Fourier Acceleration in the $SU(N) \times SU(N)$ Chiral Model

MPhys Project

Pablo Morandé

Abstract

This study investigates the two-dimensional $SU(N) \times SU(N)$ Principal Chiral Model- a theory that captures some of the Quantum Chromodynamics' key properties in the lattice for general N by employing a Fourier Accelerated algorithm. The employment of Fourier Acceleration leads to an order of magnitude speed up in terms of the Integrated Autocorrelation Time and plain efficiency even at small correlation lengths. This improvement allows us to verify the main properties of the model with modest computing capacities at a level of accuracy exceeding previous lattice calculations of this model.

School of Physics And Astronomy
The University of Edinburgh
United Kingdom
2023-2024

Personal Statement

Preface

> A mind needs books as a sword
> needs a whetstone, if it is to
> keep its edge.
>
> George R. R. Martin, A Game
> of Thrones

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Pablo Morandé, 2024

Acknowledgements

I would like to deeply thank both of my supervisors, Dr Roger Horsely and Dr Brian Pendleton, for their patience and guidance throughout the project.

Contents

Chapter 1

Introduction

> I have learned all kinds of things from my many mistakes. The one thing I never learn is to stop making them.
>
> Joe Abercrombie, Last Argument of Kings

The Standard Model of Particle Physics is the current theory that best describes all interactions between particles (not including gravity). Despite its incredible success, there exists compelling evidence of the existence of physics beyond it, such as the existence of dark matter [1]. Consequently, intensive efforts are dedicated to scrutinizing and testing the Standard Model in the search for new physics. One particularly rich sector in this regard is Quantum Chromodynamics (QCD), the theory of the strong force, i.e., the theory of quarks and gluons.

The search for new physics encompasses research on both ends of the energy spectrum. In the high-energy regime, the search for new physics is centred on discovering new particles in collider experiments, while in the low-energy regime, it is focused on observing small deviations in some Standard Model observables [1]. Therefore, the latter approach relies on making both precise theoretical predictions and experimental measurements at low energies. In this case, we are interested in the problems associated with the theoretical calculations: at low energies QCD becomes strongly coupled, leading to the breakdown of perturbation theory. Nevertheless, it is possible to obtain predictions of QCD in this energy range thanks to a fully non-perturbative framework known as Lattice Field Theory. This is an ab initio numerical approach that relies on the discretization of the continuous spacetime into a discrete lattice to simulate QCD.

Over the years, improvements in computation and theory have brought Lattice QCD to a point where it can be used to measure Standard Model observables with great accuracy. For instance, lattice QCD is currently being used in the determination of the hadronic contribution to the muon's anomalous moment $g_\mu - 2$. Lattice calculations could help in resolving the current tension between the experimental value and the theoretical prediction of this quantity [2].

Despite the success of Lattice Field Theory, one important caveat is the computational power it requires. In order to improve the quality of results obtained by this framework, the lattice needs to reproduce the continuum, which requires the use of finer and larger lattices at the

1

expense of computational cost. Furthermore, the algorithms used in Lattice Field Theory become highly inefficient as the continuum is approached; this problem is commonly referred to as critical slowdown. Several algorithmic strategies exist to address critical slowdown, such as Multi Grid Monte Carlo [3] and Fourier Acceleration. The latter has been successfully employed on different models [4; 5] and will be the topic of this report.

Implementing Fourier Acceleration directly in Lattice QCD poses significant challenges due to QCD's inherent complexity, although advancements are being made in this area [6]. In this report, we will focus on implementing and analysing the effect of Fourier Acceleration on a simpler theory known as the two-dimensional $SU(N) \times SU(N)$ Chiral Model for general N. This model shares with QCD some of its key properties like asymptotic freedom, which means that it retains the essence of QCD while being computationally easier to work with. By studying the effect of Fourier Acceleration in this model, we hope to establish a basis for studying more complicated theories and, eventually, QCD.

The objective of this project is twofold: firstly, to analyse and verify the main properties of the Principal Chiral Model in the lattice, and secondly, to compare the performance of the Fourier Accelerated HMC with the traditional HMC for this model.

This report will be structured as follows. In Chapter 2, we will introduce Lattice Field Theory, critical slowdown and how Fourier Acceleration can be used to address this problem. Chapter 3 will provide an overview of $SU(N) \times SU(N)$ Principal Chiral Model in two dimensions and will explain how Hybrid Monte Carlo and Fourier Acceleration are implemented in this model. In Chapter 4 we will present the results obtained by employing Fourier Acceleration in the study of the Chiral Model. We will finish with a detailed comparison between the performance of the Fourier Accelerated Hybrid Monte Carlo and its non-accelerated version. Lastly, Chapter 5 will suggest different options for future work and conclude the report. All the code developed for this project is available in [7].

Background

> In the beginning, there was nothing, which exploded.
>
> Terry Pratchett, Lords and Ladies

This chapter aims to briefly introduce the main tools used in Lattice Field Theory. After addressing the important questions of what Lattice Field Theory is and why we need to use it, the chapter moves to explain Hybrid Monte Carlo (HMC), the main algorithm used in Lattice Field Theory and some of the data analysis techniques used to study these theories. The chapter concludes with a conceptual introduction to Fourier Acceleration, the main technique used in this report.

2.1 Field Theories in the Lattice

In Quantum Field Theory, the focus is on computing the Path Integral, which gives all the information regarding the observables. That is, if O is an observable, its expectation value relative to the ground state of the theory $|0\rangle$ is given by:

$$\langle 0| O |0\rangle \equiv \langle O \rangle = \frac{1}{Z} \int \mathcal{D}\phi O \exp\{iS\} \tag{2.1}$$

$$Z = \int \mathcal{D}\phi \exp\{iS\} \tag{2.2}$$

where ϕ is the field of the theory, and the measure $\int \mathcal{D}\phi$ is essentially an integral over all possible field configurations. On the other hand, S is the action governing each particular field theory. For example, the action for a free scalar field theory in d dimensions is given by:

$$S = \int dx^d \left(\frac{1}{2}\partial^\mu \phi \partial_\mu \phi - \frac{m^2}{2}\phi(x)^2 \right) \tag{2.3}$$

The first step towards the estimation of the Path Integral is the so-called "Wick rotation" [8], which effectively involves letting $t = -i\tau$; more formally, a Wick rotation is an analytic continuation to imaginary time. Under this change, $iS \to -S_E$ where the subscript E stands for Euclidean since the Wick rotation involves a change from the Minkowski metric to the Euclidean one.

For instance, in the case of the free field theory, the Euclidean action corresponding to a Wick rotation in this case is given by:

$$S_E = \int dx^d \frac{1}{2}(\partial_\mu \phi)^2 + \frac{m^2}{2}\phi(x)^2 \tag{2.4}$$

After the rotation, the Path Integral is given by:

$$Z = \int \mathcal{D}\phi \exp\{-S_E\} \tag{2.5}$$

If the Euclidean action is real, then the Path Integral can be interpreted as a weighted sum over all possible field configurations, where the statistical weight of each configuration would be given by $\exp\{-S_E\}$. This idea will be further explored in the following sections, but first, we need to deal with the sum over all possible configurations.

2.1.1 The Lattice

In a continuous space-time, the integral $\int \mathcal{D}\phi$ in Eq. 2.5 is infinitely dimensional, i.e. it represents the integral over an infinite number of variables. This happens because $\int \mathcal{D}\phi$ represents the sum over all possible field configurations, and in order to include all of them, it is necessary to integrate over the infinite number of degrees of freedom of the field ϕ in the continuum. Consequently, it is impossible to numerically estimate this integral directly as no device can handle infinite degrees of freedom. To overcome this challenge, we need to discretize the space-time of the theory, which renders the integral finite-dimensional.

In order to discretize the theory, we start with a cubic-like region of spacetime with d volume $V = L^d$. Then, we divide each dimension into N slices such that, in total, we obtain N^d points. $x_\mu = n_\mu a$ where $\{n_\mu = 0, 1...N-1, \forall \mu\}$ $a = \frac{L}{N}$ (known as the lattice spacing) with n ranging from 0 to $N-1$. In summary, our lattice Λ is defined as the set of points given by:

$$\Lambda = \{x = (x_0, \underline{x}) \in aZ^d | 0 \leq x_\mu \leq a(N-1)\} \tag{2.6}$$

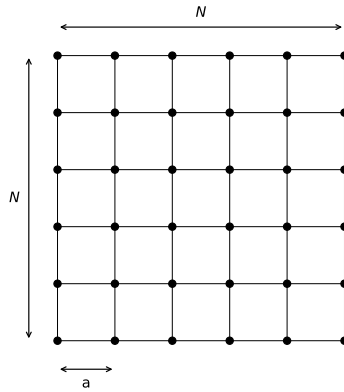Fig. 2.1 offers a visualisation of a 2-dimensional lattice.

Fig. 2.1: A digaramatic representation of a 2-dimensional lattice. a denotes the lattice spacing, and N is the number of points in each direction. Adapted from [9].

Under this discretization, the field $\phi(x)$ becomes a discrete set of values. We denote such set as ϕ:

$$\phi = \{\phi_x, x \in \Lambda\}, \phi_x = \phi(x) \tag{2.7}$$

Additionally, we impose periodic boundary conditions in all directions of space-time which turns the region into a toroid:

$$\phi_x = \phi_{x+Na\hat{\mu}}, \forall \mu \tag{2.8}$$

where $\hat{\mu}$ is a unit vector aligned with the μ axis. The measure of the Path Integral becomes:

$$\int \mathcal{D}\phi \to \int \Pi_{x \in \Lambda} d\phi_x \tag{2.9}$$

Then Eq. 2.1 turns into:

$$\langle O \rangle_{V,a} = \frac{\int \Pi_{x \in \Lambda} d\phi_x O \exp\{-S_{E,a,V}\}}{\int \Pi_{x \in \Lambda} d\phi_x \exp\{-S_{E,a,V}\}} \tag{2.10}$$

with $S_{E,a,V}$ being a discrete version of the action of the field theory. The true quantity of interest $\langle O \rangle$ in the continuum corresponds to the infinite volume and zero lattice spacing limit of Eq. 2.10:

$$\langle O \rangle = \lim_{V \to \infty, a \to 0} \langle O \rangle_{V,a} \tag{2.11}$$

Hence, to estimate $\langle O \rangle$ we calculate $\langle O \rangle_{V,a}$ and we extrapolate to the continuum limit by considering larger and finer lattices.

We now have a product of integrals over a finite number of degrees of freedom. However, while now finite, the number of degrees of freedom is still very large (as this effectively controls how close we get to the continuum limit), so we still need a technique to reliably perform this calculation. To avoid notation cluttering, we will henceforth drop the V, a subscripts.

2.1.2   Monte Carlo Integration

The problem is now being able to compute the finite-dimensional integral of Eq. 2.10. This is no easy task since the most simple methods used to approximate one or two-dimensional integrals, like the trapezoid rule, perform badly when faced with a high-dimensional integral like the one in Eq. 2.10. Therefore, numerically estimating $\langle O \rangle$ requires a different approach which scales well with the dimensionality of the integral.

We begin by reaffirming the connection of the Path Integral with statistics. One can see that if the action is real, then the function $\mathcal{P}(\phi) = \frac{\exp\{-S_E(\phi)\}}{Z}$ is positive, that is, $\mathcal{P}(\phi) > 0$. Furthermore, $\mathcal{P}(\phi)$ is normalized under the measure : $\int \Pi_{x \in \Lambda} d\phi_x$. This suggests that $\mathcal{P}(\phi)$ can be interpreted as a probability distribution function, which determines the distribution of the multidimensional variable ϕ. In this context, $O(\phi)$ is also a random variable, and we can identify $\langle O(\phi) \rangle = E[O(\phi)]$, where E indicates the expectation value in a statistical sense.

In order to solve our problem, it is helpful to understand the behaviour of expectation values in high-dimensional spaces. Consider, then, the expectation value of an arbitrary function f, which is a function of a stochastic variable q distributed according to the distribution $\pi(q)$:

$$E[f] = \int dq f(q)\pi(q) \tag{2.12}$$

We are interested in determining the region of the $q-$space that yields the dominant contribution to Eq. 2.12 in the case that the q-space is high-dimensional. To keep the discussion for a general function f, it is commonly assumed that the target function f is sufficiently uniform such that its variation does not contribute much to the integrand [10]. This assumption allows us to focus on the effects of the distribution $\pi(q)$ and the geometry of the space itself.

In high-dimensional spaces, the probability density tends to concentrate around its mode, and this effect becomes accentuated as the dimensionality increases [10]. Nevertheless, it would be naive to conclude that the neighbourhood of the mean gives the dominant contribution to the expectation value since this does not take into account the effect of the volume dq.

As the dimension of the space is increased, the volume within any localized neighbourhood dramatically shrinks relative to the volume of the space outside that neighbourhood [10]. Consequently, the small volume of the neighbourhood of the mode prevents this region from becoming dominant in the expectation value despite its high density. Conversely, the vanishing density of the region far away from the mean prevents it from making a major contribution to the expectation value despite its large volume contribution. Therefore, the dominant contribution to Eq. 2.12 is given by the region where both effects equilibrate, known as the Typical Set. The opposite behaviour of the density and the volume in high dimensional spaces are depicted in Fig. 2.2 along with the Typical Set.
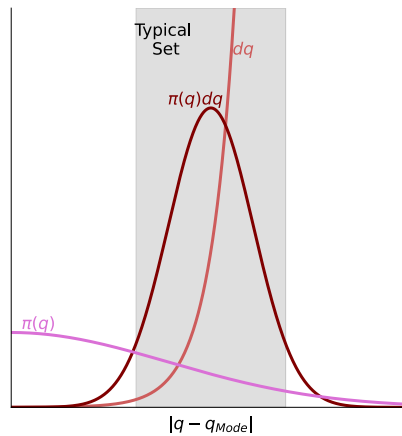

Fig. 2.2: The Figure shows the opposite behaviour of the density $\pi(q)$, which concentrates around the mode, and the volume dq, which is exponentially larger outside the neighbourhood of the mode (in high dimensional spaces). The Typical Set is the region where both effects equilibrate (grey area). The figure was adapted from [10].

Since the opposite behaviour of the density and the volume becomes more pronounced with increasing dimensionality, the Typical Set becomes narrower.

In conclusion, the dominant contribution to expectation values in high dimensional space is given by the Typical Set, a region in which $\pi(q)dq$ is maximised and which becomes smaller as the dimensionality is increased. Returning to our initial problem of estimating the expectation value $\langle O \rangle$, we need a method that is able to correctly identify the Typical Set of the $\phi-$space, which becomes smaller as the continuum limit is approached. The following section will explain how Markov Chains can be used for that purpose.

2.1.2.1 Markov Chains

A Markov Chain is a sequence of random variables ϕ_i in which the next state of the chain depends only on its current state and not. Therefore, a Markov Chain is defined through its transition probability, which is the probability of the next state being ϕ^* if the chain is in the state ϕ and is usually denoted as $\mathcal{T}(\phi^*|\phi)$. If the transition probability is chosen carefully, then Markov Chains can be used to take samples from the typical set and to estimate expectation values such as Eq. 2.10.

In general, the probability distribution of ϕ along the chain will not converge to a particular distribution. However, it will converge to $\mathcal{P}(\phi)$ (known as the target distribution) if the following conditions are met:

1. Invariance of the target distribution

   We require that the transition probability does not affect the target distribution. That is, if the chain is generating samples from the target distribution, the following samples will continue to be sampled from it. This condition can be expressed as:

   $$\mathcal{P}(\phi^*) = \int \mathcal{D}\phi \mathcal{P}(\phi)\mathcal{T}(\phi^*|\phi) \tag{2.13}$$

   A sufficient (but not necessary condition) to satisfy Eq. 2.13 is known as Detailed Balance, which can be summarized as;

   $$\mathcal{P}(\phi)\mathcal{T}(\phi^*|\phi) = \mathcal{P}(\phi^*)\mathcal{T}(\phi|\phi^*) \tag{2.14}$$

2. Ergodicity

   Ergodicity refers to the ability of the chain to reach any allowed state by $\mathcal{P}(\phi)$ in a finite number of steps starting from an arbitrary initial state. This condition allows the Markov Chain to converge to the Typical Set regardless of the initial conditions.

If these two conditions are satisfied, then the state space explored by the Markov Chain will eventually converge to the Typical Set of $\mathcal{P}(\phi)$ [10], so the chain will give samples ϕ_i corresponding to the dominant contribution region on any expectation value of the form:

$$\langle O \rangle = E[O] = \int \mathcal{D}\phi O(\phi)\mathcal{P}(\phi) \tag{2.15}$$

Once the chain has been given enough time, such that the sequence of samples ϕ_i effectively represents the typical set[1], it is possible to estimate Eq. 2.15 with:

$$\hat{O}_N = \frac{1}{N} \sum_i O(\phi_i) \tag{2.16}$$

This estimator is guaranteed to converge to the correct value as the Law of Large numbers applies to Markov Chains [11], that is:

$$\lim_{N \to \infty} \hat{O}_N = E[O] \tag{2.17}$$

Furthermore, a version of the Central Limit Theorem also applies to this process, which allows us to establish statistical uncertainties for our estimator when the number of samples is finite.

Theorem 1. *After a Markov Chain has equilibrated, the estimator $\hat{O}_N$ will satisfy [11]:*

$$\hat{O}_N - \mathcal{N}(E[O], \frac{\sigma_O^{MCMC}}{\sqrt{N}}) \tag{2.18}$$

The theorem states that the estimator will be distributed as a normal distribution centred in its asymptotic value and with a standard deviation equal to $\sigma = \frac{\sigma_O^{MCMC}}{\sqrt{N}}$ [11]. Furthermore, it is possible to express the standard deviation as:

$$\sigma = \frac{\sigma_O^{MCMC}}{\sqrt{N}} = \sqrt{\frac{Var[O]\tau_{IAT}}{N}} \tag{2.19}$$

where τ_{IAT} is the Integrated Autocorrelation Time (IAT), and it appears in the equation to take into account the autocorrelation of the samples in the Markov Chain. The method of calculating the IAT is detailed in Section 2.3.1.

In conclusion, if we are able to generate a Markov Chain whose target distribution is $\mathcal{P}(\phi) = \frac{\exp\{-S_E\}}{Z}$ then we will be able to estimate $\langle O \rangle$ defined in Eq. 2.10 through the estimator $\hat{O}_N$ and, perhaps even more importantly, we will be able to provide errors (in a statistical sense) to our estimates. The process of using Markov Chains to estimate these integrals is known as Markov Chain Monte Carlo (MCMC).

It is important to note that the convergence of Markov Chain Monte Carlo and the applicability of the Central Limit Theorem is, in general, more complicated than the overview given in this report. More details can be found in [12].

2.2   Hybrid Monte Carlo

We now need an algorithm that can be used to create the Markov Chain that explores the Typical Set of a given probability distribution. These algorithms typically operate in

---

[1]The chain typically needs some time to equilibrate, the samples generated in the equilibration process are excluded from the analysis.

a two-stage process: the proposal of a new field configuration ϕ, followed by its evaluation for acceptance or rejection. This is performed in a way that the algorithm satisfies detailed balance, making the target distribution invariant under the transition probability.

HMC is the most commonly used algorithm to generate the Markov Chain in Lattice Field Theory. It was initially proposed in 1987 in [13], and it aims to be able to propose global updates to the field configurations - proposing a change to all the components of ϕ simultaneously- are updated at once) with high acceptance rates such that usually, the new configurations are accepted. In HMC, one defines an auxiliary momentum field π conjugate to ϕ and introduces a Hamiltonian H defined as:

$$H \equiv \frac{1}{2}\pi \cdot \pi + S(\phi) \tag{2.20}$$

Furthermore, a fictitious or computer time is introduced such that π and ϕ satisfy Hamilton's equations in such time, that is:

$$\dot{\phi} = \frac{\partial H}{\partial \pi} \tag{2.21}$$

$$\dot{\pi} = -\frac{\partial H}{\partial \phi} = -\frac{\partial S}{\partial \phi} \tag{2.22}$$

In order to propose a new field configuration, first the momentum π is chosen from a distribution $P_G(\pi)$, which is usually taken to be $P_G \propto \exp\{\frac{1}{2}\pi \cdot \pi\}$. The tuple formed by (ϕ, π) forms a point in phase space, which is then evolved a fixed amount in fictitious time t using Hamilton's Equations, such that the probability of proposing a new phase space configuration $((\phi^*, \pi^*))$ is given by:

$$\mathcal{P}_P(\phi, \pi \rightarrow \phi^*, \pi^*) = \delta[(\phi^*, \pi^*) - (\phi(t), \pi(t))] \tag{2.23}$$

where $(\phi(t), \pi(t))$ is the solution to Hamilton's Equations with boundary conditions:

$$(\phi(0), \pi(0)) = (\phi, \pi) \tag{2.24}$$

The new configuration in phase space can now be either accepted or rejected with an acceptance probability equal to:

$$P_A(\phi, \pi \rightarrow \phi^*, \pi^*) = min(1, \exp\{-\delta H\}) \tag{2.25}$$

$$\delta H = H(\phi^*, \pi^*) - H(\phi, \pi) \tag{2.26}$$

The introduction of the auxiliary momentum means that in order to recover the distributions over ϕ only, we need to marginalize over π, that is, the transition probability from configuration ϕ to ϕ^* is given by:

$$\mathcal{T}(\phi^*|\phi) = \int \mathcal{D}\pi \mathcal{D}\pi^* P_G(\pi) \, \mathcal{P}_P(\phi, \pi \rightarrow \phi^*, \pi^*) P_A(\phi, \pi \rightarrow \phi^*, \pi^*) \tag{2.27}$$

We now show that HMC satisfies the detailed balance condition, such that the target distribution is reached by the Markov Chain. In particular, we follow the proof presented in [13].

In order to show this, we need the dynamics to be time reversible and to preserve volumes in phase space. The first condition implies that $\mathcal{P}_P(\phi, \pi \to \phi^*, \pi^*) = \mathcal{P}_P(\phi^*, -\pi^* \to \phi, -\pi)$ and the latter guarantees that $\mathcal{D}\phi\mathcal{D}\pi = \mathcal{D}\phi^*\mathcal{D}\pi^*$. Both are immediately satisfied with Hamilton's equations.

The following identity is needed:

$$\exp\{-H(\phi, \pi)\}min(1, \exp\{-\delta H\}) = \exp\{-H(\phi^*, \pi^*)\}min(\exp\{\delta H\}, 1) \qquad (2.28)$$

Given this identity and that $\mathcal{P}(\phi)\mathcal{P}_G(\pi) \propto \exp\{-H(\phi, \pi)\}$ it follows that:

$$\mathcal{P}(\phi)\mathcal{P}_G(\pi)P_A(\phi, \pi \to \phi^*, \pi^*) = \mathcal{P}(\phi^*)\mathcal{P}_G(\pi^*)P_A(\phi^*, \pi^* \to \phi, \pi) \qquad (2.29)$$
$$= \mathcal{P}(\phi^*)\mathcal{P}_G(-\pi^*)P_A(\phi^*, -\pi^* \to \phi, -\pi) \qquad (2.30)$$

where the last step uses the invariance of H under the change $\pi \to -\pi$. We can now multiply both sides by $\mathcal{P}_P(\phi, \pi \to \phi^*, \pi^*)$ and we integrate out the final and initial momenta to obtain:

$$\int \mathcal{D}\pi\mathcal{D}\pi^*\mathcal{P}(\phi)\mathcal{P}_G(\pi)P_A(\phi, \pi \to \phi^*, \pi^*)\mathcal{P}_P(\phi, \pi \to \phi^*, \pi^*) \qquad (2.31)$$

$$= \int \mathcal{D}\pi\mathcal{D}\pi^*\mathcal{P}(\phi^*)\mathcal{P}_G(-\pi^*)P_A(\to \phi^*, -\pi^* \to \phi, -\pi)\mathcal{P}_P(\phi^*, -\pi^* \to \phi, -\pi) \qquad (2.32)$$

where we have used the time reversibility of the dynamics. Noting that the measure is invariant under $\mathcal{D}\pi\mathcal{D}\pi^* = \mathcal{D}(-\pi)\mathcal{D}(-\pi)^*$ and identifying the definition of the transition probabilities we arrive at the detailed balance condition:

$$\mathcal{P}(\phi)\mathcal{T}(\phi^*|\phi) = \mathcal{P}(\phi^*)\mathcal{T}(\phi|\phi^*) \qquad (2.33)$$

This proves that the transition probability leaves invariant the probability distribution of the ϕ field. In order for the Markov Chain to converge to the desired distribution from any starting point, it also needs to be ergodic. A discussion around the ergodicity of HMC is provided in [14].

2.2.1   HMC in practice

If the time evolution in HMC could be performed analytically, then the Hamiltonian would be exactly conserved along the phase space trajectory, always resulting in the proposed configurations being accepted. However, this is usually not the case, and one needs to resort to approximations. In particular, one breaks the continuous fictitious time evolution into N discrete updates in which the field and conjugate momentum are evolved a total of dt in time (commonly known as the integration step size), such that $t = N \cdot dt$.

The leapfrog integrator is a common scheme used to approximate the time evolution in HMC. In particular, this integrator has the property of preserving time reversal and conserving volumes in phase space (symplectic) for any values of dt, N. This is essential since we require these two properties to achieve detailed balance.

The different steps of the leapfrog integration are as follows:

One begins with a half-step in momentum, that is:

$$\pi\left(\frac{dt}{2}\right) = \pi(0) - \frac{dt}{2}\frac{\partial S(\phi(0))}{\partial\phi(0)} \tag{2.34}$$

Then one performs a full step in the fields:

$$\phi = \phi(0) + dt\pi\left(\frac{dt}{2}\right) \tag{2.35}$$

after these two initial updates, the following sequential updates are repeated from $n = 1$ up to $n = N - 1$:

$$\pi\left(ndt + \frac{dt}{2}\right) = \pi\left(ndt - \frac{dt}{2}\right) - dt\frac{\partial S(\phi(ndt))}{\partial\phi(ndt)} \tag{2.36}$$

$$\phi((n+1)dt) = \phi(ndt) + dt\pi\left(ndt + \frac{dt}{2}\right) \tag{2.37}$$

Finally one performs a half step in momentum so that π ends at the same time as ϕ:

$$\pi(t) = \pi\left(\left(N - \frac{1}{2}\right)dt\right) - \frac{dt}{2}\frac{\partial S(\phi(t))}{\partial\phi(t)} \tag{2.38}$$

Furthermore, it can be shown that the leapfrog integrator has local errors (between two updates) of order $\mathcal{O}(dt^3)$ and a global error of order $\mathcal{O}(dt^2)$ [13] which yields a better approximation than other techniques with a global error of $\mathcal{O}(dt)$ such as Euler's Method.

HMC (along with the leapfrog algorithm) can be used to generate a sequence of samples of the field ϕ, which will represent the Typical Set of the $\phi-$space. These samples can then be used to estimate the expected value of any given observable. Algorithm 1 summarizes the different steps in the HMC algorithm.

Algorithm 1 Hybrid Monte Carlo

1: **procedure** GENERATE CHAIN(N, N_t, dt)
2:     $\phi_0 = Initialize\ \phi_0$
3:     **for** $n = 1\ Up\ to\ n = N$ **do**
4:         $\phi_n = UPDATE(\phi_{n-1}, N_t, dt)$
5: **procedure** UPDATE(ϕ, N_t, dt)
6:     $\pi \leftarrow Array\ of\ N\ random\ conjugate\ Momenta\ distributed \propto \exp\left\{-\frac{\pi^2}{2}\right\}$
7:     $H = \frac{1}{2}\sum \pi_i^2 + S(\phi$
8:     $\pi(N_t dt), \phi(N_t dt) = leapfrog\_integration(\pi, \phi, N_t, dt)$
9:     $H_{new} = \frac{1}{2}\sum \pi(N_t dt)_i^2 + S(\phi(N_t dt))$
10:     $r \leftarrow Generate\ Uniform\ Random\ Number\ [0,1]$
11:     **if** $r \leq min\left(1, \exp\{-\left(H_{new} - H\right)\}\right)$ **then return** $\phi(N_t dt)$
12:     **elsereturn** ϕ

2.3 Data Analysis in MCMC for Field Theories

2.3.1 Error Estimation and autocorrelations

In Section 2.1.2.1, it was stated that, given the necessary conditions, a version of the Central Limit Theorem applies to Ergodic Markov Chains with a stationary distribution $\mathcal{P}(\phi)$. We will now give more details on the σ^{MCMC} appearing in such a version of the Central Limit Theorem and how it can be estimated.

Let us assume initially that the sequence of N samples in the Markov Chain ϕ_i are uncorrelated such that the samples $O_i = O(\phi_i)$ are also uncorrelated. In this case σ^{MCMC} can be easily computed:

$$\sigma^{MCMC} = \lim_{N\to\infty} N \cdot Var[\hat{O}_N] \tag{2.39}$$

In the case of uncorrelated variables, the variance of the sum is given by the sum of the variances, giving:

$$\sigma^{MCMC} = \lim_{N\to\infty} N \cdot Var[\hat{O}_N] = \lim_{N\to\infty} \cdot\frac{1}{N} Var[\sum_i O_i] = Var[O] \tag{2.40}$$

In this case, σ^{MCMC} is given by the usual variance of the observable of interest, which can be estimated using the usual unbiased estimator:

$$\hat{Var}[O] = \frac{1}{N-1} \sum_i (O_i - \hat{O}_N)^2 \tag{2.41}$$

In the case that the different samples are correlated, the expression gets more complicated as the variance of the sum is not equal to the sum of the variances as one needs to take into account the correlation between the variables:

$$Var[\hat{O}_N] = \frac{1}{N^2} \sum_i Var[O_i] + 2\frac{1}{N^2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Cov[O_j, O_j] \tag{2.42}$$

Where $Cov[O_j, O_j]$ is the covariance of the variables O_i, O_j, which is defined as:

$$Cov[O_j, O_j] \equiv (O_i - E[O_i])(O_j - E[O_j]) \tag{2.43}$$

Given that all the samples are distributed according to the target distribution, the first term is independent of i. Furthermore, the second term can be expressed as a single sum, giving:

$$Var[\hat{O}_n] = \frac{1}{n^2} Var[O_1] \cdot n + 2\frac{n}{n^2} \sum_{k=1}^{n-1} (1 - \frac{k}{n}) Cov[O_i, O_{i+k}] \tag{2.44}$$

It is common to define $Cov[O_i, O_{i+k}] = \gamma_k$, known as autocovariance. The correction $\frac{k}{n}$ is usually ignored as usually N is large, and due to the exponential nature of γ_k, this contribution is suppressed in the region where k is close to n [15]:

$$Var[\hat{O}_N] = \frac{1}{N} Var[O_1] + \frac{2}{N} \sum_{k=1}^{N-1} \gamma_k \tag{2.45}$$

12

After multiplying by n and taking the limit $n \to \infty$ we obtain:

$$\sigma^2_{MCMC} = Var[O] + 2\sum_k^\infty \gamma_k \tag{2.46}$$

Factorising $Var[O]$ out of this expression gives the standard definition of the IAT:

$$\sigma^2_{MCMC} = Var[O]\left(1 + 2\sum_k^\infty \frac{\gamma_k}{Var[O]}\right) = Var[O]\tau_{IAT} \tag{2.47}$$

Now that the definition of the IAT is clear, it is time to understand how to estimate it. On the one hand, it is possible to reliably estimate the autocorrelation $\rho_k \equiv \frac{\gamma_k}{Var[O]}$ through the estimator given by:

$$\hat{\rho}_k = \frac{1}{\hat{Var}[O](N-k)} \sum_{t=1}^{N-k} (O_t - \hat{O}_N)(O_{k+t} - \hat{O}_N) \tag{2.48}$$

It would be tempting to conclude that the estimator of the IAT $\hat{\tau}_{IAT}$ could be simply given by [16]:

$$\hat{\tau}_{IAT} = 1 + 2\sum_i^n \hat{\rho}_k \tag{2.49}$$

However, this would be a mistake since the variance of this estimator does not go to 0 as the number of samples is increased. This effect arises from the fact that the sum in the estimator suffers from a signal-noise problem for high values of k. The inclusion of the noise-dominated region in the sum distorts the value of τ_{IAT} given by this estimator and makes its variance of order 1 [16].

In order to avoid this problem, it is common to introduce a cut in the sum of Eq. 2.49 at some $M < N$ so that the inclusion of the noisy region is avoided. The standard choice is to choose M such that $M \geq 4\tau_{IAT}$ [16]. The estimator of the IAT is now given by:

$$\hat{\tau}_{IAT} = 1 + 2\sum_i^M \hat{\rho}_k \tag{2.50}$$

The variance of this estimator is given by [16]:

$$Var[\hat{\tau}_{IAT}] \approx \frac{2(2M+1)}{N}\tau_{IAT} \tag{2.51}$$

In order to estimate it, one replaces the theoretical value τ_{IAT} with its estimator [16].

2.3.2 Critical Slowdown

The integrated autocorrelation time measures the correlation between configurations in a Markovian process. As the error on an observable is proportional to $\sqrt{\frac{\tau_{IAT}}{N}}$, the effect of

the autocorrelations in the calculations is to reduce the number of effective measurements to $N_{eff} = \frac{N}{\tau}$. In the case of HMC and other algorithms, autocorrelations tend to rapidly increase as the continuum limit is approached, causing the IAT to diverge as the lattice spacing, a, is reduced [17]. Furthermore, the precise asymptotic relationship between the lattice spacing and the IAT is given by [17]:

$$\tau_{IAT} \approx a^{-z} \tag{2.52}$$

where z varies between theories and algorithms. Effectively, this means that the number of samples needed to attain a certain degree of statistical certainty increases rapidly as we approach the continuum limit. This problem is usually referred to as a critical slowdown.

In order to obtain a qualitative explanation of why critical slowdown arises in HMC, we need to examine the Fourier Decomposition of the field ϕ. The discretization of the Theory imposes a maximum and a minimum in the wavelength of the modes that can be included in such expansion, as the minimum allowed wavelength one can fit in the lattice is proportional to the lattice spacing, and the maximum is proportional to its volume. As one gets closer to the continuum limit, the ranges of allowed wavelengths increase in the continuum limit $a \to 0$ and $V \to \infty$.

The origin of the problem is that HMC treats both low wavelength (high energy) and high wavelength (low energy) modes equally, as there is a single integration step size in the molecular dynamics evolution of the fields and momentum. It turns out that the low wavelength modes set a strict upper limit on the integration step size so that they can be integrated completely, which means that the more physical high wavelength modes barely evolve in each molecular dynamics step (due to the step size being small), ending up with highly correlated configurations for a fixed number of integration steps N_t. [6]

2.3.3 Jackknife Method

The estimation of the error on any observable relies on the calculation of τ_{IAT}. However, the generalization of the method outlined above to calculate the IAT for quantities of the form $O = f(\langle A \rangle)$ is complicated (although possible [18]). We now introduce the Jackknife method, which provides a simple framework capable of estimating the errors on these observables.

As before, we have a set of N Monte Carlo measurements of a primary observable A, which we label as $\Omega = \{A_1, A_2, ...A_N\}$. The data is divided into M blocks with length l such that $l >> \tau_{IAT}$ to get rid of autocorrelations; we denote each of these blocks as $\Omega_i = \{A_{l \cdot i}, A_{(l \cdot i+1), A_{2l \cdot i}}\}$. Then, the average of A for all the data without the block i is calculated. This process is repeated for all M blocks, which generates M new samples of A, known as Jackknife samples, denoted as A_i^J:

$$A_i^J = \frac{1}{N - l} \sum_m A_m \tag{2.53}$$
$$A_m \in \Omega / \Omega_i$$

Furthermore, the Jackknife average and the usual average estimators are given by:

$$\hat{A}^J = \frac{1}{M} \sum_i A_i^J \tag{2.54}$$

$$\hat{A} = \frac{1}{N} \sum_i A_i \tag{2.55}$$

It turns out that an unbiased estimator for the expectation value of A is [18]:

$$E[A] = \langle A \rangle \approx M\hat{A} - (M-1)\hat{A}^J \tag{2.56}$$

The error on A is estimated by calculating the deviation of the set A_i^J of the average value $\hat{A}^J$:

$$\hat{\sigma}_A^2 = \frac{M-1}{M} \sum_i (A_i^J - \hat{A}^J)^2 \tag{2.57}$$

If, instead, we are interested in the error of a quantity like $f(\langle A \rangle)$, the procedure is similar; for each of the Jackknife samples A_i^J we calculate $f_i^J = f(A_i^J)$ and then the estimate on $f(A)$ and its error is simply [18]:

$$f(A) \approx Mf(\hat{A}) - \frac{M-1}{M} \sum_i f_i^J \tag{2.58}$$

$$\hat{\sigma}_f^2 = \frac{M-1}{M} \sum_i \left(f_i^J - \frac{1}{M} \sum_l f_l^J \right)^2 \tag{2.59}$$

Therefore, we now have a method which allows us to estimate the statistical errors of quantities of the form $f(\langle A \rangle)$

2.3.4 Correlations and Spectrum

One important class of observables in Field Theory are correlation functions, which allows us to study the bound spectrum (eigenstates of the Hamiltonian) of the theory of interest. Let $\mathcal{O}$ be an observable in the lattice with 0 expectation value $\langle \Omega | \mathcal{O} | \Omega \rangle = 0$ (this value can always be subtracted from a general observable). In general, $\mathcal{O}$ will be a function of the field ϕ, so its value depends both on the spatial position $\underline{x}$ and on the Euclidean time τ. It is possible to project out the spatial part by summing over it in the lattice:

$$\mathcal{O}(\tau) = \sum_{\underline{x}} \mathcal{O}(\tau, \underline{x}) \tag{2.60}$$

This is done to fix the state's (physical) momentum $\underline{p}$ produced by $\mathcal{O}(0) | \Omega \rangle$ to 0 [19]. Then, the temporal correlation function is defined as:

$$C(\tau) = \langle \Omega | \mathcal{O}^\dagger(\tau)\mathcal{O}(0) | \Omega \rangle = \langle \mathcal{O}(\tau)\mathcal{O}(0) \rangle \tag{2.61}$$

Where $\mathcal{O}^\dagger(\tau)$ is the hermitian conjugate of the operator $\mathcal{O}$. The time evolution operator gives the time evolution in Euclidean time of any operator:

$$\mathcal{O}^\dagger(\tau) = \exp\{H\tau\}\mathcal{O}^\dagger(0) \exp\{-H\tau\} \tag{2.62}$$

15

where H is the Hamiltonian of the Theory (not to be mistaken with the Hamiltonian in HMC). Using this relation and introducing a sum over the complete set of states of the Hamiltonian between the two operators leads to:

$$C(\tau) = \sum_n \langle \Omega | \mathcal{O}(\tau) | n \rangle \langle n | \mathcal{O}(0) | \Omega \rangle = \sum_n | \langle \Omega | \mathcal{O}(0) | \Omega \rangle |^2 \exp\{-(E_n - E_\Omega)\tau\} \quad (2.63)$$

Furthermore, since $|0\rangle = |\Omega\rangle$ and $E_\Omega = 0$ (everything is measured relative to the vacuum) and the Operator has been chosen such that $\langle \Omega | \mathcal{O}(0) | \Omega \rangle = 0$ this leads to:

$$C(\tau) = \sum_{n=1} a_n \exp\{-E_n\tau\} = \sum_{n=1} a_n \exp\{-m_n\tau\} \quad (2.64)$$

Given that the momentum has been forced to be 0, we have that $E_n = m_n$, the mass of the states. In the large time limit, this relation is reduced to:

$$C(\tau) \approx a_1 \exp\{-m_1\tau\} \quad (2.65)$$

This makes it possible to extract the mass of the lightest state of the theory if the values of $C(\tau)$ are known (extracted from lattice simulation). In practice, one needs to take into account the periodic boundary conditions which modify Eq. 2.64:

$$C(\tau) = \sum_n b_n \cosh\left(m_n(\tau - \frac{L}{2})\right) \quad (2.66)$$

In lattice calculations, one does not directly obtain m_1 but rather the dimensionless quantity am_1, which vanishes in the continuum limit. It is common to define the correlation length in lattice units as follows:

$$\xi = \frac{1}{am_1} \quad (2.67)$$

This quantity diverges as $a \to 0$, so it quantifies how close we get to the continuum (the larger ξ, the better).

If one wants to access different states of the spectrum, one must choose operators that only couple with those states. That is, if state $|n\rangle$ is of interest, we want an operator which satisfies $\mathcal{O}|\Omega\rangle \approx |n\rangle$ as closely as possible (and more importantly that does not couple with lighter states). Usually, one selects a range of operators with the properties of the state(s) of interest and solves a generalized eigenvalue problem.

2.4   Fourier Acceleration

We now introduce the main technique used in this report in order to reduce critical slowdown: Fourier Acceleration. In order to motivate it, we follow [20] in which this technique was originally proposed.

Consider a scalar free field theory, whose action can be expressed as:

$$S(\phi) = \phi \cdot (m^2 - \partial^2) \cdot \phi = \phi \cdot K \cdot \phi \to K = (m^2 - \partial^2) \quad (2.68)$$

We use the dot notation to hide all the sums (or integrals in the continuum) in the action. In order to implement HMC, we define a conjugate momenta π and define the Hamiltonian H:

$$H = \frac{1}{2}\pi \cdot \pi + S(\phi) \tag{2.69}$$

This Hamiltonian leads to the following Equations of Motion for the field and momenta in fictitious time t:

$$\dot{\pi} = -(m^2 - \partial^2) \cdot \phi \tag{2.70}$$
$$\dot{\phi} = \pi \tag{2.71}$$

If we expand ϕ in terms of plane wave solutions with momentum p, then these Equations of Motion lead to the dispersion relation given by:

$$w(p) = p^2 + m^2 \tag{2.72}$$

That is, the evolution rate depends on the momentum of each mode. The modes with higher momentum evolve at a higher rate than those with a lower one. The upper bound on the integration step size dt of leapfrog integration is severely limited by the high-momentum modes, and in general, the maximum dt is inversely proportional to the maximum possible frequency [20]. On the other hand, each mode has a period $T \propto \frac{1}{w(p)}$. Therefore, the number of steps required for the mode with lowest momentum to complete a whole cycle is given by:

$$\frac{T}{dt} = \frac{w(p_{max})}{w(p_{min})} \tag{2.73}$$

Critical slowdown occurs when this ratio diverges, as this would mean that the slow modes barely evolve for a fixed number of steps. This is problematic since the interesting physics is encoded in these slow modes [20]. With the current dispersion relation, $w(p)^2 = p^2 + m^2$, it is clear that the ratio of Eq. 2.73 when we approach the continuum limit, which causes p_{max} to diverge and p_{min} to vanish.

A possible way of avoiding critical slowdown is to modify the dynamics (Hamiltonian) such that this ratio does not diverge [20]. This technique is commonly referred to as Fourier Acceleration (FA). If instead of defining the Hamiltonian as in Eq. 2.69, we define it as:

$$H = \frac{1}{2}\pi \cdot K^{-1}(M) \cdot \pi + S(\phi) \tag{2.74}$$
$$K^{-1}(M) = (M^2 - \partial^2)^{-1} \tag{2.75}$$

Where K^{-1} is the inverse of the kernel appearing in the action (an operator acting on the conjugate momentum). This modification leads to the altered Equations of Motion:

$$\dot{\pi} = -(m^2 - \partial^2) \cdot \phi \tag{2.76}$$
$$\dot{\phi} = (M^2 - \partial^2)^{-1} \cdot \pi \tag{2.77}$$

If we now repeat the expansion of the field ϕ in terms of plane waves, then these Equations of Motion lead to the dispersion relation given by:

$$w(p) = \frac{p^2 + m^2}{p^2 + M^2} \tag{2.78}$$

If M is chosen to be equal to m, then the evolution rate w becomes independent of p, meaning that all modes evolve at an equal rate. Therefore, this modification completely avoids critical slowdown as the ratio of Eq. 2.73 no longer diverges as the continuum limit is approached.

The price of avoiding critical slowdown comes from the Equations of Motion being more difficult to solve. Fourier Space is the most efficient way of working with them since the Kernel in this space [20], hence the name Fourier Acceleration. Furthermore, the modification of the Hamiltonian complicates the initial sampling of the canonical momentum, which is also efficiently performed in Fourier Space.

2.4.1 FA in Discrete Scalar Free Field Theory

In order to illustrate how FA HMC can be implemented in general, we will explicitly show the steps needed to be followed in the case of the discrete free scalar theory in d dimensions; for convenience, we will work in units of lattice spacing (effectively setting $a = 1$) since it is not relevant to our discussion.

We begin with traditional HMC, for which the conjugate momenta ϕ_x and the Hamiltonian are introduced:

$$H = \frac{1}{2} \sum_x \pi_x^2 + S(\phi) \tag{2.79}$$

In turn, this leads to the Equations of Motion given by:

$$\dot{\pi}_x = -\frac{\partial S}{\partial \phi_x} \tag{2.80}$$

$$\dot{\phi}_x = \pi_x \tag{2.81}$$

Upon discretisation, it is possible to express the action of this theory as:

$$S = \sum_x \left(m^2 \phi_x^2 - \sum_\mu (\phi_x \phi_{x+\mu} + \phi_x \phi_{x-\mu} - 2\phi_x^2) \right) \tag{2.82}$$

We can rewrite it in a form that resembles Eq. 2.68 in order to identify the discrete version of the Kernel:

$$S = \sum_{x,y} \phi_x K_{x,y} \phi_y \rightarrow K = \sum_\mu (2\delta_{x,y} - \delta_{x,y+\mu} - \delta_{x,y-\mu}) + m^2 \delta_{x,y} \tag{2.83}$$

The Fourier transform and its inverse are defined as:

$$\tilde{f}(k) = \mathcal{F}[f(x)]_k = \sum_x f(x) \exp\left\{-2\pi i \frac{k \cdot x}{L}\right\} \tag{2.84}$$

$$f(x) = \mathcal{F}^{-1}[\tilde{f}(k)]_x = \frac{1}{L^d} \sum_k \tilde{f}(k) \exp\left\{2\pi i \frac{k \cdot x}{L}\right\} \tag{2.85}$$

Some algebra reveals that the Fourier Transform of the Kernel is given by:

$$\tilde{K}_{k,k'} = \delta_{k,k'} \left(\sum_\mu \sin\left(\frac{\pi k_\mu}{L}\right) + m^2 \right) \tag{2.86}$$

This, as expected, indicates that the kernel is diagonal (meaning that we can label it by a single index k) in Fourier space, making its inversion trivial.

In order to implement Fourier Acceleration, we modify the momentum term of the Hamiltonian (known as the Kinetic Term), which leads to the modification of the Equations of Motion:

$$H = \frac{1}{2} \sum_x \pi_x^2 + S(\phi) \rightarrow \frac{1}{2} \sum_{x,y} \pi_y K_{x,y}^{-1} \pi_x + S(\phi) \tag{2.87}$$

$$\dot{\phi}_x = \pi_x \rightarrow \sum_y K_{x,y}^{-1} \pi_y \tag{2.88}$$

The Equation of Motion is now more complicated. However, it is possible to work in Fourier Space, where the inverse kernel is diagonal, and then convert the result back to configurational space:

$$\dot{\tilde{\phi}}_k = \tilde{K}_k^{-1} \tilde{\pi}_k \rightarrow \dot{\phi}_x = \mathcal{F}^{-1}[\tilde{K}_k^{-1}\tilde{\pi}_k]_x \tag{2.89}$$

On the other hand, the sampling of the initial momentum is also modified due to the introduction of the kernel in the kinetic term (the term that depends on the momentum) of the Hamiltonian. This is, again, easier to do in Fourier Space.

Expressing the Hamiltonian in terms of $\tilde{\pi}$ reveals how the momentum is distributed in Fourier Space:

$$H = \frac{1}{2L^d} \sum_k \frac{|\tilde{\pi}_k|^2}{\sum_\mu \sin\left(\frac{\pi k_\mu}{L}\right) + M^2} + S(\phi) \tag{2.90}$$

Hence, $\tilde{\pi}$ needs to be sampled from a distribution given by:

$$\prod_k \exp\left\{ -\frac{1}{2L^d} \frac{|\tilde{\pi}_k|^2}{\sum_\mu \sin\left(\frac{\pi k_\mu}{L}\right) + M^2} \right\} = \prod_k \exp\left\{ -\frac{1}{2L^d} \frac{|\tilde{\Pi}_k|^2}{\sum_\mu \sin\left(\frac{\pi k_\mu}{L}\right) + M^2} \right\} \tag{2.91}$$

In the last step, we have introduced the real-valued object $\tilde{\Pi}$. Given that π is real, $\tilde{\pi}$ only has L^d independent degrees of freedom. In order to sample correctly $\tilde{\pi}$ we organise those L^d real values into the object $\tilde{\Pi}$ with the additional constrain $\sum_k |\tilde{\pi}_k|^2 = \sum_k |\tilde{\Pi}_k|^2$.

The process of sampling the momentum is the following: The components of $\tilde{\Pi}$ are each sampled from a Gaussian distribution with 0 mean and standard deviation equal to $\sqrt{\frac{L^d}{\tilde{K}_k^{-1}}}$, then those components are used to reconstruct $\tilde{\pi}$ and finally the inverse Fourier Transform is taken to obtain π. The precise relation between $\tilde{\Pi}$ and $\tilde{\pi}$ for the two-dimensional case is presented in the appendix.

$SU(N) \times SU(N)$ Chiral Model

> It's a simple idea, but also stupid. Thing is, when stupid ideas work, they become genius ideas. We'll see which way this one falls.
>
> Andy Weir, Project Hail Mary

This chapter begins with an introduction to the $SU(N) \times SU(N)$ Principal Chiral Model and a brief motivation of our focus on such a model. Then, a review of the basic properties of Lie Groups is presented, as these are essential to understanding the model. Finally, the last two sections introduce our implementation of HMC and Fourier Acceleration for the Chiral Model on the Lattice.

3.1 The Model

In the Continuum, the two-dimensional Principal Chiral Model is defined through the Lagrangian:

$$\mathcal{L} = \frac{1}{T} Tr\{\partial_\mu U \partial_\mu U^\dagger\} \tag{3.1}$$

where $U(x,t) \in SU(N)$ and T is the coupling constant. The action is invariant under the field's simultaneous left and right multiplication by two independent $SU(N)$ matrices. This means the model is invariant under the transformation given by:

$$U(x) \to \Lambda_L U(x) \Lambda_R \tag{3.2}$$

where both $\Lambda_{L,R} \in SU(N)$. Therefore, we say that the model has a global $SU(N) \times SU(N)$ symmetry.

Moreover, some quantities are known for these models in the continuum limit. For instance, the bound state mass spectrum is given by [21]:

$$M_r = M \frac{\sin\left(\frac{r\pi}{N}\right)}{\sin\frac{\pi}{N}} \tag{3.3}$$

where M is the mass of the lightest state and r goes from 1 up to $N-1$. This theory's bound is quite restrictive, allowing only $N-1$ states, many of which have degenerate masses. For

instance, for $SU(3)$, only two bound states exist with equal masses. One of the goals of the project will be to verify the spectrum for $SU(4)$, which has two independent mass ratios and has not been done previously in the literature.

broader relevance of the two-dimensional $SU(N) \times SU(N)$ Principal Chiral Model stems from its shared properties with non-abelian gauge theories, notably Quantum Chromodynamics (QCD). These similarities include the fields being elements of $SU(N)$, confining and asymptotic freedom [21]. These shared properties ensure that the Chiral Model captures the essence of QCD while remaining computationally manageable.

The usual discretization of the action of this model is given by [21]:

$$S = -\beta N \sum_{x,\mu>0} Tr\{U_x^\dagger U_{x+\mu} + U_{x+\mu}^\dagger U_x\} \tag{3.4}$$

β is defined as $\beta = \frac{1}{NT}$.

3.1.1 The $SU(N)$ Group

The Special Unitary Group SU(N) is a compact and simple real Lie group of dimension (i.e. number of generators) equal to $N^2 - 1$. In its defining or fundamental representation, the group is represented by the group of $N \times N$ unitary matrices with determinants equal to one. The associated real Lie algebra to this group is denoted as $\mathfrak{su}(N)$, and its fundamental representation corresponds to the vector space spanned by the set $N \times N$ hermitian and traceless matrices. Given a particular representation, it is possible to define the group's generators in that representation as a basis of the Lie algebra (basis in the context of vector spaces).

The different elements of the group (connected to the Identity) in any representation are obtained through the exponentiation of the Lie Algebra:

$$U = \exp\{i\alpha^a T_a\} \tag{3.5}$$

where T_a are the generators of the $\mathfrak{su}(N)$ Lie algebra in such representation and a is being summed over. We will now only focus on the defining or fundamental representation of the $SU(N)$ group. In the case of $SU(2)$ and $SU(3)$, a possible choice of generators for their associated Lie algebras are the Pauli and the Gell-Mann matrices, respectively. It is possible to generalise the Gell-Mann for the general $SU(N)$. In particular, we need a set of $N^2 - 1$ matrices $\{T_i^N | i = 1, ..., N-1\}$ (all of which are $N \times N$) which are traceless, hermitian and form a basis. The process to obtaining these for general N is described in [22] and can be checked in Appendix ??

3.2 Hybrid Monte Carlo

To implement the HMC algorithm, we need to define a conjugate momenta to the field U and a Hamiltonian, which dictates the Equations of Motion of the field and the momenta in

the fictitious time t. Note that t does not denote the time position in the lattice but rather the fictitious time.

As mentioned before, the field U_x corresponds to $SU(N)$ elements in its fundamental representation, and they can be expressed as:

$$U_x = \exp\{i\alpha^i(x,t)T_i\} \tag{3.6}$$

where $\alpha^i(x,t) \in \mathbb{R}$ are the degrees of freedom of the field, and as usual, the T_i are the generators of the group (the Generalized Gell-Mann Matrices in our case). Therefore, it is possible to define the conjugate momenta p_x^i as the time derivative of $\alpha^i(x,t)$:

$$p^i(x,t) = \frac{d\alpha^i(x,t)}{dt} \tag{3.7}$$

$$p_x = p^i(x,t)T_i \tag{3.8}$$

This already fixes the time evolution of U_x and $U_x^\dagger$ in terms of p_x:

$$\frac{d}{dt}U_x = \dot{U}_x = ip_x U_x \tag{3.9}$$

$$\frac{d}{dt}U_x^\dagger = \dot{U}_x^\dagger = -iU_x^\dagger p_x \tag{3.10}$$

We can now define a Hamiltonian $H = T + S$ (with S being the model's action) to obtain the Equations of Motion of the momentum. To obtain these equations, we need to require that:

$$\frac{d}{dt}H = \dot{H} = \dot{T} + \dot{S} = 0 \tag{3.11}$$

The Kinetic Term T (the one that depends on the conjugate momenta) needs to be defined such that the Equation of motion $\alpha^i(x,t)$ is given by Hamilton's Equations:

$$T = \frac{1}{2K}\sum_x Tr\{p_x p_x\} = \frac{1}{2K}\sum_x p_i(x,t)p_j(x,t)Tr\{T_iT_j\} = \frac{1}{2}\sum_x p_i(x,t)p_i(x,t) \tag{3.12}$$

where K is defined from the condition: $Tr\{T_iT_j\} = K\delta_{ij}$. Finally, obtaining the Equations of Motion for the momenta relies on finding $\dot{T} + \dot{S} = 0$. The first term is easy:

$$\dot{T} = \sum_x p_i(x,t)\dot{p}_i(x,t) = \frac{1}{K}\sum_x Tr\{p_x\dot{p}_x\} \tag{3.13}$$

However, obtaining a similar expression for $\dot{S}$ requires more effort:

$$\dot{S} = -\beta N \sum_{x,\mu>0} Tr\{\dot{U}_x^\dagger U_{x+\mu} + U_x^\dagger \dot{U}_{x+\mu} + \dot{U}_{x+\mu}^\dagger U_x + U_{x+\mu}^\dagger \dot{U}_x\}$$

$$= i\beta N \sum_{x,\mu>0} Tr\{p_x\left((U_{x+\mu} + U_{x-\mu})U_x^\dagger - h.c\right)\} \tag{3.14}$$

22

The last step requires relabelling the sum over x by $x \to x + \mu$ in the second term to reduce the expression. Requiring now the condition of Eq. 3.11 leads to the identification of the Equation of Motion for p_x.

$$\dot{p}_x = -i\beta KN \sum_{\mu>0} \left((U_{x+\mu} + U_{x-\mu}) U_x^\dagger - h.c \right) \tag{3.15}$$

Having now the two Equations of Motion for p_x and U_x, we can proceed to integrate them using the leapfrog integration to evolve the field and the momenta in fictitious time. The discrete version of Eq. 3.9 is:

$$U(x, t + dt) - U(x, t) = ip\left(x, t + \frac{dt}{2}\right) U(x, t) dt \tag{3.16}$$

Up to the first order, this corresponds to the exponential update given by:

$$U(x, t + dt) = \exp\left\{ ip\left(x, t + \frac{dt}{2}\right) dt \right\} U(x, t) \tag{3.17}$$

Likewise, the discrete version of Eq. 3.15 is given by:

$$p\left(x, t + \frac{dt}{2}\right) = p\left(x, t - \frac{dt}{2}\right) + dt\dot{p}(x, t) \tag{3.18}$$

The HMC parameters dt, N_t are chosen such that $dt \cdot N_t = 1$ and the Acceptance rate of the runs (for a given value of β, L) lies within $[0.65, 0.85]$.

3.2.1 Group Features of the HMC

3.2.1.1 Removal of the Trace

The first potential issue is that the conjugate momenta p_x are elements of the algebra $\mathfrak{su}(N)$ for any possible time t, which implies that they are $N \times N$ hermitian and traceless matrices. This implies that its time derivative must also be traceless, that is, $Tr\{\dot{p}_x\} = 0$. However, the Equation of Motion for p_x in Eq. 3.15 only guarantees this constraint for the case of $SU(2)$ as the trace of U_x in this case is always real. Therefore, for general N, we will need to explicitly remove the trace in Eq. 3.15 to guarantee the traceless condition of $\dot{p}_x$ and hence of p_x at any time t.

To remove the trace of an $N \times N$ matrix M, we need to subtract the Identity multiplied by the trace and divided by N, that is:

$$M \to M - \frac{Tr\{M\}}{N} \mathbb{I} \tag{3.19}$$

3.2.1.2 Exponential of the Algebra

Being able to perform the Exponential Update of Eq. 3.16 relies on us being able to calculate the exponential of the algebra (as $p(x, t + \frac{\epsilon}{2})\epsilon \in \mathfrak{su}(N)$). In the case of $SU(2)$ it is possible

to express the group elements as $U \equiv \exp\{i\alpha^i \sigma_i\} = \cos(\alpha)\mathbb{I} + i\sin(\alpha)(\hat{\alpha}^i \sigma_i)$, where we define $\alpha = \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2}$ and $\hat{\alpha}_i = \frac{\alpha_i}{\alpha}$. It is also possible to reexpress this in terms of four real numbers c_0, c_1, c_2, c_3, c_4 as: $U = c_0\mathbb{I} + c_i\sigma^i$ (with i going from 1 to 4). The connection to the previous coefficients is trivial as $c_0 = \cos(\alpha), c_i = \sin\alpha\hat{\alpha}_i$ and they satisfy the constraint $c_0^2 + c_1^2 + c_2^2 + c_3^2 = 1$. This allows us to easily perform the exponential update in this case.

However, it is not possible to extend this simple procedure for general N, so another method must be introduced. One possibility is to define the matrix exponential through its Taylor series, which has an infinite radius of convergence:

$$\exp\{M\} = \sum_{n=0}^{\infty} \frac{(M)^n}{n!} \tag{3.20}$$

To approximate the exponential of a matrix, we can truncate the Taylor series at some big n. This approach can be combined with the property :

$$\exp\left\{\frac{M}{k}k\right\} = \left(exp\left\{\frac{M}{k}\right\}\right)^k \tag{3.21}$$

The trick is to choose a relatively big k to make the Taylor series converge faster, and then we take the matrix power of the result. In general, this procedure reduces the time required to estimate the exponential. Using the Taylor series to estimate the exponential update increases the computational resources needed compared to the procedure for $SU(2)$.

3.2.1.3   Reunitarisation

Finally, it is crucial to consider that the field U_x will start to move away from the $SU(N)$ group manifold due to the error introduced by approximating the exponential update as a truncated Taylor series. Even in the case of $SU(2)$, small numerical errors in the computation will eventually compound and cause the same effect. Consequently, it is necessary to project U_x back into the group manifold. However, as the fields are displaced from the group by a small amount in each Molecular Dynamics step, it will not be necessary to project back to the group in each iteration of the HMC. The exact number of iterations between each projection will be kept as a parameter.

3.3   Fourier Acceleration

In order to implement Fourier Acceleration for this model, we begin by identifying the kernel of the action:

$$S = -2\beta N \sum_{x,\mu>0} ReTr\{U_x^{\dagger}U_{x+\mu}\} \tag{3.22}$$

We would like to write the action in the form:

$$S = \beta N \sum_{x,y} ReTr\{U_x^{\dagger}K_{x,y}U_y\} \tag{3.23}$$

where $K_{x,y}$ is the kernel we are trying to identify. Comparing the two equations, it can be shown that a valid choice for the kernel is given by:

$$K_{x,y} = \sum_{\mu > 0} (2\delta_{x,y} - \delta_{x,y+\mu} - \delta_{x,y-\mu}) \tag{3.24}$$

Which is similar to the Kernel obtained in the free field theory. In Fourier Space, the Kernel is diagonal, and it is given by:

$$\tilde{K}_{k,k'} = \sum_{\mu} (2 - e^{\frac{2\pi i}{L} k \cdot \mu} - e^{\frac{2\pi i}{L} k \cdot \mu}) \delta_{k,k'} = \sum_{\mu} 4 \sin^2 \left(\frac{\pi k \cdot \mu}{L} \right) \delta_{k,k'} \tag{3.25}$$

In order to make the inverse kernel well-defined for all values of k, we introduce the acceleration mass parameter such that the inverse Kernel in Fourier space is given by:

$$\tilde{K}_{k,k'}^{-1}(M) = \frac{\delta_{k,k'}}{\sum_{\mu} 4 \sin^2(\frac{\pi k \cdot \mu}{L}) + M^2} \tag{3.26}$$

This is the same Kernel as in the Scalar Free Field Theory case. In order to implement Fourier Acceleration, one introduces the inverse kernel in the kinetic term of the Hamiltonian such that:

$$T = \frac{1}{2K} \sum_{x,y} Tr\{p_x K_{x,y}^{-1}(M) p_y\} = \frac{1}{2} \sum_{x,y} p_x^i K_{x,y}(M)^{-1} p_y^i \tag{3.27}$$

In turn, this modifies the equations of motion of the field[1], which are now given by:

$$\dot{\alpha}^i(x,t) = \sum_y K_{x,y}^{-1} p^i(y,t) \tag{3.28}$$

$$\dot{U}_x = i \sum_y K_{x,y}^{-1} p_y U_x \tag{3.29}$$

As in the case of the free field theory, these equations are easier to solve in Fourier space. In the leapfrog integration, the field is updated according to:

$$U_x(t + dt) = \exp\left\{ i \mathcal{F}^{-1}[\tilde{K}_k^{-1}(M) \tilde{p}_k(t)]_x dt \right\} U_x(t) \tag{3.30}$$

On the other hand, it is possible to express the kinetic term in Fourier space:

$$T = \frac{1}{L^2} \sum_k \frac{|\tilde{p}_k^i|^2}{4 \sin^2(\frac{\pi k \cdot \mu}{L}) + M^2} \tag{3.31}$$

Which reveals how the conjugate momenta is distributed in Fourier Space. Given that the Kernel is identical to the Free Field case, it is possible to use the same framework outlined in Section 2.4.1 in order to sample $\tilde{p}_k^i$ and, in turn, reconstruct p_x^i.

---

[1]the momentum suffers no modifications since no alterations were made to the action part of the Hamiltonian.

Results

This chapter will review the FA HMC algorithm's results for the $SU(N) \times SU(N)$ principal Chiral Model. After briefly assessing our implementation's validity, we will examine the values obtained for different observables for several values of N. The energy density, the correlation length, and the mass spectrum will be some of the quantities examined. Finally, we will conclude with a comparison of the performance of the FA HMC against the non-accelerated version of the algorithm.

4.1 Model Verification: Energy Density

We begin our analysis of the $SU(N) \times SU(N)$ Chiral Model by looking at the energy density e, which will allow us to verify our implementation of the FA algorithm. This observable is commonly defined as [21]:

$$e = 1 - \frac{1}{2D}\frac{\partial f}{\partial \beta} \tag{4.1}$$

$$f = \frac{1}{VN^2}\log(Z) \tag{4.2}$$

where V is the volume of the Lattice, and D is the number of dimensions of our model. It is easy to see that $\frac{\partial f}{\partial \beta} = \frac{1}{VN^2}\frac{\partial}{\partial \beta}\log(Z) = \frac{1}{VN^2}\frac{1}{Z}\frac{\partial}{\partial \beta}(Z) = \frac{1}{VN^2}\langle\frac{-S}{\beta}\rangle$. Therefore, the final expression for e in the Lattice is given by:

$$e = 1 - \frac{1}{2DVN^2}\left\langle -\frac{S}{\beta}\right\rangle \tag{4.3}$$

Expansions of the energy density in powers of β are available for the regions of low β (strong coupling) and high β (weak coupling) for all possible N (these are available in [21; 23]). This allows for a comparison between the values produced in our simulation and the expected ones from the asymptotic expansions, providing a method to validate our implementation.

Furthermore, since e is independent of the volume, the finite size effects are expected to be small. Hence, this quantity can be calculated in small lattices, making it a prime candidate for quickly verifying our implementation.

Fig. 4.1 shows our results obtained for the energy density for the cases $N = 2, 3, 6$ in a 16^2 lattice. Both the strong and the weak coupling expansions have been included in each plot to assess the validity of our results.
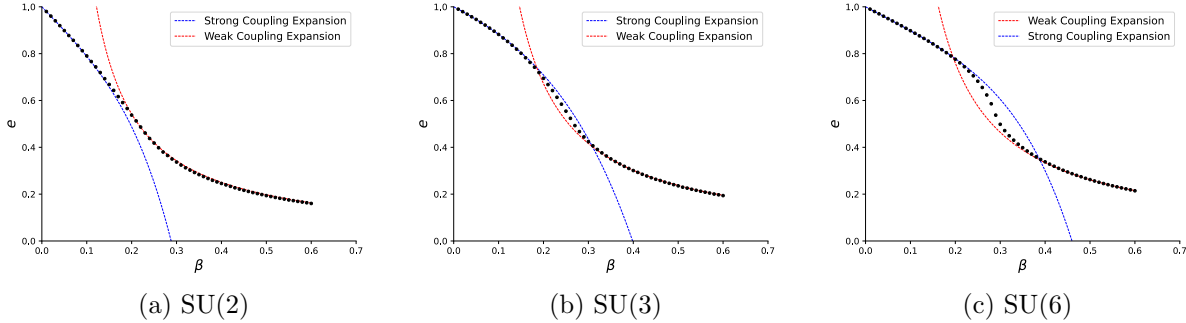

(a) SU(2)   (b) SU(3)   (c) SU(6)

Fig. 4.1: Energy Density for different N, the lines indicate the asymptotic prediction. A 16^2 Lattice was used to obtain the results, 10^4 thermal runs were performed before taking any measurements, and 10^4 measurements were performed to obtain each data point. Moreover, errors are too small to be visible. The data agrees well with the predicted asymptotic expansions.

For all the explored values of N, our results agree remarkably well with the values predicted by the asymptotic expansions in their applicability regions. Furthermore, the statistical errors for this observable are significantly small, suggesting that it does not suffer from large auto-correlations.

In addition, our simulation offers insight into the transition region between the two asymptotic expansions. This is the region that changes the most when N is varied. In Fig. 4.1, it can be seen that the transition between the two asymptotic expansions becomes more pronounced as N increases. This feature hints at a phase transition in the $N \to \infty$ limit, as we expect that the slope eventually becomes a discontinuity in the energy density.

4.1.1   The Phase Transition and Heat Capacity

To study in more detail the change in the energy density shape as N is increased, it is helpful to define a heat capacity C_V as the derivative of the energy density with respect to the coupling T:

$$C_V \equiv \frac{1}{N} \frac{\partial}{\partial T} e = -\beta^2 \frac{\partial}{\partial \beta} e \tag{4.4}$$

It is possible to obtain an expression for the heat capacity C_V as a function of quantities obtainable through lattice calculations starting with the definition of e in Eq. 4.3. After

some manipulation, the heat capacity can be expressed as:

$$C_V = \frac{\beta^2}{2DVN^2} \left(\left\langle \left(\frac{S}{\beta} \right)^2 \right\rangle - \left\langle \frac{S}{\beta} \right\rangle^2 \right) \tag{4.5}$$

Results obtained from lattice calculations for different N are presented in Fig. 4.2. These results confirm both that there is a peak in the region where the model transitions between the strong and the weak coupling expansions in all the investigated values of N and that such a peak becomes more pronounced as N is increased. Furthermore, the location of the peak is more or less stable, but it seems to drift to slightly higher β values when N is increased (the higher the N, the less noticeable the displacement). Therefore, our results indicate that in the limit $N \to \infty$, the heat capacity will diverge at some finite critical value β^*, indicating a phase transition.

It is important to note that as the values were calculated in small lattices, they are prone to finite volume effects, which in this case are not as small as in the case of the energy density (this was discovered in the late stages of the project so recalculating all the points for several N was unfeasible). If the quantities are recalculated in bigger lattices, it is usually found that the overall height of the peak diminishes for all N values (some points were recalculated to confirm this). Nevertheless, we are more interested in the general behaviour of the heat capacity than in its precise values, so the analysis stands.
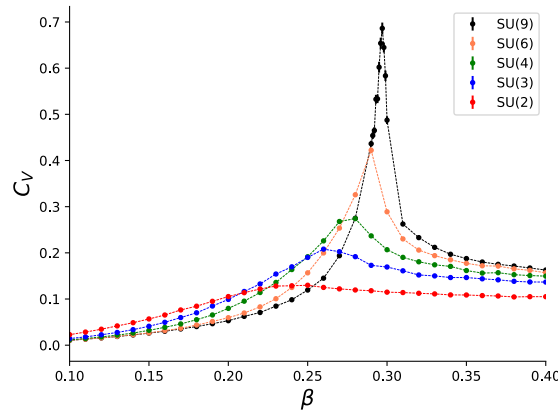

Fig. 4.2: Heat capacity for different N. A 16^2 Lattice was used to obtain the results, and 10^5 measurements were performed for each point with 10^4 initial thermalisation runs. The acceleration mass was fixed to 0.1 for all β values. As N increases the peak becomes more pronounced suggesting a divergence in the $N \to \infty$ limit.

This analysis of the energy density and the heat capacity of the Principal Chiral Model is by no means new. The same analysis was carried out in [21] with similar results; it is, however (at least to the best of our knowledge), the first time that this analysis is carried out with an FA HMC algorithm. All in all, the analysis of the energy density and the heat capacity allows us to validate our implementation of the FA HMC algorithm for the Principal Chiral Model.

4.2   The Correlation Length

The analogous quantity in this model of the correlation function introduced in Section 2.3.4 is the wall-to-wall correlation function defined as:

$$C_{ww}(d) = \frac{1}{V} \left\langle \sum_{x,y,\tau} 2ReTr(U_{x,\tau}U_{y,\tau+d}^{\dagger}) \right\rangle \qquad (4.6)$$

It is expected that this function follows the usual shape:

$$C_{ww}(d) = \sum_{i} A_i \cosh\left(m_i\left(d - \frac{N}{2}\right)\right) \qquad (4.7)$$

The correlation length, ξ, corresponds to the inverse of the mass of the lightest state m_i in units of lattice spacing[1], such that in the continuum limit, this quantity will diverge. The symmetry of the correlation function means that no new information is encoded in the region $(\frac{L}{2}, L]$, so the data in this region can be used to double the statistics of the first half.

It is possible to extract the correlation length directly from the correlation function by making a two-parameter fit (A_0, m_0) in the region where the ground state dominates, that is, in the region $d >> 0$ where the correlation function is, up to good approximation, given only by the first term in the sum of Eq. 4.7. It is important to note that the overall multiplicative factor of C_{ww} is irrelevant as it affects all the states equally; hence, it only scales the A_i equally.

Another possibility is to consider multiple points of the correlation function to obtain the following:

$$m_{eff}(d) = \cosh^{-1}\left(\frac{C_{ww}(d+1) + C_{ww}(d-1)}{2C(d)}\right) \qquad (4.8)$$

where m_{eff} will tend to the constant value m_i in the region where the first contribution of the sum in Eq. 4.7 dominates. This latter method has the advantage of being a one-parameter linear fit, which means that the value $\frac{\chi^2}{d.o.f.}$ can be used to assess the quality of the fit without any risks, but with the disadvantage of being noisier than the direct correlation function.

4.2.1   Fit Procedure

Given the importance of the correlation length value for this report's subsequent sections, we designed a precise fit procedure to extract this value from our lattice data that works well for all values of N. This section will outline and motivate our method.
Before specifying the method of our choice to obtain the correlation length of this model, it is useful to examine the main properties of the effective mass plots as they offer an interesting insight into which method should be adopted. Fig. 4.3 shows an example of the effective mass plot for the $SU(2)$ model and $\beta = 0.3$.

---

[1]Note that m_i in this section actually corresponds to the dimensionless quantity am_i and not directly the physical mass, the factor of a is omitted to simplify the discussion.
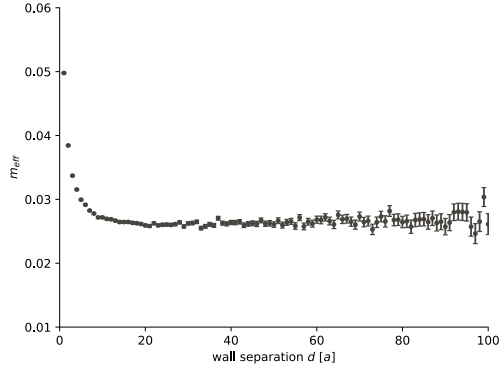
Fig. 4.3: Effective mass plot for $\beta = 0.316675$ in the SU(2) model. The simulation was run in a 400^2 Lattice with 10^4 thermalization runs and 10^5 measurements. Three distinct regions can be identified on the plot.

Three different regions exist in the effective mass plot:

1. Initial region in which the effective mass is decaying

2. Intermediate region in which the effective mass is constant

3. Final region in which the effective mass is dominated by noise.

These three different regions are present in all effective mass plots, regardless of the values of β, N chosen. Regions 2 and 3 were expected, as the former is the result we are seeking, and the latter arises due to a common signal-to-noise problem in correlation functions. Region 1 was unexpected and requires a more detailed analysis. The initial decaying effective mass is exactly what one expects in Lattice Field Theory due to the contamination of excited states in the signal for small distances. However, the spectrum of the $SU(N) \times SU(N)$ Chiral Model is exactly known, and the absence of excited states in $SU(2)$ should prevent this contamination. It should be noted that the exact solution is only valid in the continuum. Thus, this decaying region is most likely due to the discretization of the theory, a lattice artefact.

Furthermore, the decaying region is more stretched as β is increased, meaning that the separation d^* in units of lattice spacing at which the effective mass is a constant is increased as we approach the continuum limit. Fig. 4.4 shows exactly this behaviour by superimposing three effective mass plots for three different values of β in the SU(2) case. Section 4.5 gives a more detailed explanation of this effect.
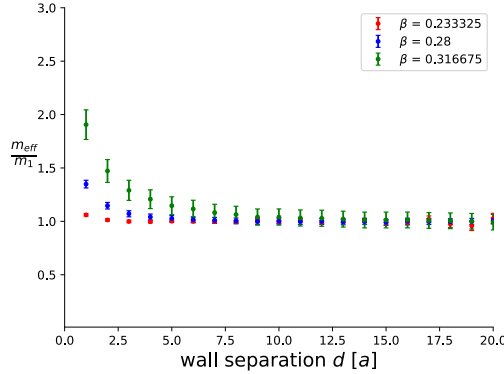
Fig. 4.4: Normalized effective mass plot for three different β values in the $SU(2)$ model. The size of the Lattices were 96^2 and 224^2 and 400^2 for $\beta = 0.233325$, 0.28 and 0.316675 respectively. The initial decaying region becomes stretched with increasing β.

If one directly plots the wall-to-wall correlation function, it may not be immediately clear how the initial region affects the final results as the deviation from the expected result (from a cosh relation in this case) is not as significant as the deviation present in the effective mass plot. Nevertheless, it should be noted that these effects impact the outcome, as the values of the parameters obtained by fitting the wall-to-wall data to a cosh heavily depend on the fit window (whether or not the initial points are included).

Moreover, it is hard to establish the correct fit window for the wall-to-wall correlation data since, in general, all the reduced chi-squared values $\chi_{red}^2 = \frac{\chi^2}{d.o.f.}$ produced by such fits are extremely low (around 0.01, even when including the points at low separations). There are a number of reasons that could explain this. Firstly, the cosh fit is a non-linear fit, meaning that the reduced chi-squared values might not directly indicate the fit's quality. Secondly, we are only doing uncorrelated fits, which might lead to an underestimation of the final value of χ^2 due to the correlated nature of the data points. However, in order to make correlated fits, we need to estimate the covariance matrix, which is a notoriously difficult task. Consequently, we will uniquely use the effective mass plots [2] to extract the correlation length.

In order to eliminate any possible human bias when choosing the fitting region, we used an adapted version of the Akaike Information Criterion (AIC) taken from [24]. AIC is usually used to compare the quality between different models for the same data, but in [24], it was adapted to account for different fitting ranges within the same model. In our case, we only have one model for the data, so this procedure was used to combine different fits with different fit windows, weighing them according to their quality and the number of points used in each of them.

In particular, in the adapted version of the AIC introduced in [24] for one-parameter fits,

---

[2]The χ^2 value associated with this plots is more sensitive to the fitting region and $\frac{\chi^2}{d.o.f.}$ can be used to assess the fit quality without risk.

the fit i out of a total of M fits is assigned a weight proportional to:

$$P_i \propto \exp\left\{-\frac{\chi_i}{2} - k_i - n_i\right\} \tag{4.9}$$

where χ_i is the chi-squared value associated with the fit, k_i is the number of fit parameters (equal to one in this case), and n_i is the number of points in the data set not used in the fit. Finally, the P_i's are normalized such that their sum is equal to unity. Then, the parameter value and its error are given by:

$$a = \sum_i P_i \cdot a_i \tag{4.10}$$

$$\sigma^2 = \sum_i P_i \cdot \sigma_i^2 + \sum_i P_i \cdot a_i^2 - \left(\sum_i P_i \cdot a_i\right)^2 \tag{4.11}$$

Given two fits with similar quality (χ^2), this version of AIC favours the one that includes more points such that most importance is given to the fits with good quality and bigger fitting regions. In practice, we select several starting and finishing points, and we weigh each of these fits using AIC. This allows us to minimize the bias when selecting the fitting window. This procedure was used to extract the correlation length of the $SU(N) \times SU(N)$ principal Chiral Model for different values of N, β; the results of this process are analyzed in the following section.

4.2.2 Correlation Length Results

In this section, we will present a small subset of the results concerning the correlation length ξ obtained through our simulations for several values of N. In particular, we will only present the results that can be compared to values obtained from different studies ([4; 21]) for these quantities.

We begin with $SU(2)$. In this case, it is possible to compare our values of ξ with the values obtained in [4], which also uses an FA HMC algorithm. Our results for small correlation lengths $(1 < \xi < 5)$ are all within $1 - \sigma$ of the values in this study, but small discrepancies appear at high ξ. These values are compared in Table 4.1.

β	L	ξ	ξ Ref [4]
0.233325	96	5.579(8)	5.542(6)
0.25	96	7.845(6)	7.9262(96)
0.266675	160	11.451(7)	11.407(13)
0.283325	160	16.898(15)	16.928(18)
0.3	224	25.50(3)	25.41(3)

Table 4.1: Comparison of correlation length values for SU(2) model with the data reported in [4]. All the points are separated by more than $2 - \sigma$. A possible interpretation is provided in the text.

The small discrepancies between the values at high correlations can be explained by the differences in the methods used to extract ξ. In [4], ξ was extracted directly by fitting the wall-wall correlation function to a cosh function starting from the point at 0 separation up to a cutoff determined by the reduced chi-squared metric[3]. However, as explained in the previous section, this would entail including points contaminated by the effect of lattice artefacts, as seen from the effective mass plots. In our case, we used the effective mass plot to extract ξ as it is more sensible to these effects, and care was taken not to include them in our fits. As these effects are minimal in the case of short correlation lengths, it would explain why our results agree in this region.

On the other hand, the errors in the estimates are of the same order, meaning that the two methods of extracting ξ yield a similar precision.

For the general case of SU(N), we turn to the values obtained by [21]. In this case, our values agree within uncertainties with the ones obtained in this study. Some values for $SU(3)$ are compared in Table 4.2:

β	L	ξ	ξ Ref [21]
0.225	24	1.895(6)	1.89(3)
0.27	48	4.805(10)	4.89(4)
0.29	82	8.02(2)	8.08(13)
0.3	90	10.44(2)	10.48(12)
0.315	120	15.42(3)	15.8(4)

Table 4.2: Correlation Length values obtained for the SU(3) model. Our data agrees well with the values obtained in [21]. Only one of the presented values is outside the $1 - \sigma$ agreement.

The agreement between the different data sets for both small ($1 < \xi < 5$) and large ($5 < \xi < 15$) for several $N = 3, 6, 9$ confirm the validity of the procedure outlined in Section 4.2.1.

4.3 Mass Spectrum

It was mentioned in Section 3.1 that the bound state spectrum of the Principal Chiral model in the continuum is given by:

$$M_r = M \frac{\sin\left(\frac{r\pi}{N}\right)}{\sin\left(\frac{\pi}{N}\right)} \tag{4.12}$$

This expression does not provide the exact values of the masses but rather the exact ratios between the excited state masses and the lightest state mass ($M = M_1$).

---

[3]Since this is a non-linear fit, the determination of the goodness of the fit with the reduced chi-squared metric is dubious.

It is possible to extract the masses of the excited states by looking at two-point functions of observables that couple with those states. The first two operators, which couple with the states with $r = 2, r = 3$, are given by [21]:

$$O^{(2)}_{abcd} = U_{ab}U_{cd} - U_{ad}U_{cb}$$

$$O^{(3)}_{abcdef} = U_{ab}U_{cd}U_{ef} - U_{ab}U_{cf}U_{ed} -$$
$$U_{ad}U_{cf}U_{eb} + U_{ad}U_{cd}U_{ef} + U_{af}U_{cb}U_{ed} - U_{af}U_{cd}U_{eb}$$

Which share the same transformation properties as the excited states they couple with.

In this section, we will now use one of these operators to analyse the spectrum of $SU(3)$ and $SU(4)$[4]. Beginning with $SU(3)$, the continuum solution predicts the existence of two distinct states which have equal masses. In order to test the degeneracy of the two existing states, we use the operator $O^{(2)}$

The procedure used to extract the mass of the excited state from the lattice data is exactly the same as the one outlined in Section 4.2.1. The ratio between the mass of the ground state and the first excited states for the $SU(3)$ model is presented in Table 4.3.

β	L	$\frac{M_2}{M}$
0.18	18	1.002(4)
0.27	48	0.994(4)
0.29	82	0.998(2)
0.3	90	0.998(2)
Continuum		1

,

Table 4.3: Ratio between the ground state mass and the first excited state mass for the $SU(3)$ at several β values

Our results show that even at short correlation ($\beta = 0.18$ corresponds to a correlation length of approximately 1) lengths, the ratio between the masses of these two states is equal to unity within errors, which confirms the degeneracy.

Continuing with $SU(4)$, according to the continuum exact solution, there are 3 different states but only two independent mass ratios, as the second excited state has the same mass as the ground state. Therefore, it is possible to use the same operator $O^{(2)}$ to extract the mass of the first excited state. The ratio between the mass of the ground state and the first excited states for the $SU(4)$ model is presented in Table 4.4.

---

[4]As a side note, if one tries to calculate in the Lattice the two-point function with any of these observables in the $SU(2)$ model it obtains 0.

β	L	$\frac{M_2}{M}$
0.225	24	1.509(7)
0.29	82	1.439(5)
0.3	90	1.435(15)
0.31	100	1.421(9)
Continuum		$\sqrt{2} \approx 1.41421356$

Table 4.4: Ratio between the ground state mass and the first excited state mass for the $SU(4)$ at several β values

Our results show that, in this case, the ratio only becomes close to the continuum prediction at relatively large correlation lengths ($\beta = 0.31$ in the $SU(4)$ model corresponds to $\xi \approx 9.3$). To our knowledge, this is the first time a lattice analysis has verified the $SU(4)$ spectrum.

This analysis can be expanded to study the different excited states of models with higher N. For instance, in [21], the two mass ratios of $SU(6)$ were obtained using $O^{(2)}$ and $O^{(3)}$, which couple with the first and second excited state respectively. Nevertheless, the observables coupling with higher excited states involve a higher number of matrix products, making them more expensive to calculate.

4.4   Asymptotic Scaling

We are now in a position to study more rigorously how the continuum limit is approached in this theory and to test how well it is approached in our simulation. While in more simple models, such as the Scalar Free Field Theory, the lattice spacing appears explicitly in the action, in the case of more complicated models, the Lattice spacing a behaviour is encoded in the coupling T (or β).

Consider a physical observable P, with a value in the continuum equal to P_0. The value of P obtained in the Lattice will depend on both the lattice spacing and the strength of the coupling, making it a function of both variables $P(T, a)$. However, as we take the continuum limit, we want this value to match P_0 and thus become independent of a. This condition means that the coupling constant T is not a constant at all but needs to be adjusted as $a \to 0$ to keep lattice value $P(T(a), a)$ constant and equal to P_0. [25]

It is possible to express this condition as:

$$\frac{dP(T(a), a)}{d\ln(a)} = \left(\frac{\partial}{\partial \ln a} + \frac{\partial T}{\partial \ln a} \frac{\partial}{\partial \ln a} \right) P(T(a), a) = 0 \tag{4.13}$$

The factor in front of the second term tells us exactly how the coupling should be varied as $a \to 0$. This term (multiplied by -1) is known as the lattice beta function.

$$-\frac{\partial T}{\partial \ln a} = \beta_L(T) \tag{4.14}$$

It should be noted that the factor of minus reflects the fact that the energy scale is proportional to $\frac{1}{a}$. The beta function is then expanded in a power series of T, giving:

$$\beta_L = -\beta_0 T^2 - \beta_1 T^3 + \mathcal{O}(T^4) \tag{4.15}$$

Where the parameters β_0, β_1 are determined by perturbation theory and require a two-loop calculation. In the case of the Principal Chiral Model, these parameters are equal to:

$$\beta_0 = \frac{N}{8\pi} \tag{4.16}$$

$$\beta_1 = \frac{N^2}{128\pi^2} \tag{4.17}$$

Integrating now equation 4.14 yields the two-loop relation between a and T:

$$a\Lambda_{L,2l} = \sqrt{\frac{8\pi}{NT}} \exp\left\{-\frac{8\pi}{NT}\right\} \tag{4.18}$$

were $\Lambda_{L,2l}$ is an integration constant (known as $\lambda-$parameter) which fixes $a = a_1$ for a given $T = T_1$. Calculations up to three loops would give small corrections to this relation.

If we invert this relation, we see that taking the continuum limit leads to the vanishing of the coupling. This is the lattice version of Asymptotic Freedom.

One way to test how well the continuum limit is approached is by using asymptotic scaling. This involves ensuring that the ratio between the physical masses and the Λ-parameter (in any regularization scheme) is constant in the continuum [21]. This is because physical masses are renormalized group invariants, which means they follow similar relationships as Eq. 4.18 [26]. To achieve this, we need the following equation to hold true as the continuum limit is approached:

$$M = C_L \Lambda_L \tag{4.19}$$

Here, C_L is a constant. The Bethe ansatz solution for the continuum model allows us to calculate the exact ratio between the physical masses and the Λ-parameter in the $\bar{M}S$ (Minimal Subtraction) scheme. We can also calculate the ratio between the Λ-parameters in the Lattice and $\bar{M}S$ regularization schemes. Combining these ratios (given in [21]) allows us to obtain an expression for C_L.

$$\frac{M}{\Lambda_L} = \sqrt{\frac{256N^2}{e\pi}} \sin\left(\frac{\pi}{N}\right) \exp\left\{\pi\frac{N^2-2}{2N^2}\right\} \tag{4.20}$$

Hence, we can test the prediction of Eq. 4.20, and therefore asymptotic freedom, by inspecting the ratio $\frac{M}{\Lambda_{L,2l}} \equiv= \frac{1}{\xi}\sqrt{8\pi\beta}\exp\{-8\pi\beta\}$ to confirm that it converges to the correct value as $\beta \to \infty$. The results of our lattice simulations for several values of N are shown in Fig. 4.5 (blue points).
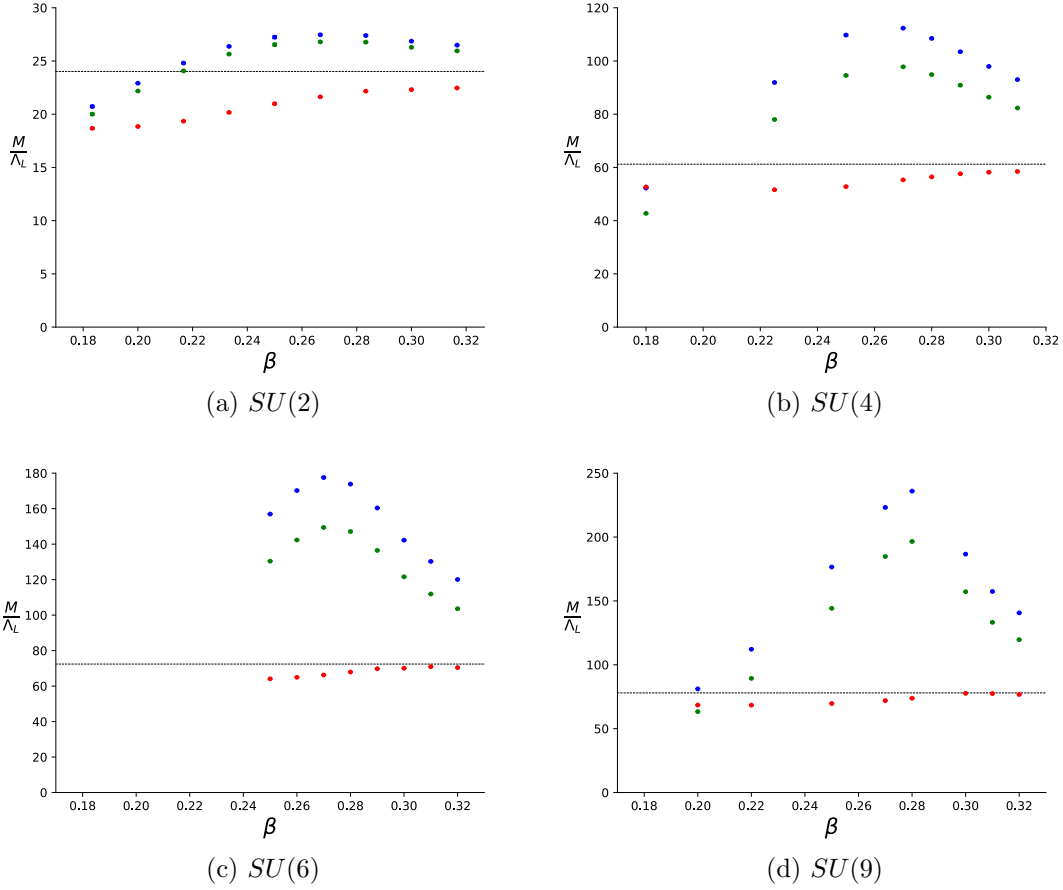
(a) $SU(2)$ (b) $SU(4)$

(c) $SU(6)$ (d) $SU(9)$

Fig. 4.5: The plots shows the Asymptotic Scaling plot for different N values. Each point entailed 10^5 measurement runs with 10^4 initial thermalization runs. The blue points indicate the Mass over Λ ratio in the regular scheme up to two loops, and the green points up to three loops. On the other hand, the red points indicate the ratio in the energy scheme. The continuum prediction is represented by the dashed line in each case.

Our results show that before starting to converge to the expected value, there is a significant overshoot in the mass over Λ_L ratio caused by a dip in the β function, which becomes more pronounced as N increases. Furthermore, this overshoot occurs in the same region as the heat capacity peak (transition between strong and weak coupling expansions), suggesting a connection between the two phenomena [21].

In [21], the energy scheme is introduced to try to fill the dip in the β-function and get a more accurate description of asymptotic scaling. In this scheme, the coupling T is redefined:

$$T_E \rightarrow \frac{8N}{N^2 - 1} e(T) \tag{4.21}$$

$$\beta_E \rightarrow \frac{1}{NT_E} \tag{4.22}$$

were e is the energy density of Section 4.1. The same expression is obtained for the Λ-parameter in this scheme up to two loops (the three-loops coefficient would be different for

the two schemes):

$$a\Lambda_{E,2l} = \sqrt{8\pi\beta_E}\exp\{-8\pi\beta_E\} \tag{4.23}$$

In order to test asymptotic scaling in this scheme, we look at the ratio defined by:

$$\left.\frac{M}{\Lambda_{L,2l}}\right|_E \equiv \frac{1}{\xi a\Lambda_{E,2l}}\frac{\Lambda_E}{\Lambda_L} \tag{4.24}$$

Which should converge to the same constant given in Eq. 4.20; the lattice results of this ratio are shown in Fig. 4.5 (red points). The success of the energy scheme over the regular one is evident in the cases of $N = 4, 6, 9$, the point at $\beta = 0.32$ deviates approximately 5% of the continuum prediction in the $N = 4$ in the energy scheme, and this deviation decreases to 2% and 1% for the cases of $N = 6, 9$ respectively for the same value of β. This suggests that not only is this scheme more successful than the regular one (in which the deviations at $\beta = 0.32$ are ...) but that it becomes more successful as N is increased. These improvements become even more noticeable when the mass over Λ ratio is analysed as a function of ξ as $\beta = 0.32$ corresponds to $\xi = ...$ for $N = 4, 6, 9$ respectively, which suggests that the onset of scaling appears at shorter correlation lengths for increasing N. In Fig. 4.5, we also show the three-loop calculation of the ratio in the regular scheme (green points); while it does improve the results, the corrections are small, which suggests that perturbation theory alone cannot explain the failure of the regular scheme and the success of the energy one.

On the other hand, in the $N = 2$ case, the energy scheme does not seem to yield any improvement over the regular one, as the deviation from the continuum limit is about 10% in both cases at $\beta = 0.32$[5]. This agrees with the findings presented in [4] in which the $SU(2)$ model was analysed. One possible explanation of the failure of the energy scheme on this occasion might come from Fig. 4.2 in which the heat capacity was presented. The energy scheme effectively removes the peak visible in this figure as the redefinition of the coupling in terms of the energy density renders the heat capacity constant. As mentioned before, there is a strong connection between the existence of the peak in the heat capacity and the dip in the $\beta-$function, so that removing it is supposed to fill the dip. However, as Fig. 4.2 reveals, in the $SU(2)$ case, the peak is much less pronounced to begin with, so it might explain the energy scheme not changing the convergence of the mass over Λ ratio. Furthermore, the heat capacity shown in Fig. 4.2 was calculated in 16^2 lattices, making them prone to finite volume size effects. Our results show that the peak is further reduced when this quantity is recalculated for $SU(2)$ with lattices satisfying $L \geq 8.5\xi$.

Additionally, there is one more difference in the $SU(2)$ case (visible in Fig. 4.1): the two energy density asymptotic expansions for this model never cross, giving a smoother transition between the two. These arguments motivate the failure of the $SU(2)$ energy scheme but do not completely explain it, and a more detailed study is required. A good start would be to look at the heat capacity at larger correlation lengths to confirm the existence of the maximum in the $SU(2)$ case once the finite volume effects have been accounted for.

---

[5]In both cases the mass over Λ ratio approaches the correct value.

4.5   Lattice Artifact

We now perform a more detailed analysis of the initial decaying region in the effective mass plot of the $SU(2)$ model. In order to do so, we begin by using the following ansatz to parametrize the decaying region.

$$m_{eff} = m_0 + g \exp\{-\gamma d\} \tag{4.25}$$

where m_0 is the fundamental mass of the system obtained by fitting the effective mass to a constant in the appropriate region, and g and γ are parameters to be determined. The goal of this parametrization is to estimate the critical distance, d^*, at which one can consider the effective mass constant. In order to do so we treat $g \exp\{-\gamma d\}$ as a small perturbation compared to the value of m_0, that is:

$$m_{eff} = m_0(1 + \delta) \tag{4.26}$$

$$\delta = \frac{g}{m_0} \exp\{-\gamma d\} \tag{4.27}$$

We can now decide on a sufficiently small value of δ and solve for d^*:

$$d^* = \frac{1}{\gamma} \log \frac{g}{m_0 \delta} \tag{4.28}$$

For this analysis, we chose $\delta = 0.01$ (other choices would also be valid), such that the perturbation is $1/100$ of the final value of the mass.

We used our lattice data to extract the two parameters g, γ for different values of β. The AIC procedure was used to reduce the bias on the selection of the fit window as it was observed that for larger correlations the first point does not agree well with the model. This effect could be due to the presence of higher-order corrections dominating this region. In turn, the values of these parameters were then used to calculate d^* as a function of beta; these results are shown in 4.6.
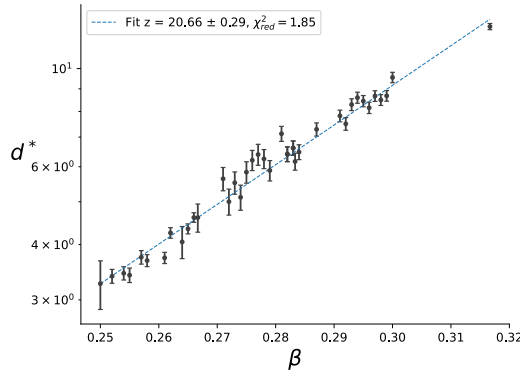

Fig. 4.6: Critical distance as a function of β for $SU(2)$. The data is presented as in a log scale to showcase its evident exponential nature. The dashed line represents the fit $g\beta^z$, with the value of z present in the plot.

Our analysis on d^* reveals that the decaying region becomes more stretched as β increases since the value of d^* provides an estimate of the extension of this region. At first, this might seem irreconcilable with this effect being a lattice artifact because such effects should disappear in the continuum limit ($\beta \to \infty$). However, the presented value of d^* is given in lattice units a, so what should vanish in the continuum limit ($\xi \to \infty$) is not d^* but rather the physical distance ad^*, that is we need:

$$\lim_{a \to 0} ad^*(a) \to 0 \tag{4.29}$$

In the previous section, we explained the precise relationship between the lattice spacing a and the coupling through the $\beta-$function. Therefore, it is possible to regard a as a function of β and replace the $a \to 0$ limit with $\beta \to \infty$:

$$\lim_{\beta \to \infty} a(\beta)d^*(\beta) \to 0 \tag{4.30}$$

In order to study if this is satisfied, we analysed the behaviour of d^* as a function of β. In particular, d^* was modelled as an exponential and the data fitted to the following ansatz:

$$d^*(\beta) = g \exp\{z\beta\} \tag{4.31}$$

This fit is visible in Fig. 4.6, and it gave a value of $z = 20.66 \pm 0.29$ with a healthy reduced chi-squared of $\chi^2_{red} = 1.85$ suggesting that our model is correct.

On the other hand, from the previous section, we know that, at least asymptotically, the lattice spacing follows:

$$a = \frac{1}{\Lambda_L}\sqrt{8\pi\beta}\exp\{-8\pi\beta\} \tag{4.32}$$

Where Λ_L is an integration constant. For the product $a \cdot d^*$ to banish in the $\beta \to \infty$ limit, we need the overall exponent of the product to be negative. That is, we require $z - 8\pi < 0$. This condition is satisfied for our value of z even within several sigma, confirming that the physical distance vanishes in the continuum limit.

In conclusion, our analysis suggests that the decaying region in the effective mass of $SU(2)$ plot is a lattice artifact.

4.6   Performance Analysis

One of the main goals of this project is to examine the effect of Fourier Acceleration and compare it with traditional HMC in the $SU(N) \times SU(N)$ Principal Chiral Model. In this section, we will explore the effect of Fourier Acceleration in our computation in terms of efficiency.

Given that by using FA, we aim to reduce autocorrelations, it is sensible to measure the algorithm's efficiency in terms of the integrated autocorrelation time τ_{IAT} of a specific observable. For the purpose of this report, we will use the susceptibility defined as:

$$\chi = \frac{1}{V}\sum_{x,y} TrU_xU_y^\dagger \tag{4.33}$$

The susceptibility is chosen for our analysis as it is dominated by slow momentum modes [27], making it particularly sensible to critical slowdown. Different choices of observables are possible, but those would lead to different (and probably less impressive) results.

In order to compare the performance of the different algorithms, we need to define a cost function for a given observable $\mathcal{O}$.

$$Cost_\mathcal{O} = \frac{\text{What you give}}{\text{What you obtain}} = \frac{\text{Time}}{N_{eff}} = \frac{Time}{N}\tau_{IAT} \tag{4.34}$$

where Time is the overall run time of the algorithm and N_{eff} is the effective number of configurations generated in that time. The last step uses the definition of the effective number of configurations $N_{eff} = \frac{N}{\tau_{IAT}}$ with N being the total number of configurations. Given that the effective number of configurations is defined as $\frac{N}{\tau_{IAT}}$ (with N being the total number of configurations). The overall scale of the cost is irrelevant, being defined only up to a proportionality constant.

It is important to note that our definition of cost differs from the one used in [4] for a similar study in the restricted case of $SU(2)$. This study's cost function is proportional to $\sqrt{\tau_{IAT}}$ and the acceptance probability. However, it is only the error on the observable that scales as $\sqrt{\tau_{IAT}}$ and not the cost of generating the configurations. Additionally, the acceptance probability is already encoded in the value of τ_{IAT}.

4.6.1 The Acceleration Mass Parameter

The inverse Kernel used in the Fourier Acceleration Algorithm for the $SU(N) \times SU(N)$ model is defined in Fourier space as:

$$\tilde{K}_{k,k'}^{-1} = \frac{\delta_{k,k'}}{\sum_\mu 4\sin\left(\pi\frac{k_u}{L}\right) + M^2} \tag{4.35}$$

This is the same Kernel that one obtains in the free field theory. However, in this case, there is no evident value for the acceleration mass M as this parameter was introduced to make the Kernel well-defined for all k. It is important to note that the algorithm's validity is independent of the choice of M, but different values of this parameter will attain different acceleration rates. The effect of the acceleration mass parameter on the FA's efficiency was examined using the cost function. For this analysis, the cost function can be simplified to:

$$Cost_\mathcal{O} = n_{MD}\tau_{IAT}(\mathcal{O}) \tag{4.36}$$

Where n_{MD} is the number of steps in the leapfrog integrator for a fixed total integration time of $t = n_{MD}\delta t = 1$. The choice of n_{MD} for each value of the acceleration mass is calibrated using the same procedure as the case for non-accelerated HMC.

All the cost values were normalized by the cost obtained by setting the acceleration mass to the lightest mass of the theory in lattice units. This choice was motivated by studies on different theories [5], which found that this value gave a close to optimal performance. The results obtained for the $SU(2)$ case are visible in Fig. 4.7.
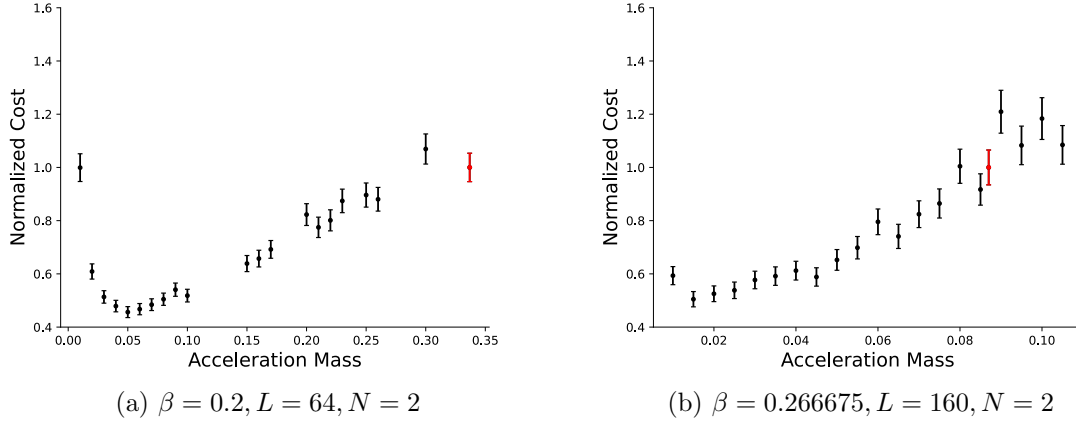
(a) $\beta = 0.2, L = 64, N = 2$

(b) $\beta = 0.266675, L = 160, N = 2$

Fig. 4.7: The plots show the normalized cost against the acceleration mass for the $SU(2)$ model and different values of β. The red value indicates the reference mass. The plots indicate that finding the optimal acceleration mass leads to a factor of two improvement for $SU(2)$

Our results show that the cost is minimized at an acceleration mass value smaller than the lightest mass of the theory (at a given β), and after this value, the cost increases. The results also reveal that this remains true even when β is increased. However, the potential benefit is approximately only a factor of two which suggests that hat expending a considerable amount of resources to determine the optimal value is not worthwhile. A similar analysis was carried out in [4] for this model, but our results differ due to the use of a different cost function. Nevertheless, the conclusion is the same: setting the acceleration mass to $M = \frac{1}{\xi}$ yields a reasonable acceleration, and the search for the optimal parameter is not justified.
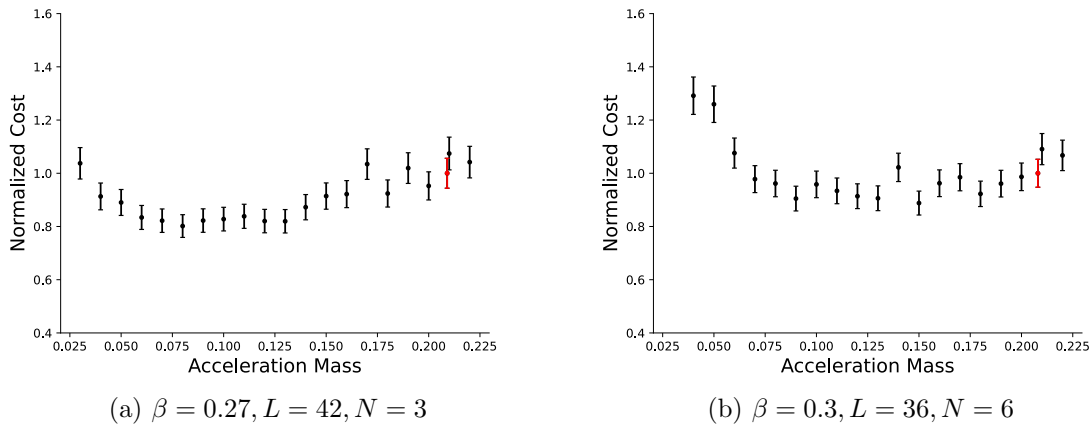

(a) $\beta = 0.27, L = 42, N = 3$

(b) $\beta = 0.3, L = 36, N = 6$

Fig. 4.8: The plots show the normalized cost against the acceleration mass for the $SU(N)$ model and different values of N, β. The flat cost function suggests that the acceleration mass does not have a great impact on the acceleration rate.

The same analysis was repeated for $N = 3, 6$ to see if there was any variation of the results

with N. Fig. 4.8 shows the normalized cost as a function of the accelerated mass for different $N = 3, 6$ at approximately the same correlation length. These results suggest that as N is increased, the cost function becomes flatter, implying that the benefit of determining the optimal effective mass decreases with increasing N. These results strengthen our previous statement that the fundamental mass provides a near-optimal acceleration and that the search for the optimal value is not justified. Guided by these results, the acceleration mass was always set to the inverse correlation length in lattice units for the performance analysis between FA HMC and non-accelerated HMC.

4.6.2 Critical Slowdown

We begin the comparison between HMC and FA HMC by analysing the effect of acceleration on the IAT and, therefore, on critical slowdown. In general, it is expected that (at least asymptotically) the IAT follows:

$$\tau_{IAT} \approx a^{-z} \approx \xi^z \tag{4.37}$$

where the value of the dynamic critical exponent depends on the algorithm. Thus, the comparison between the performance of HMC and FA HMC revolves around the precise value of z for each algorithm, as it will determine the exact behaviour of the IAT when approaching the continuum limit ($\xi \to \infty$).

In order to estimate z for both the non-accelerated and accelerated versions of HMC, we estimated the IAT of the susceptibility χ for different values of the correlation length, and we fitted these to the ansatz:

$$\tau_{IAT} = b\xi^z \tag{4.38}$$

where both b, z were fit parameters.

The process of extracting b, z presented some difficulties. Firstly, we found that the value of b heavily depends on the acceptance rate. Consequently, points measured with different acceptance rates will not follow the same curve, and incorporating them into the fit leads to a wrong estimation of z. Although our current calibration protocol tries to make all the runs fall within an acceptance rate of $[0.65, 0.85]$, it was found that this was not a sufficiently small interval to make all points fall within a single curve. In order to solve this, we manually selected the points generated through simulation runs with an acceptance rate of ≈ 0.7 and discarded the rest. Secondly, the values at small correlation lengths did not agree with the ansatz (particularly for FA HMC), so the fit window was varied such that the reduced chi-squared was close to one. This does not affect the quality of the results since we are interested in the continuum limit $\xi \to \infty$, making the values with lower ξ less important. The results obtained through this process can be seen in Fig. 4.9
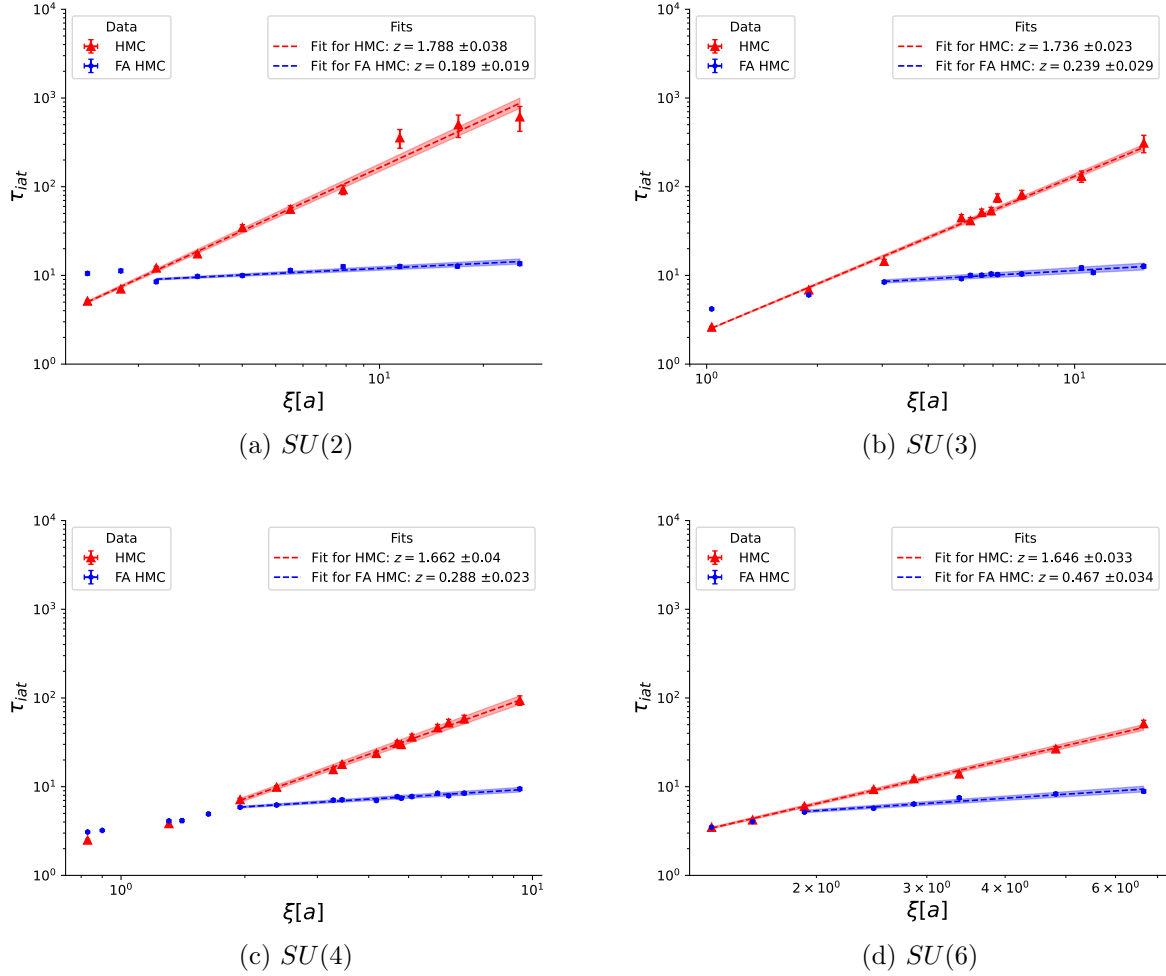
43

Fig. 4.9: The plots show the autocorrelation time τ_{IAT} as a function of the Correlation length for several $SU(N)$ choices in both accelerated and non-accelerated versions of the algorithm (log-log scale). 10^5 measurements were taken for each data point, with an initial 10^4 thermalization runs. The shaded region indicates the error on the fit parameters.

The results show that FA HMC is capable of mitigating the effect of critical slowdown for $N = 2, 3, 4, 6$. Even for the small correlation lengths considered in this report, the FA improves the IAT by at least one order of magnitude. This effect would only be increased if higher correlation lengths were considered. The $SU(2)$ value $z = 0.189 \pm 0.019$ is compatible with the value of $z = 0.21(1)$ found in [4]. On the other hand, the $SU(3)$ value $z = 0.24 \pm 0.03$ is noticeably smaller than the value $z = 0.45 \pm 0.02$ found in [28] using Multigrid Monte Carlo (MMC)[6] in this model. This speaks to the success of FA over MMC for this model.

The dependence of z on N is interesting in both HMC and FA HMC. In the HMC case, the slight variation on z seems to suggest that the algorithm's performance increases with N. However, the opposite effect is observed in the FA case. Additionally, the final $SU(6)$ value

---

[6]MMC is a different algorithm that also tries to address critical slowdown.

for the HMC is within 10% of its $SU(2)$ value, but in the case of the FA HMC, the z for $SU(6)$ is 2.5 times bigger than the z of the $SU(2)$ one, making this variation more noticeable than the HMC one.

FA's effectiveness in addressing critical slowdown appears to decrease with increasing N. However, it is also possible that this effect may be partially due to the fact that we are studying small correlation lengths. It would be interesting to look at higher correlation lengths to see if this effect is still present or if it is attenuated. Until then, the values of z obtained in this paper could be taken as an upper bound of the true values.

4.6.3   Generalized Cost

The advantage of the FA HMC over the non-accelerated version relative to the IAT is clear from the results of the previous section. However, since the FA involves a more complicated time evolution of the field and momenta, it is necessary to consider the run time when comparing the overall efficiency of the two algorithms. In particular, the FA HMC needs to perform several Fourier Transforms, which, in principle, could lead to the HMC version being more efficient.

For this purpose, we have measured the cost of both algorithms at several values of the correlation length ξ following the definition of Eq. 4.34, which takes into account both the run time and the IAT. The ratio between the cost of HMC over FA HMC is presented in Fig. 4.10. Given that the main variable that controls the run-time is the lattice size L^2 we only present the values which satisfy the condition $L \approx 8.5\xi$ of all of our data points. This allows for a fair comparison, and it makes the selected data points fall into a single curve.
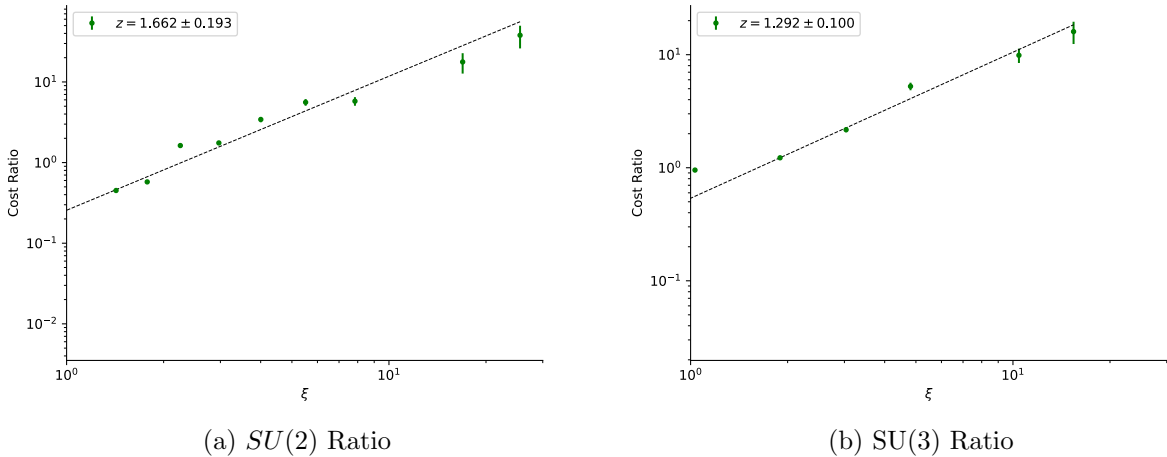

(a) $SU(2)$ Ratio   (b) SU(3) Ratio

Fig. 4.10: Ratio of the cost of HMC over FA HMC for several $N = 2, 3, 4$ as a function of the correlation length ξ. The data is fitted to the ansatz Cost $= g\xi^z$, and the values of z are presented in the plots. The plots indicate the efficiency of the FA HMC relative to the non-accelerated version.

Our results show that in all instances $(SU(2), SU(3), SU(4))$, the FA version of the algorithm

is superior to its non-accelerated version (indicated by the positive z values). Furthermore, even for the relatively small correlation lengths analysed in this report, the FA version leads to an order-of-magnitude improvement in terms of cost. This effect would only be magnified if longer correlation lengths are explored.

On the other hand, the extracted values of z vary from $z = ...$ for SU(2) up to $z = ...$ for SU(4), which suggests that the FA improvement gets weaker with higher N. This should come as no surprise, as the IAT results in the previous section already indicated this. However, it is possible that at least part of this effect comes from the small correlation lengths used in the analysis and that it gets better once larger simulations are performed.

Overall, the results indicate the success of the FA HMC algorithm over the standard HMC in terms of efficiency for the $SU(N) \times SU(N)$ Principal Chiral Model, and it leads to speed-ups of a factor of 10 for relatively small correlation lengths.

Chapter 5

Conclusion

> Per aspera ad astra. I'd heard a
> variety of translations, but the
> one I liked best was Through
> the thorns, to the stars.
>
> M.L. Rio, If We Were Villains

5.1 Limitations and Future Work

This space will be used to acknowledge some of the limitations of our study and to identify the areas in which it could be improved by future work.

The most evident limitation of our work is that with our current implementation, we are limited to the study of modest values of ξ in which we are only starting to see the onset of scaling. This limitation appears due to the fact that we are restricting ourselves to lattices which satisfy $L \gtrsim 8.5\xi$ as prescribed in [21] in order to reduce the finite volume effects to sub-per-cent levels. However, this greatly limits our ability to explore large values of ξ as large lattices are required (which, in turn, demand prohibitive amounts of computing power). One possibility around this problem would be to use the framework introduced in [29] to extrapolate quantities to the infinite volume limit by taking measurements at different lattice volumes. This technique was successfully used in [28] to study the $SU(3)$ Principal Chiral Model using Multi-Grid Monte Carlo. Using this framework would allow us to analyse the effect of FA at higher correlation lengths.

Another way of addressing this issue would be to increase our computing power using GPUs, but this would require a major modification of the code as it would need to be adapted for this purpose. However, the potential benefit of the additional computing capacity would allow us to directly study larger lattices and ξ, which would justify this effort.

Any of these two possibilities (or both in combination) would allow us to study higher values of the correlation length, improving all of the results obtained in this project. The analysis of asymptotic scaling and of the efficiency of FA HMC over traditional HMC would particularly benefit from it.

5.2   Conclusion

Several points have been addressed throughout this project. Initially, Lattice Field Theory and the critical slowdown problem when approaching the continuum limit were explained. Moreover, Fourier Acceleration was introduced as a potential solution to addressing critical slowdown.

Then we moved to the main topic of this project: The $SU(N) \times SU(N)$ Principal Chiral Model in two dimensions. In particular, we explained the motivation behind the study of this model as a proxy for QCD and how HMC and FA HMC can be implemented to simulate it. The remainder of the report was centred on presenting the results obtained through our simulations of the Chiral Model using FA HMC. In particular, the main properties of the Chiral Model, such as the mass spectrum and asymptotic freedom, were outlined and verified through lattice calculations.

Additionally, some of the questions raised in [4] about the $SU(2)$ model were addressed. We were able to motivate the failure of the energy scheme in the $SU(2)$ through its relatively small heat capacity peak. Furthermore, it was also possible to verify that the initial decaying region in the effective mass plots behaves like a lattice artifact.

The project's final part was centred on measuring the potential advantage of using FA HMC over traditional HMC. To achieve this, the performance between the two versions of the algorithm were compared through the IAT and directly through the cost function, which also accounts for run time. In both cases, it was found that FA HMC outperforms HMC, and even for small correlation lengths, one finds speed-ups of an order of magnitude. On the other hand, it was also observed that the performance of the FA HMC worsens when N is increased. Several strategies to improve these results were also presented.

In conclusion, this report has analyzed in great detail the main properties of the $SU(N) \times SU(N)$ Principal Chiral Model and has established FA HMC's success over traditional HMC in this model for $N = 2, 3, 4, 6$.

Bibliography

[1] A. Aleksejevs, S. Barkanova, S. Wu, and V. Zykunov. New Physics Search with Precision Experiments: Theory Input. *Nucl. Part. Phys. Proc.*, 273-275:2249–2252, 2016.

[2] Hartmut Wittig. Progress on $(g-2)_\mu$ from Lattice QCD. In *57th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, 6 2023.

[3] Jonathan Goodman and Alan D. Sokal. . *Phys. Rev. D*, 40:2035–2071, Sep 1989.

[4] Roger Horsley, Brian Pendleton, and Julian Wack. Hybrid Monte Carlo simulation with Fourier acceleration of the $N{=}2$ principal chiral model in two dimensions. *Physics Letters B*, 849:138429, February 2024.

[5] Simon Catterall and Sergey Karamov. Testing a Fourier-accelerated Hybrid Monte Carlo algorithm. *Physics Letters B*, 528(3–4):301–305, March 2002.

[6] Chulwoo Jung and Norman H. Christ. Riemannian Manifold HMC with fermions. In *40th International Symposium on Lattice Field Theory*, 1 2024.

[7] Pablo Morandé. Mphys Code. https://github.com/pmorande27/MPhysProject, 2024.

[8] Christof Gattringer and Christian B. Lang. *The path integral on the lattice*, pages 1–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[9] Andreas Wipf. Scalar Fields at Zero and Finite Temperature. In Andreas Wipf, editor, *Statistical Approach to Quantum Field Theory: An Introduction*, Lecture Notes in Physics, pages 85–110. Springer International Publishing, Cham, 2021.

[10] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo, July 2018. arXiv:1701.02434 [stat].

[11] Charles J. Geyer. Introduction to Markov Chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*, pages 3–48. Chapman and Hall/CRC, New York, 1 edition, May 2011.

[12] Galin L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299 – 320, December 2004.

[13] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, September 1987.

[14] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the Geometric Ergodicity of Hamiltonian Monte Carlo, November 2018. arXiv:1601.08057 [stat].

[15] T.P. Straatsma, H.J.C. Berendsen, and A.J. Stam. Estimation of statistical errors in molecular simulation calculations. *Molecular Physics*, 57(1):89–95, January 1986. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00268978600100071.

[16] Neal Madras and Alan D. Sokal. The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *Journal of Statistical Physics*, 50(1):109–186, January 1988.

[17] Guido Cossu, Peter Boyle, Norman Christ, Chulwoo Jung, Andreas Jüttner, and Francesco Sanfilippo. Testing algorithms for critical slowing downTesting algorithms for critical slowing down. *EPJ Web of Conferences*, 175:02008, 2018. arXiv:1710.07036 [hep-lat].

[18] Ulli Wolff. Monte Carlo errors with less errors. *Comput. Phys. Commun.*, 156:143–153, 2004. [Erratum: Comput.Phys.Commun. 176, 383 (2007)].

[19] Marc Wagner, Stefan Diehl, Till Kuske, and Johannes Weber. An introduction to lattice hadron spectroscopy for students without quantum field theoretical background. 10 2013.

[20] Simon Duane, Richard Kenway, Brian J. Pendleton, and Duncan Roweth. Acceleration of gauge field dynamics. *Physics Letters B*, 176(1):143–148, August 1986.

[21] Paolo Rossi and Ettore Vicari. Two-dimensional SU (N) × SU (N) chiral models on the lattice. *Physical Review D*, 49(3):1621–1628, February 1994.

[22] Reinhold A Bertlmann and Philipp Krammer. Bloch vectors for qudits. *Journal of Physics A: Mathematical and Theoretical*, 41(23):235303, May 2008.

[23] Arunabha Guha and S. C. Lee. Improved mean field studies of $SU(N)$ chiral models and comparison with numerical simulations. *Nuclear Physics B*, 240(2):141–170, September 1984.

[24] Alexander M. Segner, Andreas Risch, and Hartmut Wittig. Precision Determination of Baryon Masses including Isospin-breaking. 12 2023.

[25] Christof Gattringer and Christian B. Lang. *Pure gauge theory on the lattice*, volume 788 of *Lecture Notes in Physics*. Springer, Berlin, Heidelberg, 2010.

[26] J. Shigemitsu and J. B. Kogut. A study of Λ parameters and crossover phenomena in $SU(N) \times SU(N)$ σ models in two dimensions. *Nuclear Physics B*, 190(2):365–411, August 1981.

[27] E. Dagotto and J. B. Kogut. Testing accelerated stochastic algorithms in two dimensions: The SU(3) × SU(3) spin model. *Nuclear Physics B*, 290:451–468, January 1987.

[28] Gustavo Mana, Andrea Pelissetto, and Alan D. Sokal. Multigrid Monte Carlo simulation via XY Embedding II. Two-dimensional SU(3) principal chiral model. *Physical Review D*, 55(6):3674–3741, March 1997.

[29] Sergio Caracciolo, Robert G. Edwards, Sabino J. Ferreira, Andrea Pelissetto, and Alan D. Sokal. New method for the extrapolation of finite-size data to infinite volume. *Nuclear Physics B - Proceedings Supplements*, 42(1–3):749–751, April 1995.

Simulation tests

To verify our implementation of the FA HMC for the $SU(N) \times SU(N)$ principal Chiral Model, we performed some sanity checks that should be passed by an HMC algorithm. In particular, these tests should be passed by any implementation of the HMC using a second-order integrator for the field evolution. Given that the Fourier Acceleration of the HMC only modifies the dynamics, it should not affect the validity of the tests since these hold independently of the particular dynamics of the model within the HMC. While the results of these tests on our implementation will be explicitly shown for the case of SU(2) and the particular case of $\beta = 0.3, L = 16$, they were also performed for different N and values of the coupling constant and Lattice volume.

Firstly, looking at the exponential of the expectation value of the difference between the Hamiltonian values in each time step is helpful. The theoretical value of this quantity is one and is very sensible to fluctuations, so it can be used as a test to verify the adequacy of our implementation. Furthermore, it is expected that the value is one regardless of the stepsize in the leapfrog integration (once a reasonable acceptance probability has been achieved). On the other hand, if we look at $\langle \Delta H \rangle$ itself, for a second-order integrator such as the leapfrog or Verlet, it is expected that $\langle \Delta H \rangle \propto \Delta \tau^4$. Therefore, checking that this relationship is satisfied constitutes another valid check for our implementation. We have condensed these three tests in the left panel of Fig 1. Furthermore, it is possible to look at the acceptance rate to verify our implementation as theory predicts that for an HMC algorithm, the acceptance rate as a function of $\langle \Delta H \rangle$ should follow a curve given by: $P_{acc} = erfc\{\frac{\sqrt{\langle \Delta H \rangle}}{2}\}$. Validating this relationship can be used as a test of our implementation. Fig 1 shows the relevant plots for these tests.
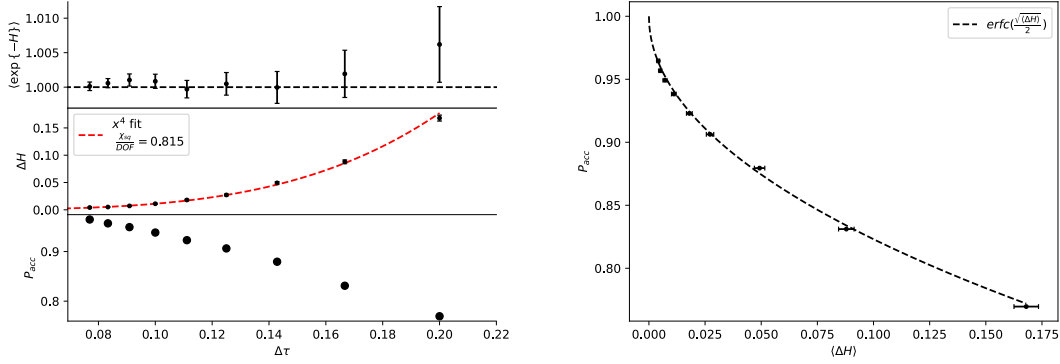
Fig. 1: The Fig shows different tests for the validity of the FA HMC. The first plot shows the value of $\langle\exp\{-\Delta H\}\rangle$ for different step sizes of the leapfrog integrator. The value needs to be close to one within errors. The second plot shows the $\langle\Delta H\rangle$ value for different step sizes, following a $\Delta\tau^4$ curve. The final plot gives the acceptance probability against the step sizes. The plot on the right panel shows the acceptance probability of the FA HMC algorithm against $\langle\Delta H\rangle$, and the dashed line indicates the theoretical prediction for the HMC algorithm.

Firstly, from the first plot on the left panel, it can be seen that $\langle\exp\{-\Delta H\}\rangle$ does average around the expected value of 1 within statistical errors. Secondly, the second plot on the left panel shows that a one-parameter fit of the form $A\Delta\tau$ gives a good description of the data with a reduced χ^2 value close to 1. Finally, the third plot on the left panel gives the acceptance probability as a function of the step size, and it is given as a reference to verify that it is high enough for the results of the two previous plots to be meaningful. On the other hand, the right panel plot shows the acceptance probability as a function of $\langle\Delta H\rangle$ and the expected result is presented as a dashed line. It is clear from the plot that the data is in good agreement with the theoretical result. Therefore, it is reasonable to conclude that our implementation of the FA HMC passes all the proposed tests.

$SU(N)$ **Generators**

We will now describe the procedure outlined in [22] to obtain a set of matrices which can be used as generators for the $SU(N)$ group in its fundamental representation for general N. This is set of matrices satisfying the following conditions:

   I Traceless, i.e., $Tr\{T_i^N\} = 0$

  II Hermiticity, i.e, $\left(T_i^N\right)^\dagger = T_i^N$

 III The set $\{T_i^N\}$ forms a basis.

 IV Orthogonality. i.e., $Tr\{T_i^N T_j^N\} = K\delta_{ij}$ (N not summed over)

It is useful to define $E_{j,k}^N$ as the $N \times N$ matrix with only the element j, k equal to one and all the others equal to zeros. We then define three types of elements.

The first type is formed by symmetric matrices given by

$$S_{j,k}^N = E_{k,j}^N + E_{j,k}^N \tag{1}$$

with the indices j, k defined as $1 \leq j < k \leq N$, which for fixed N gives a total of $\frac{N(N-1)}{2}$ different S type matrices.

The second type is formed by antisymmetric matrices defined through:

$$A_{j,k}^N = -i(E_{j,k}^N - E_{k,j}^N) \tag{2}$$

with again the indices defined as $1 \leq j < k \leq N$, which for fixed N gives again a total of $\frac{N(N-1)}{2}$ different A type matrices.

Finally, the last type is formed by a set of diagonal matrices obtained by:

$$D_l = \sqrt{\frac{2}{l(l+1)}} \left(\sum_{j=1}^{l} E_{j,j} - lE_{l+1,l+1}\right) \tag{3}$$

with the index l going from 1 to $N - 1$, giving $N - 1$ elements.

It can be proven (proof given in [22]) that the set formed by the collection of all A, S and D matrices (which account for a total of $N^2 - 1$ elements) satisfy all properties $I - IV$ and therefore serves as a set of Generalized Gell-Mann matrices. It is easy to show that for $N = 2, 3$, this procedure produces the Pauli and Gell-Mann matrices, respectively. Furthermore, with this procedure the orthogonality condition is: $Tr\{T_i^N T_j^N\} = 2\delta_{i,j}$ so $K = 2$ irrespective of N.