Baryogenesis Constraints on the Electroweak Phase Transition in Two-Higgs-Doublet Models

Gordon Jenkins

A thesis submitted for the degree of Ph.D.
in the Department of Physics and Astronomy
at the University of Glasgow, Glasgow G12 8QQ

September 1997

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

This thesis is dedicated to Mum and Dad,
for their love, encouragement and support
in everything I have achieved.

Acknowledgements

Firstly, I would like to thank my supervisor, Andrew Davies, for his guidance in my work, for the many hours he spent educating me, and especially for his patience during my extended period of writing up.

I would also like to thank the other members of staff in the particle theory group at Glasgow for their time and the benefit of their experience and understanding. I am particularly grateful for the many discussions with Gordon Moorhouse and Colin Froggatt which helped me steer a clearer course to these conclusions.

Lastly, I must thank my friends and fellow students at Glasgow, from departments in every faculty, for the long hours they spent keeping me sane and for the occasional pint.

Declaration

This thesis describes the work I undertook as a postgraduate student in the Department of Physics and Astronomy at Glasgow University between 1991 and 1995. Chapters 1 through 3 act as an introduction to the work contained in the thesis and describe the background to the various concepts covered. These are drawn from historical and contemporary literature. Chapters 4 through 8, and appendix A are my own work – chapter 4 was written in collaboration with Gordon Moorhouse and the same series of results was published by Kastening [24] while the work was in progress; chapters 5 and 6 are entirely my own work; chapter 7 presents the results of my numerical work described in chapter 6; chapter 8 was written in collaboration with Gordon Moorhouse and Colin Froggatt who provided the data on which I worked; appendix A was initiated by a conversation with Gordon Moorhouse who was good enough to review the work as I produced it. All of the work herein was undertaken in collaboration with my supervisor, Andrew Davies.

Abstract

The preservation of the baryon asymmetry of the universe (BAU) after the electroweak phase transition ($EWPT$) is endangered by baryon-number violating sphaleron transitions. Limits on the sphaleron transition rate in order to protect the BAU require a strongly first order $EWPT$ and lead to an upper mass limit on the Standard Model Higgs scalar mass of around 60GeV. This is in direct disagreement with the experimental lower limit on the SM Higgs mass, which is now greater than 60GeV.I consider the problem of avoiding this disagreement by extending the scalar sector of the SM to include a second complex scalar Higgs doublet. My primary aim is to locate a region of the parameter space which can sustain the first order $EWPT$ required for baryogenesis and satisfy the experimental bounds on Higgs masses.

The background to the electroweak phase transition and baryogenesis are discussed briefly. The formalism of the effective potential is introduced and explained. The evaluation of the effective potential is described in terms of the loop expansion, and the high temperature approximation of the 1-loop expansion is developed. Higher order corrections are included by introducing temperature dependent effective masses, equivalent to including "daisy" diagrams in the 1-loop effective potential. I also include the suppression of the zero mode longitudinal components of the $W^{\pm}$ and Z^0 bosons due to the development of a Debye "screening" mass, which leads in turn to suppression of their contributions to the $T\mathrm{Tr}\,[M^3]$ term in the effective potential by a factor of 2/3. The results obtained for the general form of the effective potential are applied to the case of the two Higgs doublet scalar sector. The tree level potential chosen for the two Higgs case is taken to be as general as possible, within the requirements of gauge invariance and renormalisability. A complex coupling term ($\mu_3^2 = \mu_R^2 + i\mu_I^2$) is included in the tree level potential to explicitly violate CP-symmetry, which is a requirement of some baryogenesis scenarios. The CP-violating term is introduced in a manner sufficient to adequately suppress any possible flavour changing neutral currents (FCNC's).

I derive conditions on the coupling parameters of the effective potential in order

to constrain the parameters to a physically acceptable basis space. The strength of the $EWPT$ is measured by the parameter $v(T_c)/T_c$, which is constrained to preserve the BAU. ($v(T_c)$ is the VEV of the scalar fields at the critical temperature, T_c.) Lastly, an upper limit is imposed on $m_i(T_c)/T_c$ (where $m_i(T_c)$ are the mass eigenvalues of all particles involved in the effective potential) to ensure that the high temperature approximation to the 1-loop effective potential is valid. The critical temperature of the phase transition is discussed for the general form of the potential and an explicit expression is derived for T_c.

Analytical evaluation of the potential is not possible due to the complicated form of the most general neutral scalar mass matrix. Hence, I develop and describe a numerical method to sample the physical basis parameter space at random in a search for the location and structure of a BAU-preserving region. This method pays particular attention to the behaviour of the μ_3^2 term by working on a grid of discrete μ_R^2, μ_I^2 values. An extensive search of the basis parameter space is conducted by sampling over one quarter of a million basis space points on this grid.

The results of the numerical search are presented, including the size of the baryogenesis region relative to the basis space, the structure of the region as a function of (μ_R^2, μ_I^2), the mass spectrum of the region, and the critical temperature and strength of the electroweak phase transition are measured. The effects of the constraints on $v(T_c)/T_c$ and $m_i(T_c)/T_c$ are considered in detail in order to understand which structures of the baryogenesis region are physical and which are due to the limited validity of the approximations.

The search concludes that a baryogenesis region does exist in the general two Higgs-doublet model, and that the size of the region is around 5% of the basis space sampled. The mass spectrum in this region is easily sufficent to avoid the mass bounds from experimental searches. However, the structure of the region is difficult to understand analytically, due to the lack of expressions for $v(T_c)$ and $m_i(T_c)$. The region is insensitive to increases in the experimental lower mass bounds, but is particularly sensitive to the exact value of the limit on the order parameter $v(T_c)/T_c$, with an increase of only 10% cutting the baryogenesis region

by around $60 - 70\%$. The complex coupling μ_3^2 is found to weaken the average strength of the $EWPT$ while simultaneously increasing the proportion of the basis space found to lie in the baryogenesis region.

A brief study of the the minimal supersymmetric model is made as a special case of the general two Higgs-doublet model and essentially the same numerical techniques are used. The renormalistion group equations (RGE) must be used to run the gauge couplings at electroweak scales, Λ_{EW}, up to the SUSY scale, Λ_{susy}, where the $MSSM$ defines the scalar couplings in terms of the electroweak gauge couplings, thereby reducing the number of free parameters in the effective potential. The RGE are also used to run the scalar couplings from SUSY scales down to Λ_{EW}. The number of free parameters in the $MSSM$ is now only 2, so the search technique is changed to methodically search for a baryogenesis region in μ_R^2 at discrete values of β_0. The resulting search finds a disappointingly small baryogenesis region, the majority of which is disfavoured by the existence of the heavy top quark.

I conclude from the results that electroweak baryogenesis is not attractive in the $MSSM$, but that the general two Higgs-doublet model looks much more promising. However, the baryogenesis region, even in the general case, is still small and very sensitive to $v(T_c)/T_c$. More detailed understanding of the limit on the order parameter for the $EWPT$ would help to decide the fate of electroweak baryogenesis with two scalar doublets, and an improved range of validity for the expansion of the effective potential would allow definitive study into regions currently beyond the reach of these techniques.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The baryon asymmetry of the universe (BAU) was long considered as a fundamental constant which should be introduced into cosmological models as an initial condition. However, it was first suggested by Sakharov [1] in the late 1960's that the BAU might be generated by baryon-number violating processes in the early universe. In the three decades since Sakharov's suggestion, various scenarios for baryogenesis have been imagined and investigated, culminating in the present interest in the electroweak phase transition ($EWPT.$)

The $EWPT$ has interesting possibilities for baryogenesis, which will be discussed in chapter 2 ("Background to the study".) However, in the early 1990's it was becoming apparent that the scenarios for baryogenesis at the $EWPT$, or preserving the BAU in the post-$EWPT$ universe, were no longer in agreement with the Standard Model (SM). The problem lay in the predicted mass of the SM Higgs boson – in order to accomodate a suitable first order $EWPT$ in the Standard Model, the Higgs boson had to be **lighter** than about 60GeV [11, 14, 15, 12, 17], while the latest experimental searches gave a **lower** limit which was also about 60GeV [18].

This disagreement between theory and experimental observations was the problem tackled by the work described in this thesis.

1.1 Aim of the thesis

The aim of this thesis is to consider the possibility of circumventing the disagreement between the theoretical upper limit and experimental lower limit on the Higgs mass. To this end, I will consider an extension to the Standard Model scalar sector – the addition of a second Higgs scalar doublet. (This is a common extension to consider since minimal supersymmetric models also have two scalar doublets.) It is hoped that the extra degrees of freedom provided by the second doublet will allow the model to accomodate a first order phase transition with Higgs masses larger than the experimental lower limit.

Having located the region of coupling parameter space which supports the strong first order phase transition, its size and structure will be investigated – for example, the scalar mass spectrum – in order to understand how the two Higgs doublet model is able to avoid the experimental Higgs mass limits.

1.2 A summary of the thesis

This work is aimed at saving the BAU – however it might be generated – from being washed out by sphaleron transitions **after** the electroweak phase transition. I do not consider the production of the BAU, although I do give a brief overview of the suggested mechanisms in chapter 2. In particular I do not discuss the dynamics of the phase transition, bubble nucleation rates, and so on.

In an attempt to avoid the disagreement between experiment and theory on the Higgs boson mass, I consider a scalar sector with two complex doublets. The two doublet model has several attractive features. For example, the mass differences between the up-type and down-type quarks can be explained by coupling each type to only one Higgs doublet. Then the mass differences are explainable in terms of the vacuum expectation values of the scalar doublets. I include the top and bottom quark in the effective potential and couple them to the Higgs doublets in this fashion.

The two doublet model also allows the introduction of explicitly CP-violating

terms. Extra sources of CP-violation are attractive in many baryogenesis scenarios. Such a term is introduced in the scalar potential while being careful to avoid flavour changing neutral currents.

Also, supersymmetric models are popular due to their many attractive features. The minimal supersymmetric model contains two Higgs doublets, and this is considered as a special case of the more general work.

The most general possible form of the scalar potential is developed, within the restrictions of gauge-invariance and renormalisability. It has been common, in the literature on two scalar doublet models, to consider special cases with restrictions imposed on the coupling parameters [14, 4] for the purposes of simplifying calculations or by invoking arguments on the likely areas of interest in the parameter space. I have opted instead to make a thorough search of the parameter space, using the most general form of the potential throughout. The region of the search is limited by developing constraints from fundamental physical arguments – eg., the mass spectrum must be physical; the vacuum state must conserve EM; the potential must be bounded below.

In order to determine the nature of the phase transition and the critical temperature, and to examine the scalar mass spectrum, I introduce the temperature dependent effective potential and consider the loop expansion at the 1-loop term. I also include higher order corrections which introduce a temperature dependence in the effective masses of the scalar fields.

Since I am interested in the preservation of the BAU **after** the phase transition, the temperature of interest to me is that of the universe just after the phase transition occurs. Hence, I define the "critical" temperature as that at which the curvature of the potential at the origin first becomes negative, and derive an analytical expression for this temperature in the general case. It is also noted that numerical studies have suggested that this definition of T_c suffers from problems in the accuracy of the expansion of the effective potential near the origin, particularly at temperatures near the phase transition. With this in mind, a possible correction to the temperature calculation is investigated in Appendix A

I impose constraints which ensure that the sphaleron transition rate after the

phase transition will not wash out the BAU, and which support BAU generation in $EWPT$ scenarios. The sphaleron rate after the $EWPT$ must be much less than the expansion rate of the universe at the epoch of the $EWPT$. This is equivalent to saying that the mass of the sphaleron, which is proportional to the VEV of the Higgs fields, must be large compared to the critical temperature. It turns out that the VEV of the Higgs field at $T = T_c$ must be greater than about half the VEV at zero T, and hence the transition must be strongly first order.

I will describe a numerical search of the parameter space confined by the constraints mentioned above, looking for a region which can accomodate a strongly first order EW phase transition capable of preserving the BAU after the phase transition. This involves developing a computational model of the effective potential which allows numerical minimisation of the potential at the critical temperature in order to assess the order of the phase transition and suppression of the sphaleron rate. The work was extended to consider minimal SUSY as a special case – this involves setting the couplings in the model at the SUSY scale and running them down to electroweak scale by using the renormalisation group equations. This followed work by Moorhouse, Froggatt and Knowles [23], who provided the coupling parameter data sets used in this part of the study.

Finally, I draw some conclusions from the results and briefly consider the future of this topic.

Chapter 2

Background to the study

2.1 A brief history of electroweak baryogenesis

In the observed universe, there exists an outstanding agreement between the empirical measurements of relative abundances of the light elements and the theoretically derived values. This is a major triumph for both cosmology (ie., the hot big bang) and nuclear theory. However, it is also observed that the material universe consists almost entirely of baryonic matter, as opposed to anti-baryonic matter (anti-matter) – on large scales, there have been no observations of the enormous energy outputs which would result from the meeting of large bodies of matter and anti-matter; and on local scales it is obvious that our immediate surroundings are almost entirely baryonic, not anti-baryonic. The source of this baryon/anti-baryon asymmetry has remained a puzzle for particle theorists, but its size plays an important role in the calculations of the light element abundances. For the outstanding agreement mentioned above, it is necessary to "fine tune" the size of the asymmetry – usually quantified in terms of the baryon-to-entropy ratio, $n_B/s \approx (4-6) \times 10^{-11}$, or the baryon-to-photon ratio, $n_B/n_\gamma \approx (3-4) * 10^{-10}$. [3]

A necessity for fine tuning is often seen as a lack in understanding of a theory – a sign that there is a more fundamental truth to be found underlying our present thought. For this reason it is attractive to explain the source of the "fine-tuned" baryon-to-entropy ratio in cosmology, and this led Sakharov, in 1970 [1], to sug-

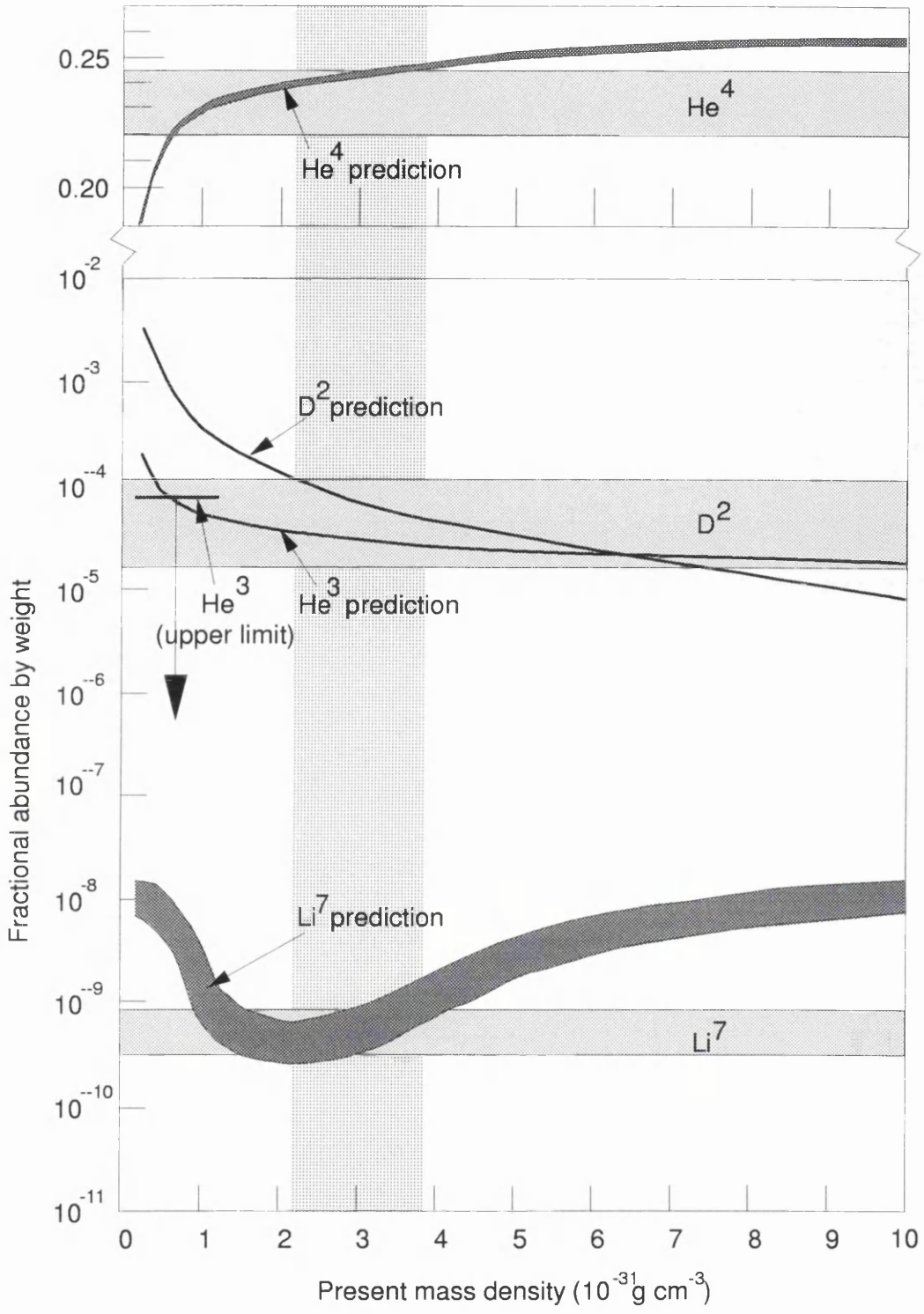

Figure 2.1: Predicted and observed mass densities of the light elements synthesised in the Big Bang (excluding Hydrogen.) [46] The horizontal shaded bands represent the ranges of observed densities; the He^3 observation is only an upper limit, represented by the short horizontal line with a downwards arrow. The thick curves represent the predicted densities, with errors, for various values of the present mass density of the universe. The shaded vertical bar represents the region of present mass densities where the predictions agree with **all** observations.

gest that the baryon/anti-baryon asymmetry could be produced through baryon number violating transitions in the early universe, the size of the asymmetry being determined dynamically. He described the following set of conditions which would be necessary for the dynamical production of a baryon asymmetry –

1. Baryon number violation
2. C and CP violation
3. Departure from thermal equilibrium

The first requirement is obvious – without baryon number (B) violating transitions then the B-symmetry is exact and we cannot hope to produce an asymmetry. It was 1976 before t'Hooft demonstrated [2] that the axial anomaly (or Adler-Bell-Jackiw/ABJ anomaly) in the Standard Model (SM) leads to B-violation. Since the violation is anomalous, it is due to non-perturbative effects and it is impossible to draw a Feynman graph which manifestly violates B-conservation. Nevertheless, the B-symmetry is not exact in SM physics.

The second condition is necessary in order to bias the direction of baryon number violation, thereby producing a net baryon number. The electroweak theory violates C (since the left and right-handed sectors of the theory have entirely different structures), and CP-violation is introduced in the Yukawa couplings via phases in the Cabibbo-Kobayashi-Maskawa matrix. However, CP-violation via the CKM matrix requires amplitudes with a quark from each generation, otherwise the phases can be transformed away, and this leads to CP-violation only at higher orders in perturbation theory. This level of CP-violation is too small to explain the observed size of the baryon/anti-baryon asymmetry. Hence, there is an interest in models which can introduce further sources of CP-violation via the Higgs sector.

Lastly, in thermal equilibrium the transitions which increase baryon number occur at exactly the same rate as their reverse transitions. In fact, in thermal equilibrium, the relative abundances of baryons and anti-baryons will depend only on their masses, and by CPT invariance all particle/anti-particle pairs have identical masses. Hence, in thermal equilibrium, the relative abundances of particles and their anti-particle counterparts are identical. However, if the universe departs sig-

nificantly from thermal equilibrium then the reaction/reverse-reaction rates are no longer balanced and it becomes possible to generate a non-zero baryon number.

As stated above, the SM already satisfies requirements 1 and 2. However, shortly after t'Hooft's demonstration of the SM baryon number violation by instantons, it remained to be seen whether condition 3 could be satisfied in SM physics. And in any case, the rates of B-violating processes by zero temperature instanton effects were entirely negligible.

In the mid-eighties it was suggested [6, 13, 8] that standard model B-violation was likely to proceed rapidly in a high temperature universe. In these arguments, Klinkhamer and Manton introduced the 'sphaleron' – a static solution of the classical field equations, which bridges between topologically different vacuua with different baryon number and different Chern-Simons number, as we shall discuss below. At temperatures higher than the mass of the sphaleron, it was argued (notably by Kuzmin et al) [6, 13, 8] that the universe would regularly cross over the barrier, entering a new vacuum configuration with new B number, or new $B + L$ number to be more specific.

With the possibility of producing a net baryon number in the early universe, the electroweak phase transition ($EWPT$) suddenly became interesting as a source of Sakharov's departure from thermal equilibrium. A first order phase transition might provide the required non-equilibrium, leaving behind a universe with a net baryon number, fixed by the dynamics of the phase transition and the subsequent rate of possible sphaleron transitions. A great deal of work has ensued in the areas of high temperature field theory, the $EWPT$ and sphaleron transitions. (See reference list.)

2.2 Sakharov's conditions in electroweak physics

Here we will briefly consider each of the conditions required for baryogenesis, as enumerated by Sakharov, and how these conditions can be met.

2.2.1 Baryon number violation

The necessity of baryon number violation in any baryogenesis process is obvious – if B is conserved, and has an initial value of 0, then it is impossible to develop a non-zero baryon number over time. In such a scenario, the universe must be "born" with an initial baryon asymmetry, but the input of such free parameters in a theory is not generally popular among theorists, as already mentioned. It is much more satisfying to be able to explain the source of these apparent free parameters and baryogenesis is such an attempt.

The original models of baryon-number violating processes involve physics at the Grand Unified Theory (GUT) scale; ie., $E \approx 10^{16}$ GeV. In these models, baryon violation tends to be explicit (as opposed to via anomalous processes) and involves the exchange of very massive particles. The energies involved can be roughly bounded below by considering the results of experimental observations of proton decay, which is predicted in these GUT scenarios – no proton decay has ever been detected and the lifetime of the proton has been shown to be $\tau_p \geq 10^{32}$ yr [Particle data group, 1990]. For a point process decay at low energy, the lifetime goes as $\tau_p \approx \Lambda^4/M_p^5$, where Λ is the typical energy scale of baryon number violation[10]. This results in a lower bound on the typical energy scale of $\Lambda \geq 10^{16}$ GeV. There are other objections to the GUT models as sources of the baryon asymmetry, such as their prediction of the existence of magnetic monopoles which also remain unobserved.

However, as already stated, it has been known since the mid-70's that electroweak physics also violates baryon number, through the ABJ anomaly, which comes about through the chiral coupling in electroweak theory. Since the theory classically conserves baryon (or, strictly, the baryon plus lepton) number, the baryon number current is divergenceless, and hence the classical Euler-Lagrange equations give

$$\partial_\mu J_B^\mu = 0 \tag{2.2.1}$$

The introduction of quantum corrections results in the baryon number current

developing a divergence of the form

$$\partial_\mu J^\mu_B = \partial_\mu K^\mu_{CS}(W,Z) \tag{2.2.2}$$

and similarly for the lepton current. This equation introduces the Chern-Simons current, K_{CS}. The Chern-Simons number is a topological number related to the winding number of the gauge field configuration.

As a result of eqn.(2.2.2), the change in baryon number (and lepton number) is equal to the change in Chern-Simons number, and hence the W and Z fields can dynamically evolve $B+L$, since

$$\frac{d}{dt}Q_B = \frac{d}{dt}Q_{CS}(W,Z) \tag{2.2.3}$$

Note that since the expressions for the change in B and L are identical, then $B-L$ is conserved, despite the fact that $B+L$ is not conserved.

The vacuum state of a system is the state which minimises the potential energy functional of the system. It has been shown [6] that the potential energy as a functional of field configurations is periodic and that it is possible to cross the barriers between adjacent vacuua via *sphalerons*. Sphalerons are static solutions to the field equations, and have one negative mode – hence they correspond to a saddle point in the potential energy functional and are the lowest point at which the field configurations can cross the barriers between vacuua. However, the Chern-Simons number is not invariant under the transformations between these vacuua – in fact, a sphaleron transition between two adjacent vaccua always results in $\Delta N_{CS} = \pm 1$. This corresponds to a change in $B+L$ of $N_f \times \Delta N_{CS} = \pm N_f$ as we move between vacuua, where N_f is the number of lepton families in the model.

2.2.2 C/CP-violation

In order to understand why CP-violation is necessary in baryogenesis, consider any process which violates baryon number - let us say, for definiteness, that the process increases B by 1. Under CP-transformation, such a process becomes its anti-particle equivalent, with all particles replaced by their anti-particle partners

- it's "anti-process", let's say. The anti-process will decrease B by 1. If CP-symmetry holds, then both of these processes will have the same cross-section, and they will occur at the same rate. Hence, even though the process violates baryon number, the combined rate of change of baryon number due to any CP-conserving process and it's anti-process will be zero.

Hence, CP-violation is a requirement of baryogenesis in order to bias the rate of B-violation in the positive direction.

The minimal standard model violates CP symmetry, as evidenced by the 2π decay of the long-lived component of a K^0 beam. This was long considered to be a CP-odd eigenstate. However, in an experiment by Christenson *et al* [45], a kaon beam consisting entirely of long-lived states was observed to decay to 2 pions, with a branching ratio of order 10^{-3}. Hence, the symmetry is violated by a factor of about 1 part in 1000. The source of the asymmetry is thought to be a phase between the EW interactions of the quarks and their Higgs interactions. The quark couplings are described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix. With 3 generations of quarks the mixing angles in the CKM can develop a physically observable phase, leading to observable CP-violation. However, if any of the mixing angles vanishes, or any two of the quark masses (of the same charge) are degenerate, then it is possible to make phase transformations in the quark fields so that there is no physically observable phase. Hence, all 3 generations must be involved in any CP-violating process. Involving all 3 generations of quarks leads to small cross-sections for such processes and hence to small levels of CP asymmetry. [3]

The typical estimate of the magnitude of the CP-asymmetry at the EWPT is given by

$$\begin{aligned} \epsilon \sim & \frac{\sin\theta_{12}\sin\theta_{23}\sin\theta_{13}}{(T_c)^{12}} \\ & \times(m_u^2-m_c^2)(m_u^2-m_t^2)(m_c^2-m_t^2)(m_d^2-m_s^2)(m_d^2-m_b^2)(m_s^2-m_b^2) \end{aligned} \tag{2.2.4}$$

where θ_{ij} are the quark mixing angles, m_i are the quark masses, and T_c is the critical temperature of the EWPT, which is considered as the typical scale at the transition. [3] The resultant asymmetry is of order 1 part in 10^{20}, which is very

small by any measure.

The BAU , being dependent on CP-violation, will be limited by such a small asymmetry - the level of CP-violation is about 10 orders of magnitude too small to explain the observed BAU of $n_B/s \sim (4-6)\times 10^{-11}$. For this reason it is attractive to introduce a further source of CP-violation, thereby allowing the possibility of achieving the observed level of the BAU. This is possible in extensions of the standard model, and we consider a CP-violating term in our scalar potential.

2.2.3 Thermal non-equilibrium

In thermal equilibrium, the proportions of the constituent particles of a system is constant, and reactions and their corresponding reverse reactions occur at the same rate. [1] Now, consider any process which violates baryon number. For definiteness, let us say that the process increases B by 1. The reverse process will decrease B by 1. In thermal equilibrium the average rate of the process will be exactly matched by its reverse process. Hence, in thermal equilibrium, even with B-violation present, the combined rate of change of baryon number due to any process and its reverse process, will be zero. Hence, the system must move out of thermal equilibrium in order to unbalance the reaction rates and bias the rate of B-violation in the positive direction.

This argument is fairly convincing, but baryogenesis requires C/CP-violation as well as thermal non-equilibrium, and this introduces a complication. The exact equivalence of rates of reactions and their reverse reactions is known as detailed balance. However, with CP-violation present, detailed balance is lost.

To see this, follow an example by Dolgov [5]. Consider a decay process which violates B-conservation, $X \to qq$, and its CP "anti-process" $\overline{X} \to \overline{qq}$. With CP-violation, the reaction rates of these processes can be written as

$$\begin{aligned} \Gamma(X \to qq) &= (1+\epsilon)\Gamma_0 \\ \Gamma(\overline{X} \to \overline{qq}) &= (1-\epsilon)\Gamma_0 \end{aligned} \tag{2.2.5}$$

[1] It will be demonstrated shortly below that this latter assertion on reaction rates is not necessarily true.

with $\epsilon > 0$. Under CPT-transformation, the rates will be invariant and hence

$$\begin{aligned} \Gamma(qq \to X) &= (1-\epsilon)\Gamma_0 \\ \Gamma(\overline{qq} \to \overline{X}) &= (1+\epsilon)\Gamma_0 \end{aligned} \tag{2.2.6}$$

So, the reverse reactions do not exactly balance their corresponding reactions – ie., detailed balance has been lost – and the bias of B-violation induced by the CP-violation is in the same direction for both the original processes and their inverses.

However, it still holds true that the particle number distributions in thermal equilibrium are determined only by their mass, even without detailed balance. The distributions are saved by realising that, since CPT is a perfect symmetry, the total decay cross sections of X and $\overline{X}$ must be equal, and hence there must be other decay channels open to X, otherwise the process could not violate CP. It is the interaction of the products of all available decay channels which maintains the equilibrium relationship between mass and the particle distributions. [5]

Hence, in order to bias the particle number distributions in favour of positive baryon number, the system must move out of thermal equilibrium.

2.3 The role of the phase transition

Systems tend towards thermal equilibrium over time. However, in first order phase transitions the discontinuous changes in the system can force it out of equilibrium. In the case of the EWPT, the order of the transition depends on the position of the secondary minimum of the scalar effective potential as it becomes the true vacuum. If the secondary minimum lies too close to the origin as the phase transition begins, then the transition will be only weakly first order and will not be strong enough to bias the baryon-violating processes sufficiently to generate the observed BAU. *A priori* it is possible that the secondary minimum might appear at the origin and move continuously away from the origin as T decreases - in this case the transition is actually second order and the universe will remain in thermal equilibrium. Either scenario is disastrous for baryogenesis.

In order to quantify the strength of a first order EWPT and determine a limiting strength for the development of the BAU, we turn to sphalerons. After the phase transition, in regions of the broken symmetry phase, the BAU is endangered by sphaleron transitions. If the rate of fractional change in baryon number is greater than the expansion rate of the universe, then the broken phase will return to thermal equilibrium and the BAU will be wiped out. Hence we can use the expansion rate of the universe to determine an upper limit for the sphaleron transition rate.

The rate of the sphaleron transitions will be dependent on the mass of the sphaleron configuration relative to the energy (or temperature) of the particle heat bath. The mass has been derived and is often quoted (eg., [4, 5, 6]) –

$$M_{\rm sph} = \frac{4\pi B(\lambda) v(T)}{g_W} \tag{2.3.7}$$

where $v(T)$ is the Higgs VEV at temperature T, and the function $B(\lambda)$ has been numerically calculated to lie in the range $1.56(\lambda = 0)$ to $2.72(\lambda \to \infty)$. [6]

The fractional change in baryon number per unit time due to sphaleron transitions in the broken phase at the critical temperature is

$$\begin{aligned} \Gamma_{\rm sph} &= \frac{1}{Q_B}\frac{dQ_B}{dt} \\ &\approx T_c \exp\left(-\frac{M_{\rm sph}(T_c)}{T_c}\right) \end{aligned} \tag{2.3.8}$$

The condition which limits the strength of an acceptable phase transition is hence

$$\Gamma_{\rm sph} \leq H \approx \frac{T_c^2}{M_{\rm Pl}} \approx 10^{-17} T_c \tag{2.3.9}$$

where H is the Hubble constant, the rate of expansion of the universe, and $M_{\rm Pl}$. Hence the sphaleron mass is constrained to be large with respect to the critical temperature, namely

$$\frac{M_{\rm sph}(T_c)}{T_c} \geq 45 \tag{2.3.10}$$

By considering the expression (2.3.7) for the sphaleron mass, we can constrain the strength of the first order phase transition with the limit

$$\frac{v(T_c)}{T_c} \geq \frac{45 g_W}{4\pi B(\lambda)} \tag{2.3.11}$$

This limit defines the minimum strength of the EWPT which is sufficient to suppress the sphaleron rate and avoid baryon number wash-out in the broken symmetry phase. We will impose this limit in order to estimate the proportion of the parameter space of the two-Higgs-doublet effective potential which can sustain electroweak baryogenesis.

2.4 Anomalous B-violation

When the ABJ anomaly first came to light, it provided a serious problem for particle theorists since it rendered the standard model unrenormalisable. Renormalisability relies on the Ward identities (and their non-Abelian extensions) which are designed to conserve the gauge currents. The anomaly violated the Ward identities in certain Feynman graphs with axial vertices– for example, the triangular graph shown below. We will not consider the details of the anomaly here, except to point out that the problems with renormalisation were solved by cancellation. There are contributions to the triangle graph from each type of fermion, and the charge structure of the fermions in each generation results in exact cancellation of the anomaly.

The cancellation of the ABJ anomaly ensures that the gauge currents are conserved, thereby saving the renormalisability of the Standard Model. However, baryon number is not a gauge quantum number, and the 4-divergence of the baryon-number current is left non-zero. This is naively unexpected since the classically derived 4-divergence will certainly be zero. But it is the consideration of quantum corrections that leads to the discovery of the anomaly and to the related baryon-number violation.

The resulting 4-divergence of the B-current (or L-current) [20] can be written

$$\partial_\mu J_B^\mu = \frac{g^2 N_f}{16\pi^2} \mathrm{Tr}\left[F\tilde{F}\right] \tag{2.4.12}$$

where F is the field strength of the electroweak gauge fields, W and Z, and N_f is the number of flavours.

Since the baryon-number current is not divergenceless, the change in baryon

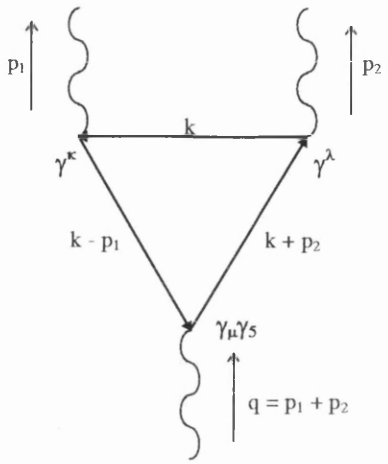

Figure 2.2: A VVA triangle diagram – diagrams such as these are the source of the ABJ anomaly.

number can be determined by integrating (2.4.12) over all space –

$$\Delta_B = \frac{g^2 N_f}{16\pi^2} \int d^4x \mathrm{Tr}\left[F\tilde{F}\right] \tag{2.4.13}$$

At this point it becomes useful to point out that $\mathrm{Tr}\left[F\tilde{F}\right]$ can be written as a divergence, giving

$$\Delta_B = N_f \int d^4x \partial_\mu K^\mu \tag{2.4.14}$$

where $K^\mu = g^2/8\pi^2 \epsilon^{\mu\nu\alpha\beta} \mathrm{Tr}\left[A_\nu(\partial_\alpha A_\beta - ig2/3 A_\alpha A_\beta)\right]$ is the Chern-Simons current.

Converting (2.4.13) to a surface integral over a 3-sphere of infinite radius, and noting that finite action field configurations must have $F = 0$ at $|\mathbf{x}| \to \infty$, then it can be shown

$$\Delta_B = N_f \left(N_{CS}(t = +\infty) - N_{CS}(t = -\infty)\right) = N_f \Delta N_{CS} \tag{2.4.15}$$

Here, N_{CS} is the Chern-Simons number, from which the Chern-Simons current derives. The Chern-Simons number is a topological number, derived by considering the mappings from 4-space to the configuration space. We will not pursue this definition further. It is sufficient for us to depict how N_{CS} varies with field configuration and potential. Diagram 2.3 represents the variation in Chern-Simons

number with field configuration. The minima correspond to vacuum configurations; as the configuration develops away from the vacuum, the potential energy increases, resulting in barriers between vaccua. The sphaleron is the lowest energy configuration which bridges the potential barrier between two adjacent vacuua. The sphaleron has one negative mode, corresponding to a saddle point in the barrier.

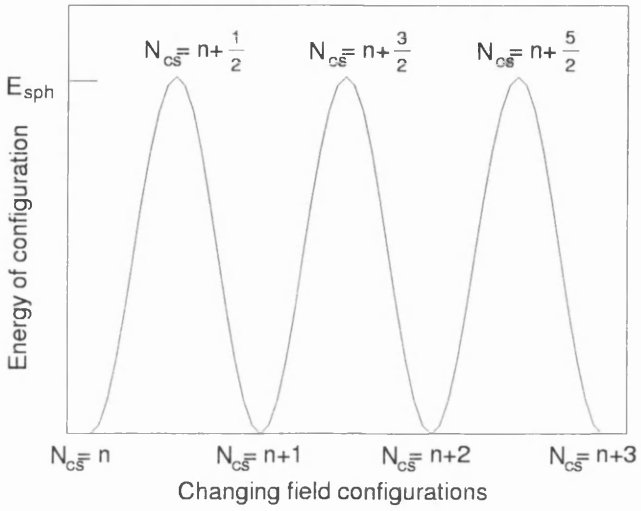

Figure 2.3: Representation of the variation of potential energy and Chern-Simons number with field configurations. The tops of the barriers represent sphaleron configurations, with N_{cs} of half integer values, and the minima represent vacuua, with integer N_{cs}.

The vacuum configurations, where V is minimised, all have integer values of N_{CS}, and hence moving from one vacuum to another results in an integer change in N_{CS}, and hence an integer change in $B + L$.

2.5 Baryogenesis scenarios

This thesis concentrates on the conditions in the broken symmetry phase immediately after the phase transition occurs. The sphaleron rate can be predicted in the broken phase by calculating the strength of the transition, the critical tem-

perature, size of the VEV, and so on. I impose conditions on the rate in order to preserve any baryon number which exists in the immediate aftermath of the EWPT. This requirement is an essential constraint on the EWPT itself, and is independent of the detailed scenario for the production of the BAU during or before the phase transition. Hence I do not investigate the baryogenesis scenarios in any detail, such as the levels of CP-asymmetry present, nor attempt to estimate the size of the BAU resulting from any particular model.

The aim of my method is to study the parameter space with a view to determining the possibility of preserving a BAU, however generated, after the EWPT in the two-Higgs-doublet model. As explained above, this does not require a study of the detailed scenarios, but I describe them briefly here for completeness.

2.5.1 Electroweak scenarios

Cosmological phase transitions are expected to proceed via the process of bubble nucleation, assuming that the transition is first order [42]. At high temperature the universe is in the symmetric phase. The universe expands and cools, eventually approaching the critical temperature, T_c. As local temperature fluctuations reach T_c, the phase transition begins locally and bubbles of the broken phase form. Initially the free energy released by the formation of a bubble is too small to overcome the surface tension in the bubble wall. These bubbles evaporate, so the universe remains largely in the symmetric phase and will supercool. Eventually, at some $T < T_c$, the fluctuations in free energy become large enough to defeat the surface tension and the bubbles can grow indefinitely, filling the universe until it lies entirely in the broken phase.

The phase transition is actually occuring at the bubble wall, the barrier between the symmetric and broken phases. Hence, it is in this region that the plasma is forced out of thermal equilibrium. So, it is C/CP-violating interactions with the bubble wall which will fulfill the Sakharov conditions. There are various models of the processes involved in baryogenesis at the bubble wall, but these fall into two categories – the adiabatic (or thick wall) and non-adiabatic (thin wall)

regimes.

The rate at which particles reach their equilibrium densities will depend on the strength of the interactions involved. If the rate of equilibration for the fastest interactions (ie. the strong interaction) is much slower than the rate of change of the Higgs VEV during the phase transition, then the bubble wall will be thin compared to the mean free path of the plasma. In this case, particles colliding with the bubble wall are scattered like quantum mechanical particles incident on a potential barrier. Due to CP-asymmetry, the reflection co-efficients of particles and anti-particles may be different, leading to a net charge (not B) in front of the bubble wall, in a process known as "charge transport". This charge asymmetry can bias anomalous B-violating processes in front of the bubble wall, leading to production of a net baryon number in this region. This passes through the wall into the broken phase, where the sphaleron rate is assumed to have been suppressed sufficiently to preserve the baryon asymmetry.

In the case where the phase transition is slow compared to the fastest strong and electroweak interactions (ie., the rate of change of the Higgs VEV is slow), the plasma will almost remain in thermal equilibrium. However, there are quasistatic chemical potentials introduced by the long equilibration times of the slowest interactions, such as the sphaleron interactions in the symmetric phase. A slowly varying Higgs field can drive charge production [42] by effectively introducing a new fermion kinetic term which splits the particle/anti-particle energy levels, leading to free energy minimisation at non-zero charge. The existence of the chemical potentials due to the slowly equilibrating quantum numbers, in the presence of the Higgs-developed charge asymmetry, leads to a chemical potential for baryon number. The baryons produced by this mechanism within the bubble wall enter the broken phase and become stable since the broken phase is assumed to preserve baryon number.

The details in both of these regimes are model-dependent. Notice, however, that in all cases the preservation of the BAU until the present epoch relies on the suppression of the sphaleron rate in the broken phase immediately after the phase transition.

2.5.2 Non-electroweak scenarios

There are many other proposed mechanisms for baryogenesis beyond those at the EWPT. These range from explicit B-violation in grand unified theories to more exotic topological effects.

Grand unified theories (GUT's) predict explicit B-violation via decays of heavy particles. However, there are several problems with GUT's which considerably weaken the case for these models. Principal among these are the predictions of proton decay and heavy relic particles, and the problem of cosmological re-heating.

GUT's predict proton decay, mediated by super heavy particles. The decay of a proton has never been experimentally observed and the lifetime is now constrained by $\tau_p \geq 10^{32}$yr, leading to a bound of $\Lambda \geq 10^{16}$GeV on the mass of the mediating particle. Such a heavy particle is difficult to incorporate even in GUT models. GUT's also predict heavy relics such as the magnetic monopole. Again, no monopole has ever been observed.

The lack of evidence for such heavy particles might not be considered surprising at today's experiments. However, possibly the strongest evidence against GUT's comes from cosmology. The arrival of the inflationary universe offered solutions to many of the problems of modern cosmology, particularly the homogeneity of the universe, and has been very successful. However, any BAU generated at the GUT scale will be massively diluted by the exponential expansion of the inflation epoch. The universe does undergo re-heating after inflation ends, but the temperature reached generally falls well below the GUT scale. Hence it is difficult to maintain a GUT scale BAU in the inflationary models. The success of inflation in describing the universe is difficult to ignore while the present GUT models cannot boast the same. Hence, baryogenesis at GUT scales becomes more difficult to convincingly sustain.

Even if a workable GUT model is found, any BAU it generates may be threatened by sphaleron transitions at a later epoch. Hence, the study of the electroweak scenarios might still play an important role even in GUT baryogenesis. One possible get-out clause here, however, is a GUT-developed $B - L$. Since the standard

model conserves $B - L$, while violating $B + L$, the BAU would not be wiped out. On balance, however, it seems that GUT theories raise more questions for baryogenesis than they answer.

There are other models which introduce heavy particles with baryon violating decays. Such models suffer many of the same problems as the GUT's, particularly avoiding dilution by inflation. One example is heavy gravitino decay. The gravitino is the super-symmetric partner of the graviton. In some models, gravitinos are very heavy, $O(10TeV)$ or more, and very weakly interacting. To avoid the number density dilution effects of inflation, they must be generated in the re-heating of the post-inflationary universe. Since they are so weakly interacting, they are out of thermal equilibrium at temperatures lower than their mass, and the decay temperature will be of order 1MeV, which is very late and hence is not threatened by inflation or sphaleron transitions. However, to produce a significant number density of such heavy particles the re-heating temperature must approach the GUT scale and, perhaps more problematically, the gravitinos decay during the nucleosynthesis epoch and can distort the standard nucleosynthesis model. As mentioned in the introduction, nucleosynthesis is one of the great achievements of modern nuclear physics and cosmology and it is difficult to tamper with it.

There are more exotic mechanisms suggested, particularly in the field of topological defects. Domain walls, strings and monopoles can radiate heavy particles with baryon non-conserving decays, for example, or can act as catalysts for baryon non-conservation. Also, black holes can contain any number of baryons but the number cannot be measured from the external universe, so they are able to "hide" the anti-baryons, leaving a baryon excess. There are proposed mechanisms to allow black holes to radiate a surplus of baryons, with or without baryon charge conservation. Consider a heavy boson with baryon number conserving decay channels to one light and one heavy baryon/anti-baryon, with CP-violation leading to a surplus of the light baryons over light anti-baryons, and a corresponding surplus of heavy anti-baryons over heavy baryons [5].

$$\Gamma_A \to L\overline{H} - \Gamma_A \to H\overline{L} \;=\; \Gamma_A\delta \qquad (2.5.16)$$

The black hole radiates A bosons, and they decay as above. However, the probability of re-capture of radiated particles by the black hole is dependent on the mass, and the heavy bosons are more likely to be captured. Hence, the flux of particles from the black hole consists of more light particles than heavy, and the light particles carry a surplus of baryons over anti-baryons. Hence, the universe develops a positive BAU, while the black holes hide the corresponding negative asymmetry.

There are many other proposed mechanisms for generating the BAU. However, many of these mechanisms – the black hole example above, say – are designed purely with baryogenesis as their aim, and hence lack a convincing argument for their existence. The baryogenesis of the electroweak and GUT scenarios, on the other hand, and of the topological defects to some extent, developed as consequence of attempts to explain the wider aspects of particle physics and cosmology. Hence they are more appealing, and the electroweak scenarios look particularly strong in the light of recent study.

2.6 An intuitive approach to non-perturbative effects

The violation of baryon number by anomalous currents is a fairly opaque mechanism. Being non-perturbative, it is impossible to explicitly describe an interaction (or Feynman diagram) which violates baryon number. However, there is a model due to Schwinger which is helpful in developing a physical picture of how anomalies can lead to violation of symmetries conserved in the classical Lagrangian. I will describe this picture here [7, 10] to complement the brief mathematical description above.

2.6.1 The Schwinger Model

In order to develop an intuitive understanding of how non-perturbative effects can violate baryon number, even in a theory with no B-violation in the Feynman

graphs, let us consider a model due to Schwinger. Take 2-dimensional QED, coupled to a massless fermion with charge e. Since the fermion is massless, its energy and momentum are simply related by $E = \pm p$. Let us also assume that the system is localised in a length L, and that we have anti-periodic boundary conditions. These assumptions are introduced simply to clarify the arguments by providing distinct states for the fermion, with no degeneracy at $E = 0$. So, the momentum is quantized in units of h/L, and since the Dirac equation has negative energy solutions, as well as positive, the energy levels of the fermion can be depicted

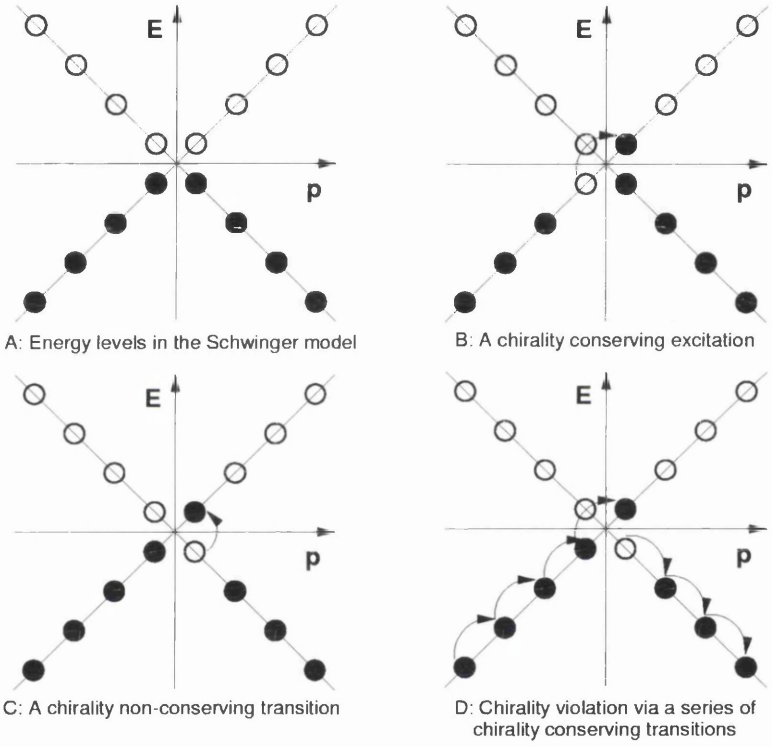

Figure 2.4: The energy and momentum levels in the Schwinger model (A). Although the model only allows chirality conserving transitions (B), chirality is violated by an infinite series of allowed transitions (D).

The chiral symmetry, $\psi \rightarrow \exp(i\theta\gamma^5)\psi$, holds for this Lagrangian, and the states with $E = \pm p$ have chirality ± 1. The symmetry is, however, anomalous.

It is easily seen that the lowest possible energy configuration here is the one

with all of the negative energy states filled – ie. with a full Dirac sea – and all of the positive energy states empty. Following Dirac's intuitive description of pair-production using the Dirac sea concept, a physical particle corresponds to an excitation of the vacuum, while an anti-particle can be described as an empty state in the Dirac sea. So, imagine a left-moving, negative energy particle being excited into a right-moving, positive energy state. This will leave behind a 'hole' in the Dirac sea, with equal but opposite energy to the particle. This hole corresponds to a right-moving,positive energy anti-fermion – this is the intuitive picture of pair production. Note that the excited particle remained on the same diagonal in diagram 2.4(B), and so the particle and anti-particle have equal but opposite chiral charge, and hence chirality is conserved.

However, if the particle was excited from one of the diagonals onto the other, the resulting process violates chiral charge, as seen in diagram 2.4(C). This chiral charge violating process appears then to be impossible, but in fact this effect can be produced by an infinite series of chirality-conserving transitions. To see this, consider the Dirac sea in a uniform electric field, directed to the right, as depicted in diagram 2.4(D). The fermions in the sea will shift as their momentum is changed. The left-moving fermions will move up the $E = +p$ diagonal, and the right-moving fermions will move down the $E = -p$ diagonal.

Consider firstly the $E = +p$ diagonal. If the electric field is allowed to act for long enough, so that the change in momentum is exactly h/L, then one particle will break the surface of the Dirac sea to occupy the first, right-moving positive energy level. However, the hole it leaves behind will be immediately filled by the particle excited up from the second energy level in the sea, and that hole filled from the third level, and so on - it is tempting at this point to say *ad infinitum.* Similarly, on the $E = -p$ diagonal, we imagine a hole being left in the first negative energy level, while the fermions are shifted into lower energy states even infinitely deep in the Dirac sea.

The remainder of this chirality-breaking argument relies on the words *ad infinitum.* If the Dirac sea is finitely deep, then our argument will collapse since we would leave a hole at the bottom of the Dirac sea on the $E = +p$ diagonal, and

we would be unable to shift fermions down into the sea on the $E = -p$ diagonal – ie., we are relying on the concept of an infinitely deep Dirac sea. Infinities are a long-standing problem in particle physics and it is well known that a theory with infinities requires to be regularised. If we introduce a cut-off scale $(-\Lambda)$ in the depth of the Dirac sea, then the Schwinger model and regularisation insist that we must be able to ignore the effects of the anti-particle "hole" as we let Λ tend to infinity. Following detailed analysis, it turns out that any well behaved regulator we introduce into a chiral gauge model must break the chiral symmetry. It is this fact that allows us to safely continue in the Schwinger model, free of worries about the behaviour of the infinities. The net effect has been to produce a non-zero change in the chiral charge of the universe. And yet this has come about through a series of chirality-conserving transitions, and hence it will be impossible to draw a Feynman diagram for the transition just described.

This model can be generalised to allow the spatial extent to tend to infinity, and allow arbitrary electric fields. The final result is the same – a non-zero change in the chiral charge of the universe but no explicit chirality changing Feynman diagrams. We have demonstrated that non-perturbative effects are responsible for breaking the chiral symmetry in a theory which classically conserves the symmetry.

2.6.2 Violating baryon number

The Schwinger model has demonstrated intuitively that it is possible for non-perturbative effects to break a symmetry which is apparently perfect at tree-level. It has not, however, demonstrated how the non-perturbative effects bring about baryon number violation.

In the Schwinger model, chirality was violated but particle number was invariant. It is a general result that, in gauge theories with vector coupling, only chiral symmetries are anomalous [40]. However, if we introduce chiral couplings, then we can also introduce non-chiral anomalies. $SU(2)$ electroweak theory is such a theory – only left-handed fermions couple to the gauge fields. In terms of the Schwinger model, this corresponds to decoupling the right-moving, negative

energy fermions from the electric field. Then the net effect of the transitions described above is to produce only a fermion, violating both chirality and particle number.

If the Schwinger model is extended to include extra species of massless, charged fermions, then each species is shifted in the Dirac sea independently of the others. Drawing analogy again with $SU(2)$, this would result in independent shifts of $SU(2)$ doublets. Hence, a unit shift of the Dirac sea would result in the production of one electron or neutrino, and one up or down-type quark of each colour. So, with the three families of the standard model, we produce changes in baryon and lepton number of $\Delta B = \Delta L = \pm 3$. Note that $B + L$ is violated, but $B - L$ is not.

Chapter 3

The Effective Potential

In a study of a phase transition, we are principally interested in the vacuum state or other meta-stable states which form false vacuua. An understanding of the nature and subsequent mechanics of the phase transition comes from an understanding of how these states develop with time and temperature.

For example, consider a model with a temperature-dependent potential. Let us say that, as the universe cools, the potential develops a secondary, local minimum, at some distance from the global minimum. If this secondary minimum (or false vacuum) becomes degenerate with the global minimum (true vacuum), then the population of both states becomes equally likely. At this point, there will be a potential barrier between the two vacuua, but it is possible for transitions to occur between them via tunnelling processes or thermal fluctuations which carry the configuration over the barrier. If the secondary minimum drops below the original vacuum, as T develops with time, it will become the true vacuum and the phase transition from the original vacuum state to the new vacuum state will be first order since we have a discontinuous change in the field strength between the two states. The rate of this phase transition, however, will clearly be determined by the height of the potential barrier which will suppress the rate of transition processes.

In a classical Lagrangian, the vacuum state is given by the minimum of the potential V, where V is the negative sum of the non-derivative terms in the Lagrangian. However, we are working with quantised theories and so we will want

to consider the quantum corrections to the Lagrangian and its potential. Hence, we look to the **effective action** – this is a field theory extension of the classical definition of the action as the integral of the Lagrangian along the trajectory of the system. From the effective action, we can develop the concept of the **effective potential.**

In order to clearly understand the roots of the effective potential, and how it is useful in the study of spontaneously broken symmetries, I will develop the formalism in the typical manner of the literature. I will then go on to consider the loop expansion technique and the inclusion of finite-temperature effects, before deriving the form of the effective potential in the 2-Higgs-doublet model.

3.1 Formalism of the effective potential, V_{eff}

Following a simple and commonly used example in the literature (eg. [29, 20, 40]), I will develop the formalism and define the effective potential in a theory with one scalar field $\phi(x)$, with Lagrangian density $\mathcal{L}(\phi, \partial_\mu \phi)$. Introducing a source $J(x)$ transforms the Lagrangian to

$$\mathcal{L} \to \mathcal{L} + J(x)\phi(x) \tag{3.1.1}$$

The effective action for this Lagrangian is a functional given by

$$\Gamma(\phi_c(x)) = W(J(x)) - \int d^4x J(x)\phi_c(x) \tag{3.1.2}$$

where ϕ_c is the classical field.

$W(J)$ is the connected generating functional – that is to say, it generates the connected Green's functions for the theory. It is defined in terms of the transition amplitude from the vacuum state at $t \to -\infty$, to the vacuum state at $t \to +\infty$, in the presence of the source J –

$$e^{iW(J)} = \langle O^+ | O^- \rangle_J \tag{3.1.3}$$

and can be written as a functional Taylor series

$$W(J) = \sum_n \frac{1}{n!} \int d^4x_1 \ldots d^4x_n G^{(n)}(x_1, \ldots, x_n) J(x_1) \ldots J(x_n) \tag{3.1.4}$$

The $G^{(n)}$ here are the connected Green's functions. In other words, $G^{(n)}$ is the sum of all connected Feynman graphs with n external legs. Hence, W generates the $G^{(n)}$ via

$$\frac{\partial^n W}{\partial J^n} = G^{(n)} \tag{3.1.5}$$

In particular, the classical field is defined as

$$\phi_c(x) = \frac{\partial W\left(J(x)\right)}{\partial J(x)} = \left[\frac{\langle O^+|\phi(x)|O^-\rangle}{\langle O^+|O^-\rangle}\right]_J \tag{3.1.6}$$

The effective action can also be expanded as a functional Taylor series –

$$\Gamma(\phi_c) = \sum_n \frac{1}{n!}\int d^4x_1 \ldots d^4x_n \Gamma^{(n)}(x_1 \ldots x_n)\phi_c(x_1)\ldots\phi_c(x_n) \tag{3.1.7}$$

It can be shown that Γ is the generating functional for the one-particle-irreducible (1PI) Green's functions. Hence, the $\Gamma^{(n)}$ are the sum of all 1PI Feynman graphs with n external legs. Note that the 1PI graphs are conventionally evaluated with vanishing external momenta.

Now that we know what the effective action represents, in terms of Feynman graphs, it is useful to consider another expansion. An expansion in terms of powers of momentum is equivalent to an expansion in terms of the derivatives of ϕ_c around the point ϕ_c = constant. This has the form

$$\Gamma = \int d^4x \left[-V(\phi_c(x)) + \frac{1}{2}(\partial_\mu\phi_c)^2 Z(\phi_c) + \ldots\right] \tag{3.1.8}$$

In the integrand above, by comparison of (3.1.8) with (3.1.7), it is clear that the n^{th} derivative of $V(\phi)$ is just the sum of all 1PI Feynman graphs with n external legs and vanishing external momenta.

$V(\phi_c(x))$ is known as the **effective potential** – the potential as derived from the effective action. I will refer to the effective potential as V_{eff}. Clearly, at tree level (ie. considering the non-derivative terms of $V_{\text{eff}}(\phi_c)$), this will reduce to the classical potential – the sum of the non-derivative terms from the classical Lagrangian.

Consider the conditions necessary for spontaneous symmetry breakdown. We require the field to develop a non-zero vacuum expectation value (VEV), even for $J(x) = 0$ – ie. $\partial\Gamma/\partial\phi_c = 0$ for some **non-zero** $\phi_c = \langle\phi\rangle$. But since we will

insist that the VEV is translationally invariant (in order to preserve conservation of linear momentum) this will reduce to $dV/d\phi_c = 0$, for some non-zero $\langle\phi\rangle$. The stability of the vacuum clearly requires that $\langle\phi_c\rangle$ lies at a minimum in V_{eff}.

The effective potential can be expanded around any point by redefining the quantum field

$$\phi \rightarrow \phi - \hat{\phi} \tag{3.1.9}$$

which leads to a corresponding redefinition of the classical field

$$\phi_c \rightarrow \phi_c - \hat{\phi}_c \tag{3.1.10}$$

With $\hat{\phi}_c = \langle\phi_c\rangle$, we are expanding about the vacuum state. Following the typical definitions of the renormalised couplings gives

$$\begin{aligned} \mu_R^2 \equiv & \quad -\Gamma^2(p_i = 0) = \left|\frac{d^2 V_{\text{eff}}}{d\phi_c^2}\right|_{\phi_c=0} \\ \lambda_R \equiv & \quad -\Gamma^4(p_i = 0) = \left|\frac{d^4 V_{\text{eff}}}{d\phi_c^4}\right|_{\phi_c=0} \end{aligned} \tag{3.1.11}$$

(This definition can be expanded to handle more complex theories, with analogous results.) By defining the couplings in this way, a shift in the quantum (classical) field leaves the form of these expressions unchanged. So, for example, if the field is redefined as $\phi_c \rightarrow \phi_c' - \langle\phi\rangle$, then the couplings are given by the expressions (3.1.11) **evaluated at** $\langle\phi\rangle$. In particular, this applies to the mass spectrum of the model – the expressions for the masses can be derived generically and the actual values are determined simply by evaluating at $\langle\phi_c\rangle$. Similarly, determining the value of $\langle\phi_c\rangle$ is simply a question of minimising V_{eff}, but the couplings will be the renormalised values.

3.2 V_{eff} – the loop expansion

The effective potential is a useful tool for studying the vacuum states in theories with spontaneous symmetry breaking. However, V_{eff} is an infinite summation of 1PI Feynman graphs, which is a computational impossibility with present techniques. Hence there is a need for an appropriate technique of approximating the effective potential.

The most commonly used approximation to V_{eff} is known as the **loop expansion** – an expansion in terms of the number of loops appearing in the Feynman graphs. Hence, the n^{th} order term in the expansion is the sum of all 1PI Feynman graphs containing n loops. We will demonstrate here that the loop expansion is equivalent to a power series expansion in terms of an expansion parameter.

Consider introducing a parameter α into the definition of the Lagrangian, thus

$$\mathcal{L} \to \alpha^{-1}\mathcal{L} \tag{3.2.12}$$

This introduces a factor of α^{-1} for every vertex in a Feynman graph, and a factor α for every internal propagator, since we are working with 1PI graphs which are evaluated with no propagators for the external legs. So, the power of α in any graph is given by

$$P = I - V \tag{3.2.13}$$

where I is the number of internal lines, and V is the number of vertices.

The number of loops in a graph will be equal to the number of independent momentum integrations. Each internal line contributes one such momentum; each vertex imposes a condition which constrains the number of independent momenta; and there is one overall condition for conservation of total momentum and energy. Hence, the number of independent momenta, and the number of loops, is given by

$$L = I - V + 1 = P + 1 \tag{3.2.14}$$

So, a loop expansion is equivalent to an expansion in powers of α, which is what we set out to show.

Since the units of the Lagrangian density are those of Planck's constant, and the Lagrangian is multiplied overall by $\hbar$, the loop expansion can be described as a power series expansion in $\hbar$. Hence, the use of the loop expansion in making perturbative studies of the effective potential is justified.

Another important point to note here is that the expansion parameter multiplies the whole of the Lagrangian. Hence, we are free to perform shifts on the fields, with no effect on the form of the loop expansion. We are also able to split

the Lagrangian into free and interacting parts, and to work with the expansion of each independently. This ensures that, at any order of the loop expansion, we are able to make use of these properties of the effective potential which are useful in surveying the vacuum states.

3.3 The effects of finite temperature

We begin by stating without proof [20] that a field theory at finite (non-zero) temperature behaves as a thermodynamic ensemble of finite-T Green's functions. This leads to a definition of finite-T operators as a thermal average of the zero-T operators, known as the Gibbs' average.

$$\Theta_\beta(x_1,\ldots,x_2) = \frac{\mathrm{Tr}e^{-\beta H}\Theta(x_1,\ldots,x_2)}{\mathrm{Tr}e^{-\beta H}} \tag{3.3.15}$$

where $\beta = 1/T$, and we are using units where Boltzman's constant has been set to 1.

Consider the numerator on the RHS of eqn(3.3.15). We can generally write this as

$$\mathrm{Tr}\left(e^{-\beta H}A(x_1,t_1)B(x_2,t_2)\ldots C(x_n,t_n)\right) \tag{3.3.16}$$

In the Schrodinger representation, $A(x_1,t_1) = e^{iHt_1}A(x_1,0)e^{-iHt_1}$. Hence we can write eqn(3.3.16) as

$$\mathrm{Tr}\left(e^{-\beta H}e^{iHt_1}A(x_1,0)e^{-iHt_1}e^{iHt_2}B(x_2,0)\ldots e^{iHt_n}C(x_n,0)e^{-iHt_n}\right) \tag{3.3.17}$$

Using the fact that $e^{-\beta H}e^{\beta H} = 1$, and the cyclic property of trace, we then have

$$\mathrm{Tr}\left(e^{-\beta H}e^{H(it_1+\beta)}A(x_1,0)e^{-iHt_1}\ldots e^{iHt_n}C(x_n,0)e^{-H(it_n+\beta}\right) \tag{3.3.18}$$

Comparing this with eqn(3.3.16), we realise that setting $it_j \to it_j + \beta$ has no effect on the operator Θ. Hence, the finite temperature Green's functions obey the same equations as those at zero-T, but they have periodic boundary conditions in Euclidean time. (This contrasts with the "usual", zero-T causal boundary conditions at $t_j = \pm\infty$.)

These results were determined for Bose particles. With Fermi statistics, the corresponding boundary conditions are anti-periodic. In both cases, the period is clearly β.

Hence, to derive the finite-T form for the effective potential, we should replace

$$\int d^4x_E \to \int_0^\beta d\tau \int d^3x \tag{3.3.19}$$

where $\tau = it$ is the Euclidean time; or, in terms of momentum,

$$k_0 \to 2\pi nT(\text{bosons}), k_0 \to (2n+1)\pi T(\text{fermions}) \tag{3.3.20}$$

which is a statement of the boundary conditions at finite-T, and

$$\int dk_0 \to i2T\sum_n \tag{3.3.21}$$

3.4 Calculation of V_{eff}

I will proceed to demonstrate how the effective potential is evaluated: first at zero-T in a simple scalar model; then we will introduce a finite temperature to the theory; and lastly we will extend these techniques to the standard model with 2 Higgs doublets.

Let us take an example Lagrangian, for definiteness.

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - g\phi^n/n! \tag{3.4.22}$$

The tree-level potential is just the negative sum of the non-derivative terms in the Lagrangian - $V_0 = g\phi_c^n/N!$.

The one loop potential is the sum of all graphs with 1 loop, r vertices, and $(n-2)$ external lines on each vertex. Each internal line contributes a factor of $(k_m^2 + i\epsilon)^{-1}$. The external lines contribute $\phi_c^{(n-2)}$. Each vertex contributes $g/(n-2)!$, the $(n-2)!$ being a combinatoric since interchanging the external legs will have no effect. We must integrate over the momentum in the loop and introduce a $1/2r$ combinatoric factor since the loop is an r-sided polygon and has $2r$ symmetries under rotation and reflection.

Hence,

$$\begin{aligned} V_1 &= \frac{1}{n!}g\phi_c^n + i\int \frac{d^4k}{(2\pi)^4}\sum_r \frac{1}{2r}\left(\frac{g\phi_c^{n-2}/(n-2)!}{k^2+i\epsilon}\right)^r \\ &= \frac{1}{n!}g\phi_c^n + \frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\ln\left(1+\frac{g\phi_c^{n-2}}{(n-2)!k^2}\right) \end{aligned} \tag{3.4.23}$$

Note that, defining $U(\phi_c) \equiv g\phi_c^n/n!$, then this can be written

$$= U(\phi_c) + \frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\ln\left(1+\frac{U''(\phi_c)}{k^2}\right) \tag{3.4.24}$$

Although we have only demonstrated the form above to hold in one particular example, it does in fact hold true for any polynomial with no derivative terms. In generalising this form, it is also important to take note of the number of degrees of freedom in the model. In the 1-loop diagrams, there will be one loop counted for each degree of freedom of the fields, N. Hence, the general form of the 1-loop potential is

$$= U(\phi_c) + \frac{N}{2}\int \frac{d^4k}{(2\pi)^4}\ln\left(1+\frac{U''(\phi_c)}{k^2}\right) \tag{3.4.25}$$

This integral is divergent, so we must introduce a cut-off at $k^2 = \Lambda^2$. The resulting integration gives

$$V(\phi_c) = U(\phi_c) + \frac{N\Lambda^2}{32\pi^2}U'' + \frac{N(U'')^2}{64\pi^2}\left[\ln\left(\frac{U''}{\Lambda^2}\right) - \frac{1}{2}\right] + \text{constant} \tag{3.4.26}$$

The cut-off terms, $1/(32\pi^2)[U''\Lambda^2 - (U'')^2\ln(\Lambda)]$ must be absorbed into the renormalised parameters of the Lagrangian. Hence, U'' and $(U'')^2$ must contain only terms of the same order as those in U. This clearly limits U to be a quartic polynomial at most.

Re-writing the integral as

$$\frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\left\{\log\left(\frac{1}{k^2}\right) + \log\left(k^2+U''\right)\right\} \tag{3.4.27}$$

and, introducing finite-temperature effects as described in (3.3.19,3.3.20,3.3.21), we have

$$\text{constant} + \frac{N_bT}{2}\sum_{n=-\infty}^{\infty}\int \frac{d^3k}{(2\pi)^3}\log\left(\underline{k}^2 + (2\pi nT)^2 + U''\right) \tag{3.4.28}$$

for bosons, and

$$\text{constant} + \frac{N_f T}{2} \sum_{n=-\infty}^{\infty} \int \frac{d^3k}{(2\pi)^3} \log\left(\underline{k}^2 + (2n+1)\pi T^2 + M_f^2\right) \tag{3.4.29}$$

for fermions, where N_b (N_f) are the number of bosonic (fermionic) degrees of freedom, and M_f^2 is the sum of the fermion squared-masses. After a change of variables [30, 20], this can be written in the form

$$\begin{aligned} V_{T=0} \;+\; & \frac{N_b T^4}{2\pi^2} \int_0^\infty dx x^2 \log\left\{1 - \exp\left(-\sqrt{\frac{U''}{T^2} + x^2}\right)\right\} \\ - \; & \frac{N_f T^4}{2\pi^2} \int_0^\infty dx x^2 \log\left\{1 + \exp\left(-\sqrt{\frac{M_f^2}{T^2} + x^2}\right)\right\} \end{aligned} \tag{3.4.30}$$

where the first integral relates to the boson contribution, the second integral relates to the fermion contribution. Unfortunately, these integrals cannot be evaluated analytically and we are forced to resort to approximations.

To summarise what has been demonstrated :

- the effective potential was defined *exactly*
- V_{eff} was calculated to 1-loop *exactly*
- the expression for V_{eff} at 1-loop must be calculated using approximations.

The fact that V_{eff} has not only been approximated by expanding to first order in a power-series expansion, but also by approximating the calculation of that first order term, may seem disconcerting – ie., can the calculation be trusted with so many approximations? We will see shortly that this is indeed unsatisfactory and that higher order Feynman graphs must be included in the calculations (going beyond the 1-loop expansion.) However, these higher order corrections are introduced with a formalism very similar to the form developed above.

3.5 The effective potential with 2 Higgs-doublets

The results of the preceding sections can now be applied to the scalar sector with 2 Higgs-doublets. First, the most general form of the tree-level potential will be described, and then the expansion to one loop at high temperature. I will then incorporate the $W^\pm$ and Z^0 bosons and the heaviest fermions into the potential,

and I will demonstrate a need for higher order corrections to V_{eff}, introducing the most important of these.

3.5.1 The general tree-level potential

Consider a scalar sector with two compex scalar doublets, which can be defined as

$$\Phi_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi_4\end{pmatrix}, \qquad \Phi_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_5 + i\phi_6 \\ \phi_7 + i\phi_8\end{pmatrix} \tag{3.5.31}$$

The ϕ_i are real scalar fields.

In developing the most general form of V_0, we must bear in mind the requirements of renormalisability. This leads us, as stated already, to restrict ourselves to a quartic polynomial at most. The most general quartic polynomial is

$$\begin{aligned} V_0 \;=\; & -\mu_1^2\Phi_1^\dagger\Phi_1 - \mu_2^2\Phi_2^\dagger\Phi_2 + \lambda_1\left(\Phi_1^\dagger\Phi_1\right)^2 + \lambda_2\left(\Phi_2^\dagger\Phi_2\right)^2 \\ & +\lambda_3\left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_2^\dagger\Phi_2\right) + \lambda_4\left(\Phi_1^\dagger\Phi_2\right)\left(\Phi_2^\dagger\Phi_1\right) \\ & +\lambda_5\left(\Phi_1^\dagger\Phi_2\right)^2 + \lambda_6\left(\Phi_2^\dagger\Phi_1\right)^2 \\ & +\text{terms of odd-numbered order in } \Phi_i \\ & +\text{terms of form } \Phi_i^m\Phi_j^n \text{ where } m,n \text{ are odd} \end{aligned} \tag{3.5.32}$$

(The signs used here are conventional and will be convenient later.)

Note that there are terms "missing" here – there are no terms which are of odd-numbered order, and no terms containing odd-numbered powers of the Φ_i. Such terms are considered dangerous for the following reason [20].

In the two-doublet model, the Yukawa interaction of the $Q = -1/3$ quarks might look like

$$\mathcal{L}_Y = h_{ij}^1\overline{\Psi_i}\Psi_j\Phi_1 + h_{ij}^2\overline{\Psi_i}\Psi_j\Phi_2 \tag{3.5.33}$$

where i and j are the generation indices. The mass matrix is then

$$M_{ij} = h_{ij}^1 v_1 + h_{ij}^2 v_2 \tag{3.5.34}$$

Diagonalising M will not generally diagonalise h^1 and h^2. Hence, the Yukawa interactions are not flavour diagonal and flavour changing neutral currents (FCNC's)

are possible. These are not observed and would lead to large mass differences in the kaon system. The acceptable levels of FCNC's are tightly constrained by the small $K_L - K_S$ mass difference [20] which is observed, and hence it is usual to impose a discrete symmetry to avoid FCNC's. Let Φ_1 be the doublet which couples to the up-type fermions. [1] Then the Lagrangian is defined to be invariant under the translation

$$\Phi_1 \to -\Phi_1 \quad , \quad (u_i)_R \to -(u_i)_R \tag{3.5.35}$$

where the $(u_i)_R$ are the right-handed, up-type fields.

This disqualifies any terms with odd powers of Φ_i, such as $(\Phi_1^\dagger \Phi_2)$ or $(\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_2)$, thereby explaining the limited number of terms which appear in the general quartic given above.

Most of the coupling parameters can be constrained in the following manner. $\Phi_i^\dagger \Phi_i$ must be real, and the potential must also be real, so we can constrain $\mu_1^2, \mu_2^2, \lambda_1, \lambda_2, \lambda_3$ to be real also. Since $\Phi_i^\dagger \Phi_j = (\Phi_j^\dagger \Phi_i)^\dagger$, then λ_4 is also constrained to be real. The remaining couplings can be complex, *a priori*, but can still be constrained further as shall be seen shortly.

Consider the terms with couplings λ_5, λ_6. Since V_0 must be real, then the sum of the $\lambda_{5,6}$ terms must also be real. Applying the properties of complex conjugates – in particular, $(ab)^* = a^* b^*$ – then we must have $\lambda_5 = \lambda_6^*$. However, at this point, λ_5 may still be complex.

As discussed earlier, in chapter 2, it is considered desirable to introduce a further source of CP-violation, in addition to the Kobayashi-Maskawa couplings. In order to do this, a term is introduced which breaks the discrete symmetry imposed above to avoid FCNC's. The terms introduced are of the form

$$-\mu_3^2 \Phi_1^\dagger \Phi_2 - \mu_4^2 \Phi_2^\dagger \Phi_1 \tag{3.5.36}$$

Note that, although these terms break the discrete symmetry explicitly, they do not introduce FCNC's at tree-level. Hence, they are said to break the symmetry

[1] As mentioned in the introductory chapter, it is attractive to couple the up-type and down-type quarks each to only one of the doublets. This is discussed in section 3.5.3.

softly and the FCNC's introduced are suppressed since they only occur at higher orders [20]. The quartic terms which would also break the discrete symmetry (with odd-numbered powers of Φ_i, as described above) would introduce FCNC's at tree-level – these would not be suppressed and so they are not acceptable. Hence, I do not allow such terms in the potential.

Following the same line of argument used for $\lambda_{5,6}$, it is easy to demonstrate that

$$\mu_3^2 = \left(\mu_4^2\right)^* \tag{3.5.37}$$

Again, *a priori*, this still allows μ_3^2 to be complex.

Consider the violation of CP-conservation in the potential. The CP-transformation of scalar fields has the form

$$\Phi_i \to e^{i\theta_i}\Phi_i^* \tag{3.5.38}$$

Hence, the terms $\Phi_i^\dagger\Phi_i$ are trivially invariant, and it follows that the only possible sources of CP-violation in the potential described above are the terms

$$\lambda_5\left(\Phi_1^\dagger\Phi_2\right)^2 + \lambda_5^*\left(\Phi_2^\dagger\Phi_1\right)^2 - \mu_3^2\left(\Phi_1^\dagger\Phi_2\right) - (\mu_3^2)^*\left(\Phi_2^\dagger\Phi_1\right) \tag{3.5.39}$$

Without loss of generality, we can choose λ_5 to be real. In this case, by considering the CP-transformation of these terms and comparing with the expression given above, some simple manipulation demonstrates that CP-invariance requires μ_3^2 to be real or pure imaginary. In other words, by insisting on λ_5 real, we introduce CP-violation by allowing complex values of μ_3^2.

Hence, without loss of generality, we have constrained all of the coupling parameters in the tree-level potential to be real, except for the CP-violating coupling μ_3^2, which we will allow to be complex.

So, our definitive general form for the tree-level potential is

$$
\begin{aligned}
V_0 \;=\; & -\mu_1^2\left(\Phi_1^\dagger\Phi_1\right) - \mu_2^2\left(\Phi_2^\dagger\Phi_2\right) + \lambda_1\left(\Phi_1^\dagger\Phi_1\right)^2 + \lambda_2\left(\Phi_2^\dagger\Phi_2\right)^2 \\
& +\lambda_3\left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_2^\dagger\Phi_2\right) + \lambda_4\left(\Phi_1^\dagger\Phi_2\right)\left(\Phi_2^\dagger\Phi_1\right) \\
& +\frac{\lambda_5}{2}\left(\left(\Phi_1^\dagger\Phi_2\right)^2 + \left(\Phi_2^\dagger\Phi_1\right)^2\right) \\
& -\mu_3^2\left(\Phi_1^\dagger\Phi_2\right) - \left(\mu_3^2\right)^*\left(\Phi_2^\dagger\Phi_1\right) \qquad (3.5.40)
\end{aligned}
$$

with $\mu_1^2, \mu_2^2, \lambda_i$ all real, and μ_3^2 complex.

3.5.2 The potential at 1-loop

The 1-loop expansion is, from eqn (3.4.30),

$$
\begin{aligned}
V_1 = V_{T=0} \;+\; & N_b\frac{T^4}{2\pi^2}\int_0^\infty dx x^2 \log\left\{1 - \exp\left(-\sqrt{\frac{U''}{T^2} + x^2}\right)\right\} \\
-\; & N_f\frac{T^4}{2\pi^2}\int_0^\infty dx x^2 \log\left\{1 + \exp\left(-\sqrt{\frac{M_f^2}{T^2} + x^2}\right)\right\} \qquad (3.5.41)
\end{aligned}
$$

where the first integral is the bosonic contribution, the second is the fermionic contribution, and U_b'', U_f'' are the second derivatives of the bosonic and fermionic parts of the tree level potential. [2]

Actually, the second derivative of the tree level potential is a matrix of derivatives of the form $\partial^2/(\partial\phi_i\partial\phi_j)$, and its eigenvalues represent the mass spectrum of the model. Hence, the matrix is known as the **mass matrix**. So, we are more correctly interested in the trace of the mass matrix,

$$
\begin{aligned}
U'' \;&\rightarrow\; \mathrm{Tr}\left(\frac{\partial^2}{\partial\phi_i\partial\phi_j}V_0\right) \\
&=\; \frac{\mathrm{Tr}\,(M_0^2)}{T^2} \\
&=\; \sum_i m_i^2 \qquad (3.5.42)
\end{aligned}
$$

where M_0^2 is the tree level mass matrix and m_i^2 are its eigenvalues. The expressions above (3.5.41) were determined in the case of a simple model with one scalar field

[2] Note that there are no fermion fields in the tree level potential described in the previous section. However, as mentioned above, I will eventually introduce the heaviest fermions into the effective potential and hence it will be useful here to consider both cases.

– the integral represents the integral over the loop momentum for this one scalar field. However, we now have 8 scalar fields and hence it is more correct to write

$$V_1 = V_{T=0} + \sum_i (N_b)_i \frac{T^4}{2\pi^2} \int_0^\infty dx x^2 \log\left\{1 - \exp\left(-\sqrt{\frac{m_i^2}{T^2} + x^2}\right)\right\} - \sum_j (N_f)_j \frac{T^4}{2\pi^2} \int_0^\infty dx x^2 \log\left\{1 + \exp\left(-\sqrt{\frac{m_j^2}{T^2} + x^2}\right)\right\} \quad (3.5.43)$$

where the summation over i (j) is the summation over the boson (fermion) fields, and $(N_b)_i, (N_f)_j$ are the numbers of degrees of freedom associated with each mass eigenvalue. I will drop the summation for the purposes of clarity in the derivation which follows, and re-introduce it in the final result. Note that the mass matrix will play an important role in the higher order corrections discussed in the next section.

I will now evaluate V_1 using a high-temperature expansion. For the moment I will simply assume that the high-T approximation holds. Having developed the expansion, I will go on to justify its use.

For small m_i/T (ie., for temperatures much larger than the mass spectrum of the model), V_1 can be expanded in a Taylor series –

$$V_1 = V_1|_{y^2=0} + \left.\frac{\partial V_1}{\partial y^2}\right|_{y^2=0} y^2 + \left.\frac{\partial^2 V_1}{\partial y^4}\right|_{y^2=0} y^4 + \ldots \quad (3.5.44)$$

where $y^2 = M_0^2/T^2$. Note that the Taylor series is actually an expansion in m_i/T.

Let us consider the bosonic contributions first. The first term is

$$\begin{aligned} V_1|_{y^2=0} &= N_b \frac{T^4}{2\pi^2} \int_0^\infty dx x^2 \ln\left(1 - e^{-x}\right) \\ &= -N_b \frac{\pi^2 T^4}{90} \end{aligned} \quad (3.5.45)$$

Although this term is temperature dependent, its effect can be removed by redefinition of zero potential and hence it will generally be ignored hereafter.

The second term is

$$\begin{aligned} \left.\frac{\partial V_1}{\partial y^2}\right|_{y^2=0} y^2 &= y^2 N_b \frac{T^4}{4\pi^2} \int_0^\infty dx \frac{x^2}{(x^2+y^2)^{\frac{1}{2}}} \frac{1}{\exp\left(\sqrt{x^2+y^2}\right) - 1} \\ &= y^2 N_b \frac{T^4}{4\pi^2} \int_0^\infty dx \frac{x}{e^x - 1} \end{aligned} \quad (3.5.46)$$

Hence,

$$\left.\frac{\partial V_1}{\partial y^2}\right|_{y^2=0} y^2 = N_b \frac{T^2}{24} M_0^2 \tag{3.5.47}$$

Evaluation of the next term in (3.5.44) is considerably more complicated. Consider the derivative part of the third term –

$$\frac{\partial^2 V_1}{\partial y^4} = -N_b \frac{T^4}{8\pi^2} \int_0^\infty \frac{dx}{(x^2+y^2)^{\frac{1}{2}}} \frac{1}{\left[exp\left((x^2+y^2)^{\frac{1}{2}}\right)-1\right]} \tag{3.5.48}$$

At $x = y = 0$, the integrand has a pole and a logarithmic singularity. Dolan and Jackiw [30] overcome these difficulties by introducing a regularisation factor $x^{-\epsilon}$ into the integral and then allowing $\epsilon \to 0$. The resulting expression is

$$\begin{aligned} \left.\frac{\partial^2 V_1}{\partial y^4}\right|_{y^2=0} y^4 &= -N_b \frac{T^4 y^3}{12\pi} - N_b \frac{T^4 y^4}{64\pi^2} \log\left(y^2\right) + \ldots \\ &= -N_b \frac{T}{12\pi} M_0^3 - N_b \frac{1}{64\pi^2} M_0^4 \log\left(\frac{M_0^2}{T^2}\right) + \ldots \end{aligned} \tag{3.5.49}$$

Note that the M_0^3 term is negative – it is this term which allows a secondary minimum to develop at non-zero $\langle \phi_c \rangle$.

Hence, combining the first 3 terms in the expansion, and re-introducing the summation over field indices, gives

$$\begin{aligned} V_1(T) &= V_b(T=0) - \frac{N_b \pi^2 T^4}{90} + \frac{N_b T^2}{24} \mathrm{Tr}\left([M_0]_b^2\right) - \frac{N_b T}{12\pi} \mathrm{Tr}\left([M_0]_b^3\right) \\ &\quad - \frac{N_b}{64\pi^2} \mathrm{Tr}\left([M_0]_b^4\right) \mathrm{Tr}\left(\log \frac{([M_0]_b^2)}{T^2}\right) + \ldots \end{aligned} \tag{3.5.50}$$

where $V_b(T=0)$ and $[M_0]_b$ are the relevant bosonic parts, and N_b is the number of bosonic fields. Following essentially the same process for the fermion contribution returns

$$\begin{aligned} V_1(T) &= V_f(T=0) - \frac{7N_f \pi^2 T^4}{720} + \frac{N_f T^2}{48} \mathrm{Tr}\left([M_0]_f^2\right) \\ &\quad + \frac{N_f}{64\pi^2} \mathrm{Tr}\left([M_0]_f^4\right) \mathrm{Tr}\left(\log \frac{([M_0]_f^2)}{T^2}\right) + \ldots \end{aligned} \tag{3.5.51}$$

where the subscript f refers to fermionic parts.

Note that there is no M_0^3 term from the fermions – they are not involved in generating the symmetry breaking vacuum. The $-TM^3$ term for bosons is the

dominant term arising from the infrared divergence of the integral in (3.5.48). It is easier to understand why there is no equivalent fermion term by considering the prescription for working with field theories at finite temperature. As stated earlier, the following substitutions should be made –

$$k_0 \to 2\pi nT(\text{bose}), \quad k_0 \to (2n+1)\pi T(\text{fermi}), \quad \int dk_0 \to i2T\sum_n \tag{3.5.52}$$

Hence,

$$\begin{aligned} \int \frac{dk_0}{k_0} &\to i2T\sum_n \frac{1}{2n\pi T} \quad \text{(bosons)} \\ \int \frac{dk_0}{k_0} &\to i2T\sum_n \frac{1}{(2n+1)\pi T} \quad \text{(fermions)} \end{aligned} \tag{3.5.53}$$

and so the bosonic integrals have a pole for the zero frequency modes ($n = 0$), while the fermions never have a pole.

Hence, the pole leads to an infrared divergence and results in the $-TM^3$ term due to the zero frequency, or long range, bosonic fields. The fermions do not display the same behaviour and so the first order $EWPT$ derives purely from long range bosonic interactions.

3.5.3 Including $W^{\pm}, Z^0$ and the heaviest fermions

Since the $W^{\pm}$ and Z^0 bosons are the weak force carriers, the critical temperature of the EWPT is likely to be of order $M_{W/Z}$. Hence, Feynman diagrams with $W^{\pm}$ and Z^0 bosonic loops are likely to be important in the effective potential near the phase transition. Also, the top quark mass is now known to be of order 175GeV, and hence the top quark fermionic loops are also likely to make an important contribution. For the sake of thoroughness, I will also include the effects of the next heaviest fermions, the bottom quark and the τ lepton, although their contributions will be very much smaller than their heavier counterparts.

There is an important point to note in the way which I have chosen to couple the quarks to the scalar doublets. There are good reasons to consider coupling the up-type and down-type quarks differently. By coupling up-type only to Φ_1,

say, and down-type to Φ_2, then the differences between masses of the up and down-type quarks can be explained in terms of the two different VEV's of the scalar doublets. This is an attractive feature which is also a natural requirement of the minimal supersymmetric model. Hence, we will join the up-type quarks to Φ_1 and the down-type to Φ_2. [3]

The masses of these particles are derived from the VEV of the scalar field and have the forms

$$\begin{aligned} m_W = gv_0/2 \quad , \quad m_Z = \sqrt{g^2 + g'^2}v_0/2 \\ m_t = g_t v_1/\sqrt{2} \quad , \quad m_b = g_b v_2/\sqrt{2} \\ m_\tau = g_\tau v_2/\sqrt{2} \end{aligned} \tag{3.5.54}$$

where v_0 is the overall VEV of the scalar fields; v_1 and v_2 are the VEV's of Φ_1 and Φ_2; g and g' are the $SU(2)$ and $U(1)$ gauge couplings; and the g_i are the appropriate Yukawa couplings. [4]

There is one loop diagram for each degree of freedom in the bosonic and fermionic fields. For each of the EW vector bosons there are 3 degrees of freedom corresponding to the 3 spin states. For each of the quarks, there are 12 degrees of freedom – particle/anti-particle; spin up/down; and 3 colours; giving $2\times2\times3 = 12$. And for the leptons, there are 4 degrees of freedom – particle/anti-particle and spin up/down.

Hence, the contributions of the $W^\pm, Z^0$, and 3 heaviest fermions to the trace of the zero-temperature mass matrix are

$$6m(v)_W^2 + 3m(v)_Z^2 + 12[m(v)_t^2 + m(v)_b^2] + 4m(v)_\tau^2 \tag{3.5.55}$$

where the $m(v)_i$ are the masses evaluated at the VEV of the scalar fields.

[3] It is more common in the literature to couple the down-type quarks to Φ_1. However, for historical reasons, I have chosen to couple up-type to Φ_1. I work consistently with this convention throughout the thesis and still define the angle β – the phase between the VEV's of the two doublets – as $\tan(\beta) = v_{up-type}/v_{down-type}$, thereby maintaining consistency with the conventional definition of β.

[4] Recall that the up-type fermions couple to Φ_1, with VEV v_1, and the down-type to Φ_2, with VEV v_2.

3.5.4 Higher order corrections

Since the early days of the study of symmetry restoration, it has been realised [31, 30] that there are important corrections required when using the 1-loop expansion of the effective potential. Consider again eqn(3.5.41) – if U'' is negative, then the integral becomes complex and so too does $V_1(T)$. Since the potential cannot be complex, this indicates that the 1-loop approximation is not adequate under these conditions. In terms of the high-T expansion, $U'' < 0$ is equivalent to having a negative eigenvalue in M_0^2, which will result in a complex eigenvalue for M_0^3 – so the complex term will be proportional to M^3T.

The critical temperature can be defined as the highest temperature at which one of the eigenvalues of the mass matrix (curvature matrix) becomes negative at the origin, turning the origin into a saddle point from which the fields may run in the direction of the negative curvature towards the new minimum. But if the potential cannot be evaluated reliably under exactly these conditions, then the accuracy of the T_c calculation is also called into question.

However, it has been shown by Takahashi [38, 39] that the complex term proportional to m^3T is exactly cancelled by infrared terms in the ring diagrams which are the next to leading order terms beyond the 1-loop potetial. (The inclusion of these terms will be discussed below.) [5]

These arguments can be viewed in various other forms [30, 31, 38, 9, 35], usually as considerations of infrared divergence problems. The dominant terms in the 1-loop expansion result from infrared divergences and it is well recognised that finite temperature field theories with massless particles suffer from severe infrared problems. But near the origin, the W and Z bosons are very light, and for temperatures near T_c so too is at least one of the Higgs bosons (which is equivalent to the argument above that U'' becomes negative at or near T_c, depending on definition of critical temperature.) Hence, the higher order infrared terms (the ring diagrams) should be included in the effective potential.

It has been shown that the effect of the ring diagrams is to introduce a tem-

[5] As will be mentioned later in the discussion of numerical work, I will use this argument to ignore complex terms arising in evaluation of the effective potential and critical temperature.

perature dependent effective mass [9, 38]. In order to clearly see how this works, let us consider the potential of one self-interacting bosonic field –

$$V(\phi) = -\frac{\mu^2}{2}\phi^2 + \frac{\lambda\phi^4}{4!} \tag{3.5.56}$$

Temperature dependence is introduced into the 1-loop potential, as already seen, in the form

$$V(\phi,T) \;=\; V_0 + \frac{M_0^2T^2}{24} - \frac{M_0^3T}{12\pi} + \ldots \tag{3.5.57}$$

where M_0^2 is the second derivative of the tree level potential, namely $-\mu^2+\lambda\phi^2/2$. However, expanding (3.5.57) and considering the second order terms gives

$$V(\phi,T) \;=\; \frac{\phi^2}{2}\left(-\mu^2 + \frac{\lambda T^2}{24}\right) + \ldots \tag{3.5.58}$$

and so we see that, at finite temperature, there is an effective dimensional coupling $(-\mu^2 + \lambda T^2/24)$. Hence, the effective mass matrix also carries a temperature dependency but this is not reflected in (3.5.57) – the expansion is not self-consistent.

By defining $M^2(T)$ as the second derivative of the temperature-dependent potential, $V''(\phi,T)$, and substituting $M^2(T)$ for M_0^2 into the M^3T term in (3.5.57), the expansion is made consistent. As already stated, it has been demonstrated [9, 38] that this is equivalent to including the ring diagrams (or "daisy diagrams") in the expansion – see figure 3.1.

Hence, the resulting form of the expansion is

$$V = V_0 + \frac{T^2}{24\pi^2}\mathrm{Tr}M_0^2 - \frac{T}{12\pi}\mathrm{Tr}M^3(T) + \ldots \tag{3.5.59}$$

where

$$\begin{aligned}
\left[M_0^2\right]_{ij} &= \frac{\partial^2 V_0}{\partial\phi_i\partial\phi_j}, \\
\left[M^2(T)\right]_{ij} &= \frac{\partial^2 V_T}{\partial\phi_i\partial\phi_j} \\
\text{where} \qquad V_T &= V_0 + \frac{T^2}{24\pi^2}\mathrm{Tr}M_0^2,
\end{aligned} \tag{3.5.60}$$

and so on.

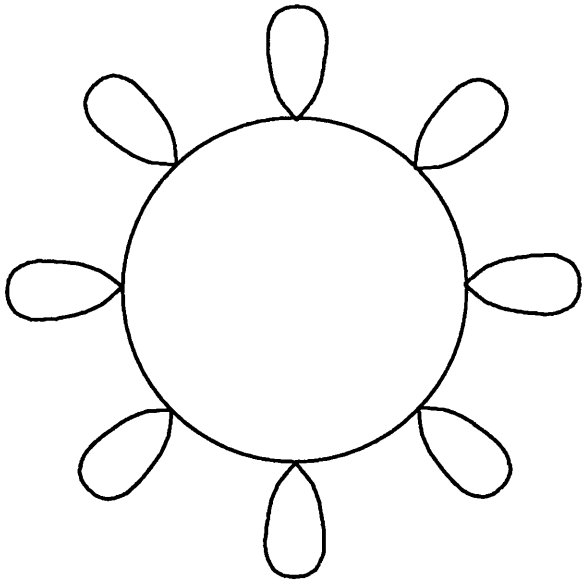

Figure 3.1: A typical ring (or "daisy") diagram. These diagrams form the next-to-leading order contribution to the loop expansion of the effective potential.

A thorough examination of the question of higher-order corrections is found in a paper by Dine, *et al.*[35] In particular, they identified an important correction to the m^3T term which had gone unnoticed until that point.

From expressions (3.5.55,3.5.50), the contribution of the $W^{\pm}$ boson is

$$2 \times 3 \times \left(-\frac{\pi^2 T^4}{90} + \frac{m_W^2(v)T^2}{24} - \frac{m_W^3(v)T}{12\pi} + \ldots \right) \tag{3.5.61}$$

The numerical factors of 2 and 3 correspond to the degrees of freedom – two charge states, and two transverse and one longitudinal polarisations.

The cubic term arises due to the zero frequency modes when calculating the loop diagrams. The vector field propagator at zero frequency looks like

$$\begin{aligned} D_{00}(\underline{k}) &= \frac{1}{\underline{k}^2 + m_W(v)^2} \\ D_{ij}(\underline{k}) &= \frac{\delta_{ij} - (k_i k_j)/k^2}{k^2 + m_W(v)^2} \end{aligned} \tag{3.5.62}$$

where D_00 is the longitudinal or Coulomb propagator, and the D_{ij} are the transverse propagators. However, the Coulomb field at zero frequency develops the Debye mass, $m_D^2 \sim g^2T^2$ – this is analogous to the development of an effective

mass, due to screening effects, in photons propagating in a plasma. The result is a correction to the D_{00} propagator –

$$D_{00}(\underline{k} \to 0) = \frac{1}{\underline{k}^2 + m_W(v)^2 + m_D^2} \tag{3.5.63}$$

The Debye squared-mass is much greater than the squared-mass of the W-boson, so the propagator is heavily suppressed and the contribution of the longitudinal mode to the loop diagram effectively vanishes. The same situation exists for the Z-boson, and hence its longitudinal contribution also disappears.

The zero frequency transverse modes, on the other hand do not develop such a mass [36]. Hence, the contribution of the W and Z bosons to the cubic term in the 1-loop effective potential is actually reduced by a factor of 2/3, and becomes

$$2 \times -\frac{T}{12\pi}\left(2m_W^3(v) + m_Z^3(v)\right) \tag{3.5.64}$$

where the first numerical factor of 2 represents the two transverse modes.

The effect of this correction is to reduce the contribution to the $-TM^3$ term, thereby reducing the strength of the phase transition. This in turn lowers the Higgs mass upper bound, leaving electroweak baryogenesis impossible in the minimal Standard Model.

3.6 Summary of V_{eff} with two Higgs doublets

For reference purposes, and for clarity, the results of this chapter as applied to the two Higgs-doublet model are reproduced here.

I described the most general form of the tree-level potential, subject to the following conditions:

– the imposition of a discrete symmetry , $\Phi_1 \to -\Phi_1, (u_i)_R \to -(u_i)_R$, where the $(u_i)_R$ are the right-handed, up-type fermion fields. This symmetry avoids experimentally unobserved flavour changing neutral currents (FCNC's).

– the inclusion of an explicitly CP-violating term, $-\mu_3^2(\Phi_1^\dagger\Phi_2) + h.c.$, which softly breaks the discrete symmetry above. This will not introduce FCNC's at tree-level, and hence the FCNC's can be suppressed in agreement with the experimental limits.

The resulting expression is

$$\begin{aligned} V_0 &= -\mu_1^2\left(\Phi_1^\dagger\Phi_1\right) - \mu_2^2\left(\Phi_2^\dagger\Phi_2\right) + \lambda_1\left(\Phi_1^\dagger\Phi_1\right)^2 + \lambda_2\left(\Phi_2^\dagger\Phi_2\right)^2 \\ &\quad +\lambda_3\left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_2^\dagger\Phi_2\right) + \lambda_4\left(\Phi_1^\dagger\Phi_2\right)\left(\Phi_2^\dagger\Phi_1\right) \\ &\quad +\frac{\lambda_5}{2}\left[\left(\Phi_1^\dagger\Phi_2\right)^2 + \left(\Phi_2^\dagger\Phi_1\right)^2\right] \\ &\quad -\mu_3^2\left(\Phi_1^\dagger\Phi_2\right) - \left(\mu_3^2\right)^*\left(\Phi_2^\dagger\Phi_1\right) \end{aligned} \tag{3.6.65}$$

with $\mu_1^2, \mu_2^2, \lambda_i$ all real, and μ_3^2 complex.

At high temperature, defined by $m_i(T)/T < 1$, where $m_i(T)$ are the mass eigenvalues of the scalar fields (this will be discussed in more detail in chapter 6), the 1-loop expansion of the effective potential can be approximated by

$$\begin{aligned} V_1(T) &= V_0 - \frac{N_b\pi^2T^4}{90} - \frac{7N_f\pi^2T^4}{720} \\ &\quad +\frac{N_bT^2}{24}\mathrm{Tr}\left([M_0]_b^2\right) + \frac{N_fT^2}{48}\mathrm{Tr}\left([M_0]_f^2\right) \\ &\quad -\frac{N_bT}{12\pi}\mathrm{Tr}\left([M_0]_b^3\right) + \ldots \end{aligned} \tag{3.6.66}$$

where the subscripts b, f refer to bosonic and fermionic parts, and the mass-matrix M_0 is defined as $[M_0]_{ij} = \partial^2 V_0/\partial\phi_i\phi_j$.

The effects of the vector bosons and the heaviest fermions (top and bottom quarks, and the tau lepton) were taken into account by including their masses in the trace of the mass-matrix. Hence, the $T^2\mathrm{Tr}(M^2)$ term has the following form

$$\begin{aligned} &\frac{N_bT^2}{24}\mathrm{Tr}\left([M_0]_b^2\right) + \frac{N_fT^2}{48}\mathrm{Tr}\left([M_0]_f^2\right) \\ &= \frac{T^2}{24}\left\{-4(\mu_1^2+\mu_2^2) + (6\lambda_1 + 2\lambda_3 + \lambda_4 + 3g_t^2 + \frac{9}{4}g^2 + \frac{3}{4}g'^2)\sum_{i=1}^{4}\phi_i^2\right. \\ &\quad \left. + (6\lambda_2 + 2\lambda_3 + \lambda_4 + 3g_b^2 + g_\tau^2 + \frac{9}{4}g^2 + \frac{3}{4}g'^2)\sum_{i=5}^{8}\phi_i^2\right\} \end{aligned} \tag{3.6.67}$$

I also introduced a higher order correction to the potential by defining a temperature-dependent mass matrix, $M(T)^2$, viz:

$$\left[M(T)^2\right]_{ij} = \frac{\partial^2}{\partial\phi_i\partial\phi_j}\left(V_0 + \frac{N_bT^2}{24}\mathrm{Tr}\left([M_0]_b^2\right) + \frac{N_fT^2}{48}\mathrm{Tr}\left([M_0]_f^2\right)\right) \tag{3.6.68}$$

This matrix is introduced at the $T\mathrm{Tr}(M^3)$ term, and is equivalent to adding ring (or daisy) diagrams to the 1-loop effective potential. The resulting expression is

$$-\frac{N_b T}{12\pi}\mathrm{Tr}\left([M_0]_b^3\right) = -\frac{T}{12\pi}\left\{\sum_{i=1}^{8}\left[m_i^2(T)\right]^{\frac{3}{2}} + 4\left[\frac{g^2}{4}\sum_{i=1}^{8}\phi_i^2\right]^{\frac{3}{2}} + 2\left[\frac{g^2+g'^2}{4}\sum_{i=1}^{8}\phi_i^2\right]^{\frac{3}{2}}\right\} \tag{3.6.69}$$

where the $m_i^2(T)$ are the eigenvalues of $M(T)^2$, which generally cannot be derived analytically and must be determined numerically. Note that the fermion masses do **not** appear here since the fermion fields do not develop a linearly T-dependent term.

The full forms of the mass-matrices M_0^2 and $M(T)^2$ can be found in appendix B.

Chapter 4

The coupling parameter space

Having introduced the most important tool in the study of the EWPT – namely, the effective potential – and developed the most general form of the potential in a two-Higgs-doublet scalar sector, I will now consider the parameter space of the model.

There are 9 distinct coupling parameters in the general form of the effective potential, as demonstrated in the tree-level potential V_0 –

$$\begin{aligned} V_0 &= -\mu_1^2\Phi_1^\dagger\Phi_1 - \mu_2^2\Phi_2^\dagger\Phi_2 + \lambda_1\left(\Phi_1^\dagger\Phi_1\right)^2 + \lambda_2\left(\Phi_2^\dagger\Phi_2\right)^2 \\ &\quad +\lambda_3\left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_2^\dagger\Phi_2\right) + \lambda_4\left|\Phi_1^\dagger\Phi_2\right|^2 + \frac{\lambda_5}{2}\left\{\left(\Phi_1^\dagger\Phi_2\right)^2 + \left(\Phi_2^\dagger\Phi_1\right)^2\right\} \\ &\quad -\mu_3^2\Phi_1^\dagger\Phi_2 - \mu_3^{2*}\Phi_2^\dagger\Phi_1 \end{aligned} \tag{4.0.1}$$

However, there are only 8 free parameters since there is one experimental constraint which must be imposed, namely

$$v_0 = 246 \quad \text{GeV} \tag{4.0.2}$$

Hence, there is an 8-dimensional parameter space to be investigated.

I will not be interested in the entire parameter space. It is reasonable to expect that there are regions in the parameter space where the effective potential does not satisfy the most basic physical requirements (such as being bounded below), and hence is not acceptable. We will see in this chapter that such regions do exist. Hence it is reasonable to restrict attention to the region where the effective

potential is physically acceptable. Such a region will be bounded by imposing physical constraints on the effective potential. This region will form the basis space for further study, and it will be referred to as "the basis space" henceforth.

Within the basis space, I am particularly interested in those regions which support models of electroweak baryogenesis at the $EWPT$ (ie., first order phase transition) and preserve any baryon asymmetry after the phase transition is complete. Such a region will be bounded by imposing the requirements of baryogenesis and BAU-preservation at the $EWPT$ – namely a strongly first order phase transition – and the limits on sphaleron mass at the critical temperature of the phase transition. The volume of this region relative to the volume of the basis space, and the characteristics of the region, are the primary interests of this study. This region will be referred to as the "baryogenesis region" or "baryon preserving region".

In this chapter, I will consider appropriate representations of the parameter set. In particular, it is necessary to derive the relationship between the coupling parameters in eqn(4.0.1) and the VEV of the scalar fields, v_0, in order to impose the constraint (4.0.2). I will develop physical constraints on the coupling parameters in order to identify the boundaries of the basis space. I will also impose the requirements of electroweak baryogenesis and sphaleron mass limits in order to locate the baryogenesis region.

4.1 Constraining the basis parameter space

Let us first develop some of the notation to be used in describing the basis parameter space. Then let us consider symmetries in the effective potential which will allow us to restrict the region of study without loss of generality. I will follow this with an argument which constrains the position of the global minimum of the effective potential – ie., constrains the VEV's of the scalar fields.

4.1.1 Notation

The parameter set, as it appears in eqn(4.0.1) above, does not allow the direct imposition of the condition in eqn(4.0.2). The general form of the potential is a complicated expression and the overall VEV of the scalar fields is dependent on the VEV's of the 8 scalar fields $\phi_{1...8}$ as defined in eqn. (3.5.31) However, the number of field variables required in the potential can be reduced by making appropriate rotations. Φ_1 can always be rotated into the the simple form shown here, but then Φ_2 will still generally have 4 non-zero fields, thus–

$$\langle\Phi_1\rangle = \begin{pmatrix} 0 \\ v_3 \end{pmatrix} \qquad \langle\Phi_2\rangle = \begin{pmatrix} v_5 + iv_6 \\ v_7 + iv_8 \end{pmatrix} \tag{4.1.3}$$

However, it will be demonstrated later in this chapter that it is a necessary condition in the vacuum state that the VEV's of the two doublets be parallel. Hence we can generally reduce the number of fields we are working with to only those 3 which develop non-zero VEV's, as shown here –

$$\langle\Phi_1\rangle = \begin{pmatrix} 0 \\ v_3 \end{pmatrix} \qquad \langle\Phi_2\rangle = \begin{pmatrix} 0 \\ v_7 + iv_8 \end{pmatrix} \tag{4.1.4}$$

Note that the variables v_i have been labelled according to their corresponding scalar fields and will be used specifically to refer to the VEV's of those fields. I will also write the potential in terms of the fields themselves, using the 3 variables σ_i, defined below

$$\Phi_1 = \begin{pmatrix} 0 \\ \sigma_3 \end{pmatrix} \qquad \Phi_2 = \begin{pmatrix} 0 \\ \sigma_7 + i\sigma_8 \end{pmatrix} \tag{4.1.5}$$

I will also often use the notation $(\sigma_1, \sigma_2, \theta)$, where

$$\Phi_1 = \begin{pmatrix} 0 \\ \sigma_1 \end{pmatrix} \qquad \Phi_2 = e^{i\theta}\begin{pmatrix} 0 \\ \sigma_2 \end{pmatrix} \tag{4.1.6}$$

and this has a corresponding notation for the VEV's, (v_1, v_2, θ_0). It should be clear by the context which notation is being used and I will not normally explain which is in use.

Another parameter which is commonly used is the phase angle between the two scalar doublets. This is defined as

$$\tan(\beta) = \frac{\sigma_2}{\sigma_1} \tag{4.1.7}$$

with a corresponding vacuum expectation value of $\tan(\beta_0) = v_2/v_1$.

4.1.2 Symmetries in the effective potential

There are several symmetries present in the general form of the scalar potential. These symmetries will allow us to place restrictions on the ranges of the parameter space search, and help to develop the details of constraints on the basis parameter space. So, let us begin with the tree level potential in the following form —

$$\begin{aligned} V_0 &= -\frac{\mu_1^2}{2}\sigma_1^2 - \frac{\mu_2^2}{2}\sigma_2^2 + \frac{\lambda_1}{4}\sigma_1^4 + \frac{\lambda_2}{4}\sigma_2^4 + \frac{1}{4}[\lambda_3 + \lambda_4 + \lambda_5\cos(2\theta)]\,\sigma_1^2\sigma_2^2 \\ &\quad -\frac{1}{2}\left[\mu_R^2\cos(\theta) - \mu_I^2\sin(\theta)\right]\sigma_1\sigma_2 \end{aligned} \tag{4.1.8}$$

I will state the symmetries at tree level without proof since they are fairly obvious and simple to check. I will then discuss how the symmetries extend to higher order terms in the effective potential.

V_0 is invariant under

$$\begin{aligned} &\sigma_1 \to -\sigma_1, \quad \sigma_2 \to -\sigma_2 \\ &\sigma_i \to -\sigma_i, \qquad\qquad \theta \to \pi + \theta \\ &\qquad\qquad \mu_R^2 \to -\mu_R^2, \quad \theta \to \pi - \theta \\ &\qquad\qquad \mu_I^2 \to -\mu_I^2, \quad \theta \to -\theta \end{aligned} \tag{4.1.9}$$

where σ_i represents σ_1 *or* σ_2.

It is not obvious, *a priori*, that these symmetries will also hold in the higher order terms, since each of these terms arises from derivatives of all of the lower order terms in the effective potential and taking derivatives will change the order of σ_i^n terms, for example. However, since the symmetries hold in the tree level potential then they must also hold in each term of the second derivative matrix (because the curvature of the potential must be of the same size and sign at any two points related by any of the symmetries above.) Since the symmetries will hold term by term in the curvature matrix, they must also hold in the trace of the matrix and hence we see that the symmetries also hold at the $V_0 + T^2/24\mathrm{Tr}\,[M^2]$ level. We can then follow a similar line of argument to show that these symmetries

will hold at the $V_0 + T^2/24\mathrm{Tr}\,[M^2] - T/(12\pi)\mathrm{Tr}\,[M^3]$ level, and so on.

By combining the symmetries above, it is always possible to set the signs of v_1, v_2, μ_R^2 and μ_I^2 by adjusting the value of θ and without loss of generality. In order to simplify the constraints, I will insist that

$$v_1, v_2, \mu_R^2, \mu_I^2 \geq 0 \tag{4.1.10}$$

by convention. Note that this implies we must have β_0 (the phase angle between the VEV's of the first and second doublets) in the first quadrant.

What is not so clear is the resulting effect of the convention in eqn.(4.1.10) on θ_0. I will now explain that, **under this convention**, θ_0 must lie in the fourth quadrant.

4.1.3 Restricting θ_0

Eqn.(4.1.8) above can be written as two parts, one of which is dependent on θ, the other independent of θ. Let us define the θ–dependent part to be V_θ. Then we have

$$V_\theta = \frac{1}{4}\Big\{\lambda_5\sigma_1\sigma_2\cos(2\theta) - 2\mu_R^2\cos(\theta) + 2\mu_I^2\sin(\theta)\Big\}\sigma_1\sigma_2 \tag{4.1.11}$$

Naively, one might suggest that we can minimise the potential by simply minimising both the θ–dependent and θ–independent parts of the potential individually. However, they are not independent of one another since both parts are dependent on (σ_1, σ_2), and hence this naive approach will generally fail. On the other hand, for any given values of (σ_1, σ_2) we can minimise the contribution of V_θ to the potential by varying θ. Hence it is certainly true that θ_0 is just the value of θ which minimises V_θ with $(\sigma_1, \sigma_2) = (v_1, v_2)$.

Following this argument, it is clear that minimisation of the tree level potential V_0, with respect to θ, is exactly equivalent to solving

$$\frac{\partial V_\theta}{\partial\theta} = \Big(-2\lambda_5\sin(2\theta)\,v_1v_2 + 2\mu_R^2\sin(\theta) + 2\mu_I^2\cos(\theta)\Big)\,v_1v_2 = 0 \tag{4.1.12}$$

This is important to the argument here because, assuming that eqn(4.1.12) has valid solutions [1] then θ_0 always lies at the global minimum of V_θ, with v_1 and v_2

[1] We shall see that there are **always** valid solutions.

held constant. So, any range of θ in which V_θ must achieve its minimum must contain θ_0.

Following this line of argument, I will show below that

$$\theta_0 \in \left[-\frac{\pi}{2}, 0\right] \tag{4.1.13}$$

by considering the signs of the terms in eqn.(4.1.12) **with the convention developed in eqn.(4.1.10) of the last section.**

The expression for θ_0 reduces to

$$\mu_R^2 \sin(\theta_0) + \mu_I^2 \cos(\theta_0) - \lambda_5 \sin(\theta_0)\cos(\theta_0) v_1 v_2 = 0 \tag{4.1.14}$$

Unfortunately, this cannot be solved analytically [2] but we can determine which quadrant contains θ_0.

So, let us first consider whether or not there exist legitimate solutions to eqn.(4.1.12). Re-write in the following form

$$\mu^2 \cos(\theta + \alpha) = \lambda_5 v_1 v_2 \sin(2\theta) \tag{4.1.15}$$

where $\mu^2 = +\sqrt{\mu_R^4 + \mu_I^4}$, and $\tan(\alpha) = -\mu_R^2/\mu_I^2$, using simple trigonometrical identities.

Note that $\mu_R^2, \mu_I^2 \geq 0$, so $\tan(\alpha) < 0$, and that $\mu_R^2 \sin(\theta) + \mu_I^2 \cos(\theta)$ is increasing at $\theta = 0$, so $\alpha \in [-\pi/2, 0]$.

Since α is in the 4^{th} quadrant, the LHS of eqn.(4.1.15) is monotonically increasing in the 4^{th} quadrant, and certainly changes sign in the 4^{th} quadrant.

The RHS of eqn.(4.1.15) is equal to zero at $-\pi/2$ and 0, and non-zero everywhere else in the 4^{th} quadrant (except for $\lambda_5 = 0$, in which case the RHS is zero everywhere.)

Since the RHS of eqn.(4.1.15) is zero at $-\pi/2$ and 0, and continuous in the range $[-\pi/2, 0]$, and the LHS is continuous in the same range and certainly changes sign at some point in the range, then the two sides are certainly equal for some $\theta \in [-\pi/2, 0]$. (See figure 4.1.) Hence, there is always a solution of eqn.(4.1.15) in the 4^{th} quadrant.

[2] Actually, we can solve analytically, by rearranging the expression as a quartic equation in $\cos(\theta)$ and solving using the analytic solutions to quartic polynomials. However, the resulting expressions are so complex and long-winded as to be unusable.

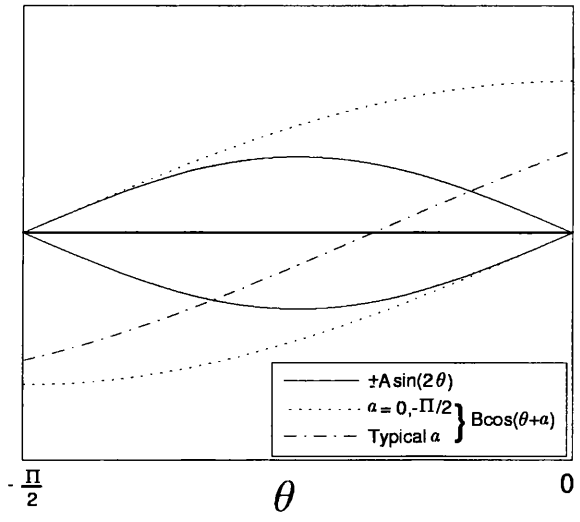

Figure 4.1: $\pm A\sin(2\theta)$ and $B\cos(\theta+\alpha)$ plotted for θ and α in the 4th-quadrant. The cosine is plotted for $\alpha = 0, -\pi/2$ and a typical value $\alpha = -1$. It is easy to see from this diagram that, with α in this range, these two curves **must** intersect in the 4th-quadrant.

However, this only proves that we have a stationary point of some description in the 4^{th} quadrant. It must also be shown that this is the global minimum.

Consider again V_θ – to minimise with respect to θ we must minimise the expression

$$\lambda_5\sigma_1\sigma_2\cos(2\theta) - 2\mu_R^2\cos(\theta) + 2\mu_I^2\sin(\theta) \tag{4.1.16}$$

This can be re-written as

$$\lambda_5\sigma_1\sigma_2\cos(2\theta) + 2\mu^2\cos(\theta+\alpha) \tag{4.1.17}$$

by using trigonometrical identities, giving $\mu^2 = +\sqrt{\mu_I^4 + \mu_R^4}$ and $\tan(\alpha) = \mu_I^2/\mu_R^2$. Since $\mu_I^2, \mu_R^2 \geq 0$, then $-\mu_R^2\cos(\theta) + \mu_I^2\sin(\theta)$ is rising at $\theta = 0$, and so $\alpha \in [-\pi, -\pi/2]$.

Note that the first term, proportional to $\cos(2\theta)$, is symmetrical about $\theta = 0$, while the second term, proportional to $\cos(\theta+\alpha)$, is not. Consider θ_0 – there are 4 points, $\theta_i \in [-\pi, \pi], i = 1\ldots4$, say, such that $\cos(2\theta_i) = \cos(2\theta_0)$. Thanks to the symmetry in $\cos(2\theta)$, these 4 points lie one in each quadrant, so let us number

them accordingly – θ_i lies in the i^{th}-quadrant. In terms of θ_4, the points are given by

$$\begin{aligned} \theta_1 &= -\theta_4 \\ \theta_2 &= \pi + \theta_4 \\ \theta_3 &= -\pi - \theta_4 \end{aligned} \tag{4.1.18}$$

and $\theta_0 = \theta_j$, for some j between 1 and 4.

Since the $\cos(2\theta)$ term is identical at all four θ_i, the term $2\mu^2 \cos(\theta + \alpha)$ must be the deciding factor in identifying which of the θ_i is the minimum. Now, we already have α in the 3^{rd} quadrant, and hence $\cos(\theta_i + \alpha)$ will be smallest for $i = 4$ for **any** four points related like the those in (4.1.18). (Consider figure 4.2 to see this.)

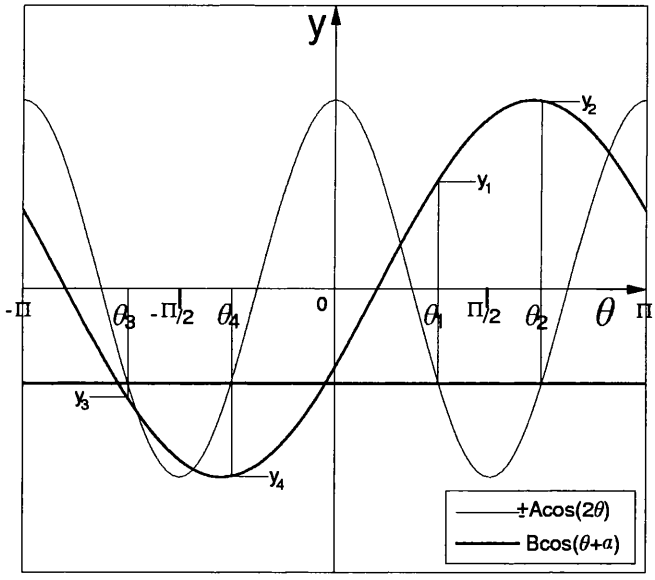

Figure 4.2: Plots of $\pm A\cos(2\theta)$ and $B\cos(\theta+\alpha)$ for a typical α in the 3^{rd}-quadrant. The $\theta_i, i = 1\ldots4$ are 4 typical points, one in each quadrant, where $\pm A\cos(2\theta_i)$ are identical. The y_i are the values of $B\cos(\theta_i + \alpha), i = 1\ldots4$. For α in the third quadrant, y_4 is always smaller than $y_i, i = 1\ldots3$. Hence, the global minimum of $\pm A\cos(2\theta) + B\cos(\theta + \alpha)$ must lie in the 4^{th}-quadrant when $\alpha \in (-\pi, -\pi/2)$.

Hence, in the sign convention $v_1, v_2, \mu_R^2, \mu_I^2 > 0$, it is true that θ_0 is constrained

to lie in the 4^{th}-quadrant – ie.,

$$\theta_0 \in [-\pi/2, 0] \tag{4.1.19}$$

4.1.4 Conditions for a physically acceptable potential

It has already been noted in this chapter that we cannot analytically determine the boundaries of the baryon preserving regions. Hence the need for a numerical search of the parameter space.

However, it must also be noted that there are regions of the parameter space, lying outwith the baryogenesis regions, where the potential becomes completely unphysical – eg., where the potential becomes unbounded below and thereby has no global minimum or stable vacuum. It is obvious that the baryogenesis regions must be wholly contained within regions where the potential *is* physically acceptable (I will refer to these regions as "physical regions.") These physical regions will form the basis parameter space and I will be specifically interested in what proportion of the physical region is contained in the baryogenesis region. In order to restrict attention to the physical region, we must decide what constitutes a physically acceptable potential.

The simplest approach to deciding what constitutes a physically acceptable potential is to decide when a potential is unphysical. The next few sections will explain the various ways in which the potential can become unphysical and derive the conditions which form the boundaries between the physical and unphysical regions. Of course, such conditions have been considered and derived before in many references. However,to my knowledge they have never been derived in the most general two-Higgs-doublet model before this work. [3]

4.1.5 Stability of the vacuum

When the potential is unbounded below then there is no stable minimum and, hence, no stable vacuum. The fields would tend to roll ever further from the

[3]The general conditions were reported by Kastening in [24] while this work was in progress, providing a useful check on my own work and indicated an error in one of my derivations.

origin and take on infinitely large VEV's, resulting in an infinite energy density in the vacuum. This is, of course, entirely unacceptable. So, under what conditions does the potential take on these unfortunate characteristics?

It is useful to prove the following result at this point –

Let

$$\Phi_1 = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad \Phi_2 = \begin{pmatrix} v_3 \\ v_4 \end{pmatrix} \tag{4.1.20}$$

$$\text{So,} \Phi_1^\dagger \Phi_1 = |v_1|^2 + |v_2|^2, \quad \Phi_2^\dagger \Phi_2 = |v_3|^2 + |v_4|^2, \quad \Phi_1^\dagger \Phi_2 = \overline{v_1} v_3 + \overline{v_2} v_4 \tag{4.1.21}$$

Then we have

$$\begin{aligned} & \left(\Phi_1^\dagger \Phi_1\right)\left(\Phi_2^\dagger \Phi_2\right) - \left(\Phi_1^\dagger \Phi_2\right)\left(\Phi_2^\dagger \Phi_1\right) \\ = \; & |v_2|^2 |v_3|^2 + |v_1|^2 |v_4|^2 - v_1 v_4 \overline{v_2 v_3} - \overline{v_1 v_4} v_2 v_3 \\ = \; & (v_1 v_4 - v_2 v_3)\,\overline{(v_1 v_4 - v_2 v_3)} \\ = \; & |v_1 v_4 - v_2 v_3|^2 \qquad \geq 0 \end{aligned} \tag{4.1.22}$$

So, the result is

$$(\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) \geq (\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1) \tag{4.1.23}$$

This will be put to use in the following series of derivations.

Let us now introduce a notation to be used in deriving the necessary conditions on the parameters of the potential. The notation will be

$$\Phi_1^\dagger \Phi_1 = A, \quad \Phi_2^\dagger \Phi_2 = B, \quad \Phi_1^\dagger \Phi_2 = X + iY \tag{4.1.24}$$

where A, B, X, Y are all real, and applying this to the result (4.1.23) above leads to the condition $AB \geq (X^2 + Y^2)$.

The behaviour of the potential at large fields depends only on the quartic terms. For the potential to be bounded below requires that the quartic terms are positive at large fields or, equivalently, that the gradient of the quartic terms should be positive (negative) at large positive (negative) fields. So, it is necessary that

$$\lambda_1 A^2 + \lambda_2 B^2 + \lambda_3 AB + (\lambda_4 + \lambda_5) X^2 + (\lambda_4 - \lambda_5) Y^2 \; > \; 0 \tag{4.1.25}$$

Firstly, consider the two directions $A = 0$ and $B = 0$. It is quickly realised that the following conditions are necessary

$$\begin{aligned} \lambda_1 &> 0 \\ \lambda_2 &> 0 \end{aligned} \qquad (4.1.26)$$

Now, consider the last two terms in (4.1.25). If $(\lambda_4 \pm \lambda_5) \geq 0$, then the sum of these terms is certainly bounded below by zero since $X^2, Y^2 \geq 0$. However, if one of, or both, $(\lambda_4 \pm \lambda_5) < 0$, then the sum of these terms is bounded below by $\min(\lambda_4 + \lambda_5, \lambda_4 - \lambda_5)AB$, since $X^2, Y^2 \leq AB$ by condition (4.1.23). So, the condition for stability of the vacuum at large fields reduces to

$$\left(\sqrt{\lambda_1}A - \sqrt{\lambda_2}B\right)^2 + \left(\lambda_3 + \lambda_{min} + 2\sqrt{\lambda_1\lambda_2}\right)AB \;>\; 0 \qquad (4.1.27)$$

where $\lambda_{min} = \min(0, \lambda_4 + \lambda_5, \lambda_4 - \lambda_5)$ However, when $A/B = \sqrt{\lambda_2/\lambda_1}$, then it is necessary that $\lambda_3 + \lambda_{min} + 2\sqrt{\lambda_1\lambda_2} > 0$.

This can be re-written to provide the necessary condition

$$\min(\lambda_3, \lambda_3 + \lambda_4 + \lambda_5, \lambda_3 + \lambda_4 - \lambda_5) + 2\sqrt{\lambda_1\lambda_2} > 0 \qquad (4.1.28)$$

4.1.6 Conservation of E.M.

The true vacuum is the state which provides the spontaneous breaking of the $SU(2)_L \times U(1)_Y$ symmetry which is required in the Higgs' mechanism, but must not break the $U(1)_{em}$ symmetry. It can be shown [41] that a charge conservation law of the form $\partial_\mu j^\mu(x) = 0$ implies that the vacuum is uncharged. Hence, the broken symmetry vacuum must be uncharged. If the doublets are aparallel in the vacuum then one of them must have a non-zero VEV in its charged fields, leading to the breaking of electromagnetic charge conservation. So, conservation of EM charge is equivalent to the condition that the VEV's of the two doublets must be parallel. So, it is necessary to derive the conditions on the parameters of the potential such that the global minimum must lie in some direction with the two doublets parallel.

In order to simplify the derivation of the necessary conditions, it is useful to

rewrite the tree level potential in the following form –

$$\begin{aligned} V' &= -\mu_1^2\left\{A-\frac{v_1^2}{2}\right\}-\mu_2^2\left\{B-\frac{v_2^2}{2}\right\} \\ &+\lambda_1\left\{A^2-\frac{v_1^4}{4}\right\}+\lambda_2\left\{B^2-\frac{v_2^4}{4}\right\}+\lambda_3\left\{AB-\frac{v_1^2v_2^2}{4}\right\} \\ &+\lambda_4\left\{X^2+Y^2-\frac{v_1^2v_2^2}{4}\right\}+\frac{\lambda_5}{2}\left\{(X+iY)^2+(X-iY)^2-\cos(2\theta_0)\frac{v_1^2v_2^2}{2}\right\} \\ &-\mu_3^2(X+iY)-\mu_3^{2\star}(X-iY)+\left(\mu_R^2\cos(\theta_0)-\mu_I^2\sin(\theta_0)\right)v_1v_2 \end{aligned}$$

where the global minimum is **defined** to lie at the point

$$\Phi_1=\frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_1\end{pmatrix}\qquad \Phi_2=\frac{\exp(i\theta_0)}{\sqrt{2}}\begin{pmatrix}0\\ v_2\end{pmatrix} \tag{4.1.29}$$

Then clearly $V' \geq 0$ for all fields Φ_1 and Φ_2 and it is possible to derive necessary conditions on the parameters of the potential by considering the sign of the tree level potential at various points away from the global minimum.

Consider the point

$$\Phi_1=\frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_1\end{pmatrix}\qquad \Phi_2=\frac{\exp(-i\theta_0)}{\sqrt{2}}\begin{pmatrix}0\\ -v_2\end{pmatrix} \tag{4.1.30}$$

All terms in V', except the soft symmetry–breaking term, are zero at this point. The soft breaking term becomes

$$\begin{aligned} &\{\mu_R^2\cos(\theta_0)+\mu_I^2\sin(\theta_0)\}v_1v_2+\{\mu_R^2\cos(\theta_0)-\mu_I^2\sin(\theta_0)\}v_1v_2 \\ =\;& 2\mu_R^2\cos(\theta_0)v_1v_2 \end{aligned} \tag{4.1.31}$$

and hence, in our sign convention, it is necessary that

$$cos(\theta_0)\geq 0 \tag{4.1.32}$$

Similarly, by considering the point where $\theta_0 \to -\theta_0$ and the v_i remain unchanged, the soft symmetry-breaking term is the only changed term again. In our sign convention, this leads to the result

$$\begin{aligned} \{-2\mu_I^2\sin(\theta_0)\}v_1v_2 &\geq 0 \\ \to\qquad \sin(\theta_0) &\leq 0 \end{aligned} \tag{4.1.33}$$

These two results combine to constrain θ_0 to the fourth quadrant (as has already been demonstrated.)

Consider now the case

$$\Phi_1 = \frac{1}{\sqrt{2}}\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{4}}\begin{pmatrix}0\\v_2\end{pmatrix}, \Phi_2 = \frac{\exp(i\theta_0)}{\sqrt{2}}\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{1}{4}}\begin{pmatrix}0\\v_1\end{pmatrix} \tag{4.1.34}$$

Then all terms in V' involving $\Phi_i^\dagger\Phi_j$ go to zero and the resulting potential has the form

$$\begin{aligned}
V' &= -\frac{\mu_1^2}{2}\left\{\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{2}}v_2^2 - v_1^2\right\} - \frac{\mu_2^2}{2}\left\{\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{1}{2}}v_1^2 - v_2^2\right\} \\
&\quad + \frac{\lambda_1}{4}\left\{\left(\frac{\lambda_2}{\lambda_1}\right)v_2^4 - v_1^4\right\} + \frac{\lambda_2}{4}\left\{\left(\frac{\lambda_1}{\lambda_2}\right)v_1^4 - v_2^4\right\} \\
&= -\frac{\mu_1^2}{2}\left\{\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{2}}v_2^2 - v_1^2\right\} - \frac{\mu_2^2}{2}\left\{\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{1}{2}}v_1^2 - v_2^2\right\} \\
&= -\frac{1}{2}\left\{\lambda_1 v_1^2 + \frac{\Sigma_+}{2}v_2^2 - \gamma_1\frac{v_2}{v_1}\right\}\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{4}}\left\{\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{4}}v_2^2 - \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{1}{4}}v_1^2\right\} \\
&\quad -\frac{1}{2}\left\{\lambda_2 v_2^2 + \frac{\Sigma_+}{2}v_1^2 - \gamma_1\frac{v_1}{v_2}\right\}\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{1}{4}}\left\{\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{1}{4}}v_1^2 - \left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{4}}v_2^2\right\} \\
&= \frac{1}{2}\left\{\sqrt{\lambda_1\lambda_2} - \frac{1}{2}\left(\Sigma_+ - \frac{2\gamma_1}{v_1v_2}\right)\right\}\left\{\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{1}{4}}v_1^2 - \left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{4}}v_2^2\right\}^2 \\
&\geq 0
\end{aligned} \tag{4.1.35}$$

where $\Sigma_+ = \lambda_3 + \lambda_4 + \lambda_5\cos(2\theta_0)$.

Hence, the necessary condition is

$$\left(\Sigma_+ - \frac{2\gamma_1}{v_1v_2}\right) \leq 2\sqrt{\lambda_1\lambda_2} \tag{4.1.36}$$

The last necessary condition required is most simply derived by considering the physical mass spectrum.

4.1.7 Reality of the physical masses

Since we have defined that we are working at the global minimum then it must be true that the curvature matrix, evaluated at the minimum, is semi-positive definite — ie. the mass spectrum has only non-negative masses. We have 8 scalar

fields and hence the mass (ie. curvature) matrix is an 8×8 matrix but it can be arranged in a block diagonal form as shown below —

$$M^2 = \begin{pmatrix} M^2_{1256} & O_4 \\ O_4 & M^2_{3478} \end{pmatrix} \tag{4.1.37}$$

where the submatrices M^2_{1256} and M^2_{3478} are formed from the bases $\{\phi_1, \phi_2, \phi_5, \phi_6\}$ and $\{\phi_3, \phi_4, \phi_7, \phi_8\}$ respectively, and have the forms

$$M^2_{1256} = \begin{pmatrix} a_{1256} & 0 & c_{1256} & d_{1256} \\ 0 & a_{1256} & -d_{1256} & c_{1256} \\ c_{1256} & -d_{1256} & b_{1256} & 0 \\ d_{1256} & c_{1256} & 0 & b_{1256} \end{pmatrix} \tag{4.1.38}$$

$$M^2_{3478} = \begin{pmatrix} a_{3478} & e_{3478} & f_{3478} & g_{3478} \\ e_{3478} & b_{3478} & -g_{3478} & h_{3478} \\ f_{3478} & -g_{3478} & c_{3478} & j_{3478} \\ g_{3478} & h_{3478} & j_{3478} & d_{3478} \end{pmatrix} \tag{4.1.39}$$

For completeness, the full form of these matrices can be found in appendix B.

We can easily diagonalise the matrix M^2_{1256}, as seen in appendix B, which leads to the eigenvalues

$$\begin{aligned} M^2_{\chi^\pm} &= 0 \\ M^2_+ &= -\frac{v_0^2}{2}\left[\lambda_4 + \lambda_5\cos(2\theta_0) - \frac{\mu_R^2\cos(\theta_0) - \mu_I^2\sin(\theta_0)}{v_1 v_2}\right] \end{aligned} \tag{4.1.40}$$

where $M^2_{\chi^\pm}$ correspond to the would-be Goldstone bosons $\chi^\pm$ and M^2_+ corresponds to the charged scalars $H^\pm$.

Since M^2_+ must be positive, then we have the condition

$$\begin{aligned} \lambda_4 + \lambda_5\cos(2\theta_0) &\leq \tfrac{2\mu_R^2\cos(\theta_0) - \mu_I^2\sin(\theta_0)}{v_1 v_2} \\ \rightarrow \qquad (\lambda_4 + \lambda_5\cos(2\theta_0)) &\leq \tfrac{2\gamma_1}{v_1 v_2} \end{aligned} \tag{4.1.41}$$

Since μ_R^2, μ_I^2, v_1 and v_2 are all positive, and θ_0 lies in the fourth quadrant, then the RHS of this inequality is certainly positive. Mathematically, the value of γ_1 can be placed in the range

$$\min\left(\mu_R^2, \mu_I^2\right) \leq \gamma_1 \leq 1.41\max\left(\mu_R^2, \mu_I^2\right) \tag{4.1.42}$$

but it should be remembered that the value of θ_0 is determined by the physical requirements of the model and not simple mathematical possibilities — eg., if $\mu_I^2 = 0$, then $\theta_0 = 2\pi$ and the lower limit given above can then appear very loose. Nevertheless, these limits are useful in understanding the effect of the constraint in eqn.(4.1.41). (Since μ_R^2 and μ_I^2 will be restricted to be less than $22.5 \times 10^3 GeV^2$ in our computations, then γ_1 is certainly always less than about $30 \times 10^3 GeV^2$. For given values of μ_R^2 and μ_I^2, the minimal values of γ_1 occur for θ_0 around $3\pi/2$ or 2π.)

The largest possible value of $(v_1 v_2)$ is $v_0^2/2$, which is of the order $30 \times 10^3 GeV^2$. This maxmimum value occurs for values of β_0 around $\pi/4$. The minimum value of $(v_1 v_2)$ in our allowed range lies at $\beta_0 = \pi/2$, which would result in the RHS of (4.1.41) becoming very large in any case (except, of course, in the case of $\mu_3^2 = 0$, where θ_0 is no longer a required parameter of the model.)

Hence, at large values of β_0, and in cases of large $|\mu_3^2|$ with θ_0 lying in the interior of the fourth quadrant, the RHS of the condition will tend to be larger than the upper limit of the LHS of the inequality. In these circumstances, the condition is clearly very weak. However, for smaller $|\mu_3^2|$, and especially with β_0 around the region of $\pi/4$, the condition can become quite tight.

4.1.8 Necessary and sufficient conditions for a physical potential

In all, 7 necessary conditions have been listed — eqns (4.1.26) (two separate conditions), (4.1.28),(4.1.32),(4.1.33),(4.1.36), and (4.1.41). Kastening [24] has shown that these same conditions are also *sufficient* for the existence of a stable vacuum with the correct symmetry-breaking properties.

In determining whether a particular point in the parameter space lies in the basis space, it is essential to check that it satisfies each of these conditions.

The reader should note that the symmetry-breaking conditions have been derived at small fields but the vacuum stability conditions depend on behaviour at large fields. Hence, conditions (4.1.26) and (4.1.28) might be derived using the

renormalisation group equation to provide "running" parameters, ensuring that the vacuum is stable at least up to the scale of new physics — the SUSY scale, say. However, such detailed considerations might be unjustified considering the uncertainties which exist in the numerical values of the limits which should be imposed on $m(T_c)/T_c$ and $v(T_c)/T_c$, which will be discussed later.

4.2 Constraints on the baryogenesis region

Within the basis space I hope to find a BAU-preserving region (or baryogenesis region.) This region must allow a first order $EWPT$, and must sufficiently suppress the sphaleron transition rate in the broken symmetry phase to preserve any BAU generated during or before the $EWPT$.

Immediately after the phase transition, the broken symmetry phase is at the critical temperature, T_c, of the phase transition, or very close to it. The sphaleron rate in the broken phase depends on the mass of the sphaleron relative to the temperature of the plasma. As will be explained in chapter 5, I will define the critical temperature to be the maximum temperature at which one of the curvature eigenvalues of the effective potential goes to zero at the origin. This is actually the lower limit of the critical temperature and so it defines the latest time by which the sphaleron rate must be suppressed.

As for any thermal process, the likelihood of a thermal fluctuation generating the sphaleron configuration is controlled by a Boltzmann term, $\exp(-M_{sph}/T)$, where M_{sph} is the sphaleron mass. The rate is approximately given by [4],

$$\Gamma_{\rm sph} = \frac{1}{Q_B}\frac{dQ_B}{dt} \tag{4.2.43}$$

$$nonumber \approx T_c \exp\left(-\frac{M_{\rm sph}(T_c)}{T_c}\right) \tag{4.2.44}$$

and must be much less than the expansion rate of the universe immediately after the electroweak phase transition, which is given by the Hubble "constant"

$$H \approx \frac{T_c^2}{M_{Pl}} \approx 10^{-17} T_c \tag{4.2.45}$$

where M_{Pl} is the Planck mass. So,

$$\exp\left(-\frac{M_{sph}(T_c)}{T_c}\right) \ll 10^{-17} \approx \exp(2.3 \times -17)$$
$$\Rightarrow \quad \frac{M_{sph}(T_c)}{T_c} \gg 39.1 \tag{4.2.46}$$

Hence, $M_{sph}(T_c)/T_c > 40$ for sufficient suppression of the sphaleron rate. A more precise calculation of the rate by Shaposhnikov [11] yields an even tighter constraint

$$\boxed{\frac{M_{sph}(T_c)}{T_c} \geq 45} \tag{4.2.47}$$

The sphaleron mass is given by (eg. [6, 4, 11, 14, 13])

$$M_{sph} = \frac{4\pi B(\lambda)}{g} v(T) \tag{4.2.48}$$

where g is the $SU(2)$ gauge coupling, $v(T)$ is the Higgs VEV at temperature T, and $B(\lambda)$ ranges from $B(0) = 1.56$ to $B(\infty) = 2.72$ and is dependent on the coupling constants of the scalar fields. Replacing the mass in the M/T limit by the expression above gives

$$\boxed{\frac{v(T_c)}{T_c} \geq 1.1} \tag{4.2.49}$$

This limit determines the minimum strength of the first order phase transition which will be required to suppress the sphaleron rate. Unfortunately, since no explicit expression for $v(T_c)$ is available to us in the most general form of the two-Higgs-doublet effective potential, this constraint is as much as we are able to determine about the boundary of the baryogenesis region. Hence, the limit must be applied numerically in a computational search of the parameter space in order to find the baryogenesis region.

Chapter 5

The Critical Temperature

5.1 Definition of the Critical Temperature

The critical temperature of a phase transition is one of its most fundamental properties. However, the definition of T_c in the EWPT is not a clear cut question.

How should the EWPT critical temperature be defined? In the simplest terms, this is the temperature at which the phase transition occurs. However, from the brief discussion of the various models of the phase transition and their dynamical considerations in chapter 2, it is clear that the beginning of the phase transition is open to interpretation.

In particular, for the typical model of transition by bubble nucleation, there is some temperature T_1, say, at which the true and false vacuua become degenerate – transitions into the false vacuum become energetically favourable for temperatures below T_1, but the rate is suppressed by the potential barrier between the two vacuua. At some lower temperature T_{SC} say, (SC - sub-critical) the bubbles first begin to nucleate but these are initially sub-critical and collapse under their own surface tension, delaying the onset of the true phase transition and leading to a supercooled universe. At some temperature $T_2 < T_{SC}$, the bubbles release enough free energy to overcome the surface tension and begin to expand indefinitely – this would be considered the start of the true phase transition. It is expected that eventually the origin turns over to become a saddle point at temperature T_3. At this temperature the fields are free to move away from the origin with no

suppression due to a potential barrier. At this temperature it seems unlikely that bubble nucleation will be required, the transition will proceed quickly and so the phase transition will end at this time if it has not already done so.

I will now consider each of these cases in more detail.

(i). The phase transition might be said to occur at the temperature T_1 at which the global minimum of the scalar potential (or the true, symmetric vacuum) becomes degenerate with the secondary, local minimum (or false, asymmetric vacuum.) At higher temperatures the false vacuum will lie at a higher potential than the true vacuum, and population of the false vacuum is not energetically favourable. At T_1 and below, the transition to the new, broken symmetry vacuum is favoured. However, there will be a potential barrier between these two minima, requiring the fields to tunnel through the barrier, or be carried over the top by thermal fluctuations, resulting in a suppressed transition rate.

Hence, T_1 is the maximum temperature at which the phase transition can begin but the rate suppression will tend to delay the onset of bubble nucleation and the "true" critical temperature is likely to lie somewhere below T_1.

Evaluation of T_1 is not simple. The potential at the origin is easily given as a function of T; this must be equated with the potential in the false vacuum and solved for T_1. However, in order to determine the potential in the false vacuum it is first necessary to locate it. This means minimising the potential with respect to the VEV's of the fields. (The techniques used are described in chapter 6.) In the fully general form of V_{eff},which I will be using, the minimisation cannot be done analytically and hence T_1 must be evaluated numerically.

(ii). At some temperature $T_{SC} < T_1$ the suppression of the bubble nucleation rate due to the potential barrier is weak enough to allow the process to begin. However, as mentioned above, initially these bubbles are too small for their free energy to overcome the surface tension in the bubble walls and the bubbles simply collapse on themselves. As the temperature drops further, the universe becomes supercooled [1] and the free energy of the typical nuclear bubble increases. Eventu-

[1] which leads to a departure from thermal equilibrium, as required in Sakharov's conditions for baryogenesis - see chapter 2.

ally, at some $T_2 < T_{SC}$, the bubbles reach the critical size to overcome the surface tension and they are able to expand indefinitely – the phase transition has now certainly begun. The bubbles expand to fill the entire universe, at which point the phase transition is complete.

A detailed definition of the temperature at which the nucleating bubbles become critical and can expand indefinitely is model-dependent. Hence, it requires a study of the bubbles of the new vacuum, which I have already explained is beyond the scope of this thesis. Hence, we can do no more than constrain T_2 to a range of values. Certainly, $T_2 < T_{SC} < T_1$ – the simplest lower bound on T_2 is the temperature at which a saddle point develops in the potential at the origin, as described now.

(*iii*). It is certain that the scalar fields have a non-zero VEV at zero temperature, as witnessed by the massive gauge bosons. (Assuming that the Higgs mechanism properly explains particle masses.) Hence, the effective potential is almost certainly not a local minimum at the origin and is hence a local maximum or saddle point. However, at high temperature the scalar field is expected to have no VEV, and the spontaneously broken symmetry is restored. So, there is certainly some temperature T_3 at which the effective potential at the origin changes from a minimum to a saddle point in at least one direction in the scalar field space.

The saddle point represents an unstable equilibrium and the fields will be energetically free to follow the route of the negative curvature away from the origin towards the new minimum (which obviously must lie below the origin otherwise it could not be a saddle point.) Since there is no potential barrier to suppress the rate of this transition it will proceed quickly and the phase transition will complete quickly if it has not already completed.

Hence, T_3 forms the lower bound on the critical temperature. It might be possible, in any given model, to put a tighter lower bound on T_2 by considering the bubble wall structure and propagation, but T_3 is dependent only on the parameters of the effective potential itself and, as it turns out, can be derived analytically, at least in a perturbative fashion.

These three situations are represented in diagram (5.1).

We must decide on one of these definitions as our critical temperature. T_1 does not make a good choice for our purposes since it is generally higher than the temperature at which the phase transition actually occurs and we are interested in conditions immediately after the phase transition. Certainly it will be true that a first order phase transition will start at some $T_2 > T_3$ and will finish (if it has not already done so) at T_3, since there is no longer any suppression of the transition at this point. This thesis is specifically interested in the possibility of avoiding the wash-out of any baryon asymmetry – however generated – in the immediate aftermath of the phase transition. This requires the study of the sphaleron rate in the broken phase – ie., **inside** the bubbles of new phase, **after** the bubble wall has passed. However, the bubble wall is where the phase transition actually occurs, and the temperature in the bubble interiors will continue to fall as the universe expands. So, I am interested in temperatures which lie in the range $T_3 \leq T \leq T_2$.

I will use the definition

$$T_c = T_3, \tag{5.1.1}$$

for the following reasons. T_3 forms an approximate lower bound on the critical temperature, and hence forms a bound for the suppression of the sphaleron rate – the sphaleron rate must be suppressed sufficiently to preserve the baryon asymmetry by the time $T = T_3$ since the transition will be complete by the time the universe cools to T_3. Also, as will be demonstrated in the following sections, T_3 can be approximated analytically purely by considerations of the effective potential and without resorting to the dynamics of any particular baryogenesis scenario in the bubble walls. This analytical derivation also gives some physical feel for the critical temperature.

5.2 Derivation of the critical temperature

T_c is the highest temperature at which the curvature of the scalar potential becomes negative in some direction *at the origin,* ie. at which the curvature matrix, evaluated at the origin, becomes non-positive-definite. Hence, we evaluate the curvature matrix and note that it is considerably simplified at the origin as compared

to the general form, allowing us to diagonalise the matrix analytically, which is not possible for the general form. Having determined the diagonal form, or essentially determined the eigenvalues, we can go on to evaluate T_c.

Let us begin by defining the curvature matrix in question. I will call the matrix C_0, and define it as

$$[C_0]_{ij} = \frac{\partial^2}{\partial\phi_i\partial\phi_j}\left\{V_0 + \frac{T^2}{24} \times \mathrm{Tr}\left[M^2(\phi,T)\right]\right\}\Big|_{origin} \tag{5.2.2}$$

Note that only terms up to the M^2T^2 term are considered. This approach was suggested by Dolan and Jackiw [30] who pointed out that the M^3T term can introduce an imaginary element into the calculation. (The existence and possible cancellation of these imaginary terms in the potential was discussed in chapter 3, section 3.5.)

Then C_0 has the form

$$C_0 = \begin{pmatrix} A & 0 & 0 & 0 & -\gamma_1 & \gamma_2 & 0 & 0 \\ 0 & A & 0 & 0 & -\gamma_2 & -\gamma_1 & 0 & 0 \\ 0 & 0 & A & 0 & 0 & 0 & -\gamma_1 & \gamma_2 \\ 0 & 0 & 0 & A & 0 & 0 & -\gamma_2 & -\gamma_1 \\ -\gamma_1 & -\gamma_2 & 0 & 0 & B & 0 & 0 & 0 \\ \gamma_2 & -\gamma_1 & 0 & 0 & 0 & B & 0 & 0 \\ 0 & 0 & -\gamma_1 & -\gamma_2 & 0 & 0 & B & 0 \\ 0 & 0 & \gamma_2 & -\gamma_1 & 0 & 0 & 0 & B \end{pmatrix} \tag{5.2.3}$$

where

$$\begin{aligned} A &= -\mu_1^2 + \frac{T^2}{12}(6\lambda_1 + 2\lambda_3 + \lambda_4) + \frac{T^2}{16}\left(3g^2 + g'^2 + 4g_t^2\right) \\ B &= -\mu_2^2 + \frac{T^2}{12}(6\lambda_2 + 2\lambda_3 + \lambda_4) + \frac{T^2}{16}\left(3g^2 + g'^2 + 4g_b^2 + \frac{4}{3}g_\tau^2\right) \\ \gamma_1 &= \mu_R^2\cos(\theta) - \mu_I^2\sin(\theta) \\ \gamma_2 &= \mu_R^2\sin(\theta) + \mu_I^2\cos(\theta) \end{aligned} \tag{5.2.4}$$

This can be rotated to the form

$$C_0 = \begin{pmatrix} S & 0_4 \\ 0_4 & S \end{pmatrix} \tag{5.2.5}$$

where S is the matrix

$$S = \begin{pmatrix} A & -\gamma_1 & 0 & \gamma_2 \\ -\gamma_1 & B & -\gamma_2 & 0 \\ 0 & -\gamma_2 & A & -\gamma_1 \\ \gamma_2 & 0 & -\gamma_1 & B \end{pmatrix} \tag{5.2.6}$$

Clearly then, C_0 has *at most* 4 distinct eigenvalues, which are just the eigenvalues of S. I derive these eigenvalues by considering the characteristic equation of S, namely $\det[S - xI_4] = 0$. This gives

$$\left(x^2 - (A+B)x - \left(\gamma_1^2 + \gamma_2^2\right) + AB\right)^2 = 0 \tag{5.2.7}$$

and so we obtain

$$x = \frac{(A+B) \pm \sqrt{(A-B)^2 - 4\left(\gamma_1^2 + \gamma_2^2\right)}}{2} \tag{5.2.8}$$

However, this form is not conducive to determining the critical temperature. Recall that we are interested in finding the highest temperature at which one of these eigenvalues is zero. Clearly, the smaller eigenvalue will go to zero first (ie., the one with the negative square root term.) Hence we can proceed by writing

$$\begin{aligned} (A+B)^2 = &\quad (A-B)^2 - 4\left(\gamma_1^2 + \gamma_2^2\right) \\ \Rightarrow \quad AB + \left(\mu_R^2 + \mu_I^2\right) &= 0 \end{aligned} \tag{5.2.9}$$

If we re-write A in the form $A = -\mu_1^2 + A'T^2$, and similarly for B, then eqn(5.2.9) reduces to

$$A'B'T^4 - \left(\mu_1^2 B' + \mu_2^2 A'\right) T^2 + \left(\mu_R^2 + \mu_I^2\right) + \mu_1^2 \mu_2^2 = 0 \tag{5.2.10}$$

which leads to

$$T^2 = \frac{\left(\mu_1^2 B' + \mu_2^2 A'\right) \pm \sqrt{\left(\mu_1^2 B' - \mu_2^2 A'\right)^2 + 4A'B'\left(\mu_R^2 + \mu_I^2\right)}}{2A'B'} \tag{5.2.11}$$

Hence, there are four distinct possibilities for the critical temperature. However, two of them will be negative (assuming T^2 above is positive, which I will discuss in a moment), which cannot be considered physical. So, the definition of T_c will be the positive temperature such that

$$T_c^2 = \max\left\{ \frac{\left(\mu_1^2 B' + \mu_2^2 A'\right) \pm \sqrt{\left(\mu_1^2 B' - \mu_2^2 A'\right)^2 + 4A'B'\left|\mu_3^2\right|^2}}{2A'B'} \right\} \tag{5.2.12}$$

Note that we must ensure that T_c is physical by considering the possibility of T^2 above becoming negative, leading to a purely imaginary temperature, or the term inside the square root becoming negative, leading to a complex temperature. It should be clearly understood, however, that these constraints say nothing about whether or not the parameter set in question describes a physically realistic potential, only whether the origin is always (or never) a minimum of the potential. If T_c is not real, and the origin is always at least a local minimum, then it is possible that the phase transition might still proceed by tunnelling and bubble nucleation, and it would certainly be first order, though possibly very weakly.

I will complete this section with a comment on the critical temperature in the special cases where μ_I^2 and/or μ_R^2 zero, which are the cases generally considered in the literature. When $\mu_I^2 = 0$, then the critical temperature will look almost exactly the same as the definition above, with $|\mu_3^2|^2$ replaced by μ_R^2. When μ_R^2 is also zero, then the solutions for critical temperatures reduce drastically to the forms $-\mu_{1[2]}^2 + A'[B']T^2 = 0$ Hence, the introduction of a $\mu_3^2\Phi_1^\dagger\Phi_2$ + h.c. term has an effect on the critical temperature. Note, however, that there are still four mathematical solutions for T, it is still possible to have unphysical critical temperatures and the matrix C_0 would still have two distinct eigenvalues. This can be explained by the properties of the matrix S which is defined above and discussed in considerable detail in appendix A on an analytical attempt to make corrections to the critical temperature.

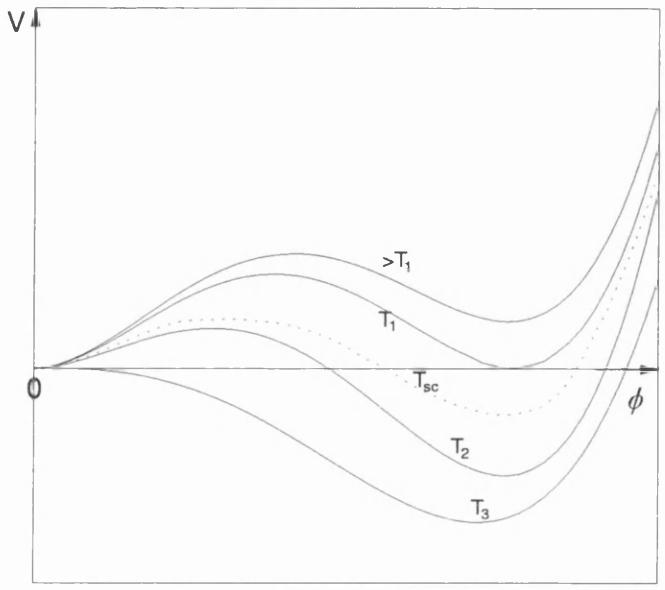

Figure 5.1: Representation of the true and false vacuua at temperatures around the critical temperature. At temperatures above T_1 the secondary minimum lies above the true vacuum. At T_1, the two minima become degenerate. At some T_{sc} below T_1 bubbles of the broken phase begin to condense but they are sub-critical size and collapse on themselves. At T_2 the bubbles become critical and can expand indefinitely to fill the universe – this is the temperature at which the phase transition is first able to proceed to its conclusion and hence can be considered as the true critical temperature. At T_3 the curvature at the origin has just reached zero and it is no longer a minimum – fields can now roll freely down to the new minimum.

Chapter 6

Computational methods

As pointed out already, particularly in chapter 4, we are forced to resort to computational methods in order to investigate the parameter space of the general two-Higgs-doublet model. For example, the form of the most general tree-level potential results in a neutral mass matrix which cannot be usefully diagonalised analytically, forcing us to turn to numerical evaluation of the scalar mass spectrum. Also, since the expressions for the mass-squared eigenvalues are not available, the $\mathrm{Tr}(M^3)$ term cannot be given explicitly. In any event, in the most general form, with the temperature-dependent masses inserted, the potential is a long, complicated expression which is virtually impenetrable when it comes to minimisation. Hence, numerical minimisation techniques are also required when determining the nature of the phase transition.

Having been forced to turn away from analytical techniques, a survey of the parameter space requires a numerical search in order to identify the baryogenesis region and its structure and properties. Hence, it is necessary to develop a methodology for the search, and to decide what information should be extracted from the numerical computations. Armed with this information it is possible to take the results of the preceding chapters and to develop a computational method to describe the baryogenesis region.

6.1 Searching the scalar sector parameter space

The initial aim is to investigate the parameter space in the extended scalar sector in a search for a region, or regions, which will avoid the washout of a baryon asymmetry by sphaleron transitions after the phase transition. This requirement is equivalent to insisting on a first order phase transition. Hence, at any point in the parameter space, we should be able to determine the nature of the $EWPT$ and measure the rate of sphaleron transitions after the phase change.

Consider the most general scalar potential possible, containing 9 parameters, as already stated. Only one of these parameters is known – ie., the VEV of the scalar fields is experimentally measured as $v_0 = 246\text{GeV}$. This leaves 8 free parameters, and hence an 8-dimensional parameter space with which to work. Visualisation of such a space is essentially impossible, of course.

In chapter 4, analytical conditions were derived on the parameters of the tree-level effective potential which form the boundaries of the physical regions of the parameter space. These regions will form the basis space for the numerical search. However, due to the complicated form of the high-T approximation to the ring-corrected 1-loop expansion of the effective potential, it is not possible to draw conclusions on the boundaries of the baryogenesis region beyond the simple requirement of a strongly first order phase transition, as described in eqn(4.2.49) – namely, $v(T_c)/T_c \geq 1.1$. The application of all of these conditions will be discussed shortly.

Also, since the temperature-dependent mass spectrum can only be determined numerically, it is not possible to analytically ensure that the high-temperature expansion of the 1-loop potential will hold. Hence, the validity of the expansion must be checked as part of the numerical process. The details of this condition will also be described shortly.

These requirements – application of constraints to identify the basis space; application of numerical techniques to determine the nature of the $EWPT$; validation of the high-T expansion; the search of an 8-dimensional space – are the minimum requirements to determine whether there does exist a baryogenesis re-

gion within the physical basis space. However, it is also of interest to investigate the size and shape of the baryogenesis region – this will allow a discussion of the relative ease with which the two Higgs-doublet model is able to avoid the experimental lower bounds on the Higgs mass, and how the model is able to evade this bound. So, it is important for the computational model not only to indicate the existence of a baryogenesis region but to deliver details of the coupling parameters within the region.

Hence, I developed a random sampling approach to the investigation of the parameter space, applying the various constraints – fundamental physical requirements, $EWPT$ requirements and validity of the approximations – at each point in a random sample, and recording details of those which lie in the baryogenesis region. The methodology was initially envisaged as follows –

1. Choose a point at random in the parameter space, ensuring that the point meets the conditions defining the basis parameter space.

2. Derive the details of the potential at this point, determining the nature of the $EWPT$ and whether the high-T approximations are valid.

3. Hence determine this point's suitability for electroweak baryogenesis. Details of suitable points are recorded for further analysis.

This is clearly a simplistic view of what is likely to be encountered in such a process. It must take account of a good deal more than is immediately apparent here. For example, we require an appropriate method of selecting random points in the parameter space while implementing the appropriate physical conditions. The nature of the $EWPT$ relies on the VEV of the potential at T_c, so numerical minimisation of the potential is required. Evaluation of the potential for the purposes of minimisation will require numerical evaluation of the mass-squared eigenvalues in order to determine the mass-cubed eigenvalues. Also, as discussed in chapter 4, the value of θ_0 (the vacuum expectation value of the phase angle between the two Higgs doublets at zero temperature) is located in the 4^{th}-quadrant, but cannot be determined analytically, requiring yet another numerical solution.

Clearly, the numerical investigation of the two Higgs doublet parameter space is a computationally intensive procedure.

Also note that the constraints derived above, which define the physical regions of the parameter space, will be fairly undemanding – they avoid the disastrous consequences of a bottomless potential, or a vacuum which violates EM charge conservation, but they do not take account of more subtle expectations. For example, we expect the theory to be weakly coupled at the electroweak scale, so the λ_i should be small. This is not one of the constraints derived in chapter 4 because it derives from our expectations of electroweak scale physics rather than fundamental requirements.

On a purely mathematical footing, the region of the search in the parameter space must be limited in order to maintain the validity of the high-T expansion and because we cannot numerically search an infinitely large space.

So, it is also important to take these factors into consideration when designing the search criteria. All of the points raised above will be discussed in detail in the remainder of this chapter.

6.1.1 Imposing the experimental constraint on v_0

The first constraint to be imposed is the experimentally observed VEV of the scalar fields – $v_0 = 246\text{GeV}$. In the "typical" notation of the scalar potential, as described in chapter 3 the parameters of the potential are considered to be the coupling parameters μ_i^2, λ_i, μ_R^2 and μ_I^2 – v_0 does not appear explicitly as a parameter. So we must begin by determining relationships between v_0 and the parameters in the "typical" notation (which will hereafter be referred to as the coupling parameters or coupling parameter set). Only then is it possible to impose this constraint.

At the minimum of the potential, the derivatives with respect to the field variables must be zero, leading to the following conditions –

$$-\mu_1^2 + \lambda_1 v_1^2 + \frac{\lambda_3 + \lambda_4 + \lambda_5 \cos(2\theta_0)}{2} v_2^2 - \gamma_1 \frac{v_2}{v_1} = 0 \qquad (6.1.1)$$

$$-\mu_2^2 + \lambda_2 v_2^2 + \frac{\lambda_3 + \lambda_4 + \lambda_5 \cos(2\theta_0)}{2} v_1^2 - \gamma_1 \frac{v_1}{v_2} = 0 \qquad (6.1.2)$$

$$\mu_R^2 \sin(\theta_0) + \mu_I^2 \cos(\theta_0) - \lambda_5 \sin(\theta_0)\cos(\theta_0) v_1 v_2 = 0 \qquad (6.1.3)$$

Using these conditions we can derive expressions for $\mu_{1,2}^2$ in terms of the VEV's of the scalar fields and the other couplings. This is a useful re-parameterisation for various reasons. Firstly, it allows us to impose the experimental measurement of the VEV v_0 by recalling that the phase between the VEV's of the two doublets is conventionally referred to as β, giving

$$v_1 = v_0 \sin(\beta_0) \qquad v_2 = v_0 \cos(\beta_0) \tag{6.1.4}$$

Secondly, although the dimensionless couplings (λ_i) are expected to be small and hence can be constrained typically by $|\lambda_i| \ll 1$, the dimensionful couplings (μ_i^2) are not so obviously restricted. The re-parameterisation afforded by the expressions above will allow us to avoid restricting the $\mu_{1,2}^2$ artificially with some estimated upper limit, and replace this problem with that of defining a sensible range for the parameters describing the position of the zero-temperature minimum, β_0 and θ_0. I have already explained that θ_0 is restricted to the 4^{th}-quadrant (chapter 4), and I will explain below how β_0 can be restricted.

However, in solving the conditions above, note that they contain 12 different parameters between them, including v_0, β_0 and θ_0. Since v_0 is fixed, I may choose any 8 other parameters to be free and then solve for the remaining 3. There are many possibilities here but it might be useful to solve for $\mu_{1,2}^2$, so I will not choose these as free parameters.

It seems somewhat inconsistent to derive one of the λ_i and choose the other four. It is also desirable to impose small coupling restrictions on the λ_i, so I will use these as free parameters.

This leaves us to consider μ_R^2, μ_I^2 and the angles β_0 and θ_0. There are strong arguments, which will be discussed later, to restrict the range of β_0, so it makes an obvious choice for another free parameter.

Again, it seems inconsistent to choose the value of one of μ_R^2 and μ_I^2, and derive the other. However, since these couplings act as a possible source of extra CP-violation over and above the Yukawa interaction, I am particularly interested in studying the effects of these terms on the baryon-preserving nature of the $EWPT$. Hence it is useful to deliberately choose values for μ_R^2 and μ_I^2 (unlike the other

dimensionful couplings $\mu_{1,2}^2$) and so I choose these two as the last free parameters.

This leaves us to derive the angle θ_0. As already stated in chapter 4, θ_0 cannot be determined analytically and so a numerical solution must be obtained.

So, I will re-parameterise the model using the expressions above. The new parameter set will be referred to as the "computational" parameter set and consists of

$$\beta_0, \lambda_{1...5}, \mu_R^2, \mu_I^2 \tag{6.1.5}$$

and, of course, $v_0 = 246GeV$.

6.1.2 Limiting the region of the search

The search will obviously be limited to part of the physical region of the parameter space. There are two sets of limits to be imposed – those derived in chapter 4, which are the fundamental restrictions necessary and sufficient to make the potential physically acceptable; and those imposed by our expectations for electroweak physics, such as weak coupling for the scalar fields at the scale of the $EWPT$. Let us consider these latter conditions first.

We expect the scalar couplings to be weak, so the dimensionless couplings in the tree level potential should be small. However, defining an upper limit to "small" can be rather arbitrary. It is typically taken to mean less than 1/2 [32, 4], but the $SU(2)$ gauge coupling $g \approx 2/3$, for example.

Hence, I have decided on an upper limit for the small couplings of 1.

This is essentially as arbitrary as any other upper limit but it may be interesting to allow couplings larger than or of similar size to g for the following reason. The aim of this search is to investigate the possibility of avoiding the experimental lower limits on the Higgs mass, and relatively strong couplings may be a factor in this by increasing masses. If couplings of order 1 really are too strong, then the baryogenesis regions are unlikely to stretch out as far as this because the scalar masses will tend to be too large to maintain the validity of the high-T expansion. In this case, the extra volume contained in the basis space will tend to diminish the relative size of the baryogenesis region compared to the basis space as a whole.

This should be borne in mind when considering the results of the search.

So, the ranges of λ_i for the search are restricted to

$$0 \leq \lambda_{1,2} \leq 1, \qquad -1 \leq \lambda_{3,4,5} \leq 1 \tag{6.1.6}$$

As mentioned earlier, the value of β_0 can also be constrained. Since I have defined Φ_1 as the doublet which couples to the up-type quarks, then the top mass is given by $m_t = g_t v_1/\sqrt{2}$. To avoid having to make g_t overly large, then it is desirable to have $v_1 > v_2$, or $\tan(\beta) \geq 1$. A more thorough examination of this limit is described in a paper by this author in collaboration with Davies, Froggatt and Moorhouse [32]. Requiring the perturbative two Higgs-doublet model to be valid up to grand unification scales, $\Lambda \approx 10^{15}(GeV)$ say, then the top quark mass is restricted to be less than or equal to its quasi-fixed point value as determined by the renormalisation group equations. The quasi-fixed point is given by $m_t^{FP} \sim 230 sin(\beta_0)$Gev, giving an upper limit on the top quark mass of 150GeV for $\tan(\beta_0) = 0.9$. Since the top quark mass is now thought to be of order 150 GeV or greater, I take

$$\tan(\beta_0) \geq 0.9 \tag{6.1.7}$$

Consider now the fundamental physical conditions derived in chapter 4 to define the boundaries of the physical region. Implementation of these conditions is fairly straightforward.

The two conditions on λ_1 and λ_2 – $\lambda_{1,2} \geq 0$ – are implemented in the model by simply restricting them both to the range $0 \leq \lambda_{1,2} \leq 1$, as already stated in condition (6.1.6).

The two conditions which are used to place θ_0 in the fourth quadrant – $\sin(\theta) < 0, \cos(\theta) > 0$ – are imposed by restricting the search for solutions for θ_0 (see chapter 4) to the fourth quadrant.

The other three conditions are (4.1.28,4.1.36,4.1.41)

$$\begin{aligned}
\min(\lambda_3, \lambda_3 + \lambda_4 + \lambda_5, \lambda_3 + \lambda_4 - \lambda_5) + 2\sqrt{\lambda_1 \lambda_2} &> 0 \\
\left(\Sigma_+ - \frac{2\gamma_1}{v_1 v_2}\right) &\leq 2\sqrt{\lambda_1 \lambda_2} \\
\lambda_4 + \lambda_5 \cos(2\theta_0) &\leq \frac{2\mu_R^2 \cos(\theta_0) - \mu_I^2 \sin(\theta_0)}{v_1 v_2}
\end{aligned}$$

These are all difficult to impose at the time of choosing the parameter values. Hence their validity is simply checked for each point chosen in the random sample. (In fact, condition (4.1.41) is tested numerically by checking that all four physical mass-squared values are positive – this overkill provided a safety net in case of numerical or coding problems.)

The last free parameters in the computational parameterisation are μ_R^2 and μ_I^2. Since I am particularly interested in the effects of these couplings on the baryon-preserving nature of the $EWPT$, it is desirable to have more control over their values in the search. In order to do this, a grid of 16×16 equally spaced values in the μ_R^2, μ_I^2-plane will be used, such that

$$\begin{aligned} 0 \leq \quad \mu_R^2 \quad &\leq 2.25 \times 10^4 \\ 0 \leq \quad \mu_I^2 \quad &\leq 2.25 \times 10^4 \end{aligned} \tag{6.1.8}$$

The upper limits on these ranges were chosen to give maximum values of μ_R, μ_I of the same order as the top mass and critical temperature of the phase transition.

The method used to choose and count points in the basis space and baryogenesis regions will be described later.

6.2 Calculations at the critical temperature

Up to this point in the choice of parameters and imposition of constraints I have implicitly been working only at tree level since the constraints were derived from knowledge of the physical vacuum in the present universe, essentially at zero temperature. However, it is the properties of the scalar potential near the critical temperature which are required to determine the nature of the EWPT and whether the potential can support baryogenesis. Hence, all further calculations and conditions described in this chapter are carried out at high temperature.

Clearly, the critical temperature, its definition and calculation, are of major consequence to the determination of the nature of the potential and the resulting $EWPT$. Hence, a chapter has already been devoted to the description of this part of the work and an analytical expression for the critical temperature was

developed. This expression is simply evaluated for each given point in the search sample in order to determine the critical temperature for the $EWPT$ in the scalar sector described at the point in question.

6.2.1 Minimisation of the scalar potential at $T = T_c$

We are unable to diagonalise the tree level mass matrix because we cannot diagonalise the submatrix formed from the neutral scalar fields (see appendix B for further details.) Hence, we are also unable to diagonalise the mass matrix with the first loop correction term in the potential, ie. the "T^2" term. So, it is necessary to resort to numerical evaluation of the mass eigenvalues at the T^2 level in order to calculate the $\mathrm{Tr}\,[M^3]$ term. Obviously, then, there is no choice but to perform a numerical minimisation of the potential since evaluation of the potential is not analytical.

One of the most notorious problems in numerical minimisation is finding a **global** minimum – how can one be sure that the minimum found is not simply local? We face this problem now since we must find the global minimum of the scalar effective potential at the critical temperature. Described below is the procedure developed to perform this minimisation and some of problems encountered in finding a stable minimisation procedure – the discussion of this problem helps to highlight a source of error in the evaluation of the potential.

The minimisation routine used is provided in the National Algorithm Group (NAG) library. It is designated "E04JBF" and a full description of the procedure (and minimisation techniques in general) can be found in the NAG library documentation. For my purposes here, the important points are the following

– the routine requires an initial estimate of the position of the minimum

– the routine does not require an evaluation of the derivatives of the function being minimised

– the convergence criteria can be controlled by the user.

The first point will be discussed presently. The second point is important because we are not able to provide an analytical evaluation of the derivatives since

we cannot determine the $\mathrm{Tr}(M^3)$ term explicitly. The last point will be addressed in the next section as part of the discussion of problems with the minimisation method.

The numerical minimisation routine requires a first estimate of the position of the minimum, from which it follows the route of fastest descent in an attempt to find the minimum. Such an initial point is not necessarily an estimate of the position of the minimum. We might try instead to provide some point from which we feel confident that the route of fastest descent will lead to the minimum – eg. in the case of minimisation at zero temperature (or, equivalently, at tree level) we know that the minimum lies in the quadrant with $v_1, v_2 > 0$ and $\theta_0 \in [3\pi/2, 2\pi]$, and hence we might choose some point lying in the direction $v_1 = v_2 \equiv v$ with $\theta_0 = 7\pi/8$ and set $v = v_0/\sqrt{2}$. Since this point lies in "the centre" of the quadrant where we know the minimum must lie, we might feel confident that the route of fastest descent from this point will lead to the minimum although we have not attempted to estimate where the minimum lies within the quadrant.

However, it is very difficult to see how such an initial point might be reliably chosen beyond the tree level due to the complexity of the problem. In particular, since I wish to minimise the potential at $T = T_c$, where the temperature dependent parts of the potential certainly play an important role, it seems unlikely that one could rely on the position of the minimum at zero T as a good estimate for that at T_c. (Indeed, the constraints on θ_0 were also derived at zero-T and we cannot guarantee that θ_{T_c} is similarly constrained to lie in the 4^{th}-quadrant.) So, I developed a procedure involving multiple minimisations, each started from a different inital point which is chosen at random. The best minimum found by these multiple attempts is taken as the best estimate of the global minimum.

Minimisation of the potential is a numerically intensive calculation because of the need for repeated evaluations of the potential, each of which involves numerical diagonalisation problems. Hence, it is desirable to perform as few minimisations as possible in the search for the global minimum.

To this end, I undertook a simple study of the number of randomly chosen initial points required to achieve the best approximation to the global minimum.

The study consisted of running and re-running the search with the same "random" sample in each run (the random number generator used is pseudo-random and can be forced to repeat a sample.) The number of initial points used in determining the minimum of the potential was varied for each run and the results compared to determine the optimal number of initial points. It was found that using more than 3 initial points resulted in no appreciable improvement in the depth of the minima detected. Hence, the minimisation procedure is based on 3 attempted minimisations, each starting from a different point, all of which have $v_1(T), v_2(T)$ located at field values between 0 and 200 GeV, and with $\beta(T)$ between 0 and 2π.

6.2.2 Problems with minimisation

A subtle problem was encountered in the original attempts to minimise the effective potential. The numerical minimisation procedure was found to return a point as its best estimate of the minimum but also returned an error status – failure to meet all of the conditions for a minimum even though no lower point can be found. This led me to make numerical investigations of the potential at several points in the parameter space, using the computer model to evaluate the potential on a grid of points in $(\sigma_1, \sigma_2, \theta)$-space. These investigations were centred on the "failed" points returned by the numerical minimisation procedure in order to determine why the procedure was failing. The resulting figures demonstrated very clearly that the minimisation procedure had indeed found a minimum and it became clear that there was a lesson to be learned herein.

Evaluating the potentials on ever finer grids it eventually became apparent that there were several local minima located in the regions being investigated. However, this fact could not be reconciled with the form of the potential. The local minima were found to lie in a very small region around the point originally returned by the minimsation procedure – the width of the region being several orders of magnitude smaller than the values of the fields in the region. This implies that the potential should have a high frequency component in its form but that is clearly not the case since it is a multi-dimensional, fourth order polynomial. So,

it appeared that some uncertainty in the evaluation of the potential must be the source of the apparent rapid variation behaviour.

Since the evaluation of the potential derives from a largely analytical form, there appear to be only two possible sources of this uncertainty. The first is the numerical precision of the machine itself. However, since the calculations involved in the evaluation of the potential are mostly simple evaluations of polynomials, and were performed using double precision FORTRAN, we can rely on them to approximately 14 significant figures while the variations observed in the potential were at a level many orders of magnitude larger than this. This leads us to the second possible source of error – the evaluation of the $\mathrm{Tr}\,[M^3]$ term. This relies on numerical estimation of the eigenvalues of the curvature matrix at the point of evaluation. However, the procedure used to find these eigenvalues is known to converge to the eigenvalues slowly (only linearly, in fact) and to converge even more slowly when there are repeated eigenvalues or zero eigenvalues. Of course, there are repeated eigenvalues and even repeated zero eigenvalues in the curvature matrix evaluated at the global minimum. Hence, the evaluation of the $\mathrm{Tr}\,[M^3]$ term is likely to be considerably less accurate than expected from arguments on machine precision.

It is a feature of minimisation routines in general that they must be given a series of conditions on which to base the decision to halt the search. These conditions compare the value of the function and its (estimated) derivatives at the test-point and a group of "nearby" points. When the function is found to be lower on the test-point than on its neighbours, and the derivatives also have appropriate values (in particular, the matrix of second derivatives is positive definite,) AND the "nearby" points are within a stated distance of the test-point, then the routine will halt its search. When we have a well-behaved function, which can be evaluated to accuracies close to the machine precision, then the tolerances set on the distances between the test-point and its neighbours can be very small. However, when uncertainties in the evaluations are much larger than the machine precision then the tolerances need to be much larger than we might at first expect. In the case in hand, it was apparent that we should try relaxing the test conditions for the

minimum by increasing the tolerances.

Tests of the minimisation procedure with larger tolerances proved very successful – the procedure no longer complained of apparent failure and the minima returned were the same as those found before. However, it became apparent that the optimal values for the tolerances varied over the μ_R^2, μ_I^2-grid. Hence, the final version of the minimisation procedure adapts the tolerances according to the values of μ_R^2 and μ_I^2.

6.2.3 Determination of the nature of the EWPT

Having minimised the effective potential at the critical temperature we are now in a position to determine the nature of the $EWPT$. With the global minimum found to lie at $(v_1(T_c), v_2(T_c), \theta(T_c))$ then, as described in chapter 4, the condition for a first order phase transition is $v(T_c)/T_c > 1.1$ where $v(T_c)^2 = v_1(T_c)^2 + v_2(T_c)^2$.

Since the derivation of this limit does involve approximations, it will be relaxed slightly to give

$$\frac{v(T_c)}{T_c} > 1 \tag{6.2.9}$$

Having imposed this relatively relaxed limit when running the numerical search, it is a simple process to consider the effects of tightening the limit when analysing the results.

The nature of the $EWPT$ is easily checked in the numerical search using this limit since T_c and the VEV of the field at T_c are evaluated at every point in the sample. Points at which the $EWPT$ satisfies this criteria are defined to lie in the baryogenesis region and will be counted and recorded as such. Points which lie in the basis space but fail to provide a strongly first-order phase transition are counted as part of the basis space but their details are not recorded.

6.2.4 Validity of the high-temperature expansion

As pointed out in chapters 3 and 4, the expansions used in evaluating the effective potential are only valid in the high temperature limit, which is normally considered to be $m(T)/T \ll 1$.

The high-T expansion was used to approximate the integrals in the exact 1-loop potential as given in eqn(3.4.30). They have the form

$$I_{\pm}\left(\frac{m^2}{T^2}\right) = \pm\int_0^{\infty} dx x^2 \log\left\{1 \mp \exp\left(-\sqrt{\frac{m^2}{T^2}+x^2}\right)\right\} \tag{6.2.10}$$

Singleton [19] fits a 10^{th}-order polynomial to the integrals in order to make numerical evaluations of the potential. Comparing these fits with the high temperature expansion shows that the validity of the expansion typically holds to within 10% for $m(T)/T$ as high as 1.6.

Hence, I have pushed the limit of the high-T expansion up to

$$\frac{m_i(T_c)}{T_c} < 1.6 \quad \forall \quad i = 1,\ldots,8; +\text{top quark} \tag{6.2.11}$$

The $m_i(T_c)$ are calculated numerically as the eigenvalues of the curvature matrix at the global minimum and hence, as explained in the previous section, the values do contain uncertainties due to the inherently slow convergence of the method used in the NAG library routine to solve the eigenvalue problem. However, these uncertainties are at least one or two orders of magnitude smaller than the eigenvalues, except in the case of very small eigenvalues which are, in any case, going to meet the condition of eqn(6.2.11). So, these uncertainties will not have any noticeable effect on the number cases passing or failing on condition 6.2.11.

6.3 Summary of the modelling procedure

I can now provide a summary of how the parameter space search is performed.

1 – Choose a point in the parameter space – these points are chosen on a grid in the (μ_R^2, μ_I^2)-plane. The grid is a 16×16 square, with spacing of $1.5 \times 10^3 \text{GeV}^2$ in both directions, giving a range of $0 \rightarrow 22.5 \times 10^3 \text{GeV}^2$ in both μ_R^2 and μ_I^2. A set of n points is examined at each point on the grid. The points are chosen in the form of the computational parameter set, described earlier in this chapter. They are then converted to the coupling parameter set, by solving for μ_1^2, μ_2^2 and θ_0 using the expressions (6.1.1).

2 – Check that the chosen point lies in the physical region – I only wish to consider points which lead to physically acceptable potentials. Hence I check that the chosen point meets the criteria which define the physically acceptable region of the parameter space.

NB! If the chosen point is found to lie in the unphysical region then it is immediately abandoned and **is not included** in the count of the number of points examined. Hence, all points counted do lie in the basis space.

3 – Determine the critical temperature – as described in chapter 5.

4 – Minimise the effective potential at the critical temperature – described above in section 6.2.1.

5 – Check the validity of the high temperature expansion – equivalent to imposing the condition (6.2.11). If the high-T expansion is found to be invalid then I abandon the point but I **do** include it in the count of the number of points examined since it does lie in the physical basis space.

6 – Determine the nature of the phase transition – the global minimum at $T = T_c$ must meet the condition (4.2.49) in order for the phase transition to be first order and to suppress the sphaleron rate after the phase transition.

Points which are found to be physical, to validate the use of the high temperature expansion and to provide a first order phase transition are recorded for later analysis.

The unphysical, physical and baryogenesis regions of the parameter space have already been defined. Note, however, that there is, in fact, a fourth type of region – the region, contained in the physical region, where the high temperature expansion is invalid. This region, which will be referred to as the "invalid region", cannot be studied by the methods developed in this chapter and so it is unknown what proportion of this region might support baryogenesis. Points which are found to lie in the invalid region **are** counted, as explained in part 5 above, and hence we must be clear that the proportion of the physical basis space which is found to support baryogenesis is actually a lower limit.

Chapter 7

Results of the numerical search of the parameter space

As stated in the introductory chapter, the primary aim of this thesis is to investigate the possibility of avoiding disagreement between the theoretical upper limit and experimental lower limit on the Higgs boson mass. The disagreement arises in the Standard Model because the requirement of a strongly first order electroweak phase transition (for the purposes of baryogenesis) can only be accomodated with a light Higgs, of mass $\sim$ 60GeV or less, while the experimental SM-Higgs searches have $M_H > 60\text{GeV}$.

In considering the two Higgs-doublet extension to the Standard Model, it was hoped that the extra degrees of freedom provided by the second scalar doublet would allow a strongly first order $EWPT$ even with relatively heavy Higgs particles. The experimental lower bounds are also weakened in the two Higgs-doublet scalar sector, and so the SM problem is being attacked on two fronts – hopefully allowing the Higgs masses to rise while driving the experimental limits down.

Hence, our first area of interest in these results is the existence of a baryogenesis region – is there a region in the basis parameter space of the two Higgs-doublet model where the $EWPT$ is first order? Secondly, and just as importantly, does this region (or part of it) lie above the experimental bounds on the two doublet Higgs masses?

If we find a viable region for the preservation of the BAU in the aftermath of

the $EWPT$, it becomes interesting to consider the size and structure of the region. The size of the region as a proportion of the basis parameter space gives a measure of the difficulty (or ease) of sustaining baryogenesis in the two Higgs-doublet model. The structure of the region will indicate what factors contribute positively to the preservation of the BAU, and what factors are likely to make the hypothesis difficult to sustain. In particular, the explicit CP-violating term in the tree-level potential is an attractive feature of the model for many baryogenesis scenarios and it will be interesting to consider how the size of the CP-violating coupling affects the strength of the phase transition and the size of the baryogenesis region.

With some knowledge of the size, location and structure of the baryogenesis region, it should be possible to derive some physical understanding of how the two doublet model manages (or fails) to accomodate the baryogenesis hypothesis in agreement with experiment.

Before proceeding to present the results, let us briefly consider the experimental bounds on Higgs masses in the two scalar doublet model. Also, please note that many graphs which are too large to appear in the text have been placed together at the end of the chapter.

7.1 Experimental bounds with two Higgs-doublets

The experimental mass lower bound on the SM Higgs boson is well established at over 60GeV. However, with two scalar doublets present the situation becomes considerably more complicated and the experimental bounds are not so clear. [32, 44]

For the charged scalar particles, which might be produced in reactions such as $e^+e^- \to Z \to H^+H^-$, a lower bound on the mass can be clearly established at 42GeV. However, the neutral sector does not succumb so easily to analysis. The most commonly treated case is that of the CP-invariant potential (ie., μ_3^2 pure real or imaginary in our notation.) In this case there are two CP-even scalars (defined as h and H, the lighter and heavier scalars respectively) and one CP-odd pseudo-scalar (defined as A.)

The CP-even states mix ϕ_3 and ϕ_7 with a mixing angle α. There is a reduction in the rate of $Z \to hZ^\star$ by a factor of $\sin^2(\beta - \alpha)$ over the Standard Model case. This allows for lower values of the lighter scalar mass m_h while remaining within experimental limits on the rate. In this case, the bound can be viewed as a curve in the m_h, m_A-plane. However, the intension in this work is to allow complex μ_3^2 and hence the potential will generally not be CP-invariant. This introduces three mixing angles into the neutral sector, $(\alpha_1, \alpha_2, \alpha_3)$ say. The mass eigenstates are no longer CP-even or CP-odd and the h, H and A labels are no longer applicable. The experimental bound becomes a surface in m_1, m_2, m_3-space, where m_i are the neutral masses.This is a complicated bound to impose in detail and so I propose to make a simplification. In the vein of the CP-invariant case, where the neutral masses were generally allowed to be lower than in the one-doublet case, I will impose the same lower limit on all of the Higgs masses. I will take this limit to be 40GeV which is approximately the lower limit on the charged masses and is well established, as mentioned earlier.

This simplification goes against the grain of the general nature of the rest of this study. However, it will be seen that the results justify this simplification since the baryogenesis region extends to masses well beyond the limit imposed even in the Standard Model case, and very little of the mass spectrum of the baryogenesis region lies in the neighbourhood of these limiting values.

7.2 The existence of a baryogenesis region

As described in chapter 6, the search of the parameter space took place on a grid of μ_R^2, μ_I^2 values. There are 16×16 equally spaced points in the grid, extending from $0 \to 22.5 \times 10^3 \text{GeV}^2$. There are 1000 sample points chosen at random within the physical basis space at **each** point on this grid. This gives a total of 256,000 points sampled in the basis space, which is enough to provide reasonably reliable statistics.

The results of the search for a baryogenesis region are summarised in graphs 7.1. These show the number of sample points found to lie in the baryogenesis

	Without experimental cuts	With experimental cuts
Size of basis space sample	$16 \times 16 \times 10^3 = 256,000$	$16 \times 16 \times 10^3 = 256,000$
Size of baryogenesis region	12833	12732
%-age of basis space in baryogenesis region	$5.01\% \pm 0.04\%$	$4.97\% \pm 0.04\%$

Table 7.1: Total size of sample in basis space and baryogenesis region. (Uncertainty determined as $\pm\sqrt{n}$, where n is the number of points counted.)

region for each point on the μ_R^2, μ_I^2-grid. The number is given as a fraction of the total number of points sampled at each point in the μ_R^2, μ_I^2-plane.

The first point to note is that a baryogenesis region does exist. As can be seen in table 7.1, of over one quarter million sampled points in the basis space, **12833** (12732 after experimental cuts) points were found to lie in the baryogenesis region – so, the region occupies about 5% of the basis space.

Note that the effects of the experimental Higgs mass bounds on the overall size of the baryogenesis region were very small – only 0.04% of the sampled points were cut from the region, which is less than 1% of its size. The detailed effects of the experimental cuts can be seen in graph 7.2. This depicts the number of cases cut at each point on the μ_R^2, μ_I^2-grid.

There appears to be a notable localisation of the cut cases near $\mu_3^2 = 0$. However, assuming a typical $\pm\sqrt{n}$ uncertainty, and the maximum number of cases cut at any point on the grid being 9, the error bars on this graph would be large and it is difficult to draw any firm conclusions. However, it will be demonstrated later, in the discussion of the mass spectrum of the baryogenesis region (section 7.4.1), that the Higgs masses tend to increase with μ_3^2. This would explain the bias towards small μ_3^2 in the experimental cuts.

Turning attention again to graphs 7.1, consider the shape of the baryogenesis region as a function of μ_R^2, μ_I^2. This clearly shows that the introduction of the explicitly CP-violating term $\mu_3^2(\Phi_1^\dagger\Phi_2)+h.c.$ has increased the available parameter

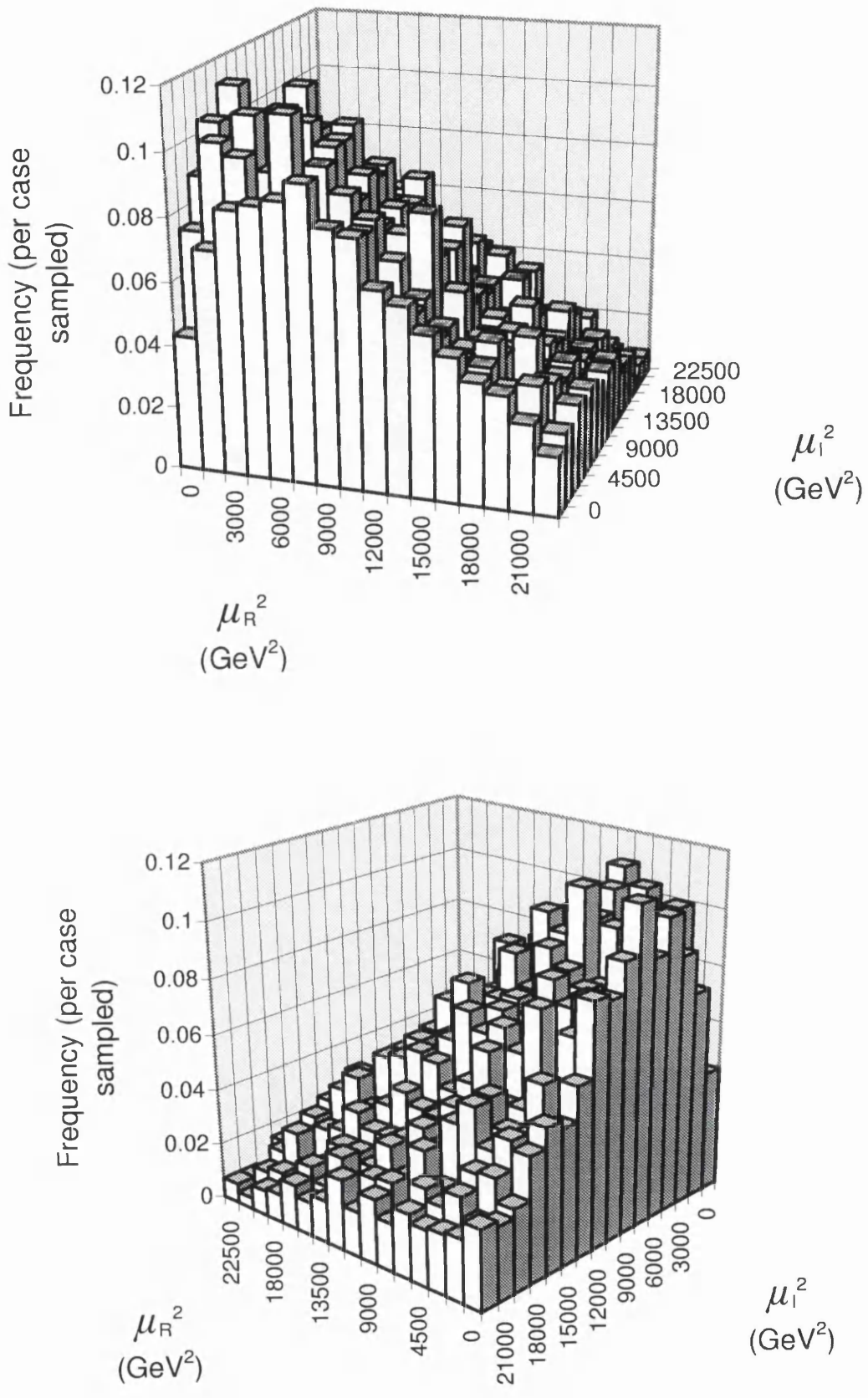

Figure 7.1: Number of cases found to lie in the baryogenesis region plotted against μ_R^2, μ_I^2. Presented as a fraction of the total number of points in sample (1000) at each point in μ_R^2, μ_I^2-grid.

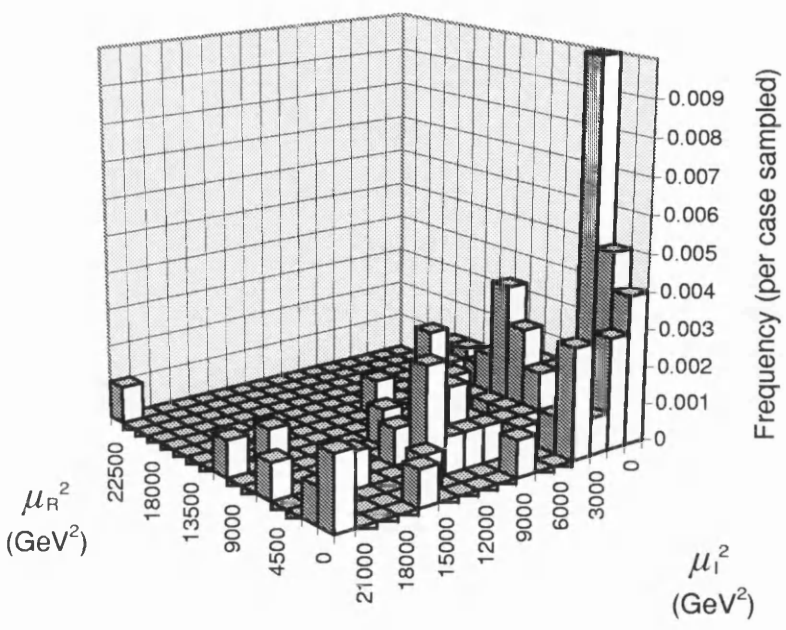

Figure 7.2: Number of cases cut from the baryogenesis region by the experimental lower mass bounds, plotted against μ_R^2, μ_I^2. Presented as a fraction of the total number of points in sample (1000) at each point in μ_R^2, μ_I^2-grid.

space to support baryogenesis. This is encouraging for electroweak baryogenesis with two scalar doublets, and particularly for those scenarios which require new sources of CP-violation.

It is interesting to consider the distribution of baryogenesis-supporting cases in more detail. Note in particular that baryogenesis is especially favoured by relatively small μ_3^2 and that $\mu_3^2 = 0$ is heavily suppressed by comparison. The rapid fall-off in BAU-preserving cases at large μ_3^2 might be explained by the increase in Higgs masses with μ_3^2 (section 7.4.1) which will tend to take this region of the parameter space beyond the area of validity of the high-temperature expansion. In order to make an authoritative statement on the source of this behaviour it is necessary to understand how the baryogenesis constraint of a strongly first order $EWPT$, and the purely mathematical limit on m/T, shape the baryogenesis region

depicted in graphs 7.1. To this end, it is useful to make a brief study of the basis parameter space which will help to explain what part each of these constraints plays in forming the baryogenesis region.

7.3 The physical basis space

In order to produce data for the basis space it is simply necessary to run the computational model with no cuts on $m(T_c)/T_c$ and $v(T_c)/T_c$. These constraints were developed purely from considerations of the validity of the high-temperature expansion and the requirement of a first order phase transition, and as such they play no part in determining whether a given set of parameters lies in the basis space. A set of data was calculated in the following manner – 50 basis space points were sampled at each point on the μ_R^2, μ_I^2-grid and the resulting details of the effective potential and *EWPT* ($v/T, mT$, mass spectrum, *etc.*) were recorded for each point. Where it proves useful or revealing, this "physical data set" will be compared with the "baryogenesis data set".

The major strength of this approach is in revealing the effects of the $m(T_c)/T_c$ and $v(T_c)/T_c$ constraints. It is important to understand how the effects of these two constraints differ from one another since the condition on $v(T_c)/T_c$ is truly physical – it has been developed by considering the order of the phase transition – while the condition on $m(T_c)/T_c$ represents only a limit in the mathematical approximations. Hence, behaviour controlled by the $v(T_c)/T_c$ constraint can be considered as the true physical nature of the model, while behaviour governed by the $m(T_c)/T_c$ condition is simply a limit in the validity of the mathematics.

It is also interesting to compare the effects of the various conditions on vacuum stability and the symmetry-breaking pattern of the vacuum state, which form the boundaries of the basis space (see chapter 4). The relative importance of these constraints is simply studied by counting the number of points which fail under each of the basis space constraints while we are in the process of choosing the basis sample. This study can be used to explain the observed distributions of the coupling constants in the physical data set.

I will now discuss how the physical constraints, developed in chapter 4 as boundaries of the basis space, shape the distributions of the coupling constants in the basis space. Then I will go on to consider how the v/T and m/T constraints shape the baryogenesis region.

7.3.1 Effects of the basis space constraints

The distributions of the λ_i in the basis space are reproduced in graph 7.8. All of these distributions can be explained in terms of the constraints (4.1.26,4.1.28, 4.1.32,4.1.33,4.1.36,4.1.41) derived in chapter 4. The appropriate conditions will be reproduced here for the sake of convenience.

Let us turn first to the distributions of λ_1 and λ_2. Recall that both of these couplings are constrained to be positive – hence the sudden cut-off in frequency as the couplings become negative. The observed steady increase in frequency of physical cases as both λ_1 and λ_2 increase can be attributed to conditions (4.1.28) and (4.1.36), namely

$$\min(\lambda_3, \lambda_3 + \lambda_4 + \lambda_5, \lambda_3 + \lambda_4 - \lambda_5) > -2\sqrt{\lambda_1\lambda_2} \qquad (7.3.1)$$

$$\left(\Sigma_+ - \frac{2\gamma_1}{v_1 v_2}\right) \leq 2\sqrt{\lambda_1\lambda_2} \qquad (7.3.2)$$

$(\lambda_3 + \lambda_4 \pm \lambda_5)$ lies in the range $[-3, 3]$; since $\gamma_1/(v_1 v_2) > 0$, and Σ_+ lies in the range $[-3, 3]$, the left-hand side of the second condition above lies in a range $[-x, 3]$, where $-x < -3$; and $2\sqrt{\lambda_1\lambda_2}$ lies in the range $[0, 2]$. Hence these conditions will tend to be more easily satisfied at larger λ_1 and λ_2. This leads to the decreased frequencies at small λ_1 and λ_2 which are seen in graph 7.8.

Also, since both constraints are symmetrical in λ_1 and λ_2, their effect on the respective distributions will be identical. This will result in similar distributions, as observed.

The distributions of λ_3 and λ_4 are similar and it is seen that both rise steadily over their entire range of $[-1, 1]$. To explain these distributions, turn to condition (4.1.41) which ensures that the charged mass-squared is positive — ie.,

$$\lambda_4 + \lambda_5 \cos(2\theta_0) - \frac{2\gamma_1}{v_1 v_2} \leq 0 \qquad (7.3.3)$$

This constrains the LHS of condition (7.3.2) above to be less than 1. This is easily seen by re-writing (7.3.2) in the form

$$\lambda_3 + \left(\lambda_4 + \lambda_5 \cos(2\theta_0) - \frac{2\gamma_1}{v_1 v_2}\right) \leq 2\sqrt{\lambda_1 \lambda_2} \tag{7.3.4}$$

and noting that the term in brackets was constrained to be negative above and that λ_3 lies in the range $[-1, 1]$.

Since the RHS of (7.3.2) is contained by $[0, 2]$, and we have constrained the LHS to be less than 1 by using condition (7.3.3), it is to be expected that (7.3.2) will not be a tight constraint and should have relatively small impact on the shape of the basis space. In the computational model, (7.3.1) was imposed before (7.3.2) and the tables 7.2 demonstrate that, as expected, condition (7.3.2) plays a relatively minor role after the imposition of the tighter constraint in (7.3.1).

Having seen that condition (7.3.2) plays a minor role in shaping the physical region in the computational model, the suppression of cases with larger negative λ_3 and λ_4 values should be put down to the effects of condition (7.3.1). In the same way that this constraint depressed the frequency of smaller values of λ_1 and λ_2, it will also suppress larger negative values of λ_3 and λ_4.

Lastly, note that the distribution of λ_5 is essentially symmetrical about zero. Considering the conditions above (7.3.1, 7.3.3), the possibility of a symmetrical behaviour in the distribution of λ_5 becomes apparent. In (7.3.1), for example, the "minimum function" term can be written as $\lambda_3 + \min(0, \lambda_4 - |\lambda_5|)$. The inclusion of the modulus brackets around λ_5 make it clear that the behaviour of λ_5 in this expression will not depend on its sign – ie., will be symmetrical about zero. In (7.3.3), λ_5 is paired with $\cos(2\theta_0)$. With θ_0 in the fourth quadrant then $\cos(2\theta_0)$ lies in the range $[-1, 1]$. Hence, the sign of the λ_5 term can once again be positive or negative, depending on the value of θ_0, and so the effect on the distribution of λ_5 is to make it symmetrical about zero.

Condition (7.3.2) has a $\lambda_5 \cos(\theta_0)$ term and hence will not contribute a symmetrical factor to the distribution of λ_5. However, as already seen, this constraint does not play a major role in the formation of the basis space and so its effect will be secondary in the distribution graphs.

The conditions on λ_1, λ_2 and θ_0 are imposed at the time of choosing the parameters and, hence, in the computational model these conditions are never tested as such. Therefore they are never seen to fail. However, a comparison of the failure rates on tests of the other conditions demonstrate their relative importance in shaping the physical region and/or the "overlap" in their effects. Tables 7.2 give the failure rates for the conditions (7.3.1) and (7.3.2) in a typical run of the computational model, with 150 sampled points in the basis space at each point on the μ_R^2, μ_I^2-grid. Condition (7.3.3), ensuring a real, positive charged mass, was **never** seen to fail after imposing the others.

Assuming that there is an overlap in the effects of the individual conditions, these failure rates are likely to depend on the order in which the conditions are tested. However, it certainly appears that the conditions which ensure stability of the vacuum – (eqns (7.3.1,7.3.2) along with $\lambda_1, \lambda_2 \geq 0$ – are enough to ensure the conservation of electromagnetic charge and physically acceptable masses.

The distribution of β_0 in the basis space appears in graph 7.3.

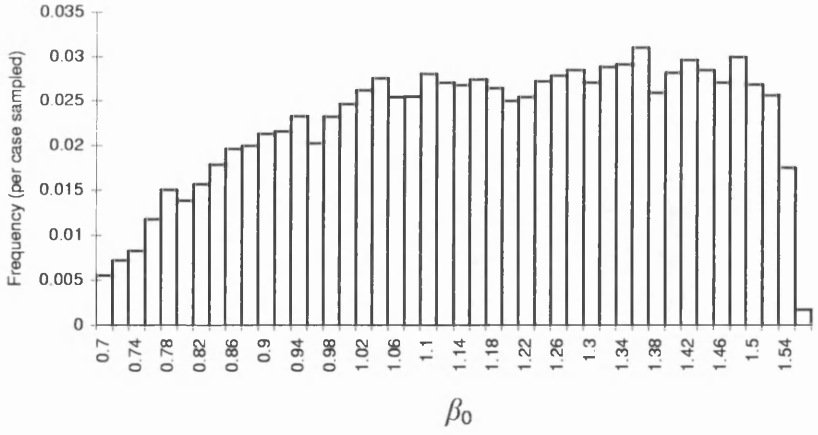

Figure 7.3: Distribution of β_0 in the basis parameter space. The data are presented as a fraction of the total number of β_0 values measured.

This distribution is relatively flat across the allowed range, except for a wide dip for $0.7 < \beta_0 < 1$, and a narrow, deep drop close to $\beta_0 = \pi/2$.

Number of failures for condition 7.3.1							
		$\mu_I^2(\text{GeV}^2)$					
		0	4500	9000	13500	18000	22500
μ_R^2 $(\text{GeV})^2$	0	1022	423	273	245	246	244
	4500	265	254	278	218	214	219
	9000	284	226	203	255	229	231
	13500	221	208	185	189	203	229
	18000	246	185	230	222	199	235
	22500	214	167	225	228	189	213

Number of failures for condition 7.3.2							
		$\mu_I^2(\text{GeV}^2)$					
		0	4500	9000	13500	18000	22500
μ_R^2 $(\text{GeV})^2$	0	24	29	18	14	18	21
	4500	20	27	24	27	25	29
	9000	17	19	22	16	22	25
	13500	21	19	24	20	20	19
	18000	27	23	15	23	29	18
	22500	24	22	24	24	22	21
		0	4500	9000	13500	18000	22500

Table 7.2: The number of points which failed conditions (7.3.1,7.3.2) in a typical run of the computational model for a sample size of 150 points at each point on the μ_R^2, μ_I^2-grid. (The values of μ_R^2 and μ_I^2 have been restricted to make the table clearer. The conclusions are still clear from this restricted data set.)

The dip at smaller angles is generated by the condition (7.3.2). Since $\tan(\beta_0) = v_1/v_2$, then $v_1 v_2 = v_0^2 \sin(2\beta_0)/2$, which will maximise for $\beta_0 = \pi/4 \approx 0.78$. Hence the term $-2\gamma_1/v_1 v_2$ will tend to be suppressed for β_0 in the range $0.7 \rightarrow 1$, making it more difficult to satisfy condition (7.3.2) in this region. Also, since $v_0^2/2 \approx 3 \times 10^4 \mathrm{GeV}^2$, all but the very largest values of γ_1 will be suppressed heavily.

7.3.2 Relative importance of the v/T,m/T constraints

The condition $m_i(T_c)/T_c < 1.6$, where the $m_i(T_c)$ are the masses of the particle content of the model, is a limit on the validity of the high-temperature expansion of the 1-loop effective potential. Hence, it does not describe any physics. So, the role of m/T in shaping the baryogenesis region should be understood before attempting to explain its structure. Graphs 7.4 show the number of points in the physical data set which failed the v/T condition only, the m/T condition only, or both conditions. It is the relative importance of the v/T and m/T constraints which determines whether the shape of the baryogenesis region can be considered as a physical phenomenon. Note that points in the sample which fail on both the v/T and m/T conditions should be considered as failures under m/T since the v/T value should be considered unreliable. So, in the following discussion I will consider the m/T cut to include failure under both conditions.

Fortunately, graphs 7.4 demonstrate that the v/T constraint dominates over the m/T cut in a large area of the μ_R^2, μ_I^2-grid. In particular, at small $|\mu_3^2|$ (less than $10 \times 10^3 \mathrm{GeV}^2$), the v/T constraint cuts $60 - 80\%$ of the sample points, while the m/T condition cuts less than 30%. **We conclude that, for small $|\mu_3^2|$, the shape of the baryogenesis region is physical.**

The v/T and m/T constraints are of roughly equal importance for $|\mu_3^2|$ around $18 \times 10^3 \mathrm{GeV}^2$, and above this the m/T cut begins to dominate. This is not surprising as we will see (section 7.4.1) that the scalar masses tend to increase with $|\mu_3^2|$, which will tend to increase the value of m/T.

The dominance of m/T in shaping the baryogenesis region at large $|\mu_3^2|$ ex-

plains the rapid fall-off in graphs 7.1, as was suggested in the preceeding section. However, the fall-off in baryon-preserving cases begins in the region of $|\mu_3^2|$ about 6000 GeV^2 which is dominated by the physically relevant v/T cut. So, the fall-off might well have a physical driving force as well as the limits of the mathematical procedures. Evidence for this physical driving force can be found in section 7.5, on the strength of the electroweak phase transition.

7.3.3 The role of v/T and m/T in the mass spectra

Analytical diagonalisation of the neutral scalar mass matrix is not possible in the most general form of the effective potential, even at zero temperature. (See appendix B.) Even if it was possible, we would still be limited in what can be learned, in an analytical fashion, about the mass spectrum. This is because the detailed shape of the spectrum derives from the effects of the constraints on the ratios $m(T_c)/T_c$ and $v(T_c)/T_c$, which are imposed at the critical temperature. It has already been stated that no useful analytical approach has been found for determining the minimum of the potential in the most general two Higgs-doublet model, and hence the v/T and m/T constraints are applied numerically.

However, in the preceding section, it was seen how these constraints act as a function of (μ_R^2, μ_I^2). The v/T constraint does dominate over a large area of the μ_R^2, μ_I^2-grid, implying that the shape of the baryogenesis region discovered in the numerical search is the result of physical effects rather than mathematical limits (at least in the region where v/T dominates.) In considering the scalar mass spectrum in the baryogenesis region, it will be important to reconsider the m/T and v/T cuts as a function of the scalar masses – otherwise we cannot determine whether the mass spectrum has been shaped by physics or mathematical validity limits.

The graphs 7.5 and 7.6 show the relative importance of the v/T and m/T cuts in the physical data set (as described in section 7.3) for each of the scalar masses.

From these graphs it is clear that the mass distributions will be shaped by two factors. The v/T constraint tends to dominate at low masses and m/T dominates

at large masses. This is not surprising – the large masses are in more danger of exceeding the m/T limit. However, in the case of the lightest neutral scalar, there is no clearly dominant constraint at any mass. This is easily explained by the way the cuts for m/T are made – if **any** of the scalar masses exceeds the mass limit, then the case is considered invalid. Hence, even the lightest examples of the light neutral scalar might be cut by the presence of a particularly massive heavy scalar, and this tends to spread the effects of the m/T condition over the entire range of masses for the lightest neutral scalar. The heaviest scalar will only fail the m/T condition when it is heavy, so the distribution of cuts is more localised.

However, apart from the lightest neutral Higgs, the shape of the mass distributions will be defined at their leading edge by v/T, and at their trailing edge by m/T.

7.4 Structure of the baryogenesis region

7.4.1 The scalar mass spectrum

Consider now the mass distributions in the baryogenesis region, as presented in graphs 7.7. These graphs show the frequencies of measured masses in the baryogenesis region for the five physical scalar particles at $T = 0$. These five particles are the charged scalars (denoted M_+ – both positively and negatively charged particles have the same mass), and the lightest, mid-weight, and heaviest neutral scalars (denoted M_1, M_2, M_3 respectively.) The graphs are derived from the baryogenesis data set; ie., the cases found to lie in the baryogenesis region.

The first important point to note from these distributions is that the masses extend well beyond the experimental lower bound, which is more encouraging news for electroweak baryogenesis. Their ranges are summarised in table 7.3.

The mass data for the graphs have **not** been cut to remove the experimentally disallowed cases but instead the cases are displayed and the experimental bound is shown. Even for the lightest neutral scalar, the masses lie in a range which extends above 200GeV and only a very small fraction of cases would be cut by

	Mass (GeV)		
	Lower limit	Modal value	Upper limit
Charged	40	200	300
Light neutral	40	140	240
Mid-weight neutral	60	190	300
Heavy neutral	90	250	300

Table 7.3: Scalar mass ranges in baryogenesis region after experimental cuts.

the experimental bound. Extending the experimental bound – even to go so far as doubling it to 80GeV – would have very little impact on the total size of the baryogenesis region. **This justifies the simple approach taken in applying the experimental bound.**

In order to understand the shapes of the distributions, it would be helpful to have explicit expressions for the masses. The expressions for the physical ($T = 0$) scalar masses can be considered separately for the charged and neutral sectors. The charged sector can be diagonalised in the general form but the neutral sector cannot. The charged sector provides masses

$$\begin{aligned} M^2_{\chi^\pm} &= 0 \\ M^2_{H^\pm} &= -\frac{v^2}{2}\left[\lambda_4 + \lambda_5\cos(2\theta_0) - \frac{2\left(\mu_R^2\cos(\theta_0) - \mu_I^2\sin(\theta_0)\right)}{v_1 v_2}\right] \end{aligned} \tag{7.4.5}$$

(7.4.6)

where the $\chi^\pm$ are the charged would-be Goldstone bosons "swallowed" by the $W^\pm$ bosons.

The neutral sector, in the simplified case with real μ_3^2, provides

$$\begin{aligned} M^2_{\chi^0} &= 0 \\ M^2_A &= \left(\mu_R^2 - \lambda_5 v_1 v_2\right)\frac{v^2}{v_1 v_2} \\ M^2_{H,h} &= \lambda_1 v_1^2 + \lambda_2 v_2^2 + \mu_R^2\frac{v^2}{2v_1 v_2} \\ &\quad \pm\sqrt{\left(\lambda_1 v_1^2 - \lambda_2 v_2^2 + \mu_R^2\frac{(v_2^2 - v_1^2)}{2v_1 v_2}\right)^2 + \left(\Sigma_+ v_1 v_2 - \mu_R^2\right)^2} \end{aligned} \tag{7.4.7}$$

where χ^0 is the neutral would-be Goldstone boson "swallowed" by the Z^0 boson; A is the pseudo-scalar; and H, h are the heavier and lighter scalars. (Recall that with μ_3^2 complex, the scalar/pseudoscalar distinction is no longer possible and the scalars will only be identified by their relative masses - lightest, mid-weight and heaviest.)

Graphs 7.5 and 7.6 show the mass spectrum in the basis space (on the curves marked "All cases") and hence the steeply rising leading edges in the mass distributions of the baryogenesis region could be explained in part by the same behaviour which appears in the basis space. However, about 95% of the sample in the basis space is removed by the cuts in v/T and m/T, and hence it is more believable that the v/T constraint is the primary factor in shaping this part of the distribution.

Part of the explanation might lie in the CP-violating parameter μ_3^2. Considering the expressions above for the physical scalar masses we see that the masses tend to increase with $|\mu_3^2|$. Admittedly, the expressions above for the neutral masses do not include complex μ_3^2, but the analytical masses do increase with μ_R^2 and the graphs 7.10 and 7.11 demonstrate numerically that the mean scalar masses increase with $|\mu_3^2|$ in the more general complex case.

7.5 Strength of the electroweak phase transition

The strength of the phase transition is the most important factor in determining the suitability of a region for preservation of the BAU. Since $v(T_c)$ cannot be determined analytically, it is necessary to derive an understanding of the behaviour of the transition order parameter $v(T_c)/T_c$ from the numerical search.

Graph 7.12 shows the mean scalar VEV in the baryogenesis region as a function of μ_R^2, μ_I^2. Graph 7.13 shows the mean critical temperature in the baryogenesis region against (μ_R^2, μ_I^2).

It has already been seen that the physical masses tend to rise with $|\mu_3^2|$ (section 7.4.1) and that this can be explained analytically for the charged scalars (and semi-analytically for the neutral scalars.) The rise in mean T_c can also be explained

analytically.

The expression for the critical temperature is

$$T_c^2 = \max\left\{\frac{(\mu_1^2 B' + \mu_2^2 A') \pm \sqrt{(\mu_1^2 B' - \mu_2^2 A')^2 + 4A'B'|\mu_3^2|^2}}{2A'B'}\right\} \tag{7.5.8}$$

where

$$\begin{aligned} A' &= 6\lambda_1 + 2\lambda_3 + \lambda_4 + 3g^2 + g'^2 + 4g_t^2 \\ B' &= 6\lambda_2 + 2\lambda_3 + \lambda_4 + 3g^2 + g'^2 + 4g_b^2 + \frac{4}{3}g_\tau^2 \end{aligned} \tag{7.5.9}$$

A' and B' will tend to positive since only the λ_3 and λ_4 terms can be negative. Hence, the μ_R^2 and μ_I^2 dependency of T_c^2 above will tend to increase T_c with $|\mu_3^2|$.

However, for $(A'B') < 0$, it is possible for μ_3^2 to decrease the critical temperature, so the rate of increase in the mean critical temperature will be slowed by this possibility. Certainly, the mean value of T_c rises more slowly with $|\mu_3^2|$ than the mean value of the charged or mid-weight neutral Higgs masses (graphs 7.10,7.11.) Hence the masses "catch up" with T_c as $|\mu_3^2|$ increases, and the m/T limit becomes stronger, as observed.

Comparison of the mean VEV and mean T_c graphs leads to a prediction of the effects of μ_3^2 on the strength parameter. The range of $\overline{|v(T_c)|}$ over the μ_R^2, μ_I^2-grid is only $175 \rightarrow 205$GeV, while $\overline{T_c}$ ranges from $150 \rightarrow 210$GeV. Hence, it is expected that the phase transition will get **weaker** with $|\mu_3^2|$. This behaviour can be observed in graph 7.14.

This behaviour might be considered unexpected, since μ_3^2 tends to increase the scalar masses, thereby increasing their contribution to the $T\mathrm{Tr}\,[M^3]$ term and strengthening the first order $EWPT$. However, it is clear that μ_3^2 tends to scale up all of the energies involved in the effective potential – the masses, the critical temperature, and the VEV of the scalar fields. It is the detailed interplay between all of these factors which determines the strength of the phase transition.

Note that the decreasing strength of the phase transition will act as the physical driving force to reduce the size of the baryogenesis region at larger μ_3^2, as suggested in subsection 7.3.2.

7.5.1 Sensitivity to limits on v/T

Consider the graph 7.15, which shows the distribution of the transition strength parameter as a fraction of the total number of cases sampled in the baryogenesis region. The distribution is heavily biased towards weak phase transitions. There is also a long tail which stretches out well beyond $v/T = 1.5$, but it is barely visible on the scale of the graph.

This distribution poses a serious threat to electroweak baryogenesis in the two Higgs-doublet models. The constraint imposed on v/T was chosen to be 1 for this work. However, this is not a precise figure. Indeed, the derivation of the limit presented in chapter 6 suggested a lower limit of 1.1, which was then eased in order to allow us to study the effects of varying the constraint in just this way. If we did force $v/T > 1.1$, then graph 7.15 shows that about 70% of the baryogenesis region would now become invalid.

The conclusion in much of the literature is that the electroweak phase transition is only ever weakly first order, whether it is tested in the Standard Model or its extensions. The same conclusion has been reached here, and in light of the weight of work suggesting that the phase transition is weak it is perhaps not surprising that the distribution is so heavily biased. This is disappointing news, however, for electroweak baryogenesis.

7.6 Conclusions

The results of this numerical search provide a great deal of encouragement for electroweak baryogenesis in two Higgs-doublet models. This is tempered, however, by the disappointingly weak phase transition and the sensitivity to the exact limit on the strength parameter $v(T_c)/T_c$.

A baryogenesis region was found, with a volume of 5% that of the basis space. (The volumes of the basis space and the baryogenesis region were determined using a simple Cartesian metric - ie., the 8 free parameters represent 8 orthogonal dimensions.) This is a fairly small proportion of the basis space but it is not negligible. Also, it must be remembered that this is only a lower limit to the

size of the region since it reaches beyond the boundaries of validity for the high-temparature expansion of V_{eff}. How much more of the space might be included in the baryogenesis region by a study using a more effective expansion can only be guessed at, but there is evidence (the weakening of the phase transition with $|\mu_3^2|$) that the baryogenesis region will not extend indefinitely beyond the region of validity of this study.

The scalar mass spectrum of the baryogenesis region is also encouraging since it reaches well beyond the experimental lower bounds on the Higgs mass in two-doublet models. The region is also very insensitive to increases in the experimental bounds. It has been seen that the leading edges of the mass spectra are controlled by the physical limit on v/T, but the trailing edges are shaped by the mathematical m/T limit. So, once again, the mass spectra might extend even further if the region of low temperatures ($m/T > 1.6$) can be penetrated.

The inclusion of the CP-violating term with complex coupling μ_3^2 has played a major part in the formation of the baryogenesis region. The graphs 7.1 demonstrate that non-zero μ_3^2 is favoured in the baryogenesis region. However, these graphs also demonstrate that this enhancing effect is dependent on the size of $|\mu_3^2|$ and not on the fact that it is complex. (The graphs are roughly rotationally symmetric about $|\mu_3^2| = 0$.) Hence, the fact that this term can be CP-violating does not appear to be a major factor.

The μ_3^2 term tends to increase the critical temperature, the Higgs masses and the VEV of the scalar fields but we have seen that the higher rate of increase in T_c (over that in $v(T_c)$) with $|\mu_3^2|$ results in a weakening of the phase transition, somewhat unexpectedly. However, the extra freedom provided by the inclusion of μ_3^2 more than compensates this weakening and results in a non-negligible baryogenesis region.

Unfortunately, the phase transition remains generally quite weak, as shown in the distribution graph 7.15. This bias towards weaker transitions leaves the baryogenesis region under threat from the details of the limit on the strength parameter $v(T_c)/T_c$, which could dramatically cut the volume of the region with shifts of less than 10%.

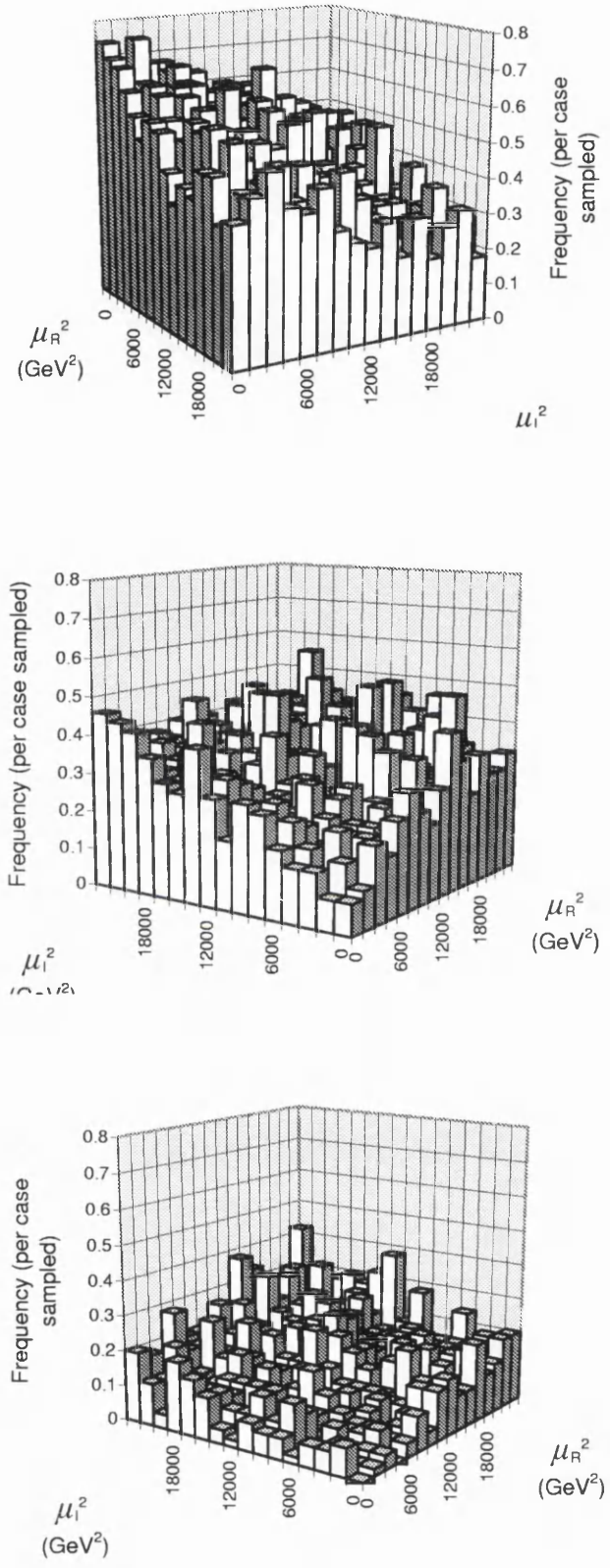

Figure 7.4: Number of cases in the physical data set cut from the baryogenesis region by the v/T limit alone (top), the m/T alone (middle), and both limits (bottom). Data presented as a fraction of the number points sampled (50) at each point in the μ_3^2, μ_I^2-grid. Note that the first graph is viewed from a different

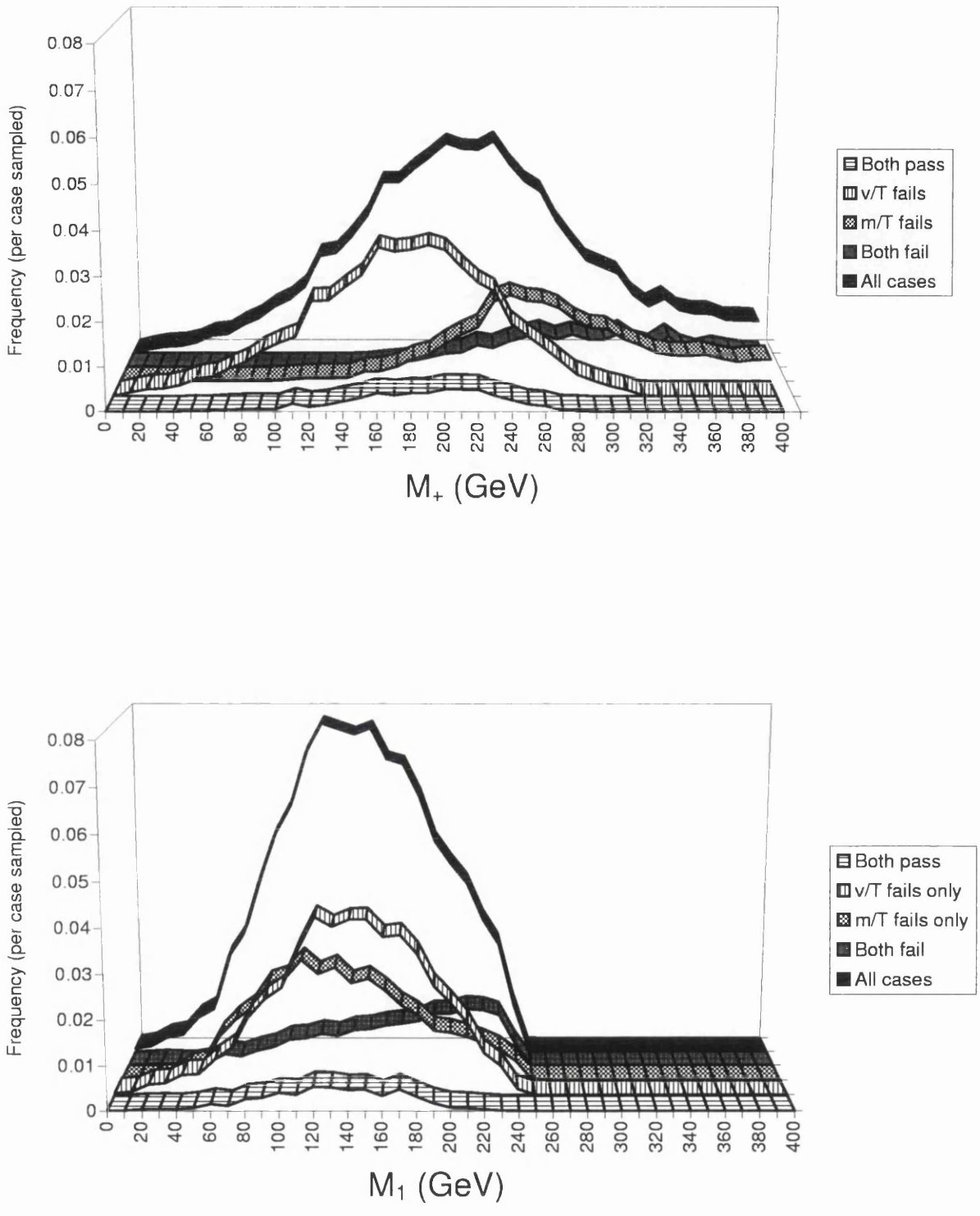

Figure 7.5: Relative importance of m/T and v/T cuts in the structure of the scalar mass spectrum for charged and lightest neutral Higgs masses. Data presented as fraction of total number of masses measured.

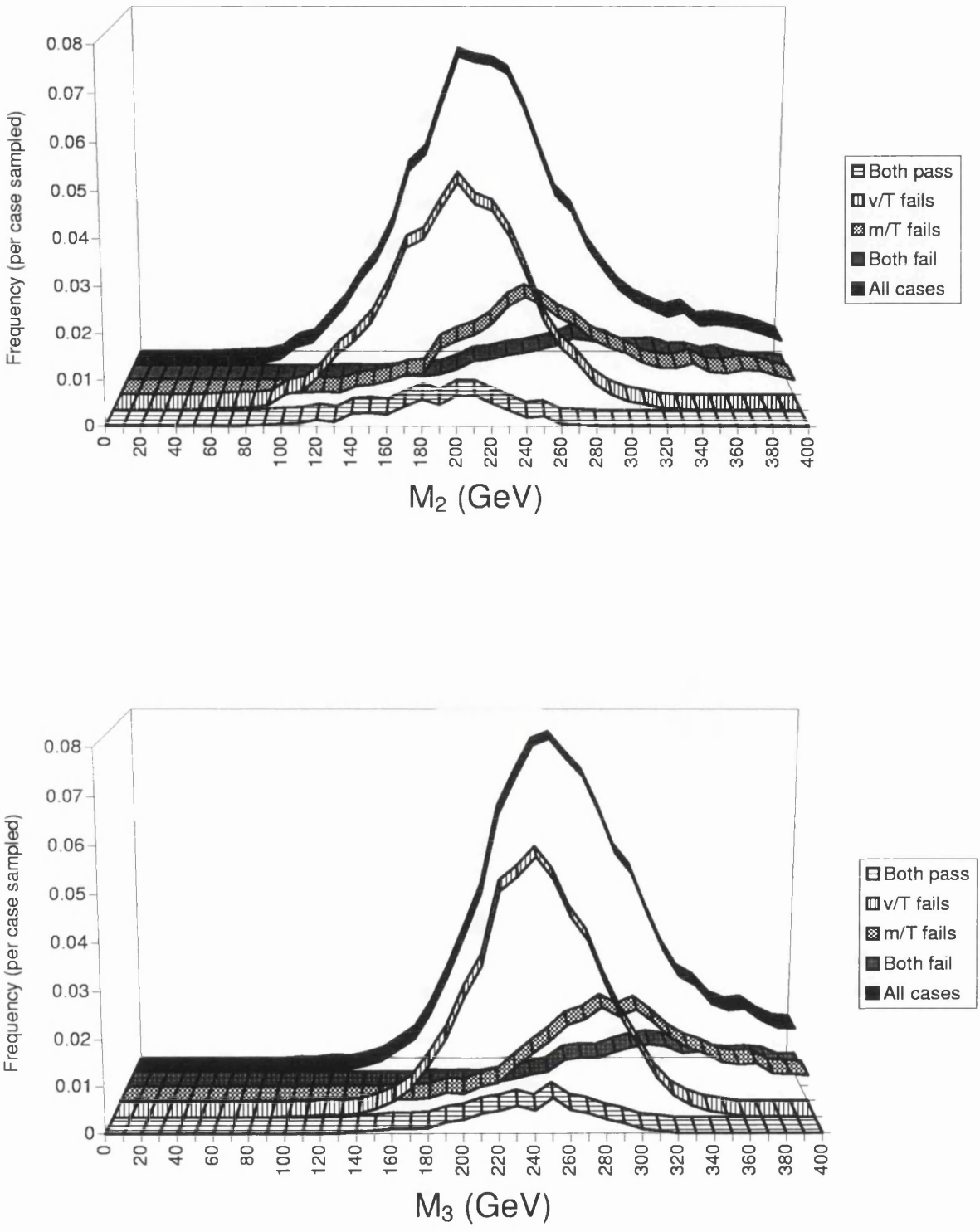

Figure 7.6: Relative importance of m/T and v/T cuts in the structure of the scalar mass spectrum for the mid-weight and heaviest neutral Higgs masses. Data presented as fraction of total number of masses measured.

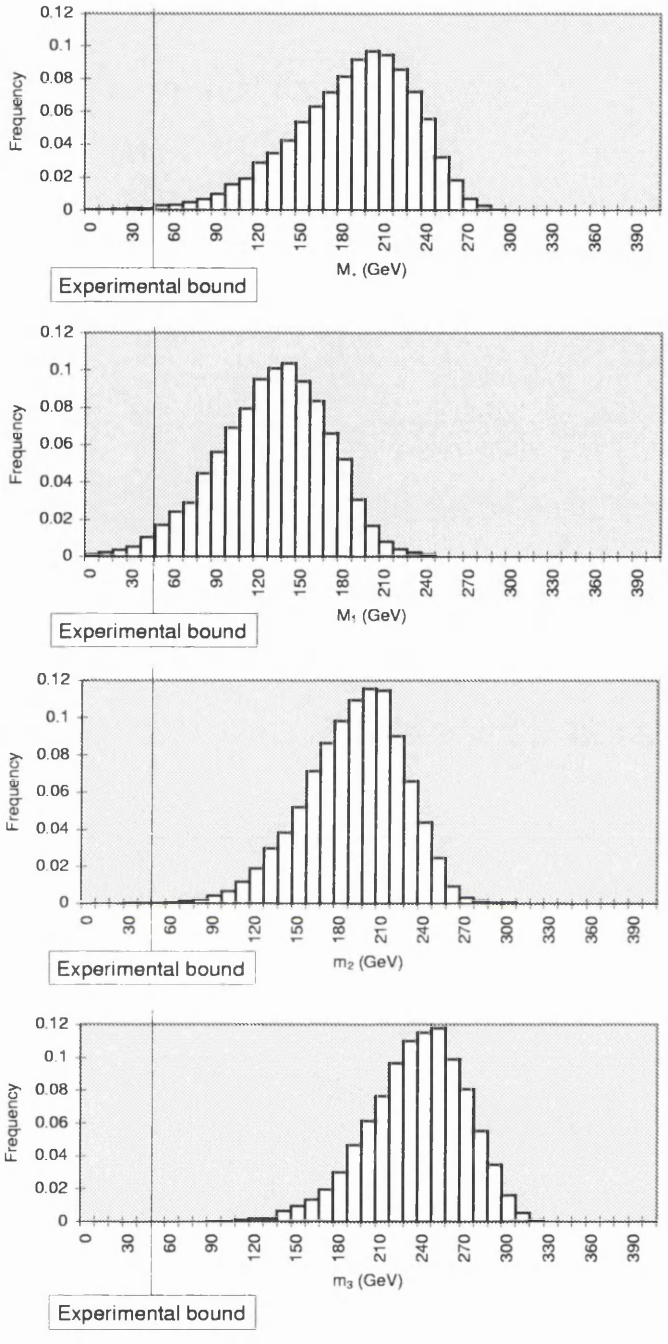

Figure 7.7: The scalar mass distributions in the baryogenesis region. The scalars are denominated as follows – M_+, charged mass; M_1, M_2, M_3, the lightest, mid-weight, and heaviest neutral scalars respectively. The data are presented as fractions of the total number of masses measured. The distributions were derived from the baryogenesis data set, defined in section 7.3.

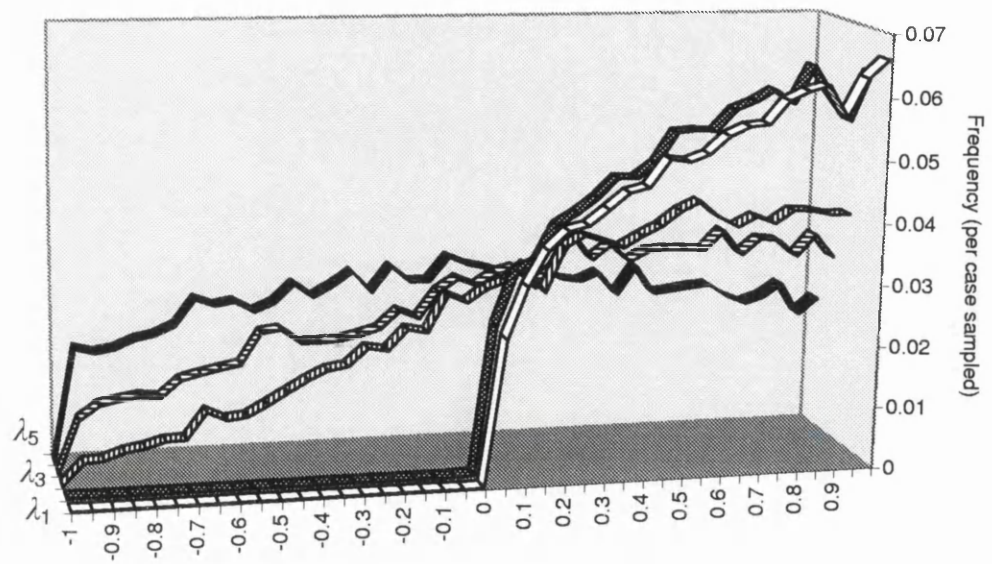

Figure 7.8: Distributions of the λ_i in the basis space. The data are presented as a fraction of the total number of λ_i measured.

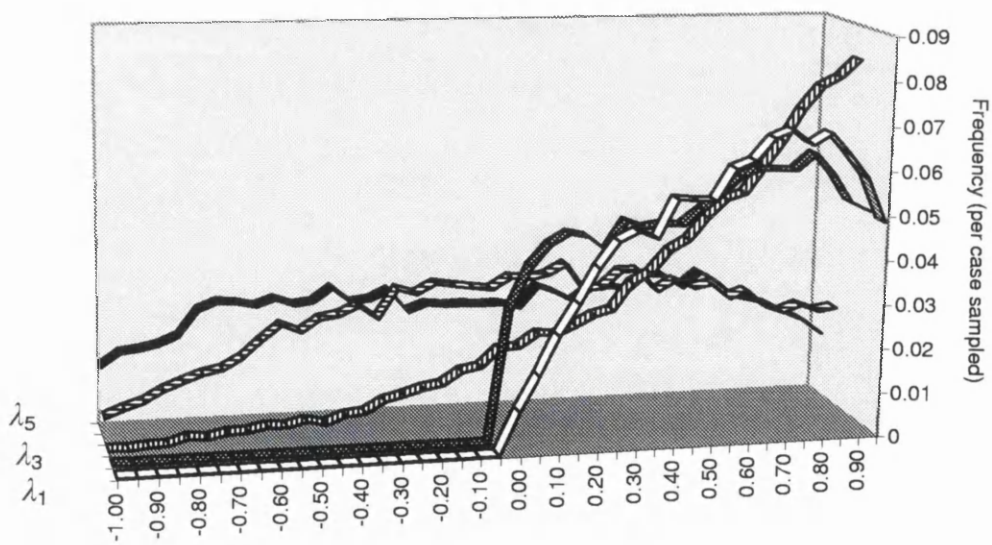

Figure 7.9: Distributions of the λ_i in the baryogenesis region. The data are presented as a fraction of the total number of λ_i measured.

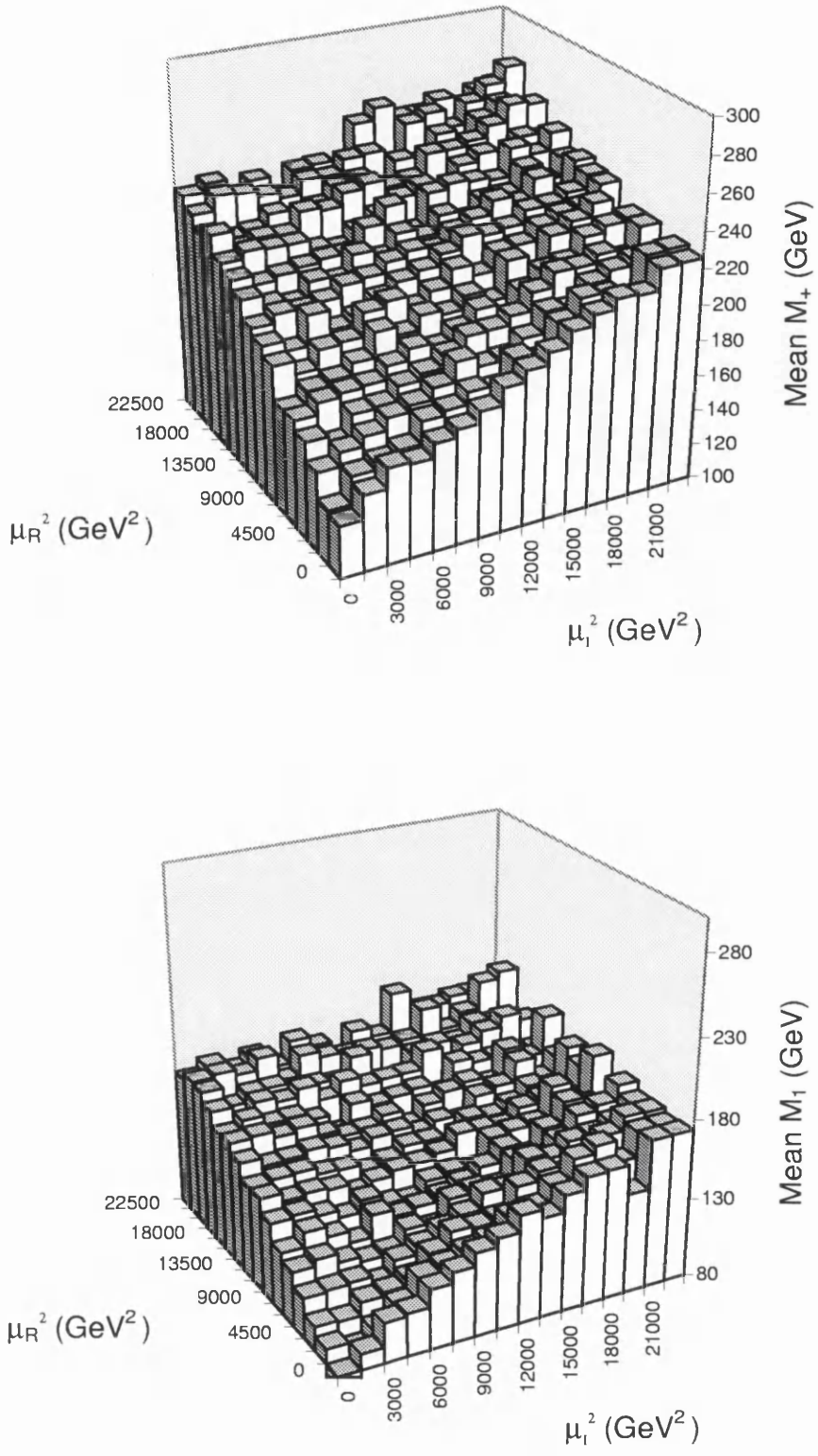

Figure 7.10: The mean masses of the charged (M_+) and lightest neutral Higgs (M_1) plotted against (μ_R^2, μ_I^2). The data are presented as a fraction of the total number of basis space points sampled at each point on the μ_R^2, μ_I^2-grid.

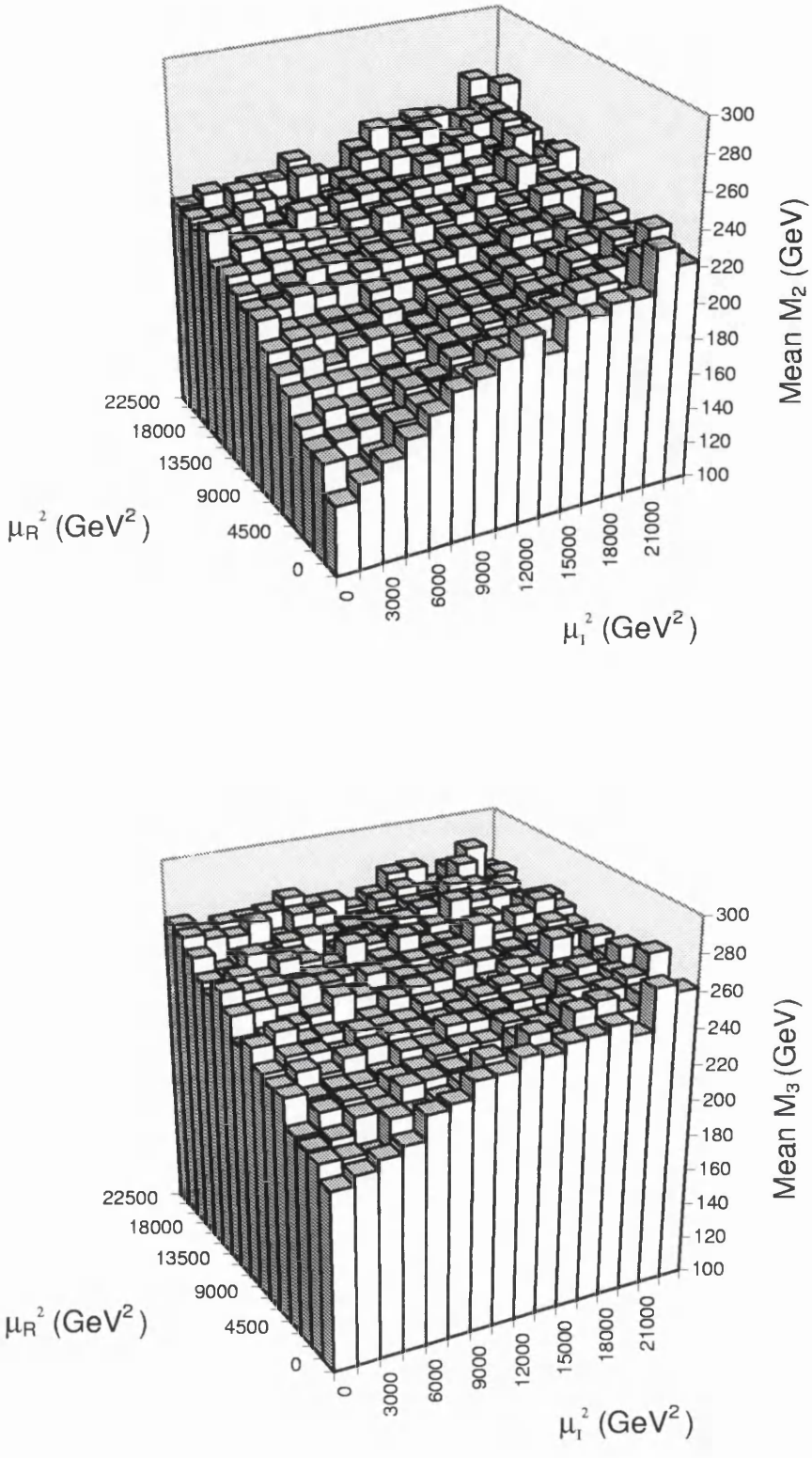

Figure 7.11: The mean masses of the mid-weight (M_2) and heaviest neutral Higgs (M_3) plotted against (μ_R^2, μ_I^2). The data are presented as a fraction of the total number of basis space points sampled at each point on the μ_R^2, μ_I^2-grid.

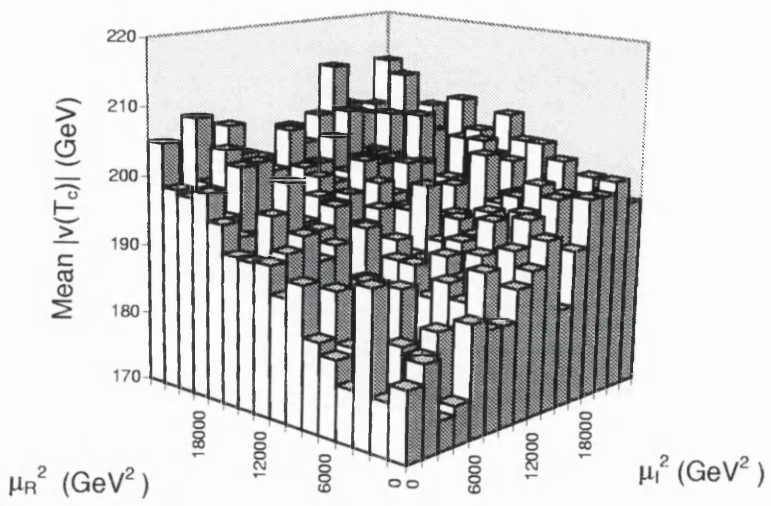

Figure 7.12: Mean VEV of the scalar fields in the baryogenesis region, plotted as a function of (μ_R^2, μ_I^2).

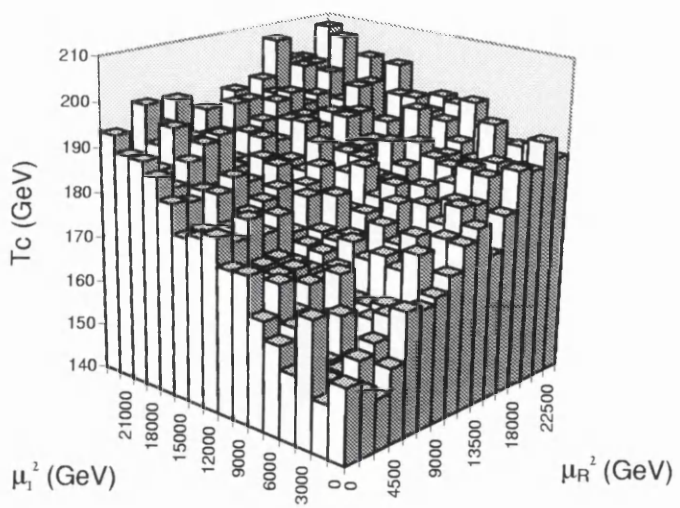

Figure 7.13: Mean T_c of the scalar fields in the baryogenesis region, plotted as a function of (μ_R^2, μ_I^2).

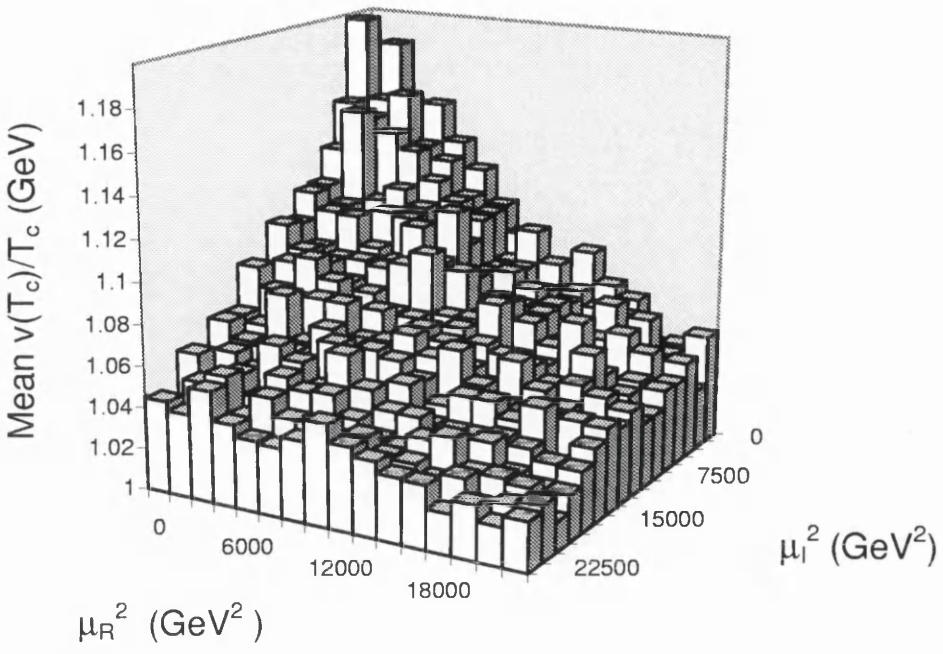

Figure 7.14: Mean value of the phase transition strength parameter $v(T_c)/T_c$ in the baryogenesis region as a function of (μ_R^2, μ_I^2).

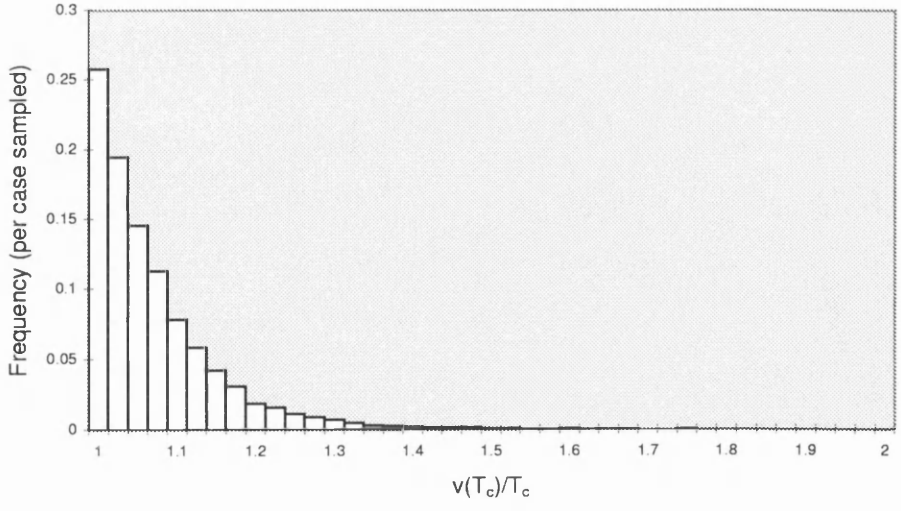

Figure 7.15: Distribution of the phase transition order parameter $v(T_c)/T_c$ in the baryogenesis region.

Chapter 8

Special case - the minimal supersymmetric model

The minimal supersymmetry model ($MSSM$) is a special case of the general two Higgs-doublet model. Minimal SUSY requires two Higgs doublets in the scalar sector [32, 33, 34] and also requires the up-type and down-type quarks to couple separately to only one doublet, as we prescribed in chapter 3. (This was made a requirement in our model for the purposes of agreement with the $MSSM$, as well as the possibility of explaining the differences between the up-type and down-type quark masses in terms of the scalar doublet VEV's.)

Supersymmetry can constrain some of the coupling constants in the effective potential to be well-defined functions of the gauge couplings at the supersymmetry scale, Λ_{susy} say. In turn , this will constrain the low energy values of these couplings. The number of free couplings in the effective potential is thereby reduced and the $MSSM$ becomes a special case of the work in the preceeding chapters.

The work in this chapter is an application of the methods and tools I have developed for the study of the two Higgs-doublet effective potential. It is also a continuation of work originally undertaken by Moorhouse, Froggatt and Knowles [23], where they investigated quasi-fixed point solutions of the renormalisation group equations (RGE) in minimal SUSY. They kindly provided the sets of coupling parameters used in this chapter. In what follows, I will explain qualitatively how these parameter sets are produced but I will not consider the mathematical

details since they were beyond the scope of the work.

8.1 The MSSM-restricted effective potential

Supersymmetric models predict another cosmological phase transition at energies much higher than the electroweak phase transition. In the same fashion as the $EWPT$, at temperatures higher than Λ_{susy} there is an explicit symmetry (supersymmetry), and at lower temperatures it is spontaneously broken. We will consider the simplest possible SUSY-breaking case, with a complete spectrum of supersymmetric particles and perfect supersymmetry above the SUSY scale, Λ_{susy}. It is assumed that there is a "desert" between the electroweak and SUSY scales – ie., no new physics comes into play before reaching the SUSY scale. Hence, below Λ_{susy} there will only be the known particles and the Higgs scalars.

The $MSSM$ stipulates the following forms for the coupling constants of the effective potential **at the supersymmetry scale** [23]–

$$\begin{aligned} \lambda_1^s = \lambda_2^s &= \frac{(g_s^2 + g_s'^2)}{8} \\ \lambda_3^s &= (g_s^2 - g_s'^2)4 \\ \lambda_4^s &= \frac{-g_s^2}{2} \\ \lambda_5^s &= 0 \end{aligned} \tag{8.1.1}$$

where g_s and g_s' are the electroweak gauge couplings **at the supersymmetry scale.** In other words, we must determine the gauge couplings at the SUSY scale before we can determine the λ_i. This is achieved via the renormalisation group equations, which describe how the coupling constants run with energy. Hence, it is possible to estimate the gauge couplings at Λ_{susy} by running g and g' up from their electroweak values to their SUSY values. (Note that g_s and g_s' will depend on the choice of Λ_{susy}.)

It is now possible to determine the SUSY scale values for the couplings of the effective potential, λ_i^s. Note that λ_1 and λ_2 are identical at this scale, and that λ_5 is necessarily zero. In order to apply these couplings in the effective potential at the

EWPT it is necessary to use the RGE to run the couplings down to electroweak energies.

The λ_i are now all defined parameters of the effective potential. Since λ_5 is identically zero at SUSY scales, it will remain so at electroweak energies. Thus it is possible to define μ_3^2 to be real without loss of generality. Note that this removes the explicit CP-violation in the tree level potential.

Of the 8 free parameters described in the general case in previous chapters – namely $\lambda_i, \beta_0, \mu_R^2, \mu_I^2$ – we now have six of these prescribed. Hence, there appear to be only 2 free parameters in the model – μ_R^2 and β_0. However, there is actually a third free parameter in this special case – the supersymmetry breaking scale. This is a free parameter in the determination of the coupling constants to be used in the effective potential and is not a parameter of V_{eff} itself. Hence, it has no effect on the details of the computational methods developed earlier in this thesis.

8.2 Application of the supersymmetry restrictions to the numerical methods

The application of the supersymmetry restrictions on the effective potential in our numerical model is relatively straightforward. The λ_i are now prescribed by the RGE, and μ_I^2 is zero, so we have only a 2-dimensional parameter space and the use of a random sample search technique is not so important now. The parameter space available is small enough to make a thorough and methodical search of the space. Hence the random sampling approach was abandoned, and instead μ_R^2 is varied over a range of values for given β_0.

Since the μ_3^2 coupling is now restricted to be real, θ_0 is also restricted to 2π for $\mu_R^2 \neq 0$, and $3\pi/2$ for $\mu_R^2 = 0$. So solving for θ_0 is not required in the numerical model.

The coupling parameter sets were provided, at the electroweak scale, for various β. They are described below in table 8.1. The top mass was taken to be 175GeV.

SUSY constrained coupling parameter data sets at $\Lambda_{susy} = 1\text{TeV}$						
$\tan(\beta_0)$	λ_1	λ_2	λ_3	λ_4	λ_5	β_0
0.5	0.1495	0.0636	0.0617	-0.2111	0	0.464
0.6	0.7001	0.0648	0.0534	-0.2104	0	0.54
0.7	0.4713	0.0648	0.0569	-0.2104	0	0.611
0.8	0.3496	0.0648	0.059	-0.2103	0	0.675
0.9	0.2787	0.0647	0.0605	-0.2103	0	0.733
1.0	0.2868	0.0637	0.0575	-0.2109	0	0.785

Table 8.1: Coupling parameter sets derived via the RGE in the minimal supersymmetric model. ($m_t = 175\text{GeV}$, $\Lambda_{susy} = 1\text{TeV}$)

The SUSY scale is unknown, although it has been popular in recent years to consider relatively low energies. Various SUSY-breaking energies were investigated in the range 0.5TeV to 2.5TeV. The conclusions in each case, however, were essentially the same. Hence, as a typical example of the results obtained, let us set $\Lambda_{susy} = 1\text{TeV}$.

For each of the parameter sets in table 8.1 the computational model was run for μ_R^2 in the range 0GeV^2 to 10^4GeV^2. All physical constraints, and the $v(T_c)/T_c$ and $m_i(T_c)/T_c$ limits were imposed as before.

8.3 Supersymmetry results

The results of the search for a baryogenesis region in the minimal supersymmetric model are summarised in the graphs 8.1 and 8.2.

Generally speaking, it proved difficult to find a baryogenesis region in the *MSSM*. Graphs 8.1 show the lines of constant β_0 which were found to lie in the baryogenesis region. They describe how the scalar masses vary with μ_R^2 – note that the masses increase with $|\mu_3^2|$, as seen in the non-SUSY results of chapter 7. The kinks in the graph of the lightest neutral scalar are simply the points at which two of the mass eigenvalues become degenerate.

The graphs clearly show that the viable ranges of μ_R^2 are fairly restricted. It

was also found that the baryogenesis region was restricted in $\tan(\beta_0)$. For $\tan(\beta_0)$ greater than about 1 there is no baryon preserving region, and large values of $\tan(\beta_0)$ also tend to force the lightest scalar mass down towards the experimental lower bound (as seen in graph 8.1.)

Interestingly, the $EWPT$ is rather strong in the cases studied, as seen in graph 8.2. However, it is at its weakest for large β_0, which explains why it becomes difficult to find a baryogenesis region for $\tan(\beta_0) > 1$.

The transition is much stronger at smaller $\tan(\beta_0)$, and it is the m/T constraint which cuts most of the μ_R^2 range out of the baryogenesis region in this regime. Hence, with an improved approximation to the 1-loop potential it might be possible to reliably find BAU-preserving regions at even lower $\tan(\beta_0)$. However, small $\tan(\beta_0)$ is disfavoured by the large top quark mass of 175GeV. In particular, assuming there is a desert beyond the $MSSM$ scale, up to the GUT scale say, then the Yukawa RGE's imply that the top quark mass must be less than the infrared fixed point value, m_t^{FP} [32]. For a GUT scale of $\Lambda_{GUT} = 10^{16}$GeV, then $m_t^{FP} \approx 200\sin(\beta_0)$GeV, and $\tan(\beta_0) < 1$ would then force the top mass below $140GeV$.

8.4 MSSM conclusions

We have seen the specialisation of the general two Higgs-doublet model developed in earlier chapters to include the constraints present in the minimal supersymmetric model. This constrains the parameters of the two Higgs-doublet effective potential by relating them to the electroweak gauge couplings at the SUSY scale. This drastically reduces the number of free parameters in the effective potential, leaving only μ_R^2 and β_0.

The search for an $MSSM$ baryogenesis region was conducted for various values of the SUSY scale Λ_{susy}. The results of that search are represented here by a typical example with $\Lambda_{susy} = 1$TeV. A small baryogenesis region can be found as depicted in graphs 8.1 with an apparently healthy $EWPT$ at small $\tan(\beta_0)$. However, the small $\tan(\beta_0)$ region is disfavoured by the large top mass of around

175GeV. This agrees with the results of Espinosa *et al.* [34] who found a small baryogenesis region beyond the limits of the basis space in our work, but did not report any results for the small $\tan(\beta_0)$ regime.

However, more recent results, using perturbative [48] and lattice [49] techniques, have found a window in the MSSM parameter space. By introducing a light right-handed stop with mass of about $150 - 200$GeV, a window was found for $\tan\beta \approx 2$ with $m_{h^0} \leq 85$GeV. Since we assume no SUSY particles below the SUSY scale of 1TeV in the results presented here, these latest searches lie outside the range of our search. It is interesting to note that $\tan\beta \approx 2$ is not endangered by the heavy top mass.

In conclusion, our results do not hold out much hope for electroweak baryogenesis in the minimal supersymmetric model with SUSY scales of around 1TeV. On the other hand, the more positive results of other searches suggest that relatively low SUSY–breaking scales might still yield a satisfactory EWPT capable of sustaining the BAU. However, the results of chapter 7 suggest that the extra degrees of freedom available in the general two Higgs-doublet model make the preservation of the BAU easier than in MSSM models, and next to minimal supersymmetric theories with two scalar doublets (and further extensions) should also be considered.

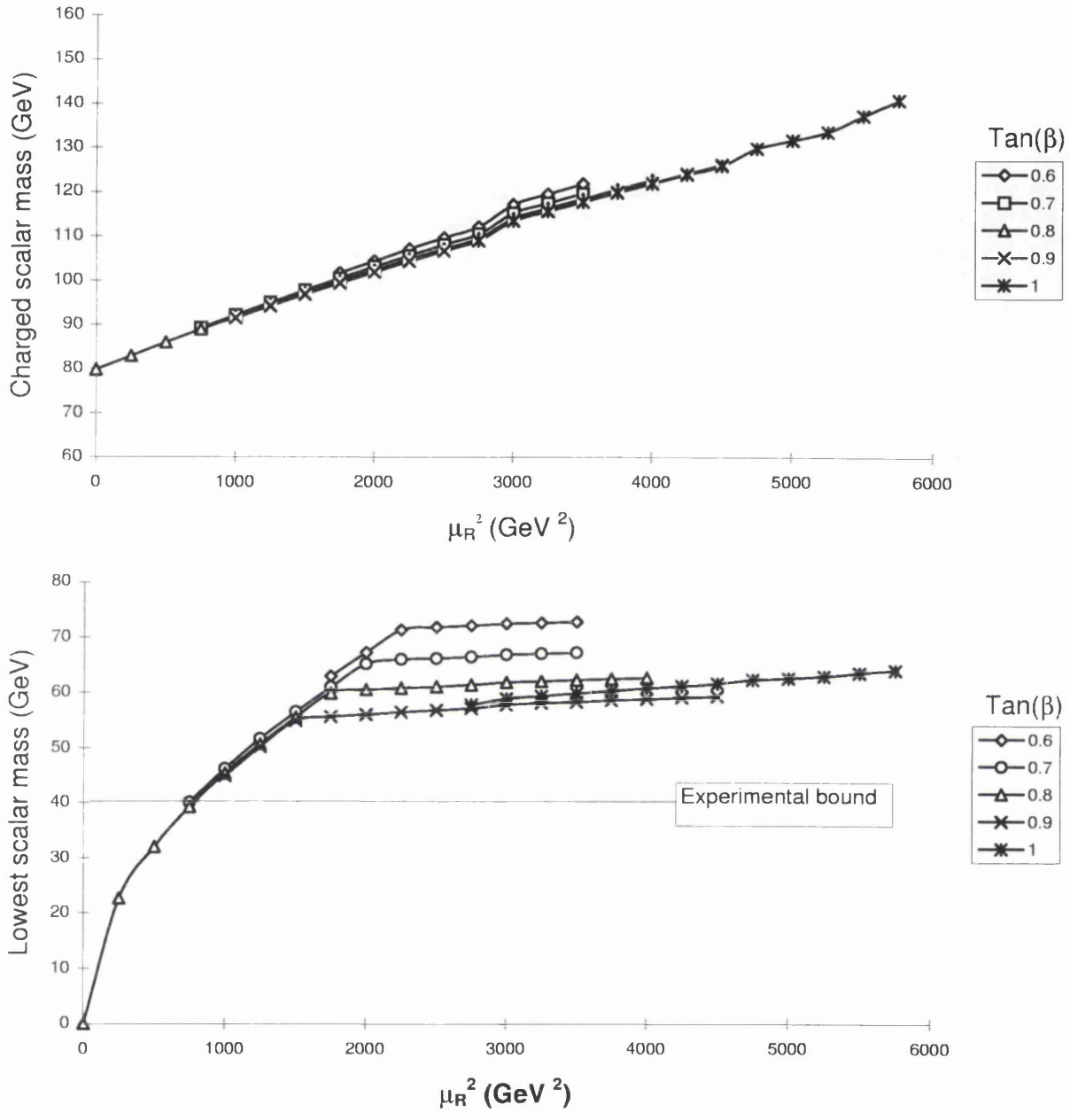

Figure 8.1: The masses of the charged and lightest neutral Higgs in the baryogenesis region of the minimal supersymmetric model. The data are plotted for various values of β_0.

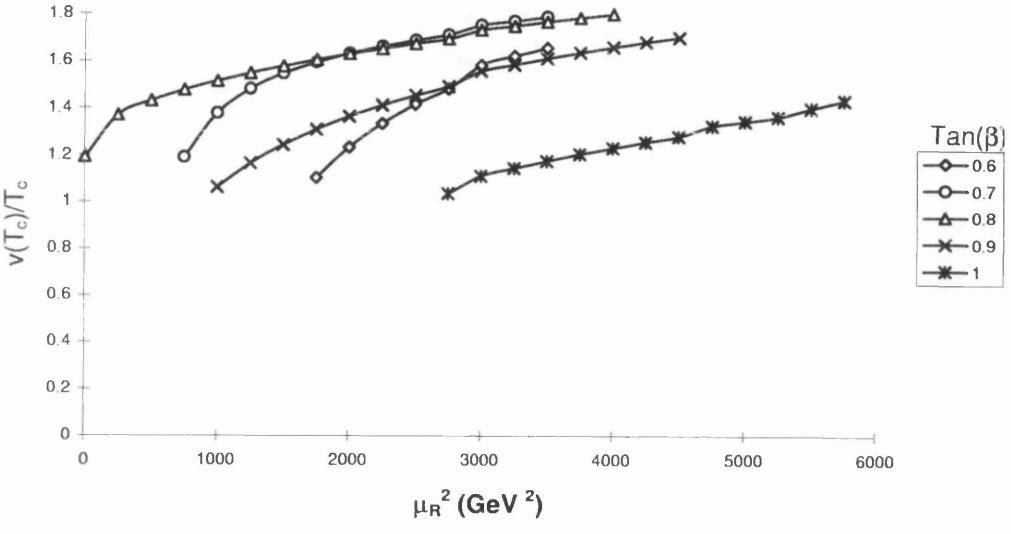

Figure 8.2: $v(T_c)/T_c$ plotted as a function of μ_R^2 for various values of β_0 in the baryogenesis region of the MSSM model

Chapter 9

Conclusion

The work in this thesis set out to consider the possibility of saving baryogenesis at the electroweak scale from the rise of the experimental lower bound on the Standard Model Higgs mass. Broadly speaking, the results suggest that the two-Higgs-doublet models **are** capable of incorporating the heavy Higgs scalars required for a strongly first order electroweak phase transition. Overall, 5% of the basis space sampled was found to support electroweak baryogenesis, and within the baryogenesis region the scalar masses were found to lie well beyond the reach of the experimental bounds – even the lightest neutral scalar extends to over 200GeV, and has a modal mass value of around 140GeV.

The decision to consider the effective potential in the most general form possible for two scalar doublets has proved fruitful, since the inclusion of the CP-violating, complex coupling μ_3^2 has important effects in expanding the size of the baryogenesis region. It is possible that this generality has contributed to the encouraging conclusions reached here, which are not matched by similar parameter space searches.

For example, the search performed by Turok and Zadrozny [4] in 1991 had several differences from the search undertaken here. Their potential was developed in the unitary gauge, which only included the 5 physical scalars (and not the would-be Goldstones), thereby introducing problems in their mass matrices. The Landau gauge used here avoids this problem. They also restricted the number of free coupling parameters by introducing symmetries in the λ_i and $\mu_{1,2}^2$ for the pur-

poses of algebraic simplicity, also justifying this with arguments on likely regions of interest in the parameter space. However, I impose no such restrictions. Turok was also more restrictive in the range of parameters in the basis space, failed to include the correction for the contribution of the W and Z bosons described by Dine *et al.* [35] and, perhaps most importantly, considered a CP-invariant potential. The inclusion of the complex μ_3^2 coupling enhances the volume of the baryogenesis region found in our two Higgs-doublet model and tends to force the scalar masses higher.

In Turok's conclusions, he finds a large baryogenesis region (but his basis space was smaller, which possibly biases the definition of large and small) but with an upper limit on the lightest Higgs mass of only 120GeV. The lightest Higgs scalar in our model can range up to 240GeV, doubling the limit found by Turok.

Last year (1996), another similar search was performed by Cline and Lemieux [37]. They do report a significant increase in the baryogenesis regions for small, non-zero μ_3^2, although they claim this is spurious and is caused by the appearance of a "double minimum", one of which is a false vacuum – for this reason they do not believe that their results corroborate our own here. However, they specifically follow the prescription of Turok and Zadrozny for the symmetrisation of their coupling parameters and do not consider complex μ_3^2, so it is difficult to see that their results are directly comparable in any case. However, in Cline's conclusion they do find encouraging results for the existence of electroweak baryogenesis in two Higgs-doublet models and can achieve heavy scalars up to 300GeV, except for the case of large real μ_3^2. These results are not in disagreement with my own conclusions.

The more restrictive parameter space of the minimal supersymmetric model did not fair so well in the search for electroweak baryogenesis. Although a baryogenesis region was found, it was only possible to sustain the preservation of the BAU at small $\tan(\beta_0)$, which is strongly disfavoured by the existence of the heavy top mass at 175GeV. These results are in agreement with the general conclusions drawn by Espinosa *et al.* [34], although they were working in a different region of the basis space from the search made in chapter 8. Combining the conclu-

sions of both searches, performed in different regions of the parameter space, it seems unlikely that the electroweak phase transition can support baryogenesis in the minimal supersymmetric models. However, as reported in chapter 8, both of these searches used SUSY scales of 1TeV or more, and recent results [48, 49] suggest that, with lower SUSY–breaking scales, a window can be found in the MSSM paramter space which evades the Higgs experimental mass bounds and the constraints on $\tan\beta$ introduced by a heavy top quark. Hence, the MSSM cannot be competely ruled out in baryogenesis scenarios by the results presented here.

9.1 The future of baryogenesis in two Higgs-doublet models

Although the results for the general two Higgs-doublet models were positive, there is a dangerous sensitivity to the v/T limit imposed to ensure a strong first order $EWPT$. A better understanding of the exact value of this strength parameter is required in order to be sure of the 5% baryogenesis region found here.

However, an even more dangerous threat might arise from non-perturbative effects. The perturbative approach used in my work is known to have problems in exactly those regions in which we are interested – ie., at temperatures close to the critical temperature, and for small field values. Since we defined the critical temperature in terms of the curvature of the effective potential at the origin (ie., small fields), it is possible that the critical temperature calculations are not accurate.

Early numerical investigations in this problem, notably by Shaposhnikov [47] suggested that the non-perturbative effects play a very important role in the details of the $EWPT$ and that the perturbative calculations were no longer to be trusted. In particular, he performed lattice studies of the $EWPT$ in the Standard Model and found strong first order phase transitions but with critical temperatures close to, but less than, the critical temperature derived from the perturbative calculations. (It was these suggestions which first led me to consider

the T_c correction described in Appendix A.) These results suggested that the BAU was sustainable in the SM after all. However, Shaposhnikov's later work in this area [50] has refined the lattice calculations and there is now a rough agreement between the lattice and effective potential approaches. Hence, the SM's reprieve was short-lived and it certainly seems that the SM has once again proved unsuitable to sustain the BAU — we must continue to look for suitable extensions of the model if we want a consistent description of baryogenesis.

There are undoubtedly some difficult problems to overcome in the use of the perturbative expansions of the effective potential. In particular, the results here show how sensitive the baryogenesis region is to the exact value of the limit imposed on the sphaleron rate, with small changes capable of almost wiping out the BAU entirely. At the same time, an extension of the region of validity of the expansion could stretch the boundaries of the baryogenesis region. Hence, with a better understanding of the limit on the strength parameter, or of the high-temperature expansion limit, it might be possible to extend and consolidate the perturbative results.

Appendix A

Attempted analytical correction to T_C

This appendix describes an attempt to include the $T\mathrm{Tr}\,[M^3]$ in the calculation of T_c. Since T_c is calculated by considering the curvature of the potential at the origin, where the curvature matrices are easily diagonalised analytically, it was hoped that the correction to the curvature eigenvalues due to inclusion of $T\mathrm{Tr}\,[M^3]$ might be determined using a perturbative approach to solving the characteristic equation of the curvature matrix. However, as will be described below, the attempt was eventually abandoned due to uncertainties about the validity of the small amounts arguments.

The effort did produce some interesting results on the structure of the curvature matrix as evaluated at the origin. Hence, it will be described here, for the information of anyone interested in possible corrections to T_c, and to act as a record of work undertaken during this project.

A.1 Formalism of the perturbative correction

In chapter 5, T_c was defined to be the largest temperature at which one of the eigenvalues of the matrix

$$C_{ij}^2 = \frac{\partial^2 V_2}{\partial\phi_i \partial\phi_j}$$

$$= \frac{\partial^2}{\partial\phi_i\partial\phi_j}\left\{V_0 + \frac{T^2}{24}\mathrm{Tr}\left[M^2(\phi,T)\right]\right\} \qquad (1.1.1)$$

goes to zero **when evaluated at the origin**. This procedure ignores the term proportional to $\mathrm{Tr}\,[M^3]$, as discussed in chapter 5, and this has been the typical approach to evaluation of T_c in the literature to date because of possible problems with imaginary terms. However, it has been suggested lately, as mentioned in the conclusions chapter, that lattice calculations point towards a poor evaluation of T_c as one of the main drawbacks of the perturbative approach. So, we suggest including the next term in the series when making the evaluation of T_c. It is exactly this term which provides the possibility of a first order phase transition and, by neglecting it, it seems at least possible that we are neglecting an important factor in T_c.

To include the term $T/12\pi\mathrm{Tr}\,[M^3]$ in the evaluation of T_c, we must determine a correction to the matrix C^2 defined above. Denote this by C'^2 and define it thus

$$C'^{\,2}_{ij} = \frac{-T}{12\pi}\frac{\partial^2}{\partial\phi_i\partial\phi_j}\left\{\mathrm{Tr}\left[M^3(\phi,T)\right]\right\} \qquad (1.1.2)$$

Since it has already been made clear that $\mathrm{Tr}\,[M^3(\phi,T)]$ cannot be calculated explicitly, it is necessary to proceed in the following manner. Write $\mathrm{Tr}\,[M^3(\phi,T)] = \sum_{i=1}^{8}\eta_i^3(\phi,T)$ where $\eta_i^2(\phi,T)$ are the eigenvalues of $M^2(\phi,T)$. This may be re-written in the form $\mathrm{Tr}\,[M^3(\phi,T)] = \sum_{i=1}^{8}\left[\eta_i^2(\phi,T)\right]^{\frac{3}{2}}$. Proceed to differentiate this form

$$\begin{aligned}\frac{\partial}{\partial\phi_i}\left(\mathrm{Tr}\left[M^3(\phi,T)\right]\right) &= \frac{\partial}{\partial\phi_i}\left\{\sum_{k=1}^{8}\left[\eta_k^2(\phi,T)\right]^{\frac{3}{2}}\right\} \\ &= \sum_{k=1}^{8}\frac{3}{2}\left(\eta_k^2(\phi,T)\right)^{\frac{1}{2}}\frac{\partial}{\partial\phi_i}\left(\eta_k^2(\phi,T)\right) \qquad (1.1.3)\end{aligned}$$

and

$$\frac{\partial^2}{\partial\phi_i\partial\phi_j}\left(\mathrm{Tr}\left[M^3(\phi,T)\right]\right) = \qquad (1.1.4)$$

$$\begin{aligned}\sum_{k=1}^{8}\Bigg\{&\frac{3}{4}\left(\eta_k^2(\phi,T)\right)^{-\frac{1}{2}}\frac{\partial}{\partial\phi_i}\left(\eta_k^2(\phi,T)\right)\frac{\partial}{\partial\phi_j}\left(\eta_k^2(\phi,T)\right) \\ &+\frac{3}{2}\left(\eta_k^2(\phi,T)\right)^{\frac{1}{2}}\frac{\partial^2}{\partial\phi_i\partial\phi_j}\left(\eta_k^2(\phi,T)\right)\Bigg\}\end{aligned} \qquad (1.1.5)$$

At this point we make two observations.Firstly, note that all terms in the elements of $M^2(\phi,T)$ are of even power in the fields ϕ_i. Hence, all terms in the first derivatives of $M^2(\phi,T)$ and $\eta_i^2(\phi,T)$ must be odd powered in ϕ_i. Secondly, note that we are only interested in evaluating the correction matrix C'^2 *at the origin*, where all scalar fields are zero, and hence we can ignore the terms with factors of $\frac{\partial}{\partial\phi_j}(\eta_k^2(\phi,T))$.

So, the correction matrix will effectively reduce to

$$C'^{\,2}_{ij} = -\frac{T}{12\pi}\sum_{k=1}^{8}\left\{\frac{3}{2}\left(\eta_k^2(\phi,T)\right)^{\frac{1}{2}}\frac{\partial^2}{\partial\phi_i\partial\phi_j}\left(\eta_k^2(\phi,T)\right)\right\} \tag{1.1.6}$$

There is still a problem – there are no explicit expressions for the eigenvalues $\eta_k^2(\phi,T)$ since we cannot diagonalise the matrix $M^2(\phi,T)$ in general (at least, not such that the expressions obtained are realistically useable.) However, the ultimate aim of this line of reasoning is to include the $\mathrm{Tr}\,[M^3]$ term in the curvature matrix at the origin, in order to calculate T_c. Hence, we only require the expression $\frac{\partial^2}{\partial\phi_i\partial\phi_j}(\eta_k^2(\phi,T))$, and don't have to worry about determining the $\eta_k^2(\phi,T)$ explicitly.

We are able to diagonalise $M^2(\phi,T)$ **at the origin**, since this is just the definition of the matrix C^2. Let $\rho_k^2(T)$ be the eigenvalues of C^2. Then

$$\det\left[C^2 - \mathrm{diag}\left(\rho_k^2(T)\right)\right] = 0 \tag{1.1.7}$$

by definition. This expression contains no field-dependence (since it is evaluated at the origin) but, if we were to include the field-dependent parts of the curvature matrix ($F^2(\phi,T)$, say) and the eigenvalues ($\epsilon_k(\phi,T)$, say), then we would have

$$\det\left[C^2 + F^2(\phi,T) - \mathrm{diag}\left(\rho_k^2(T) + \epsilon_k(\phi,T)\right)\right] = 0 \tag{1.1.8}$$

Please take care with the notation here – we might strictly have written $(\rho_k^2(T) + \epsilon_k^2(\phi,T))$ above but we are going to work extensively with powers of $\epsilon_k(\phi,T)$ and have chosen the notation for clarity.

$F^2(\phi,T)$ is simply defined as

$$F_{ij}^2(\phi,T) = M_{ij}^2(\phi,T) - C_{ij}^2(T) \tag{1.1.9}$$

and we have expressions for $C^2_{ij}(T)$ and $\rho^2_k(T)$ for all i,j,k in $1,\ldots,8$. Hence, in principle, we can evaluate $\epsilon_k(\phi,T)$. However, we already know that this will be unworkable and so we suggest making an expansion of the determinant above, with appropriate small value approximations.

A.2 Expansion of determinant of the curvature matrix

Now,

$$\begin{aligned} & \det\left[C^2 + F^2(\phi,T) - \mathrm{diag}\left(\rho^2_k(T) + \epsilon_k(\phi,T)\right)\right] \\ = \; & \det\left[C^2 - \mathrm{diag}\left(\rho^2_k(T)\right)\right] + A_{ij}(T)\Psi_{ij} + B_k(T)\epsilon_k(\phi,T) + \ldots \\ = \; & A_{ij}(T)\Psi_{ij} + B_k(T)\epsilon_k(\phi,T) + \ldots \\ \equiv \; & 0 \end{aligned} \tag{1.2.10}$$

where we are using the notation of summation over the indices i,j, and Ψ_{ij} is simply the second order field-dependent part of F^2_{ij}.

The ultimate aim is to evaluate $\epsilon_k(\phi,T)$ at the origin, so we can consider the ϕ_i to be small. If we also assume that $\epsilon_k(\phi=0,T)$, the corrections to the eigenvalues, are small compared to $\rho^2_k(T)$, then we can make small value approximations and say

$$\begin{aligned} A_{ij}(T)\Psi_{ij} + B_k(T)\epsilon_k(\phi,T) &\approx 0 \\ \Rightarrow \epsilon_k(\phi,T) &= -\frac{A_{ij}(T)}{B_k(T)}\Psi_{ij} \end{aligned} \tag{1.2.11}$$

So, it only remains to evaluate $A_{ij}(T)\Psi_{ij}$ — the sum of terms, in the determinant above, which are second order in ϕ_i — and $B_k(T)$ — the co-efficient of $\epsilon_k(\phi,T)$ in the determinant. We will shortly see that the problem is not actually quite this simple but is still tractable, so let us continue with this line of argument.

Consider the matrix $C^2_0(T) = C^2(T) - \mathrm{diag}(\rho^2_k(T))$ in block diagonal form as explained previously in the discussion of T_c.

$$C^2_0(T) \equiv C^2(T) - \mathrm{diag}\left(\rho^2_k(T)\right) = \begin{pmatrix} S & O_4 \\ O_4 & S \end{pmatrix} \tag{1.2.12}$$

where

$$S = \begin{pmatrix} A-\rho_k^2(T) & -\gamma_1 & 0 & \gamma_2 \\ -\gamma_1 & B-\rho_k^2(T) & -\gamma_2 & 0 \\ 0 & -\gamma_2 & A-\rho_k^2(T) & -\gamma_1 \\ \gamma_2 & 0 & -\gamma_1 & B-\rho_k^2(T) \end{pmatrix} \tag{1.2.13}$$

We will obtain the second order field-dependent terms by simply evaluating the determinants of the matrices

$$\left[X_{ij}^2\right]_{kl} = \left[C_0^2(T)\right]_{kl} + \Psi_{ij}\delta_{ik}\delta_{jl} \tag{1.2.14}$$

(Please note that X_{ij}^2 is a matrix, not an element of a matrix. The indices i, j label the matrix, not its elements.)

In principle, this will lead to 64 terms. However, a short discussion of the properties of $C_0^2(T)$ will drastically reduce this number. Firstly, let us note that the only terms in $\det[X_{ij}^2]$ which are proportional to Ψ_{kl} are just those from the k,l–th signed co-factor of X_{ij}^2, which is exactly equivalent to the k,l–th signed co-factor of C_0^2. So, we can proceed by determining the signed co-factors of C_0^2. Secondly, recall that the determinant of a matrix can be defined as the sum of all possible terms (with appropriate signs) which can be formed as the product of one, and only one, element from each row and each column of the matrix. This definition will prove useful in heavily reducing the workload.

Begin by considering the k,l–th co-factors, with the k,l–th element lying in one of the off-diagonal zero-blocks. Having removed the k–th row and l–th column, we are left with a matrix which no longer has block diagonal form, and it is impossible to choose one, and only one, element from each row and each column without choosing at least one zero. Hence, every term in these co-factors will be zero and so each of these co-factors are also zero. This accounts for 32 of the possible 64 terms mentioned above!

When we consider the k,l–th co-factor, with the k,l–th element lying in one of the main diagonal S–blocks, then we are left with a block diagonal form after removing the k–th row and l–th column. Note, however, that we can make use of the symmetry in $C_0^2(T)$ here. For example, the $(1,1)$–co-factor will be equivalent

to the $(5,5)$–co- factor, and so on. Also, the $(1,1)$–co-factor of C_0^2 will simply be the $(1,1)$–co-factor of S multiplied by the determinant of S, and so on. So, we should evaluate the determinant of S and all of its signed co-factors.

The determinant of S is

$$\det[S] = \left(AB - (A+B)\rho_k^2(T) + \rho_k^4(T) + \gamma_1^2 + \gamma_2^2\right)^2 \tag{1.2.15}$$

and the matrix of co-factors is

$$S_c = \begin{pmatrix} B - \rho_k^2(T) & \gamma_1 & 0 & -\gamma_2 \\ \gamma_1 & A - \rho_k^2(T) & \gamma_2 & 0 \\ 0 & \gamma_2 & B - \rho_k^2(T) & \gamma_1 \\ -\gamma_2 & 0 & \gamma_1 & A - \rho_k^2(T) \end{pmatrix} \times \left(AB - (A+B)\rho_k^2(T) + \rho_k^4(T) + \gamma_1^2 + \gamma_2^2\right) \tag{1.2.16}$$

where $[S_c]_{ij}$ is the i,j–th co-factor of S.

Now, the determinant of C_0^2 will clearly be the square of the determinant of S. But, **by definition**, the determinant of C_0^2 is zero. Hence, we conclude that the determinant of S is also zero and then

$$\left(AB - (A+B)\rho_k^2(T) + \rho_k^4(T) + \gamma_1^2 + \gamma_2^2\right) = 0 \tag{1.2.17}$$

This is the factor which appears in all of the entries of S_c and so we see that ALL of the co-factors of $C_0^2(T)$ *and* S are zero. In the notation of eq.(1.2.10), this indicates that $A_{ij}(T) = 0$ for all i,j. What about $B_k(T)$? Well, the co-efficient of $\epsilon_k(\phi,T)$ is just a special case of the work we have done. The terms proportional to $\epsilon_k(\phi,T)$ in the expansion in eq.(1.2.10) are just those from the determinants of the matrices

$$\left[Y_k^2\right]_{ij} = \left[C_0^2(T)\right]_{ij} - \epsilon_k(\phi,T)\delta_{ki}\delta_{kj} \tag{1.2.18}$$

and so Y_k^2 will behave like X_{kk}^2, and the co-efficient of $\epsilon_k(\phi,T)$ will be the sum of the diagonal co-factors of $C_0^2(T)$. Hence, $B_k(T) = 0$ also!

A.3 The expansion to second order

Where does this leave the approach to approximating $\epsilon_k(\phi,T)$? These results on $A_{ij}(T)$ and $B_k(T)$ mean that there are no first order corrections to the determi-

nant of the curvature matrix evaluated at the origin but it does not imply that $\epsilon_k(\phi, T)$ is zero — simply that we must go to higher orders in our expansion of the determinant in order to calculate $\epsilon_k(\phi, T)$. We will now go on to show that we must, in fact, expand the determinant to fourth order in small values before we find non-zero co-efficients. In order to clarify the arguments that we are about to go through, we will develop a notation for working with sub-matrices of $C_0^2(T)$ which will allow us to discuss co-factors and terms within co-factors, and so on. We will develop this notation by considering the co-efficients of the terms in the determinant which are fourth order in the fields or, equivalently, second order in Ψ_{ij}.

What will the co-efficient of $\Psi_{ij}\Psi_{kl}$ be? A little thought reveals that it is the (appropriately signed) determinant of the submatrix $[\overline{C_0^2(T)}]_{\{ik/jl\}}$ formed by removing the i–th and k–th rows, and j–th and l–th columns of $C_0^2(T)$. We will simplify the notation for the submatrix to $[\overline{C^2}]_{\{ik/jl\}}$ and will later extend it to consider the co-efficients of terms like $\Psi_{ij}\Psi_{kl}\Psi_{mn}$ by writing $[\overline{C^2}]_{\{ikm/jln\}}$.

So, we want to know the determinants of all possible matrices $[\overline{C^2}]_{\{ik/jl\}}$. Note, first of all, that we cannot have $i = k$ or $j = l$ because the determinant never contains terms with two factors from the same row or column. Having noted this, we can go on to show that ALL of these determinants are zero! Again, we can resort to simple arguments, as opposed to direct calculation, to prove this. There are only 5 situations to consider —

1. Both (i, j) and (k, l) are in the same S–block. Then $[\overline{C^2}]_{\{ik/jl\}}$ has a block diagonal form, with a complete S–block as one of the main diagonal blocks. Hence the determinant will contain a factor of $\det[S]$ which is identically zero. So, all determinants of this form are zero.

2. (i, j) and (k, l) are in different S–blocks. Then $[\overline{C^2}]_{\{ik/jl\}}$ has a block diagonal form again, but now the two blocks on the main diagonal are both submatrices of S and the determinant will be a product of two of the co-factors of S. Since these co-factors are all zero, the determinant will also be zero. So, all determinants of this form are zero.

3. Both (i, j) and (k, l) are in the same zero-block. Then $[\overline{C^2}]_{\{ik/jl\}}$ does not

have a block diagonal form and it is impossible to choose one, and only one, element from each row and column of the $[\overline{C^2}]_{\{ik/jl\}}$ without choosing one to be zero. Hence the determinant will always be zero.

4. (i,j) and (k,l) are in different zero-blocks. Then $[\overline{C^2}]_{\{ik/jl\}}$ has a block diagonal structure exactly like that in case(2), and so the determinant is always zero.

5. (i,j) and (k,l) can lie one in each type of block; ie. (i,j) in one of the S-blocks and (k,l) in one of the zero-blocks, or vice-versa. Then $[\overline{C^2}]_{\{ik/jl\}}$ will always have a structure which is not block diagonal and we have the same situation as in case(3). So, the determinant is always zero.

This covers all possibilities and so we see that all co-efficients of $\Psi_{ij}\Psi_{kl}$ are zero. Again, the co-efficient of $\epsilon_k^2(\phi,T)$ is a special case of the results we have just given and so there is also no co-efficient of $\epsilon_k^2(\phi,T)$. However, the case of the co-efficient of $\epsilon_k(\phi,T)\Psi_{ij}$ deserves comment. Recall that $\epsilon_k(\phi,T)$ only ever appears in the diagonal terms of the curvature matrix. If Ψ_{ij} appears in the same diagonal term (ie. $i=j=k$) then clearly we can never form a term proportional to $\epsilon_k(\phi,T)\Psi_{kk}$ in the determinant and so we have no co-efficient for such terms. Every other possible placement of $\epsilon_k(\phi,T)$ and Ψ_{ij} is exactly equivalent to one of the cases described in cases(1–5) above and again we will have no co-efficient for these terms.

We conclude that there are no non-zero terms in the expansion of eq(1.2.10) which are of second order in small values. We proceed to the third order terms.

A.4 The expansion to third order

We now consider the co-efficients of terms like $\Psi_{ij}\Psi_{kl}\Psi_{mn}$. Extending the ideas we have already seen, these co-efficients will simply be the determinants of the matrices $[\overline{C^2}]_{\{ikm/jln\}}$. We claim now that all of these determinants are zero and will prove this in a manner similar to the proof given in the last section. Notice now that we cannot allow any pair of i,k,m to be equal, nor any pair of j,l,n.

1. We might have all of (i,j), (k,l) and (m,n) in the same S-block. Then

$[\overline{C^2}]_{\{ikm/jln\}}$ has a block diagonal structure with S as one of the diagonal blocks. This introduces a factor $\det[S]$ into the determinant, which is identically zero, and hence all determinants of this form are zero.

2. We might have two of (i,j), (k,l) and (m,n) in the same S-block and the third in the other S-block. Then $[\overline{C^2}]_{\{ikm/jln\}}$ has a block diagonal structure with a submatrix of S as one of the diagonal blocks. This introduces one of the co-factors of S as a factor in the determinant and these co-factors are all zero. Hence, all determinants of this form are zero.

3. We might have all of (i,j), (k,l) and (m,n) in the same zero-block. Then $[\overline{C^2}]_{\{ikm/jln\}}$ does not have a block diagonal form and it is impossible to choose one, and only one, element from each row and each column of $[\overline{C^2}]_{\{ikm/jln\}}$ without choosing one of them to be zero. Hence, all determinants of this form will be zero.

4. We might have two of (i,j), (k,l) and (m,n) in the same zero-block and the third in the other zero-block. Then $[\overline{C^2}]_{\{ikm/jln\}}$ does not have a block diagonal form and the determinant will be zero by the same line of argument used in case(3). Hence, all determinants of this form will be zero.

5. We might have two of (i,j), (k,l) and (m,n) in the same S-block and the third in a zero-block. Then $[\overline{C^2}]_{\{ikm/jln\}}$ does not have a block diagonal form and the determinant will be zero, as in cases(3,4). Hence, all determinants of this form will be zero.

6. We might have two of (i,j), (k,l) and (m,n) in the same zero-block and the third in an S-block. Then $[\overline{C^2}]_{\{ikm/jln\}}$ does not have a block diagonal form and the determinant will be zero, as in cases(3–5). Hence, all determinants of this form will be zero.

7. We might have two of (i,j), (k,l) and (m,n) in different S-blocks, (ie. one in each S-block), and the third in a zero-block. Then $[\overline{C^2}]_{\{ikm/ljn\}}$ does not have a block diagonal form and the determinant will be zero, as in cases(3–6). Hence, all determinants of this form will be zero.

8. We might have two of (i,j), (k,l) and (m,n) in different zero-blocks, (ie. one in each zero-block), and the third in an S-block. Then $[\overline{C^2}]_{\{ikm/ljn\}}$ does not have a block diagonal form and the determinant will be zero, as in cases(3–7).

Hence, all determinants of this form will be zero.

Consideration of the co-efficients of $\epsilon_k^3(\phi,T)$, $\epsilon_k^2(\phi,T)\Psi_{ij}$ and $\epsilon_k(\phi,T)\Psi_{ij}\Psi_{kl}$ follow on directly from cases(1–8) and the discussion at the end of the last section. Again, all of these co-efficients are identically zero.

So, we have proved, as pointed out in section A.3, that we must go to fourth order in small values, in our expansion of the determinant of the curvature matrix at the origin, before we are able to evaluate $\epsilon_k(\phi,T)$.

A.5 The expansion to fourth order

We come now to the fourth order terms like $\Psi_{ij}\Psi_{kl}\Psi_{mn}\Psi_{pq}$. The co-efficients of these terms are not all zero, although most of them are. However, there are so many non-zero terms that this would be effectively impossible to handle if we did not make the following observation. We are ultimately interested in evaluating the terms $\frac{\partial^2}{\partial\phi_i\partial\phi_j}(\epsilon_k(\phi,T))$ *at the origin.* So, we are only interested in terms in $\epsilon_k(\phi,T)$ which are of second order in the scalar fields. In the fourth order expansion of the determinant of the curvature matrix at the origin we have,formally

$$\epsilon_k^4+\epsilon_k^3\Psi_{ij}+\epsilon_k^2\Psi_{ij}^2+\epsilon_k\Psi_{ij}^3+\Psi_{ij}^4=0 \tag{1.5.19}$$

Since we will evaluate the second derivatives of $\epsilon_k(\phi,T)$ at the origin, we can be sure that the Ψ_{ij} are as small as we like and hence it is justifiable to say that the terms of 2nd or higher order in Ψ are negligibly small and so

$$\epsilon_k^4+\epsilon_k^3\Psi_{ij}=0 \tag{1.5.20}$$

This leads to

$$\epsilon_k=-\Psi_{ij} \tag{1.5.21}$$

but remember that this is a formal representation of the true expression, and that there coefficients present on both sides.

The non-vanishing co-efficients are generated by the following arrangements of $(i,j),(k,l),(m,n),(p,q)$ –

1. Each of $(i,j),\ldots,(p,q)$ in a different block of the curvature matrix.

2. Any two of $(i,j),\ldots,(p,q)$ in each of the S blocks.

Every other arrangement of rows and columns results in a zero co-efficient in the same fashion as the examples of the previous sections. Even with only these few cases, the resulting number of co-efficients is enormous (numbered in the thousands) and, even after the observation that we are only interested in the co-efficients of ϵ^4 and $\epsilon^3\Psi_{ij}$, the number of distinct co-efficients still numbers in the hundreds.

This enormous number of co-efficients leaves the perturbative expansion on dangerous ground. The aim of this exercise was to determine an analytical expression for the correction of the critical temperature, and although hundreds of co-efficients can be manipulated by computer algebra packages (MAPLE was used here) the resultant expressions will be unfathomable to the reader and their accuracy difficult to check.

For this reason, it seems pointless to continue with this line of argument and the correction to T_c was never implemented.

Appendix B

The zero-T mass matrices

B.1 The charged sector mass matrix

The full form of the zero temperature charged sector scalar mass matrix is –

$$
\begin{aligned}
M^2_{1256} &= \begin{pmatrix} a_{1256} & 0 & c_{1256} & d_{1256} \\ 0 & a_{1256} & -d_{1256} & c_{1256} \\ c_{1256} & -d_{1256} & b_{1256} & 0 \\ d_{1256} & c_{1256} & 0 & b_{1256} \end{pmatrix} \\
a_{1256} &= -\mu_1^2 + \lambda_1 v_1^2 + \lambda_3/2 v_2^2 \\
b_{1256} &= -\mu_2^2 + \lambda_2 v_2^2 + \lambda_3/2 v_1^2 \\
c_{1256} &= \lambda_4/2 v_1 v_2 + \lambda_5/2 \cos(2\theta_0) v_1 v_2 - \mu_R^2 \cos(\theta_0) + \mu_I^2 \sin(\theta_0) \\
d_{1256} &= \mu_R^2 \sin(\theta_0) + \mu_I^2 \cos(\theta_0) - \lambda_5 \sin(2\theta_0) v_1 v_2
\end{aligned} \tag{2.1.1}
$$

This sector diagonalises readily. First, form the characteristic equation

$$
\det\left[M^2_{1256} - m^2 I_4\right] = 0 \tag{2.1.2}
$$

to give

$$
\left(m^4 - (a+b)m^2 + ab - (c^2+d^2)\right)^2 = 0 \tag{2.1.3}
$$

where the indices 1256 have been dropped for clarity. This returns the solutions for the squared masses, m^2, as

$$
m^2 = \frac{a+b}{2} \pm \frac{\sqrt{(a-b)^2 + 4(c^2+d^2)}}{2} \tag{2.1.4}
$$

Having substituted from the full form of the matrix we arrive at the masses

$$\begin{aligned} M^2_{\chi^\pm} &= 0 \\ M^2_{H^\pm} &= -\frac{v^2}{2}\left[\lambda_4 + \lambda_5\cos(2\theta_0) - \frac{2\left(\mu_R^2\cos(\theta_0) - \mu_I^2\sin(\theta_0)\right)}{v_1 v_2}\right] \end{aligned} \tag{2.1.5}$$

where the $\chi^\pm$ are the would-be charged Goldstone bosons, "swallowed" by the $W^\pm$ bosons, and the $H^\pm$ are the charged Higgs bosons.

B.2 The neutral sector mass matrix

The full form of the zero temperature neutral sector mass matrix is –

$$M^2_{3478} = \begin{pmatrix} a_{3478} & e_{3478} & f_{3478} & g_{3478} \\ e_{3478} & b_{3478} & -g_{3478} & h_{3478} \\ f_{3478} & -g_{3478} & c_{3478} & j_{3478} \\ g_{3478} & h_{3478} & j_{3478} & d_{3478} \end{pmatrix}$$

$$\begin{aligned} a_{3478} &= -\mu_1^2 + 3\lambda_1 v_1^2 + \frac{1}{2}\left(\lambda_3 + \lambda_4 + \lambda_5\cos(2\theta_0)\right)v_2^2 \\ b_{3478} &= -\mu_1^2 + \lambda_1 v_1^2 + \frac{1}{2}\left(\lambda_3 + \lambda_4 - \lambda_5\cos(2\theta_0)\right)v_2^2 \\ c_{3478} &= -\mu_2^2 + 3\lambda_2 v_2^2 + \frac{1}{2}\left(\lambda_3 + \lambda_4 + \lambda_5\cos(2\theta_0)\right)v_1^2 \\ d_{3478} &= -\mu_2^2 + \lambda_2 v_2^2 + \frac{1}{2}\left(\lambda_3 + \lambda_4 - \lambda_5\cos(2\theta_0)\right)v_1^2 \\ e_{3478} &= \lambda_5/2\sin(2\theta_0)v_2^2 \\ f_{3478} &= \left(\lambda_3 + \lambda_4 + \lambda_5\cos(2\theta_0)\right)v_1 v_2 - \mu_R^2\cos(\theta_0) + \mu_I^2\sin(\theta_0) \\ g_{3478} &= -\lambda_5\sin(2\theta_0)v_1 v_2 + \mu_R^2\sin(\theta_0) + \mu_I^2\cos(\theta_0) \\ h_{3478} &= \lambda_5\cos(2\theta_0)v_1 v_2 - \mu_R^2\cos(\theta_0) + \mu_I^2\sin(\theta_0) \\ j_{3478} &= -\lambda_5/2\sin(2\theta_0)v_1^2 \end{aligned} \tag{2.2.6}$$

This sector does not diagonalise explicitly in the general case where μ_3^2 can be complex. However, it will diagonalise readily in the CP-conserving case where μ_3^2 is pure real or pure imaginary. In this case, the matrix simplifies to the form

$$M^2_{3478} = \begin{pmatrix} a & 0 & e & 0 \\ 0 & b & 0 & f \\ e & 0 & c & 0 \\ 0 & f & 0 & d \end{pmatrix} \tag{2.2.7}$$

The characteristic equation is

$$\det\left[M^2_{3478} - m^2 I_4\right] = 0 \tag{2.2.8}$$

which factorises to the form

$$\left(m^4 I_4 - (am^2 + cm^2)I_4 + (ac - e^2)\right)\left(m^4 I_4 - (bm^2 + dm^2)I_4 + (bd - f^2)\right) = 0 \tag{2.2.9}$$

Using the relationships which define the position of the minimum (see chapter 6, eqns (6.1.1,...,6.1.3)) -

$$-\mu_1^2 + \lambda_1 v_1^2 + \frac{\lambda_3 + \lambda_4 + \lambda_5 \cos(2\theta_0)}{2} v_2^2 - \gamma_1 \frac{v_2}{v_1} = 0 \tag{2.2.10}$$

$$-\mu_2^2 + \lambda_2 v_2^2 + \frac{\lambda_3 + \lambda_4 + \lambda_5 \cos(2\theta_0)}{2} v_1^2 - \gamma_1 \frac{v_1}{v_2} = 0 \tag{2.2.11}$$

$$\mu_R^2 \sin(\theta_0) + \mu_I^2 \cos(\theta_0) - \lambda_5 \sin(\theta_0)\cos(\theta_0) v_1 v_2 = 0$$

the forms of the elements in M^2_{3478} can be simplifed before substituting into the characteristic equation. The resulting eigenvalues are readily found to be

$$\begin{aligned} M^2_{\chi^0} &= 0 \\ M^2_A &= \left(\mu_R^2 - \lambda_5 v_1 v_2\right)\frac{v^2}{v_1 v_2} \\ M^2_H, M^2_h &= \lambda_1 v_1^2 + \lambda_2 v_2^2 + \mu_R^2 \frac{v^2}{2 v_1 v_2} \\ &\quad \pm \sqrt{\left(\lambda_1 v_1^2 - \lambda_2 v_2^2 + \mu_R^2 \frac{(v_2^2 - v_1^2)}{2 v_1 v_2}\right)^2 + (\Sigma_+ v_1 v_2 - \mu_R^2)^2} \end{aligned} \tag{2.2.12}$$

where $M^2_{\chi^0} = 0$, the would-be Goldstone boson is "swallowed" by the Z^0 boson.

References

[1] A.D. Sakharov, JETP Lett, **6** (1967) 24

[2] G. 't Hooft, Phys Rev Lett, **37** (1976) 8

G. 't Hooft, Phys Rev, **D14** (1976) 3432

[3] G. Farrar and M. E. Shaposhnikov, Phys Rev Lett **70** (1990) 283

[4] N. Turok and J. Zadrozny, Phys Rev Lett, **65** (1990) 2331

N. Turok and J. Zadrozny, Nucl Phys, **B358** (1991) 471

N. Turok and J. Zadrozny, Nucl Phys, **B369** (1992) 729

[5] A.D. Dolgov, Physics Reports, **222** (1992) 309

[6] N.S.Manton, Phys Rev, **D28** (1983) 2019

F.R.Klinkhammer and N.S.Manton, Phys Rev, **D30** (1984) 2212

B. Kastening, R.D. Peccei and X.Zhang, Phys Lett **B266** (1991) 413

[7] P. B. Arnold, Talk presented at the 1990 Theoretical Advanced Study Institute in Elementary Particle Physics

[8] P. Arnold and L. McLerran, Phys Rev **D36** (1987) 581

[9] P. Arnold, Phys Rev, **D46** (1992) 2682

[10] L. McLerran, pre-print TPI-MINN-92/65-T (1992)

[11] M.E. Shaposhnikov, JETP Lett **44** (1986) 465

[12] M.E.Shaposhnikov, Pis'ma Zh. Eksp. Teor. Fiz **44**, No.8 (1986), 364

[13] V.A. Kuzmin, V.A. Rubakov and M.E.Shapsohnikov, Phys Lett **B155** (1985) 36

[14] A.I.Bochkarev and M.E. Shaposhnikov, Mod Phys Lett **A2** (1987) 417

[15] A.I.Bochkarev, S.V. Kuzmin and M.E.Shaposhnikov, Phys Rev **D43** (1991) 369

[16] L.McLerran, M.Shaposhnikov, N.Turok and M.Voloshin, Phys Lett **B256** (1991) 451

[17] D.E. Brahm, pre-print CALT-68-1810 (1992) (for Proceedings, *XXVI* International Conference on High Energy Physics, August 1996)

[18] ALEPH, DELPHI, L3 and OPAL Collaborations, as presented by M. Davier, Proceedings of the International Lepton-Photon Symposium and Europhysics Conference on High Energy Physics, eds. S. Hegerty, K. Potter and E. Quercigh (Geneva 1991)

[19] Robert L. Singleton Jr., SLAC Report **380** (1991)

[20] M. Sher, Phys Rep **179** (1989) 273

[21] Kevin E. C. Benson, Phys Rev **D48** (1993) 2456

[22] Peter Arnold, Phys Rev **D46** (1992) 2628

[23] C. D. Froggatt, I.G. Knowles and R.G.Moorhouse, Phys Lett, **B249** (1990) 273

C. D. Froggatt, I.G. Knowles and R.G.Moorhouse, Nucl Phys, **B386** (1992) 63

[24] Boris Kastening, pre-print UCLA/92/TEP/22

[25] Laurence G. Yaffe, hep-ph/9512265 (1995)

[26] Neil Turok, Electroweak Physics and the Early Universe, Edited by Romao and Freire, Plenum Press, New York 1994

[27] M. Dine, O. Lechtenfeld, B. Sakita, W. Fischler and J. Polchinski, Nucl Phys **B342** (1990) 381

[28] G.W.Anderson and L.J.Hall, Phys Rev, **D45** (1992) 2685

[29] S. Coleman and E. Weinberg, Phys Rev **D7** (1973) 1888

[30] L. Dolan and R. Jackiw, Phys Rev **D9** (1974) 3320

[31] D.A. Kirzhnitz and A.Linde, Ann Phys (NY) **101** (1976) 195

[32] A.T.Davies, C.D. Froggatt, G. Jenkins and R.G. Moorhouse, Phys Lett **B336** (1994) 464

[33] J. Gunion, H.E. Haber, G.L.Kane and S. Dawson, The Higgs' Hunters Guide (Addison-Wesley, Reading MA, 1990)

J. Gunion and H.E.Haber, Nucl Phys **B272** (1986) 1

M.Dine, P. Huet and R. Singleton, Nucl Phys **B375** (1992) 625

S. Myint, Phys Lett **B287** (1992) 325

G.F.Giudice, Phys Rev **D45** (1992) 3177

[34] J.R. Espinosa, M. Quiros and F.Zwirner, Phys Lett **B314** (1993) 206; **B307** (1993) 106

A.Brignole, J.R.Espinosa, M.Quiros and F.Zwirner, CERN pre-print: CERN-TH 7057/93

[35] M. Dine, P. Huet, R. Leigh and A.Linde, Phys Rev **D46** (1992) 550

[36] A.D. Linde, Rep. Prog. Phys **42** (1979) 389

A.D. Linde, Phys Lett **93B** (1980) 327

D.J. Gross, R.D.Pikarski and L.G.Yaffe, Rev Mod Phys **53** (1981) 1

[37] J.M. Cline and P.A.Lemieux, hep-ph/9609240 (1996)

[38] M.E.Carrington, pre-print TPI-MINN-91/48-T (1991)

[39] K. Takahashi, Z. Phys **C26** (1985) 601

[40] H. Ryder, An Introduction to Field Theory, (1983)

[41] C. Itzykson and J.L. Zuber, Quantum Field Theory (McGraw-Hill) 512

[42] A. G. Cohen, D.B. Kaplan and A.E. Nelson, Phys Lett **B263** (1991) 86

[43] V.A.Rubakov and M.E.Shaposhnikov, pre-print CERN-TH/96-13 (1996) (alternatively known as INR-913/96 (1996), hep-ph/9603208 (1996))

[44] D. Decamp et al., Phys Lett **B246** (1990) 308; Phys Lett **B265** (1991) 475

O. Adriani et al, Phys Lett **B294** (1992) 457; P. Abreu et al., Nucl Phys **B272** (1992) 3

D. Buskulic et al., Phys Lett **B313** (1993) 299,312

G. Altarelli, CERN pre-print: CERN-TH7045/93

[45] Christenson, Phys Rev Lett **13** (1964) 138

[46] Alan H. Guth, The Inflationary Universe, Jonathan Cape, London (1997) 102

[47] M. Shaposhnikov, CERN pre-print CERN-TH.6918/93

K. Kajantie, K. Rummukainen and M. Shaposhnikov, CERN pre-print CERN-TH.6901/93

[48] M. Carena, M. Quiros and C.E.M. Wagner, Phys Lett **B380** (1996) 81

R. Delepine, J. M. Gerard, R. Gonzales Felipe and J. Weyers, Phys Lett **B386** (1996) 183

[49] J.M. Cline and K. Kainulainen, Nucl Phys **B482** (1996) 73

M. Laine, Nucl Phys **B481** (1996) 43

[50] K. Kayantie, M. Cline, K. Rummukainen and M. E. Shaposhnikov, Nucl Phys **B458** (1996) 90;

M. E. Shaposhnikov, Nucl Phys **B466** (1996) 189