

THE INSTANTON LIQUID AND THE AXION

by

Olivier Wantz

Homerton College

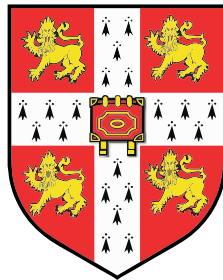
This dissertation is submitted for the degree of

Doctor of Philosophy

University of Cambridge

Department of Applied Mathematics

and Theoretical Physics



Supervisor: Prof. E.P.S. Shellard

25th September 2009

To my parents

**The Instanton Liquid
and
The Axion**

Olivier Wantz

Abstract

The ultimate goal of this thesis is to improve our understanding of the cosmology of axions. Axions couple to QCD instantons and these non-perturbative effects are modeled within the framework of the interacting instanton liquid model (IILM). The thesis describes the significant advances made within the IILM in order to study the quark-gluon plasma in realistic parameter regimes. In particular, a determination of the temperature-dependent axion mass in the IILM lays the foundation for a critical reevaluation and update of present cosmological axion constraints.

We develop grand canonical Monte Carlo routines to study topological fluctuations in the quark-gluon plasma. The model is calibrated against the topological susceptibility at zero temperature, in the chiral regime of physical quark masses. A numerical framework to derive interactions among the pseudo-particles is developed that is in principle exact, and is used to cure a pathology in the presently available finite temperature interactions.

The IILM reduces field theory to a molecular dynamics description, and we show that, quite generically, the dynamics for non-trivial backgrounds in the presence of light quarks is reminiscent of a strongly associating fluid. To deal with the well-known difficulty in simulating ionic fluids, we develop advanced algorithms based on Biased Monte Carlo techniques.

We study the IILM at finite temperature in the quenched and unquenched sector, with due diligence to a consistent thermodynamic limit. Of particular interest is chiral symmetry breaking and the temperature dependence of the topological susceptibility, and we study in detail the effects of instanton–anti-instanton pairs. Our determination of the topological susceptibility provides, for the first time, a well-motivated axion mass for all temperatures.

The misalignment mechanism for axion production is studied in detail, solving the evolution equations exactly in a radiation dominated FRW universe with the full temperature dependence of the effective degrees of freedom taken into account. Improved constraints in the classic and anthropic axion window are derived. We generalise the latter to large angle fine-tuning by including in the isocurvature contribution to the cosmic microwave background

radiation the full anharmonic axion potential effects. Finally, we reexamine bounds from axion string radiation in the thermal scenario to complete a comprehensive update of all cosmological axion constraints.

Acknowledgements

This thesis would not have been possible without the help and input from several people.

On the academic front, I want to thank my supervisor Paul Shellard for supporting me all along this rather long journey. The project has certainly taken on a different slant than initially intended. Straying rather wide and leaving his area of expertise, I would still get valuable advice. I hope that his patience was rewarded to some extent when we returned to and collaborated on axions eventually.

A large portion of this thesis does not belong to mainstream physics and it was not always easy to get feedback on these topics. I am very much indebted to Ron Horgan for taking the time to get acquainted with the instanton liquid and give advice on general issues concerning QCD and the renormalisation program.

In my endeavour to relate the results from the instanton liquid to the lattice, I was fortunate that Matt Wingate and Ron Horgan would always be available to clarify some subtle points. In particular, through Matt's feedback on some of my data I was able to find a rather nasty bug in my code.

Nick Dorey was kind enough to walk me through the renormalisation group improvement that is used in the instanton measure, and clarify some other issues related to renormalisation.

Finally, I am largely indebted to Pietro Faccioli who shared with me many of his insights into the instanton liquid during numerous phone calls, and kept me motivated through his continuous interest in my work.

Many hours have been spent coding up the tools to run simulations of the instanton liquid. As such I would rely heavily on computers, especially on the

resources provided by the national cosmology super-computer, COSMOS. I'd like to thank Paul for giving me the opportunity to work with these facilities, and I'm grateful for the brilliant support provided by Stuart Rankin, Victor Travieso and Andrey Kaliazin.

On the financial side, I'd like to acknowledge the support from STFC grant PPA/S/S2004/03793 and an Isaac Newton Trust European Research Studentship. Simulations were performed on the COSMOS supercomputer (an Altix 4700) which is funded by STFC, HEFCE and SGI.

Last but not least, I would like to thank the many friends I have made in Cambridge for their continuous support. My social life very much revolved around Homerton college, far away from DAMTP, where I could engage in all kinds of activities and forget about the harsh realities of finishing a PhD. I'd particularly like to mention Hamid Hazby who was a constant companion during the biggest challenge I have faced so far. Thank you very much for your friendship.

This section wouldn't be complete unless I mentioned my family, and in particular my parents. Thank you so much for keeping faith in me.

Preface

The work presented in this thesis has been (or is to be) published as the following series of papers:

- O. Wantz, “*The topological susceptibility from grand canonical simulations in the interacting instanton liquid model: zero temperature calibrations and numerical framework*”,
Nuclear Physics B 829 (2010) 48-90,
[arXiv:0907.3506]: Chapter 3.
- O. Wantz, “*The topological susceptibility from grand canonical simulations in the interacting instanton liquid model: strongly associating fluids and biased Monte Carlo*”
Nuclear Physics B 829 (2010) 91-109,
[arXiv:0907.4133]: Chapter 4.
- O. Wantz and E.P.S. Shellard, “*The topological susceptibility from grand canonical simulations in the interacting instanton liquid model: chiral phase transition and axion mass*”,
Nuclear Physics B 829 (2010) 110-160,
[arXiv:0908.0324]: Chapter 5.
- O. Wantz and E.P.S. Shellard, “*Axion cosmology revisited*”,
[arXiv:0910.1066]: Chapter 7.

The work presented in this thesis is my own but I have, as a matter of style, chosen to retain the use of the first person plural throughout.

This dissertation is not substantially the same as any that I have submitted, or am submitting, for a degree, diploma or other qualification at any other university. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

Signed:

Dated:

Contents

Acknowledgements	i
Preface	iii
1 Introduction	1
2 QCD, the Strong CP problem and Axions	7
2.1 Classical Yang-Mills theory	9
2.1.1 Gauge fixing	11
2.1.2 Vacuum structure	12
2.2 Quantum theory and QCD	14
2.2.1 Euclidean formulation, θ -vacuum and instantons	15
2.2.2 Perturbation theory	20
2.3 Strong CP problem	28
2.3.1 Peccei Quinn Mechanism and Axions	31
3 The IILM at zero temperature	33
3.1 Saturating the path integral	37
3.2 Interactions in the IILM	41
3.2.1 Gluonic interactions	43
3.2.2 Quark Interactions	48
3.3 Numerical Implementation	53
3.3.1 Interpolation and asymptotic matching	53
3.3.2 Monte Carlo	60
3.3.3 Fermionic determinant	63
3.4 Different Ensembles	64
3.5 Fixing parameters	74

3.5.1	Quenched case	74
3.5.2	Unquenched case	76
3.6	Conclusion	81
4	Biased Monte Carlo	88
4.1	Markov Chain Monte Carlo	90
4.2	Biased Monte Carlo	93
4.3	Toy Model	98
4.4	Conclusions	111
5	IILM at finite temperature	115
5.1	Interactions in the IILM at finite temperature	118
5.2	Numerical Implementation	124
5.2.1	Interpolation and asymptotic matching	124
5.2.2	Biased Monte Carlo	132
5.2.3	Fermionic determinant	145
5.3	Quenched simulation	148
5.4	Chiral symmetry restoration	149
5.5	Topological susceptibility and axion mass	157
5.5.1	Topological susceptibility	157
5.5.2	Temperature-dependent axion mass	161
5.6	Conclusion	167
6	Concordance Cosmology	171
6.1	General Relativity	172
6.2	Concordance cosmology	174
6.2.1	Symmetry and FRW universe	175
6.2.2	Expanding universe	178
6.2.3	Inflation	179
6.2.4	The dark sector	183
6.3	Thermodynamics in FRW	186
7	Axions	194
7.1	Axion potential	196
7.2	Axion cosmology	202
7.2.1	Misalignment mechanism	204

7.2.2	Anthropic window, isocurvature and fine-tuning bounds	211
7.2.3	Axion strings	219
7.3	Conclusion	226
8	Conclusions and Outlook	229
A	Interactions at zero temperature	237
A.1	Gluonic Interactions	237
A.1.1	Exact Interactions	239
A.1.2	Asymptotic Interactions	243
A.2	Fermionic Interactions	247
A.2.1	Exact Interactions	247
A.2.2	Asymptotic Interactions	247
B	Interactions at finite temperature	249
B.1	Gluonic Interactions	249
B.1.1	Exact Interactions	251
B.1.2	Asymptotic Interactions	263
B.2	Fermionic Interactions	267
B.2.1	Exact Interactions	268
B.2.2	Asymptotic Interactions	269
C	Cholesky decomposition update	271
C.1	Canonical Moves	271
C.2	Insertion	272
C.3	Deletion	273

List of Figures

2.1	Gauge fixing	12
2.2	Vacuum structure	15
2.3	Self-energy	28
3.1	Bosonic interactions: equal sizes	44
3.2	Bosonic interactions: unequal sizes	46
3.3	Fermionic overlap	50
3.4	Fermionic interaction	51
3.5	Large separation approximation	55
3.6	Small separation approximation	56
3.7	Interaction approximation (1)	58
3.8	Interaction approximation (2)	59
3.9	Color orientation dependence	66
3.10	Free energy	68
3.11	Mass dependence: Instanton number	69
3.12	Mass dependence: Topological susceptibility	70
3.13	Mass dependence: Instanton size	71
3.14	Mass dependence: Interaction	72
3.15	Topological charge fluctuations in the quenched sector	75
3.16	Chiral condensate	79
3.17	Smallest Dirac eigenvalue	82
3.18	Topological charge fluctuations in the unquenched sector	83
3.19	Instanton mean size vs volume	84
4.1	Interaction	99
4.2	Bonding box (1)	101
4.3	Bonding box (1)	102

4.4	Thermalisation history (1)	104
4.5	Thermalisation history (2)	105
4.6	Thermalisation history (3)	107
4.7	Thermalisation history (4)	108
4.8	Configuration snapshots	110
4.9	Interaction strength dependence	112
5.1	Incorrect thermodynamic limit	116
5.2	Bosonic interactions (1)	120
5.3	Bosonic interactions (2)	121
5.4	Fermionic overlap	123
5.5	Fermionic interaction	125
5.6	Interpolation grid	127
5.7	't Hooft potential approximation	128
5.8	Large separation approximation	129
5.9	Small separation approximation	130
5.10	Interaction approximation	131
5.11	Interaction vs. R_t and R_s	134
5.12	Plasma screening factors	139
5.13	Thermalisation history (low temperature)	141
5.14	Thermalisation history (critical temperature)	142
5.15	Thermalisation history (high temperature)	143
5.16	Configuration snapshots	146
5.17	Topological susceptibility in the quenched sector	149
5.18	Chiral susceptibility	153
5.19	Instanton–anti-instanton concentration	154
5.20	Instanton density	155
5.21	Quark condensate	156
5.22	Non-interacting sub-ensemble	158
5.23	Topological susceptibility in the unquenched sector	159
5.24	Unbonded instantons	160
5.25	Topological susceptibility and instanton density: screening parameter dependence	162
5.26	Axion mass from the topological susceptibility	165

6.1	WMAP temperature fluctuations	175
6.2	Sloan Digital Sky Survey	176
6.3	Spaces of constant curvature	177
6.4	Hubble diagram	180
6.5	Horizon problem	182
6.6	Concordance cosmology	185
6.7	Scale factor vs. temperature	188
6.8	Effective relativistic degrees of freedom	191
7.1	Axion mass (1)	201
7.2	Axion mass (2)	203
7.3	Axion angle oscillations	206
7.4	Relation between f_a and θ_a	210
7.5	Difference between analytic and numerical results	212
7.6	Axion parameter space (2d)	217
7.7	Axion parameter space (3d)	218
7.8	Axion string radiation uncertainties	223

List of Tables

3.1	Ansätze	43
3.2	Mass dependence	63
3.3	Set of quark masses	68
3.4	Mass dependence: summary	73
4.1	Autocorrelation times	106
4.2	Dependence on bonding box	111
5.1	Bonding box parameters and dependence	147

Chapter 1

Introduction

The strong interactions are very hard to tackle analytically in general. A notable exception is given by hard processes, which can be accurately described by perturbation theory. This is due to the ultra-violet Gaussian fix point, i.e. asymptotic freedom. Therefore, perturbation theory provides a self-consistent framework from the practical point of view. By this we mean that low-order computations are under control, and we neglect questions relating to the formal convergence of the perturbation theory; instead we just take for granted that the perturbation series might be asymptotic.

Many interesting features that QCD is believed to display can, however, not be studied with these tools. In particular, the strong interactions are believed to display a number of interesting, non-perturbative phenomena, among which are the confinement/deconfinement transition, spontaneous P and CP violation and chiral symmetry restoration. The hadronic regime can be addressed by chiral perturbation theory in a systematic fashion. It relies very strongly on the large symmetry group of chiral transformations, but cannot be derived from QCD from first principles, i.e. by integrating out the hard modes in the path integral. The ultimate systematic tool that is at our disposal at the moment is lattice gauge theory, see [134] for a thorough introduction. Over the last decade this field has made much progress, driven by increasing computing resources and improved algorithms, see [51] for an account of these developments. Unquenched simulations of QCD with physical quark masses can finally be studied on large lattices, and we can investigate the properties pertaining to the non-perturbative regime.

To get a more qualitative understanding of the QCD vacuum, computer simulations are perhaps too blunt a tool, although eventually any qualitative model will have to reproduce lattice results. Given a phenomenological model, inspired by experiments or lattice data itself, specific operators can be constructed to scan the (practically) infinite dimensional configuration space of the lattice and corroborate or refute said model. The insights gained will eventually influence the lattice community, and the loop of scientific progress closes.

In this thesis we will use instanton methods as a phenomenological model of the QCD vacuum [62, 167, 168, 169]. The idea is indeed old [31, 32], and has been validated by lattice studies for some time now [41, 199]. The reason to develop instanton based models at all rests on the expectation that they should describe well the chiral regime of QCD, which in turn follows from the connection between the Dirac zero modes and the quark condensate, i.e. the Casher-Banks relation, on the one hand, and the connection between the Dirac zero modes and the topology of the vacuum, i.e. the index theorem, on the other hand.

Instanton models are based on a combination of semi-classical expansions [182] and variational approaches [61, 62]. Taking this variational path integral as a starting point, Shuryak investigated what has become known as the interacting instanton liquid model (IILM) [170, 171]. In [157] many bulk properties were computed and seen to be consistent with the available lattice data and phenomenology. Some recent studies [76, 46] corroborate the earlier results that the IILM rather accurately describes the chiral properties of QCD, i.e. that instantons are the dominant degrees of freedom as far as the chiral regime of QCD is concerned.

At finite temperature the picture is less clear, and comparatively few studies of the instanton ensemble have been performed. It is, however, generally believed that the instanton liquid based on Harrington–Shepard (HS) calorons [93] does not lead to confinement. On the heuristic level this can be understood because these calorons have trivial holonomy, the order parameter for the deconfinement phase transition.

Over the last decade more general Euclidean solutions have been derived [119, 121, 120]. These possess more parameters and allow for non-trivial

holonomy. Thus, interest has been stirred again as to whether instanton models based on these more general calorons can play an instrumental role in the deconfinement transition.

In essence the IILM describes the non-perturbative regime of QCD by approximating the soft fluctuations by the collective degrees of freedom of the instanton background field. In the process, the infinite dimensional path integral is reduced to a finite dimensional partition function describing ‘particles’ in a box. The interactions that generate dynamics in the model follow from the fact that the background of instantons is not a classical solution to the equations of motion.

It is straightforward, in principle, to derive the interactions between instantons; we just have to evaluate the action over the background ansatz. On a practical level, the task is far from trivial as no analytic closed-form expressions can be given. To parametrise the interactions is tricky because of the rather high dimensionality of the moduli space that describes the collective coordinates of individual instantons. Such parametrisations are available for HS calorons [172], but they lead to unphysical behaviour. Studying the large volume dependence of the IILM in order to fit light, physical quarks into the simulation box, it will become clear that these interactions do not allow for a thermodynamic limit.

In light of the even more complicated moduli space of the non-trivial holonomy calorons, we will set up a more systematic framework to derive the interactions, that will allow us to incorporate the latter in the future. The framework we will develop in chapter 3 will rely on a combination of interpolation of the exact interactions and matched asymptotic expansions. We will stick to the HS calorons, and validate our model against the available data at finite temperature from earlier studies [157]. After studying the various systematic effects of the model, we will push the IILM into the physical quark mass regime. We will use recent lattice data [37, 38] to set the one dimensional parameter of the model. This will allow us to set the quark masses to their physical values.

After calibrating the model with the zero temperature simulations, the results in the finite temperature regime will be real predictions. It turns out, however, that simulating the IILM with light quark masses is compu-

tationally non-trivial. As we will discuss at length in chapter 4, the quark interaction will favour the formation of instanton–anti-instanton molecules. We will show that the IILM behaves as a strongly associating fluid. The latter have been extensively studied in chemical engineering and computational chemistry, and the key technique to simulate these systems is Biased Monte Carlo. We will review the basics underlying Monte Carlo methods, paying due attention to the concept of proposal probabilities. The latter are instrumental in devising powerful algorithms to deal with the highly localised regions of strong interactions. We will argue that, quite generically, the combination of localised backgrounds and light quarks will lead to systems that behave like associating fluids.

It has long been advocated that instanton–anti-instanton pairs play an important part in chiral symmetry restoration [99, 100, 158]. We will study this aspect in much detail in chapter 5 and give conclusive evidence that these molecules drive the chiral phase transition, within the IILM based on HS calorons.

Our main concern, however, will be to derive the temperature dependence of the topological susceptibility. The topological susceptibility is a key parameter of QCD and has been investigated in many lattice studies. It is a chiral property of QCD and can thus be expected to be well described by the IILM. Comparatively few studies have addressed this quantity within the IILM [173]. One reason is that, so far, the IILM is based on a canonical ensemble, whereas the most natural setting for the investigation of topological fluctuations is the grand canonical ensemble. We will use grand canonical Monte Carlo methods, which simulate directly in the grand canonical ensemble.

The ultimate purpose, and the initial motivation, for studying the topological susceptibility relies on the fact that it gives the temperature dependent mass of the axion. In the hadronic regime the axion mass can be computed within chiral perturbation theory. At high temperatures, above the phase transition, it can be reliably estimated by instanton methods in the so-called dilute gas approximation. The transition between the two regimes has so far not been addressed rigorously. Our determination of the topological susceptibility will for the first time give an axion mass for all temperatures and,

in particular, provide a well-motivated interpolation between the zero and finite temperature regimes.

Axions are the standard candidate to solve the strong CP problem¹, and provide one of the simpler extensions of the Standard Model of particle physics. Its coupling to QCD is a purely quantum mechanical effect, arising from the anomalous axial current. The fundamental origin of the axion field is not settled, however, if it exists at all. Originally it was introduced as the Goldstone boson of the spontaneously broken, anomalous, Peccei-Quinn symmetry [146, 145], tied to the electro-weak sector. Failure to detect such a particle has lead to the invisible axion, whose energy scale is a priori arbitrary, but large enough to evade laboratory and astrophysical bounds.

The axion scale is, however, also constrained from above through cosmological considerations. In fact, the very weak interactions make it one of the best motivated candidates for cold dark matter. The axion can be created through many cosmological processes. We will focus mainly on the misalignment mechanism, which is the appropriate production channel for the anthropic axion [126, 188, 184]. This axion has a very large decay constant, comparable to the GUT or string scale, and is thus much better motivated from a theoretical point of view than just any arbitrary invisible axion model [181].

We will derive the contributions from the classical evolution of the zero mode oscillations by solving the field equations exactly in chapter 7. In particular, we will include the correct number of relativistic degrees of freedom, and provide accurate fitting formulas for the latter.

In the anthropic region, the PQ symmetry is broken before inflation and the latter induces isocurvature fluctuations in the axion field. These are very tightly constrained by the temperature perturbations in the cosmic microwave background radiation. We will generalise this bound by including anharmonic effects from the axion potential.

If the PQ symmetry breaks after inflation, the axion evolution becomes much more complicated due to the emergence of a string network, as a con-

¹Other often quoted mechanisms include a massless quark, see the discussion in section 2.3, or spontaneous CP breaking, see for instance [56].

sequence of the Kibble mechanism. In fact, axion radiation from this string network supersedes the density of axions produced through the misalignment mechanism, and is thus the most relevant contribution in the “classic axion window”. We will update these axion string bounds by using our new axion mass. Since the time these constraints were first derived, the cosmological parameters have been determined to much higher precision, and we also take the opportunity to update them in that respect.

Chapter 2

QCD, the Strong CP problem and Axions

Since the 1970's physicists believe that hadron physics is described by a non-Abelian gauge field theory, quantum chromodynamics or QCD for short. The excitations of this field are quarks and gluons, not the experimentally observed hadrons, build from these fundamental degrees of freedom. That hadrons have a substructure was beautifully realised already in Gell-Mann's eightfold way [79], the constituent quark model. The full story in QCD is, however, more complicated: on top of the three 'valence' quarks, QCD predicts the constant creation and annihilation of quark-anti-quark pairs, the so-called sea quarks. Including the coupling to the electro-weak sector, these sea-quarks can even change their flavour before returning to the vacuum. This rich vacuum structure, which is a direct consequence of the field concept underlying gauge theories, goes far beyond the static constituent quark model, and gives rise, for instance, to the rather counterintuitive phenomenon of asymptotic freedom: counterintuitive because QCD was always supposed difficult to tackle analytically, being strongly coupled, yet the theory predicts a regime of weak coupling at high energies. Due to asymptotic freedom, the high energy regime of QCD can be studied perturbatively¹.

The fundamental degrees of freedom of QCD, quarks and gluons, have never been observed directly. It is postulated that at low energies quarks and

¹At higher orders, perturbation theory breaks down, however, due to magnetic instabilities at $T \neq 0$, see for instance [103].

gluons are bound into colour-neutral hadrons, i.e. are confined. Even though a compelling theoretical proof is still lacking, QCD is believed to explain confinement. There is certainly a large body of evidence from lattice QCD to support the claim. Even more importantly, experiments give indirect evidence in favour of QCD, e.g. the measurement of the cross-section of $e^+ - e^-$ annihilation to hadrons as compared to muons or the decay of the neutral pion into two photons, which are sensitive to the number of colours; deep inelastic scattering experiments gave evidence for a weakly interacting pointlike substructure in hadrons, the so-called partons, later to be identified with quarks. For more details, see [149].

Taking asymptotic freedom at face value, baryonic matter under extreme conditions², such as during the early universe or in neutron stars, has little resemblance with its hadronic realisation at low energies where only colour-neutral states can propagate freely. The connection between these regimes cannot be studied perturbatively. There is, however, compelling evidence from the lattice community that, as temperature and/or pressure is raised, QCD will undergo a phase transition to a new form of matter, the so-called quark-gluon plasma in which quarks and gluons are deconfined, i.e. can propagate freely. Whether this transition is of first order has direct consequences for cosmology because of the large amount of entropy that is produced [98]. The large experimental efforts undertaken at RHIC and LHC specifically aim at unravelling the mysteries of matter under such extreme conditions. For more details see [207].

As mentioned above, lattice gauge theory provides probably the best method to study non-perturbative effects in gauge theories. Interestingly, lattice studies of the pure gauge sector show that gluodynamics also exhibits a confinement/deconfinement phase transition, and it is conceivable that a better understanding of Yang-Mills theory alone will give us key insights into the confinement mechanism.

QCD is very hard to tackle analytically in the hadronic phase because the coupling is strong and a simple perturbative treatment is not possible. That is not to say that no analytic progress can be made at all. Given some physical intuition, different approaches have been devised; some more

²High temperature and/or pressure.

rigorous than others in that they can be ‘derived’ or at least well motivated from the fundamental QCD Lagrangian: chiral perturbation theory, heavy quark theory, sum rules, $1/N_c$ expansion, see for instance [177, 139] for more details. A different approach, that we will follow in this thesis, is based on the idea that already classical Yang-Mills theory holds the key to the non-perturbative QCD vacuum. In particular, that the proper starting point for a perturbative treatment is not based on trivial vacuum configurations; instead we should embrace the very non-linear nature of the field equations and the topologically non-trivial field space. This method is based on a combination of the semi-classical approximation and the variational approach, and will be discussed in the next chapter.

2.1 Classical Yang-Mills theory

Originally, Yang and Mills devised their theory to describe the nucleon. At the time the almost mass degenerate proton and neutron, with the mass splitting due to electromagnetic effects, were thought to be just two different realisations of the same particle, the Heisenberg nucleon, in much the same way as a spin up and spin down electron are only two different states of the electron. This similarity prompted the name isospin³, also with group $SU(2)$. Once we agree on labelling the proton as spin up, say, no more freedom is left. At the time isospin was thought to be conserved, and nature invariant under rotations in isospin space. Yang and Mills found that within a field theory description the choice to label the isospin states should in principle be allowed to change from one space-time point to another [96]. Thus they gauged $SU(2)$ and arrived at a theory similar to Maxwell’s electromagnetism⁴, however with charged ‘photons’.

Nowadays, we know that isospin is not an exact symmetry⁵, and so it cannot be gauged in a mathematically consistent way. However, non-Abelian gauge theories have become a central ingredient in our understanding of

³From isotopic spin.

⁴The principle of gauge invariance, i.e. that a symmetry operation can be performed at different locations independently, is due to Weyl’s work on electromagnetism. Incidentally electromagnetism is the prototype, and simplest, gauge theory based on the group $U(1)$.

⁵The different up and down quark masses explicitly break isospin.

the strong and weak interactions: they are described by $SU(3)$ and $SU(2)$ gauge groups respectively. They are also instrumental in physics beyond the Standard Model, for instance in grand unified theories.

The Lagrangian formalism is the most straightforward to develop models because the action has to be a scalar under any symmetry transformation the theory is supposed to possess. In the present case, the action has to be invariant under local $SU(N)$ gauge transformations. Therefore, it can be constructed by patching together into singlets the elementary building blocks that transform under irreducible representations.

One of these building blocks is the covariant derivative

$$D_\mu \equiv \partial_\mu - iA_\mu, \quad (2.1)$$

with $A_\mu \equiv A_\mu^a T^a$, where A_μ^a is the gauge field and the generators T^a span the Lie algebra $su(N)$; here it is given in the fundamental representation. In general, the generators are normalised according to $\text{Tr } T^a T^b = \frac{1}{2} \delta^{ab}$, with commutators $[T^a, T^b] = i f^{abc} T^c$. The derivative operator (2.1) transforms covariantly by definition and therefore according to

$$A_\mu \rightarrow U(A_\mu + i\partial_\mu)U^\dagger, \quad (2.2)$$

with $U = \exp(i\alpha^a T^a) \in SU(N)$. The covariant derivative can be used to build new covariantly transforming quantities. One such follows from the commutator of the covariant derivative with itself

$$i[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu] \equiv F_{\mu\nu}. \quad (2.3)$$

In components, $F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + f^{abc} A_\mu^b A_\nu^c$. Demanding that the field equations be second order, we can build a kinetic term for the gluons from $F_{\mu\nu}$,

$$L_{\text{YM}} = -\frac{1}{2g^2} \text{Tr } F_{\mu\nu} F^{\mu\nu}, \quad (2.4)$$

where the normalisation follows from a comparison with Maxwell's theory, and g is the strong coupling constant. The field equations follow from varying

this action and are given by

$$D_\mu G^{\mu\nu} = 0, \quad (2.5)$$

where, here, D_μ is the covariant derivative in the adjoint representation, i.e.

$$D_\mu = \partial_\mu - i[A_\mu, \cdot]. \quad (2.6)$$

In terms of the adjoint covariant derivative, an infinitesimal gauge transformation can be written as

$$\delta A_\mu = D_\mu \alpha. \quad (2.7)$$

2.1.1 Gauge fixing

The gauge symmetry is really a redundancy in the description of the system. In particular, field space in terms of the gauge potential is very large and contains many copies of the same physical configuration, related of course through gauge transformations. This leads to equivalence classes, which can be labelled by a set of gauge-invariant observables. Choosing one such representative from a given class and applying all possible gauge transformations to it gives rise to the notion of gauge orbit; the gauge orbit thus comprises all configurations that belong to the same equivalence class.

To pass to the Hamiltonian formalism, we need to relate the generalised momenta to the generalised velocities, and this mapping needs to be one-to-one. Due to the redundancies in the Lagrangian treatment this will not in general be the case. The standard technique is to fix the gauge so that the mapping becomes faithful.

Starting from one representative of each equivalence class, we can generate the initial configuration space by applying all possible gauge transformations. A useful picture to have in mind is that the total configuration space has the structure of a bundle over the equivalence classes with the gauge group as fibre. Gauge fixing then amounts to choosing a cross-section in that bundle space. Note that by definition this corresponds to total gauge fixing, i.e. the prescription gives a unique representative from each equivalence class, see Fig. 2.1. In practice however, the gauge might not be totally fixed.

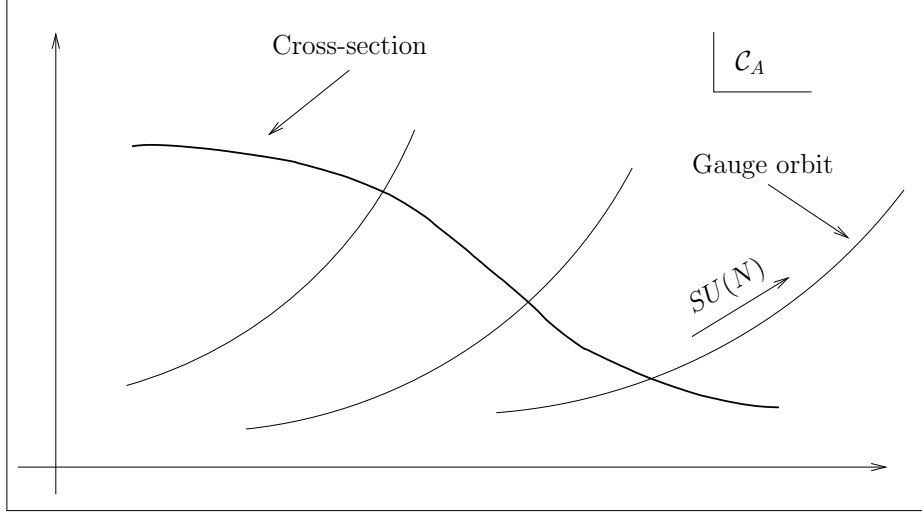


Figure 2.1: The configuration space \mathcal{C}_A of a gauge theory can be foliated by gauge orbits. This, effectively, turns field space into a bundle over equivalence classes; the fibre, running along the orbit, is given by the gauge group. Fixing the gauge completely amounts to choosing some cross-section. The latter is the (unique) solution to $F[A] = 0$, with F the gauge choice functional.

Many different gauge choices exist, e.g. temporal gauge $A_0 = 0$, axial gauge $A_i = 0$, Lorentz gauge $\partial_\mu A^\mu = 0$, or background gauge $D_\mu A^\mu = 0$, where the covariant derivative is with respect to a classical background field A_{cl} by which the original gauge field has been shifted $A \rightarrow A + A_{cl}$ and acts now on the quantum field A . In general, the gauge choice can be parametrised by $F[A] = 0$, whose (unique) solution describes a cross-section through configuration space. Typically, some residual gauge freedom is left over though, and $F[A] = 0$ does not possess a unique solution.

2.1.2 Vacuum structure

Non-Abelian gauge theories have a topologically non-trivial classical vacuum structure. The Hamiltonian density of the Yang-Mills theory, in $A_0 = 0$

gauge⁶, is given by

$$H = \frac{1}{2g^2} \left(\dot{A}_i^a \dot{A}_i^a + \frac{1}{2} F_{ij}^a F_{ij}^a \right) = \frac{1}{2g^2} (E_i^a E_i^a + B_i^a B_i^a), \quad (2.8)$$

which is manifestly positive, and we defined $E_i^a \equiv \dot{A}_i^a$, $B_i^a \equiv \frac{1}{2} \epsilon_{ijk} F_{jk}^a$. From this we clearly see that time independent pure gauge configurations, $A_i^a = i/g U \partial_i U^\dagger$ with $\dot{U} = 0$, minimise H . If we were to couple matter to this theory (in a $SU(N)$ -invariant way), the global gauge rotations would lead to currents through Noether's theorem. These symmetry currents are assumed to fall off fast enough at spatial infinity, which translates into $A_\mu \rightarrow 0$ or $U(|\vec{x}| \rightarrow \infty) \rightarrow \text{const.}$ This compactifies \mathbb{R}^3 to S^3 , and therefore the classical vacua are characterised by mappings $S^3 \rightarrow S^3$. The second S^3 follows from the fact that the group manifold of $SU(2)$ is topologically equivalent to S^3 , and because $SU(2) \in SU(N)$. These mappings are non trivial, and the corresponding vacua fall into distinct (homotopic) equivalence classes, labelled by the winding number

$$n(U) = -\frac{1}{24\pi^2} \int_{S^3} dV \text{Tr} (\epsilon_{ijk} U \partial_i U^\dagger U \partial_j U^\dagger U \partial_k U^\dagger), \quad (2.9)$$

which is invariant under homeomorphisms [42], is an integer and defines a homomorphism because $n(U_1 U_2) = n(U_1) + n(U_2)$ ⁷. In particular, the above construction shows that the vacuum manifold of Yang-Mills theory is not simply connected.

By means of the stereographic projection we can pull the integrand back to \mathbb{R}^3 , and n can then be related to the Chern-Simons current. In formulas,

$$n = \frac{1}{16\pi^2} \int d^3x K_0, \quad (2.10)$$

$$K_\mu = 4\epsilon_{\mu\nu\delta\gamma} \text{Tr} \left[A^\nu \left(\partial^\gamma A^\delta - \frac{2i}{3} A^\gamma A^\delta \right) \right]. \quad (2.11)$$

⁶Although this does not fix the gauge completely (the residual gauge freedom consists of time-independent gauge transformations), it is sufficient to get a one-to-one mapping between the generalised momenta and velocities and, hence, to write down the Hamiltonian.

⁷Which is easily proved by using the cyclic property of the trace, anti-hermiticity of $U \partial U^\dagger$ and the unitarity constraint $U(\partial U^\dagger) = -(\partial U)U^\dagger$.

Gauge transformations are called large unless they satisfy $n = 0$, in which case they are called small. A representative of a large gauge transformation with $n = 1$ is given by

$$U = \exp \left(i\pi \frac{\vec{x}\vec{\sigma}}{\sqrt{\vec{x}^2 + \rho^2}} \right), \quad (2.12)$$

where σ_i are the Pauli matrices and ρ is an arbitrary constant. From the homomorphism property of the winding number it follows that U^m belongs to the m -sector. Thus, large gauge transformations relate the different classical vacua, i.e. $n \rightarrow n+m$. The latter are physically equivalent, being mere gauge copies of the trivial vacuum.

Non-vacuum configurations have higher energy density, and their winding number will no longer be guaranteed to be an integer. Because of the potential barrier between integer winding numbers, the vacuum manifold cannot be simply connected⁸. This is the more pedestrian way to convince ourselves that the vacuum manifold is topologically non-trivial. Since the different classical vacua are physically equivalent, the winding number is sometimes said to correspond to motion along a circle which encloses the topological obstruction, or ‘hole’, in function space, see [161, 177].

2.2 Quantum theory and QCD

The topologically non-trivial structure of the classical vacua that we have been describing is reminiscent of the motion of a particle on an upright circle in a gravitational field, and is schematically indicated in Fig. 2.2. In the classical theory the configurations labelled by the winding number n give equivalent groundstates. However, in quantum mechanics these classical configurations become smeared out to form the states $|n\rangle$, the so-called n -vacua; they correspond to wave-functionals that are peaked around the respective classical minima. Using the analogy of the quantum behaviour of a particle in a such a potential, we know that the true quantum vacuum has a very different structure due to tunnelling. We will now investigate this issue for

⁸Transformations that change n as well as stay on the vacuum manifold cannot be homotopies, because the latter, by definition, belong to the trivial sector.

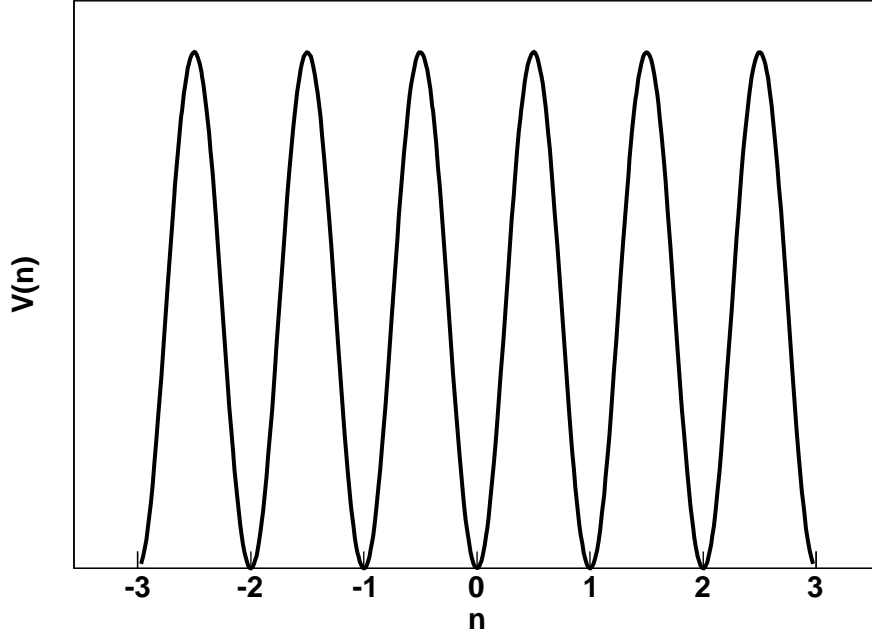


Figure 2.2: The direction in function space that encloses the topological obstruction is given by the winding number n . The energy density is minimised for integer n , and has necessarily potential barriers, V , between these vacuum configurations. The configurations labelled by n are physically equivalent as they are related by large gauge transformations.

gluodynamics.

2.2.1 Euclidean formulation, θ -vacuum and instantons

To understand the vacuum structure of a field theory, it turns out to be advantageous to study the Euclidean sector, because the Euclidean theory can be used to describe tunnelling phenomena semi-classically, see [42, 161, 27].

The continuation to imaginary time, i.e. $x_M^i \rightarrow x_E^i$ and $x_M^0 \rightarrow -ix_E^4$ in order to evade the mass poles, induces the following changes for the gluon field:

$$A_i^M \rightarrow A_i^E, \quad (2.13)$$

$$A_0^M \rightarrow iA_4^E. \quad (2.14)$$

Note that, quite generally, the zero component of a contravariant index gets a factor of $-i$, whereas the zero component of a covariant index gets a factor of i . Defining the Euclidean tensors to have a lower index⁹, we pull down the upper index and include an extra minus sign; this takes care of the negative definite metric and we need not bother anymore about the difference between co- and contra-variant tensors.

The Euclidean action becomes $S_M \rightarrow iS_E$ with¹⁰

$$S_E = \int d^4x \frac{1}{2g^2} \text{Tr} F_{\mu\nu}^E F_{\mu\nu}^E. \quad (2.15)$$

In the semi-classical approximation to the path integral that we will develop in chapter 3, finite action configurations are most important since they build the starting point for a non-trivial perturbative expansion. Actually, finite action configurations form a set of measure zero [42], but since in the semi-classical expansion we restrict the function space to the neighbourhood of finite action configurations by definition, this fact is of no consequence.

To get finite action configurations on \mathbb{R}^4 , we demand that the field becomes pure gauge at infinity. These boundary conditions define a mapping from infinity, which is topologically equivalent to S^3 , into the gauge group. Following the same logic as in the previous section, the finite action configurations define mappings $S^3 \rightarrow S^3$. We have already learned that these fall into distinct equivalence classes, which in this case are labelled by another winding number, the topological charge,

$$Q_E = \int d^4x \frac{1}{16\pi^2} \int d^4x \text{Tr} F_{\mu\nu}^E \tilde{F}_{\mu\nu}^E, \quad (2.16)$$

where $\tilde{F}_{\mu\nu}^E \equiv \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} F_{\rho\sigma}^E$ is the dual field strength. This winding number is gauge invariant, in contrast to n , and takes values in \mathbb{N} .

A note on the relation to the Minkowski counterparts is in order. The Minkowski dual field strength is defined by $\tilde{F}_{\mu\nu}^M \equiv \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} F_M^{\rho\sigma}$. Under analytic

⁹Or an upper index; as long as we stay consistent with one choice, the results will be equivalent.

¹⁰Note that the integration measure has naturally an upper index, i.e. $dx^0 \rightarrow -idx^4$.

continuation¹¹ we then have

$$\begin{aligned} F_M^{\mu\nu} \tilde{F}_{\mu\nu}^M &\rightarrow -i F_{\mu\nu}^E \tilde{F}_{\mu\nu}^E, \\ \partial^\mu K_\mu^M &\rightarrow -i \partial_\mu K_\mu^E. \end{aligned} \quad (2.17)$$

It can be shown¹² that the topological charge is related to (2.11) by

$$Q_E = \int d^4x \partial_\mu K_\mu^E, \quad (2.18)$$

which holds equally well in Minkowski space due to the transformations (2.17). Using Gauss' theorem we find that

$$Q_M = \int d^3x [K_0^M(t = +\infty) - K_0^M(t = -\infty)] = n(t = +\infty) - n(t = -\infty), \quad (2.19)$$

and we worked in Minkowski space-time and in the $A_0^M = 0$ gauge to make contact with the winding of the classical n -vacua. Clearly these finite action configurations interpolate between the classical vacua, and they are identified as tunnelling transitions [27]. One heuristic way to convince ourselves that these are tunnelling phenomena is to note that they do not cost any (Minkowski) energy. To see it, note that for self-dual configurations we have that $\vec{E}_M^a = i\vec{B}_M^a$, which annihilates the Hamiltonian density. From now on we will keep to the Euclidean theory and drop the M - and E -index.

In the classical theory each n -vacuum provided a suitable groundstate, even though it is not gauge invariant. In the quantum theory we have just learned that the n -vacua are no longer disconnected. Therefore the superposition principle forces us to superpose the different n -vacua.

It turns out that the rather large Hilbert space can be reduced through a super-selection rule. Note that the operator that implements gauge transformations commutes with the Hamiltonian and is unitary. We can therefore divide the Hilbert space into sectors labelled by the phase of the gauge transformation operator. In particular, there are different vacuum states $|\theta\rangle$ that change under m -sector gauge transformations into $U_m|\theta\rangle = e^{-i\theta m}|\theta\rangle$. Such a

¹¹ $\tilde{F}_{0i}^M \rightarrow \tilde{F}_{0i}^E, \tilde{F}_{ij}^M \rightarrow -i\tilde{F}_{ij}^E, K_0^M \rightarrow -K_0^E, K_i^M \rightarrow iK_i^E$.

¹²This is straightforward; the only 'tricky' part involves exploiting the invariance of the ϵ pseudo-tensor under cyclic permutations.

state can be constructed from the n -vacua according to

$$|\theta\rangle = \sum_{n=-\infty}^{+\infty} e^{in\theta} |n\rangle, \quad (2.20)$$

where we use that $U_m |n\rangle = |n+m\rangle$.

A gauge invariant operator O has no overlap between these smaller Hilbert spaces labelled by θ , as follows from

$$0 = \langle \theta' | [O, U_m] | \theta \rangle = \left(e^{-i\theta m} - e^{-i\theta' m} \right) \langle \theta' | O | \theta \rangle. \quad (2.21)$$

Since physical operators need to be gauge invariant, no process can change θ ; it is a fundamental constant of our universe. All of this is beautifully described in [177]. Following [152], the path integral is defined through the vacuum-to-vacuum transition

$$Z(\theta) = \langle \theta | \exp(-iHt) | \theta \rangle = \sum_q e^{iq\theta} \sum_n \langle n+q | \exp(-iHt) | n \rangle. \quad (2.22)$$

Remembering that the n -vacua are all equivalent, the last term defines the usual gauge invariant path integral $Z_q \equiv \sum_n \langle n+q | e^{-iHt} | n \rangle$, with the restriction that only fields with topological charge $Q = q$ are included. This then leads to a new path integral, summed over all Q sectors, that incorporates the topological θ -term

$$Z \equiv \int [dA] \exp \left[- \int d^4x \left(\frac{1}{2g^2} \text{Tr} F_{\mu\nu} F_{\mu\nu} + i \frac{\theta}{16\pi^2} \text{Tr} F_{\mu\nu} \tilde{F}_{\mu\nu} \right) \right]. \quad (2.23)$$

Within the semi-classical approach, the minimal action configuration corresponds to the most probable tunnelling event. Note that the action can be rearranged into the following form,

$$S = \pm \frac{8\pi^2}{g^2} Q + \frac{1}{8g^2} \int d^4x (F_{\mu\nu}^a \mp \tilde{F}_{\mu\nu}^a)^2, \quad (2.24)$$

which shows that the action is minimised, in a given equivalence class, by (anti-)self-dual configurations, i.e. $\tilde{F}_{\mu\nu}^a = \pm F_{\mu\nu}^a$. Thus, such configurations are solutions to the field equations but can be obtained much more easily

from these first order differential equations.

Perhaps the most famous such solution is the BPST instanton [22]. It belongs to the $Q = 1$ sector and is $SO(4)$ symmetric. The latter condition implies that the gauge potential is of the form

$$A_\mu = iU\partial_\mu U^\dagger f(x^2), \quad (2.25)$$

with $f(x^2) \rightarrow 1$ for $|x| \rightarrow \infty$, to comply with the boundary condition. At the origin we require f to cancel the singularity that is inevitably present because of the fact that U is defined only on the 3-sphere at infinity. A representative from the $Q = 1$ equivalence class is given by

$$U = \frac{x_4 + ix_i \sigma_i}{\sqrt{x^2}}. \quad (2.26)$$

Using this form, it is straightforward to show that at infinity

$$A_\mu^a = \frac{2}{x^2} \underbrace{(\delta_{a\mu} \delta_{\nu 4} + \epsilon_{a\mu\nu} - \delta_{a\nu} \delta_{\mu 4})}_{\equiv \eta_{a\mu\nu}} x_\nu, \quad (2.27)$$

and $\eta_{a\mu\nu}$ are the 't Hooft symbols that entangle space and group indices. The self-duality equation leads to the simple differential equation

$$f'x^2 - f + f^2 = 0, \quad f' = \frac{df(s)}{ds}, \quad (2.28)$$

which, together with the boundary conditions for f , leads to the BPST instanton solution,

$$A_\mu^a = 2\eta_{a\mu\nu} \frac{x_\nu}{x^2 + \rho^2}, \quad (2.29)$$

where ρ is a constant of integration. The winding at infinity can be traded for a winding at the origin by means of a singular gauge transformation, given by $V = U^\dagger$.

A slightly more general solution is given by 't Hooft's multi-instanton

ansatz, in singular gauge, and reads

$$A_\mu^a = -\bar{\eta}_{a\mu\nu} \partial_\nu \log \Pi, \quad (2.30)$$

$$\Pi = 1 + \sum_{i=1}^N \frac{\rho_i^2}{(x - x_i)^2}. \quad (2.31)$$

This ansatz can be used to construct the finite temperature caloron solution of Harrington–Shepard (HS) [93]. The latter is the generalisation of the BPST instanton to finite temperature, and is crucial for studying the restructuring of the QCD vacuum at finite temperature in the semi-classical approach. Finite temperature field theory can be described by a Euclidean formulation, with bosonic (fermionic) field subject to periodic (anti-periodic) boundary conditions in the time direction, with extent $\beta = 1/T$, see for instance [103]. The HS solution is explicitly realised by lining up the BPST instanton periodically along the time direction, and has the form

$$\Pi = \frac{\pi \rho^2}{\beta r} \frac{\sinh \frac{2\pi r}{\beta}}{\cosh \frac{2\pi r}{\beta} - \cos \frac{2\pi t}{\beta}}. \quad (2.32)$$

The most general solution to the Yang–Mills field equations follows from the ADHM construction [12]. Explicit solutions, believed to be important for the semi-classical treatment of the confinement/deconfinement transition, have recently been derived within the full generality of the ADHM formalism: the so-called non-trivial holonomy, or KvBLL, calorons [119, 121, 120, 123].

2.2.2 Perturbation theory

The path integral as defined in (2.23), with or without θ , is ill-defined perturbatively because we sum over equivalent configurations an infinite amount of times, and because of ultraviolet (UV) divergences.

Since we do not perform any perturbative computations in the remainder, except for a low-order one where no complications arise, we will be rather brief.

Gauge fixing

We know already, in principle, how to solve the first problem, namely by restricting the path integral to include only one representative of each gauge orbit, i.e. by gauge fixing. To perform perturbation theory, we need to define the propagator, i.e. the inverse of the quadratic part of the kinetic term. Although the quadratic part is no longer gauge invariant under the full group in general, it will still be invariant under the centre¹³. This restricted gauge group leads to gauge zero modes for the propagator and the inverse does not exist. As long as the gauge fixing breaks the centre completely perturbation theory can be defined.

Note that boundary conditions play an important role with regard to the residual gauge freedom. For instance in the Abelian case, Lorentz gauge does not fix the gauge completely but has a residual gauge freedom for gauge parameters that are harmonic, i.e. $A_\mu \rightarrow A_\mu + \partial_\mu \alpha$ with $\partial^2 \alpha = 0$. However, the boundary condition that $A_\mu \equiv 0$ at infinity on \mathbb{R}^4 , restricts $\alpha = \text{const}$ and thus lifts the gauge freedom completely.

Gauge fixing of the Lagrangian path integral (2.23) is achieved through the Fadeev-Popov technique¹⁴, which amounts to inserting a factor of unity in the path integral; in formulas,

$$1 \equiv \Delta[A_\mu] \int [dU] \delta[F[A_\mu^U]], \quad (2.33)$$

where A_μ^U describes the gauge orbit, and the cross-section over the bundle that singles out a representative of each gauge orbit is given by F . The path integral is then transformed into

$$Z \rightarrow \left(\int [dU] \right) \int [dA] \Delta[A] \delta[F[A]] \exp(-S - i\theta Q), \quad (2.34)$$

and the infinite volume factor $\int [dU] = \prod_x V^{SU(N)}$ is irrelevant in expectation values.

Since we need a constraint for each A^a , F is a vector in the adjoint representation, i.e. $F = F^a T^a$. The factor of Δ is defined by (2.33), and can

¹³The centre is defined by $Z_G \equiv \{z \in G | zg = gz, \forall g \in G\}$.

¹⁴We follow [152], which should be consulted for details.

be shown to be formally given by

$$\Delta[A_\mu] = \det \left. \frac{\delta F[A_\mu^U]}{\delta U} \right|_{U=1}, \quad (2.35)$$

where we used the gauge invariance of Δ to choose A such that $F[A] = 0$. More precisely, we have that

$$\Delta[A_\mu] = \det M^{ab}, \text{ with } M^{ab} \equiv \left. \frac{\delta F^a[A_\mu^\alpha]}{\delta \alpha^b} \right|_{\alpha^b=0}. \quad (2.36)$$

To perform the functional derivative, we use that $\delta A_\mu = D_\mu \alpha$ for an infinitesimal gauge transformation. Thus

$$M^{ab}(x, y) = \int d^4 z \frac{\delta F^a[A(x)]}{\delta A_\mu^c(z)} (\partial_\mu^z \delta_{bc} + f^{bcd} A_\mu^d(z)) \delta^4(z - y). \quad (2.37)$$

By introducing a set of auxiliary anti-commuting scalar fields, so-called Fadeev-Popov ghosts, the determinant can be put in the form of an, a priori, non-local Lagrangian, i.e.

$$\Delta[A_\mu] = \int [dc][d\bar{c}] \exp \left(- \int d^4 x d^4 y \bar{c}^a(x) M^{ab}(x, y) c^b(y) \right). \quad (2.38)$$

A standard choice is $F^a[A_\mu] = \partial_\mu A_\mu^a$, i.e. Lorentz gauge. In this case $\frac{\delta F^a[A(x)]}{\delta A_\mu^c(z)} = \partial_\mu^x \delta^4(x - z) = -\partial_\mu^z \delta^4(z - x)$ and, integrating by parts, we get $M^{ab}(x, y) = \partial_\mu D_\mu^a \delta^4(x - y)$; this then leads to a local action.

In chapter 3 we will be discussing QCD in the presence of instantons, and a convenient choice will be the so-called background gauge¹⁵

$$F[A] = D_\mu A_\mu, \quad (2.39)$$

in which case we find that

$$M(x, y) = \frac{\delta F}{\delta \alpha} = (D_\mu D_\mu - i D_\mu [A_\mu, \cdot]) \delta^4(x - y), \quad (2.40)$$

$$= D_\mu D_\mu \delta^4(x - y) - i [A_\mu, D_\mu \delta^4(x - y)], \quad (2.41)$$

¹⁵The covariant derivative is with respect to the classical background and acts on the quantum field.

where in the second line we made use of the background condition.

Renormalisation

A local field theory is generally plagued by UV divergences, and needs to be regularised and then renormalised. Regularisation tames the short-wave fluctuations. Standard techniques include a finite cutoff in the momentum space integrals, introduction of auxiliary fields with non-standard kinetic terms (Pauli-Villars), analytic continuation (dimensional regularisation) or a discrete space-time (lattice gauge theories). The second step, renormalisation, consists of shifting the parameters of the theory in such a way as to give finite results when the regulator is removed. Here, again, various prescriptions can be used, e.g. mass-shell renormalisation, momentum subtraction, or minimal or modified minimal subtraction (\overline{MS}) to name a few. For details see [44].

In super- and renormalisable theories, the parameters that appear at tree-level together with wave-function renormalisation are sufficient to render every computation finite. Even though we cannot infer the underlying bare parameters, the theory contains the same amount of free parameters as the bare theory. On the other hand, in non-renormalisable theories new interactions, and new experimental input, needs to be taken into account at higher orders in perturbation theory. In finite theories, finally, we can compute the shifts that the quantum interactions induce between the measured and bare parameters of the theory. Even though not described by field theory, a useful heuristic picture to keep in mind in this respect is the modification of the physical electron mass in condensed matter due to its interactions with the phonons. Here we could also renormalise the electron mass and parametrise the theory with the effective, observed electron mass but we never need to regularise the theory.

The bare and renormalised quantities are related through renormalisation constants Z_i . In the case of Yang-Mills theory these are the coupling constant and the field amplitudes for the gauge fields and the ghosts. The different regularisation and renormalisation prescriptions implicitly define these parameters. On a heuristic level we can view these parameters as the coordinates of some manifold that describes the theory. The different prescriptions

can then be viewed as choosing different coordinates on this manifold. It is important that predictions within different schemes can be related to each other. This is particularly important in the lattice community, that usually converts results into the \overline{MS} scheme to compare with other approaches. Such a procedure runs under the name of matching, see for instance [51].

The standard prescription to include quarks is through minimal coupling

$$L_m = \bar{\psi} (\not{D} + m) \psi, \quad (2.42)$$

here in the Euclidean formulation¹⁶. This adds two more renormalisation constants, for the quark mass and field amplitude. In the so-called renormalised perturbation theory the renormalised fields and coupling constants are used to perform perturbation theory, and the shifts between the latter and the original bare parameters are treated as perturbations that depend on the renormalisation constants. The latter are adjusted order by order to maintain finite results.

It can be shown that QCD is renormalisable and that gauge invariance survives in the quantum theory. This puts strong constraints on the form of the renormalised Lagrangian; it takes the form [152]¹⁷

$$L = \frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a + \bar{\psi} (\not{D} + m) \psi + \bar{c}^a \partial_\mu D_\mu c^a + L_{\text{c.t.}}, \quad (2.43)$$

where we used Lorentz gauge, and $L_{\text{c.t.}}$ is the counter-term Lagrangian, i.e.

$$\begin{aligned} L_{\text{c.t.}} = & (Z_2 - 1) \bar{\psi} \not{D} \psi - (Z_1 - 1) i g \bar{\psi} \not{A} \psi + (Z_0 - 1) m \bar{\psi} \psi + (\tilde{Z}_2 - 1) \bar{c}^a \partial^2 c^a \\ & + (\tilde{Z}_1 - 1) g \bar{c}^a f^{abc} A_\mu^b \partial_\mu c^c + \frac{1}{4} (Z_3 - 1) (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)^2 \\ & - \frac{1}{2} (Z_1^{\text{YM}} - 1) g (\partial_\mu A_\nu^a) f^{abc} A_\mu^b A_\nu^c + \frac{1}{4} (Z_1^{\text{YM}2} Z_3^{-1}) f^{abc} f^{ade} A_\mu^b A_\nu^c A_\mu^d A_\nu^e. \end{aligned} \quad (2.44)$$

¹⁶More details are given in section 2.3.

¹⁷More details can be found in [44]

The relation between the bare and renormalised parameters is given by

$$A_{B\mu}^a = Z_3^{1/2} A_\mu^a, \quad (2.45)$$

$$\psi_B = Z_2^{1/2} \psi, \quad (2.46)$$

$$c_B^a = \tilde{Z}_2^{1/2} c^a, \quad (2.47)$$

$$m_B = m Z_0 / Z_2, \quad (2.48)$$

and gauge invariance leads to the following relation among the renormalisation constants,

$$Z_3^{1/2} g_B / g = Z_1^{\text{YM}} / Z_3 = Z_1 / Z_2 = \tilde{Z}_1 / \tilde{Z}_2 = \text{finite}. \quad (2.49)$$

This last relation states that even after renormalisation, the same coupling constant gives the interaction strength for the quark-gluon, and 3- and 4-gluon vertex. In particular, the coupling constant follows uniquely from the gluon interaction and does not depend on the particle content in the sense that any matter particles couple to gluons through g .

Renormalisation group and matching

As described in the last subsection, there is considerable arbitrariness in the renormalisation program. This can be used to find particularly convenient schemes. For instance, it allows us to rearrange the perturbation series in expansion parameters that might lead to faster convergence.

A very convenient prescription in this respect is \overline{MS} , which defines the theory at an arbitrary scale μ . This scale is then chosen such that the perturbation series is well-behaved and in particular such that large logarithms, due to higher order contributions, are small, see [200] for a detailed discussion. This leads to the so-called running coupling constant and masses, and the evolution of these free parameters with respect to μ is encoded in the renormalisation group equations (RGE's); μ is often called the RGE scale.

In the Wilsonian picture of renormalisation the scale μ can be viewed as a floating cutoff which separates the soft from the hard modes. The latter have been integrated out and produce the logarithms that combine with the coupling constants (and masses); due to renormalisability, their sole effect

is to change these free parameters. In effect, then, the running coupling constants encode the fluctuations that have already been integrated out.

The RGE scale is an artefact from the renormalisation program and, in particular, the bare parameters, with fixed regulator, are independent of it. The latter observation allows us to derive equations that encode the renormalisation flow. From the relations between bare and renormalised parameters, i.e.

$$g_B = Z_g^{-1}(\mu)g(\mu), \quad (2.50)$$

$$m_B = Z_m^{-1}(\mu)m(\mu), \quad (2.51)$$

the following RGE's can be derived:

$$\mu \frac{dg}{d\mu} = g \frac{d \ln Z_g}{d \ln \mu} \equiv \beta \quad (2.52)$$

$$\frac{d \ln m}{d \ln \mu} = \frac{d \ln Z_m}{d \ln \mu} \equiv \gamma_m. \quad (2.53)$$

In a mass-independent renormalisation scheme, such as \overline{MS} , the renormalisation constants only depend on g , and thus β and γ only depend on g ; this allows for a simple solution:

$$\ln \frac{\mu}{\mu_0} = \int_{g_0}^g \frac{d\tilde{g}}{\beta(\tilde{g})}, \quad (2.54)$$

$$m(g) = m(g_0) \exp \left(\int_{g_0}^g \frac{\gamma_m(\tilde{g})}{\beta(\tilde{g})} d\tilde{g} \right), \quad (2.55)$$

$$= m(\mu_0) \exp \left(\int_{\mu_0}^{\mu} \frac{\gamma_m(\tilde{\mu})}{\beta(\tilde{\mu})} d \ln \tilde{\mu} \right) = m(\mu). \quad (2.56)$$

The RG functions β and γ_m can be computed in perturbation theory. It turns out that the first two coefficients of the β function and the first coefficient of γ_m are universal and gauge invariant [44]. In particular, the first coefficient of the β function is negative, so that the coupling constant decreases with increasing μ ; this is asymptotic freedom. At one-loop, the running coupling is given by

$$\alpha_s(\mu) = \frac{4\pi}{b_0 \ln \mu^2 / \Lambda}, \quad (2.57)$$

where $\alpha_s = g^2/4\pi$ and Λ is the QCD scale.

Another quantity we will be interested in is the quark condensate $\langle \bar{\psi}\psi \rangle$. It is possible to write down RGE's for such local operators. The renormalised bi-quark operator is related to the bare one by $(\bar{\psi}\psi)_B = Z_{\bar{\psi}\psi}^{-1} \bar{\psi}\psi$. We know that in the Lagrangian this term appears multiplied by m , with their combination adding up to $m_B(\bar{\psi}\psi)_B$. Thus $\gamma_{\bar{\psi}\psi} = -\gamma_m$, and the quark condensate runs inversely to the quark masses. The coefficient $\gamma_{\bar{\psi}\psi}$ is called the anomalous dimension of the local bi-quark operator.

In chapter 3 we need to relate our results for the quark condensate to those from the lattice. The latter is quoted in \overline{MS} whereas the scheme that we will use is Pauli-Villars (PV) together with a subtraction scheme analogous to \overline{MS} . To disentangle the contribution from the RGE running, we will compare the local operators at a common scale. We have just learned that $m\bar{\psi}\psi$ is RGE-invariant, so that we can work either with the quark condensate or the mass to run to the common scale; we choose the latter.

To match the PV and \overline{MS} masses, we will use the pole mass to relate. From the form of the full propagator

$$S_F = \frac{1}{\not{p} - m - \Sigma} = \frac{1}{\not{p}(1 - \Sigma_V) - m(1 + \Sigma_m)}, \quad (2.58)$$

where Σ is the quark self-energy, we can infer that the pole mass is given by

$$m_{\text{pole}}^2 = m^2 \left(\frac{1 + \Sigma_m(m_{\text{pole}})}{1 - \Sigma_V(m_{\text{pole}})} \right)^2. \quad (2.59)$$

As discussed in the previous subsection, different schemes will give different parametrisations but the results must be equivalent. Therefore, at one-loop¹⁸, we have that

$$\begin{aligned} m_{PV} (1 + \Sigma_m^{PV}(m_{PV}^2) + \Sigma_V^{PV}(m_{PV}^2)) \\ = m_{\overline{MS}} \left(1 + \Sigma_m^{\overline{MS}}(m_{\overline{MS}}^2) + \Sigma_V^{\overline{MS}}(m_{\overline{MS}}^2) \right), \end{aligned} \quad (2.60)$$

and the self-energy is given by Fig. 2.3.

¹⁸Maintaining manifest gauge-invariance in Yang-Mills theories using Pauli-Villars regularisation is not straightforward beyond one-loop.

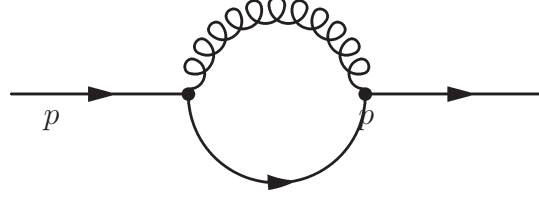


Figure 2.3: Feynman diagram needed to compute the difference between the $\overline{\text{MS}}$ and PV scheme at one-loop.

This is a textbook computation, [151]. After subtracting off the divergences, we end up with

$$\Sigma_{\text{PV}} = \frac{\alpha_s}{2\pi} C(3) \left\{ -2m + \frac{1}{4} \not{p} + \int dx (2m - (1-x)\not{p}) \ln \frac{\mu^2}{xm^2 - x(1-x)p^2} \right\}, \quad (2.61)$$

and

$$\Sigma_{\overline{\text{MS}}} = \frac{\alpha_s}{2\pi} C(3) \left\{ -m + \frac{1}{2} \not{p} + \int dx (2m - (1-x)\not{p}) \ln \frac{\mu^2}{xm^2 - x(1-x)p^2} \right\}. \quad (2.62)$$

Relating both through the pole mass and using that $C(3) = 4/3$, we get

$$m_{\overline{\text{MS}}} = m_{\text{PV}} \left(1 - \frac{\alpha_s}{2\pi} \frac{5}{3} \right). \quad (2.63)$$

2.3 Strong CP problem

As we've seen before, quarks are included through minimal coupling. In the case of QCD we couple N_f quarks in the fundamental representation to $SU(3)$ gluodynamics. Staying in the Euclidean sector, the quark Lagrangian reads

$$L_m = \bar{\psi} (\not{D} + M) \psi, \quad (2.64)$$

where M is the mass matrix. The Euclidean γ -matrices are chosen to be hermitian and to satisfy $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$; they are given by $\gamma_0 = \gamma_M^0$ and $\gamma_i = -i\gamma_M^i$. Note that $\bar{\psi}$ transforms as ψ^\dagger in Euclidean space.

If the mass matrix is proportional to the identity, the Lagrangian has a global $U(N_f)$ symmetry. If $M = 0$ we can rotate the left and right handed parts of ψ separately, due to the larger $U(N_f)_L \otimes U(N_f)_R$ symmetry. For quark masses small compared to the dynamically generated scale Λ , the massless Lagrangian is a good starting point and the quark masses can be treated as perturbations. This is the basis for chiral perturbation theory.

If the $U(N_f)_L \otimes U(N_f)_R$ symmetry was realised explicitly in nature, the hadrons should come in parity doublets, see for instance [152]. Such a multiplet structure is not seen experimentally, and the vacuum cannot be chirally symmetric. The vacuum need not preserve all symmetries however, and it turns out that the axial part of the chiral symmetry is spontaneously broken. This gives rise to massless states, the Goldstone bosons, carrying the quantum numbers of the broken generators. Since our symmetry was explicitly broken by the u , d and s quark masses, the bosons are only nearly massless states, and indeed such states can be identified with the pseudo-scalar octet.

This accommodates the spontaneous breaking of $SU_A(N_f)$. There is, however, no particle in the spectrum that would correspond to the Goldstone boson associated with the spontaneous breakdown of $U(1)_A$. This is the old $U(1)_A$ problem: “where is the 9’t h Goldstone boson?” [42].

The first step to solving this problem was the realisation that the axial current is anomalous, i.e. that it is not conserved quantum mechanically. This is a consequence of the famous Adler-Bell-Jackiw triangle anomaly [23, 2] and takes the form [152]

$$\partial_\mu j_\mu^5 = 2 \frac{g^2}{32\pi^2} F_{\mu\nu}^a \tilde{F}_{a\mu\nu}, \quad (2.65)$$

for each quark that is rotated.

Secondly, non-conservation of the current in itself is not sufficient to break the symmetry. A classical symmetry survives quantisation if its associated charge commutes with the Hamiltonian, i.e. if it is time-independent. Under

an axial rotation the charge changes by

$$\Delta Q^5 = \int dt \frac{d}{dt} Q^5 = \int d^4x \partial_t j_0^5(x) = 2Q, \quad (2.66)$$

where we neglected the contribution from $\partial_i j_i^5$ because by Gauss's theorem it only contributes a surface term which is assumed to vanish. We see that the axial charge is conserved as long as the gauge sector does not support topologically non-trivial field configurations. We have discussed at length that QCD has a non-trivial vacuum structure. Thus, we don't expect a Goldstone boson since $U_A(1)$ was no symmetry to start with. This solves the old $U(1)$ problem.

The very same non-trivial vacuum structure induces, however, a new problem because the θ term breaks CP explicitly¹⁹, whereas nature is very nearly CP invariant, so that $\theta \approx 0$ or $\theta \approx \pi$ ²⁰. The θ -term gives an electric dipole moment to the neutron [17, 45], which is tightly constrained experimentally [15] and results in the bound

$$\theta < 10^{-9}. \quad (2.67)$$

This is the new CP problem: why is θ so small?

The CP problem can be solved in QCD if one quark is massless. To see it, note that on the level of the path integral the anomaly stems from the non-invariance of the measure under axial transformations $\psi \rightarrow \exp(i\alpha\gamma_5)\psi$, i.e.

$$Z \rightarrow \int [dA][d\psi][d\bar{\psi}] \exp[-S - i(\theta + 2\alpha)Q]. \quad (2.68)$$

In the massless case we can conclude that the θ angle is unphysical as it can be rotated away by an axial transformation. For massive quarks, θ remains a physical parameter since it will be shifted into phases of the mass matrix; indeed, $\theta + \arg \det M = \text{const}$ under this $U_A(1)$.

There has been some controversy relating to this possibility [102, 78]: the non-zero quark mass used in chiral perturbation theory could in fact be an instanton induced 'soft' mass, even though the fundamental current mass is

¹⁹Under P we have that $F_{0i}^a \rightarrow -F_{0i}^a$ with the rest unchanged so that $F_{\mu\nu}^a \tilde{F}^{a\mu\nu} \rightarrow -F_{\mu\nu}^a \tilde{F}^{a\mu\nu}$. Under C $F_{\mu\nu}^a \rightarrow -F_{\mu\nu}^a$ so that $F_{\mu\nu}^a \tilde{F}^{a\mu\nu} \rightarrow F_{\mu\nu}^a \tilde{F}^{a\mu\nu}$.

²⁰Bear in mind that the quantum theory is defined via the path integral and $e^{i\pi} = e^{-i\pi}$.

zero, see for instance the discussion in [114, 115]. Recent lattice studies to determine the current quark masses seem to rule out this possibility [130]. Also, the electro-weak sector is known to break CP, and is well parametrised by the CKM matrix, generating complex quark mass phases. A massless quark could, however, be used to remove these phases from the Lagrangian²¹, and explicit CP breaking through the CKM mechanism would no longer work. Finally, assuming that CP is explicitly broken in the electro-weak sector, it seems natural that it is broken at tree-level in QCD too. There are, however, other approaches, see for instance [122].

Actually, the axial current gets anomalous contributions from all the gauge fields that the quarks couple to. In particular, other vacuum angles could be generated by the electro-weak sector. These are, however, unphysical: the vacuum angle for $U_Y(1)$ does not exist because the vacuum structure is trivial, i.e. only field configurations in the $Q = 0$ sector exist; the vacuum angle arising from the $SU_L(2)$ gluodynamics can be rotated away by the anomalous $U_B(1)$ transformation²².

2.3.1 Peccei Quinn Mechanism and Axions

Probably the most popular solution to the strong CP problem is provided by the Peccei-Quinn mechanism [146, 145]. Here, an extra chiral $U(1)_{PQ}$ is introduced to rotate the vacuum angle to zero, in much the same way as a massless quark allowed us to achieve this. Since such a symmetry is not observed in the spectrum, $U(1)_{PQ}$ is assumed to be spontaneously broken. The chiral phase gives rise to a pseudo-scalar Goldstone boson, the axion. In fact, the chiral symmetry is anomalous and the axion is not massless.

All models have the generic feature that the axion only couples deriva-

²¹Firstly, choose different $U_A^f(1)$ on the individual quark fields q_f to remove the complex phases. Secondly, the so generated θ -term can be rotated away by an axial transformation on the massless quark.

²²From the four anomalous (global) symmetries $U_e(1) \times U_\mu(1) \times U_\tau(1) \times U_B(1)$ we can construct three conserved currents and retain $U_B(1)$, say, as anomalous. Note that the rotations on the quark fields have to be identical across generations because $SU(3)$ is flavour-blind. Now, $U_B(1)$ is a vector symmetry, i.e. not broken by mass terms, so that the classical part of the Lagrangian remains invariant and the measure transformation gives only a contribution for the chiral gauge symmetries $SU_L(2)$ and $U_L(1)$ in the form of θ -terms; $SU(3)$, being a vector symmetry, does not contribute. See for instance [14] for details.

tively to matter and the only non-derivative coupling is to the topological charge

$$\mathcal{L}_a^{non-der.} = i \frac{\phi_a}{f_a} \frac{g^2}{32\pi^2} F_{\mu\nu}^a \tilde{F}_{\mu\nu}^a, \quad (2.69)$$

where ϕ_a is the axion field and f_a the axion decay constant. Thus, the axion adds another CP-violating contribution to the QCD Lagrangian unless it exactly cancels the θ term.

Indeed, the PQ mechanism solves the strong CP problem because the effective potential for the (homogenous) axion field has a CP-conserving minimum [113]

$$e^{-V V_{\text{eff}}(\phi)} = \left| \int [dA] \det(\not{D} + M) e^{-S + i(\theta + \frac{\phi}{f_a})\mathcal{Q}} \right|, \quad (2.70)$$

$$\leq \int [dA] \det(\gamma_\mu D_\mu + M) e^{-S} \left| e^{i(\theta + \frac{\phi}{f_a})\mathcal{Q}} \right|, \quad (2.71)$$

$$= e^{-V V_{\text{eff}}(\phi = -f_a \theta)}, \quad (2.72)$$

and thus $\langle \theta + \phi_a/f_a \rangle = 0$, where V the 4-dimensional volume. Note that the effective potential is periodic. Shifting the axion field, we can define $\theta + \phi/f_a \rightarrow \theta_a$, with θ_a the axion angle. Thus, the PQ mechanism effectively trades θ , a free parameter, for a dynamical field that evolves to its CP-conserving minimum.

We will study the axion mass in great depth in chapter 5 and the cosmological implications of the axion effective potential will be discussed in more detail in chapter 7.

Chapter 3

The IILM at zero temperature

In the strong coupling regime, the degrees of freedom appearing in the fundamental Lagrangian are not manageable in a perturbative treatment. One strategy is to reparametrise the theory in terms of degrees of freedom that do allow a systematic perturbative expansion, e.g. chiral perturbation theory, describing the low energy regime of QCD in terms of baryon and meson fields and based on the symmetries of the QCD Lagrangian; or exact dualities, as in the work of Seiberg and Witten. A different approach is to identify non-perturbative field configurations of the fundamental degrees of freedom. It provides genuine strong coupling information; a priori, though, it is plagued by the same shortcomings as normal perturbation theory in that higher order corrections are not under control. A posteriori, the method can provide a systematic expansion if the solitonic degrees of freedom dynamically generate a scale at which the coupling is sufficiently small to allow for a systematic perturbative treatment, the semi-classical expansion.

First written down in [22], instantons are self-dual solutions to the full Yang-Mills equations of motion and are thought to be the dominant non-perturbative fluctuations in the low energy limit of QCD, e.g. [164, 165, 166] from the phenomenological and [199, 41] from the lattice perspective. In his seminal work [182], 't Hooft computed the quantum fluctuations around an instanton background and this calculation has formed the basis for many studies of the structure of the QCD vacuum, e.g. the pioneering papers [31] and [32] that made strong use of the dilute gas approximation. The latter suffers from an infrared divergence due to the unbounded integral over the

instanton size collective coordinate. In the early works it was cut off by hand. In [61] [62] the IR problem was dealt with ‘dynamically’ by introducing the concept of an interacting instanton ensemble. The authors showed that the most important configurations are not exact multi-instanton solutions but an ensemble of instantons and anti-instantons. To this end they saturated the path integral with such an ensemble, and, using a variational principle, could show that it provides a lower bound for the exact partition function; within a mean field approximation, good agreement to phenomenology and existing lattice data was achieved.

Taking this variational path integral as a starting point, Shuryak investigated what has become known as the interacting instanton liquid model (IILM) [170, 171]. In [157] many bulk properties were computed and seen to be consistent with the available lattice data and phenomenology. Some recent studies [76] [46] corroborate the earlier results that the IILM rather accurately describes the chiral properties of QCD, i.e. that instantons are the dominant degrees of freedom as far as the chiral regime of QCD is concerned; the latter study confirmed the value of the mean instanton size $\bar{\rho}$, a key parameter of the IILM, to be around 1/3 fm, a result anticipated by Shuryak through phenomenological considerations and supported by lattice studies, e.g. [148]. The average instanton size provides an estimate for the scale of typical fluctuations in this model and thus defines the floating cutoff for the IILM to be around [46]

$$\mu \approx \frac{1}{\bar{\rho}} \approx 600 \text{ MeV},$$

and probably justifies the semi-classical expansion¹, i.e. $\alpha_s(\mu) \lesssim 1^2$.

The IILM is based on a superposition of instantons in singular gauge. This gauge is used because the potentials fall off rapidly enough at infinity for the integrals to converge. However, the gauge singularities at the centres of the instantons do lead to divergences for exactly overlapping pairs; in

¹Strictly speaking, in the semi-classical expansion we study the response of the vacuum to a classical background that is a solution to the classical equations of motion; by definition, the quantum fluctuations arise only at second order. The classical configurations we will study have generally a linear response which is, however, neglected so as to suppress quantum-induced forces that would otherwise distort the background. From an operational point of view then, we are performing a semi-classical computation.

²Note that in multi-loop computations the natural, and smaller, expansion parameter is α_s/π

practice the integrals blow up as the instantons approach each other and overlap ever more strongly, and this is interpreted as a repulsive interaction. Such a repulsive interaction is indeed needed to separate the topological from the perturbative fluctuations, which have already been integrated out.

As mentioned above, the IILM works well in describing the chiral properties of QCD, but it fails to reproduce confinement. The failure to do so might be traced back to the fact that in the singular gauge ensemble the instantons do not overlap strongly. To increase this overlap, a regular instanton ensemble has been proposed that does lead to confinement [136]. Naively, such an ansatz will lead to an infrared divergent action because of the slow fall-off of regular gauge instantons. In the above mentioned paper the instantons are placed in a finite box which also serves as cutoff for spacetime integrals. This regularises the infrared divergence, and it was established that a well-defined thermodynamic limit exists³. In the end we want to go beyond zero temperature and do not know how to extend the regular gauge ensemble to finite temperature; therefore we will stick to the singular gauge ensemble.

The topological susceptibility is a key parameter of QCD and has been investigated in many lattice studies. Comparatively few studies have addressed this quantity within the IILM [173]. One reason is that, so far, the IILM is based on a canonical ensemble. Although one can extract the topological susceptibility from the canonical ensemble through the decay of correlators, e.g. see [173] for IILM and [37, 38] for lattice simulations, it is most natural to use the grand canonical ensemble to study the topological susceptibility. Recent investigations of the IILM in the grand canonical ensemble [66, 65] were based on canonical simulations and a fugacity expansion, while we will set up a grand canonical IILM that uses grand canonical Monte Carlo simulations; see also [55] for a ‘mean-field’ study of the grand-canonical ensemble. In full QCD the topological susceptibility is a chiral quantity, and therefore we might expect it to be well modelled in the IILM; the rather large amount of available data makes the former a good candidate to fix units within the IILM. In this chapter we will use the topological susceptibility together with results from chiral perturbation theory to choose physical quark masses and set the scale μ . Another advantage of the grand canonical ensemble is that

³Screening of the long-range interactions is thought to cure the infrared problem.

it automatically reaches the equilibrium density whereas in the canonical ensemble the density has to be inferred from the minimum of the free energy.

We found it necessary to re-derive the instanton interaction because, with the original interactions for the ratio ansatz at finite temperature [172], the instanton ensemble does not possess a thermodynamic limit. We believe the cause to be a deficiency of the fitting formulas for the interactions. Specifically the instanton–instanton interaction, Eq. (3.11) in [172], contains a term that decays very slowly with instanton separation R ,

$$\ln \left(1 + \frac{\beta}{R} \right) \xrightarrow{R \rightarrow \infty} \frac{\beta}{R}, \quad (3.1)$$

and is not integrable. In their original paper the authors do discuss this long-range interaction and point out that they found the $O(1/R)$ dyon–dyon behaviour for a wide range of intermediate separations. It might well be that the interactions are still well described by this ansatz for the simulation boxes used in subsequent numerical investigations, e.g. [157], but for studying the large volume behaviour it is not appropriate. Instead of improving the fitting formulas, we decided to use direct numerical integration⁴. Of course it is not feasible to integrate during the actual simulations because multidimensional integrals are rather time consuming to compute. Instead our strategy consists of precomputing the interactions on a grid and then use look-up tables during the simulations. We believe that this scheme might also be successfully implemented for more complicated backgrounds, and eventually for an ensemble of non-trivial holonomy calorons [119, 121, 120], for which good fitting formulas will be even harder to come by.

In section 3.1 we will review the standard strategy used to write down the partition function for an ensemble of background gauge fields, i.e. the semi-classical approximation. This is rather similar in spirit to the lattice formulation of [196]. It serves the purpose to highlight which simplifying assumptions are eventually made for the actual simulations and hints at how we can go beyond these. We will then re-derive the interactions for the so-called ratio ansatz, used to construct multi-instanton backgrounds from individual instantons, in section 3.2 and compare it with other available

⁴The difficulty of getting good analytical formulas at finite temperature is already pointed out in [172].

ansätze. In section 3.3 we present the numerical framework we have set up to deal with the simulations. Given that different ansätze are available, we will study their effect on some bulk properties in section 3.4 and we endeavour to get a handle on systematic uncertainties inherent in this approach. Finally we fix the free parameters of the model and summarise our results in section 3.5.

3.1 Saturating the path integral

The method presented here is a combination of the variational and the semi-classical approach. It is not limited to QCD or instantons. Rather, we want to consider here the general computational strategy of the method, paying special attention to the low frequency part of the fluctuation spectrum; the reason is, of course, that only the low energy part is strongly sensitive to the non-trivial classical background. Details pertaining to the variational approach, gauge fixing and renormalisation can be found in the original papers [61, 62].

Depending on the theory we are interested in, we choose an appropriate field configuration, not necessarily a solution to the classical equations of motion, that is expected to play a dominant role in the phenomena under investigation. In QCD, for instance, different field configurations have been proposed to explain confinement, instantons, merons or BPS monopoles and one can, in principle, set up an IILM type path integral for any such background.

We approximate the fundamental path integral by saturating the partition function with the chosen classical background and we compute the quantum fluctuations around it:

$$Z[J] = \int [d\phi] \exp(-S[\phi_c + \phi]). \quad (3.2)$$

Since we can only perform Gaussian integrals, the computational rationale is the same as in ordinary perturbation theory, i.e. after expanding around ϕ_c , the action consists of a quadratic term, from which the propagator is derived, plus higher order terms that are treated as perturbations. The

technical complication, in this case, is that the kernel of the quadratic term is a non-trivial function of space, and, therefore, the functional determinant and the propagator are very hard to come by analytically and in general will not be available. For the class of backgrounds that are superpositions of classical solutions, for which analytic results exist, progress can be made without too much numerical effort as we will now discuss.

In general the classical background possesses (quasi) zero modes which render the path integral ill-defined if treated in the Gaussian approximation. The method of collective coordinates trades the zero mode expansion coefficients for the coordinates of the (quasi) moduli-space. The collective coordinates can be interpreted as those degrees of freedom whose quantum mechanics (on this non-trivial space) approximates the low-energy dynamics of the fundamental theory.

In what follows we will use the practical approach of discarding the linear term $\delta S[\phi]/\delta\phi|_{\phi=\phi_c}$ because it would induce a shift in the background field. The reason it appears is linked to the fact that the background is in general not a solution to the equations of motion. There exist systematic methods to cope with this issue, like the streamline [16] and valley [10] techniques, however, they have not been used for instanton–anti-instanton pairs at finite temperature and are therefore not directly applicable to our case of interest. Within this approximation the ‘free’ generating functional, coupled to an external classical source J , is given by

$$\begin{aligned} Z[J] &= e^{-S_c} \int_M d^N \xi \sqrt{g} \int [d\phi] \exp \left(-\frac{1}{2} \phi \cdot \frac{\delta^2 S}{\delta \phi^2} \cdot \phi + J \cdot \phi \right), \\ &= e^{-S_c} \int_M d^N \xi \sqrt{g} \int [d\phi'] \exp \left(-\frac{1}{2} \phi' \cdot \frac{\delta^2 S}{\delta \phi^2} \cdot \phi' + \frac{1}{2} J \cdot G \cdot J \right), \quad (3.3) \\ &= e^{-S_c} \int_M d^N \xi \sqrt{g} \left(\det \frac{\delta^2 S}{\delta \phi^2} \right)^{-1/2} \exp \left(\frac{1}{2} J \cdot G \cdot J \right), \end{aligned}$$

where we defined $\phi' = \phi - G \cdot J$. The moduli-space M is of dimension N , and g_{ij} its metric [185], equivalent to the Jacobian induced by the change to the collective coordinates. The classical source J is orthogonal to the tangent space of M , i.e has no overlap with the zero modes of $\delta^2 S/\delta \phi^2$, as is the propagator $G = (\delta^2 S/\delta \phi^2)^{-1}$.

For the sake of simplicity, we assume that the exact solution has one zero mode η_1 related to the collective coordinate γ . We start with the $N = 1$ case; to trade the eigenfunction η_1 for the collective coordinate γ , we write the full field $\phi_c + \phi$ in two equivalent forms⁵, using the eigenfunctions of $\delta^2 S / \delta \phi^2$ at zero and finite γ ,

$$\begin{aligned} \phi_c(x, 0) + \phi(x) &= \phi_c(x, 0) + \sum_{n=1}^{\infty} \zeta_n \eta_n(x, 0), \\ &= \phi_c(x, \gamma) + \sum_{n=2}^{\infty} \bar{\zeta}_n \eta_n(x, \gamma) + O(\gamma^2). \end{aligned} \quad (3.4)$$

This can be rearranged to write (omitting the x -dependence for notational clarity)

$$\begin{aligned} \phi(\{\gamma, \bar{\zeta}\}) &= \phi_c(\gamma) - \phi_c(0) + \sum_{n=2}^{\infty} \bar{\zeta}_n \eta_n(\gamma), \\ &= \sum_{m=1}^{\infty} \left[\int d^n x \left(\phi_c(\gamma) - \phi_c(0) + \sum_{n=2}^{\infty} \bar{\zeta}_n \eta_n(\gamma) \right) \eta_m(0) \right] \eta_m(0), \end{aligned} \quad (3.5)$$

where we used the fact that $\eta(x, 0)$ forms a complete basis. To compute the Jacobian for the variable change $\{\zeta_n\} \rightarrow \{\gamma, \bar{\zeta}_m\}$, we need the following partial derivatives

$$\frac{\partial \zeta_n}{\partial \gamma} = \int \left(\partial_\gamma \phi_c(\gamma) + \sum_{m=2}^{\infty} \bar{\zeta}_m \partial_\gamma \eta_m(\gamma) \right) \eta_n(0), \quad (3.6)$$

$$\frac{\partial \zeta_n}{\partial \bar{\zeta}_m} = \int \eta_m(\gamma) \eta_n(0), \quad (3.7)$$

evaluated at $\gamma = 0$. This gives

$$\left. \frac{\partial \zeta_n}{\partial \gamma} \right|_{\gamma=0} = \int \left(\partial_\gamma \phi_c(0) \eta_n(0) - \phi(\bar{\zeta}) \partial_\gamma \eta_n(0) \right), \quad (3.8)$$

$$\left. \frac{\partial \zeta_n}{\partial \bar{\zeta}_m} \right|_{\gamma=0} = \delta_{mn}, \quad (3.9)$$

⁵This follows [116].

where in the first line we used $\int \partial_\gamma \eta_m \eta_n = -\int \eta_m \partial_\gamma \eta_n$ and substituted $\phi(\bar{\zeta}) = \phi(\{\gamma = 0, \bar{\zeta}\}) = \sum_{m=2}^{\infty} \bar{\zeta}_m \partial_\gamma \eta_m(\gamma)$. Clearly $\phi(\bar{\zeta}) \perp \eta_1$. Note the occurrence of the ϕ part. This will lead to new interactions which have no classical counterpart but are purely quantum mechanical; to 1-loop order, we are allowed to discard it and (3.6) simplifies to

$$\left. \frac{\partial \zeta_n}{\partial \gamma} \right|_{\gamma=0} = \int \partial_\gamma \phi_c(0) \eta_n(0). \quad (3.10)$$

The matrix of derivatives has the following structure

$$\begin{pmatrix} \int \partial_\gamma \phi_c \eta_1 & 0 & 0 & \cdots \\ \int \partial_\gamma \phi_c \eta_2 & 1 & 0 & \cdots \\ \int \partial_\gamma \phi_c \eta_3 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3.11)$$

The Jacobian thus gives the well known result $J = \int \partial_\gamma \phi_c \eta_1$. In the above situation, we actually have the further simplification that $\int \partial_\gamma \phi_c \eta_n = 0, n > 1$; however, for the Jacobian this is irrelevant.

We now turn to the case of arbitrary N . Even though we can no longer be sure that the background has N exact zero modes, the N lowest lying modes are quasi-zero modes, i.e. they turn into exact zero modes in the dilute limit and are related to the collective coordinates γ_i . We get the following matrix, where we already dropped the higher order terms,

$$\begin{pmatrix} \int \partial_{\gamma_1} \phi_c \eta_1 & \cdots & \int \partial_{\gamma_N} \phi_c \eta_1 & 0 & \cdots \\ \int \partial_{\gamma_1} \phi_c \eta_2 & \cdots & \int \partial_{\gamma_N} \phi_c \eta_2 & 0 & \cdots \\ \int \partial_{\gamma_1} \phi_c \eta_3 & \cdots & \int \partial_{\gamma_N} \phi_c \eta_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ \int \partial_{\gamma_1} \phi_c \eta_N & \cdots & \int \partial_{\gamma_N} \phi_c \eta_N & 0 & \cdots \\ \vdots & \vdots & \vdots & 1 & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots \end{pmatrix}. \quad (3.12)$$

From the matrix structure, the Jacobian is given by $\det(\int \partial_{\gamma_n} \phi_c \eta_m)$.

Now, we do not know the set $\{\eta\}$ of exact low lying eigenfunctions. However we can approximate it by constructing an orthonormal set of the known

single particle zero modes that descend from the exact solutions used to build up the background field. With a slight abuse of notation, we substitute $\eta \rightarrow \bar{\eta} = O_B \eta$; O_B is the matrix that generates an orthonormal basis from the original set $\{\eta\}$ of single particle zero modes. The Jacobian is then given by

$$\det \left(\int \partial_{\gamma_n} \phi_c \bar{\eta}_m \right) = \det \left(\int \partial_{\gamma_n} \phi_c \eta_m \right) \det O_B. \quad (3.13)$$

The high-frequency eigenvalues, encoded in the determinant of the fluctuation operator $\delta^2 S / \delta \phi^2$ are assumed to be N -fold degenerate, and so the fluctuation determinant factorises.

In QCD we also need to introduce quarks and treat their interactions with the background field. In the case where the Dirac operator admits quasi-zero modes we can approximate the low-frequency part in the same way as for the gluonic case; the high-frequency fluctuations will again be assumed to factorise. As for the Jacobian, we do not know the exact set of low-lying eigenfunctions for the superposition, but we approximate it by constructing an orthonormal set of the exact single particle zero modes ξ_n , i.e. $\bar{\xi} = O_F \xi$. The Dirac operator, truncated to that subspace, is then given by

$$(\bar{\mathcal{D}} + m)_{\text{low}} = \overline{(\bar{\mathcal{D}} + m)}_{ij} |\bar{\xi}_i\rangle \otimes \langle \bar{\xi}_j| = \left(\bar{\mathcal{D}}_{ij} + m \delta_{ij} \right) |\bar{\xi}_i\rangle \otimes \langle \bar{\xi}_j|, \quad (3.14)$$

with $\bar{\mathcal{D}}_{ij} = \langle \bar{\xi}_i | \bar{\mathcal{D}} | \bar{\xi}_j \rangle$. The matrix of overlaps is related to the single particle zero mode overlaps by

$$\bar{\mathcal{D}}_{\text{low}} = \bar{\mathcal{D}} = O_F^\dagger \mathcal{D} O_F, \quad (3.15)$$

with $\mathcal{D}_{ij} = \langle \xi_i | \mathcal{D} | \xi_j \rangle$. Note that this is not a similarity transformation because O_F is not unitary.

3.2 Interactions in the IILM

We will now turn to instantons in QCD. In this chapter, we will only discuss BPST instantons [22]. In terms of the 't Hooft potential⁶

$$\Pi(x, \{y, \rho\}) = \frac{\rho^2}{r^2}, \quad (3.16)$$

⁶Actually $1 + \Pi$ is the 't Hooft potential.

with $r^2 = (x - y)^2$, the BPST instanton in singular gauge is given by

$$A_\mu^a = -O^{ab}\zeta_{\mu\nu}^b \frac{\partial_\nu \Pi(x, \{y, \rho\})}{1 + \Pi(x, \{y, \rho\})}, \quad (3.17)$$

with $\zeta_{\mu\nu}^b = \bar{\eta}_{\mu\nu}^b$ for instantons, $\zeta_{\mu\nu}^b = \eta_{\mu\nu}^b$ for anti-instantons and η the 't Hooft symbols. The collective coordinates are: y the centre, ρ the size and O the colour orientation in the adjoint representation.

The simplest background configuration is the sum ansatz, as used for instance in [61]. It was shown in [168] that the sum ansatz produces an unphysical amount of repulsion; this is due to the fact that the field strength actually diverges at the individual centres, and is in sharp contrast to the individual singular gauge instanton whose field strength is finite at the centre⁷. In this case, the author therefore proposed a different ansatz, inspired by 't Hooft's multi-instanton form, that stays finite at the centre of the instantons, and dubbed it the ratio ansatz. It is given by

$$A_\mu^a = -\frac{\sum_i O_i^{ab}\zeta_{\mu\nu}^b \partial_\nu \Pi_i(x, \{y_i, \rho_i\})}{1 + \sum_i \Pi_i(x, \{y_i, \rho_i\})}. \quad (3.18)$$

In what follows we will refer to R_E as the interactions or the ensemble generated by the ratio ansatz. We will compare the predictions from R_E with those of the streamline ansatz S [190] and another 'hybrid' ratio-sum ansatz R_H [157]. This is summarised in Table 3.1.

Phenomenological considerations have lead to the conclusion that the QCD vacuum consists of a dilute ensemble of instantons; a fact corroborated by lattice studies and self-consistency checks within the IILM. Diluteness and the localised nature of instantons render negligible contributions other than

⁷Note, however, that the field strength of the individual singular gauge instanton is not continuous at the centre and is only defined on the punctured Euclidean space. Incidentally, the winding about this singular point corresponds to the winding at infinity of the regular instanton.

Ansatz	Description
R_E	Interactions for ratio ansatz as derived in this paper.
R_H	Gluonic interactions are derived from the ratio ansatz whereas the quark overlaps use a sum ansatz, [157].
S	Interactions have been derived from the so-called streamline ansatz. These are only available at zero temperature, [157] [190].

Table 3.1: Several ansätze for the classical background field have been proposed. The following table summarises what they will be referred to throughout the rest of this thesis.

two-body interactions⁸, given here for an instanton–anti-instanton pair,

$$A_\mu^a = -\frac{\bar{\eta}_{\mu\nu}^a \partial_\nu \Pi_1(x, \{x_1, \rho_1\}) + O^{ab} \eta_{\mu\nu}^b \partial_\nu \Pi_2(x, \{x_2, \rho_2\})}{1 + \Pi_1(x, \{x_1, \rho_1\}) + \Pi_2(x, \{x_2, \rho_2\})}, \quad (3.19)$$

with $O = O_1^t O_2$. The formulas for like-charged pairs follow trivially from the above.

3.2.1 Gluonic interactions

The complete classical gluonic interaction is given by the sum over all the possible pairings. It is clear from the structure of (3.19) that the colour degrees of freedom can be completely factorised out. After some lengthy algebra the result for the squared field strength can be written in the form

$$\begin{aligned} F_{\mu\nu}^a F_{\mu\nu}^a &= I + (\text{Tr } O^t O + (\bar{\eta} O \eta)_{\mu\nu\mu\nu}) J + (\bar{\eta} O \eta)_{\rho\mu\rho\nu} I_{\mu\nu} \\ &+ (\bar{\eta} O \eta)_{\mu\rho\nu\sigma} I_{\mu\rho\nu\sigma} + (\eta O^t O \eta)_{\mu\rho\nu\sigma} J_{\mu\rho\nu\sigma} + (\bar{\eta} O \eta)_{\alpha\mu\alpha\rho} (\bar{\eta} O \eta)_{\beta\nu\beta\sigma} K_{\mu\rho\nu\sigma}. \end{aligned} \quad (3.20)$$

The different contributions are given in appendix A.1.

Factorising out the single instanton contributions and the coupling con-

⁸A note on terminology: whenever we use the word interaction, we mean a quantity ‘normalised’ to the dilute gas, i.e. we subtract the dilute gas counterpart if the term naturally occurs in the exponential, as in the classical gluonic interactions, or we divide by the dilute gas counterpart if the interaction is a pre-exponential factor, as in the gluonic Jacobian or the Dirac determinant.

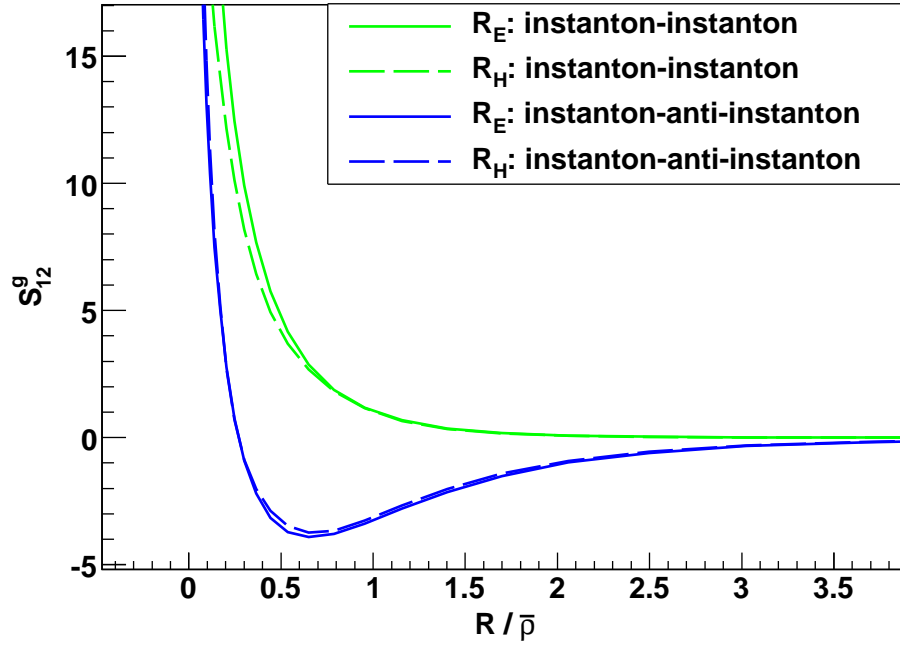


Figure 3.1: For instantons with equal sizes the interaction of R_E agrees very well with R_H for oppositely charged instantons. There is a slight discrepancy for like-charged instantons, in that the repulsion is a bit steeper in the R_E case. (We have set $\bar{\rho} = \sqrt{\rho_1^2 + \rho_2^2}$.)

stant, the classical interaction between instantons is given by

$$S_{12}^g/S_0 \equiv V_{12} \equiv (S[A]/S_0 - 2), \quad (3.21)$$

$$S[A] = \frac{1}{4g^2} \int F_{\mu\nu}^a F_{\mu\nu}^a, \quad (3.22)$$

where $S[A]$ is the action of the background gauge fields and $S_0 = 8\pi/g^2$ that of a single instanton.

The functional dependence on the size parameters is of the form $\sqrt{\rho_1\rho_2}$ in R_H , whereas in the R_E ansatz the sizes enter in the combination $\sqrt{\rho_1^2 + \rho_2^2}$ ⁹; this is in agreement with [61]¹⁰. For equal sizes both ansätze lead to the same $\sqrt{\rho_1\rho_2}$ behaviour, so that we will use this regime to validate our results.

For equal sizes the agreement with R_H is very good, see Fig. 3.1. However, for unequal sizes there are noticeable differences, see Fig. 3.2. The discrepancy follows from the different functional dependence on the size parameters.

We will now comment on the assumptions that have so far been made in the practical implementation of the IILM and that we will adopt in this study too. These are, on the one hand, the approximation of the high-frequency quantum interaction by an inverse running coupling constant evaluated at the scale of the mean instanton size and, on the other hand, the neglect of the Jacobian that introduces the collective coordinates and represents the low-frequency quantum interaction.

In the single instanton case the quantum fluctuations lead to charge renormalisation and the coupling constant is replaced by the running coupling at the scale given by the instanton size [182], $S_0(\rho) = 8\pi/g^2(\rho)$. The same calculation has never been performed for a pair. The high-frequency part of the fluctuation determinant could in principle be estimated along the lines of [54], as suggested in [61]. This approach seems to fit in well with the strategy used so far, as it would basically consist of deriving expressions like (3.20) for higher order gauge invariant operators; but it is not an exact computation since it neglects the lower part of the high-frequency spectrum. On the analytical side this approach allows us to estimate that the (high-

⁹At least in the parameter regions of large and small separations, see appendix A.1.2.

¹⁰The authors of [61] have considered the sum ansatz; for large separations, however, every ansatz is equivalent to the sum ansatz.

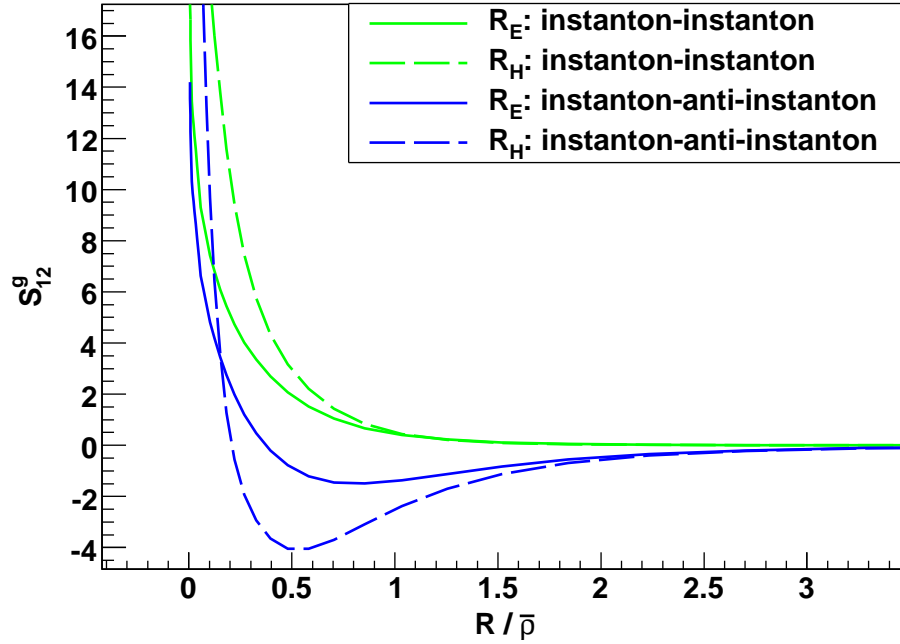


Figure 3.2: For unequal size parameters, e.g. $\rho_1/\rho_2 = 3$ in this case, large differences start to become apparent. The reason is that the dependence on the sizes is more complicated than the functional form $\sqrt{\rho_1\rho_2}$ used in R_H . Note that the attractive well is much deeper in the R_H case which will eventually lead to a denser ensemble. (We have set $\bar{\rho} = \sqrt{\rho_1^2 + \rho_2^2}$.)

frequency) quantum interaction is subdominant to the classical interaction: instantons are localised and thus these higher order operators are subdominant for well separated instantons. The variational approach has indeed shown that the system stabilises in a rather dilute state and that the classical interaction dominates over the quantum interactions [61]. In the latter paper the authors argue then that the quantum interaction can be estimated by modulating the (total) classical interaction with the inverse running coupling constant, a slowly varying function of the background field, at the scale $\bar{\rho}$. We will adopt the parametrisation put forth in [168, 157] that estimates the scale of the running coupling constant on a pair-by-pair basis and uses the geometrical mean of the sizes to set this scale. The full gluonic interaction is then given by

$$S_{12}^g = S_0(\sqrt{\rho_1 \rho_2}) V_{12}. \quad (3.23)$$

The Jacobian (3.12) is positive by definition¹¹. After normalising it to the dilute gas contribution and placing it into the exponential we can therefore interpret it as a (low-frequency) quantum interaction. The normalised Jacobian matrix is of the form $J = \mathbb{I} + \mathcal{O}$, where \mathcal{O} represents the overlap between different instantons and the unit matrix ‘self’-overlaps. For large separations, \mathcal{O} becomes small, and we can expand the determinant. The leading term will be given by $\text{Tr } \mathcal{O}^2$. A rough estimate of the large distance behaviour suggests that the asymptotic power-law decay is $O(1/R^6)$, with R the separation between the pair. This is a faster decay than the well-known dipole–dipole interaction that follows from the classical action. For strong overlaps the matrix will become approximately degenerate, and the determinant small, because the matrix elements of the pair with the other instantons will be roughly equal¹². From the point of view of the effective interaction this might be regarded as a repulsive core. For complete degeneracy the singularity should be logarithmic because one singular value will tend to zero as the rank of the matrix decreases by one. The repulsion will thus be of the form

$$\ln J_{12}^{\text{sing}} \sim \ln \left(\frac{R^2}{\rho_1^2 + \rho_2^2} \right), \quad (3.24)$$

¹¹It is a mere variable transformation.

¹²We assume that, other than the centres, the sizes and colour orientations of the pair are approximately equal.

with proportionality factor of order $O(1) - O(10)$ because, as we argued, the degeneracy is due to one overlapping pair and should not get contributions from other instantons. In (3.3.1) we will discuss the small separation asymptotic behaviour for the ratio ansatz; the analytical expressions are given in appendix A.1.2, and we note that the singular behaviour is also repulsive and logarithmic,

$$I_{IA}^{\text{sing}} \sim \ln \left(1 + \frac{\rho_I^2 + \rho_A^2}{R^2} \right), \quad (3.25)$$

with a proportionality factor that is again of order $O(1) - O(10)$. In the intermediate region it is harder to estimate the Jacobian interaction, but the logarithm should make its contribution subdominant. Also, the classical interaction is boosted by the quantum contribution through charge renormalisation. We conclude that the Jacobian interaction is probably negligible compared to the classical interactions. Note that within the moduli-space approximation it should be incorporated for consistency; it is straightforward to compute and will be included in the future. In this work, however, we will focus solely on the classical interaction.

Finally, note that for superpositions that do not lead to a repulsion classically, as for instance the streamline ansatz S , a repulsive interaction has been added by hand [157]. The need for such a term follows from the fact that strongly overlapping instanton–anti-instanton pairs are indistinguishable from perturbative fluctuations, since their topological charges have practically cancelled each other off; but perturbative modes have already been integrated out and should no longer be present in the IILM. The repulsion then allows for a separation between instantons and perturbative fluctuations by introducing an obstruction for strongly overlapping pairs to form. The IILM might provide its own mechanism to achieve this through the Jacobian and render unnecessary the need to add a classical repulsive core by hand.

3.2.2 Quark Interactions

The quark interaction arises from (3.14), as is clear from our discussion in section 3.1, and is purely quantum mechanical. As for the gluonic interaction, some further approximations have been used in the literature; we will adopt these, albeit rephrased slightly differently.

We will assume that the single instanton zero modes $\{\xi\}$ form a functional orthonormal basis, i.e. we neglect contributions arising from non-vanishing overlaps among the ξ_i . With this in mind, the finite dimensional low-frequency Dirac operator is then given by

$$(\mathcal{D} + m)_{ij} = \langle \xi_i | \mathcal{D} + m | \xi_j \rangle = \mathcal{D}_{ij} + m\delta_{ij}. \quad (3.26)$$

To reiterate, we attribute the diagonal mass term to the requirement of orthonormality¹³ rather than the degree of dilution of the instanton ensemble, e.g. [139]. On the practical level this is irrelevant in as far as we recover the same determinantal interaction as used in previous works.

The quark zero mode, in singular gauge and in the chiral representation, is given by [84]

$$\xi_I = \frac{1}{2\pi\rho_I} \sqrt{1 + \Pi_I} \not{D} \frac{\Pi_I}{1 + \Pi_I} \begin{pmatrix} U_I \varphi \\ 0 \end{pmatrix}, \quad (3.27)$$

$$\xi_A = \frac{1}{2\pi\rho_A} \sqrt{1 + \Pi_A} \not{D} \frac{\Pi_A}{1 + \Pi_A} \begin{pmatrix} 0 \\ U_A \varphi \end{pmatrix}, \quad (3.28)$$

with $\varphi_{\alpha a} = \epsilon_{\alpha a}$, normalised according to $\epsilon_{12} = 1$. Finally, U_i is the 3×3 colour matrix describing the collective coordinates for the colour embedding; it is related to the adjoint representation by $O^{ab} = 1/2 \text{Tr}(U \tau^a U^\dagger \tau^b)$, with $U = U_I^\dagger U_A$.

The Dirac operator, as defined above, is anti-hermitian. Eventually we need to diagonalise it, but, since readily available routines work with hermitian matrices, we display here the matrix elements of $i\not{D}$. Within the ratio ansatz, the matrix elements $T_{IA} = \int \xi_I^\dagger i\gamma_\mu \not{D}_\mu \xi_A$ are as follows

$$T_{IA} = \int d^4x \frac{1}{4\pi^2 \rho_I \rho_A} \frac{1}{2} \text{Tr}(U \tau_\beta^+) I_\beta. \quad (3.29)$$

The concrete realisation of I_β is given in appendix A.2.

The rather large difference between R_E and R_H , see Fig. 3.3, is due to the fact that the latter use a sum ansatz. The ratio ansatz was introduced

¹³Writing $H = H_{ij} |\psi_i\rangle \otimes \langle \psi_j|$ makes only sense if $\{\psi_i\}$ forms an orthonormal system, given the scalar product $\langle \cdot | \cdot \rangle$.

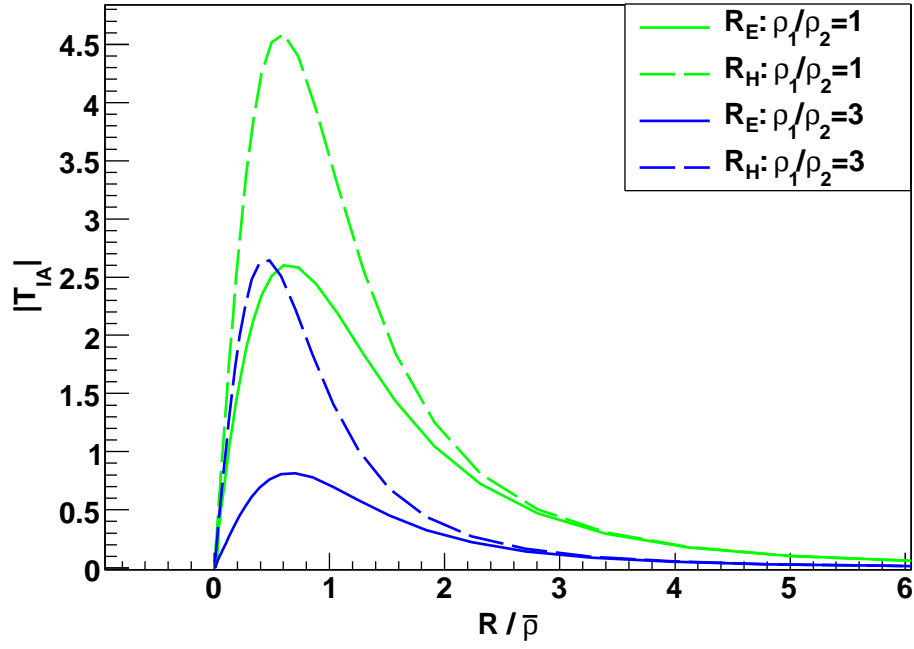


Figure 3.3: The relatively large discrepancy is due to the fact that R_E uses the full ratio ansatz in the Dirac operator whereas R_H uses the sum ansatz. (We have set $\bar{\rho} = \sqrt{\rho_1^2 + \rho_2^2}$.)

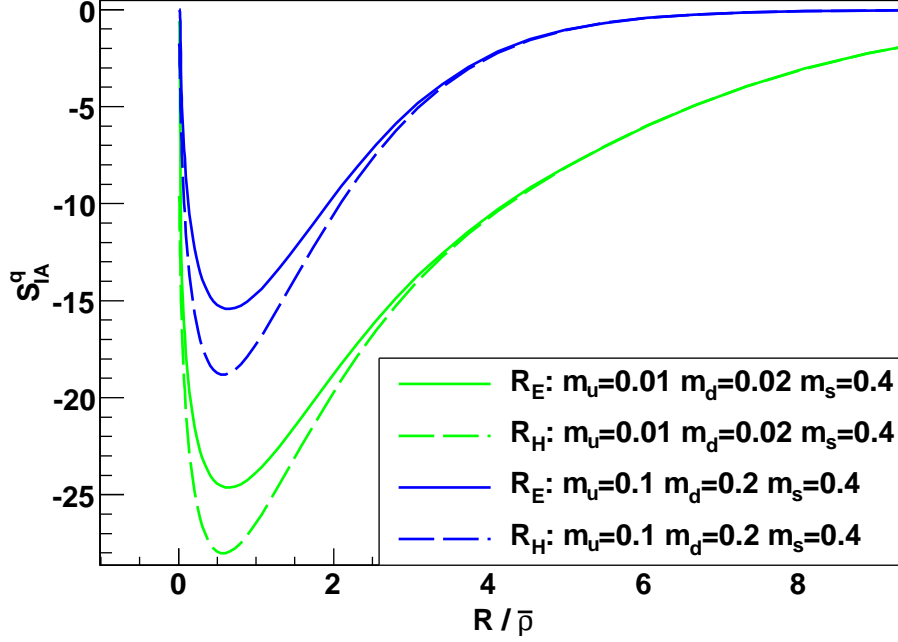


Figure 3.4: On the level of the effective interaction, the difference between R_E and R_H is not as pronounced, i.e. the relative difference has decreased substantially. We can clearly see that light quark masses lead to a stronger attractive interaction between instantons and anti-instantons. Note that the relative difference between the ansätze R_E and R_H does not seem to depend strongly on the quark masses. The instantons have been set up with equal sizes. (We have set $\bar{\rho} = \sqrt{\rho_1^2 + \rho_2^2}$.)

to remove the unphysical divergence in the field strength; no such problem afflicts the overlap matrix elements. On top of that the quark determinant is a pre-exponential factor and as an effective interaction the extra logarithmic factor should make it rather insensitive to its exact form, see [172]. Within our numerical framework, the full ratio ansatz does not produce any additional overhead and has the merit to be more consistent with the gluonic interactions. We have checked that upon neglect of the contributions special to the ratio ansatz, i.e. simplifying the overlaps so as to recover the sum ansatz, our results agree very well with those of R_H , apart from the aforementioned discrepancy in the instanton size parametrisation. As for the gluonic interactions, the colour matrices could again be completely factorised out.

Note that the Dirac operator only connects instantons to anti-instantons due to the extra γ -matrix factor as compared to the mass operator, which vanishes between instantons and anti-instantons. Therefore the quark fluctuation operator has the following form

$$m\mathbb{I} - i \begin{pmatrix} 0 & T \\ T^\dagger & 0 \end{pmatrix}, \quad (3.30)$$

with T the $N_I \times N_A$ matrix of overlaps T_{IA} , and N_I (N_A) is the number of instantons (anti-instantons); the 0-matrices are $N_I \times N_I$ and $N_A \times N_A$ dimensional, respectively; finally, \mathbb{I} is the identity operator on the quasi-zero mode space of dimension $(N_I + N_A) \times (N_I + N_A)$. To diagonalise $i\mathcal{D}$, it suffices to know the singular-value-decomposition of T . The left and right singular vectors, ψ^L and ψ^R , are defined by

$$T\psi_n^R = \lambda_n\psi_n^L, \quad (3.31)$$

$$T^\dagger\psi_n^L = \lambda_n\psi_n^R. \quad (3.32)$$

The singular eigenvalues λ_n are always positive. The kernel of the Dirac operator is spanned by the $\lambda = 0$ singular eigenvectors, ψ^K , of either T or T^\dagger , depending on whether $N_I < N_A$ or $N_I > N_A$. We can then construct the eigenvalue decomposition of the Dirac operator. The non-zero eigenvalue part has the following eigensystem

$$\left\{ \left[\lambda_n, \begin{pmatrix} \psi_n^L \\ \psi_n^R \end{pmatrix} \right], \left[-\lambda_n, \begin{pmatrix} -\psi_n^L \\ \psi_n^R \end{pmatrix} \right] \middle| n \in \{1, \dots, \min(N_I, N_A)\} \right\}. \quad (3.33)$$

Finally, the kernel is spanned by the eigensystem

$$\left\{ \left\{ \left[0, \begin{pmatrix} \psi_n^K \\ 0 \end{pmatrix} \right] \middle| n \in \{1, \dots, N_A - N_I\} \right\}, \quad N_I < N_A, \right. \\ \left. \left\{ \left[0, \begin{pmatrix} 0 \\ \psi_n^K \end{pmatrix} \right] \middle| n \in \{1, \dots, N_I - N_A\} \right\}, \quad N_I > N_A. \right\} \quad (3.34)$$

Note that the non-zero eigenvalues come in pairs. Together with the zero

eigenvalues, the determinant of the Dirac operator can be written as

$$\det(i\mathcal{D}) = m^{|Q|} \prod_n^{\min(N_I, N_A)} (m^2 + \lambda_n^2), \quad (3.35)$$

with $Q = N_I - N_A$ the topological charge. If we are only interested in the determinant, and not so much in the eigensystem, this can be put in the equivalent form

$$m^{|Q|} \begin{cases} \det(TT^\dagger + m^2), & Q < 0 \\ \det(T^\dagger T + m^2), & Q > 0. \end{cases} \quad (3.36)$$

Upon placing this term into the exponential, the normalised determinant of quark zero mode overlaps leads to an effective interaction. The normalisation consists of dividing (3.36) by $m^{N_I+N_A}$. The quark interaction is thus given by

$$S_{N_f}^q = - \sum_{n=1}^{N_f} \begin{cases} \ln \det(TT^\dagger + m_n^2) - N_I \ln m_n^2, & Q < 0 \\ \ln \det(T^\dagger T + m_n^2) - N_A \ln m_n^2, & Q > 0 \end{cases}, \quad (3.37)$$

with N_f the number of active quark flavors. Note that the quark interaction is always attractive. This follows from the fact that we can write the overlap matrix for each flavour as $\mathbb{I} + \frac{T^2}{m_n^2}$, and this form makes it explicit that the determinant is bounded from below by unity because the smallest eigenvalue is easily seen to satisfy $\lambda_{\min} \geq 1$.

This exhausts the interactions in the IILM because the fluctuation operator of the ghost part is positive definite and its lack of zero modes prevents the construction of the low frequency part of the spectrum within the moduli-space approximation. We are thus left with the high-frequency part which, as in the other cases, is assumed to factorise and cannot lead to interactions.

3.3 Numerical Implementation

3.3.1 Interpolation and asymptotic matching

The decoupling of the colour degrees of freedom is a computational benefit: by using global $SO(4)$ transformations, without loss of generality, we place

the first instanton at the origin and the second along the z -direction. The initial orientational dependence is then factored out of the integrand and combines with the colour matrices as in [61]. These integrations are too time consuming to perform during actual simulations; instead, they are computed beforehand to fill interpolation tables that are, in turn, used during the simulations. The interpolation grid is three-dimensional, and depends on ρ_1 , ρ_2 and $R = |x_1 - x_2|$. For numerical stability we choose to use simple linear interpolation.

A uniform grid can, of course, only extend over a finite region and we must decide which portion of the parameter space to cover. We took the single instanton moduli-space measure as a guide for the size grid because, suitably normalised, it can be interpreted as a probability density. We choose the lower limit, $\rho_{\min} \approx \frac{2}{30}\Lambda^{-1}$, to be a fairly small quantile¹⁴. Here, Λ is the scale at which QCD starts to become strongly coupled. The upper limit is set to $\rho_{\max} = \Lambda^{-1}$. Larger instantons cannot be treated consistently in the IILM because it uses perturbation theory, which breaks down below Λ .

We believe that these choices cover the relevant parameter space, and we sample the sizes from the interval $[\rho_{\min}, \rho_{\max}]$. As a consistency check we monitored the actual size distribution and did not find any evidence for a significant weight at the edges of the sample interval. We therefore conclude that this procedure is well-defined.

The classical gluonic interaction in the ratio ansatz suffers from gauge singularities that prevent us from extending the grid down to vanishingly small instanton separations, $R \rightarrow 0$. The opposite limit, $R \rightarrow \infty$, cannot be covered either unless we use a non-uniform measure on \mathbb{R}_+ . In principle this would seem like the most elegant approach, however, it is not feasible practically because the numerical integration becomes inaccurate at larger separations; the only remedy would be to set very small error tolerances for the numerical integrations, but that is computationally prohibitive. Therefore, we decided to use matching formulas for both the large and small separation regimes.

The rationale is not to derive accurate formulas in absolute terms but to get the absolute value from the interpolation results at a matching point R_m .

¹⁴It corresponds to less than the millionth quantile.

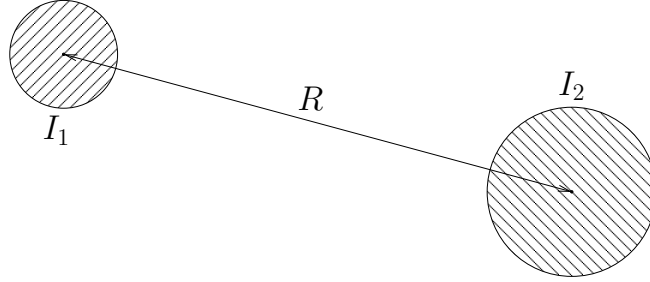


Figure 3.5: The instantons I_1 and I_2 are so far apart that, within the shaded region that give the dominant contribution to the field strength of each, the partner's field strength is roughly constant and fixed at $x_\mu - R_\mu \approx -R_\mu$. We can then safely extend the integration region to be all of \mathbb{R}^4 , with a negligible error due to the rather strong localisation of the individual instantons.

The matching formulas are thus to be understood as accurate in a relative sense, i.e. the asymptotic interactions f_{asy} should behave, asymptotically, like the exact numerical interactions f_{ex} . This ensures that we reproduce the correct fall-off or singularity behaviour. Thus, we compute the interactions according to

$$f(R) = f_{\text{asy}}(R) \frac{f_{\text{ex}}(R_m)}{f_{\text{asy}}(R_m)}, \quad (3.38)$$

whenever they fall out of the grid. Since the localisation of the instantons is set by the sizes, it is natural for the matching point to be proportional to the former. Eventually, the exact proportionality factor follows from an ‘optimisation’ procedure, given that we aim for the interpolated interactions to be correct at the one percent level.

The full gluonic interaction consists of different pieces that are added together, (3.20). We could use (3.38) for these subinteractions term by term but it turns out that such a matching is numerically rather unstable. Thus, even though we are only interested in asymptotic relations, we need a systematic procedure that insures that the different asymptotic subinteractions are added up with the correct magnitude relative to each other.

For the large separation case we want the instantons to be so far apart from each other that within the region in which the field strength for I_1 is strong the field strength of I_2 hardly changes: we can approximate $x_\mu - R_\mu \approx$

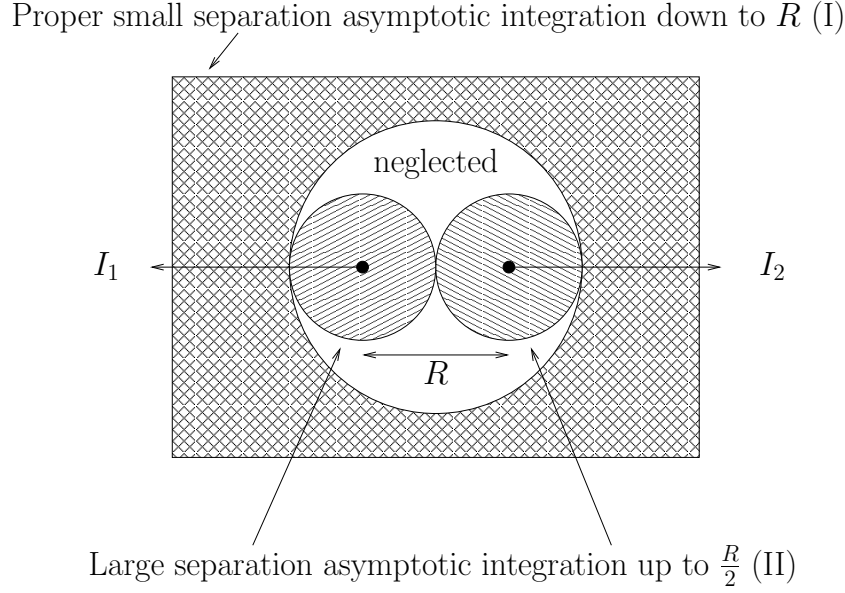


Figure 3.6: The instantons I_1 and I_2 are strongly overlapping. We approximate the integral by, first, integrating over I_1 keeping I_2 fixed at R_μ , as in the large separation case but with upper limit $R/2$; to this we add the analogous contribution from I_2 . Secondly, the, possibly, singular behaviour is picked up by integrating from infinity down to R , and approximating the arguments to be $x_\mu - R_\mu/2 \approx x_\mu$ and $x_\mu + R_\mu/2 \approx x_\mu$, respectively.

$-R_\mu$ ¹⁵. Since the field strength is negligible at and beyond R_μ , we can safely extend the integration region to cover all of \mathbb{R}^4 . The field strength of I_2 behaves as a constant, and we can use the rather simple rational expression for the interaction in terms of the 't Hooft potential to find exact results. We add to this the analogous contribution from $I_1 \leftrightarrow I_2$. The configuration is illustrated in Fig. 3.5.

We shall call this the zeroth order approximation, and it is clear that, to this order, terms odd in derivatives of I_i will vanish due to $O(4)$ symmetry. However, it turns out numerically that, upon combining all the different terms from (3.20), the zeroth order terms are sufficient. Our formulas are given in appendix A.1.2.

In principle, we can compute the neglected terms by going to first order, i.e. $g(x - R) \approx g(-R) + x_\mu \partial_\mu g(-R)$, or beyond. Such higher order contribu-

¹⁵We use a translation to place I_1 at the origin.

tions will typically no longer converge on \mathbb{R}^4 . It seems natural to cut them off at R , and this will generally lead to logarithms, $\ln(1 + R^2/(\rho_1^2 + \rho_2^2))$, together with rational functions. However, in contrast to the fitting formulas of [157], the Taylor expansion of our asymptotic formulas produce only power-law like decays for large separations; in addition they fall off more strongly than the zeroth order terms.

We now turn to the case of asymptotically small separations. A typical situation is depicted in Fig. 3.6. The rationale is to split the integration into 2 regions.

- I The far-field region beyond both centres, placed symmetrically around the origin; we approximate the arguments by $x_\mu \pm R_\mu/2 \approx \pm x_\mu$.
- II The region around each instanton up to $R/2$, with R the pair separation. We integrate around $\pm R_\mu/2$ keeping the arguments of the partner instanton fixed at $x_\mu \mp R_\mu/2 \approx \mp R_\mu/2$. This is similar to the large separation case, but here we only integrate up to $R/2$.

Region I accounts for possible singularities. After adding up all the different subinteractions, the singularities from region I dominate the total interaction. Since we need the region II approximations anyway in the large separation case, it does not represent any extra overhead to use them as well in the small separation limit.

In Figs. 3.7 and 3.8 we plot the exact and approximate result for the gluonic, quark and total interaction. Note that some subinteractions in the gluonic sector are poorly approximated by the zeroth order asymptotic matching formulas in region I. However, those terms that do exhibit singularities completely dominate, and the total gluonic interaction is well approximated for all separations. The quark overlap consists of just one term, which is not well described in region I. We estimated that the correct asymptotic behaviour can be obtained at second order. However, this is unnecessary since the quark interaction is bounded in that region and the gluonic interaction completely dominates. Combining the gluonic and fermionic interactions, we see that the total pair interaction is accurate on the one percent level.

Note that we use interpolation up to very strong overlaps; also, instantons rarely enter region I during simulations because it makes up a very small part

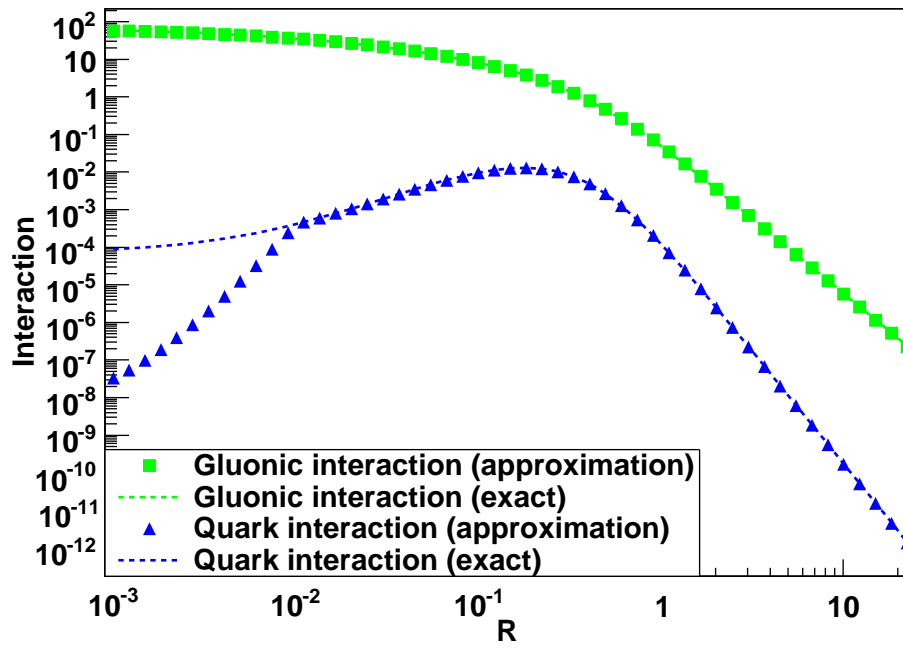


Figure 3.7: The gluonic interaction is well approximated by the combination of interpolation and asymptotic matching. The quark overlap is very poorly approximated by the zeroth order small separation asymptotic formula; it tends to zero with too high a power as compared to the exact result. The correct behaviour can in principle be obtained from higher orders, and we've estimated that the second order contribution will suffice. In practice, the quark interaction in this region is completely irrelevant as compared to the gluonic interaction.

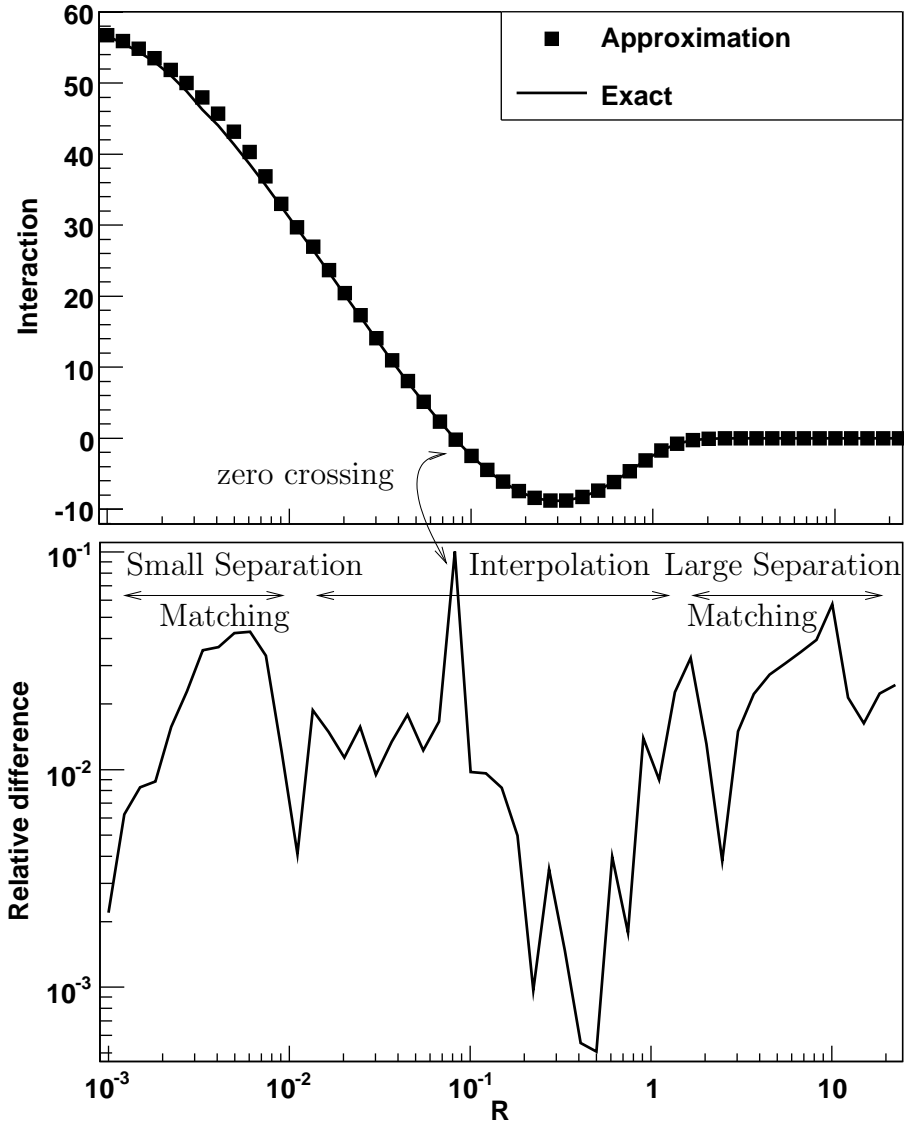


Figure 3.8: The total pair interaction is accurate on the one percent level. Note that the error in region I from the quark interaction is completely negligible. The spike in the bottom plot is due to the sensitivity to zero crossings.

of the total volume box. In any case, region I overlaps will almost certainly be rejected in Monte Carlo moves, and therefore quantities that are computed solely from the quark interaction, as for instance the quark condensate, are very insensitive to large errors in region I. Ultimately, this is the reason why we chose the interpolation grid to cover such strong overlaps.

3.3.2 Monte Carlo

Previous studies, lattice results and phenomenology indicate that the instanton ensemble is fairly dilute. Therefore, we organise the partition function into a dilute gas measure times the exponential of interactions,

$$Z = \sum_{N_I, N_A}^{\infty} \frac{1}{N_I!} \frac{1}{N_A!} \prod_i^{N_I} d(\rho_i) \prod_j^{N_A} d(\rho_j) \exp(-(S^g + S^q)) , \quad (3.39)$$

$$\equiv \sum_{N_I, N_A}^{\infty} \frac{1}{N_I!} \frac{1}{N_A!} Z_{N_I, N_A} , \quad (3.40)$$

$$S^g = \sum_{ij} S_{ij}^g , \quad (3.41)$$

where S_{ij}^g is given in (3.23) and S^q is given in (3.37). We follow [157] and use one-loop accuracy for the charge renormalisation factor that modulates the classical gluonic interaction, i.e. $S_0(\sqrt{\rho_1 \rho_2}) \rightarrow \beta_1(\sqrt{\rho_1 \rho_2})$, with β_1 given below in (3.46). Although not really consistent, the single instanton density is given at two-loop in order to replace the pre-exponential bare by the running coupling constants [61]; the former were induced by the transformation to collective coordinates. The two-loop single instanton measure is then given by

$$d(\rho) = d_0^g(\rho) d_0^q(\rho)^{N_f} , \quad (3.42)$$

$$d_0^g(\rho) = C_{N_c} \rho^{-5} \beta_1(\rho)^{2N_c} \exp \left[-\beta_2(\rho) + \left(2N_c - \frac{b'}{2b} \right) \frac{b'}{2b} \frac{\ln \beta_1(\rho)}{\beta_1(\rho)} \right] , \quad (3.43)$$

$$d_0^q(\rho) = m\rho \exp \left(-\frac{1}{3} \ln m\rho + \frac{\frac{1}{3} \ln m\rho + 2\alpha - (6\alpha + 2\beta)(m\rho)^2 + 2A_1(m\rho)^4 - 2A_2(m\rho)^6}{1 - 3(m\rho)^2 + B_1(m\rho)^4 + B_2(m\rho)^6 + B_3(m\rho)^8} \right) . \quad (3.44)$$

For the quark term, d_0^q , we use the generalisation of 't Hooft's [182] result valid for arbitrary mass [59]. The different terms in d_0^q are given by

$$C_{N_c} = \frac{0.466 e^{-1.679 N_c}}{(N_c - 1)!(N_c - 2)!}, \quad (3.45)$$

$$\beta_1(\rho) = -b \ln(\rho \Lambda), \quad b = \frac{11}{3} N_c - \frac{2}{3} N_f, \quad (3.46)$$

$$\beta_2(\rho) = \beta_1(\rho) + \frac{b'}{2b} \ln \left(\frac{2}{b} \beta_1(\rho) \right), \quad b' = \frac{34}{3} N_c^2 - \frac{13}{3} N_c N_f + \frac{N_f}{N_c}. \quad (3.47)$$

Note that the above has been derived in Pauli-Villars regularisation.

Being an interacting many-body system, the partition function cannot be evaluated analytically, and we choose Monte Carlo methods to cope with it numerically. More precisely, we will use the Metropolis algorithm to sample the important integration regions of the partition function. This is, of course, all well known, but it seems appropriate to introduce the, possibly less known, Monte Carlo moves corresponding to insertion and deletion of instantons needed for grand canonical simulations.

Following the usual strategy of imposing detailed balance, the simplest insertion/deletion algorithm consists of randomly placing an instanton in the box and randomly selecting an instanton to be removed. Imposing detailed balance and considering the case of an instanton, we arrive at

$$\frac{1}{V} p_{N_I, N_A}^{\text{eq}} \mathcal{A}_{N_I, N_I+1} = \frac{1}{N_I + 1} p_{N_I+1, N_A}^{\text{eq}} \mathcal{A}_{N_I+1, N_I}. \quad (3.48)$$

As usual, $p_{N_I, N_A}^{\text{eq}} = Z_{N_I, N_A} / Z^{16}$ is the probability to be in the state $\{N_I, N_A\}$. The acceptance probability \mathcal{A}_{ij} is implicitly defined through (3.48), and the

¹⁶Note that we neglect the factorial terms in the definition of the equilibrium probability density p^{eq} because they are an artifact as far as the measure is concerned. They have been introduced to render the integration volume simple, i.e. the product of the single instanton moduli-spaces M^{N_I} . During the integration process all the permutations of a given set of coordinates are generated, but, since the instantons are indistinguishable, they really correspond to only one configuration. To correct for this overcounting, we then have to divide by a factor of $N_I!$. The important point is that for the transition probabilities these factorial factors are irrelevant.

Metropolis algorithm defines it to have the following form, [70],

$$\mathcal{A}_{N_I, N_I+1} = \min(1, \mathcal{A}), \quad (3.49)$$

$$\mathcal{A}_{N_I+1, N_I} = \min(1, \mathcal{A}^{-1}). \quad (3.50)$$

Plugging this into (3.48) we finally arrive at

$$\mathcal{A} = \frac{V}{N_I + 1} \frac{p_{N_I+1, N_A}^{\text{eq}}}{p_{N_I, N_A}^{\text{eq}}}. \quad (3.51)$$

The difference to ordinary Monte Carlo moves, as used in the canonical ensemble¹⁷, is that the proposal probabilities do not cancel and the transition matrix is not symmetric. In this specific case, the proposal probability for an insertion is $\mathcal{P}_P^{\text{ins}} = 1/V$, corresponding to the probability to place the instanton randomly within the box, whereas the proposal probability for a deletion is $\mathcal{P}_P^{\text{del}} = 1/(N_I + 1)$, corresponding to the probability to select an instanton among the $N_I + 1$ available.

When we perform the standard updates, it is easy to monitor the acceptance rates and tune the the proposal probabilities to achieve good rates, i.e. 50% say. For the move described by (3.51) we do not have a parameter to tune though. At $T = 0$, this is not really a big issue because it turns out that the acceptance rate is ≈ 0.4 , even for the rather small quark masses that we will use. This is still acceptable and does not really justify the overhead of more sophisticated update algorithms.

Also, note that such an insertion/deletion step is a fairly large change as compared to the normal coordinate updates, and so these grand canonical moves actually help to sweep through phase space more quickly.

Finally, we need to decide how many grand canonical moves we perform per coordinate update, that is we need to fix the a-priori-probabilities δ_C and δ_{GC} . We found that, for $T = 0$, the ensemble is not sensitive at all to this parameter, see Table 3.2. Since we will ultimately be interested in computing the topological susceptibility, we will aim to achieve low autocorrelation times for the instanton number, $N = N_I + N_A$, and topological charge, $Q = N_I - N_A$, i.e. we will perform rather more insertion/deletion moves than less. In

¹⁷That is, updates for the positions, sizes and colour orientations

δ_C	δ_{GC}	$\langle N \rangle$	ξ_N	$\langle Q^2 \rangle$	ξ_{Q^2}	$\langle S_{\text{int}} \rangle$	ξ_{int}
0.5	0.5	101.4(6)	2500	2.4(1)	200	-5.005(5)	1200
0.6	0.4	100.8(6)	2100	2.30(5)	130	-5.0010(5)	1400
0.8	0.2	102.6(6)	4000	2.35(4)	130	-5.010(2)	2400
0.9	0.1	102.1(6)	5000	2.5(1)	270	-5.016(3)	3700

Table 3.2: The sample size is roughly equivalent for each set, with 200 independent configurations generated according to the autocorrelation time ξ_N . Considering some bulk properties, we see that the sampling does not really depend on the a-priori-probabilities δ_i . Even though the autocorrelation times are only rough estimates, we will take the data at face value and choose $\delta_C = 0.6$ and $\delta_{GC} = 0.4$ for the remaining simulations.

practice we perform canonical moves only 60% of the time.

3.3.3 Fermionic determinant

As mentioned in the introduction, we want to study the IILM for ‘physical’ quark masses. In that case, we must make sure that the simulation box is large to be insensitive to finite size effects. In the lattice community it is common practice to use a box length that corresponds to 4 – 5 times the wavelength of the lightest propagating degree of freedom, which is the pion. In practice, we want to circumvent the need for extremely large boxes by studying the thermodynamic limit, $V \rightarrow \infty$.

As compared to fitting formulas, our combination of interpolation and asymptotic matching results in a rather substantial computational overhead. This is particularly so in the quenched case. For unquenched simulations the situation is less drastic as the computationally most demanding part is the evaluation of the determinant and/or the determination of the eigensystem of the Dirac operator. Increasing the simulation box, i.e. increasing the number of instantons, this becomes the bottleneck to large volume simulations.

The Monte Carlo changes are, however, of a rather simple form, changing only one column of the overlap matrix T at a time. We can therefore use decomposition update techniques to reduce the complexity from $O(N^3)$ to $O(N^2)$.

For the update step we only need to evaluate the determinant (3.36). Given the fact that $m^2 + TT^\dagger$, respectively $m^2 + T^\dagger T$, is a positive defi-

nite hermitian matrix, the fastest evaluation will be achieved by using the Cholesky decomposition. An added bonus is that the Cholesky decomposition and its algorithm are known to be very stable.

Focusing on $M^2 = m^2 + TT^\dagger = LDL^\dagger$, an update $T' = T + \Delta T$ can be written as two rank 1 updates for M^2 , of the form

$$M'^2 = M^2 + \Phi\Phi^\dagger - \Psi\Psi^\dagger, \quad (3.52)$$

with Φ, Ψ vectors. Details are given in appendix C, where we also discuss more efficient ways to deal with adding and removing instantons, and the corresponding updates. The Cholesky decomposition can be updated efficiently when it only changes by rank 1 matrices, that is transformations of the form

$$M'^2 = L'D'L'^\dagger = M^2 + \alpha zz^\dagger = L(D + \alpha ww^\dagger)L^\dagger, \quad (3.53)$$

where $Lw = z$. The algorithms then compute the decomposition of $D + \alpha ww^\dagger = \tilde{L}\tilde{D}\tilde{L}^\dagger$, which can be achieved in $O(N^2)$ because D is diagonal. Furthermore, the matrix \tilde{L} has a special form which allows an efficient matrix multiplication, $L' = L\tilde{L}$, in $O(N^2)$. Details can be found in [81]. The algorithm we use in practice is known to be unstable for downgrading, $\alpha < 0$, unless the resulting matrix, M'^2 , is known to be positive definite. Since upgrading, $\alpha > 0$, is always stable, it is important to perform the two consecutive updates in the order given by (3.52).

In general we will be performing grand canonical simulations, and need to keep track of two decompositions, one for $m^2 + TT^\dagger$ and one for $m^2 + T^\dagger T$. Furthermore, we deal with 3 active quarks so that each Monte Carlo update entails $2 \cdot 2 \cdot 3 = 12$ rank 1 updates. We find that for an ensemble with 100 instantons and 100 anti-instantons we achieve a computational gain of a factor of 2 as compared to the full Cholesky decomposition.

3.4 Different Ensembles

To be predictive, the IILM should not depend too sensitively on the chosen ansatz. Given the fact that, for instance, the streamline and the ratio ansatz have quite different functional forms for overlapping pairs, the insensitivity

of the model to specific background ansätze can only be determined a posteriori. On a heuristic level, we expect insensitivity to emerge if the ensemble stabilises in a rather dilute form so that the precise functional form of the repulsion is irrelevant. The large separation limit is a priori unproblematic because all ansätze are constructed such that, asymptotically, they approach the simple sum ansatz, i.e. $A = A_1 + A_2$, with A the gauge field.

First, we will frame our discussion on the pair interactions. The total effective interaction of a pair of oppositely charged partners is given by

$$S_{12} = S_0(\sqrt{\rho_1\rho_2})V_{12} - \sum_{n=1}^{N_f} \ln \left(\frac{|T_{12}|^2 + m_n^2}{m_n^2} \right). \quad (3.54)$$

Identically charged pairs only feel the gluonic interaction as $T_{II} = T_{AA} = 0$. As expected, the ratio and streamline ansatz are markedly different only for strongly overlapping pairs, see Fig. 3.9, where strongly overlapping pairs are characterised by $R \leq \sqrt{\rho_1^2 + \rho_2^2}$.

In the quenched case, we notice that the R_E ansatz has a higher absolute minimum as compared to the R_H ansatz, occurring roughly for the same separation. So we expect the ensemble to become slightly more dilute because it will not be as favourable, energetically, for instantons to come close. For unequal sizes, however, the repulsion is weaker in the R_E case which would favour a denser ensemble, as less volume is excluded. The streamline ansatz will lead to a substantially more dilute system because the core repulsion is broader, excluding more volume for the instantons to move through.

In the unquenched case, the difference in the absolute interaction strength is much more pronounced between R_E and R_H . We therefore expect that the R_E ansatz should be quite a bit more dilute as compared to the R_H ansatz. Considering that the streamline ansatz has a deeper minimum than the R_E ansatz, the former will favour instantons to come closer. However, it has more excluded volume. Both trends work in opposite directions, and there is a possibility that they lead to roughly identical ensembles, at least on the level of the instanton density.

We will address these issues in more detail by performing canonical simulations and minimising the free energy. This follows closely [157], and also serves to validate our code against their results. Note that the simulations

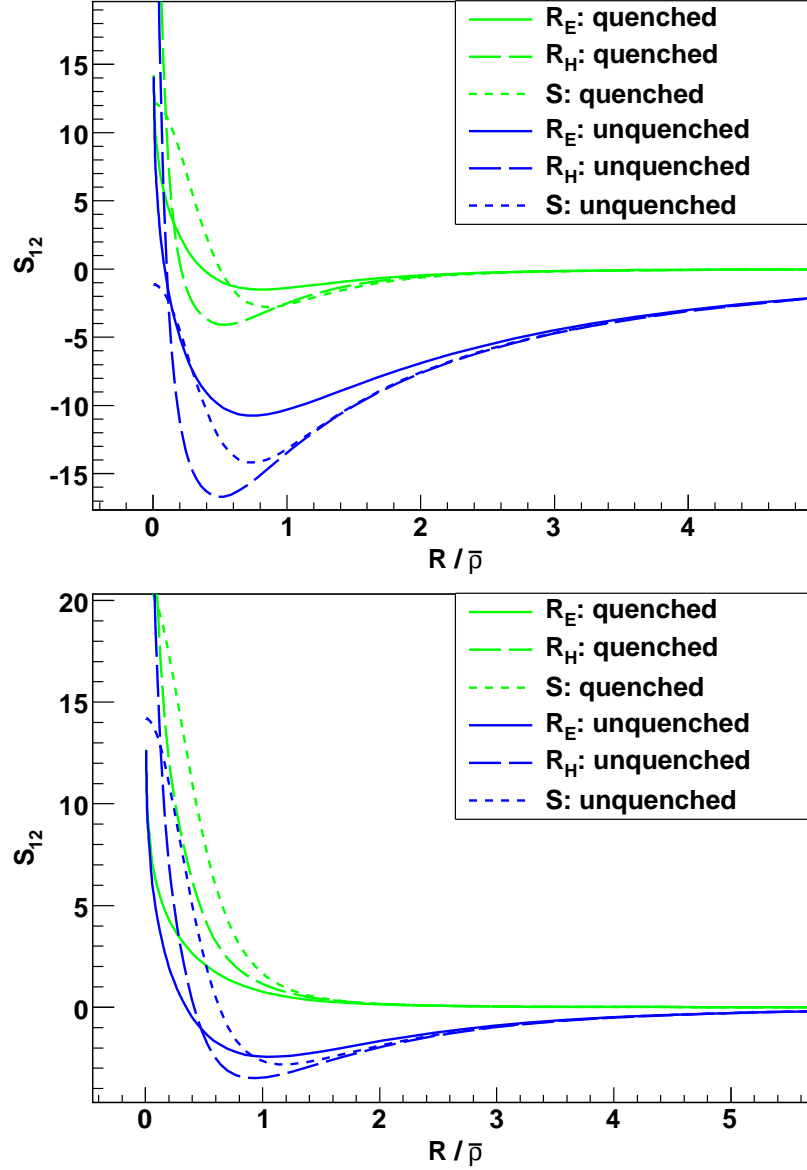


Figure 3.9: *Top*: Most attractive colour orientation. *Bottom*: Random colour orientation. For both graphs the instantons have different size parameters. We expect the R_H ensemble to be denser than the R_E ensemble because the attraction well is deeper, whereas the excluded volume due to repulsion is not that different. Along the same lines, the S ansatz should lead to a rather more dilute system in the quenched case. For unquenched simulations, the deeper attraction well, and the steeper and broader repulsion, of the S interactions might lead to an ensemble roughly equivalent to the R_E one.

are performed in the topologically trivial sector, for which $N_I = N_A$, i.e. the topological charge $Q = 0$. In Fig. 3.10 we plot the free energy $F = -\ln Z/V$ against the instanton density $n = (N_I + N_A)/V$. As expected from our considerations of the pair interactions, in the quenched case the R_E and R_H ansätze are only slightly different, with R_E leading to a slightly denser ensemble. Also, the interactions stored in that ensemble¹⁸ are a bit lower, again as could be anticipated from the pair interactions. The S ansatz leads to a much more dilute instanton ensemble, and our data reproduces well that of [157].

Ultimately, we will be interested in smaller quark masses. It is clear from (3.54) and Fig. 3.4 that smaller masses increase the quark interaction strength as compared to the gluonic counterpart, which stays constant and is responsible for the core repulsion. From a purely energetic point of view, smaller quark masses should then lead to denser ensembles. However, we clearly see in Fig. 3.10 that the ensembles become more dilute. The reason is that the small quark masses enter the instanton size distribution; in turn, the density, in the dilute gas limit, is entirely set by the size distribution, i.e. $n = 2 \int d\rho d(\rho)$. Bringing it into the action, we can interpret the size distribution as the energy cost needed to insert an instanton into the box. This is a well-known fact, namely that small quark masses suppress instanton contributions to the QCD vacuum because the different topological vacua become equivalent in the limit of vanishing quark masses; phrased differently, the energy barrier has disappeared, and only field configurations with topological charge $Q = 0$ survive. In the dilute gas approximation this leads to the disappearance of instantons altogether. As Fig. 3.10 shows, this is not true for an interacting instanton ensemble, where the instanton density converges to a finite limit as the quark mass is lowered¹⁹. The results from Fig. 3.10 also show that the R_E ansatz generates an ensemble that differs more and more from the R_H ansatz, as was anticipated from our considerations of the pair interactions. We can also clearly see that the R_E ensemble does not converge to the S ensemble.

¹⁸The difference in the free energies is directly related to the difference of the interaction per instanton

¹⁹Remember that the simulations take place in the topologically trivial sector, i.e. $N_I = N_A$ or $Q = 0$.

	m_u	m_d	m_s
M_1	0.1	0.1	0.7
M_2	0.05	0.05	0.3
M_3	0.012	0.022	0.44

Table 3.3: We use three different sets of quark masses and investigate how the instanton liquid depends on them.

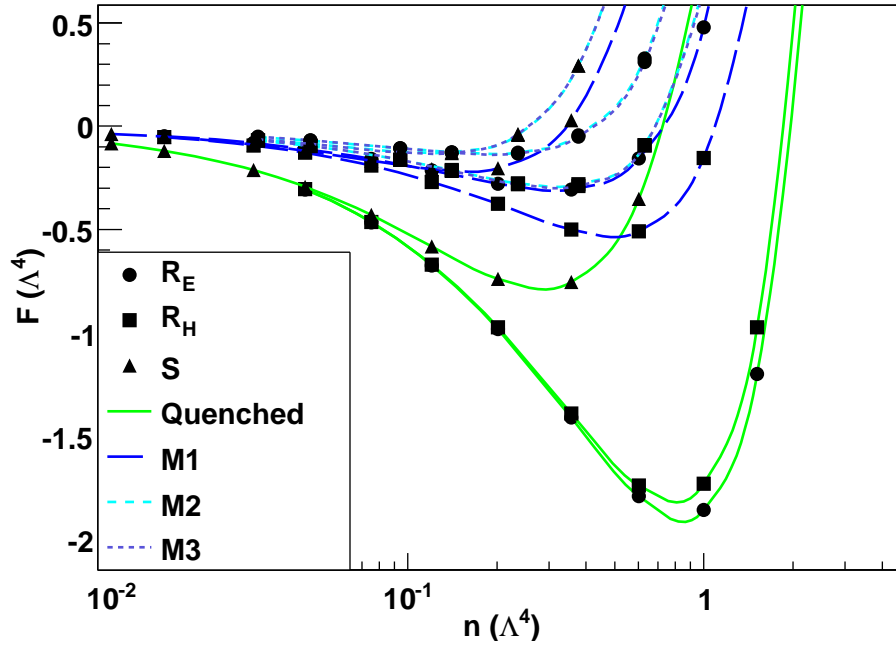


Figure 3.10: The different quark masses are summarised in Table 3.3. The simulations were performed in the topologically trivial sector, i.e. $N_I = N_A$. For the quenched and the M_1 case we fixed $N = 64$ as in [157]. The other two unquenched simulations have $N = 200$. We clearly see that small quark masses suppress instanton contributions to the QCD vacuum, but also that there exists a finite limit for the instanton density as the quark masses vanish; this is in contrast to the dilute gas approximation which suppresses instanton contributions completely for zero quark masses. For unquenched simulations the free energy for the R_E ensemble roughly agrees with that of the S ensemble, although the equilibrium densities are rather different. Still, the approximate equality between the free energies might be interpreted as evidence of an approximate equivalence between both ensembles for bulk properties, e.g. equivalent pressure, since it is directly related to the free energy. However, the R_E liquid does not seem to converge towards the S ensemble as we lower the quark masses.

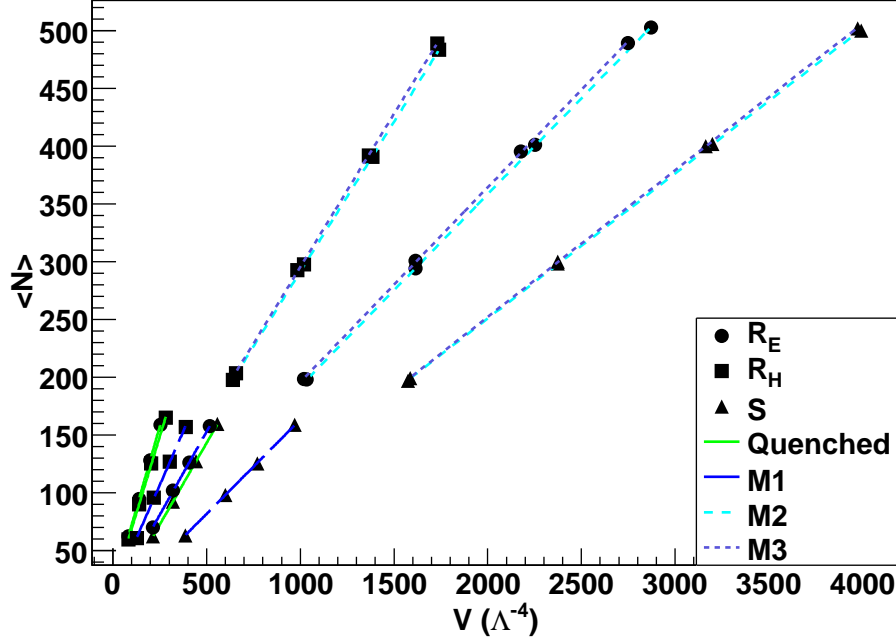


Figure 3.11: As anticipated from the canonical study, the instanton number for both ratio ansätze is very similar in the quenched case. There seems to exist a finite limit for the instanton density as the quark masses vanish, and instantons will be present in the QCD vacuum even in the chiral limit; this is in sharp contrast to dilute gas approximations.

So far we have framed the discussion essentially in terms of the instanton density. To investigate the similarities and differences in more detail, we will look at a few bulk properties and their dependence on the different ansätze in the thermodynamic limit and the grand canonical ensemble. We clearly see how the density decreases with the quark masses, but approaches a finite limit, Fig. 3.11. As we have discussed before, the quark masses will suppress fluctuations to inequivalent topological sectors and in the limit of vanishing masses only the trivial $Q = 0$ sector will survive. The fluctuations between topological sectors are encoded in the topological susceptibility $\chi = \langle Q^2 \rangle / V$, which vanishes with the quark masses, see Fig. 3.12. Both the instanton number and the topological charge fluctuations exhibit a nice scaling with the volume.

We will now turn to an intensive quantity, the mean instanton size $\bar{\rho}$ Fig. 3.13. For the quenched and the first unquenched, M_1 , simulations,

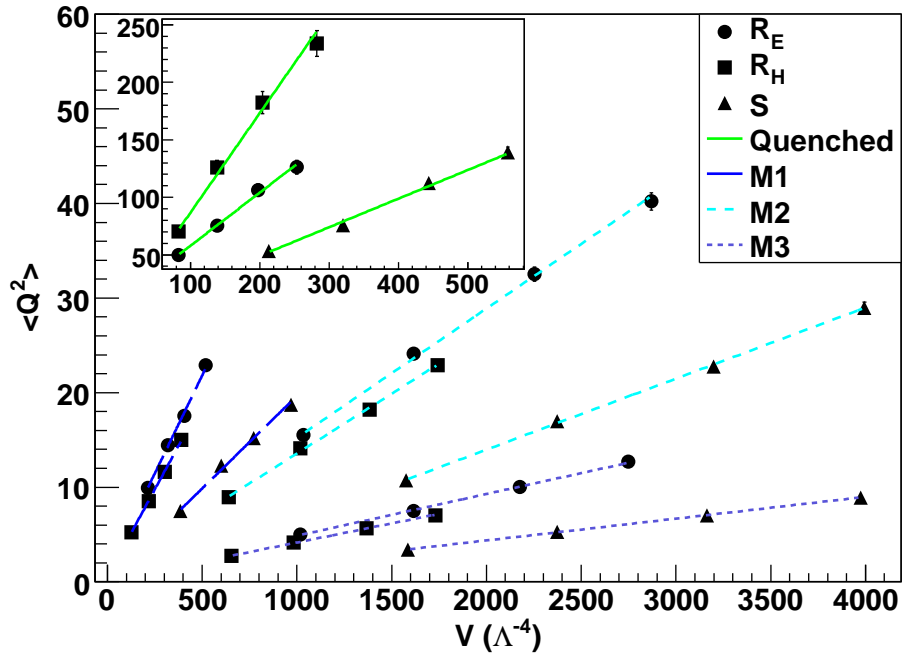


Figure 3.12: The topological susceptibility, the slope of the graphs, is very sensitive to the quark masses. It is screened by small quark masses and will vanish in the chiral limit. This is expected as QCD with massless quarks does not have topologically inequivalent vacua; in this case the so-called θ parameter is not physical and can be rotated away by a chiral rotation of the quark fields. See also [173] for another work on the topological susceptibility in the ILM.

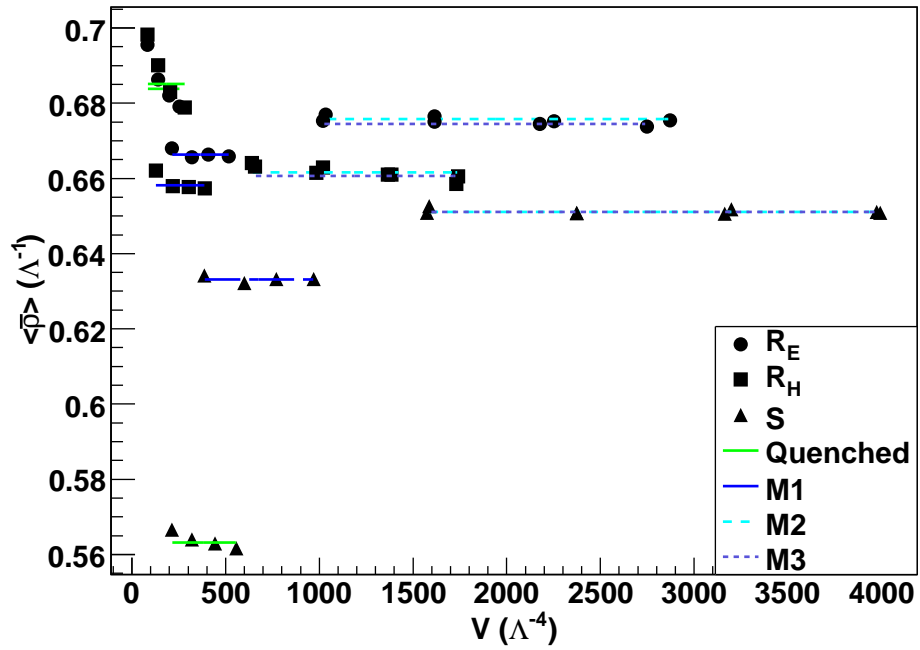


Figure 3.13: For the quenched and the M_1 simulations, which have been fixed to $N \approx 64$ for the smallest volume, the simulation boxes are still too small, as can be seen by the systematic drift. For the other two unquenched simulations ($N \approx 200$ for the smallest volume) we are much closer to the thermodynamic limit, although there are still systematic deviations for the R_H ansatz. In any case, the different ansätze give rather similar results. Also the mean instanton size approaches a unique limit for small quark masses.

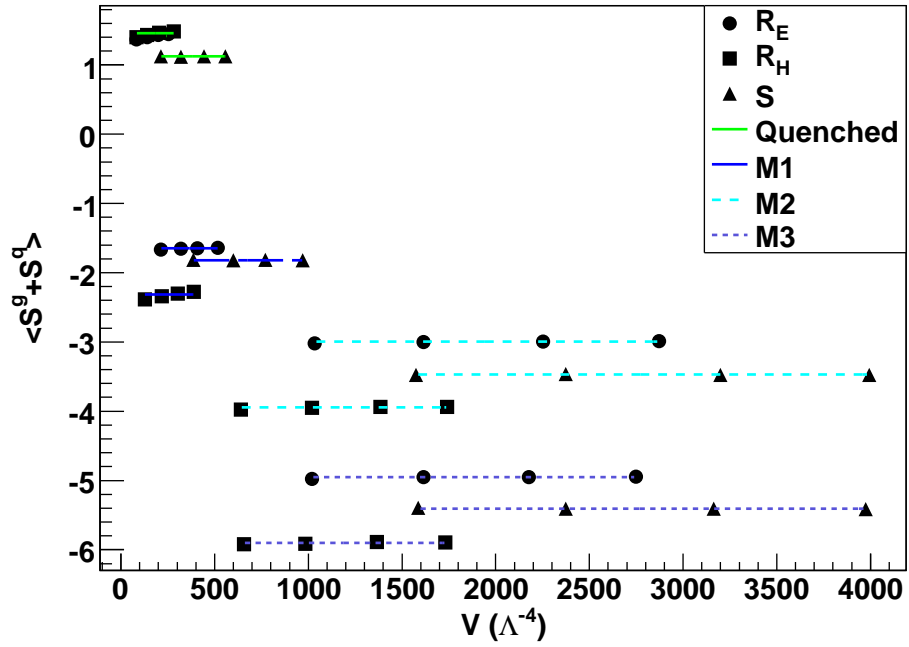


Figure 3.14: As anticipated from the pair interaction considerations, the interactions are very similar in the quenched sector for R_E and R_H . For full simulations, the differences between the ansätze stay constant as the quark masses vary, with R_H leading to the strongest attractive interactions, as was expected. Note that the pair interactions are less sensitive to finite size effects compared to the mean instanton size.

	n	χ	$\langle \rho \rangle$	$\langle S^g + S^q \rangle / V$
Quenched	0.567(1) [R_E]	0.46(3)	0.6837(2)	1.420(1)
	0.532(1) [R_H]	0.86(5)	0.6850(1)	1.455(1)
	0.282(1) [S]	0.24(1)	0.5631(1)	1.122(1)
M_1	0.288(2)	0.041(1)	0.6662(2)	-1.648(1)
	0.370(2)	0.0369(8)	0.6581(3)	-2.311(2)
	0.163(1)	0.0196(6)	0.6331(3)	-1.818(3)
M_2	0.1660(7)	0.0136(4)	0.6757(2)	-2.996(1)
	0.259(1)	0.0126(2)	0.6615(2)	-3.945(2)
	0.1255(8)	0.0075(2)	0.6510(2)	-3.472(2)
M_3	0.1686(8)	0.00440(7)	0.6744(2)	-4.954(1)
	0.265(1)	0.00403(7)	0.6606(2)	-5.902(2)
	0.1269(5)	0.00230(5)	0.6511(2)	-5.407(2)

Table 3.4: Thermodynamic extrapolations for the instanton density, the topological susceptibility, the mean instanton size and the mean interaction. The data has been obtained from Figs. 3.11, 3.12, 3.13 and 3.14 respectively.

which have been calibrated to achieve $N \approx 64$ for the smallest volume as in the canonical simulations, we see that the simulation boxes are not large enough, even though the density and the charge fluctuations seem to suggest otherwise, i.e. display good scaling with V . For the other two unquenched simulations, tuned to $N \approx 200$ for the smallest volume, we have reached volume sizes large enough to perform a thermodynamic limit. It is worth noticing that the mean instanton size is a rather robust quantity, and does neither depend strongly on the ansatz nor on the quark masses. This makes it a good quantity to use when comparing data from different ensembles, e.g. [46] where the authors establish that the scale at which the IILM is operating is given by the inverse of the mean instanton size. Another intensive quantity, the interaction per instanton, is less sensitive to finite size effects, see Fig. 3.14. The data shows that the weaker repulsion of the R_E ansatz as compared to the R_H ansatz dominates over the deeper attractive well of the latter. Therefore, the total interaction in the R_E ensemble is slightly lower, leading to a denser system. The stronger repulsion for the S interactions leads to more excluded volume; this, in turn, leads to lower interactions and a more dilute ensemble. These conclusions are in agreement with the direct measurement of the instanton density Fig. 3.11.

We are mostly interested in unquenched results, and the following comments relate this sector. From the data of Table 3.4, we can infer that results for $\chi^{1/4}$ have a 12% systematic ansatz dependence. The topological susceptibility is surprisingly similar for the R_E and R_H ansatz in the unquenched sector. The mean instanton size is indeed a rather robust quantity and only affected on the 3% level by these systematics. Also, note that the mass dependence on $\langle\rho\rangle$ is rather small, with differences not larger than 5%. The instanton interactions and $n^{1/4}$ agree within 20%, and the latter converges to a fixed limit as the quark masses vanish.

3.5 Fixing parameters

3.5.1 Quenched case

In the quenched case, the IILM has only one freely adjustable constant, the lambda parameter Λ . We need one observable, from the lattice say, to fix it. Different approaches can be chosen. In the early works, Λ was determined by fixing the instanton density to 1 fm^{-4} at $T = 0$. To compare this with the lattice is not straightforward as the classical instanton content is convoluted with the quantum mechanical fluctuations. With the discovery of the KvBLL calorons, there is a renewed interest in studying the topological structures on the lattice, see for instance [29]. Since the topological susceptibility is well measured on the lattice and is easily accessible within the IILM, it is a natural candidate. The lambda parameter is then given by

$$\Lambda = \sqrt[4]{\frac{\chi_{\text{lat}}}{\chi_{\text{IILM}}}}. \quad (3.55)$$

We will use $\chi_{\text{lat}}^{1/4} = 193 \text{ MeV}$, [60]. The topological susceptibility in the IILM is extracted from Fig. 3.15 by using the definition

$$\chi_{\text{top}} = \lim_{V \rightarrow \infty} \frac{\langle Q^2 \rangle}{V}. \quad (3.56)$$

This yields $\Lambda = 234(1) \text{ MeV}$. The error is purely statistical. The instanton density turns out to be $n = 0.543 \Lambda^4 = 1.02(2) \text{ fm}^{-4}$, fairly close to the

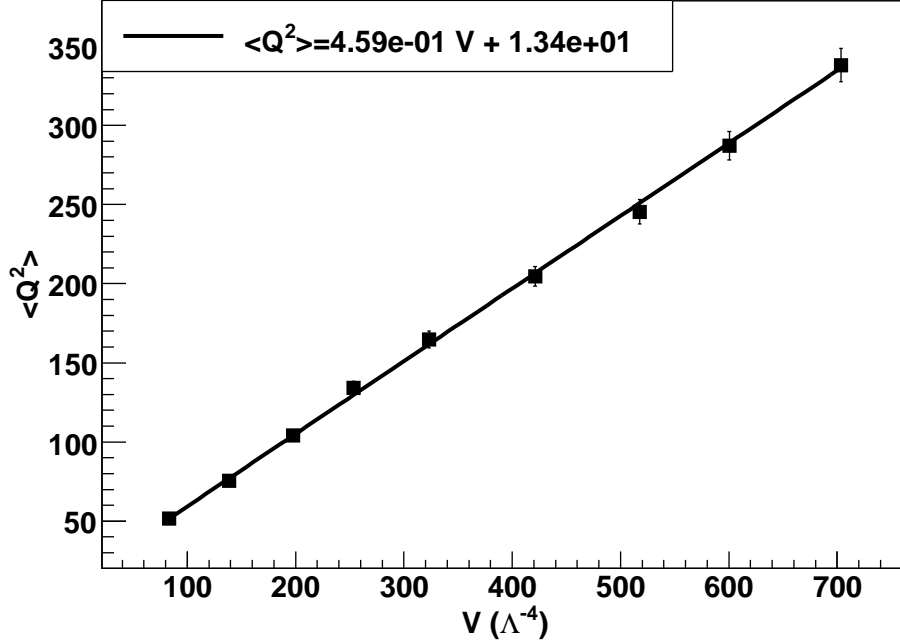


Figure 3.15: The fluctuations of the topological charge $\langle Q^2 \rangle$ show a nice linear dependence with the volume box V , as it should be for an extensive quantity. From this we infer the topological susceptibility $\chi_{\text{top}} = \lim_{V \rightarrow \infty} \frac{\langle Q^2 \rangle}{V}$.

usually quoted phenomenological value of $n = 1 \text{ fm}^{-4}$. We find that even for these larger volumes the mean instanton size is still evolving towards lower values, as in Fig. 3.13. The largest volume then leads to the upper bound $\bar{\rho} < 0.57 \text{ fm}$. Using a simple fit to $\bar{\rho} = \bar{\rho}_\infty + \alpha V^{-0.25}$ to extrapolate to the asymptotic value, we find $\bar{\rho}_\infty \approx 0.53 \text{ fm}$; this is rather large compared to the phenomenological value of $\bar{\rho} \approx 0.33 \text{ fm}$.

To estimate the systematic error due to the dependence on the ansatz, we will use the data from Table 3.4. The fact that we take a fourth root reduces the rather large differences in χ_{IILM} to about 15% for Λ , i.e. $\Lambda = 234(35) \text{ MeV}$. Our value has been obtained through simulations in PV regularisation. To compare it with lattice data, we will convert it to the $\overline{\text{MS}}$ scheme, [94], $\Lambda_{\overline{\text{MS}}}/\Lambda_{\text{PV}} = \exp(-1/22)$. This gives $\Lambda_{\overline{\text{MS}}} = 224(33)$ and compares well with the lattice result $\Lambda_{\overline{\text{MS}}} = 259(20)$ [82].

3.5.2 Unquenched case

We want to use realistic quark masses. These are fairly small, and one must worry whether such light degrees of freedom will fit into the simulation box. The usual approach, used in the lattice community and also in work on the IILM [46], is to compute the pion mass from a set of unphysical quarks and to fix the volume box such that $Lm_\pi > 5$; chiral perturbation theory can then be used to extrapolate to physical masses. Ultimately, the lattice wants to test the predictions of chiral perturbation theory as well, and in recent years, the computing power and, most importantly, the algorithms have improved to such an extent that physical quark mass simulations are becoming feasible; however, these are still immensely costly simulations, and $2 + 1$ flavour simulations were rare until recently.

We follow a rather more modest rationale by simply demanding that the quark mass be at least so small as to be comparable to the lowest eigenvalue of the Dirac operator, $\langle \lambda_{\min} \rangle$, see Fig. 3.17. This sets the smallest box we use in our simulations. We then use ever larger volumes and extrapolate to the thermodynamic limit.

In [46] the lambda parameter²⁰ is fixed by computing the meson and nucleon masses, through current correlators of the interpolating fields and their asymptotic spatial decay, and by comparing them with the available lattice data. This study established that the IILM is compatible with the predictions of chiral perturbation theory. We will take this for granted in what follows.

In order to fix Λ , we could still use the topological susceptibility as it is routinely measured on the lattice. However, the topological susceptibility depends strongly on the quark masses, see Fig. 3.12. We can get rid of the mass dependence by using chiral perturbation theory and computing the chiral condensate $\langle \bar{q}q \rangle$ ²¹. The chiral condensate has been studied within chiral perturbation theory and, more recently, it has been precisely determined on the lattice [39, 38, 37]. We will take it to be $\langle \bar{q}q \rangle_0^{\overline{\text{MS}}}(\mu = 2 \text{ GeV}) = 250 \text{ MeV}$.

²⁰Actually, the authors fixed the mean instanton size. But it is trivial to relate the latter to the lambda parameter.

²¹To reiterate, we implicitly rely on the fact that the IILM is describing well the chiral properties of QCD, as has been checked in numerous studies, the most convincing being [46].

To extract the chiral condensate from the IILM, we will use the procedure adopted in [39, 38, 37]: we compute the topological susceptibility for different sets of quark masses and extrapolate to the chiral limit, $m_i \rightarrow 0$. The condensate can then be determined by chiral perturbation theory [125],

$$\begin{aligned}\chi &= m_{\text{eff}} \langle \bar{q}q \rangle_0 + O(m^2), \\ m_{\text{eff}} &= \left(\sum_n^{N_f} \frac{1}{m_n} \right)^{-1}.\end{aligned}\tag{3.57}$$

The chiral condensate has an anomalous dimension and, therefore, depends on the scale. Furthermore, the IILM is set up with a PV regulator, whereas the quoted result is computed in dimensional regularisation. It is well known that within an unphysical renormalisation scheme such as $\overline{\text{MS}}^{22}$ the results depend on the regulator (for unphysical quantities like masses, coupling constants and amplitudes). We therefore need to compute the finite counterterms that relate the PV to the $\overline{\text{MS}}$ regularised results. Using the result (2.63) we derived in section 2.2.2, we find that $\langle \bar{q}q \rangle_0^{\text{PV}}(\mu = 2 \text{ GeV}) \approx 244 \text{ MeV}$. This is a one-loop result. The two-loop correction can be estimated very roughly to be on the 10% level as is typical for computations around the scale of $\mu = 2 \text{ GeV}^{23}$.

We will define the scale of the IILM by $\mu_\Lambda = \Lambda/\bar{\rho}$, as suggested in [46], and determine Λ from the self-consistency equation

$$\langle \bar{q}q \rangle_0^{\text{PV}}(\mu_\Lambda) = \Lambda^3 \langle \bar{q}q \rangle_0^{\text{IILM}},\tag{3.58}$$

where we run the chiral condensate $\langle \bar{q}q \rangle_0^{\text{PV}}(\mu)$ at one loop. To that order, there is no difference between schemes and we can use the $\overline{\text{MS}}$ results, e.g. [191].

To get an estimate of the quark mass ratio dependence, we have used two

²²t Hooft's computation of the one-loop instanton measure, using Pauli-Villars regularisation, is also unphysical because, instead of poles, logarithms of the regulator mass are subtracted.

²³Strictly speaking, we should use the two-loop result because the simulations in the IILM have been obtained using the two-loop improved instanton measure. However, Pauli-Villars regularisation is not straightforward for non-Abelian gauge theories beyond the one-loop level, and we do not have the expertise to embark on this endeavour. In any case, the difference should still be on the 10% level.

different sets of quark masses, one inspired by the chiral perturbation theory and the other by the quark masses extracted from the lattice [131]. The two sets have the following ratios

$$\frac{m_i}{m_j} = \begin{cases} 1 : 1.83 : 36.7 & (M_1) \\ 1 : 2.32 : 45.0 & (M_2) \end{cases}. \quad (3.59)$$

For each set we perform 5 simulations with ever smaller absolute masses, see Fig. 3.16. This data is fitted to (3.57) to extract the chiral condensate. The results for the two sets agree on the 1σ level, and we can argue that the chiral condensate depends only weakly on the quark mass ratios, given that the latter vary by roughly 25%, see (3.59). This is as it should be since the exact chiral condensate does not depend on the quark masses at all. From an operational point of view, the robustness against quark mass ratios²⁴ makes the chiral condensate a good quantity to set Λ .

Solving (3.58) we find that the lambda parameter is given by

$$\Lambda_i = \begin{cases} 401(5)(40)(15) \text{ MeV} \\ 389(6)(40)(15) \text{ MeV} \end{cases}, \quad (3.60)$$

where the errors follow from the fit, $\langle \bar{q}q \rangle_0^{\text{PV}}$ and the systematic on χ . This leads to an overall error of 44 MeV, or roughly 11%, and is strongly dominated by the one-loop result for $\langle \bar{q}q \rangle_0^{\text{PV}}$. Since we run at one loop in the self-consistency equation (3.58), such a large error is certainly realistic, if not underestimated.

We found that running at two-loop²⁵ gives results consistent with (3.60). Using the prescription of [46], the scale and the mean instanton size for the

²⁴i.e. taking the limit from different directions in quark mass space.

²⁵We use β -functions and anomalous dimensions from the $\overline{\text{MS}}$ scheme, [191], since we do not know them for PV regularisation. However, both regularisations are thought to give roughly similar results, for instance $\langle \bar{q}q \rangle_0^{\text{PV}}$ and $\langle \bar{q}q \rangle_0^{\overline{\text{MS}}}$ agree on the 3% level at one-loop and $\mu = 2 \text{ GeV}$, and for the purpose of estimating errors in the one-loop running this procedure should be fine.

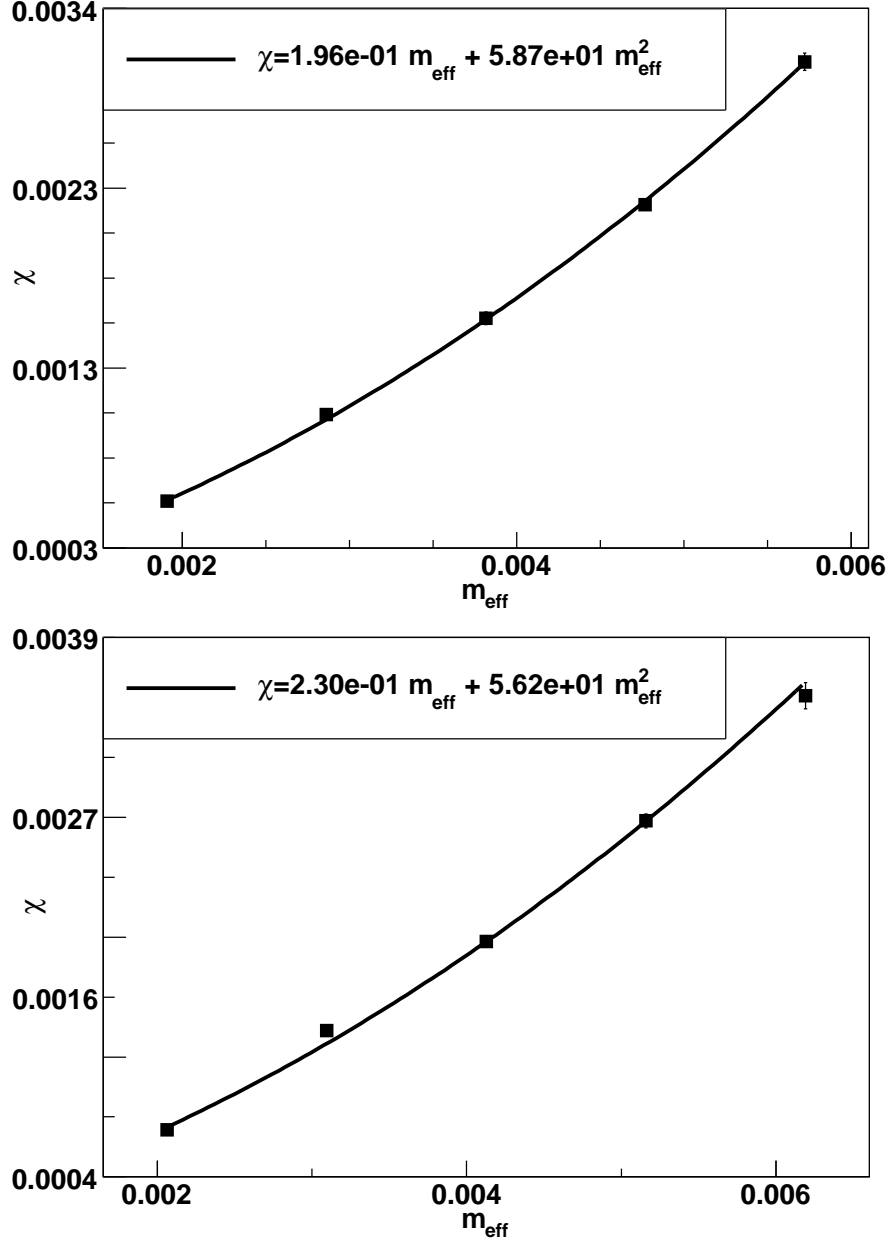


Figure 3.16: Computing the topological susceptibility allows for the extraction of the chiral condensate by using chiral perturbation theory, (3.57). The rationale is the same as used in recent lattice studies to extract the chiral condensate [39, 38, 37]. In order to get a rough estimate on the systematic error introduced by the chiral limit, two sets of masses have been used. The upper plot corresponds to the set M_1 and the lower plot to M_2 , as given in (3.59). The chiral condensate for both mass ratios agrees on the 1σ level, and we conclude that it depends only weakly on the quark mass ratios.

IILM is²⁶

$$\mu_{\Lambda_i} = \begin{cases} 598(65) \text{ MeV} \\ 580(64) \text{ MeV} \end{cases} \quad \rho_{\Lambda_i} = \begin{cases} 0.33(3) \text{ fm} \\ 0.34(4) \text{ fm} \end{cases} . \quad (3.61)$$

This is in very good agreement with the precision study [46]. Given that both works use chiral properties for the calibrations, the nice overlap is probably not totally unexpected.

Note that (3.60) is a prediction for the lambda parameter with 3 active quark flavours. To compare our result with experimental data, we run down the coupling constant $\alpha_s^{\overline{\text{MS}}} = 0.117(2)$ [64] from M_Z to μ_Λ and convert it to a lambda parameter. This is a rather big difference in scales and it is appropriate to use two-loop running, although not entirely consistent when we compare it to the one-loop result (3.60). To deal with threshold effects, we use the *Mathematica* package *RunDec*, [36]. The conversion between the $\overline{\text{MS}}$ and PV lambda parameters is given by [94], [11]

$$\Lambda_{\text{PV}} = \Lambda_{\overline{\text{MS}}} \exp \left(\frac{1}{22 - 4N_f/3} \right) . \quad (3.62)$$

This leads to $\Lambda_{\text{PV}}^{(3)} = 325(40)$ and the IILM result agrees on the 1σ level. Trusting the perturbative running down to the rather low scale μ_Λ is a leap of faith. However, earlier studies have seen good agreement between IILM and lattice predictions for physical quantities, such as meson masses, and so the agreement between the lambda parameters might not just be a fluke.

To determine the physical quark masses, we will use (3.57) rewritten in terms of the pion mass

$$\chi = m_\pi^2 f_\pi^2 \frac{m_u m_d}{(m_u + m_d)^2} + O \left(\frac{1}{m_s} \right) = \begin{cases} (77.4 \text{ MeV})^4 \\ (75.9 \text{ MeV})^4 \end{cases} . \quad (3.63)$$

We used $m_\pi = 135 \text{ MeV}$ and $f_\pi = 93 \text{ MeV}$. Together with the fits, Fig. 3.16, we can compute the corresponding quark masses. We convert them into $\overline{\text{MS}}$ masses at 2 GeV, run at one-loop, in order to compare them more easily with

²⁶Remember that in the unquenched case the instanton size is fairly independent of the quark masses.

other sources. Our results are

$$m_i^{\text{PV}}(\mu = 0.6 \text{ GeV}) = \begin{cases} 2.2(2) & 4.0(4) & 80(9) \\ 1.9(2) & 4.4(4) & 87(10) \end{cases} \text{ (MeV)}, \quad (3.64)$$

$$m_i^{\overline{\text{MS}}}(\mu = 2 \text{ GeV}) = \begin{cases} 1.9(2) & 3.4(5) & 69(11) \\ 1.7(2) & 3.8(5) & 74(11) \end{cases} \text{ (MeV)}. \quad (3.65)$$

The errors include an estimate from the 2-loop running. These masses compare well with the particle data group masses [63], i.e. $m_u = 1.5 - 3.3 \text{ MeV}$, $m_d = 3.5 - 6.0 \text{ MeV}$ and $m_s = 70 - 130 \text{ MeV}$, and to the lattice masses [131], i.e. $m_u = 1.9(2) \text{ MeV}$, $m_d = 4.4(3) \text{ MeV}$ and $m_s = 87(6) \text{ MeV}$.

Very large volume simulations are expensive even in the IILM, and the thermodynamic limit has been performed only on four volumes, in the range $2 \lesssim Lm_\pi \lesssim 3$. Even though the data has displayed a nice scaling with the volume, it is important to check whether the thermodynamic limit was consistent. To this end we will run large volume simulations, $Lm_\pi \in [2.11, 3.7]$, for the particular set of physical masses inspired by chiral perturbation theory: in dimensionless units, $m_u = 0.00546$, $m_d = 0.01001$ and $m_s = 0.2002$. This will allow us to estimate the systematic error introduced by performing the thermodynamic on the set of smaller simulation boxes.

The thermodynamic limit on the topological susceptibility turns out to be rather insensitive, see Fig. 3.18. However, the mean instanton size does not converge to a constant even for the largest volumes, see Fig. 3.19. It is a rather lucky fact that the mean instanton size does not vary much in absolute terms. The slope is clearly decreasing and we might estimate the convergence to occur somewhere in the range $\bar{\rho} \in [0.68, 0.66]$. A fit to $\bar{\rho} = \bar{\rho}_\infty + \alpha V^{-0.25}$ gives $\bar{\rho}_\infty = 0.6720(5) \Lambda^{-1} = 0.33(3) \text{ fm}$, in good agreement with the phenomenological value. The instanton density turns out to be $n = 1.7(7) \text{ fm}^{-4}$, and like the topological susceptibility displays a nice thermodynamic limit.

3.6 Conclusion

With the discovery of the new non-trivial holonomy calorons [119, 121], [123], there is renewed interest in studying the role of non-trivial field configurations in QCD, especially their role in the confinement/deconfinement phase

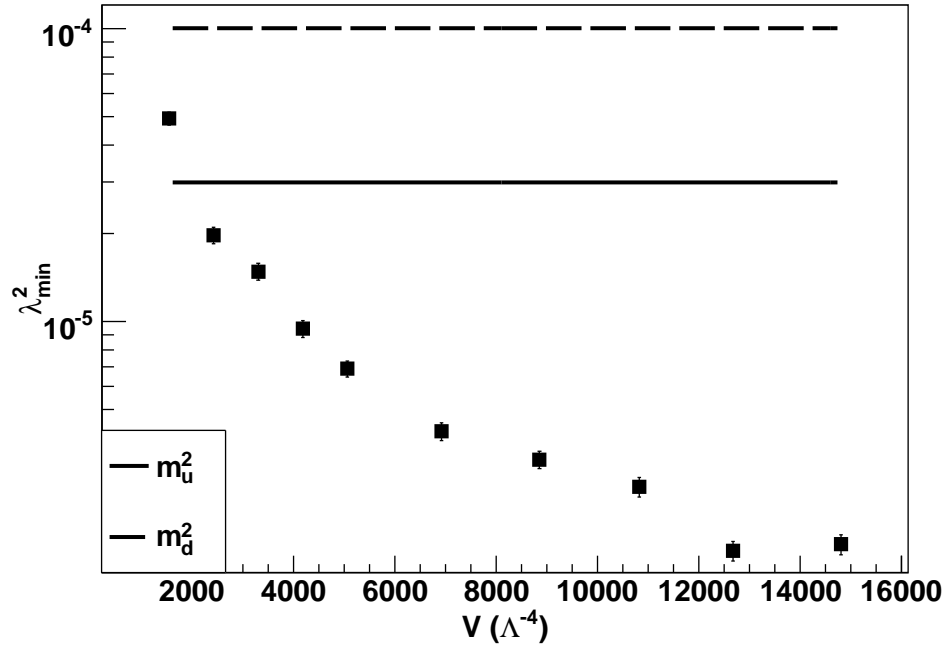


Figure 3.17: The simulations are performed for different simulation boxes. The average of the smallest Dirac eigenvalue, $\langle \lambda_{\min} \rangle$, is smaller than the quark masses for all but the smallest simulation box.

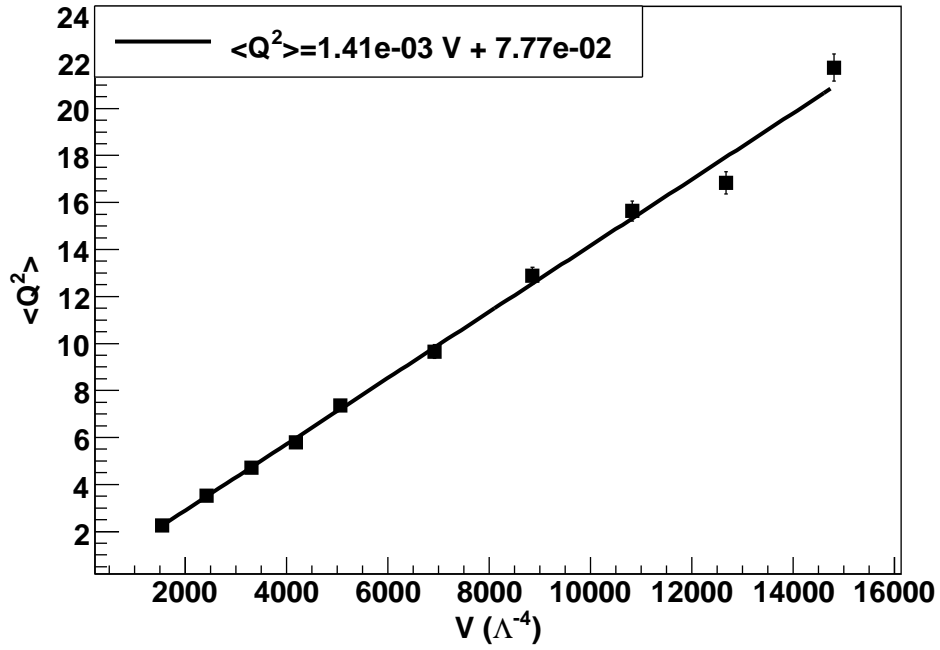


Figure 3.18: The fluctuations of the topological charge $\langle Q^2 \rangle$ show a nice linear dependence with the volume box V , as it should be for an extensive quantity. Applying the thermodynamic limit to the 4 smallest volumes yields a topological susceptibility that agrees on the 1σ level with the corresponding result using all available volumes. The mean instanton number ranges from $\langle N \rangle \approx 200$ to $\langle N \rangle \approx 1600$.

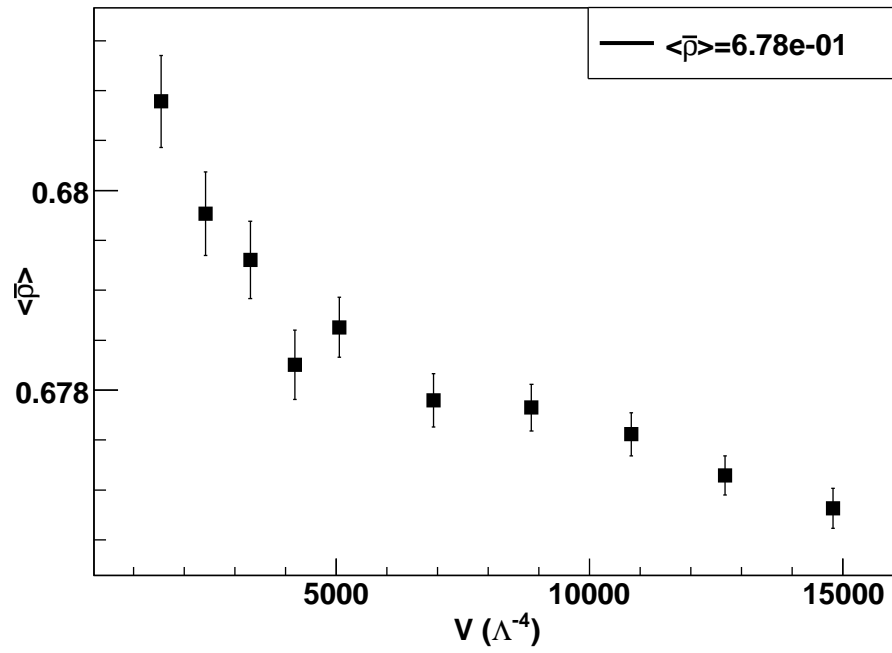


Figure 3.19: Even for the largest volumes the mean instanton size $\bar{\rho}$ is still decreasing and does not seem to converge to a constant. We are rather lucky that, although the effect is clearly systematic, the variation of $\bar{\rho}$ is small in absolute terms.

transition. A lattice based approach [80] is well suited for the pure gauge sector because it easily incorporates many-body instanton interactions in the classical action. However, the introduction of fermions will be plagued by the same problems that full lattice computations face. Most notably, the computations will become very costly.

A different approach, pioneered by Shuryak, Diakonov and Petrov, is to formulate the instanton liquid in the continuum, see for instance [157]. This approach suffers from the fact that it is not straightforward to include many-body interactions. Incidentally, only two-body effects are taken into account. The strongly localised profiles of instantons and the, a posteriori, fact that the instanton ensemble is rather dilute make this a viable working premise. The big advantage of the continuum formulation is the ease with which quarks can be incorporated, see also [197]. From this perspective, both the lattice and the continuum models complement each other rather well. Meanwhile, different publications have investigated the confining nature of other backgrounds, such as regular gauge instantons and merons [124, 136], [196] at zero temperature.

So far the continuum models used explicit analytic formulas for the interactions. They have been obtained through asymptotic considerations and fits to numerical evaluations of the classical action. We noted that these formulas do not possess a thermodynamic limit at finite temperature. More importantly perhaps, the more complex moduli-space of the non-trivial holonomy calorons probably demand a more systematic approach. In this paper we have set up a framework which we believe is numerically well-defined, can be extended to more complicated backgrounds and does not suffer from the parametrisation bias introduced implicitly through analytical formulas motivated by symmetry arguments and fits. The price to pay is a larger numerical overhead, which in our framework comes about through look-up tables and asymptotic matching formulas.

We have found that the analytic formulas of [157] agree very well with our interactions at zero temperature. Especially for the case of equal instanton sizes, where strong symmetry arguments support the analytic formulas, the agreement can be seen as a validation of the numerics. In general, however, the interactions of both schemes differ; the differences are especially

pronounced for the quark overlaps because in this paper we use the full ratio ansatz whereas a sum ansatz is used in [157]. Shifting the point of view, we considered the formulas of [157] to be another valid scheme; together with the streamline ansatz we studied the dependence of bulk properties of the IILM on the choice of these three rather different interactions and found that this introduces a systematic effect which depends on the quantity under consideration, but was generally rather large, up to 20%.

The IILM has been shown to be compatible with the chiral properties of QCD, see for instance [46]. A key chiral property, the topological susceptibility, has not been studied extensively within the IILM, see however [173]. One reason might be that the IILM was so far set up for simulations in the canonical ensemble whereas the topological susceptibility is most naturally studied in the grand canonical ensemble. We have enlarged the Monte Carlo moves to incorporate insertion/deletion steps in order to simulate an open ensemble. Apart from technical problems related to book-keeping issues this is rather straightforward.

A major incentive for this work was to investigate the regime of physical quark masses. In order to deal with such light quarks, rather large volumes need to be considered. We dealt with this issue by, first, reducing the complexity of the algorithm from $O(N^3)$ to $O(N^2)$; this is achieved by rewriting the updates in a form suitable for fast matrix modifications. Secondly, we study the thermodynamic limit and monitor some bulk quantities to guarantee a consistent large volume extrapolation.

The topological susceptibility is easy to compute in the IILM and has been studied extensively on the lattice. It represents a natural candidate to fix units in the IILM. For quenched simulations we found rather good agreement between the IILM and the lattice in this way. We are, however, mainly interested in the unquenched case. Instead of using the topological susceptibility directly, we decided to use the chiral condensate, extracted through the topological susceptibility and chiral perturbation theory, to set units. The reason we decided against a direct use of the topological susceptibility is that it depends strongly on the quark masses. We found that the chiral condensate has a very weak dependence on the chiral limit, and use it to set the units in the unquenched sector. We achieve good agreement with previous work and

also with experimental data on the strong coupling constant α_s . Using chiral perturbation theory, we are able to determine physical quark masses. These turn out to compare well with experimental bounds and lattice simulations.

Finally, we investigated the uncertainties introduced by the large volume extrapolations, and found that our procedure, of bounding the volume in such a way that the quark masses are smaller than the smallest Dirac eigenvalues, allows for a systematic thermodynamic limit.

Chapter 4

Biased Monte Carlo

In chapter 3, we have set out to perform simulations in the interacting instanton liquid model (IILM) with ‘physical’ quark masses. Using these input parameters for the finite temperature simulations, we have found that the Monte Carlo simulations become rather inefficient. The reason lies in the formation of instanton–anti-instanton pairs. They provide the mechanism for chiral symmetry restoration within the IILM [99], [100], [158]. In another numerical study [157], no technical problems were encountered because the quark mass parameters were large enough for ordinary Monte Carlo to work well. But as the masses decrease, the interactions become stronger and random sampling starts to run into trouble. Furthermore, in the deconfined phase screening sets in: the temperature fluctuations obstruct the formation of coherent field configurations that exceed the screening length, so that instanton sizes are cutoff at $\rho \lesssim 1/T$. Since the interactions in the IILM follow from the overlap of the instanton profiles, smaller sizes lead to shorter-ranged forces between pairs.

The strong attraction is a generic feature of the IILM at finite temperature. Normalising the low frequency determinant to the dilute gas, the quark effective interaction is of the form

$$S^q = -\ln \det (\mathbb{I} + T^2/M^2) . \quad (4.1)$$

Depending on the topological charge, the determinant is over the subspace spanned by the N_I or N_A zero modes: $T^2 = T^\dagger T$ if $N_I > N_A$, and $T^2 = T T^\dagger$

otherwise. The matrix M^2 is diagonal and of the same dimension as T^2 ; its non-vanishing elements are given by the squared quark mass. Crucially, it is an attractive interaction because the determinant is bounded from below by unity. It is clear from the above expression that the quark interaction becomes ever stronger as quark masses decrease.

The feature of strong and short-ranged interactions is thus not confined to the specific case of the trivial holonomy calorons [93] that we have been studying, but is a general characteristic of non-trivial backgrounds; in particular, it will play a role for the newly found non-trivial holonomy calorons [119, 121] and [123]. The latter might be the correct degrees of freedom to explain the confinement/deconfinement phase transition [80], [53], the lack of which is a major shortcoming of the current IILM.

Systems with strong and localised interactions have been investigated for a long time in chemical engineering and computational chemistry, and are known to present computational challenges. They run under the name of strongly associating, or ionic, fluids. The technical problems that Monte Carlo methods face are two-fold:

- The small relative volume of attraction makes ordinary Monte Carlo updates miss them most of the time.
- Once a pair is formed, ordinary Monte Carlo moves can get stuck in these configurations because of the large energy difference.

From an algorithmic point of view, it means that the acceptance rates to reach or leave the regions of phase space corresponding to instanton–anti-instanton molecules are very low. This would not be a problem if we had an infinite amount of computer time, but, in practice, it leads to very long autocorrelation times. This, in turn, induces a strong dependence on initial conditions and, in the extreme case of very poor mixing between correlated and uncorrelated states, the system might stay fixed in one of those ‘phases’ during the available computer time. For all practical purposes, we lose ergodicity because we end up with samples that are either devoid or dominated by pairs, none of which is a representative sample.

The solution is to develop algorithms which explicitly sample the attraction centres and are able to break up pairs with high acceptance rates. To

sample these volumes, we need to, first, construct these regions and, then, to define a measure over them. This is a non-trivial geometrical problem; it also depends very strongly on the specific problem. Considering the dipole character of the instanton interactions, this seems like a daunting task.

In recent years, general purpose algorithms have been developed which do not rely on an accurate construction of the union of all attraction volumes but on the different ways that a given, simple interaction box can be reached, e.g. the Unbonding–Bonding algorithm [205]. In its original form, it has been given for a canonical ensemble. We will adapt it to our needs for grand canonical simulations.

As argued above, the quark interactions will generically lead to strong interactions for small quark masses, and any non-trivial background will lead to screening effects at finite temperature. To avoid unnecessary complications and background dependent features, e.g. the colour-orientation dependence for calorons, we will use a toy model that mimics the interactions of (4.1).

In section 4.1 we will quickly review the idea behind Markov chain Monte Carlo. We will then discuss biased Monte Carlo in section 4.2, and set up the framework for dealing with strong and short-ranged interactions in the grand canonical ensemble. Finally, we will present a toy model in section 4.3, and benchmark the biased moves from section 4.2 within that setting.

4.1 Markov Chain Monte Carlo

The phase space for a general statistical mechanical system will be very high-dimensional for a large number of particles. Usual integration rules over a grid are not well adapted to evaluate the defining partition function because, for such high dimensional integrals, the number of grid points would be overwhelmingly large.

Markov chain Monte Carlo techniques are the method of choice¹ because they can ‘find’ the relevant regions that dominate the partition function. It is achieved by preferentially sweeping those parts of parameter space that have a large measure. The configurations produced by such algorithms form a Markov chain $\{X_n\}$, whose dynamics is governed by the transition prob-

¹If the integrand is positive definite

abilities $p_{ij} = \mathbb{P}(X_{n+1} = x_j | X_n = x_i)$. The well-developed mathematics of Markov chains allow us to rely on powerful theorems that guarantee convergence to an equilibrium distribution. For these theorems to hold, the Markov chain must satisfy a certain number of assumptions, which we will take for granted in what follows². On a heuristic level, we can use our intuition from classical mechanics to assess whether they are fulfilled; the rationale is that these assumptions correspond to ergodic motion through phase space, i.e. the trajectory, induced by p_{ij}^n , of a fictitious particle must pass through any region of phase space, any number of times. A classic example of a motion that does not meet these requirements is a random walk with drift³.

It can be shown that $p_{ij}^n \rightarrow p_j^{\text{eq}}$, where p^{eq} is the equilibrium distribution. The theorems guarantee that the limit actually corresponds to the unique invariant distribution. Invariance of the equilibrium distribution means that

$$\sum_i p_i^{\text{eq}} p_{ij} = p_j^{\text{eq}}. \quad (4.2)$$

Under these same assumptions, ergodic theorems can be proved. They relate ensemble averages to time averages of paths through phase space, with a specific path given by the Markov chain $\{X_n\}$,

$$\frac{1}{N} \sum_{i=1}^N f(X_i) \rightarrow \langle f \rangle = \sum_{i \in I} p_i^{\text{eq}} f(x_i), \quad (4.3)$$

and are of tremendous practical importance.

In our case, we know p^{eq} , essentially the integrand of the partition function, and we want to construct p_{ij} that converges to this equilibrium distribution. Using the convergence theorems, this is a well posed problem if we demand that p^{eq} is invariant for p_{ij} . It is not hard to see⁴ that (4.2) is

²The assumptions are irreducibility, positive recurrence and aperiodicity, see for instance [138].

³In this case, the Markov chain is not positive recurrent. Strictly speaking, this does not prevent convergence, but the equilibrium distribution will no longer be guaranteed to be invariant, a fact that is crucial for the ergodic theorems and the practical applications that follow from it.

⁴We only need to remember that $\sum_i p_{ji} = 1$, i.e. we go to some state with probability one.

fulfilled if we impose the stronger condition of detailed balance,

$$p_i^{\text{eq}} p_{ij} = p_j^{\text{eq}} p_{ji}. \quad (4.4)$$

This deceptively simple looking equation lies at the heart of all Monte Carlo simulations. Note that the p_{ij} 's are not unique and that there is a large amount of freedom in choosing them. This redundancy can be used to accelerate convergence.

If there are different paths that connect states i and j , $p_{ij} = \sum_a^N p_{ij}^a$, (4.4) can be met by imposing the stronger condition

$$p_i^{\text{eq}} p_{ij}^a = p_j^{\text{eq}} p_{ji}^a. \quad (4.5)$$

This is sometimes called super-detailed-balance [72]. If there are very many different paths, $N \gg 1$, the implementation can become rather tedious, and super-detailed-balance might be the only viable option. Even if there are a manageable number of paths, the book-keeping needed to relate them might be overwhelming.

It is important to note that the transition probability is really the product of the proposal probability \mathcal{P}_{ij} and the acceptance probability \mathcal{A}_{ij} . This observation lies at the heart of biased Monte Carlo techniques: we can tweak the proposal probabilities to increase acceptance rates. Typically, this leads to asymmetric transition probabilities. The latter can, however, also be found in ordinary Monte Carlo: consider for instance a simple insertion/deletion step for grand canonical simulations; the proposal probability for an insertion corresponds to the probability to place the particle within the simulation box, whereas for a deletion it gives the probability to choose a particle already in the box. Clearly, these will be different in general.

Finally, let us mention that there are different prescriptions to satisfy (4.4). The most straightforward is the heatbath algorithm. It consists of using as transition probabilities the equilibrium distribution itself, i.e. $p_{ij} = p_j^{\text{eq}}$. It clearly satisfies (4.4). This approach only works in very simple cases, but for those it usually converges faster than alternative algorithms. On a heuristic level, this seems obvious since we take full advantage of our knowledge of p^{eq} . A general purpose algorithm, and the one we will focus on, is the

Metropolis algorithm [72], which defines the acceptance probability by

$$\mathcal{A}_{ij} = \min \left(1, \frac{p_j^{\text{eq}} \mathcal{P}_{ji}}{p_i^{\text{eq}} \mathcal{P}_{ij}} \right). \quad (4.6)$$

It is not hard to see that it satisfies (4.4).

4.2 Biased Monte Carlo

But for the simplest statistical systems, ordinary Monte Carlo techniques can become rather inefficient, if not downright inadequate: inefficiency manifests itself by long autocorrelation times, and inadequacy stems from an inability to sample the relevant regions accurately. For finite computer time, and hence finite-sized samples, we basically lose ergodicity.

By ‘ordinary’ Monte Carlo moves we mean random sampling. A typical example is updating the position of a particle: it is straightforward to implement such a move by adding a random increment to the particle’s current position. This random increment is typically drawn from a fixed volume with uniform measure. Thus, \mathcal{P}_{ij} is constant for all updates and cancels in (4.6).

For strong and short-ranged interaction, such a random increment will be very inefficient. Either we choose it so small that we can sample the interaction region, in which case very many sweeps are needed to move through phase space; this leads to very long autocorrelation times. Or we choose a large enough increment, so as to sweep quickly through phase space, but thereby missing the interaction regions most of the time; the sample will most likely not be representative. Furthermore, the acceptance probability for random sampling is typically given by the ratio of two Boltzmann factors, and does only depend on the energy difference. Due to the strong interaction, the acceptance probability will be very low for moves that attempt to leave the attraction centres. The Markov chain can become trapped in these energy-dominated configurations, which leads again to a non-representative sample.

The way out is to use importance sampling. It is designed to sample those parts of phase space that dominate the partition function. It might help to get a rough criterion for importance sampling. Remember that in our

case Monte Carlo methods try to evaluate the integral given by the partition function. If the integrand is peaked so strongly that the exact integral is well approximated by the integral over some localised patches, the algorithm will need to preferentially sample these parts of phase space. Thus, random sampling will be inefficient if

$$V < \Delta V \exp(-H(\Delta V)), \quad (4.7)$$

where V is the volume of the simulation box, and $\Delta V \ll V$ is the small region where the interaction is very strong.

As mentioned in the introduction, in chemical engineering and computational chemistry the issues of low acceptance rates have long been known: they are very low in polymer physics simulations, due to conformational obstructions; or in simulations of ionic fluids, due to strong short-ranged interactions. The latter is of immediate interest to us because the IILM at finite temperature displays the same characteristics. To treat such systems correctly, new algorithms based on biased Monte Carlo have been developed [194]. We can take advantage of these well tested techniques in IILM simulations at finite temperature.

Recently, efficient and general purpose algorithms have been developed [34] [205]. In these algorithms, the focus is not on an accurate construction of the union of all the interaction regions, which is a difficult, and problem-dependent, geometric task, but on the individual interaction regions and all the possible routes that lead to the same final state. In [34], the algorithm is further simplified by using super-detailed-balance (4.5), whereas the algorithm in [205] does not rely on this stronger condition, and was shown to converge faster. We use the latter scheme, the Unbonding–Bonding algorithm (UB).

The UB algorithm starts by defining a bonding region. It does not necessarily need to be the exact physical bonding region, but a strong departure from it will not be very useful. A list is made of those instantons that are in at least one bonding region of an anti-instanton and the same is done for the anti-instantons. There are N_I^B (N_A^B) bonded instantons (anti-instantons). $N_I^B(i_I)$ is the number of anti-instantons that instanton i_I is bonded to, and analogously for $N_B^B(i_A)$. In what follows, i_I means instanton i_I , but can also

stand for the state the instanton i_I is in; unprimed quantities are evaluated before the move, whereas primed ones denote the same quantity after the move.

We will now focus on the UB for instantons; the case for anti-instantons is then obvious. The bonding move consists of choosing uniformly an instanton and an anti-instanton, and placing the instanton in the bonding region of the anti-instanton with flat measure. The unbonding move consists of choosing one of the bonded instantons and to place it randomly in the simulation box; again both steps are performed with a flat measure for the bonded instantons and the simulation box respectively. This leads to the following transition probabilities

$$\mathcal{P}_{i_I(i'_I, i_A)}^B = \frac{1}{N_I} \frac{1}{N_A} \frac{1}{V_{i_A}}, \quad (4.8)$$

$$\mathcal{P}_{i i'}^U = \frac{1}{N_I^B} \frac{1}{V}, \quad (4.9)$$

where (i_I, i_A) is a bonded instanton–anti-instantion pair, and V_{i_A} is the bonding region of anti-instanton i_A . The UB algorithm now adds up all possible routes that lead to the same final state i' , bonding and unbonding; super-detailed-balance is, however, used with respect to the unbiased displacement move. The forward and backward proposal probabilities are then given by

$$\mathcal{P}_{i_I i'_I} = \sum_{i_A}^{N_I'^B(i_I)} \mathcal{P}_{i_I(i'_I, i_A)}^B + \delta_{i_I}^B \mathcal{P}_{i_I i'_I}^U, \quad (4.10)$$

$$\mathcal{P}_{i'_I i_I} = \sum_{i_A}^{N_I^B(i_I)} \mathcal{P}_{i'_I(i_I, i_A)}^B + \delta_{i'_I}^B \mathcal{P}_{i'_I i_I}^U, \quad (4.11)$$

with $\delta_i^B = 1$ if i is bonded and $\delta_i^B = 0$ otherwise. We assume that bonding and unbonding moves have the same a-priori-probability, $1/2$, which we omitted because it cancels out anyway.

Since we want to perform grand canonical simulations, we also need in-

sertion and deletion moves. The unbiased moves are given by

$$\mathcal{P}_{N_I, N_I+1}^{\text{ub}} = \frac{1}{V}, \quad (4.12)$$

$$\mathcal{P}_{N_I+1, N_I}^{\text{ub}} = \frac{1}{N_I + 1}. \quad (4.13)$$

The corresponding biased insertions and deletions will be constructed along the lines of the UB algorithm, either by placing the instanton i_I into the bonding region of an anti-instanton, or by removing the bonded instanton i_I . More precisely, insertions consist of choosing an anti-instanton i_A , placing the instanton i_I uniformly within the bonding box V_{i_A} and finally summing over all possible anti-instantons that could have been chosen to reach this same final state; to delete an instanton, we choose uniformly from the list of bonded instantons N_I^B . In formulas,

$$\mathcal{P}_{N_I, N_I+1}^{\text{b}} = \sum_{i_A}^{N_I^B(i_I)} \frac{1}{N_A} \frac{1}{V_{i_A}}, \quad (4.14)$$

$$\mathcal{P}_{N_I+1, N_I}^{\text{b}} = \frac{\delta_i^B}{N_I^B}. \quad (4.15)$$

In this case, it does not produce much overhead to combine the biased and unbiased moves. More importantly, we found that acceptance rates could be boosted upon combining biased and unbiased insertions/deletions. To add them up, we need to specify the relative weights, the a-priori-probability for biased updates p_b . The full insertion/deletion proposal probabilities are

$$\mathcal{P}_{N_I, N_I+1} = p_b \mathcal{P}_{N_I, N_I+1}^{\text{b}} + (1 - p_b) \mathcal{P}_{N_I, N_I+1}^{\text{ub}}, \quad (4.16)$$

$$\mathcal{P}_{N_I+1, N_I} = p_b \mathcal{P}_{N_I+1, N_I}^{\text{b}} + (1 - p_b) \mathcal{P}_{N_I+1, N_I}^{\text{ub}}. \quad (4.17)$$

To reiterate, a similar factor for the a-priori-probabilities of canonical moves was tacitly omitted before because we chose to follow the original implementation of the UB algorithm, which does not mix biased and unbiased particle updates: its use of super-detailed-balance implies that the a-priori-probabilities give an overall multiplicative factor that drops out.

Other than good mixing between bonded and unbonded structures, cluster moves have been argued to be important to accelerate convergence to-

wards the equilibrium distribution [142]. For the IILM, we assume that instanton–anti-instanton pairs are dominating in this respect. We therefore augment our list of single particle moves by pair-displacements and pair-insertions and -deletions.

Whereas we made sure to have a uniform distribution among the N_I^B bonded instantons for the UB-type moves discussed so far, regardless of whether they were bonded many times, we naturally demand uniformity in the number of pairs N_P for the pair-moves.

The pair-displacements consist of collective translations of the pair, by an increment $\delta \in v_C$, and of internal displacements, $\delta \in v_I$, of one of the pair’s constituents. The latter are rejected if the displaced instanton leaves the bonding box. The proposal probabilities for these moves are given by

$$\mathcal{P}^C = \frac{1}{N_P} \frac{1}{v_C}, \quad (4.18)$$

$$\mathcal{P}^I = \frac{1}{N_P} \frac{1}{v_I}, \quad (4.19)$$

We use super-detailed-balance for the internal displacement because the acceptance rates could not be boosted by including unbiased and/or UB moves; the additional overhead, required to go beyond super-detailed-balance, is thus not justified.

The pair-insertions and -deletions, set up without recourse to super-detailed-balance, are built from biased and unbiased moves, whose probabilities are summed up in the end. A biased pair-insertion consists of placing, uniformly, either an instanton i_I or an anti-instanton i_A in the simulation box; the partner is then positioned randomly within the bonding box. Note that the probabilities for both possibilities are added up. The unbiased pair-insertion consists of placing, uniformly, both an instanton i_I and an anti-instanton i_A in the simulation box. For a biased pair-deletion we select an instanton and anti-instanton uniformly from the pairs N_P , whereas for an unbiased pair-deletion we select an instanton and anti-instanton randomly from N_I and N_A respectively. The final proposal probabilities are then given

by

$$\mathcal{P}_{N,N+1} = p'_b \delta_{i_I i_A}^P \left(\frac{1}{2} \frac{1}{V} \frac{1}{V_{i_I}} + \frac{1}{2} \frac{1}{V} \frac{1}{V_{i_A}} \right) + (1 - p'_b) \frac{1}{V} \frac{1}{V}, \quad (4.20)$$

$$\mathcal{P}_{N+1,N} = p'_b \frac{\delta_{i_I i_A}^P}{N'_P} + (1 - p'_b) \frac{1}{N'_I} \frac{1}{N'_A}. \quad (4.21)$$

The factor $\delta_{i_I i_A}^P$ ensures that the biased contribution to the proposal probability is added only if i_I and i_A are paired, i.e. $\delta_{i_I i_A}^P = 1$ if instanton i_I and anti-instanton i_A are paired and $\delta_{i_I i_A}^P = 0$ otherwise. The a-priori-probabilities p'_b can be chosen to tune the acceptance rates even further.

Given these different proposal probabilities, together with the trivial case of unbiased displacements, we can compute the acceptance probability through (4.6).

Finally, we need to specify the increment volumes v_i and the bonding boxes V_i , and we have to decide upon the different a-priori-probabilities p_b , p'_b and p_i . The latter give the probability to perform a specific update and have, so far, been omitted because they will always cancel in the acceptance probabilities. These choices will depend on the problem at hand, and some fine-tuning runs cannot be avoided to set these parameters.

In practice, we have finite-sized samples, and the outcome of the simulations will depend on these parameters. The differences will vanish with increasing sample size; this observation provides a straightforward means to test whether the biased updates have been implemented correctly. For computer-intensive simulations, large samples are prohibitive, and it is important that the dependence is rather weak if the algorithms are to be useful.

4.3 Toy Model

Now that we have set up the update moves using biased Monte Carlo techniques, we will test their performance compared to random sampling. We will use a two-dimensional toy model that nevertheless will mimic the IILM. The advantage of the toy model is that it will be computationally very cheap. Also, we can adjust the parameters freely to get a rough estimate for those regions in parameter space where pair formation will be important and, hence,

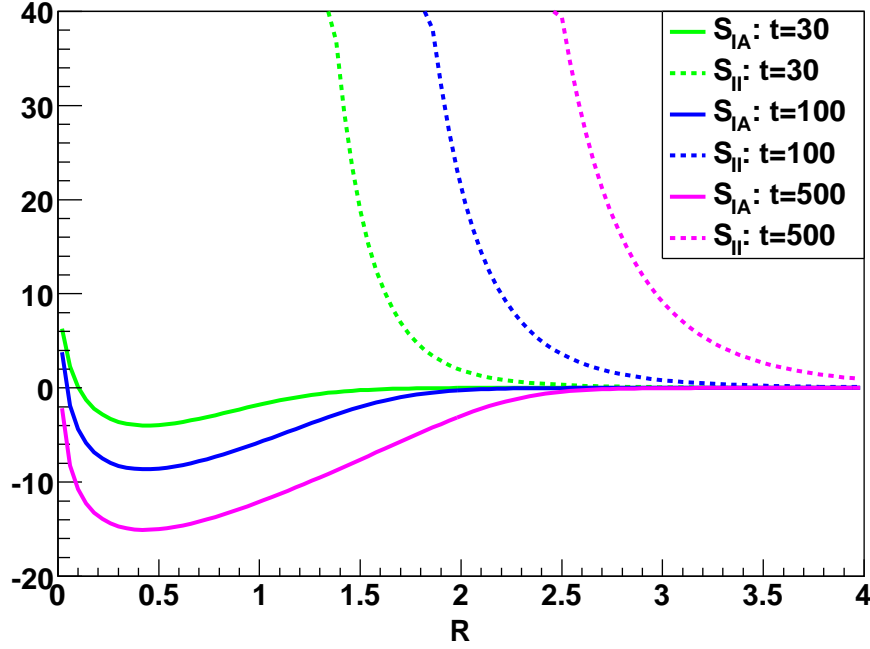


Figure 4.1: The solid lines represent the interaction for an instanton–anti-instanton pair. The repulsion for like-charged pairs is represented by dashed lines. It is chosen so strong that clusters larger than simple pairs are strongly disfavoured. We see that as the temperature t is raised the attractive well deepens. In this graph $\alpha = 2$.

biased Monte Carlo essential.

The interactions in the IILM have the following form

$$S_{IA}^{\text{int}} = \ln \left(1 + \frac{1}{R^2} \right) - \frac{1}{1 + R^2} - N_f \ln \left(1 + e^{-2\pi R} R^2 \frac{T^2}{m^2} \right), \quad (4.22)$$

$$S_{II}^{\text{int}} = \ln \left(1 + \frac{1}{R^2} \right) - \frac{1}{1 + R^2}. \quad (4.23)$$

The logarithmic repulsion in the pair separation R , given here in units of the inverse temperature T , is typical for the IILM⁵. It combines with the rational function to produce the large separation decay. The instanton–anti-instanton pairs feel an additional attraction due to the quark wavefunction overlaps. Their R^2 dependence follows from ‘strong’ overlaps, whereas the exponential describes the large separation behaviour. For our purposes, the product of

⁵In the ratio ansatz, see chapter 3.

both will be a good enough approximation. The quark mass is given by m , the number of flavours by N_f . As for the IILM, S_{II} and S_{IA} have the same fall-off behaviour.

In two dimensions, this isotropic interaction results in a rather dense ensemble because the repulsion is mild and the attractive wells are comparatively deep. More importantly, these interactions will favour large clusters of instantons and anti-instantons. Such a behaviour does not correspond to the situation that we expect from the IILM, namely the formation of rather isolated pairs. In the IILM, the interaction is orientation dependent, and large conglomerates of instantons cannot form because most of the relative orientations within the cluster would be repulsive. To achieve this same effect of pair formation with our toy model, we add an extra ‘entropic’ repulsion term to S_{int}^{II} .

The exponential decay for the quark overlaps starts for $e^{-2\pi R t^2} \approx 1$; we have set $t = T/m$, i.e. the temperature in units of the quark mass. For separations smaller than $R_0 = \ln t/\pi$, the attraction wells are rather deep, and we choose the repulsion to become strong in order to obstruct the formation of large clusters. The functional form of this ‘entropic’ repulsion is rather unimportant, and for simplicity we choose it to be given by

$$S_{II}^{\text{ent}} = \left(\frac{\alpha R_0}{R} \right)^8. \quad (4.24)$$

It turns out that choosing $\alpha = 2-3$ is sufficient to prevent cluster formation. For dilute ensembles, the precise value is rather irrelevant.

Now that we have fixed the interactions, see also Fig. 4.1, we define our system by the following partition function

$$Z = \sum_{N_I, N_A}^{\infty} \int dx^{N_I+N_A} \frac{d^{N_I}}{N_I!} \frac{d^{N_A}}{N_A!} \exp(-S_{\text{int}}(x)), \quad (4.25)$$

$$S_{\text{int}} = \sum_{i < j}^{N_I} S_{ij}^{\text{int}} + S_{ij}^{\text{ent}} + \sum_{i < j}^{N_A} S_{ij}^{\text{int}} + S_{ij}^{\text{ent}} + \sum_{i_I i_A}^{N_I, N_A} S_{i_I i_A}^{\text{int}}, \quad (4.26)$$

where the free parameters of the model are d , t and the simulation box V . In the full model, instantons have a size ρ , and quantum effects are such that

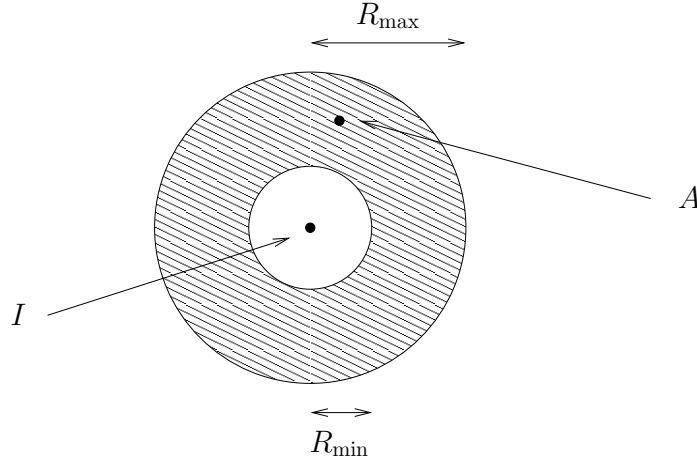


Figure 4.2: For the biased Monte Carlo moves to work well, the bonding box should be adapted to the interactions. Isotropy suggests the best form to be in the shape of an annulus. The shaded region corresponds to the bonding box. A priori R_{\min} and R_{\max} are free parameters that can be fine tuned to achieve good mixing, i.e. low autocorrelation times.

small sizes are disfavoured, $d(\rho) = \rho^\beta$ with $\beta > 0$ [182]; the free parameter d plays that role.

Neglecting interactions, d determines the density of the ensemble,

$$\frac{\langle N \rangle}{V} = 2d. \quad (4.27)$$

This is the dilute gas result. Once interactions are included, the system will be more dilute if the total interaction is repulsive and less dilute otherwise; this is adjusted by the parameter t which captures the combined effects of temperature and quark masses, see (4.22).

In order to use the biased Monte Carlo framework set up in section 4.2, we need to decide on the bonding box to use. Given that the interaction is isotropic, we will choose an annulus as bonding box, see Fig. 4.2. It is natural to fix the free parameters R_{\min} and R_{\max} so as to achieve low autocorrelation times.

We perform biased simulations for three sets of bonding boxes, and compare them with unbiased runs. The natural box size extends to $R_{\max} = R_0$, the point at which the quark interaction starts to vanish exponentially. The

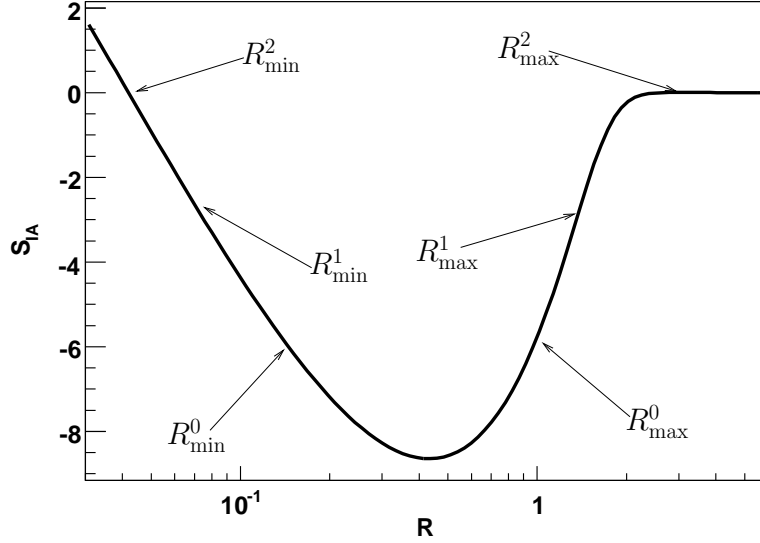


Figure 4.3: We choose three different sets for (R_{\min}^i, R_{\max}^i) . A small interval that samples the strong interaction region, a ‘natural’ one that sample most of the attractive well and a larger box that extends beyond what would naturally look like the bonding region.

corresponding lower edge R_{\min} is determined by $S_{IA}^{\text{int}}(R_{\min}) \approx S_{IA}^{\text{int}}(R_{\max})$. Given the relatively large values for t , we can solve this approximately by

$$R_{\min} \approx \exp \left(-\frac{1}{6} S_{IA}^{\text{int}}(R_{\max}) - \frac{2}{3} \ln t \right). \quad (4.28)$$

We also choose a larger interval with $R_{\max} = 2R_0$ and a smaller one that samples predominantly the very strong interaction region. In the latter case, we choose R_{\min} and R_{\max} according to $S_{IA}^{\text{int}}(R_{\max}) \approx S_{IA}^{\text{int}}(R_{\min}) \approx \frac{1}{3} \min_R S_{IA}^{\text{int}}$. A typical setup is displayed in Fig. 4.3.

Before we start the simulations, it is worth mentioning one more technical point, related to the choice for the equilibrium probability density. We claim that it is given by

$$p^{\text{eq}} = \frac{d^{N_I} d^{N_A} e^{-S_{\text{int}}}}{Z}. \quad (4.29)$$

The slightly subtle point we want to raise is that in (4.29) we neglect the Boltzmann-counting factors for indistinguishable particles. We would like to argue that they are an artifact. These terms are introduced so as to allow for

an unconstrained integration region, $x_i^a \in [0, L^a]$, instead of the complicated multi-dimensional region given by $\{x_i^a | x_1^a < x_2^a \cdots < x_{N_j}^a\}$, with a labelling a component. The result of the integrations is corrected by dividing by the factorials because $\cup_{\sigma} \{x_i^a | x_{\sigma(1)}^a < x_{\sigma(2)}^a \cdots < x_{\sigma(N_j)}^a\} = [0, L^a]^{N_j}$, where σ is a permutation. Therefore, the Boltzmann-counting factors are not really part of the weight [1]. A different, but equivalent, point of view is to keep the factorials in the weight and to sum the equilibrium probability, with the factorials included, over all permutations $N_i!$ that correspond to mere relabellings of the particles in the initial state [156]. The permutations cancel off the factorials, and we are left again with (4.29). The transition probabilities P_{ij} are similarly summed over all initial and final state permutations, and the resulting factor of $N_i!N_j!$ trivially cancels out of (4.4). The simplest way to understand the need to sum over permutations, lack of which will violate detailed balance, is to realise that, taking indistinguishability at face value, the states are really equivalence classes⁶. Since we actually use a specific representative but the result should be independent of it, we need to sum over all members of the given equivalence class. The upshot is that, for all practical purposes, we treat the different particles as distinguishable!

The parameter t is set to the values given in Fig. 4.1. The dilution parameter d is chosen small enough so that we end up with a rather dilute system because the biased Monte Carlo moves, set up in section 4.2, will not work well for dense ensembles. We are interested in how the system behaves if d is further decreased. We can envisage three different equilibrium states: either the ensemble will be dominated by pairs, or equilibrates in a mixture of paired and unpaired instantons, with roughly equal weight, or settles in an uncorrelated state with no pairs.

For the simulations that follow, we use the ‘natural’ bonding box, $R_{\max} = R_0$. We always use hot initial conditions, i.e. place the instantons and anti-instantons randomly throughout the volume. As we can see from Fig. 4.4, the smallest value of t leads to equivalent equilibrium states for both biased and unbiased Monte Carlo. However, as we increase t , and lower d in order to maintain a dilute ensemble, ordinary Monte Carlo needs a very long time to reach equilibrium compared to biased Monte Carlo, see Fig. 4.5. Since

⁶The group by which we factor out is the permutation group

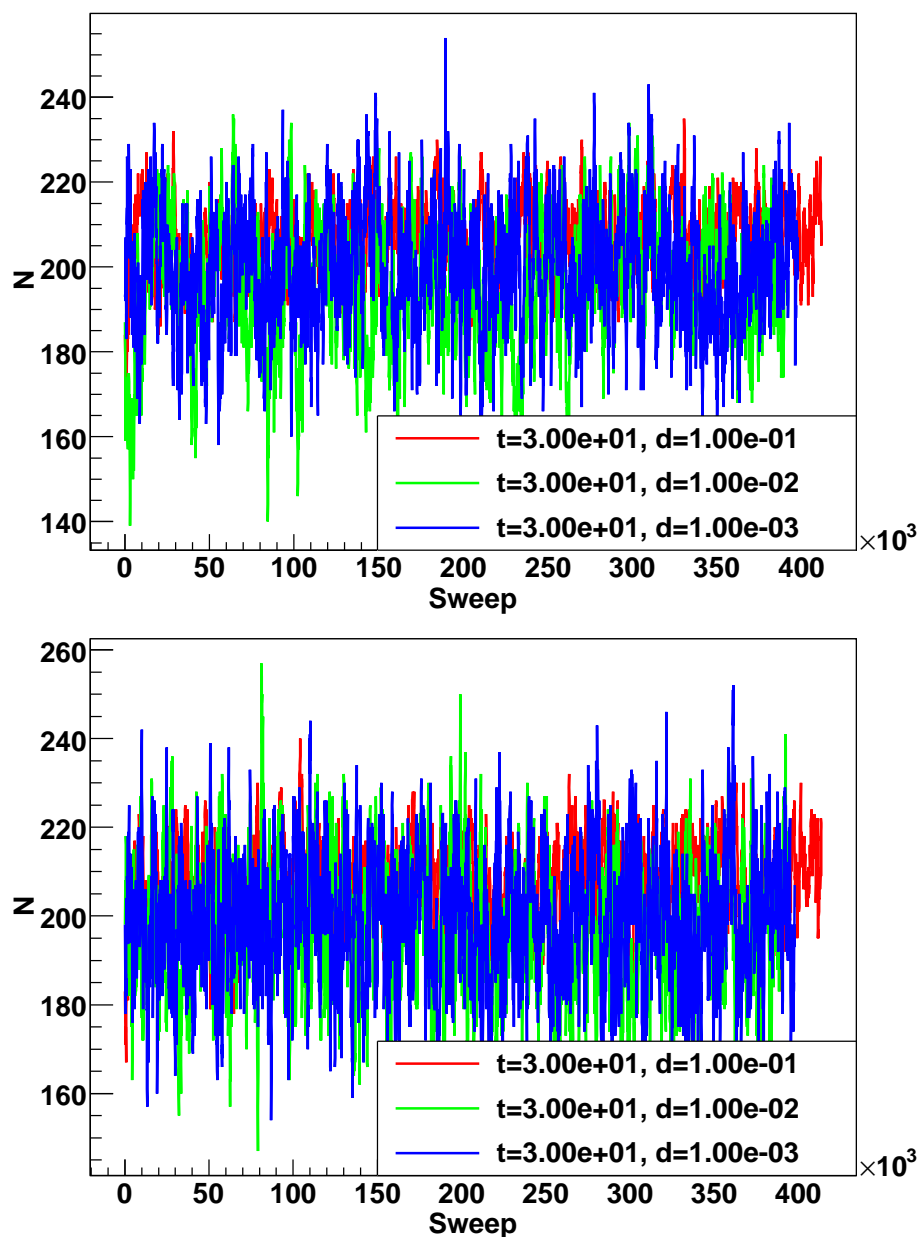


Figure 4.4: Choosing a rather small parameter for t , i.e. a shallow attraction well, the biased (bottom) and unbiased (top) simulations have thermalised over the same time scale. The overhead that importance sampling introduces is not necessary provided that the autocorrelation times are similar. In the present case we found that this is indeed so.

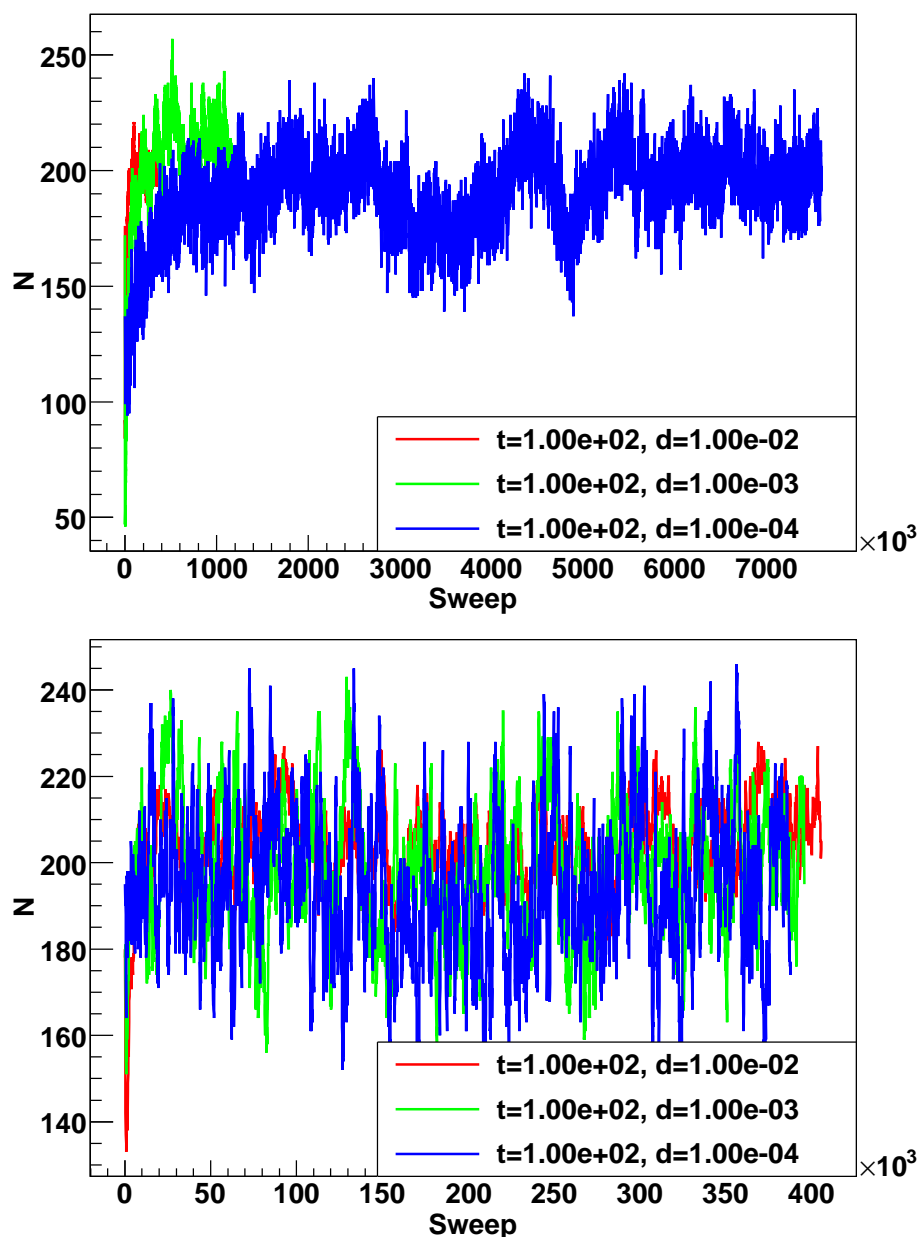


Figure 4.5: As the interaction well is deepened, i.e. for larger t , unbiased Monte Carlo algorithms (top) start to run into trouble. The thermalisation process takes longer and longer as we increase the dilution of the system, i.e. by decreasing d . Importance sampling (bottom) thermalises much faster: notice the rather large difference in the number of sweeps it takes to reach equilibrium for the blue ($d = 10^{-4}$) curve.

MC	d	$\langle N \rangle$	ξ
biased	10^{-2}	202.4(2)	7940
	10^{-3}	199.7(4)	7302
	10^{-4}	193.1(4)	2857
unbiased	10^{-2}	202.0(1)	41058
	10^{-3}	200.2(1)	75115
	10^{-4}	196.8(2)	140683

Table 4.1: We estimate the autocorrelation times ξ for the instanton number N as a function of the Monte Carlo algorithm and the dilution parameter d . The temperature/mass parameter is $t = 100$, i.e. it corresponds to the data in Fig. 4.5. We clearly see that as the density of the system is lowered, i.e. by decreasing d , the autocorrelation times for unbiased simulations become very large compared to the unbiased ones.

we start with a random distribution, the long thermalisation process can be attributed to the fact that it takes longer and longer to locate the attraction centres as the system becomes ever more dilute. The correlation between the number of instantons N and the number of pairs P is clearly visible in Fig. 4.6.

After equilibration, we can estimate autocorrelation times. We find that the autocorrelation time in the instanton number N is much larger for random sampling simulations, see Table 4.1. The reason is that random insertions and deletions are suppressed: insertions because they generally do not form pairs, and deletions because they most likely try to break up pairs; this leads to low acceptance probabilities.

The toy-model is computationally rather cheap, which allows for very long thermalisation sweeps. During such a long equilibration process, helped by the low dimensional configuration space, random sampling does eventually form sufficient pairs to converge to the same state as importance sampling. In higher dimensional simulations, as for instance in the IILM, the interaction regions have much lower ‘entropy’⁷, and it will be harder for ordinary Monte Carlo simulations to find the regions of phase space that lead to pair formation. Also, realistic systems are computationally more expensive⁸, and long

⁷By which we mean that $\Delta V/V \ll 1$, with ΔV the bonding volume.

⁸For instance, the IILM in the unquenched case needs to evaluate determinants, a

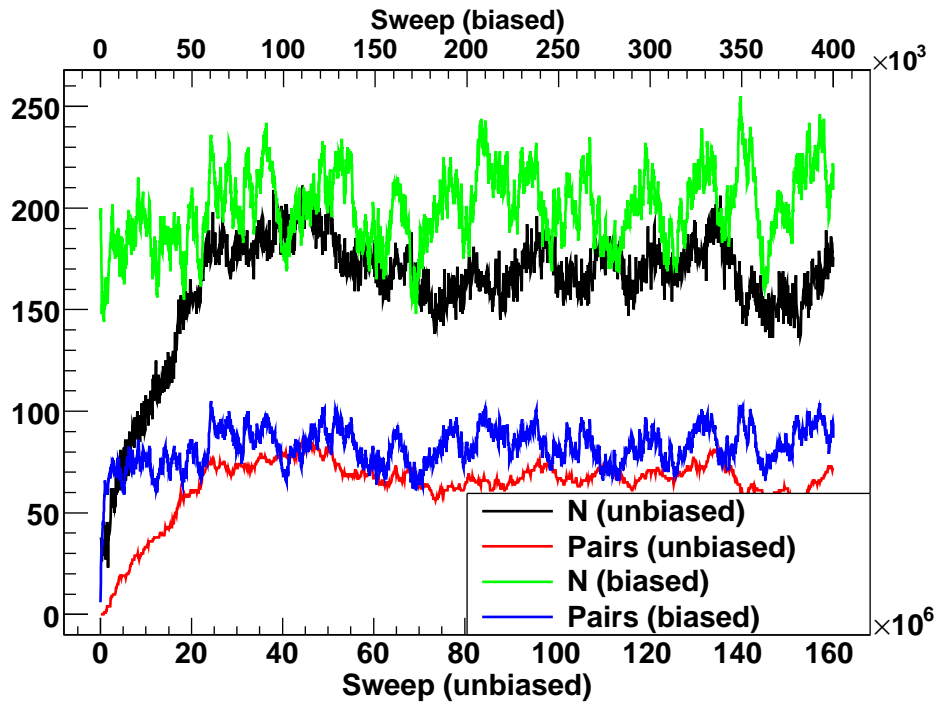


Figure 4.6: The long thermalisation time is directly correlated with the slow process of forming pairs in the unbiased case. The correlation between pairs and instanton number is also seen in the biased simulations; here thermalisation is fast (note the different scale, i.e. the upper axis) because the algorithm explicitly samples the possible bonding sites.

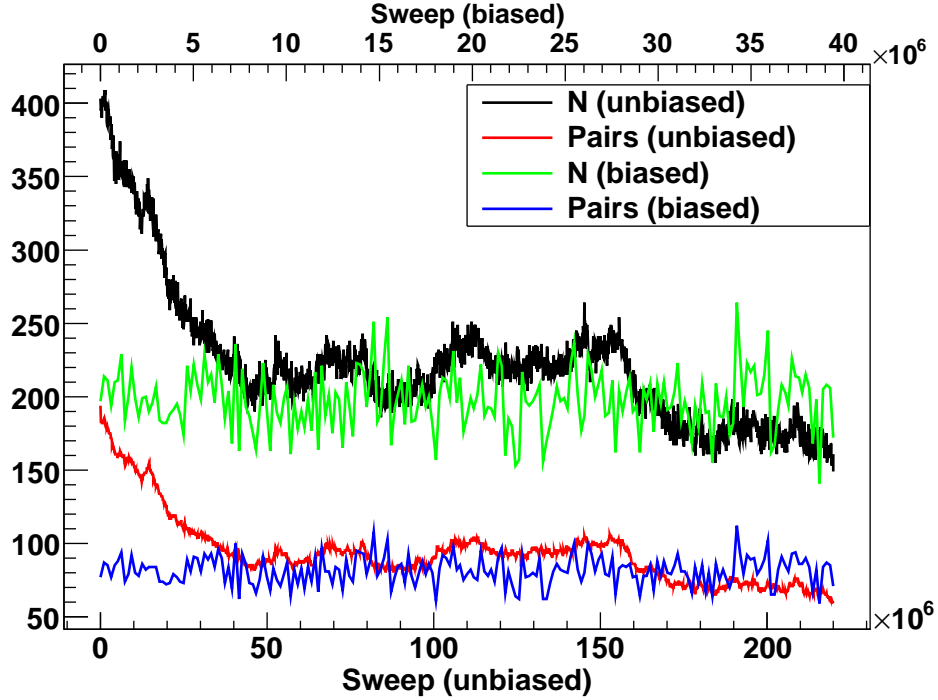


Figure 4.7: Random sampling is not well suited to deal with cold initial conditions: we started the thermalisation run with a configuration made up of 200 pairs. Random sampling has trouble to break up the excess pairs. The inefficiency can be traced back to the fact that the algorithm depends solely on the energy difference, which is big when destroying pairs in a ‘naive’ way. This leads to low acceptance probabilities and hence long thermalisation sweeps. The biased simulations equilibrate very quickly because the algorithm is tailored to deal with bonded instantons.

runs are prohibitive. Biased Monte Carlo techniques become unavoidable in these situations.

The hot initial condition manifests itself by a rapid drop in instanton numbers early in the thermalisation process, due to the lack of pairs; this can even be seen in the case of biased simulations, although the algorithm creates pair much more quickly, see Fig. 4.5. We expect that cold initial conditions, i.e. starting off with pairs, will also be problematic for random sampling techniques because they rely solely on the energy difference between states: in see Fig. 4.7 we examine how ordinary Monte Carlo copes with an ensemble with excess pairs.

time-consuming task.

From Figs. 4.5 and 4.7 it is clear that biased Monte Carlo techniques vastly outperform random sampling. The computational advantage follows from our knowledge of the structures that can form, e.g. instanton–anti-instanton molecules in the IILM [99], and the ability to construct algorithms that take full advantage of that knowledge.

So far we have restricted the simulations to a single bonding box, which is a free parameter a priori. We have run every data set with the two other bonding boxes, see Fig. 4.3. We have found that the results agree on the 2σ level. In Table 4.2, we show such data for one specific point in parameters space $\{d, t\}$. This particular set, with large t and very small d , is probably representative for more realistic simulations in that the unbiased simulations have failed to converge to the true equilibrium state; we have lost ergodicity because those regions of phase space that correspond to bonded pairs is not sampled adequately.

The weak dependence on the bonding box allows us to dismiss systematic effects, and we can base our choice entirely on achieving low autocorrelation times. The data presented in Table 4.2 would suggest that the two smallest bonding boxes are equivalent but the bulk of the computations suggest the lowest autocorrelation times are achieved with the smallest bonding box.

It is interesting to see how the ensemble behaves as we tune the two free parameters d and t separately. We have found that the system will always evolve to a random state with a negligible amount of pairs upon decreasing d for any fixed t . This is as it should be because small d favour a dilute system: from (4.27) we see that V increases for decreasing d and constant number of instantons; the strong interaction region ΔV is unaffected by d , and the inequality (4.7) will be violated for large enough V ; thus, random sampling is efficient and the ensemble equilibrates in an uncorrelated state. We illustrate this in Fig. 4.8.

Similarly, we find that for fixed dilution d and increasingly large interaction strength t the system tends to favour energy-dominated configurations. We therefore expect that the topological susceptibility should drop to zero since pairs do not contribute to the charge fluctuations. The instanton density, on the other hand, increases because the energy-dominated configurations favour a denser ensemble. This in turn implies that the topo-

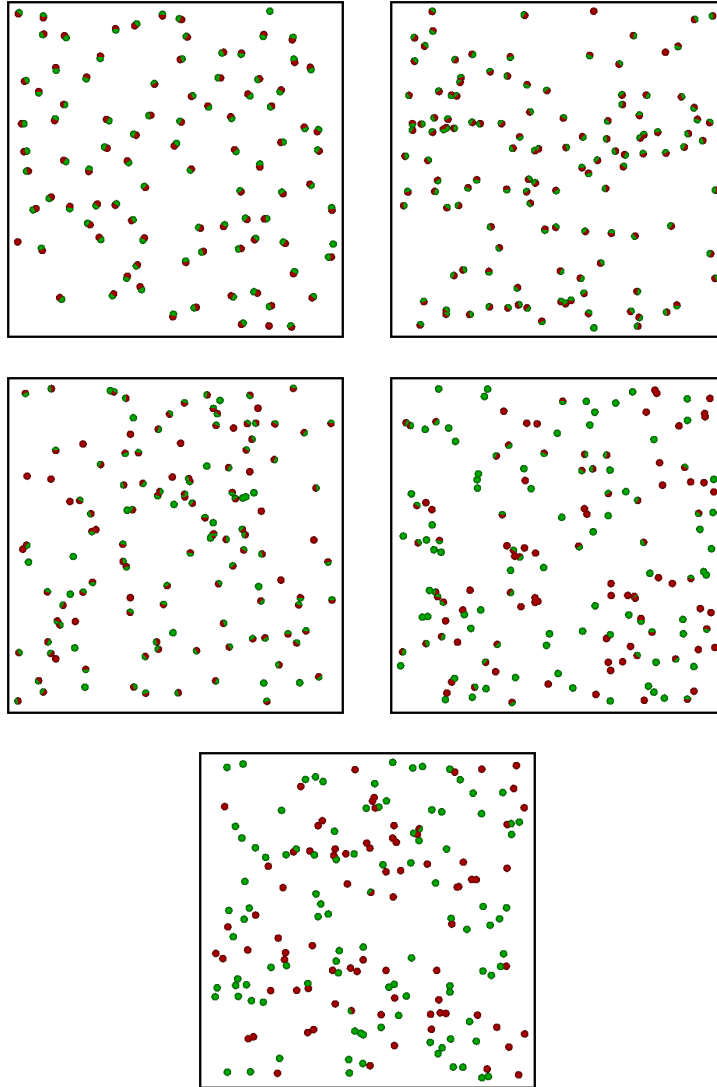


Figure 4.8: The degree of dilution increases from left to right, top to bottom; the temperature-mass parameter is fixed at $t = 300$. As d is decreased fewer instantons are paired up with anti-instantons. For the last box the system has equilibrated in a random state.

Bonding Box (R_0)	$\langle N \rangle$	$\langle Q^2 \rangle$	$\langle S_{\text{int}} \rangle$
(0.102, 0.66)	194.6(4) [1998]	134(4) [1790]	-2.323(5) [1083]
(0.046, 1.00)	194.7(4) [1907]	133(3) [1212]	-2.328(7) [1812]
(0.031, 2.00)	195.5(3) [4629]	137(3) [2607]	-2.340(7) [5475]
unbiased	131.8(3) [887]	129(5) [322]	$-0.4(6) 10^{-5}$ [545]

Table 4.2: This data is from a simulation with $t = 500$ and $d = 10^{-7}$, and for these parameters the unbiased Monte Carlo simulation equilibrates at a substantially lower value for the total number of instantons N . Note also the low interaction, from which we conclude that no pairs have formed. The results for the different bonding boxes agree well at the 2σ level. The autocorrelations are given in square brackets. The two smallest boxes give rather similar autocorrelation times but we have found that in most of the runs the smallest box leads to the fastest convergence.

logical susceptibility will be boosted because the topological fluctuations per unit volume have increased. Both effects work in opposite directions, but it turns out that the topological susceptibility decreases overall, although rather slowly. This is demonstrated in Fig. 4.9. In particular, the topological susceptibility is lower than the dilute gas approximation would suggest.

In QCD, the level of dilution and the interaction strength change simultaneously, such as to increase both with increasing temperature. Therefore, we cannot draw any direct conclusions from this work, other than to note that with increasing fermion number the screening, and hence the dilution, decreases whereas the fermionic interactions increase. It is therefore not clear a priori whether in QCD the IILM will remain in a highly correlated molecular phase after the chiral phase transition.

4.4 Conclusions

We have argued that, quite generically, light quarks in non-trivial backgrounds will induce strong and short-ranged interactions at finite temperature. The Dirac operator is positive definite and bounded, and leads to an

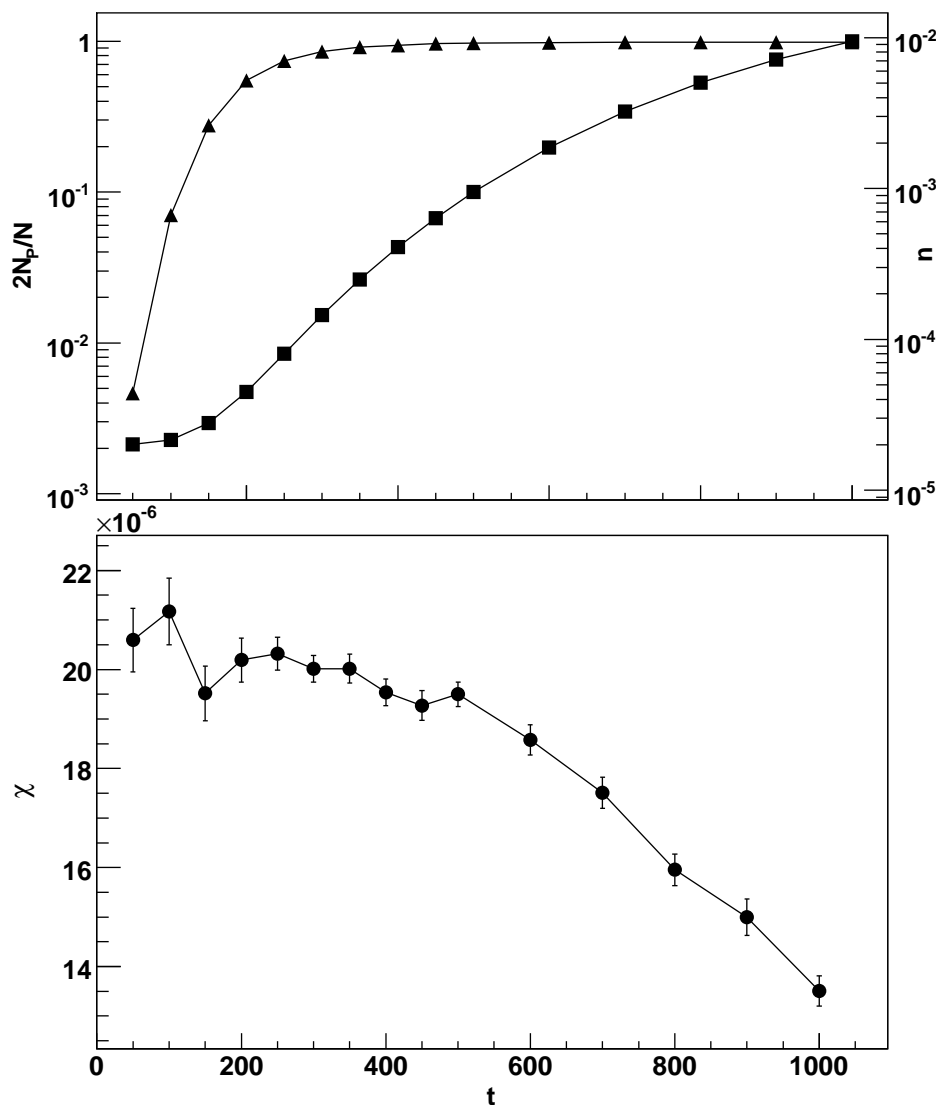


Figure 4.9: As the interaction strength t is increased, the proportion of pairs to instantons increases until basically all instantons are in pairs. Pairs do not contribute to topological fluctuations, reducing the topological susceptibility. However, the instanton density increases rather strongly with t , resulting in a much denser ensemble. This implies a larger topological susceptibility, because the fluctuations per unit volume increase. Both effects are clearly antagonistic. In this case, it turns out that the topological susceptibility decreases, although rather slowly.

attraction between instantons that becomes stronger with decreasing quark mass. Plasma effects at finite temperature constrain the spatial extent of the non-trivial backgrounds to be below the screening length; since the quark interactions are induced by the overlap of quark wavefunctions, centred on top of the classical gauge fields, it follows that the interactions become more and more short-ranged as the temperature rises.

Systems with strong and short-ranged interactions are known as strongly associating fluids in the literature of computational chemistry and chemical engineering. These fluids are hard to simulate and need special techniques. Using the concepts of the general purpose Unbonding–Bonding algorithm, we extended the biased scheme to grand canonical simulations. Following suggestions from the literature, we augmented the Monte Carlo steps by pair moves. The input from previous studies of the IILM is crucial as these have identified the instanton–anti-instanton pairs to be important at finite temperature. This knowledge has allowed us to implement biased Monte Carlo techniques that specifically address the problems faced with random sampling.

The defining characteristic of the system is its strong and short-ranged interactions. We therefore have decided to test importance sampling on a toy-model, roughly displaying the features of the IILM. However, we have decided to leave out any orientation dependence because it is not necessarily a generic feature for the non-trivial backgrounds and would only introduce further complications in the Monte Carlo steps.

The simulations show that random sampling becomes very inefficient if the ‘temperature’ is raised or the ‘mass’ is lowered. We could run the toy-model for very long times and check that both the biased and unbiased simulations give equivalent results. For one particular set of parameters we actually found that ordinary Monte Carlo was not able to reach the correct equilibrium state; in this sense we lost ergodicity because the sample was not representative, having missed that region of phase space that corresponds to pair formation. We expect this to become a much more severe problem for the IILM because the higher dimensionality of the latter will make it much harder to sample correctly the phase space volume leading to pair formation. Also, the simulations will be much more expensive, and long runs will be

prohibitive.

We found that the parameters setting the bonding box have virtually no impact on the results, and that bulk properties agree on the 2σ level. This is as it should be because a strong dependence on these free parameters would introduce some new systematics and render the method impractical. We can therefore choose the parameters that give the smallest autocorrelation times.

Finally, we found that as the degree of dilution is increased, with fixed interaction strength, entropy-dominated configurations will eventually give the bulk contribution to the partition function, and the system will settle in a random state. The opposite equilibrium state, namely a highly correlated ensemble of pairs, is reached by increasing the interaction strength while holding the dilution parameter fixed; in that case the small entropy of these configurations is largely compensated by the gain in total energy, and the partition function is saturated by these energy-dominated states. These are generic features. In the IILM, however, the interactions become stronger and the system more dilute simultaneously as the temperature is increased. Whether the system ends up in a random ensemble or stays in the molecular phase after the chiral symmetry is restored depends on which quantity grows stronger and cannot be decided by the results presented in this study. It is, however, interesting to point to the possibility of a smaller topological susceptibility in the molecular phase as compared to the random phase. In light of the ultimate goal of this series of papers, this could lead to an axion mass that is rather different from the standard one computed in the dilute gas approximation.

Chapter 5

IILM at finite temperature

In this chapter we continue the study of the interacting instanton liquid model (IILM) at finite temperature, with an ancillary goal being to improve our understanding of the temperature dependence of the axion mass. For the first time, we will be able to give a well-motivated axion mass that covers all temperatures down to $T = 0$.

In chapter 3 we set up the formalism underlying the IILM and developed a numerical framework to compute the interactions given by an arbitrary background ansatz. Simulations at zero temperature were performed with the so-called ratio ansatz to determine the parameters that enter the model: the lambda parameter, Λ , and the quark masses, m_q .

We will investigate the IILM at finite temperature based on the caloron solution of Harrington and Shepard [93]. Using as input the physical parameters we determined at zero temperature, we will study chiral symmetry restoration, based on the ideas of instanton–anti-instanton molecule formation [99], [100], [158], and determine the topological susceptibility.

In order to deal with light, physical quark masses, we study the thermodynamic limit. As mentioned in chapter 3, we found that the interactions derived previously in [172] are deficient in this respect. Although the old and new interactions agree rather well at zero temperature, this is no longer the case at $T \neq 0$ because of an unphysical behaviour for the instanton–instanton interaction that only decays as $O(1/R_s)$, with R_s the spatial instanton separation, see Eq. (3.11) in [172]. This long-range interaction prohibits a thermodynamic limit as it is not integrable, see Fig. 5.1. In their paper, the authors

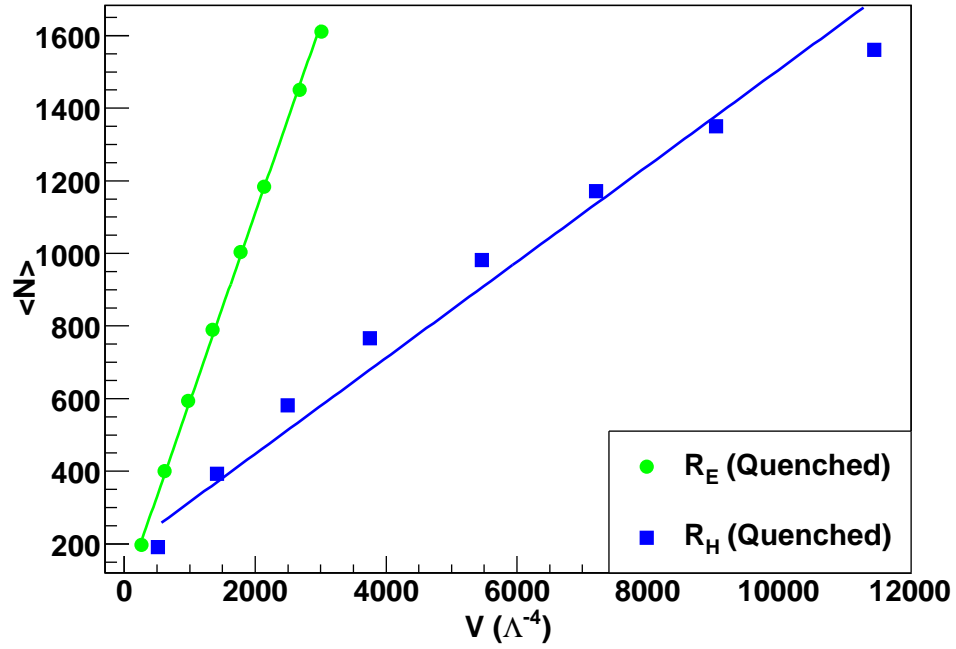


Figure 5.1: The R_H ansatz does not lead to a well-defined thermodynamic limit at finite temperature, whereas the R_E ansatz does exhibit the correct linear scaling of an extensive quantity. Here we display results from quenched simulations, but the same problems persist with dynamical quarks. The instanton number N shows the strongest response to the unphysical interactions.

do discuss this long-range interaction and report they found the $O(1/R_s)$ dyon–dyon behaviour for a wide range of intermediate separations¹. It might well be that for the simulation boxes used in a subsequent numerical investigation (see [157]) the interactions are still well described by this ansatz, i.e. that the spatial extent of the box is bounded by these intermediate separations. However, for studying the large volume behaviour this ansatz is not appropriate. Removing this particular part of the interactions, we were able to retrieve a well-defined thermodynamic limit.

Apart from this deficiency, it seems obvious that at finite temperature it will be much harder to find a good parametrisation for the action of the background ansatz because the underlying $O(4)$ symmetry is broken and the constituent gauge fields are more complicated. This fact was already pointed out in [172]. We improve the existing interactions by extending the formalism set up in chapter 3 to finite temperature; as we will see, the only difficulties are of a technical nature.

In this chapter we will achieve our initial aim of computing the temperature dependent axion mass. So far the mass was computed within a dilute gas approximation which breaks down at low temperatures. The connection between the high temperature regime, where the dilute gas becomes ever more accurate, and the zero temperature result, obtained through chiral perturbation theory, has been performed in a rather crude manner up to date: either by unsmoothed matching [186, 13] or by an ad hoc interpolation prescription [117]. Our determination of the axion mass will for the first time give a well-motivated interpolation between the zero and finite temperature regimes. Comparison with lattice data will allow for a critical evaluation of the systematic uncertainties. In particular, considerations regarding the anthropic axion with large decay constant [126, 188, 184] are potentially very sensitive to the non-perturbative effects of the QCD phase transition, when their mass becomes sizable.

In section 5.1 we will re-derive the finite temperature interactions for the ratio ansatz. We will then discuss, in section 5.2, the new elements that finite temperature introduces in the numerical framework. After these technical preliminaries we will have a short investigation of the topological

¹The gauge potential of a dyon is of the form $A_\mu^a \propto \bar{\eta}_{\mu\nu}^a x_\nu / r_{3d}^2$.

susceptibility in the quenched sector in section 5.3, before we discuss the main numerical results of the unquenched IILM regarding chiral symmetry restoration, section 5.4, and the topological susceptibility and axion mass, section 5.5.

5.1 Interactions in the IILM at finite temperature

In terms of the 't Hooft potential $1 + \Pi$, the Harrington–Shepard caloron [93], an infinite sequence of singular gauge BPST instantons [22] along the Euclidean time direction, is given by

$$A_\mu^a = -O_i^{ab} \zeta_{\mu\nu}^b \frac{\partial_\nu \Pi(x, \{y, \rho\})}{1 + \Pi(x, \{y, \rho\})}, \quad (5.1)$$

where O is the colour matrix in the adjoint representation of the embedding $SU(2) \rightarrow SU(3)$, and $\zeta_{\mu\nu}^b = \bar{\eta}_{\mu\nu}^b$ ($\zeta_{\mu\nu}^b = \eta_{\mu\nu}^b$) for instantons (anti-instantons); η are the 't Hooft symbols. The 't Hooft potential has the following form

$$\Pi(x, \{y, \rho\}) = \frac{\pi \rho^2}{\beta r} \frac{\sinh \frac{2\pi r}{\beta}}{\cosh \frac{2\pi r}{\beta} - \cos \frac{2\pi t}{\beta}}, \quad (5.2)$$

with $r^2 = (\vec{x} - \vec{y})^2$ and $t = x^4 - y^4$; the collective coordinates are: y the centre, ρ the size and O the colour orientation. At finite temperature bosonic quantities such as Π are periodic in the Euclidean time direction with period $\beta = 1/T$. Note that Π approaches the zero temperature instanton potential in the singular gauge for $\beta \rightarrow \infty$.

We will use as background the ratio ansatz and, as for the zero temperature case, only consider two-body interactions. The gauge field is then given by

$$A_\mu^a = -\frac{\sum_i O_i^{ab} \zeta_{\mu\nu}^b \partial_\nu \Pi_i(x, \{y_i, \rho_i\})}{1 + \sum_i \Pi_i(x, \{y_i, \rho_i\})}, \quad (5.3)$$

with $O = O_1 O_2$. This pair interaction has been derived in [172], and we will refer to it as R_H ; the corresponding forces will be denoted by R_E . However, in studying the volume dependence of various quantities, we noticed that

these interactions did not allow for a thermodynamic limit. The problem can be traced back to the log term in the instanton–instanton interaction, (3.11) in [172], only decaying like $O(1/R_s)$ for large R_s , where R_s is the spatial separation of the pair. Note that this term is attributed to the dyon–dyon interaction for intermediate separations, $\beta \ll R_s \ll \rho^2/\beta$, in the high temperature limit.

The ratio ansatz has the same functional form in terms of Π as for zero temperature, and so we can use result from chapter 3 to write

$$\begin{aligned} F_{\mu\nu}^a F_{\mu\nu}^a = & I + (\text{Tr } O^t O + (\bar{\eta} O \eta)_{\mu\nu\mu\nu}) J + (\bar{\eta} O \eta)_{\rho\mu\rho\nu} I_{\mu\nu} \\ & + (\bar{\eta} O \eta)_{\mu\rho\nu\sigma} I_{\mu\rho\nu\sigma} + (\eta O^t O \eta)_{\mu\rho\nu\sigma} J_{\mu\rho\nu\sigma} + (\bar{\eta} O \eta)_{\alpha\mu\alpha\rho} (\bar{\eta} O \eta)_{\beta\nu\beta\sigma} K_{\mu\rho\nu\sigma}. \end{aligned} \quad (5.4)$$

The different terms are given in appendix B.1. Due to charge renormalisation the action, $S[A] = \frac{1}{4g^2} \int F_{\mu\nu}^a F_{\mu\nu}^a$, acquires a quantum contribution in the form of the running coupling constant. The classical interaction is given by

$$S_{12}^g/S_0 \equiv V_{12} \equiv (S[A]/S_0 - 2), \quad (5.5)$$

where $S_0 = 8\pi/g^2$ is the single instanton action. The quantum effects substitute g in S_0 for the running coupling constant; the RG scale is estimated by the geometric mean $\sqrt{\rho_1 \rho_2}$, as proposed in [168, 157].

First, we will look at the dependence of the interactions on R_t , the instanton separation in the (imaginary) time direction. Compared to the zero temperature case, the differences are substantial even for pairs with equal sizes. The major difference comes from the fact that the R_H interactions are not periodic. For unequal sizes, the difference is even more pronounced as was the case at zero temperature, see Fig. 5.2. The reason is again that the functional form on the instanton sizes in R_H is not general enough.

The interaction between instantons and anti-instantons does not lead to as close a match between R_E and R_H as in the $T = 0$ case, but is still qualitatively similar for not too large separations. In the instanton–instanton case, however, we have found significant differences, see Fig. 5.3: we do find the dyon–dyon behaviour for intermediate distances but for separations $R_s \gg \rho^2/\beta$ the functional dependence changes into an integrable $O(R_s^{-4})$ for R_E whereas the non-integrable dyon–dyon interaction persists in the R_H

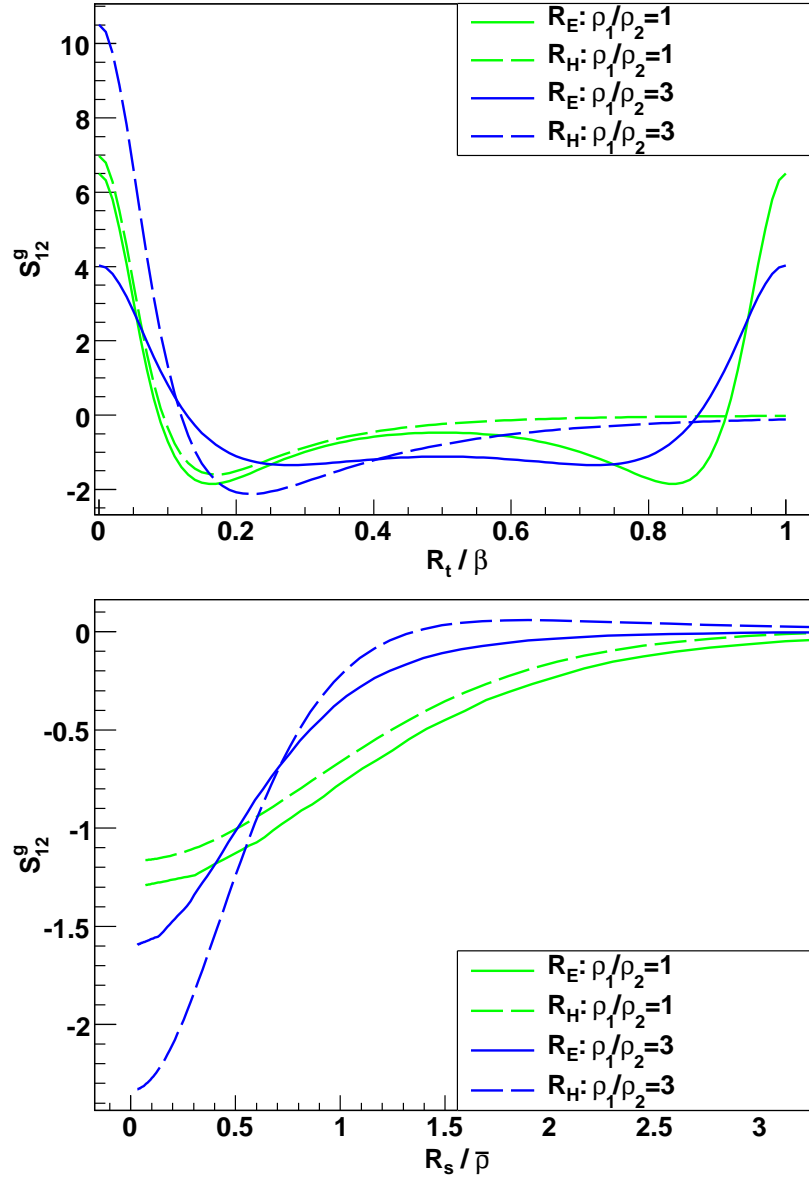


Figure 5.2: The major difference follows from the fact that R_H interactions are not periodic in R_t . This is clearly a deficiency of the analytical formulae because it follows directly from (5.4) that the interactions should have period β . For spatial separations, the main differences occur for unequal size parameters, e.g. $\rho_1/\rho_2 = 3$ in this case. The reason is that the dependence on the sizes is more complicated than the functional form, $\sqrt{\rho_1\rho_2}$, used in R_H . (We have set $\bar{\rho} = \sqrt{\rho_1^2 + \rho_2^2}$.)

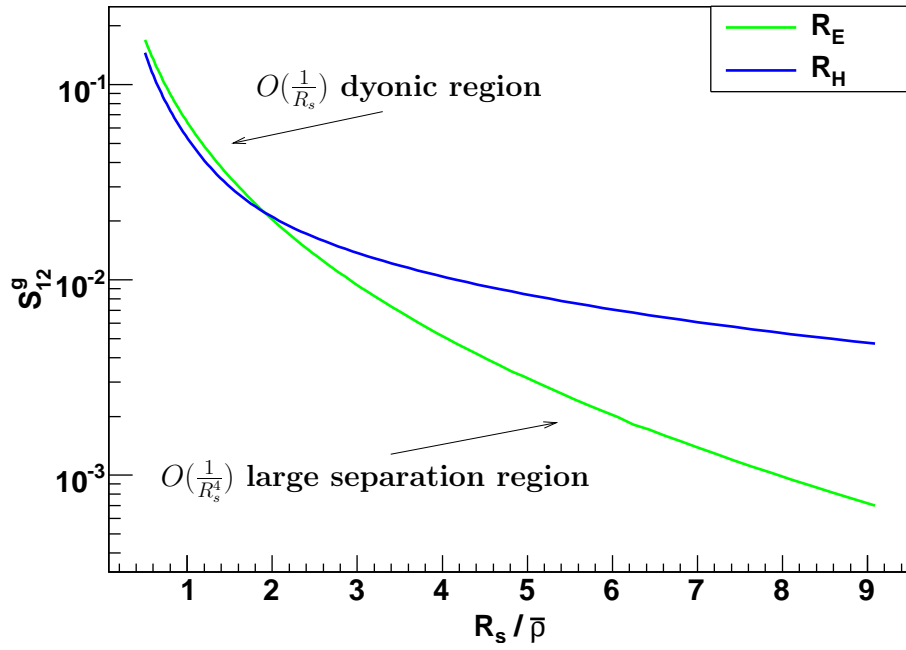


Figure 5.3: The instanton–instanton interaction behaves very differently for R_H and R_E . In the high temperature, large size limit we do see the dyon–dyon behaviour but for larger separations, $R_s \gg \rho^2/\beta$, the interaction decays much faster, $O(R_s^{-4})$. This fall-off behaviour is integrable and we can study the thermodynamic limit, in contrast to the purely dyonic interaction. (We have set $\bar{\rho} = \sqrt{\rho_1^2 + \rho_2^2}$.)

ansatz. That the large separation fall-off should be different from the dyonic regime is clear from the discussion in [83] where the authors show that the caloron field develops a dipole-like character in the far-field region. Note that this leads to a three-dimensional dipole–dipole interaction for an instanton–anti-instanton pair, but that the R_H ansatz retains the zero temperature (four-dimensional) dipole–dipole interaction².

The fermionic interaction follows from the quark wave function overlaps

$$(\not{D} + m)_{ij} = \langle \xi_i | \not{D} + m | \xi_j \rangle = \not{D}_{ij} + m\delta_{ij}. \quad (5.6)$$

Even though the set of eigenfunctions $\{\xi_i\}$ is generally not an orthonormal basis, the extra contributions are neglected; effectively, we treat $\{\xi_i\}$ as being orthonormal which leads to the diagonal mass term. The fermionic zero mode at finite temperature is given by [84], [83], [172],

$$\xi_I = \frac{1}{2\pi\rho_I} \sqrt{1 + \Pi_I} \not{D} \frac{\chi_I}{1 + \Pi_I} \begin{pmatrix} U_I \varphi \\ 0 \end{pmatrix}, \quad (5.7)$$

$$\xi_A = \frac{1}{2\pi\rho_A} \sqrt{1 + \Pi_A} \not{D} \frac{\chi_A}{1 + \Pi_A} \begin{pmatrix} 0 \\ U_A \varphi \end{pmatrix}, \quad (5.8)$$

$$\chi = \Pi \frac{\cos \frac{\pi t}{\beta}}{\cosh \frac{\pi r}{\beta}}, \quad (5.9)$$

with $\varphi_{\alpha a} = \epsilon_{\alpha a}$, normalised according to $\epsilon_{12} = 1$, and U_i the collective coordinates for the colour embedding in the fundamental representation.

The overlaps $T_{IA} = \int \xi_I^\dagger i \not{D} \xi_A$ have a slightly more complicated form than their $T = 0$ counterparts and are given by

$$T_{IA} = \int_{\mathbb{R}^3 \times S^1} \frac{1}{4\pi^2 \rho_I \rho_A} \left(\frac{1}{2} \text{Tr} (U \tau_\beta^+) I_\beta - \frac{i}{2} \text{Tr} (U \tau_\beta^+ \tau_a) \bar{\eta}_{\mu\alpha}^a J_{\beta\mu\alpha} + \frac{i}{2} \text{Tr} (U \tau_a \tau_\beta^+) \eta_{\mu\alpha}^a K_{\beta\mu\alpha} \right). \quad (5.10)$$

The different contributions can be found in appendix B.2.

²instanton–instanton pairs decay slightly faster, at $O(1/R_s^4)$, however, and their interaction is not of dipole type; note that this happens at $T = 0$ too, where instanton–anti-instanton pairs exhibit a dipole–dipole interaction but instanton–instanton pairs don't [61].

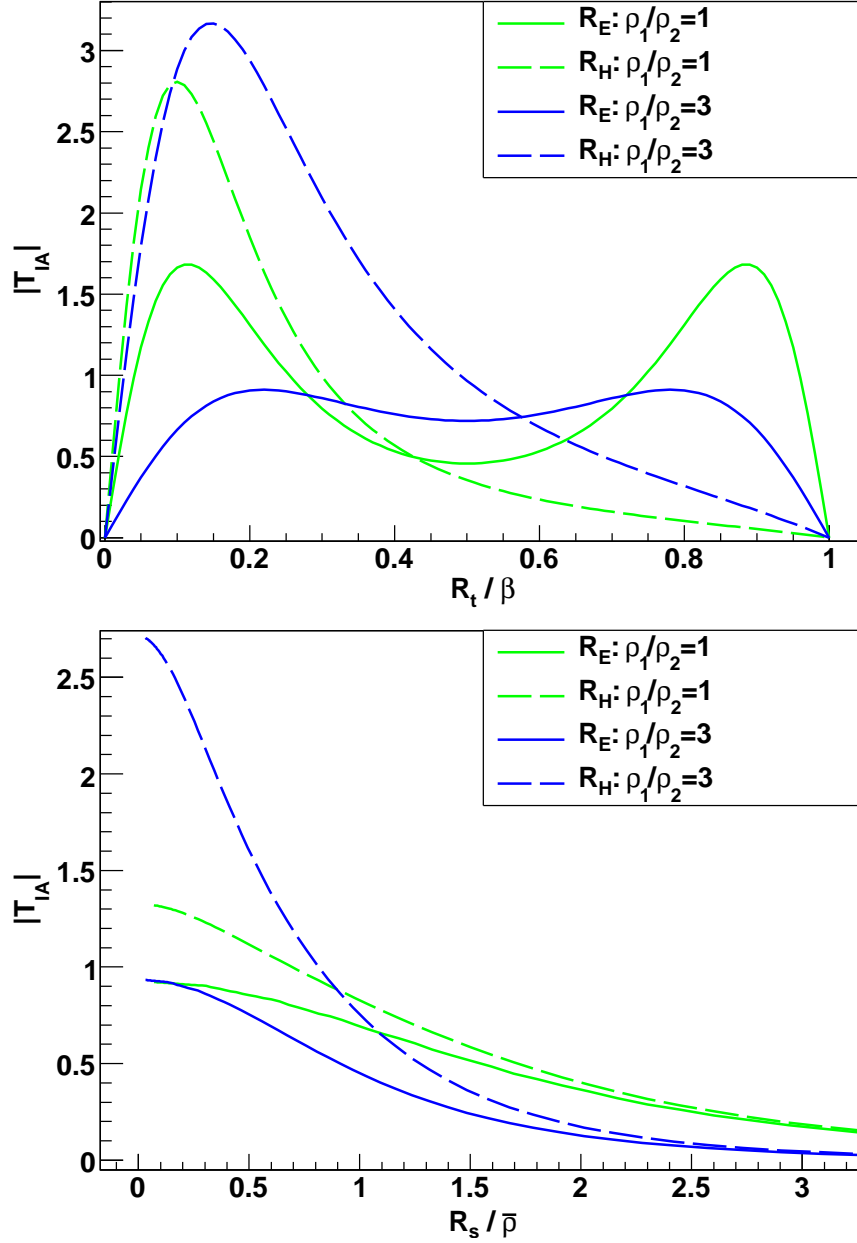


Figure 5.4: The large discrepancy between R_E and R_H for both temporal and spatial separations is due to the fact that R_E uses the full ratio ansatz in the Dirac operator whereas for R_H a sum ansatz is used. Periodicity, which follows directly from (5.6), is lacking for overlaps in the R_H ansatz but is realised in the R_E ansatz. (We have set $\bar{\rho} = \sqrt{\rho_1^2 + \rho_2^2}$.)

The difference between the R_E and R_H interactions is quite large for temporal and spatial separations, see Fig. 5.4. However, it is not straightforward to compare both ansätze because the R_H overlaps are computed on a different background [172], the sum ansatz, whereas we use the full ratio ansatz in R_E . As in the gluonic case, the quark overlaps are not periodic for the R_H ansatz as opposed to those of R_E . As always, unequal sizes increase the differences between the R_H and R_E ansätze even more. For large separations, when the ratio ansatz becomes indistinguishable from the sum ansatz, and for equal sizes, we find very good agreement between R_H and R_E .

The total interaction, after normalising to the dilute gas approximation, is given by

$$S_{\text{int}} = \sum_{\text{pairs } (i,j)} S_0(\sqrt{\rho_i \rho_j}) V_{ij} - \sum_{n=1}^{N_f} \begin{cases} \ln \det(\mathbb{I} + \frac{TT^\dagger}{m_n^2}) & , Q < 0 \\ \ln \det(\mathbb{I} + \frac{T^\dagger T}{m_n^2}) & , Q > 0 \end{cases} , \quad (5.11)$$

where $Q = N_I - N_A$ is the topological charge and N_f the number of quark flavours. For details see chapter 3.

Even though the analytic formulas of the R_H ansatz are not periodic, this is not really a major shortcoming because periodicity can be realised by folding back the instantons into the fundamental interaction region $R_t \in [-\beta/2, \beta/2]$. We demonstrate this for the total interaction in Fig. 5.5. We can clearly see that the R_E ansatz is intrinsically periodic.

5.2 Numerical Implementation

5.2.1 Interpolation and asymptotic matching

As for zero temperature, we will need to set up a grid for the numerical evaluation of the two-body interactions. The look-up tables will depend on four variables: the spatial separation R_s , the temporal separation R_t and the two sizes ρ_1 and ρ_2 . The colour degrees of freedom O , or equivalently U , have been completely factored out and can be treated exactly.

As we have seen in the introduction to this chapter, the dyon–dyon regime is characterised by a fairly slow fall-off. However, not all bosonic interactions in appendix B.1.1 decay this slowly; rather they can be grouped according to

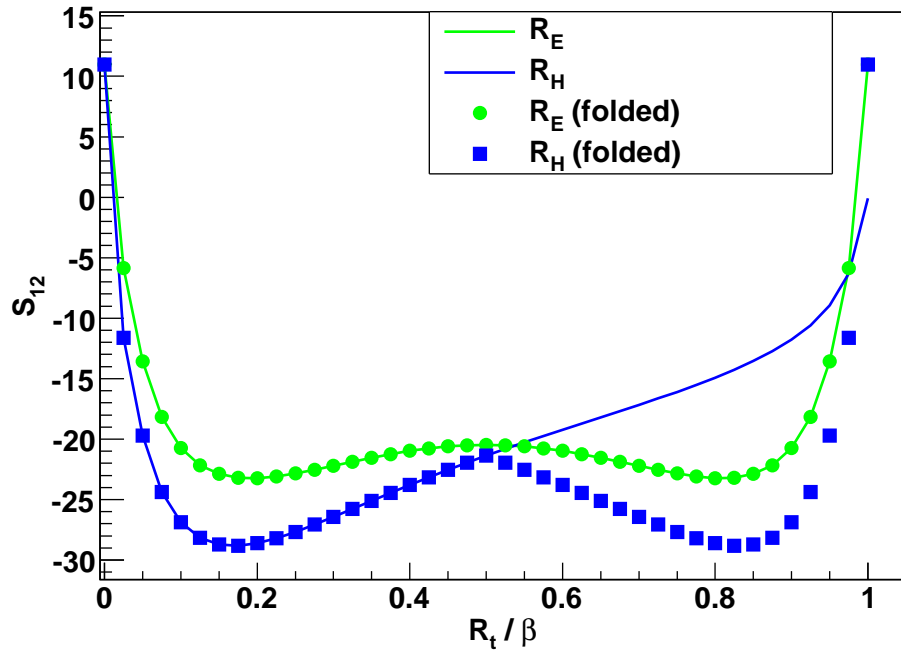


Figure 5.5: Folding back the temporal separation into the fundamental interaction region, $R_t \in [-\beta/2, \beta/2]$, we explicitly retrieve periodic interactions for R_H . The R_E ansatz is intrinsically periodic.

polynomial and exponential decay. The fermionic overlaps fall-off exponentially, but slightly more slowly than their bosonic counterparts. Therefore, we found it advantageous to define three different grids:

- For the polynomial grid, the maximal separation R_s^{\max} will depend on $\tilde{\rho} \equiv \pi\rho^2/\beta$, the natural size parameter for separations beyond β . Note that for very small sizes $R_s^{\max} < \beta$ the prescription for R_s^{\max} should switch over to the $T = 0$ case; this is implemented by setting $R_s^{\max} = \max(\alpha_0^p \rho, \alpha_T^p \pi\rho^2/\beta)$.
- The exponential decay sets in at $R_s \approx \beta$. To accommodate very small instantons we set again $R_s^{\max} = \min(\alpha_0^{eg} \rho, \alpha_T^{eg} \beta)$.
- The grid for quark overlaps is set up accordingly, i.e. $R_s^{\max} = \min(\alpha_0^{eq} \rho, \alpha_T^{eq} \beta)$.

The constants α_i^j are fine-tuned so as to achieve fast and stable numerical integrations and a good matching to the analytic expressions used for separations beyond R_s^{\max} . In the temporal direction the grid is bounded by $|R_t| = \beta/2$; care needs to be taken again for small instantons and we set $|R_t^{\max}| = \min(R_s^{\max}, \beta/2)$.

The size distribution is supported on 30 grid points and for each pair (ρ_i, ρ_j) the $R_s - R_t$ plane consists of $30 \cdot 29$ nodes; this leads to a total of roughly 1 million interpolation points. The grid is ‘logarithmic’ in the size and R_s direction. A typical $R_s - R_t$ plane is shown in Fig. 5.6.

The minimal and maximal sizes supported by the grid are fixed as in the $T = 0$ case, i.e. they are chosen so small, respectively large, that they correspond to a very low, respectively high, quantile of the normalised instanton density. As we will see in section 5.2.2, at finite temperature the instanton density becomes temperature dependent, reflecting the screening of coherent background fields. Instead of computing a grid for every temperature we run simulations at, we can exploit the following scaling transformations: under coordinate rescaling, $x_\mu \rightarrow \alpha x_\mu$,

$$V(\rho_i, R_j, \beta) = V(\rho_i/\beta, R_j/\beta), \quad (5.12)$$

$$T(\rho_i, R_j, \beta) = \frac{1}{\beta} T(\rho_i/\beta, R_j/\beta). \quad (5.13)$$

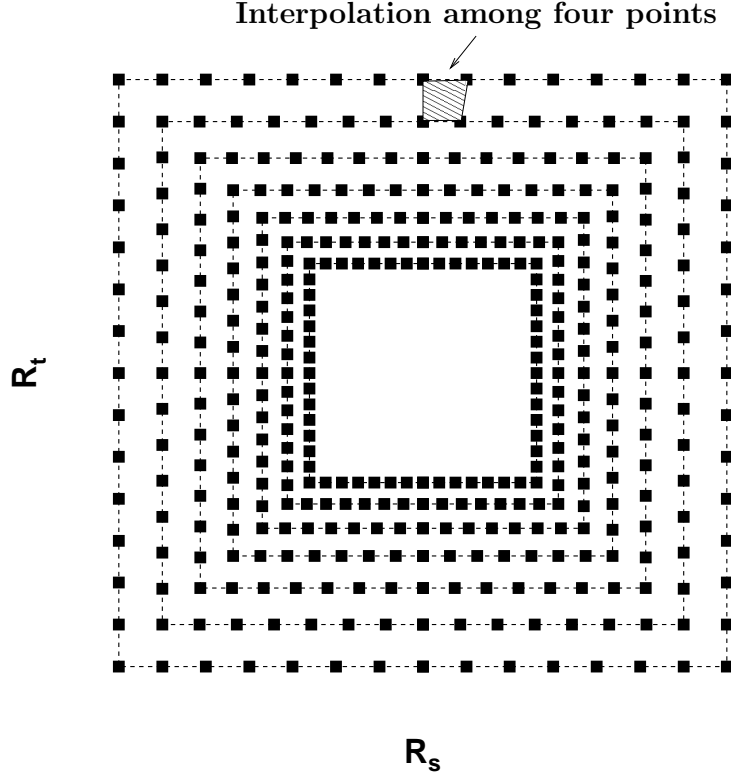


Figure 5.6: It turns out that the interactions are more or less identical on the $R_s - R_t$ plane, when expressed in units of $\rho_1^2 + \rho_2^2$. We therefore choose the grid to be irregular, i.e. the points on the $R_s - R_t$ plane are different for different values of the sizes. Since instantons are localised field configurations, the separation between grid points grows geometrically in the R_s direction. In the ρ directions the mesh is regular, but still ‘logarithmic’. The inner and outer regions are matched to asymptotic expansions: for small separations the matching is performed radially from the origin because in this region the interactions are roughly $O(4)$ symmetric; for large separations we match along the R_s direction with constant R_t because the interactions become t -independent. (Note that for graphical reasons not all 870 grid points are displayed and that we actually display $\log R_s$. Also the actual grid only contains the $R_s > 0$ points.)

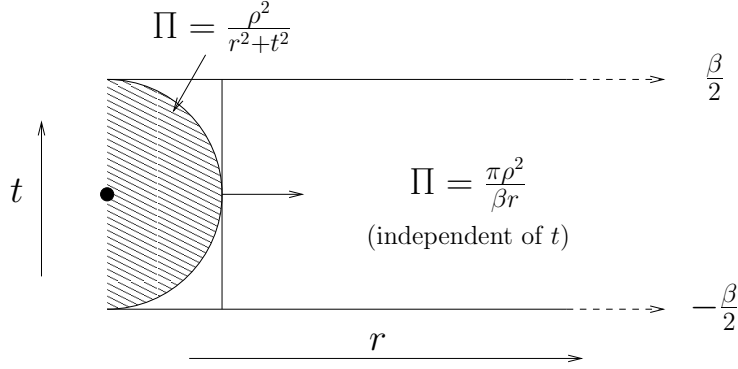


Figure 5.7: In order to derive simple analytical expressions in the limit of small and large pair separations, we approximate the 't Hooft potential to have a simple functional form in different integration regions. This procedure works well for instantons with sizes that do not exceed β . As we will see the plasma screening effects limit instanton sizes to be rather small compared to β .

These can be used to transform the grid and interactions to different temperatures. We just need to make sure to choose the original sizes large enough to accommodate the low temperature behaviour, i.e. $\rho_{\max} = \Lambda$. We defined the grid and the interactions at $\beta = 1$ with $\rho \in [0.01, 1.6]$.

As for the zero temperature case, the matching consists in deriving asymptotic formulas f_{asy} that are patched on to the numerically integrated interactions according to

$$f(R) = f_{\text{asy}}(R) \frac{f_{\text{ex}}(R_m)}{f_{\text{asy}}(R_m)}, \quad (5.14)$$

where R_m is the matching point. For details see chapter 3. At $T \neq 0$, however, we match differently for small and large separations. In the former case the interactions will be approximately $O(4)$ symmetric and the matching is performed in an $O(2)$ symmetric way on the $R_s - R_t$ plane, i.e. along a ray connecting the origin with (R_s, R_t) . For large separations the interactions become t -independent and we match along a line of constant R_t .

To derive the large separation asymptotic formulas, we consider two different integration regions, see Fig. 5.7. Therefore, we effectively approximate integrals by

$$\int_{S^1 \times \mathbb{R}^3} \approx \int_{S^3 \times [0, \beta/2]} + \beta \int_{S^2 \times [\beta/2, \infty]}, \quad (5.15)$$

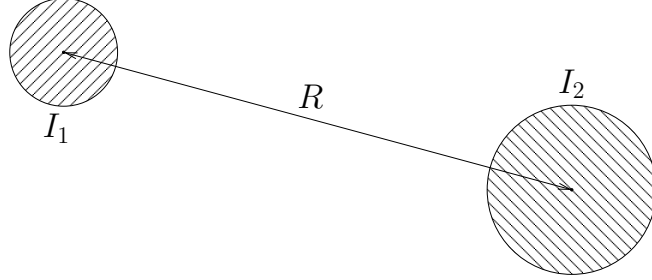


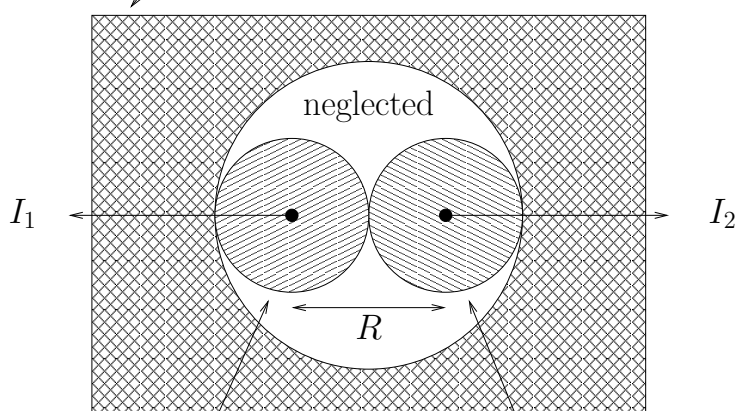
Figure 5.8: The instantons I_1 and I_2 are so far apart that within the shaded region that give the dominant contribution to the field strength of each, the other's field strength is roughly constant and fixed at $x_\mu - R_\mu \approx -R_\mu$. We can then safely extend the integration region to be all of $S^1 \times \mathbb{R}^3$, with a negligible error due to the rather strong localisation of the individual instantons. The functional form of the relevant 't Hooft potential changes from $\Pi = \rho^2/r_{4d}^2$ to $\Pi = \pi\rho^2/\beta r_{3d}$ at $r_{4d} = \beta/2$, see Fig. 5.7.

where we exploit the respective spherical symmetry of the 't Hooft potentials. In both regions we only have to deal with rational functions which can be easily integrated exactly. Apart from this extra complication the strategy is the same as in the zero temperature case and is illustrated in Fig. 5.8. The results are given in appendix B.1.2 and B.2.2.

The main objective for small separations is to capture adequately the singularity structure, for which it is sufficient to restrict ourselves to the $T = 0$ region. To capture contributions that do not blow up, we add the large separation contributions truncated at the pair separation (R_t, R_s) . The results are given in appendix B.1.2 and B.2.2. Apart from considering different functional forms for the 't Hooft potential this is the same procedure as for the $T = 0$ asymptotic interactions and is illustrated in Fig. 5.9.

As in the $T = 0$ case, the quark overlaps in the small separation region are qualitatively incorrect. However, this region is again dominated by the gluonic repulsion. The latter is not as well approximated as in the $T = 0$ case but the agreement is still on the 10% level and, most importantly, qualitatively correct, see Fig. 5.10. In contrast, the large separation asymptotic formulas work well.

Proper small separation asymptotic integration down to R



Large separation asymptotic integration up to $\frac{R}{2}$

Figure 5.9: The instantons I_1 and I_2 are strongly overlapping. We approximate the integral by first integrating over I_1 keeping I_2 fixed at R_μ as in the large separation case, see Fig. 5.8, but with upper limit $R/2$; to this we add the analogous contribution from I_2 . The possibly singular behaviour is picked up by integrating from infinity down to R and approximating the arguments by $x_\mu - R_\mu/2 \approx x_\mu$ and $x_\mu + R_\mu/2 \approx x_\mu$, respectively.

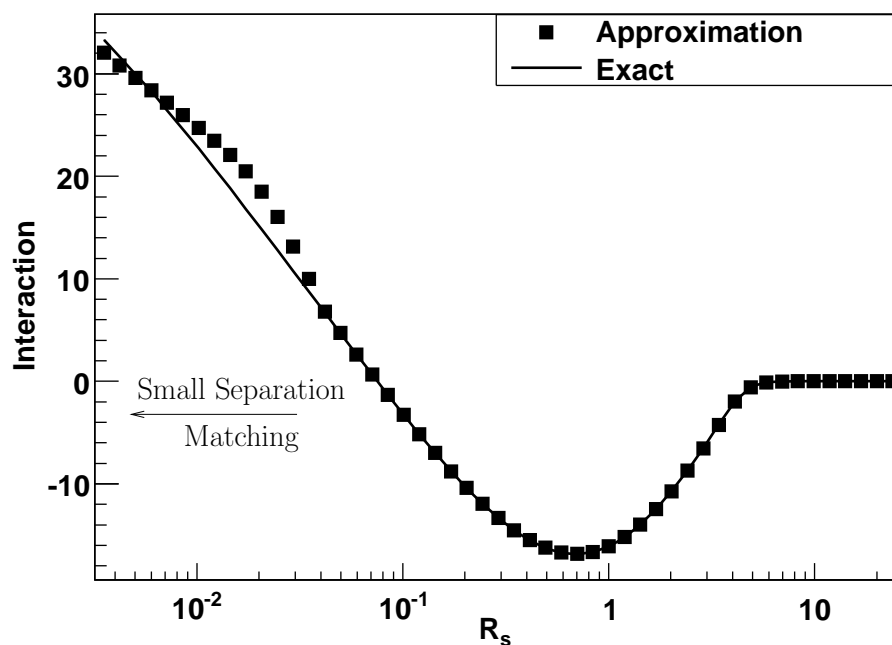


Figure 5.10: The small separation asymptotic approximation is only qualitatively correct. It overestimates the correct interaction for moderately small separations because it underestimates the quark contribution. For very small separations, where the exact quark interaction is indeed negligible, the gluonic approximation underestimates the exact result. However, the grid covers rather small separations so that the asymptotic interactions for strongly overlapping instanton–anti-instanton pairs are rarely needed. In those cases where it is needed we get a qualitatively correct repulsion. (The temporal separation R_t has been chosen very small in order to see the repulsion).

5.2.2 Biased Monte Carlo

The IILM is defined by its partition function,

$$Z = \sum_{N_I, N_A}^{\infty} \frac{1}{N_I!} \frac{1}{N_A!} \prod_i^{N_I} d(\rho_i) \prod_j^{N_A} d(\rho_j) \exp(-S_{\text{int}}), \quad (5.16)$$

with S_{int} defined in (5.11) and the single instanton density given by $d(\rho) = d_0(\rho)d_T(\rho)$. The zero temperature contribution d_0 is given in (3.42). The finite temperature term describes the screening in the plasma of coherent field excitations exceeding the inverse temperature scale β , and is given by [83] [157],

$$d_T(\rho) = \exp \left[-\frac{1}{3}(2N_c + N_f)(\pi\rho T)^2 - \left(1 + \frac{N_c}{6} - \frac{N_f}{6}\right) \left(-\ln(1 + (\pi\rho T)^2/3) + \frac{0.15}{(1 + 0.15(\pi\rho T)^{-3/2})^8} \right) \right]. \quad (5.17)$$

Apart from book-keeping related technicalities, the major challenge is to adequately simulate instanton–anti-instanton molecules, the structures responsible for the chiral symmetry restoration within the IILM [99], [100], [158]. Detailed numerical studies at finite temperature have been performed in [157]. For the mass parameters used in these latter simulations no technical problems were encountered. However, for the small quark masses that we determined in the $T = 0$ simulations, we found that standard Monte Carlo techniques faced severe problems with the strong attraction between instanton–anti-instanton pairs. In chapter 4 we have argued that this technical problem will occur quite generally for the semi-classical expansion with small quark masses. There we have set up the framework to deal with these algorithmic issues by adapting techniques from chemical engineering and computational chemistry and developed for the study of strongly associating fluids.

The instanton–anti-instanton pair interactions become ever stronger and more localised for small quark masses, and random sampling methods will generically miss these configurations. We therefore need to preferentially sample the attraction centres. This can be achieved by resorting to biased Monte Carlo schemes. They exploit the large redundancy in devising transi-

tion probabilities P_{ij} that satisfy detailed balance,

$$P_i^{\text{eq}} P_{ij} = P_j^{\text{eq}} P_{ji}, \quad (5.18)$$

where the equilibrium distribution P^{eq} is given by the partition function. Specifically, the degeneracy follows from the fact that the transition probability consists of two parts, the proposal probability, \mathcal{P}_{ij} , and the acceptance probability, \mathcal{A}_{ij} ,

$$P_{ij} = \mathcal{P}_{ij} \mathcal{A}_{ij}. \quad (5.19)$$

To satisfy (5.18), the acceptance probability is given by

$$\mathcal{A}_{ij} = \min \left[1, \frac{P_j^{\text{eq}} \mathcal{P}_{ji}}{P_i^{\text{eq}} \mathcal{P}_{ij}} \right], \quad (5.20)$$

which is the Metropolis prescription.

Recently, efficient and general purpose algorithms have been developed to achieve this importance sampling; we will use [205], the Unbonding–Bonding algorithm (UB). This algorithm focuses not on the union of all the interaction regions, a complicated and case-specific geometrical problem, but on the individual interaction regions and all the possible routes that lead to the same final state.

For the IILM, we choose the bonding region to consist of the two disconnected valleys of strong interaction, see Fig. 5.11, and take them to be hypercubes for simplicity. The precise size of the box is a free parameter and will be chosen to achieve the fastest convergence. We will check that the results do not depend on the bonding box parameters, within statistical uncertainties. We will now present the transition probabilities for the different types of updates; details can be found in chapter 4.

The forward and backward transition probabilities for the UB moves of

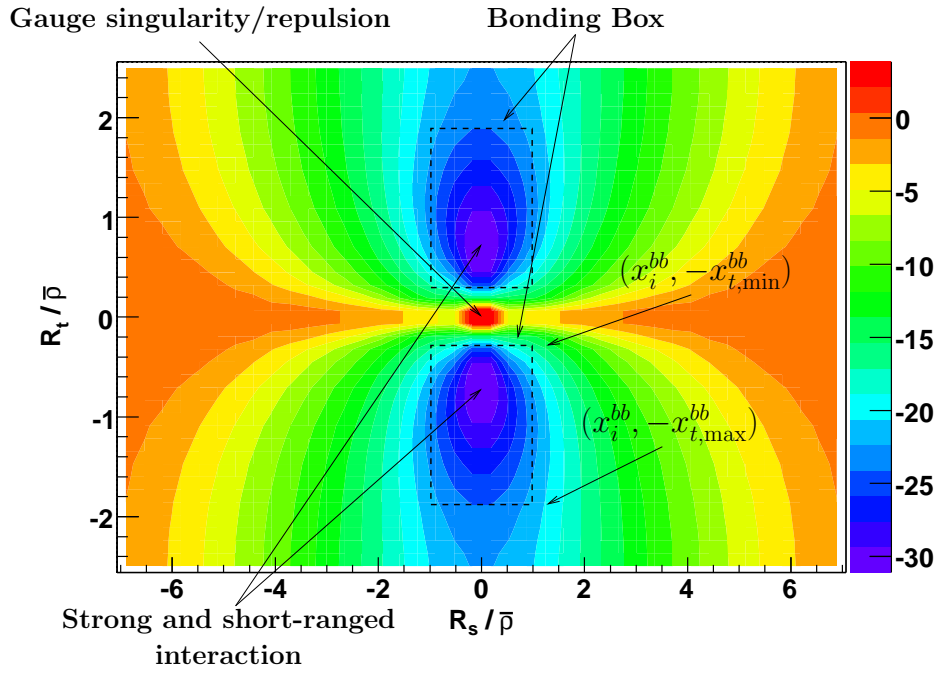


Figure 5.11: The strongest interaction is achieved for the colour orientation $U = \mathbb{I}$ and located in the two troughs. For simplicity we will choose the bonding box to consist of the two hypercubes displayed by dashed lines.

instanton i are given by

$$\mathcal{P}_{ii'} = \sum_j^{N_I'^B(i)} \mathcal{P}_{i(i',j)}^B + \delta_i^B \mathcal{P}_{ii'}^U, \quad (5.21)$$

$$\mathcal{P}_{i'i} = \sum_j^{N_I^B(i)} \mathcal{P}_{i'(i,j)}^B + \delta_{i'}^B \mathcal{P}_{i'i}^U, \quad (5.22)$$

$$(5.23)$$

with $\delta_i^B = 1$ if i is bonded and $\delta_i^B = 0$ otherwise. We denote by N_I^B (N_A^B) the number of bonded instantons (anti-instantons) and $N_I^B(i)$ is the number of anti-instantons that instanton i is bonded to, and analogously for $N_B^B(i)$. Unprimed quantities are evaluated before the move whereas primed ones denote the same quantity after the move. The individual bonding and unbonding transition probabilities are given by

$$\mathcal{P}_{i(i',j)}^B = \frac{1}{N_I} \frac{1}{N_A} \frac{1}{V_j}, \quad (5.24)$$

$$\mathcal{P}_{ii'}^U = \frac{1}{N_I^B} \frac{1}{V}. \quad (5.25)$$

The bonding move consists of choosing uniformly an instanton and an anti-instanton, and to place the instanton i uniformly in the bonding region V_j of the anti-instanton j . The unbonding move consists of choosing uniformly one of the bonded instantons and to place it uniformly in the simulation box.

Insertion and deletion will be constructed along the lines of the UB algorithm by either placing the instanton i into the bonding region of an anti-instanton or removing the bonded instanton i . Adding up all possible paths, including the unbiased one, we get

$$\mathcal{P}_{N_I N_I'} = p_b \sum_j^{N_I'^B(i)} \frac{1}{N_A} \frac{1}{V_j} + (1 - p_b) \frac{1}{V}, \quad (5.26)$$

$$\mathcal{P}_{N_I' N_I} = p_b \frac{\delta_i^B}{N_I'^B} + (1 - p_b) \frac{1}{N_I'}. \quad (5.27)$$

Here p_b is the a-priori-probability to perform biased moves.

It has also been argued that cluster moves are important to achieve good

mixing of the Markov chain [142]. We assume that $I - A$ pairs are the dominant clusters that form. The pair displacements consist of translating the pair as a whole or to displace one of the constituents within the bonding region of its partner, and the resulting transition probabilities are obvious. The pair insertions and deletions are given by

$$\mathcal{P}_{NN'} = \frac{\delta_{i_I i_A}^P}{4} \frac{1}{V} \left(\frac{1}{V_{i_I}} + \frac{1}{V_{i_A}} \right) + \frac{1}{2} \frac{1}{V} \frac{1}{V}, \quad (5.28)$$

$$\mathcal{P}_{N'N} = \frac{1}{2} \frac{\delta_{i_I i_A}^P}{N'_P} + \frac{1}{2} \frac{1}{N'_I} \frac{1}{N'_A}, \quad (5.29)$$

where N_P is the number of pairs. The last terms in each line follow from randomly and independently inserting/deleting an instanton and an anti-instanton. We found that inclusion of these unbiased moves enhances the acceptance rates at high temperature when we approach the dilute gas limit.

So far we have only discussed the spatial arrangements of the instantons. However, the interaction also depends on the colour degrees of freedom [72]. This further technicality can be dealt with by adapting techniques for orientation dependent forces in molecular dynamics simulations. To this end we define the measure in colour space by the Boltzmann factor of the pair interaction. This ensures that U is chosen so as to increase the interaction. More precisely, given the position of the instanton within the bonding box of its partner anti-instanton, we want to sample the following equilibrium distribution:

$$\mathcal{P}(U_I | x_I, x_A, U_A) = \frac{\exp \left(-S_{\text{int}}^{\text{pair}}(U_I, U_A, x_I, x_A) \right)}{\int dU_I \exp \left(-S_{\text{int}}^{\text{pair}}(U_I, U_A, x_I, x_A) \right)}, \quad (5.30)$$

where dU_I is the Haar measure over $SU(3)$. Note that we neglect the influence of the neighbouring instantons and anti-instantons; we could include them but since the fermionic interaction involves a determinant the computation would become rather costly.

We cannot sample this distribution analytically, but we can sample it exactly within a Monte Carlo scheme [72]. Using the fact that the Haar measure is invariant under group composition and also that the interactions (5.4) and (5.10) depend on $U = U_I^\dagger U_A$, it is natural to work with U . To

sample (5.30), we choose a set of N different colour matrices uniformly over $SU(3)$, i.e. according to the Haar measure. We select one among those N matrices according to the probability

$$p_i = \frac{e^{-S_i^{\text{pair}}}}{\sum_j^N e^{-S_j^{\text{pair}}}}, \quad (5.31)$$

with $S_i^{\text{pair}} = S_{\text{int}}^{\text{pair}}(U_i, x_I, x_A)$. Now, remember that the transition probability is given by a proposal and acceptance probability, and that the former includes any a-priori-probabilities for generating the proposed state. Thus, the orientational bias includes the generation probability of the set of matrices, and is given by

$$\mathcal{P}_i = \left(\frac{1}{V_{SU(3)}} \right)^{2N-1} N p_i. \quad (5.32)$$

The selection probability and the first N volume factors are obvious. The additional $N - 1$ volume factors need further explanation. Given the form (5.20) for the acceptance probability, apart from the N matrices used in (5.31), we need to generate another $N - 1$ trial matrices that combine with the matrix selected in the forward move to form the set of N matrices necessary to evaluate the orientational bias, (5.31), in the backward move. Finally, the extra enhancement of N is necessary because we are not so much interested in U_i being generated at the i th trial but rather that the value U_i is selected. Thus we must marginalise over the label i : adding up all the permutations of the set of trial matrices modulo the permutations of the subset of the $N - 1$ matrices that are not selected produces the extra N factor. Of the $2N - 1$ volume factors $2N - 2$ will cancel in general³ and so we can simplify (5.32) to

$$\mathcal{P}_i = \frac{e^{-H_i}}{\frac{V_{SU(3)}}{N} \sum_j^N e^{-H_j}}, \quad (5.33)$$

which we recognise as an approximation to (5.30); note, however, that within the MC scheme is exact, i.e. the results do not suffer from any discretisation errors. Including the orientation bias amounts to multiplying the bonding

³One might think that all $2N - 1$ volume factors drop out, but in general we might add new moves that select U with a rule different from (5.31); in such a case we only need $2N - 2$ trial matrices, i.e. $N - 1$ from either the forward and the backward move.

box volume factors V_j by (5.33).

We found that the interactions are rather self-similar when expressed in units of $\bar{\rho} = \sqrt{\rho_I^2 + \rho_A^2}$. Thus, the bonding region depends on both particles that make up the pair through a function of their sizes. We choose the functional form of the size dependence to be given by

$$V_j \rightarrow V_{ij} = V_{\text{UB}} (\rho_i^2 + \rho_j^2)^2 . \quad (5.34)$$

By construction, V_{UB} is independent of the specific pair, and we will tune it to achieve fast convergence.

The IILM does not lead to confinement; from a heuristic point of view this follows from the simple fact that the Harrington–Shepard calorons have a vanishing Polyakov loop, the order parameter of the phase transition. It has been argued in [163] that at low temperatures the instanton density should only depend very weakly on temperature and that, in particular, the result (5.17) is only applicable in the plasma, i.e. deconfined, phase. In the hadronic phase the fundamental degrees of freedom are not screened and the heat bath consists mainly of pions. We will follow the prescription put forward in [157] and include in (5.17) a phenomenological term that mimics the transition from the confined to the deconfined phase. We have decided to investigate this with the following two choices, see Fig. 5.12,

$$\begin{aligned} d_T^1(\rho) &= d_0(\rho) d_T(\rho)^{\alpha_T} , \\ d_T^2(\rho) &= d_0(\rho) (1 + \alpha_T (d_T(\rho) - 1)) , \\ \alpha_T &= \frac{1}{2} \left(1 + \tanh \frac{T - T_*}{\Delta T} \right) . \end{aligned} \quad (5.35)$$

The functional form d_T^1 for the transition between hadronic and plasma phase has been used in [157]. The functional form d_T^2 is chosen on the basis that, intuitively, it should lead to a more symmetrical behaviour around T_* . With these two choices we will get a rough idea on the systematics due to these phenomenological terms. We’ve introduced two new parameters into the IILM; they will be fixed by comparing our results to available lattice data.

So far these modified screening factors are purely phenomenological. However, the more general non-trivial holonomy calorons [119, 121, 120] [123],

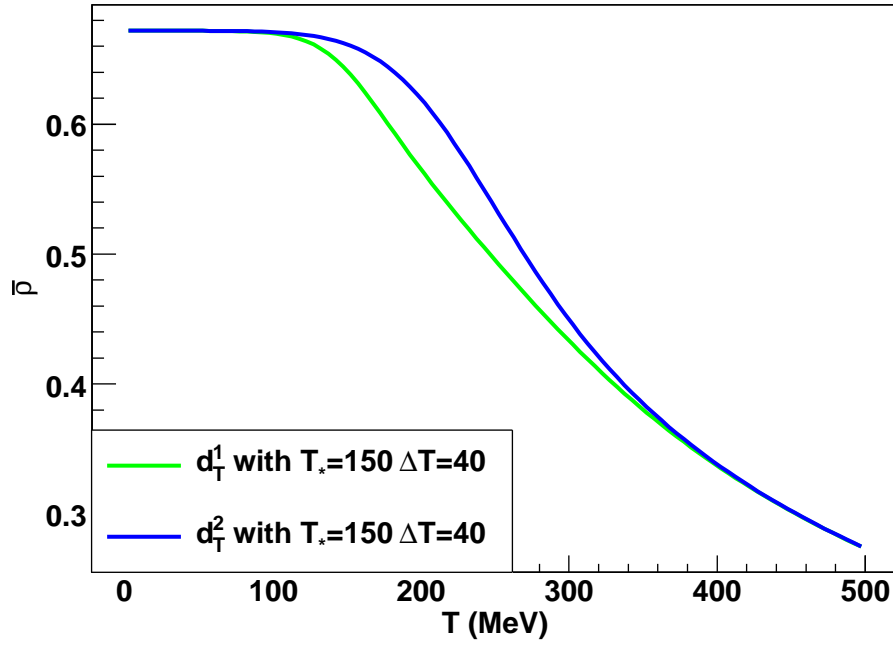


Figure 5.12: The screening factors (5.35) lead to constant mean instanton sizes in the confined phase, below T_* , as has been found from lattice investigations [40]. As expected, the factor d_2 leads to a slower and more symmetrical switch-on of the screening.

that might play a crucial role in driving the confinement/deconfinement phase transition, have been shown to lead to mean instanton sizes that depend very weakly on temperature in the confined phase [80]. Given the assumptions of that paper, the change of the Polyakov loop with temperature is responsible for the different screening behaviour of instantons in the confined and deconfined phase: in the former, the equilibrium state corresponds to an ensemble of maximally non-trivial holonomy calorons which are not screened, whereas the latter consists of trivial holonomy Harrington–Shepard calorons which are strongly screened. We can therefore interpret the modified screening factors (5.35) as an effective description of the change of the Polyakov loop as the system evolves through the phase transition. Generalising the IILM to include these calorons in the future, we might be able to describe the plasma screening effects self-consistently.

Note that for low temperatures the IILM should not be very different from the $T = 0$ case, which was well dealt with by ordinary Monte Carlo. We can use this regime to test the more advanced Monte Carlo scheme against unbiased simulations. In Fig. 5.13 we plot the thermalisation history at such low temperatures for two different set of quark masses; the latter are given by the physical masses determined in chapter 3 and the second set has the same dimensionless strange quark mass and identical light quarks with a value of ten times the physical up quark. We can clearly see that the histories for the biased and unbiased simulations are very similar.

As the temperature is increased, ordinary Monte Carlo fails for the set of physical quark masses. For intermediate temperatures, around T_* , there is a substantial difference in the results from biased and unbiased simulations. Even for the set of larger masses there are differences visible to the naked eye, see Fig. 5.14. For these temperatures instanton–anti-instanton pair formation, the mechanism that drives the chiral symmetry restoration in the IILM, is supposed to be very important. We find that the number of pairs is indeed quite different with $\langle N_P \rangle \approx 20$ for biased and $\langle N_P \rangle \approx 7$ for unbiased simulations and physical quarks. For the larger quark masses we find that the difference in the number of pairs between biased, $\langle N_P \rangle \approx 16$, and unbiased, $\langle N_P \rangle \approx 11$, simulations is less pronounced, as expected.

At temperatures well inside the plasma phase, ordinary Monte Carlo fails

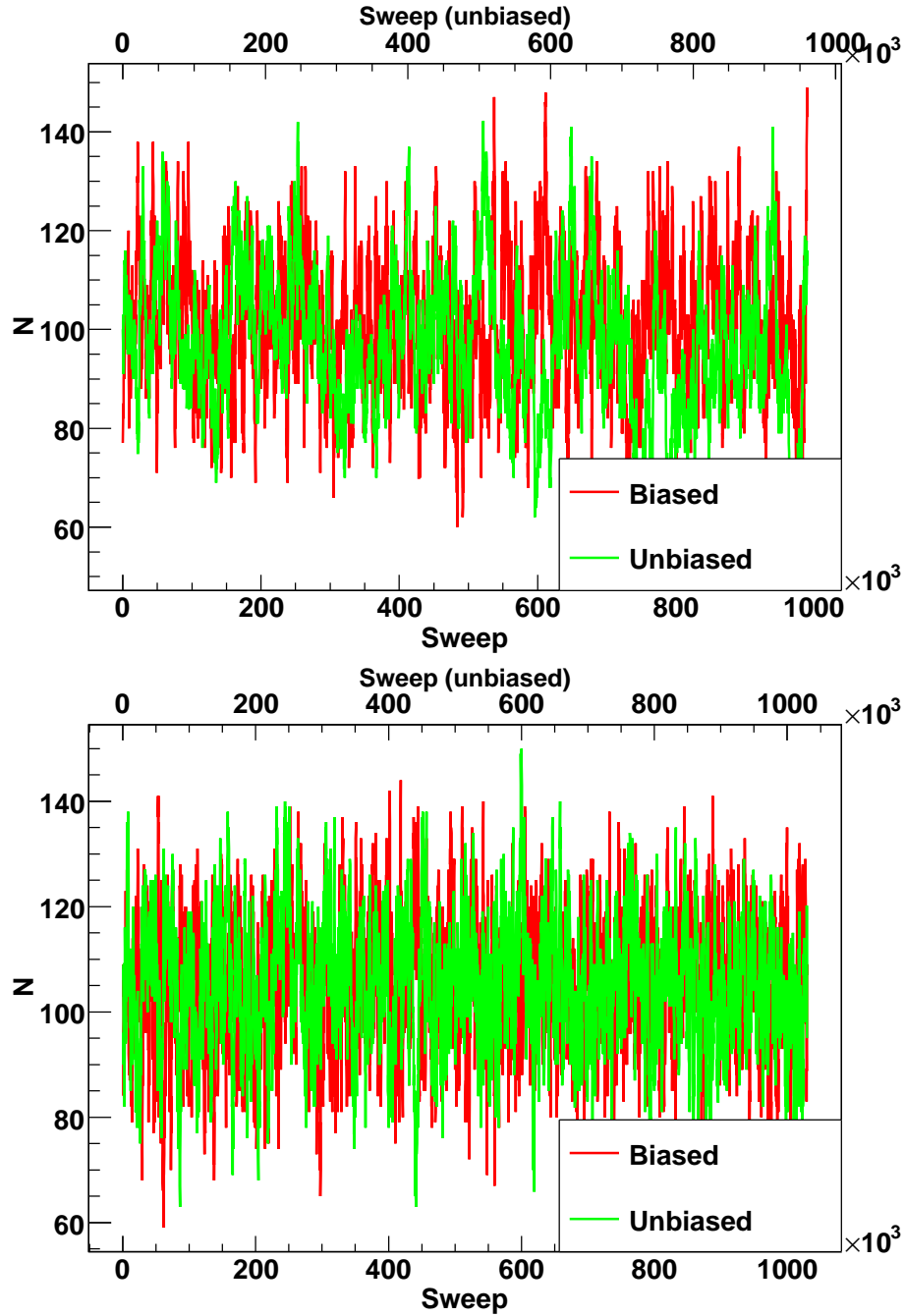


Figure 5.13: In the hadronic phase, i.e. without the plasma screening factor (5.17), biased and unbiased simulations agree very well. This is to be expected because we know from the $T = 0$ simulations that ordinary Monte Carlo leads to good sampling. These low- T simulations serve thus to check the biased Monte Carlo scheme.

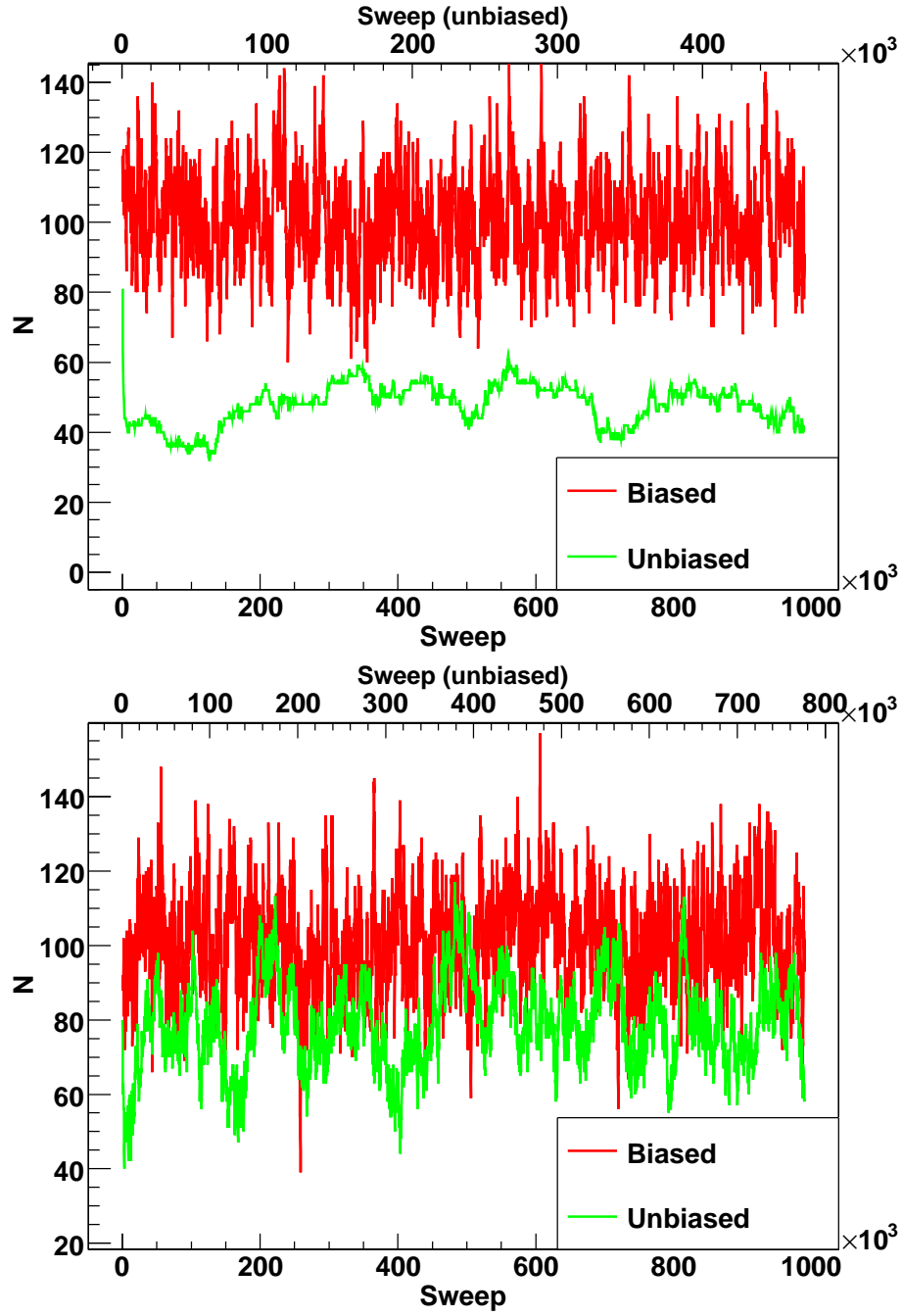


Figure 5.14: Around the confinement/deconfinement transition, modeled by (5.35), the formation of instanton–anti-instanton pairs is important as it drives the chiral symmetry restoration in the IILM. We can clearly see that for the small, physical quark masses (top) ordinary Monte Carlo fails to sample the ensemble correctly. For the large quark masses the effect is much less pronounced but there is a clear systematic difference in the mean instanton number. It can be attributed again to fewer pairs.

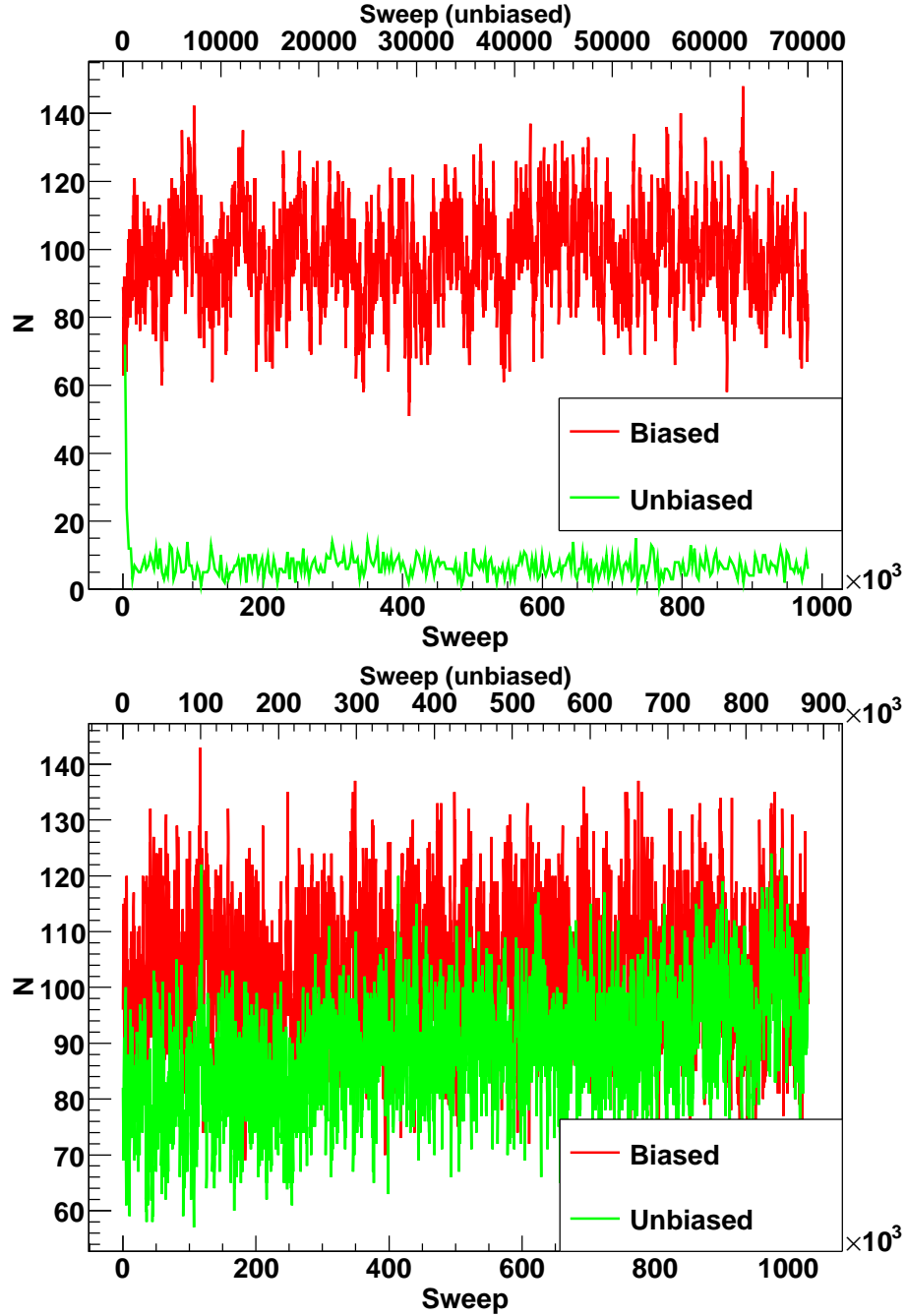


Figure 5.15: In the plasma phase, ordinary Monte Carlo fails drastically for the physical quark masses (top). This trend will, however, not persist to higher temperatures. Depending on the quark masses the energy-dominated pair configurations will be outweighed in the partition function by entropy-dominated configurations at sufficiently high temperatures. For larger masses this happens at lower temperatures (bottom).

completely to describe the system accurately for physical quark masses, see Fig. 5.15. In this case the number of pairs is very different, $\langle N_P^{\text{biased}} \rangle \approx 18$ and $\langle N_P^{\text{unbiased}} \rangle \approx 1$. Note, however, that the number of pairs has decreased for the higher temperatures.

Simulations at larger masses converge (slowly) to the correct equilibrium distribution. Actually, ordinary Monte Carlo perform better at high temperature than around T_* for these larger quark masses. The reason is that the ensemble equilibrates in a very dilute state, so that there are only a few instanton–anti-instanton pairs, $\langle N_P^{\text{biased}} \rangle \approx 4.5$ and $\langle N_P^{\text{unbiased}} \rangle \approx 1.5$. Note, again, that the pair concentration has dropped with respect to temperatures around T_* . It turns out that, quite generally, the molecule concentration drops with increasing temperature, because the energy-dominated pair configurations have a smaller and smaller entropy and are outweighed by the entropy-dominated configurations; the latter correspond to a truly random ensemble of instantons at a high enough temperature. This will happen for physical quark masses as well, albeit at higher temperatures.

To be a bit more precise, consider the contribution to the partition function of an instanton–anti-instanton pair, which we approximate by its dilute gas limit and by the contribution from the bonding box. The latter dominates, and hence biased simulations might be needed, if

$$1 < \left(\int d(\rho) \right) \Delta V^{\overline{\text{exp}}}(-S_{\text{int}}). \quad (5.36)$$

The contribution from the two instanton measures have cancelled on both sides, and we used $V = N/2 \int d$ for the dilute gas. The overbar is meant to indicate the average over the bonding box, size and colour orientations.

Using the 1-loop formula for the instanton measure and choosing one specific size with $T\rho = \text{const}$, this leads to

$$0 < \left[- \left(7 + \frac{N_f}{3} \right) + \left(11 - \frac{2}{3} N_f \right) (-V_{12}) + 2N_f - 4 \right] \log T + \text{const}. \quad (5.37)$$

We only display explicitly the temperature dependence for the one-loop instanton measure (first term), the gluonic interaction (second term), the quark interaction (third term) and the scaling with temperature of the bonding box

(fourth term). It turns out that the gluonic contribution is bounded by 5 for $N_f = 3$. Hence the $\log T$ term is negative and the dilute gas approximation becomes better and better as the temperature increases, see also Fig. 5.16. For higher temperatures, when more quark flavours become active, the trend seems to be reversed. Note, however, that this conclusion completely neglects the high-frequency quantum interaction, which should be investigated in the future to study this issue further.

The UB algorithm is only useful if the results do not depend strongly on the precise implementation, i.e. the actual definition of the bonding box. We checked this by using different bonding boxes for three different temperatures, i.e. above, around and below T_* , see Table 5.1. We found that the results agree very well for temperatures below T_* . This corroborates our expectations that biased Monte Carlo is not essential in that regime. For a modest sample of 200 independent configurations⁴, we also find rather good agreement for the higher temperatures where random sampling fails. We have also checked the dependence on the a-priori-probabilities⁵; we find again that the results only depend weakly on these choices for the small sample size we have used. Autocorrelation times, however, depend much more strongly on the bonding box; this allows us to fine-tune the parameters to achieve efficient sampling. The bonding box B_4 will be used for the final simulation.

5.2.3 Fermionic determinant

We argued that for sufficiently high temperatures, the IILM goes over to a dilute ensemble. It turns out that for high dilution the numerical manipulation of the quark overlaps becomes unstable because the overlap matrix becomes nearly degenerate. To render the evaluation of the fermionic interaction stable we need to decompose the overlap matrix into smaller non-degenerate blocks.

To achieve this, we truncate the interactions to zero above a certain cut-off; it is a function of the pair sizes and determines which instantons interact

⁴This relates to $\langle N \rangle$; the autocorrelation time for $\langle Q^2 \rangle$ was always smaller, effectively leading to a larger sample size. The better agreement on the latter quantity corroborates the expectation that the differences among bonding boxes vanish for infinite sample size.

⁵The data in Table 5.1 corresponds to one particular choice for the a-priori-probabilities.

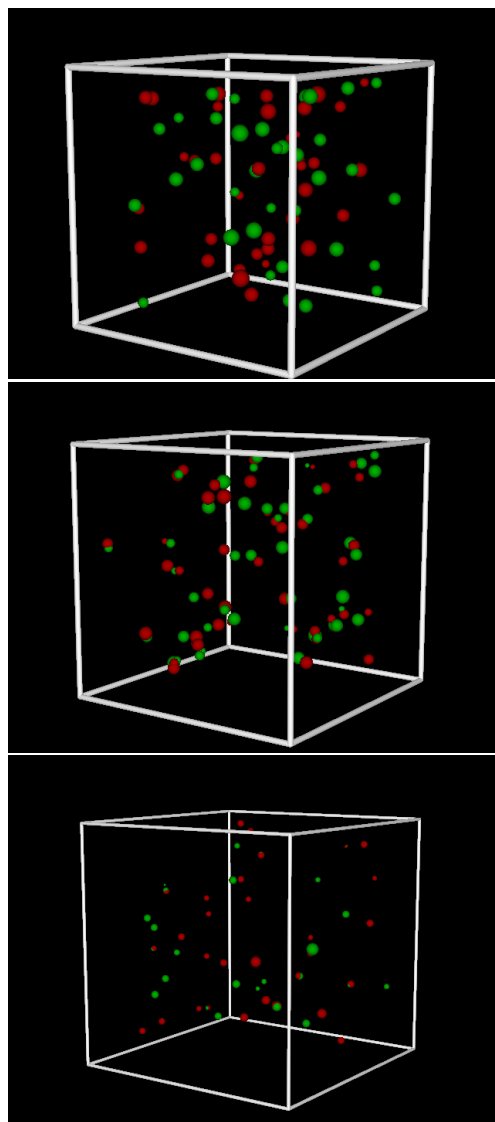


Figure 5.16: We display a (spatial) snap-shot of a typical configuration. For low temperature (top) spatial correlations are not very pronounced and the system equilibrates in a ‘random’ state. In the region $T \approx T_*$ (middle) a higher concentration of instanton–anti-instanton pairs can be seen, signalling the restoration of chiral symmetry in the IILM. At higher temperatures (bottom) the ensemble becomes dilute: the energy gain in pair-formation is outweighed by the large entropy gain of a random distribution of instantons throughout the box.

	B_1	B_2	B_3	B_4	B_5	B_6
$\langle N \rangle$	99(1)	100(1)	97(1)	102(1)	98(1)	98(1)
$\langle Q^2 \rangle$	0.81(1)	0.81(1)	0.81(1)	0.78(1)	0.81(1)	0.779(9)
$\langle \bar{u}u \rangle$	230(9)	219(8)	222(9)	217(9)	211(9)	225(8)
$\langle N \rangle$	105(1)	101(1)	96.7(9)	99.3(8)	98.6(9)	95.3(9)
$\langle Q^2 \rangle$	0.36(1)	0.31(2)	0.39(2)	0.30(2)	0.347(8)	0.349(6)
$\langle \bar{u}u \rangle$	84(8)	68(6)	80(8)	69(7)	75(7)	72(7)
$\langle N \rangle$	102.0(9)	106(1)	99.2(9)	106(1)	108(1)	102.2(9)
$\langle Q^2 \rangle$	6.9(3)	7.0(3)	7.0(2)	6.7(3)	6.8(2)	6.9(1)
$\langle \bar{u}u \rangle$	1320(30)	1300(30)	1300(40)	1200(30)	1310(40)	1210(30)

	B_1	B_2	B_3	B_4	B_5	B_6
x_i^{bb}	0.4	0.4	0.4	0.6	0.23	0.17
$x_{t,\min}^{bb}$	0.2	0.25	0.3	0.2	0.35	0.4
$x_{t,\max}^{bb}$	2	1.8	1.6	2	1.2	1

Table 5.1: The temperature increases from top to bottom. We find rather good agreement between the different bonding boxes. The sample has a modest size of 200 independent configurations for $\langle N \rangle$. The topological susceptibility has smaller autocorrelation times and a correspondingly larger sample size. The better agreement for $\langle Q^2 \rangle$ can be used to argue that the differences between the bonding boxes vanish with increasing sample size, as expected. The bottom table gives the dimensions, defined in Fig. 5.11, for the different bonding boxes B_i .

with each other. Given this cutoff, we can build clusters of pairwise interacting instantons. The overlap matrix T_{IA} is then decomposed into a direct sum according to these clusters. Not only does this render the numerical manipulations stable, but it also leads to a dramatic speed increase. Incidentally, we also use a cutoff for the gluonic interactions to get a better scaling with the number of instantons.

5.3 Quenched simulation

The main purpose of this section is to investigate how well the phenomenological screening factor (5.35) can model the confinement/deconfinement phase transition. To this end, we compare the topological susceptibility, for various values of the free parameters, to lattice data [6]. Normalising to the $T = 0$ result, we get $\Lambda = 206(8)\text{MeV}$.

We performed many simulations with different sets $\{T_*, \Delta T\}$. We report a few of these in Fig. 5.17, which is representative of the general findings. Namely, that the screening factors do not capture the lattice data well. In particular, we found that the IILM result for the topological susceptibility decays too fast: the power-law-like behaviour, e.g. $\chi \propto T^{-8}$ from the dilute gas approximation, at temperatures above T_* is not compatible with the available lattice data. The screening factor d_T^1 generically underestimates the lattice result for rather moderate temperatures. Only higher T_* could remedy this; however, for such values the topological susceptibility is too high in the low temperature regime. This can also be seen in Fig. 5.17 for the second screening factor d_T^2 : it has a gentler switch-on of the plasma screening effects, and also leads to rather large values for the topological susceptibility at low temperature.

We actually found that without any plasma screening, and therefore also in the low temperature region, the IILM produces a susceptibility that rises with temperature. This is not what the lattice predicts, namely an almost constant χ . The quenched IILM at finite temperature does not seem to give a good description of the pure Yang-Mills sector: that it fails around the phase transition could be expected since the crude phenomenological screening terms can only be expected to be a rough effective description;

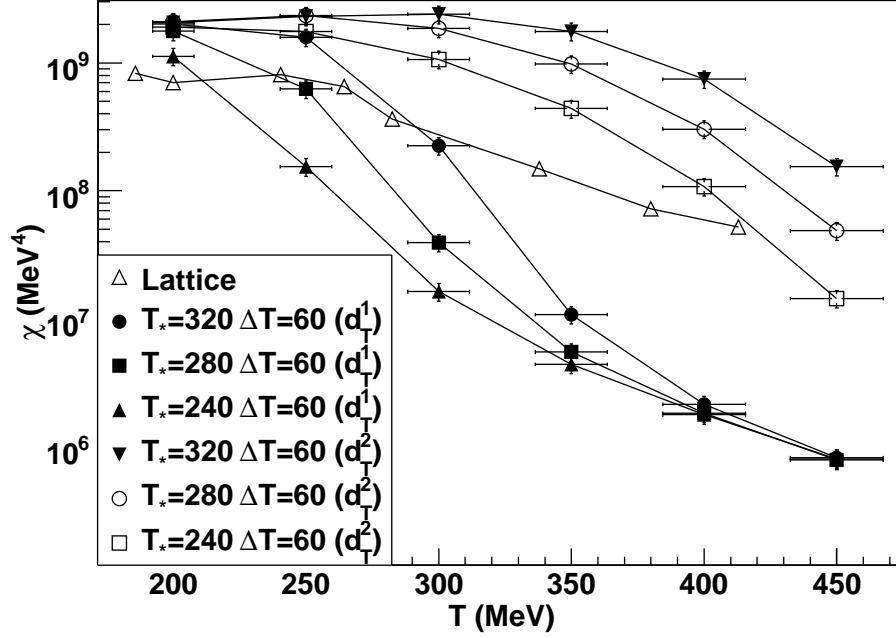


Figure 5.17: A striking feature of the quenched simulation is that the topological susceptibility decays generically too fast, as compared to lattice data. At low temperatures the IILM predicts a rising topological susceptibility, in contrast to the lattice data, which is roughly constant. The screening factor d_T^1 underestimates the lattice result, whereas d_T^2 tends to overestimate it at temperatures around T_* . Once the plasma screening is fully effective, both lead to too fast decays: the dilute gas approximation gives an approximate power-law of $\chi \propto T^{-8}$ which is not compatible with the available lattice data.

however, we find that the IILM fails to qualitatively describe the pure gauge dynamics at low and high temperature.

Given the fact that the dilute gas of non-trivial holonomy calorons [80] gave rather encouraging results, it might be that these are essential to model QCD in the quenched case. In particular, it will be interesting to find out whether these degrees of freedom can capture the smaller decay at high temperature or whether other degrees of freedom are dominant.

5.4 Chiral symmetry restoration

Compared to the quenched sector, full QCD has the additional property of (softly broken) chiral symmetry. Based on it, powerful analytical approaches

like chiral perturbation theory have been developed. A strong point of the IILM is its rather accurate description of the chiral properties of QCD. As compared to the quenched sector, this additional structure can be expected to improve the agreement between the IILM and the lattice at finite temperature.

We will repeat the analysis of the previous section for the unquenched IILM. Instead of the topological susceptibility, which we want to compute after all, we will use the chiral susceptibility [9] to estimate the parameters T_* and ΔT . Note that there is a controversy with regard to the lattice results for the critical temperature [52, 104, 8] with differences on the order of 20 MeV.

The chiral susceptibility is defined by

$$\chi_{\bar{q}q} = \partial_m^2 \ln Z_{\text{QCD}} = \int \int \langle \bar{\psi}\psi(x) \bar{\psi}\psi(y) \rangle - \left(\int \bar{\psi}\psi(x) \right)^2, \quad (5.38)$$

with Z_{QCD} the QCD partition function. It is actually composed of a connected and a disconnected part, and the latter is most sensitive to the chiral transition [105]. In terms of propagators both parts are given by

$$\chi_{\bar{q}q}^d = \left\langle \left(\int \text{Tr } S_A(x, x) \right)^2 \right\rangle_A - \left(\left\langle \int \text{Tr } S_A(x, x) \right\rangle_A \right)^2, \quad (5.39)$$

$$\chi_{\bar{q}q}^c = - \left\langle \int \text{Tr } S_A(x, y) S_A(y, x) \right\rangle_A + \chi_{\bar{q}q}^d, \quad (5.40)$$

and the expectation values are over the gluon fields. These formulas make it clear that the disconnected part is the main order parameter of the transition. Within the IILM, the propagator is the sum of a low and a high frequency part; the latter is usually just approximated by the free massive propagator. As such it gives an infinite contribution to the disconnected part and has to be subtracted at $T = 0$. A finite contribution remains at $T \neq 0$; it can be derived from the free partition function [103] and turns out to be small compared to the low-frequency contribution, and we will ignore it. This is expected, as the chiral transition is driven by the dynamics of the interacting instanton ensemble, and so the contribution from the free propagator should be negligible. By the same reasoning, we will ignore the cross terms between the low frequency and the free propagator. It is then straightforward to

compute the chiral susceptibility in the IILM. The propagator, truncated to the finite dimensional low frequency part, is given by

$$S_A(x, y) = (\not{D} + m)_{\text{low}}^{-1} = \sum_n \frac{1}{m - i\lambda_n} \xi_n(x) \xi_n^\dagger(y), \quad (5.41)$$

where λ are the eigenvalues of the matrix defined in (5.10); note that we used the fact that the set $\{\xi_n\}$ of zero modes is assumed orthonormal. The susceptibilities in the IILM are then given by

$$\chi_{\bar{q}q}^d = \left\langle \left(\sum_n \frac{1}{m - i\lambda_n} \right)^2 \right\rangle - \left(\left\langle \sum_n \frac{1}{m - i\lambda_n} \right\rangle \right)^2, \quad (5.42)$$

$$\chi_{\bar{q}q}^c = -\left\langle \sum_n \left(\frac{1}{m - i\lambda_n} \right)^2 \right\rangle + \chi_{\bar{q}q}^d. \quad (5.43)$$

For notational simplicity we have omitted the subscript A on the expectation value, but the average is again over the gauge fields. Remembering that the non-zero eigenvalues are paired due to chiral transformations, we find that

$$\chi_{\bar{q}q}^d = \langle (\bar{q}q - \langle \bar{q}q \rangle)^2 \rangle, \quad \bar{q}q = \frac{|\nu|}{m} + \sum_n \frac{2m}{m^2 + \lambda_n^2}, \quad (5.44)$$

$$\chi_{\bar{q}q}^c = -\left\langle \frac{|\nu|}{m^2} + \sum_n \frac{2(m^2 - \lambda_n^2)}{(m^2 + \lambda_n^2)^2} \right\rangle + \chi_{\bar{q}q}^d, \quad (5.45)$$

where $|\nu|$ is the number of zero modes and is related to the topological charge Q through the index theorem, $\nu = Q$.

Now, the chiral susceptibility in [9] is isospin symmetric, whereas we have broken it explicitly, i.e. $m_u \neq m_d$. In order to obtain effectively isospin symmetric results, we will compute

$$\chi_{\bar{q}q} = \frac{1}{2} \frac{\partial^2 Z}{\partial \bar{m}^2}, \quad (5.46)$$

with $\bar{m} = \frac{1}{2}(m_u + m_d)$ the mean quark mass and $m_u/m_d = \text{const.}$ This leads to

$$\bar{m}^2 \chi_{\bar{q}q} = \frac{1}{2} m_u^2 \chi_{\bar{u}u}^c + \frac{1}{2} m_d^2 \chi_{\bar{d}d}^c + m_u m_d \langle (\bar{u}u - \langle \bar{u}u \rangle)(\bar{d}d - \langle \bar{d}d \rangle) \rangle. \quad (5.47)$$

It turns out that the chiral susceptibility, with all plasma screening effects removed, peaks around $T_c \approx 120 \text{ MeV}$, and decays very slowly afterwards. Screening effects in the IILM tend to increase the dilution of the ensemble, and thus they will be responsible for the faster decay at higher temperature. Incidentally, with screening effects included down to very low temperatures, below T_c , the peak in the chiral susceptibility disappears completely; this gives a first constraint on T_* .

The lattice data shows a peak at $T_c \approx 160 \text{ MeV}$ ⁶ and no tuning of the parameters $\{T_*, \Delta T\}$ will shift the chiral phase transition in the IILM towards the lattice result.

To fix the free parameters, we simply demand that the IILM result for the chiral susceptibility does not exceed the lattice result at high temperature, i.e. that the rapid decay due to screening effects should be around $T \approx 160 \text{ MeV}$, and that its peak remains intact. We've computed the chiral susceptibility for four sets of parameters to estimate the systematics, see Fig. 5.18. The large errors are due to the large systematics in Λ that is used to set dimensions. By construction, the data below $T_c \approx 120 \text{ MeV}$ is very robust as the screening effects are almost completely negligible, and the location of the peak remains fairly unaffected. Above T_c the results do not depend too strongly on the parameters $\{T_*, \Delta T\}$, apart for the first curve (solid circles) which does seem to have too big a T_* -value. Given the phenomenological terms, see (5.35), and the constraint $T_* > 120 \text{ MeV}$, the IILM predicts a chiral phase transition at $T_c \approx 120 \text{ MeV}$.

One of the successful predictions of the IILM has been chiral symmetry restoration based on the formation of instanton–anti-instanton molecules [99], [100], [158]: through pairing up of instantons and anti-instantons, and the strongly localised quark wavefunctions, the Dirac operator spectrum develops a gap and the Casher-Banks relation tells us that the quark condensate vanishes, i.e. chiral symmetry is restored.

Naively we would expect that the pair concentration is highest around the phase transition. This is, however, not what we find and, interestingly, the number of pairs seems to drop at the phase transition. In fact, the

⁶The newer analysis in [8] results in a shift towards lower temperatures by about 5 MeV. This is not a significant change with regard to the IILM prediction of $T_c \approx 120 \text{ MeV}$ and we will ignore it in what follows.

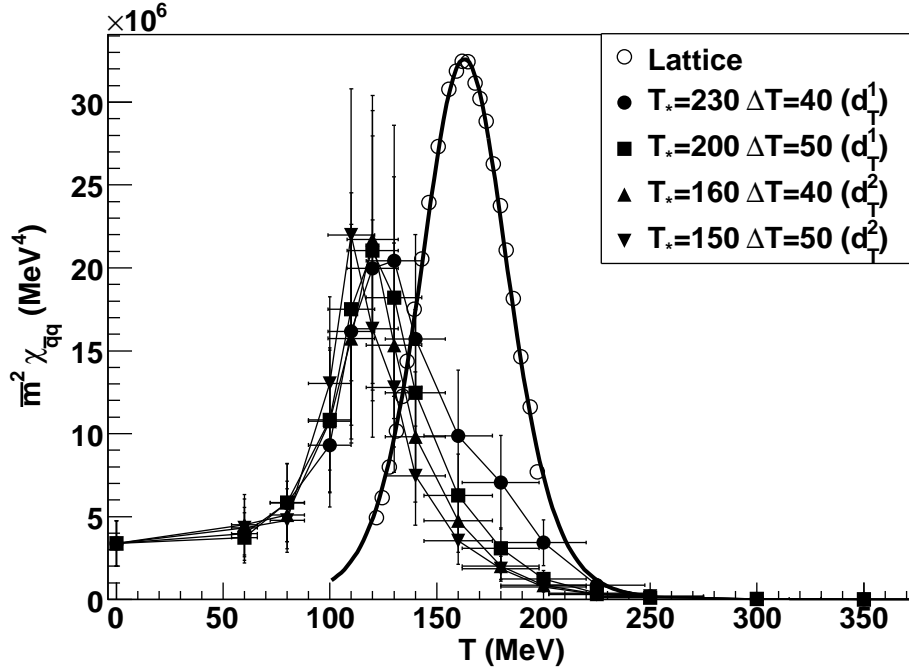


Figure 5.18: The chiral susceptibility in the IILM predicts a chiral phase transition at $T_c \approx 120$ MeV. Given the constraint that $T_* > 120$ MeV this is a robust result. The different parameter sets were chosen such that the plasma screening effects are effectively fully enhanced at about 160 MeV, the transition predicted by the lattice result (note that we use the continuum extrapolation of [9]). The first curve (solid circles) falls slightly short of this expectation. Taking the errors at face value (including those for the lattice result) the IILM is off by roughly 3σ . Considering the newer analysis [8] that shifts the lattice result to lower temperatures, the agreement improves slightly.

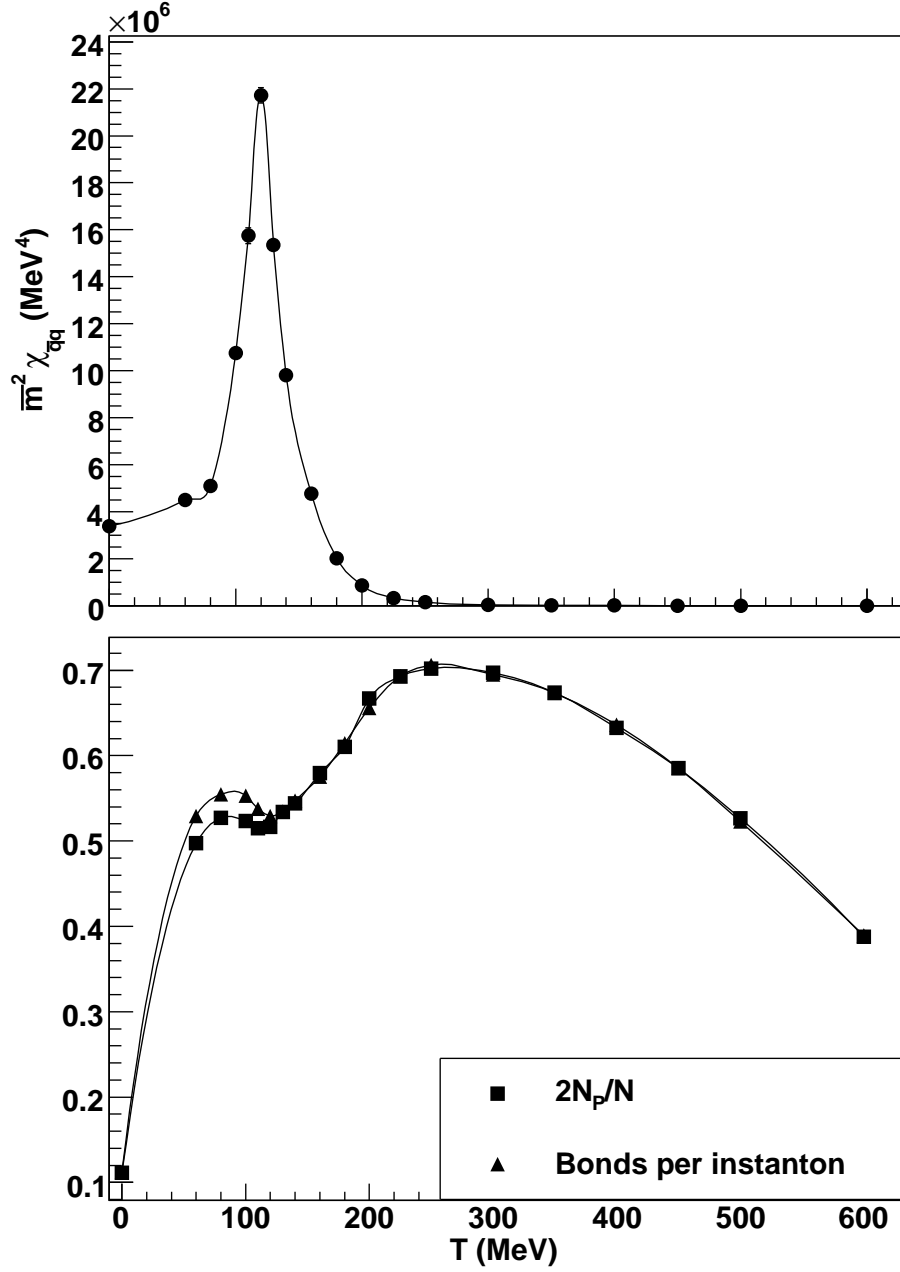


Figure 5.19: Contrary to expectation, the pair concentration does not peak at the chiral phase transition, although it is large with 50%, but, interestingly dips at the phase transition. Also, it keeps growing to 70% at roughly $T \approx 300$ MeV; beyond that point it decays.

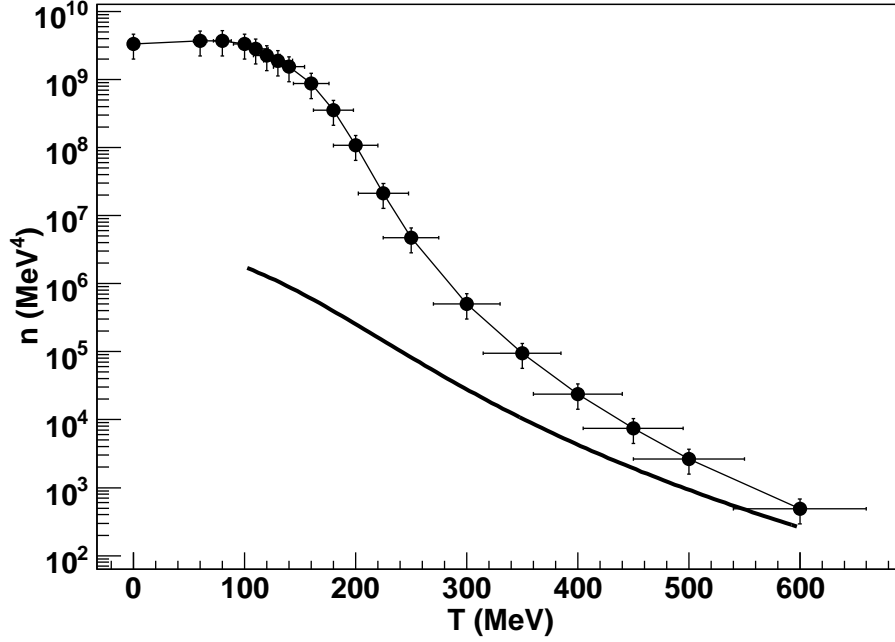


Figure 5.20: The instanton density compared to the dilute gas result. The data clearly shows that even for temperatures as high as $T = 400$ MeV the ensemble is not compatible with a dilute gas. The higher density in the IILM is due to the energy stored in pairs.

concentration of molecules keeps growing beyond T_c , see Fig. 5.19, and at T_c only half of the instantons are paired up into molecules. The maximum concentration of 70% is reached at $T \approx 300$ MeV, beyond which it starts to decay rather slowly. Note that even at $T = 400$ MeV the system is still far from a dilute random gas of individual instantons, see Fig. 5.20. It is, however, important to note that the identification of pairs is correlated to our specific definition of a bonding box, and even though the Monte Carlo results did not depend on the definition of the box, the number of instanton–anti-instantons does certainly depend on it.

Despite the fact that the ensemble is distinct from a random gas, the quark condensate tracks the dilute gas approximation already for moderately high temperatures, see Fig. 5.21. The reason is that the Dirac eigenvalues for pairs are large compared to the quark masses; this in turn follows from the fact that pairs have lined up along the time direction and thus their separation becomes smaller as the temperature increases. Therefore they

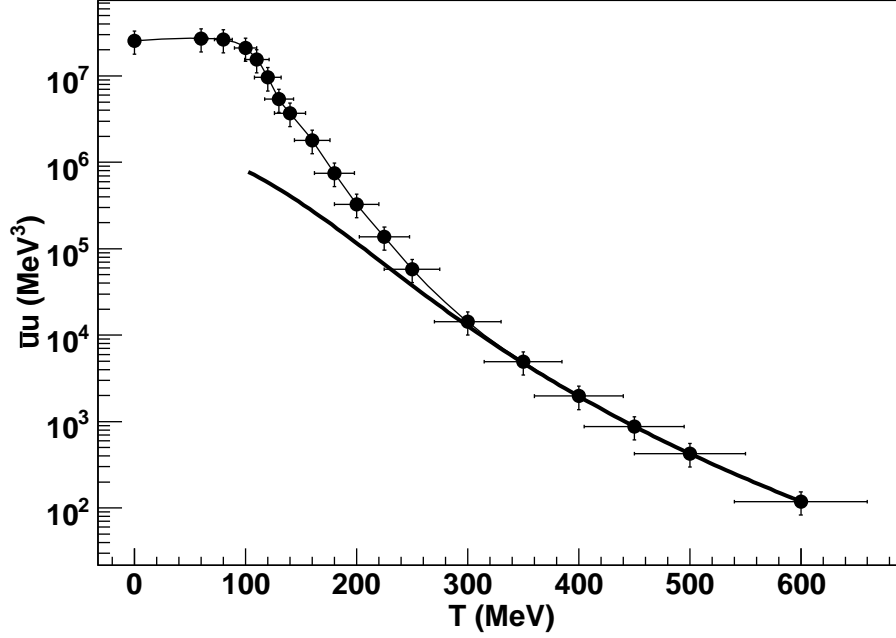


Figure 5.21: Although at $T \approx 300$ MeV the pair concentration is high, the quark condensate already tracks the dilute gas approximation (solid line).

are negligible compared to the zero modes. Bearing in mind the cluster decomposition of the quark overlap matrix T_{IA} , the quark condensate is given by

$$\langle \bar{q}q \rangle \approx \frac{1}{m_q} \sum_{\text{clusters}} \langle |\nu| \rangle \equiv \frac{\langle |\nu_{\text{eff}}| \rangle}{m_q}. \quad (5.48)$$

Note that this has almost the form of the dilute gas result, for which $T_{IA} = 0$ and hence $\lambda_n = 0$,

$$\langle \bar{q}q \rangle|_{\text{DGA}} = \frac{\langle N \rangle / V}{m_q}. \quad (5.49)$$

It turns out that $\langle |\nu_{\text{eff}}| \rangle$ has a temperature dependence that follows the dilute gas approximation for $\langle N \rangle$, see Fig. 5.22. The straightforward explanation that $|\nu_{\text{eff}}|$ is approximately equal to the number of unbonded instantons did not hold up to scrutiny, especially $\langle \# \text{unbonded} \rangle$ does not behave as $\langle N/V \rangle|_{\text{DGA}}$. Thus the population of unbonded instantons is still too large. But the clustering is defined through cutoffs in the interactions; these cutoffs can therefore also be used to identify a truly non-interacting sub-ensemble in the IILM. By definition, this population behaves according to the dilute gas

approximation. We find that its density becomes important at $T \approx 300$ MeV which explains the behaviour of the quark condensate.

We can conclude that, quite generally, the quark condensate is rather insensitive to the details of the molecule population. Presumably, only its concentration determines the effective number of zero modes ν_{eff} , respectively the density of the non-interacting instantons.

5.5 Topological susceptibility and axion mass

5.5.1 Topological susceptibility

The topological susceptibility approaches the dilute gas approximation for temperatures above approximately 250 MeV, and, like the quark condensate, does not seem to be affected by the high concentration of instanton–anti-instanton molecules, see Fig. 5.23.

Guided by the ideas of [100] that the IILM leads to a mixture of a highly correlated and a random component, it is natural to identify the unbonded instantons as the dominant contribution to the topological susceptibility; the molecules, at least at zeroth order, do not lead to charge fluctuations and thus cannot account for the topological susceptibility. For a random ensemble N_r it follows that

$$\chi \propto \langle N_r^2 \rangle - \langle N_r \rangle^2 \propto \langle N_r \rangle. \quad (5.50)$$

We found, however, that the population of unbonded instantons is far too large to be responsible for the small topological susceptibility, under the assumption that N_{ub} is a completely random ensemble, see Fig. 5.24. Again, as in the case of the quark condensate, it is the non-interacting instantons, not the unbonded instantons, that fit the bill. Note that for higher temperatures, $T \approx 600$ MeV, where the pair concentration is still quite high, the concentration of unbonded and non-interacting instantons becomes equal and there is no more distinction between the two. For lower temperatures a coupling between unbonded and bonded instantons still exists.

Note that, in the chiral limit, the topological susceptibility is related to the quark condensate through chiral perturbation theory. It is therefore consistent that the condensate and the topological susceptibility behave rather

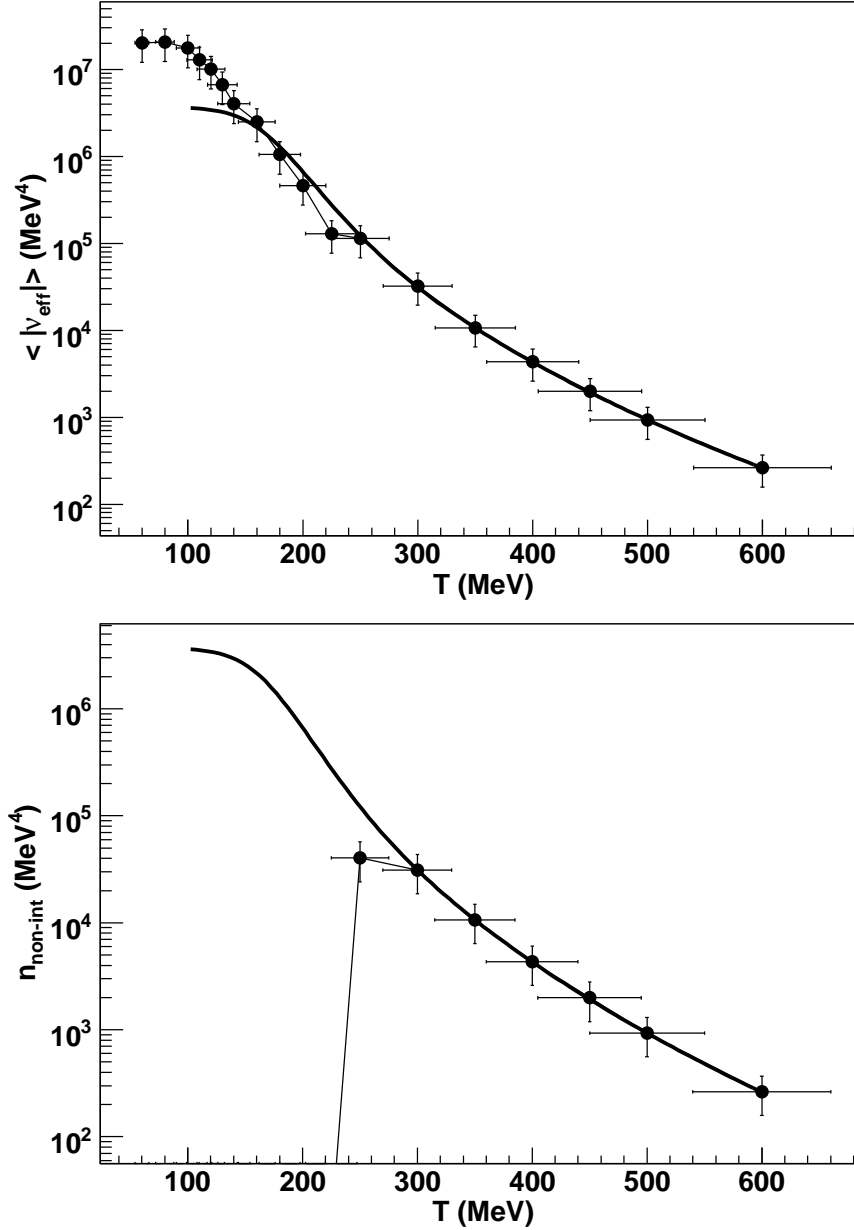


Figure 5.22: It turns out that the effective number of zero modes, $|\nu_{\text{eff}}|$, follows the temperature dependence of the dilute gas approximation for the density (solid line). Together with the observation that the condensate is dominated by ν_{eff} at high temperatures, see text, this explains why the condensate has a temperature dependence that follows closely the dilute gas result. At temperatures around $T \approx 300$ MeV we can attribute this effect to a sub-population of non-interacting instantons.

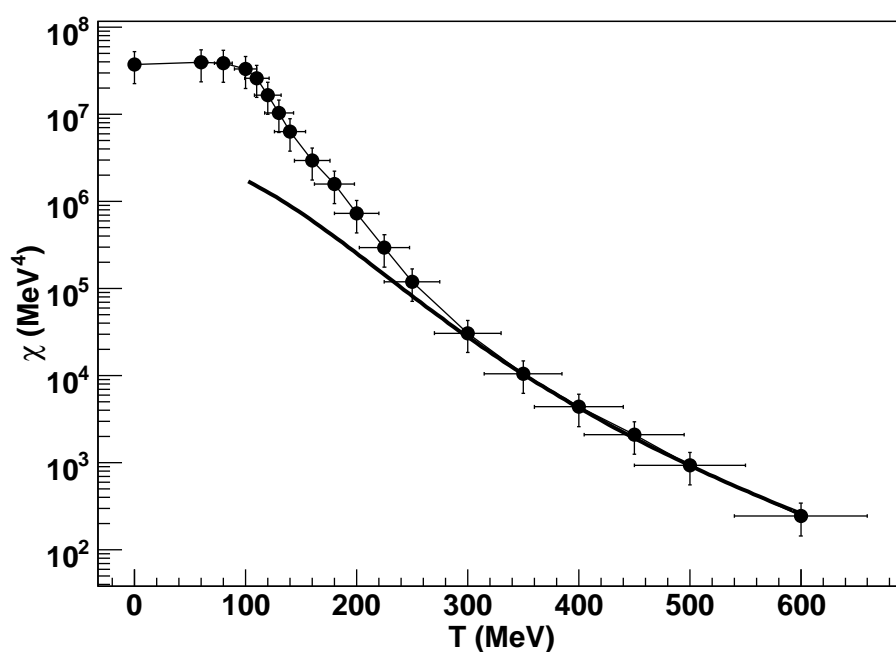


Figure 5.23: The IILM reproduces the topological susceptibility in the dilute gas approximation for temperatures where the ensemble is still far from dilute. We found that this phenomenon cannot be attributed to a population of unbonded instantons. Rather, it can be explained by the population of non-interacting instantons, Fig. 5.22, which is smaller than that of the unbonded instantons. At higher temperatures the distinction between unbonded and non-interacting instantons becomes ever more unimportant.

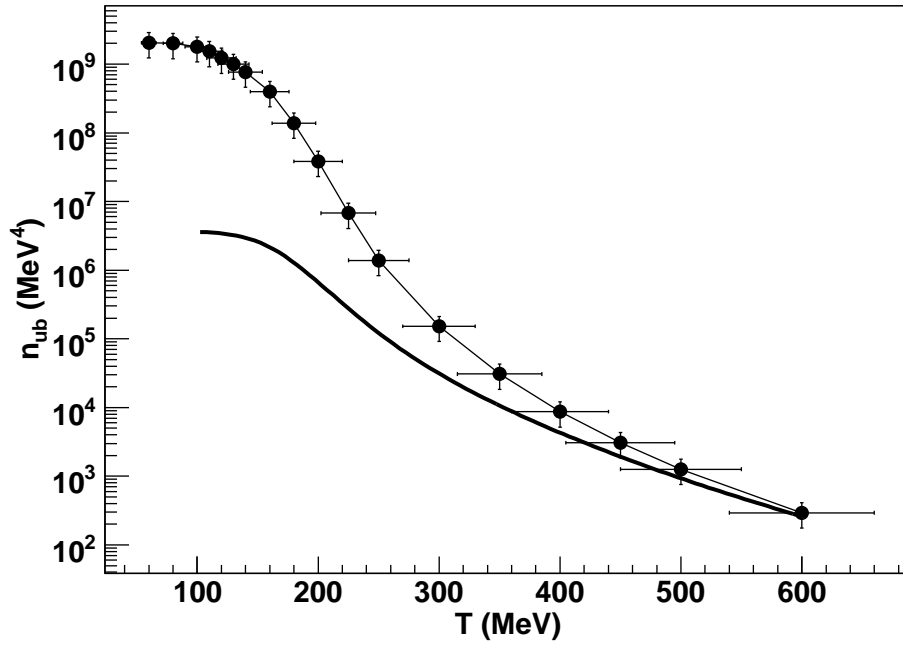


Figure 5.24: For moderately high temperatures, the population of unbonded instantons cannot be responsible for the topological susceptibility (solid line); we have checked that $\langle N_{ub}^2 \rangle - \langle N_{ub} \rangle^2 \approx \langle N_{ub} \rangle$. The unbonded instantons still interact too strongly with the highly correlated instanton–anti-instanton molecules. For higher temperatures, the unbonded instantons become indistinguishable from the non-interacting instantons that can account for the topological susceptibility. The latter instantons play the role of the random sub-ensemble along side the ‘molecular’ ensemble, following the ideas of [100].

similarly.

Finally, it is worth noting that the topological susceptibility, and also the quark condensate, are not overly sensitive to the free parameters T_* and ΔT , see Fig. 5.25.

5.5.2 Temperature-dependent axion mass

In order to solve the strong CP problem, Peccei and Quinn introduced a new field into the Standard Model of particle physics [146, 145]. It was soon realised that this field gives rise to a new light particle, a pseudo-Goldstone boson, the axion [206, 201]. At the time it seemed natural to tie the axion to the electro-weak scale but laboratory experiments have ruled out such an axion. Leaving the scale of the PQ field f_a free, and large to evade the previously mentioned constraints, the so-called ‘invisible’ axions were born, e.g. [57, 208, 112, 162].

Through their weak couplings to ordinary matter, invisible axions can play the role of a dark matter candidate. They have a rich phenomenology, and can be produced through a variety of production channels: the thermal scenario (similar to WIMP production), cosmic string decay (axions are the Goldstone boson of a spontaneously broken $U_{\text{PQ}}(1)$ symmetry) and the so-called misalignment mechanism. The latter is essentially the PQ mechanism: due to the anomalous $U_{\text{PQ}}(1)$ symmetry an axion mass term is generated through the coupling of the axion to the topological charge

$$\mathcal{L}_{a-g} \propto \frac{\phi}{f_a} \frac{g^2}{32\pi^2} F_{\mu\nu}^a \tilde{F}_{\mu\nu}^a. \quad (5.51)$$

This term combines with the vacuum angle θ and the axion field can be shifted to get rid of θ , i.e. the vacuum angle becomes a dynamical field, the axion angle θ_a . This is the key insight because now we can evoke the principle of least action to argue that $\theta_a \rightarrow 0$ dynamically to solve the strong CP problem.

Indeed, integrating out the gluons, we can determine the axion mass from the effective action,

$$\exp(-V V_{\text{eff}}(\phi)) = \int [dA] \det(\mathcal{D} + M) \exp(-S_g - S_{a-g}), \quad (5.52)$$

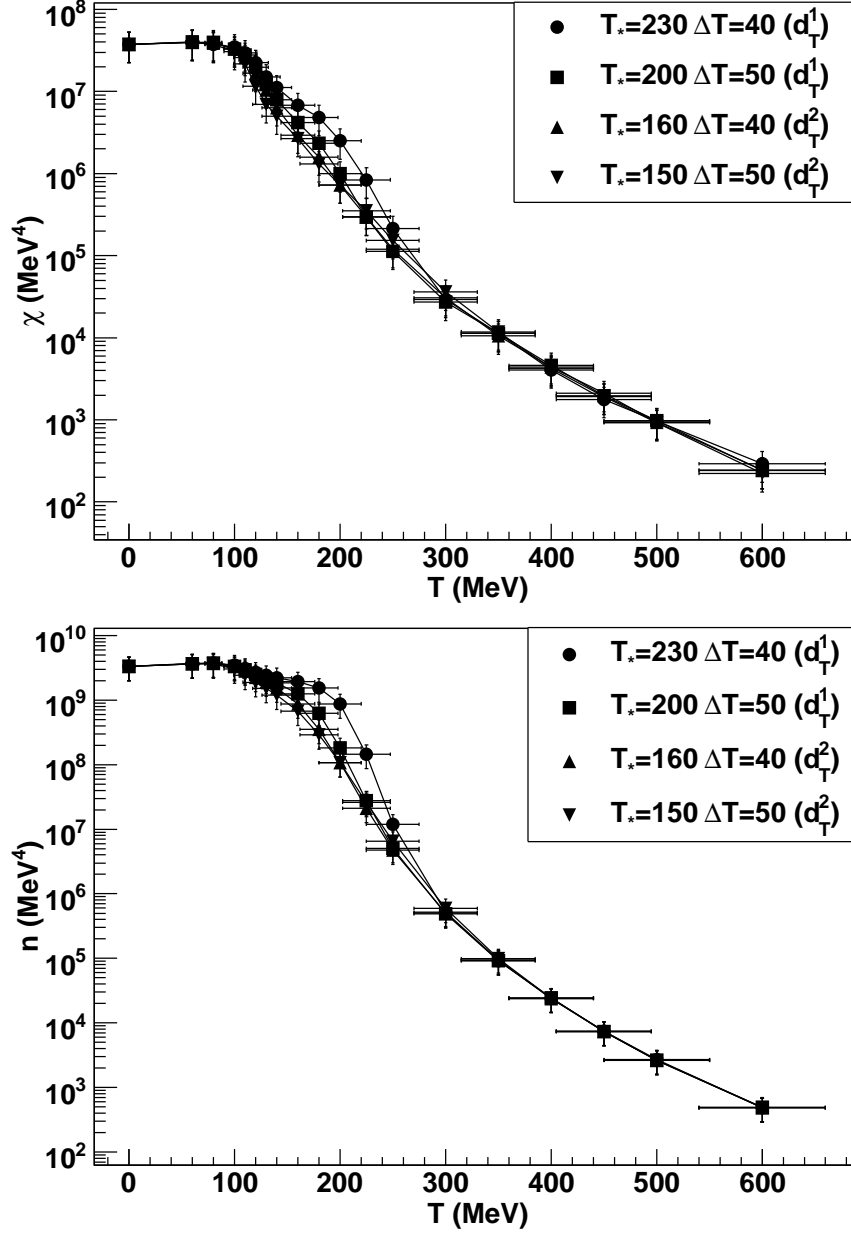


Figure 5.25: Both the topological susceptibility and the density are fairly constant for low temperatures, except for a very gentle growth from $T = 0$ towards a maximum at around $T = 100$ MeV. In contrast to the quenched case, the IILM is not as sensitive to the free parameters, $\{T_*, \Delta T\}$, that effectively describe the phase transition. The first curve (solid circles) could be argued to be unphysical as it delays the decay in a rather unnatural fashion.

according to

$$m_\phi^2 = \left. \frac{\partial^2 V_{\text{eff}}}{\partial \phi^2} \right|_{\phi=0} = \frac{\chi|_{\theta=0}}{f_a^2}, \quad (5.53)$$

and the axion eventually evolves to its minimum at $\theta_a = 0$ [189]. It also demonstrates that the mass for the QCD axion is set by the topological susceptibility.

Of all the production channels, the misalignment mechanism is most sensitive to the axion mass. From above it is clear that the axion mass is inherently a non-perturbative problem. At high temperatures, the dilute gas approximation to the instanton ensemble can be used. However, at lower temperatures it breaks down. From our determination of the topological susceptibility we can for the first time give a well-motivated axion mass that covers all temperatures down to $T = 0$.

In Fig. 5.26 we display the axion mass together with a fit and its error range mainly due to the error in Λ . The data suggests that the axion mass turns into the dilute gas approximation rather quickly; around $T \approx 300$ MeV the differences are negligible, and the fit takes this into account. We have seen in the previous sections that the IILM predicts a phase transition that is slightly too low. Using the lattice data for the phase transition temperature T_c , we also include a tentative fit to what the ultimate lattice data might look like, and again we impose the dilute gas limit at moderately high temperatures. It is worth noting that we currently do not have lattice data available that could corroborate such a result; remember that in the quenched case the lattice data did not behave according to the dilute gas result at high temperatures! Given the state of lattice calculations it should not take too much longer before a comprehensive lattice study with physical quark masses and across a wide range of temperatures will be available to give the exact axion mass.

Within the IILM some progress can be made by including the non-trivial holonomy calorons; a dilute gas study in the quenched sector has given encouraging results [80] that these degrees of freedom might play a role in the confinement/deconfinement transition. It would be very interesting to investigate their role in the unquenched sector, where we might expect less dramatic qualitative changes for chiral quantities, such as the topological

susceptibility, because chiral properties are reasonably well modeled by the IILM. However, we certainly expect a closer agreement with the lattice results if the QCD vacuum is indeed dominated by non-perturbative, topological fluctuations.

A second improvement with regard to finite temperature effects is the implication of quantum interactions on the plasma screening effects. In particular, we have found within a toy-model that the topological susceptibility can change qualitatively if the screening effects become subdominant to the quark zero mode interactions. This would favour a higher concentration of molecules and therefore reduce the non-interacting instanton density that sets the axion mass. Specifically, it would lead to a faster decay of the axion mass.

The IILM result has a slight rise at low temperatures. We do not have enough low- T data to determine the shape of this rise; however, in the gauge sector we found that the topological susceptibility had a roughly linear dependence on T . We will therefore constrain the fit at low temperatures to be a first order polynomial in temperature. It is given by

$$m_a^2 f_a^2 = 1.46 \cdot 10^{-3} \Lambda^4 \frac{1 + 0.50 T/\Lambda}{1 + (3.53 T/\Lambda)^{7.48}}, \quad 0 < T < 0.45, \quad (5.54)$$

where $\Lambda = 400 \text{ MeV}$, and the errors will be mainly due to the uncertainties in Λ .

Inspired by the lattice result for the chiral phase transition [9], we include a fit that delays the decay of the topological susceptibility until $T_c \approx 160 \text{ MeV}$. As mentioned above, there is a rather large disagreement between different lattice collaborations, and the chiral phase transition could occur at higher temperatures still. We assume again that the dilute gas limit will be recovered rather quickly above the phase transition. This assumption is hard to justify quantitatively given the lack of lattice data in that regime; the dilute gas result is really the best estimate we have at the moment. For the lattice inspired fit to smoothly connect with the dilute gas limit at around $T \approx 200 \text{ MeV}$, we patch together two different rational functions with the help of a washed-out step function, $\alpha(T) = \frac{1}{2} \left(1 - \tan \frac{T/\Lambda - 0.40}{0.075} \right)$. The result

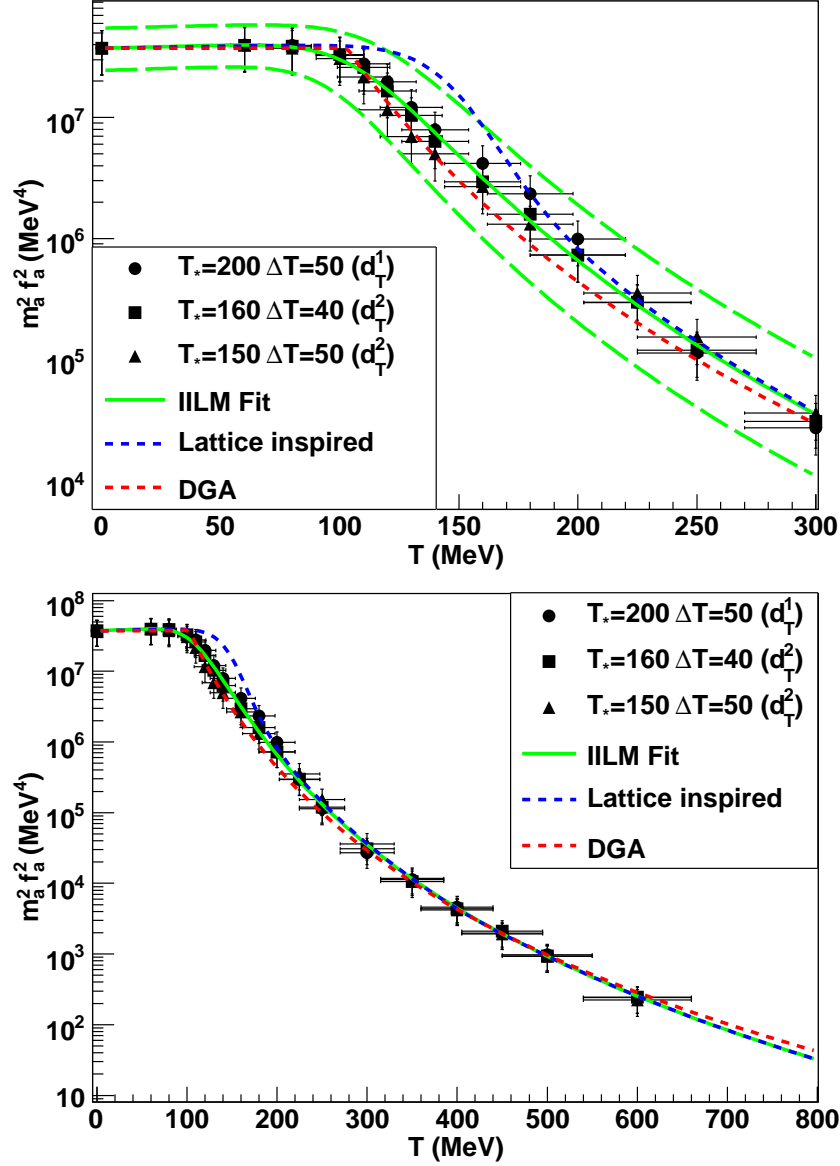


Figure 5.26: The mass for the QCD axion follows from the topological susceptibility, $m_a^2 f_a^2 = \chi$. The fit goes over to the dilute gas approximation for moderately high temperatures, in accordance to the IILM data. Note that the large errors are mostly due to the large uncertainties in the determination of Λ , used to set dimensions. We also include a similar fit that is slightly shifted towards higher temperatures to mimic the phase transition as seen on the lattice, and a simple power-law approximation to the dilute gas limit (DGA), see (5.59). We see that such a simple power-law approximation to the full result certainly has its merit as it is fairly accurate given the analytic simplicity.

is

$$m_a^2 f_a^2 = 1.46 \cdot 10^{-3} \Lambda^4 \left(\frac{1 + 0.3 T/\Lambda}{1 + (2.5 T/\Lambda)^{8.3}} \right)^{\alpha(T)} (1 + (3.4 T/\Lambda)^{7.4})^{\alpha(T)-1}, \quad (5.55)$$

for the temperature range $0 < T < 0.41 \text{ GeV}$.

These fits cover the low temperature regime, bounded by what we call T_{DGA} . Given our aim to improve on the current axion mass computations, we will also be more systematic for high temperatures and take the effects of quark thresholds into account. To that end, remember that we really are using the language of effective field theory: the results are given as a function of $g^{(3)}$, the strong coupling for $N_f = 3$ active flavours. It is well known that in order for S-matrix elements to be smooth, the ‘free’ parameters, such as masses and coupling constants, become ‘discontinuous’ [202, 143, 144, 155]; actually, the parameters are not discontinuous but belong to different theories that are matched at quark mass thresholds, i.e. $g^{(3)} = g^{(4)} + O((g^{(4)})^2)$. Since part of the quantum effects in the instanton density are two-loop improved, we expect that such ‘discontinuities’ do arise. In the present case the coupling in the exponential is the only place that displays two-loop improvements, and, at the threshold, the dilute gas approximations are related to each other by

$$\chi^{(4)} = \chi^{(3)} \exp(\gamma_{3,4}). \quad (5.56)$$

The factor $\gamma_{3,4}$ is essentially given by the Dirac determinant of the newly active quark flavour. In practice we just determine $\gamma_{3,4}$ such that the dilute gas approximations are smooth across the quark threshold.

At high temperatures the dilute gas result is almost given by a power-law. We found that the corrections could be straightforwardly taken into account by a higher order polynomial in log-log space; therefore the fit takes the form of an exponential of a polynomial in logarithms. We find

$$m_a^2 f_a^2 = \Lambda^4 \begin{cases} \exp \left[d_0^{(3)} + d_1^{(3)} \ln \frac{T}{\Lambda} + d_2^{(3)} \left(\ln \frac{T}{\Lambda} \right)^2 + d_3^{(3)} \left(\ln \frac{T}{\Lambda} \right)^3 \right], & N_f = 3 \\ \exp \left[d_0^{(4)} + d_1^{(4)} \ln \frac{T}{\Lambda} + d_2^{(4)} \left(\ln \frac{T}{\Lambda} \right)^2 \right], & N_f = 4 \\ \exp \left[d_0^{(5)} + d_1^{(5)} \ln \frac{T}{\Lambda} + d_2^{(5)} \left(\ln \frac{T}{\Lambda} \right)^2 \right], & N_f = 5 \end{cases}, \quad (5.57)$$

valid in the region $T_l^{(i)} \leq T < T_h^{(i)}$ and the different parameters are given by

N_f	$d_0^{(N_f)}$	$d_1^{(N_f)}$	$d_2^{(N_f)}$	$d_3^{(N_f)}$	$T_l^{(N_f)}$	$T_h^{(N_f)}$
3	-15.6	-6.68	-0.947	0.555	T_{DGA}	1.2
4	15.4	-7.04	-0.139		1.2	4.2
5	-14.8	-7.47	-0.0757		4.2	100

(5.58)

The γ_{N_f-1, N_f} -factors have already been absorbed into $d_0^{(N_f)}$; they are $\gamma_{3,4} = 0.444$ and $\gamma_{4,5} = 1.54$.

We conclude by giving a very simple approximation to the dilute gas result in the form of a power-law, as in earlier work [186, 13],

$$m_a^2 = \frac{\alpha_a \Lambda^4}{f_a^2 (T/\Lambda)^n}, \quad (5.59)$$

where $n = 6.68$ and $\alpha = 1.68 \cdot 10^{-7}$, from (5.58), and compare well with the more recent study [13]; it compares well with [13]. We believe it is a rather lucky coincidence that such a simple fit, based solely on the high temperature regime, still gives such a good overall approximation to the much more elaborate result of the IILM simulations, see Fig. 5.26. The small qualitative differences are that the power-law approximation: overshoots the IILM result at high temperatures, due to the wrong running of the QCD β -function, and underestimates the axion mass at low temperature where it is by construction constant whereas it reaches the $T = 0$ limit from above in the IILM. Given its analytic simplicity and its unexpectedly good agreement, such a simple power-law has certainly its merit. Whether such a conclusion pertains to an improved IILM including the more general calorons, and ultimately to the lattice, remains an open question.

5.6 Conclusion

We have been able to improve on the finite temperature interactions in the IILM. The numerical framework we set up in chapter 3 could successfully be implemented at finite temperature as well, and well-defined interactions that lead to a consistent thermodynamic limit have been derived.

Using these improved interactions, we investigated the IILM at finite temperature with light, physical quark masses. For these small quark masses we have found that the usual ‘random’ Monte Carlo sampling is very inefficient and can even break down. Using the results from chapter 4, where we introduced biased Monte Carlo techniques and, in particular, adapted the Unbonding–Bonding algorithm to the grand canonical ensemble, including a biasing scheme to deal with the orientation-dependent interaction of the IILM, we could run efficient simulations at finite temperature. We have found that the screening factors, from single instanton quantum fluctuations, will lead to a dilute, random ensemble at high enough temperatures. There is, however, a possibility that this trend could be reversed when more quarks become active; this does seem rather unnatural though. We want to point out that the high-frequency quantum interactions from overlapping instantons might be important to settle this question.

In the Yang-Mills sector we found rather poor overlap of the IILM and the lattice data. Most strikingly, we found that the IILM failed to reproduce the lattice data, even in a qualitative manner, in the regions where the phenomenological factors that mimic the phase transition in the IILM are unimportant, namely at low and high temperatures. To model the pure gauge sector, the IILM might have to be generalised to include the non-trivial holonomy calorons. Ultimately this is needed in any case because, as for the zero temperature IILM, confinement is still lacking and these more general degrees of freedom, the KvBLL calorons [119, 121, 120, 123], might play a crucial⁷.

In the unquenched sector, we investigated the chiral susceptibility and the quark condensate to gauge the free parameters introduced by the phenomenological screening factors; we found that the IILM is not overly sensitive to these, given some mild restrictions. The IILM unambiguously leads to chiral symmetry restoration at $T_c = 120$ MeV, slightly too low as compared to lattice data.

We investigated the population of instanton–anti-instanton molecules and found that, rather surprisingly, the maximum concentration does not occur

⁷Recent lattice studies see evidence of the lumpy structure characteristic for an ensemble of these new caloron solutions [101].

at T_c but rather at higher temperatures. A large population of instanton pairs prevailed to fairly high temperatures, indicating that the dilute gas limit of the instanton ensemble is only reached far beyond the critical temperature. However, the quark condensate and the topological susceptibility behave according to the dilute gas result much earlier. We could attribute this behaviour to a sub-ensemble of non-interacting instantons, distinct from the unbonded instantons that still interact considerably with the instanton–anti-instanton molecule population. At higher temperature, the distinction between unbonded and non-interacting instantons becomes irrelevant.

Given the topological susceptibility, we have presented a fit to the axion mass. We paid due attention to extrapolate the axion mass to higher temperature by including threshold effects due to heavier quarks that become active as the temperature rises. The main improvement, however, is a real computation of the low temperature axion mass that matches smoothly to high temperatures. Considerations of the anthropic axion for which θ and f_a are considered free parameters, one ‘environmental’ and the other fundamental, need knowledge of the axion mass for all temperatures, which this work provides. Comparison with lattice data leaves open the intriguing possibility that the high temperature axion mass does not behave according to the dilute gas result based on Harrington–Shepard calorons. A considerably different fall-off behaviour of the axion mass will change the cosmological bounds decidedly: the classic axion misalignment scenario, where θ_a is set by its rms fluctuations at the time of symmetry breaking, would get a weaker upper bound. This conclusion relies on our findings from the pure gauge simulations, where the lattice topological susceptibility did not fall off as quickly as predicted by the IILM. The unquenched case is, however, significantly different and chiral symmetry, which has proved powerful at zero temperature, might constrain the discrepancies between the lattice and the IILM. Especially, a higher instanton–anti-instanton molecule density, possibly due to the non-trivial holonomy calorons or weaker screening effects for the strongly overlapping pairs (the effective size of such a pair will be smaller), would lead to stronger upper bounds and could possibly rule out the classic axion window. This was one motivation for the present work, but within the present IILM this expectation could not be corroborated. Further

investigations are necessary to settle this issue. It might turn out, however, that the non-interacting sub-population that sets the axion mass in the IILM will not change considerably; such robustness within the IILM could be interpreted as evidence that the axion mass is rather insensitive to the details of the QCD phase transition. Given the advance of lattice QCD simulations, the ultimate axion mass determination might be available in the near future.

Chapter 6

Concordance Cosmology

The birth of the modern, quantitative discipline of cosmology can arguably be posited to have begun with the introduction of General Relativity (GR) by Einstein in the early 20th century. In pre-relativistic physics space-time was fixed and eternal, the background stage on which physical processes took place. At the time, the known interactions unfolding on this stage were Newtonian gravity and Maxwell's theory of electromagnetism. And just as these laid out the theoretical frameworks for astronomy, and electricity and magnetism, GR finally provided physicists with a dynamical theory to describe space-time itself. As such it gave the community, for the first time, a means to speculate about the origins of the universe within a precise mathematical framework.

Einstein himself did not embrace a dynamical universe at first. He introduced the famous cosmological constant to constrain the universe to remain static. Although not convincing from a theoretical point of view, as the solution is unstable, Einstein only gave up on his static universe when Hubble's observations provided experimental evidence for the cosmic expansion. Extrapolating backwards in time, the universe must have begun in a singularity, the Big Bang. It is generally seen as the beginning of the universe, when space and time, and matter were created. Extrapolating back over billions of years was, and is, a reckless procedure and alternative cosmologies were proposed, the most famous being the steady state theory by Bondi, Gold, and Hoyle in the late 40's. Only with the discovery of the cosmic microwave background radiation (CMBR) in the 60's, a definite prediction of the Big

Bang theory and very hard to reconcile with the ideas of the steady state theory, did the scales shift decisively in favour of the former. For an account of the other observational probes in favour of the Big Bang see for instance the discussion in [117], some of which will be alluded to in the next sections.

Taking the paradigm of the Big Bang at face value, the conditions of the early universe must have been extreme, i.e. very high temperature and pressure. Such environments are hardly found anymore in the universe around us, and are only achievable to some extent in laboratory experiments. This realisation led to a convergence of two disparate branches of physics: relativity and cosmology, and particle physics. The combined efforts of these communities, both theoretical and experimental, have lead to what is often called the era of precision cosmology, and created the field of particle cosmology.

The reason that particle physics can hope to address the early universe certainly relies heavily on the paradigm of unification: it is generally believed that at higher energies the different interactions will unify and the universe will become ever more symmetric, and hence more simple to describe. In such a scenario, phase transitions necessarily occur during the evolution of the universe, which can in principle be detected, validating or refuting models beyond the Standard Model of particle physics. On the other hand, the fundamental theories give the promise of theoretically investigating the original singularity at the Big Bang, if it exists at all. This has led again to alternative scenarios to the Big Bang, such as pre-Big Bang cosmologies [77] and cyclic universes [178], which future experimental observations will test, providing us with more insight into the origin of our universe.

6.1 General Relativity

Einstein's great insights are that the (apparently) fundamental gravitational force is really fictitious, i.e. it is the acceleration that test particles undergo due to their motion on a non-trivial manifold, and that the metric structure of space-time is influenced by the matter content itself. This feedback mechanism renders general relativity highly non-linear; the Einstein equations look,

however, deceptively simple

$$G_{\mu\nu} = \frac{8\pi}{m_{Pl}^2} T_{\mu\nu}. \quad (6.1)$$

Here, $G_{\mu\nu}$ is the Einstein and $T_{\mu\nu}$ the energy-momentum tensor, and m_{Pl} is the Planck mass. Note that the right hand side is not just a source term because it typically involves the metric tensor implicitly.

In order to define distances on a manifold, a metric $g_{\mu\nu}$ must be introduced. In (pseudo-)Riemannian geometry¹, the setting for general relativity, $g_{\mu\nu}$ provides a preferred covariant derivative operator ∇_μ . On a mixed tensor T of rank (m, n) it acts as (in components)

$$\nabla_\mu T_{\sigma_1 \dots \sigma_n}^{\rho_1 \dots \rho_m} = \partial_\mu T_{\sigma_1 \dots \sigma_n}^{\rho_1 \dots \rho_m} + \sum_i^m \Gamma_{\mu\nu}^{\rho_i} T_{\sigma_1 \dots \sigma_n}^{\rho_1 \dots \nu \dots \rho_m} - \sum_i^n \Gamma_{\mu\sigma_i}^\nu T_{\sigma_1 \dots \sigma_i \dots \sigma_n}^{\rho_1 \dots \rho_m}, \quad (6.2)$$

where $\Gamma_{\mu\nu}^\rho$ are the Christoffel symbols. They are uniquely determined through the condition of metric compatibility, which states that the metric should not change under parallel transport, i.e. $\nabla_\rho g_{\mu\nu} = 0$. The Christoffel symbols are then given by

$$\Gamma_{\mu\nu}^\rho = \frac{1}{2} g^{\rho\sigma} (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}). \quad (6.3)$$

The covariant derivative is constructed so as to preserve geometric quantities; in other words it transforms covariantly. Therefore, just as in the gauge theory case, the covariant derivative can be used to define new quantities transforming covariantly under the group of diffeomorphisms. In particular, the kinetic term is again constructed from the covariant derivative alone, see for instance [198]. In component form the Einstein tensor is given by

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}, \quad (6.4)$$

where $R_{\mu\nu}$ is the Ricci tensor and $R = R_\mu^\mu$ the Ricci scalar. The former is defined as a contraction of the Riemann tensor, i.e. $R_{\mu\nu} \equiv R_{\mu\rho\nu}^\rho$; the latter

¹Riemannian geometry has a positive metric, whereas in GR the metric is of Lorentzian signature, see for instance [135].

encodes the intrinsic curvature of the metric space, and is given by

$$R_{\mu\nu\rho}^{\sigma} = \partial_{\nu}\Gamma_{\mu\rho}^{\sigma} - \partial_{\mu}\Gamma_{\nu\rho}^{\sigma} + \Gamma_{\mu\rho}^{\alpha}\Gamma_{\alpha\nu}^{\sigma} - \Gamma_{\nu\rho}^{\alpha}\Gamma_{\alpha\mu}^{\sigma}. \quad (6.5)$$

Eq. (6.4) can be shown to follow from varying the Einstein-Hilbert action

$$S_{EH} = \frac{m_{Pl}^2}{16\pi} \int d^4x \sqrt{-g} R, \quad (6.6)$$

where g is the determinant of $g_{\mu\nu}$.

A useful rule of thumb to convert equations from Minkowski space-time to a space-time with generic metric $g_{\mu\nu}$ is to replace $\partial \rightarrow \nabla$. This ensures that the equations are generally covariant, the mathematical embodiment of the equivalence principle². Theories constructed accordingly are said to be minimally coupled. Two often used examples are the perfect fluid, whose energy momentum tensor reads

$$T_{\mu\nu} = \rho u_{\mu} u_{\nu} + p(u_{\mu} u_{\nu} - g_{\mu\nu}), \quad (6.7)$$

with ρ the energy density and p the pressure measured in the rest frame, and the scalar field with Lagrangian

$$L = \frac{1}{2} \nabla_{\mu} \phi \nabla^{\mu} \phi - V(\phi). \quad (6.8)$$

In the latter case the equations of motion read

$$\nabla_{\mu} \nabla^{\mu} \phi + V'(\phi) = 0. \quad (6.9)$$

6.2 Concordance cosmology

Cosmology is the study of the universe as a whole, and aims to explain the dynamical evolution of the cosmos from the time of the Big Bang until times of large scale structure formation. The subsequent stages in the evolution of the universe, e.g. galaxy and solar system formation and dynamics, are described by astrophysics.

²See [198] for more details and ‘exceptions’ to this rule.

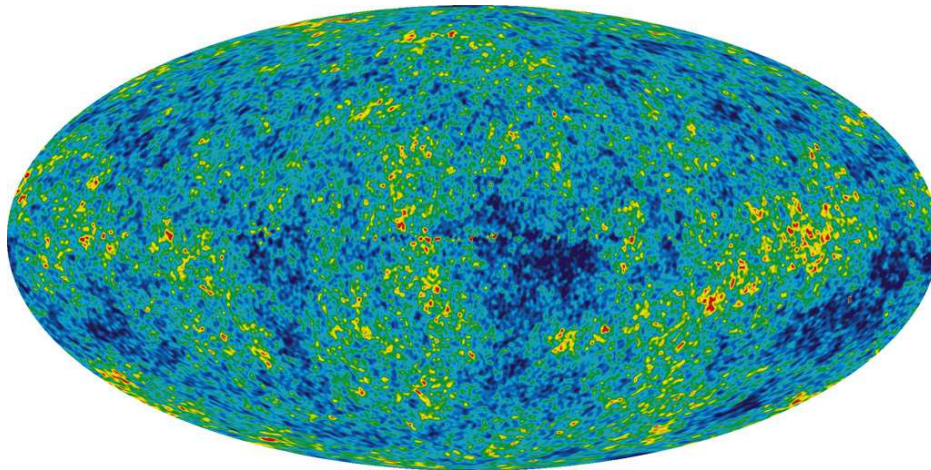


Figure 6.1: The CMBR temperature fluctuation map from the WMAP year 5 data release, from <http://map.gsfc.nasa.gov>. The fluctuations, encoded by a blue (colder) and red (hotter) colour scheme, are on the order of $\delta T/T < 10^{-4}$, demonstrating the very high degree of symmetry of the universe. Incidentally, it is also the most precise determination of the absolute CMBR temperature, $T_\gamma = 2.725(2)$ K from the FIRAS experiment, http://lambda.gsfc.nasa.gov/product/cobe/firas_overview.cfm.

6.2.1 Symmetry and FRW universe

Experimental observations of the temperature fluctuations in the CMBR, see Fig. 6.1, or of the large scale structure of the universe around us, for instance from high redshift galaxy surveys, see Fig. 6.2, suggest that our position is in no way special: the Copernican principle extends to cosmological scales, and in this context is sometimes called the cosmological principle. In concordance cosmology, this empirical fact of apparent isotropy and homogeneity of the observable universe³, can be explained by inflation, the very rapid expansion of the universe after the Big Bang that stretched out initial inhomogeneities.

Isotropy and homogeneity lead to a Friedmann-Robertson-Walker (FRW) universe, with infinitesimal line element

$$ds^2 = dt^2 - a(t)^2 h_{ij} dx^i dx^j, \quad (6.10)$$

³Mathematically, we can formulate the cosmological principle as the requirement that translations and rotations are isometries of the background space-time, see [198] for details.

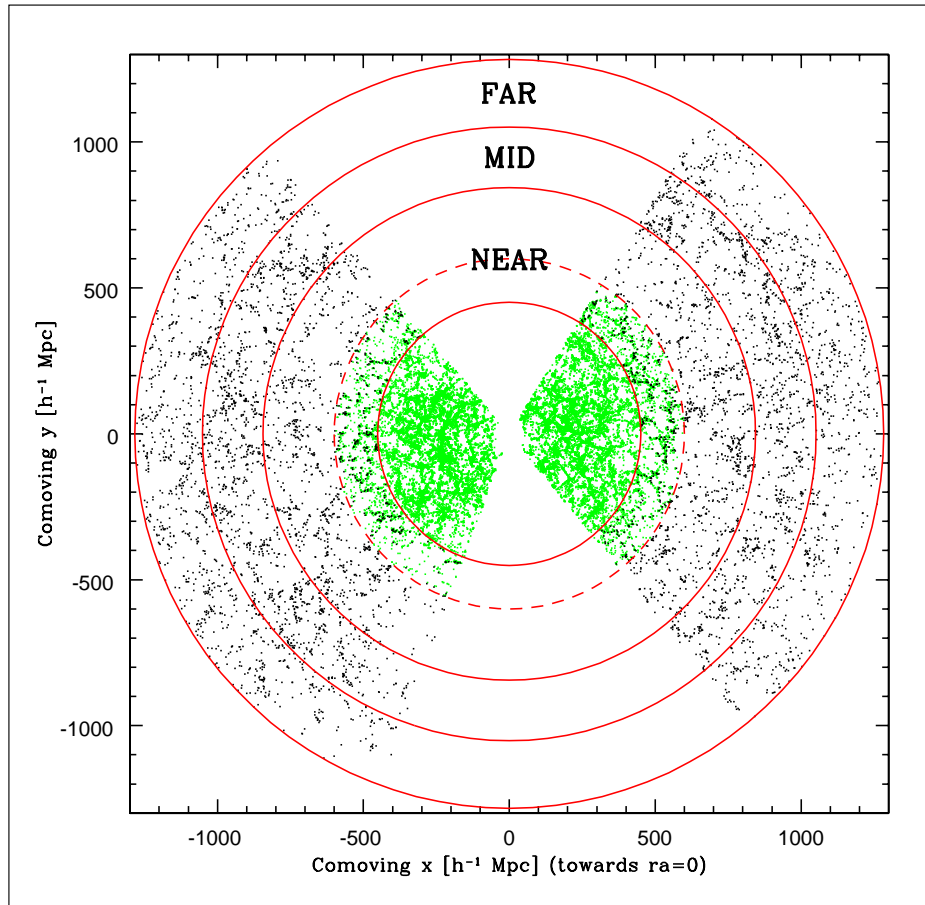


Figure 6.2: Distribution of luminous red galaxies (LRGs) in the Sloan Digital Sky Survey, from [183], gives evidence for the homogeneity and isotropy of the universe around us. The power spectrum extracted from large scale structure, especially baryon acoustic oscillations, can be used to test cosmological models and determine cosmological parameters; the Sloan Digital Sky Survey results for the year 5 data release is given in [147].

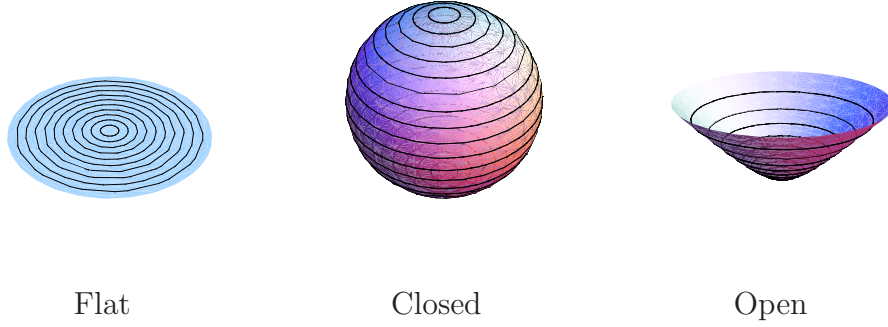


Figure 6.3: The spatial metric structure of FRW cosmologies is constrained to be one of three possible spaces of constant curvature, see [198]. We show here 2-dimensional examples: the disc, the sphere and the hyperboloid. Note that the hyperboloid $z = \sqrt{R^2 + x^2 + y^2}$ is embedded in a three dimensional Minkowski space with line element $dl^2 = dz^2 - dx^2 - dy^2$.

where $a(t)$ is the scale factor of the universe and h_{ij} is the spatial metric. The latter is restricted to represent a space of constant curvature of which there exist only three: flat (zero curvature), closed (positive curvature) and open (negative curvature), see Fig. 6.3. The time parameter t is often called cosmic time.

The Christoffel symbols are given by

$$\Gamma_{ij}^0 = \frac{\dot{a}}{a} h_{ij}, \quad (6.11)$$

$$\Gamma_{j0}^i = \frac{\dot{a}}{a} \delta_j^i, \quad (6.12)$$

which lead to the following expressions for the Ricci tensor

$$R_{00} = -3 \frac{\ddot{a}}{a}, \quad (6.13)$$

$$R_{ij} = \left(\frac{\ddot{a}}{a} + 2 \frac{\dot{a}^2}{a^2} + 2 \frac{k}{a^2} \right) h_{ij}. \quad (6.14)$$

From this the Ricci scalar follows

$$R = -6 \left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2} \right), \quad (6.15)$$

and $k = -1, 0, 1$ parametrises the spatial curvature (negative, flat, positive).

The dynamics of a FRW universe filled with a perfect fluid (6.7) follows from (6.1)

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi}{3m_{Pl}^2}\rho + \frac{\Lambda}{3} - \frac{k}{a^2}, \quad (6.16)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3m_{Pl}^2}(\rho + 3p) + \frac{\Lambda}{3}, \quad (6.17)$$

or equivalently

$$H^2 = \frac{8\pi}{3m_{Pl}^2}\rho + \frac{\Lambda}{3} - \frac{k}{a^2}, \quad (6.18)$$

$$\dot{\rho} = -3H(\rho + p), \quad (6.19)$$

where $H \equiv \frac{\dot{a}}{a}$ is the Hubble parameter. The parameter $\Lambda = (8\pi/m_{Pl}^2)\rho_\Lambda$ is the cosmological constant⁴. The equation of state, $p = w\rho$, closes the system. The Friedmann equation (6.16) connects the spatial curvature to the energy density of the universe, i.e.

$$\frac{k}{a^2} = H^2(\Omega - 1), \quad (6.20)$$

$$\Omega = \frac{\rho + \rho_\Lambda}{\rho_c}, \quad (6.21)$$

$$\rho_c = \frac{3H^2 m_{Pl}^2}{8\pi}, \quad (6.22)$$

and thus the universe is

- open if $k = -1$ or equivalently if $\Omega < 1$,
- flat if $k = 0$ or equivalently if $\Omega = 1$,
- closed if $k = +1$ or equivalent if $\Omega > 1$.

6.2.2 Expanding universe

General relativity applied to cosmology allowed for the first time a dynamical description of the universe. The constraints of homogeneity and isotropy

⁴In the Einstein-Hilbert action (6.6) replace $R \rightarrow R + 2\Lambda$.

lead to very simple equations, and the whole dynamics is encoded in the scale factor $a(t)$. As alluded to in the introduction, Einstein did not embrace this fact and introduced the cosmological constant to render the universe static: by fine-tuning Λ so as to maintain a vanishing right hand side for the Friedmann equation.

Taking the dynamical nature of the universe at face value, galaxies that we observe should display a Hubble drift

$$v = H_0 d, \quad (6.23)$$

where v is the velocity and d the distance of the galaxy; H_0 is the Hubble parameter today. Only after Hubble was able to corroborate the dynamical nature of the universe by his experimental measurements [97] of the galaxy recession velocities, according to (6.23), did Einstein embrace the concept of an evolving universe, and famously said the introduction of the cosmological constant to have been his greatest blunder.

The Hubble parameter H_0 is a very important experimental input as it enters almost all cosmological computations. Its measurement is, however, not a simple one, and until recently the experimental uncertainties were quite large. This has changed with the Hubble Space Telescope Key Project, which has measured $H_0 = 72(8) \text{ km sec}^{-1} \text{ Mpc}^{-1}$ [71], see also Fig. 6.4.

6.2.3 Inflation

The standard hot Big Bang cosmology faced a number of, now, classic shortcomings [117]:

- **The flatness-oldness problem:** the simple observation that our universe is so old, roughly 14 billion years, gave a first rather conclusive hint that our universe has the critical density $\Omega = 1$. This is because a universe this old cannot have been very different from flat as it would have either collapsed or undergone heat death. It is of course very encouraging that this anthropic reasoning has been put on firmer grounds by other measurements, most notably by the CMBR which gives the best evidence that our universe is very nearly flat. However, $\Omega \approx 1$ is dynamically unstable. To see it, we use (6.17), (6.20) and the equation

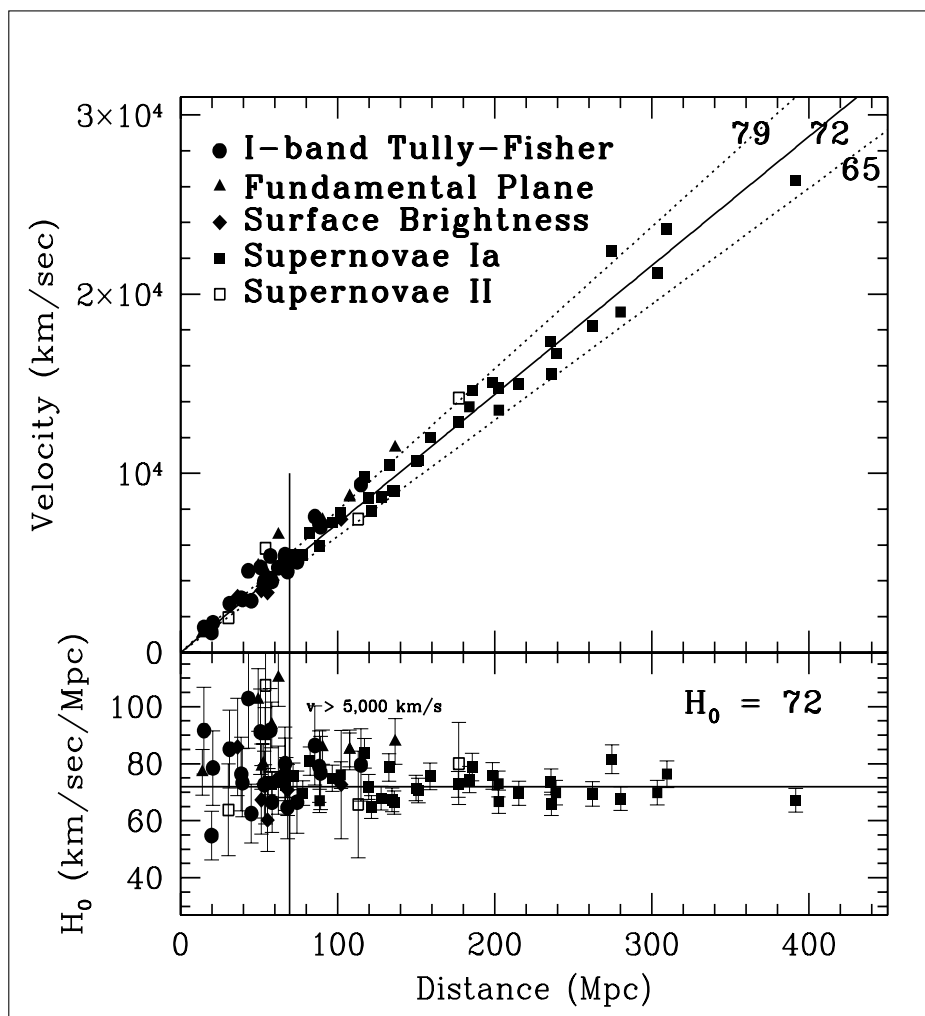


Figure 6.4: Final results from the Hubble Space Telescope Key Project to determine H_0 from galaxy recession speeds and redshifts [71]. A nice discussion about distances in cosmology and their experimental determination can be found in [204].

of state to get

$$\frac{d\epsilon}{\epsilon(1+\epsilon)(1+3\omega(\epsilon))} = \frac{da}{a}, \quad (6.24)$$

where $\epsilon = \Omega - 1$. To lowest order in ϵ , we find that a small initial deviation from flatness is amplified according to

$$\epsilon = \epsilon_0 \left(\frac{a}{a_0} \right)^{1+3\omega(0)}, \quad (6.25)$$

as long as $1 + 3\omega > 0$. Thus, to have a nearly flat universe today, just after the Big Bang the universe must have been very much closer to critical. Therefore, the flatness problem is a fine tuning problem. For different variations of the flatness problem see [127].

- **The horizon-smoothness problem:** measurements of the temperature fluctuations in the CMBR from all directions show only tiny variations. In a standard radiation and matter dominated universe this fact cannot be explained, see Fig. 6.5. It is not difficult to estimate how many causally connected patches at the time of last scattering have evolved, i.e. Hubble-expanded, into our present-day observable universe⁵. To this end, note that during radiation- and matter-domination the horizon evolves as $d_H \propto t$. The number of patches is then

$$N_p = \frac{V_H(t_0)}{V_H(t_{ls})} \left(\frac{a_{ls}}{a_0} \right)^3. \quad (6.26)$$

Using the connection between time and temperature during matter-domination, i.e. $t \propto T^{-3/2}$, and that during most of the universe's evolution $T \propto 1/a$, we can estimate that $N_p \approx (T_{ls}/T_0)^{3/2} = O(10^4)$. Thus, in standard cosmology the isotropic CMBR is a mystery.

- **Unwanted relics:** in the Standard Model of particle physics, and even to a greater extent for physics beyond the Standard Model, a number of phase transitions occur as the universe cools down after the Big Bang. During these phase transitions topological defects, such as magnetic monopoles, cosmic strings or domain walls, can be formed, see for in-

⁵After the time of last-scattering these patches do not interact anymore, and the correlations imprinted on the CMBR at that time remain fixed, modulo Hubble-expansion.

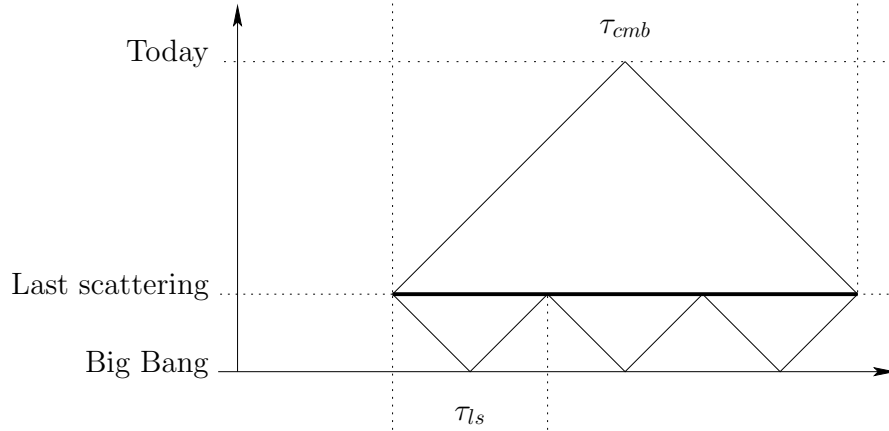


Figure 6.5: The Big Bang event is causally disconnected in standard radiation/matter dominated FRW cosmologies. In the conformal coordinates displayed here, this is encoded in the non-compact spatial extent of the initial singularity, leading to a particle horizon, see [198]. Therefore, at the time of last scattering not all regions of the, then, observable universe could have been in thermal equilibrium. Since that epoch the photons have free-streamed, in particular no interactions have been able to wash out the signatures set up during the time of recombination. Hence, we expect to see the induced “super-horizon” inhomogeneities in the CMBR, the lack of which has been a major puzzle. By pushing upwards the event of last scattering in the conformal diagram, we eventually encounter a situation where the CMBR that we measure today has originated from one causally connected primordial region. This stretching of the conformal diagram can be achieved at early times within the framework of inflation.

stance [193] for a comprehensive reference. Often a sizable abundance of these defects cannot be reconciled with standard cosmology, and a mechanism is needed to explain their absence.

- **Small structure:** although the universe is very smooth on large scales, we also need a way to explain the origin of small fluctuations that eventually, through gravitational instability, have evolved into galaxies. In particular, the theory should predict the temperature fluctuations in the CMBR.

It turns out that the theory of inflation, a very rapid expansion of the universe shortly after the Big Bang, provides a solution to all these problems.

From (6.17) we see that an accelerated expansion can be achieved provided $\rho + 3p < 0$ ⁶. In the standard scenario this is achieved by having a field stuck in a high energy state; as long as this field does not develop enough kinetic energy, it behaves essentially like vacuum energy, i.e. $p = -\rho$. As can be seen from (6.25), a flat universe becomes an attractor; as explained in Fig. 6.5 inflation also solves the horizon problem, and, provided no phase transitions leading to cosmologically unviable topological defects occur, the unwanted relics have been diluted away. Perhaps the most appealing aspect of inflation is that it provides a mechanism to explain structure formation through the amplification of quantum mechanical fluctuations, see for instance [58].

The precise mathematical implementation of this idea has, however, not yet been found. The simplest models, using single or multiple scalar fields, are plagued by some fine-tuning issues themselves. Also, scalar fields have so far eluded experimental detection. A theoretical motivation/justification to work with scalar fields might well come from fundamental theories such as string theory, where such fields abound and can sometimes be given a geometric meaning, which can arguably be seen as an improvement. Alternative cosmological scenarios, such as the cyclic universe [178], endeavour to connect high energy inflation, operating in the early universe, to the recently discovered low energy inflation, the so-called dark energy, that dominates the late time evolution of our universe [150, 153].

6.2.4 The dark sector

Even though we still lack a comprehensive implementation of inflation, the inflationary paradigm has become part of mainstream cosmology. The coincidences in standard hot Big Bang cosmology that inflation is able to explain gives it a strong theoretical support. Taking the paradigm of inflation at face value, we end up in particular with a critical universe, $\Omega = 1$. Precision measurements of the CMBR together with high redshift supernovae data and baryon acoustic oscillations all give experimental evidence for a flat universe.

The naive guess would have been that baryons make up the totality of the energy content of the universe, but it turns out that we live in a quite different

⁶We neglect the low-energy Λ -term that is usually identified with the late-time acceleration.

universe. First of all, it has been known for a very long time that baryons only account for a tiny fraction of Ω , obtained for instance from studying galaxy rotation curves to estimate the average galaxy mass. Extending these measurements beyond the luminous core of the galaxies, it became clear that most galaxies are surrounded by a non-luminous halo of matter, i.e. dark matter. Although substantially larger than the baryon contribution, the dark matter component of the universe is still not sufficient to explain a flat universe. This astrophysical evidence for dark matter, on galactic and super-galactic scales (from studying the motion of galactic clusters), is supplemented by evidence for dark matter on cosmological scales.

The above mentioned precision measurements, i.e. fits to the power-spectrum of the temperature fluctuations in the CMBR [118] and to the power spectrum of large scale structure (baryon acoustic oscillations) [147], and the determination of the deceleration parameter from high redshift supernovae data [154], give the best estimate for the amount of dark matter and dark energy, and lead to the astonishing fact that the latter, a homogeneous substance permeating the universe, contributes more than 70% to the energy content of the universe, see Fig. 6.6. Not much is known about dark energy. Possible explanations include vacuum energy, i.e. $\Lambda \propto \langle T_{00} \rangle$ together with anthropic reasoning to explain its tiny value [203] and the ‘modern’ version in terms of the string theory landscape [180], a very light homogeneous scalar field reminiscent of a low-energy inflaton, i.e. quintessence, see for instance [128], or alternative cosmology like the cyclic universe [178] among others.

Compared to dark energy, we know a lot about dark matter. It obviously has to be stable, or at least meta-stable on cosmological time scales. It turns out that the origin of the light elements, i.e. H, D, He, Li , is cosmological. Their abundances can be computed in Big Bang Nucleosynthesis (BBN), and are sensitive to the baryon density. From the experimental observations we can infer that dark matter cannot be baryonic. Large scale N -body simulations of galaxy formation, sensitive to astrophysical details of the halo structure, prefer dark matter to be cold, i.e. non-relativistic at the time of structure formation. The same conclusion follows from cosmological probes of large scale structure formation.

The fact that laboratory experiments have not detected particles with

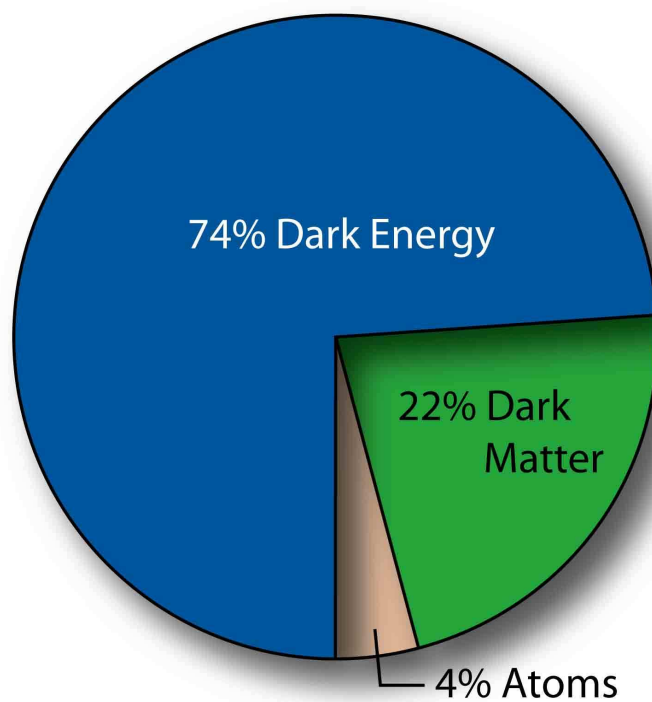


Figure 6.6: Energy composition of our universe, from <http://map.gsfc.nasa.gov>. Note that the very fact that we can write down unambiguously the different components as percentages, follows from $\Omega = 1$ which holds for a spatially flat universe.

the relevant characteristics gives strong evidence that dark matter should be weakly interacting. Such hypothetical particles are not produced copiously in the primordial plasma, and therefore need to be heavy to make up dark matter. Such weakly interacting massive particles (WIMP's) in fact constitute the prime candidate for cold dark matter, as they generically are non-relativistic during structure formation. Canonical examples are a heavy (sterile) neutrino, the lightest supersymmetric particle (LSP) e.g. the neutralino, the gravitino in supergravity; see for instance [14] for a textbook treatment. Taking the paradigm of string theory seriously, the heavier Kaluza-Klein modes from compactification of extra dimensions provide another class of WIMPs. For a comprehensive review on dark matter, discussing experimental evidence and constraints on dark matter, and a wide range of candidates, see for instance [26].

The new particle does not need to be a WIMP, however. If it is produced non-thermally, new possibilities open up. The prime example is the invisible axion that was never in thermal equilibrium and can be produced by coherent field oscillations or, topologically, by axionic string radiation. We will study the axion at length in the next chapter.

6.3 Thermodynamics in FRW

The early universe was filled with a hot plasma of particles and anti-particles, e.g. quarks, gluons, leptons, possibly some hypothetical dark matter particles, to name but a few. After recombination, matter cooled into a neutral atomic gas. The vast number of particles present make a statistical treatment mandatory. In the early universe most interaction rates were high enough to keep the cosmos in thermal equilibrium, and we can use thermodynamics. In particular, as long as the universe stays in thermal equilibrium, the evolution will be adiabatic.

During the cosmological evolution some particles will decouple from the heat bath. Particles decouple whenever their interaction rates with the heat reservoir drop below the Hubble expansion rate, at which stage their abundance is frozen in, and the particle density simply Hubble redshifts. See [117] for details.

The universe is possibly the best isolated system we can think of. As such the first law of thermodynamics is given by

$$dU = -p dV , \quad (6.27)$$

where p is the pressure, V the volume and U the energy of the system. Assuming that we stay in thermal equilibrium, the first (6.27) and the second law of thermodynamics,

$$T dS = dU + p dV , \quad (6.28)$$

with T the temperature of the universe, imply that the universe evolves adiabatically, i.e. $dS = 0$, as expected on physical grounds. It can be shown [117] that the entropy density is given by

$$s = \frac{S}{V} = \frac{\rho + p}{T} , \quad (6.29)$$

and $\rho = U/V$ is the energy density. In particular,

$$S = a^3 s = \text{const} . \quad (6.30)$$

Therefore, an adiabatically evolving universe has a specific relation between the temperature and scale factor, see Fig. 6.7. It allows us to relate cosmic time to the temperature of the plasma. This relationship holds as long as the universe remains in thermal equilibrium. Typical situations when (6.30) is violated are first order phase transition, because of the latent heat that is produced in the process; or an out-of-equilibrium decay of a heavy particle, an example of which is provided by the inflaton: at the end of inflation the universe is very cold, and has very low entropy, but the out-of-equilibrium decay of the inflaton produces all the known particles in the form of a hot plasma, which has very high entropy.

The connection with the microscopic degrees of freedom follows through the phase-space density function $f(x, p)$. In thermal equilibrium it takes the

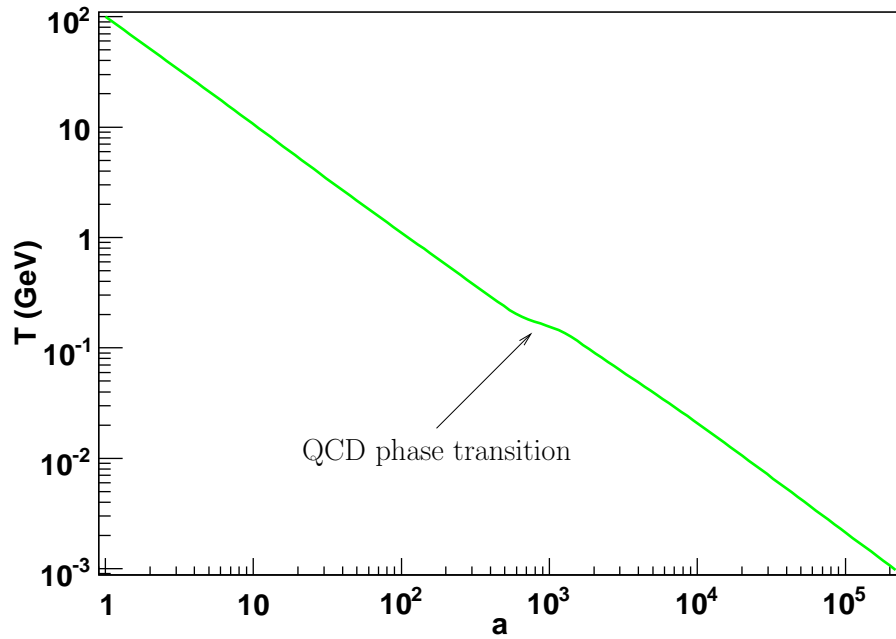


Figure 6.7: In an adiabatically evolving universe the scale factor and the temperature are related through the condition of constant entropy. Given the knowledge of the effective degrees of freedom $g_{*,S}$, it amounts to solving an implicit equation. The QCD phase transition occurs at around $T_{\text{QCD}} \approx 180 \text{ MeV}$, when the number of hadronic excitations rises very sharply, and $g_{*,S}$ is almost discontinuous; the would-be latent heat ‘reheats’ the universe, which is clearly seen in the graph.

well-known form⁷

$$f(p) = \frac{1}{\exp \frac{E}{T} \pm 1}, \quad (6.31)$$

with the upper sign for fermions and the lower sign for bosons. Note that by definition the equilibrium density function cannot depend on the spatial coordinates. The particle and energy density, and pressure are defined through

$$n = \frac{g}{(2\pi)^3} \int d^3p f(p), \quad (6.32)$$

$$\rho = \frac{g}{(2\pi)^3} \int d^3p E f(p), \quad (6.33)$$

$$p = \frac{g}{(2\pi)^3} \int d^3p \frac{p^2}{3E} f(p), \quad (6.34)$$

where $E = \sqrt{p^2 + m^2}$ and g is the degeneracy factor that gives the intrinsic degrees of freedom, e.g. $g = 2 \cdot 3 = 6$ for a quark. Summing over all the particle species, we get the total energy and entropy density

$$\rho = \frac{\pi^2}{30} g_{*,R} T^4, \quad (6.35)$$

$$s = \frac{2\pi^2}{45} g_{*,S} T^3, \quad (6.36)$$

where we normalise by the contribution from the photons, i.e. $T = T_\gamma$ ⁸. This defines the effective number of (relativistic) degrees of freedom $g_{*,i}$; they are given by

$$g_{*,R} = \sum_i \left(\frac{T_i}{T} \right)^4 \frac{15g_i}{\pi^4} \int_0^\infty dx \frac{\sqrt{x^2 + y_i^2}}{\exp \sqrt{x^2 + y_i^2} + (-1)^{Q_i^f}}, \quad (6.37)$$

$$g_{*,S} = \sum_i \left(\frac{T_i}{T} \right)^3 \frac{45g_i}{4\pi^4} \int_0^\infty dx \frac{x^2 \sqrt{x^2 + y_i^2} \left(1 + \frac{1}{3} \frac{x^2}{x^2 + y_i^2} \right)}{\exp \sqrt{x^2 + y_i^2} + (-1)^{Q_i^f}}, \quad (6.38)$$

where T_i is the temperature of species i , $y_i = m_i/T_i$, and $Q^f(\text{fermion}) = 1$

⁷We have neglected the chemical potential, which turns out to be negligible compared to the temperatures of interest.

⁸We choose the photon field, instead of some other relativistic species, because it will be the only relativistic field at sufficiently low temperatures.

and $Q^f(\text{boson}) = 0$. The energy and entropy densities are dominated by the contribution from relativistic species.

When the temperature of the plasma drops below some particle mass, their contribution to the entropy and radiation density drops to zero fairly rapidly. To compute these threshold effects correctly, the relevant phase-space integrals have to be evaluated, which is not possible in closed form in general. Following the analysis of [43], we include all the known hadrons⁹, up to a mass of 3 GeV, in the low temperature regime and match g_* smoothly to the plasma phase. Using a rough estimate of the available lattice data sensitive to the confinement/deconfinement transition [104], e.g. the Polyakov loop, we will use $T_c = 180$ MeV; given the spread in these results and also the cross-over nature of the transition [9], we apply the smoothing over a range of $\Delta T = 20$ MeV. This is slightly different from [43] which used $\{T_c = 180 \text{ MeV}, \Delta T = 5 \text{ MeV}\}$; the differences are small in any case.

If a species decouples, it will no longer be in contact with the heat bath, and its temperature will simply redshift due to the Hubble expansion. This is in contrast to the evolution of the plasma temperature which follows from entropy conservation, i.e. from (6.30). Thus, if a species becomes non-relativistic, it transfers its entropy only to those particles that are still in equilibrium with the heat bath, and reheats the plasma in the process. This leads to the different temperatures T_i for the decoupled fields.

In standard cosmology neutrinos decouple before e^\pm -annihilation, and so have a slightly lower temperature than the photons, $T_\nu = (\frac{4}{11})^{1/3} T_\gamma$ today. This result is valid at $T \ll 1$ MeV. For a general temperature T , the neutrino temperature is given by¹⁰

$$T_\nu = T \left(\frac{g_{*,S}(T)}{g_{*,S}(T_\nu^d)} \right)^{1/3}, \text{ if } T < T_\nu^d, \quad (6.39)$$

where $T_\nu^d \approx 1$ MeV is the neutrino decoupling temperature. It is a consequence of the fact that for decoupled (relativistic) particles such as neutrinos $T_\nu^i/T_\nu^f = a_f/a_i$, where the ratio of the scale factors follows from the adiabatic evolution of the universe. In fact, (6.39) is an implicit equation in

⁹“with strong experimental evidence”, see [7]; we use a multiplicity of $g_J = 1$ if the angular momentum is not known

¹⁰Assuming a sharp decoupling transition.

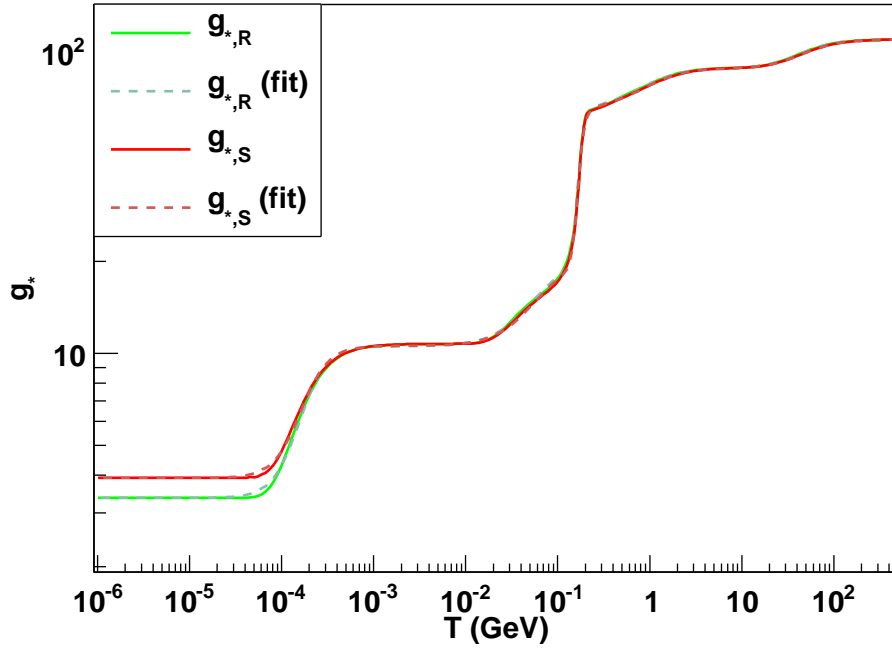


Figure 6.8: The effective degrees of freedom $g_{*,R}$ and $g_{*,S}$ are given for the temperature range up to $T \approx 100$ GeV. The decoupling of the neutrinos is included and manifests itself in the differences between $g_{*,R}$ and $g_{*,S}$ after e^\pm annihilation, when $T_\nu \neq T_\gamma$. We followed closely [43], but included some minor changes to take into account the better understanding of the QCD phase transition that recent lattice studies have put forth, see main text; the differences are small in any case. We have determined fits by using a sequence of smoothed step functions, see main text. As seen from the graph, the fits are rather accurate, generally below the 1%.

T_ν since $g_{*,S}(T)$ depends on it. With only neutrinos decoupling, it is easy to solve, and the result is (6.39) with the substitution $g_{*,S} \rightarrow g_{*,S} - g_\nu$. In order to reproduce the correct analytic ratio for low temperature, we have set $T_\nu^d = 5 \text{ MeV}$. This then leads to Fig. 6.8.

Evaluation of the exact numerical integration is fast. However, in chapter 7 we need to solve a system of ODE's and, in the process, relate the scale factor to the photon temperature to compute the axion mass. To integrate these equations in a reasonable amount of time, we have approximated $g_{*,i}$ by fits that are sums of smoothed-out step functions in log-log space. These fits are plotted in Fig. 6.8 and are generically accurate below the 1% level, except at the QCD phase transition and the e^\pm -annihilation where the accuracy is temporarily only around 4%. The fits are given by

$$g_{*,i} = \exp \left[a_0^i + \sum_{j=1}^5 a_{j,1}^i \left(1.0 + \tanh \frac{t - a_{j,2}^i}{a_{j,3}^i} \right) \right], \quad t = \log \frac{T}{1 \text{ GeV}}. \quad (6.40)$$

The corresponding parameters are

j	1	2	3	4	5
a_0^R	1.21				
$a_{j,1}^R$	0.572	0.330	0.579	0.138	0.108
$a_{j,2}^R$	-8.77	-2.95	-1.80	-0.162	3.76
$a_{j,3}^R$	0.682	1.01	0.165	0.934	0.869
a_0^S	1.36				
$a_{j,1}^S$	0.498	0.327	0.579	0.140	0.109
$a_{j,2}^S$	-8.74	-2.89	-1.79	-0.102	3.82
$a_{j,3}^S$	0.693	1.01	0.155	0.963	0.907

(6.41)

To our knowledge, such fits to g_* covering a large temperature range have not been given in the literature before. This can be traced to the fact that g_* is a slowly varying function for most temperatures; the two major exceptions are at e^\pm -annihilation and the QCD phase transition. That said, any investigation that covers a wide range of temperatures can benefit from our approximations to g_* ; we will use them to study the anthropic axion window in chapter 7. In addition, it is very easy to continue the exact g_*

to higher temperatures by including the particle content from one's favorite beyond the SM extension, and amend the fits by including some more 'step'-functions.

Chapter 7

Axions

Axions are still one of the best motivated cold dark matter candidates. Initially invented to solve the strong CP problem (“why is the QCD vacuum angle so small?”, i.e. $\theta < 10^{-9}$, see chapter 2 for details), it was soon realised by Weinberg [201] and Wilczek [206] that the Peccei-Quinn mechanism [146, 145] gave rise to a very-light pseudo-scalar Goldstone boson. In order to retain renormalisability, Peccei and Quinn introduced a new chiral symmetry, $U(1)_{PQ}$, on the quark and Higgs fields, that is spontaneously broken. This implies the existence of a new particle, a would-be pseudo-Goldstone boson, the axion; it receives a mass due to instantons because $U(1)_{PQ}$ is anomalous. In the original papers, the axion was incorporated in the electroweak sector but laboratory experiments soon ruled out such a light boson with GeV coupling. This gave rise to the so-called invisible axion models [162, 112, 57, 208], that a priori are not tied to any known energy scale. To constrain them, it was realised that such extremely weakly interacting particles could provide a new cooling mechanism for stars. The invisible axions have typically very weak couplings to ordinary matter. On the one hand, this makes their experimental detection difficult but, on the other hand, provides us with a well-motivated dark matter candidate. Refer to past reviews [113, 35, 187] for further details.

The axion has a rich phenomenology in that it can be produced thermally or non-thermally. The thermal production channel is the standard scenario for most WIMP’s [117]. Recently it was shown that the thermal axion cannot

contribute the dominant dark matter component of the universe [91]¹. The axion can also be produced non-thermally: after the spontaneous breaking of the PQ symmetry, the axion lives in a $U(1)$ vacuum manifold; such a broken field supports the formation of topological strings [48, 49, 50, 47, 19] [92, 86, 85], whose radiation produce axions. Finally, axions can be produced non-thermally through the so-called misalignment mechanism: at the QCD phase transition non-perturbative effects generate a mass, and the axion field relaxes to its minimum, which is precisely the PQ mechanism², invented to solve the strong CP problem. The oscillation around its minimum produce a coherent state of zero mode axions, i.e. a Bose-Einstein condensate [175]. This last production scenario is potentially sensitive to the QCD effects, i.e. the axion mass.

Because of the anomalous $U_{\text{PQ}}(1)$ symmetry, the axion has a two gauge boson interaction and can thus decay into two photons; such processes are used to look for axions experimentally, e.g. in solar axion searches and vacuum birefringence experiments. The former is one of the more stringent astrophysical constraints, the strongest coming from the analysis of the supernova 1987A neutrino flux which would be affected by axions. It gives a lower bound for the axion decay constant, $f_a \gtrsim 10^9 \text{ GeV}$. See [122] for a recent, comprehensive set of review articles.

In section 7.1 we discuss the effective axion potential, and essentially focus on the temperature dependent axion mass. We very briefly review our determination of the mass in the framework of the interacting instanton liquid model (IILM) from chapter 5. In section 7.2 we reexamine the cosmology of the vacuum realignment production mode in light of this new mass function: We solve the cosmological evolution equations numerically, and compare the results to the standard analytic approximation, identifying regimes in which present estimates are and are not robust. To make the numerics self-contained, we include the correct effective degrees of freedom for the entropy and radiation density and derived in chapter 6. For completeness,

¹The thermal axion bound follows from $(m_a^{th}/130 \text{ eV}) < \Omega_c g_*^{dec}/10$, and is saturated for $m_a^{th} \approx 15 \text{ eV}$. This bound is, however, excluded by the new astrophysics bound $m_a < 0.01 \text{ eV}$. Thus, $\Omega_a^{th} = \Omega_c(0.01/15) \approx 0.001 \Omega_c$.

²The θ angle, a free parameter, is replaced by a dynamical field that evolves to its CP-conserving minimum.

we also review and update constraints from axion string radiation.

7.1 Axion potential

As discussed in chapter 2, all axion models have the generic feature that the axion only couples derivatively to matter and the only non-derivative coupling is to the topological charge

$$\mathcal{L}_a^{\text{non-der.}} = i \frac{\phi_a}{f_a} \frac{g^2}{32\pi^2} F_{\mu\nu}^a \tilde{F}_{\mu\nu}^a, \quad (7.1)$$

where ϕ_a is the axion field and f_a the axion decay constant. This combines with the θ angle, and it can be shown that the effective potential for the axion field,

$$e^{-V_{\text{eff}}(\phi)} = \int [dA] \det(\gamma_\mu D_\mu + M) e^{-S + i(\theta + \frac{\phi}{f_a})\mathcal{Q}}, \quad (7.2)$$

has a CP-conserving minimum at $\langle \theta + \phi_a/f_a \rangle = 0$, where V the 4-dimensional volume. Note that the effective potential is periodic and that our computation is performed in the Euclidean theory. At finite temperature, real-time configurations, sphaleron transitions, might also give contributions. However, as was shown in [133], these classical field configurations do not affect the axion mass at leading order, basically because the classical field equations do not depend on θ . Shifting the axion field, we will define $\theta + \phi/f_a \rightarrow \theta_a$, with θ_a the axion angle. This is the PQ solution to the strong CP problem: θ , a free parameter, is traded for a dynamical field that evolves to its CP-conserving minimum.

Evaluation of (7.2) allows us, in principle, to determine the axion effective potential. We can of course not hope to get exact, analytic formulas since the effective potential includes strong coupling QCD effects. Ultimately, the lattice will be able to compute the axion effective potential exactly. However, there are technical problems since the action is complex and cannot be studied directly by simple Monte Carlo methods. This is the same problem that arises in lattice gauge theory with a finite baryon density, i.e. a non-zero chemical potential. In this chapter we will study the axion potential in the

framework of the ILM set up in chapters 3 and 5.

To summarise, in the ILM we saturate the partition function (7.2) with suitably chosen background configurations. In the non-interacting limit, the so-called dilute gas approximation (DGA), used so far for the axion mass determination, the path integral factorises into one-instanton contributions

$$Z_1 = \int d\rho d_0(\rho) d_T(\rho), \quad (7.3)$$

where $d \equiv d_0 d_T$ is the instanton measure. The zero temperature part d_0 is given by (3.42). At finite temperature, electric Debye screening prohibits large scale coherent field configurations to exist in the plasma, i.e. fields with a correlation length $\xi \gtrsim 1/T$. The finite temperature part to the instanton measure d_T is given in (5.17).

Given Z_1 , the dilute gas approximation follows immediately as

$$Z = \sum_{N_I, N_A} \frac{1}{N_I! N_A!} Z_1^{N_I + N_A} \exp(i\theta_a(N_I - N_A)), \quad (7.4)$$

where we have already included the axion angle through its non-derivative coupling to the topological charge, which is approximated by $Q = N_I - N_A$ in the background of N_I instantons and N_A anti-instantons. This sum is easily computed, and we find the effective potential

$$V_{\text{eff}}(\theta_a) = -2 \int d(\rho) \cos \theta_a. \quad (7.5)$$

That the potential had to be periodic follows from the defining QCD path integral (7.2). In general, the potential will have a much more complicated form than this simple cosine, see [89, 87, 88, 73, 74, 75]. Still, one can estimate that the first few terms of a harmonic expansion should describe the axion potential rather accurately [113], and it is custom for axion cosmology to work with a simple cosine.

Note that the effective potential is negative at its CP conserving minimum. The reason is that the one-instanton contribution is really normalised to the perturbative result, which by definition has zero energy. Instantons, interpreted as tunnelling effects between the n -vacua, restructure the vacuum

and give rise to the energetically lower true θ -vacuum.

At temperatures below the QCD phase transition, the dilute gas approximation breaks down. In that regime chiral perturbation theory can however be used to derive the classic result for the zero temperature axion mass

$$m_a^2 f_a^2 = m_\pi^2 f_\pi^2 \frac{m_u m_d}{(m_u + m_d)^2}, \quad (7.6)$$

with corrections of order one that depend on the precise model. The above axion mass is exact in the framework of the simplest hadronic axion [162, 112], where only a new super-heavy $SU(2) \times U(1)$ singlet Dirac spinor carries PQ charge; after integrating out the heavy field, we are left with the typical non-derivative coupling to the QCD topological charge. The low energy effective QCD Lagrangian, i.e. chiral perturbation theory, including the axion, is also used to derive the couplings of axions with pions and the weak sector, needed in the thermal production scenario [91]. Above the phase transition these coupling follow from the fundamental Lagrangian [132].

This same result can also be derived from the defining QCD path integral by noting that

$$m_a^2 f_a^2 = \frac{\partial^2 V_{\text{eff}}}{\partial \theta_a^2} \equiv \chi, \quad (7.7)$$

where $\chi = \lim_{V \rightarrow \infty} \frac{\langle Q^2 \rangle}{V}$ is the topological susceptibility. It can be shown that χ is related to the quark condensate [125] and, using chiral perturbation theory, this can be transformed into (7.6). Here, only the axion carries PQ charge and it interacts with matter derivatively, together with the topological coupling to the gauge sector. This axion field is the same as the physical axion field in the effective Lagrangian approach, i.e. the propagation eigenstate.

The important point to note is that the axion mass is essentially given by the topological susceptibility, a quantity that is routinely measured on the lattice. With the recent progress of lattice algorithms and increased computing power, physical quark mass simulations are finally feasible and the lattice will soon be able to provide us with the best estimate for the temperature dependent axion mass.

Beyond the DGA, the IILM saturates the path integral with an ansatz for the multi-instanton background configurations, and treats the low-frequency fluctuations ‘exactly’ while still assuming a factorisation of the high frequency

gluon, quark and ghost spectrum. The ILM partition function is defined by

$$Z_{\text{ILM}} = \sum_{N_I, N_A} \frac{1}{N_I! N_A!} \int \prod_{i=1}^{N_I+N_A} d\gamma_i d(\rho_i) e^{-S_{\text{int}}}, \quad (7.8)$$

$$S_{\text{int}} = \sum_{\text{pairs } (i,j)} S_0(\sqrt{\rho_i \rho_j}) V_{ij} - \sum_{q=1}^{N_F} \begin{cases} \ln \det(\mathbb{I} + \frac{TT^\dagger}{m_q^2}) & , Q < 0 \\ \ln \det(\mathbb{I} + \frac{T^\dagger T}{m_q^2}) & , Q > 0 \end{cases}. \quad (7.9)$$

The integration is over the collective coordinates, which are the positions, sizes and the colour embedding matrices. The classical gluonic two-body interaction is given by V_{ij} ; it receives a contribution from the high frequency fluctuations through charge renormalisation. The latter is approximated by the one-instanton action, $S_0 = 8\pi/g^2$, with the running coupling evaluated at $\sqrt{\rho_i \rho_j}$. The low frequency quark determinant is approximated by the finite dimensional subspace of quasi zero modes $\{\xi_n\}$, i.e. $T_{IA} = \langle \xi_I | \not{D} | \xi_A \rangle$. The details of the zero temperature calibrations are given in chapter 3. In particular we determined, self-consistently, physical quark masses. Finite temperature simulations have been performed to compute the topological susceptibility in chapter 5, and to derive the axion mass: the low temperature axion mass is displayed in Fig. 7.1 and given by

$$m_a^2 f_a^2 = 1.46 \cdot 10^{-3} \Lambda^4 \frac{1 + 0.50 T/\Lambda}{1 + (3.53 T/\Lambda)^{7.48}}, \quad 0 < T < 0.45, \quad (7.10)$$

where $\Lambda = 400 \text{ MeV}$; the high temperature mass, including threshold effects³, can be parametrised as

$$m_a^2 f_a^2 = \Lambda^4 \begin{cases} \exp \left[d_0^{(3)} + d_1^{(3)} \ln \frac{T}{\Lambda} + d_2^{(3)} \left(\ln \frac{T}{\Lambda} \right)^2 + d_3^{(3)} \left(\ln \frac{T}{\Lambda} \right)^3 \right], & N_f = 3 \\ \exp \left[d_0^{(4)} + d_1^{(4)} \ln \frac{T}{\Lambda} + d_2^{(4)} \left(\ln \frac{T}{\Lambda} \right)^2 \right] & , N_f = 4 \\ \exp \left[d_0^{(5)} + d_1^{(5)} \ln \frac{T}{\Lambda} + d_2^{(5)} \left(\ln \frac{T}{\Lambda} \right)^2 \right] & , N_f = 5 \end{cases}, \quad (7.11)$$

and the different parameters are given in (5.58). We also give a very simple approximation to the dilute gas result in the form of a power-law, as in earlier

³The quark thresholds are treated within the effective field theory language, where decoupling is enforced by hand and continuity is achieved through matching conditions.

work [186, 13],

$$m_a^2 = \frac{\alpha_a \Lambda^4}{f_a^2 (T/\Lambda)^n}, \quad (7.12)$$

where $n = 6.68$ and $\alpha = 1.68 \cdot 10^{-7}$, from (5.58); it compares well with [13]. We believe it is a coincidence that such a simple fit, based solely on the high temperature regime, still gives such a good overall approximation to the much more elaborate result of the IILM simulations, see Fig. 7.2.

In chapter 5 we found that the instanton ensemble is very distinct from a non-interacting system. Corroborating earlier ideas on the instanton liquid at finite temperature, we found a population of instanton–anti-instanton molecules and a non-interacting remnant. The molecules do not lead to charge fluctuations and, hence, the axion mass is determined by the random sub-ensemble. It turns out that the latter have a concentration that just matches the dilute gas approximation. We believe this is an unfortunate coincidence; in particular, we have found within a toy-model that, depending on the interaction and screening effects, a different high temperature behaviour can occur: for stronger interactions the molecule concentration can become higher so that the non-interacting sub-ensemble acquires a lower density, and hence a lower axion mass, compared to the dilute gas estimate. A crude argument within the IILM gave evidence that at higher temperatures, with more active quark flavours, the fermionic interactions might outweigh the screening effects and the molecule concentration could increase. For temperatures below the charm or even the bottom threshold, the molecule concentration will, however, decrease as the screening effects dominate over the interactions. That the molecule concentration would not depend monotonically on the temperature seems unnatural. Below the charm threshold, where we find that the molecule density decreases, weaker screening effects could alter this trend: corrections to the factorised high frequency quantum interactions could indeed induce weaker screening because overlapping instantons have effectively a smaller size; note that at zero temperature such quantum interactions were estimated to be subdominant but this has not been repeated for the finite temperature case.

In the pure gauge sector the IILM is not able to accurately describe the topological susceptibility as obtained from lattice simulations, whereas

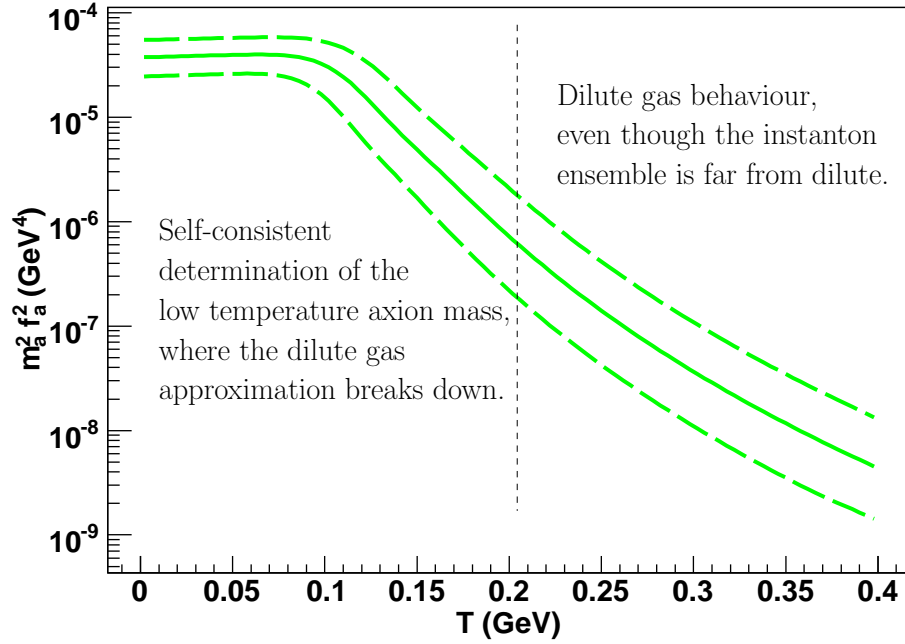


Figure 7.1: The mass for the QCD axion follows from the topological susceptibility, $m_a^2 f_a^2 = \chi$. The fit goes over to the dilute gas approximation for moderately high temperatures $T \approx 400$ MeV, in accordance with the IILM data. Note that the large errors are mostly due to the large uncertainties in the determination of Λ , used to set dimensions.

a dilute gas model of non-trivial holonomy calorons fared rather better [80]. In particular, the IILM predicts a topological susceptibility that decays too fast. These more general degrees of freedom should therefore be included into the IILM in the future to improve on the quenched sector and to investigate whether they also lead to significant changes in the unquenched case. We might expect the implications to be smaller because of chiral symmetry, which is successfully implemented in the IILM and believed to play a crucial role in the strong dynamics. In the unquenched case, we don't have lattice data available to check for a qualitatively different behaviour of the IILM in the high temperature region. Given the progress of lattice simulations in the physical regime, this issue will be settled in the near future.

Despite these uncertainties, our investigation might be evidence that the axion mass is fairly insensitive to the details of the instanton ensemble and that the dilute gas approximation might prove to be a reliable estimate, even though the instanton ensemble is certainly not non-interacting.

7.2 Axion cosmology

Now that we have determined the mass function for the axion, we can turn to the cosmological implications. For the misalignment mechanism to produce a dominant axion contribution, we assume that the axion is created before the end of inflation; otherwise axionic string radiation will produce the bulk of the axion population [48, 49, 50, 47, 19]⁴. Thus when the PQ symmetry is spontaneously broken, we have different initial angles, the misalignment angles, in the causally disconnected regions. Later, inflation sets in and stretches these patches to cosmological sizes such that throughout the observable universe the same misalignment angle prevails. Once instanton effects set in, the axion becomes massive and starts to oscillate.

Only the homogeneous part of the axion field is relevant for the misalignment mechanism. We are interested in the regime when the axion starts rolling, which happens around the QCD scale when the universe is radiation dominated. Recent lattice data suggests that the QCD phase transition is

⁴Note that there is still some controversy with regard to this statement, and other numerical work [92, 86] finds the axion contribution from string radiation and decay to be comparable to the contribution from the vacuum realignment production mode.

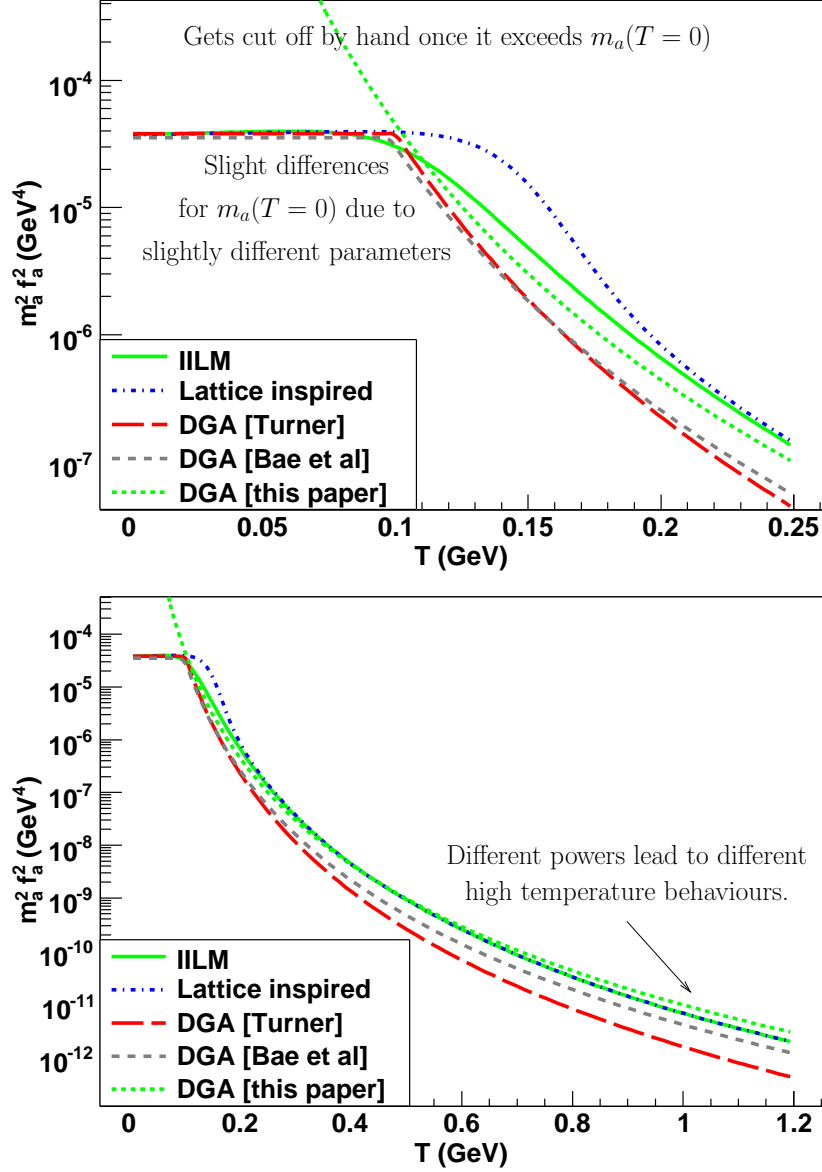


Figure 7.2: Shown are the mass for the QCD axion from IILM simulations (7.10), from a lattice inspired fit that uses the IILM mass shifted towards higher temperatures to mimic the phase transition at $T_c^{lat} \approx 160$ MeV, from the classic dilute gas approximation (DGA) by Turner [186] and its update by Bae et al. [13], and from the DGA derived in this paper (7.12). The simple power-law DGA axion masses are cut off by hand once they exceed $m_a(T=0)$ and give a surprisingly good approximation to the full IILM result. The differences that persist to high temperatures arise from the slightly different quark masses.

a cross-over rather than a sharp phase transition [9]. In particular, it is not of first order, and no latent heat is produced. Furthermore, we assume that no exotic particles beyond those in the Standard Model decouple relativistically. Thus, the universe evolves adiabatically and we leave aside models with additional entropy production; see for instance the discussion in [69].

Since we will also be interested in assessing the accuracy of the standard analytic results for the misalignment mechanism, we will take care to include the correct number of degrees of freedom $g_{*,R}$ and $g_{*,S}$, defined in (6.37) and (6.38) and displayed in Fig. 6.8. In particular, this allows us to derive a specific relation between the temperature and the scale factor, Fig. 6.7. The details are given in chapter 6.

7.2.1 Misalignment mechanism

As usual in standard cosmology, the universe will be described by a flat FRW metric [117], with cosmological parameters given by the concordance of the best available data (we take WMAP5+BAO+SN [118]). For the temperature regions of interest we can restrict ourselves to radiation and axions, in which case Einstein's equations are given by

$$H^2 = \frac{1}{3M_P^2} \left(\frac{\pi^2}{30} g_{*,R} T^4 + f_a^2 \left(\frac{1}{2} \dot{\theta}_a^2 + m_a^2(T) (1 - \cos \theta_a) \right) \right) \quad , \quad (7.13)$$

$$\ddot{\theta} + 3H\dot{\theta}_a + m_a^2(T) \sin \theta_a = 0 \quad , \quad (7.14)$$

where M_P^2 is the reduced Planck mass. Note that the effective axion potential has been shifted so that non-perturbative effects do not lead to a non-vanishing vacuum energy⁵.

The dynamics of the axion evolution consists of three qualitatively different stages: First, as long as its Compton wavelength is below the Hubble scale, the axion is effectively massless; the Hubble friction enforces a constant axion field in this case. Secondly, once the axion mass becomes comparable to the Hubble scale, at a time when $m_a \approx 3H$ holds, the axion feels the pull of its mass and starts to roll towards its minimum at $\theta_a = 0$. Finally, after a few oscillations the axion evolution is indistinguishable from pressureless

⁵Note that there exist theories that combine another axion-like field to entangle the dark matter and the dark energy sector [137, 114].

matter and the axion number per comoving volume is conserved. These three regimes are illustrated clearly for an explicit numerical solution in Fig. 7.3.

The physics underlying the misalignment mechanism is based on the fact that the energy redshifts with time, and that the Hubble dilution starts once the oscillations in the axion zero mode begin. Consequently, the total Hubble redshift increases with m_a . This leads to the a priori counterintuitive behaviour that light axions, i.e. high f_a , contribute more to the energy balance than heavy axions.

Analytical progress can be made by noting that

$$\dot{\rho}_a = 2m_a\dot{m}_a(1 - \cos\theta_a) - 3H\dot{\theta}_a^2, \quad (7.15)$$

where we have made use of the equation of motion for the axion field (7.14). We assume that over one oscillation a , H and m_a do not change much, i.e. adiabatic evolution. Furthermore, we consider times when the axion field has been Hubble redshifted for long enough so that anharmonic effects are negligible, in which case the axion behaves as a damped harmonic oscillator with constant coefficients. For such a system $\overline{\theta_a^2} = m_a^2 \overline{a^2}$, where the expectation value is an average over one oscillation. This leads to

$$\frac{\overline{\rho_a a^3}}{m_a} = \text{const}. \quad (7.16)$$

Eq. (7.16) allows us to easily compute the energy in the axion field today

$$\rho_a(\text{today}) = \rho_a(T) \frac{m_a(\text{today})}{m_a(T)} \frac{s(\text{today})}{s(T)}, \quad (7.17)$$

where T is the temperature when we reached the asymptotic behaviour predicted by (7.16), $s = \frac{2\pi^2}{45} g_{*,S} T^3$ is the entropy density and $g_{*,S}$ is the effective number of degrees of freedom shown in Fig. 6.8; we have neglected the overbar for simplicity. This can then be compared to the present critical energy density, which constrains the axion mass. Note that apart from the time average, a spatial averaging is implicitly understood. In the formulas that follow the coarse-grained θ_a now stands for the effective axion angle

$$\theta_a^{\text{eff}} = \sqrt{\langle \theta_a^2 \rangle} = \sqrt{\langle \theta_a \rangle^2 + \sigma_{\theta_a}^2}, \quad (7.18)$$

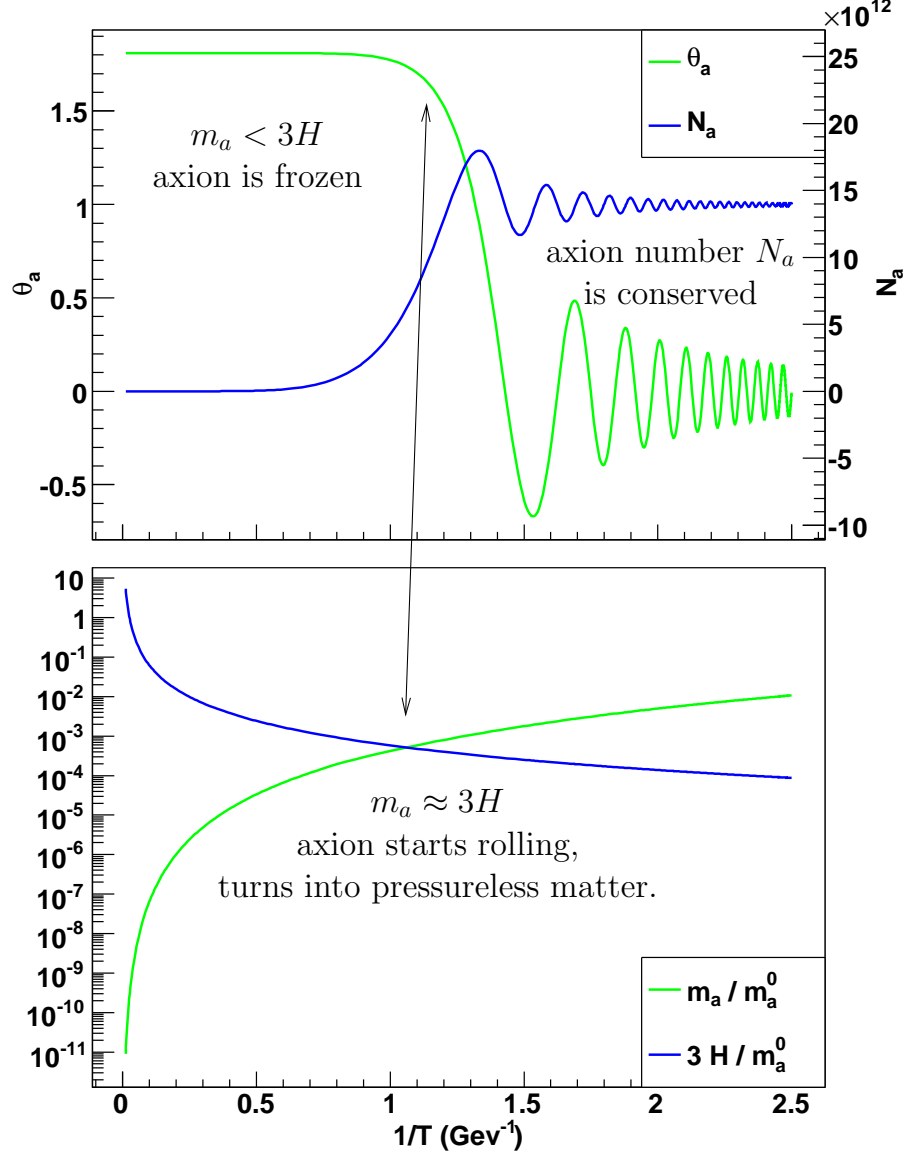


Figure 7.3: As long as the axion Compton wavelength is well outside the horizon, the axion zero mode is frozen; this corresponds to the late-time solution of (7.14) with m_a neglected. The axion starts to feel the pull of its mass at $m_a \approx 3H$, and evolves to its minimum at $\theta_a = 0$, i.e. the PQ mechanism to solve the strong CP problem. After a few oscillations the axion number per comoving volume stays constant as long as the axion mass and the scale factor change slowly (adiabatic approximation). This is then used to extrapolate the result to today.

where σ_{θ_a} represents contributions from quantum fluctuations and classical inhomogeneities. (This notation for θ_a is used except where explicitly stated otherwise or in expectation values.).

As alluded to in the introduction, the axions produced from the heat bath have been ruled out as the dominant cold dark matter contribution [91]. The remaining window for larger values of f_a , bounded from below by astrophysical considerations, implies tiny couplings of axions to ordinary matter so that the axions will stay out of equilibrium with respect to the cosmic plasma; anharmonic effects that couple higher modes can also be ignored by the same reasoning. In addition, axions have very small velocity dispersions⁶; in particular, axions are non-relativistic after recombination, i.e. during structure formation. Therefore, axions provide a generic cold dark matter candidate, whose density is constrained by observation to be $\Omega_c \leq 0.23$ [118].

We will follow the early paper [186] and the recent update [13] to derive an analytic formula for the axion abundance which we then compare to a full numerical solution (using the same updated parameters). We will use the simple power-law approximation (7.12) for the axion mass from the dilute instanton gas; it reaches the $T = 0$ mass at $T \approx 100$ MeV, converges to the full IILM mass at $T \approx 400$ MeV and overestimates it slightly for higher temperatures, as can be seen in Fig. 7.2. Note that the coefficient $\alpha_a \propto m_u m_d m_s$. The axion starts rolling at

$$T_a^{n+4} \approx 0.04 \frac{\Lambda^{n+4} \alpha_a m_{Pl}^2}{g_{*,R} f_a^2}, \quad T \gtrsim 103 \text{ MeV} \quad (7.19)$$

$$T_a^4 \approx 0.04 \frac{\chi m_{Pl}^2}{g_{*,R} f_a^2}, \quad T \lesssim 103 \text{ MeV}. \quad (7.20)$$

Assuming that the number of axions per comoving volume is conserved at T_a , we can use (7.17) to get

$$\rho_a = \frac{45 s_0}{4\pi^2} \left\{ \begin{array}{l} (0.2 m_{Pl})^{-\frac{6+n}{4+n}} g_{*,R}^{-\frac{1+n/2}{4+n}} \chi^{\frac{1}{2}} \Lambda^{-1} \alpha_a^{-\frac{1}{4+n}} f_a^{\frac{6+n}{4+n}} \theta_a^2 \\ (0.2 m_{Pl})^{-\frac{3}{2}} g_{*,R}^{-\frac{1}{4}} \chi^{\frac{1}{4}} f_a^{\frac{3}{2}} \theta_a^2 \end{array} \right. . \quad (7.21)$$

⁶The zero-mode has, by definition, none. Non-zero-mode contributions, due to inhomogeneities in the axion field set up during the PQ phase transition and due to string decay, can be shown to be small (see, for example, Sikivie's review article in [122]).

Note that the axion abundance is fairly insensitive to changes in α_a due to its small exponent. In particular, the large uncertainties in the quark masses have only a minor effect on the final results. We therefore expect this estimate of the energy density ρ_a to differ only slightly relative to the results of [13], say. The additional T -dependence due to $g_{*,R}$ and $g_{*,S}$ will also be weak, again due to rather small exponents. Using $\Lambda = 0.4 \text{ GeV}$, $\chi = (78.2 \text{ MeV})^4$ and $T_\gamma = 2.725 \text{ K}$ we get

$$T \approx \begin{cases} 1.46 \text{ GeV } g_{*,R}^{-0.094} \left(\frac{10^{12} \text{ GeV}}{f_a} \right)^{0.19} \\ 122 \text{ GeV } g_{*,R}^{-0.25} \left(\frac{10^{12} \text{ GeV}}{f_a} \right)^{0.5} \end{cases}, \quad (7.22)$$

and

$$\Omega_a h^2 \approx \begin{cases} 0.54 g_{*,R}^{-0.41} \theta_a^2 \left(\frac{f_a}{10^{12} \text{ GeV}} \right)^{1.19} \\ 0.0064 g_{*,R}^{-0.25} \theta_a^2 \left(\frac{f_a}{10^{12} \text{ GeV}} \right)^{1.5} \end{cases}. \quad (7.23)$$

These analytic results can be improved by taking anharmonic effects into account [13]. The upshot is that we can incorporate these through the substitution $\theta_a^2 \rightarrow \theta_a^2 f(\theta_a)$, where

$$f(\theta_a) = \frac{4\sqrt{2}}{\pi\theta_a^2} \int_0^{\theta_a} d\theta \sqrt{\cos \theta - \cos \theta_a}. \quad (7.24)$$

It follows from the adiabatic invariant $I = \oint p_\theta d\theta$, where p_θ is the canonical conjugate to θ . The Lagrangian is $L = a^3(\frac{1}{2}\dot{\theta}^2 - m_a^2(1 - \cos \theta))$, and it is important to note that to derive (7.24) we keep a and m_a constant in time. We have that $f(\theta_a) \rightarrow 1$ as $\theta_a \rightarrow 0$; it decreases monotonically to $f(\pi) \approx 0.516$. In this regime, $\theta_a \rightarrow \pi$, the adiabatic condition breaks down because the axion behaves like an inflaton, and the scale factor a no longer varies slowly. However, analytic progress can be made in this limit if one takes into account that the above quantity needs to be combined with an estimate for the time when the axion actually starts rolling (and the inflationary phase is over). This occurs no longer at $m_a \approx 3H$ [129, 179, 195]. In the full numerical set-up, we always propagate the solution into the harmonic regime before we use (7.17) to extrapolate to today, and so the correction factor $f(\theta_a)$ is not required.

To take the temperature dependence of $g_*(T)$ into account, we must solve

the implicit equations (7.19) and (7.20). We find that neglect of the temperature dependence of g_* leads to an error of 10% in the analytic computation. On top of that the analytic computation assumes that the adiabatic regime is reached when $m_a = 3H$; this is a further source of error relative to the full numerical result.

Imposing the dark matter constraint $\Omega_c \approx 0.23$, we get a relation between the two parameters f_a and θ_a . The analytic approximation compares well with the full numerical result, except for very large or very small initial misalignment angles, see Fig. 7.4. At large angles the adiabatic condition breaks down. At small angles the DGA approximation to the axion mass is constant whereas the full IILM mass decreases slowly towards $m_a(T=0)$, i.e. the IILM dynamics starts rolling slightly earlier so that the Hubble dilution acts for longer which in turn leads to a smaller Ω_a at a given f_a ; equivalently, to reach Ω_c a slightly larger f_a is needed, as depicted in Fig. 7.4.

If the PQ symmetry breaks spontaneously after inflation, the correlated domains with a given misalignment angle are not stretched to macroscopic sizes and a strong spatial dependence remains. In ref. [186], these fluctuations are averaged in the observable universe to find the root-mean-square fluctuations $\theta_a = \frac{\pi}{\sqrt{3}}$. This was then taken to be the initial condition for an estimate of the zero momentum mode axions. While this approach to axion production in the thermal scenario is flawed (as we shall discuss), it provides a useful benchmark with which to compare competing estimates. Adopting these initial conditions, from the full numerical results we obtain the important bound

$$f_a \leq 2.8(2) 10^{11} \text{ GeV} \quad \text{or} \quad m_a \geq 21(2) \mu\text{eV}. \quad (7.25)$$

This canonical result can be compared directly with our analytic modelling, as well as other estimates in the literature, to check accuracy.

In Fig. 7.5, the errors of the analytic results with the DGA axion masses are compared to the numerical result across the full range of f_a . The numerical solution has f_a and θ_a chosen such that $\Omega_a = 0.23$. Except at large and small f_a , the differences with the analytic models using the same parameters are of order $O(1)$. It is clear in this regime that the relative abundance is not strongly dependent on the exact shape of the axion mass. We also com-

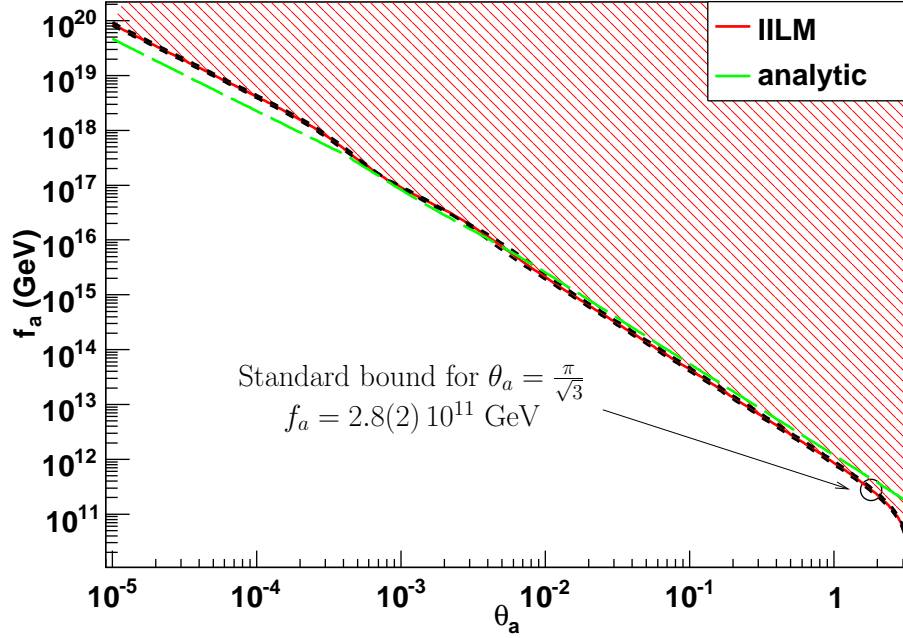


Figure 7.4: The anthropic axion is defined through its relation between f_a and θ_a given a fixed Ω_a . We display here the result for the case that axions form the dominant dark matter component of the universe, i.e. $\Omega_a = 0.23$. The short-dashed lines correspond to the uncertainties in $f_a(\theta_a)$ from the systematic errors in Λ , which is the dominant source of uncertainties. We also include the result of the analytic computation (7.21). Over many orders of magnitude the agreement is very good. More pronounced differences only show up at $\theta_a \rightarrow 0$ and $\theta_a \rightarrow \pi$. The latter is due to the fact that the adiabatic condition is not fulfilled because the potential becomes very flat and acts like a source of inflation, i.e. a rapidly changing scale factor. The differences between the analytic and numerical data at small θ_a follow from the different functional form of the axion masses: in the former case the axion mass is constant whereas in the latter it is slightly rising with temperature (see main text).

pared the full numerical results for the DGA masses and found again that differences are only pronounced in the region of the QCD phase transition where the different mass ansätze differ considerably. We checked against the full numerical solution for the lattice-inspired mass function too, see Fig. 7.2, and found less than 5% variations for $f_a < 10^{15}$ GeV.

While revealing the regimes in which analytic estimates go astray, for the most part the results are in good agreement with this simple DGA treatment. It is encouraging to note that estimates of the axion density are relatively insensitive to the detailed nature of the axion mass function, affirming the validity of the previous literature. However, this does not mean that a simple treatment will automatically give agreement with (7.25); this requires appropriate normalisation, updated parameter choices and a careful treatment of g_* , as attested by the significant differences between quoted bounds.

7.2.2 Anthropic window, isocurvature and fine-tuning bounds

The discussion so far has been purely classical. In order to discuss the anthropic window, where we fine-tune θ_a at large and small f_a , we need to take into account the quantum mechanical fluctuations of the axion field induced by inflation. Like any other massless field, the axion receives quantum mechanical fluctuations during the quasi de Sitter evolution of the early universe, i.e. $\sigma_\phi = \frac{H_I}{2\pi}$ or in terms of the axion angle

$$\sigma_{\theta_a} = \frac{H_I}{2\pi f_a}. \quad (7.26)$$

Apart from the spatial averages, these quantum mechanical effects need also be taken into account in the coarse grained equations; they are dominant for the anthropic scenario where inflation has smoothed out the ‘classical’ inhomogeneities, and makes sense only if the PQ symmetry is broken before the end of inflation. The fluctuations in the massless axion field then lead to isocurvature perturbations in the cosmic microwave background radiation (CMBR). The ensuing constraints apply only if the PQ symmetry has not been restored after inflation; this could happen if the quantum mechanical fluctuations or the reheat temperature are too large, i.e. if

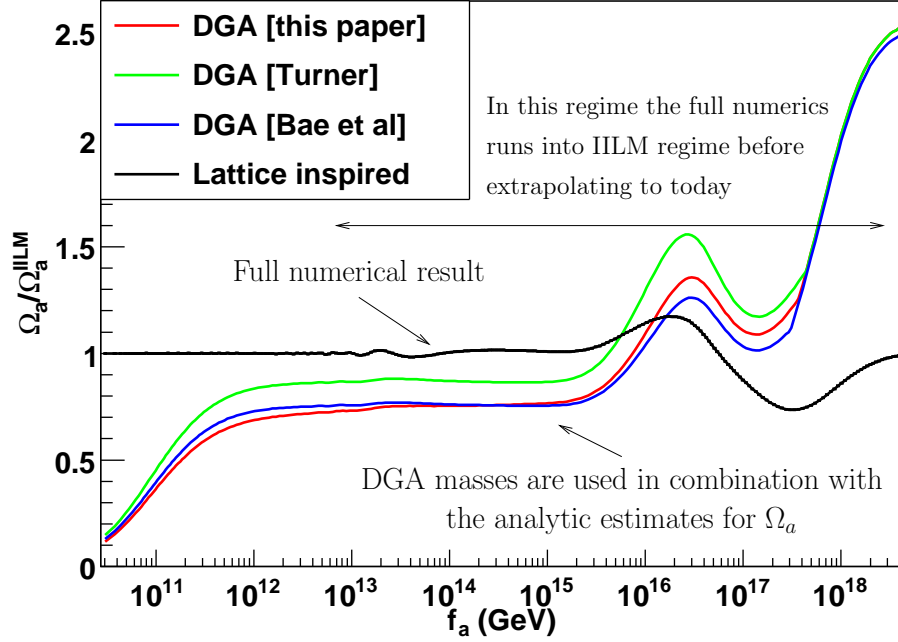


Figure 7.5: To analyse the differences between the different determinations of the axion abundance, we compute Ω_a on a fixed set of $\{f_a, \theta_a\}$ derived from the full numerical determination with the IILM axion mass, given that $\Omega_a = \Omega_c$. We see that the analytic results are off by more than a factor of 2 in the regimes of small and large axion angles; for the case when inflation has not operated after PQ symmetry breaking, i.e. for $f_a < 2.8 \cdot 10^{11}$ GeV, we can see that the numerical and analytic results are different by a factor of 3 (and more if the axion is not the dominant dark matter component). For the main part of parameter space the discrepancy is smaller but systematically an underestimate. The full numerical data using the lattice inspired axion mass are very similar for most f_a (by construction), and differ only for large f_a , as might have been expected. It is noteworthy that the numerical determination runs into the proper IILM axion mass already for rather low f_a 's before it extrapolates to today.

$$f_a < \max \left[\frac{H_I}{2\pi}, T_{\text{RH}} \right].$$

As mentioned previously, the fluctuations lead to an effective axion angle, $\theta_a^{\text{eff}} = \sqrt{\langle \theta_a^2 \rangle} = \sqrt{\langle \theta_a \rangle^2 + \sigma_{\theta_a}^2}$, and will be used in the coarse grained evolution equations. It follows that the effective angle is bounded by the root-mean-square fluctuations

$$\theta_a^2 \geq \sigma_{\theta_a}^2 = \frac{H_I^2}{(2\pi f_a)^2}. \quad (7.27)$$

We will follow refs [30, 24, 95, 90, 195] to put constraints on the PQ scale f_a and on the inflationary scale H_I . By definition, isocurvature perturbations do not change the energy density, i.e. $\delta\rho = 0$. For a massless field such as the axion, fluctuations do not change the potential energy density. In addition we have assumed that the quantum mechanical fluctuations are small, i.e. $f_a > H_I$ so that the PQ symmetry is not restored, and the gradient energy density is negligible compared to the inflationary energy scale, for details see ref. [24]. Thus, inflationary axion fluctuations are indeed isocurvature which leads to

$$\delta\rho = 0 = \delta\rho_a + \sum_{i \neq a} \delta\rho_i + \delta\rho_r. \quad (7.28)$$

Assuming that all other fields have adiabatic perturbations, i.e. $\frac{\delta\rho_i}{\rho_i} = \frac{3}{4} \frac{\delta\rho_r}{\rho_r}$, we get a relation between the temperature fluctuation, i.e. $\frac{\delta\rho_r}{\rho_r} = 4 \frac{\delta T}{T}$, and the axion fluctuation

$$\frac{\delta T}{T} = - \frac{\rho_a}{3 \sum_{i \neq a} \rho_i + 4\rho_r} \frac{\delta n_a}{n_a}. \quad (7.29)$$

It is customary to define the entropy perturbation

$$S_i \equiv \frac{\delta(n_i/s)}{n_i/s} = \frac{\delta n_i}{n_i} - 3 \frac{\delta T}{T}, \quad (7.30)$$

where $s \propto T^3$ is the entropy density and n_i the number density of particle species i . It is clear that for adiabatic perturbations $S_i = 0$ by definition. At the time the relevant perturbations are set, the universe is radiation dominated, and it follows from (7.29) that the radiation perturbations are negligible with respect to axion fluctuations. The initial condition is thus given by $S_a = \frac{\delta n_a}{n_a}$ and once this mode leaves the horizon it remains constant.

The relevant scales cross back inside the horizon during matter domina-

tion, i.e. $\rho_r \ll \rho_a$; it then follows from (7.29) that

$$\left(\frac{\delta T}{T}\right)_{\text{iso}} = -\frac{6}{15} \frac{\Omega_a}{\Omega_m} S_a, \quad (7.31)$$

where we have added the Sachs-Wolfe contribution and Ω_m is the total matter abundance. The fraction of isocurvature to total temperature fluctuations has been constrained to [118]

$$\alpha_a \equiv \frac{\langle (\delta T/T)_{\text{iso}}^2 \rangle}{\langle (\delta T/T)_{\text{tot}}^2 \rangle} < 0.072, \quad (7.32)$$

and $(\delta T/T)_{\text{tot}}^{\text{rms}} \approx 1.1 \cdot 10^{-5}$, where the sum is taken over the first few low-lying multipoles.

For the large f_a we are interested in, the axion dependence on the energy density is well approximated by $\rho_a \propto \theta_a^2$, where in this case θ_a is not the effective axion angle. Assuming a Gaussian distribution for the axion angle perturbation, we find that

$$\langle S_a^2 \rangle = \left\langle \left(\frac{\theta_a^2 - \langle \theta_a^2 \rangle}{\langle \theta_a^2 \rangle} \right)^2 \right\rangle = \frac{2\sigma_{\theta_a}^2 (2\theta_a^2 - \sigma_{\theta_a}^2)}{\theta_a^4}. \quad (7.33)$$

From this it follows that the isocurvature fraction is given by

$$\alpha_a = \frac{4}{25} \frac{\Omega_a^2/\Omega_m^2}{\langle (\delta T/T)_{\text{tot}}^2 \rangle} \frac{2\sigma_{\theta_a}^2 (2\theta_a^2 - \sigma_{\theta_a}^2)}{\theta_a^4}, \quad (7.34)$$

which we can rewrite into a constraint equation for H_I :

$$\left(\frac{H_I}{2\pi}\right)^4 - 2(\theta_a f_a)^2 \left(\frac{H_I}{2\pi}\right)^2 + \frac{\tilde{\alpha}_a}{2\Omega_a^2} (\theta_a f_a)^4 > 0. \quad (7.35)$$

All explicit numerical factors have been absorbed into

$$\tilde{\alpha}_a = 0.072 \frac{25}{4} \langle (\delta T/T)_{\text{tot}}^2 \rangle \Omega_m^2 \approx 4 \cdot 10^{-12} \ll 1. \quad (7.36)$$

Of the two possible solutions, the larger one is in conflict with (7.27), so that

the isocurvature constraint becomes

$$H_I < \frac{\sqrt{\tilde{\alpha}_a}\pi}{\Omega_a}\theta_a f_a < \frac{\sqrt{\tilde{\alpha}_a}\pi}{\Omega_a}\theta_a(f_a)f_a \approx \frac{6.3 \cdot 10^{-6}}{\Omega_a}\theta_a(f_a)f_a, \quad (7.37)$$

where $\theta_a(f_a)$ follows from Fig. 7.4. Note that $\theta_a/\Omega_a \propto \Omega_a^{-\frac{1}{2}}$, so that for fixed f_a the bound becomes weaker if the axions make up only a fraction of the dark matter content of the universe.

Actually, (7.37) is an implicit equation for H_I since the effective axion angle depends on it. However, in this regime the difference between θ_a and $\langle\theta_a\rangle$ is negligible. To see this, we rewrite (7.37) in terms of $\langle\theta_a\rangle$, and get

$$H_I < \frac{\sqrt{\tilde{\alpha}_a}\pi}{\Omega_a}\langle\theta_a\rangle f_a \left(1 - \frac{\tilde{\alpha}_a}{4\Omega_a^2}\right)^{-\frac{1}{2}} \approx \frac{\sqrt{\tilde{\alpha}_a}\pi}{\Omega_a}\langle\theta_a\rangle f_a. \quad (7.38)$$

This, in turn, can be used to estimate that $\sigma_a \ll \langle\theta_a\rangle$. In the anthropic regime the dependence of θ_a on H_I is totally negligible, and we can fine-tune the effective axion angle to $\theta_a \approx \langle\theta_a\rangle(1 - 10^{-10})$.

If we want to extend the analysis to all of the anthropic window, i.e. $\theta_a = O(1)$, we need to take into account anharmonic effects. We are greatly helped by the fact that $\sigma_a \ll \langle\theta_a\rangle$: it allows us to Taylor expand the true axion potential around $\langle\theta_a\rangle$ and we find that

$$\langle S_a^2 \rangle = \frac{\sin^2\langle\theta_a\rangle(\sigma_{\theta_a}^2 - \sigma_{\theta_a}^4) + \frac{1}{2}\cos^2\langle\theta_a\rangle\sigma_{\theta_a}^4}{(1 - \cos\langle\theta_a\rangle)^2 + \frac{1}{4}\cos^2\langle\theta_a\rangle\sigma_{\theta_a}^4 + (1 - \cos\langle\theta_a\rangle)\cos\langle\theta_a\rangle\sigma_{\theta_a}^2(1 - \frac{1}{4}\sigma_{\theta_a}^2)}, \quad (7.39)$$

where we expanded to fourth order, and we assume again that $\delta\theta_a$ is Gaussian with mean zero. For small $\sigma_{\theta_a} \ll \langle\theta_a\rangle \ll 1$ this goes over into (7.33). For large misalignment angles the behaviour becomes rather different, and for $\langle\theta_a\rangle \rightarrow \pi$ we see that the entropy perturbation tends to $\langle S_a^2 \rangle \rightarrow \frac{1}{8}\sigma_{\theta_a}^4$ in contrast to the $\langle S_a^2 \rangle \rightarrow \frac{4}{\pi^2}\sigma_{\theta_a}^2$ for the harmonic case. Since the perturbation has to be small, the solution for σ_{θ_a} in the anharmonic case will be much larger and the constraint on H_I consequently much weaker. The solution satisfies again $\sigma_a \ll \langle\theta_a\rangle$, and the approach is self-consistent.

Even though the fluctuation is very small, the regime $\theta_a \rightarrow \pi$ can only be achieved if H_I is further constrained, as we will now show. Fine-tuning in the effective axion angle is hindered at large H_I by quantum fluctuations.

Assuming that $\langle\theta_a\rangle = \pi$, we obtain the following bound on H_I

$$H_I < \sqrt{8\pi\xi}\pi f_a, \quad (7.40)$$

where ξ is defined by $\xi \equiv \pi - \theta_a$ and encodes the degree of fine-tuning. It is clear then that for axions on the top of the potential, this bound supersedes the isocurvature bound. This leads to a further restriction on the allowed parameter space shown in Fig. 7.6 (bottom left). This cut-off provides an interesting boundary in parameter space suggesting that inflationary axions have a maximum possible mass of approximately $m_a \lesssim 1 \text{ meV}$. Inflationary scenarios usually consider small axion masses in the range $m_a \lesssim 10 \mu\text{eV}$, but we see here with anharmonic tuning that it is possible to have $\Omega_a = 0.23$ with masses at and above that of the thermal dark matter axion (see below). Naturalness arguments tend to disfavour inflation models with a low energy scale, but we note that the weakening of the isocurvature constraint at $\theta_a \approx \pi$ opens up the inflationary window by several orders of magnitude relative to naive expectations. It is possible to have these heavier axions with an inflation scale approaching $H \approx 10^9 \text{ GeV}$. The fact that two alternative scenarios – thermal and inflationary – can produce axions with masses $m_a = 0.1\text{--}1 \text{ meV}$ might strengthen the case for experimental searches in this parameter regime.

Note that although the parameter space is really three-dimensional, i.e. in $\langle\theta_a\rangle$, f_a and H_I , we only plot the H_I – f_a plane as these are the fundamental parameters, whereas $\langle\theta_a\rangle$ is naturally seen as an ‘environmental’ feature of our location in the universe after inflation. Recall also that without inflation the causal patches with different $\langle\theta_a\rangle$ stay microscopic all the way to today with $\theta_a \rightarrow \frac{\pi}{\sqrt{3}}$, so again the axion angle is not required as an extra parameter. The anthropic window bounds can be weakened by choosing $\theta_a < \theta_a(f_a)$, however, if the axion is to be the dominant dark matter candidate then the parameter space is truly 2-dimensional and collapses to the H_I – f_a plane.

We have mentioned previously, that the natural axion angle to be used in the evolution equations is the effective axion angle that takes into account root-mean-square fluctuations. A priori this entangles H_I and $\langle\theta_a\rangle$. However, we have seen that for a large part of the anthropic window θ_a does not really depend on H_I . On the upper green curve in Fig. 7.6, that is if the axion is to

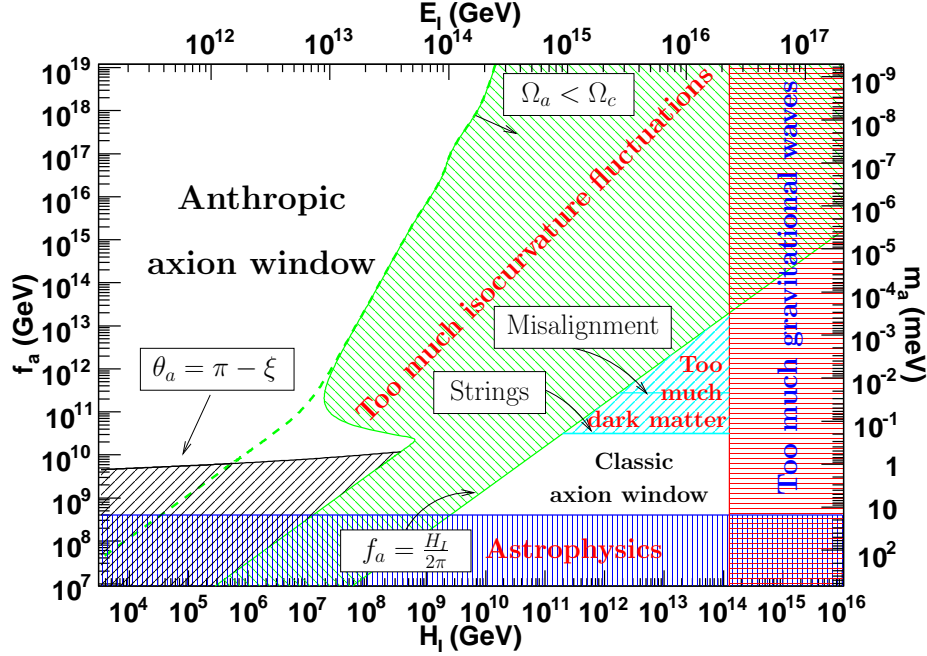


Figure 7.6: The allowed parameter space in the H_I – f_a plane is plotted in white; the inflationary energy scale is defined by $E_I \equiv (H_I m_{Pl} / \sqrt{8\pi/3})^{1/2}$. The green curve in the upper left corner follows from the isocurvature constraint (7.37), when the axion is the dominant dark matter candidate. If axions provide only a fraction of the dark matter content of the universe, the bound becomes weaker. For lower f_a , i.e. larger θ_a , the anharmonic effects become important and the bound on H_I weakens because anharmonic effects lead to smaller perturbations. For $\theta_a \rightarrow \pi$, the dependence of H_I on θ_a can no longer be neglected and leads to the black curve. The lower green curve gives the lower bound for isocurvature production (very inefficient reheating is assumed [95]); beneath this curve, the axion angle is spatially varying (with root-mean-square fluctuation $\theta_a = \pi/\sqrt{3}$). The cyan wedge is excluded as it would lead to too much dark matter from axion string radiation. The bound from the non-detection of gravitational waves, i.e. $r < 0.22$ [118], leads to the upper bound on the inflationary scale $H_I < 1.26 \cdot 10^{14}$ GeV. Finally, f_a is bounded from below by astrophysical considerations, i.e. axion emission from stars; we use $f_a > 4 \cdot 10^8$ GeV, see [122] chapter 3.

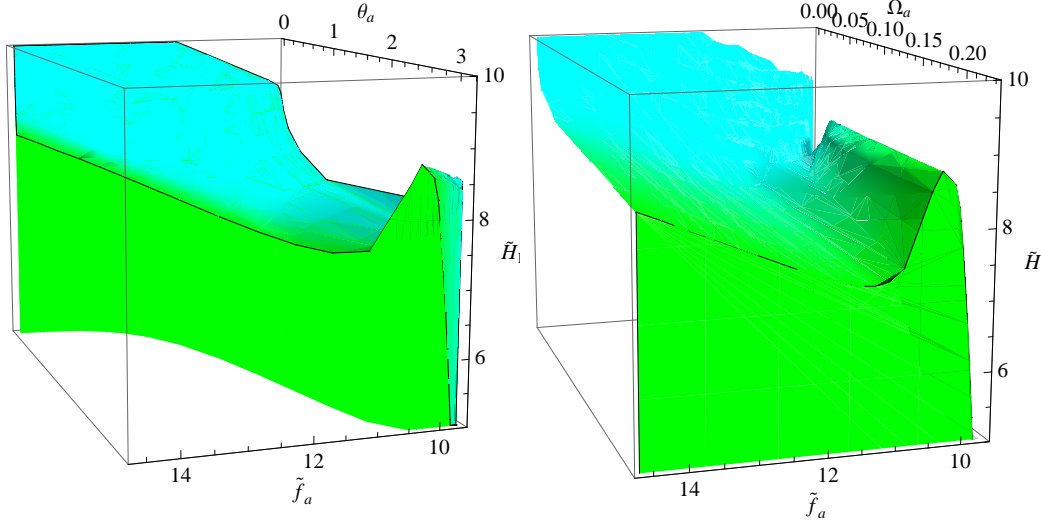


Figure 7.7: The allowed parameter space in the anthropic window. The dark matter axion lives on the green face. Although not as clearly visible as for the right plot, the projection of the green face from the left plot into the H_I – f_a plane corresponds to the anthropic window displayed in Fig. 7.6. The colour shading is calibrated to the axion density; this is clear for the right plot but is also true for the left plot. We see that the parameter space does not depend sensitively on Ω_a in the range $\Omega_a \in [\Omega_c, 0.1 \Omega_c]$; for smaller densities H_I starts to grow rapidly and eventually we reach the regime $\langle \theta_a \rangle \ll \sigma_{\theta_a} \ll 1$ where the isocurvature bound can no longer be fulfilled. This gives another bound in the anthropic window, although a rather uninteresting one since the axion density has become totally negligible at that point.

be the dominant dark matter candidate, then f_a and $\langle \theta_a \rangle$ are in the one-to-one correspondence defined through Fig. 7.4. Again, on the black curve in Fig. 7.6, $\langle \theta_a \rangle$ is very close to π and θ_a depends solely on H_I for axions that contribute $\Omega_a = \Omega_c$.

If we allow the axion to contribute only a fraction of the dark matter content to the universe, the parameter space in the anthropic region becomes truly 3-dimensional. For $\theta_a < \theta_a(f_a)$ the axion density drops, and smaller effective angles are in one-to-correspondence with $\Omega_a < \Omega_c$ for fixed f_a . Note that the relation $\Omega_a = \Omega_a(f_a, H_I, \langle \theta_a \rangle)$ can be inverted and allows us to trade $\langle \theta_a \rangle$ for Ω_a . In Fig. 7.7 we plot the available parameter space in the anthropic window.

It was somewhat surprising that $f_a(\theta_a)$ varies very slowly as $\pi - \theta_a \rightarrow 0$; indeed in that regime the axion field behaves like an inflaton, and one might have expected that $f_a \rightarrow 0$ rather fast. This is in agreement with the earlier numerical work by [186] and analytic computations [129, 179, 195]. It follows from this that the axion potential is not flat enough to support a prolonged inflationary period. In this regime, the fine-tuning in θ_a to many digits is reminiscent of the fine-tuning for the couplings of the inflationary potential to produce sufficient e-folds.

7.2.3 Axion strings

For completeness, let us take a closer look at the “classic axion window”, that is, the thermal scenario in which the universe reheats to temperatures above the PQ symmetry breaking scale, $T > f_a$. There are a number of misconceptions in the literature about the importance of misalignment production of axions from this thermal state. At high temperature, the thermal axion distribution θ_a will fluctuate randomly in time and space around the circle $0 \rightarrow 2\pi$. An rms average with $\langle \theta \rangle = \pi/\sqrt{3}$ away from the minimum at $\theta_a = 0$ is assumed to be the thermal initial condition for a misalignment zero momentum mode (7.25) which begins to oscillate when the axion mass switches on. Unfortunately, this generically underestimates axion production in the thermal scenario because it does not quantitatively account for the full spectrum of axion momentum states which are excited by the time of the mass ‘switch on’.

Axion production in the thermal scenario is in fact dominated by radiation from axion strings which inevitably form via the Kibble mechanism [111] when the $U_{\text{PQ}}(1)$ symmetry breaks. These strings radiate axions during their scale invariant evolution on all subhorizon scales and then annihilate and disappear at axion mass switch on (as we will explain in more detail below). The important point is that axion strings at late times can effectively radiate low-frequency axions in the range $H < \omega < m_a$ (for $f_a \ll 10^{16}$ GeV). Given the same total energy radiated into axions, a spectrum concentrated at low momenta will lead to a higher number of axions; thus the more low-energy axions are produced, the tighter the constraints from string radiation will become. This issue still remains controversial [176]. When the mass switches

on, it is fairly straightforward to show that the resulting density of these non-relativistic axions inevitably exceeds those naively estimated from the zero momentum mode. This discussion follows the review given in ref. [19, 21], updating the string constraint using new data.

Axion strings are global strings which possess strong long-range forces, with a (renormalised) energy per unit length given by

$$\mu \approx 2\pi f_a^2 \ln(L/\delta), \quad (7.41)$$

where $\delta \sim f_a^{-1}$ is the string core width and the typical radius of curvature is $L \sim t$. At late times near axion mass switch on, the logarithm $\ln(L/\delta) \approx 70$ for $f_a \sim 10^{11}\text{GeV}$. Here, we essentially have a renormalisation of the bare string energy density $\mu_0 = 2\pi f_a^2$ by radially integrating out the effect of the axion field that winds by 2π around the string. Despite the nomenclature, with such a large $\mu \gg \mu_0$, the axion string is in fact highly localised with more than 95% of its energy in only 0.1% of the volume through which it traverses; it behaves to lowest order like a (local) Nambu string.

The evolution of a network of cosmic strings is non-pathological (whether local or global), because their evolution is scale-invariant with their curvature radius growing as $L = \gamma t$ and the string density scaling as $\rho = \zeta \mu/t^2$, that is, in proportion to the background cosmological density. In the radiation era, we have $\zeta \equiv \gamma^{-2} \approx 13$, as determined by large-scale numerical simulations [25, 5]. The string network consists of both a population of long strings, that carry most of the energy, and small (subhorizon) loops which are created through long string reconnections,

$$\rho_{\text{str}} = \rho_\infty + \rho_\ell \equiv \frac{\mu\zeta}{t^2} + \mu \int \ell n(\ell, t) d\ell, \quad (7.42)$$

where $n(\ell, t)d\ell$ is the number density of loops in the range ℓ to $\ell + d\ell$ at the time t . By conservation of energy for the string network, one can calculate the loop number density (see the discussion in e.g. [193]):

$$n(\ell, t)d\ell = \frac{\zeta g(1 - \langle v^2 \rangle) \alpha^{1/2}}{\ell^{5/2} t^{3/2}}, \quad (7.43)$$

where α describes the typical loop size at creation $\ell = \alpha t$, the rms string ve-

locity is $\langle v^2 \rangle^{1/2} \approx 0.65$, and the relativistic boost factor $g \approx 1/\sqrt{2}$ accounts for centre-of-mass kinetic energies at loop creation (with $\nu_0 \equiv g\langle v^2 \rangle \approx 0.28$ reliably measured in simulations). Significant uncertainty remains concerning the loop creation size α given the complex nature of small scale structure on strings and the limited dynamic range available in simulations. The most recent and largest scale simulations indicate that the largest loops are produced on length scales $\alpha \approx 0.1$ and below, but with a fairly flat loop production function which extends down to smaller scales [141].

Loops oscillate periodically and decay fairly rapidly into axions. Axion radiation is primarily emitted in low frequencies, that is, in the lowest harmonics of the fundamental frequency of the loop oscillation $\omega_0 = 4\pi/L$ (the loop period is $T = L/2$). We note that axion radiation by strings can be treated very similarly to radiation into gravitational waves [193]. Decomposing the loop radiation power P per oscillation into a spectrum P_n for each harmonic $n\omega_0$ we expect

$$P = \sum_n P_n, \quad P_n \propto n^{-q} \quad (n \gg 1), \quad \text{with } q \geq 4/3. \quad (7.44)$$

Here, the spectral index $q = 4/3$ can be calculated analytically for loops with cusps, but realistically this is expected to be greater than $4/3$ because of radiative backreaction, with an effective maximum n_* . Given the dominance of the low harmonics in these spectra, the following results are relatively insensitive to the details at large n . The typical integrated power per oscillation is described by Γ_a , defined in the continuum limit with a radiation spectrum $g(\ell\omega)$ as

$$P = \Gamma_a f_a^2 = f_a^2 \int g(x) dx \quad \text{with} \quad \frac{dP_\ell(\omega)}{d\omega} = f_a^2 \ell g(\ell\omega). \quad (7.45)$$

The typical loop radiation rate Γ_a can be estimated analytically for specific loop trajectories (see, for example, ref. [3]), but it is measured numerically from simulation loop trajectories to be $\Gamma_a \approx 65$ [4].

Due to these radiative losses into axions, a loop will shrink linearly from its original size at creation $\ell_i = \alpha t_i$ as

$$\ell = \ell_i - \kappa(t - t_i) \quad \text{with} \quad \kappa = \Gamma_a/\mu \approx \Gamma_a/2\pi f_a^2 \ln(\ell/\delta). \quad (7.46)$$

Given $\ln(\ell/\delta) \approx 70$ for the energy scales and times of interest, this means that the loop backreaction rate $\kappa \approx 0.15$ and a loop will oscillate about 10-15 times before vanishing. This loop decay will modify and cut-off the loop distribution given above (7.43), becoming

$$n(\ell, t)d\ell = \frac{\zeta\nu_0\alpha^{1/2}(1 + \kappa/\alpha)^{3/2}}{(\ell + \kappa t)^{5/2}t^{3/2}}. \quad (7.47)$$

Given the spectral assumptions (7.45), we can integrate over the loop distribution 7.47 to obtain a spectrum for the number density of axions n_a (for $\alpha \lesssim \kappa$) [21]

$$\frac{dn_a}{d\omega} = \frac{1}{\omega} \frac{d\rho_a}{d\omega} = \frac{4\Gamma_a\zeta\nu_0\alpha^{1/2}(1 + \kappa/\alpha)^{3/2}}{3\omega^2\kappa^{3/2}t^2} \left[1 - \left(1 + \frac{\alpha}{\kappa} \right)^{-3/2} \right]. \quad (7.48)$$

Integrating down to the lowest frequencies $\omega = 4\pi/\alpha t$ emitted at a time t , we obtain the total axion number density [21]

$$n_a = \frac{\Gamma_a\zeta\nu_0}{3\pi t} \left[\left(1 + \frac{\alpha}{\kappa} \right)^{3/2} - 1 \right], \quad (7.49)$$

where the prefactor has a numerical value $\Gamma\zeta\nu_0/3\pi \approx 31$ with moderate uncertainties ($\pm 30\%$). The most important uncertainty in the expression (7.49) is clearly the loop size parameter α measured relative to the backreaction rate κ . In Fig. 7.8, we see the strong dependence of the axion string bound on the ratio α/κ having imposed the dark matter constraint $\Omega_a \leq 0.23$. Recent string simulations suggest $\alpha \lesssim 0.1$ implying $\alpha/\kappa \lesssim 0.7$, but what is the lower limit or, rather, the appropriate range for α ?

The lower cutoff for the loop production scale α depends on the nature of small-scale structure on long strings, since it is unlikely that loops can be produced on scales below which the string network becomes smooth. Numerical and analytic studies of long string backreaction indicate that it is somewhat weaker than loop backreaction because of geometric effects and the typical amplitude and velocity of long string oscillations. The loop radiation rate κ is then replaced by the long string rate [20]

$$\kappa_\infty \sim \frac{\pi^2}{8} \ln(t/\delta) \sim 0.02, \quad (7.50)$$

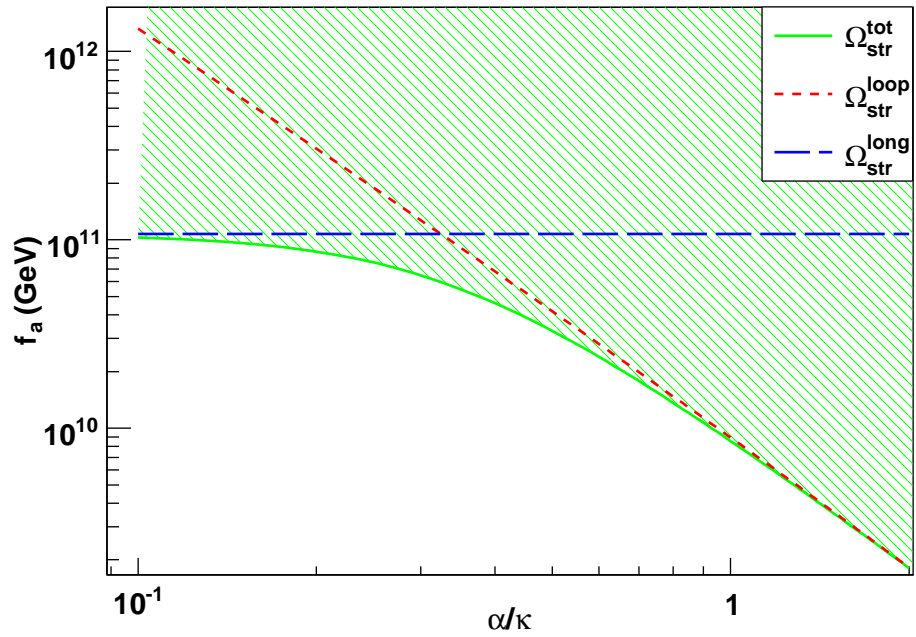


Figure 7.8: Effect of the loop creation ratio $r \equiv \alpha/\kappa$ on the dark matter axion constraint. If $\alpha(t)$ the loop creation size at time t is larger than $\kappa(t)$, the loop radiation backreaction scale, then the constraint is stronger and conversely for $\alpha < \kappa$. Note that in the second case with $\alpha \ll \kappa$, the dominant contribution arises from direct long string radiation, which again exceeds misalignment production.

(refer also to related studies of gravitational waves from long strings in ref. [140]). The expression (7.50) indicates a lower limit on the loop production size $\ell \gtrsim \kappa_\infty t$, thus with the loop size ratio lying in a fairly narrow range $0.1 \lesssim \alpha/\kappa \lesssim 0.7$. Nevertheless, it is the larger loops with $\alpha/\kappa \sim 0.5$ which are expected to dominate the axion contribution, given the nature of the spectral weighting underlying (7.49).

There is also a significant contribution coming from direct radiation from long strings which, given (7.50), has a spectral radiation rate [21, 20]

$$\frac{dn_a^\infty}{d\omega} = \frac{1}{\omega} \frac{d\rho_a}{d\omega} \approx \frac{\pi^3 f_a^2 \zeta}{8\kappa_\infty \omega^2 t^2}. \quad (7.51)$$

This integrates to produce the additional axion number density

$$n_a^\infty = \frac{\pi^2 f_a^2 \zeta}{32t}, \quad (7.52)$$

where we have assumed that the radiation spectrum is peaked at the back-reaction scale $\omega \approx 4\pi/\kappa_\infty t$ (second harmonic). This assumption is certainly a conservative underestimate because there will be significant radiation from the strings on the correlation length and above, up to the horizon scale $\omega \gtrsim H$, i.e. effectively contributing more non-relativistic axions at mass ‘switch on’.

Taken together, the sum of the axions produced by loops and long strings is shown in Fig. 7.8, indicating that this is a significant stronger constraint than the misalignment estimate even if loops were very small $\alpha/\kappa \ll 0$. Given that present numerical simulations indicate that a significant proportion of loops are produced on the largest scales in the available range $0.1 \lesssim \alpha/\kappa \lesssim 0.7$, we take the effective value $\alpha/\kappa = 0.5 \pm 0.2$ to obtain an updated axion string constraint

$$f_a \lesssim 3.2_{-2}^{+4} 10^{10} \text{ GeV}, \quad m_a \gtrsim 0.20_{-0.1}^{+0.2} \text{ meV}, \quad (7.53)$$

that is, we predict that if thermal axions are the predominant dark matter then they will have a mass near $m_a \approx 200 \mu\text{eV}$ (revising slightly upward previous estimates from strings [21]). We emphasise that the axion string constraint (7.53) is a conservative bound which is an order of magnitude

stronger than the misalignment estimate (7.25).

Finally, we comment on additional contributions, uncertainties and alternative views of the thermal scenario. At the QCD phase transition, the tilting axion potential causes the axion field θ to relax toward the minimum $\theta = 0$, assisted on long wavelengths ($\omega \sim H$) by Hubble damping – as for the zero momentum mode in inflationary scenarios. However, the presence of a string implies a topological obstruction to this relaxation, so the field instead localises its variations $0 \rightarrow 2\pi$ into domain walls connecting strings to others of the opposite orientation [174, 192]. This correlation of strings within the network, as well as the intercommuting properties of strings colliding with walls, leads to the rapid demise of the hybrid network as demonstrated numerically [159, 160]. Self-intersections of loops bounding domain walls are inevitable until the system breaks up into tiny loops $\ell \lesssim m_a^{-1}$ and wall tension effects become negligible. These loops can decay by radiating (massive) axions and gravitational waves. Estimates of the axion number density n_a^{dw} produced by this process are crude but suggest a contribution comparable only to that from misalignment [33]. Axion production through hybrid network formation deserves more detailed analysis particularly for nonrelativistic modes with $H \lesssim \omega \lesssim m_a$.

The key uncertainty in the string constraint above remains the typical loop production size α for a global string network. This is difficult to probe numerically for two reasons. First, Nambu string simulations are progressing in resolution but they do not at present include the realistic radiative backreaction necessary for describing global strings. The understanding of α is improving but how it cuts off on small scales must be estimated analytically. Secondly, alternative field theory simulations do not have the dynamic range needed to address the loop production issue; with $\ln(t/\delta) \sim 3 \ll 70$, simulated global strings remain in a strongly damped regime unlike cosmological axion strings. It has been suggested that the spectrum of loop and long string radiation could be ‘flat’ (see, for example, [176] and references therein), rather than dominated by the lowest harmonics as in the standard picture presented above (7.45). The proposed ‘flat’ spectrum entails producing equal radiation power over many orders of magnitude $t \lesssim \omega \lesssim f_a$ (up to 30), effectively suppressing the importance of the long wavelength modes

on which the bound (7.53) is based. The ‘flat’ spectrum yields an axion string constraint roughly matching the misalignment bound (7.25) [85], but a detailed critique of this scenario is offered elsewhere [21].

7.3 Conclusion

We have presented a temperature-dependent axion mass, based on instanton methods, that is valid for all temperatures. The transition between the high- and low-temperature regime is well-motivated and computed within the same model, the IILM, in contrast to some ad hoc procedure to connect the two.

Even though the IILM does not explain confinement, chiral symmetry restoration is incorporated and the model can be expected to give qualitatively correct results for the axion mass; this relies on the fact that it is related to the QCD topological susceptibility which in turn is a chiral quantity. We note that chiral symmetry restoration is indeed seen in the IILM, although at a slightly lower temperature. Given the discovery of the more general KvBLL calorons, that may play an important role in the confinement/deconfinement transition, we expect to improve our understanding of the axion potential in the future by incorporating these new degrees of freedom into the IILM.

Using this new axion mass, we solved numerically the axion evolution equations in the concordance FRW cosmology. It turns out that the analytic approximations used previously differ by a factor of 2 – 3. This is unexpectedly good agreement, considering the crude determination of the axion mass within the dilute gas approximation. We believe it to be the result of a coincidence, that the extrapolation of the high temperature DGA axion mass fairly closely follows the full IILM result around the phase transition. Conversely, this correspondence can be interpreted as evidence that the axion mass determination is fairly robust. This is also seen from the rather small differences between the numerical results between the IILM mass and the lattice inspired mass.

We want to draw to attention that the IILM is a model of QCD, and as such it needs to be checked against lattice data, say. In that light, there remains the possibility that the true high temperature axion mass is different from the IILM prediction. This could lead to qualitatively different results:

On the one hand, a softer decay will lead to weaker constraints. A more abrupt mass switch on, on the other hand, would tighten the constraints and potentially close the classic axion window. The ideas of molecule formation within the IILM, and the subsequent stronger suppression of the axion mass in the plasma phase, have initially prompted this investigation. Within the current IILM this is not realised, but it is not ruled out either. The lattice community is performing realistic QCD simulations directly at the physical quark masses, and we can expect a reliable axion mass determination to follow from that data in future.

To get an accurate estimate across the whole f_a axis, we included the correct temperature dependence between the scale factor and the plasma temperature which follows from the conservation of entropy. To that end we computed the full phase-space integral to get the temperature dependence for the effective degrees of freedom $g_{*,S}$ and $g_{*,R}$, following closely [43]. Additionally, we provide accurate fitting formulas, which to our knowledge have not been presented in the literature previously.

In the classic axion window, where the PQ symmetry breaks only after inflation, a quantitative analysis of the misalignment yields the new updated $f_a < 2.8(2) 10^{11} \text{ GeV}$, $m_a > 21(2) \mu\text{eV}$. While the misalignment axion is a useful benchmark for comparing alternative estimates, it is far exceeded in density by axions from string radiation which yields the more stringent string bound $f_a \lesssim 3.2^{+4}_{-2} 10^{10} \text{ GeV}$. Experimental searches for a thermal dark matter axion should focus on masses around $m_a \sim 200 \mu\text{eV}$.

The anthropic axion window, which entangles θ_a and f_a , is constrained by investigating the production of isocurvature perturbations. The rather small contribution of these to the CMBR power-spectrum places strong bounds on f_a and the inflationary scale H_I . This bound is strong enough so that quantum fluctuations are totally negligible, i.e. $\sigma_a \ll \langle \theta_a \rangle$, for almost all axion angles. The anthropic window allows a dark matter axion that can have any initial misalignment angle, provided the inflationary scale is low enough. In light of naturalness in fundamental theory, most interest has focused on the constraints for large f_a .

Nevertheless, taking the whole of the anthropic range seriously reveals some interesting possibilities. In particular, we investigated the region with

large misalignment angles, especially $\theta_a \rightarrow \pi$. In this regime the axion potential can no longer be described by a parabola, and we need to take the anharmonic effects into account. A self-consistent solution follows for the regime with small fluctuations, in which case we can Taylor-expand the axion potential around $\langle\theta_a\rangle$; the anharmonic effects are then encoded in the anharmonic dependence on $\langle\theta_a\rangle$. These effects lead to smaller isocurvature perturbations as compared to the harmonic case and, consequently, lead to weaker constraints on the inflationary scale.

Fine-tuning the initial misalignment angle to π , we find that the quantum fluctuations must be even more stringently constrained. If the axion provides the dominant component to the dark matter content of the universe and if we take at face value an inflationary period after the spontaneous breaking of the PQ symmetry, then this places the strongest constraint on f_a and H_I at large initial misalignment angle. We note the intriguing possibility of a dominant dark matter axion with a mass $m_a \sim 200 \mu\text{eV}$ which is consistent with either the thermal or the inflationary scenarios. This model independence provides extra motivation for experimental searches around this mass range. Anthropic tuning near $\theta \approx \pi$ allows inflationary dark matter axions ($\Omega_a = 0.23$) to have masses as high as $m_a \leq 1 \text{ meV}$, but no higher because quantum fluctuations restrict the fine tuning (for $H_I \geq 10^4 \text{ GeV}$).

Finally, the isocurvature and quantum fluctuation constraints become weaker if the axion is not the dominant dark matter candidate, and we displayed the ensuing 3-dimensional parameter space.

Chapter 8

Conclusions and Outlook

The Peccei-Quinn mechanism still provides the best motivated solution to the strong CP model, and inevitably introduces axions into the particle zoo. From the theoretical side axions are appealing because they can also provide for cold dark matter in a very natural way, even though they were not designed for it. Among the different cold dark matter candidates, axions are therefore rather well motivated. In recent years axions have become much more popular again, due to their generic appearance in models for physics beyond the Standard Model [181, 18]. These axions have typically very large decay constants and so may well provide a precise implementation for an anthropic axion within a larger theoretical framework. As such the misalignment mechanism to axion production is most important. This channel is, however, also the most sensitive to non-perturbative effects from the strong interactions, encoded in the axion mass and its temperature dependence. This work was largely motivated by our perceived need to improve the determination of and investigate the uncertainties relating to the temperature-dependent axion mass.

We have decided to tackle this question by using instanton based models, namely the interacting instanton liquid model. In this model, interactions between the pseudo-particles are taken into account, and computations can be performed at all temperatures, most notably also in the vacuum. This goes far beyond the usual treatment of the axion mass in the dilute gas approximation, which can only be trusted beyond the deconfinement phase transition. Anthropic axions, however, are sensitive to the strong interactions

at lower temperatures.

The IILM was pioneered by Shuryak and, Diakonov and Petrov. Only two-body effects are taken into account, but the strongly localised profiles of instantons and the, a posteriori, fact that the instanton ensemble is rather dilute make this a viable working premise. The big advantage of the continuum formulation is the ease with which quarks can be incorporated, a definite advantage over lattice based approaches. The latter, on the other hand, can handle the gluonic interactions in a much more straightforward way, especially when incorporating many-body effects. Since our main interest will be in full QCD with light quarks, i.e. in chiral properties, the correct formulation of the fermionic part of the model will be paramount and so the continuum formulation is appropriate.

Compared to a formulation on the lattice, it is not straightforward to derive the gluonic interactions. Only in special cases with high symmetry can good approximations in closed-form be given. These are generally obtained through asymptotic considerations and fits to numerical evaluations of the classical action. We noted that the available formulas do not possess a thermodynamic limit at finite temperature. In chapter 3 we have set up a framework which we believe is numerically well-defined, can be extended to more complicated backgrounds and does not suffer from the parametrisation bias introduced implicitly through analytical formulas motivated by symmetry arguments and fits. The price to pay is a larger numerical overhead, which in our framework comes about through look-up tables and asymptotic matching formulas.

We have found that the analytic formulas of [157] agree very well with our interactions at zero temperature, and this regime can be seen as a validation of our numerical procedure for determining the interactions. In particular, there are no qualitative differences for the zero temperature interactions. On a quantitative level, however, our interactions differ. We exploited this fact, and considered the formulas of [157] to be another valid scheme. We studied the dependence of bulk properties of the IILM on these different parametrisation, and found that this introduces a systematic effect which depends on the quantity under consideration, e.g. the instanton density is most sensitive and the mean size the least, but was generally rather large,

up to 20%.

We could corroborate earlier results, e.g. [46], that IILM is compatible with the chiral properties of QCD. By simulating directly in the grand canonical ensemble, we were able to study the topological susceptibility in much more detail than has been done before.

In light of our ultimate goal with regard to the axion mass, a major incentive for this work was to investigate the regime of physical quark masses. As is well-known, it is computationally expensive to simulate light degrees of freedom. In essence, this boils down to using rather large boxes. To simulate the latter in a reasonable amount of time, we reduced the complexity of the algorithm from $O(N^3)$ to $O(N^2)$ by rewriting the updates in a form suitable for fast matrix modifications; and by studying the thermodynamic limit to guarantee a consistent large volume extrapolation.

The topological susceptibility represents a natural candidate to fix units in the IILM. For quenched simulations we found rather good agreement between the IILM and the lattice. In the unquenched case we have used the topological susceptibility to infer the chiral condensate, which has a very weak dependence on the chiral limit, and is therefore better suited for setting units. Our results on the strong coupling constant and the quark masses are in good agreement with experimental bounds and lattice simulations.

Extending our analysis to finite temperature, we found that the Monte Carlo updates were utterly inappropriate: the simulations did no longer produce representative samples, generically equilibrating into a dilute state with a few 'unbreakable' instanton–anti-instanton pairs. The latter would be present in random numbers, producing widely fluctuating results. This prompted a thorough investigation of the Monte Carlo routines against the initial conditions, and we have traced the problem to the strong and local interactions induced by light quarks. In particular, we have argued that this phenomenon is quite generic in localised backgrounds with light degrees of freedom.

Systems with strong and short-ranged interactions are known as strongly associating fluids, and their simulation presents computational challenges. Drawing on the expertise from the fields of computational chemistry and chemical engineering, we developed new algorithms based on Biased Monte

Carlo. The new routines are largely inspired by the Unbonding–Bonding algorithm, and we generalised the latter to the grand canonical ensemble and included pair updates.

To test the algorithm, we have simulated a simple toy-model, and found that random sampling becomes very inefficient if the temperature is raised or the mass is lowered. In particular, we found that simple Monte Carlo algorithms could effectively violate ergodicity. Tuning the free parameters of the model, we found regions in the parameter space which displayed a high temperature behaviour different from that of a dilute gas.

Using the numerical framework set up in chapter 3, we have been able to improve on the finite temperature interactions, which we have shown to lead to a consistent thermodynamic limit. We have generalised the Monte Carlo routines set up in chapter 4 to deal with the specific characteristics of calorons, i.e. the orientation dependence originating from the non-trivial entanglement of the colour and space-time structure.

We have found that the IILM at high temperature differs in a qualitative way from the lattice results in the pure gauge sector. In particular, the topological susceptibility decays too quickly with temperature. This gives evidence that the IILM based on HS calorons turns into a dilute gas too quickly.

In the unquenched sector, we investigated the chiral susceptibility and the quark condensate and found that the IILM unambiguously leads to chiral symmetry restoration at $T_c = 120$ MeV, which is slightly too low as compared to lattice data. The good qualitative agreement can possibly be accounted for by chiral symmetry which is correctly implemented in the IILM. We did not have lattice data available to compare the high temperature regime; such data can be expected in the near future, at which point we can check whether the IILM is in qualitative agreement with the lattice.

We investigated in detail the relevance of instanton–anti-instantons. In particular, a large population of instanton pairs is present even at fairly high temperatures, indicating that the dilute gas limit of the instanton ensemble is only reached far beyond the critical temperature. However, the quark condensate and the topological susceptibility behave according to the dilute gas result much earlier. We could attribute this behaviour to a sub-ensemble

of non-interacting instantons, distinct from the unbonded instantons that still interact considerably with the instanton–anti-instanton molecule population.

We have shown that the axion couples to the topological charge and that its mass is given by the topological susceptibility. Using our results on the latter, we have presented a fit to the axion mass. We improved the high temperature axion mass by including threshold effects due to heavier quarks that become active as the temperature rises. This gives for the first time a well-motivated axion mass for all temperature regions. In particular, we have been able to compute in a consistent fashion the transition between the high and low temperature regimes, which were derived so far within different frameworks, i.e. the dilute instanton gas and chiral perturbation theory, and connected in a rather ad hoc and arbitrary way.

In chapter 7 we have used this axion mass to investigate its cosmological implications. We have mainly focused on the misalignment mechanism. It is well-known that the axion evolution is adiabatic after a few oscillations of the zero mode around its minimum, which we exploited to propagate the results forward. The non-trivial, i.e. non-adiabatic, evolution has been performed numerically in a radiation dominated FRW universe, appropriate for the axion parameter space of interest. The radiation dominated universe is treated exactly in that we use the correct, temperature-dependent relativistic degrees of freedom derived in chapter 6. We have found that the original analytic approximations, based on the old axion mass determinations, are in unexpectedly good agreement with our full numerical results. We believe this to be a lucky coincidence, considering the crude determination of the axion mass within the dilute gas approximation and its extrapolation to lower temperatures.

In the axion parameter space where the Peccei-Quinn symmetry breaks only after inflation, the axion evolution is considerably more complicated. In particular, the Kibble mechanism leads to cosmic strings and the assumption of homogeneity underlying the misalignment mechanism is only approximately satisfied if the network is dilute; which it is. We have updated the original results on the axion string network evolution to include the new axion mass and to incorporate the most recent cosmological parameters, which are substantially different from those used at the time. In this standard

thermal scenario, or classic axion window, we found $f_a < 2.8(2) 10^{11} \text{ GeV}$, $m_a > 21(2) \mu\text{eV}$ from the misalignment contribution. The latter is, however, subdominant to the contribution from axionic string radiation, whose bound reads $f_a \lesssim 3.2_{-2}^{+4} 10^{10} \text{ GeV}$, allowing for a thermal dark matter axion with mass $m_a \sim 200 \mu\text{eV}$.

The anthropic axion can only exist if the Peccei-Quinn symmetry breaks before inflation, and when the reheat temperature or quantum mechanical fluctuations do not restore the PQ symmetry subsequently. In the anthropic window the misalignment production channel is dominant and the misalignment angle becomes a new, environmental parameter. The dark matter constraint entangles θ_a and f_a . Inflation induces isocurvature fluctuations in the cosmic microwave background radiation, and intertwines the axion parameter space with the inflationary one.

We have found that the isocurvature constraint from the CMBR places strong bounds on f_a and H_I . Still, the anthropic window allows a dark matter axion that can have any initial misalignment angle, provided the inflationary scale is low enough. In light of naturalness in fundamental theory, most interest has focused on the constraints for large f_a , i.e. fine-tuned to $\pi = 0$. To extend the existing analysis to large axion angles for which the usual harmonic potential no longer provides an adequate approximation, we have developed a Taylor-expansion approach to incorporate the anharmonic effects. The latter lead to smaller isocurvature perturbations as compared to the harmonic case and, consequently lead to weaker constraints on the inflationary scale. In particular, anthropic tuning near $\theta \approx \pi$ allows inflationary dark matter axions ($\Omega_a = 0.23$) to have masses as high as $m_a \leq 1 \text{ meV}$, but no higher because quantum fluctuations restrict the fine tuning (for $10^4 \lesssim H_I \lesssim 10^9 \text{ GeV}$). In fact, the fine-tuning constraint for such low energy axion models places the strongest bounds on f_a and H_I if the axion is the dominant dark matter candidate. Note that both the anthropic misalignment and the thermal production channel have an overlapping allowed axion mass region in the range $0.1 - 1 \text{ meV}$. This should strengthen the case for enhancing the experimental effort to search for axion dark matter axions in the mass range $0.1 - 1 \text{ meV}$. Finally, for a subdominant dark matter axion the constraints change and the parameters space becomes 3-dimensional, because there is no longer a strict

functional relationship between f_a and θ_a .

These conclusions follow from the data that the present formulation of the IILM has provided. However, it is important to bear in mind that the IILM is only a model of QCD, and that we must also carefully compare it with available experimental or lattice data, say. As mentioned throughout the thesis, chiral properties of QCD are expected to be well described by the IILM at least at zero temperature; this has been checked by many studies so that we can be rather confident about our calibrations at zero temperature. At finite temperature the picture is slightly different. Firstly of all it has long been known that the IILM does not provide a mechanism for confinement. Secondly, in this thesis we have also shown that the topological susceptibility in the quenched case is qualitatively different from lattice results. The latter is no longer a chiral observable, so the earlier expectation for good agreement might not be justified in this case. In the unquenched case, we did not have lattice data available to adequately compare the IILM prediction. As mentioned before, this should be possible in the near future.

The discovery of the new non-trivial holonomy calorons [119, 121], [123] raises questions about whether the IILM would not be improved if based on these, especially since recent lattice studies indeed see evidence of the lumpy structure characteristic for an ensemble of these new caloron solutions [101]. Note that such an ensemble seems to give a better qualitative agreement [80] around the confinement/deconfinement phase transition. The latter study is based on a lattice formulation of the pure gauge instanton liquid, and it will be interesting to see whether the continuum model can corroborate these results; furthermore, the continuum formulation allows for a rather simple treatment of quarks and might be better suited to study the unquenched sector. We believe that the numerical framework set up in chapter 3 is well adapted to deal with these more complicated backgrounds.

The need to address more clearly the validity of the IILM at finite temperature has a direct implication for our predictions concerning the axion. In particular, it is possible that the high temperature dependence of the axion mass is qualitatively different: in our toy model, in chapter 4, we found that the high temperature limit need not be a non-interacting gas. In the IILM, such a behaviour is in principle possible if the screening effects do not

dominate over the quark overlap interactions when the quantum interactions are taken into account. It will be interesting to see how the quark overlaps behave if the non-trivial holonomy calorons are introduced. On a heuristic level, overlapping pairs have a smaller effective size and so the screening should be reduced. A considerably different fall-off behaviour of the axion mass will change the cosmological bounds decidedly: a softer decay, as seen in gluodynamics, will lead to weaker constraints; a more abrupt mass switch on, e.g. due to a higher instanton–anti-instanton molecule density, would lead to stronger upper bounds and could possibly rule out the classic axion window. The lattice community is performing realistic QCD simulations directly at the physical quark masses, and we can expect the best axion mass determination to follow from that data in the near future.

Other physical questions that can be addressed within such a microscopic model of the QCD vacuum are related to on-going investigations of the spontaneous breaking of P and CP in the strong interactions. It has been argued that around the phase transition meta-stable states could be exited which will break P and CP spontaneously [108]. One might hope that instanton models can shed some light on the mechanism for the formation of these P-odd bubbles as the topological susceptibility plays an important role [109]. Within the IILM, these excited vacuum domains would correspond to regions of non-zero topological charge, and can effectively be modeled by introducing explicitly a non-zero vacuum angle. This leads to charge separation [65, 106, 110, 107], which can in principle be studied in heavy-ion-collisions. Strong CP breaking might also lead to interesting cosmological implications: there are proposals to use the QCD phase transition and the CP-breaking bubbles to explain baryogenesis [28], the formation of the galactic magnetic fields [67] and even dark matter [209, 68].

Appendix A

Interactions at zero temperature

A.1 Gluonic Interactions

The ratio ansatz for an instanton–anti-instanton pair is defined by

$$A_\mu^a = -\frac{\bar{\eta}_{\mu\nu}^a \partial_\nu \Pi_1(x, \{x_1, \rho_1\}) + O^{ab} \eta_{\mu\nu}^b \partial_\nu \Pi_2(x, \{x_2, \rho_2\})}{1 + \Pi_1(x, \{x_1, \rho_1\}) + \Pi_2(x, \{x_2, \rho_2\})}, \quad (\text{A.1})$$

where $\Pi(x, \{y, \rho\}) = \frac{\rho^2}{r^2}$ and $O = O_1^t O_2$, with O_i the respective colour embeddings. A global colour rotation has been performed to bring the gauge potential into this form, which is irrelevant since the action is gauge invariant. Instanton–instanton and anti-instanton–anti-instanton pairs differ by having either only $\bar{\eta}$ or η in the above formula. A brute force computation then gives

$$\begin{aligned} F_{\mu\nu}^a F_{\mu\nu}^a = & I + (\text{Tr } O^t O + (\bar{\eta} O \eta)_{\mu\nu\mu\nu}) J + (\bar{\eta} O \eta)_{\rho\mu\rho\nu} I_{\mu\nu} \\ & + (\bar{\eta} O \eta)_{\mu\rho\nu\sigma} I_{\mu\rho\nu\sigma} + (\eta O^t O \eta)_{\mu\rho\nu\sigma} J_{\mu\rho\nu\sigma} + (\bar{\eta} O \eta)_{\alpha\mu\alpha\rho} (\bar{\eta} O \eta)_{\beta\nu\beta\sigma} K_{\mu\rho\nu\sigma}. \end{aligned} \quad (\text{A.2})$$

The different terms have the following form

$$\begin{aligned}
I = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} [(\partial_\mu \partial_\nu \Pi_1)(\partial_\mu \partial_\nu \Pi_1) + (\partial_\mu \partial_\nu \Pi_2)(\partial_\mu \partial_\nu \Pi_2)] \\
& - \frac{8}{(1 + \Pi_1 + \Pi_2)^3} [(\partial_\mu \partial_\nu \Pi_1)(\partial_\mu \Pi_1)(\partial_\nu \Pi_2) + (\partial_\mu \partial_\nu \Pi_2)(\partial_\mu \Pi_2)(\partial_\nu \Pi_1) \\
& \quad + 2(\partial_\mu \partial_\nu \Pi_1)(\partial_\mu \Pi_1)(\partial_\nu \Pi_1) + 2(\partial_\mu \partial_\nu \Pi_2)(\partial_\mu \Pi_2)(\partial_\nu \Pi_2)] \\
& + \frac{4}{(1 + \Pi_1 + \Pi_2)^4} [3(\partial_\mu \Pi_1 \partial_\mu \Pi_1)(\partial_\mu \Pi_1 \partial_\mu \Pi_2) + 3(\partial_\mu \Pi_2 \partial_\mu \Pi_2)(\partial_\mu \Pi_2 \partial_\mu \Pi_1) \\
& \quad + 3(\partial_\mu \Pi_1 \partial_\mu \Pi_1)^2 + 3(\partial_\mu \Pi_2 \partial_\mu \Pi_2)^2 + 2(\partial_\mu \Pi_1 \partial_\mu \Pi_1)(\partial_\mu \Pi_2 \partial_\mu \Pi_2) \\
& \quad + (\partial_\mu \Pi_1 \partial_\mu \Pi_2)^2] . \quad (\text{A.3})
\end{aligned}$$

$$J = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1 \partial_\mu \Pi_1)(\partial_\mu \Pi_2 \partial_\mu \Pi_2) . \quad (\text{A.4})$$

$$\begin{aligned}
I_{\mu\nu} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} (\partial_\mu \partial_\sigma \Pi_1)(\partial_\mu \partial_\sigma \Pi_2) \\
& + \frac{4}{(1 + \Pi_1 + \Pi_2)^3} [(\partial_\mu \partial_\nu \Pi_1)(\partial_\sigma \Pi_2 \partial_\sigma \Pi_2) + (\partial_\mu \partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_1) \\
& \quad - 2(\partial_\mu \partial_\sigma \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_2) - 2(\partial_\mu \Pi_1)(\partial_\sigma \Pi_1)(\partial_\nu \partial_\sigma \Pi_2) \\
& \quad - 2(\partial_\mu \partial_\sigma \Pi_1)(\partial_\sigma \Pi_1)(\partial_\nu \Pi_2) - 2(\partial_\mu \Pi_1)(\partial_\nu \partial_\sigma \Pi_2)(\partial_\sigma \Pi_2)] \\
& + \frac{4}{(1 + \Pi_1 + \Pi_2)^4} [-(\partial_\mu \Pi_1)(\partial_\nu \Pi_1)(\partial_\sigma \Pi_2 \partial_\sigma \Pi_2) \\
& \quad - (\partial_\mu \Pi_2)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_1) + 3(\partial_\mu \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_1) \\
& \quad + 3(\partial_\mu \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_2 \partial_\sigma \Pi_2) + 3(\partial_\mu \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_2)] . \quad (\text{A.5})
\end{aligned}$$

$$\begin{aligned}
I_{\mu\rho\nu\sigma} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} (\partial_\mu \partial_\nu \Pi_1)(\partial_\rho \partial_\sigma \Pi_2) \\
& + \frac{8}{(1 + \Pi_1 + \Pi_2)^3} [(\partial_\mu \Pi_2)(\partial_\rho \partial_\nu \Pi_1)(\partial_\sigma \Pi_2) + (\partial_\mu \Pi_1)(\partial_\rho \partial_\nu \Pi_2)(\partial_\sigma \Pi_1)] \\
& + \frac{8}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1)(\partial_\rho \Pi_2)(\partial_\nu \Pi_1)(\partial_\sigma \Pi_2) . \quad (\text{A.6})
\end{aligned}$$

$$J_{\mu\rho\nu\sigma} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1) (\partial_\rho \Pi_2) (\partial_\nu \Pi_1) (\partial_\sigma \Pi_2). \quad (\text{A.7})$$

$$K_{\mu\rho\nu\sigma} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1) (\partial_\rho \Pi_2) (\partial_\nu \Pi_1) (\partial_\sigma \Pi_2). \quad (\text{A.8})$$

A.1.1 Exact Interactions

When computing the look-up tables, we use global translations and rotations in \mathbb{R}^4 to place one instanton at the origin and the partner at $y'_4 = R = \sqrt{R_\mu R_\mu} = |y^{I_1} - y^{I_2}|$, where y^i are the instanton centres. The rotation will reemerge in contractions of R_μ with the colour structure, as we will now see. The relation between the position vector R_μ and $R'_\mu \equiv (0, 0, 0, R)$ is given by the following rotation matrix

$$\begin{aligned} R'_\mu &= \mathcal{O}_{\mu\nu}^t R_\nu, \\ \mathcal{O}_{\mu 4} &= \frac{R_\mu}{R}, \end{aligned} \quad (\text{A.9})$$

and the other components of the rotation matrix are irrelevant.

Note that, with the choice of R'_μ , the integrands are $O(3)$ symmetric in the subspace orthogonal to the 4-direction. Denoting the arguments of the 't Hooft potentials $\Pi(x, \{y, \rho\})$ by x_μ and $\tilde{x}_\mu \equiv x_\mu - R_\mu$, we can extract the R_μ from the integrands with help of the following formulas, which we order according to the tensor structure of the x_μ -dependence on the integrand.

$$\int x_\mu = \mathcal{O}_{\mu 4} \int x'_4. \quad (\text{A.10})$$

$$\int x_\mu x_\nu = \delta_{\mu\nu} \int x_1'^2 + \mathcal{O}_{\mu 4} \mathcal{O}_{\nu 4} \int (x_4'^2 - x_1'^2). \quad (\text{A.11})$$

$$\int x_\mu x_\nu x_\kappa = (\delta_{\mu\nu}\mathcal{O}_{\kappa 4} + \delta_{\kappa\mu}\mathcal{O}_{\nu 4} + \delta_{\nu\kappa}\mathcal{O}_{\mu 4}) \int x_1'^2 x_4' \quad (\text{A.12})$$

$$+ \mathcal{O}_{\mu 4}\mathcal{O}_{\nu 4}\mathcal{O}_{\kappa 4} \int (x_4'^3 - 3x_1'^2 x_4'). \quad (\text{A.13})$$

$$\int x_\mu x_\nu x_\kappa x_\delta = (\delta_{\mu\nu}\delta_{\kappa\delta} + \delta_{\mu\kappa}\delta_{\nu\delta} + \delta_{\mu\delta}\delta_{\kappa\nu}) \int x_1'^2 x_2'^2 \quad (\text{A.14})$$

$$+ (\delta_{\mu\nu}\mathcal{O}_{\kappa 4}\mathcal{O}_{\delta 4} + \text{perm.}) \int (x_4'^2 x_1'^2 - x_1'^2 x_2'^2) \quad (\text{A.15})$$

$$+ \mathcal{O}_{\mu 4}\mathcal{O}_{\nu 4}\mathcal{O}_{\kappa 4}\mathcal{O}_{\delta 4} \int (x_4'^4 - 6x_1'^2 x_4'^2 + 3x_1'^2 x_2'^2). \quad (\text{A.16})$$

Terms with \tilde{x} can be constructed from these. Incidentally, splitting the different integrands according to the above formulas is the most stable procedure numerically. Taking into account the antisymmetry of the 't Hooft symbols, we end up with the following integrands.

$$\begin{aligned} I = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} [(\Pi_1'')^2 + 3(\Pi_1'/r)^2 + (\Pi_2'')^2 + 3(\Pi_2'/\tilde{r})^2] \\ & - \frac{8}{(1 + \Pi_1 + \Pi_2)^3} \left[2\Pi_1''(\Pi_1')^2 + 2\Pi_2''(\Pi_2')^2 + \frac{x\tilde{x}}{r\tilde{r}} (\Pi_1''\Pi_1'\Pi_2' + \Pi_1'\Pi_2''\Pi_2') \right] \\ & + \frac{4}{(1 + \Pi_1 + \Pi_2)^4} \left[12(\Pi_1')^4 + 12(\Pi_2')^4 + 8(\Pi_1')^2(\Pi_2')^2 + 4\left(\frac{x\tilde{x}}{r\tilde{r}}\Pi_1'\Pi_2'\right)^2 \right. \\ & \quad \left. + 12(\Pi_1')^2\left(\frac{x\tilde{x}}{r\tilde{r}}\Pi_1'\Pi_2'\right) + 12\left(\frac{x\tilde{x}}{r\tilde{r}}\Pi_1'\Pi_2'\right)(\Pi_2')^2 \right]. \quad (\text{A.17}) \end{aligned}$$

Note that to achieve good numerical precision, we need to subtract the one-instanton integrands from the above before performing the numerical integration.

$$J = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\Pi_1')^2 (\Pi_2')^2. \quad (\text{A.18})$$

$$I_{\mu\nu} = \delta_{\mu\nu} \tilde{I}_{\mu\mu} + \frac{R_\mu R_\nu}{R^2} \tilde{I}_{\mu\nu}. \quad (\text{A.19})$$

$$\begin{aligned} \tilde{I}_{\mu\mu} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x_1'^2}{r^2} (\Pi_1'' - (\Pi_1'/r)) (\Pi_2'/\tilde{r}) \right. \\ & + \frac{x_1'^2}{\tilde{r}^2} (\Pi_1'/r) (\Pi_2'' - (\Pi_2'/\tilde{r})) + (\Pi_1'/r) (\Pi_2'/\tilde{r}) \\ & \left. + \frac{x_1'^2}{r\tilde{r}} \frac{x\tilde{x}}{r\tilde{r}} (\Pi_1''\Pi_2'' - \Pi_1''(\Pi_2'/\tilde{r}) - (\Pi_1'/r)\Pi_2'' + (\Pi_1'/r)(\Pi_2'/\tilde{r})) \right] \\ & + \frac{1}{(1 + \Pi_1 + \Pi_2)^3} \left[4((\Pi_1'/r)(\Pi_2')^2 + (\Pi_1')^2(\Pi_2'/\tilde{r})) \right. \\ & + \frac{x_1'^2}{r^2} (4(\Pi_1'' - (\Pi_1'/r))(\Pi_2')^2 - 8(\Pi_1')^2(\Pi_2'/\tilde{r})) \\ & + \frac{x_1'^2}{\tilde{r}^2} (4(\Pi_1')^2(\Pi_2'' - (\Pi_2'/\tilde{r})) - 8(\Pi_1'/r)(\Pi_2')^2) \\ & + \frac{x_1'^2}{r\tilde{r}} (-8\frac{x\tilde{x}}{r\tilde{r}}((\Pi_1'' - (\Pi_1'/r))(\Pi_2')^2 + (\Pi_1')^2(\Pi_2'' - (\Pi_2'/\tilde{r}))) \\ & \quad \left. - 8\Pi_1''\Pi_1'\Pi_2' - 8\Pi_1'\Pi_2''\Pi_2') \right] \\ & + \frac{1}{(1 + \Pi_1 + \Pi_2)^4} \left[-4\frac{x_1'^2}{r^2} (\Pi_1')^2(\Pi_2')^2 - 4\frac{x_1'^2}{\tilde{r}^2} (\Pi_1')^2(\Pi_2')^2 \right. \\ & \quad \left. + 12\frac{x_1'^2}{r\tilde{r}} \Pi_1'\Pi_2'((\Pi_1')^2 + (\Pi_2')^2 + \frac{x\tilde{x}}{r\tilde{r}}\Pi_1'\Pi_2') \right]. \quad (\text{A.20}) \end{aligned}$$

$$\begin{aligned}
\tilde{I}_{\mu\nu} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x_4'^2 - x_1'^2}{r^2} (\Pi_1'' - (\Pi_1'/r)) (\Pi_2'/\tilde{r}) \right. \\
& + \frac{(x_4' - R)^2 - x_1'^2}{\tilde{r}^2} (\Pi_1'/r) (\Pi_2'' - (\Pi_2'/\tilde{r})) \\
& + \frac{x_4'(x_4' - R) - x_1'^2}{r\tilde{r}} \frac{x\tilde{x}}{r\tilde{r}} (\Pi_1''\Pi_2'' - \Pi_1''(\Pi_2'/\tilde{r}) - (\Pi_1'/r)\Pi_2'' + (\Pi_1'/r)(\Pi_2'/\tilde{r})) \left. \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x_4'^2 - x_1'^2}{r^2} (4(\Pi_1'' - (\Pi_1'/r))(\Pi_2')^2 - 8(\Pi_1')^2(\Pi_2'/\tilde{r})) \right. \\
& + \frac{(x_4' - R)^2 - x_1'^2}{\tilde{r}^2} (4(\Pi_1')^2(\Pi_2'' - (\Pi_2'/\tilde{r})) - 8(\Pi_1'/r)(\Pi_2')^2) \\
& + \frac{x_4'(x_4' - R) - x_1'^2}{r\tilde{r}} (-8\frac{x\tilde{x}}{r\tilde{r}} ((\Pi_1'' - (\Pi_1'/r))(\Pi_2')^2 + (\Pi_1')^2(\Pi_2'' - (\Pi_2'/\tilde{r}))) \\
& \quad \left. - 8\Pi_1''\Pi_1'\Pi_2' - 8\Pi_1'\Pi_2''\Pi_2' \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^4} \left[-4\frac{x_4'^2 - x_1'^2}{r^2} (\Pi_1')^2(\Pi_2')^2 - 4\frac{(x_4' - R)^2 - x_1'^2}{\tilde{r}^2} (\Pi_1')^2(\Pi_2')^2 \right. \\
& \quad \left. + 12\frac{x_4'(x_4' - R) - x_1'^2}{r\tilde{r}} \Pi_1'\Pi_2'((\Pi_1')^2 + (\Pi_2')^2 + \frac{x\tilde{x}}{r\tilde{r}} \Pi_1'\Pi_2') \right] . \quad (\text{A.21})
\end{aligned}$$

$$I_{\mu\rho\nu\sigma} = \delta_{\mu\nu}\delta_{\rho\sigma}\tilde{I}_{\mu\nu\mu\nu} + \delta_{\mu\nu}\frac{R_\rho R_\sigma}{R^2}\tilde{I}_{\mu\rho\mu\sigma} . \quad (\text{A.22})$$

$$\tilde{I}_{\mu\nu\mu\nu} = 0 \quad (\text{analytically}) . \quad (\text{A.23})$$

$$\begin{aligned}
\tilde{I}_{\mu\rho\mu\sigma} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x_4'^2 - x_1'^2}{r^2} (\Pi_1'' - (\Pi_1'/r)) (\Pi_2'/\tilde{r}) \right. \\
& + \frac{(x_4' - R)^2 - x_1'^2}{\tilde{r}^2} (\Pi_1'/r) (\Pi_2'' - (\Pi_2'/\tilde{r})) \\
& \left. \frac{x_1'^2 R^2}{(r\tilde{r})^2} (\Pi_1'' - (\Pi_1'/r)) (\Pi_2'' - (\Pi_2'/\tilde{r})) \right] \\
- & \frac{8}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x_4'^2 - x_1'^2}{r^2} (\Pi_1')^2 (\Pi_2'/\tilde{r}) + \frac{(x_4' - R)^2 - x_1'^2}{\tilde{r}^2} (\Pi_1'/r) (\Pi_2')^2 \right. \\
& \left. + \frac{x_1'^2 R^2}{(r\tilde{r})^2} ((\Pi_1'' - (\Pi_1'/r)) (\Pi_2')^2 + (\Pi_1')^2 (\Pi_2'' - (\Pi_2'/\tilde{r}))) \right] \\
+ & \frac{8}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_1'^2 R^2}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2 \right]. \tag{A.24}
\end{aligned}$$

$$\tilde{J}_{\mu\rho\nu\sigma} = \delta_{\mu\nu} \frac{R_\rho R_\sigma}{R^2} \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_1'^2 R^2}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \tag{A.25}$$

$$\begin{aligned}
\tilde{K}_{\mu\rho\nu\sigma} = & [(\delta_{\mu\nu}\delta_{\rho\sigma} + \delta_{\mu\rho}\delta_{\nu\sigma} + \delta_{\mu\sigma}\delta_{\rho\nu})x_1'^2 x_2'^2 \\
& + \delta_{\mu\nu} \frac{R_\rho R_\sigma}{R^2} ((x_4' - R)^2 x_1'^2 - x_1'^2 x_2'^2) + \delta_{\rho\sigma} \frac{R_\mu R_\nu}{R^2} (x_4'^2 x_1'^2 - x_1'^2 x_2'^2) \\
& + (\delta_{\mu\rho} \frac{R_\nu R_\sigma}{R^2} + \delta_{\mu\sigma} \frac{R_\nu R_\rho}{R^2} + \delta_{\nu\rho} \frac{R_\mu R_\sigma}{R^2} + \delta_{\nu\sigma} \frac{R_\mu R_\rho}{R^2}) (x_4' (x_4' - R) x_1'^2 - x_1'^2 x_2'^2) \\
& + \frac{R_\mu R_\nu R_\rho R_\sigma}{R^4} (x_4'^2 (x_4' - R)^2 + 3x_1'^2 x_2'^2 - x_4'^2 x_1'^2 - (x_4' - R)^2 x_1'^2 \\
& - 4x_4' (x_4' - R) x_1'^2)] \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{1}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \tag{A.26}
\end{aligned}$$

A.1.2 Asymptotic Interactions

As explained in the main text, the small separation asymptotic formulas get contributions which have the same functional form as those for the large separation asymptotics; the difference lies in the integration limit. We will therefore start with the large separation formulas and leave the integrals explicit.

Large Separation

The upper integration limit z follows from variable substitution and has the the following for integration over I_1 , with I_2 held fixed,

$$z_1^2 = \frac{1 + \Pi_2}{\rho_1^2} r^2. \quad (\text{A.27})$$

Apart from the dependence of z on Π , the rational form of the 't Hooft potential allows for a complete factoring out of Π under the above mentioned variable substitution. For the large separation formulas it is understood that $z^2 \rightarrow \infty$ because the initial integration variable r extends to infinity.

The integral over I contains terms that do not mix the 't Hooft potential Π_1 and Π_2 except for the denominators. At zeroth order in our expansion, these terms can be transformed to exactly match the single instanton contributions by exploiting scale invariance. Remembering that we actually subtract the one-instanton contributions to get the interactions, we can neglect these terms altogether. We then end up with the following formulas.

$$\int I = 72\pi^2 \rho^2 \frac{\partial_\mu \Pi \partial_\mu \Pi}{(1 + \Pi)^3} \int^z \frac{s^5 ds}{(s^2 + 1)^4} + \text{sym}. \quad (\text{A.28})$$

$$\int J = 16\pi^2 \rho^2 \frac{\partial_\mu \Pi \partial_\mu \Pi}{(1 + \Pi)^3} \int^z \frac{s^5 ds}{(s^2 + 1)^4} + \text{sym}. \quad (\text{A.29})$$

$$\begin{aligned} \int I_{\mu\nu} = & 16\pi^2 \rho^2 \frac{\partial_\mu \partial_\nu \Pi}{(1 + \Pi)^2} \int^z \frac{s^3 ds}{(s^2 + 1)^3} \\ & - \left(8\pi^2 \rho^2 \delta_{\mu\nu} \frac{\partial_\sigma \Pi \partial_\sigma \Pi}{(1 + \Pi)^3} + 8\pi^2 \rho^2 \frac{(\partial_\mu \Pi)(\partial_\nu \Pi)}{(1 + \Pi)^3} \right) \int^z \frac{s^5 ds}{(s^2 + 1)^4} + \text{sym}. \end{aligned} \quad (\text{A.30})$$

At zeroth order, partial integration and the antisymmetry of the 't Hooft symbols can be used to simplify

$$I_{\mu\rho\nu\sigma} \rightarrow \frac{8}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1)(\partial_\rho \Pi_2)(\partial_\nu \Pi_1)(\partial_\sigma \Pi_2), \quad (\text{A.31})$$

with asymptotic behaviour

$$\int I_{\mu\rho\nu\sigma} = 16\pi^2 \rho^2 \delta_{\mu\nu} \frac{(\partial_\rho \Pi)(\partial_\sigma \Pi)}{(1 + \Pi)^3} \int^z \frac{s^5 ds}{(s^2 + 1)^4} + \text{sym}. \quad (\text{A.32})$$

$$\int J_{\mu\rho\nu\sigma} = 4\pi^2 \rho^2 \delta_{\mu\nu} \frac{(\partial_\rho \Pi)(\partial_\sigma \Pi)}{(1 + \Pi)^3} \int^z \frac{s^5 ds}{(s^2 + 1)^4} + \text{sym}. \quad (\text{A.33})$$

For $K_{\mu\rho\nu\sigma}$ no 't Hooft symbols can be used to exchange the index pairs $(\mu, \nu) \leftrightarrow (\rho, \sigma)$, and so we cannot simplify with a symmetry argument anymore.

$$\begin{aligned} \int K_{\mu\rho\nu\sigma} = & 4\pi^2 \rho_1^2 \delta_{\mu\nu} \frac{(\partial_\rho \Pi_2)(\partial_\sigma \Pi_2)}{(1 + \Pi_2)^3} \int^{z_1} \frac{s^5 ds}{(s^2 + 1)^4} \\ & + 4\pi^2 \rho_2^2 \delta_{\rho\sigma} \frac{(\partial_\mu \Pi_1)(\partial_\nu \Pi_1)}{(1 + \Pi_1)^3} \int^{z_2} \frac{s^5 ds}{(s^2 + 1)^4}. \end{aligned} \quad (\text{A.34})$$

Small Separation

As explained in the main text, the small separation asymptotic formulas get contributions from the large asymptotics. Also, in this case we have performed a global translation so that the instantons sit at $\pm R_\mu/2$. Therefore, in the large separation formulas we need to put $z^2 = \frac{1+\Pi}{\rho^2}(R/2)^2$.

We now turn to the proper small separation asymptotic formulas that encode the repulsion through the gauge singularity. We will again introduce an explicit upper limit for the integrals; abusing notation we will use the same letter as before, but here the meaning becomes

$$z^2 = \frac{R^2}{\rho_1^2 + \rho_2^2}, \quad z_i^2 = \frac{R^2}{\rho_i^2}. \quad (\text{A.35})$$

To derive these formulas, we approximate the arguments $x_\mu \pm R_\mu/2 \rightarrow x_\mu$. We have, therefore, explicitly restored $O(4)$ symmetry, which can be exploited to set several integrals to zero. Eventually, we arrive at

$$\begin{aligned}
\int I = 384\pi^2 \Bigg[& \frac{\rho_1^4 + \rho_2^4}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^2} \\
& - \left(\frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} + 2 \frac{\rho_1^6 + \rho_2^6}{(\rho_1^2 + \rho_2^2)^3} \right) \int_z \frac{ds}{s(s^2 + 1)^3} \\
& \frac{\rho_1^8 + \rho_2^8 + \rho_1^4 \rho_2^4 + \rho_1^6 \rho_2^2 + \rho_1^2 \rho_2^6}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4} \\
& - \int_{z_1} \frac{s^4 ds}{s(s^2 + 1)^4} - \int_{z_2} \frac{s^4 ds}{s(s^2 + 1)^4} \Bigg] . \quad (\text{A.36})
\end{aligned}$$

$$\int J = 64\pi^2 \frac{\rho_1^4 \rho_2^4}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4} . \quad (\text{A.37})$$

$$\begin{aligned}
\int I_{\mu\nu} = \delta_{\mu\nu} \Bigg[& 96\pi^2 \frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^2} \\
& - 192\pi^2 \frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^3} \\
& + 32\pi^2 \frac{\rho_1^4 \rho_2^4 + 3\rho_1^6 \rho_2^2 + 3\rho_1^2 \rho_2^6}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4} \Bigg] , \quad (\text{A.38})
\end{aligned}$$

$$\begin{aligned}
\int I_{\mu\rho\nu\sigma} = \delta_{\mu\nu} \delta_{\rho\sigma} \Bigg[& -32\pi^2 \frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^2} \\
& + 32\pi^2 \frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^3} \Bigg] . \quad (\text{A.39})
\end{aligned}$$

$$\int J_{\mu\rho\nu\sigma} = 0 . \quad (\text{A.40})$$

$$\int K_{\mu\rho\nu\sigma} = \frac{8}{3}\pi^2 (\delta_{\mu\nu} \delta_{\rho\sigma} + \delta_{\mu\rho} \delta_{\nu\sigma} + \delta_{\mu\sigma} \delta_{\nu\rho}) \frac{\rho_1^4 \rho_2^4}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4} . \quad (\text{A.41})$$

A.2 Fermionic Interactions

The Dirac overlap matrix elements are given by

$$T_{IA} = \int d^4x \frac{1}{4\pi^2 \rho_I \rho_A} \frac{1}{2} \text{Tr} (U \tau_\beta^+) I_\beta. \quad (\text{A.42})$$

Note that $\frac{1}{2} \text{Tr} (U \tau_\beta^+) \equiv iu_\beta$ is the colour four-vector used for instance in [157]. After some straightforward algebra, we find that I_β has the following form

$$I_\beta = \frac{-1}{(1 + \Pi_I + \Pi_A)(1 + \Pi_I)^{3/2}(1 + \Pi_A)^{3/2}} \left(\frac{\Pi_A}{1 + \Pi_I} (\partial_\mu \Pi_I \partial_\mu \Pi_I) \partial_\beta \Pi_A + (\partial_\mu \Pi_A \partial_\mu \Pi_A) \partial_\beta \Pi_I \right). \quad (\text{A.43})$$

A.2.1 Exact Interactions

Using the same rotations (A.9) as for the gluonic interactions to marry the space-time with the colour indices, we get

$$I_\beta = \frac{R_\beta}{R} \frac{-3}{(1 + \Pi_I + \Pi_A)(1 + \Pi_I)^{3/2}(1 + \Pi_A)^{3/2}} \left\{ \frac{x'_4}{r} \Pi'_I (\Pi'_A)^2 + \frac{x'_4 - R}{\tilde{r}} (\Pi'_I)^2 \Pi'_A \frac{\Pi_A}{1 + \Pi_I} \right\}. \quad (\text{A.44})$$

A.2.2 Asymptotic Interactions

Large Separation

In order to get rather simple formulas, we make the following additional simplification

$$1 + \Pi_I + \Pi_A \rightarrow \begin{cases} 1 + \Pi_I : & \text{Integration over } \Pi_I \\ 1 + \Pi_A : & \text{Integration over } \Pi_A \end{cases}. \quad (\text{A.45})$$

Given these further assumption, we can proceed as for the gluonic interactions. Finally, caution needs to be taken in the case of an anti-instanton because it sits at $-R_\mu$ so that $\partial_\beta \Pi_A$ generates an extra minus sign.

$$\int I_\beta = 8\pi^2 \rho_I^2 \frac{\Pi_A \partial_\beta \Pi_A}{(1 + \Pi_A)^{3/2}} \int^{z_I} \frac{s^4 ds}{(s^2 + 1)^{7/2}} - 8\pi^2 \rho_A^2 \frac{\partial_\beta \Pi_I}{(1 + \Pi_I)^{3/2}} \int^{z_A} \frac{s^4 ds}{(s^2 + 1)^{5/2}}. \quad (\text{A.46})$$

Small Separation

At zeroth order, i.e. $x_\mu \pm R_\mu/2 \rightarrow x_\mu$, the contribution to I_β vanishes because of $O(4)$ symmetry. It turns out that the large separation asymptotics falls off too quickly as $R \rightarrow 0$. However, this is not important because in this regime the gluonic interaction is dominant.

Appendix B

Interactions at finite temperature

B.1 Gluonic Interactions

The ratio ansatz for an instanton–anti-instanton pair is defined by

$$A_\mu^a = -\frac{\bar{\eta}_{\mu\nu}^a \partial_\nu \Pi_1(x, \{x_1, \rho_1\}) + O^{ab} \eta_{\mu\nu}^b \partial_\nu \Pi_2(x, \{x_2, \rho_2\})}{1 + \Pi_1(x, \{x_1, \rho_1\}) + \Pi_2(x, \{x_2, \rho_2\})}, \quad (\text{B.1})$$

$$\Pi(x, \{y, \rho\}) = \frac{\pi \rho^2}{\beta r} \frac{\sinh \frac{2\pi r}{\beta}}{\cosh \frac{2\pi r}{\beta} - \cos \frac{2\pi t}{\beta}}, \quad (\text{B.2})$$

where $O = O_1^t O_2$, with O_i the respective colour embeddings. A global colour rotation has been performed to bring the gauge potential into this form. Since we are ultimately interested in the action, we do not need to bother about it since the action is gauge invariant. Instanton–instanton and anti-instanton–anti-instanton pairs differ by having either only $\bar{\eta}$ or η in the above formula. A brute force computation then gives

$$\begin{aligned} F_{\mu\nu}^a F_{\mu\nu}^a &= I + (\text{Tr } O^t O + (\bar{\eta} O \eta)_{\mu\nu\mu\nu}) J + (\bar{\eta} O \eta)_{\rho\mu\rho\nu} I_{\mu\nu} \\ &+ (\bar{\eta} O \eta)_{\mu\rho\nu\sigma} I_{\mu\rho\nu\sigma} + (\eta O^t O \eta)_{\mu\rho\nu\sigma} J_{\mu\rho\nu\sigma} + (\bar{\eta} O \eta)_{\alpha\mu\alpha\rho} (\bar{\eta} O \eta)_{\beta\nu\beta\sigma} K_{\mu\rho\nu\sigma}. \end{aligned} \quad (\text{B.3})$$

The different terms have the following form

$$\begin{aligned}
I = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} [(\partial_\mu \partial_\nu \Pi_1)(\partial_\mu \partial_\nu \Pi_1) + (\partial_\mu \partial_\nu \Pi_2)(\partial_\mu \partial_\nu \Pi_2)] \\
& - \frac{8}{(1 + \Pi_1 + \Pi_2)^3} [(\partial_\mu \partial_\nu \Pi_1)(\partial_\mu \Pi_1)(\partial_\nu \Pi_2) + (\partial_\mu \partial_\nu \Pi_2)(\partial_\mu \Pi_2)(\partial_\nu \Pi_1) \\
& \quad + 2(\partial_\mu \partial_\nu \Pi_1)(\partial_\mu \Pi_1)(\partial_\nu \Pi_1) + 2(\partial_\mu \partial_\nu \Pi_2)(\partial_\mu \Pi_2)(\partial_\nu \Pi_2)] \\
& + \frac{4}{(1 + \Pi_1 + \Pi_2)^4} [3(\partial_\mu \Pi_1 \partial_\mu \Pi_1)(\partial_\mu \Pi_1 \partial_\mu \Pi_2) + 3(\partial_\mu \Pi_2 \partial_\mu \Pi_2)(\partial_\mu \Pi_2 \partial_\mu \Pi_1) \\
& \quad + 3(\partial_\mu \Pi_1 \partial_\mu \Pi_1)^2 + 3(\partial_\mu \Pi_2 \partial_\mu \Pi_2)^2 + 2(\partial_\mu \Pi_1 \partial_\mu \Pi_1)(\partial_\mu \Pi_2 \partial_\mu \Pi_2) \\
& \quad + (\partial_\mu \Pi_1 \partial_\mu \Pi_2)^2] . \quad (B.4)
\end{aligned}$$

$$J = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1 \partial_\mu \Pi_1)(\partial_\mu \Pi_2 \partial_\mu \Pi_2) . \quad (B.5)$$

$$\begin{aligned}
I_{\mu\nu} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} (\partial_\mu \partial_\sigma \Pi_1)(\partial_\mu \partial_\sigma \Pi_2) \\
& + \frac{4}{(1 + \Pi_1 + \Pi_2)^3} [(\partial_\mu \partial_\nu \Pi_1)(\partial_\sigma \Pi_2 \partial_\sigma \Pi_2) + (\partial_\mu \partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_1) \\
& \quad - 2(\partial_\mu \partial_\sigma \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_2) - 2(\partial_\mu \Pi_1)(\partial_\sigma \Pi_1)(\partial_\nu \partial_\sigma \Pi_2) \\
& \quad - 2(\partial_\mu \partial_\sigma \Pi_1)(\partial_\sigma \Pi_1)(\partial_\nu \Pi_2) - 2(\partial_\mu \Pi_1)(\partial_\nu \partial_\sigma \Pi_2)(\partial_\sigma \Pi_2)] \\
& + \frac{4}{(1 + \Pi_1 + \Pi_2)^4} [-(\partial_\mu \Pi_1)(\partial_\nu \Pi_1)(\partial_\sigma \Pi_2 \partial_\sigma \Pi_2) \\
& \quad - (\partial_\mu \Pi_2)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_1) + 3(\partial_\mu \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_1) \\
& \quad + 3(\partial_\mu \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_2 \partial_\sigma \Pi_2) + 3(\partial_\mu \Pi_1)(\partial_\nu \Pi_2)(\partial_\sigma \Pi_1 \partial_\sigma \Pi_2)] . \quad (B.6)
\end{aligned}$$

$$\begin{aligned}
I_{\mu\rho\nu\sigma} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} (\partial_\mu \partial_\nu \Pi_1)(\partial_\rho \partial_\sigma \Pi_2) \\
& + \frac{8}{(1 + \Pi_1 + \Pi_2)^3} [(\partial_\mu \Pi_2)(\partial_\rho \partial_\nu \Pi_1)(\partial_\sigma \Pi_2) + (\partial_\mu \Pi_1)(\partial_\rho \partial_\nu \Pi_2)(\partial_\sigma \Pi_1)] \\
& + \frac{8}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1)(\partial_\rho \Pi_2)(\partial_\nu \Pi_1)(\partial_\sigma \Pi_2) . \quad (B.7)
\end{aligned}$$

$$J_{\mu\rho\nu\sigma} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1) (\partial_\rho \Pi_2) (\partial_\nu \Pi_1) (\partial_\sigma \Pi_2). \quad (\text{B.8})$$

$$K_{\mu\rho\nu\sigma} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1) (\partial_\rho \Pi_2) (\partial_\nu \Pi_1) (\partial_\sigma \Pi_2). \quad (\text{B.9})$$

B.1.1 Exact Interactions

When computing the look-up tables we use global translations and rotations in \mathbb{R}^3 , and periodicity in S^1 , to place one instanton at the origin and the partner at $y'_3 = R_s = \sqrt{R_i R_i} = |y_s^{I_1} - y_s^{I_2}|$ and $y'_4 = R_t = |y_t^{I_1} - y_t^{I_2}|$; y^i are the instanton centres. The rotation will reemerge in contractions of R_i with the colour structure as we will now see. The relation between the spatial position vector R_i and $R'_i \equiv (0, 0, R_s)$ is given by the following $O(3)$ rotation matrix

$$R'_i = \mathcal{O}_{ij}^t R_j, \quad (\text{B.10})$$

$$\mathcal{O}_{i3} = \frac{R_i}{R_s}, \quad (\text{B.11})$$

and the other components of the rotation matrix are irrelevant.

Note that with the choice of R'_i the integrands are $O(2)$ symmetric in the subspace orthogonal to the 3-direction. Denoting the spatial arguments of the 't Hooft potentials, $\Pi(x, \{y, \rho\})$, by x_i and $\tilde{x}_i \equiv x_i - R_i$, we can extract the R_i from the integrands with help of the following formulas, which we order according to the tensor structure in the x_i -dependence of the integrand

$$\int x_i = \mathcal{O}_{i3} \int x'_3. \quad (\text{B.12})$$

$$\int x_i x_j = \delta_{ij} \int x_1'^2 + \mathcal{O}_{i3} \mathcal{O}_{j3} \int (x_3'^2 - x_1'^2). \quad (\text{B.13})$$

$$\int x_i x_j x_k = (\delta_{ij} \mathcal{O}_{k3} + \delta_{ki} \mathcal{O}_{j3} + \delta_{jk} \mathcal{O}_{i3}) \int x_1'^2 x_3' \quad (\text{B.14})$$

$$+ \mathcal{O}_{i3} \mathcal{O}_{j3} \mathcal{O}_{k3} \int (x_3'^3 - 3x_1'^2 x_3'). \quad (\text{B.15})$$

$$\int x_i x_j x_k x_h = (\delta_{ij} \delta_{kh} + \delta_{ik} \delta_{jh} + \delta_{ih} \delta_{kj}) \int x_1'^2 x_2'^2 \quad (\text{B.16})$$

$$+ (\delta_{ij} \mathcal{O}_{k3} \mathcal{O}_{h3} + \text{perm.}) \int (x_3'^2 x_1'^2 - x_1'^2 x_2'^2) \quad (\text{B.17})$$

$$+ \mathcal{O}_{i3} \mathcal{O}_{j3} \mathcal{O}_{k3} \mathcal{O}_{h3} \int (x_3'^4 - 6x_1'^2 x_3'^2 + 3x_1'^2 x_2'^2). \quad (\text{B.18})$$

Insertion of \tilde{x} can be constructed from these. Incidentally splitting the different integrands according to the above formulas is the most stable procedure numerically. Taking into account the antisymmetry of the 't Hooft symbols, we end up with the following integrands

$$\begin{aligned} I = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[(\Pi_1'')^2 + 2(\Pi_1'/r)^2 + 2(\dot{\Pi}_1')^2 + (\ddot{\Pi}_1')^2 \right. \\ & \left. + (\Pi_2'')^2 + 2(\Pi_2'/\tilde{r})^2 + 2(\dot{\Pi}_2')^2 + (\ddot{\Pi}_2')^2 \right] \\ & - \frac{8}{(1 + \Pi_1 + \Pi_2)^3} \left[2\Pi_1''(\Pi_1')^2 + 4\dot{\Pi}_1' \dot{\Pi}_1 \Pi_1' + 2\ddot{\Pi}_1(\dot{\Pi}_1)^2 \right. \\ & \quad + 2\Pi_2''(\Pi_2')^2 + 4\dot{\Pi}_2' \dot{\Pi}_2 \Pi_2' + 2\ddot{\Pi}_2(\dot{\Pi}_2)^2 \\ & \quad + \dot{\Pi}_1' \Pi_1' \dot{\Pi}_2 + \dot{\Pi}_1 \dot{\Pi}_2' \Pi_2' + \ddot{\Pi}_1 \dot{\Pi}_1 \dot{\Pi}_2 + \dot{\Pi}_1 \ddot{\Pi}_2 \dot{\Pi}_2 \\ & \quad \left. + \frac{x\tilde{x}}{r\tilde{r}} \left(\Pi_1'' \Pi_1' \Pi_2' + \Pi_1' \Pi_2'' \Pi_2' + \dot{\Pi}_1' \dot{\Pi}_1 \Pi_2' + \Pi_1' \dot{\Pi}_2' \dot{\Pi}_2 \right) \right] \\ & + \frac{4}{(1 + \Pi_1 + \Pi_2)^4} \left[12((\Pi_1')^2 + (\dot{\Pi}_1')^2)^2 + 12((\Pi_2')^2 + (\dot{\Pi}_2')^2) \right. \\ & \quad + 8((\Pi_1')^2 + (\dot{\Pi}_1')^2)((\Pi_2')^2 + (\dot{\Pi}_2')^2) + 4\left(\frac{x\tilde{x}}{r\tilde{r}} \Pi_1' \Pi_2' + \dot{\Pi}_1 \dot{\Pi}_2\right)^2 \\ & \quad \left. + 12((\Pi_1')^2 + (\dot{\Pi}_1')^2 + (\Pi_2')^2 + (\dot{\Pi}_2')^2) \left(\frac{x\tilde{x}}{r\tilde{r}} \Pi_1' \Pi_2' + \dot{\Pi}_1 \dot{\Pi}_2 \right) \right]. \quad (\text{B.19}) \end{aligned}$$

Note that to achieve good numerical precision, we need to subtract the

one-instanton integrands from the above before performing the numerical integration

$$J = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} ((\Pi'_1)^2 + (\dot{\Pi}_1)^2)((\Pi'_2)^2 + (\dot{\Pi}_2)^2). \quad (\text{B.20})$$

$$I_{\mu\nu} = \delta_{ij} \tilde{I}_{ii} + \frac{R_i R_j}{R_s^2} \tilde{I}_{ij} + \frac{R_i}{R_s} \tilde{I}_{it} + \frac{R_j}{R_s} \tilde{I}_{tj} + \tilde{I}_{tt}. \quad (\text{B.21})$$

$$\begin{aligned} \tilde{I}_{ii} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x_1'^2}{r^2} (\Pi_1'' - (\Pi'_1/r)) (\Pi'_2/\tilde{r}) \right. \\ & + \frac{x_1'^2}{\tilde{r}^2} (\Pi'_1/r) (\Pi_2'' - (\Pi'_2/\tilde{r})) + (\Pi'_1/r) (\Pi'_2/\tilde{r}) \\ & + \frac{x_1'^2}{r\tilde{r}} \left(\frac{x\tilde{x}}{r\tilde{r}} (\Pi_1''\Pi_2'' - \Pi_1''(\Pi'_2/\tilde{r}) - (\Pi'_1/r)\Pi_2'' + (\Pi'_1/r)(\Pi'_2/\tilde{r})) + \dot{\Pi}_1'\dot{\Pi}_2' \right) \Big] \\ & + \frac{1}{(1 + \Pi_1 + \Pi_2)^3} \left[4((\Pi'_1/r)((\Pi'_2)^2 + (\dot{\Pi}_1)^2) + ((\Pi'_1)^2 + (\dot{\Pi}_2)^2)(\Pi'_2/\tilde{r})) \right. \\ & + \frac{x_1'^2}{r^2} (4(\Pi_1'' - (\Pi'_1/r))((\Pi'_2)^2 + (\dot{\Pi}_2)^2) - 8(\Pi'_1)^2(\Pi'_2/\tilde{r})) \\ & + \frac{x_1'^2}{\tilde{r}^2} (4((\Pi'_1)^2 + (\dot{\Pi}_1)^2)(\Pi_2'' - (\Pi'_2/\tilde{r})) - 8(\Pi'_1/r)(\Pi'_2)^2) \\ & + \frac{x_1'^2}{r\tilde{r}} (-8\frac{x\tilde{x}}{r\tilde{r}} ((\Pi_1'' - (\Pi'_1/r))(\Pi'_2)^2 + (\Pi'_1)^2(\Pi_2'' - (\Pi'_2/\tilde{r}))) \\ & - 8\dot{\Pi}_1'\dot{\Pi}_2'\Pi'_2 - 8\dot{\Pi}_1\Pi'_1\dot{\Pi}_2' - 8(\Pi_1''\Pi'_1 + \dot{\Pi}_1'\dot{\Pi}_1)\Pi'_2 \\ & \left. - 8\Pi'_1(\Pi_2''\Pi'_2 + \dot{\Pi}_2'\dot{\Pi}_2') \right] \\ & + \frac{1}{(1 + \Pi_1 + \Pi_2)^4} \left[-4\frac{x_1'^2}{r^2} (\Pi'_1)^2((\Pi'_2)^2 + (\dot{\Pi}_2)^2) - 4\frac{x_1'^2}{\tilde{r}^2} ((\Pi'_1)^2 + (\dot{\Pi}_1)^2)(\Pi'_2)^2 \right. \\ & \left. + 12\frac{x_1'^2}{r\tilde{r}} \Pi'_1\Pi'_2((\Pi'_1)^2 + (\dot{\Pi}_1)^2 + (\Pi'_2)^2 + (\dot{\Pi}_2)^2 + \frac{x\tilde{x}}{r\tilde{r}} \Pi'_1\Pi'_2 + \dot{\Pi}_1'\dot{\Pi}_2') \right]. \quad (\text{B.22}) \end{aligned}$$

$$\begin{aligned}
\tilde{I}_{ij} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x_3'^2 - x_1'^2}{r^2} (\Pi_1'' - (\Pi_1'/r)) (\Pi_2'/\tilde{r}) \right. \\
& + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2} (\Pi_1'/r) (\Pi_2'' - (\Pi_2'/\tilde{r})) \\
& + \frac{x_3'(x_3' - R_s) - x_1'^2}{r\tilde{r}} \left(\frac{x\tilde{x}}{r\tilde{r}} (\Pi_1''\Pi_2'' - \Pi_1''(\Pi_2'/\tilde{r}) - (\Pi_1'/r)\Pi_2'' + (\Pi_1'/r)(\Pi_2'/\tilde{r})) \right. \\
& \quad \left. \left. + \dot{\Pi}_1'\dot{\Pi}_2' \right) \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x_3'^2 - x_1'^2}{r^2} (4(\Pi_1'' - (\Pi_1'/r))((\Pi_2')^2 + (\dot{\Pi}_2')^2) \right. \\
& \quad - 8(\Pi_1')^2(\Pi_2'/\tilde{r})) \\
& + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2} (4((\Pi_1')^2 + (\dot{\Pi}_1')^2)(\Pi_2'' - (\Pi_2'/\tilde{r})) - 8(\Pi_1'/r)(\Pi_2')^2) \\
& + \frac{x_3'(x_3' - R_s) - x_1'^2}{r\tilde{r}} \left(-8\frac{x\tilde{x}}{r\tilde{r}} ((\Pi_1'' - (\Pi_1'/r))(\Pi_2')^2 + (\Pi_1')^2(\Pi_2'' - (\Pi_2'/\tilde{r}))) \right. \\
& - 8\dot{\Pi}_1'\dot{\Pi}_2'\Pi_2' - 8\dot{\Pi}_1\Pi_1'\dot{\Pi}_2' - 8(\Pi_1''\Pi_1' + \dot{\Pi}_1'\dot{\Pi}_1)\Pi_2' \\
& \quad \left. - 8\Pi_1'(\Pi_2''\Pi_2' + \dot{\Pi}_2'\dot{\Pi}_2) \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^4} \left[-4\frac{x_3'^2 - x_1'^2}{r^2} (\Pi_1')^2((\Pi_2')^2 + (\dot{\Pi}_2')^2) \right. \\
& \quad - 4\frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2} ((\Pi_1')^2 + (\dot{\Pi}_1')^2)(\Pi_2')^2 \\
& + 12\frac{x_3'(x_3' - R_s) - x_1'^2}{r\tilde{r}} \Pi_1'\Pi_2'((\Pi_1')^2 + (\dot{\Pi}_1')^2 + (\Pi_2')^2 + (\dot{\Pi}_2')^2 \\
& \quad \left. + \frac{x\tilde{x}}{r\tilde{r}} \Pi_1'\Pi_2' + \dot{\Pi}_1'\dot{\Pi}_2) \right]. \quad (\text{B.23})
\end{aligned}$$

$$\begin{aligned}
\tilde{I}_{it} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x'_3}{r} \left(\frac{x\tilde{x}}{r\tilde{r}} (\Pi_1'' - (\Pi_1'/r)) \dot{\Pi}_2' + \dot{\Pi}_1' \ddot{\Pi}_2 \right) \right. \\
& \left. \frac{x'_3 - R_s}{\tilde{r}} (\Pi_1'/r) \dot{\Pi}_2' \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x'_3}{r} \left(4\dot{\Pi}_1' ((\Pi_2')^2 + (\dot{\Pi}_2)^2) \right. \right. \\
& - 8 \frac{x\tilde{x}}{r\tilde{r}} (\Pi_1'' - (\Pi_1'/r)) \dot{\Pi}_2 \Pi_2' - 8\dot{\Pi}_1' (\dot{\Pi}_2)^2 - 8\Pi_1' \left(\frac{x\tilde{x}}{r\tilde{r}} \Pi_1' \dot{\Pi}_2' + \dot{\Pi}_1 \ddot{\Pi}_2 \right) \\
& - 8(\Pi_1'' \Pi_1' + \dot{\Pi}_1' \dot{\Pi}_1) \dot{\Pi}_2 - 8\Pi_1' (\dot{\Pi}_2' \Pi_2' + \ddot{\Pi}_2 \dot{\Pi}_2) \Big) \\
& \left. + \frac{x'_3 - R_s}{\tilde{r}} (4((\Pi_1')^2 + (\dot{\Pi}_1)^2) \dot{\Pi}_2' - 8(\Pi_1'/r) \dot{\Pi}_2 \Pi_2') \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x'_3}{r} \left(-4\dot{\Pi}_1 \Pi_1' ((\Pi_2')^2 + (\dot{\Pi}_2)^2) \right. \right. \\
& + 12\Pi_1' \dot{\Pi}_1 ((\Pi_1')^2 + (\dot{\Pi}_1)^2 + (\Pi_2')^2 + (\dot{\Pi}_2)^2 + \frac{x\tilde{x}}{r\tilde{r}} \Pi_1' \Pi_2' + \dot{\Pi}_1 \dot{\Pi}_2) \Big) \\
& \left. - 4 \frac{x'_3 - R_s}{\tilde{r}} ((\Pi_1')^2 + (\dot{\Pi}_1)^2) \dot{\Pi}_2 \Pi_2' \right] . \quad (\text{B.24})
\end{aligned}$$

$$\begin{aligned}
\tilde{I}_{ti} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x'_3}{r} \dot{\Pi}'_1(\Pi'_2/\tilde{r}) \right. \\
& \left. \frac{x'_3 - R_s}{\tilde{r}} \left(\frac{x\tilde{x}}{r\tilde{r}} \dot{\Pi}'_1(\Pi''_2 - (\Pi'_2/\tilde{r})) + \ddot{\Pi}_1 \dot{\Pi}'_2 \right) \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x'_3}{r} (4\dot{\Pi}'_1((\Pi'_2)^2 + (\dot{\Pi}_2)^2) - 8\dot{\Pi}_1 \Pi'_1(\Pi'_2/\tilde{r})) \right. \\
& + \frac{x'_3 - R_s}{\tilde{r}} \left(4((\Pi'_1)^2 + (\dot{\Pi}_1)^2) \dot{\Pi}'_2 - 8 \left(\frac{x\tilde{x}}{r\tilde{r}} \dot{\Pi}'_1 \Pi'_1 + \ddot{\Pi}_1 \dot{\Pi}_1 \right) \Pi'_2 \right. \\
& - 8 \frac{x\tilde{x}}{r\tilde{r}} \dot{\Pi}_1 \Pi'_1(\Pi''_2 - (\Pi'_2/\tilde{r})) - 8(\dot{\Pi}_1)^2 \dot{\Pi}'_2 \\
& \left. \left. - 8(\dot{\Pi}'_1 \Pi'_1 + \ddot{\Pi}_1 \dot{\Pi}_1) \Pi'_2 - 8\dot{\Pi}_1(\Pi''_2 \Pi'_2 + \dot{\Pi}'_2 \dot{\Pi}_2) \right) \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^4} \left[-4 \frac{x'_3}{r} \dot{\Pi}_1 \Pi'_1((\Pi'_2)^2 + (\dot{\Pi}_2)^2) \right. \\
& \frac{x'_3 - R_s}{\tilde{r}} \left(-4((\Pi'_1)^2 + (\dot{\Pi}_1)^2) \dot{\Pi}_2 \Pi'_2 \right. \\
& \left. \left. + 12\dot{\Pi}_1 \Pi'_2((\Pi'_1)^2 + (\dot{\Pi}_1)^2 + (\Pi'_2)^2 + (\dot{\Pi}_2)^2 + \frac{x\tilde{x}}{r\tilde{r}} \Pi'_1 \Pi'_2 + \dot{\Pi}_1 \dot{\Pi}_2) \right) \right] . \quad (B.25)
\end{aligned}$$

$$\begin{aligned}
\tilde{I}_{tt} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x\tilde{x}}{r\tilde{r}} \dot{\Pi}'_1 \dot{\Pi}'_2 + \ddot{\Pi}_1 \ddot{\Pi}_2 \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^3} \left[4\ddot{\Pi}_1((\Pi'_2)^2 + (\dot{\Pi}_2)^2) + 4((\Pi'_1)^2 + (\dot{\Pi}_1)^2) \ddot{\Pi}_2 \right. \\
& - 8 \left(\frac{x\tilde{x}}{r\tilde{r}} \dot{\Pi}'_1 \dot{\Pi}_2 \Pi'_2 + \ddot{\Pi}_1 (\dot{\Pi}_2)^2 \right) - 8 \left(\frac{x\tilde{x}}{r\tilde{r}} \dot{\Pi}_1 \Pi'_1 \dot{\Pi}'_2 + (\dot{\Pi}_1)^2 \ddot{\Pi}_2 \right) \\
& \left. - 8(\dot{\Pi}'_1 \Pi'_1 \dot{\Pi}_2 + \ddot{\Pi}_1 \dot{\Pi}_1 \dot{\Pi}_2) - 8(\dot{\Pi}_1 \dot{\Pi}'_2 \Pi'_2 + \dot{\Pi}_1 \ddot{\Pi}_1 \dot{\Pi}_2) \right] \\
& + \frac{1}{(1 + \Pi_1 + \Pi_2)^4} \left[-4(\dot{\Pi}_1)^2((\Pi'_2)^2 + (\dot{\Pi}_2)^2) - 4((\Pi'_1)^2 + (\dot{\Pi}_1)^2)(\dot{\Pi}_2)^2 \right. \\
& \left. + 12\dot{\Pi}_1 \dot{\Pi}_2((\Pi'_1)^2 + (\dot{\Pi}_1)^2 + (\Pi'_2)^2 + (\dot{\Pi}_2)^2 + \frac{x\tilde{x}}{r\tilde{r}} \Pi'_1 \Pi'_2 + \dot{\Pi}_1 \dot{\Pi}_2) \right] . \quad (B.26)
\end{aligned}$$

$$\begin{aligned}
I_{\mu\rho\nu\sigma} = & \delta_{ij}\delta_{kh}\tilde{I}_{ijij} + \delta_{ij}\frac{R_kR_h}{R_s^2}\tilde{I}_{ikih} + \delta_{kh}(\delta_{\nu t}\frac{R_i}{R_s}\tilde{I}_{iktk} + \delta_{\mu t}\frac{R_j}{R_s}\tilde{I}_{tkjk}) \\
& + \delta_{\mu t}\delta_{\nu t}(\delta_{kh}\tilde{I}_{tktk} + \frac{R_kR_h}{R_s^2}\tilde{I}_{tkth}). \quad (\text{B.27})
\end{aligned}$$

$$\tilde{I}_{ijij} = 0 \quad (\text{analytically}). \quad (\text{B.28})$$

$$\begin{aligned}
\tilde{I}_{ikih} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x_3'^2 - x_1'^2}{r^2}(\Pi_1'' - (\Pi_1'/r))(\Pi_2'/\tilde{r}) \right. \\
& + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2}(\Pi_1'/r)(\Pi_2'' - (\Pi_2'/\tilde{r})) \\
& \left. \frac{x_1'^2 R_s^2}{(r\tilde{r})^2}(\Pi_1'' - (\Pi_1'/r))(\Pi_2'' - (\Pi_2'/\tilde{r})) \right] \\
- & \frac{8}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x_3'^2 - x_1'^2}{r^2}(\Pi_1')^2(\Pi_2'/\tilde{r}) + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2}(\Pi_1'/r)(\Pi_2')^2 \right. \\
& \left. + \frac{x_1'^2 R_s^2}{(r\tilde{r})^2}((\Pi_1'' - (\Pi_1'/r))(\Pi_2')^2 + (\Pi_1')^2(\Pi_2'' - (\Pi_2'/\tilde{r}))) \right] \\
+ & \frac{8}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_1'^2 R_s^2}{(r\tilde{r})^2}(\Pi_1')^2(\Pi_2')^2 \right]. \quad (\text{B.29})
\end{aligned}$$

$$\begin{aligned}
\tilde{I}_{iktk} = \tilde{I}_{tkjk} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x'_3}{r} \dot{\Pi}'_1(\Pi'_2/\tilde{r}) + \frac{x'_3 - R_s}{\tilde{r}} (\Pi'_1/r) \dot{\Pi}'_2 \right. \\
& \left. + \frac{x_1'^2 R_s}{r \tilde{r}^2} \dot{\Pi}'_1(\Pi''_2 - (\Pi'_2/\tilde{r})) - \frac{x_1'^2 R_s}{r^2 \tilde{r}} (\Pi''_1 - (\Pi'_1/\tilde{r})) \dot{\Pi}'_2 \right] \\
& - \frac{8}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x'_3}{r} \dot{\Pi}_1 \Pi'_1(\Pi'_2/\tilde{r}) + \frac{x'_3 - R_s}{\tilde{r}} (\Pi'_1/r) \dot{\Pi}_2 \Pi'_2 \right. \\
& + \frac{x_1'^2 R_s^2}{r \tilde{r}^2} (\dot{\Pi}'_1(\Pi'_2)^2 + \dot{\Pi}_1 \Pi'_1(\Pi''_2 - (\Pi'_2/\tilde{r}))) \\
& \left. - \frac{x_1'^2 R_s^2}{r^2 \tilde{r}} ((\Pi'_1)^2 \dot{\Pi}'_2 + (\Pi''_1 - (\Pi'_1/\tilde{r})) \dot{\Pi}_2 \Pi'_2) \right] \\
& + \frac{8}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_1'^2 R_s}{r \tilde{r}^2} \dot{\Pi}_1 \Pi'_1(\Pi'_2)^2 - \frac{x_1'^2 R_s}{r^2 \tilde{r}} (\Pi'_1)^2 \dot{\Pi}_2 \Pi'_2 \right]. \tag{B.30}
\end{aligned}$$

$$\begin{aligned}
\tilde{I}_{tktk} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\ddot{\Pi}_1(\Pi'_2/\tilde{r}) + (\Pi'_1/r) \ddot{\Pi}_2 + \frac{x_1'^2}{r^2} (\Pi''_1 - (\Pi'_1/r)) \ddot{\Pi}_2 \right. \\
& \left. + \frac{x_1'^2}{\tilde{r}^2} \ddot{\Pi}_1(\Pi''_2 - (\Pi'_2/\tilde{r})) - 2 \frac{x_1'^2}{r \tilde{r}} \dot{\Pi}'_1 \dot{\Pi}'_2 \right] \\
& - \frac{8}{(1 + \Pi_1 + \Pi_2)^3} \left[(\Pi'_1/r) (\dot{\Pi}_2)^2 + (\dot{\Pi}_1)^2 (\Pi'_2/\tilde{r}) \right. \\
& + \frac{x_1'^2}{r^2} ((\Pi''_1 - (\Pi'_1/r)) (\dot{\Pi}_2)^2 + (\Pi'_1)^2 \ddot{\Pi}_2) \\
& + \frac{x_1'^2}{\tilde{r}^2} ((\dot{\Pi}_1)^2 (\Pi''_2 - (\Pi'_2/\tilde{r})) + \ddot{\Pi}_1 (\Pi'_2)^2) \\
& \left. - 2 \frac{x_1'^2}{r \tilde{r}} (\dot{\Pi}'_1 \dot{\Pi}_2 \Pi'_2 + \dot{\Pi}_1 \Pi'_1 \dot{\Pi}'_2) \right] \\
& + \frac{8}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_1'^2}{r^2} (\Pi'_1)^2 (\dot{\Pi}_2)^2 + \frac{x_1'^2}{\tilde{r}^2} (\dot{\Pi}_1)^2 (\Pi'_2)^2 - 2 \frac{x_1'^2}{r \tilde{r}} \dot{\Pi}_1 \Pi'_1 \dot{\Pi}_2 \Pi'_2 \right]. \tag{B.31}
\end{aligned}$$

$$\begin{aligned}
\tilde{I}_{tkth} = & \frac{4}{(1 + \Pi_1 + \Pi_2)^2} \left[\frac{x_3'^2 - x_1'^2}{r^2} (\Pi_1'' - (\Pi_1'/r)) \ddot{\Pi}_2 \right. \\
& + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2} \ddot{\Pi}_1 (\Pi_2'' - (\Pi_2'/\tilde{r})) - 2 \frac{x_3'(x_3' - R_s) - x_1'^2}{r\tilde{r}} \dot{\Pi}_1' \dot{\Pi}_2' \left. \right] \\
& - \frac{8}{(1 + \Pi_1 + \Pi_2)^3} \left[\frac{x_3'^2 - x_1'^2}{r^2} ((\Pi_1'' - (\Pi_1'/r)) (\dot{\Pi}_2)^2 + (\Pi_1')^2 \ddot{\Pi}_2) \right. \\
& + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2} ((\dot{\Pi}_1)^2 (\Pi_2'' - (\Pi_2'/\tilde{r})) + \ddot{\Pi}_1 (\Pi_2')^2) \\
& \left. - 2 \frac{x_3'(x_3' - R_s) - x_1'^2}{r\tilde{r}} (\dot{\Pi}_1' \dot{\Pi}_2 \Pi_2' + \dot{\Pi}_1 \Pi_1' \dot{\Pi}_2') \right] \\
& + \frac{8}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_3'^2 - x_1'^2}{r^2} (\Pi_1')^2 (\dot{\Pi}_2)^2 + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2} (\dot{\Pi}_1)^2 (\Pi_2')^2 \right. \\
& \left. - 2 \frac{x_3'(x_3' - R_s) - x_1'^2}{r\tilde{r}} \dot{\Pi}_1 \Pi_1' \dot{\Pi}_2 \Pi_2' \right]. \quad (\text{B.32})
\end{aligned}$$

$$\begin{aligned}
J_{\mu\rho\nu\sigma} = & \delta_{ij} \frac{R_k R_h}{R_s^2} \tilde{J}_{ikih} + \delta_{kh} (\delta_{\nu t} \frac{R_i}{R_s} \tilde{J}_{iktk} + \delta_{\mu t} \frac{R_j}{R_s} \tilde{J}_{tkjk}) \\
& + \delta_{\mu t} \delta_{\nu t} (\delta_{kh} \tilde{J}_{tktk} + \frac{R_k R_h}{R_s^2} \tilde{J}_{tkth}). \quad (\text{B.33})
\end{aligned}$$

$$\tilde{J}_{ikih} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_1'^2 R_s^2}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \quad (\text{B.34})$$

$$\begin{aligned}
\tilde{J}_{iktk} = \tilde{J}_{tkik} = & \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_1'^2 R_s^2}{r\tilde{r}^2} \dot{\Pi}_1 \Pi_1' (\Pi_2')^2 \right. \\
& \left. - \frac{x_1'^2 R_s^2}{r^2 \tilde{r}} (\Pi_1')^2 \dot{\Pi}_2 \Pi_2' \right]. \quad (\text{B.35})
\end{aligned}$$

$$\tilde{J}_{tktk} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_1'^2}{r^2} (\Pi_1')^2 (\dot{\Pi}_2)^2 + \frac{x_1'^2}{\tilde{r}^2} (\dot{\Pi}_1)^2 (\Pi_2')^2 - 2 \frac{x_1'^2}{r\tilde{r}} \dot{\Pi}_1 \Pi_1' \dot{\Pi}_2 \Pi_2' \right]. \quad (\text{B.36})$$

$$\tilde{J}_{tkth} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \left[\frac{x_3'^2 - x_1'^2}{r^2} (\Pi_1')^2 (\dot{\Pi}_2)^2 + \frac{(x_3' - R_s)^2 - x_1'^2}{\tilde{r}^2} (\dot{\Pi}_1)^2 (\Pi_2')^2 - 2 \frac{x_3'(x_3' - R_s) - x_1'^2}{r\tilde{r}} \dot{\Pi}_1 \Pi_1' \dot{\Pi}_2 \Pi_2' \right]. \quad (\text{B.37})$$

$$\begin{aligned} K_{\mu\rho\nu\sigma} = & (\delta_{ij}\delta_{kh} + \delta_{ik}\delta_{jh} + \delta_{ih}\delta_{kj}) \tilde{K}_{ikik} + \delta_{ij} \frac{R_k R_h}{R_s^2} \tilde{K}_{ikih} + \delta_{kh} \frac{R_i R_j}{R_s^2} \tilde{K}_{ikjk} \\ & + (\delta_{ik} \frac{R_j R_h}{R_s^2} + \delta_{ih} \frac{R_j R_k}{R_s^2} + \delta_{jk} \frac{R_i R_h}{R_s^2} + \delta_{jh} \frac{R_i R_k}{R_s^2}) \tilde{K}_{iijh} \\ & + \frac{R_i R_j R_k R_h}{R_s^4} \tilde{K}_{ikjh} \\ & + \delta_{\mu t} \delta_{kh} \frac{R_j}{R_s} \tilde{K}_{tkjk} + \delta_{\mu t} \delta_{jh} \frac{R_k}{R_s} \tilde{K}_{tkjj} + \delta_{\mu t} \delta_{jk} \frac{R_h}{R_s} \tilde{K}_{tjjh} + \delta_{\mu t} \frac{R_j R_k R_h}{R_s^3} \tilde{K}_{tkjh} \\ & + \delta_{\rho t} \delta_{jh} \frac{R_i}{R_s} \tilde{K}_{itjj} + \delta_{\rho t} \delta_{ih} \frac{R_j}{R_s} \tilde{K}_{itji} + \delta_{\rho t} \delta_{ij} \frac{R_h}{R_s} \tilde{K}_{itih} + \delta_{\rho t} \frac{R_i R_j R_h}{R_s^3} \tilde{K}_{itjh} \\ & + \delta_{\mu t} \delta_{\rho t} \delta_{jh} \tilde{K}_{ttjj} + \delta_{\mu t} \delta_{\rho t} \frac{R_j R_h}{R_s^2} \tilde{K}_{ttjh} + \delta_{\mu t} \delta_{\sigma t} \delta_{jk} \tilde{K}_{tjjt} + \delta_{\mu t} \delta_{\sigma t} \frac{R_j R_k}{R_s^2} \tilde{K}_{tkjt} \\ & + \delta_{\mu t} \delta_{\nu t} \delta_{kh} \tilde{K}_{tktk} + \delta_{\mu t} \delta_{\nu t} \frac{R_k R_h}{R_s^2} \tilde{K}_{tkth} + \delta_{\rho t} \delta_{\sigma t} \delta_{ij} \tilde{K}_{itjt} + \delta_{\rho t} \delta_{\sigma t} \frac{R_i R_j}{R_s^2} \tilde{K}_{itjt} \\ & + \delta_{\mu t} \delta_{\nu t} \delta_{\sigma t} \frac{R_k}{R_s} \tilde{K}_{tktt} + \delta_{\rho t} \delta_{\nu t} \delta_{\sigma t} \frac{R_i}{R_s} \tilde{K}_{ittt} + \delta_{\mu t} \delta_{\rho t} \delta_{\nu t} \delta_{\sigma t} \tilde{K}_{tttt}. \quad (\text{B.38}) \end{aligned}$$

$$\tilde{K}_{ikik} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_1'^2 x_2'^2}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \quad (\text{B.39})$$

$$\tilde{K}_{ikih} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{(x'_3 - R_s)^2 x_1'^2 - x_1'^2 x_2'^2}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \quad (\text{B.40})$$

$$\tilde{K}_{ikjk} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_3'^2 x_1'^2 - x_1'^2 x_2'^2}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \quad (\text{B.41})$$

$$\tilde{K}_{ijjh} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x'_3(x'_3 - R_s)x_1'^2 - x_1'^2 x_2'^2}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \quad (\text{B.42})$$

$$\begin{aligned} \tilde{K}_{ikjh} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} & \left(x_3'^2 (x'_3 - R_s)^2 + 3x_1'^2 x_2'^2 - x_3'^2 x_1'^2 \right. \\ & \left. - (x'_3 - R_s)^2 x_1'^2 - 4x'_3(x'_3 - R_s)x_1'^2 \right) \frac{1}{(r\tilde{r})^2} (\Pi_1')^2 (\Pi_2')^2. \end{aligned} \quad (\text{B.43})$$

$$\tilde{K}_{tkjk} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x'_3 x_1'^2}{r\tilde{r}^2} \dot{\Pi}_1 \Pi_1' (\Pi_2')^2. \quad (\text{B.44})$$

$$\tilde{K}_{tkjj} = \tilde{K}_{tjjh} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{(x'_3 - R_s)x_1'^2}{r\tilde{r}^2} \dot{\Pi}_1 \Pi_1' (\Pi_2')^2. \quad (\text{B.45})$$

$$\begin{aligned} \tilde{K}_{tkjh} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} & \left(x'_3(x'_3 - R_s)^2 - x_3'^2 x_1'^2 \right. \\ & \left. - 2(x'_3 - R_s)x_1'^2 \right) \frac{1}{r\tilde{r}^2} \dot{\Pi}_1 \Pi_1' (\Pi_2')^2. \end{aligned} \quad (\text{B.46})$$

$$\tilde{K}_{itjj} = \tilde{K}_{itji} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x'_3 x_1'^2}{r^2 \tilde{r}} (\Pi_1')^2 \dot{\Pi}_2 \Pi_2'. \quad (\text{B.47})$$

$$\tilde{K}_{itih} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{(x'_3 - R_s)x_1'^2}{r^2 \tilde{r}} (\Pi'_1)^2 \dot{\Pi}_2 \Pi'_2. \quad (\text{B.48})$$

$$\begin{aligned} \tilde{K}_{itjh} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} & \left(x'_3(x'_3 - R_s)^2 - x'_3 x_1'^2 \right. \\ & \left. - 2(x'_3 - R_s)x_1'^2 \right) \frac{1}{r^2 \tilde{r}} (\Pi'_1)^2 \dot{\Pi}_2 \Pi'_2. \end{aligned} \quad (\text{B.49})$$

$$\tilde{K}_{ttjj} = \tilde{K}_{tjjet} = \frac{8}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_1'^2}{r \tilde{r}} \dot{\Pi}_1 \Pi'_1 \dot{\Pi}_2 \Pi'_2. \quad (\text{B.50})$$

$$\tilde{K}_{ttjh} = \tilde{K}_{tkjt} = \frac{8}{(1 + \Pi_1 + \Pi_2)^4} \frac{x'_3(x'_3 - R_s) - x_1'^2}{r \tilde{r}} \dot{\Pi}_1 \Pi'_1 \dot{\Pi}_2 \Pi'_2. \quad (\text{B.51})$$

$$\tilde{K}_{tktk} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_1'^2}{\tilde{r}^2} (\dot{\Pi}_1)^2 (\Pi'_2)^2. \quad (\text{B.52})$$

$$\tilde{K}_{tkth} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{(x'_3 - R_s)^2 - x_1'^2}{\tilde{r}^2} (\dot{\Pi}_1)^2 (\Pi'_2)^2. \quad (\text{B.53})$$

$$\tilde{K}_{itit} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_1'^2}{r^2} (\Pi'_1)^2 (\dot{\Pi}_2)^2. \quad (\text{B.54})$$

$$\tilde{K}_{itjt} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} \frac{x_3'^2 - x_1'^2}{r^2} (\Pi'_1)^2 (\dot{\Pi}_2)^2. \quad (\text{B.55})$$

$$\tilde{K}_{tktt} = \frac{4}{(1 + \Pi_1 + \Pi_2)^4} \frac{x'_3 - R_s}{\tilde{r}} (\dot{\Pi}_1)^2 \dot{\Pi}_2 \Pi'_2. \quad (\text{B.56})$$

$$\tilde{K}_{ittt} = \frac{4}{(1 + \Pi_1 + \Pi_2)^4} \frac{x'_3}{r} \dot{\Pi}_1 \Pi'_1 (\dot{\Pi}_2)^2. \quad (\text{B.57})$$

$$\tilde{K}_{tttt} = \frac{2}{(1 + \Pi_1 + \Pi_2)^4} (\dot{\Pi}_1)^2 (\dot{\Pi}_2)^2. \quad (\text{B.58})$$

B.1.2 Asymptotic Interactions

As explained in the main text, the small separation asymptotic formulas get contributions which have the same functional form as those for the large separation asymptotics but with different explicit integration limits, see Fig. 5.9. We will therefore start with the large separation formulas and not evaluate the integrals explicitly.

Large Separation

As explained in Fig. 5.7, the 't Hooft potential is approximated by the $T = 0$ BPST form for small arguments up to $r_{4d} = \sqrt{x_\mu x_\mu} = \beta/2$; for larger arguments we use the $r_{3d} \rightarrow \infty$ asymptotic form of the $T \neq 0$ potential. These simple rational expressions allow us to factor out completely the 't Hooft potential of the partner instanton, whose argument is kept fixed, except for its contribution to the integration limit. For fixed I_2 , and depending on the integration region, this leads to the two integration limits

$$z_{4d}^2 = \frac{1 + \Pi_2}{\rho_1^2} r_{4d}^2, \quad (\text{B.59})$$

$$z_{3d} = \frac{1 + \Pi_2}{\tilde{\rho}_1} r_{3d}. \quad (\text{B.60})$$

We have defined $\tilde{\rho} = \pi\rho^2/\beta$. In the following we will encounter the integrals $\int^z \equiv \int_0^z$ and $\int_z \equiv \int_z^\infty$.

Finally, for $r_{3d} > \beta/2$ the approximate 't Hooft potential is independent of t . This will lead to some indices only running over the spatial set $i = \{1, 2, 3\}$ whenever we use roman letters.

The integral over I contains terms that do not mix the 't Hooft potential Π_1 and Π_2 except for the denominators. Just as in the $T = 0$ case, these can be transformed to exactly match the single instanton contributions by exploiting scale invariance and independence of the action on β . Remembering that we actually subtract the one-instanton contributions to get the interactions, we can neglect these terms altogether. We then end up with the following formulas

$$\begin{aligned} \int I = & 72\pi^2 \rho^2 \frac{\partial_\mu \Pi \partial_\mu \Pi}{(1 + \Pi)^3} \int^{z_{4d}} \frac{s^5 ds}{(s^2 + 1)^4} \\ & + \left(32\pi\beta\tilde{\rho} \frac{\partial_\mu \Pi \partial_\mu \Pi}{(1 + \Pi)^3} + \frac{16}{3}\pi\beta\tilde{\rho} \frac{\partial_i \Pi \partial_i \Pi}{(1 + \Pi)^3} \right) \int_{z_{3d}} \frac{s^2 ds}{(s + 1)^4} + \text{sym.} \quad (\text{B.61}) \end{aligned}$$

$$\begin{aligned} \int J = & 16\pi^2 \rho^2 \frac{\partial_\mu \Pi \partial_\mu \Pi}{(1 + \Pi)^3} \int^{z_{4d}} \frac{s^5 ds}{(s^2 + 1)^4} \\ & + 8\pi\beta\tilde{\rho} \frac{\partial_\mu \Pi \partial_\mu \Pi}{(1 + \Pi)^3} \int_{z_{3d}} \frac{s^2 ds}{(s + 1)^4} + \text{sym.} \quad (\text{B.62}) \end{aligned}$$

$$\begin{aligned} \int I_{\mu\nu} = & 16\pi^2 \rho^2 \frac{\partial_\mu \partial_\nu \Pi}{(1 + \Pi)^2} \int^{z_{4d}} \frac{s^3 ds}{(s^2 + 1)^3} \\ & - \left(8\pi^2 \rho^2 \delta_{\mu\nu} \frac{\partial_\sigma \Pi \partial_\sigma \Pi}{(1 + \Pi)^3} + 8\pi^2 \rho^2 \frac{(\partial_\mu \Pi)(\partial_\nu \Pi)}{(1 + \Pi)^3} \right) \int^{z_{4d}} \frac{s^5 ds}{(s^2 + 1)^4} \\ & + \left(16\pi\beta\tilde{\rho} \frac{\partial_\mu \partial_\nu \Pi}{(1 + \Pi)^2} - \frac{32}{3}\pi\beta\tilde{\rho} \frac{\partial_i \partial_\nu \Pi}{(1 + \Pi)^2} \right) \int_{z_{3d}} \frac{s ds}{(s + 1)^3} \\ & - \left(\frac{16}{3}\pi\beta\tilde{\rho} \delta_{ij} \frac{\partial_\sigma \Pi \partial_\sigma \Pi}{(1 + \Pi)^3} + 16\pi\beta\tilde{\rho} \frac{(\partial_\mu \Pi)(\partial_\nu \Pi)}{(1 + \Pi)^3} \right. \\ & \quad \left. - 16\pi\beta\tilde{\rho} \frac{(\partial_i \Pi)(\partial_\nu \Pi)}{(1 + \Pi)^3} \right) \int_{z_{3d}} \frac{s^2 ds}{(s + 1)^4} + \text{sym.} \quad (\text{B.63}) \end{aligned}$$

At zeroth order, partial integration and the antisymmetry of the 't Hooft

symbols can be used to simplify

$$I_{\mu\rho\nu\sigma} \rightarrow \frac{8}{(1 + \Pi_1 + \Pi_2)^4} (\partial_\mu \Pi_1) (\partial_\rho \Pi_2) (\partial_\nu \Pi_1) (\partial_\sigma \Pi_2). \quad (\text{B.64})$$

With asymptotic behaviour

$$\begin{aligned} \int I_{\mu\rho\nu\sigma} &= 16\pi^2 \rho^2 \delta_{\mu\nu} \frac{(\partial_\rho \Pi)(\partial_\sigma \Pi)}{(1 + \Pi)^3} \int^{z_{4d}} \frac{s^5 ds}{(s^2 + 1)^4} \\ &\quad + \frac{32}{3} \pi \beta \tilde{\rho} \delta_{ij} \frac{(\partial_\rho \Pi)(\partial_\sigma \Pi)}{(1 + \Pi)^3} \int_{z_{3d}} \frac{s^2 ds}{(s + 1)^4} + \text{sym}. \end{aligned} \quad (\text{B.65})$$

$$\begin{aligned} \int J_{\mu\rho\nu\sigma} &= 4\pi^2 \rho^2 \delta_{\mu\nu} \frac{(\partial_\rho \Pi)(\partial_\sigma \Pi)}{(1 + \Pi)^3} \int^{z_{4d}} \frac{s^5 ds}{(s^2 + 1)^4} \\ &\quad + \frac{8}{3} \pi \beta \tilde{\rho} \delta_{ij} \frac{(\partial_\rho \Pi)(\partial_\sigma \Pi)}{(1 + \Pi)^3} \int_{z_{3d}} \frac{s^2 ds}{(s + 1)^4} + \text{sym}. \end{aligned} \quad (\text{B.66})$$

For $K_{\mu\rho\nu\sigma}$ no 't Hooft symbols can be used to exchange the index pairs $(\mu, \nu) \leftrightarrow (\rho, \sigma)$, and so we cannot simplify with a symmetry argument anymore

$$\begin{aligned} \int K_{\mu\rho\nu\sigma} &= 4\pi^2 \rho_1^2 \delta_{\mu\nu} \frac{(\partial_\rho \Pi_2)(\partial_\sigma \Pi_2)}{(1 + \Pi_2)^3} \int^{z_1^{4d}} \frac{s^5 ds}{(s^2 + 1)^4} \\ &\quad + 4\pi^2 \rho_2^2 \delta_{\rho\sigma} \frac{(\partial_\mu \Pi_1)(\partial_\nu \Pi_1)}{(1 + \Pi_1)^3} \int^{z_2^{4d}} \frac{s^5 ds}{(s^2 + 1)^4} \\ &\quad + \frac{8}{3} \pi \beta \tilde{\rho}_1 \delta_{ij} \frac{(\partial_\rho \Pi_2)(\partial_\sigma \Pi_2)}{(1 + \Pi_2)^3} \int_{z_1^{3d}} \frac{s^2 ds}{(s + 1)^4} \\ &\quad + \frac{8}{3} \pi \beta \tilde{\rho}_2 \delta_{kh} \frac{(\partial_\mu \Pi_1)(\partial_\nu \Pi_1)}{(1 + \Pi_1)^3} \int_{z_2^{3d}} \frac{s^2 ds}{(s + 1)^4}. \end{aligned} \quad (\text{B.67})$$

Small Separation

As explained in Fig. 5.9, the small separation asymptotic formulas get contributions from the large asymptotics. Also, in this case we have performed a global translation so that the instantons sit at $\pm R_\mu/2$. For the terms in \int^z the integration limit is given by $z^2 = \frac{1+\Pi}{\rho^2} (R/2)^2$. The $T \neq 0$ terms given

by the \int_z integrals are to be interpreted as $\theta_H(R - \beta/2) \int_z$, i.e. they only contribute if the separation R is bigger than $\beta/2$. In practice, these terms do not contribute because they are covered by the look-up tables. The proper small separation asymptotic formulas, that encode the repulsion through the gauge singularity, are then given by the $T = 0$ formulas, which we repeat here for convenience.

Introducing another explicit upper limit z

$$z^2 = \frac{R^2}{\rho_1^2 + \rho_2^2}, \quad z_i^2 = \frac{R^2}{\rho_i^2}, \quad (\text{B.68})$$

and approximating the arguments $x_\mu \pm R_\mu/2 \rightarrow x_\mu$, we arrive at

$$\begin{aligned} \int I = 384\pi^2 \Bigg[& \frac{\rho_1^4 + \rho_2^4}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^2} \\ & - \left(\frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} + 2 \frac{\rho_1^6 + \rho_2^6}{(\rho_1^2 + \rho_2^2)^3} \right) \int_z \frac{ds}{s(s^2 + 1)^3} \\ & \frac{\rho_1^8 + \rho_2^8 + \rho_1^4 \rho_2^4 + \rho_1^6 \rho_2^2 + \rho_1^2 \rho_2^6}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4} \\ & - \int_{z_1} \frac{s^4 ds}{s(s^2 + 1)^4} - \int_{z_2} \frac{s^4 ds}{s(s^2 + 1)^4} \Bigg]. \quad (\text{B.69}) \end{aligned}$$

$$\int J = 64\pi^2 \frac{\rho_1^4 \rho_2^4}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4}. \quad (\text{B.70})$$

$$\begin{aligned} \int I_{\mu\nu} = \delta_{\mu\nu} \Bigg[& 96\pi^2 \frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^2} \\ & - 192\pi^2 \frac{\rho_1^2 \rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^3} \\ & + 32\pi^2 \frac{\rho_1^4 \rho_2^4 + 3\rho_1^6 \rho_2^2 + 3\rho_1^2 \rho_2^6}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4} \Bigg]. \quad (\text{B.71}) \end{aligned}$$

$$\int I_{\mu\rho\nu\sigma} = \delta_{\mu\nu}\delta_{\rho\sigma} \left[-32\pi^2 \frac{\rho_1^2\rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^2} + 32\pi^2 \frac{\rho_1^2\rho_2^2}{(\rho_1^2 + \rho_2^2)^2} \int_z \frac{ds}{s(s^2 + 1)^3} \right]. \quad (\text{B.72})$$

$$\int J_{\mu\rho\nu\sigma} = 0. \quad (\text{B.73})$$

$$\int K_{\mu\rho\nu\sigma} = \frac{8}{3}\pi^2(\delta_{\mu\nu}\delta_{\rho\sigma} + \delta_{\mu\rho}\delta_{\nu\sigma} + \delta_{\mu\sigma}\delta_{\nu\rho}) \frac{\rho_1^4\rho_2^4}{(\rho_1^2 + \rho_2^2)^4} \int_z \frac{ds}{s(s^2 + 1)^4}. \quad (\text{B.74})$$

B.2 Fermionic Interactions

The Dirac overlap matrix elements are given by

$$T_{IA} = \int_{\mathbb{R}^3 \times S^1} \frac{1}{4\pi^2 \rho_I \rho_A} \left(\frac{1}{2} \text{Tr} (U \tau_\beta^+) I_\beta - \frac{i}{2} \text{Tr} (U \tau_\beta^+ \tau_a) \bar{\eta}_{\mu\alpha}^a J_{\beta\mu\alpha} + \frac{i}{2} \text{Tr} (U \tau_a \tau_\beta^+) \eta_{\mu\alpha}^a K_{\beta\mu\alpha} \right). \quad (\text{B.75})$$

Some straightforward algebra leads to

$$I_\beta = \frac{-3}{(1 + \Pi_I + \Pi_A)(1 + \Pi_I)^{1/2}(1 + \Pi_A)^{1/2}} \left[\frac{\Pi_A}{1 + \Pi_I} \left((\partial_\mu \Pi_A \partial_\mu \chi_A) - \frac{(\partial_\mu \Pi_I \partial_\mu \Pi_I) \chi_A}{1 + \Pi_I} \right) \left(\partial_\beta \chi_A - \frac{(\partial_\beta \Pi_A) \chi_A}{1 + \Pi_A} \right) + \left((\partial_\mu \Pi_A \partial_\mu \chi_A) - \frac{(\partial_\mu \Pi_A \partial_\mu \Pi_A) \chi_I}{1 + \Pi_A} \right) \left(\partial_\beta \chi_I - \frac{(\partial_\beta \Pi_I) \chi_I}{1 + \Pi_I} \right) \right]. \quad (\text{B.76})$$

$$J_{\beta\mu\alpha} = \frac{1}{(1 + \Pi_I + \Pi_A)(1 + \Pi_I)^{1/2}(1 + \Pi_A)^{1/2}} \frac{\Pi_A}{1 + \Pi_I} (\partial_\mu \chi_I \partial_\alpha \Pi_I) \left(\partial_\beta \chi_A - \frac{(\partial_\beta \Pi_A) \chi_A}{1 + \Pi_A} \right). \quad (\text{B.77})$$

$$K_{\beta\mu\alpha} = \frac{1}{(1 + \Pi_I + \Pi_A)(1 + \Pi_I)^{1/2}(1 + \Pi_A)^{1/2}} (\partial_\mu \chi_A \partial_\alpha \Pi_A) \left(\partial_\beta \chi_I - \frac{(\partial_\beta \Pi_I) \chi_I}{1 + \Pi_I} \right). \quad (\text{B.78})$$

B.2.1 Exact Interactions

The three contributions give rise to two different colour structures which can be combined with the help of the colour four-vector $iu_\beta \equiv \frac{1}{2} \text{Tr}(U\tau_\beta^+)$, used for instance in [157]; we have that

$$T_{IA} = i \int_{\mathbb{R}^3 \times S^1} \left(u_b \frac{R_b}{R_s} \tilde{T}_s + u_4 \tilde{T}_t \right), \quad (\text{B.79})$$

with

$$\begin{aligned} \tilde{T}_s = & \frac{-1}{(1 + \Pi_I + \Pi_A)(1 + \Pi_I)^{1/2}(1 + \Pi_A)^{1/2}} \\ & \left[\frac{x'_3}{r} \left\{ \frac{\Pi_A}{1 + \Pi_I} \left(\dot{\chi}_A - \frac{\dot{\Pi}_A \chi_A}{1 + \Pi_A} \right) (\dot{\chi}_I \Pi'_I - \chi'_I \dot{\Pi}_I) \right. \right. \\ & + 3 \left(\chi'_I - \frac{\Pi'_I \chi_I}{1 + \Pi_I} \right) \left(\Pi'_A \chi'_A + \dot{\Pi}_A \dot{\chi}_A - \frac{((\Pi'_A)^2 + (\dot{\Pi}_A)^2) \chi_A}{1 + \Pi_A} \right) \Big\} \\ & + \frac{x'_3 - R_s}{\tilde{r}} \left\{ \left(\dot{\chi}_I - \frac{\dot{\Pi}_I \chi_I}{1 + \Pi_I} \right) (\dot{\chi}_A \Pi'_A - \chi'_A \dot{\Pi}_A) \right. \\ & \left. \left. + 3 \frac{\Pi_A}{1 + \Pi_I} \left(\chi'_A - \frac{\Pi'_A \chi_A}{1 + \Pi_A} \right) \left(\Pi'_I \chi'_I + \dot{\Pi}_I \dot{\chi}_I - \frac{((\Pi'_I)^2 + (\dot{\Pi}_I)^2) \chi_I}{1 + \Pi_I} \right) \right\} \right]. \quad (\text{B.80}) \end{aligned}$$

$$\begin{aligned}
\tilde{T}_t = & \frac{1}{(1 + \Pi_I + \Pi_A)(1 + \Pi_I)^{1/2}(1 + \Pi_A)^{1/2}} \\
& \left[\frac{x'_3(x'_3 - r_s) + 2x_1'^2}{r\tilde{r}} \left\{ \frac{\Pi_A}{1 + \Pi_I} \left(\chi'_A - \frac{\Pi'_A \chi_A}{1 + \Pi_A} \right) (\dot{\chi}_I \Pi'_I - \chi'_I \dot{\Pi}_I) \right. \right. \\
& \quad \left. \left. + \left(\chi'_I - \frac{\Pi'_I \chi_I}{1 + \Pi_I} \right) (\dot{\chi}_A \Pi'_A - \chi'_A \dot{\Pi}_A) \right\} \right. \\
& - 3 \frac{\Pi_A}{1 + \Pi_I} \left(\dot{\chi}_A - \frac{\dot{\Pi}_A \chi_A}{1 + \Pi_A} \right) \left(\Pi'_I \chi'_I + \dot{\Pi}_I \dot{\chi}_I - \frac{((\Pi'_I)^2 + (\dot{\Pi}_I)^2) \chi_I}{1 + \Pi_I} \right) \\
& \left. - 3 \left(\dot{\chi}_I - \frac{\dot{\Pi}_I \chi_I}{1 + \Pi_I} \right) \left(\Pi'_A \chi'_A + \dot{\Pi}_A \dot{\chi}_A - \frac{((\Pi'_A)^2 + (\dot{\Pi}_A)^2) \chi_A}{1 + \Pi_A} \right) \right]. \quad (\text{B.81})
\end{aligned}$$

B.2.2 Asymptotic Interactions

Large Separation

At $T \neq 0$ the quark zero mode has an additional factor $\chi \propto \exp(-r/\beta)$. Therefore we completely neglect the contributions from $r > \beta/2$, and recover the $T = 0$ formulas. For convenience we display the final results again here; details are given in section 3.3.1 and appendix A.2. The only difference with the $T = 0$ formulas is that the upper limit $z \propto \beta/2$ instead of infinity

$$\begin{aligned}
\int I_\beta = & 8\pi^2 \rho_I^2 \frac{\Pi_A \partial_\beta \Pi_A}{(1 + \Pi_A)^{3/2}} \int^{z_I} \frac{s^4 ds}{(s^2 + 1)^{7/2}} \\
& - 8\pi^2 \rho_A^2 \frac{\partial_\beta \Pi_I}{(1 + \Pi_I)^{3/2}} \int^{z_A} \frac{s^4 ds}{(s^2 + 1)^{5/2}}. \quad (\text{B.82})
\end{aligned}$$

At zeroth order, we have that $\int J_{\beta\mu\alpha} = \int K_{\beta\mu\alpha} = 0$. For $J_{\beta\mu\alpha}$ this follows from the fact that within the $r < \beta/2$ region we approximate $\chi \rightarrow \Pi$ and so the integration over I_I yields zero because of the anti-symmetry of the 't Hooft symbols. Integration over I_A vanishes too because of $O(4)$ symmetry.

Small Separation

At zeroth order, i.e. $x_\mu \pm R_\mu/2 \rightarrow x_\mu$, the contribution to I_β vanishes because of $O(4)$ symmetry. It turns out the large separation asymptotics falls off too

quickly as $R \rightarrow 0$. However this is not important because in this regime the gluonic interaction is dominant.

Appendix C

Cholesky decomposition update

In this appendix we look in detail at how the structure suitable for the Cholesky decomposition update comes about. We will also see that insertion can be performed faster whereas deletions will be the most costly.

C.1 Canonical Moves

Upon updating instanton I , we have that $T \rightarrow T + \Delta T$, with $(\Delta T)_{ij} = \delta_{iI} \xi_j^*$. This induces the following changes

$$(T^\dagger T)_{ij} \rightarrow (T^\dagger T)_{ij} + T_{iI}^\dagger \xi_j^* + \xi_i T_{Ij} + \xi_i \xi_j^*, \quad (\text{C.1})$$

$$\psi_i \equiv T_{Ii}^*, \quad (\text{C.2})$$

$$\phi_i \equiv \xi_i + \psi_i, \quad (\text{C.3})$$

$$T^\dagger T \rightarrow T^\dagger T + \phi \phi^\dagger - \psi \psi^\dagger. \quad (\text{C.4})$$

$$(TT^\dagger)_{ij} \rightarrow (TT^\dagger)_{ij} + \delta_{iI} \xi_k^* T_{kj}^\dagger + T_{ik} \xi_k \delta_{Ij} + \delta_{iI} \delta_{jI}, \quad (\text{C.5})$$

$$\psi_i \equiv \frac{1}{|\xi|} (T\xi)_i, \quad (\text{C.6})$$

$$\phi_i \equiv \delta_{Ii} |\xi| + \psi_i, \quad (\text{C.7})$$

$$T^\dagger T \rightarrow T^\dagger T + \phi \phi^\dagger - \psi \psi^\dagger. \quad (\text{C.8})$$

Changes in an anti-instanton will have analogous formulas.

C.2 Insertion

We focus on inserting an instanton. Insertion of an anti-instanton is then similar. Since in the code we always add an instanton at the end of the arrays, an insertion corresponds to adding a row to T .

$$T \rightarrow \begin{pmatrix} T \\ \xi^\dagger \end{pmatrix}, \quad (\text{C.9})$$

$$TT^\dagger \rightarrow \begin{pmatrix} T \\ \xi^\dagger \end{pmatrix} \begin{pmatrix} T^\dagger & \xi \end{pmatrix} = \begin{pmatrix} TT^\dagger & T\xi \\ \xi^\dagger T^\dagger & \xi^\dagger \xi \end{pmatrix}. \quad (\text{C.10})$$

On the level of the Cholesky decomposition

$$L \rightarrow \begin{pmatrix} L & 0 \\ \chi^\dagger & 1 \end{pmatrix}, \quad (\text{C.11})$$

$$D \rightarrow \begin{pmatrix} D & 0 \\ 0 & d \end{pmatrix}, \quad (\text{C.12})$$

$$LDL^\dagger \rightarrow \begin{pmatrix} LDL^\dagger & LD\chi \\ \chi^\dagger DL^\dagger & \chi^\dagger D\chi + d \end{pmatrix}. \quad (\text{C.13})$$

Remembering that the insertion also adds a mass term in the diagonal, we have to solve the following system

$$\begin{cases} LD\chi = T\xi \\ d = \xi^\dagger \xi + m^2 - \chi^\dagger D\chi \end{cases}, \quad (\text{C.14})$$

which can be solved in $O(N^2)$ by using backsubstitution. The case for $T^\dagger T$ is simply given by

$$T^\dagger T \rightarrow \begin{pmatrix} T^\dagger & \xi \end{pmatrix} \begin{pmatrix} T \\ \xi^\dagger \end{pmatrix} = T^\dagger T + \xi \xi^\dagger, \quad (\text{C.15})$$

which is a rank-1 update.

C.3 Deletion

We focus again on an instanton. Deletion will be a two step process. We first delete the last instanton and then swap it with that instanton that has been selected for deletion. The swapping is similar to a canonical move, where now ξ is given by the difference between the last instanton and the selected instanton.

The proper deletion part is given by

$$TT^\dagger \rightarrow \begin{pmatrix} TT^\dagger & 0 \\ 0 & 0 \end{pmatrix}, \quad (\text{C.16})$$

$$L \rightarrow \begin{pmatrix} L & 0 \\ 0 & 1 \end{pmatrix}, \quad (\text{C.17})$$

$$D \rightarrow \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}. \quad (\text{C.18})$$

The $T^\dagger T$ part is again simply related to a rank-1 update because, upon rearranging the result from the insertion part, we get

$$T^\dagger T \rightarrow T^\dagger T - \xi \xi^\dagger. \quad (\text{C.19})$$

Bibliography

- [1] Private communication with David Kofke.
- [2] S. Adler. Axial-Vector Vertex in Spinor Electrodynamics. *Phys. Rev.*, 177:2426, 1969.
- [3] B. Allen, P. Casper, and A. Ottewill. Analytic results for the gravitational radiation from a class of cosmic string loops. *Phys. Rev.*, D50:3703–3712, 1994.
- [4] B. Allen and E. P. S. Shellard. Gravitational radiation from cosmic strings. *Phys. Rev.*, D45:1898–1912, 1992.
- [5] B. Allen and E.P.S. Shellard. COSMIC STRING EVOLUTION: A NUMERICAL SIMULATION. *Phys. Rev. Lett.*, 64:119–122, 1990.
- [6] B. Allés, M. D’Elia, and A. Di Giacomo. Topological susceptibility at zero and finite T in SU(3) Yang-Mills theory. *Nucl. Phys. B*, 494:281–292, 1997.
- [7] C. Amsler et al. Review of particle physics. *Phys. Lett.*, B667:1, 2008.
- [8] Y. Aoki, S. Borsanyi, S. Durr, Z. Fodor, S. Katz, S. Krieg, and K. Szabo. The QCD transition temperature: results with physical masses in the continuum limit II. *JHEP*, 06:088, 2009.
- [9] Y. Aoki, Z. Fodor, S.D. Katz, and K.K. Szabo. The QCD transition temperature: results with physical masses in the continuum limit. *Phys.Lett. B*, 643:46–54, 2006.
- [10] H. Aoyama and H. Kikuchi. A new valley method for instanton deformation. *Nucl. Phys. B*, 369:219–234, 1992.

- [11] A. Armoni, M. Shifman, and G. Veneziano. QCD Quark Condensate from SUSY and the Orientifold Large-N Expansion. *Phys. Lett. B*, 579:384–390, 2004.
- [12] M. Atiyah, N. Hitchin, V. Drinfeld, and Y. Manin. Construction of instantons. *Phys. Lett.*, A65:185–187, 1978.
- [13] K. Bae, J. Huh, and J. Kim. Update of axion CDM energy. *JCAP*, 0809:005, 2008.
- [14] D. Bailin and A. Love. *Cosmology in Gauge Field Theory and String Theory*. Institute of Physics (Graduate Student Series in Physics), 2004.
- [15] C. A. Baker et al. An improved experimental limit on the electric dipole moment of the neutron. *Phys. Rev. Lett.*, 97:131801, 2006.
- [16] I.I. Balitsky and A.V. Yung. Collective-coordinate method for quazero modes. *Phys. Lett. B*, 168:113–119, 1986.
- [17] V. Baluni. CP Violating Effects in QCD. *Phys. Rev.*, D19:2227–2230, 1979.
- [18] T. Banks and M. Dine. The cosmology of string theoretic axions. *Nucl. Phys.*, B505:445–460, 1997.
- [19] R. Battye and E.P.S. Shellard. Axion String Constraints. *Phys. Rev. Lett.*, 73:2954–2957, 1994.
- [20] R. A. Battye and E. P. S. Shellard. Radiative back reaction on global strings. *Phys. Rev.*, D53:1811–1826, 1996.
- [21] R. A. Battye and E. P. S. Shellard. Recent perspectives on axion cosmology. 1997.
- [22] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin. Pseudoparticle solutions of the Yang-Mills equations. *Phys. Lett. B*, 59:85–87, 1975.

- [23] J. Bell and R. Jackiw. A PCAC puzzle: $\pi^0 \rightarrow \gamma \gamma$ in the sigma model. *Nuovo Cim.*, A60:47–61, 1969.
- [24] M. Beltran, J. Garcia-Bellido, and J. Lesgourgues. Isocurvature bounds on axions revisited. *Phys. Rev.*, D75:103507, 2007.
- [25] D. Bennett and F. Bouchet. High resolution simulations of cosmic string evolution. 1. Network evolution. *Phys. Rev. D*, 41:2408, 1990.
- [26] G. Bertone, D. Hooper, and J. Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005.
- [27] K. Bitar and S.-J.Chang. Vacuum Tunneling of Gauge Theory in Minkowski Space. *Phys. Rev.*, D17:486, 1978.
- [28] R. Brandenberger, I. Halperin, and A. Zhitnitsky. Baryogenesis with QCD Domain Walls. In *COSMO 98*, 1998. hep-ph/9809471.
- [29] F. Bruckmann, C. Gattringer, E.M. Ilgenfritz, M. Müller-Preussker, A. Schafer, and S. Solbrig. Quantitative comparison of filtering methods in lattice QCD. *Eur. Phys. J. A*, 33:333–338, 2007.
- [30] S. Burns. Isentropic and isocurvature axion perturbations in inflationary cosmology. 1997.
- [31] C.G. Callan, R.F. Dashen, and D.J. Gross. Toward a theory of the strong interactions. *Phys.Rev. D*, 17:2717–2763, 1978.
- [32] C.G. Callan, R.F. Dashen, and D.J. Gross. A theory of hadronic structure. *Phys.Rev. D*, 19:2826–1855, 1979.
- [33] Sanghyeon Chang, C. Hagmann, and P. Sikivie. Studies of the motion and decay of axion walls bounded by strings. *Phys. Rev.*, D59:023505, 1999.
- [34] B. Chen and J.I. Siepmann. A Novel Monte Carlo Algorithm for Simulating Strongly Associating Fluids: Applications to Water, Hydrogen Fluoride, and Acetic Acid. *J. Phys. Chem. B*, 104:8725–8734, 2000.

- [35] H.-Y. Cheng. The Strong CP Problem Revisited. *Phys. Rept.*, 158:1, 1988.
- [36] K. Chetyrkin, J. Kühn, and M. Steinhauser. RunDec: a Mathematica package for running and decoupling of the strong coupling and quark masses. *Comput. Phys. Commun.*, 133:43–65, 2000.
- [37] T. Chiu, S. Aoki, H. Fukaya, S. Hashimoto, T. Hsieh, T. Kaneko, H. Matsufuru, J. Noaki, K. Ogawa, T. Onogi, and N. Yamada. Topological susceptibility in 2-flavor lattice QCD with fixed topology. *PoS Lattice 2007*, 068, 2007.
- [38] T. Chiu, S. Aoki, S. Hashimoto, T. Hsieh, T. Kaneko, H. Matsufuru, J. Noaki, T. Onogi, and N. Yamada. Topological susceptibility in (2+1)-flavor lattice QCD with overlap fermion. *PoS Lattice 2008*, 072, 2008.
- [39] T. Chiu, T. Hsieh, and P. Tseng. Topological susceptibility in (2+1) flavors lattice QCD with domain-wall fermions. *Phys. Lett. B*, 671:135–138, 2008.
- [40] M. Chu and S. Schramm. Instanton content of finite temperature QCD matter. *Phys. Rev. D*, 51:4580–4586, 1995.
- [41] M.C. Chu, J. Grandy, S. Huang, and J. W. Negele. Evidence for the Role of Instantons in Hadron Structure from Lattice QCD. *Phys. Rev. D*, 49:6039–6050, 1994.
- [42] S. Coleman. *Aspects of Symmetry*. Cambridge University Press, 1995.
- [43] T. Coleman and M. Roos. Effective degrees of freedom during the radiation era. *Phys. Rev.*, D68:027702, 2003.
- [44] J.C. Collins. *Renormalization*. Cambridge University Press, 1984.
- [45] R. Crewther, P. Di Vecchia, G. Veneziano, and E. Witten. Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics. *Phys. Lett.*, B88:123, 1979.

- [46] M. Cristoforetti, P. Faccioli, M.C. Traini, and J.W. Negele. Exploring the Chiral Regime of QCD in the Interacting Instanton Liquid Model. *Phys.Rev. D*, 75:034008, 2007.
- [47] A. Dabholkar and J. Quashnock. PINNING DOWN THE AXION. *Nucl. Phys.*, B333:815, 1990.
- [48] R. Davis. Goldstone bosons in string models of galaxy formation. *Phys. Rev. D*, 32:3172, 1985.
- [49] R. Davis. Cosmic axions from cosmic strings. *Phys. Lett. B*, 180:225, 1986.
- [50] R. Davis and E.P.S. Shellard. Do axions need inflation. *Nucl. Phys. B*, 324:167, 1989.
- [51] T. DeGrand and C. DeTar. *Lattice Methods for Quantum Chromodynamics*. World Scientific.
- [52] C. Detar and R. Gupta. Toward a precise determination of T_c with 2+1 flavors of quarks. *PoS*, LAT2007:179, 2007.
- [53] D. Diakonov, N. Gromov, V. Petrov, and S. Slizovskiy. Quantum weights of dyons and of instantons with non-trivial holonomy. *Phys.Rev. D*, 70:036003, 2004.
- [54] D. Diakonov, V. Petrov, and A. Yung. Quasiclassical expansion of Yang-Mills heat kernels and approximate calculation of functional determinants. *Phys.Lett. B*, 130:385–388, 1983.
- [55] D. Diakonov, M. Polyakov, and C. Weiss. Hadronic matrix elements of gluon operators in the instanton vacuum. *Nucl. Phys.*, B461:539–580, 1996.
- [56] M. Dine. TASI Lectures on the strong CP problem. *hep-ph/0011376*.
- [57] M. Dine, W. Fischler, and M. Srednicki. A simple solution to the strong CP problem with a harmless axion. *Phys. Lett. B*, 104:199, 1981.
- [58] S. Dodelson. *Modern Cosmology*. Academic Press, 2003.

- [59] G.V. Dunne, J. Hur, C. Lee, and H. Min. Calculation of QCD Instanton Determinant with Arbitrary Mass. *Phys.Rev. D*, 71:085019, 2005.
- [60] S. Durr, Z. Fodor, C. Hoelbling, and T. Kurth. Precision study of the SU(3) topological susceptibility in the continuum. *JHEP*, 04:055, 2007.
- [61] D.I. Dyakonov and V.Yu. Petrov. Instanton-based vacuum from the Feynman variational principle. *Nucl. Phys. B*, 245:259–292, 1984.
- [62] D.I. Dyakonov and V.Yu. Petrov. A theory of light quarks in the instanton vacuum. *Nucl. Phys. B*, 272:457–489, 1986.
- [63] C. Amsler et al. *Phys. Lett. B*, 667:1, 2008.
- [64] K. Hagiwara et al. Quantum Chromodynamics. *Phys. Rev. D*, 66:010001–1, 2002.
- [65] P. Faccioli. Strong CP breaking and quark-antiquark repulsion in QCD, at finite θ . *Phys. Rev. D*, 71:091502, 2005.
- [66] P. Faccioli, D. Guadagnoli, and S. Simula. The Neutron Electric Dipole Moment in the Instanton Vacuum: Quenched Versus Unquenched Simulations. *Phys.Rev.D*, 70:074017, 2004.
- [67] M. Forbes and A. Zhitnitsky. Primordial galactic magnetic fields from domain walls at the QCD phase transition. *Phys. Rev. Lett.*, 85:5268–5271, 2000.
- [68] M. Forbes and A. Zhitnitsky. WMAP Haze: Directly Observing Dark Matter? 78:083505, 2008.
- [69] P. Fox, A. Pierce, and S. Thomas. Probing a QCD string axion with precision cosmological measurements. 2004.
- [70] D. Frankel and B Smit. *Molecular Simulation: From Algorithm to Applications*. Academic Press, second edition edition, 2002.
- [71] W. Freedman et al. Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. *Astrophys. J.*, 553:47–72, 2001.

- [72] D. Frenkel and B. Smit. *Understanding Molecular Simulation*. Academic Press, 2002.
- [73] T. Fugleberg, I. Halperin, and A. Zhitnitsky. Domain walls and theta dependence in QCD with an effective Lagrangian approach. *Phys. Rev.*, D59:074023, 1999.
- [74] G. Gabadadze and M. Shifman. Vacuum structure and the axion walls in gluodynamics and QCD with light quarks. *Phys. Rev.*, D62:114003, 2000.
- [75] G. Gabadadze and M. Shifman. QCD vacuum and axions: What’s happening? *Int. J. Mod. Phys.*, A17:3689–3728, 2002.
- [76] A. M. García-García and J. C. Osborn. Chiral phase transition and Anderson localization in the Instanton Liquid Model for QCD. *Nucl. Phys. A*, 770:141–161, 2006.
- [77] M. Gasperini and G. Veneziano. Pre - big bang in string cosmology. *Astropart. Phys.*, 1:317–339, 1993.
- [78] J. Gasser and H. Leutwyler. Quark Masses. *Phys. Rept.*, 87:77–169, 1982.
- [79] M. Gell-Mann. The Eightfold Way: A Theory of strong interaction symmetry. CTSL-20.
- [80] P. Gerhold, E.M. Ilgenfritz, and M. Müller-Preussker. An $SU(2)$ KvBLL caloron gas model and confinement. *Nucl. Phys. B*, 760:1–37, 2007.
- [81] P. Gill, G. Golub, W. Murray, and M. Saunders. Methods for Modifying Matrix Factorizations. *Mathematics of Computations*, 28:505–535, 1974.
- [82] M. Göckeler, R. Horsley, A. Irving, D. Pleiter, P. Rakow, G. Schierholz, and H. Stüben. A Determination of the Lambda Parameter from Full Lattice QCD. *Phys. Rev. D*, 73:014513, 2006.

- [83] D.J. Gross, R.D. Pisarski, and L.G. Yaffe. QCD and instantons at finite temperature. *Rev. Mod. Phys.*, 53:43–80, 1981.
- [84] B. Grossman. Zero energy solutions of the Dirac equation in an N-pseudoparticle field. *Phys. Lett. A*, 61:86–88, 1977.
- [85] C. Hagmann, Sanghyeon Chang, and P. Sikivie. Axion radiation from strings. *Phys. Rev.*, D63:125018, 2001.
- [86] C. Hagmann and P. Sikivie. Computer simulations of the motion and decay of global strings. *Nucl. Phys. B*, 363:247, 1991.
- [87] I. Halperin and A. Zhitnitsky. Anomalous effective Lagrangian and theta dependence in QCD at finite N(c). *Phys. Rev. Lett.*, 81:4071–4074, 1998.
- [88] I. Halperin and A. Zhitnitsky. Axion potential, topological defects and CP-odd bubbles in QCD. *Phys. Lett.*, B440:77–88, 1998.
- [89] I. Halperin and A. Zhitnitsky. Can Theta/N dependence for gluodynamics be compatible with 2π periodicity in Theta? *Phys. Rev.*, D58:054016, 1998.
- [90] J. Hamann, S. Hannestad, G. Raffelt, and Y. Wong. Isocurvature forecast in the anthropic axion window. *JCAP*, 0906:022, 2009.
- [91] S. Hannestad, A. Mirizzi, and G. Raffelt. New cosmological mass limit on thermal relic axions. *JCAP*, 0507:002, 2005.
- [92] D. Harari and P. Sikivie. On the evolution of global strings in the early universe. *Phys. Lett. B*, 195:361, 1987.
- [93] B.J. Harrington and H.K. Shepard. Periodic Euclidean solutions and the finite-temperature Yang-Mills gas. *Phys. Rev. D*, 17:2122–2125, 1978.
- [94] A. Hasenfratz and P. Hasenfratz. The scales of euclidean and hamiltonian lattice QCD. *Nucl. Phys. B*, 193:210–220, 1981.

- [95] M. Hertzberg, M. Tegmark, and F. Wilczek. Axion Cosmology and the Energy Scale of Inflation. *Phys. Rev.*, D78:083507, 2008.
- [96] K. Huang. *Quarks, Leptons & Gauge Fields (2nd edition)*. World Scientific, 1992.
- [97] E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proc. Nat. Acad. Sci.*, 15:168–173, 1929.
- [98] W. Hwang. Some thoughts on the cosmological QCD phase transition. *Int. J. Mod. Phys.*, A23:4757–4777, 2008.
- [99] E. Ilgenfritz and E. Shuryak. Chiral symmetry restoration at finite temperature in the instanton liquid. *Nucl. Phys. B*, 319:511–520, 1989.
- [100] E. Ilgenfritz and E. Shuryak. Quark-induced correlations between instantons drive the chiral phase transition. *Phys. Lett. B*, 325:263–266, 1994.
- [101] E.M. Ilgenfritz, M. Müller-Preussker, and D. Peschka. Calorons in $SU(3)$ lattice gauge theory. *Phys.Rev. D*, 71:116003, 2005.
- [102] D. Kaplan and A. Manohar. Current Mass Ratios of the Light Quarks. *Phys. Rev. Lett.*, 56:2004, 1986.
- [103] J. I. Kapusta. *Finite-temperature field theory*. CUP, 1993.
- [104] F. Karsch. Recent lattice results on finite temperature and density QCD, part II. *PoS*, LAT2007:015, 2007.
- [105] F. Karsch and E. Laermann. Susceptibilities, the specific heat and a cumulant in two flavor QCD. *Phys. Rev. D*, 50:6954, 1994. Private communication with Frithjof Karsch.
- [106] D. Kharzeev. Parity violation in hot QCD: why it can happen, and how to look for it. *Phys. Lett. B*, 633:260–264, 2006.
- [107] D. Kharzeev, L. McLerran, and H. Warringa. The effects of topological charge change in heavy ion collisions: Event by event P and CP violation. *Nucl. Phys. A*, 803:227–253, 2008.

- [108] D. Kharzeev, R. Pisarski, and M. Tytgat. Possibility of spontaneous parity violation in hot QCD. *Phys. Rev. Lett.*, 81:512–515, 1998.
- [109] D. Kharzeev, R. Pisarski, and M. Tytgat. Aspects of parity, CP, and time reversal violation in hot QCD. In *SEWM 2000*, 2000. hep-ph/0012012.
- [110] D. Kharzeev and A. Zhitnitsky. Charge separation induced by P-odd bubbles in QCD matter. *Nucl. Phys. A*, 797:67–79, 2007.
- [111] T. W. B. Kibble. Topology of Cosmic Domains and Strings. *J. Phys.*, A9:1387–1398, 1976.
- [112] J. Kim. Weak-interaction singlet and strong CP invariance. *Phys. Rev. Lett.*, 43:103, 1979.
- [113] J. Kim. Light pseudoscalars, particle physics and cosmology. *Phys. Rep.*, 150:1, 1987.
- [114] J. Kim. QCD axion and quintessential axion. 2003.
- [115] J. Kim. Axion as a CDM component. 2007.
- [116] H. Kleinert. *Path Integral in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*. World Scientific, 2006.
- [117] E.W. Kolb and M.S. Turner. *The Early Universe*. Westview Press, 1990.
- [118] E. Komatsu et al. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Astrophys. J. Suppl.*, 180:330–376, 2009.
- [119] T.C. Kraan and P. van Baal. Exact T -duality between Calorons and Taub-NUT spaces. *Phys. Lett. B*, 428:268–276, 1998.
- [120] T.C. Kraan and P. van Baal. Monopole Constituents inside $SU(N)$ Calorons. *Phys. Lett. B*, 435:389–395, 1998.
- [121] T.C. Kraan and P. van Baal. Periodic Instantons with non-trivial Holonomy. *Nucl. Phys. B*, 533:627–659, 1998.

- [122] M. Kuster, G. Raffelt, and B. Beltrán, editors. *Axions*. Springer, 2008.
- [123] K. Lee and C. Lu. $SU(2)$ Calorons and Magnetic Monopoles. *Phys. Rev. D*, 58:025011, 1998.
- [124] F. Lenz, J.W. Negele, and M. Thies. Confinement from merons. *Phys. Rev. D*, 69:074009, 2004.
- [125] H. Leutwyler and A. Smilga. Spectrum of Dirac operator and role of winding number in QCD. *Phys. Rev. D*, 46:5607–5632, 1992.
- [126] A. Linde. Inflation and Axion Cosmology. *Phys. Lett.*, B201:437, 1988.
- [127] A. Linde. *PARTICLE PHYSICS AND INFLATIONARY COSMOLOGY*. Harwood Academic Publishers, 1990.
- [128] E. Linder. The Dynamics of Quintessence, The Quintessence of Dynamics. *Gen. Rel. Grav.*, 40:329–356, 2008.
- [129] D. H. Lyth. Axions and inflation: Sitting in the vacuum. *Phys. Rev.*, D45:3394–3404, 1992.
- [130] Q. Mason, H. Trottier, R. Horgan, C. Davies, and G. Lepage. High-precision determination of the light-quark masses from realistic lattice QCD. *Phys. Rev.*, D73:114501, 2006.
- [131] Q. Mason, H. Trottier, R. Horgan, C. Davies, and G. Lepage. High-precision determination of the light-quark masses from realistic lattice QCD. *Phys. Rev. D*, 73:114501, 2006.
- [132] E. Masso, F. Rota, and G. Zsembinski. On axion thermalization in the early universe. *Phys. Rev.*, D66:023004, 2002.
- [133] L. McLerran, E. Mottola, and M. Shaposhnikov. SPHALERONS AND AXION DYNAMICS IN HIGH TEMPERATURE QCD. *Phys. Rev.*, D43:2027–2035, 1991.
- [134] I. Montvay and G. Münster. *Quantum Fields on a Lattice*. Cambridge University Press, 1997.

- [135] M. Nakahara. *Geometry, Topology and Physics*. Institute of Physics (Graduate Student Series in Physics), 2003 (second edition).
- [136] J.W. Negele, F. Lenz, and M. Thies. Confinement from Instantons or Merons. *Nucl. Phys. Proc. Suppl.*, 140:629, 2005.
- [137] H. P. Nilles. Hidden sector axions: Physics and cosmology. 2003.
- [138] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.
- [139] M.A. Nowak, M. Rho, and I. Zahed. *Chiral Nuclear Dynamics*. World Scientific, 1996.
- [140] Ken D. Olum and J. J. Blanco-Pillado. Radiation from cosmic string standing waves. *Phys. Rev. Lett.*, 84:4288–4291, 2000.
- [141] Ken D. Olum and Vitaly Vanchurin. Cosmic string loops in the expanding universe. *Phys. Rev.*, D75:063521, 2007.
- [142] G. Orkoulas and A.Z. Panagiotopoulos. Free energy and dphase equilibria for the restricted primitive model of ionic fluids from Monte Carlo simulations. *J. Chem. Phys.*, 101:1452–1459, 1994.
- [143] B. Ovrut and H. Schnitzer. The Decoupling Theorem and Minimal Subtraction. *Phys. Lett. B*, 100:403–406, 1981.
- [144] B. Ovrut and H. Schnitzer. Effective Field Theory in Background Gauge. *Phys. Lett. B*, 110:139–142, 1982.
- [145] R. Peccei and H. Quinn. Constraints imposed by CP conservation in the presence of pseudoparticles. *Phys. Rev. D*, 16:1791, 1977.
- [146] R. Peccei and H. Quinn. CP conservation in the presence of pseudoparticles. *Phys. Rev. Lett.*, 38:1440, 1977.
- [147] W. Percival et al. The shape of the SDSS DR5 galaxy power spectrum. *Astrophys. J.*, 657:645–663, 2007.
- [148] M. Garcia Perez, O. Philipsen, and I.O. Stamatescu. Cooling, Physical Scales and Topology. *Nucl. Phys. B*, 551:293–313, 1999.

- [149] D.H. Perkins. *Introduction to High Energy Physics*. Cambridge University Press, 4th edition (hardback) edition, 2000.
- [150] S. Perlmutter et al. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *Astrophys. J.*, 517:565–586, 1999.
- [151] M. Peskin and D. Schroeder. *An Introduction to Quantum Field Theory*. Perseus Books, 1995.
- [152] S. Pokorski. *Gauge Field Theories*. Cambridge University Press, 2000.
- [153] A. Riess et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astron. J.*, 116:1009–1038, 1998.
- [154] A. Riess et al. New Hubble Space Telescope Discoveries of Type Ia Supernovae at $z > 1$: Narrowing Constraints on the Early Behavior of Dark Energy. *Astrophys. J.*, 659:98–121, 2007.
- [155] G. Rodrigo and A. Santamaria. QCD matching conditions at thresholds. *Phys. Lett. B*, 313:441–446, 1993.
- [156] L.F. Rull, G. Jackson, and B. Smit. The condition of microscopic reversibility in Gibbs ensemble Monte Carlo simulations of phase equilibria. *Mol. Phys.*, 85:435–447, 1995. Thanks to Berend Smit for pointing out this paper.
- [157] T. Schäfer and E.V. Shuryak. Interacting instanton liquid in QCD at zero and finite temperatures. *Phys.Rev. D*, 53:65226542, 1996.
- [158] T. Schäfer, E.V. Shuryak, and J.J.M. Verbaarschot. Chiral phase transition and instanton-anti-instanton molecules. *Phys. Rev. D*, 51:1267–1281, 1995.
- [159] E. P. S. Shellard. AXIONIC DOMAIN WALLS AND COSMOLOGY. In *Liege 1986, Proceedings, Origin and early history of the universe* 173-180. (see Conference Index).
- [160] E. P. S. Shellard. *Quantum Effects in the Early Universe*. PhD thesis, University of Cambridge, 1987.

- [161] M. Shifman. *ITEP Lectures on Particle Physics and Field Theory*. World Scientific, 1999.
- [162] M. Shifman, A. Vainshtein, and V. Zakharov. Can confinement ensure natural CP invariance of strong interactions? *Nucl. Phys. B*, 166:493, 1980.
- [163] E. Shuryak and M. Velkovsky. Instanton density at finite temperatures. *Phys. Rev. D*, 50:3323–3327, 1994.
- [164] E.V. Shuryak. The role of instantons in quantum chromodynamics (I). Physical vacuum. *Nucl. Phys. B*, 203:93–115, 1982.
- [165] E.V. Shuryak. The role of instantons in quantum chromodynamics (II). Hadronic structure. *Nucl. Phys. B*, 203:116–139, 1982.
- [166] E.V. Shuryak. The role of instantons in quantum chromodynamics (III). Quark-gluon plasma. *Nucl. Phys. B*, 203:140–156, 1982.
- [167] E.V. Shuryak. Toward the quantitative theory of the instanton liquid (I). Phenomenology and the method of collective coordinates. *Nucl. Phys. B*, 302:559–579, 1988.
- [168] E.V. Shuryak. Toward the quantitative theory of the instanton liquid (II). The $SU(2)$ gluodynamics. *Nucl. Phys. B*, 302:574–598, 1988.
- [169] E.V. Shuryak. Toward the quantitative theory of the instanton liquid (III). Instantons and light fermions. *Nucl. Phys. B*, 302:599–620, 1988.
- [170] E.V. Shuryak. Instantons in QCD (I). Properties of the "instanton liquid". *Nucl. Phys. B*, 319:521–540, 1989.
- [171] E.V. Shuryak. Instantons in QCD (II). Correlators of pseudoscalar and scalar currents. *Nucl. Phys. B*, 319:541–569, 1989.
- [172] E.V. Shuryak and J.J.M. Verbaarschot. QCD Instantons at finite temperature. *Nucl. Phys. B*, 364:255–282, 1991.
- [173] E.V. Shuryak and J.J.M. Verbaarschot. Screening of the topological charge in a correlated instanton vacuum. *Phys. Rev. D*, 52:295–306, 1995.

- [174] P. Sikivie. Of Axions, Domain Walls and the Early Universe. *Phys. Rev. Lett.*, 48:1156–1159, 1982.
- [175] P. Sikivie and Q. Yang. Bose-Einstein Condensation of Dark Matter Axions. 2009.
- [176] Pierre Sikivie. Axion cosmology. *Lect. Notes Phys.*, 741:19–50, 2008.
- [177] A. Smilga. *Lectures on Quantum Chromodynamics*. World Scientific, 2001.
- [178] P. Steinhardt and N. Turok. Cosmic evolution in a cyclic universe. *Phys. Rev.*, D65:126003, 2002.
- [179] Karl Strobl and Thomas J. Weiler. Anharmonic evolution of the cosmic axion density spectrum. *Phys. Rev.*, D50:7690–7702, 1994.
- [180] L. Susskind. The anthropic landscape of string theory. 2003.
- [181] P. Svrcek and E. Witten. Axions in string theory. *JHEP*, 06:051, 2006.
- [182] G. 't Hooft. Computation of the quantum effects due to a four-dimensional pseudoparticle. *Phys. Rev. D*, 14:3432–3448, 1976.
- [183] M. Tegmark et al. Cosmological Constraints from the SDSS Luminous Red Galaxies. *Phys. Rev.*, D74:123507, 2006.
- [184] Max Tegmark, Anthony Aguirre, Martin Rees, and Frank Wilczek. Dimensionless constants, cosmology and other dark matters. *Phys. Rev.*, D73:023505, 2006.
- [185] D. Tong. TASI Lectures on Solitons. *hep-th/0509216*.
- [186] M. Turner. Cosmic and local mass density of invisible axions. *Phys. Rev. D*, 33:889, 1986.
- [187] M. Turner. Windows on the Axion. *Phys. Rept.*, 197:67–97, 1990.
- [188] M. Turner and F. Wilczek. Inflationary Axion Cosmology. *Phys. Rev. Lett.*, 66:5–8, 1991.

- [189] C. Vafa and E. Witten. Parity conservation in quantum Chromodynamics. *Phys. Rev. Lett.*, 53:535, 1984.
- [190] J.J.M. Verbaarschot. Streamlines and conformal invariance in Yang-Mills theories. *Nucl. Phys. B*, 362:33–53, 1991.
- [191] J. Vermaseren, S. Larin, and T. Ritbergen. The 4-loop quark mass anomalous dimension and the invariant quark mass. *Phys. Lett. B*, 405:327–333, 1997.
- [192] A. Vilenkin and A. E. Everett. Cosmic Strings and Domain Walls in Models with Goldstone and PseudoGoldstone Bosons. *Phys. Rev. Lett.*, 48:1867–1870, 1982.
- [193] A. Vilenkin and E.P.S. Shellard. *Cosmic Strings and Other Topological Defects*. Cambridge University Press, 1994.
- [194] D.P. Visco and D.A. Kofke. Modeling the Monte Carlo simulation of associating fluids. *J. Chem. Phys.*, 110:5493–5502, 1999.
- [195] L. Visinelli and P. Gondolo. Dark Matter Axions Revisited. *Phys. Rev.*, D80:035024, 2009.
- [196] M. Wagner. Classes of confining gauge field configurations. *Phys. Rev. D*, 75:016004, 2007.
- [197] M. Wagner. Fermions in the pseudoparticle approach. *Phys. Rev.*, D76:076002, 2007.
- [198] R.M. Wald. *General Relativity*. The University of Chicago Press, 1984.
- [199] Z.Q. Wang, X.F. Lu, and F. Wang. Dilute liquid of instanton and its topological charge dominate the QCD vacuum. *AIP Conf. Proc.*, 865:242–247, 2006.
- [200] S. Weinberg. *The quantum theory of fields (Vol 2)*. Cambridge University Press.
- [201] S. Weinberg. A new light boson? *Phys. Rev. Lett.*, 40:223, 1978.
- [202] S. Weinberg. Effective Gauge Theories. *Phys. Lett. B*, 91:51–55, 1980.

- [203] S. Weinberg. Anthropic Bound on the Cosmological Constant. *Phys. Rev. Lett.*, 59:2607, 1987.
- [204] S. Weinberg. *Cosmology*. Oxford University Press, 2008.
- [205] S. Wierzchowski and S.A. Kofke. A general-purpose biasing scheme for Monte Carlo simulation of associating fluids. *J. Chem. Phys.*, 114:8752–8762, 2001.
- [206] F. Wilczek. Problem of strong P and T invariance in the presence of Instantons. *Phys. Rev. Lett.*, 40:279, 1978.
- [207] K. Yagi, T. Hatsuda, and Y. Miake. *Quark-Gluon Plasma*. Cambridge University Press, paperback edition edition, 2008.
- [208] A. Zhitnitsky. On Possible Suppression of the Axion Hadron Interactions. (In Russian). *Sov. J. Nucl. Phys.*, 31:260, 1980.
- [209] A. Zhitnitsky. Cold Dark Matter as Compact Composite Objects. *Phys. Rev. D*, 74:043515, 2006.