Application of nested sampling in statistical physics: the Potts model

Manuel J. Pfeifenberger[a,*], Wolfgang von der Linden[b]

[a] *Erich Schmid Institute of Materials Science, Austrian Academy of Sciences, 8700 Leoben, Austria*
[b] *Institute of Theoretical and Computational Physics, Graz University of Technology, 8010 Graz, Austria*

Abstract

We present a systematic benchmark study of the nested sampling algorithm on the basis of the Potts model. This model exhibits a first order phase transition for $q > 4$ at the critical temperature. The numerical evaluation of the partition function and thermodynamic observables, which involves high dimensional sums of sharply structured multimodal density functions, represents a major challenge to most standard numerical techniques, such as Markov Chain Monte Carlo. Nested sampling, on the other hand, is particularly suited for such problems. In this paper we will employ both, nested sampling and thermodynamic integration to evaluate the partition function of the Potts model. In both cases individual moves are based on Swendsen-Wang updates. A autocorrelation time analysis of both algorithms shows that the severe slowing down of thermodynamic integration around the critical temperature does not occur in nested sampling. In addition we show, how thermodynamic variables can be computed with high accuracy from the results of a single nested sampling run, without numerical derivatives. Results for the internal energy are compared to known results obtained by means of a multi-canonical simulation. Eventually an approach for a parallel implementation of nested sampling is presented and analysed in detail.

Keywords: Nested sampling, Potts model, Thermodynamic integration, Multi-canonical simulation, Parallel nested sampling, Partition function

1. Introduction

Monte Carlo (MC) simulations are the most important instruments when it comes to evaluating integrals in a high dimensional phase space. For such problems diverse methods, for example simulated tempering [1] or multicanonical simulations [2], [3], are available. Basically they are trying to enhance the efficiency of the MC algorithm by flattening relevant probability distributions. A conceptual completely new approach, named *nested sampling* (NESA), has been suggested by Skilling in 2004 [4]. It is a promising way for estimating high dimensional, multimodal integrals and is based on the idea of Lebesgue integration. Since its development it has already found its way into various fields of research. Especially for solving problems in statistics and Bayesian inference many applications already exist [5, 6, 7, 8, 9]. A first application in the field of statistical physics, in particular the Potts model, has been presented by Murray *et al.* [10]. The Potts model provides, despite its simple structure, a wide variety of physically interesting properties and the analytical availability of certain quantities of the two dimensional model makes it an optimal playground for testing new approaches in simulation techniques. For parameters, where the model exhibits a first order phase transition, the evaluation of the partition function poses a severe difficulty for standard MC algorithms, because at first order phase transitions the autocorrelation times can become huge even in small systems. Moreover, such systems are characterized by a double-peak structure in the probability density for the energy with a pronounced minimum in between, which is exponentially suppressed due to the interface tension. This causes severe mixing problems in most Monte-Carlo techniques [11]. In this case two well separated phase space regions need to be explored. Transitions between these regions are possible but very improbable and therefore the relative weights of the maxima will be not sampled correctly.

In his original paper [4] Skilling already mentioned the possibility of a faster exploration of the phase space via a parallel implementation of nested sampling. This idea has recently been picked up in various publications ([6, 12, 13, 14]).

In the present work the performance of nested sampling for evaluating the partition function of the Potts model is investigated in more detail. The performance of thermodynamic integration (TI), an alternative way of computing the partition function, is used to benchmark nested sampling. The paper is organized as follows: In Sec. 2 the Potts model is introduced. The MC methods employed by us to compute the partition function of the Potts model, namely thermodynamic integration, multi-canonical sim-

*Corresponding author at: Erich Schmid Institute of Materials Science, Austrian Academy of Sciences, Jahnstr. 12, 8700 Leoben, Austria

Email address: manuel.pfeifenberger@oeaw.ac.at (Manuel J. Pfeifenberger)

ulation and nested sampling are introduced in Sec. 3.1, Sec. 3.2 and Sec. 3.3 respectively. Being in the focus of our investigation nested sampling is treated in a more elaborate way. A parallel implementation of this algorithm is presented in Sec. 4.4. Results for the Potts model, as well as a performance comparison of thermodynamic integration and nested sampling are compiled in Sec. 4. Further it is outlined how thermodynamic variables can be evaluated from a single nested sampling run. Subsequently an analysis of the capability of parallel nested sampling is given. Finally the results are discussed and potential generalizations are outlined.

2. Potts model

One of the most investigated models in statistical physics is the Potts model [15]. Its Hamiltonian, without any external field, reads as follows

$$H(\mathbf{s}) = -J \sum_{\langle i,j \rangle} \left(\delta_{s_i,s_j} - 1\right) = -J\big(N_{\mathrm{eq}}(\mathbf{s}) - N_{\mathrm{p}}\big)\ . \tag{1}$$

The dynamic variables s_i of the model, referred to as spin or colour, can assume the integer values between 1 and q. In Eq. (1) $N_{\mathrm{eq}}(\mathbf{s}) = \sum_{\langle i,j \rangle} \delta_{s_i,s_j}$ is the number of nearest neighbour pairs with equal spin, and N_{p} is the total number of nearest neighbour pairs in the lattice under consideration. Here the exchange coupling J is positive and site independent. The lattice indices are denoted by i and j, ranging from 1 to the number of sites N. The sum only includes nearest neighbour interactions which is denoted by $\langle i,j \rangle$. The sought-for partition function reads

$$Z_P(\beta) = \sum_{\mathbf{s}} e^{\beta J \sum_{\langle i,j \rangle} \left(\delta_{s_i,s_j} - 1\right)}\ . \tag{2}$$

Two limiting cases can easily be determined

$$Z(\beta) \underset{\beta \to 0}{\longrightarrow} q^N\ , \qquad Z(\beta) \underset{\beta \to \infty}{\longrightarrow} q\ , \tag{3}$$

which will be of interest later on. Our investigations are restricted to 2d square lattices with periodic boundary conditions. The infinite square lattice exhibits for $q \leq 4$ ($q > 4$) a second (first) order phase transition. The exact critical inverse temperature for the Potts model on a two dimensional infinite square lattice follows from self-duality of the low and high temperature region [16]

$$\beta_c J = \ln(1 + \sqrt{q}). \tag{4}$$

3. Evaluation of the partition function

Quite generally, the partition function Z associated to the thermodynamic potential (Helmhotz free energy) contains the entire thermodynamic information of a system in the canonical ensemble. It can be expressed as

$$Z = \sum\!\!\!\!\!\!\int d\mathbf{x}\, L(\mathbf{x})\, \pi(\mathbf{x}), \tag{5}$$

where $\mathbf{x}$ describes a point in a multidimensional phase space, which can either be continuous or discrete. In the following $L(\mathbf{x})$ is denoted as likelihood function and $\pi(\mathbf{x})$ as prior function. Assuming the likelihood function $L(\mathbf{x})$ shows a strong variation, then a classical Markov Chain MC needs a huge sample size to yield reasonable variances [17]. For the evaluation of Z special methods exist. Three of them will be presented in the following subsections.

3.1. Thermodynamic integration

For the derivation of thermodynamic integration (see [17]) an inverse temperature β is introduced as an an auxiliary parameter

$$Z(\beta) = \int d\mathbf{x}\, L(\mathbf{x})^\beta\, \pi(\mathbf{x}), \tag{6}$$

where $Z(0) = 1$, due to the normalized prior. In statistical physics application the inverse temperature is already part of the likelihood and needs not to be introduced artificially. The derivative of $\ln[Z(\beta)]$ with respect to β leads to

$$\begin{aligned} \frac{d}{d\beta}\, \ln[Z(\beta)] &= \int d\mathbf{x}\, \underbrace{\frac{\pi(\mathbf{x})}{Z(\beta)} L(\mathbf{x})^\beta}_{p_\beta(\mathbf{x})}\, \ln[L(\mathbf{x})] \\ &= \langle\, \ln[L(\mathbf{x})]\, \rangle_\beta\, . \end{aligned} \tag{7}$$

Here $\langle \ln[L(\mathbf{x})] \rangle_\beta$ denotes the expectation value of $\ln[L(\mathbf{x})]$ under the distribution $p_\beta(\mathbf{x})$. In statistical physics we have the relation $\ln[L(\mathbf{x})] = -U(\beta)$, where $U(\beta)$ is the internal energy. Evaluating $\langle \ln[L(\mathbf{x})] \rangle_\beta$ by MC simulation (e.g. Swendsen-Wang SW [18]) one can compute $\ln[Z(\beta)]$ using

$$\ln[Z(1)] - \ln[Z(0)] = \int_0^1 d\beta\ \langle\, \ln[L(\mathbf{x})]\, \rangle_\beta. \tag{8}$$

This integral can be approximated by a sum over discrete inverse temperature values β_i.

$$\ln(Z) \approx \sum_{i=1}^{M} \Delta\beta\ \langle\, \ln[L(\mathbf{x})]\, \rangle_{\beta_i}\ , \tag{9}$$

where $\Delta\beta = \beta_i - \beta_{i-1}$ and $\beta_0 = 0$. The integrand turns out to be rather smooth and the integral can be approximated reliably by a modest number of β-values.

Comparing Eq. (6) with the definition of the partition function $Z = \sum_{\mathbf{s}}\ e^{-\beta H(\mathbf{s})}$ we find for the Potts model

$$L(\mathbf{x})^\beta \to e^{-\beta H(\mathbf{s})}\ , \tag{10}$$

with the Hamiltonian given in Eq. (1), we can retrieve

$$\ln[L(\mathbf{x})] \to -J\, (N_{\mathrm{eq}}(\mathbf{s}) - N_{\mathrm{p}}), \tag{11}$$

where $N_{\mathrm{eq}}(\mathbf{s})$ denotes the number of nearest-neighbour pairs with the same spin-value. Employing Eq. (9) the estimate of $\ln(Z)$ for the Potts model is given by

$$\ln(Z) = -J \sum_{i=1}^{M} \left(\, \langle N_{\mathrm{eq}} \rangle_{\beta_i} - N_{\mathrm{p}}\right) \Delta\beta \tag{12}$$

with $\langle N_{\mathrm{eq}} \rangle_{\beta_i}$ being the expectation value of $D(\mathbf{s})$ at a certain inverse temperature β_i.

3.2. Multi-canonical simulation

Multi-canonical simulation (MUCA) is a method introduced by Berg and Neuhaus [2]. The underlying idea is to tune the weights of the Monte Carlo sampling in a way to flatten a certain distribution function such that the whole phase space can be reached (see [19]). In principle the distribution can be a function of any order parameter of the system [19], but as we want to obtain the partition function the energy is our order parameter of interest. The Boltzmann weight $\omega_{\rm can}(H) \propto e^{(-\beta H)}$ leads to the following canonical probability distribution for the energy

$$p_{\rm can}(H) = \frac{1}{Z_{\rm can}} n(H) e^{(-\beta H)}, \tag{13}$$

where $n(H)$ is the probability distribution of the energy and $Z_{\rm can}$ is the partition function. For the multi-canonical probability distribution an additional weight function $\omega_{\rm muca}(H)$ is introduced that defines a new probability distribution

$$p_{\rm muca}(H) \propto p_{\rm can}(H)\, \omega_{\rm muca}(H)\ , \tag{14}$$

which is used in the MC simulation. The goal is obtain a flat distribution $p_{\rm muca}(H)$ within a relevant energy window ensuring perfect mixing. The simulation with the modified weights allows to evaluate expectation values for observables under the multi-canonical distribution. Canonical expectation values are then retrieved via re-weighting. More details can be found in[2, 3], and for the estimation related to Potts models in [19]. Including the value of the partition function $Z(\beta = 0)$ it is possible to retrieve Z for arbitrary β.

3.3. Nested sampling

A sound introduction of NESA and its application to basic problems is presented in [20]. Skilling suggested to rewrite the integral in Eq. (5) as

$$Z = \int d\lambda\, X(\lambda)\ , \tag{15}$$

$$X(\lambda) := \int\!\!\!\!\!\!\diagup d\mathbf{x}\, \pi(\mathbf{x})\, \Theta[L(\mathbf{x}) > \lambda]. \tag{16}$$

The prior mass $X(\lambda)$ accumulates the prior $\pi(\mathbf{x})$ over the parameter space $\mathbf{x}$, which is subject to a constraint on the likelihood. $X(\lambda)$ is constrained to values in the range from 0 to 1, due to the positive and normalized prior $\pi(\mathbf{x})$. Furthermore, Eq. (16) ensures a monotonic decrease of $X(\lambda)$ with an increasing threshold λ. The implementation for the Potts model is outlined below. Instead of integrating over phase space $\mathbf{x}$, we integrate over the prior mass X, which leads to the simplified form of Eq. (5)

$$Z = \int_0^1 dX\, L(X)\ . \tag{17}$$

The same symbol L is used for $L(\mathbf{x})$ and $L(X)$, since the meaning is clear from the context. Details of the transformation can be found in [20]. Finally, the integral in Eq. (17) is approximated by a Riemann sum

$$Z = \sum_{n=1}^{\infty} L(X_n)\Delta X_n \quad , \quad \Delta X_n = X_n - X_{n+1}\ . \tag{18}$$

The monotonicity of $L(X)$ ensures that the sum yields a lower bound for the integral value. Equivalent

$$Z = \sum_{n=1}^{\infty} L(X_n)\, (X_{n-1} - X_n) \tag{19}$$

represents an upper bound. Skilling proposed the following algorithm to compute the $L(X_n)$ values.

Algorithm 3.1: NESA ALGORITHM$(\hat{\lambda}_n, n_{\rm max})$

input parameters: K, k, ϵ_λ
initialize $\hat{\lambda}_0 = 0$, $n = 0$,
draw K configurations $\{x_i\}$ at random from $\pi(\mathbf{x}|\hat{\lambda}_0)$
sort likelihood values $\lambda_i = L(x_i)$ in increasing order
determine the $k^{\rm th}$ smallest likelihood: $\hat{\lambda}_{n=1} := \lambda_k$
while $\hat{\lambda}_{n+1} - \hat{\lambda}_n > \epsilon_\lambda$
do
- $n \leftarrow n + 1$
- discard configurations with $\lambda_i \le \hat{\lambda}_n$
- replace them by new configurations as follows:
- **parallel**
  - start thread $j = 1, 2, ..., k$
  - draw x_n^j from $\pi(\mathbf{x}|\hat{\lambda}_n)$
- determine the $k^{\rm th}$ smallest likelihood $\hat{\lambda}_{n+1}$ of the the entire list of λ-values

set $n_{\rm max} = n$
return $(\hat{\lambda}_n, n_{\rm max})$

In the initialization it is assumed that likelihood values are not negative (hence $\hat{\lambda}_0 = 0$), which will be the case for the Potts model. During the NESA simulation K bond configurations $\mathbf{b}$ (walkers) are treated simultaneously. In each step k of the walkers, those with the smallest likelihood values are replace by new configurations, drawn from the prior subject to the constraint $L(\mathbf{b}) > \hat{\lambda}_n$. The replacement of k walkers is ideally suited for parallelization (see Sec. 4.4). The nested sampling moves in configuration space ensure that even well separated peaks of the likelihood function in configuration space are sampled correctly (see [20]). The crucial step for the nested sampling algorithm is the draw from the constrained prior probability

$$\pi(\mathbf{x}|\hat{\lambda}_{n-1}) = \frac{\pi(\mathbf{x})}{X(\hat{\lambda}_{n-1})}\, \Theta(L(\mathbf{x}) > \hat{\lambda}_{n-1})\ , \tag{20}$$

which represents the normalized prior restricted to areas, where $L(\mathbf{x})$ exceeds the λ threshold. There exist various ways to draw random configurations, also named walkers, from this prior. In the approach employed by us, to get new walker we clone valid ones by choosing k of the remaining $K - k$ walkers at random and update these clones

via MC steps. Given the n_{max} likelihood minima $\hat{\lambda}_n$, the Riemann sum in Eq. (18) can be estimated by

$$\hat{Z} = \sum_{n=1}^{n_{max}} \hat{\lambda}_n \Delta \hat{X}_n. \tag{21}$$

The prior masses $\hat{X}_n$ are random variables, where

$$l_n := -\ln(\hat{X}_n), \tag{22}$$

is Γ-distributed. The moments of $\hat{X}_n$ can therefore be computed easily. In particular for $k = 1$ (i.e. we omit one walker per update) the mean is given by [20]

$$\langle \hat{X}_n \rangle = \xi^n \ , \qquad \text{with } \ \xi := \frac{K}{K+1} \ . \tag{23}$$

In addition, when we omit k walkers per update, the mean and variance of l_1 are

$$\langle\, l_1 \,\rangle = (K-k+1)^{-1} + (K-k+2)^{-1} ... + K^{-1}, \tag{24}$$

$$\langle (\Delta l_1)^2 \rangle = (K-k+1)^{-2} + (K-k+2)^{-2} ... + K^{-2}. \tag{25}$$

Therefore, we can easily draw samples for X_n and compute estimates for the mean and the variance of Z. According to Skilling, this approach becomes problematic for extremely peaked likelihood functions, because the distribution of Z will not be Gaussian any more, but rather the distribution of $\ln(Z)$ [4]. In this case it necessary to base the inference on $\ln(Z)$.

Then an estimate for the variance of $\ln(Z)$ can be determined as follows [20]. First we employ the sample $\hat{l}_n$, drawn from the Gamma distribution, to get an estimate for $\ln(\hat{Z})$. The respective mean of the logarithmic prior masses after n steps, $l_n = n\,\langle\, l_1 \,\rangle$, yields $\ln(\bar{Z})$. An estimate for the variance of $\ln(Z)$ can then be computed via

$$\left\langle \left[\ln(\hat{Z}_K) - \ln(\bar{Z}_K) \right]^2 \right\rangle = \left\langle \ln \left(\frac{\sum_{n=1}^{n_{\max}} e^{-\hat{l}_n} \Delta \hat{\lambda}_n}{\sum_{n=1}^{n_{\max}} e^{-l_n} \Delta \hat{\lambda}_n} \right)^2 \right\rangle . \tag{26}$$

3.4. Application to the Potts model

In [10] the application of the NESA algorithm to the Potts model in the representation of Fortuin and Kasteleyn (FK) is introduced, where the spin variables are replaced by bond variables b_{ij} defined between each pair of neighbouring sites i and j. A bond variable b_{ij} is either active (1) or inactive (0). We denote the entire bond configuration on the lattice by $\mathbf{b}$. For a graphical representation each active bond b_{ij} is represented by a line connecting the sites i and j. The set of sites, connected by lines forms a cluster. An isolated site, to which no line is attached, also qualifies as cluster. In the FK representation two properties of a bond configuration $\mathbf{b}$ are of central importance, the number of active bonds

$$D(\mathbf{b}) = \sum_{\langle i,j \rangle} b_{ij} \tag{27}$$

and the number of clusters $C(\mathbf{b})$. In the FK model the distribution function for the bond variables $\mathbf{b}$ reads

$$P(\mathbf{b}) = \frac{e^{-J\beta N_{\rm p}}}{Z_P} e^{\kappa\, D(\mathbf{b})}\, q^{C(d)} \ , \tag{28}$$

where $N_{\rm p}$ is the number of pairs in the lattice, which is given by $N_{\rm p} = 2N$ for the 2d square lattice and $\kappa = \ln(e^{\beta J} - 1)$. The probability that nearest neighbours with equal spin value form an active bond is defined as

$$p_{\rm b} = 1 - e^{-\beta J} \ . \tag{29}$$

The partition function Z_P in the bond representation is equivalent to the spin-representation and it reads

$$Z_P \ = \ Z_\pi\, e^{-J\beta N_{\rm p}}\, Z_{\rm NESA} \quad , \tag{30}$$

$$Z_{\rm NESA} \ := \ \sum_{\mathbf{b}} L(\mathbf{b})\, \pi(\mathbf{b}) \ . \tag{31}$$

The likelihood function and the prior probability are defined as

$$L(\mathbf{b}) = e^{\kappa\, D(\mathbf{b})} \quad , \quad \pi(\mathbf{b}) = \frac{q^{C(d)}}{Z_\pi} \ . \tag{32}$$

In order to have a normalized prior $\pi(\mathbf{b})$ in Eq. (30) we had to introduce $Z_\pi = \sum_{\mathbf{b}} q^{C(\mathbf{b})}$ as prefactor. It is essential that the unknown prior normalization Z_π is not a function of β. To determine Z_π we can use one of the two limit cases $\beta_1 = 0$ or $\beta_2 = \infty$, for which the partition function is given in Eq. (3). First we note that $J\beta = \ln(2)$ splits the temperature into two regimes, since for $J\beta > \ln(2)$ ($J\beta \leq \ln(2)$) we have $\kappa > 0$ ($\kappa < 0$). Since nested sampling requires monotonically increasing likelihood values, $J\beta > \ln(2)$ ($J\beta \leq \ln(2)$) corresponds to increasing (decreasing) $D(\mathbf{b})$. We therefore have to perform separate NESA runs for these two temperature regimes. We are, however, only interested in $J\beta > \ln(2)$ as it includes the phase transition and the low temperature regime. It also includes the limit $J\beta \to \infty$. For this limit the exact value of Z_P is given in Eq. (3), and we rewrite Eq. (30) as

$$\ln\left(Z_\pi\right) = \ln(q) + J\beta N_{\rm p} - \ln\left(Z_{\rm NESA}(\beta \to \infty)\right) \ . \tag{33}$$

Hence, the prior normalization can be determined if we determine $Z_{\rm NESA}(\beta \to \infty)$ within NESA. How this can be done will be discussed later in Sec. 4.2.

Next we have a closer look at the likelihood definition in Eq. (32) and how it affects the NESA algorithm. As mentioned before we restrict the NESA simulation to $J\beta > \ln(2)$ for which the likelihood constraint has the form $D(\mathbf{b}) > D^*$. We want to discuss some technical details of the NESA algorithm for the Potts model closely related to the ideas outlined in [10]. First of all, we need a MC algorithm to sample from the constrained prior. Interestingly, for $\kappa = 0$, i.e. $J\beta = J\beta_3 = \ln(2)$ the distribution function of the Potts model (see Eq. (28)) coincides with the prior. So we can simply apply Swendsen-Wang for that temperature. The algorithm is then as follows. Starting from a bond configuration $\mathbf{b}$ we perform the steps

1. Identify the clusters
2. For each cluster draw a random colour $\in \{1, \ldots, q\}$ that is assigned to all spins of the cluster
3. Identify the list $\mathcal{L}$ of nearest neighbour pairs with equal spin values. Only elements of $\mathcal{L}$ can become active bonds. Let D_c be the number of bond candidates in $\mathcal{L}$.
4. For each element in $\mathcal{L}$ assign an active bond with probability $p = (1 - e^{-J\beta_3}) = \frac{1}{2}$.

The result yields a new bond configuration $\mathbf{b}'$ drawn according to the prior probability.

Before we can implement the likelihood constraint we have to get rid of likelihood degeneracy, i.e. there are many bond configurations with the same number of active bonds. As outlined in [20] the degeneracy can be lifted by augmenting the phase space by a single additional variable, x say. The walkers now consist of the bond configuration $\mathbf{b}$ and the value x. We introduce a modified likelihood defined as

$$L(\mathbf{b}, x) := e^{\kappa\left(D(\mathbf{b}) + x\, 0^+\right)} \;, \tag{34}$$

where 0^+ is an infinitesimal positive real number. For the joint distribution function we use

$$P(\mathbf{b}, x) := P(\mathbf{b})\; p(x) \;\text{ with }\; p(x) := p_u(x|x_0 = 0) \;, \tag{35}$$

where $p_u(x|x_0)$ is the PDF of a uniform random variable from the interval $(x_0, 1]$, i.e.

$$p_u(x|x_0) := \frac{1}{1 - x_0} \Theta(x_0 < x \leq 1) \;. \tag{36}$$

The additional variable x in the augmented likelihood in Eq. (34) lifts the degeneracy and has otherwise no impact on the likelihood values. The likelihood constraint in the augmented phase space is equivalent to the condition

$$D(\mathbf{b}') + x' > D(\mathbf{b}^*) + x^*. \tag{37}$$

The implementation of this constraint is now in principle an easy task. Given the threshold pair (D^*, x^*) we draw from the prior at random a new walker configuration $(\mathbf{b}', x')$. If it fulfils the likelihood constraint Eq. (37) the new configuration is accepted and it is rejected otherwise. The rejection step can become very time consuming. There is, however, a much more efficient approach to implement this very idea, by avoiding the rejection steps. According to the above rules, bonds are independently activated with probability $p = 1/2$. The number of active bonds, therefore, follows a binomial distribution $P(D'|D_c, p = \frac{1}{2})$. Due to the likelihood constraint $D' \geq D_0$, we need the truncated binomial

$$\tilde{P}(D'|D_0, D_c) := \frac{\Theta(D \geq D_0)}{Z_B(D_0, D_c)} P(D|D_c, p = \frac{1}{2}), \tag{38}$$

where $Z_B(D_0, D_c)$ is the corresponding normalization. Actually, from this distribution only $P_{D>D^*} := \tilde{P}(D > D^*|D^*, D_c)$ is required. In addition we need $P_{x<x^*} := P(x < x^*) = x^*$. By the elementary rules of probability theory we find easily that the probability that the next accepted step in the brute-force approach corresponds to $D > D^*$ is given by

$$\tilde{P}_{D>D^*} = \frac{P_{D>D^*}}{1 - (1 - P_{D>D^*}) P_{x<x^*}}. \tag{39}$$

Now, we can modify step 4 of the SW algorithm to incorporate the likelihood constraint in an rejection-less way:

4a. draw a random number r from $p_u(x|x_0 = 0)$

4b. if $r < \tilde{P}_{D>D^*}$ (i.e. $D' > D^*$)

- Determine at random the number of active bonds according to $\tilde{P}(D'|D^* + 1)$
- draw at random x' form $p_u(x'|x_0 = 0)$

4c. if $r \geq \tilde{P}_{D>D^*}$ (i.e. $D' = D^*$)

- draw at random x' from $p_u(x|x_0 = x^*)$
- set $D' = D^*$

4d. activate at random D' bonds from the list $\mathcal{L}$, resulting in the new bond configuration $\mathbf{b}'$

4e. the new walker configuration is $(\mathbf{b}', x')$

Finally it should be stressed that a single NESA run suffices to compute the partition function for all temperatures $J\beta > \ln(2)$. This is easily achieved by storing the number of active bonds $\hat{D}_n$ instead of the corresponding likelihood minima $\hat{\lambda}_n$, introduced before. Based on Eq. (32) we can determine the likelihood values $\hat{\lambda}_n = e^{\kappa \hat{D}_n}$ for all temperatures and compute the partition function according to Eq. (21).

4. Results

In this section results for the properties of nested sampling applied to the Potts model are provided. The autocorrelation times of successive bond configurations $\mathbf{b}$, evaluated by the nested sampling algorithm, are computed. The dependence of the partition function on the number of NESA steps is illustrated. Results for the partition function and its derivatives are presented and compared to data obtained via a MUCA simulation. An analysis of the computation time required by thermodynamic integration and nested sampling follows. Finally we assess the parallel implementation of the NESA algorithm.

The CPU used in this work is an Intel Core i7-3770K with 32GB RAM running at 3.5GHz. It is a quad-core CPU with hyperthreading, thus providing eight effective hardware threads.

4.1. Autocorrelation times

Successive configurations drawn from the prior as described in chapter 3.3 will exhibit autocorrelations. The impact of these autocorrelations on the nested sampling results shall be analysed in this section for the temperature range $J\beta > \ln(2)$. Given a time series of bond configurations, $\mathbf{b}^{(n)}$ for $n = 1, 2, \ldots, N$, we compute the autocorrelation in the number of active bonds $B_n := D(\mathbf{b}^{(n)})$, i.e.

$$\rho_m := \frac{1}{M} \sum_{n=1}^{M} \Delta B_{n+m} \, \Delta B_n \, , \tag{40}$$

with $M = N - m_{\max}$, and $m_{\max}$ being the maximum lag m for which the autocorrelation is computed. An elementary NESA run starts from a given initial configuration $\mathbf{b}^{(0)}$ with the corresponding likelihood $\lambda^* := L(\mathbf{b}^{(0)})$ and generates a fixed number N of SW updates $\mathbf{b}^{(n)} \to \mathbf{b}^{(n+1)}$ restricted to $L(\mathbf{b}^{(n+1)}) \geq \lambda^*$. The threshold λ^* is not modified during these steps. The sequence of bond configurations $\mathbf{b}^{(n)}$ generated during the run are used to compute the corresponding sequence $B_n = D(\mathbf{b}^{(\mathbf{n})})$ of active bonds and the autocorrelation defined in Eq. (40). Now, for one and the same initial configuration we repeat the elementary NESA run L times (they only differ in the random numbers) and average the individual autocorrelation functions ρ_m resulting in $\overline{\rho}_m$.

Next we determine the integrated autocorrelation time $\tau_{\rm int}$ by the following procedure. The average autocorrelation function $\overline{\rho}_m(t)$ is cut off at m^*, where either an increase or a negative value in $\overline{\rho}_m(t)$ occurs. This is necessary to get rid of the statistical noise in the data. Then we append a single-exponential tail to $\overline{\rho}_m$ for $m > m^*$. The parameters are determined from the second half of the remaining data ($m^*/2 < m \leq m^*$). Finally the integrated correlation time τ_{int} is computed by summing $\overline{\rho}_m(t)$ up to the cut-off value and then adding the contribution of the exponential tail, which can be expressed analytically by a geometric sum [21].

Finally we analyse the impact of the initial configuration $\mathbf{b}^{(0)}$. To this end we perform L elementary NESA runs of length $2N$, all starting from the initial bond configuration $\mathbf{b}_0$ with fixed threshold λ^* (as before) and from all L final configurations we pick the one with the least likelihood and use it as new initial configuration $\mathbf{b}^{(0)}$. For M different initial configurations thus determined we compute individually the integrated autocorrelation times. The actual numbers used are $N = L = 100$ and $M = 1000$. The histogram of the resulting $M = 1000$ integrated correlation times $\overline{\tau}_{int}$ are shown in Fig. 1.

Obviously, the mean of the distribution is roughly 1.75 with a small standard deviation. For a smaller (3×3) system the mean integrated correlation time yields for different q-values similar results

- 1.54 ± 0.26 for $q = 2$,
- 1.71 ± 0.03 for $q = 5$.

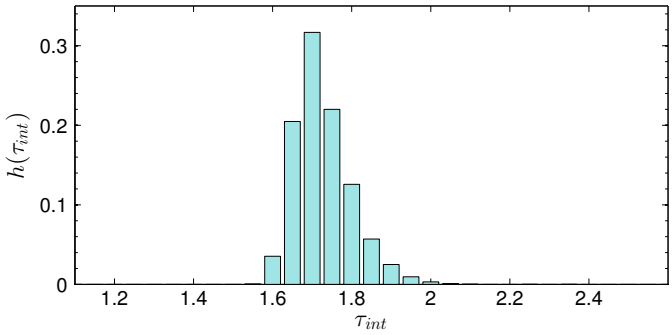

Figure 1: Histogram of the integrated correlation times $\overline{\tau}_{int}$ corresponding to the autocorrelation function in Eq. (40) for a 16×16 Potts model with $q = 10$.

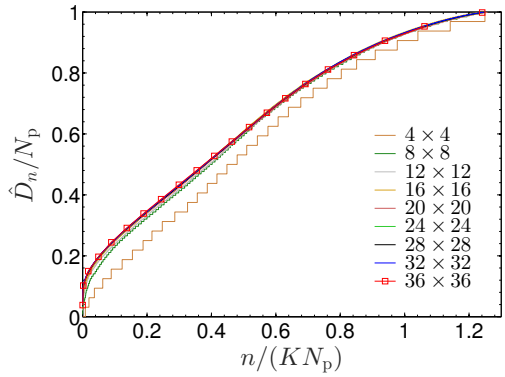

Figure 2: $\hat{D}_n$ for the Potts model with $q = 10$ and $K = 100$ for various system sizes.

The small value for the correlation times $\bar{\tau}_{int}$, almost independent of the system size and q, gives reason to assume that successive states of a walker during its evolution in NESA are not significantly correlated. Moreover, as outlined above, there is no temperature dependence in the values $\hat{D}_n$ and consequently, the autocorrelation is also independent of temperature. It should be emphasized that for NESA, i.e. the prior sampling, we employ Swendsen-Wang only for the temperature $J\beta = \ln(2)$. That is why we do not have to bother with critical slowing down, that occurs near the critical temperature $J\beta_c$.

4.2. Partition function and other thermodynamic quantities

In statistical physics we are interested in the dependence of the partition function on the inverse temperature β, as it provides the entire thermodynamic information of the system. As emphasized before, the dependence of $\ln(Z)$ on β (for $J\beta > \ln(2)$, i.e. $\kappa > 0$) can be obtained from a single NESA run, since the likelihood $e^{\kappa D(\mathbf{b})}$ is a monotonic function in the number of active bonds $D(\mathbf{b})$. Instead of generating the increasing sequence of likelihood values, we can equally well generate the corresponding increasing sequence of active bonds $\hat{D}_n$, from which the partition function can be computed for all temperatures $J\beta > \ln(2)$. The upper limit of active bonds $D_{\max}$ is given by the num-

ber of pairs N_p. In Fig. 2 we depict the fraction of active bonds, i.e. $\hat{D}_n/N_\mathrm{p}$ versus $\xi := n/(KN_\mathrm{p})$ for various system sizes. Obviously, with increasing system size the results rapidly converge towards a universal curve. An important finding is that the number of NESA steps n_max it takes to reach D_max is proportional to the number of pairs and the number of walkers.

A separate study for q-values between 1 and 10 yields the following scaling behaviour

$$n_\mathrm{max} = (a + bq + cq^2)KN_\mathrm{p}\ , \tag{41}$$

with $a = 0.606, b = 0.096, c = -0.003$ for a 8×8 system and $a = 0.599, b = 0.089, c = -0.002$ for a 16×16 system. We find that the q-dependent prefactor is nearly independent of the system size and it is a very smooth function in q. There is no distinction in the behaviour for system that have a first or second order phase transition. The key message so far is, however, that NESA needs at most $O(1)KN_\mathrm{p}$ steps. The importance of the various values $\hat{D}_n$ for the partition function, however, is not clear yet. In order to address this aspect, we discuss additional aspects of Fig. 2. We see that $\hat{D}_n$ represents a staircase with steps that have an average height of ≈ 1 and average width of $\approx K$. We denote the position, at which the ν-th step begins, by n_ν and the corresponding step height by $\hat{D}_{n_\nu}$. Based on Eq. (23) we find $\langle \Delta X_n \rangle = \xi^n(1-\xi)$ and the mean partition function can then be written as

$$\begin{aligned} \langle Z_\mathrm{NESA} \rangle &= \sum_\nu S_\nu \\ \text{with} \quad S_\nu &:= e^{\kappa \hat{D}_{n_\nu}} \left(\xi^{n_\nu} - \xi^{n_{\nu+1}}\right). \end{aligned} \tag{42}$$

The summand S_ν represents the contribution of the ν-th step to the partition function, which clearly depends on temperature. In Fig. 3 the normalized summands S_ν/S^max ($S^\mathrm{max} = \max_\mu S_\mu$) are plotted as function of the step position n_ν scaled by KN_p for three inverse temperatures, below, at, and above β_c. For $J\beta = 1$ one observes that only a small fraction of the staircase in Fig. 2 is sufficient for a converged result. For $J\beta_c$ already a significant fraction is required and for $J\beta = 4$ the last summands clearly dominate the partition function. Hence for low temperatures (large $J\beta$) an accurate estimate for the partition function is only possible, if the NESA algorithm reaches the maximal number of active bonds $\hat{D}_n = N_\mathrm{p}$. Therefore the CPU-time scales like N^2, because the number of required NESA steps is proportional to N, and each NESA step involves one Swendsen-Wang update, which explains the second factor N.

We are now in the position to determine the prior normalization Z_π within a single NESA run. According to Eq. (33) we need $Z_\mathrm{NESA}(\beta \to \infty)$. For $J\beta \to \infty$ we have $\kappa = J\beta$ and Eq. (18) yields

$$Z_\mathrm{NESA}(\beta) = \sum_{n=0}^{\infty} e^{J\beta \hat{D}_n} \Delta X_n\ . \tag{43}$$

For $J\beta \to \infty$ only the terms with the maximal value of $\hat{D}_n$ (i.e. $\hat{D}_n = N_p$) survive, which is the case for $n \geq n_\mathrm{max}$. We therefore have

$$Z_\mathrm{NESA}(\beta) \ \to\ e^{J\beta}\, S(n_\mathrm{max})\,, \tag{44}$$

$$S(n_\mathrm{max}) \ :=\ \sum_{n=n_\mathrm{max}}^{\infty} \Delta X_n\,. \tag{45}$$

The mean of ΔX_n is given by $\Delta X_n = \xi^n(1-\xi)$ with $\xi = e^{-1/K}$ (see [20]) and we find for Eq. (33)

$$\ln\left(Z_\pi\right) = \ln q + \frac{n_\mathrm{max}}{K}\ .$$

Based on our finding, $n_\mathrm{max} = \alpha N_p K$, we obtain

$$\frac{\ln\left(Z_\pi\right)}{N_p} = \frac{\ln q}{N_p} + \alpha\ ,$$

where $\alpha = O(1)$. We see that in the thermodynamic limit $\ln\left(Z_\pi\right) \to \alpha N_p$, which is consistent with the requirement that $\ln\left(Z_\pi\right)$ is an extensive quantity, as it is proportional to the free energy. Moreover, we observe that with increasing system size the distribution of n_{max} approaches a Poisson distribution. Consequently, Z_π can be determined from a single NESA run with high accuracy.

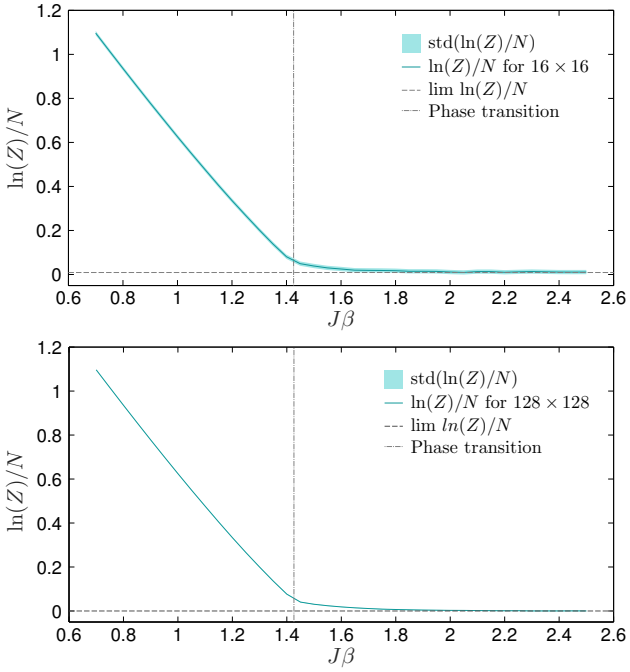

Figure 4: $\ln(Z)/N$ for a q=10 system versus $J\beta$ evaluated via nested sampling with $K = 100$ walkers. The inverse critical temperature $J\beta_c$ is marked by a vertical chain line and the exact limit value for $J\beta \to \infty$ (see Eq. (3)) is represented by a dashed line. System sizes are 16×16 (left panel) and 128×128 (right panel).

Next we turn to analysis of the partition function. Fig. 4 displays $\ln(Z)/N$ versus $J\beta$ for 16×16 and 128×128 lattices. For the 128×128 system the statistical uncertainties

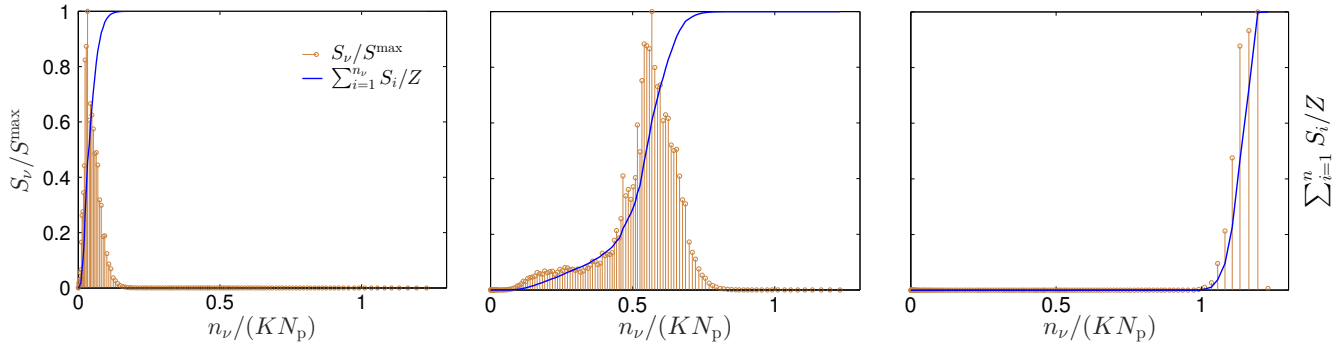

Figure 3: Summands S_ν of Eq. (42) and the corresponding cumulative sums for a 8×8 system and $q = 10$ system at different temperatures (from left to right): $J\beta = 1$, $J\beta_c \approx 1.426$, $J\beta = 4$. With increasing $J\beta$ the main contribution to the sum shifts to larger numbers of NESA updates.

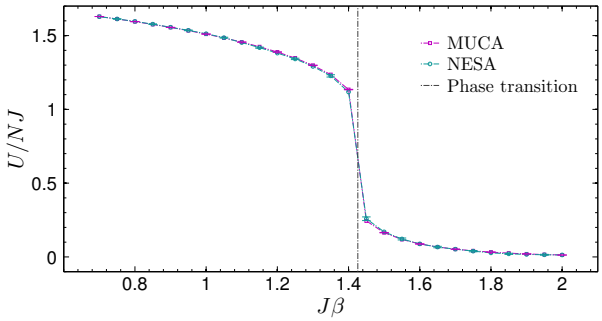

Figure 5: Internal energy U per site versus $J\beta$ evaluated via nested sampling and via multi-canonical sampling for the 20×20 and $q = 10$ Potts system. The inverse critical temperature $J\beta_c$ is marked by a vertical chain line.

are smaller than the line thickness. The inverse temperature of the phase transition for an infinite 2d square lattice as given in Eq. (4) is indicated a by a vertical line. Above $J\beta_c$, the curve approaches rapidly the limit value of $\ln(Z)/N = \ln(q)/N$ for $J\beta = \infty$, which is depicted by a dashed horizontal line.

The partition function of a 10 state Potts model was evaluated via nested sampling for grid-sizes up to of 256×256. The total computational time for the largest system was about 57 h. For an equivalent result for the same system size the thermodynamic integration method would take about 4 years.

Given $\ln(Z)$ as function of $J\beta$ for a system in the canonical ensemble, thermodynamic quantities like the Helmholtz free energy F and the internal energy U as well as the entropy S and heat capacity c_V can be deduced [22]. Starting from the estimate of the partition function in Eq. (21), expressions for the first and second derivative of $\ln(Z)$ with respect to β can be determined. Given these expressions we can evaluate the physical quantities analytically, based on the sequence of active bonds $\hat{D}_n$ from a single NESA run and the sample of prior masses. Hence we can avoid the determination of numerical derivatives and the associated errors. From Eq. (32) we have

$$\ln(L(\mathbf{b})) = \kappa\, D(\mathbf{b})\ . \tag{46}$$

Given the sequence $\hat{D}_n$ from a single NESA run and the corresponding prior masses ΔX_n we can calculate the logarithm of the partition function

$$\ln(Z_\mathrm{NESA}) = \ln\Big(\sum_n e^{\kappa \hat{D}_n}\, \Delta X_n\Big) \tag{47}$$

and the derivative with respect to β

$$\frac{\partial}{\partial\beta}\, \ln(Z_\mathrm{NESA}) = \frac{J}{1-e^{-J\beta}}\, \underbrace{\frac{\sum_n e^{\kappa \hat{D}_n}\, \hat{D}_n\, \Delta X_n}{Z_\mathrm{NESA}}}_{:=\langle D\rangle_\beta}\ . \tag{48}$$

According to Eq. (30) we find e.g. for the internal energy

$$U = \langle H\rangle = -\frac{\partial}{\partial\beta}\, \ln(Z) = JN_\mathrm{p} - \frac{J\langle D\rangle_\beta}{1-e^{-J\beta}}\ . \tag{49}$$

The mean number of active spins is related to the mean number of nearest neighbour pairs with equal spin $\langle N_\mathrm{eq}(\mathbf{s})\rangle_\beta$ via $\langle D\rangle_\beta = p_\mathrm{b}\langle N_\mathrm{eq}\rangle_\beta$, where p_b is the probability that nearest neighbour pairs of equal spin form an active bond (see Eq. (29)). Therefore, the internal energy can also be expressed as

$$U = -J\big(\langle N_\mathrm{eq}\rangle_\beta - N_\mathrm{p}\big)\ , \tag{50}$$

which is in agreement with the relation $U = -\frac{\partial}{\partial\beta}\langle H\rangle$ and the definition of the Hamiltonian in Eq. (1). The second derivative can be deduced similarly.

Fig. 5 shows the internal energy U versus $J\beta$ for a 20×20 and $q = 10$ Potts system computed via NESA and a MUCA simulation. For NESA $K = 500$ walkers have been used. The values agree excellently. For the MUCA simulation the Fortran code provided by Berg [3] is employed. For the comparison of the NESA results to the MUCA results from [3] the different definitions of β and the Hamiltonian have to be taken into account.

4.3. Performance comparison of thermodynamic integration and nested sampling

For a proper comparison of the two methods, NESA and TI, some important points need to be considered. Thermodynamic integration yields the values of the partition

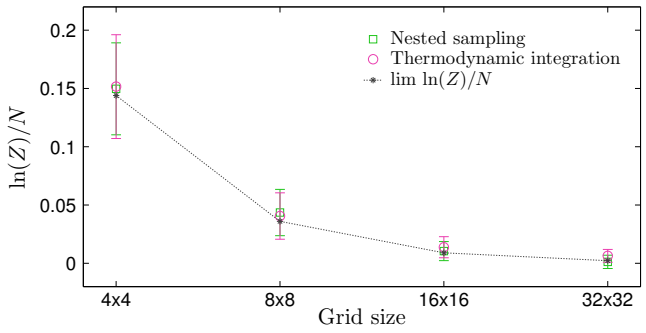

Figure 6: $\ln(Z)/N$ for the q=10 Potts model at $J\beta = 10$ computed by thermodynamic integration and nested sampling in dependence of the grid-size. The limit value of $\ln(Z)/N$ for $J\beta \to \infty$ (see Eq. (3)) is indicated by black asterisks. The dotted line serves as guide to the eye.

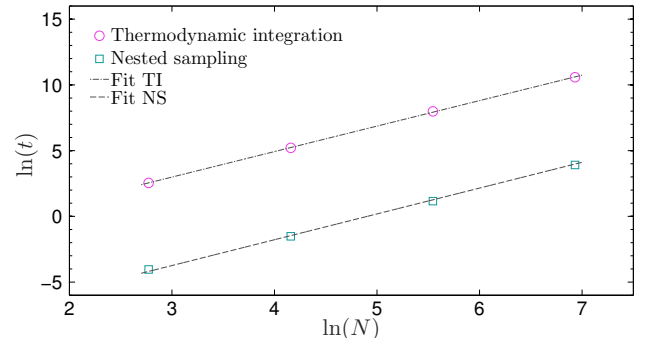

Figure 7: For the q=10 Potts model at $J\beta = 10$, the logarithmic time for computing $ln(Z)$ versus the logarithmic grid size exhibits a similar scaling for both methods.

function only at the values of the $J\beta$ steps explicitly required in Eq. (9). Remembering Eq. (12), we find that only terms contribute to the sum, if not all bonds are active. It follows, that above a certain $J\beta$ value, there will be no relevant contributions to the sum anymore. Hence it is sufficient to calculate the partition function $\ln(Z)$ with thermodynamic integration up to a value of e.g. $J\beta = 10$ for the investigation of the $q = 10$ system. Above this value no significant variation in the result of $\ln(Z)$ will occur.

The most time consuming element of both methods is the SW update. NESA requires, due to the negligible correlations, only 1 SW update for each NESA step. When employing thermodynamic integration, however, for each $J\beta$ value multiple SW updates are necessary to obtain a reliable thermodynamic expectation value $\langle \ln(L(\mathbf{b})) \rangle$. For the fair comparison of NESA and TI the number of SW updates in TI has been chosen such that the resulting error bars of $\ln(Z)$ at $J\beta = 10$ are of approximately the same size (see Fig. 6) in both methods. For system sizes ranging from 4×4 to 32×32 the autocorrelation times are computed for all $J\beta$ values (for details on the τ_{int} computation see Sec. 4.1). This evaluation enables us to correct the variances computed by thermodynamic integration at each $J\beta$ value. For NESA we used a sample of 100 walkers and 50 prior masses in the following computations. The results for $\ln(Z)/N$ computed by NESA and by TI are displayed in Fig. 6. For the q=10 Potts model at $J\beta = 10$, as shown in Fig. 7, the logarithmic time for computing $\ln(Z)$ versus the logarithmic grid size exhibits a quadratic scaling for both methods. The quadratic behaviour for NESA has already been explained in a previous section.

The agreement of the scaling laws appears stunning at first sight, because nested sampling does not require to increase the number of updates near the phase transition, there is no critical slowing down. As a matter of fact, we use one and the same NESA run for all temperatures and the the phase transition is not special in any respect. However, while TI needs to increase the number of Swendsen-Wang updates for larger systems due to rising correlation times, nested sampling needs more updates to reach the maximum likelihood value (maximum number of active bonds) necessary to compute the partition function for large $J\beta$ reliably, as is obvious from Fig. 3. This increase with system size appears to scale similarly for both methods. Eventually, the reason for the overall prefactor in the performance is that for each expectation value $\langle \ln(L(x)) \rangle_\beta$ in the thermodynamic integration, we need to evaluate multiple SW updates.

4.4. Parallel nested sampling

In this section we analyse the parallel NESA algorithm (Alg. 3.1 for $k > 1$) as proposed by Henderson *et al.* [12] in the frame of the Potts model. In the parallel NESA algorithm k walkers with the smallest likelihood values are discarded at each step. The update of each discarded walker is calculated in parallel on separate cores of the CPU. Henderson *et al.* derived the required scaling of the number of live walkers K_k with respect to k, to ensure a constant variance of $\ln(Z)$. The scaling is given by

$$K_k \approx \sqrt{k}\, K_1. \tag{51}$$

For the case of $k = 1$ we employed a sample of $K_1 = 100$ walkers. For $k > 1$, K_n is scaled according to Eq. (51). The computations for $k = 1, 2, 4$ are performed on a processor with 4 cores. The evaluation of each $\ln(Z)$ value uses 100 independent sequences of prior samples X_n. The error of $\ln(Z)$ is evaluated according to Eq. (26). Fig. 8 displays the wall-clock time of the parallel NESA evaluations for the analysed grid-sizes ranging from 4×4 to 64×64. The wall-clock time depicts the real time, which passes by for the user, while computing a task. It also includes the time needed for e.g. input and output operations. The values displayed in Fig. 8 are normalised with respect to the case of $k = 1$. Each value is the mean of five independent runs. For clarity the respective standard deviations are only displayed for the measurements at $k = 4$. One can observe that for increasing grid-sizes the increase in wall clock time, over the number of cores, shrinks. For the 64×64 system we find an increase of about 10 percent for 4 cores compared to the case of 1 core. The slight kink in

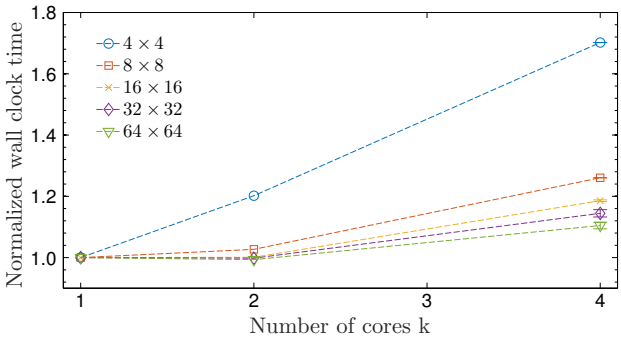

Figure 8: Normalized wall clock time required by parallel NESA for the computation of $\ln(Z)$ for the $q = 10$ Potts model at $J\beta = 10$ for various grid-sizes in dependence of the number of employed cores k.

the plot is presumably caused by reaching the full capacity of the processor, when employing all 4 cores. Although the number of walkers in the NESA runs has been scaled to yield constant variance, we find a slight decrease in the error of $\ln(Z)$ for all system sizes.

Overall the parallel NESA algorithm does not yield a distinct improvement for the evaluation of the partition function of the 10 state Potts model at $J\beta = 10$. The decreasing trend in variance as well as the trend to constant values for the wall-clock time for larger systems suggest that for systems with a more structured likelihood function, the parallel NESA would clearly be advantageous, since it involves more walkers within almost the same time. This is advantageous for multi-modal likelihood functions, where we need enough walkers per mode if we employ the present clone approach (see Sec. 3.3).

5. Summary and conclusions

In this work we have benchmarked the nested sampling algorithm in the frame of the Potts model on a 2d square lattice for different sizes. As alternative method, the established thermodynamic integration has been employed. The correlation time of thermodynamic integration, displays a massive slowing down near the critical temperature. The determination of correlation times for nested sampling yielded negligible values. Furthermore we have deduced expressions to determine thermodynamic quantities from the results of a single nested sampling run, without employing numerical derivatives.

Both methods exhibit a power law scaling of the computation time with increasing grid-sizes. Fits yielded roughly a quadratic exponent for both methods. Though, due to a much smaller prefactor, the nested sampling algorithm is about three orders of magnitude faster than thermodynamic integration. At first sight, it appears stunning that the two algorithms have the same scaling law, because nested sampling does not require to increase the number of updates near the phase transition, due to the absence of critical slowing down. However, while TI needs to increase the number of Swendsen-Wang updates for larger systems due to rising correlation times, nested sampling needs more updates to reach the maximum likelihood value (maximum number of active bonds) necessary to compute the partition function for large $J\beta$ reliably. This increase appears to scale similarly for both methods. Eventually, the reason for the overall prefactor in the performance is that for each expectation value $\langle \ln(L(x)) \rangle_\beta$ in the thermodynamic integration, we need to evaluate multiple SW updates. Another important advantage of nested sampling is, that the partition function for all temperatures is available from the results of a single run. We have compared the internal energy with results from a multi-canonical simulation [3] and found excellent agreement. Finally a parallel version of the nested sampling algorithm was investigated. We have employed 1 to 4 cores for system sizes ranging from 4×4 to 64×64. The wall clock times as well as the respective results for the partition function did not reveal a relevant improvement for the application of the parallel version of NESA for the Potts model. The reason being that parallel NESA requires more walkers for the same statistical error. On the other hand, more walkers lead to a better exploration of phase space, which is important for problems involving a multi-modal likelihood. In case of the Potts model this aspect was of minor importance, which is why parallel NESA was not really advantageous.

In summary we have found that nested sampling is able to deal efficiently with problems that exhibit first order phase transitions. The method is a promising alternative to the multi-canonical and multi-bondic algorithms, which are state of the art computational techniques for dealing with Potts-type of models. In our opinion nested sampling represents a high potential algorithm for applications in statistical physics and due to its uniqueness it deserves a place in a physicists standard repertoire of simulation techniques.

ACKNOWLEDGMENTS

We acknowledge fruitful discussions with H. G. Evertz and U. von Toussaint and the kind support of I. Murray who provided us with the code he developed for the results in [10].

References

[1] D. P. Landau, K. Binder, A guide to Monte Carlo simulations in statistical physics, Cambridge University Press, Cambridge; New York, 2009.

[2] B. A. Berg, Introduction to Multicanonical Monte Carlo Simulations, Fields Inst.Commun.26 1-24, 2000.
URL http://arxiv.org/abs/cond-mat/9909236

[3] B. A. Berg, Multicanonical simulations step by step, Computer Physics Communications 153 (3) (2003) 397–406. doi:10.1016/S0010-4655(03)00245-5.

[4] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Anal. 1 (4) (2006) 833–859, Mathematical Reviews number (MathSciNet) MR 2282208. doi:10.1214/06-BA127.

[5] S. Aitken, O. E. Akman, Nested sampling for parameter inference in systems biology: application to an exemplar

circadian model, BMC Systems Biology 7 (1) (2013) 72. doi:10.1186/1752-0509-7-72.

[6] N. S. Burkoff, C. Varnai, S. A. Wells, D. L. Wild, Exploring the Energy Landscapes of Protein Folding Simulations with Bayesian Computation, Biophys J 102 (4) (2012) 878–886. doi:10.1016/j.bpj.2011.12.053.

[7] F. Feroz, J. Skilling, Exploring Multi-Modal Distributions with Nested Sampling, AIP Conference Proceedings, 1553, pp. 106-113 (2013) 106–113doi:10.1063/1.4819989.

[8] P. Mukherjee, D. Parkinson, A. R. Liddle, A Nested Sampling Algorithm for Cosmological Model Selection, The Astrophysical Journal 638 (2) (2006) L51–L54. doi:10.1086/501068.

[9] H. Qiao, P. Pal, Generalized nested sampling for compression and exact recovery of symmetric Toeplitz matrices, in: 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2014, pp. 443–447. doi:10.1109/GlobalSIP.2014.7032156.

[10] I. Murray, D. MacKay, Z. Ghahramani, J. Skilling, Nested sampling for potts models, in: Y. Weiss, B. Schölkopf, J. C. Platt (Eds.), Advances in Neural Information Processing Systems 18, MIT Press, 2006, pp. 947–954.

[11] W. Janke, Rugged Free-Energy Landscapes: An Introduction, in: Rugged Free Energy Landscapes, no. 736 in Lecture Notes in Physics, Springer Berlin Heidelberg, 2008, pp. 1–7. doi:10.1007/978-3-540-74029-2_1.

[12] R. W. Henderson, P. M. Goggans, Parallelized nested sampling, in: AIP Conference Proceedings, Vol. 1636, AIP Publishing, 2014, pp. 100–105. doi:10.1063/1.4903717.

[13] S. Martiniani, J. D. Stevenson, D. J. Wales, D. Frenkel, Superposition Enhanced Nested Sampling, Phys. Rev. X 4 (3) (2014) 031034. doi:10.1103/PhysRevX.4.031034.

[14] W. Vanderbauwhede, S. Lewis, D. Ireland, Implementing data parallelisation in a Nested-Sampling Monte Carlo algorithm, in: 2013 International Conference on High Performance Computing and Simulation (HPCS), 2013, pp. 512–518. doi:10.1109/HPCSim.2013.6641462.

[15] F. Y. Wu, The Potts model, Rev. Mod. Phys. 54 (1) (1982) 235–268. doi:10.1103/RevModPhys.54.235.

[16] R. J. Baxter, Potts model at the critical temperature, J. Phys. C: Solid State Phys. 6 (23) (1973) L445. doi:10.1088/0022-3719/6/23/005.

[17] U. von Toussaint, Bayesian inference in physics, Rev. Mod. Phys. 83 (3) (2011) 943–999. doi:10.1103/RevModPhys.83.943.

[18] R. H. Swendsen, J.-S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, Phys. Rev. Lett. 58 (2) (1987) 86–88. doi:10.1103/PhysRevLett.58.86.

[19] K. Rummukainen, Multicanonical methods and cluster algorithms (2008).

[20] W. von der Linden, V. Dose, U. von Toussaint, Bayesian Probability Theory: Applications in the Physical Sciences, Cambridge University Press, Cambridge, 2014.

[21] C. F. Baillie, P. D. Coddington, Comparison of cluster algorithms for two-dimensional Potts models, Phys. Rev. B 43 (13) (1991) 10617–10621. doi:10.1103/PhysRevB.43.10617.

[22] R. K. Pathria, P. D. Beale, Statistical Mechanics, Academic Press, 2011.