode in $\varrho\,(V|P, T)$.

An algorithm to calculate the mean volume of the dominant mode has the following outline:

1. Generate `Npost` posterior volume samples, $\{D_k\}$.
2. Perform a Bayesian model selection over mixtures of $M \in \{1, 2, \ldots, \texttt{NGmax}\}$ Gaussians. This gives us prob $(M|\{D_k\})$. At each M assume that the maximum likelihood mixture adequately characterises the distribution prob $(\boldsymbol{\theta}|M, \{D_k\})$.
3. For $M > 1$, partition the maximum likelihood mixture into overlapping sets of Gaussians.
4. Calculate the total probability mass and mean value of each partition.
5. Record the average of the dominant partitions' mean values, weighted by prob $(M|\{D_k\})$.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_9

9.2 Algorithmic Details

The details of the algorithm described above are given in Sects. 9.2.1–9.2.4.

9.2.1 Generating Posterior Samples

The output of nested sampling [1, 2] consists of a series of configurations $\{(\mathbf{s}, \mathbf{h}_0, V)_i\}$, each with a microscopic enthalpy Y_i, and associated volume of configuration space $\chi_0(\zeta_{i-1} - \zeta_i)$, (7.12). The partition function is then approximated as

$$\Delta(\beta) \simeq \frac{1}{N! h^{3N} V_0} \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \sum_{i=1}^{N_{\text{its}}} \Delta\chi_i \, \mathrm{e}^{-\beta Y_i}. \qquad \text{(8.8 revisited)} \qquad (9.1)$$

Posterior samples are selected according to the weight they contribute to the partition function (8.8)

$$w_j = \frac{\Delta\chi_j \, \mathrm{e}^{-\beta Y_j}}{\Delta(\beta)}. \qquad (9.2)$$

That is, we select `Npost` configurations by sampling with replacement, according to the weights (9.2).

9.2.2 Bayesian Model Selection over a Mixtures of Gaussians

A mixture of M Gaussians with amplitudes $\mathbf{A} = (A_1, A_2, \ldots, A_M)$, means $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M)$ and standard deviations $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_M)$ is given by

$$G_M(x|\mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{i=1}^{M} \frac{A_i}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right]. \qquad (9.3)$$

We apply the methodology and assumptions described in Appendix A to calculate prob$(M|\{D_k\})$.

Under the assumptions made in Appendix A, prob$(M|\{D_k\})$ (A.5), (A.11) require the parameters $\boldsymbol{\theta}_0$ that maximise the log likelihood $L(\boldsymbol{\theta}) = \log(\text{prob}(\{D_k\}|M, \boldsymbol{\theta}))$, together with the value of the prior over the parameters (A.10), and the determinant of the Hessian matrix $\nabla\nabla L|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.

Defining $\boldsymbol{\theta} = (\mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma})$, the maximum likelihood parameter set $\boldsymbol{\theta}_0$ is obtained by application of the soft K-means algorithm, version 2 [3]. That algorithm is presented in Appendix B, on page 153. The soft K-means algorithm, version 2 requires a number `Npost` posterior samples, which can be generated as in Sect. 9.2.1. We found `Npost` $= 2000$ to be a good number of samples.

We have assumed that the prior over the parameters takes a constant value inside an allowed range

$$\text{prob}\,(\boldsymbol{\theta}|M) = \left[\sigma_{\max}^{M}\,(\mu_{\max}-\mu_{\min})^{M}\int_{(0,1)^M} \mathrm{d}A_1\ldots\mathrm{d}A_M\,\delta\left(1-\sum_{i=1}^{M}A_i\right)\right]^{-1}. \tag{A.10 revisited}$$

For $\mu_{\max}$ and $\mu_{\min}$ we used maximum and minimum volumes of the `Npost` posterior samples, and $\sigma_{\max}$ was taken to be $5\sigma_0$ with σ_0 the standard deviation of the posterior samples. The integrals

$$f_M = \int_{(0,1)^M} \mathrm{d}A_1\ldots\mathrm{d}A_M\,\delta\left(1-\sum_{i=1}^{M}A_i\right) \tag{9.4}$$

were evaluated using Mathematica [4].

The Hessian $\nabla\nabla L|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ was calculated as follows. The total log likelihood is given by

$$L = \sum_{k=1}^{\texttt{Npost}} L_k \tag{9.5}$$

$$L_k = \log s_k \tag{9.6}$$

where we define

$$s_k = G_M\,(x_k|\boldsymbol{\theta})\,. \tag{9.7}$$

Applying the chain rule for differentiation we obtain

$$\frac{\partial^2 L_k}{\partial\theta_i\theta_j} = -\frac{1}{s_k^2}\frac{\partial s_k}{\partial\theta_i}\frac{\partial s_k}{\partial\theta_j} + \frac{1}{s_k}\frac{\partial^2 s_k}{\partial\theta_i\theta_j}. \tag{9.8}$$

Terms in (9.8) are calculated individually as follows:

$$\frac{\partial s_k}{\partial A_i} = \frac{1}{\sqrt{2\,\pi}\sigma_i}\exp\left[-\frac{(x_k-\mu_i)^2}{2\sigma_i^2}\right] \tag{9.9}$$

$$\frac{\partial s_k}{\partial \mu_i} = \frac{A_i}{\sqrt{2\,\pi}\sigma_i^3}\,(x_k-\mu_i)\exp\left[-\frac{(x_k-\mu_i)^2}{2\sigma_i^2}\right] \tag{9.10}$$

$$\frac{\partial s_k}{\partial \sigma_i} = \frac{A_i}{\sqrt{2\,\pi}\sigma_i^4}\left([x_k-\mu_i]^2-\sigma_i^2\right)\exp\left[-\frac{(x_k-\mu_i)^2}{2\sigma_i^2}\right] \tag{9.11}$$

$$\frac{\partial^2 s_k}{\partial A_i A_j} = 0, \qquad \frac{\partial^2 s_k}{\partial A_i \mu_j} = \delta_{ij}\frac{1}{A_i}\frac{\partial s_k}{\partial \mu_i}, \qquad \frac{\partial^2 s_k}{\partial A_i \sigma_j} = \delta_{ij}\frac{1}{A_i}\frac{\partial s_k}{\partial \sigma_i} \tag{9.12}$$

$$\frac{\partial^2 s_k}{\partial \mu_i \mu_j} = \delta_{ij} \frac{1}{\sigma_i} \frac{\partial s_k}{\partial \sigma_i} \tag{9.13}$$

$$\frac{\partial^2 s_k}{\partial \sigma_i \mu_j} = \delta_{ij} \frac{(x_k - \mu_i)^2 - 3\sigma_i^2}{\sigma_i^3} \frac{\partial s_k}{\partial \mu_i} \tag{9.14}$$

$$\frac{\partial^2 s_k}{\partial \sigma_i \sigma_j} = \delta_{ij} \left[\frac{(x_k - \mu_i)^2}{\sigma_i^3} - \frac{4}{\sigma_i} + \frac{1}{x_k - \mu_i + \sigma_i} + \frac{1}{\mu_i - x_k + \sigma_i} \right] \frac{\partial s_k}{\partial \sigma_i} \tag{9.15}$$

Having calculated the probability for each number of Gaussians M, we assume that the distribution prob $(\boldsymbol{\theta}|M, \{D_k\})$ is well characterised by the parameters $\boldsymbol{\theta}_0$.[1] The parameters $\boldsymbol{\theta}_0$ were then used to partition the mixture into one or more non-overlapping sets. That process is described in Sect. 9.2.3.

9.2.3 *Partitioning Gaussians into Non-overlapping Sets*

The fact that a set of posterior samples may be better described by two Gaussians than one does not guarantee that there are two modes in the distribution. Examples of posterior distributions demonstrating this are shown in Figs. 9.1 and 9.2. In both of these cases the model selection gives higher probability to a mixture of two Gaussians than to a single Gaussian (factors of 1.2×10^2 in the solid phase, and 2.7×10^{102} for the supercritical phase, where the distribution is far from Gaussian.) However, close to the melting point there are clearly two separate phases, as shown in Fig. 9.3. Indeed, the model selection gives prob $(M = 2|\{D_k\}) = 1.3 \times 10^{252} \times$ prob $(M = 1|\{D_k\})$ for these conditions so close to the melting point. Clearly a method is required to partition Gaussians into separate modes.

We partition Gaussians into distinct modes using the following heuristic:

> Gaussians A and B have means μ_A, μ_B, and standard deviations σ_A, σ_B respectively. $\mu_A < \mu_B$.
>
> - Gaussian A is in the same partition as Gaussian B if $\mu_B - \mu_A < \sigma_A + \sigma_B$ OR there is no point $\mu_A < x < \mu_B$ with $G_M(x|\boldsymbol{\theta}) < \min(G_M(\mu_A|\boldsymbol{\theta}), G_M(\mu_B|\boldsymbol{\theta}))$.
> - If Gaussian A is in the same partition as Gaussian B, and Gaussian B is in the same partition as Gaussian C, then Gaussian A is in the same partition as Gaussian C.

[1] This is made possible by the assumption made in Appendix A, Sect. A.1, of a uniform prior over the parameters (A.10). Applying Bayes' Theorem, we see prob $(\boldsymbol{\theta}|M, \{D_k\}) =$ prob $(\{D_k\}|M, \boldsymbol{\theta}) \times \frac{\text{prob}(\boldsymbol{\theta}|M)}{\text{prob}(\{D_k\}|M)}$. With our assumption of a uniform prior for the parameters, the fraction is simply a constant.

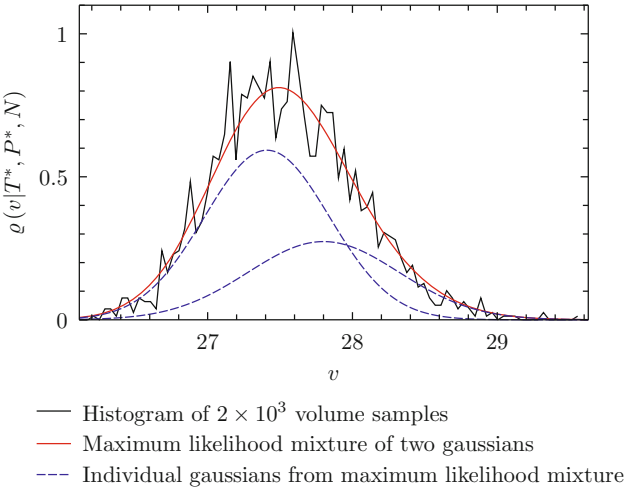

Fig. 9.1 Lennard-Jonesium periodic simulation of 64 atoms in the solid phase. $N = 64$, $T^* = 6.463 \times 10^{-1}$, $P^* = 1.758 \times 10^{-1}$. Using the parameters described in Sect. 9.2.2, prob $(M = 2|\{D_k\}) = 1.2 \times 10^2$ prob $(M = 1|\{D_k\})$. Nevertheless there is clearly one mode

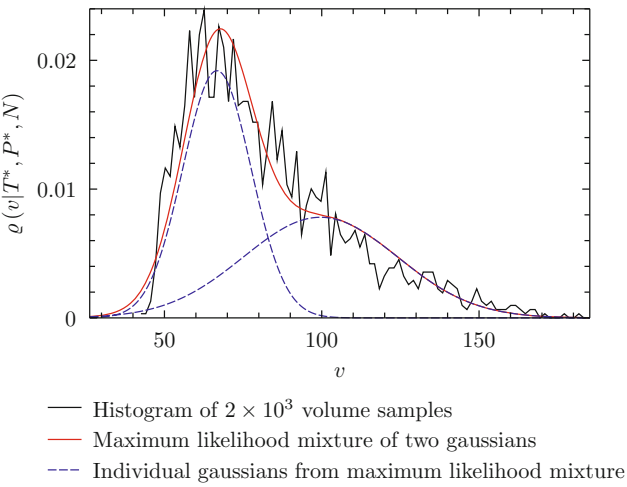

Fig. 9.2 Lennard-Jonesium periodic simulation of 64 atoms in the supercritical phase. $N = 64$, $T^* = 1.379$, $P^* = 1.758 \times 10^{-1}$. Using the parameters described in Sect. 9.2.2, prob $(M = 2|\{D_k\}) = 2.7 \times 10^{102}$ prob $(M = 1|\{D_k\})$. Although far from being a Gaussian distribution, there is clearly one mode

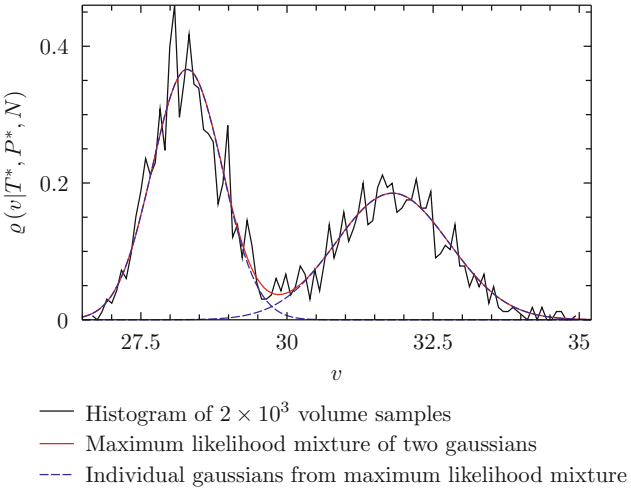

Fig. 9.3 Lennard-Jonesium periodic simulation of 64 atoms close to melting. $N = 64$, $T^* = 7.325 \times 10^{-1}$, $P^* = 1.758 \times 10^{-1}$. Using the parameters described in Sect. 9.2.2, prob $(M = 2|\{D_k\}) = 1.3 \times 10^{252}$ prob $(M = 1|\{D_k\})$. Here, close to the melting temperature, there are two clear modes

9.2.4 Calculating the Mean of the Dominant Partition

Let us begin by considering the mean of the dominant mode for a single mixture. Having partitioned the Gaussians as described in Sect. 9.2.3, we have a partition integer t_i associated with each Gaussian i. The probability mass of a partition is the sum of the amplitudes from its constituent Gaussians

$$\text{prob}\,(t_i) = \sum_{j=1}^{M} \delta_{t_i t_j} A_j . \tag{9.16}$$

In this way the partition with largest weight (the dominant partition) is easily identified. Let's call that partition integer t_M. The mean of partition t_M is

$$\mu\,(t_M) = \sum_{j=1}^{M} \delta_{t_M t_j} A_j \mu_j . \tag{9.17}$$

The result of our inference is therefore

$$\mu = \sum_{M=1}^{M_{\max}} \left[\text{prob}\,(M|\{D_k\}) \sum_{j=1}^{M} \delta_{t_M t_j} A_j \mu_j \right] . \tag{9.18}$$

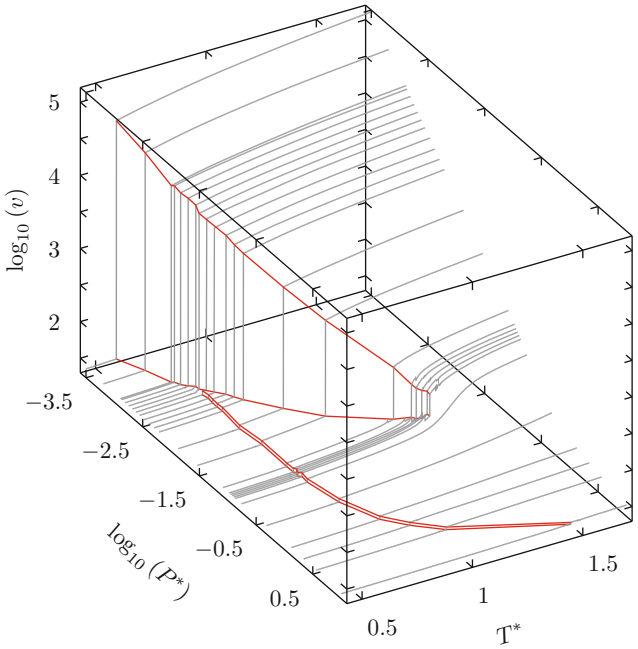

Fig. 9.4 Equation of state for Lennard-Jonesium. Axes: $T^* = {}^{k_B T}/_{\epsilon}$; $P^* = {}^{P\sigma^3}/_{\epsilon}$; $v = {}^{V}/_{N}$. These calculations were performed using 64 atoms, $r_c = 3\sigma$, and a mean-field long-range correction, as in Eq. 8.29

9.2.5 Equation of State for Lennard-Jonesium

The algorithm described in this chapter was used to calculate the equation of state of simple Lennard-Jonesium. The nested sampling calculations were exactly those described in Sect. 8.8. Since we do not have the exact pressure of the triple point in any of those calculations, no more than two partitions were thought to be required, and $M_{\max} = 2$ was used throughout. The results of these calculations are shown in Fig. 9.4.

9.3 Further Work

From the equation of state we could in principle obtain the $P - V$ phase diagram. It would be interesting to compare the results we obtain for that diagram with the literature values.

The algorithm we describe fits the volume distribution with a number of Gaussians, M. In the example calculation, a maximum of $M = 2$ Gaussians were used. Where the volume distribution is extremely non-Gaussian, the procedure produced

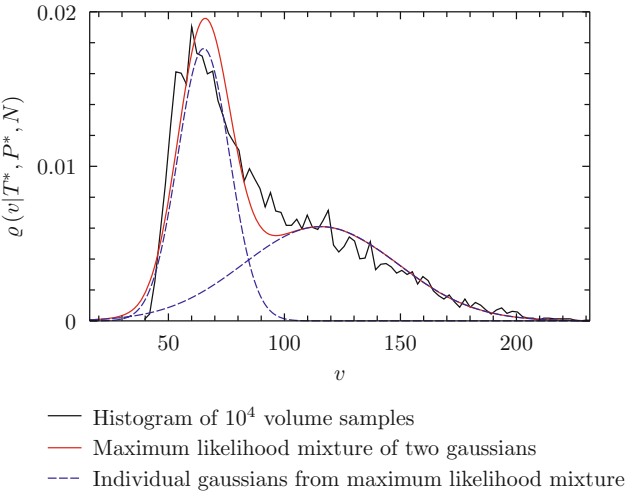

Fig. 9.5 Extremely non-Gaussian $\varrho(V)$ in the supercritical region. $N = 64$, $T^* = 1.327$, $\log_{10} P^* = -3.274$. We consider the underlying distribution for volume to have a single mode. With only two Gaussians the partitioning algorithm described in the text identifies two separate modes. A better fit would almost certainly obtained by using a Gamma mixture model

undesirable results with so few Gaussians. This occurs in the supercritical region, as illustrated in Fig. 9.5. There are two approaches that we could use: either we fit the distribution using many more Gaussians, or better still, we use a Gamma mixture model. Maximum likelihood estimation of Gamma mixture models is described in [5], and in [6] for a single gamma distribution. This ought to lead to a more automated means of resolving the critical point.

References

1. J. Skilling, Nested sampling. AIP Conf. Proc. **735**, 395 (2004)
2. J. Skilling, Nested sampling for general Bayesian computation. Bayesian Anal. **1**, 833 (2006)
3. D. MacKay, *Information Theory Inference and Learning Algorithms* (Cambridge University Press, 2003)
4. Wolfram Research, Inc., *Mathematica*, version 10.0 (Champaign, IL, 2014)
5. J. Almhana, Z. Liu, V. Choulakian, R. McGorman, A recursive algorithm for gamma mixture models, in *IEEE International Conference on Communications, 2006. ICC'06*, vol. 1 (IEEE, 2006), p.197
6. T.P. Minka, Estimating a gamma distribution. Technical Report (Microsoft Research, Cambridge, UK, 2002)

Chapter 10
Parallelising Nested Sampling

10.1 Introduction

We saw in Sect. 8.11 that a nested sampling calculation can require billions of energy evaluations. If we want to apply nested sampling to atomic potentials that are slower to evaluate than Lennard-Jonesium then it will certainly be necessary to parallelise the algorithm.

In this section we describe two approaches to parallelising nested sampling. The first [1, 2] is to parallelise over the number of iterations performed, by discarding $\mathcal{P}$ configurations at each iteration, rather than just one. Each iteration $\mathcal{P}$ samples are cloned and decorrelated, using $\mathcal{P}$ computer processors in parallel. In this way the calculation proceeds in larger steps through χ and takes less time to run.

The second approach, developed by Lívia Bartók-Pártay and myself, parallelises within the individual iterations, by parallelising over the walk length L. We detail an efficient algorithm for performing our parallelisation scheme, and evaluate the speedup, finding it to be almost perfect.

10.2 Parallelising over the Number of Iterations

Unlike standard nested sampling, where E_{lim} is updated to the highest energy in our sample set at each iteration, here it is updated to the $\mathcal{P}^{\text{th}}$ highest energy. We then record the $\mathcal{P}$ highest energies and their configurations with appropriate weightings [2]. Finally, $\mathcal{P}$ other configurations are chosen at random from our remaining samples, and the clone and decorrelate algorithm 7.3 is applied to those $\mathcal{P}$ clones in parallel.

Recall that in Sect. 7.5 we saw that for standard nested sampling, after $\mathcal{K}$ iterations $\log \chi$ is reduced from $\log \chi_i$ to $\log \chi_{i+\mathcal{K}} = \log \chi_i - 1$ with a standard deviation $\frac{1}{\sqrt{\mathcal{K}}}$. We will now calculate the standard deviation for the same reduction $\log \chi \leftarrow \log \chi - 1$, when taking steps of $\mathcal{P}$ configurations as described above.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_10

The pdf for $t = \frac{\zeta_i}{\zeta_{i-1}}$, when our live set includes $\mathcal{K}$ configurations overall is [3]

$$\varrho\left(t|\mathcal{P},\mathcal{K}\right) = \frac{\mathcal{K}!}{(\mathcal{P}-1)!\,(\mathcal{K}-\mathcal{P})!} t^{\mathcal{K}-\mathcal{P}}\left(1-t\right)^{\mathcal{P}-1}. \tag{10.1}$$

Evaluating the mean, and variance of $\log t$ we find [4]

$$\langle \log t \rangle = -\left[\sum_{i=\mathcal{K}-\mathcal{P}+1}^{\mathcal{K}} \frac{1}{i}\right] \tag{10.2}$$

$$\mathrm{Var}\left(\log t\right) = \left.\frac{d^{(\mathcal{K}-\mathcal{P}+1)}\Gamma(z)}{dz^{(\mathcal{K}-\mathcal{P}+1)}}\right|_{z=1} - \left.\frac{d^{(\mathcal{K}+1)}\Gamma(z)}{dz^{(\mathcal{K}+1)}}\right|_{z=1} \tag{10.3}$$

When parallelising over the number of iterations, Y_{lim} is updated to the lowest of the $\mathcal{P}$ highest enthalpies in our sample set, and the configuration space volume contained by the updated Y_{lim} is $\chi_i \approx \chi_0[(\mathcal{K}-\mathcal{P}+1)/(\mathcal{K}+1)]^i$, where i is the NS iteration number. For $\mathcal{P} > 1$ it is also possible to give analytic estimates of the configuration space volumes contained by the $\mathcal{P}-1$ higher enthalpy values between $Y_{\mathrm{lim}}^{(i-1)}$ and $Y_{\mathrm{lim}}^{(i)}$ [2]. Thus one may consider the configurational entropy contained at fractional numbers of NS enthalpy levels.

After a number of enthalpy levels

$$n_\Delta = \left(\sum_{i=\mathcal{K}-\mathcal{P}+1}^{k} \frac{1}{i}\right)^{-1} \tag{10.4}$$

the expectation of the logarithm of the configuration space enclosed by Y_{lim} decreases by 1:

$$\langle \log \chi_i - \log \chi_{i+n_\Delta} \rangle = -1. \tag{10.5}$$

If we assume that it is possible to draw exact random samples in each iteration, then it can be shown that, after the same number of enthalpy levels n_Δ, the variance of $\Delta \log \chi = \log \chi_i - \log \chi_{i+n_\Delta}$ is given by

$$\mathrm{Var}(\Delta \log \chi) = \left.\frac{d^{(\mathcal{K}-\mathcal{P}+1)}\Gamma(z)}{dz^{(\mathcal{K}-\mathcal{P}+1)}}\right|_{z=1} - \left.\frac{d^{(\mathcal{K}+1)}\Gamma(z)}{dz^{(\mathcal{K}+1)}}\right|_{z=1} \tag{10.6}$$

where $\Gamma(z)$ is the gamma function. The standard deviation, $\left[\mathrm{Var}(\Delta \log \chi)\right]^{\frac{1}{2}}$, represents the rate at which uncertainty in $\log \chi$ accumulates during a nested sampling calculation. For a serial calculation ($\mathcal{P} = 1$), $\left[\mathrm{Var}(\Delta \log \chi)\right]^{\frac{1}{2}} = \frac{1}{\sqrt{\mathcal{K}}}$.

Figure 10.1 shows how the ratio of $\left[\mathrm{Var}(\Delta \log \chi)\right]^{\frac{1}{2}}$ for parallel and serial NS, $R = [\mathrm{Var}(\Delta \log \chi)]^{\frac{1}{2}} \div \frac{1}{\sqrt{\mathcal{K}}}$, depends on $\frac{\mathcal{P}}{\mathcal{K}}$. R represents the relative rate at which uncertainty in $\log \chi$ accumulates during parallel and serial calculations. One

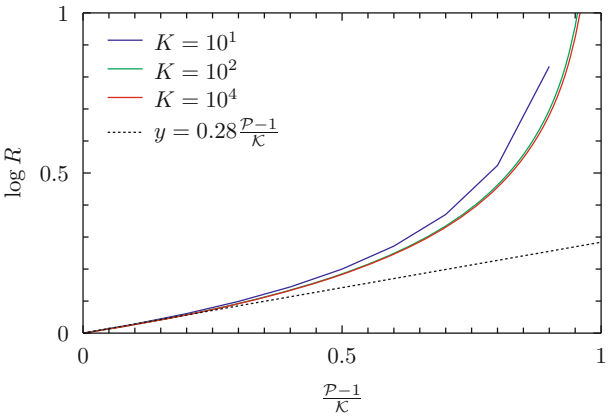

Fig. 10.1 Ratio of the rates, R, at which uncertainty in $\log \chi$ accumulates for runs parallelised by removing and walking $\mathcal{P} > 1$ as compared to serial run with $\mathcal{P} = 1$, as a function of scaled number of configurations removed $\frac{\mathcal{P}-1}{\mathcal{K}}$, for several values of $\mathcal{K}$. Dashed line indicates linear trend for $\mathcal{P} - 1 \ll \mathcal{K}$

can see that $\log R$ converges for $\mathcal{K} > 10^2$, and for $\frac{\mathcal{P}}{\mathcal{K}} \leq 0.25$, $\left[\mathrm{Var}(\Delta \log \chi)\right]^{\frac{1}{2}} \sim \exp\left(0.28\frac{\mathcal{P}-1}{\mathcal{K}}\right)\frac{1}{\sqrt{\mathcal{K}}}$. For larger $\frac{\mathcal{P}-1}{\mathcal{K}}$, R increases more rapidly.

10.3 Parallelising Within Each Iteration

Instead of parallelising over the number of iterations, we propose a method of parallelising within each iteration, with the aim of avoiding the cumulative error shown in Fig. 10.1.

In the serial algorithm, each new clone is decorrelated using a trajectory of walk length L. In our parallelised algorithm, new clones are decorrelated using a trajectory of walk length $\frac{L}{\mathcal{P}}$, and $(\mathcal{P} - 1)$ other random configurations are also (independently) propagated through $\frac{L}{\mathcal{P}}$ steps. Thus, rather than propagating a single configuration for L steps, we propagate $\mathcal{P}$ configurations for $\frac{L}{\mathcal{P}}$ steps.

It is possible to derive the variance of the total number of MC steps that is applied to a sample between when it is cloned by copying, and when it is eventually written as output, L'.

$$\mathrm{Var}(L') = L^2 \frac{\mathcal{P} - 1}{\mathcal{P}} \tag{10.7}$$

While unlike the case of parallelising over the number of iterations, there are no analytic results for the error due to the variability of L', several observations can be made. One is that the square root of the variance of L' (except for the serial case of $\mathcal{P} = 1$) is almost as large as L itself. Another is that the scaling of $\mathrm{Var}(L')$ with the

number of extra parallel tasks $\mathcal{P} - 1$ is polynomial, unlike the exponential scaling of the error R with extra tasks for parallelising over the number of iterations. It is unclear, however, how this variability in walk length will affect the error in the results, although the empirical observation is that this effect does not appear to be strong. Nevertheless, the serial case results in the lowest uncertainty in estimates of configuration space volumes.

10.3.1 Details of Our Implementation

The MCMC strategy described in Sect. 8.5 makes abundant use of single atom moves. When computing the change in energy after such a move it is often not necessary to recompute the interaction energies between all atoms. Rather, most of the interaction energies may remain unchanged, such that it is only necessary to compute a few terms. In this case it is efficient to store the individual interaction energies, updating only a few at a time. However, the number of such interaction energies can quickly become very large, and sending large arrays between processors using Message Passing Interface (MPI) can be slow, depending on the network connecting the processors. For this reason we developed an implementation with minimal use of MPI.

In our implementation, each processor stores $\frac{\mathcal{K}}{\mathcal{P}}$ configurations locally. This number of local configurations should be the same for each processor. At the end of each iteration, each processor identifies the local configuration with highest energy, and a single `MPI_ALLREDUCE` call is made to identify which processor stores the highest energy value overall. Next, a specific processor sends a cloned configuration to the processor storing the maximum energy. The "sending" processor is chosen at random, and picks the configuration to send at random from the set of configurations it stores. The configuration received is taken to be the "cloned" configuration, and will be propagated by the receiving processor, using $\frac{L}{\mathcal{P}}$ steps. Each of the other processors (including the "sending" processor) picks a configuration at random from its local set, and propagates that configuration through $\frac{L}{\mathcal{P}}$ steps.

I performed 8 sets of 16 simulations of 64 Lennard-Jonesium particles, at a reduced pressure $\log_{10}\left(P^* \equiv \frac{P\sigma^3}{\epsilon}\right) = -1.194$, with $\mathcal{K} = 640$ and $L = 4160$. Each set of 16 calculations was performed with a different value of $\mathcal{P}$. The mean run time τ in seconds for these calculations is shown in Fig. 10.2. I fitted the average run times with a function $\tau = A \times \mathcal{P}^B + C$, and found the fixed run time of the algorithm C to be less than 1% of the serial run time A, even for Lennard-Jonesium. The exponent B was found to be $B = -0.9799 \pm 0.0050$, which is extremely close to perfect parallelisation.

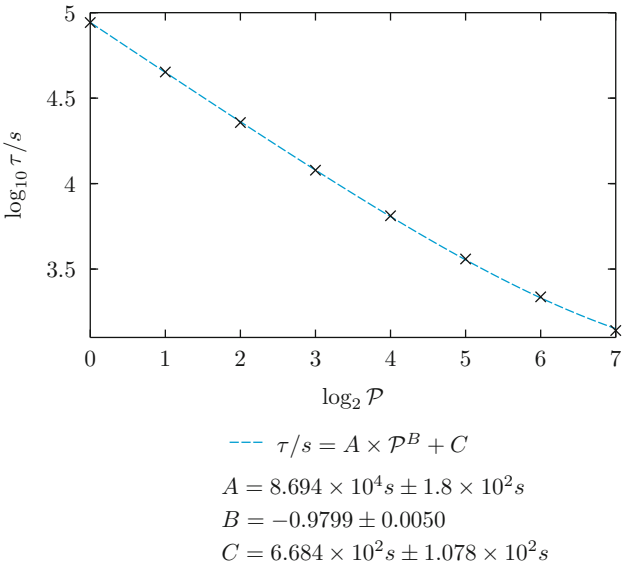

Fig. 10.2 Parallelisation by walk length leads to almost perfect speed up. That mean fixed run time of the algorithm C was found to be less than 1% of the serial run time A. This is particularly surprising for Lennard-Jonesium where a single energy evaluation is almost as quick as it could possibly be. The exponent B was found to be $B = -0.9799 \pm 0.0050$, which is extremely close to perfect parallelisation

10.4 Summary

We assessed the performance of two parallelisation schemes for nested sampling: parallelisation over the number of iterations and parallelisation over the walk length, within each iteration.

We found that parallelising over the number of iterations increases the rate at which error is introduced to our estimate of $\chi(U)$. This error rate is approximately exponential in $\frac{\mathcal{P}-1}{\mathcal{K}}$ for $\frac{\mathcal{P}-1}{\mathcal{K}} \leq 0.25$, and much greater for large values.

To address this, we proposed a method of parallelising over the walk length, within each iteration. In this algorithm $\mathcal{P}$ processors decorrelate $\mathcal{P}$ configurations, each for a $\frac{L}{\mathcal{P}}$ steps. The cloned configuration is always one of the $\mathcal{P}$ configurations to be decorrelated. We found the variance of the total number of MC steps that is applied to a sample between when it is copied by cloning, and when it is eventually written as output, L', to be $\mathrm{Var}(L') = \langle L \rangle^2 \frac{\mathcal{P}-1}{\mathcal{P}}$. For $\mathcal{P} \gg 1$, the standard deviation of L' is almost as large as the total walk-length L. Nevertheless, we observe empirically that this variability in walk length does not strongly affect the error in the results.

Finally, we gave a detailed account of our implementation, and measured its performance. We performed sixteen calculations at each of $\mathcal{P} \in \left\{2^0, 2^1, 2^2, \ldots, 2^7\right\}$, with a total walk length $L = 4160$ and $\mathcal{K} = 640$. Fitting the mean run times with

a function $\tau = A \times \mathcal{P}^B + C$, we found that the fixed calculation costs C were less than 1% of the serial run time, $A = 8.694 \times 10^4\text{s} \pm 1.8 \times 10^2\text{s}$. This is surprising for Lennard-Jonesium, where one might have expected the fixed costs to be a larger proportion of the run time, since the energy evaluations are very fast. We found the exponent to be $B = -0.9799 \pm 0.0050$, which is extremely close to perfect parallelisation.

10.5 Further Work

It would be enlightening to make a comparison of the two parallelisation schemes, similar to Fig. 10.1. This would require many calculations using the "parallelisation within each iteration" method, at each number of processors, up to at least $\frac{\mathcal{P}-1}{\mathcal{K}} = 0.5$. For each number of processors, the error, relative to serial calculations, would be obtained. These could then be compared to the theoretical curve for the relative error in the method of "parallelisation over the number of iterations". Repeating this process at a range of pressures would constitute a fair comparison of the two methods.

References

1. N.S. Burko, C. Várnai, S.A. Wells, D.L. Wild, Exploring the energy landscapes of protein folding simulations with Bayesian computation. Biophys. J. **102**, 878 (2012)
2. S. Martiniani, J.D. Stevenson, D.J. Wales, D. Frenkel, Superposition enhanced nested sampling. Phys. Rev. X **4**, 031034 (2014)
3. S. Martiniani, Masters thesis (unpublished), Cavendish Laboratory, University of Cambridge, 2013
4. Wolfram Research, Inc., *Mathematica*, version 10.0 (Champaign, IL, 2014)

Part III
Molecular Dynamics Nested Sampling

Chapter 11
Introduction

Up to this point, we have described how nested sampling [1, 2] can be performed using a random walk to decorrelate samples in the "clone and decorrelate" algorithm. 7.3. In this section we explore the use of Hamiltonian dynamics as an alternative means of decorrelating cloned configurations. Using Hamiltonian dynamics to produce independent samples is far from a new idea [3–9]. Indeed, a method of using Hamiltonian dynamics for this purpose in nested sampling has already been developed by John Skilling [10, 11] and is already being applied in acoustics [12]. We return to this in Chapter. 12.

The problem we are interested in is as follows: the need to quickly explore atomic configuration spaces in condensed phases to produce configurations according to a known distribution. For this problem, random moves of the kind described in Chap. 8 are poorly suited. In condensed phases atoms are crowded together. If we choose an atom and attempt to displace it in a random direction, it will almost always come into close proximity with a neighbour. The interaction between atoms is strongly repulsive at close proximity, so this proposed configuration will have a very high potential energy, and be rejected. To achieve rapid decorrelation, we require a success rate of approximately 0.23 [9]. In a condensed phase, this requires a very small step size for atomic displacements.

The motivation for using Hamiltonian dynamics to update the configuration is that, rather than moving independently, the atoms should flow cooperatively with no sudden spikes in the potential energy. Such a motion should allow the system to decorrelate more quickly. In this section we describe a version of the "Hamiltonian Monte Carlo" (HMC) algorithm [9], adapted to nested sampling for atomistic simulation.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_11

11.1 Hamiltonian Monte Carlo for the Canonical Distribution

In Hamiltonian Monte Carlo we begin by considering a phase space $\{\mathbf{q}, \mathbf{p}\}$, rather than a configuration space $\{\mathbf{q}\}$ alone. In the traditional case, Hamiltonian Monte Carlo is used to draw samples from the canonical distribution.

11.1.1 The Canonical Distribution, Revisited

In the canonical distribution, $\mathbf{q}$ and $\mathbf{p}$ are distributed according to the joint distribution

$$\varrho(\mathbf{q}, \mathbf{p}) = \frac{1}{N!h^{3N}} \frac{\mathrm{e}^{-\beta E(\mathbf{q},\mathbf{p})}}{Z(N, \beta, V)}. \tag{4.38 revisited}$$

Here $E(\mathbf{q}, \mathbf{p})$ is the Hamiltonian for our system. The Hamiltonian typically separates into potential and kinetic energy terms

$$E(\mathbf{q}, \mathbf{p}) = K(\mathbf{p}) + U(\mathbf{q}). \tag{3.1 revisited}$$

This in turn makes $\mathbf{q}$ and $\mathbf{p}$ statistically independent, each following separate canonical distributions

$$\varrho(\mathbf{q}, \mathbf{p}) = \frac{1}{N!h^{3N}Z} \exp\left[-\beta U(\mathbf{q})\right] \exp\left[-\beta K(\mathbf{p})\right]. \tag{11.1}$$

11.1.2 Why Hamiltonian Dynamics?

What is it about Hamiltonian dynamics that make it suitable for sampling? The answer has three parts.

1. The dynamics is reversible. This is useful for proving that transitions of the Markov chain leave the target distribution invariant.
2. Phase space volume preservation. The volume of a phase space element $\mathrm{d}\mathbf{q} \wedge \mathrm{d}\mathbf{p}$ is conserved by Hamiltonian dynamics. The significance of this is that we do not have to account for any non-symmetrical change in volume when accepting or rejecting proposals in Metropolis updates.
3. Conservation of the Hamiltonian. Exact conservation of the Hamiltonian guarantees the acceptance of new proposals. Approximate conservation of the Hamiltonian makes the probability of acceptance for new proposals close to unity.

Any dynamics that meet these criteria is a fair alternative to Hamiltonian dynamics.

In practice, exactly integrating a continuous dynamics is often not possible, and the dynamics must be integrated approximately by taking discrete time steps. The crucial quality of any integrator therefore is that it has these desirable qualities: reversibility, phase space preservation, and conservation of the Hamiltonian. Conservation of the Hamiltonian can not be guaranteed exactly, so it is important that the rate at which error is accrued be low and that it go to zero when the step size is taken to the infinitesimal limit.

11.1.3 Hamiltonian Monte Carlo Sampling Step

Hamiltonian Monte Carlo samples from the joint distribution prob $(\mathbf{q}, \mathbf{p})$. A Hamiltonian Monte Carlo step has two parts.

1. Hamiltonian update
   (a) Propagate $(\mathbf{q}, \mathbf{p})$ by L steps of leapfrog integration with step size ϵ. This takes us to state $\left(\mathbf{q}', \mathbf{p}'\right)$.
   (b) Accept or reject $\left(\mathbf{q}', \mathbf{p}'\right)$ according to the Metropolis-Hastings rule

$$\text{acc}\left[(\mathbf{q}, \mathbf{p}) \rightarrow \left(\mathbf{q}', \mathbf{p}'\right)\right] = \min\left(1, \mathrm{e}^{-\beta\left[U\left(\mathbf{q}'\right) - U(\mathbf{q}) + K\left(\mathbf{p}'\right) - K(\mathbf{p})\right]}\right). \quad (11.2)$$

   If the leapfrog integration were perfect, with infinitesimal step size, then the energy of the states $(\mathbf{q}, \mathbf{p})$ and $\left(\mathbf{q}', \mathbf{p}'\right)$ would be identical, and every step would be accepted. Finite step size allows the total energy to change slightly in the course of a trajectory. This Metropolis-Hastings update ensures that the proposed state $\left(\mathbf{q}', \mathbf{p}'\right)$ is visited according to the canonical distribution (11.1). If the proposal is rejected, we return to the original state $(\mathbf{q}, \mathbf{p})$ but we do not reset the clock. Let our final state be $(\mathbf{q}^*, \mathbf{p}^*)$.

2. Select a fresh momentum $\mathbf{p}^{**}$ according to the correct joint distribution

$$\text{prob}\left(\mathbf{q}^*, \mathbf{p}^{**}\right) = \text{prob}\left(\mathbf{p}^{**} | \mathbf{q}^*\right) \times \text{prob}\left(\mathbf{q}^*\right). \quad (11.3)$$

   By definition, $(\mathbf{q}^*, \mathbf{p}^{**})$ is a correct sample from the target joint distribution. In cases where $\mathbf{q}$ and $\mathbf{p}$ are statistically independent, such as the canonical distribution, $\mathbf{p}^{**}$ is chosen from its independent distribution.

The fresh state is $(\mathbf{q}^*, \mathbf{p}^{**})$. We have seen that $\mathbf{q}$ and $\mathbf{p}$ are statistically independent in the canonical distribution (11.1). Therefore, for the purposes of sampling quantities in configuration space, we can associate the full mass of each sample with its configuration $\mathbf{q}$.

References

1. J. Skilling, Nested sampling. AIP Conf. Proc. **735**, 395 (2004)
2. J. Skilling, Nested sampling for general Bayesian computation. Bayesian Anal. **1**, 833 (2006)
3. S. Duane, A.D. Kennedy, B.J. Pendleton, D. Roweth, Hybrid Monte Carlo. Phys Lett. B **195**, 216 (1987)
4. E. Fermi, J.R. Pasta, S.M. Ulam, Studies of nonlinear problems. LASL Report LA-1940 (1955)
5. B.J. Alder, T. Wainwright, Studies in molecular dynamics. I. General method. J. Chem. Phys. **31**, 459 (1959)
6. A. Rahman, Correlations in the motion of atoms in liquid argon. Phys. Rev. **136**, A405 (1964)
7. D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Computational science series (Elsevier Science, 2001)
8. D. MacKay, *Information Theory Inference and Learning Algorithms* (Cambridge University Press, 2003)
9. R.M. Neal, MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo* (CRC Press, New York, NY, 2011), p. 113
10. J. Skilling, Bayesian computation in big spaces-nested sampling and Galilean Monte Carlo, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 31st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 1443 (AIP Publishing, 2012), p. 145
11. F. Feroz, J. Skilling, Exploring multi-modal distributions with nested sampling (2013), arXiv:1312.5638
12. P. Goggans, R.W. Henderson, N. Xiang, Using nested sampling with Galilean Monte Carlo for model comparison problems in acoustics. Proc. Meet. Acoust. **19**, 055089 (2013)

Chapter 12
Hamiltonian Monte Carlo for Nested Sampling

In order to use Hamiltonian Monte Carlo in Nested Sampling, we expand our configuration space to include momenta $\mathbf{p}$ conjugate to the coordinates $\mathbf{q}$. Hamiltonian Monte Carlo draws new samples from the joint distribution prob $(\mathbf{q}, \mathbf{p})$. For nested sampling we would ideally like to draw configuration samples alone. To this end we might seek to construct a separable joint distribution prob $(\mathbf{q}, \mathbf{p}) =$ prob $(\mathbf{q})$ prob $(\mathbf{p})$, so that we can map the full mass of each sample to $\mathbf{q}$, as in the canonical distribution. This is the approach of John Skilling, in "Galilean Monte Carlo" [1, 2], which we now describe.

12.1 Galilean Monte Carlo

In Galilean Monte Carlo, we define a second potential energy function

$$U'\left(\mathbf{q}|U_{\mathrm{lim}}\right) = \begin{cases} 0, & U\left(\mathbf{q}\right) \leq U_{\mathrm{lim}} \\ \infty, & \text{Elsewhere} \end{cases} \tag{12.1}$$

The canonical distribution (11.1) for $U'\left(\mathbf{q}|U_{\mathrm{lim}}\right)$ is just

$$\varrho\left(\mathbf{q}\right)\varrho\left(\mathbf{p}\right) = \begin{cases} \frac{1}{N!h^{3N}Z}\exp\left[-\beta K\left(\mathbf{p}\right)\right], & U\left(\mathbf{q}\right) \leq U_{\mathrm{lim}} \\ 0, & \text{Elsewhere.} \end{cases} \tag{12.2}$$

This distribution over $\mathbf{q}$ (12.2) is exactly what we require for nested sampling: namely it is constant for $U\left(\mathbf{q}\right) \leq U_{\mathrm{lim}}$ and zero outside this region. Nested sampling can therefore be performed using the potential energies alone. Hamiltonian dynamics on

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_12

the potential energy function $U'(\mathbf{q}|U_{\text{lim}})$ corresponds to motion in a straight line with elastic reflections off the walls of the "well". The dynamics proposed by Skilling is as follows.

When no reflection occurs, position is updated using steps of constant time step ϵ.

$$\left[\mathbf{q}, \mathbf{p}\right] \rightarrow \left[\mathbf{q}' = \mathbf{q} + \epsilon\mathbf{p}, \mathbf{p}\right] \tag{12.3}$$

When the proposed coordinates $\mathbf{q}'$ enter the region of zero probability $U(\mathbf{q})' > U_{\text{lim}}$ a reflection must instead have occurred. Such (rejected) proposals are replaced by a reflection step

$$\left[\mathbf{q}, \mathbf{p}\right] \rightarrow \left[\mathbf{q}'' = \mathbf{q} + \mathbf{p} + \mathbf{p}', \mathbf{p}' = \mathbf{p} - 2\widehat{\mathbf{n}}\left(\widehat{\mathbf{n}} \cdot \mathbf{p}\right)\right]. \tag{12.4}$$

The unit vector $\widehat{\mathbf{n}}$ is the normalised gradient of $U(\mathbf{q})$ at the (rejected) point $\mathbf{q}' = \mathbf{q} + \epsilon\mathbf{p}$. If this reflection takes the configuration inside the good region $\left(U\left(\mathbf{q}''\right) \leq U_{\text{lim}}\right)$ then this update is accepted. If the reflection proposal $\mathbf{q}''$ is also outside of the allowed region then the only plausible move satisfying detailed balance is for the system to double back on its trajectory

$$\left[\mathbf{q}, \mathbf{p}\right] \rightarrow \left[\mathbf{q}, \mathbf{p}' = -\mathbf{p}\right]. \tag{12.5}$$

In the condensed phases of atomistic systems most atomic displacements lead to collisions and there are relatively few "soft" degrees of freedom along which appreciable displacement can be made. In order to effectively bias exploration along those soft degrees of freedom we would like to make use of gradient information (atomic forces) at every timestep, while Galilean Monte Carlo only uses gradient information when we leave the allowed region. This leads us to "Molecular dynamics nested sampling" (MDNS).

12.2 Molecular Dynamics Nested Sampling in the Total Hamiltonian

Here, we relax the desirable condition that $\mathbf{q}$ and $\mathbf{p}$ be statistically independent, so as to gain the fast decorrelation of Hamiltonian Monte Carlo. We seek to distribute samples according to the joint distribution

$$\text{prob}\left(\mathbf{q}, \mathbf{p}|E_{\text{lim}}\right) = \begin{cases} \frac{1}{\Phi(E_{\text{lim}})}, & E\left(\mathbf{q}, \mathbf{p}\right) \leq E_{\text{lim}} \\ 0, & \text{Elsewhere} \end{cases} \tag{12.6}$$

which is suitable for performing nested sampling in the total Hamiltonian $E(\mathbf{q}, \mathbf{p})$. The integrated density of states $\Phi\left(E_{\text{lim}}\right)$ is taken exactly from the microcanonical ensemble (4.3)

$$\Phi(N, E, V) = \int_{-\infty}^{E} \Omega\left(N, E', V\right) \mathrm{d}E'. \tag{4.3 revisited}$$

Hamiltonian Monte Carlo is performed largely as described in Sect. (11.3). Since we seek the Hamiltonian bounded uniform distribution (12.6), the Metropolis-Hastings rejection (11.2) is replaced by the rule

$$\mathrm{acc}\left[(\mathbf{q}, \mathbf{p}) \to \left(\mathbf{q}', \mathbf{p}'\right)\right] = \begin{cases} 1, & E\left(\mathbf{q}', \mathbf{p}'\right) \leq E_{\mathrm{lim}} \\ 0, & \text{Elsewhere} \end{cases} \tag{12.7}$$

The position $\mathbf{q}$ and momenta $\mathbf{p}$ are not statistically independent. Therefore, when generating fresh momenta $\mathbf{p}$ we choose a momentum vector uniformly from the region

$$K(\mathbf{p}) \leq E_{\mathrm{lim}} - U(\mathbf{q}). \tag{12.8}$$

Generating a fresh momentum vector for a system of N particles can be performed as follows.

1. Select a point from the surface of a unit sphere in $3N$-dimensions. This can be performed using the standard method [3, 4] of generating a vector of $3N$ Gaussian distributed random numbers,[1] then normalising the vector. Let us call this final vector $\widehat{\mathbf{r}}$.
2. Generate a radius r according to the polynomial distribution

$$\mathrm{prob}(r) = \begin{cases} 3Nr^{3N-1}, & r \leq 1 \\ 0, & \text{Elsewhere.} \end{cases} \tag{12.9}$$

   This can be performed using Algorithm 12.1.
3. The kinetic energy has the form

$$K(\mathbf{p}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i}. \tag{3.2 revisited}$$

   In this case, the final momentum is given by

$$\mathbf{p}^{**} = r\mathbf{S}\widehat{\mathbf{r}} \tag{12.10}$$

$$S_{ij} = \delta_{ij}\left(2m_i\left[E_{\mathrm{lim}} - U(\mathbf{q})\right]\right)^{\frac{1}{2}}. \tag{12.11}$$

[1] These Gaussian distributed random numbers should have mean 0.

```
subroutine polynomial_rand
                                ! Generates a random number
                                ! x distributed as
                                ! prob(x) = x^expo : 0 ≤ x ≤ maximum
x=ranf()**(1.0/(expo+1.0))
x=rand_out*maximum

end subroutine
```

Algorithm 12.1 Generating a polynomial distributed random variate with non-negative power, below a maximum value. Samples from the pdf (12.9). `ranf()` is a random number uniform in [0,1].

12.3 Nested Sampling in the Total Hamiltonian at Fixed *N* and *P*

As in Sect. 8, we will perform nested sampling at constant pressure. In this section we describe sampling under conditions of fixed pressure and fixed simulation cell shape.

It is common to take the "microscopic enthalpy" $H(\mathbf{q}, \mathbf{p}, P, V)$ to be that introduced by Hans C. Andersen in his famous paper [5]

$$H_{\mathrm{A}} = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{q}, V) + PV + \frac{p_V^2}{2W_{\mathrm{A}}}. \tag{12.12}$$

Here the traditional phase space $\{(\mathbf{q}, \mathbf{p})\}$ has been expanded to include volume as an additional coordinate with conjugate momentum p_V. The constant W_{A} has dimensions energy $\times$ time$^2 \div$ volume^2. Promoting V to a dynamic variable facilitates the design of molecular dynamics schemes at fixed pressure.

Several molecular dynamics schemes have been published in the literature for exploring the fixed isothermal—isobaric ensemble with a fixed shape simulation cell. All these schemes build upon the work of Andersen [5]. Here I use the MTK equations of motion [6] since they can be shown to give the correct pressure virial theorem and work virial [7]. Here, Martyna *et al.* consider a Hamiltonian of the same form

$$H_{\mathrm{MTK}} = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{q}, V) + PV + \frac{p_\epsilon^2}{2W_{\mathrm{MTK}}}. \tag{12.13}$$

However, unlike Andersen, Martyna et al. construct a non-Hamiltonian dynamics that nevertheless has the desirable qualities of Hamiltonian dynamics required for HMC. In this dynamics, volume is evolved through a new variable $\epsilon = \frac{1}{3}\log\left(\frac{V}{V_0}\right)$, where $\epsilon = \frac{p_\epsilon}{W_{\mathrm{MTK}}}$. This is further described in Sect. 12.3.1.

The partition function for conditions of fixed NPT and fixed simulation cell shape is

$$\Delta(N, P, \beta) = \frac{1}{N! h^{3N} V_0} \int_0^\infty \mathrm{d}\Phi \, \mathrm{e}^{-\beta H_{\mathrm{MTK}}(\Phi)} \div \int_0^\infty \mathrm{d}p_V \, \mathrm{e}^{-\frac{\beta p_\epsilon^2}{2W_{\mathrm{MTK}}}} \tag{12.14}$$

$$= \frac{1}{N! h^{3N} V_0} \left(\frac{2\beta}{\pi W_{\mathrm{MTK}}} \right)^{\frac{1}{2}} \int_0^\infty \mathrm{d}\Phi \, \mathrm{e}^{-\beta H_{\mathrm{MTK}}(\Phi)} \tag{12.15}$$

where the integrated density of states Φ, and density of states Ω are given by

$$\Phi(N, H, P) = \int_{-\infty}^{H} \Omega\left(N, H', P\right) \mathrm{d}H' \tag{4.62 revisited}$$

$$\Omega\left(N, H', P\right) = \frac{1}{N! h^{3N} V_0} \int_0^\infty \mathrm{d}p_\epsilon \int_0^\infty \mathrm{d}V \int \mathrm{d}^N \mathbf{p} \int_{D(V)} \mathrm{d}^N \mathbf{q} \, \delta\left(H' - H_{\mathrm{MTK}}\left[\mathbf{q}, V, \mathbf{p}, p_\epsilon; P\right]\right). \tag{12.16}$$

As usual in nested sampling, we require samples with enthalpies H distributed as the density of states (12.16) below some limit H_{lim}. This can be achieved by distributing the samples according to the joint distribution

$$\mathrm{prob}\left(\mathbf{q}, V, \mathbf{p}, p_V | H_{\mathrm{lim}}, P\right) = \begin{cases} \frac{1}{\Phi(H_{\mathrm{lim}})}, & H_{\mathrm{MTK}}\left(\mathbf{q}, V, \mathbf{p}, p_\epsilon, P\right) \leq H_{\mathrm{lim}} \\ 0, & \text{Elsewhere.} \end{cases} \tag{12.17}$$

12.3.1 MTK Dynamics

We follow the integration scheme described by Yu et al. in their paper "Measure-preserving integrators for molecular dynamics in the isothermal—isobaric ensemble derived from the Liouville operator" [8]. The MTK equations of motion [6] for the isoenthalpic-isobaric ensemble are[2]

[2]Particular care must be taken when evaluating the double summation in P_{int}. For each atom i we consider interactions with each atom $j \in \{i+1, \ldots, N\}$ and all periodic images of j. If the simulation cell is so small that self-images are to be considered, interactions between atom i and its own periodic images incur a factor of a half due to double counting.

$$
\begin{aligned}
\dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m_i} + \frac{p_\epsilon}{W_{\mathrm{MTK}}}\mathbf{q}_i \\
\dot{\mathbf{p}}_i &= -\frac{\partial U}{\partial \mathbf{q}_i} - \left(1 + \frac{1}{N}\right)\frac{p_\epsilon}{W_{\mathrm{MTK}}}\mathbf{p}_i \\
\dot{V} &= \frac{3V}{W_{\mathrm{MTK}}}p_\epsilon \\
\dot{p_\epsilon} &= 3V\,(P_{\mathrm{int}} - P) + \frac{1}{N}\sum_i \frac{\mathbf{p}_i^2}{m_i} \\
P_{\mathrm{int}} &= \frac{1}{3V}\left[\sum_i^N \frac{\mathbf{p}_i^2}{m_i} - \sum_{j>i}\sum_i^N \frac{\partial U}{\partial \mathbf{q}_{ij}}\cdot\mathbf{q}_{ij}\right].
\end{aligned}
\tag{12.18}
$$

The constant W_{MTK} has dimensions energy $\times$ time^2. The value of this constant is discussed in Sect. 12.3.2.

The MTK equations of motion (12.18) cannot be derived from the Hamiltonian (12.13) but nevertheless conserve the numerical value of H_{MTK}. They are also reversible and preserve the volume of a phase space element in the expanded space. These good qualities are properly maintained in the Liouville-operator based integrator of Yu et al. [8], where the global error in a long trajectory of M steps, each of length $\mathrm{d}t$, is of order $\mathrm{d}t^2$. The integrator is therefore suitable for use in Hamiltonian Monte Carlo, as explained in Sect. 11.2.

12.3.2 *Choice of W_{MTK}*

Any fixed choice of W_{MTK} will lead to resonances in the dynamics of the cell when the available kinetic energy is tuned to excite its motion. In order to perform sampling in a predictable way, the cell mass must be tuned according to the total kinetic energy. An optimal choice for W_{MTK} is proposed to be [9]

$$
W_{\mathrm{MTK}} = (3N + 1)\,k_{\mathrm{B}}T\,\tau_{\mathrm{b}}^2. \tag{12.19}
$$

Here τ_{b} is a free parameter, specifying the time scale of motions of the cell.

Changing W_{MTK} necessarily changes the joint distribution (12.17) and the density of states (12.16). One would think that this would destroy the validity of the calculation. In fact, the sole dependence on temperature (12.19) is a subtly chosen one. For any value of temperature, W_{MTK} takes a particular constant value. This ensures that for each temperature, high probability values of H are sampled with the same cell mass W_{MTK}. The correct heat capacity is recovered by applying Eqs. (8.27) and (8.28) using the total enthalpy values $\{H_i\}$ from MDNS in place of $\{Y_i\}$ from standard MCMC nested sampling, then subtracting k_{B} from the result.

We now turn our attention to updating W_{MTK} in practice. Since we are sampling from the density of states, without thermodynamic temperature, we substitute for T using the virial theorem

$$k_{\mathrm{B}}T = \frac{2\langle K\rangle}{3N+1} \tag{12.20}$$

$$\Rightarrow W_{\mathrm{MTK}} = 2\langle K\rangle\tau_{\mathrm{b}}^2 \tag{12.21}$$

The value of $\langle K\rangle$ was re-assessed, and W_{MTK} updated once every $\frac{\mathcal{K}}{2}$ iterations (where $\mathcal{K}$ is the number of independent copies of the system used in the NS calculation). The value of W_{MTK} typically decreases in the course of the calculation, as expected, since W_{MTK} has dimensions energy $\times$ time^2 and Φ_{lim} is reduced. When this happens, the kinetic energy term $\frac{p_\epsilon^2}{2W_{\mathrm{MTK}}}$ increases. At this point we re-scaled the momenta $\mathbf{p}$, p_ϵ for each system, to retain their enthalpy values.

12.3.3 Hamiltonian Monte Carlo Scheme

The motion of V and p_ϵ occurs on a much slower timescale than that of the atoms. We found that any randomisation of p_ϵ destroyed the decorrelation of the system. Therefore, only atomic velocities were randomised during the HMC update. For this reason, it was necessary to make use of a slightly more elaborate form of the HMC update [10]. The Hamiltonian Monte Carlo update used in this thesis is as follows

The initial state is $(\mathbf{q}, \mathbf{p}, V, p_\epsilon)$.

1. Hamiltonian update
   (a) Propagate $(\mathbf{q}, \mathbf{p}, V, p_\epsilon)$ for L steps of the integrator due to Yu et al. [8] with step size dt. This takes us to state $\left(\mathbf{q}^{(1)}, \mathbf{p}^{(1)}, V^{(1)}, p_\epsilon^{(1)}\right)$.
   (b) Negate all momenta.

$$\left(\mathbf{q}^{(1)}, \mathbf{p}^{(1)}, V^{(1)}, p_\epsilon^{(1)}\right) \rightarrow \left(\mathbf{q}^{(1)}, \mathbf{p}^{(2)} = -\mathbf{p}^{(1)}, V^{(1)}, p_\epsilon^{(2)} = -p_\epsilon^{(1)}\right)$$

   This step is important because it makes the deterministic update reversible, so that it satisfies detailed balance. This step was not included in Sect. (11.3) for two reasons: first because kinetic energy is an even function so negating $\mathbf{p}$ and p_ϵ does not affect the probability of accepting the move as given in Eq. (12.7), and second, because all momenta were discarded when choosing a fresh momentum $\mathbf{p}^{**}$ in the last part of the update. It is included here because we will only randomise the atomic momenta, $\mathbf{p}$. Negating the momenta is reversible, and since H_{MTK} is an even function of the momenta, this move leaves H_{MTK} unchanged.
   (c) Accept or reject $\left(\mathbf{q}^{(1)}, \mathbf{p}^{(2)}, V^{(1)}, p_\epsilon^{(2)}\right)$ with probability

$$\mathrm{acc}\left[(\mathbf{q},\mathbf{p},V,p_\epsilon)\to\left(\mathbf{q}^{(1)},\mathbf{p}^{(2)},V^{(1)},p_\epsilon^{(2)}\right)\right]=\begin{cases}1, & H_{\mathrm{MTK}}\left(\mathbf{q}^{(1)},\mathbf{p}^{(2)},V^{(1)},p_\epsilon^{(2)}\right)\le H_{\mathrm{lim}}\\ 0, & \text{Elsewhere}\end{cases} \tag{12.22}$$

The small error due to finite step size may cause the value of H_{MTK} to exceed H_{lim}, but there is no penalty for H_{MTK} decreasing or increasing by less than this amount. If the proposal is rejected, we return to the original state $(\mathbf{q},\mathbf{p},V,p_\epsilon)$ but we do not reset the clock. Let our final state be $\left(\mathbf{q}^{(3)},\mathbf{p}^{(3)},V^{(3)},p_\epsilon^{(3)}\right)$.

2. Negate the momenta once again.

$$\left(\mathbf{q}^{(3)},\mathbf{p}^{(3)},V^{(3)},p_\epsilon^{(3)}\right)\to\left(\mathbf{q}^{(3)},\mathbf{p}^{(4)}=-\mathbf{p}^{(3)},V^{(3)},p_\epsilon^{(4)}=-p_\epsilon^{(3)}\right)$$

If the proposal from (a) was accepted in (c), this negation positions the state at the end of the trajectory with momenta in the forward facing direction. If the proposal from (a) was rejected, this reversal positions the state at the initial point $(\mathbf{q},-\mathbf{p},V,-p_\epsilon)$, so that the trajectory will double back on itself.

3. Select fresh *atomic* momenta $\mathbf{p}^{(4)}$ according to the correct joint distribution

$$\begin{aligned}\mathrm{prob}\left(\mathbf{q}^{(3)},\mathbf{p}^{(4)},V^{(3)},p_\epsilon^{(3)}\right)&=\mathrm{prob}\left(\mathbf{p}^{(4)}|\mathbf{q}^{(3)},V^{(3)},p_\epsilon^{(3)}\right)\times\mathrm{prob}\left(\mathbf{q}^{(3)},V^{(3)},p_\epsilon^{(3)}\right)\\&=\mathrm{prob}\left(\mathbf{p}^{(4)}|\mathbf{q}^{(3)},V^{(3)},p_\epsilon^{(3)}\right)\times\mathrm{const.}\end{aligned} \tag{12.23}$$

This is performed exactly as described in Sect. 12.2.

The final state is $\left(\mathbf{q}^{(3)},\mathbf{p}^{(4)},V^{(3)},p_\epsilon^{(3)}\right)$.

12.3.4 Initialisation

As in nested sampling with MCMC, we impose a boundary on configuration space that is well inside the ideal gas region. Unlike the MCMC case, we did not find it convenient to impose a minimum density, and imposed a maximum enthalpy H_0 instead. A relatively system-independent value for H_0 is $N\times 2000\mathrm{eV}$.

The method we found convenient for generating $\mathcal{K}$ samples according to (12.17) with $H_{\mathrm{lim}}=H_0$ is as follows.

1. Estimate the ground state enthalpy of the potential, H_{min}. This can be performed for an atomic fluid by choosing any perfect close-packed configuration (FCC, for example) and calculating the minimum configurational enthalpy of that configuration for 100 values of the box side length, corresponding to atomic radii in the range (0.1Å, 10.0Å).
2. Set $V=0.4\times\frac{(H_0-H_{\mathrm{min}})}{P}$. This allows some interatomic interactions within the initial atomic configurations.

3. Atomic coordinates q_{ij} are univariate random numbers on $\left[0, V^{\frac{1}{3}}\right)$.
4. Minimise the configurational enthalpy of each configuration with respect to $\mathbf{q}$ to below $\frac{H_0 - H_{\min}}{2}$. We applied molecular dynamics to each configuration, starting from zero momenta, $\mathbf{p}$. Whenever the configurational enthalpy increased, the coordinates were reset by one step to their most recent minimum enthalpy value, and the momenta were reset to zero before the trajectory was allowed to continue. The time step dt was chosen so that it would take an average of 10^4 steps for an atom to traverse a distance equal to the mean distance between atoms, assuming that the total kinetic energy per atom is $\frac{H_0 - H_{\min}}{1000N}$. Occasionally a configuration became trapped, repeatedly taking failed steps. Those configurations were replaced at the end with other configurations that had not become trapped, chosen at random. When all configurations had $Y(\mathbf{q}, V, P) < \frac{H_0 - H_{\min}}{2}$ a further 10 steps of minimisation were applied to each configuration. This ensured that all configurations, even those with initial Y in the good region, took some minimisation steps.
5. Estimate an initial value for W_{MTK} using Eq. (12.21), assuming an initial kinetic energy $K = \frac{H_0 - H_{\min}}{2}$.
6. Set $\mathbf{p} = \mathbf{0}$, $p_V = 0$.
7. Obtain a step size dt that sets an average acceptance rate of `acc_rat_cautious` $\pm$ 5% for the Metropolis acceptance step in HMC updates. (`acc_rat_cautious` is a parameter of the calculation, and between 0 and 1.)
8. Select fresh atomic momenta $\mathbf{p}$ as described in Sect. 12.2.
9. Perform an initial `roam_eq` $\times \frac{1}{4}$ molecular dynamics steps for each configuration. (`roam_eq` is a parameter of the calculation and corresponds to the total number of steps taken when initialising each starting configuration.) Each trajectory consists of a series of HMC updates as described in Sect. 12.3.3, with L steps for each update.
10. Revise the value for W_{MTK} using Eq. (12.21).
11. Revise the step size dt that sets an average acceptance rate of `acc_rat_cautious` $\pm$ 5% for the Metropolis acceptance step in HMC updates.
12. Perform a further `roam_eq` $\times \frac{3}{4}$ molecular dynamics steps for each configuration, as in step 9.
13. Further revise the step size dt that sets an average acceptance rate of `acc_rat_cautious` $\pm$ 5% for the Metropolis acceptance step in HMC updates.

12.3.5 HMC Acceptance Rate

The time step dt is set to fix the acceptance rate of HMC proposals to some value `acc_rat` $\pm$ 5%. In the gas phase, collisions between atoms are extremely penetrat-

ing, and the system enters very small regions of high curvature in $U(\mathbf{q})$. Motion through such regions must be integrated with a much smaller time step than motion through free space. Accommodating both time steps would require an adaptive time step. At the same time, the regions corresponding to highly penetrating collisions are a very small fraction of configuration space, surely only relevant at exceptionally high temperatures and pressures. For this reason, we do not employ an adaptive step length, preferring to allow the integration to fail in such regions. In this event the total energy will change almost discontinuously to some very high value that is then rejected by the HMC update.

What then controls the acceptance rate of HMC proposals in the gas phase? Throughout the gas phase the density is reduced on average by a constant amount each iteration, and the evolution is fast. In the rest of the nested sampling calculation it is more than sufficient to update the time step every $\frac{\mathcal{K}}{2}$ iterations (every time Φ halves), but in the gas phase this is not the case. After $\frac{\mathcal{K}}{2}$ iterations the acceptance rate has dropped notably and the exploration has become very poor. There are two solutions to this problem. The first, to update the step size more often at the beginning of the run; the second, to set `acc_rat` to some high value for the first part of the run. Re-evaluating dt consumes additional energy and force evaluations and we found that setting `acc_rat` $= 95\%$, exploration in the gas phase was excellent. Therefore we chose the second method.

The acceptance rate was set to `acc_rat_cautious` $= 95\% \pm 5\%$ until an interval of $\frac{\mathcal{K}}{2}$ iterations passed between successive dt updates, in which the maximum recorded value of the difference between the instantaneous enthalpy in any timestep and H_{lim}, $H(t) - H_{\text{lim}}$, was not greater than $\langle H_{\text{lim}} - H \rangle$, as calculated from the $\mathcal{K}$ samples at the end of that interval. This corresponds to a definition of the point at which failures of the dynamics integrator first stop. For the second (major) part of the calculation, the acceptance rate was set to `acc_rat_confident` $\pm 5\%$.

Figure 12.1 compares the heat capacities calculated using MDNS to the converged value from standard (MCMC) nested sampling. Using the "two stage" program for `acc_rat` (right column) we find that the calculation converges for virtually any value of `acc_rat_confident`. The fixed `acc_rat` calculations (left column) sample the gas phase poorly for `acc_rat` ≤ 0.65, and fail for `acc_rat` ≤ 0.35. The change of acceptance ratio corresponds to the temperature indicated by the dotted blue line. In particular, the calculation with constant `acc_rat` $= 0.2$ fails, whereas with the two-stage acceptance rate, the evaporation peak is converged and the melting peak is "close" to convergence.

Figure 8.11 suggests that, for MCMC nested sampling, it is more challenging to converge the melting transition than the evaporation transition. Convergence of the melting transition therefore seems like the sensible measure by which to compare convergence of the MCMC and HMC nested sampling methods. We repeated the calculation 16 times at each of `acc_rat_confident` $\in \{0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$ and found the maximum value of C_P. The mean and standard deviations of those 16 values are shown for each value of `acc_rat_confident` in Fig. 12.2. Phase space exploration is least good for high rejection rates. Results for

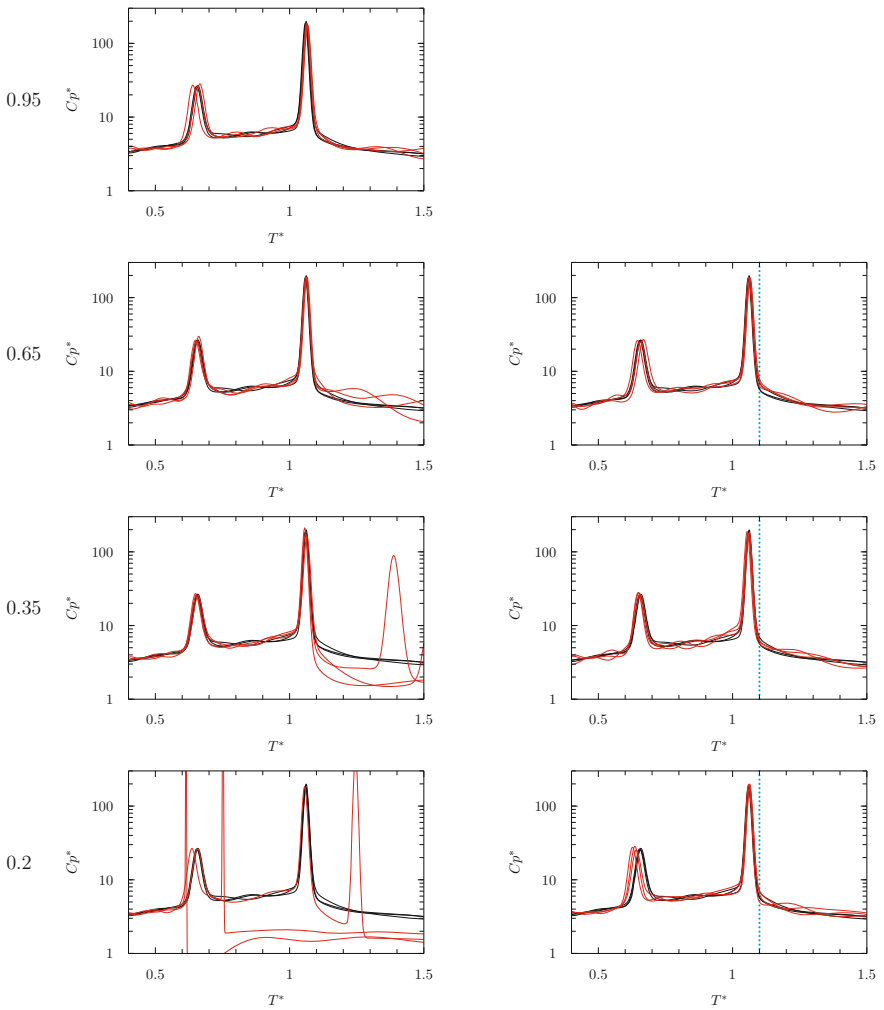

Fig. 12.1 Behaviour of MDNS with a single HMC acceptance rate (left column) and two part acceptance rate (right column). The HMC acceptance rate used for each row is shown to the left. In the right hand column, an acceptance rate 0.95 was used for the first part of the calculation, until the point where no failures of the integrator were detected (see the text). That region corresponds to temperatures to the right of the dotted blue line. The red lines show the results of three MDNS calculations with $\mathcal{K} = 640$ live samples. Decorrelating fresh samples was performed using a single HMC update of walk length $L = 65$ steps. The black lines show three converged results with standard (MCMC) NS, using $\mathcal{K} = 640$ live samples, and a walk length $L = 165120$. We see that with a single acceptance rate, the integration fails in the gas region for `acc_rat` < 0.65. With a two part acceptance rate, the convergence is good for `acc_rat` ≥ 0.35. In particular, note the dramatic failure of calculations with fixed `acc_rat` $= 0.2$. In contrast, the two part acceptance rate calculation is correct for the evaporation peak, and close to convergence for the melting transition. These heat capacities were calculated for Lennard-Jonesium (64 atoms, periodic simulation) at fixed pressure, with $\log_{10} \frac{P\sigma^3}{\epsilon} = -1.194$ and a fixed shape (cubic) simulation cell. The potential used is described in Sect. 12.3.7

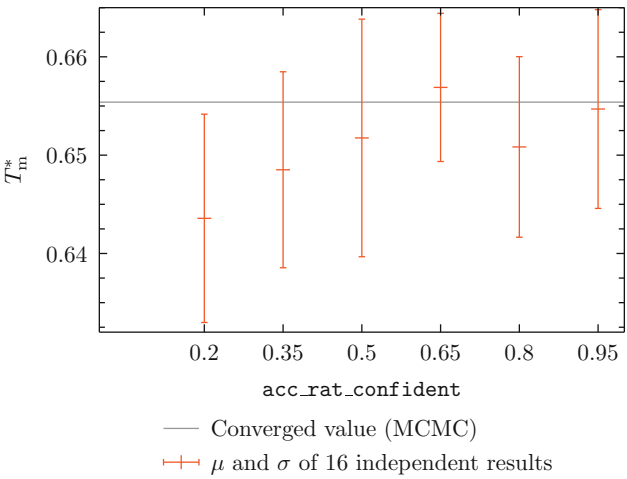

Fig. 12.2 Convergence of melting temperature T_m^* with respect to `acc_rat_confident`. Phase space exploration is worse for high rejection rates (low acceptance rates). Nevertheless, the result is robust with respect to `acc_rat_confident` for `acc_rat_confident` ≥ 0.5. Interestingly, the lowest variance in this sample set is for `acc_rat_confident` $= 0.65$, which is the theoretically most efficient value for sampling from the canonical distribution for a potential energy that is the sum of $3N + 1$ independent variables [11]. The standard deviation for T_m^* ranges from 1.15% to 1.85% of the melting temperature, with 1.54% at `acc_rat_confident` $= 0.95$. We conclude that the two-stage acceptance ratio is required for `acc_rat` < 0.95, but that the simpler, single-stage method is almost as good, using `acc_rat` $= 0.95$. We therefore advise users to employ the simpler scheme, with single value `acc_rat` $= 0.95$

`acc_rat_confident` ≥ 0.5 are robust with respect to `acc_rat_confident`. Optimal performance in this sample set was observed for `acc_rat_confident` $=$ 0.65. Interestingly, this is also the theoretically most efficient value for sampling the canonical distribution with HMC for a potential energy that is the sum of $3N + 1$ independent variables [11]. For `acc_rat_confident` < 0.95, the two-stage method is clearly required. However, the improvement in performance over the single-stage acceptance ratio with `acc_rat` $= 0.95$ is marginal: The standard deviation of melting temperatures is 1.54% of T_m^* for `acc_rat` $= 0.95$ (single value), and 1.15% of T_m^* for the two-stage acceptance rate with `acc_rat_confident` $= 0.65$. We consider this to be an insufficient advantage to motivate the more convoluted scheme, and propose the simpler single acceptance rate `acc_rat` $= 0.95$ throughout.

12.3.6 Thermal Distributions

In Sect. 7.2 we described the mechanism by which is is possible to use all the output from nested sampling to calculate thermal distributions. Coordinates and momenta are statistically independent in the isobaric–isoenthalpic distribution, and typically

one would like to produce thermal distributions over the coordinates alone. However, in performing nested sampling in the total Hamiltonian we have instead produced samples from the joint distribution in the isoenthalpic-isobaric ensemble (12.17). Unpicking the relative contribution to the configurational density of states would be a very noisy process. However, it is valid to sample phase-space by choosing samples according to their thermalised weights (12.24) in the isobaric—isothermal distribution (4.59).

$$w_i \propto \Delta\Phi_i \mathrm{e}^{-\beta H_i} \tag{12.24}$$

Since the isobaric—isothermal distribution is separable, we can then take the coordinates of those samples as proper thermal samples from configuration space alone. This is equivalent to mapping the full thermalised weight of each sample to its coordinates, and thus we may process the results as before.

12.3.7 Comparison of MDNS with MCMC

The heat capacity of a periodic system of $N = 64$ Lennard-Jonesium particles was calculated using HMC and MCMC nested sampling methods. We noted in Sect. 12.3.5 that Fig. 8.11 suggests that it is most difficult to converge the melting peak. We therefore compare the behaviour of the two methods for this problem.

Simulation details

MCMC calculations with walk lengths $\frac{L_{\mathrm{MCMC}}}{N} \in \{8, 16, 32, 64, 128\}$ were compared to HMC calculations with a factor N fewer steps: $L_{\mathrm{HMC}} \in \{8, 16, 32, 64, 128, 256\}$.

The potential used was the "truncated and shifted" Lennard-Jonesium potential, where the pair-potential goes continuously to 0 at r_c with a discontinuous first derivative, (12.25).

$$\begin{aligned} U\left(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N\right) &= \sum_{i \geq j} u_{ij} \\ u_{ij} &= \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] - u_\mathrm{c}, & r_{ij} \leq r_\mathrm{c} \\ 0, & \text{Elsewhere} \end{cases} \\ u_\mathrm{c} &= 4\epsilon\left[\left(\frac{\sigma}{r_\mathrm{c}}\right)^{12} - \left(\frac{\sigma}{r_\mathrm{c}}\right)^{6}\right] \end{aligned} \tag{12.25}$$

We used parameters $\epsilon = 0.1\mathrm{eV}$, $\sigma = 3\text{Å}$ and $r_\mathrm{c} = 3\sigma$ for the potential. The reduced pressure was set to $\log_{10} P^* = -1.194$. Sixteen independent serial calculations were performed for each walk length.

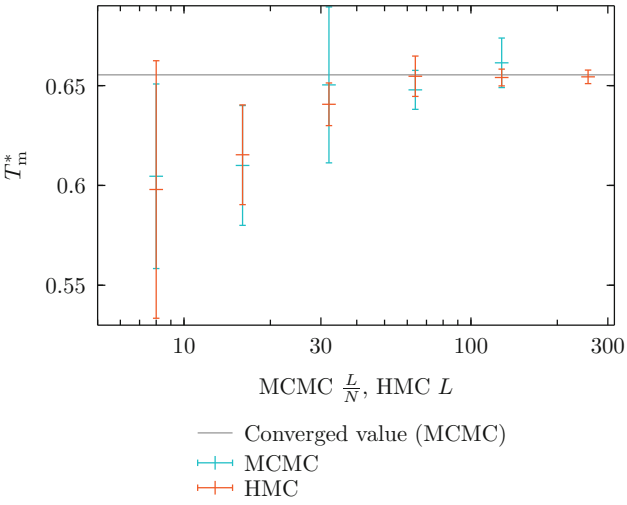

Fig. 12.3 Convergence of melting temperature T_m^* with respect to walk length L in HMC and MCMC. Sixteen (serial) calculations were performed for each parameter value. Here we show the mean value of the maximum of T_m^*, together with the standard deviation of those 16 results. Note that for MCMC, the horizontal axis corresponds to a factor $N = 64$ times as many walk steps in decorrelation, compared to HMC. HMC requires only $L = 65$ steps to obtain the correct answer with a standard deviation 1.54%, and that error decreases rapidly when L is increased: 0.64% for $L = 128$, and 0.52% for $L = 256$. For MCMC, the correct result is within a standard deviation (6.01%) for $L = 32 \times N$. It is not clear that such a result can be said to be converged. At longer walk lengths, the converged value is consistently within a standard deviation of the mean, although the convergence is much slower than for HMC. We conclude that for nested sampling in the total Hamiltonian with 64 atoms in the simulation cell, HMC requires about a factor of 64 fewer walk steps to decorrelate fresh samples

For MCMC, the calculations were performed with zero lattice shear or lattice stretch MC moves. Otherwise, calculations were performed as described in Sect. (8.8.1).

For HMC, calculations were initialised using $H_0 = N \times 161.25\text{eV}$. The "cell timescale" τ_b was set to 10 fs. The lattice mass W_{MTK} and time step dt were updated in that order every $\frac{K}{2}$ iterations (corresponding to a reduction of log Φ by $-\frac{1}{2}$). A single value for the HMC acceptance rate `acc_rat` $= 0.95$ was used throughout each calculation. Decorrelating each fresh sample was performed using a single HMC update.

Figure 12.3 compares the results of these calculations.

Results

Figure 12.3 compares convergence of the predicted melting point with MCMC and HMC nested sampling. The horizontal axis corresponds to a factor of $N = 64$ more energy evaluations for MCMC as compared to HMC. Even with a factor 64 fewer energy evaluations, HMC is better converged. HMC requires only $L = 65$ steps to

obtain the correct melting temperature with a standard deviation 1.54%, and that error decreases rapidly when L is increased to 0.64% for $L = 128$, and 0.52% for $L = 256$. For MCMC, the correct result is within a standard deviation (6.01%) for $L = 32 \times 64$. It is not clear that such a result can be said to be converged. At longer walk lengths, the converged value is consistently within a standard deviation of the mean, although the convergence is much slower than for MDNS.

A single HMC update was used to decorrelate samples. If that single update is unsuccessful, the coordinates return to the values belonging to their clone, p_ϵ is negated, and $\mathbf{p}$ randomised. At long walk lengths, the convergence of HMC may therefore have been even better had the decorrelation been divided into shorter parts. Clearly the HMC updates are a far more effective means of decorrelating configurations than a random walk.

We should state, that by performing nested sampling in total phase space, we have doubled the dimensions of the space we must characterise. This creates an additional factor of 2 in the number of iterations required for an equivalent calculation. Nevertheless, we feel that the factor of 2 in the number of iterations is not important when compared to the factor of $\frac{1}{64}$ reduction of the number of energy evaluations that are required.

12.4 Summary

We developed a new method for decorrelating configurations using Hamiltonian Monte Carlo. Calculations were performed at fixed pressure, using the MTK dynamics with a fixed shape (cubic) simulation cell. Instead of performing nested sampling in configuration space, the approach was instead to perform nested sampling in phase space. We found this method to be more than 64 times as efficient as the MCMC algorithm for calculating the melting temperature, in a simulation with 64 atoms.

The MTK dynamics promote the cell volume to a dynamical degree of freedom, with an associated "mass" W_{MTK}. To avoid resonances of the cell dynamics at a particular kinetic energy, Martyna et al. propose [9] that W_{MTK} be made proportional to the temperature. As explained in Sect. 12.3.2, this choice is a subtle one: changing the value of W_{MTK} changes the partition function, but changing W_{MTK} according to the temperature ensures that our calculations are correct at any given temperature.

12.5 Further Work

I would like to include anisotropic changes in the unit cell matrix $\mathbf{h}_0$ for HMC in the total Hamiltonian. Martyna et al. [6] note that having distinct masses for the isotropic and anisotropic motion of the simulation cell enhanced exploration. Unfortunately the integrator they propose is rather complex. We have preferred to follow the elegant scheme of Yu et al. [8]. Yu et al. detail integrators for both isotropic

cell motion, and anisotropic cell motion. For anisotropic cell motion, the scheme they propose has only a single mass for both motions of the simulation cell, and in our exploratory calculations we did indeed find the anisotropic exploration to be rather slow. Hamiltonian Monte Carlo is also not well suited to hard boundaries, such as the minimum cell height criterion introduced in Sect. 8.4. One possible solution to the poor anisotropic decorrelation would be to develop a cell mass tensor. This would allow for two masses within a scheme closely related to [8]. A simpler solution would be to integrate the lattice-stretch and lattice-shear MC moves described in Sect. 8.5 into the HMC scheme, by performing both HMC and MC moves.

References

1. J. Skilling, Bayesian computation in big spaces-nested sampling and Galilean Monte Carlo, in Bayesian Inference and Maximum Entropy Methods, in *Science and Engineering: 31st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. vol. 1443 (AIP Publishing, 2012), p. 145
2. F. Feroz, J. Skilling, Exploring multi-modal distributions with nested sampling, arXiv:1312.5638
3. G. Marsaglia et al., Choosing a point from the surface of a sphere. Ann. Math. Stat. **43**, 645 (1972)
4. M.E. Muller, A note on a method for generating points uniformly on n-dimensional spheres. Commun. ACM **2**, 19 (1959)
5. H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature. J. Chem. Phys. **72**, 2384 (1980)
6. G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms. J. Chem. Phys. **101**, 4177 (1994)
7. M. Tuckerman, *Statistical Mechanics and Molecular Simulations*, (Oxford University Press, 2008)
8. T.-Q. Yu, J. Alejandre, R. López-Rendón, G.J. Martyna, M.E. Tuckerman, Measure-preserving integrators for molecular dynamics in the isothermal-isobaric ensemble derived from the Liouville operator. Chem. Phys. **370**, 294 (2010)
9. G.J. Martyna, M.L. Klein, M. Tuckerman, Nosé-Hoover chains: the canonical ensemble via continuous dynamics. J. Chem. Phys. **97**, 2635 (1992)
10. A.M. Horowitz, A generalized guided Monte Carlo algorithm. Phys. Lett. B **268**, 247 (1991)
11. R.M. Neal, MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo* (CRC Press, New York, NY, 2011), p. 113

Part IV
Conclusion

Chapter 13
Summary and Further Work

13.1 Summary

Nested sampling is an algorithm for calculating the integrated density of states. Phase transitions are easily located by finding the peaks of response functions such as the heat capacity. In contrast to methods for comparing specific phases, nested sampling unbiasedly explores all of configuration space. Consequently, the phenomena it reveals are not limited to expected phases or known phase transitions.

We developed the application of nested sampling to systems under constant pressure conditions. The method was applied to standard Lennard-Jonesium, a binary Lennard-Jonesium "alloy" and the NPB embedded atom model potential for aluminium. In each of these systems nested sampling revealed unexpected results: in the binary alloy we discovered the order-disorder transition in the solid region, with two different ordered phases I and II, stable at low and high pressures respectively; in aluminium, we found that the NPB embedded atom model potential incorrectly describes the solid phase, the hexagonal close packed structure being unstable at zero Kelvin for all pressures; and we found the Widom line for maxima of the heat capacity in the supercritical region of all three systems.

We identified a small system size effect in the fully flexible simulation cell formulation with periodic boundary conditions, whereby the most probable low temperature structure of any crystal has a thin simulation cell. Such thin simulation cells introduce fictitious periodicity, and lead to simulated systems of effectively reduced dimension. Such systems can not approximate an atomic fluid in three dimensions. We saw that by introducing a "minimum cell height", it is possible to rigorously exclude unphysical simulation cells.

We developed an algorithm for identifying the mean volume of the dominant phase at fixed pressure and temperature, from the output of nested sampling. By combining results from nested sampling calculations at a range of pressures, we obtained the equation of state of Lennard-Jonesium: $\phi(P, T, V) = 0$.

We reviewed the published method for parallelising nested sampling, "parallelisation over the number of iterations". We showed that, relative to serial nested sam-

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_13

pling, the error in this method is at least exponential in $\frac{\mathcal{P}-1}{\mathcal{K}}$ where $\mathcal{P}$ is the number of processors used, and $\mathcal{K}$ is the number of configurations used in the calculation. We proposed a new method of parallelisation: "parallelisation within each iteration, over the walk length". This new scheme makes the total number of MC steps that is applied to a sample between when it is cloned by copying, and when it is eventually written as output, L', a random variable. While there are no analytic results for the error due to the variability of L', we were able to show that for $\mathcal{P} \gg 1$ the standard deviation of L' was almost as large as $\langle L' \rangle$. Nevertheless, empirically this does not appear to strongly affect the results of a nested sampling calculation, while "parallelising over the number of iterations" introduces an exponential scaling of the error in $\{\chi_i\}$ with respect to $\mathcal{P}$.

We developed a version of nested sampling based on Hamiltonian Monte Carlo, using MTK dynamics at constant pressure. Here, nested sampling was performed in the total Hamiltonian. We compared the efficiency of this method to our earlier Markov chain Monte Carlo method of nested sampling. Both methods were used to calculate the melting transition of Lennard-Jonesium at a reduced pressure $\log_{10} P^* = -1.194$. We found that, for a simulation of 64 atoms, the Hamiltonian Monte Carlo method required a factor of 64 fewer energy evaluations as compared to the MCMC method for a similar level of convergence.

In all of the constant pressure methods we explored, the volume of the simulation cell is allowed to vary. For classical dynamics at constant pressure, the simulation cell volume was promoted to a dynamical degree of freedom, with its own mass. This naturally leads to resonances between the atomic and cell motions at particular kinetic energies, and to avoid this we used a temperature dependent mass for the volume. Varying the volume mass changes the density of states and partition function, but by setting the mass according to temperature, we find that the sampling is consistent at each temperature and that phase transitions are correctly predicted.

13.2 Further Work

Most of the computational expense in nested sampling occurs in the "clone and decorrelate" algorithm. This algorithm first creates a "hole and a bump" in the uniform distribution, then re-equilibrates that same uniform distribution. It would be transformative to create a version of nested sampling in which this did not occur. One strategy might be to copy, with some probability, from the recent high energy configuration at each iteration. The difficulty for any such approach would be avoiding bad sampling in multimodal potentials.

For the equation of state, it would be interesting to compare the predicted volumes at phase transitions to those from the literature. Improved results close to the critical point should also be obtained by using a gamma mixture model, instead of a Gaussian mixture model for the volume distribution.

Anisotropic shape changes of the simulation cell could be included in the HMC nested sampling approach by performing some additional simulation cell shape Monte Carlo moves from the Markov chain Monte Carlo algorithm described in Part II.

Appendix A: Model Selection for Gaussian Mixtures

We have some independent data points drawn from an unknown distribution that is possibly multimodal. We want to model that distribution as a Gaussian mixture. In this section we see how to calculate a probability distribution over the number of Gaussians included in the mixture.

A mixture of M Gaussians with amplitudes $\mathbf{A} = (A_1, A_2, \ldots, A_M)$, means $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M)$ and standard deviations $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_M)$ is given by

$$G_M\left(x|\mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma}\right) = \sum_{i=1}^{M} \frac{A_i}{\sqrt{2\,\pi}\sigma_i} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right]. \tag{A.1}$$

For simplicity, we represent the parameters $\mathbf{A}$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ collectively as $\boldsymbol{\theta}$. The data points $\{D_k\}$ are drawn independently from the unknown distribution. We want to calculate the probability prob $(M|\{D_k\})$.

From Bayes' theorem (2.18) we have

$$\text{prob}\left(M|\{D_k\}\right) = \frac{\text{prob}\left(\{D_k\}|M\right)\text{prob}\left(M\right)}{\text{prob}\left(\{D_k\}\right)} \tag{A.2}$$

The denominator prob $(\{D_k\})$ will cancel when we normalise the probabilities, and prob (M) is just the prior reflecting the information we have to start with. Therefore, we turn our attention to prob $(\{D_k\}|M)$. Applying marginalisation (2.11) followed by the product rule (2.15), we have

$$\text{prob}\left(\{D_k\}|M\right) = \int \mathrm{d}\boldsymbol{\theta}\ \text{prob}\left(\{D_k\}, \boldsymbol{\theta}|M\right) \tag{A.3}$$

$$= \int \mathrm{d}\boldsymbol{\theta}\ \text{prob}\left(\{D_k\}|\boldsymbol{\theta}, M\right)\text{prob}\left(\boldsymbol{\theta}|M\right) \tag{A.4}$$

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0

The term prob $(\boldsymbol{\theta}|M)$ is our prior for the parameters $\boldsymbol{\theta}$ in a mixture of M Gaussians. Substituting from (A.4) into (A.2) and normalising, we find

$$\begin{array}{r} \text{prob}\,(M|\,\{D_k\}) = \frac{1}{Z} \times \text{prob}\,(M) \int \mathrm{d}\boldsymbol{\theta}\ \text{prob}\,(\{D_k\}\,|\boldsymbol{\theta}, M)\ \text{prob}\,(\boldsymbol{\theta}|M) \\ Z = \sum_M \text{prob}\,(M) \int \mathrm{d}\boldsymbol{\theta}\ \text{prob}\,(\{D_k\}\,|\boldsymbol{\theta}, M)\ \text{prob}\,(\boldsymbol{\theta}|M). \end{array} \tag{A.5}$$

Assuming that the data points were drawn independently,

$$\text{prob}\,(\{D_k\}\,|\boldsymbol{\theta}, M) = \prod_k G_M\,(D_k|\boldsymbol{\theta}). \tag{A.6}$$

The integral in (A.5) could be performed in a number of ways. One way would be to use Monte Carlo integration, although the integral may well be dominated by a small region of configuration space, proportional to e^{-M}. This would tend to make the integral prohibitively expensive to compute. We could certainly compute the integral using nested sampling [1], [2] (see Part II), however in the next section we describe some simplifying assumptions that allow fast evaluation of the integral.

A.1 Simplifying Assumptions: Uniform Priors And A Harmonic Approximation

We might reasonably assign a uniform prior for M on the set $M \in \{1, 2, \ldots, M_{\max}\}$. Similarly, we can take a uniform prior for $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ on the ranges $0 \leq \sigma_M \leq \sigma_{\max}$, $\mu_{\min} \leq \mu_M \leq \mu_{\max}$, together with a prior $\delta\left(1 - \sum_M A_M\right)$ on the range $0 \leq A_M \leq 1$.

Next, we assume that at each value of M, prob $(\{D_k\}\,|\boldsymbol{\theta}, M)$ has a single mode, of Gaussian form with mean equal to the value of $\boldsymbol{\theta}$ that maximises that likelihood. The optimal $\boldsymbol{\theta}$ corresponds to the minimum of χ^2 where

$$-\frac{\chi^2}{2} = \log\left(\text{prob}\,(\{D_k\}\,|\boldsymbol{\theta}, M)\right) \tag{A.7}$$

$$= \sum_k \log\left(G_M\,(D_k|\boldsymbol{\theta})\right). \tag{A.8}$$

In other words, we have made a harmonic approximation for χ^2

$$\begin{array}{r} \chi^2\,(\boldsymbol{\theta}) \approx \chi^2_{\min} + \frac{1}{2}\,(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T\,\nabla\nabla\chi^2\,(\boldsymbol{\theta}_0)\,(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ \chi^2_{\min} = \chi^2\,(\boldsymbol{\theta}_0)\,. \end{array} \tag{A.9}$$

Our uniform prior for $\boldsymbol{\theta}$ corresponds to

$$\text{prob}\,(\boldsymbol{\theta}|M) = \left[\sigma_{\max}^{M}\,(\mu_{\max}-\mu_{\min})^{M}\int_{(0,1)^M} \mathrm{d}A_1\ldots\mathrm{d}A_M\,\delta\left(1-\sum_{i=1}^{M}A_i\right)\right]^{-1}. \tag{A.10}$$

Thus the integral in (A.5) is a multivariate Gaussian integral, and can be evaluated as

$$\int \mathrm{d}\boldsymbol{\theta}\ \text{prob}\,(\{D_k\}\,|\boldsymbol{\theta},\,M)\text{prob}\,(\boldsymbol{\theta}|M) = \text{prob}\,(\boldsymbol{\theta}|M)\,\frac{M!\,(4\,\pi)^{\frac{3M}{2}}}{\sqrt{\det\left(\nabla\nabla\chi^2\right)}}\exp\left(-\frac{\chi^2_{\min}}{2}\right). \tag{A.11}$$

The interested reader is referred to [3] where a very similar exposition is given at greater length.

References

[1] J. Skilling, "Nested sampling," *AIP Conference Proceedings*, vol. 735, p. 395, 2004.

[2] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Analysis*, vol. 1, p. 833, 2006.

[3] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*, ser. Oxford science publications. Oxford University Press, 2006.

Appendix B: Soft K-Means Algorithm, Version 2

The soft K-means algorithm, version 2 [1] is an algorithm for optimising the likelihood of a Gaussian mixture with respect to its parameters, given a number of samples drawn from a distribution. A Gaussian mixture of M Gaussians is given by

$$G_M\left(x|\mathbf{A},\boldsymbol{\mu},\boldsymbol{\sigma}\right)=\sum_{i=1}^{M}\frac{A_i}{\sqrt{2\,\pi}\sigma_i}\exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right]. \qquad \text{(A.1 revisited)}$$

Since we seek to approximate a pdf, we impose the condition that the mixture is normalised

$$\sum_{i}^{M}A_i=1. \qquad \text{(B.1)}$$

The soft K-means algorithm, version 2 has two steps:

Assignment step. We assign a "responsibility" array $r_k^{(n)}$. Here the index n refers to a sample, and the index k to a Gaussian in the mixture (A.1). In one dimension, the responsibilities are calculated thus:

$$r_k^{(n)}=\frac{\frac{A_k}{\sqrt{2\,\pi}\sigma_k}\exp\left[-\frac{\left(\mu_k-x^{(n)}\right)^2}{2\sigma_k^2}\right]}{\sum_{k'}\frac{A_k'}{\sqrt{2\,\pi}\sigma_k'}\exp\left[-\frac{\left(\mu_k'-x^{(n)}\right)^2}{2\sigma_k'^2}\right]}. \qquad \text{(B.2)}$$

Update step. The parameters of each Gaussian A_k, μ_k, σ_k^2 are adjusted to match the data points that it is responsible for.

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0

$$\mu_k = \frac{\sum_n r_k^{(n)} x^{(n)}}{R_k} \tag{B.3}$$

$$\sigma_k = \frac{\sum_n r_k^{(n)} \left(\mu_k - x^{(n)}\right)^2}{R_k} \tag{B.4}$$

$$A_k = \frac{R_k}{\sum_k R_k} \tag{B.5}$$

where R_k is the total responsibility of Gaussian k

$$R_k = \sum_n r_k^{(n)}. \tag{B.6}$$

These two steps are repeated until convergence is obtained, typically in a few dozen iterations. The amplitude update (B.5) guarantees that the amplitudes satisfy the normalisation condition (B.1).

To initialise the algorithm, amplitudes may all be set to $\frac{1}{M}$ and standard deviations can be drawn randomly from some range, for example $\sigma \sim U(0, 5\sigma_0)$ where σ_0 is the standard deviation of the sample set. Initial values for the means $\{\mu_k\}$ can be generated by choosing the values of M random samples.

Repeating the algorithm a handful of times from different initial conditions is usually enough to produce a very close fit to the global maximum.

Reference

[1] D. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.