

High precision measurements of interfaces' free energy

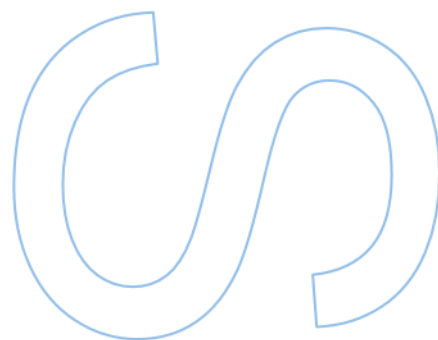
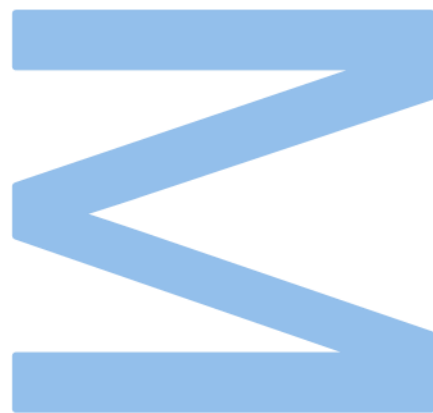
David Capitão Lima

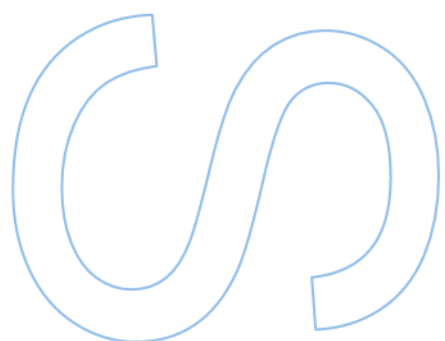
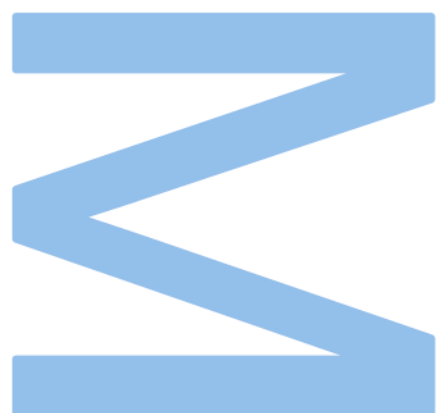
Mestrado em Física

Departamento de Física e Astronomia
2022

Orientador

João Manuel Viana Parente Lopes, FCUP - CFP





(Artigo 14.º do Código Ético de Conduta Académica da U.Porto)

Declaração de Honra

Eu, David Capitão Lima, inscrito no Mestrado em Física da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

David Capitão Lima

Porto, 19 de janeiro de 2023

UNIVERSIDADE DO PORTO

FACULDADE DE CIÊNCIAS

MESTRADO EM FÍSICA

**High precision measurement of interfaces’
free energy**

David Capitão Lima

Supervisor

Prof. Dr. João Manuel Viana Parente Lopes

*A thesis submitted in fulfilment of the requirements
for the degree of Master of Science*

at

Departamento de Física e Astronomia

Porto, December of 2022

[This page was intentionally left blank]

“Man stands face to face with the irrational. He feels within him his longing for happiness and for reason. The absurd is born of this confrontation between the human need and the unreasonable silence of the world.”

-Albert Camus

Contents

Acknowledgements	x
Abstract	xii
Resumo	xiv
Introduction	xvi
List of Figures	xxiii
1. Interface free energy - from strings to Ising 3D	1
1.1. Nambu-Goto action / capillary wave model	1
1.2. Polyakov Lagrangian	3
1.3. Energy spectrum of a closed string	8
1.4. Partition function	9
1.5. $1/\mathcal{A}$ expansion for finite strings	10
2. Monte Carlo	12
2.1. General properties	12
2.1.1. Importance sampling	12
2.1.2. Markov process	13
2.1.3. Ergodicity	14
2.1.4. Detailed balance	15
2.2. Metropolis-Hastings algorithm	17
2.3. Cluster algorithms	18
2.3.1. Wolff algorithm	18
3. 3D Ising model with continuous wall coupling	21
3.1. Phenomenology of the domain wall	23
3.1.1. Bubble removal procedure	23
3.1.2. Number of walls	23
3.1.3. Position of the wall	24

3.2.	$J_{\text{Wall}} = -1$ scenario	28
3.2.1.	Renormalization group analysis for the KT transition of the two-dimensional XY model	29
4.	Non-equilibrium methods	31
4.1.	Jarzynski Equality	31
4.1.1.	Proof for deterministic evolution[1]	32
4.1.2.	Proof for Monte Carlo evolution	33
4.1.3.	Snap case ($N=1$)	34
4.1.4.	Adiabatic case ($N \rightarrow \infty$)	35
4.2.	Application to Ising 3D	35
4.3.	Normalizing flows	36
4.3.1.	Definition and inner working	37
4.3.2.	Affine layers	38
4.3.3.	Normalizing flows as a Probability Flow Monte Carlo	39
4.3.4.	NF for Free energy measurement	40
4.3.5.	Two particle case	41
4.3.6.	Hubbard-Stratonovitch transformation	42
5.	Multicanonical algorithm	46
5.1.	General idea	47
5.1.1.	Generic Wang Landau	48
5.2.	Entropy competition approach	50
5.3.	Independent bins statistical moments	52
6.	Algorithms' efficiency study	54
6.1.	Jarzynski	54
6.2.	Normalizing Flows	56
6.3.	Multicanonical	58
6.3.1.	Wang-Landau	58
6.3.2.	Entropy competition	60
6.3.3.	Independent bins momentum	62
7.	Data vs NG expansion	66
7.1.	From partition ratios to interface free energy	66
7.2.	Universal and non-universal contributions	68
7.3.	c_{-3} computation	69

8. Conclusions and Outlook	71
A. Markov Matrix properties	73
B. Statistical Iterators	75
B.1. Average	75
B.2. Variance	76
B.3. Correlation	77
C. Equations of motion for Polyakov Lagrangian and its equivalence with NG	80
D. Backpropagation for a simple network	82
E. Normalizing Flows Architecture	84
F. Free energy data	86
Bibliography	86

Acknowledgements

As one reads the work presented throughout this thesis, it is important to keep in mind the importance of collaboration in scientific work, be it direct or more subtle. And so, I take this opportunity to thank the people who have, in some way or another, contributed to the work presented in this document, but also to those who have made the past five years truly unforgettable.

First, to my supervisor, Prof. João Lopes, whose tireless dedication to anything he sets his mind onto was an awe-inspiring sight. It was a rewarding, despite very challenging, endeavour to work with such a creative mind during the last two years. I hope to carry the many advices you gave me onto the future. I would also like to thank José Matos, whose many brilliant insights are sprinkled through this work, and without which I certainly wouldn't have completed my Masters. Still in the context of this work, I would like to thank Dr. João Pedro Pires for the helpful hand he selflessly lent throughout this year.

To Inductiva Research Labs, in particular to Luís Sarmiento and David Carvalho, for accepting me into their small family and by teaching me many things about Deep Learning.

To Tiago Gonçalves and João Pinho; we were brave enough to face the Physics Masters together, and after tackling this challenge there is nothing that can take us down.

To all the friends that accompanied me through these years. Our adventures and stories are treasures I truly cherish. A particular thanks to Pedro Sousa and Tiago Souto, my partners in crime from the day one, Leonardo Hügens, Bruno “BMW” Mota, Porfírio Fernandes, to Alexandre Ribeiro for the numerous moonlight lit conversations and to Pedro Páscoa, for showing me that there are many ways to see the world.

To Beatriz, for everything. Mere words are not enough to express how much you mean to me. I can't wait to face what the future holds for us.

Finally, and most importantly, to my brother and parents, for their support and for teaching me all the things that matter. I hope I can make you proud.

[This page was intentionally left blank]

Abstract

In this thesis we explored different algorithms to precisely measure the interface free energy of the Ising model in three dimensions and understand its deviations from the prediction provided by an effective, capillary wave (CW) description. To this effect, we developed and tested a plethora of algorithms that can be categorized into either the non-equilibrium transformations (NET) class or into the multicanonical category. We found that the method with the best precision / computation time ratio was the Wang-Landau algorithm. Along that finding, we computed the first deviation from the free energy expansion predicted by the CW model for transversal square interfaces. We found that our results obey the expected constraints, but are not consistent between them.

[This page was intentionally left blank]

Resumo

Nesta tese exploramos diferentes algoritmos para medir com precisão a energia livre de uma interface do modelo de Ising em três dimensões e compreender os seus desvios da previsão dada or uma descrição efetiva em termos de ondas capilares. Para este efeito, desenvolvemos e testamos uma plethora de algoritmos, que podem ser categorizados quer numa classe de transformações de não-equilíbrio ou na categoria de algoritmos multicanónicos. Descobrimos que o método com a melhor precisão em termos do tempo de computação é o Wang-Landau. Também calculamos o primeiro termo de expansão não universal para uma série de diferentes tamanhos de rede. Concluimos que os resultados obedecem a condições conhecidas, mas não são auto-consistentes.

[This page was intentionally left blank]

Introduction

The confinement of quark-antiquark pairs in quantum chromodynamics (QCD) is seen to arise from the existence of color flux tubes in the confined phase of the gauge theory which can be modeled as thin vibrating strings[2, 3, 4, 5, 6]. Since such strings break translational invariance, the Goldstone theorem guarantees that massless Nambu-Goldstone bosons will emerge from the theory[7, 8]. It is possible[9] to integrate out the massive fields and reach an action valid for strings that fluctuate at long wavelengths: an *effective string theory* (EST) description (see refs. [10, 11, 12, 13] for a detailed review on this topic). Direct simulations on lattice QCD showed good results for this approximation[14, 15]. Later, people have come to realize that ESTs enjoy what is described as *low-energy universality*, arising from the symmetry constraints imposed by Poincaré invariance of the target space. Therefore, the Nambu-Goto[3, 4] (NG) action, fully determined by the area of the string’s worldsheet, proves to be the most important contribution in the regime of long and thin flux tubes, with all the other terms appearing as sub-leading corrections¹.

In order to indirectly study the properties of this EST, a lot of attention was devoted to understanding the properties of interfaces in three-dimensional systems dominated by massless excitations[18]. Such objects, known as *fluid interfaces*, can be described by field theoretical methods, which yield universal results, that are independent of microscopic detail. One of the simplest realizations of a two-dimensional interface is the domain wall in the 3D Ising model.

The Ising model, proposed to explain the ferromagnetic properties of materials, was first studied by Ernst Ising at the 1920’s [19]. In the one-dimensional model, Ising found the absence of a phase transition, and quickly rushed to the conclusion that the model lacked a phase transition at higher dimensions. Later on, the two-dimensional case was demonstrated to have a phase transition[20], which was then completely characterized by Onsager’s analytic solution[21]. For higher dimensions, the model also exhibits a critical point, that is well described by mean field theory above three dimensions. Unfortunately, an exact description of the 3D critical point still evades physicists to this day. The Ising model is the absolute poster child of Statistical Physics, and more than one hundred years after its inception, it is still widely studied, with a copious amount of computational work already published[22, 23].

Regarding 2D interfaces, we can study the domain wall that separates regions of up and down magnetization at low temperature ($\beta < \beta_C$) in the ferromagnetic phase. Such interfaces

¹It must be noted, however, that this approximation breaks down for some few select cases[16, 17].

appear when we impose anti-periodic boundary conditions along one of the directions. In the 90s, an impressive amount of work was devoted to understand the behavior of such an object [24, 25, 26, 27, 28, 29, 30], and the domain wall was shown to undergo a Kosterlitz-Thouless phase transition at $\beta_R \sim 2\beta_C$. The specifics of this transition goes beyond the scope of this thesis, but suffices to say that the behavior at $\beta_C < \beta < \beta_R$ is described precisely by the fluid interpretation.

The effective description of a rough interface is the one provided by the *capillary wave model*[31] (CW), a model that assumes an effective Hamiltonian proportional to the variation in the surface area. This is precisely the rationale behind NG and, in fact the two models are shown to be equivalent[31]. This means that we can perform measurements of the interfaces' free energy in the Ising model and, by studying its deviation from the prediction of the CW model, obtain insights about the EST behavior.

Objectives

For 30 years now, physicists have been struggling to measure the non-universal deviations to the Nambu-Goto action[15, 14, 32]. In this thesis, with the help of continuous improvements of CPU technology and the use of previously unchecked multicanonical approaches, we aim to break this paradigm for the 3D Ising model, through high-precision measurements of the interfaces' free energy and of the first non-universal correction to the Nambu-Goto / CW model. The free energy of the interface can be probed by comparing the free energy of a system with periodic boundary conditions with one having anti-periodic boundary conditions in one direction.

Thesis outline

We start, chapter 1, by deriving the universal contributions that arise from the Nambu-Goto expansion. We follow the approach presented in ref. [33]. It is also presented the perturbative expansion to which we will stake our data against.

In chapter 2, we outline the general principles behind Monte Carlo equilibrium simulations, that will be prevalent throughout the rest of the work.

In chapter 3 we will study the phenomenology of the domain walls in the Ising 3D model when we promote the wall coupling to a continuous value.

Then, in chapter 4, we describe the first class of algorithms we employ to compute the interfaces' free energy, consisting of the Jarzynski method[1] and the normalizing flow approach. We cast them together since both of them rely on the properties of non-equilibrium transformations between the periodic and anti-periodic Ising models. Moreover, it was recently shown that they belong to a more general class of algorithms: *Stochastic Normalizing Flows*[34]. The

Jarzynski approach is similar to [14], but with a distinct Monte Carlo update. The free energy computation with the Normalizing Flow is a novel approach.

We proceed by outlining three original multicanonical algorithms, that configure the second class of employed techniques.

We finish by presenting and comparing the results of the different algorithms, chapter 6, and by, employing the best of them, computing the first non-universal coefficient, chapter 7, contrasting it with previous results from the literature[35, 36, 37].

[This page was intentionally left blank]

List of Figures

1.1.	Example of the promotion of the particle's worldline to a string's worldsheet. Even further generalizations to higher dimensions are possible if we introduce the concept of a brane. Image from [38].	2
1.2.	Example of an open (A) and closed (B) string world-sheet	5
2.1.	A Wolff step for a finite temperature β builds a cluster (shaded region) and flips it. Image from [39].	19
3.1.	Possible 2D Ising model configuration with periodic conditions alongside the y -direction and anti-periodic ones alongside the x -direction and system size 512×512 . An interface forms between domains of up (white region) and down (black region) spins. The two possible metrics to determine the wall position are when m_z , defined at equation 3.3 equals zero; or when γ_z , defined at equation 3.2, yields its maximum.	22
3.2.	Example of three steps of the bubble removal procedure for a simple configuration of two-dimensional Ising model. The SW algorithm identifies the largest cluster and flips it.	24
3.3.	Distribution for the number of walls for a system with $J_{wall} = 0.80$, $L_z = 96$ and $L_x = L_y = 30$ at $\beta = 0.223102$. There are, as expected, only configurations with an odd number of walls. The yellow curve corresponds to an exponential fit to the data.	25
3.4.	Chemical potential μ with respect to the coupling wall constant for a system with parameters: $L_z = 96$ and $L_x = L_y = 32$. Despite the high error associated with each decay coefficient, one can distinguish a nearly constant behavior of μ up to $J_{wall} \sim 0.6$ and a monotonic increase afterwards.	26
3.5.	Wall position histogram for a system with $\beta = 0.223102, L_z = 96, L_x = L_y = 40$ for three different values of J_{wall} . It is important to note that the y-axis does not have the same scale on the three figures. In the middle figure, we compare our data with the uniform distribution. There are clearly three distinct behaviors - the wall is either more probable to be near $z = 0$ for $J_{wall} > -1$; to be uniformly distributed within the system or to be suppressed at the edges.	26

3.6.	Γ for a system with $\beta = 0.223102, L_z = 96, L_x = L_y = 40$ in function of J_{wall} . Γ decreases until $J_{wall} = -1$, that corresponds to the uniform distribution. After that, there is a slight increase, because the distribution is not quite uniform (walls are suppressed at z_0), but it never reaches the same values as before, due to the distribution still being uniform at the bulk.	27
3.7.	Γ for a system with $\beta = 0.223102, L_z = 96$, in function of J_{wall} for different $L_x = L_y = L$. Γ takes greater values for greater L , except at the fixed point for $J_{wall} = -1$, since $\Gamma = \frac{1}{L_z}$, regardless of the transversal area.	27
4.1.	Normalizing flows between a prior distribution $q_0(z)$ and a target distribution $q_N(\phi)$	37
4.2.	Measured free energy of the first 40 NF eras for the two particle system for a batch of 1000 independent measurements. The horizontal line corresponds to the analytical result. Each era consists of 100 epochs, and each epoch corresponds to a backpropagation update of the NF weights due to the forward pass of a batch of 640 configurations sampled from the $\lambda = 0$ case. As we can see, the results converge quickly to the true value. Further training of the Normalizing Flow don't allow the precision to exceed the fifth decimal place.	42
5.1.	Rough estimation for $\beta(J)$ for the system with $L_x = L_y = 16, L_z = 96$ and two realizations of $f(J)$, with the corresponding generated histograms. The intersection between $\beta(J)$ and $f_i(J)$ governs the localization of the histogram. Both f_1 and f_2 have the same slope m but different b 's.	51
6.1.	On the left plots we have the measurement of ΔF for a sample of 50000 initial configurations and one Wolff sweep between time steps for different values of N_{steps} for the direct (periodic $\rightarrow$ anti-periodic Ising model) and inverse update. The results tend to the value presented at [14]. The right plots represent the distribution $\rho(-\beta W)$ (blue) and $e^{-\beta W} \rho(-\beta W)$ (orange), for the distinct N_{steps} for the direct update.	55
6.2.	Uncertainty and deviation from the expected result for three different approaches to the Jarzynski algorithm. The blue line corresponds to one Wolff sweep between two consecutive updates; the orange line corresponds to ten Wolff sweeps. Both of them have a low sample of initial configurations ($N_{samples} = 96$). The green line is similar to the blue approach but has greater $N_{samples}$	56

6.3.	Distribution (blue histogram) of $\log \Xi \equiv -\beta\{H(g_N(z)) - H_0(z) - Q\}$ with z sampled for the periodic Ising model for different epochs of the Normalizing Flow training. The red line corresponds to the results predicted by the Jarzynski algorithm. The blue line corresponds to $\log\langle\Xi\rangle$. There is a clear dissonance between the expected result and the obtained one, and such discrepancy is clearly induced by the presence of some Ξ with manifestly high values.	57
6.4.	Rough $\chi(J)$ (blue) and the corrected $\chi_{\text{corr}}(J)$ for a system with $L_x = L_y = 32$ and $L_z = 96$. This graph is obtained with a standard Wang-Landau algorithm for $\epsilon_{\text{final}} = 2^{-12}$. Both the value $\chi(1)$ and $\chi_{\text{corr}}(1)$ are fixed at zero, since the WG function is defined up to a constant.	59
6.5.	Mean histogram $H(J)$, left plot, and the associated error for each bin, $\delta H(J)$, right-down plot, for a system with $L_x = L_y = 32$ and $L_z = 96$ with the $\chi(J)$ of Fig. 6.4.	60
6.6.	$\beta(J)$ and corresponding error for a system with size $8 \times 8 \times 16$. The error bars are too small to be visible in the first subfigure.	61
6.7.	$\mu_1 \equiv \langle j \rangle$ and $\mu_2 \equiv \langle j^2 \rangle - \frac{\Delta^2}{12}$ for $L_x = L_y = 32$ and $L_z = 96$	63
6.8.	Results for $\frac{dS}{dJ}$ and $\frac{d^2S}{dJ^2}$ for $L_z = 96$ and $L_x = L_y = 32$	64
6.9.	Difference between the two different ways to compute $\frac{dS}{dJ}$ (corrected Wang-Landau and IBM).	65
7.1.	$F^{(1)}(L)$ (up) and $F^{(1)}(L) - F^{(2)}(L)$ (down) for a system with $L_z = 96$, $\beta = \beta_{\text{chosen}}$ and $L_x = L_y = L$	67
7.2.	$F^{(2)}(A = L^2)$ for various L_z values. The transversal area has $L_x = L_y = L^2$. Both graphs represent the same data but the second one is in a log-log scale.	68
7.3.	MSE with respect to proposed σ 's (a)). $F^{(2)}$ data for various square transversal areas with $L_z = 96$ and correspondent fit. b) and c) present the same data, but with a regular and log-log plot, respectively	69
7.4.	Dots correspond to our non-universal data for the various system sizes. The continuous lines are the fit obtained with Eq. 7.6.	70
D.1.	Architecture of the NN employed in the appendix.	82

E.1. FCNN architecture employed in all the layers of the two particles Normalizing Flow. The first layer has ReLU activation functions and the final layer has Tanh activation functions. The first layer takes as input the non-active coordinate, i.e., the coordinate that will not change by application of the layer. The values for the s and t function can be computed at once, and correspond to the two outputs of the NN. The expressivity of such functions is controlled by the number of nodes present in the hidden layer - the more nodes the more expressive the functions become. The weights associated with each node are updated through the backpropagation algorithm, and aim to minimize the loss function 4.33.	85
E.2. CNN architecture employed in all the layers of the Ising model. The input is the non-active sublattice. The shape of the sublattice is the same as the one of the lattice, but half of the values are null. The output layer returns the values for the s and t function. We utilize a CNN since we have a configuration with local interactions. A detailed review of this kind of architectures can be found at [40].	85

[This page was intentionally left blank]

1. Interface free energy - from strings to Ising 3D

Before we can understand the non-universal corrections of the EST picture, we need first to compute the Nambu-Goto contributions, in order to be able to recognize deviations.

In this chapter we will start by studying the Nambu-Goto action, calculating its energy spectrum (Sec. 1.2, which follows the analysis present at [33]), that leads naturally to the partition function (Sec. 1.4). We finish by computing the expansion for long strings (Sec. 1.5). This is by no means a thorough analysis, and it is presented in this thesis for the sake of completion; a more detailed account can be found at [33].

1.1. Nambu-Goto action / capillary wave model

Nambu-Goto is the simplest operator compatible with the constraints of the effective string theory of a long string[9]. Although widely used to study quantum gravity (string theory) we are more interested in its role as a generalization of the point particle action to a string. Then, let us review the action of a point particle, which we can write as

$$S = -m \int ds, \quad (1.1)$$

with m as the rest mass. The integral corresponds to the length of the particle path in the space-time, an one dimensional object.

We now need to generalize this notion of length of space-time to two dimensions, so it is to be expected that we are going to promote the length (volume of a 1D object) to the area (volume of a 2D object). It is natural then to write, if we consider h as the metric induced on the worldsheet,

$$S_{NG} = \sigma \int d^2\xi \sqrt{\det h}, \quad (1.2)$$

with σ as the string tension and $\int d^2\xi \sqrt{\det h}$ as the worldsheet's area.

The worldsheet, a 2-dimensional manifold Σ , can be parametrized by $\{\xi^0, \xi^1\}$, and is embedded in a d -dimensional target space, with a metric g_{ij} and parametrized by coordinates $\{X^0, \dots, X^d\}$.

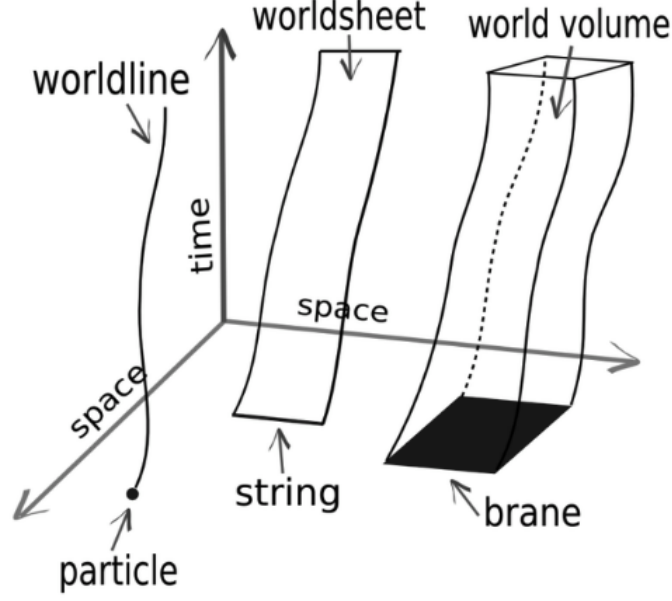


Figure 1.1.: Example of the promotion of the particle's worldline to a string's worldsheet. Even further generalizations to higher dimensions are possible if we introduce the concept of a brane. Image from[38].

The induced metric onto the worldsheet is

$$h_{ab} = \frac{\partial X^i}{\partial \xi^a} \frac{\partial X^j}{\partial \xi^b} g_{ij}, \quad (1.3)$$

and is its determinant that appears in eq. (1.2). These quantities are diffeomorphism invariant, i.e., are invariant under arbitrary differentiable coordinate transformations on the worldsheet. This induces a gauge-like freedom in the parametrization of our target space, that can be circumvented by choosing the so-called static gauge,

$$\begin{aligned} X^0 &= \xi^0 \\ X^1 &= \xi^1 \end{aligned} \quad (1.4)$$

such that the target space is parametrized by

$$X^\mu = (\xi^0, \xi^1, \vec{X}(\xi^0, \xi^1)), \quad (1.5)$$

with $\vec{X}$ being a $d - 2$ dimensional vector. With this, we can explicitly obtain the induced metric h for an euclidean background. The computation for h_{00} yields

$$h_{00} = \frac{\partial X^i}{\partial \xi^0} \frac{\partial X^j}{\partial \xi^0} g_{ij} = \frac{\partial X^0}{\partial \xi^0} \frac{\partial X^0}{\partial \xi^0} g_{00} + \frac{\partial \vec{X}^i(\xi^0, \xi^1)}{\partial \xi^0} \frac{\partial \vec{X}^i(\xi^0, \xi^1)}{\partial \xi^0} g_{ii} = 1 + (\partial_0 \vec{X}(\xi^0, \xi^1))^2. \quad (1.6)$$

The computation for h_{11} is very similar, and

$$h_{11} = 1 + (\partial_1 \vec{X}(\xi^0, \xi^1))^2. \quad (1.7)$$

Finally, the off-diagonal terms are

$$h_{10} = h_{01} = \frac{\partial X^i}{\partial \xi^0} \frac{\partial X^j}{\partial \xi^1} g_{ij} = \frac{\partial \vec{X}^i(\xi^0, \xi^1)}{\partial \xi^0} \frac{\partial \vec{X}^i(\xi^0, \xi^1)}{\partial \xi^1} g_{ii} = \partial_0 \vec{X}(\xi^0, \xi^1) \partial_1 \vec{X}(\xi^0, \xi^1), \quad (1.8)$$

which allows us to compute the determinant

$$\det h = (1 + (\partial_0 \vec{X}(\xi^0, \xi^1))^2)(1 + (\partial_1 \vec{X}(\xi^0, \xi^1))^2) - (\partial_0 \vec{X}(\xi^0, \xi^1) \partial_1 \vec{X}(\xi^0, \xi^1))^2, \quad (1.9)$$

and re-express the action in its most usual form,

$$S_{NG} = \sigma \int_{\Sigma} d\xi^0 d\xi^1 \left(1 + (\partial_0 \vec{X})^2 + (\partial_1 \vec{X})^2 + \partial_0 \vec{X} \cdot \partial_0 \vec{X} \partial_1 \vec{X} \cdot \partial_1 \vec{X} - (\partial_0 \vec{X} \cdot \partial_1 \vec{X})^2 \right)^{1/2}. \quad (1.10)$$

We remark that this action has the following symmetries:

1. Isometry of flat spacetime (d -dimensional Poincaré group)

$$X'^{\mu}(\xi^0, \xi^1) = \underbrace{\Lambda_{\nu}^{\mu}}_{\text{Lorentz transformation}} X^{\nu}(\xi^0, \xi^1) + \underbrace{a^{\mu}}_{\text{translations}}; \quad (1.11)$$

2. Diffeomorphism invariance on the worldsheet, that is, performing a coordinate transformation $\xi^0, \xi^1 \rightarrow \xi'^0(\xi^0, \xi^1), \xi'^1(\xi^0, \xi^1)$, we have

$$X'^{\mu}(\xi'^0, \xi'^1) = X^{\mu}(\xi^0, \xi^1). \quad (1.12)$$

1.2. Polyakov Lagrangian

Computations with the explicit NG-action are quite cumbersome, which suggests the introduction of a new action. This is done by putting forward a new world-sheet metric γ_{ab} that will act as a background field. This is the Polyakov action[41], expressed by

$$S_P[X, \gamma] = \frac{\sigma}{2} \int_{\Sigma} d\xi^0 d\xi^1 \sqrt{-\det \gamma} \gamma^{ab} \partial_a X^\mu \partial_b X_\mu. \quad (1.13)$$

This action is equivalent to the Nambu-Goto action. That can be seen by computing γ 's equation of motion and plugging them back into eq. (1.13). The detailed computations are present at Appendix C.

This action has the following symmetries:

1. Isometry of flat spacetime (d -dimensional Poincaré group)

$$X'^\mu(\xi^0, \xi^1) = \underbrace{\Lambda^\mu_\nu}_{\text{Lorentz transformation}} X^\nu(\xi^0, \xi^1) + \underbrace{a^\mu}_{\text{translations}}; \quad (1.14)$$

2. Diffeomorphism invariance on the worldsheet. Performing a coordinate transformation $\xi^0, \xi^1 \rightarrow \xi'^0(\xi^0, \xi^1), \xi'^1(\xi^0, \xi^1)$, we have

$$X'^\mu(\xi'^0, \xi'^1) = X^\mu(\xi^0, \xi^1); \quad (1.15)$$

3. Weyl invariance, i.e., invariance under a local rescaling of the world-sheet metric

$$\begin{aligned} X'^\mu(\xi^0, \xi^1) &\rightarrow X^\mu(\xi^0, \xi^1) \\ \gamma'_{ab}(\xi^0, \xi^1) &\rightarrow \exp(2\omega(\xi^0, \xi^1)) \gamma_{ab}(\xi^0, \xi^1) \end{aligned} \quad (1.16)$$

for any arbitrary $\omega(\xi^0, \xi^1)$.

Actions 1.2 and 1.13 define 2D conformal field theories on the string world-sheet Σ . The Polyakov action, for example, describes a massless scalar field X^μ coupled to the metric γ_{ab} , since its motion equation is

$$\sqrt{-\det \gamma} \partial_a \partial^a X^\mu = 0. \quad (1.17)$$

The considerations stated above presuppose any generic Σ , but since we are interested in a finite interface, we will have to take into account the behavior induced by the presence of boundary terms. Let's start by taking the coordinate region of our submanifold to be

$$0 < \xi^1 < l, \quad (1.18)$$

with this choice arising by considering ξ^1 as a “spacial” coordinate and ξ^0 as a “time” one. This describes a string propagating in time without external sources. If one is to variate again the action with respect to X^μ :

$$\delta S_P = 2\sigma \int_{-\infty}^{\infty} d\xi^0 \int_0^l d\xi^1 \sqrt{-\det \gamma} \gamma^{ab} \partial_a \delta X^\mu \partial_b X_\mu,$$

and, integrating by parts,

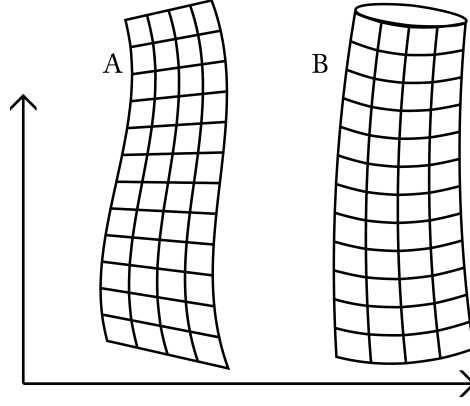


Figure 1.2.: Example of an open (A) and closed (B) string world-sheet

$$\begin{aligned} \delta S_P = & -2\sigma \int_{-\infty}^{\infty} d\xi^0 \int_0^l d\xi^1 \sqrt{-\det \gamma} \partial_a \partial^a X^\mu \delta X^\mu \\ & + 2\sigma \int_{-\infty}^{\infty} d\xi^0 \sqrt{-\det \gamma} \delta X^\mu \partial^a X^\mu \Big|_{\xi^1=0}^{\xi^1=l}, \end{aligned} \quad (1.19)$$

we may have boundary terms, that are zero if

$$\partial^a X^\mu(\xi^0, 0) = \partial^a X^\mu(\xi^1, l) = 0, \quad (1.20)$$

corresponding to vanishing derivatives of the scalar field at the boundary. The reader may recognize them as von Neumann boundary conditions that describe an open-ended string. We also have a vanishing boundary term if the string is closed, i.e., if

$$\begin{aligned} X^\mu(\tau, 0) &= X^\mu(\tau, l) \\ \gamma^{ab}(\tau, 0) &= \gamma^{ab}(\tau, l) \end{aligned} \quad (1.21)$$

Let's consider this last choice and compute its energy spectrum, since, as we will see, those will be the realizations we will realize on the lattice. Firstly, however, we need to deal with the gauge invariance of the Polyakov action.

The Polyakov action has a diffeomorphism and a Weyl symmetry. In order to compute the spectrum, we will employ an approach that starts by fixing the gauge to the so-called light cone gauge[42]. Although successful, it must be said that there are more generic, symmetry-driven ways [33] to obtain this result. Such methods will not be discussed in this thesis.

We start once again by considering the d -dimensional target space, parametrized by the coordinates $\{X^0, \dots, X^d\}$. The light-cone coordinates are

$$X^\pm = \frac{X^0 \pm X^1}{\sqrt{2}}, \quad X^i, \quad (1.22)$$

which allows the gauge to be fixed by imposing

$$\begin{aligned}\xi^0 &= X^+ \\ \partial_{\xi^1} \gamma_{\xi^1 \xi^1} &= 0 \\ \det \gamma &= -1\end{aligned}\tag{1.23}$$

with the first two conditions fixing a coordinate and a parametrization, respectively, in order to deal with the diffeomorphism invariance, and the latter fixing a scale to our string to deal with the Weyl symmetry.

So what we need to do now is to construct a set of coordinates and a γ that satisfies the required conditions. Since the metric is symmetric, we have three degrees of freedom ($\gamma_{\xi^0 \xi^0}$, $\gamma_{\xi^0 \xi^1}$ and $\gamma_{\xi^1 \xi^1}$). The determinant condition fixes one of those. By the second condition at 1.23, $\gamma_{\xi^1 \xi^1}$ only depend of ξ^0 . There are only two degrees of freedom $\gamma_{\xi^1 \xi^1}(\xi^0)$ and $\gamma_{\xi^0 \xi^1}(\xi^0, \xi^1)$, and will be in our interest to write the Polyakov action with respect to these quantities. Let's start by obtaining the inverse metric γ^{ab} with respect to the metric γ_{ab} ,

$$\begin{pmatrix} -\gamma_{\xi^1 \xi^1}(\xi^0) & \gamma_{\xi^0 \xi^1}(\xi^0, \xi^1) \\ \gamma_{\xi^0 \xi^1}(\xi^0, \xi^1) & \gamma_{\xi^1 \xi^1}^{-1}(\xi^0)(1 - \gamma_{\xi^0 \xi^1}^2(\xi^0, \xi^1)) \end{pmatrix},\tag{1.24}$$

which can be readily seen to obey to the condition

$$\det \gamma = -\gamma_{\xi^1 \xi^1}(\xi^0) \gamma_{\xi^1 \xi^1}^{-1}(\xi^0)(1 - \gamma_{\xi^0 \xi^1}^2(\xi^0, \xi^1)) - \gamma_{\xi^0 \xi^1}^2(\xi^0, \xi^1) = -1.\tag{1.25}$$

We will now manipulate the Polyakov Lagrangian until arriving at a simpler expression. Let's start by considering

$$\begin{aligned}S &= \int d\xi^0 \underbrace{\frac{\sigma}{2} \int d\xi^1 \sqrt{-\det \gamma} \gamma^{ab} \partial_a X^\mu \partial_b X_\mu}_L \\ L &= \frac{\sigma}{2} \int_0^l d\xi^1 \overbrace{\sqrt{-\det \gamma}}^1 (A + B),\end{aligned}$$

recalling that

$$\begin{aligned}\partial_{\xi^0} X^\mu &= (1, \partial_{\xi^0} X^-, \partial_{\xi^0} X^i) \\ \partial_{\xi^1} X^\mu &= (0, \partial_{\xi^1} X^-, \partial_{\xi^1} X^i)\end{aligned},\tag{1.26}$$

and

$$\begin{aligned}\partial_{\xi^0} X^\mu \partial_{\xi^0} X^\mu &= -2\partial_{\xi^0} X^- + \partial_{\xi^0} X^i \partial_{\xi^0} X^i \\ \partial_{\xi^1} X^\mu \partial_{\xi^1} X^\mu &= \partial_{\xi^1} X^i \partial_{\xi^1} X^i \\ \partial_{\xi^0} X^\mu \partial_{\xi^1} X^\mu &= -\partial_{\xi^1} X^- + \partial_{\xi^1} X^i \partial_{\xi^0} X^i\end{aligned}\tag{1.27}$$

This allows us to obtain

$$\begin{aligned} A &= -\gamma_{\xi^1 \xi^1} \partial_{\xi^1} X^i \partial_{\xi^1} X^i + 2\gamma_{\xi^0 \xi^1} \partial_{\xi^1} X^i \partial_{\xi^0} X^i + \gamma_{\xi^1 \xi^1}^{-1} (1 - \gamma_{\xi^0 \xi^1}^2) \partial_{\xi^0} X^i \partial_{\xi^0} X^i \\ B &= 2\gamma_{\xi^1 \xi^1} \partial_{\xi^0} X^- - 2\gamma_{\xi^0 \xi^1} \partial_{\xi^1} X^- \end{aligned} \quad (1.28)$$

$X^-(\xi^0, \xi^1)$ can be further broken into

$$X^-(\xi^0, \xi^1) = \underbrace{x^-(\xi^0)}_{\text{mean value of } X^-} + Y^-(\xi^0, \xi^1), \quad (1.29)$$

which promotes

$$\begin{aligned} 2\gamma_{\xi^0 \xi^1} \partial_{\xi^1} X^i &\rightarrow 2\gamma_{\xi^0 \xi^1} \partial_{\xi^1} Y^i \\ \gamma_{\xi^1 \xi^1} \partial_{\xi^0} X^-(\xi^0, \xi^1) &\rightarrow \gamma_{\xi^1 \xi^1} \partial_{\xi^0} x^-(\xi^0) + \gamma_{\xi^1 \xi^1} \partial_{\xi^0} Y^-(\xi^0, \xi^1) \end{aligned} \quad (1.30)$$

Since we know that $\gamma_{\xi^1 \xi^1}$ is independent of ξ^0 , by the condition that fixed the gauge, eq. (1.23),

$$\int_0^l d\xi^1 \gamma_{\xi^1 \xi^1} \partial_{\xi^0} Y^-(\xi^0, \xi^1) = \gamma_{\xi^1 \xi^1} \partial_{\xi^0} \int_0^l d\xi^1 Y^- = 0, \quad (1.31)$$

and we finally arrive at the Lagrangian

$$\begin{aligned} L = \sigma l \gamma_{\xi^1 \xi^1} \partial_{\xi^0} x^- + \frac{\sigma}{2} \int_0^l d\xi^1 &\left(-\gamma_{\xi^1 \xi^1} \partial_{\xi^1} X^i \partial_{\xi^1} X^i + 2\gamma_{\xi^0 \xi^1} \partial_{\xi^1} X^i \partial_{\xi^0} X^i \right. \\ &\left. + \gamma_{\xi^1 \xi^1}^{-1} (1 - \gamma_{\xi^0 \xi^1}^2) \partial_{\xi^0} X^i \partial_{\xi^0} X^i - 2\gamma_{\xi^0 \xi^1} \partial_{\xi^1} Y^i \right) \end{aligned} \quad (1.32)$$

In the closed string, we have yet an extra freedom, since the $\xi^1 = 0$ point is not uniquely determined. Most of the remaining freedom can be fixed by imposing

$$\gamma_{\xi^0 \xi^1}(\xi^1, 0) = 0, \quad (1.33)$$

and by taking Eq. 1.33 into account, the Polyakov Lagrangian for a closed string simplifies to

$$L = \sigma l \gamma_{\xi^1 \xi^1} \partial_{\xi^0} x^- + \frac{\sigma}{2} \int_0^l d\xi^1 \left(-\gamma_{\xi^1 \xi^1} \partial_{\xi^1} X^i \partial_{\xi^1} X^i + \gamma_{\xi^1 \xi^1}^{-1} \partial_{\xi^0} X^i \partial_{\xi^0} X^i \right). \quad (1.34)$$

This simplified Lagrangian finally allows us to easily obtain an equation of motion for the field $X^i(\xi^0, \xi^1)$, which reads

$$\partial_{\xi^0}^2 X^i = \underbrace{\left(\frac{\sigma l}{p^+} \right)^2}_c \partial_{\xi^1}^2 X^i, \quad (1.35)$$

and may be recognized as the wave equation. For a closed string, the solution is

$$X^i(\xi^0, \xi^1) = \underbrace{x^i + \frac{p^i}{p^+} \xi^0}_{\text{center of mass dynamics}} + \frac{i}{\sqrt{4\pi\sigma}} \sum_{n=-\infty}^{\infty} \underbrace{\left\{ \frac{\alpha_n^i}{n} \exp \left[-\frac{2\pi i n (\xi^1 - c\xi^0)}{l} \right] \right\}}_{\equiv \text{left moving excitations}} \quad (1.36)$$

$$+ \underbrace{\left\{ \frac{\tilde{\alpha}_n^i}{n} \exp \left[\frac{2\pi i n (\xi^1 + c\xi^0)}{l} \right] \right\}}_{\equiv \text{right moving excitations}}. \quad (1.37)$$

1.3. Energy spectrum of a closed string

To compute the energy spectrum, we need to quantize the Lagrangian, in a similar vein to what is done usually in Quantum Field Theories. We quantize the independent degrees of freedom (transverse oscillators (α_n and $\tilde{\alpha}_n$) and transverse (x^i, p^i) and longitudinal (x^-, p^+) centre of mass variables):

$$\begin{aligned} [x^-, p^+] &= -i \\ [x^i, p^j] &= i\delta^{ij} \\ [\alpha_m^i, \alpha_n^j] &= m\delta^{ij}\delta_{m,-n} \\ [\tilde{\alpha}_m^i, \tilde{\alpha}_n^j] &= m\delta^{ij}\delta_{m,-n} \end{aligned}$$

This is very similar to the algebra of the harmonic oscillator if we define operators $\tilde{\alpha}_m^i = \sqrt{m}a_n^i$, and by choosing the quantum numbers k as the center of mass momentum and $N(\tilde{N})$ as the level of the right(left) excitation, we can build a vacuum state $|0, 0, k\rangle$ and a general state $|N, \tilde{N}, k\rangle$. This analogy leads us to the mass formula,

$$m^2 = 4\pi\sigma \left(N + \tilde{N} - \frac{D-2}{12} \right), \quad (1.38)$$

and the momentum

$$p = \frac{2\pi}{l}(N - \tilde{N}), \quad (1.39)$$

and by taking into account the centre of mass momentum k , we arrive at the energy spectrum for the Nambu-Goto action,

$$\mathcal{E}^2(N, \tilde{N}, k) = \overbrace{4\pi\sigma \left(N + \tilde{N} - \frac{D-2}{12} \right)}^{m^2} + \overbrace{\left[\frac{2\pi}{l}(N - \tilde{N}) \right]^2}^{p^2} + k^2. \quad (1.40)$$

1.4. Partition function

Now that we have the energy spectrum of the Nambu-Goto action for the closed string, we will realize the physical system in some lattice model. String theory is, in its essence, a theory of one spatial dimensional objects that respect the symmetries presented in section 1.2, and such symmetries are broken when we discretize the theory, but can be restored if our lattice system is undergoing a second order phase transition.

The simplest choice for studying the dynamics of a 2D object on a lattice model at a critical point is the 3D Ising. This model has a well-known second order phase transition and it is easy to simulate. In this model there is a clear analogue to the worldsheet - a domain wall/interface. These appear in the broken symmetry phase, splitting domains of up and down aligned spins. A detailed study of these objects will be presented in chapter 3, but suffices here to say that we can generate domain walls by imposing anti-periodic boundary conditions along the z -direction of the simulated lattice.

The behavior of the domain wall can be properly modeled by the capillary wave model, that postulates that the energy of this object is proportional to its area. Given the similarities with NG, the working hypothesis is that the dynamics of the domain wall near the critical point will be governed by the Nambu-Goto action, up to higher order corrections[14]. There is, however, an additional term we need to add to eq. (1.40) : the minimal area of the surface. Due to the periodic boundary conditions alongside x and y -directions, the domain wall cannot be contracted into a point, and has a minimum area, that corresponds to the size of the transversal section, i.e., $L_x \times L_y$. By making the identification of the x -direction with the space dimension and the y -direction with the time dimension, this is equivalent of having a string that has a minimal length L_x , and such contribution must be added to the mass term, leading to the expression

$$\mathcal{E}(N, \tilde{N}, k) = \sigma L_x \sqrt{1 + \frac{4\pi\sigma}{\sigma^2 L_x^2} (N + \tilde{N} - \frac{D-2}{12}) + \frac{1}{\sigma^2 L_x^2} \left[\frac{2\pi}{L_x} (N - \tilde{N}) \right]^2 + \left(\frac{k}{\sigma L_x} \right)^2}. \quad (1.41)$$

We are now able to compute, by summing over all the possible N , $\tilde{N}$ and k ,

$$Z = \text{Tr}_{N, \tilde{N}, k} \left[\exp(-\underbrace{T}_{=L_y} \mathcal{E}(N, \tilde{N}, k)) \right], \quad (1.42)$$

having identified y as the time direction.

Let's recall that, by the construction of the states on the closed string, we need to sum over the degrees of freedom of the left and right excitations (N and $\tilde{N}$) and the center of mass longitudinal momentum (k), so

$$Z = \sum_{N, \tilde{N}, k} c_N c_{\tilde{N}} \left\langle N, \tilde{N}, k \left| e^{-L_y \mathcal{E}} \right| N, \tilde{N}, k \right\rangle, \quad (1.43)$$

with c_N and $c_{\tilde{N}}$ as the degeneracies of each energy level. The longitudinal direction is finite ($\equiv L_z$) and has periodic boundary conditions. This imposes a discretization for the center of mass momentum

$$k = \frac{2\pi}{L_z} n \quad n \in \mathbb{Z}. \quad (1.44)$$

We will compute the partition function in the large L_z limit, so $\sum_k \rightarrow \frac{L_z}{2\pi} \int dk$,

$$Z = \frac{L_z}{2\pi} \sum_{N, \tilde{N}} c_N c_{\tilde{N}} \int_{-\infty}^{+\infty} dk \left\langle N, \tilde{N}, k \left| e^{-L_y \mathcal{E}} \right| N, \tilde{N}, k \right\rangle. \quad (1.45)$$

To integrate over k , we perform two successive changes of variable. First, by defining $z = \frac{k}{\sigma L_x M_{N, \tilde{N}}}$, $M_{N, \tilde{N}}^2 = 1 + \frac{4\pi}{\sigma L_x^2} \left(N + \tilde{N} - \frac{D-2}{12} \right) + \left[\frac{2\pi(N-\tilde{N})}{\sigma L_x^2} \right]^2$ and $A = L_x L_y$,

$$Z = \frac{L_z}{2\pi} \sum_{N, \tilde{N}} c_N c_{\tilde{N}} \sigma L_x M_{N, \tilde{N}} \int_{-\infty}^{\infty} dz e^{-\sigma A M_{N, \tilde{N}} \sqrt{1+z^2}}, \quad (1.46)$$

and then, stipulating $\cosh(\beta) = \sqrt{1+z^2}$, we arrive at

$$\sigma L_x M_{N, \tilde{N}} \int_{-\infty}^{\infty} dz e^{-\sigma A M_{N, \tilde{N}} \sqrt{1+z^2}} = 2\sigma L_x M_{N, \tilde{N}} \int_0^{\infty} d\beta e^{-\sigma A M_{N, \tilde{N}} \cosh(\beta)} \cosh(\beta)$$

and this integral corresponds to the Bessel function of the second kind,

$$= 2\sigma L_x M_{N, \tilde{N}} K_1 \left(\sigma A M_{N, \tilde{N}} \right). \quad (1.47)$$

This gives rise to a partition function

$$Z = \frac{\sigma L_z}{\pi} \sum_{N, \tilde{N}} L_x M_{N, \tilde{N}} K_1 \left(\sigma A M_{N, \tilde{N}} \right) \quad (1.48)$$

1.5. $1/\mathcal{A}$ expansion for finite strings

For big transversal areas $A = L_x L_y$, the fact that the energy is proportional to the area will mean that $\mathcal{E}$ should go proportionally with A , with a proportionality constant σ that is given by the particular details of the theory. However, there are corrections that can appear, since the wall is not a perfectly flat slab. This corrections should grow slower than the leading σA term, given that in the limit $A \rightarrow \infty$ the wall is perfectly flat, meaning that they appear as an expansion parameter with $\frac{1}{\sigma A} \equiv \frac{1}{\mathcal{A}}$ as the expansion parameter. The partition function, eq.

(1.48) can be expanded with respect to this parameter. Performing the computations[43, 15] it is obtained, for a generic ratio $u \equiv \frac{L_y}{L_x}$,

$$Z = \left(\frac{\sigma L_z^2}{2\pi u} \right)^{\frac{1}{2}} \frac{e^{-\mathcal{A}}}{\tilde{\eta}^2} \left(1 + \sum_{n=1}^{\infty} \frac{f_n(u)}{\mathcal{A}^n} \right), \quad (1.49)$$

with

$$\tilde{\eta} = \left| \frac{\eta(iu)}{\eta(u)} \right|, \quad (1.50)$$

if we consider η as the usual Dedekind's function

$$\eta(\tau) = e^{\frac{\pi i \tau}{12}} \prod_{n=1}^{\infty} (1 - e^{2n\pi i \tau}). \quad (1.51)$$

Expanding 1.49 up to $\frac{1}{\mathcal{A}^4}$ for a square interface ($u = 1$), we arrive at the expression

$$F = 0.3914 - \frac{1}{2} \log \left(\frac{\sigma}{u} L_z^2 \right) + \mathcal{A} - \frac{0.2500}{\mathcal{A}} + \frac{0.0141}{\mathcal{A}^2} + \frac{0.1314}{\mathcal{A}^3} - \frac{2.0303}{\mathcal{A}^4}, \quad (1.52)$$

which will be the function we will test our data against. Any deviations from this behavior will constitute deviations from the pure NG behavior. Such deviations are expected to only appear at order $\frac{1}{\mathcal{A}^3}$ or greater[10]. The main goal throughout this work will be measuring this deviations, because the first detour from the universal series, that appear at order $\frac{1}{\mathcal{A}^3}$, has interesting physical properties[35] and, despite previous numerical studies[36, 37], its precise value still evades physicists to this day.

2. Monte Carlo

The Monte Carlo methods[44] in statistical physics are used to accurately predict the thermodynamic properties of complex models. Starting from a microscopic description of the desired physical systems, Hamiltonian and lattice structure, one can in principle compute any average.

This method is necessary because, for any reasonably sized system, it is simply impossible to sum over the configuration space exactly. As an example, take a system with a d-dimensional cubic lattice as an example. Recall that the partition function can be written as a sum over microstates:

$$Z = \sum_{\mu} \exp(-\beta H(\mu)), \quad (2.1)$$

with $\beta = \frac{1}{k_B T}$ and H as the model Hamiltonian. In the Ising, each lattice site corresponds to a degree of freedom with 2 distinct possible values and, if our cubic lattice has linear dimension L , the number of possible configurations is 2^{L^d} . For our typical lattice size ($L \sim 96$), we are dealing with $2^{96^3} \sim 10^{266332}$ configurations. This is an absolutely gigantic number, and constitutes a significant impediment in the computation of the partition function.

In this chapter we will explain the general features required to build a generic Markov Chain Monte Carlo (MCMC) algorithm, Sec.2.1. Then, we will discuss two prominent approaches to the sampling of Ising models - the local, Sec.2.2, and cluster algorithms, Sec.2.3. We will delve a little bit in the most used algorithms in our work.

2.1. General properties

Monte Carlo encompass a large subset of algorithms, some of which seem as distinct as day and night. However, all of them need to fulfil some core features, which we outline in this section.

2.1.1. Importance sampling

The underlying principle behind the Monte Carlo approach is the importance sampling. The concept can be illustrated with a simple example. Let's consider the integral:

$$I = \int_a^b f(x)dx, \quad (2.2)$$

which can be expressed as a mean value of $f(x)$,

$$\langle f(x) \rangle = (b-a) \int_{-\infty}^{+\infty} f(x)w(x)dx \quad (2.3)$$

if we consider $w(x)$ to be the uniform distribution defined between a and b . Since Eq. 2.3 holds, one can produce a sample of values x_i from the uniform distribution and compute the mean estimator as an approximation for the mean value of $f(x)$:

$$\frac{1}{M} \sum_{i=1}^M f(x_i) \rightarrow \langle f(x) \rangle = I. \quad (2.4)$$

Notwithstanding, let's consider that the integral we want to compute is

$$I = \int_0^L \exp(-x^4)dx,$$

with L as an arbitrary cut-off.

We can implement the same procedure as detailed above. However, the convergence rate will be rather small, since the function decays rather sharply near the origin, and we are sampling numbers that do not have a significant contribution to the integral. Still, it is important to note that:

$$I = \int_0^L \exp(-x^4) \frac{w(x)}{w(x)} dx = \left\langle \frac{\exp(-x^4)}{w(x)} \right\rangle_{w(x)}, \quad (2.5)$$

so we can utilize different probability distributions to rate the integral. In particular, the closer that probability distribution is to function whose integral we want to compute, the more success we can expect to get in estimation.

2.1.2. Markov process

Usually, to perform physical observable computations, that in Ising model are quantities like mean magnetization, heat capacity, we need to produce a random set of configurations. Such configurations, or microstates ϕ , need to follow the distribution

$$P(\phi) \propto \exp(-\beta H(\phi)). \quad (2.6)$$

This requirement gives rise to the inadequacy of producing uniformly random states, i.e., we can't go to each lattice site and flip a coin and assign spin up or down given the coin result, since that begets the same kind of problems that we saw in subsection 2.1.1 - most of the

configurations sampled with this rule will have a very small P , and will be irrelevant to the observables calculations.

Instead, we rely on the Markov process to generate configurations. This process consists in assigning a transition probability $P(\phi_\mu \rightarrow \phi_\nu)$ of generating a new state ϕ_ν , given a previous state ϕ_μ . We choose the Markov process that has the desired 2.6 asymptotic distribution.

With the previous definition, it is possible to build a Markov Matrix $\mathbb{M}$, that acts in the configuration space of our system, with the matrix element being transition probabilities, i.e.,

$$\mathbb{M}_{\mu\nu} = P(\phi_\mu \rightarrow \phi_\nu). \quad (2.7)$$

This matrix contains all the information related to each specific Markov process: Metropolis, Wolff, Swendsen-Wang, etc. Each distinct process has its specific rules to build the matrix, but all of them must obey the conditions that we will expose at 2.1.3 and 2.1.4.

Since we are dealing with probabilities,

$$\mathbb{M}_{\mu\nu} = P(\phi_\mu \rightarrow \phi_\nu) \in [0, 1] \quad (2.8)$$

and

$$\sum_\nu \mathbb{M}_{\mu\nu} = 1. \quad (2.9)$$

This last condition is required, both as a simple property of conditional probabilities, and as a general statement of probability conservation.

2.1.3. Ergodicity

The ergodicity condition forces our Markov process to allow any given configuration ϕ_μ to reach any other configuration ϕ_ν (i.e., to have a finite transition probability), within a finite number of Markov steps. This affirmation can be summed up as:

$$\forall \mu, \nu, \exists n : [\mathbb{M}^n]_{\mu\nu} > 0. \quad (2.10)$$

Since we want to produce a series of microstates that follow the Boltzmann distribution, equation 2.6, each configuration has a non-zero probability. Hence, we can't have two states that are inaccessible through a Markov process, because if we start our Markov chain at one of them, the probability associated with the inaccessible one would be null. This argument also holds for any probability distribution with non-zero probability for any microstate.

2.1.4. Detailed balance

We want our Markov process to produce, in the asymptotic limit, a set of configurations that follow the Boltzmann (or any other) probability distribution¹.

If the system is in the equilibrium distribution, then the probability flow in and out of a given configuration μ must be the same, since P_μ (stationary probability of the configuration ϕ_μ) must not change. That can be expressed as

$$\underbrace{\sum_{\nu} P_\mu \mathbb{M}_{\mu\nu}}_{\text{outwards flow}} = \underbrace{\sum_{\nu} P_\nu \mathbb{M}_{\nu\mu}}_{\text{inwards flow}}, \quad (2.11)$$

which simplifies, through equation 2.9, to

$$P_\mu = \sum_{\nu} P_\nu \mathbb{M}_{\nu\mu}. \quad (2.12)$$

This condition is enough to ensure that there is one stationary probability distribution P , that obeys to

$$\mathbb{M}P = P, \quad (2.13)$$

i.e., a eigenstate of the Markov Matrix with eigenvalue 1. This eigenvalue analysis allows us to further understand the behavior of the Markov process. If we consider the right² eigenvalues r of $\mathbb{M}$,

$$M_{ij} r_j^{(\alpha)} = \lambda_\alpha r_i^{(\alpha)}, \quad (2.14)$$

we build a basis $\{r^{(\alpha)}\}$ that can be used to decompose any initial probability distribution q ,

$$q_i = \sum_{\alpha} a_{\alpha} r_i^{(\alpha)} \quad (2.15)$$

so that the application of $\mathbb{M}$ n times is

$$(\mathbb{M}^n q)_i = \sum_{\alpha} a_{\alpha} \lambda_{\alpha}^n r_i^{(\alpha)}. \quad (2.16)$$

It can be proven (see Appendix A) that $|\lambda_{\alpha}| \leq 1$, so the only significative contribution for $n \rightarrow \infty$ corresponds to the asymptotic distribution, since it corresponds to the $\lambda = 1$ eigenvalue.

However, there are other solutions that obey to $|\lambda_{\alpha}| = 1$ - complex eigenvalues

$$\lambda_{\alpha} = \exp(i\theta_{\alpha}),$$

¹We can denominate this distribution as asymptotic, stationary, equilibrium or canonical

²There are right and left eigenvectors, since the Markov Matrix is not symmetric

that survive in the $n \rightarrow \infty$ limit. The corresponding eigenvectors exhibit dynamic equilibrium, characterized by probability distributions that obey

$$\mathbb{M}^n W = W, n \neq 1. \quad (2.17)$$

We now present a simple example of a process that exhibits dynamic equilibrium and show how to deal with this issue.

N state Markov process with dynamical equilibrium

Let's consider a system with n possible configuration ϕ_i , with a Markov process characterized by:

$$\begin{cases} P(\phi_i \rightarrow \phi_{i+1}) = 1 \\ P(\phi_i \rightarrow \phi_j) = 0 \quad \text{otherwise} \end{cases}, \quad (2.18)$$

which can be represented through the Markov matrix:

$$M_{ij} = \delta_{i,j-1}. \quad (2.19)$$

This Markov matrix satisfies 2.11, and the stationary configuration is a uniform distribution $\mathbb{U}$ over the configurations, since

$$(\mathbb{M}\mathbb{U})_i = \sum_j M_{ij} U_j = \frac{1}{n} \sum_j \delta_{i,j-1} = \frac{1}{n} = \mathbb{U}_i. \quad (2.20)$$

However, we can see that if we have a state with $p(i) = \delta_{i,n}$, the state will cycle through the states and come back to i after n iteration, i.e.,

$$\mathbb{M}^n p = p. \quad (2.21)$$

In order to get rid of the cycles, we need to ensure that each line of the Markov matrix in the large n limit, $\mathbb{M}^n$, corresponds to Π , with Π as the asymptotic distribution, which is equivalent to have $|\lambda_\alpha| < 1$ for all but one eigenvalue. In order to respect this limitations, we need to promote 2.11 to the detailed balance condition

$$P_\mu \mathbb{M}_{\mu\nu} = P_\nu \mathbb{M}_{\nu\mu}, \quad (2.22)$$

which is a statement of microscopic reversibility. This condition destroys the cyclical behavior since we are forcing that, in a single step, the probability transfered from $\mu \rightarrow \nu$ to be the same as the one transfered from $\nu \rightarrow \mu$.

2.2. Metropolis-Hastings algorithm

The Metropolis algorithm was first introduced by Nicolas Metropolis[45], to simulate hard sphere gas dynamics. The mathematical properties were later formalized by Hastings[46]. Nowadays, the Metropolis algorithm (MA) is the most common algorithm used to sample Ising configurations.

The main idea behind this algorithm is the choice of writing the transition probability $T(\mu \rightarrow \nu)$ as a product of probabilities

$$A(\mu \rightarrow \nu)S(\mu \rightarrow \nu), \quad (2.23)$$

with S as the suggestion probability and A as the acceptance probability. As these names suggest, we firstly propose a new configuration and then accept or reject it with a different probability distribution. The probabilities must be chosen to obey the detailed balance,

$$P(\mu)A(\mu \rightarrow \nu)S(\mu \rightarrow \nu) = P(\nu)A(\nu \rightarrow \mu)S(\nu \rightarrow \mu). \quad (2.24)$$

The particular choice of the Metropolis algorithm is to define $S(\mu \rightarrow \nu)$ as choosing and proposing the flip of a random spin in the lattice. The choice is uniform, for a lattice with linear size L , $S(\mu \rightarrow \nu) = \frac{1}{L^d}$.

This choice then fixes

$$e^{-\beta E_\mu} A(\mu \rightarrow \nu) = e^{-\beta E_\nu} A(\nu \rightarrow \mu)$$

$$\frac{A(\mu \rightarrow \nu)}{A(\nu \rightarrow \mu)} = e^{-\beta(E_\nu - E_\mu)}. \quad (2.25)$$

The optimal choice for the acceptance probability

$$A(\mu \rightarrow \nu) = \min(1, e^{-\beta \Delta E}). \quad (2.26)$$

We can easily see that this gives

$$\frac{A(\mu \rightarrow \nu)}{A(\nu \rightarrow \mu)} = \frac{\min(1, e^{-\beta \Delta E})}{\min(1, e^{\beta \Delta E})},$$

and since ΔE is either positive or negative, the acceptance will be 1 in the step that minimizes the energy and $e^{-\beta \Delta E}$ in the opposite one, which satisfies the detailed balance. Also, the algorithm is ergodic, since we can transit between any two configurations μ and ν with a finite number steps, flipping the spins that yield different values between μ and ν .

Usually, the time scale between two sampled configurations corresponds to L^d Metropolis steps ($\equiv$ a sweep), which is chosen since it corresponds to the decorrelation time when $\beta = 0$.

2.3. Cluster algorithms

Near the critical temperature, the correlation length ξ diverges, which gives rise to large domains of clusters with the same spin orientation. For those configurations, the local algorithms demonstrate their shortcomings, falling victims of critical slowing down [47], i.e., the number of Monte Carlo steps we need to perform from a configuration ϕ to obtain a new uncorrelated configuration ψ is much bigger near the critical temperature than in any other point. Since we want to simulate three dimensional Ising with a temperature near T_C we will introduce cluster dynamics[48, 49], with particular focus on the Wolff algorithm[50].

2.3.1. Wolff algorithm

The Wolff algorithm for a generic system S [51, 52] invariant under a global symmetry transformation r (as an example we can think in the Ising model, that is invariant under global transformations $s_i \rightarrow -s_i$) builds a connected cluster (collection of lattice sites) and apply the symmetry transformation to the cluster. This means that the local energy between two sites that belong to the cluster remains the same, changing only at the boundary of the cluster. The algorithm goes as follows:

1. Pick a random site x_{ij} with value s_i and add it to a cluster;
2. Check all the neighbors of the site x_{ij} , and add each of them individually to the cluster with probability

$$p = \max(0, 1 - \exp(-\beta\Delta E)); \quad (2.27)$$

where $\Delta E \equiv E_f - E_i$ is the energy difference between the current energy among the pair of neighbors E_i and the energy if only the one that already belongs to the cluster gets the symmetry transformation applied onto him;

3. Repeat the step 2 for all the elements of the cluster until all the neighbor states of the cluster boundary were either all rejected or are part of the cluster;
4. Take all the elements of the cluster and apply to them the symmetry transformation r .

An example of an application of a Wolff step can be seen at Fig. 2.1. For the usual Ising model, the cluster only admits spins that share the orientation with the spin that was chosen at the first place. This comes since

$$E_i(\uparrow\uparrow) = -J \quad E_f(\uparrow\uparrow) = J \quad (2.28)$$

$$E_i(\downarrow\uparrow) = J \quad E_f(\downarrow\uparrow) = -J \quad (2.29)$$

which implies that

$$p(\uparrow\uparrow) = 1 - \exp -2\beta J \quad p(\downarrow\uparrow) = 0 \quad (2.30)$$

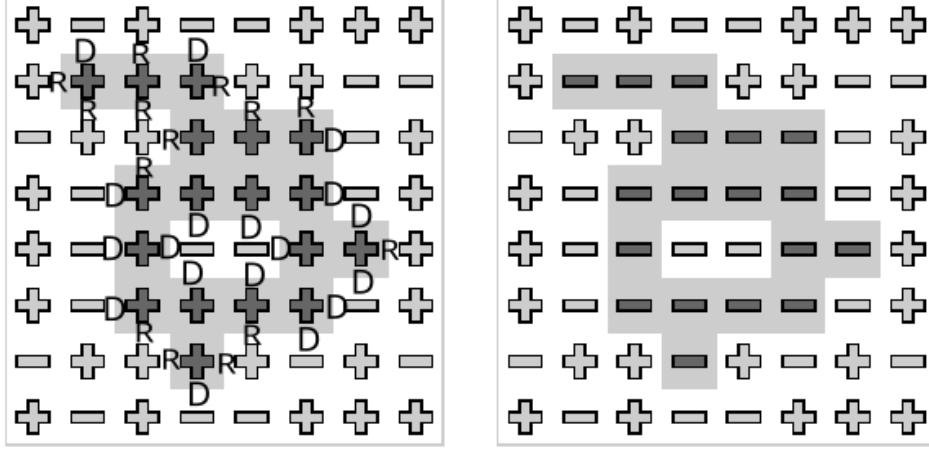


Figure 2.1.: A Wolff step for a finite temperature β builds a cluster (shaded region) and flips it. Image from [39].

since ΔE in the anti-aligned case is smaller than zero, so $\exp(-\beta\Delta E) > 1$ and equation 2.27 yields 0. However, this assumes a ferromagnetic bond. If the bond is anti-ferromagnetic, i.e., $J < 0$, that corresponds to flipping the values of the probabilities at equation 2.30. If we have both ferro and anti-ferromagnetic bonds throughout our model, the choice of p changes with the particular pair of spins we are dealing with.

Its intuitive to see that for inspection of equation 2.27 that for the Ising model the Wolff algorithm has two definite temperature limits. At high temperatures, $\beta \rightarrow 0$ and p goes to zero. This means that the cluster is a single spin that is flipped - which corresponds to the Metropolis behavior at high temperature as well. This allows us to extend the argument of the Metropolis ergodicity to the Wolff algorithm. However, at low temperatures $\beta \rightarrow \infty$ and the proposed lattice elements for the (anti-)ferromagnetic case are always accepted if they are (anti-)aligned, so we will perform flips of the full domains of aligned spins for the ferro case and checkerboard-like domains for the antiferro.

As stated before, the change in the energy of the system only occurs due to the boundary rejected spins, since inside the cluster the energy remains equal. As can be seen in the example of Fig. 2.1, the change of energy is $\Delta E = -2J(\#_D - \#_R)$ where $\#_D$ is the number of anti-parallel bonds pre-cluster flip and $\#_R$ is the number of parallel bonds. The probability of building a specific cluster is equal to the probability of accepting all the $\#_I$ inside bonds and inspecting and rejecting the parallel bonds at the boundary, i.e.,

$$T(\mu \rightarrow \nu) = p^{\#_I} (1 - p)^{\#_R}. \quad (2.31)$$

Now take into account that for the reverse step, the once parallel spins become anti-parallel

and vice-versa, so that the reverse cluster is build with probability

$$T(\nu \rightarrow \mu) = p^{\#I} (1 - p)^{\#D} . \quad (2.32)$$

We can now check how the chosen probability obeys to the detailed balance condition, since

$$\begin{aligned} P(\mu)T(\mu \rightarrow \nu) &= P(\nu)T(\mu \rightarrow \nu) \\ \frac{p^{\#I} (1 - p)^{\#R}}{p^{\#I} (1 - p)^{\#D}} &= e^{-\beta \Delta E} \\ e^{-\beta 2J(\#R - \#D)} &= e^{-\beta 2J(\#R - \#D)} . \end{aligned}$$

3. 3D Ising model with continuous wall coupling

Since we are now equipped with the powerful apparatus of Monte Carlo dynamics, we can start to explore the phenomenological aspects of the domain wall on the three-dimensional Ising model. This model lives in a lattice and the energy of a configuration ϕ is given by the Hamiltonian

$$H(\phi) = -J \sum_{\langle ij \rangle} s_i s_j, \quad (3.1)$$

whose summation runs over all the neighbors ij . The spin degrees of freedom take the value 1 or -1. The defect arises because we impose anti-periodic boundary conditions alongside a chosen direction (equivalent to a wall of antiferromagnetic J 's) and periodic alongside the others. At low temperatures, the spins will tend to be aligned in the bulk, but will favor anti-alignment alongside the wall. This leads to frustration; a generic domain of aligned spins will invert its signal when it goes through the anti-periodic boundary, which may lead different sections of the same domain to be at odds with each other. An illuminating example for the two-dimensional case can be seen at Fig. 3.1. Note that when the entire lattice is fully aligned, we consider that there is a domain wall at $z = 0$, since the energetically favorable configuration is the anti-alignment of the $s_{x,y,z=0}$ and $s_{x,y,z=L_z}$ spins. Also, since the transverse area is finite, the possibility of the existence of an odd number of domain walls must be considered.

The realization of the domain wall is usually achieved, for systems with bulk $J = 1$, with $J_{\text{wall}} = -1$, but the same arguments presented above hold true for a generic negative J_{wall} . Having said that, the behavior for intermediate, $J_{\text{wall}} \in [0, -1]$ is not so clear, and, in particular, near $J_{\text{wall}} = -1$, some energetic arguments suggest the occurrence of a delocalization of the domain wall. Such delocalization alongside the z direction is reminiscent of a Kosterlitz-Thouless (KT) transition[53]. The suggestion for a KT transition also comes from the fact that it is known that the system undergoes a transition like that for fixed $J_{\text{wall}} = -1$ by decreasing the temperature.

In this chapter, we will explore some ideas that arise from this interesting behavior. First we will expose some phenomenological facts about the domain wall in our system, sec. 3.1, and then we will figure out the existence of a KT transition at $J_{\text{wall}} = -1$, sec. 3.2.

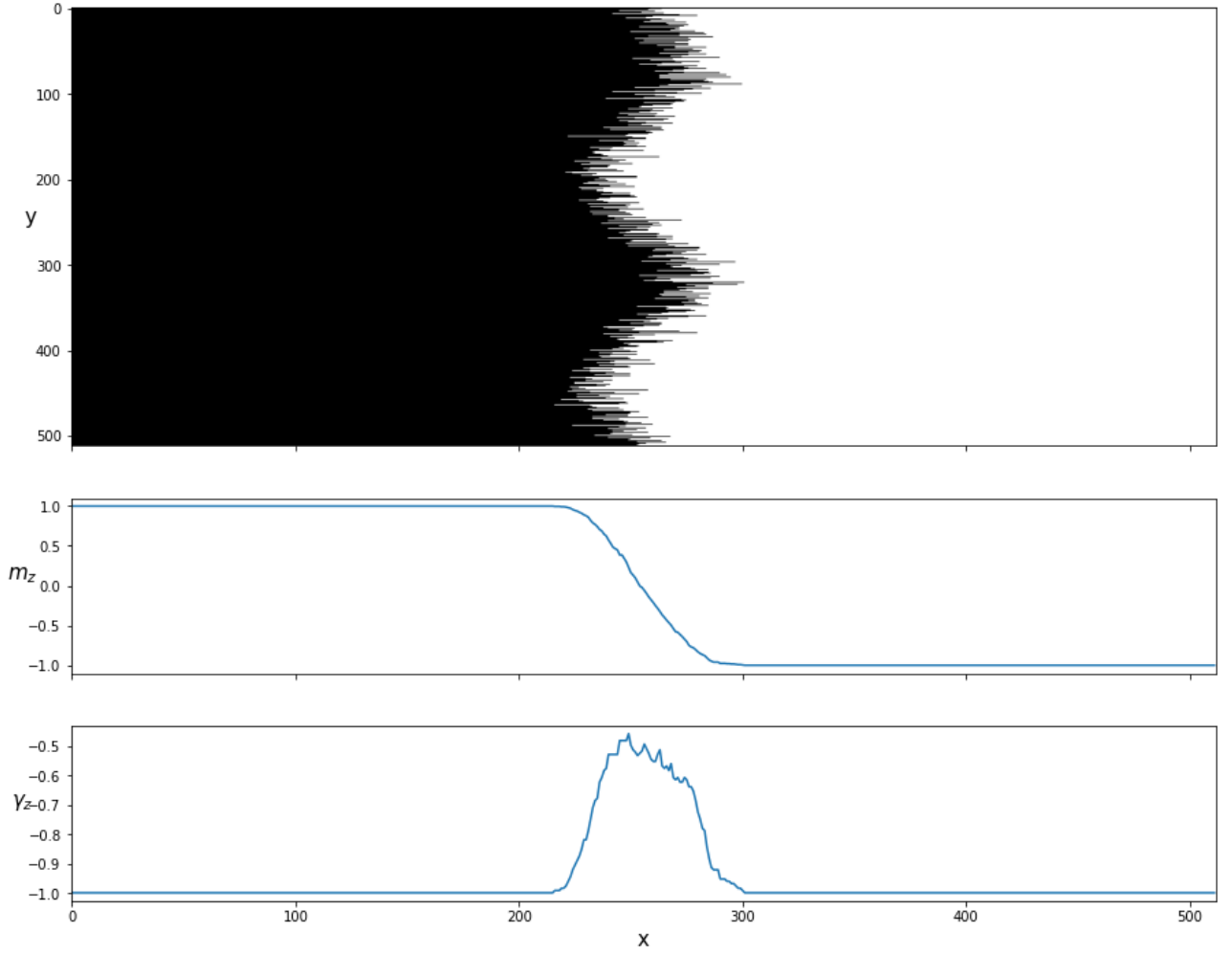


Figure 3.1.: Possible 2D Ising model configuration with periodic conditions alongside the y -direction and anti-periodic ones alongside the x -direction and system size 512×512 . An interface forms between domains of up (white region) and down (black region) spins. The two possible metrics to determine the wall position are when m_z , defined at equation 3.3 equals zero; or when γ_z , defined at equation 3.2, yields its maximum.

3.1. Phenomenology of the domain wall

There are two main ideas on how to detect the localization of the domain wall.

- Measure the mean local energy γ_z alongside a layer of fixed z . First, we define the quantity

$$\gamma_{z'} = \frac{1}{L_x L_y} \sum_{xy} s_{x,y,z=z'} s_{x,y,z=z'+1}, \quad (3.2)$$

that takes values near +1 when the spins are aligned with its forward neighbors and -1 when the opposite is true. So, we can define the position of the wall as the layer z where γ_z takes its minimal value;

- Measure the mean magnetization alongside a layer of fixed z . First, we define

$$m_z = \frac{1}{L_x L_y} \sum_{xy} s_{x,y,z=z'}, \quad (3.3)$$

that takes its absolute value near 1 when the layer has a clear preferential spin direction. We define the position of the layer by solving $m_z = 0$.

We will follow the scientific consensus[30, 54, 55], and go with the latter approach.

3.1.1. Bubble removal procedure

Usually, before the determination of the position of the domain wall, the first order bubbles are removed[29]. Such structures are clusters of same orientation that do not have an island of opposite orientation within them.

In order to remove them from our configuration, we identify, label and save the size of all the clusters of our spin configuration. This can be done through a simple Swedenson-Wang algorithm[48, 56] at $T = 0$. Then, we flip the largest cluster. We perform this bubble removal again until there is only one cluster left. An example of this procedure for a simplified configuration can be seen at figure 3.2. The results we obtained were not substantially different with or without the bubble removal.

3.1.2. Number of walls

Now that we know how to find the walls, it's instructive to study the distribution of their number for a system with a given set of parameters, Fig. 3.3. The distribution suggests

$$P(N \text{ walls}) \propto \exp(-\mu N), \quad (3.4)$$

with μ with as the chemical potential associated with one wall. This correspondence suggests a interpretation of our domain walls as defects of our system, that can be thought as particles,

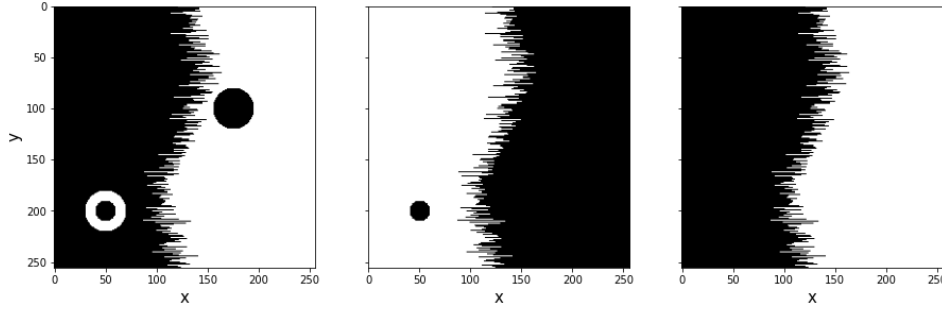


Figure 3.2.: Example of three steps of the bubble removal procedure for a simple configuration of two-dimensional Ising model. The SW algorithm identifies the largest cluster and flips it.

whose number is governed by the assigned chemical potential.

One could study, Fig. 3.4, how μ scales with the coupling J_{wall} .

3.1.3. Position of the wall

The study of $\mu(J)$ did not yield any clear conclusion about the different behaviors of our system. However, with a purely energetic argument, we do suspect that something may change as we tune J_{wall} . The reasoning is as follows (we will consider $J_{\text{wall}} < 0$ at all times) :

1. When $J_{\text{wall}} > -1$, the energy cost ($\equiv \Delta$) of a pair of anti-aligned spins, with one of them at $z = 0$ and the other at $z = L_z$, is $\Delta_{\text{wall}} = -2J_{\text{wall}}$ and in the bulk is $\Delta_{\text{bulk}} = 2$. Since $\Delta_{\text{bulk}} > \Delta_{\text{wall}}$ and there needs to be at least a domain wall, it is energetically favorable for the system to “place” the wall at $z = 0$;
2. When $J_{\text{wall}} = -1$, $\Delta_{\text{wall}} = \Delta_{\text{bulk}} = 2$. This means that it is equivalent to the system to place the wall at the boundary of at any of the L_z slices of the system;
3. When $J_{\text{wall}} < -1$, $\Delta_{\text{wall}} = -2J_{\text{wall}} > \Delta = 2$. Occurs the inverse that happened at situation 1) - the wall is prone to live in the bulk.

Since we are dealing with a system at a finite $\beta \sim \beta_C$, it is important to check, Fig. 3.5, if the above reasoning still applies.

We can see that there are clearly three different stages for our domain wall to be in. Starting at 3.5a), the histograms’ peaks flatten as the coupling constant J_{wall} decreases, until it becomes a uniform distribution. Then, the walls near $z = 0$ become suppressed.

One can define a quantitative measure, inspired by the Inverse Participation Ratio, of this behavior, if we consider $\rho(z, J_{\text{wall}})$ as the normalized histogram of the distributions of the z positions of the domain wall for a given J_{wall} ,

$$\Gamma(J_{\text{wall}}) = \sum_z \rho(z, J_{\text{wall}})^2. \quad (3.5)$$

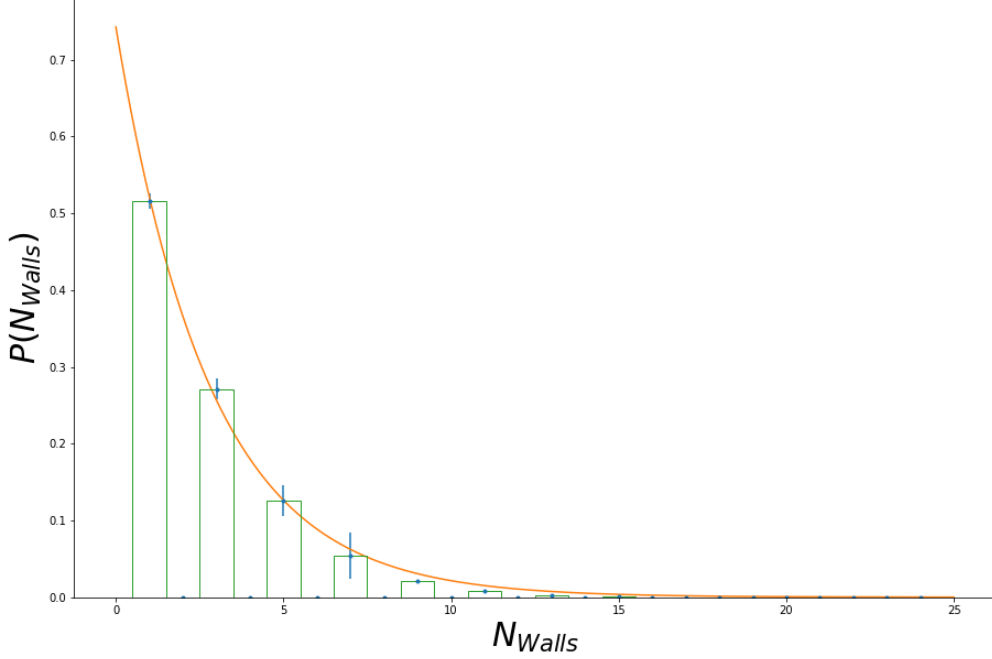


Figure 3.3.: Distribution for the number of walls for a system with $J_{wall} = 0.80$, $L_z = 96$ and $L_x = L_y = 30$ at $\beta = 0.223102$. There are, as expected, only configurations with an odd number of walls. The yellow curve corresponds to an exponential fit to the data.

Since the distribution is normalized, Γ takes extremal values at

$$\begin{cases} \rho(z, J_{wall}) = \delta_{z, z_0} & \rightarrow \Gamma = 1 \\ \rho(z, J_{wall}) = \frac{1}{L_z} & \rightarrow \Gamma = \frac{1}{L_z} \end{cases},$$

which correspond to the situation in which the wall is localized at z_0 and completely delocalized, respectively. The dependence of this metric with the coupling wall constant can be seen in figure 3.6.

We can check the Γ dependence for different transverse areas in figure 3.7. For bigger transverse areas, the energy cost to build a domain wall is higher ($\sim \sigma L^2$), so the thermal fluctuations have a reduced effect and the wall position distributions have a better defined behavior. For example, when $J_{wall} > -1$, $\rho(z)$ decays quickly for greater L . Such behavior is laid down at fig. 3.7.

All the results that we have presented in this subsection hint to a major question: Is there a phase transition at $J_{wall} = -1$? This idea is raised because at that point the distribution of the wall position is scale invariant

$$\rho(bz) = \frac{\rho(z)}{b}, \quad (3.6)$$

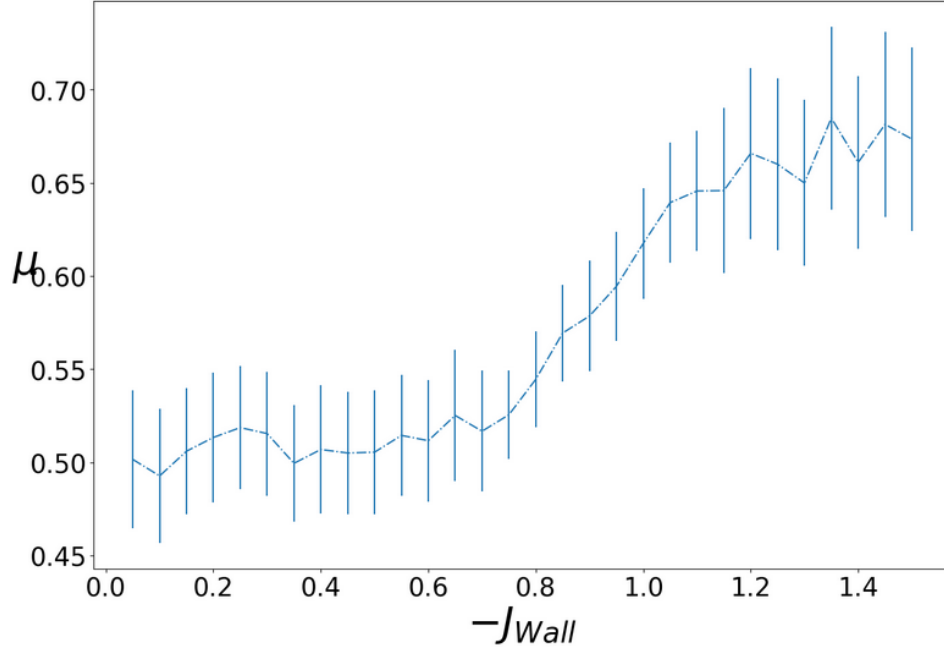


Figure 3.4.: Chemical potential μ with respect to the coupling wall constant for a system with parameters: $L_z = 96$ and $L_x = L_y = 32$. Despite the high error associated with each decay coefficient, one can distinguish a nearly constant behavior of μ up to $J_{wall} \sim 0.6$ and a monotonic increase afterwards.

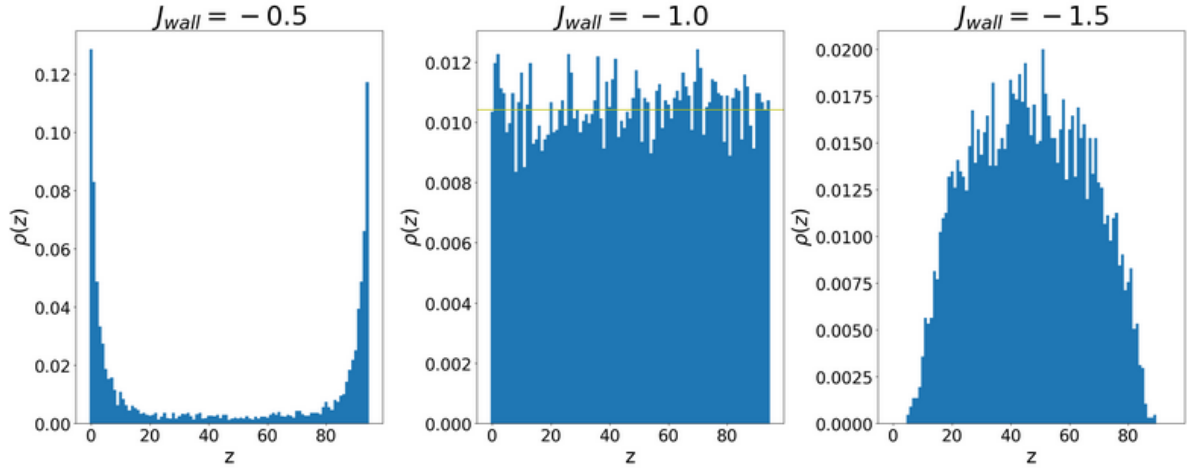


Figure 3.5.: Wall position histogram for a system with $\beta = 0.223102, L_z = 96, L_x = L_y = 40$ for three different values of J_{wall} . It is important to note that the y-axis does not have the same scale on the three figures. In the middle figure, we compare our data with the uniform distribution. There are clearly three distinct behaviors - the wall is either more probable to be near $z = 0$ for $J_{wall} > -1$; to be uniformly distributed within the system or to be suppressed at the edges.

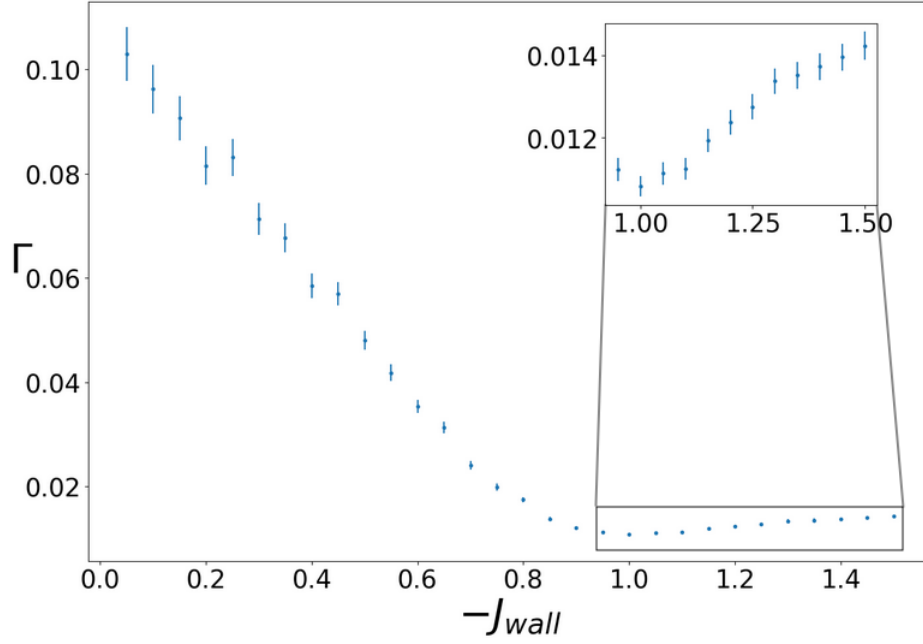


Figure 3.6.: Γ for a system with $\beta = 0.223102, L_z = 96, L_x = L_y = 40$ in function of J_{wall} . Γ decreases until $J_{wall} = -1$, that corresponds to the uniform distribution. After that, there is a slight increase, because the distribution is not quite uniform (walls are suppressed at z_0), but it never reaches the same values as before, due to the distribution still being uniform at the bulk.

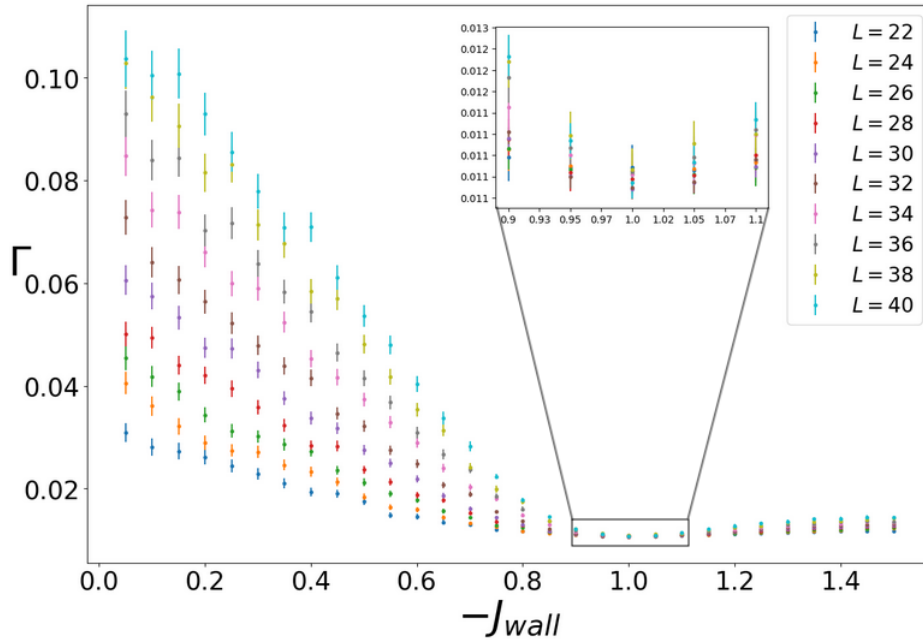


Figure 3.7.: Γ for a system with $\beta = 0.223102, L_z = 96$, in function of J_{wall} for different $L_x = L_y = L$. Γ takes greater values for greater L , except at the fixed point for $J_{wall} = -1$, since $\Gamma = \frac{1}{L_z}$, regardless of the transversal area.

since it corresponds to the uniform distribution, and because we have a delocalization of the domain wall, reminiscing of the vortex - antivortex unbinding in the XY model[53] and the entropy due to the possible localization of the wall goes with $\log(L_z)$.

3.2. $J_{\text{wall}} = -1$ scenario

It is possible to define the probability of the occurrence of a wall at a position z as

$$P(1 \text{ Wall}; z) = e^{-\beta\mu(J)}\rho(z), \quad (3.7)$$

and the z -dependence can be computed by analysis of the data from the histograms present at figure 3.5. The numerical data for $J_{\text{wall}} > -1$ yields a dependence

$$\rho(z) \propto z^{-\gamma(J_{\text{wall}}+1)}, \quad (3.8)$$

so, if we consider an effective potential $\tilde{V}_{J_{\text{wall}}}(z_1, z_2)$, that gives the strength of the interaction between a domain wall at z_1 and a defect (wall coupling equal to J_{wall}) at z_2 , we have

$$\tilde{V}_{J_{\text{wall}}}(z_1, z_2) = \frac{\gamma}{\beta} (J_{\text{wall}} + 1) \log(|z_1 - z_2|), \quad (3.9)$$

which correspond to the only relevant interaction term, since there are no defect-defect interplay due to only existing one defect in our system and there are no wall-wall interaction by definition.

We can further decompose the z -independent part as

$$e^{-\beta\mu(J)} = \underbrace{e^{-\beta\mu}}_y e^{-\beta F_{\text{wall}}(J)} e^{-\beta F_{\text{defect}}(J)} \quad (3.10)$$

with μ as the chemical potential of a domain wall in a periodic Ising model. This is the dominant term at the bulk, since the defect doesn't spread its influence to lattice spacing greater than the correlation length ξ . Also, the $F_{\text{wall}}(J)$ takes into account the possible $\log L_z$ positions at which the wall can be.

Our goal is to check if the considerations we have about the domain wall behavior are enough to, through a renormalization group analysis, conclude if the system undergoes a KT-transition at $J_{\text{wall}} = -1$. It is illuminating to briefly check a similar analysis that can be done in the XY-model.

3.2.1. Renormalization group analysis for the KT transition of the two-dimensional XY model

The XY-model[53] is defined in a square lattice with linear size L and the energy of a configuration ϕ is given by the Hamiltonian

$$H(\phi) = -J \sum_{\langle ij \rangle} \cos(\theta_i - \theta_j), \quad (3.11)$$

whose summation runs over all the neighbours ij . The spin degrees of freedom θ take the values $[0, 2\pi[$. If we impose periodic boundary conditions, at low temperatures, the stable configuration will correspond to all the spins being aligned. However, the low temperature case also admits non-uniform singular solution - vortices and anti-vortices. These topological defects have an associated integer winding number n_i , positive for a vortex and negative for an anti-vortex, that allow to express the interaction energy of a pair of defects with its center at r_i and r_j as

$$E_{\text{int}}(r_i, r_j) = -2\pi J n_i n_j \log |r_i - r_j| + E_C, \quad (3.12)$$

with E_C as the energy of the defects cores. The entropy of a defect pair is $2 \log L^2$, since each vortex has L^2 accessible sites. The energy between a vortex - anti-vortex pair is minimized when they are close to each other, i.e., are binded, which means that at low temperatures we have a “gas” of vortex - anti-vortex pairs. However, at the KT temperature $T_{\text{KT}} = \frac{\pi J}{2}$, the defects unbind and roam independently through the lattice. We will show that through a RG analysis, following the arguments present at [?].

Let's start by defining the scalar quantities

$$x \equiv 2 - \beta\pi J \quad y_0 = \exp(-\beta\mu), \quad (3.13)$$

with μ as the chemical potential of any vortex or anti-vortex. By virtue of 3.12, the probability of a configuration with a pair of vortices with $n_1 = 1$ and $n_2 = -1$ is

$$P(r_1, r_2) = y_0^2 e^{-\beta E_C} |r_1 - r_2|^{-2\pi\beta J}. \quad (3.14)$$

Given that the distance independent part is $y_0^2 e^{-\beta E_C}$, we can redefine the y quantity as $y \equiv y_0 e^{-\beta E_C/2}$. For a global rescaling of the system, $r \rightarrow br$, the probability $P(r_1, r_2)$ remains invariant. Taking into account the change in entropy, since the available sites for a vortex goes

from L^2 to $b^{-2}L^2$, we arrive at the scaling relation for the y -quantity,

$$\tilde{P}(r_1, r_2) = P(br_1, br_2) \quad (3.15)$$

$$\tilde{y}^2 |r_i - r_j|^{-2\pi\beta J} = y^2 e^{\overbrace{2 \log b^2}^{\equiv \text{entropy change}}} |br_1 - br_2|^{-2\pi\beta J} \quad (3.16)$$

$$\tilde{y}^2 = b^{4-2\pi\beta J} y^2 \quad (3.17)$$

$$\implies \tilde{y} = b^{2-\pi\beta J} y, \quad (3.18)$$

and if we consider an infinitesimal rescaling, $b = 1 + dl$, we arrive at the first differential RG equation

$$\frac{dy}{dl} = (2 - \pi\beta J) y = xy. \quad (3.19)$$

The differential equation for x is derived near the KT transition ($|x| \ll 1$). Since the presence of vortices disorders the system, x should increase with y^2 . Stability arguments [57] specific to the XY-model can be employed to show that this is the only relevant term, which leads us to the other RG equation

$$\frac{dx}{dl} = \varepsilon^2 y^2, \quad (3.20)$$

with ε as a constant.

Eq. 3.19 and 3.20 are the Kosterlitz equations, and if we can arrive to this RG equations from our model we can conclude that the system undergoes a KT transition at $J_{\text{wall}} = -1$. Anyhow, that appears to not be possible with the problem configuration as it stands. Firstly, our analogue of the defects, the domain wall and the wall of J_{wall} couplings do not have a similar standing. As we seen in the XY, the vortices and anti-vortices have similar chemical potential μ . In our case, however, the wall of J_{wall} couplings appears only once, while the domain wall have a specific μ associated. Also, the entropy of our defects goes with $\log L_z$, since the model is effectively one-dimensional. This considerations hinder our ability to reach the KT equations for the domain wall Ising.

4. Non-equilibrium methods

The first set of techniques we will employ to calculate the free energy difference between periodic and anti-periodic Ising 3D models will be covered in this chapter. Both the methods here present will share the common characteristic of connecting both models through a non-equilibrium transformation.

We will first discuss the Jarzynski algorithm [1, 58] and then derive the eponymous equality. Then, it will be shown how the algorithm has a close analogue to a seemingly very disparate class of algorithms - normalizing flows[34, 59].

In parallel, we will discuss how both the Jarzynski algorithm and a normalizing flows approach can be used to deal with our problem.

4.1. Jarzynski Equality

Jarzynski equality, first derived in the late nineties[1], established a relation between the average dissipated work during a non-equilibrium transformation and the difference in free energy between the endpoints of the aforementioned transformation.

This equality is most intriguing, since it connects an irreversible quantity, highly dependent of details of the change imposed on the system, with the free energy difference, a quantity only worth discussing at equilibrium. To be convinced of the effectiveness and usefulness of this physical fact, it's illuminating to behold its ubiquity in various fields of the natural sciences. It is used in topics so disparate as Lattice Gauge Theory[14], molecular chemistry [60, 61] and biophysics[62]. Numerous experimental tests were devised to test its validity, with positive results [62, 63].

Before proceeding to the proof of the Jarzynski equality, it's important to clarify what we mean by the "endpoints" of a system. When we make such claim, we consider a system with dependence on an external (and tunable) parameter λ . Without loss of generality, we can say that the non-equilibrium evolution of the system goes from $\lambda_{\text{in}} = 0$ to $\lambda_{\text{fin}} = 1$.

Let's take, as an illustrative example, a one dimensional single particle system with

$$H(\lambda) = \frac{p^2}{2m} + \frac{\lambda}{2}m\omega^2 x^2, \quad (4.1)$$

the application of the Jarzynski equality will allow us to obtain the free energy difference between the free particle case ($\lambda = 0$) and the harmonic potential one ($\lambda = 1$).

Let's sketch a proof that follows closely the argument presented by Jarzynski itself [1], for a isolated system evolving at a fixed temperature T . If we let the external parameter evolve from $\lambda = 0$ to $\lambda = 1$ adiabatically, we have the expression for the canonical work,

$$W_\infty = F_1 - F_0 \equiv \Delta F, \quad (4.2)$$

but what happens if we change λ at a finite rate? The system will lag behind quasi static equilibrium with the reservoir, and the total work W will depend on its initial condition[58]. What Jarzynski equality (JE) tells us is that if we have an ensemble of trajectories whose initial configurations are drawn from the equilibrium distribution at $\lambda = 0$, the work function obeys, if we average $e^{-\beta W}$ over the trajectories' ensemble,

$$\langle e^{-\beta W} \rangle = \frac{Z_{\lambda=1}}{Z_{\lambda=0}}. \quad (4.3)$$

4.1.1. Proof for deterministic evolution[1]

Let's say that we have a system governed by an Hamiltonian H_λ that depends on an external parameter λ , that is switched from $\lambda = 0$ to $\lambda = 1$ during a time t_s . The evolution of the system for an initial configuration z_0 is given by a deterministic trajectory $z(t)$. This determinism means that if we know the configuration at time t we know from which initial configuration it came from. This informs our choice of notation: $z(z_0, t)$, that corresponds to the configuration z , the result of the evolution that started at z_0 .

The initial equilibrium distribution for $\lambda = 0$ is the canonical, $f(z_0, 0) = \frac{\exp(-\beta H_0(z_0))}{Z_0}$ distribution. For $t > 0$, $f(z(z_0, t))$ will be a non-equilibrium distribution, with its evolution given by the Liouville equation,

$$\frac{\partial f}{\partial t} + \{f, H_\lambda\} = 0. \quad (4.4)$$

Without loss of generality, let's consider $\lambda = \frac{t}{t_s}$. At t_s the function $f(z(z_0, t))$ is obtained from solving equation 4.4. Each final configuration $z(z_0, t_s)$ has a function $W(z)$ associated, that corresponds to the accumulated work performed on the trajectory that ends at $z(z_0, t_s)$, i.e.,

$$W(z) = \int_0^{t_s} \dot{\lambda} \frac{\partial H_\lambda(z(z_0, t))}{\partial \lambda} dt = \int_{z_0}^z dH_\lambda = H_1(z(z_0, t_s)) - H_0(z_0). \quad (4.5)$$

This allows us to obtain the average of $\exp(-\beta W)$ in terms of the final distribution and the work function,

$$\langle \exp(-\beta W) \rangle = \int dz f(z(z_0, t_s)) \exp(-\beta W(z, t_s)); \quad (4.6)$$

this equation simplifies if we consider, by the Liouville theorem, that the phase space density is conserved along any trajectory

$$f(z(z_0, t)) = f(z_0), \quad (4.7)$$

which leads to

$$\langle \exp(-\beta W) \rangle = \int dz \frac{\exp(-\beta H_0(z_0))}{Z_0} \exp(-\beta (H_1(z) - H_0(z_0))) \quad (4.8)$$

$$= \int dz \frac{\exp(-\beta H_1(z))}{Z_0} = \frac{Z_1}{Z_0}. \quad (4.9)$$

Equation 4.9 corresponds to the Jarzynski equality.

4.1.2. Proof for Monte Carlo evolution

In this subsection, we will take the kernel ideas of the Jarzynski equality and apply them to a system that evolves under a generic Monte Carlo algorithm, with an original proof for the Jarzynski equality for this particular case.

We start by discretizing the work in N evenly spaced steps. For each time step $t_i = i/N$ the external parameter λ takes the value i/N . Let's say that the evolution at time t_i is governed by a Monte Carlo algorithm controlled by the Hamiltonian H_λ , which means that the trajectory $z(t)$ for an initial z_0 is no longer deterministic. Define the probability of a given trajectory $\{z_0, z_1, \dots, z_N\}$ as $P(z_0, z_1, \dots, z_N)$. The work accumulated in the time step i is

$$\delta W_i = H_{i+1}(z_i) - H_i(z_i), \quad (4.10)$$

then

$$\langle \exp -\beta W \rangle = \int \prod_{i=0}^N dz_i P(z_0, z_1, \dots, z_N) \exp \left(-\beta \sum_{i=0}^{N-1} \delta W_i \right). \quad (4.11)$$

Let's start by rewriting the probability of the trajectory as

$$P(z_0, z_1, \dots, z_N) = A_0(z_0) T(z_1|z_0) T(z_2|z_1) \dots T(z_N|z_{N-1}), \quad (4.12)$$

where A_i is the equilibrium distribution for $\lambda = i$ and $T(z_{i+1}|z_i)$ the transition probability from the configuration z_i to the configuration z_{i+1} given by the MC algorithm ruled by H_{i+1} .

We can reorganize the integrand of the equation 4.11,

$$\begin{aligned} & \int \prod_{i=0}^N dz_i P(z_0, z_1, \dots, z_N) \exp \left(-\beta \sum_{i=0}^{N-1} \delta W_i \right) \\ &= \int \prod_{i=0}^N dz_i A_0(z_0) \prod_i^{N-1} T(z_{i+1}|z_i) \exp(-\beta (H_{i+1}(z_i) - H_i(z_i))); \end{aligned}$$

however, recall that our Monte Carlo algorithm must obey the detailed balance condition.

This means that

$$T(z_{i+1}|z_i) \exp(-\beta H_{i+1}(z_i)) = T(z_i|z_{i+1}) \exp(-\beta H_{i+1}(z_{i+1})) \quad (4.13)$$

so

$$\langle \exp(-\beta W) \rangle = \int \prod_{i=0}^N dz_i A_0(z_0) \prod_{i=0}^{N-1} T(z_i|z_{i+1}) \exp(-\beta H_{i+1}(z_{i+1})) \exp(\beta H_i(z_i)). \quad (4.14)$$

Paying closer attention to the first term, we can see that

$$\langle \exp(-\beta W) \rangle = \int \prod_{i=0}^N dz_i \frac{\exp(-\beta H_0(z_0))}{Z_0} \exp(\beta H_0(z_0)) T(z_0|z_1) \exp(-\beta H_1(z_1)) (\dots) \quad (4.15)$$

$$= \frac{1}{Z_0} \int \prod_{i=0}^N dz_i T(z_0|z_1) \exp(-\beta H_1(z_1)) (\dots) \quad (4.16)$$

so that the only dependence on z_0 is on $T(z_0|z_1)$. Since

$$\int dz_0 T(z_0|z_1) = 1, \quad (4.17)$$

we have

$$\langle \exp(-\beta W) \rangle = \frac{1}{Z_0} \int \prod_{i=1}^N dz_i \exp(-\beta H_1(z_1)) \prod_{i=1}^{N-1} T(z_i|z_{i+1}) \exp(-\beta H_{i+1}(z_{i+1})) \exp(\beta H_i(z_i)) \quad (4.18)$$

we can iterate the above steps, succesively eliminating the dependences on z_i ; we finnaly arrive at

$$\langle \exp(-\beta W) \rangle = \frac{1}{Z_0} \int dz_N \exp(-\beta H_N(z_N)) = \frac{Z_N}{Z_0}, \quad (4.19)$$

which is again the Jarzynski equality.

Nowhere in this proof we took consideration about the speed at which λ changes, so JE holds regardless of that quantity. In order to illustrate this fact, let's study two limiting cases of this theorem, one at which the λ goes from 0 to 1 in a single iteration ($\equiv$ snap case) and one at which the λ changes adiabatically ($\equiv$ adiabatic case).

4.1.3. Snap case (N=1)

If we perform the non-equilibrium transformation in a single step,

$$W = H_1(\phi) - H_0(\phi), \quad (4.20)$$

and

$$\langle e^{-\beta W} \rangle_0 = \int e^{-\beta(H_1(\phi) - H_0(\phi))} \frac{e^{-\beta H_0(\phi)}}{Z_0} d\phi \quad (4.21)$$

$$= \int \frac{e^{-\beta H_1(\phi)}}{Z_0} d\phi \quad (4.22)$$

$$= \frac{Z_1}{Z_2} = e^{-\beta \Delta F}, \quad (4.23)$$

and Jarzynski equality holds.

Well, this is nothing more but the reweighting technique, i.e., the trick that allows us to compute the properties of a system through the equilibrium probability distribution of a different one.

4.1.4. Adiabatic case ($N \rightarrow \infty$)

If we let the number of updates of λ tend to ∞ , the system is updated in a slow, adiabatic fashion, and by virtue of equation 4.2,

$$e^{-\beta W} = e^{-\beta \Delta F}, \quad (4.24)$$

which implies that the mean

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F} \quad (4.25)$$

follows the Jarzynski equality. Note that for this limit we only need one measurement to obtain the free energy difference, which raises a question about the dependence between the rate at which λ evolves and the variance of the distribution $e^{-\beta W}$. As we will prove later, chapter 6, this is an integral feature of this method.

4.2. Application to Ising 3D

To apply the Jarzynski formulation to our specific problem, one needs to particularize the, so far, generic transformations, and, since we want to “connect” the Ising 3D model with periodic and antiperiodic boundary conditions alongside the z direction, the external parameter that we will tune will be the coupling constant of the wall at $z = 0$, i.e.,

$$H_\lambda(\phi) = - \sum_{\langle xy \rangle} J_{xy} s_x s_y, \quad (4.26)$$

with

$$J_{xy} = \begin{cases} 1 & \text{for bulk interactions} \\ 1 - 2\lambda & \text{for wall interactions} \end{cases},$$

given that we recover the usual Ising model at $\lambda = 0$ and the anti-periodic one at $\lambda = 1$. This definition of the Hamiltonian, allows us to naturally arrive at the work done to the system for a change of the wall couplings, that consists in, for a fixed spin configuration ϕ ,

$$\begin{aligned} \delta W_{i+1} &= H_{\lambda+\frac{1}{N}}(\phi) - H_\lambda(\phi) \\ &= \frac{2}{N} \sum_{\text{wall}} s_x s_y, \end{aligned} \quad (4.27)$$

in which the $\sum_{\text{wall}}$ corresponds to a summation over the spins connected through a wall bond.

Finally, it only remains to be specified which Monte Carlo dynamic will be employed to evolve our system. We will study the Ising 3D model near the critical point, so it is only logical to employ an algorithm that do not suffer greatly of critical slowing down. With that in mind, we will resort to the generic Wolff algorithm, in contrast with previous applications of this algorithm to the same system, that used the multispin coding approach[14]. Between each update we perform a Wolff sweep, that consists in the successive application of $\frac{N}{N_{\text{cluster}}}$ Wolff steps, with N_{cluster} as the average cluster size for β .

4.3. Normalizing flows

Jarzynski equality was a novel approach to this free energy computation problem. Its core tenet, the possibility of measuring an equilibrium property through a non-equilibrium transformation, was recently identified to be very similar to the philosophy behind Normalizing Flows[34].

Normalizing flows (NF)[59] are deep generative models, designed to sample from hard to deal with probability distributions.

In the past few years, there has been a series of papers[64, 65, 66, 67, 68, 69, 70, 71] that illustrate the potential of this approach to effectively sample lattice configurations. The main attraction of this method for the field of lattice gauge theory (LGT) is its good behavior near the critical point of theories[71], i.e., the ability to sample configurations near criticality without suffering from critical slowing down.

Despite the promise that this approaches may entail, the reason we are interested in this models lies with the aforementioned possibility of performing free energy difference computa-

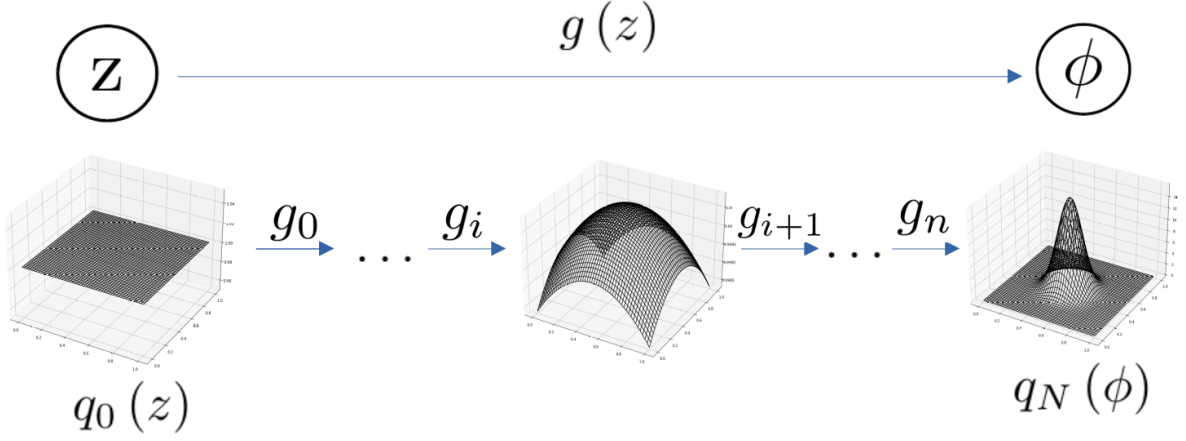


Figure 4.1.: Normalizing flows between a prior distribution $q_0(z)$ and a target distribution $q_N(\phi)$.

tions between two models[34].

In this subsection we will start by explaining with detail the inner workings of the normalizing flows, making clear why they are so coveted by the LGT field, and show how to compute ΔF with them.

4.3.1. Definition and inner working

Normalizing flows can be thought as a discrete composition of N bijective and differentiable functions (commonly called layers), i.e. a map, between two probability distributions - a base, prior, distribution to a target distribution.

This flow acts in the space of a D -dimensional configuration ($\mathbb{R}^D \rightarrow \mathbb{R}^D$), and the change in the configurations ($z \rightarrow g(z) \equiv \phi$) gives rise to the associated change in the probability distribution,

$$q_N(\phi) = q_0(g^{-1}(\phi)) |\det J(\phi)|, \quad (4.28)$$

with J as the Jacobian of the function g^{-1} . As 4.1 suggests, our function g will be thought as a composition,

$$g(z) = (g_N \circ g_{N-1} \dots \circ g_1 \circ g_0)(z), \quad (4.29)$$

with the according probability distribution change induced by the application of g_n being

$$q_{n+1}(z_{n+1}) = q_n(z_n) |\det J_n(z_n)|^{-1}, \quad (4.30)$$

for J_n as the Jacobian of g_n^{-1} and $z_{n+1} = g_n(z_n)$.

The layers g'_i s need to be highly expressible¹ if we want to correctly sample from an intractable distribution, so one of the common choices is to build each layer as a composition of trainable neural networks[59] (NN). NNs allow our model to, after a training procedure, arrive at the optimal behavior for the NF. This training requires a loss function and a protocol for the update of the NNs, whose job is to arrive at a setup of layers that minimize the loss function $\mathcal{L}$.

Given that we want $q_N(z_N)$ to resemble a distribution p , one can choose the Kullback-Liebler divergence as the loss function,

$$D_{KL}(q_N||p) = \int d\phi q_N(\phi) [\ln q_N(\phi) - \ln p(\phi)], \quad (4.31)$$

and, if we particularize by choosing the Boltzmann distribution, eq. (2.6), it simplifies

$$\begin{aligned} D_{KL}(q_N||p) &= \int d\phi q_N(\phi) [\ln q_N(\phi) + \beta H(\phi) + \ln Z] \\ &= \int d\phi q_N(\phi) [\ln q_N(\phi) + \beta H(\phi)] + \underbrace{\ln Z \int d\phi q_N(\phi)}_1, \end{aligned} \quad (4.32)$$

and since $\ln Z$ is a constant, the loss can be defined up to a constant

$$\mathcal{L}(q_N) \equiv D_{KL}(q_N||p) - \ln Z. \quad (4.33)$$

Equation 4.33 is uniquely defined by β and the hamiltonian function of the target distribution, so there is no need to sample p to compute the Kullback-Liebler divergence. This allows the NF approach to be self-consistent, and, in opposition to most of the deep learning methods, do not require a training set to compute the loss. In order to train the model, we start by drawing N samples from the prior distribution, that are subsequently sent through the NF, yielding a final distribution that can be used to compute the loss function. If we have a generic trainable parameter ω inside a NN, it can be updated through gradient descent $\omega \rightarrow \omega - \eta \frac{\partial \mathcal{L}}{\partial \omega}$, for a given learning rate η . The useful property of NN is that, by virtue of the the backpropagation algorithm[72], the term $\frac{\partial \mathcal{L}}{\partial \omega}$ can be efficiently computed by exploiting the NN structure. An example for a very simple NN architecture can be found at Appendix D.

We update the parameters of the NN until arriving at a minima of the loss function.

4.3.2. Affine layers

To apply normalizing flows to the Ising model, we need to specify how a layer transforms a lattice configuration. There is a lot of liberty in this choice. The layer configuration has only three requirements: must be invertible, have an easily computed Jacobian and a high

¹Expressible here means that the composition of layers need to be flexible enough so that the final probability distribution can be as complex as we want it to be.

expressive power[59]. Following the architecture present in [73], if we have a configuration ϕ with D lattice sites, we split it into two checkerboard-like sub-lattices, ϕ_1 and ϕ_2 , each with $D/2$ lattice sites, with the layer i acting as

$$g_i(\phi) = \begin{cases} z_1 = \phi_1 \\ z_2 = \phi_2 \otimes e^{s_i(\phi_1)} \oplus t_i(\phi_1) \end{cases}, \quad (4.34)$$

and the output thought as the recombination of the two sub-lattices z_1 and z_2 . The objects s_i and t_i are the functions that correspond to the neural networks, i.e., the trainable parameters of our flow. The output of the NNs is still $D/2$ -dimensional. We take the $s_i(\phi_1)$ and exponentiate it elementwisely. The rest of the operations are understood if we consider that $\otimes$ here denotes elementwise multiplication and $\oplus$ elementwise addition.

We can see how this choice of layer fulfils our criteria. First, check how this kind of layers are easily inverted, with

$$g_i^{-1}(z) = \begin{cases} \phi_1 = z_1 \\ \phi_2 = (z_2 \ominus t_i(z_1)) \otimes e^{-s_i(z_1)} \end{cases}. \quad (4.35)$$

The Jacobian matrix corresponds to

$$J_i(\phi) = \begin{bmatrix} \mathbb{I}_{D/2} & \cdots \\ 0 & e^{\mathbb{S}_i(\phi_1)} \end{bmatrix}, \quad (4.36)$$

with $\mathbb{S}_i(\phi_1)$ as a diagonal matrix with entries

$$(\mathbb{S}_i(\phi_1))_{jk} = (s_i(\phi_1))_j \delta_{jk}, \quad (4.37)$$

that allows us to easily compute the determinant

$$\det J_i(\phi) = \prod_{j=1}^{D/2} \left(e^{s_i(\phi_1)} \right)_j. \quad (4.38)$$

Note that an application of a layer leaves ϕ_1 invariant. Then, since a Normalizing Flow is composed of many layers, each successive pair of layers needs to transform alternate sublayers, if we want to convey changes on the entire configuration.

4.3.3. Normalizing flows as a Probability Flow Monte Carlo

So, if we have an already trained Normalizing Flow, if we sample configurations $\{\phi_0, \dots, \phi_M\}$ from the initial distribution q_0 , then $g(\phi_i)$ will constitute a sample of configurations that follow the target distribution p .

However, it is important to state that the NF may not be perfect, and q_N can be different

from p . In order to correct minor disagreements, we can build a Metropolis algorithm[66], that accepts and rejects a new configuration given a previous one with probability $\min(1, e^{-\beta\Delta E})$, with the string of suggested configurations $\{g(\phi_0), \dots, g(\phi_M)\}$.

Since two consecutive configurations are completely independent, the only thing that governs the autocorrelation is the rejection probability of the Metropolis, and it can be shown [74] that NF's that train for longer have lower rejection rates. We can then construct a Markov chain with small autocorrelation, even at the critical point. Nonetheless, the training process can be computationally expensive.

4.3.4. NF for Free energy measurement

The idea present in this subsection was firstly presented in [34], after the idea of combining deterministic transformations between two systems and the Jarzynski equality being flirted with at [75].

Let's introduce a weight function $\tilde{w}$ for a configuration ϕ

$$\tilde{w}(\phi) = \frac{\exp(-\beta H(\phi))}{Z_0 q_N(\phi)}, \quad (4.39)$$

with H as the Hamiltonian of the desired canonical distribution, Z_0 as the partition function of the prior distribution and q_N as the target probability distribution.

This definition allows us to rewrite Z as:

$$\begin{aligned} Z &= \sum_{\phi} \exp(-\beta H(\phi)) = \sum_{\phi} Z_0 q_N(\phi) \tilde{w}(\phi) = Z_0 \langle \tilde{w}(\phi) \rangle_{q_N} \\ \frac{Z}{Z_0} &= \langle \tilde{w}(\phi) \rangle_{q_N} \end{aligned} \quad (4.40)$$

Since

$$\ln \tilde{w}(z) = -\left(\beta H(g_N(z)) + \ln Z_0 + \ln q_N(g_N(z)) \right) \quad (4.41)$$

$$= -\left(\beta H(g_N(z)) + \ln Z_0 + \ln \underbrace{q_0(z)}_{\frac{e^{-\beta H_0(z)}}{Z_0}} - \underbrace{\sum_{n=0}^N \ln |\det J_n(z_n)|}_{\equiv \beta Q} \right) \quad (4.42)$$

$$\tilde{w}(z) = e^{-\beta(H(g_N(z)) - H_0(z) - Q)}, \quad (4.43)$$

we can rewrite 4.40 as

$$\frac{Z}{Z_0} = \langle \exp(-\beta\{H(g_N(z)) - H_0(z) - Q\}) \rangle_{q_0}, \quad (4.44)$$

which clearly resembles the Jarzynski equality. This allows us to sample z from an initial system governed by an Hamiltonian H_0 , send such sample through a previously trained NF and, by computing the successive values of J_n and the final set of configurations on which z ends, we can arrive at $\langle w(g(z)) \rangle_{q_0}$. The main difference between the Jarzynski procedure and this one is that the NF is a fixed, deterministic transformation. A configuration drawn from the initial probability distribution will always yield the same result after going through the NF.

The Jarzynski algorithm can be seen as an analogue of the NFs, but the layers, rather than deterministic, are probabilistic[34]. NFs become then the deterministic limit of this so called stochastic normalizing flows, with $P(\phi_{j+1}|\phi_j) = \delta(\phi_{j+1} - g_j(\phi_j))$, and the pure Jarzynski equality corresponds to the case when the probability is determined by the Monte Carlo evolution.

4.3.5. Two particle case

In order to exemplify the power of this approach, we will apply it to a very simple two particle system, which can be analytically treated, two non-interacting particles A and B in one dimension, each of them in a quadratic potential, such that the hamiltonian corresponds to:

$$H(x_A, x_B) = H_1(x_A) + H_2(x_B) = V_1(x_A - \mu_1)^2 + V_2(x_B - \mu_2)^2, \quad (4.45)$$

and to keep this example simple, let's consider $\mu_1 = -1, \mu_2 = 1$ and $V_1 = V_2 = \frac{1}{\beta}$.

Let's assume now that there is an elastic interaction between the particles,

$$V_{int}(x_A, x_B) = V(x_A - x_B)^2, \quad (4.46)$$

and our goal is to discover the free energy difference between the two cases (interacting and non-interacting one). Since this is a simple case, we can compute the partition function of each case. Let's start with the non-interacting one:

$$Z_0 = \int dx_A \int dx_B \exp(-(x_A - 1)^2 - (x_B + 1)^2) \quad (4.47)$$

$$= \left(\int dx \exp(-(x - 1)^2) \right)^2 = \pi, \quad (4.48)$$

in view that we are dealing with a simple Gaussian integral.

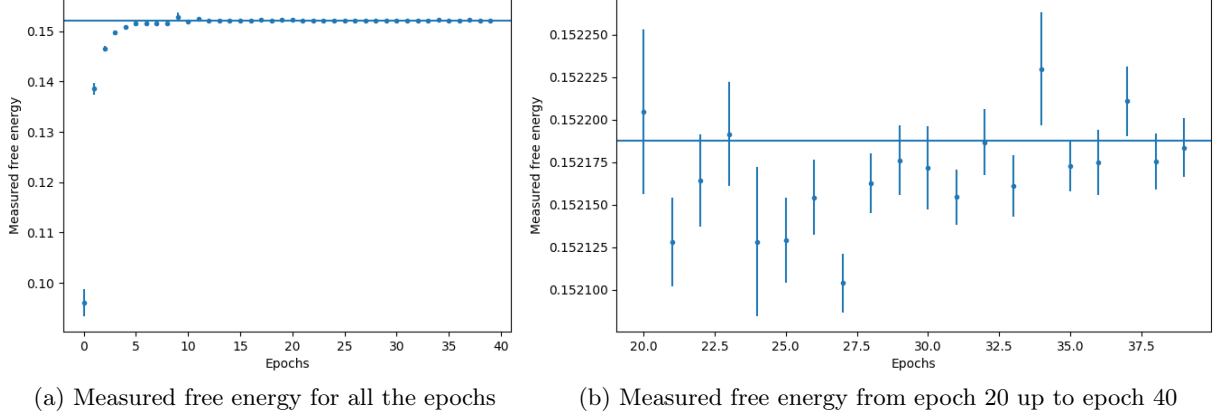


Figure 4.2.: Measured free energy of the first 40 NF eras for the two particle system for a batch of 1000 independent measurements. The horizontal line corresponds to the analytical result. Each era consists of 100 epochs, and each epoch corresponds to a backpropagation update of the NF weights due to the forward pass of a batch of 640 configurations sampled from the $\lambda = 0$ case. As we can see, the results converge quickly to the true value. Further training of the Normalizing Flow don't allow the precision to exceed the fifth decimal place.

The computation of Z_1 is a little bit trickier, since we need to first diagonalize and find the eigenvectors of the covariance matrix. The final result is:

$$Z_1 = \int dx_A \int dx_B \exp \left(- (x_A - 1)^2 - (x_B + 1)^2 - (x_A - x_B)^2 \right) = \frac{\pi}{\sqrt{3}e^{\frac{4}{3}}}, \quad (4.49)$$

which means that

$$\frac{Z}{Z_0} = \frac{1}{\sqrt{3}e^{\frac{4}{3}}}. \quad (4.50)$$

We can now build our normalizing flow according to the guidelines that were previously stated, with a detailed survey of the architecture of the NF employed in this calculation found in the Appendix E. The results can be seen at figure 4.2, and we can see that the NF approach yields a very good result for this simple case.

4.3.6. Hubbard-Stratonovitch transformation

If we want to generalize this formulation to compute the free energy difference of the periodic and anti-periodic Ising model, we need, effectively, to circumvent the glaring issue that the degrees of freedom of the Ising model are discrete, which comes into conflict with the fact that the NF returns continuous outputs.

In order to remedy that, we construct a (probabilistic) map between the usual Ising configurations and lattice realizations with continuous values at the lattice sites. This map was previously used to sample spin glass configurations with NFs at [76], and configures a Hubbard-Stratonovitch transformation of the Ising configurations.

The map goes from a regular configuration s to x , a lattice configuration with the same geometry but with continuously valued degrees of freedom, is

$$p(x|s) \propto e^{-\frac{1}{2}(x-s)^T \beta \tilde{J}}, \quad (4.51)$$

with J as the coupling matrix for the Ising model, i.e.,

$$J_{ij} = \begin{cases} -J & \text{if } i \text{ and } j \text{ are neighbours} \\ 0 & \text{otherwise} \end{cases},$$

and $\tilde{J}$ as the diagonally shifted matrix

$$\tilde{J}_{ij} = J_{ij} + \Delta \delta_{ij}, \quad (4.52)$$

where $\Delta = \max(0, \epsilon - \lambda_1(J))$. This shift is made to ensure that the probability distribution 4.51 is integrable.

With the above prescription, $p(x, s)$ becomes:

$$\begin{aligned} p(x, s) &= p(x|s)p(s) \\ &\propto \exp\left(-\frac{1}{2}(x-s)^T \beta \tilde{J}(x-s)\right) \exp\left(\frac{1}{2}s^T \beta J s\right) \\ &= \exp\left(-\frac{1}{2}x^T \beta \tilde{J} x + s^T \beta \tilde{J} x + \frac{1}{2}s^T \beta \tilde{J} s - \frac{1}{2}s^T \beta J s\right) \\ &= \exp\left(-\frac{1}{2}x^T \beta \tilde{J} x + s^T \beta \tilde{J} x + \frac{1}{2}\beta \Delta \underbrace{s^T s}_N + \frac{1}{2}s^T \beta J s - \frac{1}{2}s^T \beta J s\right) \\ &= \exp\left(-\frac{1}{2}x^T \beta \tilde{J} x + s^T \beta \tilde{J} x + \frac{1}{2}\beta \Delta N\right) \\ &\propto \exp\left(-\frac{1}{2}x^T \beta \tilde{J} x + s^T \beta \tilde{J} x\right) \end{aligned}$$

This expression for $p(x, s)$ decouples the interactions between different spins. This gives us $p(x)$ by summing over the two possible values for sat a single site:

$$\begin{aligned}
p(x) &= \sum_{\{s\}} \exp \left(-\frac{1}{2} x^T \beta \tilde{J} x + s^T \beta \tilde{J} x \right) \\
&= \exp \left(-\frac{1}{2} x^T \beta \tilde{J} x \right) \sum_{\{s\}} \exp \left(s^T \beta \tilde{J} x \right) \\
&= \exp \left(-\frac{1}{2} x^T \beta \tilde{J} x \right) \sum_{\{s\}} \prod_{ij} \exp \left(\beta \tilde{J}_{ij} s_i x_j \right) \\
&= \exp \left(-\frac{1}{2} x^T \beta \tilde{J} x \right) \prod_{ij} \sum_{\{-1,1\}} \exp \left(\beta \tilde{J}_{ij} s_i x_j \right) \\
&= \exp \left(-\frac{1}{2} x^T \beta \tilde{J} x \right) \prod_{ij} \left(\exp \left(\beta \tilde{J}_{ij} x_j \right) + \exp \left(-\beta \tilde{J}_{ij} x_j \right) \right) \\
&= \exp \left(-\frac{1}{2} x^T \beta \tilde{J} x \right) \prod_{ij} \left(2 \cosh \left(\beta \tilde{J}_{ij} x_j \right) \right) \\
&= \exp \left(-\beta \underbrace{\left(\frac{1}{2} x^T \tilde{J} x - \frac{1}{\beta} \sum_{ij} \ln 2 \cosh \left(\beta \tilde{J}_{ij} x_j \right) \right)}_{\mathcal{H}(x)} \right).
\end{aligned}$$

The induced change onto the partition function is

$$Z_x = (2\pi)^{-N/2} (\det(\beta \tilde{J}))^{-1/2} e^{N\beta\Delta/2} Z_s \quad (4.53)$$

We arrived at a complete description of a model with a slightly different Hamiltonian and probability distribution. We can perform our normalizing flow computations in the continuous picture, and then perform the inverse map to go back to the physical Ising model. Such map is

$$p(s|x) = \prod_{i=1}^N \frac{\exp \left(\beta s_i \tilde{J} x \right)}{2 \cosh \left(\beta s_i \tilde{J} x \right)}. \quad (4.54)$$

The map is the analogous for the periodic and anti-periodic Ising models. The only difference is the coefficients of the $\tilde{J}$ matrix, since in the anti-periodic case there are some $\tilde{J}_{ij} = 1$, that come from the wall couplings.

We can then start with the periodic Ising model, sample configurations that obey to 5.3, perform the HS map and then train a NF that will try to connect the output of the NF to the probability distribution of the anti-periodic Ising model. Following the procedure outlined at 4.3.4 will give us the ratio $\frac{Z_x^A}{Z_x^P}$, that can be connected to the desired quantity,

$$\frac{Z_x^A}{Z_x^P} = e^{N\beta\frac{\Delta^A - \Delta^P}{2}} \frac{\det(\beta\bar{J}^A)^{-1/2} Z^A}{\det(\beta\bar{J}^P)^{-1/2} Z^P}. \quad (4.55)$$

5. Multicanonical algorithm

As stated previously, the two main routes to the problem at hand - computation of the free energy difference between two distinct Ising 3D models - are the non-equilibrium transformations and Multicanonical approaches.

In this chapter, we will discuss in detail three Multicanonical algorithms that can be used to compute ΔF . Since all three algorithms presuppose the extension of the wall coupling constant to continuous values between -1 and 1 (the reason for that will become clearer later), it begs to be explained first how we will make such prolongation.

Our model will be governed by the Hamiltonian

$$H_{J_{wall}}(\{s\}) = - \sum_{\langle ij \rangle} J_{ij} s_i s_j, \quad (5.1)$$

with

$$J_{ij} = \begin{cases} 1 & \text{for bulk interactions} \\ J_{ij}^w & \text{for wall interactions} \end{cases}.$$

We can define a new quantity

$$\mathcal{J}_{\text{wall}} = \frac{1}{L^2} \sum_{\text{wall}} J_{ij}^w, \quad (5.2)$$

that takes value 1(-1) for the (anti-)periodic case. Since we want to perform an extension of the coupling constant, we will have $\mathcal{J}_{\text{wall}} \in [-1, 1]$. If, for example, we want $\mathcal{J}_{\text{wall}} = J$, this can be achieved either by having all the $J_{wall} = J$ (global picture) or by having $L^2(\frac{1}{2} - \frac{J}{2L^2})$ wall couplings with value 1 and $(\frac{1}{2} + \frac{J}{2L^2})L^2$ with value -1 (local picture).

There are both merits and disadvantages with each approach. The main problem with the local picture ends up being the fact that there is a limited number of available $\mathcal{J}$'s. That and the greater similarity with our Jarzynski algorithm render our choice the global picture. In the rest of this document, we will use global updates. Nonetheless, most of the arguments presented in this chapter conform to either adoption.

5.1. General idea

Since we extended the phase space to add a degree of liberty $\mathcal{J}_{\text{wall}}$, in order to specify a microstate we need both the lattice spin configuration and the value of the wall coupling constant. We can then link to each microstate the probability

$$p(\mu = (\{s\}, \mathcal{J}_{\text{wall}})) = Z^{-1} \exp(-\beta H_{J_{\text{wall}}}(\{s\}) - f(\mathcal{J}_{\text{wall}})), \quad (5.3)$$

with f as an arbitrary function that takes only as argument the coupling constant in the wall. We retrieve the canonical case when $f(x) = 0$.

With this probability associated to a microstate, one can define the probability of measuring energy E and a coupling constant J ,

$$P(E, J) = \sum_{\mu} p(\mu) \delta(E - H(\mu)) \delta(J - J(\mu)) \quad (5.4)$$

$$= Z^{-1} \exp(S(E, J) - \beta E - f(J)). \quad (5.5)$$

We can even sum over the energies, to be left only with the probability of measuring J

$$P(J) = \sum_E P(E, J) = Z^{-1} \exp(-f(J)) \sum_E \exp(S(E, J) - \beta E) \quad (5.6)$$

$$= Z^{-1} \exp(-f(J)) Z(\beta, J), \quad (5.7)$$

and let's enforce Equation 5.7 to be a uniform distribution in the domain $J \in [-1, 1]$, i.e.

$$P(J) = K \quad (5.8)$$

$$\exp(-f(J)) Z(\beta, J) = K \quad (5.9)$$

$$\implies \exp(-f(-1)) Z(\beta, -1) = \exp(-f(1)) Z(\beta, 1) \quad (5.10)$$

$$\frac{Z(\beta, -1)}{Z(\beta, 1)} = \exp(-(f(1) - f(-1))). \quad (5.11)$$

$Z(\beta, 1)$ corresponds to the partition function of the periodic Ising model and that $Z(\beta, -1)$ corresponds to the anti-periodic one. Hence,

$$\begin{aligned} \frac{Z_A(\beta)}{Z_B(\beta)} &= \exp(-(f(1) - f(-1))) \\ \exp(-\beta(\mathcal{F}_A(\beta) - \mathcal{F}_B(\beta))) &= \exp(-(f(1) - f(-1))) \end{aligned}$$

$$\mathcal{F}_A(\beta) - \mathcal{F}_B(\beta) = \frac{f(1) - f(-1)}{\beta}, \quad (5.12)$$

so we can conclude that if we force the J histogram to be flat, we can arrive at a function f , whose values at -1 and 1 allow us to compute the free energy difference. In fact, the function itself is even more powerful, since it allows us to compute the free energy difference between two Ising models with different J_{wall} .

In the following subsections, we will discuss different ways to obtain the function $f(\mathcal{J}_{wall})$. Given that the measurements will be made at fixed β , in the rest of the text we employ the notation

$$Z(\beta, J) = \exp(S(J)), \quad (5.13)$$

and will call $S(J)$ an entropy. This denomination is chosen because we can write an extended partition function as

$$\mathcal{Z}(\beta) = \sum_J Z(\beta, J) = \sum_J e^{S(J)}. \quad (5.14)$$

5.1.1. Generic Wang Landau

Given that the problem at hand requires the attainment of a function that will render our histogram flat, the Wang-Landau (WL) algorithm [77] is a natural way to approach this issue.

Foremost, one must state that Wang Landau, despite appearing in any standard Monte-Carlo application in physics textbook, is not a Markovian algorithm, so we will not check if the algorithm follows the conditions detailed in chapter 2.

Our WL algorithm consists of the following. We start at a random spin configuration ϕ , with a correspondent $\mathcal{J}_{wall}$, considering an initial guess $\chi(J) = 0, \forall J \in [-1, 1]$ and $\epsilon = 1$. The configuration evolves ($\phi \rightarrow T(\phi)$) through a Wolff step with fixed $\mathcal{J}_{wall}$. After n such evolutions, $\tilde{\phi} = T^n(\phi)$ is subject to a change in the $\mathcal{J}_{wall}$.

The dynamics of the random walk in the $\mathcal{J}_{wall}$ are:

1. We sample δJ from a uniform distribution $(-a, a)$. We then propose $J_{i+1} = J_i + \delta J$;
2. We accept the new J with probability

$$A(J_f|J_i) = \min \left(1, \exp \left(-\beta \Delta J E_{wall} - (\chi(J_f) - \chi(J_i)) \right) \right), \quad (5.15)$$

with $\Delta J E_{wall}$ as the energy change induced on the wall by changing J .

Afterwards, we add ϵ to $\chi(J_f)$. It's in here that lies the power and the non-Markovianity of the algorithm [78]. If we visit a lot of microscopic configurations with the same J' (configurations with high $S(E, J)$) we will add ϵ to $\chi(J')$ many times. Eventually, $\chi(J')$ will be so great that,

by virtue of 5.15, the algorithm will be forced to leave J' and visit other possible wall coupling values.

Also, this step breaks detailed balance, since the microscopy reversibility is lost. The f -function changes from one step to the other, which means that if

$$A(J_f|J_i) = \exp(-\beta\Delta JE + \alpha) \quad (5.16)$$

then, if we perform the reverse step in the next update,

$$A(J_i|J_f) = \exp(\beta\Delta JE - (\alpha - \epsilon)), \quad (5.17)$$

because ϵ was added to $\chi(J_f)$.

In the end, since we will have a histogram of J 's, that follows

$$H(J) = \exp(S(J) - \chi(J)), \quad (5.18)$$

and if the histogram is flat, we can identify $\chi(J)$ with the desired $f(J)$ of equation 5.7; so we need to let J walk around the domain space following the rules explained above until the histogram is flat. However, since $\epsilon = 1$, there is a limit to the precision of the f -function. Due to that, we perform the above procedure until we have an acceptably flat histogram. Then, we update $\epsilon \rightarrow \epsilon/2$ and reiterate this process until we have an infinitesimal ϵ , small enough to do not change the f -function. In this asymptotic limit, the algorithm is in fact Markovian, since 5.16 and 5.17 are rendered equivalent and the usual MCMC properties hold.

After performing this algorithm, we have a histogram distribution $H(J)$ that follows 5.18. Ideally, this histogram will be flat. Notwithstanding, if that is not the case, we can correct the f -function

$$\begin{aligned} H(J) &= \exp(S(J) - f(J)) \\ \implies f_{\text{corr}}(J) &= f(J) + \ln(H(J)), \end{aligned} \quad (5.19)$$

which allows us to associate an error to our f -function, by replacing $H(J)$ by $H(J) + \delta H(J)$ and linearizing,

$$\delta f(J) = \frac{\delta H(J)}{H(J)} + \mathcal{O}((\delta H(J))^2). \quad (5.20)$$

5.2. Entropy competition approach

The Wang Landau philosophy is to force the histogram to be flat by changing the $f(J)$ function. However, the converse approach is also possible, i.e., to assign a fixed shape to $f(J)$ and infer the entropy properties[52].

Let's first define the microcanonical β -function

$$\beta(J) = \frac{dS(J)}{dJ}, \quad (5.21)$$

and assign to the f -function the shape

$$f(J) = \frac{m}{2}J^2 + bJ. \quad (5.22)$$

If we sample from the 5.18 distribution with this 5.22, we will have a generic histogram $H(J)$. The maximum of the distribution function is given by

$$\left. \frac{dH(J)}{dJ} \right|_{\mu} = 0, \quad (5.23)$$

that can be obtained in terms of the S and f functions,

$$\begin{aligned} \frac{d}{dJ} \left(S(J) - f(J) \right) \exp(S(J) - f(J)) &= 0 \\ \left. \frac{dS(J)}{dJ} \right|_{\mu} &= \left. \frac{df(J)}{dJ} \right|_{\mu}. \end{aligned} \quad (5.24)$$

Expanding around the maximum,

$$\begin{aligned} H(J) &= \exp \left(S(\mu) - f(\mu) + \overbrace{\frac{d}{dJ} (S(J) - f(J))}^{0 \text{ by } 5.24} (J - \mu) + \frac{1}{2} \frac{d^2}{dJ^2} (S(J) - f(J)) (J - \mu)^2 \right) \\ H(J) &\propto \exp \left(-\frac{1}{2} \left(m - \left. \frac{d^2}{dJ^2} S(J) \right|_{\mu} \right) (J - \mu)^2 \right), \end{aligned} \quad (5.25)$$

we obtain a gaussian distribution, with mean value μ and variance

$$\sigma^2 = \frac{1}{m - \left. \frac{d^2}{dJ^2} S(J) \right|_{J^*}}. \quad (5.26)$$

This behavior is clearly illustrated at figure 5.1.

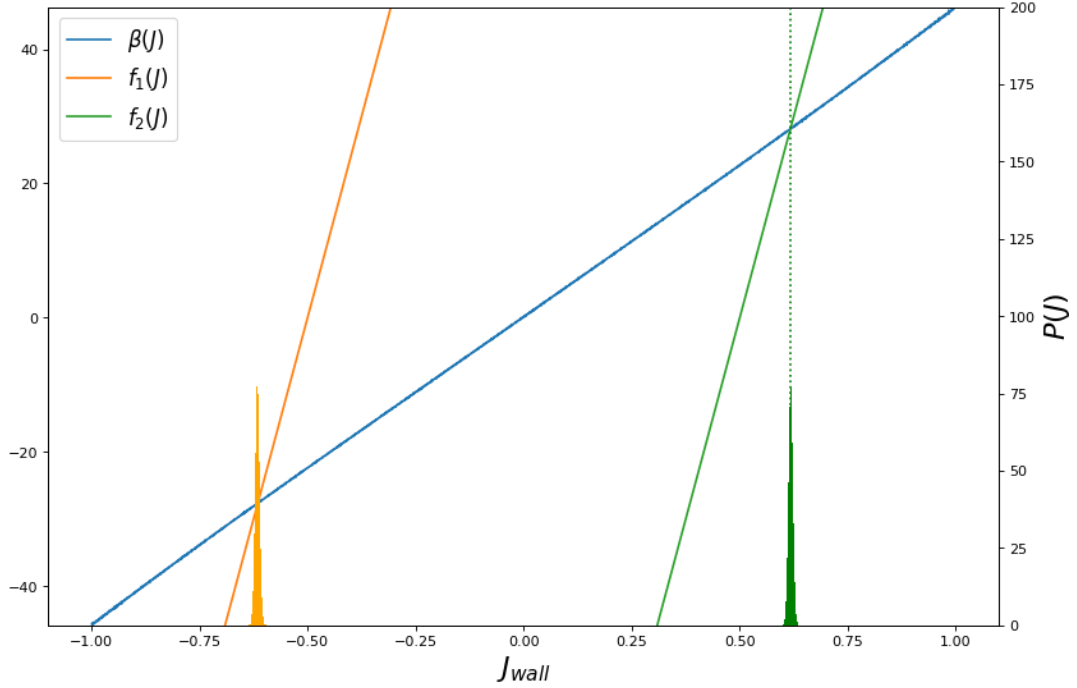


Figure 5.1.: Rough estimation for $\beta(J)$ for the system with $L_x = L_y = 16$, $L_z = 96$ and two realizations of $f(J)$, with the corresponding generated histograms. The intersection between $\beta(J)$ and $f_i(J)$ governs the localization of the histogram. Both f_1 and f_2 have the same slope m but different b 's.

Since we have a rough idea of the shape of the β -function, we can design a methodical way to find the entropy derivative at any J^* , since we can set any arbitrary array of parameters (m, b) to the f -function. For a pair (m, b) , we perform a random walk in the configuration space, changing both the spins and the wall couplings. Such random walk is governed by the acceptance probability 5.15, and will yield a histogram $H(J)$ that will follow 5.18.

With the histogram, we can compute its mean μ , and arrive, by virtue of eq. 5.24, to

$$\beta(\mu) = \left. \frac{d}{dJ} S(J) \right|_{\mu} = m\mu + b, \quad (5.27)$$

which yield the derivative at $J = \mu$.

In order to obtain the derivatives for the entirety of the domain $[-1, 1]$, we need only to tweak the parameter b in order for the intersection between $\beta(J)$ and $mJ + b$ to be at different J 's, see fig.5.1. In that sense, both μ and $\beta(\mu)$ are functions of b . Our method has an inherent error in μ , governed by the variance of the probability distribution, that propagates for the $\beta(\mu)$ computation.

Since σ decreases with m , an histogram sampling with a f -function with a very large m will

yield a very precise measurement of μ . Also, this error propagates to our value for $\beta(\mu)$, i.e., for a μ measurement with error ξ_μ we arrive at a error ξ_β ,

$$\begin{aligned} \beta &= \beta(\mu) + \xi_\beta = m(\mu + \xi_\mu) + b \\ \xi_\beta &= m\xi_\mu \end{aligned} \quad (5.28)$$

The mean value and standard deviation of ξ_β are

$$\begin{aligned} \langle \xi_\beta \rangle &= 0 \\ \langle \xi_\beta^2 \rangle &= m^2 \sigma_\mu^2 = \frac{m^2}{m - \frac{d\beta(\mu)}{dJ}} \end{aligned} \quad (5.29)$$

clearly showing that $\langle \xi_\beta^2 \rangle$ doesn't tend to zero as m increases, but instead to ∞ . The optimum value for $\langle \xi_\beta^2 \rangle$ is obtained by minimizing this value with respect to m ,

$$\frac{d\langle \xi_\beta^2 \rangle}{dm} = \frac{d}{dm} \frac{m^2}{m - \frac{d\beta(\mu)}{dJ}} = 0 \quad (5.30)$$

$$2m \left(m - \frac{d\beta(\mu)}{dJ} \right) = m^2 \wedge m \neq \frac{d\beta(\mu)}{dJ} \quad (5.31)$$

$$m = 0 \vee m = 2 \frac{d\beta(\mu)}{dJ} \quad (5.32)$$

so, if we have an estimative for the second derivative of the entropy, we can choose the m that yields the best possible estimation for the entropy derivative. Then, after performing such algorithm for an enough number of J 's it is simply a matter of integrating

$$\int_{-1}^1 \frac{dS}{dJ} dJ = S(1) - S(-1) \quad (5.33)$$

to obtain the free energy difference.

5.3. Independent bins statistical moments

An important feature of our Wang Landau approach is that it is dependent of the discretization imposed onto the J_{wall} -domain, since the function $\chi(J)$ is constant within a bin. This means that the data we obtain by sampling with $\chi(J)$ corresponds not to $\exp(S(J) - \chi(J))$ but in fact to the average over a bin,

$$H(J_i) = \int_{J_i - \frac{\Delta}{2}}^{J_i + \frac{\Delta}{2}} \exp(S(J) - \chi(J_i)) dJ, \quad (5.34)$$

and the further corrections to $S(J_i)$ with this result will take into account the density probability of the entire bin. This is usually not an issue, since $\Delta \ll 1$. However, in the extremes of the domain (near $|J| = 1$), where the function changes more sharply, and the discrepancy can be noticeable.

To circumvent this issue it is useful to measure the statistical moments of the J - distribution in each bin. Starting by expanding $S(J)$ at a single bin,

$$S(J_i + j) = S(J_i) + \left. \frac{dS}{dJ} \right|_{J_i} j + \frac{1}{2} \left. \frac{d^2 S}{dJ^2} \right|_{J_i} j^2 + \mathcal{O}(j^3), \quad (5.35)$$

taking into account up to the quadratic term, we express,

$$\begin{aligned} H(J_i) &= \int_{-\Delta/2}^{\Delta/2} dj e^{S(J_i) + aj + \frac{b}{2}j^2 - \chi(J_i)} \\ a &= \beta(J_i) \\ b &= \left. \frac{d\beta}{dJ} \right|_{J_i} \end{aligned}$$

Knowing a and b parameters leads to a continuum description in each bin. To extract that information, we measure $\langle j \rangle$ and $\langle j^2 \rangle$ inside a bin, defined as

$$\langle j^n \rangle \equiv \frac{\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} j^n \exp \left(S(J_i) + aj + \frac{b}{2}j^2 - \chi(J_i) \right) dj}{\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \exp \left(S(J_i) + aj + \frac{b}{2}j^2 - \chi(J_i) \right) dj}, \quad (5.36)$$

which are enough to uniquely determine a and b .

Also, it is possible to revise the value for $S(J_i)$, addressing the issue stated previously, since we have the integrated value $\tilde{S}$

$$\exp(\tilde{S}(J_i)) = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \exp \left(S(J_i) + aj + \frac{b}{2}j^2 \right) dj, \quad (5.37)$$

and by inverting eq. (5.37) we obtain $S(J_i)$.

6. Algorithms' efficiency study

In this chapter, we will present the results for the various algorithms we developed through this thesis, in order to figure out which of them yields more efficiently the desired ΔF values.

We start by presenting the results for the non-equilibrium approaches, starting with the Jarzynski algorithm, section 6.1, presenting the results for a three dimensional lattice size and checking the dependence of the algorithm performance with the various free parameters of our method. It is also shown how the Jarzynski algorithm yields a poor performance for smaller transition rates. Then, we show the results for the Normalizing Flows approach, section 6.2. Critically, we do not use a three dimensional lattice size. In fact, a small, two dimensional lattice is used, but the results fail to properly converge to the expected value. The problem is shown to be very similar to one that also plagues the Jarzynski approach.

Subsequently, we present the results of the various multicanonical approaches, section 6.3. The main takeaways are that the Wang-Landau, subsection 6.3.1, and entropy competition, subsection 6.3.2, approaches produce good results for the free energy computation. However, the independent bins method yields results that do not agree with the reference values, present at [14].

6.1. Jarzynski

The results presented in this section are for a system of size $18 \times 18 \times 96$, which is the smallest system in [14]. We use the same method as in [14], replacing the multi-spin coding[79] by the Wolff algorithm, see subsec. 4.2. All measurements were performed for $\beta = 0.223102$.

There are three main external parameters that we need to prescribe to the algorithm:

1. the number of samples, N_{samples} , corresponding to the amount of initial configurations that we sample from the equilibrium distribution with the initial parameters;
2. the number of steps, N_{steps} , that takes the external parameter to go from $\lambda = 0$ to $\lambda = 1$;
3. the number of Wolff sweeps, N_{sweeps} , that takes to evolve a configuration from λ_i to λ_{i+1} .

In the fig. 6.1, we analyse the dependence of the free energy difference with N_{steps} introduced to discretize the external parameter change.

As we can see in fig. 6.1, the ΔF measurements for smaller N_{steps} have a stark discrepancy with the expected results; with such difference decreasing as we increase N_{steps} . This is

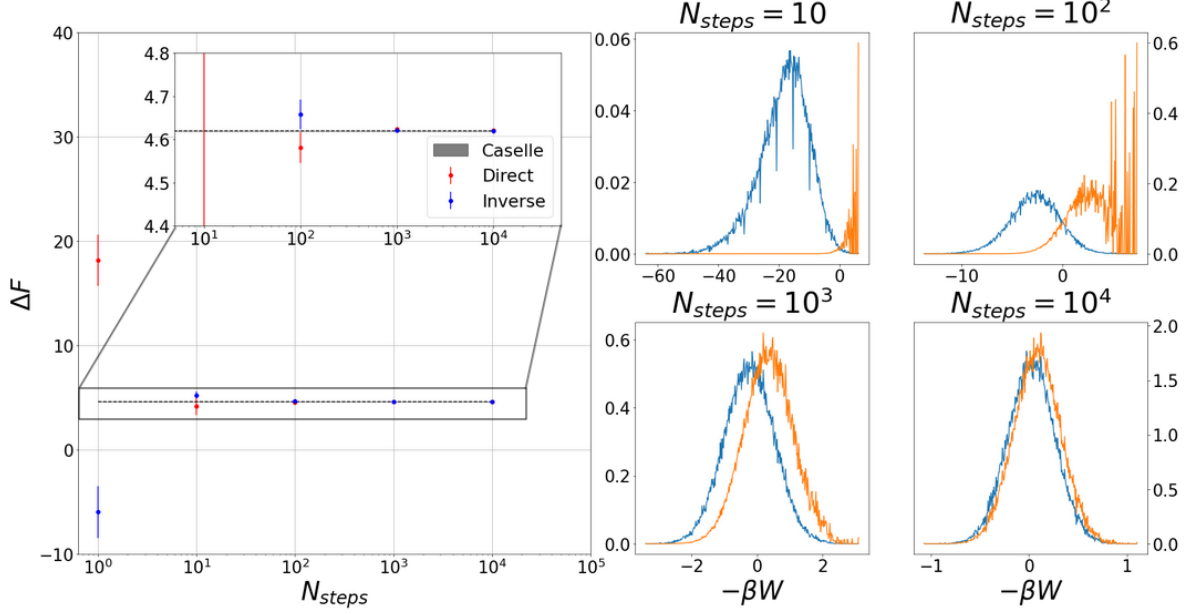


Figure 6.1.: On the left plots we have the measurement of ΔF for a sample of 50000 initial configurations and one Wolff sweep between time steps for different values of N_{steps} for the direct (periodic $\rightarrow$ anti-periodic Ising model) and inverse update. The results tend to the value presented at [14]. The right plots represent the distribution $\rho(-\beta W)$ (blue) and $e^{-\beta W}\rho(-\beta W)$ (orange), for the distinct N_{steps} for the direct update.

surprising, since the Jarzynski algorithm is independent of the speed at which we are changing the wall coupling. To understand this problem we must analyse the probability distribution of the work measurements.

Indeed, if we take a set of measurements for independent Jarzynski realizations for a fixed N_{steps} , we can have an associated distribution of $-\beta W_i$, $\rho(-\beta W)$, which allows us to compute the relevant quantity for the Jarzynski equality,

$$\langle e^{-\beta W} \rangle = \int dW e^{-\beta W} \rho(-\beta W). \quad (6.1)$$

We can see in fig. 6.1 the distribution $\rho(-\beta W)$ for different N_{steps} . We can see that rare measurements with a high value for $-\beta W$ will be disproportionally relevant to the free energy measurement. For the smaller N_{steps} 's, values of $-\beta W$ at the tail of the gaussian contribute greatly to the mean value of $e^{-\beta W}$. As we increase N_{steps} , the functions $e^{-\beta W}\rho(-\beta W)$ and $\rho(-\beta W)$ begin to overlap, which indicates a more harmonious contribution of each Jarzynski realization, i.e., the tail values don't skew the data. This rare event contamination is a non-trivial source of error, and is the reason for the poor performance of low N_{steps} measurements. The method will converge (the Jarzynski equality assures us), but needs exponentially large

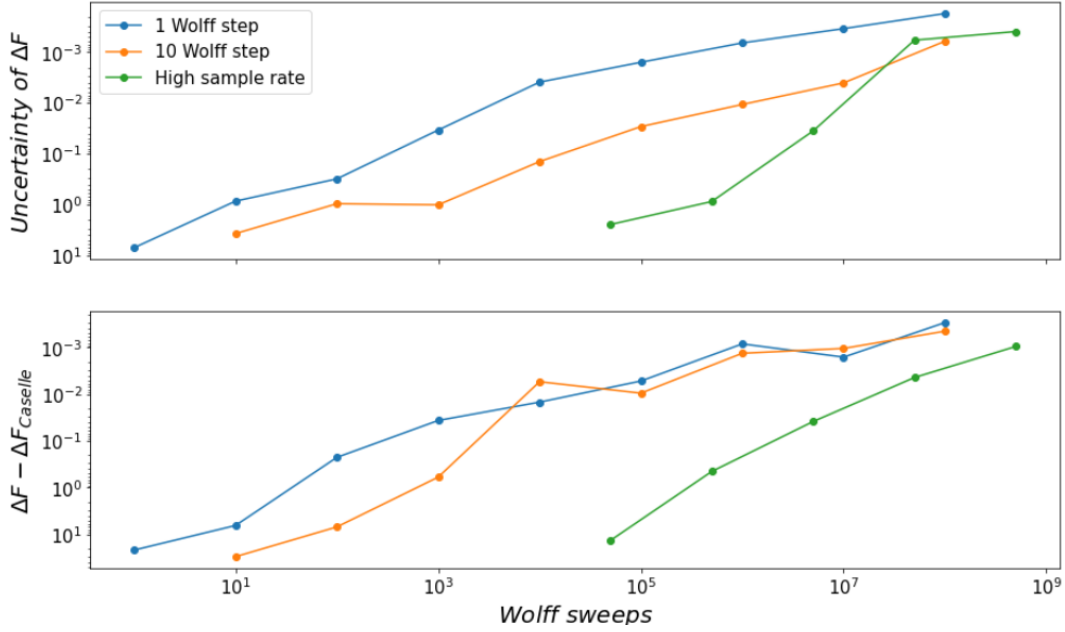


Figure 6.2.: Uncertainty and deviation from the expected result for three different approaches to the Jarzynski algorithm. The blue line corresponds to one Wolff sweep between two consecutive updates; the orange line corresponds to ten Wolff sweeps. Both of them have a low sample of initial configurations ($N_{\text{samples}} = 96$). The green line is similar to the blue approach but has greater N_{samples} .

number of samples as N_{steps} decreases.

Now that we are aware of this issue, we need to figure out the right balance between the number of transition steps, sample size and Wolff updates between each time step. The three approaches that we employ can be compared simultaneously, since the time that each method takes is governed by $\text{Wolff sweeps} = N_{\text{samples}} \times N_{\text{steps}} \times N_{\text{sweeps}}$.

The results present at fig. 6.2 show that the increase in the sample size does not yield a worthwhile increase in the precision of our measurement. The deciding factor seems to be the number of transition steps. The closer the non equilibrium transformation resembles an adiabatic process, the better results we obtain, regardless of the number of samples or the number of Wolff updates between steps.

6.2. Normalizing Flows

We implemented a Normalizing Flows architecture for a two-dimensional Ising model with size 16×16 . This is not the desired dimensionality of the system (we do not have domain walls at two dimensions), but is a toy model that can illuminate if the Normalizing Flow measurement coupled to the HS transformation can yield sensible results. The detailed architecture used in this section can be found at Appendix E.

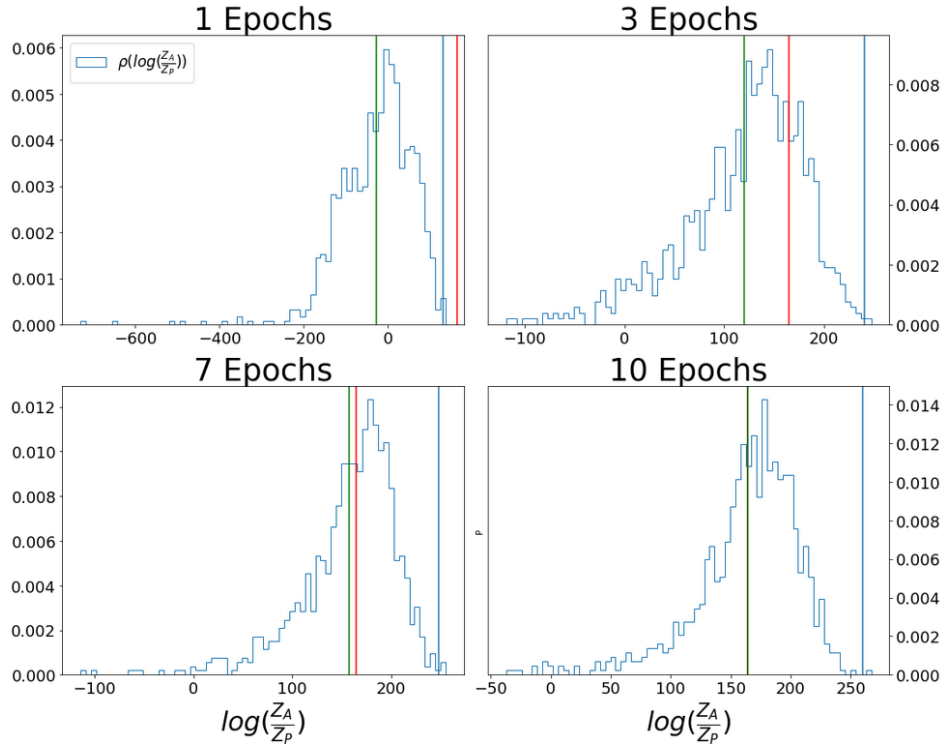


Figure 6.3.: Distribution (blue histogram) of $\log \Xi \equiv -\beta\{H(g_N(z)) - H_0(z) - Q\}$ with z sampled for the periodic Ising model for different epochs of the Normalizing Flow training. The red line corresponds to the results predicted by the Jarzynski algorithm. The blue line corresponds to $\log \langle \Xi \rangle$. There is a clear dissonance between the expected result and the obtained one, and such discrepancy is clearly induced by the presence of some Ξ with manifestly high values.

Since there are not results for the difference of free energy for the two dimensional lattice, we contrast our results with a run of Jarzynski algorithm. The data can be found at figure 6.3. It's clear to see that we have a significative (several orders of magnitude) difference between the results outputted by the Jarzynski and NF approach. The justification for that seems to be the problem of the previous section - there are a few configurations whose weight $\exp(-\beta\{H(g_N(z)) - H_0(z) - Q\})$ far exceeds in many orders of magnitude the average value, skewing the data to erroneous values.

However, there is not a clear way to circumvent this issue. Recall that for the JA, an increase of the number of transition steps solves this issue; for the NF a naive increase of the layers is not an assurance that the transformation the configuration suffers within the flow is adiabatic.

6.3. Multicanonical

The findings for the difference of free energy acquired by the multicanonical algorithms discussed in chapter 5 will be presented in this subsection. The results show, unequivocally, that, among the multicanonical algorithms, the Wang-Landau has the most success, and will be the chosen algorithm for the computation of the γ_3 coefficient.

This section will begin by summarizing the results for the Wang-Landau algorithm, subsection 6.3.1, for a system of size $32 \times 32 \times 96$, exemplifying the high degree of precision that we can achieve with this method.

Then, we present the Entropy competition, subsection 6.3.2, and independent moments' strategy findings, subsection 6.3.3, and highlighting its flaws. Indeed, entropy competition has a poor computational performance, much worse than Wang-Landau, and the independent moments' strategy yields values at the extremes of the function $f(J)$ that contradict the rest of the methods. The cause of this lack of agreement is still unclear to us.

6.3.1. Wang-Landau

Recall that the Wang Landau algorithm has two successive steps: first we obtain a rough approximation for the $\chi(J)$ -function with a standard WL procedure. As discussed before, this result is not enough, since the sampling with this $\chi(J)$ doesn't produce a flat histogram in the J 's. Then, this function is used to sample a histogram $H(J)$ that follows $\sum_E \exp(\chi(J) - \beta H(E, J))$, further correcting our data,

$$\chi_{\text{corr}}(J) = \chi(J) + \ln(H(J)). \quad (6.2)$$

For a system with size $32 \times 32 \times 96$, the rough entropy and the correction can be seen in Fig. 6.4. The correction doesn't seem relevant in a quick glance, but a zoom on a subset of J_{wall} shows that it allows for a diminution of the noise of $\frac{d\chi(J)}{dJ}$.

Must be said that our approach was not to correct $\chi(J)$ with a single histogram. For the same values of the WL rough estimate, we sample N independent histograms. We can take their average, which allows us to define a mean histogram $H_{\text{mean}}(J)$, in which each bin follows

$$H_{\text{mean}}(J_i) = \sum_n^N H_n(J_i). \quad (6.3)$$

This allows us to define an error associated with each bin: the standard deviation of the distribution $H(J_i)$, that goes as $\frac{\sigma}{\sqrt{N}}$, with σ as the error for $H(J')$. An example for mean histogram that produced the data present at fig. 6.4 can be found at fig. 6.5. The uncertainty is $\sim 10^3$ times smaller than the obtained value, and it increases near $|J| \sim 1$. Such behavior is consistent across different system sizes.

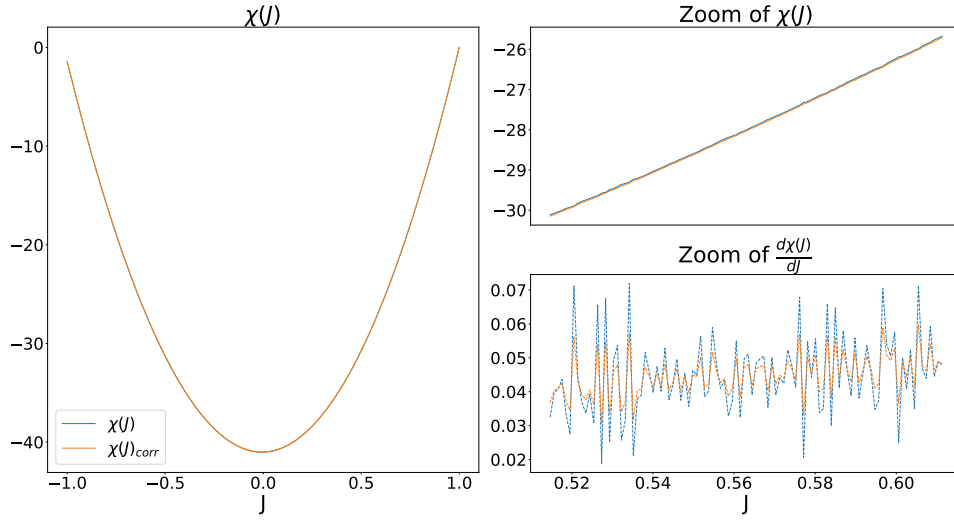


Figure 6.4.: Rough $\chi(J)$ (blue) and the corrected $\chi_{\text{corr}}(J)$ for a system with $L_x = L_y = 32$ and $L_z = 96$. This graph is obtained with a standard Wang-Landau algorithm for $\epsilon_{\text{final}} = 2^{-12}$. Both the value $\chi(1)$ and $\chi_{\text{corr}}(1)$ are fixed at zero, since the WG function is defined up to a constant.

After the correction due to 5.19, we fit the resulting function, defined at the central points of the bins, in order to extend it to the entire bin. We perform the fit with Chebyshev polynomials, since they are well defined in the interval we are interested in, $J \in [-1, 1]$. We then have an entropy, whose difference at the extremes $\chi(1) - \chi(-1)$ is the desired free energy difference (ΔF).

What is the associated uncertainty? Since

$$\delta\chi(J') = \frac{\delta H(J')}{H(J')},$$

the change between $\chi(1)$ and $\chi(-1)$ has the compounded effect of the uncertainty of each individual bin. Which implies

$$\delta\Delta F = \sqrt{\frac{1}{\varsigma} \sum_{J'} \delta\chi(J')^2}, \quad (6.4)$$

with ς as the interval of a bin.

We then have a free energy difference for this system size of

$$\Delta F = 5.97718(18). \quad (6.5)$$

This result is in good agreement with the one present at [14], and has a greater precision.

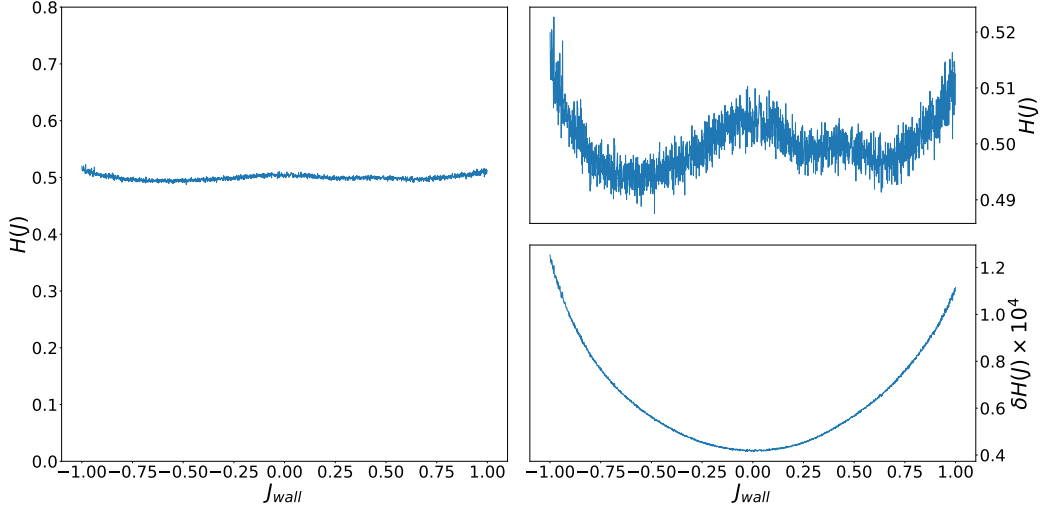


Figure 6.5.: Mean histogram $H(J)$, left plot, and the associated error for each bin, $\delta H(J)$, right-down plot, for a system with $L_x = L_y = 32$ and $L_z = 96$ with the $\chi(J)$ of Fig. 6.4.

This results were obtained with $\sim 9k$ CPU hours, which is smaller than the time we required for a bigger system with the Jarzynski algorithm. With this in mind, we will utilize this methodology to obtain the ΔF 's for various system sizes and compare the data with the theoretical prediction. We were able to perform the computation for higher values of transverse area than Cassele et al, and we have better accuracy across the board. The complete ledger of the values we obtained can be found at the Appendix F.

6.3.2. Entropy competition

Let's consider a smaller model, with $L_x = L_y = 8$ and $L_z = 16$. In order to uniquely determine a entropy competition algorithm run, we need to specify the two parameters m and b of the f -function. We have seen, equation 5.24, that the optimal parameter for m is $2 \frac{d^2}{dJ^2} S(J) \Big|_{J^*}$, so we perform a loose run of the Wang-Landau algorithm in order to work out an approximate value for the second derivative of the entropy and to estimate the range of b -values we will need to utilize in order for $\beta(J)$ to intersect $mJ + b$ at the range $J \in [-1, 1]$.

Our algorithm returns $\frac{d^2}{dJ^2} S(J) \sim 5$ for all values of J , since the function $S(J)$ is approximately quadratic. The parameter b must lie (approximately) in the interval $(-5, 5)$ for each instance of the algorithm. The results can be seen in figure 6.6a), with each point corresponding to a distinct value of b . The result is a fairly linear function.

Recall that for each instance of (m, b) , the algorithm generates a normal distribution. Each point $(J^*, \beta(J^*))$ gets the $\beta(J^*)$ by performing the mean of 10^6 uncorrelated measurements.

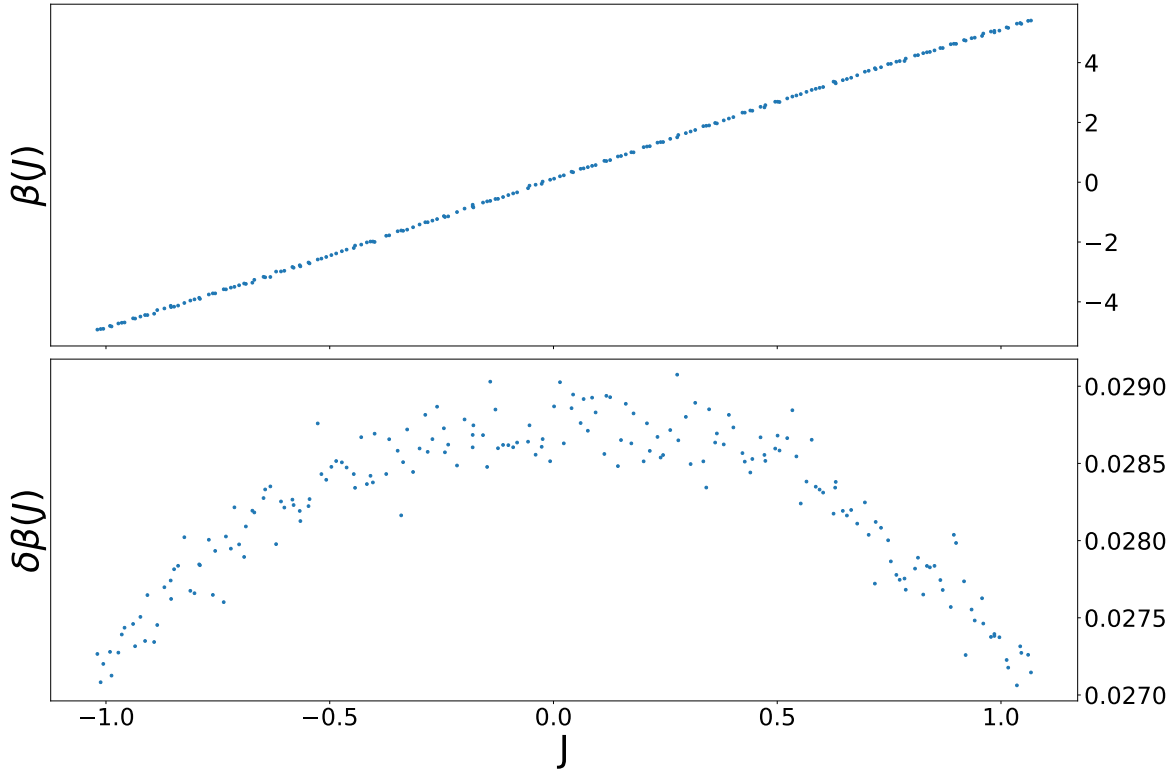


Figure 6.6.: $\beta(J)$ and corresponding error for a system with size $8 \times 8 \times 16$. The error bars are too small to be visible in the first subfigure.

The uncertainty can be seen in Fig. 6.6b) and it corresponds to the standard deviation of the corresponding gaussian distribution. The error is not constant across the measured domain. Such disparity may be caused for a change in $\frac{d^2}{dJ^2}S(J)$, since the function is not a perfect quadratic function.

We can now decompose $\beta(J)$ in Legendre polynomials[80],

$$\beta(J) = \sum_{n=0}^N a_n \mathcal{L}_n(J). \quad (6.6)$$

This choice is made because this polynomials are orthogonal

$$a_n = \int_{-1}^1 \beta(J) \mathcal{L}_n(J) dJ, \quad (6.7)$$

and $\mathcal{L}_0(J) = 1$, so we can write¹

$$\int_{-1}^1 \beta(J) dJ = \frac{a_0}{2}, \quad (6.8)$$

¹One could have chosen Chebyshev polynomials and the analysis would be similar.

which means that we can obtain the integral of $\frac{dS}{dJ}$ within the domain $[-1, 1]$ through a least squares fit for Legendre polynomials, without the need of performing a numerical integration.

This analysis allows us to predict

$$\Delta F(L_z) - \log(L_z) = 2a_0 \quad (6.9)$$

and for our system size we have

$$\Delta F(L_x = L_y = 8, L_z = 16) = 3.013 \pm 0.004. \quad (6.10)$$

Now it's only a matter of applying this rationale to the larger systems we are interested in studying. Also, note that the computation of each of the points at fig. 6.6 is independent, and follows the same algorithm, only with a different b . This, allied with the fact that the standard deviation of the gaussian distribution produced by the algorithm (an example can be seen at Fig. 5.1) goes with $\sim \frac{1}{\sqrt{N}}$, allows us to measure the time that our algorithm will take to measure $\Delta F(L_x, L_y, L_z)$. We estimate that, for $18 \times 18 \times 96$, we would need 17k CPU hours to reach the level of precision present at [14], which ranks this algorithm as slightly worse than the Jarzynski one.

6.3.3. Independent bins momentum

Let's now analyse the results from the independent bins momentum approach. Recall from Eq. 5.37 that we have a relation between the moments of J within each bin and the true entropy function $S(J)$. Since we expanded $S(J)$ around the mean point (J^*) of the bin up to $(J - J^*)^2$, we have two unknowns, which means that we measure $\langle j \rangle$ and $\langle j^2 \rangle$ in order to fully determine $\left. \frac{dS}{dJ} \right|_{J^*}$ and $\left. \frac{d^2S}{dJ^2} \right|_{J^*}$, with $\langle j^n \rangle$ given by Eq. 5.36.

Since we are considering small bins within the interval between -1 and 1 , we can assume that the distribution $\exp\left(aj + \frac{1}{2}bj^2\right)$ in the bin is a perturbation around an uniform distribution, so one would expect that the $\langle j^2 \rangle$ would be close to $-\frac{\Delta^3}{12}$. We can see an example of values for $L_x = L_y = 32$ and $L_z = 96$ in figure 6.7.

So now remains only to solve the two equations for a and b . The approach used was the inversion of a perturbative series, i.e., we promote $a \rightarrow \alpha\delta$ and $b \rightarrow \beta\delta^2$, with δ as a perturbative parameter. We assign to b a smaller value, informed by the fact that the second derivative must have a much smaller contribution than the linear component of the distribution within the bin. We will expand the equations for μ_1 and μ_2 with respect to δ up to order δ^6 .

This expansion organizes the series in order of the most relevant parameters, i.e.,

$$\langle j^n \rangle = \sum_{i=1}^5 c_i^n(\alpha, \beta) \delta^i + \mathcal{O}(\delta^6). \quad (6.11)$$

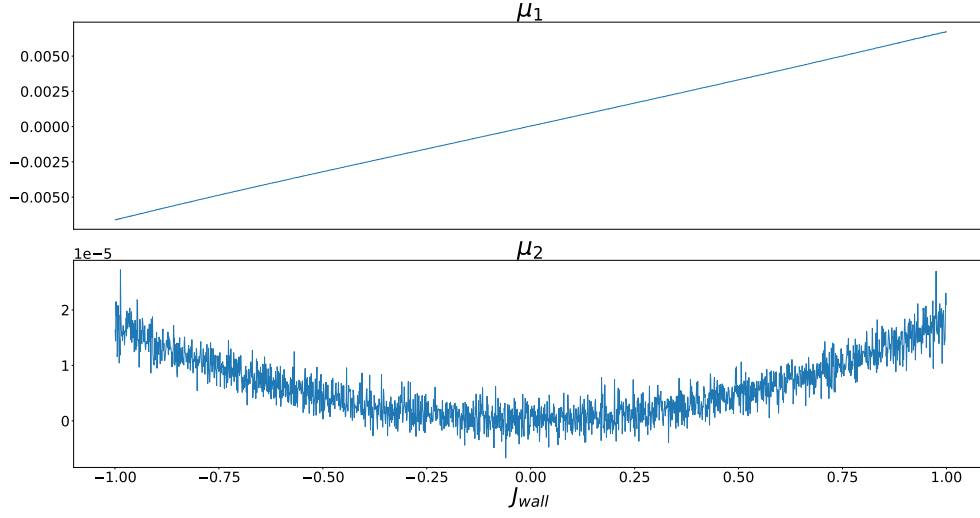


Figure 6.7.: $\mu_1 \equiv \langle j \rangle$ and $\mu_2 \equiv \langle j^2 \rangle - \frac{\Delta^3}{12}$ for $L_x = L_y = 32$ and $L_z = 96$

which allows us to recognize the most relevant contributions for the moments calculation, and, more importantly, the series 6.11 can be inverted to obtain a series for α and β

$$\begin{aligned}\alpha &= \sum_{i=1}^5 d_i(\mu_1, \mu_2) \delta^i + \mathcal{O}(\delta^6) \\ \beta &= \sum_{i=1}^5 e_i(\mu_1, \mu_2) \delta^i + \mathcal{O}(\delta^6) \end{aligned} \quad (6.12)$$

The error can be upper-bounded by the last coefficient of the 6.12 series,

$$\begin{aligned}\delta\alpha &= d_5 \\ \delta\beta &= e_5 \end{aligned} \quad (6.13)$$

These coefficients can be calculated numerically, which allows us to obtain

$$d_1 = 12\mu_1 \quad (6.14)$$

$$d_2 = 0 \quad (6.15)$$

$$d_3 = \frac{144}{5} \mu_1 (3\mu_1^2 - 5\mu_2) \quad (6.16)$$

$$d_4 = 0 \quad (6.17)$$

$$d_5 = \frac{1728}{175} \mu_1 (33\mu_1^2 - 35\mu_2) (3\mu_1^2 - 5\mu_2) \quad (6.18)$$

and

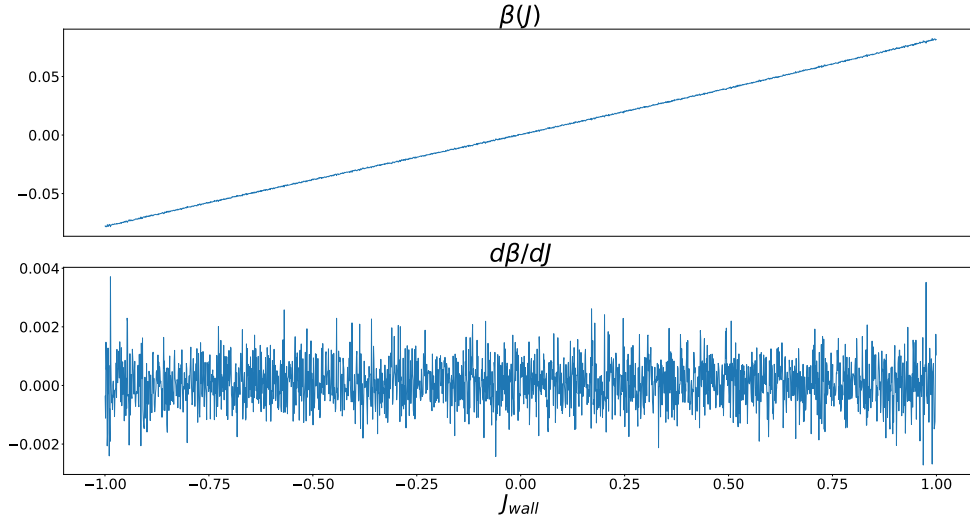


Figure 6.8.: Results for $\frac{dS}{dJ}$ and $\frac{d^2S}{dJ^2}$ for $L_z = 96$ and $L_x = L_y = 32$.

$$e_1 = 0 \quad (6.19)$$

$$e_2 = 72(-2\mu_1^2 + 5\mu_2) \quad (6.20)$$

$$e_3 = 0 \quad (6.21)$$

$$e_4 = -\frac{432}{35}(108\mu_1^4 - 280\mu_1^2\mu_2 + 125\mu_2^2) \quad (6.22)$$

$$e_5 = 0. \quad (6.23)$$

With this perturbative series, we can map (μ_1, μ_2) to $(\frac{dS}{dJ}, \frac{d^2S}{dJ^2})$, such that we can compute $\int_{-1}^1 \frac{dS}{dJ}(J)dJ$ and obtain the free energy difference.

Using the data previously presented in fig. 6.7 and the perturbative series 6.12, we obtain the results present in fig. 6.8. We can see that the function for $\beta(J)$ is very similar to μ_1 , which is not unexpected since the leading order in δ for a is $12\mu_1$. Also, $\frac{d\beta}{dJ} \ll \beta$.

Furhter recall that we have a relation between the effective entropy at the center of the bin $\tilde{S}(J_i)$ and the “true” entropy $S(J_i)$,

$$\exp(\tilde{S}(J_i) - \chi(J_i)) = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \exp\left(S(J_i) - \chi(J_i) + aj + \frac{b}{2}j^2\right) dj, \quad (6.24)$$

which we can invert with a similar approach as the outlined above, that gives us a perturbative series expansion for the corrected entropy,

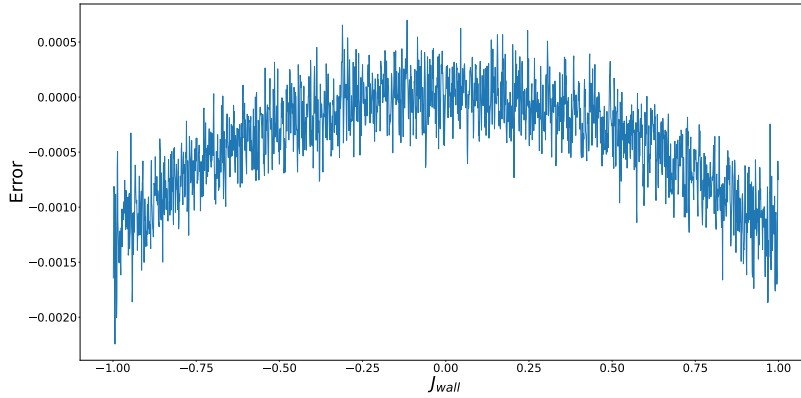


Figure 6.9.: Difference between the two different ways to compute $\frac{dS}{dJ}$ (corrected Wang-Landau and IBM).

$$S(J_i) = \tilde{S}(J_i) - 15\mu_2 - \frac{18}{7} (12\mu_1^4 - 25\mu_2^2). \quad (6.25)$$

This result is not necessary to compute the free energy, since we already have $\beta(J)$. However, we can draw a comparison between the integration of the $\beta(J)$ we obtain with the independent bins (IBM) approach and the one obtained (and further corrected with eq. (6.25)) with the Wang-Landau algorithm for the system with the same size. We check that the local derivatives do not correspond to the same values (see fig 6.9). This error compounds through the domain,

$$\Delta(\Delta F) = \int_{-1}^1 (\beta_{\text{WL}}(J) - \beta_{\text{IBM}}(J)) dJ, \quad (6.26)$$

and skew the data obtained purely from the IBM.

The reason behind this errors may be linked to the fact that we are expanding the entropy up to the second derivative. If we assigned the magnitude orders δ and δ^2 to the first and second derivative, the effects of the third derivative will be felt starting at δ^3 , which is much sooner than the precision of δ^5 that we claimed previously. Further study is required to check if this is indeed the case or the source of error comes from another unidentified issue. Such analysis will not be present in this document, since the data we collected to study the IB approach was obtained concurrently with the Wang-Landau data, and, since that approach yielded good results, we didn't pursue this issue any further.

7. Data vs NG expansion

Now that we have narrowed down the best method, the Wang Landau multicanonical algorithm, for free energy computation, we can now compute the interface free energy for various system sizes, since we are interested in the dependence of ΔF with respect to the transverse area. In this chapter we will start by presenting the results for a system with fixed $L_z = 96$, section 7.1; then we will check how can we distinguish between the universal Nambu-Goto contributions and the particular contributions of this model, section 7.2, and will finish this thesis by computing the c_{-3} expansion term and comparing it with previous results from the literature.

7.1. From partition ratios to interface free energy

It is known, see subsec. 6.3.1, that our algorithm of choice will return the ratio Z_A/Z_P between the periodic and anti-periodic Ising models. We need to bridge this quantity to the interface free energy. Indeed, let's assume that we have a finite L_z and the $L_x, L_y \rightarrow \infty$ limit. This, in the antiperiodic system, leads to the appearance of a single domain wall, given that the free energy scales with the area, and allows us to define the energy of the interface [81, 14] as¹

$$F^{(1)} = -\ln \left(\frac{Z_A}{Z_P} \right) + \ln L_z, \quad (7.1)$$

with the $\ln L_z$ term arising from the fact that the interface may exist at any point alongside the z -direction. We are not, unfortunately, in the $L_x, L_y \rightarrow \infty$ limit, and, henceforth, need to consider the possibility of an odd number of domain walls, see subsection 3.1.2. Since that is a possibility, we can arrive at an expression for Z_A [81] that assumes non-interacting and indistinguishable interfaces,

$$Z_A = \sum_n \frac{1}{n!} \exp [-n (F_{\text{wall}} - \ln L_z)], \quad (7.2)$$

with the summation running over the odd integers. The expression above is easily generalized to the periodic case. Recall that we allow the existence of multiple domain walls in the

¹Free energy is usually defined as $\mathcal{F} = -\frac{\ln Z}{\beta}$. However, we will not take into account the role of β - that factor will be absorbed into $F^{(1)}$ and $F^{(2)}$

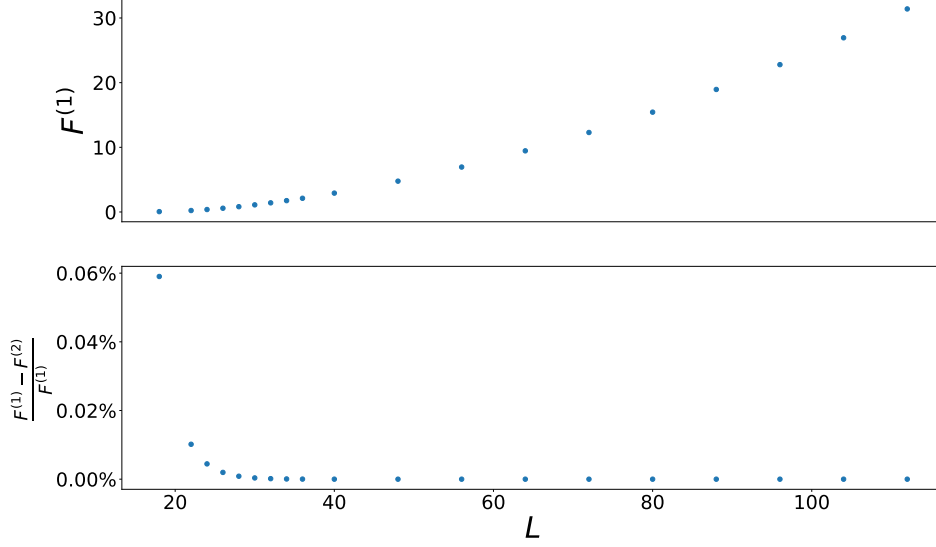


Figure 7.1.: $F^{(1)}(L)$ (up) and $F^{(1)}(L) - F^{(2)}(L)$ (down) for a system with $L_z = 96$, $\beta = \beta_{\text{chosen}}$ and $L_x = L_y = L$.

$J_{\text{wall}} = -1$ scenario, but they need to appear an odd number of times, due to the anti-periodic boundary conditions. The rationale behind the appearance of domain walls in the periodic Ising model is the same, but now we sum over the even integers, since we do not consider that same spin orientation among the spins at $z = L_z$ and $z = 0$ constitutes a domain wall. Then,

$$\frac{Z_A}{Z_P} = \frac{\sum_m \frac{1}{2m+1!} \exp[-(2m+1)(F_{\text{wall}} - \ln L_z)]}{\sum_m \frac{1}{2m!} \exp[-2m(F_{\text{wall}} - \ln L_z)]}$$

and one can identify in the numerator and the denominator with the Taylor series of sinh and cosh, respectively. By virtue of that, we can write

$$\frac{Z_A}{Z_P} = \tanh \exp[-(F_{\text{wall}} - \ln L_z)], \quad (7.3)$$

and arrive to a new definition for the domain wall energy,

$$F^{(2)} = -\ln \operatorname{arctanh} \left(\frac{Z_A}{Z_P} \right) + \ln L_z. \quad (7.4)$$

We can compare both quantities, $F^{(1)}$ and $F^{(2)}$ for a system with variable transverse area. We can see at Fig. 7.1 that the distinction between both quantities is smaller than the error bar in the studied system sizes, in particular for high transverse area systems.

The data for the various system size parameters we explored can be found at fig. 7.2.

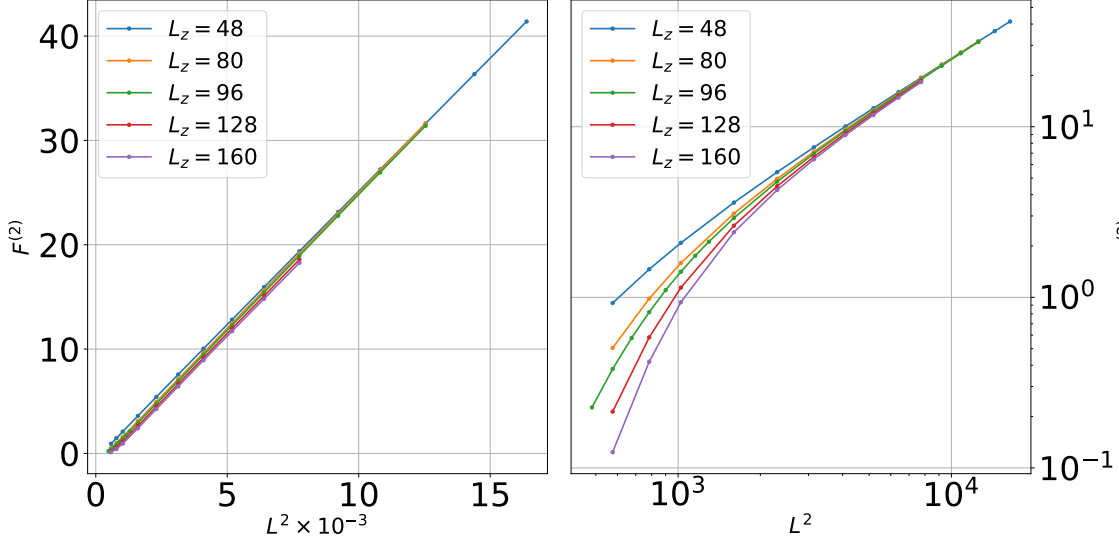


Figure 7.2.: $F^{(2)}(A = L^2)$ for various L_z values. The transversal area has $L_x = L_y = L^2$. Both graphs represent the same data but the second one is in a log-log scale.

7.2. Universal and non-universal contributions

Let's now figure out if the data we obtained, present at Fig. 7.2, is well described by Nambu-Goto at leading order. Recall that the partition function ratio, if we consider only the NG action, is, up to $\frac{1}{(L_x L_y)^4}$,

$$F^{\text{NG}} = 0.3915 - \frac{1}{2} \log(\sigma L_z^2) + A - \frac{0.2500}{A} + \frac{0.0141}{A^2} + \frac{0.1314}{A^3} - \frac{2.0303}{A^4}, \quad (7.5)$$

with $A = \sigma L_x L_y$. Note that we remain yet to specify how we define the value of the interface tension σ . Since the leading order at Eq. 7.5 is $\sigma L_x L_y$, one would assume that if we measure the free energy for large enough transverse areas we can perform a linear fit and arrive at σ . However, such approach has to be taken with a grain of salt. In fact, when we increase $L_x L_y$, if the value of L_x or L_y is much greater than L_z , the domain wall can be self-interacting, in the sense that its fluctuations will have greater amplitude than L_z . That introduces non-trivial contributions to the free energy, and the $\sigma L_x L_y$ behavior no longer holds.

Since it's not clear to pinpoint for which transverse area we start to suffer from self-interaction, we will measure σ for a fixed L_z in a different fashion. Let's remove the universal NG contribution from Eq. 7.5, with σ as a free parameter. Since we know that the non-universal terms only appear, at most, at order $\frac{1}{A^3}$, we expect

$$\Delta F - F^{\text{NG}}(\sigma) = \frac{c-3}{A^3} + \frac{c-4}{A^4}, \quad (7.6)$$

with σ once again as a free parameter. However, the theoretical terms are well defined and fixed by our theory, so if we do not choose the appropriate σ , there will appear terms in Eq. 7.6 at lower orders than $\frac{1}{A^3}$. If we have a good guess for σ , we can explore the neighborhood of such guess and verify which value corresponds to the best fit for $\Delta F - F^{\text{NG}}$ when we only allow degrees of freedom for the c_{-3} and c_{-4} terms. The proposed range for σ is informed by [14], and we choose as the fit metric the mean square error between the fit and our data. This parameter has a natural minimum, that will be defined as our $\sigma(L_z = 96)$, which allows us to obtain, see figure 7.3, $\sigma = 0.00260249(2)$ for $L_z = 96$, which is close to Caselle's value of $\sigma = 0.0026083(6)$.

With this choice for σ , we are able to remove the NG contributions and to perform the fit to $F^{(2)} - F^{\text{NG}}$. We present the data and the fit at Fig. 7.3. $\Delta F - F^{\text{NG}}$ fails to obey to Eq. 7.6 when $L > \frac{L_z}{2}$. One of the reasons may be the self-interaction of the domain wall. That suggests that we may perform the σ -finding and fit for the non - universal data for the data up to a certain cutoff square transversal area size, L_z^C .

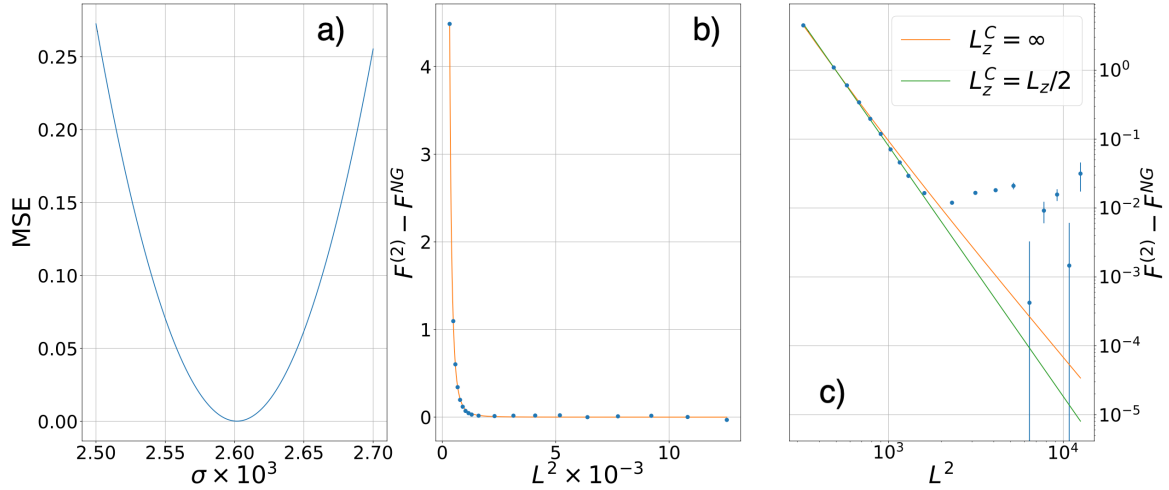


Figure 7.3.: MSE with respect to proposed σ 's (a)). $F^{(2)}$ data for various square transversal areas with $L_z = 96$ and correspondent fit. b) and c) present the same data, but with a regular and log-log plot, respectively

7.3. c_{-3} computation

Now that we have the data for a various L_z 's and a concrete way of obtaining the $\sigma(L_z)$, we can try to understand if the obtained values will be connected through some kind of finite size scaling. Unfortunately, the data doesn't support this idea. Nonetheless, we can still utilize the σ -value to remove the universal part and fit the remaining functions to Eq. 7.6. Taking into account the data for systems with $L < L_z/2$, we arrive at values for c_{-3} and c_{-4} for each

L_z	c_{-3}	c_{-4}
48	0.655(3)	1.339(7)
80	0.467(5)	2.05(1)
96	1.36(1)	1.13(1)
128	0.93(3)	2.23(9)
160	1.5(1)	2.1(3)

Table 7.1.: Expansion coefficients for the non-universal dependence of the interface free energy.

distinct L_z . The fits are represented at Fig. 7.4 and the data for the expansion coefficients is present at Table 7.1.

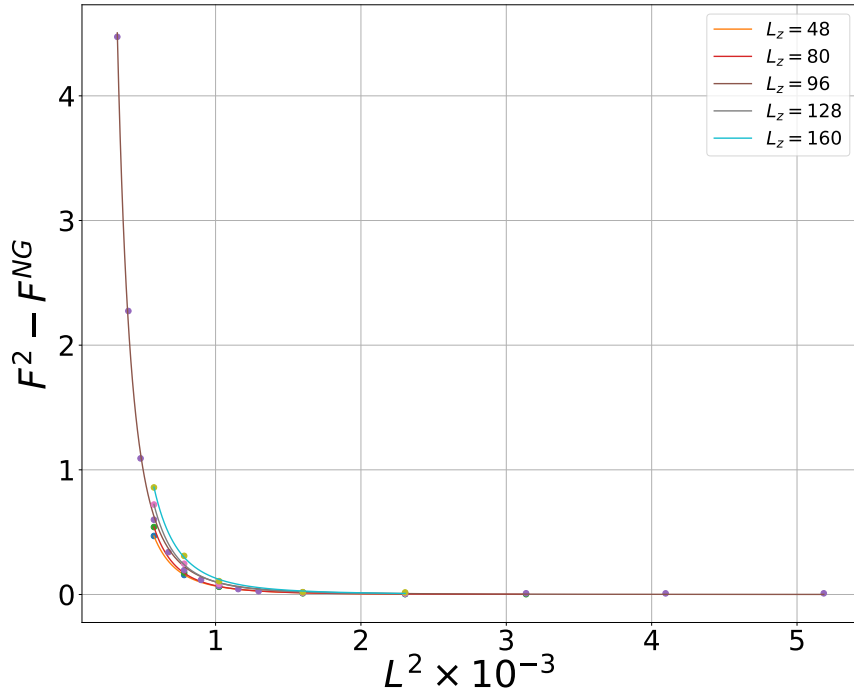


Figure 7.4.: Dots correspond to our non-universal data for the various system sizes. The continuous lines are the fit obtained with Eq. 7.6.

Once again, we obtain a set of data that do not exhibit a clear finite size scaling that may explain the disparity of the coefficient values. Nonetheless, there is a known bound [35] for the value of c_{-3} ,

$$c_{-3} \geq -0.377, \quad (7.7)$$

and we can see that all the values we got obey to Eq. 7.7.

8. Conclusions and Outlook

In this thesis we presented and studied the problem of the precise computation of the 3D Ising interface free energy, through a myriad of different algorithms, allowing us to compute new ΔF values for previously unchecked lattice sizes and to obtain the first non-universal expansion parameter.

The road that led up to this was long and winding. We started with the analytical study of the partition function of the NG model, chapter 1, which allowed us to work out the general contributions for the free energy expansion in powers of $\frac{1}{L_x L_y}$.

After a review of generic Monte Carlo algorithms, chapter 2, we studied some phenomenological properties of the domain wall for Ising models with continuous wall coupling, chapter 3. We showed that the interface undergoes delocalization at $J_{\text{wall}} = -1$, but were unable to conclusively demonstrate if such transition is in fact a KT one.

After exploring the Ising domain wall properties, we moved on to the main event: the development of new algorithms able to compute the ratio between the partition function of the periodic Ising model and the anti-periodic one. Such algorithms were classified in two major classes. The first one, chapter 4, corresponds to the non-equilibrium methods. We modify the Jarzynski algorithm presented at [14] by studying the underlying evolution of the system from multispin coding to Wolff dynamics. Despite the hopes that a cluster update would improve the simulations, we demonstrated, Fig. 6.1, that the non-equilibrium transformation from one Ising model to the other is prone to rare event contamination of the average $\langle e^{-\beta W} \rangle$. From the Jarzynski, we proceeded to the free energy computation with aid of Normalizing Flows, that belong to the same universal, SNF class. We showed how this fairly recent[34] idea works in a small system of two interacting particles, and combined NF ideas with a HS transformation of the Ising model to develop a new algorithm. We showed, however, that the substantial increase of the degrees of freedom and the need to perform a HS transformation in order to circumvent the discrete degrees of freedom problem give rise to a rare event contamination of the data. Further work is necessary to discern where, exactly, lies the problem and if new NF architectures can be used to remedy the issue.

Subsequently, chapter 5, we approached the problem of the partition ratios through a Multicanonical outlook. We developed three distinct Multicanonical algorithms. One of them, the independent bins approach, was shown to fail to accurately predict the f -function derivative at the points near the frontiers of our domain. The reasons of that failure are yet not clear to

us and further studies are necessary in order to understand where the error is coming from. The remaining two, the entropy competition and the Wang-Landau, exhibited good results, so we employed the most efficient one, WL, to make our precise measurements.

Finally, chapter 7, we analyze the free energy expansion for square transverse areas for different L_z 's. We showed that the leading order coefficients have a good agreement with the theoretical predictions. There is, however, a lack of agreement between the first non-universal coefficient between distinct L_z s, which taints our results. New measurements are required for systems with different ratios $\frac{L_x}{L_y}$ in order to infer any concrete conclusion.

A. Markov Matrix properties

It can be proven that the right eigenvalues of a Matrix matrix have norm less or equal to 1.

First we need to recall that the matrix admits left and right eigenvalues

$$M_{ij}r_j^{(\alpha)} = \lambda_\alpha r_i^{(\alpha)}$$

$$l_i^{(\beta)} M_{ij} = \lambda_\beta l_j^{(\beta)}$$

that can be show to obey to the orthogonality condition

$$\begin{aligned} l_i^{(\beta)} M_{ij} r_j^{(\alpha)} &= l_i^{(\beta)} M_{ij} r_j^{(\alpha)} \\ \lambda_\alpha l_i^{(\beta)} r_i^{(\alpha)} &= \lambda_\beta l_i^{(\beta)} r_i^{(\alpha)} \\ 0 &= (\lambda_\beta - \lambda_\alpha) l_i^{(\beta)} r_i^{(\alpha)}, \end{aligned}$$

so that

$$\lambda_\alpha = \lambda_\beta \quad l_i^{(\beta)} r_i^{(\alpha)} = \delta^{\alpha\beta}.$$

We can now choose a vector $L = \{1, \dots, 1\}$, that is a left eigenvector,

$$\sum_i L_i M_{ij} = \underbrace{\sum_i M_{ij}}_{\text{by 2.9}} = 1 = L_j,$$

so that we can always have a eigenvalue equal to one. Now we just need to show that the remaining eigenvalues obey to $|\lambda| \leq 1$.

Consider a generic left eigenvector $l^{(\alpha)}$, so that

$$\exists k \quad \forall i \quad |l_k^{(\alpha)}| \geq |l_i^{(\alpha)}|,$$

which entails for any i ,

$$\begin{aligned} |\lambda_\alpha l_i^{(\alpha)}| &= \left| \sum_j l_j^{(\alpha)} M_{ji} \right| \\ &\leq \left| \sum_j l_j^{(\alpha)} \right| M_{ji} \\ &\leq \left| \sum_j l_k^{(\alpha)} \right| M_{ji} \leq \left| \sum_j l_k^{(\alpha)} \right|, \end{aligned}$$

and in particular for $i = k$, we arrive at

$$|\lambda_\alpha| \leq 1.$$

B. Statistical Iterators

There is an inherent numerical instability associated with the computation of statistical moments of distributions with an entire dataset. When we are able to generate new statistic from sampling algorithms, it's advised to update the statistical moments as we go through the dataset.

In this section I present and derive various statistical iterators that were used through this work.

B.1. Average

We can define mean as:

$$X(N) = \frac{1}{N} \sum_{i=0}^N x_i \quad (\text{B.1})$$

Knowing that:

$$\begin{aligned} X(N+1) &= \frac{1}{N+1} \sum_{i=0}^{N+1} x_i \\ &= \frac{1}{N+1} \sum_{i=0}^N x_i + \frac{1}{N+1} x_{N+1} \end{aligned}$$

Multiplying by $\frac{N}{N}$:

$$\begin{aligned} X(N+1) &= \frac{N}{N+1} \underbrace{\left(\frac{1}{N} \sum_{i=0}^N x_i \right)}_{X(N)} + \frac{1}{N+1} x_{N+1} \\ \implies X(N+1) &= \frac{N+1-1}{N+1} X(N) + \frac{1}{N+1} x_{N+1} \\ &= X(N) + \frac{1}{N+1} (x_{N+1} - X(N)) \end{aligned} \quad (\text{B.2})$$

B.2. Variance

We can express variance as:

$$\text{var}(N) = X_2(N) - X_1(N)^2, \quad (\text{B.3})$$

in which

$$X_2(N) = \frac{1}{N} \sum_{i=0}^N x_i^2 \quad (\text{B.4})$$

$$X_1(N) = \frac{1}{N} \sum_{i=0}^N x_i. \quad (\text{B.5})$$

One can compute:

$$\begin{aligned} \text{var}(N+1) &= X_2(N+1) - X_1(N+1)^2 \\ &= \frac{1}{N+1} \sum_{i=0}^{N+1} x_i^2 - X_1(N+1)^2 \\ &= \frac{1}{N+1} \sum_{i=0}^N x_i^2 + \frac{1}{N+1} x_{N+1}^2 - X_1(N+1)^2 \\ &= \frac{N}{N+1} \left(X_2(N) - X_1(N)^2 + X_1(N)^2 \right) + \frac{1}{N+1} x_{N+1}^2 - X_1(N+1)^2 \\ &= \frac{N}{N+1} \text{var}(N) + \frac{N}{N+1} X_1(N)^2 + \frac{1}{N+1} x_{N+1}^2 - X_1(N+1)^2 \\ &= \text{var}(N) - \frac{1}{N+1} \text{var}(N) + \frac{N}{N+1} X_1(N)^2 + \frac{1}{N+1} x_{N+1}^2 - X_1(N+1)^2 \\ &= \text{var}(N) + \frac{1}{N+1} (-\text{var}(N) + \xi), \end{aligned}$$

with

$$\begin{aligned} \xi &= NX_1(N)^2 + x_{N+1}^2 - (N+1)X_1(N+1)^2 \\ &= (N+1) \left(X_1(N)^2 - X_1(N+1)^2 \right) + x_{N+1}^2 - X_1(N)^2 \\ &= (N+1) \left(X_1(N) - X_1(N+1) \right) \left(X_1(N) + X_1(N+1) \right) + \left(x_{N+1} - X_1(N) \right) \left(x_{N+1} + X_1(N) \right). \end{aligned}$$

Replacing $X_1(N+1)$ by the value of the iterator at [B.2](#) yields

$$\begin{aligned}
&= (N+1) \left(X_1(N) - X(N) - \frac{1}{N+1} (x_{N+1} - X(N)) \right) (X_1(N) + X_1(N+1)) \\
&+ (x_{N+1} - X_1(N)) (x_{N+1} + X_1(N)) \\
&= - (x_{N+1} - X(N)) (X_1(N) + X_1(N+1)) + (x_{N+1} - X_1(N)) (x_{N+1} + X_1(N)) \\
&= (x_{N+1} - X_1(N)) (x_{N+1} + X_1(N) - X_1(N) - X_1(N+1)) \\
&= (x_{N+1} - X_1(N)) (x_{N+1} - X_1(N+1)),
\end{aligned}$$

allowing us to finally write

$$var(N+1) = var(N) + \frac{1}{N+1} \left(-var(N) + (x_{N+1} - X_1(N)) (x_{N+1} - X_1(N+1)) \right). \quad (\text{B.6})$$

B.3. Correlation

Let's take as starting point the expression for covariance with lag τ :

$$cov_\tau(n) = \frac{1}{n-\tau} \sum_{i=0}^{n-\tau} x_i x_{i+\tau} - \left(\frac{1}{n-\tau} \sum_{i=0}^{n-\tau} x_i \right) \left(\frac{1}{n-\tau} \sum_{i=0}^{n-\tau} x_{i+\tau} \right) \quad (\text{B.7})$$

In order to clean the notation, let's define:

$$X_2(n) = \frac{1}{n-\tau} \sum_{i=0}^{n-\tau} x_i x_{i+\tau} \quad (\text{B.8})$$

$$X_1^\tau(n) = \frac{1}{n-\tau} \sum_{i=0}^{n-\tau} x_{i+\tau}. \quad (\text{B.9})$$

In a way that the covariance expression is

$$cov_\tau(n) = X_2(n) - X_1^\tau(n) X_1^0(n). \quad (\text{B.10})$$

Let's now try to find the expression for $cov_\tau(n+1)$,

$$cov_\tau(n+1) = \frac{1}{n+1-\tau} \sum_{i=0}^{n+1-\tau} x_i x_{i+\tau} - \left(\frac{1}{n+1-\tau} \sum_{i=0}^{n+1-\tau} x_i \right) \left(\frac{1}{n+1-\tau} \sum_{i=0}^{n+1-\tau} x_{i+\tau} \right).$$

In order to simplify, let's define $m \equiv n - \tau$:

$$\begin{aligned}
cov_\tau(n+1) &= \frac{1}{m+1} \sum_{i=0}^{m+1} x_i x_{i+\tau} - \left(\frac{1}{m+1} \sum_{i=0}^{m+1} x_i \right) \left(\frac{1}{m+1} \sum_{i=0}^{m+1} x_{i+\tau} \right) \\
&= \frac{m}{m+1} X_2(n) + \frac{1}{m+1} x_{m+1} x_{m+1+\tau} - \left(\frac{1}{m+1} \sum_{i=0}^{m+1} x_i \right) \left(\frac{1}{m+1} \sum_{i=0}^{m+1} x_{i+\tau} \right) \\
&= X_2(n) - \frac{1}{m+1} X_2(n) + \frac{1}{m+1} x_{m+1} x_{m+1+\tau} - X_1^\tau(n+1) X_1^0(n+1) \\
&= X_2(n) - X_1^\tau(n) X_1^0(n) + X_1^\tau(n) X_1^0(n) - \frac{1}{m+1} X_2(n) \\
&\quad + \frac{1}{m+1} x_{m+1} x_{m+1+\tau} - X_1^\tau(n+1) X_1^0(n+1) \\
&= cov_\tau(n) + X_1^\tau(n) X_1^0(n) - \frac{1}{m+1} X_2(n) + \frac{1}{m+1} x_{m+1} x_{m+1+\tau} - X_1^\tau(n+1) X_1^0(n+1) \\
&= cov_\tau(n) + \frac{1}{m+1} \left[(m+1) X_1^\tau(n) X_1^0(n) - X_2(n) \right. \\
&\quad \left. + x_{m+1} x_{m+1+\tau} - (m+1) X_1^\tau(n+1) X_1^0(n+1) \right] \\
&= cov_\tau(n) + \frac{1}{m+1} \left[-cov_\tau(n) + \underbrace{m X_1^\tau(n) X_1^0(n) + x_{m+1} x_{m+1+\tau} - (m+1) X_1^\tau(n+1) X_1^0(n+1)}_{\xi} \right],
\end{aligned} \tag{B.11}$$

allowing us to simplify ξ ,

$$\begin{aligned}
\xi &= m X_1^\tau(n) X_1^0(n) + x_{m+1} x_{m+1+\tau} - (m+1) X_1^\tau(n+1) X_1^0(n+1) \\
&= (m+1) \left(\underbrace{X_1^\tau(n) X_1^0(n) - X_1^\tau(n+1) X_1^0(n+1)}_{\gamma} \right) + \left(x_{m+1} x_{m+1+\tau} - X_1^\tau(n) X_1^0(n) \right), \tag{B.12}
\end{aligned}$$

and factorize γ ,

$$\begin{aligned}
\gamma &= X_1^\tau(n)X_1^0(n) - X_1^\tau(n+1)X_1^0(n+1) \\
&= \left(X_1^\tau(n) - X_1^\tau(n+1)\right)\left(X_1^0(n) + X_1^0(n+1)\right) - X_1^\tau(n)X_1^0(n+1) + X_1^\tau(n+1)X_1^0(n) \\
&= -\left(\frac{1}{m+1}(x_{m+1+\tau} - X_1^\tau(n))\right)\left(X_1^0(n) + X_1^0(n+1)\right) - X_1^\tau(n)(X_1^0(n) \\
&\quad + \frac{1}{m+1}(x_{m+1} - X_1^0(n))) + \left(X_1^\tau(n) + \frac{1}{m+1}(x_{m+1+\tau} - X_1^\tau(n))\right)X_1^0(n) \\
&= -\left(\frac{1}{m+1}(x_{m+1+\tau} - X_1^\tau(n))\right)\left(X_1^0(n) + X_1^0(n+1)\right) - X_1^\tau(n)\left(\frac{1}{m+1}(x_{m+1} - X_1^0(n))\right) \\
&\quad + \left(\frac{1}{m+1}(x_{m+1+\tau} - X_1^\tau(n))\right)X_1^0(n) \\
&= \frac{1}{m+1}\left\{-\left(x_{m+1+\tau} - X_1^\tau(n)\right)\left(X_1^0(n) + X_1^0(n+1)\right) - X_1^\tau(n)x_{m+1} + X_1^\tau(n)X_1^0(n) \right. \\
&\quad \left.+ x_{m+1+\tau}X_1^0(n) - X_1^\tau(n)X_1^0(n)\right\} \\
&= \frac{1}{m+1}\left\{-x_{m+1+\tau}X_1^0(n) - x_{m+1+\tau}X_1^0(n+1) + X_1^\tau(n)X_1^0(n) + X_1^\tau(n)X_1^0(n+1) \right. \\
&\quad \left.- X_1^\tau(n)x_{m+1} + x_{m+1+\tau}X_1^0(n)\right\} Ising \\
&= \frac{1}{m+1}\left\{-x_{m+1+\tau}X_1^0(n+1) + X_1^\tau(n)X_1^0(n) + X_1^\tau(n)X_1^0(n+1) - X_1^\tau(n)x_{m+1}\right\}. \quad (B.13)
\end{aligned}$$

Replacing B.13 at B.12 we have

$$\begin{aligned}
\xi &= -x_{m+1+\tau}X_1^0(n+1) + X_1^\tau(n)X_1^0(n) + X_1^\tau(n)X_1^0(n+1) - X_1^\tau(n)x_{m+1} + x_{m+1}x_{m+1+\tau} - X_1^\tau(n)X_1^0(n) \\
&= -x_{m+1+\tau}X_1^0(n+1) + X_1^\tau(n)X_1^0(n+1) - X_1^\tau(n)x_{m+1} + x_{m+1}x_{m+1+\tau} \\
&= \left(x_{m+1+\tau} - X_1^\tau(n)\right)\left(x_{m+1} - X_1^0(n+1)\right). \quad (B.14)
\end{aligned}$$

And finally, by replacing at B.11:

$$cov_\tau(n+1) = cov_\tau(n) + \frac{1}{n-\tau+1} \left[-cov_\tau(n) + \left(x_{m+1+\tau} - X_1^\tau(n)\right)\left(x_{m+1} - X_1^0(n+1)\right) \right] \quad (B.15)$$

C. Equations of motion for Polyakov Lagrangian and its equivalence with NG

The Polyakov Action amounts to

$$S_P[X, \gamma] = \frac{\sigma}{2} \int_{\Sigma} d\xi^0 d\xi^1 \sqrt{-\det \gamma} \gamma^{ab} \partial_a X^\mu \partial_b X_\mu, \quad (\text{C.1})$$

and the associated equations of motion can be obtained by varying the action with respect to the metric and to the field X^μ .

With respect to the field, we obtain

$$\delta_X S_P[X, \gamma] = \sigma \int_{\Sigma} d\xi^0 d\xi^1 \sqrt{-\det \gamma} \gamma^{ab} \partial_a X^\mu \partial_b \delta X_\mu \quad (\text{C.2})$$

$$= -\sigma \int_{\Sigma} d\xi^0 d\xi^1 \sqrt{-\det \gamma} \gamma^{ab} (\partial_b \partial_a X^\mu) \delta X_\mu \quad (\text{C.3})$$

$$= -\sigma \int_{\Sigma} d\xi^0 d\xi^1 \sqrt{-\det \gamma} \partial^a \partial_a X^\mu \delta X_\mu, \quad (\text{C.4})$$

so by imposing $\delta_X S_P[X, \gamma] = 0$ we get

$$\sqrt{-\det \gamma} \partial^a \partial_a X^\mu = 0, \quad (\text{C.5})$$

that corresponds to the KG equation coupled with the metric determinant.

Varying with respect to the metric yields

$$\delta_\gamma S_P[X, \gamma] = \frac{\sigma}{2} \int_{\Sigma} d\xi^0 d\xi^1 \delta \left(\sqrt{-\det \gamma} \gamma^{ab} \right) \partial_a X^\mu \partial_b X_\mu, \quad (\text{C.6})$$

and, recalling that the variation of the determinat obeys to

$$\delta \det \gamma = -\det \gamma \gamma_{ab} \delta \gamma^{ab}, \quad (\text{C.7})$$

we got

$$\delta_\gamma S_P[X, \gamma] = \frac{\sigma}{2} \int_\Sigma d\xi^0 d\xi^1 \delta \left(\sqrt{-\det \gamma} \gamma^{ab} \right) \partial_a X^\mu \partial_b X_\mu \quad (\text{C.8})$$

$$= \frac{\sigma}{2} \int_\Sigma d\xi^0 d\xi^1 \left(\left(\delta \sqrt{-\det \gamma} \right) \gamma^{ab} + \sqrt{-\det \gamma} \delta \gamma^{ab} \right) \partial_a X^\mu \partial_b X_\mu \quad (\text{C.9})$$

$$= \frac{\sigma}{2} \int_\Sigma d\xi^0 d\xi^1 \left(\frac{1}{2\sqrt{-\det \gamma}} \det \gamma_{cd} \delta \gamma^{cd} \gamma^{ab} + \sqrt{-\det \gamma} \delta \gamma^{ab} \right) \underbrace{\partial_a X^\mu \partial_b X_\mu}_{\equiv h_{ab}} \quad (\text{C.10})$$

$$= -\frac{\sigma}{2} \int_\Sigma d\xi^0 d\xi^1 \left(-\frac{1}{2} \sqrt{\det \gamma} \gamma_{ab} \gamma^{cd} h_{cd} + \sqrt{-\det \gamma} h_{ab} \right) \delta \gamma^{ab} \quad (\text{C.11})$$

$$= \frac{\sigma}{2} \int_\Sigma d\xi^0 d\xi^1 \sqrt{-\det \gamma} \left(\underbrace{-\frac{1}{2} \gamma_{ab} \gamma^{cd} h_{cd}}_{\alpha \quad T_{ab}} + h_{ab} \right) \delta \gamma^{ab}, \quad (\text{C.12})$$

so

$$h_{ab} - \frac{1}{2} \gamma_{ab} \gamma^{cd} h_{cd} = 0, \quad (\text{C.13})$$

which bot allows us to define the energy-momentum tensor and to verify the equivalence between this action and the Nambu-Goto action. Note that equation C.13 implies

$$h_{ab} = \frac{1}{2} \gamma_{ab} \gamma^{cd} h_{cd} \quad (\text{C.14})$$

$$\frac{h_{ab}}{\sqrt{-\det h}} = \frac{1}{2} \frac{\gamma_{ab} \gamma^{cd} h_{cd}}{\sqrt{-\det \left(\frac{1}{2} \gamma h_c^c \right)}} \quad (\text{C.15})$$

$$\frac{h_{ab}}{\sqrt{-\det h}} = \frac{1}{2} \frac{\gamma_{ab} h_c^c}{\sqrt{\left(\frac{h_c^c}{2} \right)^2} \sqrt{-\det \gamma}} \quad (\text{C.16})$$

$$h_{ab} = \sqrt{-\det h} \frac{\gamma_{ab}}{\sqrt{-\det \gamma}}, \quad (\text{C.17})$$

which is the key to understand the connection between Polyakov and Lagrangian actions, since by direct application of this equality onto the Polyakov action we arrive at

$$S_P[X, \gamma] = \frac{\sigma}{2} \int_\Sigma d\xi^0 d\xi^1 \sqrt{-\det \gamma} \gamma^{ab} \sqrt{-\det h} \frac{\gamma_{ab}}{\sqrt{-\det \gamma}} \quad (\text{C.18})$$

$$= \frac{\sigma}{2} \int_\Sigma d\xi^0 d\xi^1 \sqrt{-\det h} \underbrace{\gamma^{ab} \gamma_{ab}}_{d=2} \quad (\text{C.19})$$

$$S_P[X, \gamma] = \sigma \int_\Sigma d\xi^0 d\xi^1 \sqrt{-\det h} = S_N[X, h]. \quad (\text{C.20})$$

D. Backpropagation for a simple network

In this appendix we detail the backpropagation algorithm for a very basic *fully connected neural network* (FCNN). The architecture of the FCNN can be found at fig. (D.1). Each node corresponds to a value $x_i^{(l)}$, where l denotes the layer where the node lies and i the position within each layer. Our architecture, for example, has four layers, $l = 0, \dots, 3$ (one input, two hidden and one output). Each node can be obtained by

$$x_i^{(l+1)} = \sigma \left(\sum_j w_{ij}^{(l)} x_j^{(l)} \right), \quad (\text{D.1})$$

with w_{ij} as the weight that connect the j -th element of the layer l with the i -th element of the layer $l+1$ (in the pictorial representation, correspond to the lines that connect different nodes). σ represents the application of a non-linear differentiable function to the linear combination of $x^{(l)}$'s. The weights are the degrees of freedom of the network, and will be updated through the backpropagation procedure.

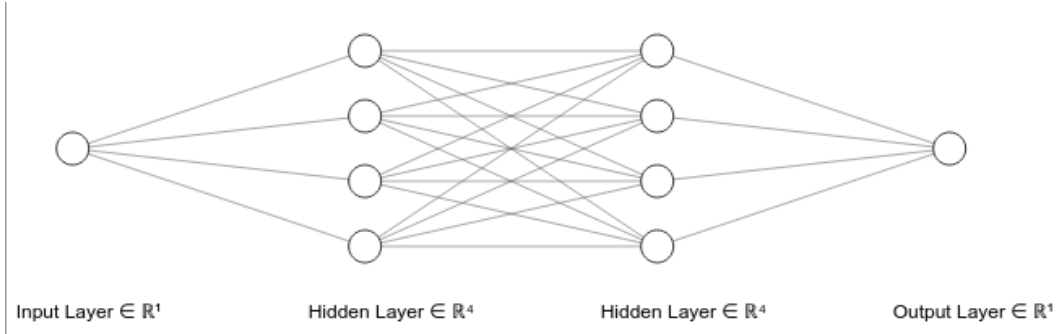


Figure D.1.: Architecture of the NN employed in the appendix.

If we feed to the network a given value $x^{(0)}$, we will obtain a final value $x^{(L)}$, that can be used to compute the loss function $\mathcal{L}(x^{(L)})$. With this value, it is possible to update the weights

$$w_{ij} \rightarrow w_{ij} + \eta \frac{\partial \mathcal{L}}{\partial w_{ij}}, \quad (\text{D.2})$$

and the power of the NN resides in the efficient computation of $\frac{\partial \mathcal{L}}{\partial w_{ij}}$. Note that the update

for the weights that connect the last hidden layer with the output layer is

$$\frac{\partial \mathcal{L}}{\partial w_{1j}^{(L-1)}} = \frac{\partial \mathcal{L}}{\partial x^{(L)}} \frac{\partial x^{(L)}}{\partial \xi^{(L)}} \frac{\partial \xi^{(L)}}{\partial w_{1j}^{(L-1)}}, \quad (\text{D.3})$$

if we identify $x_i^{(l)} = \sigma(\xi_i^{(l)}) \iff \xi_i^{(l)} = \sum_j w_{ij}^{(l-1)} x_j^{(l-1)}$. If the derivative of the loss function and the predetermined activation functions are known a priori, then the weight is easily updated. Similarly, the updates for $w_{ij}^{(L-2)}$ require the computation of

$$\frac{\partial \mathcal{L}}{\partial w_{ij}^{(L-2)}} = \underbrace{\frac{\partial \mathcal{L}}{\partial x^{(L)}} \frac{\partial x^{(L)}}{\partial \xi^{(L)}}}_{\equiv A} \underbrace{\frac{\partial \xi^{(L)}}{\partial x_i^{(L-1)}}}_{w_i^{(L-1)}} \frac{\partial x_i^{(L-1)}}{\partial \xi_i^{(L-1)}} \frac{\partial \xi_i^{(L-1)}}{\partial w_{ij}^{(L-2)}}. \quad (\text{D.4})$$

Once again the derivatives can be computed to update the weight. However, note that A was already previously obtained to update $w_{1j}^{(L-1)}$. In fact, when we update the weights between layers $l-1$ and l , we use terms that were previously computed to update the weights between the layers l and $l+1$. This means that we can start at the output layer and propagate backwards, updating iteratively the degrees of freedom of the NN according to eq. (D.2).

E. Normalizing Flows Architecture

In this appendix we detail the architecture employed in the Normalizing Flows for both the two particle case and the 2D Ising model.

Two particles

For either case, we need to start with a generator that allows us to sample from the equilibrium distribution, and for this case we are dealing with two non-interacting particles in one dimension. The probability of a microstate is

$$p(x_A, x_B) = \exp \left(-(x_A - 1)^2 - (x_B + 1)^2 \right)$$

so we need to sample each position x_i from a standard gaussian distribution, a fairly standard procedure.

Now we need to design the Normalizing Flow. First, recall that a single affine layer is defined as

$$g_i(x_A, x_B) = \begin{cases} x'_A = x_A \\ x'_B = x_B \otimes e^{s_i(x_A)} + t_i(x_A) \end{cases} \quad (\text{E.1})$$

and we need to specify the structure of the neural network functions s and t . In this simple case, we build such functions with a standard fully connected neural network (FCNN), as can be seen in figure [E.1](#). Then, we compose 16 of such layers, alternating the non-active coordinate with each successive layer.

2D Ising Model

The extension of the Normalizing Flows approach to the Ising model is quite simple. The only things we need to change are the generator procedure and the shape of the neural network. The configurations drawn from the periodic Ising equilibrium distributions are obtained with a Wolff algorithm, and after being sampled pass through the HS map, equation [4.51](#). The generic layer, splits the lattice configuration ϕ into two checkerboard-like sub-lattices, ϕ_1 and ϕ_2 , which take the role of x_A and x_B in [E.1](#). The neural network functions are convolutional neural network (CNN), whose architecture can be seen in figure [E.2](#).

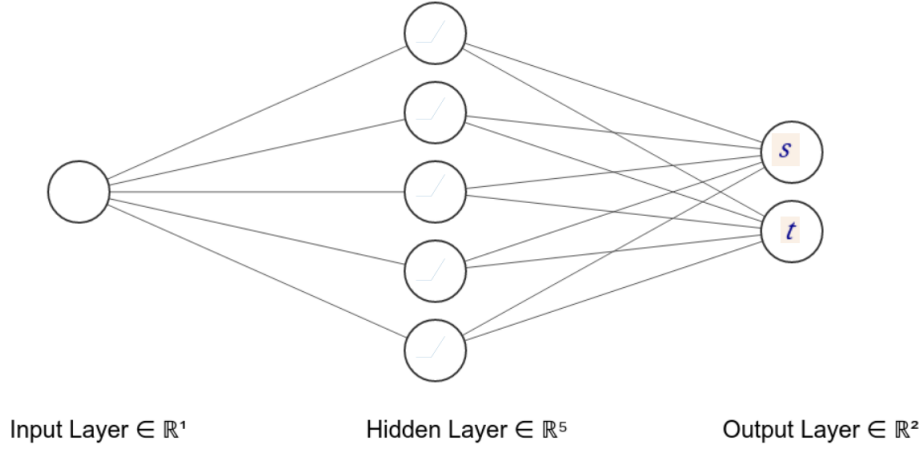


Figure E.1.: FCNN architecture employed in all the layers of the two particles Normalizing Flow. The first layer has ReLU activation functions and the final layer has Tanh activation functions. The first layer takes as input the non-active coordinate, i.e., the coordinate that will not change by application of the layer. The values for the s and t function can be computed at once, and correspond to the two outputs of the NN. The expressivity of such functions is controlled by the number of nodes present in the hidden layer - the more nodes the more expressive the functions become. The weights associated with each node are updated through the backpropagation algorithm, and aim to minimize the loss function 4.33.

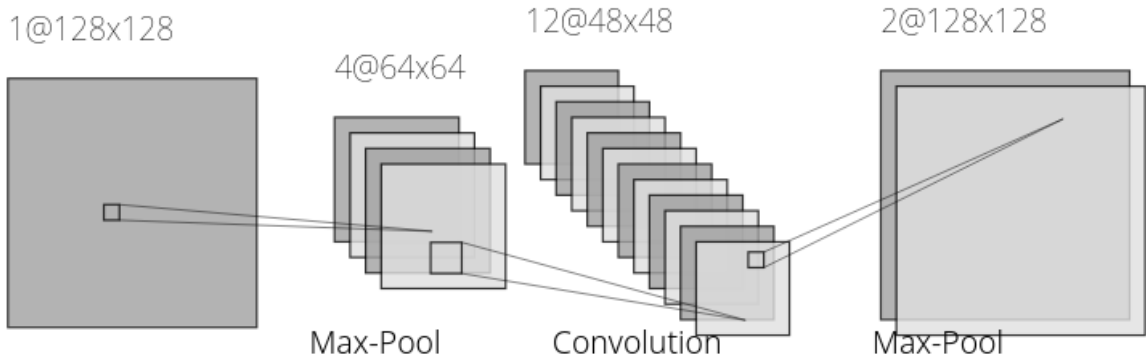


Figure E.2.: CNN architecture employed in all the layers of the Ising model. The input is the non-active sublattice. The shape of the sublattice is the same as the one of the lattice, but half of the values are null. The output layer returns the values for the s and t function. We utilize a CNN since we have a configuration with local interactions. A detailed review of this kind of architectures can be found at [40].

F. Free energy data

In this appendix we list the data we obtained for the various system sizes. Recall that all the measurements were made for $L_x = L_y$.

$L_x = L_y$	$F^{(1)}$	$\delta F^{(1)}$
24	4.8873	0.0002
28	5.3649	0.0002
32	5.9721	0.0002
40	7.4813	0.0002
48	9.3324	0.0004
56	11.5107	0.0004
64	14.0225	0.0004
72	16.857	0.001
80	20.025	0.001
88	23.532	0.001
96	27.348	0.001
104	31.500	0.002
112	36.009	0.002

(a) Data for $L_z = 80$.

$L_x = L_y$	$F^{(1)}$	$\delta F^{(1)}$
24	5.0657	0.0003
28	5.4347	0.0002
32	5.9932	0.0002
40	7.4846	0.0004
48	9.3326	0.0005
56	11.5102	0.0005
64	14.012	0.001
72	16.834	0.003
80	19.950	0.003
88	23.436	0.004

(b) Data for $L_z = 128$

$L_x = L_y$	$F^{(1)}$	$\delta F^{(1)}$
24	5.1989	0.0003
28	5.4947	0.0003
32	6.0119	0.0003
40	7.4821	0.0004
48	9.3323	0.0005
56	11.5059	0.0006
64	14.009	0.006
72	16.796	0.007
80	19.896	0.007
88	23.343	0.008

(c) Data for $L_z = 160$

Table F.1.: Data for various L_z 's with square transverse areas.

$L_x = L_y$	$F^{(1)}$	$\delta F^{(1)}$
18	4.6192	0.0002
22	4.7906	0.0003
24	4.9442	0.0001
26	5.1422	0.0004
28	5.3830	0.0005
30	5.6662	0.0006
32	5.9771	0.0001
34	6.31922	0.00087
36	6.683	0.001
40	7.4823	0.0003
48	9.3313	0.0004
56	11.5122	0.0005
64	14.0186	0.0004
72	16.857	0.001
80	20.002	0.001
88	23.511	0.002
96	27.349	0.002
104	31.499	0.002
112	35.963	0.007

(a) Data for $L_z = 96$.

$L_x = L_y$	$F^{(1)}$	$\delta F^{(1)}$
24	4.7973	0.0001
28	5.3316	0.0002
32	5.9575	0.0002
40	7.4633	0.0002
48	9.2934	0.0002
56	11.441	0.0003
64	13.9125	0.0003
72	16.699	0.001
80	19.809	0.001
88	23.245	0.002
96	27.007	0.002
104	31.084	0.002
112	35.496	0.002
120	40.225	0.003
128	45.239	0.002

(b) Data for $L_z = 48$.Table F.2.: Data for various L_z 's with square transverse areas.

Bibliography

- [1] C. Jarzynski. A nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690–2693, April 1997.
- [2] M. Lüscher. Symmetry-breaking aspects of the roughening transition in gauge theories. *Nuclear Physics B*, 180(2):317–329, March 1981.
- [3] Y Nambu. Strings, monopoles, and gauge fields. page 7.
- [4] Tetsuo Gotō. Relativistic Quantum Mechanics of One-Dimensional Mechanical Continuum and Subsidiary Condition of Dual Resonance Model. *Progress of Theoretical Physics*, 46(5):1560–1569, November 1971.
- [5] M. Lüscher, K. Symanzik, and P. Weisz. Anomalies of the free loop wave equation in the WKB approximation. *Nuclear Physics B*, 173(3):365–396, October 1980.
- [6] Joseph Polchinski and Andrew Strominger. Effective string theory. *Physical Review Letters*, 67(13):1681–1684, September 1991.
- [7] Yoichiro Nambu. Quasi-Particles and Gauge Invariance in the Theory of Superconductivity. *Physical Review*, 117(3):648–663, February 1960.
- [8] J. Goldstone. *Field Theories with "Superconductor" Solutions*. 1961.
- [9] Ofer Aharony and Matthew Dodelson. Effective String Theory and Nonlinear Lorentz Invariance. *Journal of High Energy Physics*, 2012(2):8, February 2012.
- [10] Ofer Aharony and Zohar Komargodski. The effective theory of long strings. page 26, 2013.
- [11] Gunnar S Bali. QCD forces and heavy quark bound states. *Physics Reports*, page 136, 2001.
- [12] Michele Caselle. Effective string description of the confining flux tube at finite temperature, May 2021.
- [13] Bastian B. Brandt and Marco Meineri. Effective string description of confining flux tubes. *International Journal of Modern Physics A*, 31(22):1643001, August 2016.

- [14] Michele Caselle, Gianluca Costagliola, Alessandro Nada, Marco Panero, and Arianna Toniato. Jarzynski's theorem for lattice gauge theory. *Physical Review D*, 94(3):034503, August 2016.
- [15] Michele Caselle, Martin Hasenbusch, and Marco Panero. The interface free energy: Comparison of accurate Monte Carlo results for the 3D Ising model with effective interface models. *Journal of High Energy Physics*, 2007(09):117, 2007.
- [16] Michele Caselle, Marco Panero, and Davide Vadacchino. Width of the flux tube in compact U(1) gauge theory in three dimensions. page 21, 2016.
- [17] Michele Caselle, Marco Panero, Roberto Pellegrini, and Davide Vadacchino. A different kind of string. *Journal of High Energy Physics*, 2015(1):105, January 2015.
- [18] Marco Billó, Michele Caselle, and Livia Ferro. The partition function of interfaces from the Nambu-Goto effective string theory. *Journal of High Energy Physics*, 2006(02):070–070, February 2006.
- [19] Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, February 1925.
- [20] H. A. Kramers and G. H. Wannier. Statistics of the Two-Dimensional Ferromagnet. Part I. *Physical Review*, 60(3):252–262, August 1941.
- [21] L. Onsager. A Two-Dimensional Model with an Order-Disorder Transition. *Phys. Rev.* 65, page 117, 1944.
- [22] JMN Silva. Simulação de um modelo físico simples. *Gazeta da Física*, (Gaz. Fis.), 1986.
- [23] Martin Niss. History of the Lenz–Ising model 1965–1971: The role of a simple model in understanding critical phenomena. *Archive for History of Exact Sciences*, 65(6):625–658, November 2011.
- [24] Hans Gerd Evertz, Martin Hasenbusch, Mihail Marcu, Klaus Pinn, and Sorin Solomon. Stochastic cluster algorithms for discrete gaussian (SOS) models. *Physics Letters B*, 254(1):185–191, January 1991.
- [25] Hans Evertz, Martin Hasenbusch, Mihail Marcu, Klaus Pinn, and Sorin Solomon. High precision measurement of the SOS surface thickness in the rough phase. *Journal de Physique I*, 1, December 1991.
- [26] Hans Gerd Evertz, Martin Hasenbusch, Mihai Marcu, and Klaus Pinn. The solid-on-solid surface width around the roughening transition. *Physica A: Statistical Mechanics and its Applications*, 199(1):31–39, October 1993.

- [27] M. Hasenbusch, G. Lana, M. Marcu, and K. Pinn. Cluster Algorithm for a Solid-On-Solid Model with Constraints, July 1992.
- [28] M. Hasenbusch and K. Pinn. Comparison of Monte Carlo results for the 3D Ising interface tension and interface energy with (extrapolated) series expansions. *Physica A: Statistical Mechanics and its Applications*, 203(2):189–213, February 1994.
- [29] M. Hasenbusch, S. Meyer, and M. Pütz. The Roughening Transition of the 3D Ising Interface: A Monte Carlo Study. *Journal of Statistical Physics*, 85(3-4):383–401, November 1996.
- [30] Martin Hasenbusch and Klaus Pinn. Surface Tension, Surface Stiffness, and Surface Width of the 3-dimensional Ising Model on a Cubic Lattice. *Nuclear Physics B - Proceedings Supplements*, 30:857–860, March 1993.
- [31] Vladimir Privman. Fluctuating Interfaces, Surface Tension, and Capillary Waves: An Introduction. *International Journal of Modern Physics C*, 03(05):857–877, October 1992.
- [32] M. Caselle, M. Panero, and P. Provero. String effects in Polyakov loop correlators. *Journal of High Energy Physics*, 2002(06):061–061, June 2002.
- [33] Joseph Polchinski. *String Theory: Volume 1: An Introduction to the Bosonic String*, volume 1 of *Cambridge Monographs on Mathematical Physics*. Cambridge University Press, Cambridge, 1998.
- [34] Michele Caselle, Elia Cellini, Alessandro Nada, and Marco Panero. Stochastic normalizing flows as non-equilibrium transformations. *JHEP 07*, February 2022.
- [35] Joan Elias Miró, Andrea L. Guerrieri, Aditya Hebbar, João Penedones, and Pedro Vieira. Flux Tube S -Matrix Bootstrap. *Physical Review Letters*, 123(22):221602, November 2019.
- [36] S. Dubovsky, R. Flauger, and V. Gorbenko. Flux tube spectra from approximate integrability at low energies. *Journal of Experimental and Theoretical Physics*, 120(3):399–422, March 2015.
- [37] Chang Chen, Peter Conkey, Sergei Dubovsky, and Guzmán Hernández-Chifflet. Undressing confining flux tubes with TT. *Physical Review D*, 98(11):114024, December 2018.
- [38] Serap Ogmen. *A General Look at the String Theory and Its Variations*. November 2017.
- [39] W Krauth. *Statistical Mechanics: Algorithms and Computations*. Oxford Master Series in Physics, 13.

- [40] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53, December 2021.
- [41] A. M. Polyakov. Quantum geometry of bosonic strings. *Physics Letters B*, 103(3):207–210, July 1981.
- [42] D. J. Pritchard and W. J. Stirling. QCD calculations in the light-cone gauge. *Nuclear Physics B*, 165(2):237–268, March 1980.
- [43] M. Caselle, R. Fiore, F. Gliozzi, M. Hasenbusch, K. Pinn, and S. Vinti. Rough Interfaces Beyond the Gaussian Approximation, July 1994.
- [44] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Oxford University Press, 1999.
- [45] Nicholas Metropolis and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6), 1953.
- [46] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, April 1970.
- [47] Ulli Wolff. Critical slowing down. *Nuclear Physics B - Proceedings Supplements*, 17:93–102, September 1990.
- [48] Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58(2):86–88, January 1987.
- [49] Ferenc Niedermayer. General Cluster Updating Method for Monte Carlo Simulations. *Physical Review Letters*, 61(18):2026–2029, October 1988.
- [50] Ulli Wolff. Collective Monte Carlo Updating for Spin Systems. *Physical Review Letters*, 62(4):361–364, January 1989.
- [51] Jaron Kent-Dobias and James P. Sethna. Cluster representations and the Wolff algorithm in arbitrary external fields. *Physical Review E*, 98(6):063306, December 2018.
- [52] J. Viana Lopes, Miguel D. Costa, J. M. B. Lopes dos Santos, and R. Toral. Optimized multicanonical simulations: A new proposal based on classical fluctuation theory. *Physical Review E*, 74(4):046702, October 2006.
- [53] J. M. Kosterlitz and D. J. Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6(7):1181–1203, April 1973.

- [54] John D. Weeks, George H. Gilmer, and Harry J. Leamy. Structural Transition in the Ising-Model Interface. *Physical Review Letters*, 31(8):549–551, August 1973.
- [55] K. K. Mon, D. P. Landau, and D. Stauffer. Interface roughening in the three-dimensional Ising model. *Physical Review B*, 42(1):545–547, July 1990.
- [56] Jian-Sheng Wang and Robert H. Swendsen. Cluster Monte Carlo algorithms. *Physica A: Statistical Mechanics and its Applications*, 167(3):565–579, September 1990.
- [57] H. Nishimori and G. Ortiz. *Elements of Phase Transitions and Critical Phenomena*. 2011.
- [58] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56(5):5018–5035, November 1997.
- [59] Ivan Kobyzev, Simon J. D. Prince, and Marcus A. Brubaker. Normalizing Flows: An Introduction and Review of Current Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, November 2021.
- [60] Gerhard Hummer and Attila Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences*, 98(7):3658–3661, March 2001.
- [61] Daniel K. West, Peter D. Olmsted, and Emanuele Paci. Free energy for protein folding from nonequilibrium simulations using the Jarzynski equality. *The Journal of Chemical Physics*, 125(20):204910, November 2006.
- [62] Jan Liphardt, Sophie Dumont, Steven B. Smith, Ignacio Tinoco, and Carlos Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski’s equality. *Science (New York, N.Y.)*, 296(5574):1832–1835, June 2002.
- [63] F Douarche, S Ciliberto, A Petrosyan, and I Rabbiosi. An experimental test of the Jarzynski equality in a mechanical experiment. *Europhysics Letters (EPL)*, 70(5):593–599, June 2005.
- [64] Daniel C. Hackett, Chung-Chun Hsieh, Michael S. Albergo, Denis Boyda, Jiunn-Wei Chen, Kai-Feng Chen, Kyle Cranmer, Gurtej Kanwar, and Phiala E. Shanahan. Flow-based sampling for multimodal distributions in lattice field theory. *Phys. Rev. D* 104, July 2021.
- [65] Michael S. Albergo, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, Julian M. Urban, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, and Phiala E. Shanahan. Flow-based sampling for fermionic lattice field theories. *arXiv:2106.05934 [cond-mat, physics:hep-lat]*, June 2021.

- [66] M. S. Albergo, G. Kanwar, and P. E. Shanahan. Flow-based generative models for Markov chain Monte Carlo in lattice field theory. *Physical Review D*, 100(3):034515, August 2019.
- [67] Kim A. Nicoli, Christopher J. Anders, Lena Funcke, Tobias Hartung, Karl Jansen, Pan Kessel, Shinichi Nakajima, and Paolo Stornati. Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models. *Physical Review Letters*, 126(3):032001, January 2021.
- [68] Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan. Equivariant flow-based sampling for lattice gauge theory. *Physical Review Letters*, 125(12):121601, September 2020.
- [69] Denis Boyda, Gurtej Kanwar, Sébastien Racanière, Danilo Jimenez Rezende, Michael S. Albergo, Kyle Cranmer, Daniel C. Hackett, and Phiala E. Shanahan. Sampling using $SU(N)$ gauge equivariant flows. *Physical Review D*, 103(7):074504, April 2021.
- [70] Luigi Del Debbio, Joe Marsh Rossney, and Michael Wilson. Efficient Modelling of Trivializing Maps for Lattice ϕ^4 Theory Using Normalizing Flows: A First Look at Scalability. *arXiv:2105.12481 [hep-lat]*, October 2021.
- [71] Pim de Haan, Corrado Rainone, Miranda C. N. Cheng, and Roberto Bondesan. Scaling Up Machine Learning For Quantum Field Theory with Equivariant Continuous Flows. November 2021.
- [72] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986.
- [73] Michael S. Albergo, Denis Boyda, Daniel C. Hackett, Gurtej Kanwar, Kyle Cranmer, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan. Introduction to Normalizing Flows for Lattice Field Theory, August 2021.
- [74] Luigi Del Debbio, Joe Marsh Rossney, and Michael Wilson. Machine Learning Trivializing Maps: A First Step Towards Understanding How Flow-Based Samplers Scale Up. December 2021.
- [75] Suriyanarayanan Vaikuntanathan and Christopher Jarzynski. Escorted Free Energy Simulations: Improving Convergence by Reducing Dissipation. *Physical Review Letters*, 100(19):190601, May 2008.
- [76] Gavin S. Hartnett and Masoud Mohseni. Self-Supervised Learning of Generative Spin-Glasses with Normalizing Flows, January 2020.

- [77] Fugao Wang and D. P. Landau. An efficient, multiple range random walk algorithm to calculate the density of states. *Physical Review Letters*, 86(10):2050–2053, March 2001.
- [78] JMVP Lopes. *Métodos de Monte Carlo Para Sistemas Com Desordem e Interação de Longo Alcance*. PhD thesis, FCUP, 2006.
- [79] Laurence Jacobs and Claudio Rebbi. Multi-spin coding: A very efficient technique for Monte Carlo simulations of spin systems. *Journal of Computational Physics*, 41(1):203–210, May 1981.
- [80] George B. Arfken, Hans J. Weber, and Frank E. Harris. Chapter 15 - Legendre Functions. In George B. Arfken, Hans J. Weber, and Frank E. Harris, editors, *Mathematical Methods for Physicists (Seventh Edition)*, pages 715–772. Academic Press, Boston, January 2013.
- [81] M. Billo, M. Caselle, and L. Ferro. Universal behaviour of interfaces in 2d and dimensional reduction of Nambu-Goto strings. *Nuclear Physics B*, 795(3):623–634, June 2008.