



UNIVERSITÄT
LEIPZIG

Master of Science

Generalized Microcanonical Population Annealing for First Order Phase Transitions

Till Pfaff

2024

Computational Quantum Field Theory (CQT)
Institut für Theoretische Physik
Leipzig University

Primary supervisor:	Prof. Dr. Wolfhard Janke
Secondary supervisor:	Denis Gessert
Date:	September 18, 2024

Abstract

This thesis introduces and evaluates modifications to population annealing (PA), specifically the momentum and mixed ensemble, which are derived from the microcanonical ensemble. There, the total energy \mathcal{E} is the control parameter, instead of the inverse temperature β . We particularly test the new methods on their performance on simulating first-order phase transitions which is an area where canonical ensemble PA struggles.

We apply these methods to the Ising, Potts, and Blume-Capel models. The data indicated that the momentum PA performs comparably to canonical PA in the Ising model but outperforms it in the Potts model by successfully suppressing the first-order phase transition. Additionally, the mixed ensemble approach provides a method for annealing isolated variables in a multiple term Hamiltonian, allowing effective simulations of the Blume-Capel model.

Contents

1	Introduction	1
2	Theory and Methods	2
2.1	First- and Second-Order Phase Transitions	2
2.2	Monte Carlo Algorithms	3
2.3	Population Annealing	6
2.3.1	Choice of Annealing Schedule	8
2.3.2	Estimation of Canonical Observables	9
2.3.3	Estimation of the Partition Function	10
2.4	Correlations	10
2.5	Microcanonical Ensemble	12
2.5.1	Ceiling Ensemble	13
2.5.2	Momentum Ensemble	13
2.5.3	Role of Weight Function Exponent	14
2.5.4	Mixed Ensemble	15
2.6	Histogram Reweighting Techniques	15
2.7	Use of the Ensembles in Population Annealing	17
2.8	Models	17
2.8.1	Ising Model	17
2.8.2	Potts Model	18
2.8.3	Blume-Capel Model	19
3	Results	21
3.1	Ising Model	21
3.1.1	Comparison to Exact Solution	21
3.1.2	Family Observables	21
3.2	Potts Model	24
3.2.1	Family Observables	26
3.2.2	Weight Function Exponent	26
3.3	Blume-Capel Model	29
4	Conclusion and Outlook	31

Contents

Appendices	33
A Derivation of the Variance of the Number of Copies	33
B Approximation of the Histogram Width in the Momentum Ensemble . . .	34
Acknowledgements	35
Bibliography	36
Declaration of Authorship	38

List of Figures

2.1	Visualization of the algorithm. The different colors represent the statistically independent initial replicas, while the differing shades stand for the decorrelating effects of the MCMC routine.	7
2.2	Phase diagram of the Blume-Capel model [24].	19
3.1	(a) (Canonical) specific heat for $L = 32$ estimated from the different ensembles using Metropolis update algorithm and $\theta = 1, R = 5000$, weight exponent $\alpha = 1$, compared to the exact solution (dashed line). (b) Deviation $\epsilon(E) = (E - E_{\text{exact}})/E_{\text{exact}}$ from the exact solution, the dashed line shows the pseudo-critical point. (c,d) $\theta = 20$, respectively.	22
3.2	(a) Canonical energy distribution $P(e)_{\beta_c}$ at the (exact) pseudo-critical point for $L = 32$ estimated from the different ensembles using Metropolis update algorithm and $\theta = 1, R = 5000$, weight exponent $\alpha = 1$, compared to the exact solution (dashed line). (b) $\theta = 20$, respectively.	22
3.3	(a) Average family size ρ_t for different ensembles for the $L = 32$ Ising model using Metropolis update algorithm and $\theta = 1, R = 5000$. (b) $\theta = 20$ respectively.	23
3.4	(a) Culling fractions obtained from the exact solution [20]. (b) Comparison of the observed ($L = 32, R = 5000$) value of ρ_t to the optimal value obtained from the exact solution in the ceiling ensemble.	24
3.5	(a) Canonical average energy in $q = 10, L = 32$ Potts model estimated in the three ensemble methods with $R\theta f \approx \text{const}$. (b) Specific heat. (c) Average magnetization. (d) Magnetic susceptibility.	25
3.6	Canonical energy distribution $P(e)_{\beta_c}$ at the (exact) pseudo-critical point $\beta_c = 1.423$ [25] for $q = 10, L = 32$, estimated from the different ensembles using the heatbath update algorithm and $\theta = 20, R = 10000$, weight exponent $\alpha = 1$	26
3.7	(a) Average family size ρ_t for different ensembles for the $L = 32, q = 10$ Potts model using heatbath update algorithm and $\theta = 1, R = 10000$ for the canonical and momentum PA run and $R = 40000$ for the ceiling PA in order to properly resolve at large ρ_t . (b) $\theta = 20$, with $R = 10000$ and $R = 20000$, respectively.	27

3.8	(a) Average family size ρ_t for different momentum ensemble with $\alpha = 1, 25, 50$ for the $L = 32, q = 10$ Potts model using heatbath update algorithm and $\theta = 20, R = 5000$. (b) Critical energy distribution of the $L = 32, q = 10$ Potts model in the momentum ensemble with varying α , compared to canonical critical energy distribution.	27
3.9	(a) Critical energy distribution of the $q = 10$ Potts model for various system sizes in the momentum ensemble with $\alpha = 1/4L$. (b) Scaling of estimated $\tilde{C}_{\mathcal{E}_e} = N\sigma_e^2$ compared to approximation in (2.44). (c,d) $\alpha = 2L$, respectively.	28
3.10	(a,b) Canonical specific heat C of the $L = 20$ Blume-Capel model, simulated with canonical, momentum and mixed ensemble, with $\theta = 20, R = 10000$ and the same number of (adaptive) annealing steps, estimated from 16 independent PA runs for each method. The dashed line shows the transition point $(\beta_c, \Delta_c) = (2, 1.989)$. (c,d) Canonical distribution of E_Δ at the first-order transition point (β, Δ) estimated from the same simulation runs.	29
3.11	Average family size ρ_t of the same simulation runs shown in Fig. 3.10.	30
B.1	(a) Density of states $\Omega(e)$ of the $q = 10$ Potts model for various system sizes. (b) Slope of $\ln \Omega(e_c)$ with e_c in the region where $\ln \Omega$ is linear.	34

List of Abbreviations

Abbreviations	Meaning
d.o.f.	degree of freedom
DOS	Density of States
ECSD	Exponentially Critical Slowdown
l.h.s.	left hand side
r.h.s.	right hand side
MCMC	Markov Chain Monte Carlo
MC	Monte Carlo
MCS	Monte Carlo Sweep
MHR	Multi-Histogram Reweighting
RNG	Random Number Generator
PA	Population Annealing

1 Introduction

The recent advent of artificial intelligence and machine learning has pushed parallel computing hardware to new extremes: The training of the large language model GPT-3, for example, reportedly used an astounding 10,000 high-power GPUs with tens of thousands of cores each [1].

This development spurred new interest in utilizing this power in the field of statistical physics as well. In Monte Carlo simulations, a common task that benefits particularly well from parallelization is the analysis of rugged free energy landscapes, as are found, for example, in glassy systems. A widely used algorithm for this task is parallel tempering (PT) [2, 3], which is, however, inherently limited in its ability to scale in parallel. For this reason, there has been growing interest in another parallel algorithm to study rugged free energy landscapes, namely population annealing (PA).

PA does not perform well in systems with first-order transitions [4], though, as these systems involve large energy barriers [5], that the algorithm is not very well suited to overcome. Generalized ensemble algorithms such as multicanonical ensemble methods [6] and Wang-Landau sampling [7] are designed to tackle first-order transitions, but involve intricate iterative procedures.

This work uses the microcanonical ensemble in the framework of population annealing. This ensemble has been utilized before in parallel tempering [8] and also population annealing [9], which this work will generalize upon. This is achieved by introducing what we call the *momentum ensemble* [8, 10] to population annealing, which we generalize further to specifically fit it to the studied model. Our goal is to evaluate the different ensembles in context of PA and provide an initial assessment of their statistical and systematic error.

The following chapters will be structured as follows: Chapter 2 gives a review of the theoretical background, starting with phase transitions as described in statistical physics, and going over traditional Markov Chain Monte Carlo methods. Subsequently, we will introduce the population annealing algorithm and give a description of the employed ensembles and their implementation in PA. Chapter 3 presents the results obtained from applying these methods to the Ising, Potts, and Blume-Capel models. Finally, the conclusion summarizes the findings and gives a brief outlook into issues that would benefit from further examination.

2 Theory and Methods

2.1 First- and Second-Order Phase Transitions

When varying a thermodynamic variable like temperature or pressure, a system sometimes undergoes a sudden drastic change in its properties. This phenomenon is known as a phase transition. Some common examples are the conversion from liquid to solid (or gas), the sudden mixing or separating of fluids or a rapid change in the magnetization from zero to a non-zero value. Generally, the transition shifts the system from a more ordered to a less ordered phase, or vice versa.

If a system is placed in an environment with constant temperature T , and it can only exchange energy with this environment, the system is said to be in the canonical ensemble and its equilibrium distribution is governed by the Boltzmann distribution

$$p_\beta(\mathcal{H}) = \frac{1}{\mathcal{Z}_\beta} e^{-\beta\mathcal{H}}, \quad (2.1)$$

where β is the inverse temperature $1/(k_B T)$ ¹, \mathcal{H} is the system's Hamiltonian and \mathcal{Z}_β is the canonical partition function. It not only normalizes the probabilities, but can also be used to calculate all thermodynamic properties the system exhibits in the ensemble.

In the thermodynamic limit, where the number of particles N tends to infinity, phase transitions are identified by singularities in the partition function or the free energy

$$\beta F = -\ln \mathcal{Z}_\beta, \quad (2.2)$$

respectively. If the first derivative of the free energy with respect to the order parameter is discontinuous, the transition is labeled as first-order. The discontinuity has the consequence that two phases can coexist at the critical point, separated by an energy difference known as the latent heat Δe [5]. This causes fluctuations to stay localized, keeping the correlation length ξ finite at the critical point.

In contrast, if the second derivative of the free energy is discontinuous or singular, while the first derivative stays continuous, the phase transition is called continuous, or second-order. In this case, there is no phase coexistence and the system requires long-range order to

¹ k_B is the Boltzmann constant. In this work, we set $k_B = 1$.

change from one phase to the other, in other words, ξ diverges at the critical point. This loss of a characteristic length scale gives rise to power-law critical phenomena described by critical exponents and universality [11]. For example, near criticality $\xi \sim |1 - T/T_c|^{-\nu}$ and the specific heat $C \sim |1 - T/T_c|^{-\alpha}$ where ν and α are critical exponents.

For systems with a finite size L^D , there cannot be any true discontinuities since the free energy remains analytic. Thus, phase transitions are only defined in the thermodynamic limit. Still, at finite sizes, there remain clear leftovers of both first- and second-order transitions, which are differentiated from their infinite-size counterparts by dubbing them pseudo-transitions [12].

For first-order transitions, we see a double peak distribution of energies at the pseudo-critical point, with the suppressed region in between the two peaks containing states with an interface. The suppression scales $\sim \exp(-\beta\sigma L^{D-1})$, with σ being the interface tension [12]. In standard simulations of the canonical ensemble, this causes the relaxation time to grow exponentially with L^{D-1} , since the travel between the two phases is exponentially unlikely (*exponentially critical slowing down*) [10, 12]. Additionally, as a consequence of the non-vanishing latent heat, the width of the energy distribution, which is proportional to the specific heat, scales $\sim (\Delta e)^2 L^D$ at criticality [5].

At a continuous pseudo-transition, the system size is the upper bound for the diverging correlation length (finite-size scaling). Thus, $\xi \rightarrow L$, which causes $|1 - T/T_c| \sim \xi^{-\nu} \rightarrow L^{-\nu}$ and $C \sim L^{\alpha/\nu}$. Additionally, the required simulation time which scales with ξ^z will behave like L^z , where z is the dynamical critical exponent [12].

2.2 Monte Carlo Algorithms

Many of such physical systems are either impossible or computationally very hard to solve as they involve an extremely high number of microscopic degrees of freedom. Monte Carlo (MC) methods, named after the famous Casino de Monte-Carlo, circumvent this problem by utilizing large quantities of random numbers. In particular, we use MC to sample a subset of states γ from the state space of a system which is weighted by an ensemble's equilibrium probability distribution. This set of states can then be used to estimate ensemble averages containing macroscopic information.

The ensemble average of an observable \mathcal{O} is in general defined as

$$\langle \mathcal{O} \rangle_\lambda = \sum_{\{\gamma\}} \mathcal{O}(\gamma) p_\lambda(\gamma), \quad (2.3)$$

and can be estimated as an average over the sampled states. The probability distribution is defined as

$$p_\lambda(\gamma) = \frac{1}{C_\lambda} \omega_\lambda(\gamma), \quad (2.4)$$

with a state-independent constant C_λ , weight function $\omega_\lambda(\gamma)$ and a control parameter λ . The constant

$$C_\lambda = \sum_{\{\gamma\}} \omega_\lambda(\gamma) \quad (2.5)$$

is the partition function of the ensemble. Note that if p_λ is the Boltzmann distribution, λ and C coincide with the inverse temperature β and the canonical partition function Z , respectively.

For systems with a large number N of microscopic degrees of freedom (d.o.f.),

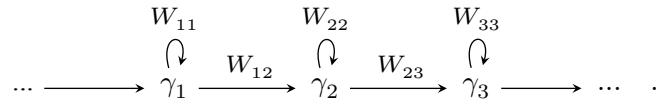
$$\gamma = (\sigma_1, \sigma_2, \dots, \sigma_N), \quad (2.6)$$

it is as good as impossible to sample from $p_\lambda(\gamma)$ directly, as the normalization requires the summation of q^N terms, where q is the number of admissible states a single σ_k may take. Simply sampling states at random, on the other hand, is only able to generate highly disordered states, as the ordered states are exceedingly rare. For any reasonable ensemble, however, we need to be able to sample the ordered states as well.

Markov Chain Monte Carlo (MCMC) provides a solution to this issue by constructing a Markov chain whose evolution follows the given probability distribution without the need of normalization. A Markov chain is a sequence or web of the q^N states the system can evolve into, connected by transition probabilities (W_{ij}). The conditions

$$\text{a) } W_{ij} \geq 0, \quad \text{b) } \sum_j W_{ij} = 1 \quad (2.7)$$

need to be met such that W contains well-defined probabilities, rendering it a transition matrix. Crucially, the value of a single W_{ij} depends explicitly only on the current state, in other words the chain has no memory of its prior evolution. A simplified excerpt of a Markov chain can be depicted as



Of course, a huge number of edges both coming from and leading to other states are omitted. However, the vast majority of these may have a transition probability equal to zero, as long as any state in the phase space can be reached within a finite number of steps (*ergodicity*).

To make the Markov chain follow the probability distribution p_λ , W must satisfy

$$Wp_\lambda = p_\lambda.^2 \quad (2.8)$$

This is known as the balance condition; it defines p_λ to be an eigenvector of W with unit eigenvalue. It can be shown³ that any ergodic transition matrix has a single eigenvalue equal to one, with all others being of strictly smaller magnitude. Thus, p_λ is a stable attractor of W and for an arbitrary probability distribution ψ ,

$$W^k \psi \xrightarrow{k \rightarrow \infty} p_\lambda. \quad (2.9)$$

In practice, MCMC algorithms are checked against the stronger condition of detailed balance,

$$W_{ij} p_\lambda(\gamma_i) = W_{ji} p_\lambda(\gamma_j), \quad (2.10)$$

which can be reduced to (2.8) by summing over i and using (2.7b).

These conditions are still quite general, so there are many algorithms that could fulfill them. Of course, we are interested an implementation that is able to provide uncorrelated samples from the equilibrium distribution with low computational effort, i.e., a small number of MC steps in between samples. A sensible unit for the temporal evolution⁴ of a system is the Monte Carlo sweep. It consists of N Monte Carlo steps such that the expected number of steps per d.o.f. is constant. Correlations in Monte Carlo decay exponentially for large timescales [12]. Thus, we can define autocorrelation time τ as the time it takes for the correlations decay by $1/e$.

The original Metropolis algorithm is an update algorithm that turns out to be a very versatile choice. Here, the transition probability is split into two steps,

$$W_{ij} = \begin{cases} f_{ij} w_{ij} & i \neq j, \\ f_{ij} w_{ij} + \sum_{i \neq j} f_{ij} (1 - w_{ij}) & i = j, \end{cases} \quad (2.11)$$

the proposal of a transition with selection probability f_{ij} and its acceptance with probability

$$w_{ij} = \min \left(1, \frac{f_{ji} p_\lambda(\gamma_j)}{f_{ij} p_\lambda(\gamma_i)} \right). \quad (2.12)$$

The transition between two arbitrary states is bound to be very unlikely. One way to solve this problem is to only propose *local* updates at any single step, i.e., transitions in only a

²Recall W is a $q^N \times q^N$ matrix and p_λ is a q^N -component vector.

³cf. Gershgorin circle theorem and Perron-Frobenius theorem.

⁴When referring to time in the realm of Monte Carlo simulations, what is actually meant is the system's development walking through the Markov chain. It is not directly related to time in the physical sense.

single d.o.f. σ_k that may simply be chosen at random from a uniform distribution. The proposed transition in the d.o.f. can then be drawn uniformly as well, though this may lead to high rejection rates if the number of admissible states q is large.

In this case, it may be more effective to draw a d.o.f. σ_k , calculate the probability weight of its q states and then draw the new state from the normalized distribution

$$w_{ij} = \frac{p_\lambda(\gamma_j)}{\sum_{\ell=0}^q p_\lambda(\gamma_\ell)}, \quad (2.13)$$

which is known as the heatbath algorithm. Both (2.12) and (2.13) may be further simplified by only including terms of p_λ involving σ_k .

For both of these algorithms (2.7) and (2.10) are fulfilled [12].

2.3 Population Annealing

Population annealing (PA) was first introduced by Hukushima and Iba [13] and was designed with rugged free energy landscapes in mind. It is a technique build upon simulated annealing, where a single MCMC simulation is gradually cooled down, in the hope of finding minima in the free energy. If there are many local minima, as is the case for example in spin glasses, a single MCMC walker will have a high risk of getting stuck in a suboptimal, metastable state. If this happens, it will not be able to equilibrate further when the temperature is lowered. This can be partly remedied by starting with a population of evolving MCMC states, that is annealed in parallel. If the population size R is very large, some population members will end up in enough minima to correctly sample the phase space, though many walkers will be stuck in metastable states. Thus, only a small fraction of the population will sample the equilibrium distribution. When performing a weighted average, only a few members will yield significant contributions, and a lot of computing power is wasted on members with low statistical weight.

The PA algorithm provides a solution to this by introducing a resampling into the annealing step, such that the low weight members are pruned from the population while high weight members are replicated. Thus, the population will stay close to the equilibrium distribution. Depending on the chosen type of resampling, the population size may fluctuate from one annealing step to another.

This is followed by a MCMC routine in order to reduce correlation introduced into the population during resampling. The annealing takes place in discrete annealing steps from the annealing schedule $(\lambda_0, \lambda_1, \dots, \lambda_f)$. The schedule may be chosen freely, as long as there is sufficient overlap between the distributions of consecutive steps. Additionally, the initial distribution p_{λ_0} should be chosen such that it is easy to sample from (e.g. by simple

sampling if $\omega_{\lambda_0} \equiv \text{const}$). The basic procedure of this algorithm is visualized in Fig. 2.1.

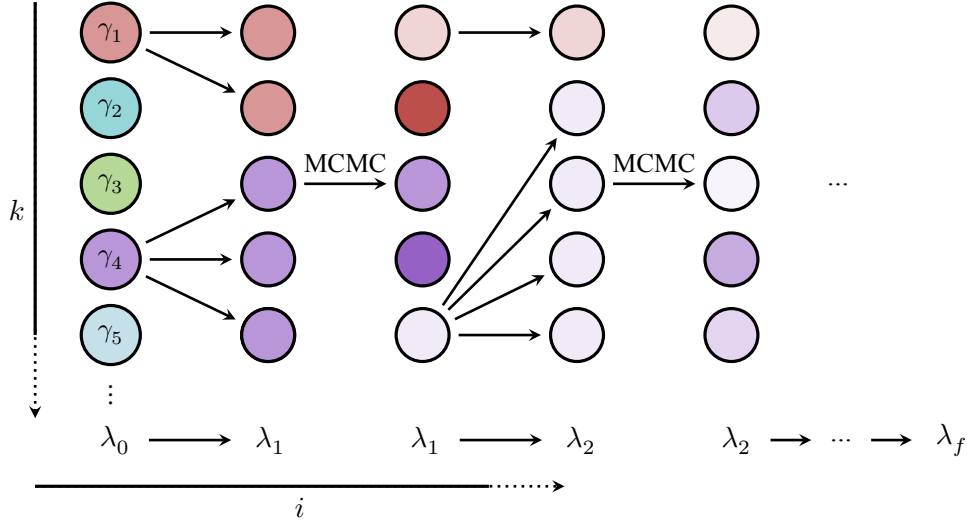


Figure 2.1: Visualization of the algorithm. The different colors represent the statistically independent initial replicas, while the differing shades stand for the decorrelating effects of the MCMC routine.

Crucially, when transitioning to the next annealing step, the population is held in equilibrium through resampling, where the population is reweighted according to the replicas' weights in the initial and target distribution.

The annealing step starts by calculating, for each replica γ_k at annealing step i , its relative weight in the new population,

$$\hat{\tau}_i^{(\gamma_k^{(i)})} = \frac{R_0}{R_{i-1}} \frac{1}{Q_i} \frac{\omega_{\lambda_i}(\gamma_k^{(i)})}{\omega_{\lambda_{i-1}}(\gamma_k^{(i)})}, \quad (2.14)$$

with the normalization factor

$$Q_i = \frac{1}{R_{i-1}} \sum_{k=0}^{R_{i-1}} \frac{\omega_{\lambda_i}(\gamma_k^{(i)})}{\omega_{\lambda_{i-1}}(\gamma_k^{(i)})}. \quad (2.15)$$

This makes $\hat{\tau}_i$ the expected number of descendants a replica should propagate to the next population.

The factor R_0/R_{i-1} introduced in (2.14) ensures that the population size tends to stay close to the initial value R_0 by lowering the overall copy probability if $R_i > R_0$ and rising it in the opposite case.

The actual number of descendants r_i , which has to be an integer, is then sampled from a distribution with $\langle r_i \rangle = \hat{\tau}_i$. There is a wide range of resampling schemes to choose from, some of which keep the population size R constant and others let it fluctuate [14]. An easy and quite effective way is nearest integer resampling, where a random number uniformly distributed on $[0, 1)$ is drawn and the number of copies is determined by

$$r_i = \begin{cases} \lfloor \hat{\tau}_i \rfloor + 1 & \text{if } \xi < q = \hat{\tau}_i - \lfloor \hat{\tau}_i \rfloor, \\ \lfloor \hat{\tau}_i \rfloor & \text{else,} \end{cases} \quad (2.16)$$

essentially taking the decimal part of $\hat{\tau}_i$ as the probability whether to round up or down.

Following the resampling, we make θ MCMC sweeps in order to reduce the correlations that are introduced by adding identical copies of replicas in the resampling step. Thereafter, measurements may be taken from the population.

If the final value, λ_f , in the annealing schedule has not been reached, proceed by subsequent annealing steps as defined above.

2.3.1 Choice of Annealing Schedule

In the basic setup, the annealing step size is uniform $\lambda_i \rightarrow \lambda_i + \Delta\lambda$ with an appropriately chosen $\Delta\lambda$ [14]. This yields an asymptotically correct sampling process but is certainly not optimal as it lets the culling fraction ϵ_i [15], which is the fraction of the population that does not survive into the next generation, fluctuate widely. This leads to a high loss of diversity in the population if ϵ is large and many unnecessary computations if ϵ is very small as in that case the histogram barely changes. Furthermore, the dependence of $\Delta\lambda$ on ϵ is highly dependent on the studied Hamiltonian and ensemble.

For this reason it is useful to choose an annealing schedule such that ϵ is constant. The expected ϵ , dependent on the annealing step size, can be estimated with

$$\hat{\epsilon}_i = \frac{1}{R_{i-1}} \sum_{k=0}^{R_{i-1}} \begin{cases} 1 - \hat{\tau}_i^{(\gamma_k^{(i)})} & \hat{\tau}_i^{(\gamma_k^{(i)})} \leq 1 \\ 0 & \text{else.} \end{cases} \quad (2.17)$$

before the actual resampling takes place. Thus, we can numerically find an annealing step such that the expected culling fraction is constant.

The histogram overlap, which is the fraction of the population that is propagated to the next generation, is clearly just the remaining fraction of the population such that the overlap is equal to $1 - \epsilon$.

2.3.2 Estimation of Canonical Observables

Generally, we are interested in averages in the canonical ensemble, i.e., determined by the Boltzmann distribution. There, the expectation value of observable \mathcal{O} at inverse temperature β is given by the expression

$$\langle \mathcal{O} \rangle_\beta = \frac{1}{\mathcal{Z}_\beta} \sum_{\{\gamma\}} \mathcal{O}(\gamma) e^{-\beta \mathcal{H}(\gamma)} = \frac{1}{\mathcal{Z}_\beta} \sum_E \Omega(E) \mathcal{O}(E) e^{-\beta E} \quad (2.18)$$

with the system's Hamiltonian $\mathcal{H}(\gamma)$. The density of states (DOS), denoted as $\Omega(E)$, can be defined as

$$\Omega(E) = \sum_{\{\gamma\}} \delta_{\mathcal{H}(\gamma), E}, \quad (2.19)$$

counting up the number of states that have energy E . Here, δ is the Kronecker delta. Notably, the DOS is a property independent of the individual ensemble, it occurs in any ensemble whose weight function is only dependent on the system's Hamiltonian ($\omega_\lambda \equiv \omega_\lambda(\mathcal{H}(\gamma))$). If we know the DOS, our (energetic) job is done as it encapsulates the complexity of the exponentially many configurations and can be used to compute the energetic observables in any ensemble.

Still, there are observables that are not directly dependent on the energy, such as the magnetization. For these we use their microcanonical average,

$$\langle \langle \mathcal{O} \rangle \rangle_E = \frac{1}{\Omega(E)} \sum_{\{\gamma\}} \mathcal{O}(\gamma) \delta_{\mathcal{H}(\gamma), E}, \quad (2.20)$$

giving us an energy dependent value that can be transformed into canonical observables with the DOS as [12]

$$\langle \mathcal{O} \rangle_\beta = \frac{1}{\mathcal{Z}_\beta} \sum_E \Omega(E) \langle \langle \mathcal{O} \rangle \rangle_E e^{-\beta E}. \quad (2.21)$$

Thus, we use the following strategy: Utilizing ensembles that allow easier sampling of the suppressed states. This allows us to estimate both the DOS and the microcanonical averages of non-energetic observables. These estimates can then be used with the r.h.s. of (2.18) to obtain canonical averages of all observables.

In PA, ensemble averages can simply be estimated as population averages, as the population is kept in equilibrium through resampling (aided by the MCMC step), such that

$$\langle \mathcal{O} \rangle_{\lambda_i} = \frac{1}{C_{\lambda_i}} \sum_{\{\gamma\}} \mathcal{O}_{\lambda_i}(\gamma) \omega_{\lambda_i}(\gamma) \approx \bar{\mathcal{O}} \equiv \frac{1}{R} \sum_{k=0}^R \mathcal{O}_k. \quad (2.22)$$

Microcanonical averages are estimated over the span of the whole annealing (recall f is the number of annealing steps) with

$$\langle\langle\mathcal{O}\rangle\rangle_E \approx \frac{1}{f} \sum_{i=0}^f \frac{1}{H_i(E)} \sum_{k=1}^{R_i} \mathcal{O}_k \delta_{\mathcal{H}(\gamma_k^{(i)}), E}, \quad (2.23)$$

where $H_i(E)$ is the histogram at step i ,

$$H_i(E) = \sum_{k=1}^{R_i} \delta_{\mathcal{H}(\gamma_k^{(i)}), E}, \quad (2.24)$$

i.e., the number of states observed with energy E .

2.3.3 Estimation of the Partition Function

PA lends itself well to the estimation of the partition function, as the normalization factors Q_i can be reinterpreted as estimators of partition function ratios. To see this, we first use (2.22) to see that Q_i is the estimator of the ensemble average

$$Q_i = \frac{1}{R_{i-1}} \sum_{k=0}^{R_{i-1}} \frac{\omega_{\lambda_i}(\gamma_k^{(i)})}{\omega_{\lambda_{i-1}}(\gamma_k^{(i)})} \approx \left\langle \frac{\omega_{\lambda_i}(\gamma^{(i)})}{\omega_{\lambda_{i-1}}(\gamma^{(i)})} \right\rangle_{\lambda_{i-1}}. \quad (2.25)$$

When plugging into (2.3), we see that

$$Q_i \approx \frac{1}{C_{\lambda_{i-1}}} \sum_{\{\gamma\}} \frac{\omega_{\lambda_i}(\gamma^{(i)})}{\omega_{\lambda_{i-1}}(\gamma^{(i)})} \omega_{\lambda_{i-1}}(\gamma) = \frac{C_{\lambda_i}}{C_{\lambda_{i-1}}}. \quad (2.26)$$

Using a telescopic product, the partition function at any step i can be expressed as

$$C_{\lambda_i} = C_{\lambda_0} \frac{C_{\lambda_1}}{C_{\lambda_0}} \dots \frac{C_{\lambda_i}}{C_{\lambda_{i-1}}} \approx C_{\lambda_0} \prod_{j=0}^i Q_j. \quad (2.27)$$

This equation allows us to estimate the partition function of any ensemble, starting from an initial known partition function C_{λ_0} and multiplying by the successive partition function ratios Q_j .

2.4 Correlations

As previously discussed, the resampling step introduces correlations. This occurs because multiple identical copies of replicas are added into the population. Correlations have significant implications for both systematic and statistical errors.

A primary consequence of these correlations is the reduction in the effective population size, denoted as R_{eff} . Since the statistical error scales $\sim 1/\sqrt{R_{\text{eff}}}$, a decrease in R_{eff} directly increases the statistical error. Additionally, the correlations cause systematic error, too. The reweighting tends to distort the histogram as the low probability tail towards lower energies is getting amplified, while it only contains a few replicas. Especially if a part of the equilibrium distribution lies behind a region that is very hard to cross by the MCMC routine, the population is no longer a representation of the equilibrium probability distribution and as such, the reweighted population will, too, be biased towards lower energies. This is the cause of hysteresis, where a phase transition is observed to occur later than it would in a truly equilibrated population, which becomes visible as two separated paths in observables like the energy and magnetization average when the annealing direction is reversed.

To gain deeper insight into these correlations, we use the approach of [15] and introduce the notion of families, which are the descendants of the (fully independent) initial replicas. By keeping track of the distribution of family sizes, we can better understand the extent of correlations, as copying replicas will naturally increase the average family size. If we let the number of descendants of initial replica \mathfrak{o} be $\mathfrak{N}_{\mathfrak{o}}$, we can find the replica averaged family size ρ_t [14] at step i by averaging over the population as

$$\rho_t = \frac{1}{R_i} \sum_{k=1}^{R_i} \mathfrak{N}_{\mathfrak{o}_k}. \quad (2.28)$$

However, keeping track of $\mathfrak{N}_{\mathfrak{o}_k}$ is quite cumbersome, so an alternative approach is to average over the families instead. Each family contributes R_i/R_0 $\mathfrak{N}_{\mathfrak{o}}$ replicas of the current population such that

$$\rho_t = \frac{1}{R_i} \sum_{\mathfrak{o}=1}^{R_0} \mathfrak{N}_{\mathfrak{o}}^2. \quad (2.29)$$

The MCMC routine is intended to decorrelate the replicas. However, if the MCMC routine fails to sufficiently decorrelate the replicas between resampling steps, the average family size ρ_t grows more rapidly. As this is particularly the case at phase transitions, ρ_t can serve as an indicator for undergoing a (badly equilibrated) phase transition.

The family size generally increases over time due to the cumulative effects of resampling, not being affected by the decorrelating effects of the MCMC step. The ratio R/ρ_t thus only provides an upper bound for R_{eff} [16].

Additionally, if the replicas are independently and identically distributed, the increase in ρ_t per annealing step is equal to the variance of the number of copies, as was shown in [14], such that

$$\rho_t^{(i)} = \rho_t^{(i-1)} + \sigma^2(r_i). \quad (2.30)$$

The variance $\sigma^2(r_i)$ is itself connected to the culling fraction, and, assuming $\hat{\tau}$ is equal for all surviving replicas, follows the form

$$\sigma^2(r_i) = \left\lfloor \frac{1}{1-\epsilon} \right\rfloor \left(\epsilon + 1 + \left\lfloor \frac{1}{1-\epsilon} \right\rfloor (\epsilon - 1) \right), \quad (2.31)$$

as we show in [Appendix A](#). Together with the observation that, by definition, $\rho_t^{(0)} = 1$, this leads to

$$\rho_t^{(i)} = 1 + \sum_{\ell=1}^i \left\lfloor \frac{1}{1-\epsilon_i} \right\rfloor \left(\epsilon_i + 1 + \left\lfloor \frac{1}{1-\epsilon_i} \right\rfloor (\epsilon_i - 1) \right). \quad (2.32)$$

2.5 Microcanonical Ensemble

In the microcanonical ensemble, the total energy $\mathcal{H}(\gamma) = \mathcal{E}$ is held constant, so all microstates with energy \mathcal{E} are equally probable. Therefore, *the total energy \mathcal{E} is the control parameter λ in the microcanonical ensemble*, such that the annealing schedule will be $(\mathcal{E}_0, \dots, \mathcal{E}_f)$. The microcanonical ensemble is defined by the weight function

$$\omega_{\mathcal{E}} = \delta_{\mathcal{H}(\gamma), \mathcal{E}}. \quad (2.33)$$

When comparing this to (2.19), it is clear that the partition function $C_{\mathcal{E}}$ of the microcanonical ensemble is simply the density of states,

$$C_{\mathcal{E}} = \Omega_{\mathcal{E}}, \quad (2.34)$$

where we used a subscript to convey that we should think of the energy as the control parameter. Thus, an implementation of MCMC in the microcanonical ensemble should mainly sample macrostates with the highest DOS, or in other words, maximize the entropy

$$S_{\mathcal{E}} = \ln \Omega_{\mathcal{E}}. \quad (2.35)$$

This is in contrast to the canonical ensemble where the MCMC routine mostly explores states near the free-energy minimum.

The expected advantage of employing the microcanonical ensemble is that it allows for straightforward sampling of states within the suppressed region by simply selecting an appropriate energy \mathcal{E} .

However, we face the issue that in PA a non-zero histogram overlap is essential. Since the microcanonical ensemble strictly enforces a constant total energy, the overlap between subsequent \mathcal{E}_i is always zero, making this form of the ensemble unsuitable for methods such as PA. In addition, it is not at all trivial finding an efficient MCMC routine that is

energy conserving. For this reason, we introduce an artificial kinetic energy term $E_{\text{kin}} > 0$ to take up the energy difference between $\mathcal{H}(\gamma)$ and \mathcal{E} . This allows configurations with $\mathcal{H}(\gamma)$ equal to *or lower* than \mathcal{E} . The modified Hamiltonian (that is held constant) is of the form

$$\tilde{\mathcal{H}}(\gamma, E_{\text{kin}}) = \mathcal{H}(\gamma) + E_{\text{kin}} = \mathcal{E}. \quad (2.36)$$

2.5.1 Ceiling Ensemble

The ceiling ensemble [9, 17] encompasses all states with $\mathcal{H}(\gamma) \leq \mathcal{E}$, each with uniform weight. Therefore, its weight function is

$$\omega_{\mathcal{E}}(\gamma) = \theta(\mathcal{E} - \mathcal{H}(\gamma)), \quad (2.37)$$

where θ is the Heaviside step function. It ensures that only states with $\mathcal{H}(\gamma) \leq \mathcal{E}$ contribute to the ensemble.

The partition function C of the ceiling ensemble is then given by the “ceiling DOS”

$$\Omega_{\mathcal{E}}^c = \sum_{\{\gamma\}} \theta(\mathcal{E} - \mathcal{H}(\gamma)), \quad (2.38)$$

where its logarithm is denoted as the “ceiling entropy” in [9]. As it is simply the number of states with $\mathcal{H}(\gamma) \leq \mathcal{E}$, it is particularly useful because it can be directly related to the density of states

$$\Omega(E \approx \mathcal{E}_i) = \Omega_{\mathcal{E}_i}^c - \Omega_{\mathcal{E}_{i+1}}^c, \quad (2.39)$$

where the approximation becomes a true equality in the case \mathcal{E}_i and \mathcal{E}_{i+1} are only a single energy step of the Hamiltonian energy apart. We can estimate $\Omega_{\mathcal{E}}^c$ with (2.27), giving us an estimator for $\Omega(E)$.

2.5.2 Momentum Ensemble

In the momentum ensemble [8, 10], the kinetic energy is introduced as originating from N one-, or three-dimensional momenta, respectively. We generalize this to $M = \alpha N$ one-dimensional momenta. This leads to a modified Hamiltonian of the form

$$\tilde{\mathcal{H}}(\gamma, p) = \mathcal{H}(\gamma) + \sum_{i=1}^M \frac{p_i^2}{2} = \mathcal{E}. \quad (2.40)$$

We can obtain the density of states associated with this modified Hamiltonian, which includes both the original system’s energy and the kinetic energy contributions from the momenta, through the inverse Laplace transform of the canonical partition function [18],

$$\tilde{\Omega}_{\mathcal{E}, M} = \mathcal{L}_{\beta}^{-1} \{ \tilde{Z}_{\beta, M} \} (\mathcal{E}), \quad (2.41)$$

where the canonical partition function is given by

$$\begin{aligned}\tilde{Z}_{\beta,M} &= \frac{1}{M! h^M} \int \sum_{\{\gamma\}} e^{-\beta \tilde{\mathcal{H}}(\gamma,p)} d^M p \\ &= \frac{1}{M! h^M} \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta p^2} dp \right)^M \sum_{\{\gamma\}} e^{-\beta \mathcal{H}(\gamma)} \\ &= \frac{1}{M! h^M} \left(\sqrt{2\pi/\beta} \right)^M \sum_{\{\gamma\}} e^{-\beta \mathcal{H}(\gamma)}.\end{aligned}$$

Evaluating the inverse Laplace transform yields the DOS of the modified Hamiltonian

$$\tilde{\Omega}_{\mathcal{E},M} = \frac{(2\pi)^{M/2}}{M! h^M} \frac{1}{\Gamma(M/2)} \sum_{\{\gamma\}} (\mathcal{E} - \mathcal{H}(\gamma))^{\frac{M-2}{2}} \theta(\mathcal{E} - \mathcal{H}(\gamma)). \quad (2.42)$$

Recalling that the DOS is the microcanonical partition function, we use this result to find the pmf

$$p_{\mathcal{E},M}(\gamma) = \frac{1}{\tilde{\Omega}_{\mathcal{E},M}} (\mathcal{E} - \mathcal{H}(\gamma))^{\frac{M-2}{2}} \theta(\mathcal{E} - \mathcal{H}(\gamma)). \quad (2.43)$$

The introduction of power-law term has two main effects: Firstly, it pushes the observed energy distribution, which we call the histogram, away from the maximally allowed energy \mathcal{E} . This can be interpreted as introducing a pressure-like effect. Secondly, it broadens the histogram which has the consequence that a wider distribution of states is sampled at any annealing step.

2.5.3 Role of Weight Function Exponent

The exponent in the weight function reflects the influence of the kinetic energy on the system's energy distribution. It can be used to dial in a certain histogram width and pressure.

Thus, it is crucial to understand how exactly the histogram width can be tuned by the parameter M .

If we approximate the scaling of the DOS for near-critical energies as $\Omega(E) \sim \exp(kE)$ (see Fig. B.1) and for $M, N \gg 1$, the histogram is very similar to a Gaussian distribution. This implies there is a second order term that dominates in the exponent. The prefactor of this term will therefore be $1/(2\sigma_E^2)$. In Appendix B, we follow this reasoning in detail and come to the conclusion that the histogram width can be approximated by

$$\sigma_E = \sqrt{\frac{M-2}{2k^2}}, \quad \text{or} \quad \sigma_e = \frac{1}{N} \sqrt{\frac{M-2}{2k^2}}, \quad (2.44)$$

respectively. In the literature, the momentum ensemble is used with $M \sim N$ and sometimes denoted as the “real” microcanonical ensemble [8], in opposition to the microcanonical ensemble of the unmodified Hamiltonian. This causes the histogram width to scale \sqrt{N} , which is similar to the canonical ensemble where $\sigma_E \sim \sqrt{NC}$. Also note that the ceiling ensemble is a special case of the momentum ensemble where M is equal to 2.

2.5.4 Mixed Ensemble

If we have a Hamiltonian that contains multiple terms, such as

$$\mathcal{H} = J_0 \mathcal{H}_0 + J_1 \mathcal{H}_1 + \dots \quad (2.45)$$

we may want to perform the annealing process with a single, isolated coordinate, J_0 , treated microcanonically. With this end in mind, we add the momenta to that term. This results in the modified Hamiltonian

$$\tilde{\mathcal{H}} = J_0 \underbrace{\left(\mathcal{H}_0 + \sum_{i=1}^M \frac{p_i^2}{2} \right)}_{=\mathcal{E}_0} + J_1 \mathcal{H}_1 + \dots \quad (2.46)$$

Then, we perform the inverse Laplace transform not over β but the coordinate J_0 , in order to find the mixed ensemble’s partition function

$$\begin{aligned} \tilde{\Gamma}_{\beta, M, \mathcal{E}_0, J_1, \dots} &= \mathcal{L}_{J_0}^{-1} \left\{ \tilde{Z}_{\beta, M, J_0, \dots} \right\} (\mathcal{E}_0) = \mathcal{L}_{J_0}^{-1} \left\{ \frac{1}{M! h^M} \left(\sqrt{2\pi/(\beta J_0)} \right)^M \sum_{\{\gamma\}} e^{-\beta \mathcal{H}} \right\} (\mathcal{E}_0) \\ &= \frac{(2\pi/\beta)^{M/2}}{M! h^M \Gamma(M/2)} \sum_{\{\gamma\}} e^{-\beta(J_1 \mathcal{H}_1 + \dots)} (\mathcal{E}_0 - \beta \mathcal{H}_0)^{\frac{M-2}{2}} \theta(\mathcal{E}_0 - \beta \mathcal{H}_0), \end{aligned} \quad (2.47)$$

and subsequently its weight function

$$\omega_{\beta, M, \mathcal{E}_0, J_1, \dots} = e^{-\beta(J_1 \mathcal{H}_1 + \dots)} (\mathcal{E}_0 - \beta \mathcal{H}_0)^{\frac{M-2}{2}} \theta(\mathcal{E}_0 - \beta \mathcal{H}_0). \quad (2.48)$$

2.6 Histogram Reweighting Techniques

In this section, we will discuss a method to retrieve the (unmodified) DOS $\Omega(E)$ from our measurements. Fortunately, Ref. [19] presents an application of multi-histogram reweighting, a method to estimate $\Omega(E)$, for canonical PA. This method can be easily generalized to an arbitrary ensemble.

The normalized histogram of the population at a given stage in the simulation is defined as

$$\widehat{P}_{\lambda_i}(E) = \frac{1}{R_i} H_{\lambda_i}(E) = \frac{1}{R_i} \sum_{k=0}^{R_i} \delta_{\mathcal{H}(\gamma_k^{(i)}), E}. \quad (2.49)$$

This histogram provides an estimate for the probability distribution of the energy at λ_i which in general is of the form

$$P_{\lambda_i}(E) = \frac{1}{C_{\lambda_i}} \Omega(E) \omega_{\lambda_i}(E). \quad (2.50)$$

Thus, the estimator can be reweighted to an arbitrary λ_0 with

$$\widehat{P}_{\lambda_0}(E) \sim \widehat{P}_{\lambda_i}(E) \frac{\omega_{\lambda_0}(E)}{\omega_{\lambda_i}(E)}, \quad (2.51)$$

as long as there is a significant overlap between the distributions (*single-histogram reweighting*). This is akin to the population reweighting in (2.14).

Furthermore, from (2.50), each histogram $\widehat{P}_{\lambda_i}(E)$ provides an estimator for the DOS given by

$$\widehat{\Omega}_i(E) = \widehat{P}_{\lambda_i}(E) \frac{C_{\lambda_i}}{\omega_{\lambda_i}(E)}. \quad (2.52)$$

We can combine these as a variance optimized estimator [12]

$$\widehat{\Omega}(E) = \frac{\sum_i \sigma_{\widehat{\Omega}_i}^{-2}(E) \widehat{\Omega}_i(E)}{\sum_i \sigma_{\widehat{\Omega}_i}^{-2}(E)} \quad \text{with} \quad \sigma_{\widehat{\Omega}_i}^2(E) = \frac{C_{\lambda_i}^2}{R_i^2 \omega_{\lambda_i}^2(E)} \sigma_{H_{\lambda_i}}^2(E). \quad (2.53)$$

If the replicas are assumed to be independent of each other, they are Poisson distributed such that [19]

$$\sigma_{H_{\lambda_i}}^2(E) = R_i \frac{1}{C_{\lambda_i}} \Omega(E) \omega_{\lambda_i}(E). \quad (2.54)$$

This leads to the method of *multi-histogram reweighting* (MHR), which combines information from multiple histograms to give a refined estimate of the DOS as

$$\widehat{\Omega}(E) = \frac{\sum_i H_{\lambda_i}(E)}{\sum_i R_i \frac{1}{C_{\lambda_i}} \omega_{\lambda_i}(E)}. \quad (2.55)$$

Table 2.1: Comparison of population annealing methods and canonical estimators.

Method	Weight Function	Control Parameter	Determination of canon. observables
Canonical PA	$e^{-\beta \mathcal{H}(\gamma)}$	β (inverse temp.)	Population average
Ceiling PA	$\theta(\mathcal{E} - \mathcal{H}(\gamma))$	\mathcal{E} (total energy)	Obtain DOS from $\Omega(E) = \Omega_{\mathcal{E}_i}^c - \Omega_{\mathcal{E}_{i+1}}^c$ + microcan. averages
Momentum PA	$(\mathcal{E} - \mathcal{H}(\gamma))^{\frac{M-2}{2}} \times \theta(\mathcal{E} - \mathcal{H}(\gamma))$	\mathcal{E} (total energy)	Obtain DOS from MHR + microcan. averages
Mixed PA	$e^{-\beta(J_1 \mathcal{H}_1 + \dots)} \times (\mathcal{E}_0 - \beta \mathcal{H}_0)^{\frac{M-2}{2}} \times \theta(\mathcal{E}_0 - \beta \mathcal{H}_0)$	\mathcal{E}_0 (single term)	as above

2.7 Use of the Ensembles in Population Annealing

This section provides a short summary of the different novel ensembles that are used to sample the state space efficiently. In Table 2.1, we present an overview that contrasts canonical PA with the microcanonical variants, in order to make it clear how these are used in the PA algorithm. Even though there are some physical intuitions behind the novel ensembles, like the notion of a non-negative kinetic energy, the momenta and energy as a conserved quantity, at the end these ensembles are used as a sleight of hand, with the goal to efficiently sample the state space and to calculate canonical averages.

2.8 Models

In this work, we study three different models, all of which are discrete spin models on a two-dimensional lattice of size L with periodic boundary conditions. They are chosen for their simplicity and due to the particular characteristics of their phase transitions.

2.8.1 Ising Model

The *Ising model*, which is one of the most studied models in all of statistical physics where it serves as a simple model of a ferromagnet. It was first introduced by Wilhelm Lenz in 1920 and solved in two dimensions by Lars Onsager in 1944.

Here, it serves as an initial test for validating the novel PA methods. Since its exact solution is solvable with a polynomial algorithm [20], we can directly compare the results to the exact values to quantify the systematic error.

The zero-field model is given by the Hamiltonian

$$\mathcal{H}(\gamma = \{\sigma_i\}) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j, \quad \sigma \in \{-1, 1\}, \quad (2.56)$$

where J is the coupling constant and $\langle ij \rangle$ indicates the sum is over the nearest neighbors on the lattice. It describes a ferromagnet for positive J , and an anti-ferromagnet when $J < 0$. Here, and in the following models, we take $J = 1$.

The infinite-size 2D square lattice Ising model exhibits a continuous phase transition at inverse critical temperature β_c , as was shown by Onsager [21] who provided an analytic solution for the partition function as well as a closed form for the critical temperature with $\beta_c = \ln(1 + \sqrt{2})/(2J)$.

In the Ising model, the energy e and magnetization m (per number of spins N) are defined as

$$e = \frac{1}{N} \sum_{\langle ij \rangle} \sigma_i \sigma_j, \quad \text{and} \quad m = \frac{1}{N} \sum_{i=1}^N \sigma_i. \quad (2.57)$$

The canonical specific heat and susceptibility in the Ising model are defined as

$$C(\beta) = N\beta^2(\langle e^2 \rangle_\beta - \langle e \rangle_\beta^2), \quad \text{and} \quad \chi = N\beta(\langle m^2 \rangle_\beta - \langle m \rangle_\beta^2). \quad (2.58)$$

2.8.2 Potts Model

The *Potts model*, introduced in 1952, provides a generalization of the Ising model to more than two spin states. Here it is significant, as is one of the simplest models that shows a first-order phase transition. Its Hamiltonian is

$$\mathcal{H}(\{\sigma_i\}) = - \sum_{\langle ij \rangle} \delta_{\sigma_i, \sigma_j}, \quad \sigma \in \{1, 2, \dots, q\}. \quad (2.59)$$

where δ is the Kronecker delta. Here, the energy e and magnetization m (per number of spins N) are defined as

$$e = \frac{1}{N} \sum_{\langle ij \rangle} \delta_{\sigma_i, \sigma_j}, \quad \text{and} \quad m = \frac{q n_{\max} - 1}{(q - 1)}, \quad (2.60)$$

where n_{\max} is the fraction of spins that point into the most common spin direction, i.e.,

$$n_{\max} = \frac{1}{N} \max_{a \in 1 \dots q} \left(\sum_{i=1}^N \delta_{a, \sigma_i} \right). \quad (2.61)$$

The canonical specific heat and susceptibility in the Potts model are defined identical as in the Ising model (see (2.58)).

Many exact properties are known for the Potts model in two dimensions for $N \rightarrow \infty$, some of which are: The critical temperature is given by $\beta_c = \ln(1 + \sqrt{q})$. For $q \leq 4$, the phase transition is continuous, and for $q > 4$ it is of first order. Notably, it is equivalent to Ising model for $q = 2$. The δ -peaks of the energy at criticality lie at [22]

$$E_{c,\pm} = 1/(1 + \sqrt{q}) (-1 \pm \tanh(\theta/2) \prod_{n=1}^{\infty} \tanh^2(n\theta)),$$

$$\text{with } \theta = \begin{cases} \text{arcosh}(\sqrt{q}/2) & \text{for } q > 4 \text{ and} \\ 0 & \text{for } q \leq 4, \end{cases}$$

where the two peaks fall together. The interfacial tension, which is a characteristic of the strength of the first-order transition, is known exactly by the formula $2\hat{\sigma}_{od} = \frac{1}{4} \sum_{n=0}^{\infty} \ln \left(\frac{1+w_n}{1-w_n} \right)$ with $w_n = \left(\frac{\sqrt{2} \cosh((n+1/2)\pi^2)}{2v} \right)^{-1}$ and $v = \ln(\sqrt{\sqrt{q}+2} + \sqrt{\sqrt{q}-2}) - \ln 2$ [23].

2.8.3 Blume-Capel Model

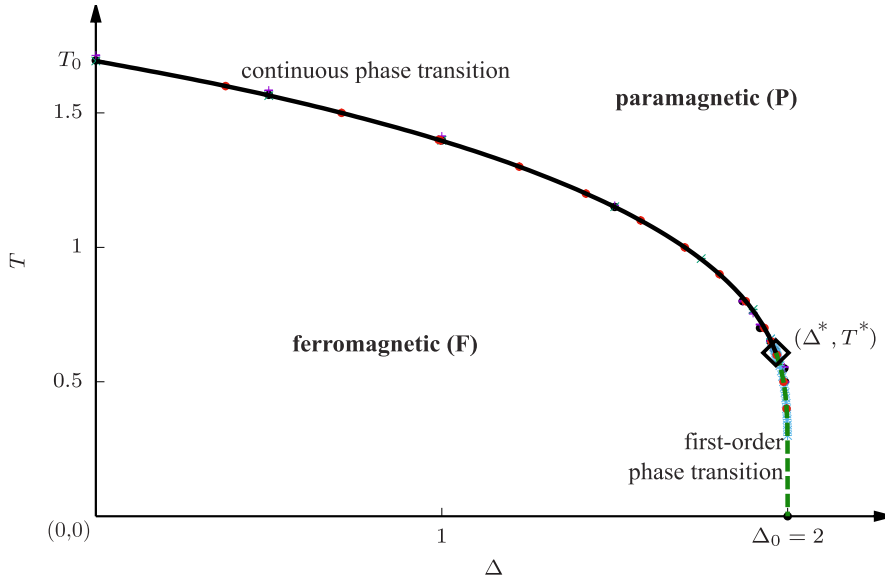


Figure 2.2: Phase diagram of the Blume-Capel model [24].

Finally, we study the *Blume-Capel model* which was independently introduced by Blume and Capel in the 1960s. It is (in zero field) described by a two-term Hamiltonian. It exhibits a phase transition that is governed by the interplay of the two terms. This makes it a suitable candidate for testing the mixed ensemble. The Hamiltonian is

$$\mathcal{H}(\{\sigma_i\}) = - \sum_{\langle ij \rangle} \sigma_i \sigma_j + \Delta \sum_{i=1}^N \sigma_i^2, \quad \sigma \in \{-1, 0, 1\}, \quad (2.62)$$

where Δ is known as the crystal-field strength. It controls the density of $\sigma = 0$ spins, such that for $\Delta \rightarrow -\infty$, the model becomes equivalent to the Ising model [24]. Fig. 2.2 shows a phase diagram in the T - Δ -plane. The Blume-Capel model features a tricritical point at $(\beta^*, \Delta^*) \approx (1.645, 1.966)$. For $\Delta < \Delta^*$, the phase transition is continuous, while for $\Delta > \Delta^*$ it is first-order. As we are interested in first-order dynamics in this work, we will focus on the transition line to the right of the tricritical point.

In this model, we are measuring the bond energy e_J , the crystal field energy e_Δ and magnetization m . They are defined as

$$e_J = \frac{1}{N} \sum_{\langle ij \rangle} \sigma_i \sigma_j, \quad e_\Delta = \frac{1}{N} \sum_{i=1}^N \sigma_i^2, \quad \text{and} \quad m = \frac{1}{N} \sum_{i=1}^N \sigma_i. \quad (2.63)$$

The specific heat is defined as [24]

$$C = -\beta N (\langle e_J e_\Delta \rangle - \langle e_J \rangle \langle e_\Delta \rangle). \quad (2.64)$$

3 Results

3.1 Ising Model

3.1.1 Comparison to Exact Solution

To test the efficiency of the different ensemble PA methods, they are tuned such that they take the same amount of computational effort. In particular, we set the product of the population size R , the number of Monte Carlo sweeps θ , and the number of annealing steps f , $R \cdot \theta \cdot f$, to be constant. This is achieved by adjusting the target culling fraction such that the canonical and momentum ensemble PA require the same number of annealing steps as the ceiling ensemble PA, as the former two allow for finer tuning of the number of annealing steps. Simultaneously, the number of sweeps and the population size are set to the same value for all methods. For the momentum ensemble, we chose $\alpha = 1$, such that $M = N$. In Fig. 3.1a, c we show the estimated canonical specific heat of the $L = 32$ Ising model from the different methods, for the number of sweeps $\theta = 1$ and $\theta = 20$, respectively. The data is averaged over 16 simulation runs, each, to produce the error bars¹. Fig. 3.1b, d show the deviation of the average energy in the same simulations from the exact solution [20]. The dashed line marks the pseudo-transition in the exact solution. For completeness, Fig. 3.2 shows the estimated probability canonical energy distribution at the pseudo-critical point $\beta_{c,L=32} = 0.4359\dots$ for the same runs, compared to the exact solution. There, one can observe that the momentum ensemble features the smallest statistical error. The reason lies in the slower traversal through the critical region.

It is evident that canonical and momentum PA perform better with a higher number of sweeps while this does not appear to be the case for ceiling PA. To gain a closer insight into the statistics, we now consider the family observables.

3.1.2 Family Observables

Fig. 3.3 shows the evolution of the average family size ρ_t ² during the same simulation runs. In the case of $\theta = 1$, there are pronounced, sudden increases in ρ_t in the canonical

¹For ease of the eye, only some of the error bars are shown

²The figures show $\rho_t - 1$ rather than ρ_t since the former allows for a nicer view in a logarithmic plot, as by definition $\rho_t^0 = 1$

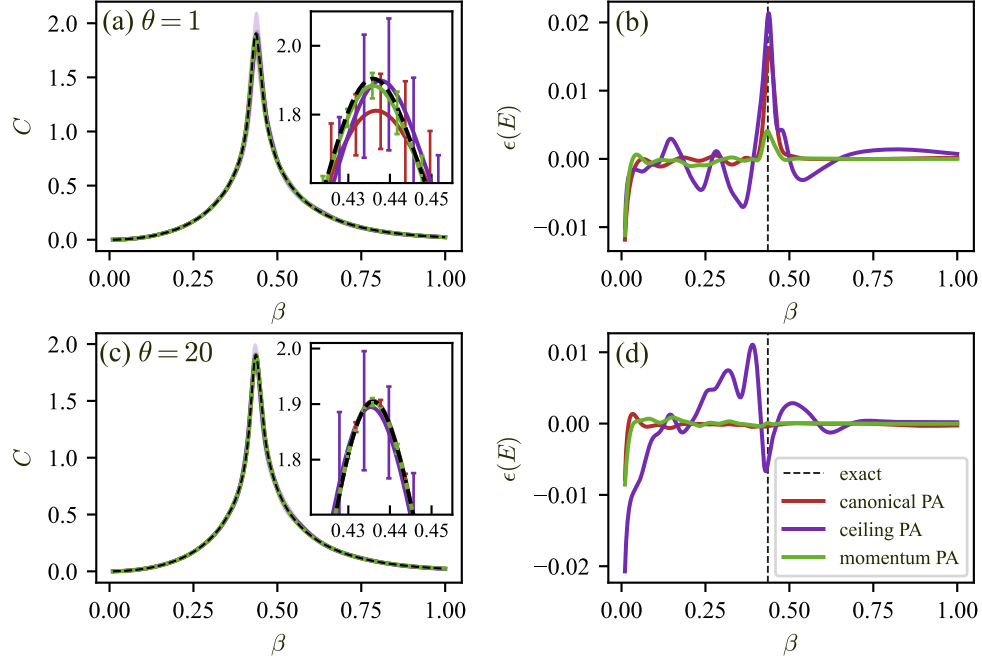


Figure 3.1: (a) (Canonical) specific heat for $L = 32$ estimated from the different ensembles using Metropolis update algorithm and $\theta = 1$, $R = 5000$, weight exponent $\alpha = 1$, compared to the exact solution (dashed line). (b) Deviation $\epsilon(E) = (E - E_{\text{exact}})/E_{\text{exact}}$ from the exact solution, the dashed line shows the pseudo-critical point. (c,d) $\theta = 20$, respectively.

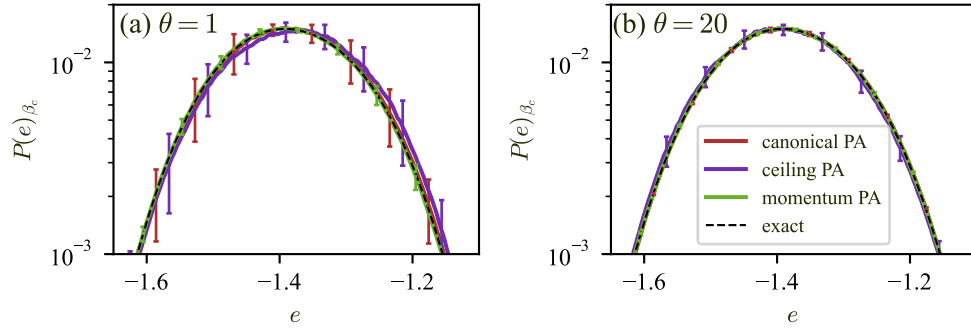


Figure 3.2: (a) Canonical energy distribution $P(e)_{\beta_c}$ at the (exact) pseudo-critical point for $L = 32$ estimated from the different ensembles using Metropolis update algorithm and $\theta = 1$, $R = 5000$, weight exponent $\alpha = 1$, compared to the exact solution (dashed line). (b) $\theta = 20$, respectively.

and momentum PA simulation runs, a sign of insufficient decorrelation by the MCMC step, which is larger for the canonical PA. This indicates an increase in the autocorrelation time. By choosing a higher number of sweeps, as in Fig. 3.3b, this effect is compensated equally well in the canonical and momentum PA run.

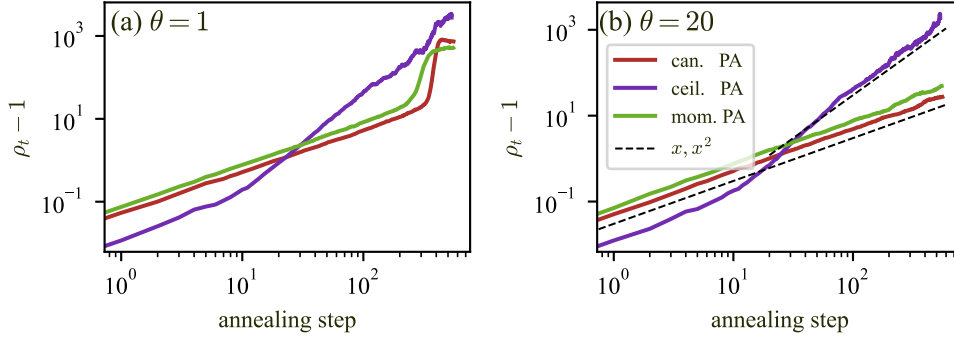


Figure 3.3: (a) Average family size ρ_t for different ensembles for the $L = 32$ Ising model using Metropolis update algorithm and $\theta = 1$, $R = 5000$. (b) $\theta = 20$ respectively.

The growth of ρ_t in the ceiling PA, however, behaves quite differently. For one thing, the slope of ρ_t is barely affected by a higher number of sweeps, indicating a short autocorrelation time. Secondly, for a large amount of the simulation run, ρ_t grows much steeper than in the other PA methods. The cause of this lies in a peculiarity of the ceiling ensemble: its culling fraction is directly tied to the DOS and cannot be freely tuned by adjusting the energy ceiling \mathcal{E} . In the ceiling ensemble, the culling fraction is given by

$$\epsilon(\mathcal{E}) = \frac{\Omega(E = \mathcal{E})}{\sum_{\mathcal{E}' \leq \mathcal{E}} \Omega(E = \mathcal{E}')} \quad (3.1)$$

This has the consequence that the culling fraction will only change in discrete steps with a minimum $\epsilon > 0$, making it impossible to freely tune the culling fraction in discrete systems. As the exact DOS is known for the Ising model, we are able to calculate $\epsilon(\mathcal{E})$, which is shown in Fig. 3.4a. As can be seen there, the minimal culling fraction reaches up to $\sim 80\%$ for a significant part of the annealing process. Fig. 3.4b shows the optimal ρ_t , reconstructed from the exact solution $\epsilon(\mathcal{E})$ with the help of (2.32). There, it becomes clear that even for perfectly uncorrelated samples, the steep incline in ρ_t cannot not be diminished.

The optimal strategy is therefore to employ an annealing schedule that stops at every energy level, such that number of annealing steps f scales with the system size N . This scaling also has the consequence that $\rho_t^{(f)}$ will also grow with N . As such, we need to scale $R \sim N$ in order to keep the buildup of correlations at bay.

For the other ensembles, the culling fraction can in fact be held constant with the help of an adaptive annealing schedule, as any value for ϵ can be dialled in. Thus, with (2.30), it is clear that ρ_t will grow linearly in the case of sufficient decorrelation, as is the case in Fig. 3.3. The scaling of the histogram width $\sim \sqrt{N}$ in both the canonical (where $\sigma_E^2 = CN$) and momentum PA (cf. (2.44) if $M \sim N$) leads to a scaling of the number of annealing steps $f \approx N/\sigma_E \sim \sqrt{N}$. Thus, the population size R likewise only needs to scale with the square root of the system size, \sqrt{N} .

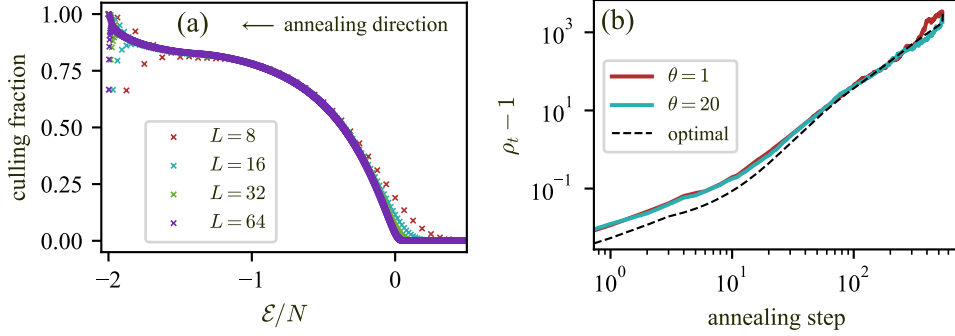


Figure 3.4: (a) Culling fractions obtained from the exact solution [20]. (b) Comparison of the observed ($L = 32$, $R = 5000$) value of ρ_t to the optimal value obtained from the exact solution in the ceiling ensemble.

3.2 Potts Model

In this section we investigate the performance of the PA algorithms under a first-order phase transition, as seen in the Potts model for $q > 4$. Fig. 3.5 shows the energy, specific heat, magnetization and magnetic susceptibility for the $q = 10$, $L = 32$ Potts model in the different PA methods, where the simulation was performed with the same computational effort for each method $R \cdot \theta \cdot f = \text{const.}$ in the same manner as for the Ising model above. The data is averaged over 16 simulation runs each, with the statistical error shown as the shaded region adjacent to the lines. As expected, the canonical PA performs poorly near the pseudo-critical first-order transition point where it has large statistical errors due to a diminished effective sample size, as well as a delayed transition, which we identify as a hysteresis effect. This occurs as a consequence of the exponentially suppressed transition probability from the disordered to the ordered phase.

In Fig. 3.6a we show the estimated density of states and Fig. 3.6b shows the canonical energy distribution near the pseudo-critical point $\beta_{c,q=10,L=32} \approx 1.423$ as estimated with data from [25]. As can be seen, the two microcanonical PA methods are both able to resolve

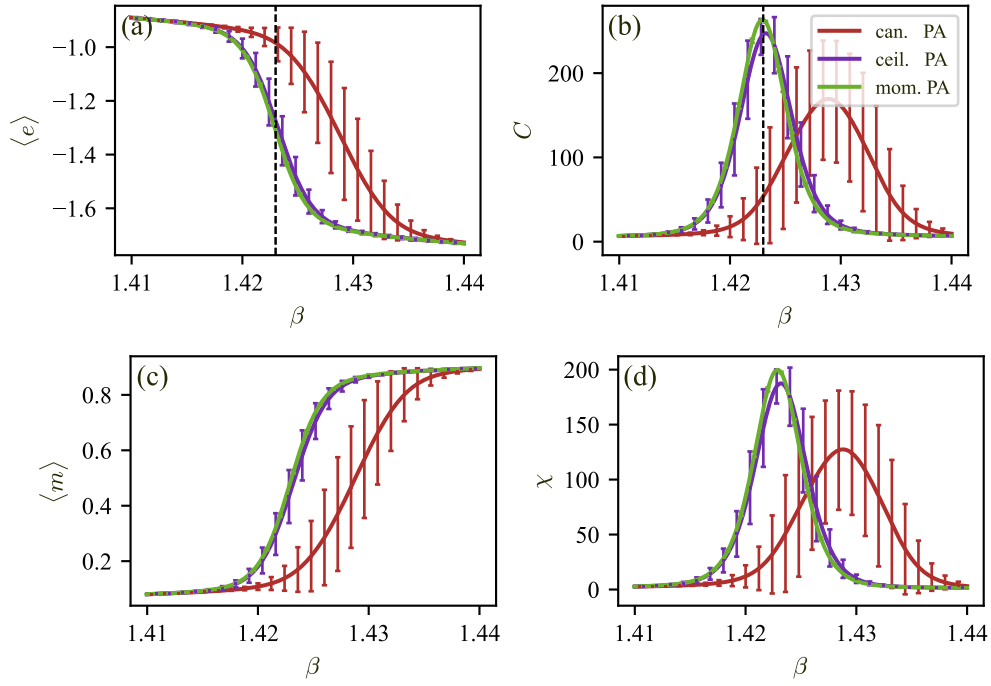


Figure 3.5: (a) Canonical average energy in $q = 10, L = 32$ Potts model estimated in the three ensemble methods with $R\theta f \approx \text{const.}$ (b) Specific heat. (c) Average magnetization. (d) Magnetic susceptibility.

the low-probability area in between the ordered and disordered phase peaks, leading to the characteristic double peak shape. The ceiling ensemble is showing higher error bars as a consequence of the higher correlated population. At the pseudo-critical point, only some of the canonical simulation runs show the transition, while others are still stuck in the ordered phase, which is the reason for the large error bars.

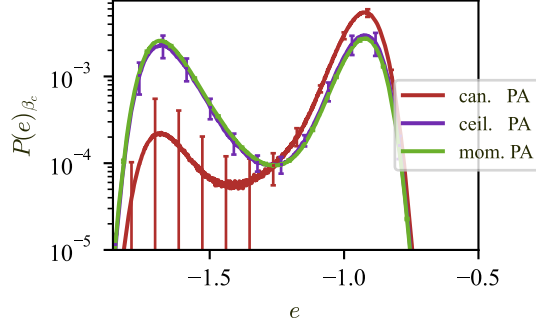


Figure 3.6: Canonical energy distribution $P(e)_{\beta_c}$ at the (exact) pseudo-critical point $\beta_c = 1.423$ [25] for $q = 10, L = 32$, estimated from the different ensembles using the heatbath update algorithm and $\theta = 20, R = 10000$, weight exponent $\alpha = 1$.

3.2.1 Family Observables

In the following family statistics (Fig. 3.7), we see a similar pattern concerning the performance. In the left panel we show ρ_t in the case of inadequate decorrelation, where we see the characteristic jump at the pseudo-phase transition for all methods, indicating the effects of the high autocorrelation time on the effective population size.

As shown in the right panel, the sudden increase can be compensated for momentum ensemble and ceiling ensemble but is unchanged for canonical ensemble, which is expected, as exponentially critical slowing down (ECSO) is occurring.

3.2.2 Weight Function Exponent

In the momentum ensemble, as described in (2.43), we have the additional parameter M , which influences the width of the histogram seen during the annealing process. Here, we set $M = \alpha N$ and investigate the role of α in the context of the Potts model first-order transition. Fig. 3.8a shows ρ_t for the $L = 32, q = 10$ Potts model for varying α . As can be seen, a higher α generally decreases the slope of ρ_t with the cost of a higher jump at the critical point.

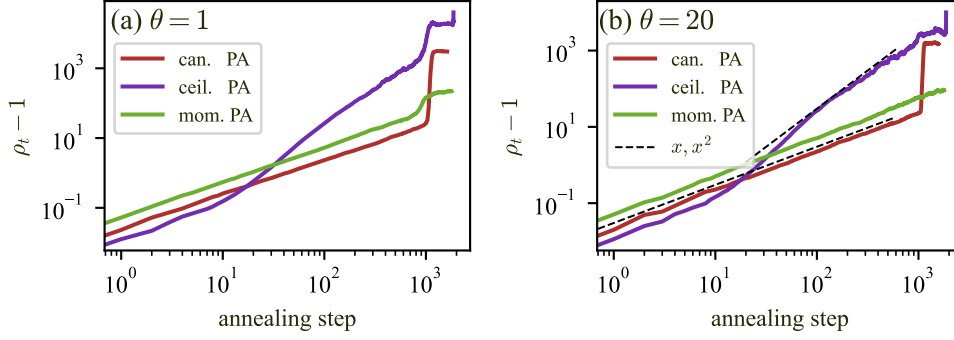


Figure 3.7: (a) Average family size ρ_t for different ensembles for the $L = 32, q = 10$ Potts model using heatbath update algorithm and $\theta = 1, R = 10000$ for the canonical and momentum PA run and $R = 40000$ for the ceiling PA in order to properly resolve at large ρ_t . (b) $\theta = 20$, with $R = 10000$ and $R = 20000$, respectively.

Whether the jump corresponds to a first-order transition with the consequence of ECSD, can be discerned from the ensemble's critical energy distribution seen during the simulation, which is shown in Fig. 3.8b. ECSD will appear as soon as the distribution is wide enough to reveal the double peak. But even before that point, the decorrelating effects of the MCMC seem to be diminished by the widening of the histogram, as there remains a small jump in ρ_t for $\alpha = 10, 25$ even though the double is not yet visible. This observation

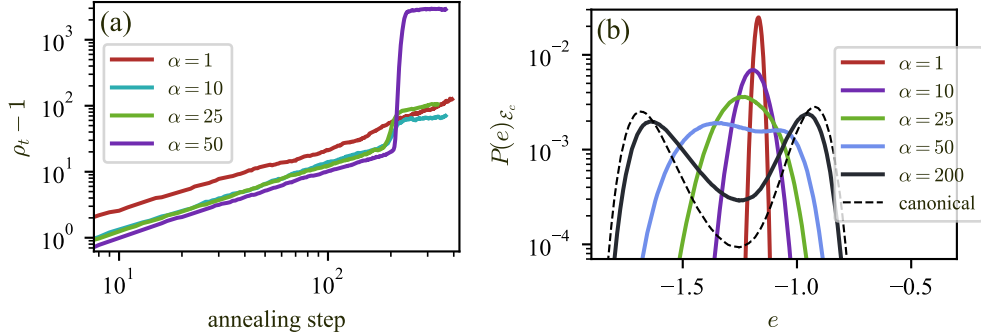


Figure 3.8: (a) Average family size ρ_t for different momentum ensemble with $\alpha = 1, 25, 50$ for the $L = 32, q = 10$ Potts model using heatbath update algorithm and $\theta = 20, R = 5000$. (b) Critical energy distribution of the $L = 32, q = 10$ Potts model in the momentum ensemble with varying α , compared to canonical critical energy distribution.

aside, we are still interested at which α the transition from a single peak to a double peak occurs in the energy distribution. Here, we investigate the hypothesis that the transition is primarily dependent on the ratio of the histogram width σ_e to the double peak distance $\Delta e_L \approx a/L + \Delta e_\infty$. To test this, we choose α such that by the approximation (2.44), it

3 Results

stays below the peak distance of the Potts model. On testing, we found that when we set $\alpha = bL$, we find a single peak for small b and a double peak for larger b . This is shown in Fig. 3.9a, c.

Another assessment of whether the transition is of first order, is to investigate the scaling of the specific heat. The specific heat $C = N\beta^2\sigma_e^2$ scales, in the case of first order transitions, $\sim N$ at the critical point β_c [5]. To assess the scaling in the momentum ensemble, we define a property resembling the specific heat, $\tilde{C} = N\sigma_e^2$. If the approximation (2.44) holds,

$$\tilde{C}_{\mathcal{E}_c} = N\sigma_e^2 \approx \frac{\alpha}{2k^2}, \quad (3.2)$$

while for a first order transition we would expect $\tilde{C}_{\mathcal{E}_c} \sim N = L^2$. The comparison between the two cases is shown in Fig. 3.9b, d.

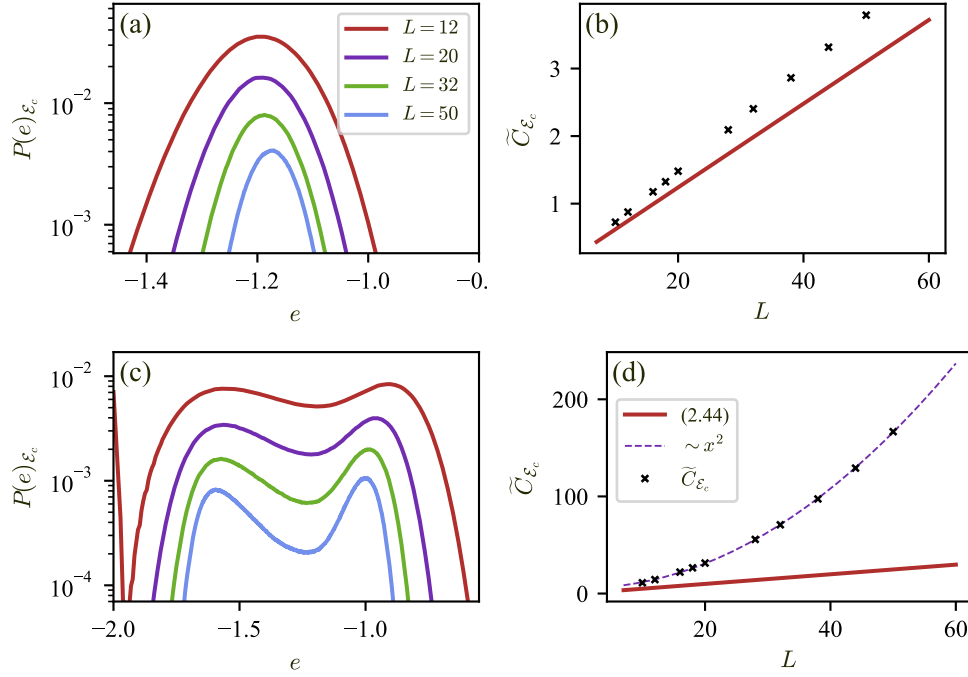


Figure 3.9: (a) Critical energy distribution of the $q = 10$ Potts model for various system sizes in the momentum ensemble with $\alpha = 1/4L$. (b) Scaling of estimated $\tilde{C}_{\mathcal{E}_c} = N\sigma_e^2$ compared to approximation in (2.44). (c,d) $\alpha = 2L$, respectively.

3.3 Blume-Capel Model

In this section we analyze the performance of the canonical, momentum, and, importantly, the mixed ensemble PA for the Blume-Capel model. This is done in order to assess the effect the two-term Hamiltonian has on the performance of the methods across the phase transition.

We use two types of canonical PA, one annealing in the inverse temperature β (with fixed Δ) and the other in the crystal field strength Δ (with fixed β , respectively), along with momentum PA in the total energy \mathcal{E} , and a mixed ensemble PA, as defined in (2.48), where the coordinate J_0 is equated with Δ .

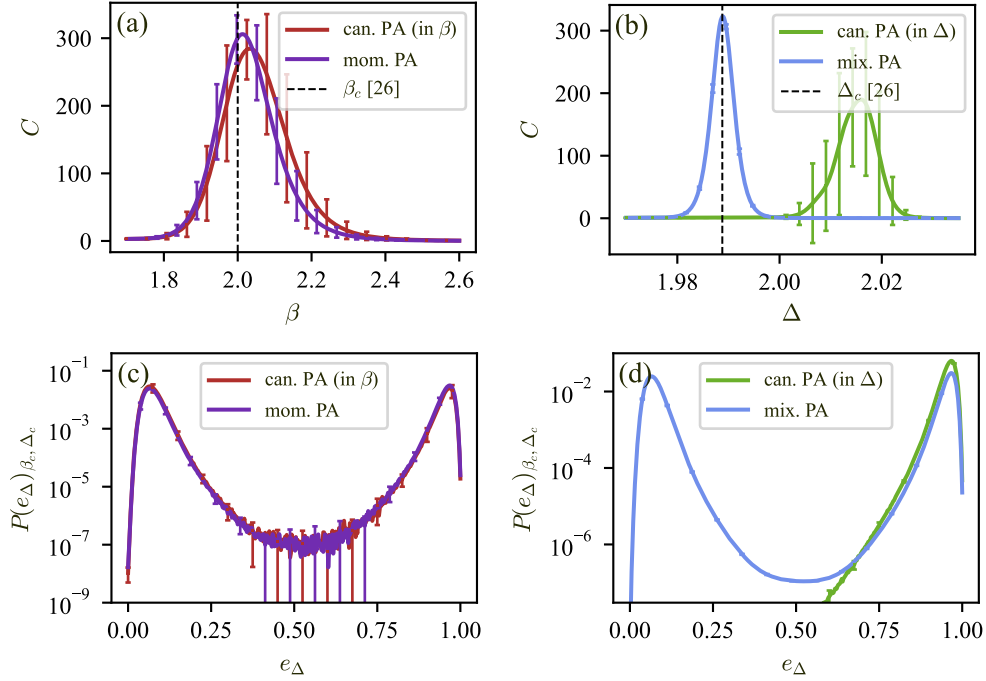


Figure 3.10: (a,b) Canonical specific heat C of the $L = 20$ Blume-Capel model, simulated with canonical, momentum and mixed ensemble, with $\theta = 20$, $R = 10000$ and the same number of (adaptive) annealing steps, estimated from 16 independent PA runs for each method. The dashed line shows the transition point $(\beta_c, \Delta_c) = (2, 1.989)$. (c,d) Canonical distribution of E_Δ at the first-order transition point (β, Δ) estimated from the same simulation runs.

We aim to cross the phase transition at $(\beta, \Delta)_{c, L=20} = (2, 1.989)$ [26], where the system undergoes a first-order transition (see Fig. 2.2 for the phase diagram). In terms of performance, canonical PA in β and momentum PA are found to perform similarly well, as

shown in Fig. 3.10a, c, which shows the estimation of the specific heat and the distribution of the crystal field energy e_Δ near the critical region. Fig. 3.11 illustrates the corresponding evolution of ρ_t .

Among the methods, the mixed ensemble PA seems to perform most effectively. It exhibits the smallest statistical error and, notably, did not display the sharp increase in ρ_t , indicating that it was the only method that was able to successfully suppress the first-order phase transition.

In contrast, the canonical PA in Δ showed the strongest delay in the phase transition. In fact, in none of the 16 runs the system had undergone the phase transition at the critical crystal field coupling Δ_c . This can be attributed to the fact that, when annealing in Δ , the phase transition is crossed over a much shorter interval compared to annealing in β so the system had not enough time to relax. This difference is evident from the specific heat peaks, which are much narrower in the latter case, indicating a more abrupt transition.

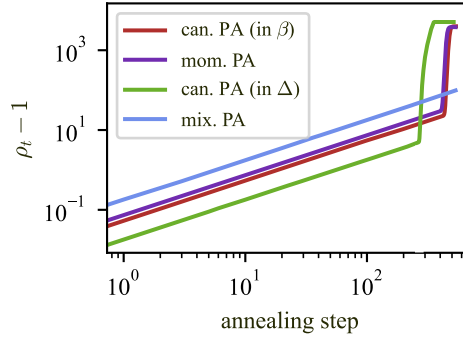


Figure 3.11: Average family size ρ_t of the same simulation runs shown in Fig. 3.10.

4 Conclusion and Outlook

In this work, we proposed and evaluated novel additions to the population annealing (PA) framework, particularly the momentum and the mixed ensemble PA methods. These methods are built upon the microcanonical ensemble, where the total energy \mathcal{E} is the control parameter. They are aimed at improving the performance of PA at first-order transitions, where canonical PA shows inadequate performance. We generalized the momentum and mixed ensemble by introducing the additional parameter α that can be used to effectively narrow or broaden the observed energy histogram such that the number of necessary annealing steps is dependent on α . All the measurements in these ensembles could be successfully transformed into canonical observables.

We performed comparisons to the exact solution for the Ising model to gauge the systematic error. Additionally, we compared the method's performance to the canonical PA simulations on the Potts and Blume-Capel model, which exhibit first-order phase transitions. We employed the replica averaged family size ρ_t to track the occurrences of inadequate decorrelation.

Our results show that momentum PA slightly outperforms canonical PA in the Ising model, while the ceiling ensemble shows a much higher increase in correlations. In the Potts model, our findings indicate that it is indeed possible to suppress the first-order phase transition during the annealing process with both the ceiling and momentum PA. By tuning the parameter α , we found that the width of the histogram is essential as to whether the double peak distribution (and therefore the first-order transition) is observed while simulating. A value of $\alpha \sim L$ is still able to suppress the first-order transition, but further analysis is needed, as we suspect the autocorrelation time to be dependent on α as well.

Our analysis of ρ_t revealed a significant difference in the scaling of both simulation time and statistical errors between ceiling and momentum PA. Both the number of annealing steps as well as the maximum ρ_t scale $\sim N$ in energy ceiling PA, while only with $\sim \sqrt{N}$ in momentum PA at a constant number of sweeps. While this makes momentum PA look like a more attractive choice of method, it remains to be tested how the autocorrelation time scales with the system size¹. It will be very insightful to study the scaling of τ_{int} with

¹Actually, ceiling PA is suspected to perform much better in this regard, at a single annealing step, the sliver of phase space that is explored, is in fact much narrower than in the other ensembles and shows no obvious scaling with the system size.

L in the different ensembles, and additionally under α . Conversely, measuring R_{eff} would examine the effects of the autocorrelation time on the population, directly.

Finally, we tested the mixed ensemble PA, which resembles momentum PA, but can perform the annealing in a single, isolated phase space coordinate. We tested this approach on the first-order transition in the Blume-Capel model, where it turned out to be very effective, while standard momentum PA performed only comparable to the traditional, canonical approach.

The implementations used in this study will benefit greatly from further optimization and enhancements to enable their use in distributed computing environments and to leverage the parallel computing power of GPUs.

In future studies, it will be very insightful to test the performance these methods on systems with rough free energy landscapes that also involve first-order transitions, e.g. systems with disorder, like bond-diluted and frustrated spin models.

Appendices

A Derivation of the Variance of the Number of Copies

To derive the variance $\sigma^2(r_i)$, we start with the observation that for an (on average) constant population size, the expectation value of the number of copies is equal to one. Additionally, it is clear that the probability $p(r_i = 0)$ is equal to ϵ , and $p(r_i > 0)$ is equal to $1 - \epsilon$, respectively. Thus, we can write the expected number of copies as

$$E[r_i] \stackrel{!}{=} 1 = \epsilon \cdot 0 + (1 - \epsilon) \tilde{\tau}_i, \quad (\text{A.1})$$

where $\tilde{\tau}_i = (1 - \epsilon)^{-1}$ is the expected number of copies, given that the replica survives. In fact, in the ceiling ensemble, τ_i is identical for all surviving replicas such that $\sigma_{\tilde{\tau}_i}^2 = 0$.

If we assume this is generally true, then for nearest-integer resampling (2.16), every surviving replica will either be copied $\lfloor \tilde{\tau}_i \rfloor + 1$ or $\lfloor \tilde{\tau}_i \rfloor$ times, where $q = \lfloor \tilde{\tau}_i \rfloor - \tilde{\tau}_i$ is the probability to make $\lfloor \tilde{\tau}_i \rfloor + 1$ copies, and in turn $(1 - q)$ is the probability to make $\lfloor \tilde{\tau}_i \rfloor$ copies. Thus, the probability distribution of the number of copies can be written as

$$p(r_i) = \begin{cases} \epsilon & r_i = 0 \\ (1 - q)(1 - \epsilon) & r_i = \lfloor \tilde{\tau}_i \rfloor \\ q(1 - \epsilon) & r_i = \lfloor \tilde{\tau}_i \rfloor + 1, \end{cases} \quad (\text{A.2})$$

with $\mu = 1$ and variance

$$\sigma^2(r_i) = \epsilon \cdot (0 - \mu)^2 + (1 - q)(1 - \epsilon) \cdot (\lfloor \tilde{\tau}_i \rfloor - \mu)^2 + q(1 - \epsilon) \cdot (\lfloor \tilde{\tau}_i \rfloor + 1 - \mu)^2, \quad (\text{A.3})$$

which after plugging in μ , q , and $\tilde{\tau}_i$ can be simplified to

$$\sigma^2(r_i) = \left\lfloor \frac{1}{1 - \epsilon} \right\rfloor \left(\epsilon + 1 + \left\lfloor \frac{1}{1 - \epsilon} \right\rfloor (\epsilon - 1) \right), \quad (\text{A.4})$$

which for $\epsilon < 0.5$ can be further simplified to

$$\sigma^2(r_i) = 2\epsilon. \quad (\text{A.5})$$

B Approximation of the Histogram Width in the Momentum Ensemble

Here, we try to quantify the effect of the weight exponent on the histogram width. For this, we approximate the DOS near the critical point as $\Omega(E) \sim \exp(kNe) = \exp(kE)$, with $e := E/N$ and a size-independent constant k , as shown in Fig. B.1.

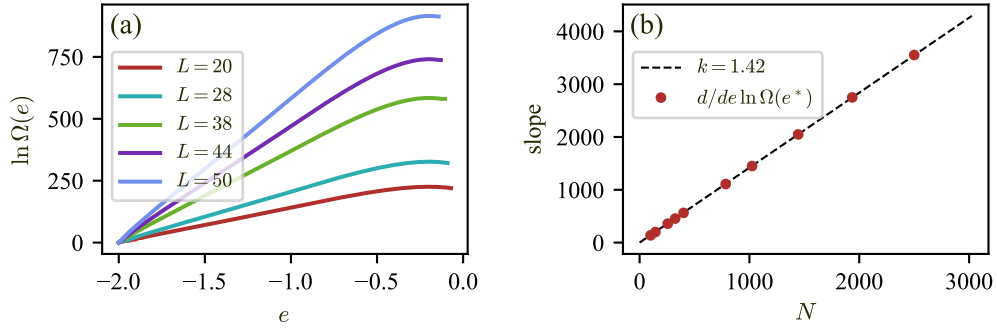


Figure B.1: (a) Density of states $\Omega(e)$ of the $q = 10$ Potts model for various system sizes. (b) Slope of $\ln \Omega(e_c)$ with e_c in the region where $\ln \Omega$ is linear.

We can write the histogram as

$$P(E) = \frac{1}{C} \Omega(E) (\mathcal{E} - E)^{\frac{M-2}{2}} \sim \exp \left[\underbrace{\left(kE + \frac{M-2}{2} \ln(\mathcal{E} - E) \right)}_{=: f(E)} \right]. \quad (\text{B.1})$$

Where we have identified the function $f(E)$. It has a global maximum at $E_0 = \frac{M-2}{2k} + \mathcal{E}$. This has the consequence that the first-order term in the Taylor expansion becomes zero if we expand around E_0 while the second order term is negative, such that

$$f(E) \approx f(E_0) + \underbrace{f'(E_0)}_{=0} (E - E_0) + \frac{1}{2} \underbrace{f''(E_0)}_{=-2k^2/(M-2)} (e - e_0)^2 + \dots \quad (\text{B.2})$$

If $N, M \gg 1$, higher order terms can be neglected and we find the approximation

$$P(e) \sim \exp \left[f(E_0) \right] \exp \left[-\frac{1}{2} \frac{(E - E_0)^2}{\sigma_E^2} \right], \quad \text{with} \quad \sigma_E^2 = \frac{M-2}{2k^2}. \quad (\text{B.3})$$

Acknowledgements

I gratefully acknowledge the suggestion of the topic as well as help and many useful discussions over the span of the thesis by Denis Gessert, Martin Weigel and Wolfhard Janke.

Bibliography

- [1] T. Brown et al., Language Models are Few-Shot Learners, in *Advances in Neural Information Processing Systems*, Vol. 33 (2020), p. 1877.
- [2] C. J. Geyer et al., Markov Chain Monte Carlo Maximum Likelihood, in *Computing science and statistics: Proceedings of the 23rd Symposium on the Interface* (1991), p. 156.
- [3] K. Hukushima and K. Nemoto, Exchange Monte Carlo Method and Application to Spin Glass Simulations, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [4] L. Barash, M. Weigel, L. N. Shchur, and W. Janke, Exploring first-order phase transitions with population annealing, *Eur. Phys. J. Spec. Top.* **226**, 595 (2017).
- [5] W. Janke, First-order phase transitions, in *Computer Simulations of Surfaces and Interfaces* (Springer, Dordrecht, 2003).
- [6] B. A. Berg and T. Neuhaus, Multicanonical ensemble: A new approach to simulate first-order phase transitions, *Phys. Rev. Lett.* **68**, 9 (1992).
- [7] F. Wang and D. P. Landau, Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [8] P. Schierz, J. Zierenberg, and W. Janke, First-order phase transitions in the real microcanonical ensemble, *Phys. Rev. E* **94**, 021301 (2016).
- [9] N. Rose and J. Machta, Equilibrium microcanonical annealing for first-order phase transitions, *Phys. Rev. E* **100**, 063304 (2019).
- [10] V. Martin-Mayor, Microcanonical approach to the simulation of first-order phase transitions, *Phys. Rev. Lett.* **98**, 137207 (2007).
- [11] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, Cambridge, 2007).
- [12] W. Janke, Monte Carlo simulations in statistical physics – From basic principles to advanced applications, in *Order, Disorder and Criticality* (2013), p. 93.
- [13] K. Hukushima and Y. Iba, Population Annealing and Its Application to a Spin Glass, *AIP Conf. Proc.* **690**, 200 (2003).
- [14] D. Gessert, W. Janke, and M. Weigel, Resampling schemes in population annealing: Numerical and theoretical results, *Phys. Rev. E* **108**, 065309 (2023).
- [15] W. Wang, J. Machta, and H. G. Katzgraber, Population annealing: Theory and application in spin glasses, *Phys. Rev. E* **92**, 063307 (2015).
- [16] M. Weigel, L. Barash, L. Shchur, and W. Janke, Understanding population annealing Monte Carlo simulations, *Phys. Rev. E* **103**, 053301 (2021).
- [17] J. Callahan and J. Machta, Population annealing simulations of a binary hard-sphere mixture, *Phys. Rev. E* **95**, 063315 (2017).

-
- [18] R. Kubo, *Statistical Mechanics* (New Holland, Amsterdam, 1965) Chap. 2.
 - [19] L. Barash, J. Marshall, M. Weigel, and I. Hen, Estimating the density of states of frustrated spin systems, *New J. Phys.* **21**, 073065 (2019).
 - [20] P. D. Beale, Exact distribution of energies in the two-dimensional Ising model, *Phys. Rev. Lett.* **76**, 78 (1996).
 - [21] L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition, *Phys. Rev.* **65**, 117 (1944).
 - [22] F. Y. Wu, The Potts model, *Rev. Mod. Phys.* **54**, 235 (1982).
 - [23] C. Borgs and W. Janke, An explicit formula for the interface tension of the 2D Potts model, *J. Phys. I* **2**, 2011 (1992).
 - [24] J. Zierenberg, N. Fytas, M. Weigel, W. Janke, and A. Malakis, Scaling and universality in the phase diagram of the 2D Blume-Capel model, *Eur. Phys. J. Spec. Top.* **226**, 789 (2017).
 - [25] A. Billoire, R. Lacaze, and A. Morel, A numerical study of finite-size scaling for first-order phase transitions, *Nucl. Phys. B* **370**, 773 (1992).
 - [26] W. Kwak, J. Jeong, J. Lee, and D. Kim, First-order phase transition and tricritical scaling behavior of the Blume-Capel model: A Wang-Landau sampling approach, *Phys. Rev. E* **92**, 022134 (2015).

Declaration of Authorship

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works of any other author, in any form, is properly acknowledged at their point of use.

Till Pfaff

Leipzig, September 18, 2024