Student Seminar Theoretical Physics

Magnetic monopoles

Arjen Baarsma
(0433764)

February 28, 2009

Universiteit Utrecht

Contents

Introduction

The aim of this text is to on the one hand give an explicit definition of monopole solutions in gauge theory and the conditions for their existence and on the other hand describe the magnetic monopoles that are predicted by Grand Unified Theories from the viewpoint of cosmology. We will first introduce some notions from algebraic topology and briefly mention a few results for Lie groups in chapter 1 before we define the notion of a monopole solution in field theory in chapter 2 and discuss the conditions for their existence. Chapter 3 will mainly be about the *magnetic* monopoles and the reason why they are predicted by GUT theories. Finally, chapter 4 will deal with observational bounds that can be put on the monopole abundance in the Universe through cosmological and astrophysical arguments and earth-bound experiments.

The reader is assumed to have some basic understanding of a number of concepts from quantum field theory and cosmology, as well as some basic knowledge of Lie groups and Lie algebras in the context of gauge theory. Furthermore, familiarity with spontaneous symmetry breaking and phase transitions in the early universe may be useful, which is the topic of the paper by Doru Sticlet. It may finally be useful to point out that throughout this text a number of constants have been set to 1, namely $c = \hbar = k_{\mathrm{B}} = 1$, and that furthermore $e^2 = 4\,\pi\,\alpha$, where e is the elemtary charge and $\alpha \simeq \frac{1}{137}$ is the fine structure constant.

1. Some preliminary mathematics

To understand the topological nature of the monopole solutions described in chapter 2 it is important to first review some notions from algebraic topology and *homotopy theory* in particular. The so-called homotopy groups objects will turn out to play a key role in the classification of monopole solutions as topological soliton, so we will define them here and give some important properties that will become important later. For a more thorough review of these objects and the proofs to most of the statements made here, as well as some other notions from algebraic topology, the reader is referred to [1].

1.1 Homotopy

All the definitions in this section have been formulated for general topological spaces, so they apply in particular to manifolds (in which case path-connected can be replaced by connected), which are the types of spaces we are interested in. In this section functions are always considered to be continuous, even when this is not explicitly mentioned. The entire theory of homotopy can also be defined in the context of smooth[1] manifolds, in which case all functions and homotopies are assumed to be smooth (infinitely many times differentiable) as it can be proven that the existence of an ordinary homotopy between two smooth functions is equivalent to the existence of a smooth homotopy. The proof is hidden deeply somewhere in [2] and is highly non-trivial.

Definition 1. *A homotopy of maps from a space X to another space Y is a continuous function $H\colon [0,1] \times X \to Y, (t,x) \mapsto H_t(x)$. We call two functions $f, g\colon X \to Y$ homotopic if and only if there exists a homotopy H such that $f = H_0$ and $g = H_1$ and we denote this by $f \sim g$.*

In words we could say that two functions are homotopic when they can be continuously deformed into each other and that the homotopy is just a description of this deformation. We can consider functions between two spaces up to homotopy (modulo continuous deformations) by splitting the space of functions from X to Y into *homotopy classes*. If we denote the class of all functions that are homotopic to some map $f\colon X \to Y$ by $[f] = \{g\colon X \to Y \mid f \sim g\}$ then it is a relatively easy exercise to

[1] By smooth we mean C^∞, i.e. infinitely many times differentiable.

show that each continuous map $g\colon X \to Y$ is contained in exactly one such class (so $[f] = [g] \Leftrightarrow f \sim g$).

For $n \geq 1$ let $I^n = [0,1]^n = \{(s_1, \ldots, s_n) \in \mathbb{R}^n \mid 0 \leq s_1, \ldots, s_n \leq 1\}$ be the closed n-dimensional unit cube and let ∂I^n be its boundary (the set of all points in I^n with at least one coordinate equal to either 0 or 1). For $n = 0$ we choose I^0 to be a point and ∂I^0 to be the empty set. Let $x_0 \in X$ furthermore be some point in a space X.

If you only consider maps that satisfy certain conditions, it is possible to demand that the homotopies satisfy the same condition, i.e. that the map H_t satisfies this condition for any t. In our case we would like to consider maps from I^n to X that map the boundary ∂I^n into the point x_0. Instead of explicitly saying this every time we will write $f\colon (I^n, \partial I^n) \to (X, x_0)$ when we have imposed this condition on the function $f\colon I^n \to X$.

Definition 2. *A homotopy of maps $(I^n, \partial I^n) \to (X, x_0)$ is a function $H\colon [0,1] \times I^n \to X, (t, y) \mapsto H_t(y)$ such that $H_t(\partial I^n) \subseteq \{x_0\}$ for all $t \in [0,1]$. We call two functions $f, g\colon (I^n, \partial I^n) \to (X, x_0)$ homotopic if and only if there exists such a homotopy for which $H_0 = f$ and $H_1 = g$ and we denote this, as before, by $f \sim g$.*

Thus two such functions are homotopic whenever they can be continuously deformed into each other without breaking the condition that ∂I^n should be mapped into the point x_0. Even though the definition has been altered somewhat, not a lot has changed. For any map $f\colon (I^n, \partial I^n) \to (X, x_0)$ we can define the homotopy class $[f] = \{g\colon (I^n, \partial I^n) \to (X, x_0) \mid f \sim g\}$ and each map $g\colon (I^n, \partial I^n) \to (X, x_0)$ is once again contained in exactly one such class (so $[f] = [g] \Leftrightarrow f \sim g$).

Definition 3. *For any non-negative integer $n \in \mathbb{Z}_{\geq 0}$ the n-th* homotopy group *of a space X with basepoint x_0, which is denoted by $\pi_n(X, x_0)$ is the set of homotopy classes of functions $(I^n, \partial I^n) \to (X, x_0)$. In other words*

$$\pi_n(X) = \{[f] \mid f\colon (I^n, \partial I^n) \to (X, x_0)\}. \tag{1.1}$$

For $n \geq 1$ the group structure on $\pi_n(X, x_0)$ is defined by the operation $*$, which in turn is defined as follows: For any two elements $[f], [g] \in \pi_n(X, x_0)$, represented by the functions $f, g\colon (I^n, \partial I^n) \to (X, x_0)$, we define $[f] * [g] := [f * g]$ with

$$(f * g)(s_1, \ldots, s_n) = \begin{cases} f(2\,s_1, s_2, \ldots, s_n) & \text{if } s_1 \leq \frac{1}{2} \\ g(2\,s_1 - 1, s_2, \ldots, s_n) & \text{if } s_1 \geq \frac{1}{2}. \end{cases} \tag{1.2}$$

Showing that $[f] * [g]$ is well-defined and again an element of $\pi_2(X, x_0)$ is a rather trivial exercise. Showing that $*$ defines a group structure is also not very difficult and will not play a very important role in the remainder of this text, so we will not prove this here. We will call a homotopy group trivial if it consists of just one element and we will then write $\pi_n(X, x_0) = \{1\}$ (since as a group it only contains the unit element).

For $n = 0$, ∂I^0 was the empty set, so functions $f\colon (I^0, \partial I^0) \to (X, x_0)$ can just be viewed as points in X and a homotopy just describes a path between points. The elements of the zeroth homotopy group $\pi_0(X, x_0)$ are therefore just the path-connected components of X. It is important to note that the zeroth homotopy group $\pi_0(X, x_0)$ is not actually a group because the group operation described above is ill-defined for it.

The first homotopy group, usually called the *fundamental group*, is probably the easiest to grasp. Since a map from the interval $[0,1]$ to the space X_0 mapping both 0 and 1

to the point x_0 is just a closed loop, the fundamental group in this case just consists of all (oriented) closed loops in X starting and ending in the point x_0 up to continuous deformations. It is interesting to note that $\pi_1(X, x_0)$ is not necessarily abelian, while every other group $\pi_n(X, x_0)$ (for $n \geq 2$) is (for a proof of this see [1]).

Since we always map the boundary ∂I^n to a single point, we might as well identify all points of ∂I^n and say that they are all the same point. This identification of the boundary gives us the n-sphere up to a homeomorphism (continuous function with a continuous inverse), so instead of looking at functions from the n-cube to X sending the boundary to x_0 we could have considered functions from the n-sphere S^n to X sending one of the poles to this point. With this alternative description we see that if $\pi_n(X, x_0) = \{1\}$, then any map from S^n to X mapping one of the poles to the point x_0 can be contracted to a point. Since we will see in a few moments that for path-connected spaces the homotopy group is independent on the base point, this means that any n-sphere embedded in X can be contracted to a point. Conversely if any n-sphere can be contracted to a point, the n-th homotopy group can be shown to be trivial.

Proposition 4. *If X is path-connected (i.e. if $\pi_0(X, x_0)$ is trivial) then $\pi_n(X, x_0)$ is isomorphic to $\pi_n(X, x_1)$ for any other point $x_1 \in X$.*

Why this must be true can easily be seen for the case $n = 1$ because for every closed loop γ starting and ending at x_0 we can define a loop starting and ending at x_1 by first traversing some path η from x_1 to x_0 (possible by path-connectedness), then following the loop γ and finally travelling back to x_1 along the path η in the opposite direction, which results an identification of elements of $\pi_1(X, x_0)$ with elements of $\pi_1(X, x_1)$. For $n \geq 2$ something very similar can be done, as is described in more detail in [1].

Any path from x_0 to x_1 gives an isomorphism between $\pi_n(X, x_0)$ and $\pi_n(X, x_1)$, but if X is not *simply connected*[2], this isomorphism may depend on the path chosen. We will often forget about the basepoint and just talk about the n-th homotopy group $\pi_n(X)$ of the space X, but we should keep in mind that it is only possible to define the homotopy group independently from the basepoint for simply connected spaces.

The following proposition will become important when discussing the homotopy groups of Lie groups

Proposition 5. *For some number of path-connected spaces $X_1, \ldots, X_N$ and any non-negative integer $n \in \mathbb{Z}_{\geq 0}$, we have that $\pi_n(X_1 \times \ldots \times X_N) \simeq \pi_n(X_1) \times \ldots \times \pi_n(X_N)$, where the symbol $\simeq$ denotes equality up to a group isomorphism.*

For a proof of this proposition, see [1]

1.2 Lie groups and homotopy

Recall the definition of a Lie group (see [3] for a review on Lie groups)

Definition 6. *A* Lie group *G is a smooth manifold that carries a group structure such that group multiplication ($G \times G \to G, (x, y) \mapsto xy$) and inversion ($G \to G, x \mapsto x^{-1}$) are both described by smooth maps.*

[2] A space X is called simply connected if $\pi_1(X) = \{1\}$

Since Lie groups are smooth manifolds by definition, the homotopy groups discussed above are in particular well-defined for connected Lie groups. We are interested in these homotopy groups because Lie groups will describe the symmetry of the systems we will consider and knowledge of their homotopy groups will turn out to be essential for understanding monopoles.

The first three homotopy groups of a number of important compact connected Lie groups have been collected in table 1.1. We see that the second homotopy group of each of these Lie groups is trivial and this is no coincidence, as it in fact turns out that the second homotopy group of *any* compact Lie group is trivial [4].

Theorem 7. *The second homotopy group of any compact Lie group G is trivial. Moreover, if G is also* semisimple *then its fundamental group is finite.*

The second result, due to Weyl, translates into the statement that every semisimple compact Lie group has a compact universal covering group [5][3]. The real definition of simple and semisimple Lie groups are a little beyond the scope of this paper, but since we will only be dealing with compact groups we can use the following proposition as a definition [3] (N.B. every simple Lie group is semisimple).

Proposition 8. *A compact connected Lie group G is semisimple if and only if its centre, $Z(G) = \{h \in G \mid g\,h\,g^{-1} = h \text{ for all } g \in G\}$, is a discrete subgroup. G is furthermore simple if and only if it has no non-trivial connected normal subgroups*[4].

Examples of (semi)simple Lie groups are easily found: Apart from $\mathrm{U}(n \geq 1)$, $\mathrm{SO}(2)$, $\mathrm{SO}(4)$ and $\mathrm{Spin}(4)$ all of the groups in table 1.1 are *simple* and the groups $\mathrm{SO}(4)$ and $\mathrm{Spin}(4)$ are (only) semisimple. The groups $\mathrm{U}(n \geq 1)$, as well as $\mathrm{SO}(2) = \mathrm{U}(1)$, on the other hand are not even semisimple, as is obvious from the fact that they have infinite fundamental groups. It is useful to note that the product of any number of semisimple Lie groups is again semisimple, but that a product including non-semisimpile Lie groups never is, so for instance $\mathrm{SU}(2) \times \mathrm{SU}(3)$ is semisimple, but $\mathrm{SU}(3) \times \mathrm{U}(1)$ is not.

	$\mathrm{U}(1)$	$\mathrm{U}(\geq 2)$	$\mathrm{SO}(2)$	$\mathrm{SO}(3)$	$\mathrm{SO}(4)$	$\mathrm{SO}(\geq 5)$	other
π_1	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\{1\}$
π_2	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$
π_3	$\{1\}$	$\mathbb{Z}$	$\{1\}$	$\mathbb{Z}$	$\mathbb{Z} \times \mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$

Table 1.1: The homotopy groups of a number of compact connected Lie groups. Here "other" stands for any of the groups $\mathrm{Spin}(n \geq 3)$, $\mathrm{SU}(n \geq 2)$, $\mathrm{Sp}(n \geq 1)$, G_2, F_4, E_6, E_7 or E_8.

A direct consequence of proposition 5 is that the n-th homotopy group of the product $G_1 \times \ldots \times G_N$ of Lie groups is given by the product of the homotopy groups of the separate groups (up to an isomorphism),

$$\pi_n(G_1 \times \ldots \times G_N) \simeq \pi_n(G_1) \times \ldots \times \pi_n(\times G_N). \tag{1.3}$$

For the remainder of this section, suppose that G is some compact connected Lie group and that $H \subseteq G$ is a closed subgroup of G. For any $g \in G$ we say that the

[3] The universal covering group $\tilde{G}$ of G is the unique simply connected Lie group with the same Lie algebra as G. What is relevant for our purposes is that any representation of G is also a representation of $\tilde{G}$.

[4] A subgroup $H < G$ is called normal if $g\,h\,g^{-1} \in H$ for all $h \in H$ and $g \in G$.

set $gH := \{gh \mid h \in H\}$ is a right coset of H in G and we call the space $G/H := \{gH \mid g \in G\}$, consisting of all right cosets, the *coset space* of H in G. The coset space inherits its topology from G: We call a subset $\{gH \mid g \in X\}$ for $X \subseteq G$ open if and only if $XH = \{gh \mid g \in X, h \in H\} \subseteq G$ is open. The space G/H even has a canonical smooth structure, making it a smooth manifold, as discussed in [3].

There exists a set of canonical maps (group homomorphisms actually) between the homotopy groups of G, H and G/H.

- The map i_*^n from $\pi_n(H)$ to $\pi_n(G)$ is relatively simple to define since any map $f\colon (I^n, \partial I^n) \to (H, x_0)$ is also a map from I^n to $G \supseteq H$ mapping the boundary ∂I^n to $x_0 \in G$, so we can just send a class $[f] \in \pi_n(H)$ to its corresponding class $[f] \in \pi_n(G)$. This map is well-defined and it respects the group structure, but it is not necessarily injective or surjective.
- The map j_*^n from $\pi_n(G)$ to $\pi_n(G/H)$ requires just a bit more work, but is almost as simple since any map $f\colon (I^n, \partial I^n) \to (G, x_0)$ defines a canonical map from I^n to G/H mapping the boundary ∂I^n to the *point* $x_0 H \in G/H$ (G/H was a coset space). If the map $\tilde{f}\colon (I^n, \partial I^n) \to (G/H, H)$ is defined as $\tilde{f}(s_1, \ldots, s_n) = f(s_1, \ldots, s_n)\, H$, then j_*^n will can be defined to send $[f] \in \pi_n(G)$ to $[\tilde{f}] \in \pi_n(G/H)$.
- There also exists a canonical map ∂_*^{n+1} from $\pi_{n+1}(G/H, H)$ to $\pi_n(H, 1)$, but this map is more complicated. For any map from $f\colon (I^{n+1}, \partial I^{n+1}) \to G/H$ there exists a map $g\colon I^{+1} \to G/H$ such that $g(\partial I^{n+1}) \subseteq H$ and such that furthermore $g(s_1, \ldots, s_n, s_{n+1}) = 1 \in H$ if at least one of the coordinates $s_1, \ldots, s_n$ is equal to either 1 or 0 or if $s_{n+1} = 0$ [5]. We can now define the function

  $$h\colon (I^n, \partial I^n) \to (H, 1), \qquad h(s_1, \ldots, s_n) = g(s_1, \ldots, s_n, 1). \tag{1.4}$$

  This defines an element $[h] \in \pi_n(H, 1)$ that turns out to be unique (independent of the choice of g), so we obtain a canonicial map from $\pi_{n+1}(G/H, H)$ to $\pi_n(H)$. By replacing H by $x_0 H$ everywhere this becomes a map from $\pi_{n+1}(G/H, x_0 H)$ to $\pi_n(x_0 H) \simeq \pi_n(H)$. For more details on this map see [6] or [7].

It turns out that these maps fit together very nicely in an *exact sequence*. An exact sequence is a sequence of maps $f_i\colon A_i \to A_{i+1}$ (for $i \in \mathbb{Z}$, the sequence is allowed to end for some minimal and maximal value of i) such that for each i the kernel of f_i is equal to the image of f_{i-1}.

Theorem 9. *The maps described above form an exact sequence of group homomorphisms*

$$\ldots \to \pi_n(G) \xrightarrow{j_*^n} \pi_n(G/H) \xrightarrow{\partial_*^n} \pi_{n-1}(H) \xrightarrow{i_*^{n-1}} \pi_{n-1}(G) \to \ldots \xrightarrow{j_*^1} \pi_1(G/H). \tag{1.5}$$

This sequence actually continues all the way to $\pi_0(G)$, but the maps will not be group homomorphisms there since π_0 is not a group.

I will not prove this theorem here, but it is actually related to theorem 4.3 from [1], as pointed out by [8]. A practical proof for the relevant part of this theorem, which is the bijectivity of the map $\pi_2(G/H) \to \pi_1(H)$ if $\pi_1(G) = \pi_2(G) = \{1\}$, can be found in [6, 7].

Corollary 10. *If G is a compact Lie group and H is a closed subgroup, then $\pi_2(G/H)$ is isomorphic to the kernel of the map $i_*^1\colon \pi_1(H) \to \pi_1(G)$. In particular, if G is also simply connected then $\pi_2(G/H)$ is isomorphic to $\pi_1(H)$*

[5] It is generally not possible to demand that also $g(s_1, \ldots, s_{n+1}) = 1$ if $s_{n+1} = 1$, which is essential!

This result follows directly from theorem 9 because any compact Lie group has $\pi_2(G) = \{1\}$, which implies that $\ker(\partial^2_*) = \mathrm{im}(j^2_*) = \{1\}$ (so ∂^2_* is injective) and $\mathrm{im}(\partial^2_*) = \ker(i^1_*)$.

A *group action* of G on a (smooth) manifold is defined through a (smooth) function $\alpha\colon G \times M \to M, (g, x) \mapsto g\,x$ such that $g(h\,x) = (g\,h)x$ and $1\,x = x$ (1 is the unit element in G). The stabiliser G_x of x in G is a closed subgroup of G defined as

$$G_x = \{g \in G \mid g\,x = x\} \tag{1.6}$$

Theorem 11. *If G acts* transitively *on M (meaning that for any two points $x, y \in M$ there exists an element $g \in G$ such that $y = g\,x$) then M is diffeomorphic to G/G_x for any $x \in M$.*

That there exists a canonical bijection between the two spaces is relatively easy to see because we can associate to any point $y \in M$ the (non-empty) set $G_{x,y}$ of group elements that map the specific point x to y and $G_{x,y}$ turns out to be an element of G/G_x. This is because for any $g \in G_{x,y}$ and any $h \in G_x$ we have $(g\,h)x = g(h(x)) = g(x) = y$, so $g\,G_x \subseteq G_{x,y}$ and conversely, if $g, g' \in G_{x,y}$ then $g^{-1}\,g' \in G_x$, so $g' = g\,(g^{-1}g') \in g\,G_x$, so we even see that $G_{x,y} = g\,G_x$ and hence that $G_{x,y} \in G/G_x$. See [3] for a more rigourous proof that takes the smooth structure into account.

2. Monopoles as topological solitons

In general, a *topological soliton* is a solution to a set of partial differential equations that cannot be deformed into a “trivial solution” because of some “topological constraint”. What exactly is meant by a trivial solution depends on the context, but the topological constraint is usually the result of boundary conditions that solutions are required to obey. By a deformation from one solution to another we mean a homotopy that starts with the first, ends up at the other and stays in the space of solutions along the way. It is important to note that what we have defined here as a topological soliton is not generally a soliton in the traditional sense.

In the context of (classical) field theory, the partial differential equations are the field equations of the theory and the trivial solutions are the vacuum solutions, i.e. the solutions that minimise the energy. There are still multiple ways to impose boundary conditions on such a system, but we will choose to demand that the total energy of the system is finite (with respect to the vacuum solutions) [6,7]. It will turn out that this condition will give rise to topological conservation laws for certain systems and thereby imply the existence of monopole solutions.

2.1 Yang-Mills-Higgs models

In the remainder of this chapter we will consider a *gauge theory* defined in $3+1$ dimensions (three spatial dimensions and one time dimension) with a local symmetry defined through a Lie group G. This gauge theory will have the following ingredients:

- An n-dimensional scalar *Higgs field* $\phi = (\phi_1, \dots, \phi_n)$, which we will assume to be real, although it is also allowed to be complex[1]. Two such fields are multiplied with each other through the inner product on $\mathbb{R}^n$.
- A compact connected Lie group G of some dimension D used to define gauge transformations, $\phi(x^\mu) \mapsto g(x^\mu)\phi(x^\mu)$. The field ϕ should transform under gauge transformations via some *unitary representation* $R(G)$ of G (by unitary we mean that elements of $R(G)$ preserve the inner product on $\mathbb{R}^n$).
- A (vector) gauge field $A_\mu = A_\mu^a t^a$ that take values in the Lie algebra $\mathfrak{g}$ of $R(G)$,

[1] We can always write a complex field as two real fields, but we do not need to because up to a few factors and complex conjugations everything also goes through if we keep it complex.

defines a covariant derivative $D_\mu\phi = \partial_\mu\phi + i\,e\,A^a_\mu t^a\phi$ and transforms under gauge transformations in such a way that $D_\nu\phi$ transforms in the same way as ϕ does.
Here the elements $t^1, \ldots, t^D$ form a set of generators for $R(G)$ such that any element of $R(G)$ can be written as a product of expressions of the form $\exp(i\,g^a t^a)$. These generators are Hermitian matrices, so we can choose them such that $\mathrm{Tr}(t^a t^b) = \delta_{ab}$ (This trace defines an inner product on the space of Hermition matrices, so we can use the Gramm-Schmidt orthogonalisation proces).

- A field-strength tensor $F_{\mu\nu} = [D_\mu, D_\nu]/(i\,e) = \partial_\mu A_\nu - \partial_\nu A_\mu + i\,e\,[A_\mu, A_\nu]$, defined in terms of the gauge field A_μ, which we can write as $F_{\mu\nu} = F^a_{\mu\nu}t^a$ as with A_μ. The Lagrangian should contain a *Yang-Mills term* $\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) = F^a_{\mu\nu}F^{a\mu\nu}$, which can be verified to be a gauge-invariant expression.
- A potential $V(\phi)$ that takes the minimal value of zero[2]. The set of all points $\phi \in \mathbb{R}^n$ for which $V(\phi)$ is zero form the *vacuum manifold* $\mathcal{M} = \{\phi \in \mathbb{R}^n \mid V(\phi) = 0\}$. We will assume that G acts transitively on $\mathcal{M}$[3]. We have seen in section 1.2 that this means that $\mathcal{M} \cong G/H$, where H is the symmetry group that remains after symmetry breaking.

With these fields we can build the following gauge invariant Lagrangian, which will define the dynamics of this *Yang-Mills-Higgs* system

$$\mathcal{L} = -\tfrac{1}{2}(D_\mu\phi)(D^\mu\phi) - V(\phi) - \tfrac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}. \tag{2.1}$$

To specify a solution we only need to specify the initial field configurations $\phi(x_i, t)$ and $A^a_\mu(x_i, t)$ at some fixed time t and their first time derivaties $\partial_0\phi(x_i, t)$ and $\partial_0 A_\mu(x_i, t)$. Given a set of such initial values, the field equations uniquely determine the time-evolution of the system and a continuous deformation of the initial values leads to a continuous deformation of the solutions on the entire space-time [7]. Since the total energy of a solution is conserved, the finite energy condition will also continue to hold at later points in time, so it suffices to look at the initial data (ϕ, $\partial_0\phi$, A_μ and $\partial_0 A_\mu$) and we are at the moment not interested in solving the equations of motion for these fields. From now on we will therefore be working at some fixed time t, unless stated otherwise.

Before moving on, we will first impose a gauge on the system. It is always possible to apply a gauge transformation to make the time component of the gauge fields vanish, so we can choose a gauge such that $A_0 = A^a_0 t^a = 0$ everywhere [7]. After choosing this gauge we are still free to make time-independent gauge transformations[4] since these do not affect A^a_0. In this gauge we get simple time derivaties $D_0\phi = \partial_0\phi$ and we furthermore find that $F_{0i} = \partial_0 A_i - \partial_i A_0 - i\,e\,[A_0, A_i] = \partial_0 A_i$. The total energy in this gauge reads $E = T + \mathcal{V}$ with

$$T = \int \mathrm{d}^3x \Big(\tfrac{1}{2}(\partial_0 A^a_i)(\partial_0 A^a_i) + \tfrac{1}{2}(\partial_0\phi)(\partial_0\phi)\Big) \tag{2.2}$$

$$\mathcal{V} = \int \mathrm{d}^3x \Big(\tfrac{1}{2}(D_i\phi)(D_i\phi) + V(\phi) + \tfrac{1}{4}F^a_{ij}F^{aij}\Big) \tag{2.3}$$

[2]We assume the minimum value to be zero because we want to measure energy with respect to the groundstate.
[3]Recall that this means that for any two elements $\phi, \psi \in \mathcal{M}$ there exists a group element $g \in G$ such that $\phi = g\,\psi$. This is a reasonable assumption since any deviation from it is coincidental and highly unlikely when quantum corrections and finite energy contributions to the potential are considered.
[4]After a time-independent gauge transformation $A_0 = 0$ becomes $g\,A_0 g^{-1} + i\,e^{-1}(\partial_0 g)g^{-1} = 0 + 0 = 0$.

It is important to note that all of these terms separately are non-negative (including the field-strength term), so in order for the total energy of the system to be finite, each of them should give a finite contribution.

2.2 Finding finite energy solutions

At this time it becomes useful to partially switch to vector notation: We will write $\mathbf{A} = (A_1, A_2, a_3)$, $\mathbf{D}\phi = (D_1\phi, D_2\phi, D_3\phi)$ and $\nabla\phi = (\partial_1\phi, \partial_2\phi, \partial_3\phi)_i$ and use a dot $(\cdot)$ to denote the inner product between two such expressions. Because we are interested in the asymptotic behaviour of ϕ and $\mathbf{A}$ towards infinity in different directions, we will write any spatial point as $r\,\hat{\mathbf{x}} \sim (r, \hat{\mathbf{x}})$ with $(r, \hat{\mathbf{x}}) \in [0, \infty[\times S^2$. We could alternatively switch to polar coordinates and write $(x_1, x_2, x_3) = r\,(\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta)$ with $r \in [0, \infty[$, $\varphi \in [0, 2\,\pi[$ and $\vartheta \in [0, \pi]$, but that would make the notation more complex.

We still have the freedom to make time-independent gauge transformations, which we can use to choose a *radial gauge* and transform away the radial component of $\mathbf{A}$ (at our fixed point in time) everywhere except a small region around the origin (say for all $r \geq 1$) [7]. Here the radial component of $\mathbf{A}(r\,\hat{\mathbf{x}})$ is defined to be $A_r(r\,\hat{\mathbf{x}}) := \hat{\mathbf{x}} \cdot \mathbf{A}$.

If the energy is to be finite, then the following expression should be too

$$E \geq \int_{\mathbb{R}^3} \left\{ \tfrac{1}{2}(\mathbf{D}\phi)^2 + V(\phi) \right\} \geq \int_1^\infty \mathrm{d}r\, r^2 \int_{S^2} \mathrm{d}\Omega \left\{ \tfrac{1}{2}(\mathbf{D}\phi)^2 + V(\phi) \right\} \equiv \int_1^\infty \mathrm{d}r\, I(r, \phi) \tag{2.4}$$

where $\mathrm{d}\Omega$ is the standard measure on S^2, which would be $\int_{S^2} \mathrm{d}\Omega = \int_0^{2\,\pi} \mathrm{d}\varphi \int_{-1}^{1} \mathrm{d}(\cos\vartheta)$ in polar coordinates. For this to converge, we in particular need the (convergent) angular integral $I(r, \phi)$ to vanish as $r \to \infty$. Since the integrand is non-negative everywhere, this can only happen if the integrand vanishes as $r \to \infty$, so for any $\hat{\mathbf{x}} \in S^2$. We therefore see that $r^2(\mathbf{D}\phi(r\,\hat{\mathbf{x}}))^2 \to 0$ and $r^2 V(\phi(r\,\hat{\mathbf{x}})) \to 0$ as $r \to \infty$.

This tells us in particular that $\hat{\mathbf{x}} \cdot r\,(\mathbf{D}\phi) = r\,(\partial_r + i\,e\,\hat{\mathbf{x}} \cdot \mathbf{A})\phi) = r\,\partial_r\phi$ vanishes as $r \to \infty$ (recall that $\hat{\mathbf{x}} \cdot \mathbf{A} = 0$), so that for any $\hat{\mathbf{x}} \in S^2$, the limit

$$\phi_\infty(\hat{\mathbf{x}}) \equiv \lim_{r\to\infty} \phi(r\,\hat{\mathbf{x}}) = \lim_{r\to\infty} \phi(r, \varphi, \vartheta) \tag{2.5}$$

necessarily exists. Because we also know that $r^2 V(\phi(r\,\hat{\mathbf{x}})) \to 0$, we find that furthermore $\phi_\infty(\hat{\mathbf{x}}) \in \mathcal{M}$ [7, 9]. This defines a new function $\phi_\infty \colon S^2 \to \mathcal{M}$ which is continuous[5].

We have seen that that in order for a field configuration $\phi(r\,\hat{\mathbf{x}})$ to give a finite energy we need $r\partial_r\phi(r\,\hat{\mathbf{x}})$ and $r^2 V(\phi(r\,\hat{\mathbf{x}}))$ to vanish at infinity and that an *asymptotic function* $\phi_\infty \colon S^2 \to \mathcal{M}$ generally exists in this case. Conversely, for any continuous function $\phi_\infty \colon S^2 \to \mathcal{M}$, there clearly exists a field configuration $\phi(r\,\hat{\mathbf{x}})$ such that $\phi(r\,\hat{\mathbf{x}}) = \phi_\infty(\hat{\mathbf{x}})$ for all $r \geq 1$.

[5] Technically, we could only conclude that the integrand converges everywhere except on a subset of S^2 of measure 0, which would also spoil the continuity of ϕ_∞. Since this exceptional situation is of no consequence for the final result, we will step over it.

For this solution that $\phi(r\,\hat{\mathbf{x}}) = \phi_\infty(\hat{\mathbf{x}}) \in \mathcal{M}$ for $r \geq 1$, so it follows that $\nabla\phi(r\,\hat{\mathbf{x}}) = (\partial_1, \partial_2, \partial_3)\phi(r\,\hat{\mathbf{x}})$ consists of three vectors tangent to $\mathcal{M}$. The assumption that G acts transitively on $\mathcal{M}$ tells us that the tangent space of $\mathcal{M}$ at $\phi(r\,\hat{\mathbf{x}})$ is spanned by the action of the Lie algebra on this point, so that an infinitesimal change in ϕ can be cancelled by an infinitesimal group action. Let $\mathfrak{h}_{\hat{\mathbf{x}}} = \{s \in \mathfrak{g} \mid s\,\phi_\infty(\hat{\mathbf{x}}) = \phi_\infty(\hat{\mathbf{x}})\}$ and $\mathfrak{h}_{\hat{\mathbf{x}}}^\perp = \{s \in \mathfrak{g} \mid \forall s' \in \mathfrak{h}_{\hat{\mathbf{x}}} \colon \operatorname{Tr}(s\,s') = 0\}$ (the trace defines an inner product on $\mathfrak{g}$), then for any $r\,\hat{\mathbf{x}} \in \mathbb{R}^3$ there exists a unique vector $\mathbf{A}(r\hat{\mathbf{x}}) = (A_1, A_2, A_3) \in (\mathfrak{h}_{\hat{\mathbf{x}}}^\perp)^3$ such that $\mathbf{A}(r\hat{\mathbf{x}})\phi(r\,\hat{\mathbf{x}}) = (A_1\phi, A_2\phi, A_3\phi)(r\,\hat{\mathbf{x}}) = i\,e^{-1}\nabla\phi(r\,\hat{\mathbf{x}})$. For this choice we see that the covariant derivative $\mathbf{D}\phi(r\,\hat{\mathbf{x}}) = \nabla\phi + i\,e\,\mathbf{A}(r\hat{\mathbf{x}})\phi(r\,\hat{\mathbf{x}}) = 0$ whenever $r \geq 1$, so the covariant derivative term only contributes a finite amount to the energy[6].

We can choose the gauge fields in such a way that the contribution from the covariant derivatives becomes finite (only the finite region with $r \leq 1$ contributes). This choice for the gauge fields actually links the behaviour of $\mathbf{A}$ for large r to that of ϕ. This is because at the point $\mathbf{y} = (y_1, y_2, y_3) = \alpha\,\mathbf{z} = (\alpha\,z_1, \alpha\,z_2, \alpha\,z_3)$ with $|\mathbf{z}| = 1$ and $\alpha > 1$

$$\nabla\phi(\alpha\mathbf{z}) = \frac{\partial}{\partial y^i}\phi(\mathbf{y}) = \frac{1}{\alpha}\frac{\partial}{\partial z^i}\phi(\alpha\mathbf{z}) = \frac{1}{\alpha}\frac{\partial}{\partial z^i}\phi(\mathbf{z}) = \frac{1}{\alpha}\nabla\phi(\mathbf{z}) \tag{2.6}$$

since $\phi(\mathbf{y}) = \phi_\infty(\mathbf{y}/|\mathbf{y}|)$ for any $\mathbf{y}$ with $|\mathbf{y}| \geq 1$, which implies that

$$\lim_{r\to\infty} r\,\mathbf{A}(r\,\hat{\mathbf{x}})\phi(r\,\hat{\mathbf{x}}) = \lim_{r\to\infty} i\,e^{-1} r\,\nabla\phi(r\,\hat{\mathbf{x}}) = \lim_{r\to\infty} i\,e^{-1}\nabla\phi(\hat{\mathbf{x}}) = i\,e^{-1}\nabla\phi(\hat{\mathbf{x}}), \tag{2.7}$$

which is finite. Therefore the gauge fields go like $1/r$ as $r \to \infty$ [7], which means that the field strength $F_{\mu\nu}$ goes like $1/r^2$ and that the square of the field strength goes like $1/r^4$. Since we are working in three spatial dimensions, this tells us that the field strength term in equation (2.3) will give a finite contribution. Note that in four or more dimensions this argument will no longer work because the integral of the Yang-Mills term, which goes like $1/r^4$ diverges then.

If we look at the expression for T in equation (2.2), we see that the energy also has a contribution from the time derivates $\partial_0 A_i^a$ and $\partial_0\phi$. These time derivatives however form a completely independent set of initial data and we can even choose $\partial_0\phi(x_i, t) = \partial_0 A_j(x_i, t) = 0$ for any x_i at the fixed time t, in which case $T = 0$. It is furthermore always possible to continuously deform any other set of initial values for the time derivatives giving a finite energy contribution into this trivial case without blowing up the energy by simply gradually scaling all time derivatives down to 0, so the choice of initial values for the time derivatives is of no consequence for the classification of solutions up to continuous deformations. We should note that this freedom in the choice of initial values for the time derivatives does not imply that we can choose a static solution because we cannot say anything about the second time derivatives.

2.3 Topological conservation laws

We have shown that for any initial field configuration for ϕ and A_i there generally exists a asymptotic function $\phi_\infty \colon S^1 \to \mathcal{M}$ and conversely that for any such function there exists a finite energy field configuration.

[6] N.B. The radial gauge is not broken: $(\hat{\mathbf{x}} \cdot \mathbf{A})\phi = \hat{\mathbf{x}} \cdot (i\,e^{-1}\nabla\phi) = 0$ (since $\phi(r\,\hat{\mathbf{x}}) = \phi_\infty(\hat{\mathbf{x}})$ implies that $\nabla\phi$ has no radial component), so $\mathbf{A} = \mathbf{A} - (\hat{\mathbf{x}} \cdot \mathbf{A})\hat{\mathbf{x}}$ since $\mathbf{A}$ was unique, so it has no radial component

Any two field configurations with the same asymptotic function ϕ_∞ can be continuously deformed into each other without blowing up the energy by first gradually resolving things around infinity (without straying farther away from the vacuum manifold) and using the gauge fields to make the angular part of the gradient vanish. After that nothing we do (smoothly) in the finite region of space that remains can make the energy blow up. The asymptotic function ϕ_∞ is therefore enough to classify solutions up to continuous deformations [9, 6].

In fact, any two field configurations can be continuously deformed into each other without blowing up the energy if and only if the associated functions $\phi_\infty \colon S^2 \to \mathcal{M}$ can be continuously deformed into each other [7]. This therefore gives us a one-one correspondence between the classes of field configurations that can be deformed into each other without blowing up the energy and the classes of asymptotic functions $\phi_\infty \colon S^2 \to \mathcal{M}$ that can be deformed into each other. With the definitions from section 1.1 we see that these are exactly the homotopy classes of maps from S^2 to $\mathcal{M}$ (without reference to a basepoint), so we can associate any class of solutions that cannot be deformed into each other with a single such *homotopy class* $[\phi_\infty]$.

We now see the second homotopy group from section 1.1 appearing naturally, as we had seen that $\pi_2(\mathcal{M})$ was non-trivial if and only if there exist maps from S^2 to $\mathcal{M}$ that cannot be deformed into a constant map. The system therefore admits topological solitons (of this type) if and only if $\pi_2(\mathcal{M})$ is non-trivial. We call topological solitons that arise by this argument *monopoles*.

If $\mathcal{M}$ is simply-connected ($\pi_1(\mathcal{M}) = \{1\}$) we can associate any solution to an element of the homotopy group $\pi_2(\mathcal{M})$ since its definition is then completely independent of the basepoint used. If this is not the case, we can only associate it to an element of the homotopy group *up to changes of the base point*. The groundstate solution however always corresponds to just the trivial element $1 \in \pi_2(\mathcal{M})$ since it is uniquely carried over to different basepoints (no group homomorphism can send an identity elements to anything other than an identity element).

This classification of solutions leads to a *topological conservation law*: Since time evolution basically defines a continuous deformation of the initial configuration that preserves the energy, we would find exactly the same class $[\phi_\infty]$ of maps from S^2 to $\mathcal{M}$ if we look at the field configurations some time later. We can therefore view the homotopy class $[\phi_\infty]$ corresponding to a solution as a conserved charge, which we appropriately call the topological charge. In some cases this charge corresponds to some other physical observable, such as the magnetic charge in the case of the 't Hooft-Polyakov monopole, which will be discussed in chapter 3.

3. Magnetic monopoles

3.1 The Dirac monopole

The well-known Maxwell equations can be consistently modified to include the possibility of particles carrying a magnetic charge [10]. After this modification, the equations look like

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \rho_{\mathrm{e}} & \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} - \mathbf{j}_{\mathrm{m}} \\ \nabla \cdot \mathbf{B} &= \rho_{\mathrm{m}} & \nabla \times \mathbf{B} &= \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j}_{\mathrm{e}}, \end{aligned} \tag{3.1}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic field respectively, ρ_{e} and $\mathbf{j}_{\mathrm{e}}$ are the electric charge density and current. Two additional terms have been added, namely the magnetic charge ρ_{m} and magnetic current $\mathbf{j}_{\mathrm{m}}$.

In analogy with the electric monopole, the magnetic field of a typical magnetic monopole of charge g is expected to look like

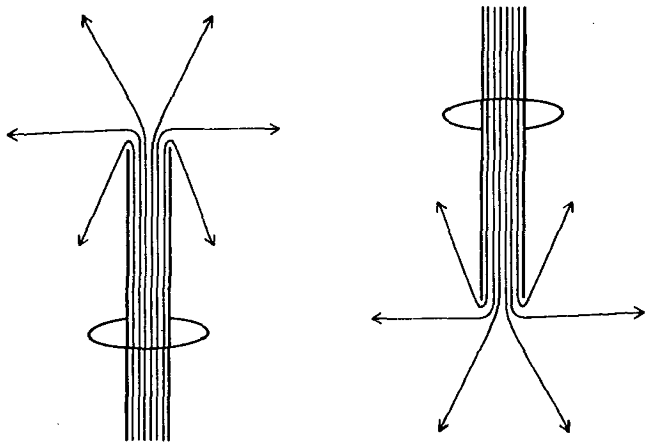
$$\mathbf{B}_{\mathrm{mm}}(\mathbf{r}) = \frac{g}{4\,\pi\,r^2}\,\hat{\mathbf{r}}, \tag{3.2}$$

which results in a total flux of

$$\Phi_{\mathrm{mm}} = \int_S \mathrm{d}\mathbf{S} \cdot \mathbf{B} = g \int_S \frac{\mathrm{d}\mathbf{S} \cdot \hat{\mathbf{r}}}{4\,\pi\,r^2} = g. \tag{3.3}$$

When this is done in the classical theory of electrodynamics, no inconsistencies arise, but for the quantum-mechanical description, the existence of a vector potential $\mathbf{A}$ such that $\mathbf{B} = \nabla \times \mathbf{A}$ is essential [6, 9, 11]. Plugging an expression of the form $\mathbf{B}_{\mathrm{mm}} = \nabla \times \mathbf{A}_{\mathrm{mm}}$ into equation (3.3) however will always yield a zero flux by Stokes' theorem so, even if we exclude the singularity at the origin, the vector potential $\mathbf{A}$ will not be well-defined.

It *is* however possible to define such a vector potential on any contractible region of space that does not contain the origin, so we can define such a potential everywhere except for the origin and some line (not necessarily straight) from the origin to infinity. We can see why this is from figure 3.1 because the field lines can vanish to infinity through this so-called *Dirac string*, which we could for instance imagine to be an infinitely long and thin solenoid (coil).



(a) One possible choice (b) Another choice

Figure 3.1: Two choices for the Dirac string. This figure originally appeared in [6].

It turns out that if the magnetic charge satisfies a simple quantisation condition, the Dirac string becomes completely undetectable [11, 12, 6]. We can for instance hope to describe electromagnetism throughout our space with two vector potentials $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ defined on the entire space apart from the negative and positive z-axis respectively (so corresponding to figure 3.1(a) and 3.1(b) respectively). In polar coordinates (r, θ, φ) (with θ the zenith angle, φ the azimuth angle and r the distance to the origin) we can for instance write down the vector potentials

$$\mathbf{A}^{(1)} = \frac{g}{4\pi r}\frac{1-\cos\theta}{\sin\theta}\mathbf{e}_\varphi \qquad \text{and} \qquad \mathbf{A}^{(2)} = \frac{-g}{4\pi r}\frac{1+\cos\theta}{\sin\theta}\mathbf{e}_\varphi, \tag{3.4}$$

where $\mathbf{e}_\varphi = (-\sin\varphi, \cos\varphi, 0)$. The first can be defined as long as $\theta \neq \pi$ and the second as long as $\theta \neq 0$ and a simple calculation furthermore shows that taking the curl of these potentials yields the magnetic field for a monopole from equation (3.1).

In addition to this, we see that on the intersection of the regions where these potentials are defined (everywhere except for the entire z-axis in this case) they are related by the gauge transformation

$$\mathbf{A}^{(2)} = \mathbf{A}^{(1)} + \nabla\alpha \qquad \text{with } \alpha = \frac{g}{2\pi}\varphi. \tag{3.5}$$

By looking at this transformation however, we immediately see a problem: While $\nabla\alpha$ may be well-defined, α is not as it increases by g if you turn around the z-axis once. We say that α is multiply-defined, so it is only defined up to the addition of some multiple of g, which is not really a problem, since α itself has not yet appeared anywhere, only its gradient $\nabla\alpha$. For different choices for the location of the Dirac strings similar results would be obtained, so to be able to glue together the vector potentials on any two contractible regions we need to allow gauge parameters that are defined up to multiples of g.

When another field ϕ, corresponding to particles with an electric charge e, is added, it should couple to the magnetic field (and also to the electric field, which we are ignoring for the moment) via the covariant derivative through $D_i\phi = \partial_i\phi + i\,e\,A_i\phi$ and transform under the gauge transformations described above as $\phi \mapsto e^{i\,e\,\alpha}\phi$ [12]. This field ϕ cannot be multiply defined since a change in phase as it rotates around the z-axis would be measurable, so we require that $e^{i\,e\,\alpha} = e^{i\,e\,(\alpha+n\,g)}$ for any $n \in \mathbb{Z}$. This holds if and only if $e\,g \in 2\,\pi\,\mathbb{Z}$.

This tells us that the charge of the monopole should be a multiple of the *Dirac charge*, $2\,\pi/e$, and that everything is still well-defined everywhere except in the origin if this is the case [11, 6, 12]. Conversely, this condition also tells us that if a single magnetic monopole with charge g exists, any electrically charged particle should have a charge that is a multiple of $2\,\pi/g$. This is called the *Dirac quantisation condition* [6, 12].

Thus, the existence of magnetic monopoles explain the observed (but unexplained) quantisation of electric charge, making the idea of the existence of magnetically charged particles extremely appealing. Nevertheless, the Dirac monopole is not completely satisfactory as we have had to cut out a point (the origin) from our space to define it. In section 3.4 we will see a more successful description of a magnetic monopole (as a topological soliton) that replaces the singularity in the origin by a (smooth) core, but can be described in exactly the same way when observed from large distances.

3.2 Grand Unified Monopoles

Grand Unified Theories (GUTs) seek to unify the electroweak and strong forces into a single fundamental interaction described by a simple compact gauge group G_{GUT} that contains the *Standard model* gauge group $G_{\text{SM}} = \mathrm{SU}(3)_{\text{c}} \times \mathrm{SU}(2)_{I_{\text{w}}} \times \mathrm{U}(1)_{\text{Y}}$ as a subgroup. The GUT itself is a linear gauge theory of roughly the type discussed in section 2.1. It has a Higgs field coupling to gauge fields through a covariant derivative, a Yang-Mills field-strength term and some gauge invariant potential. If the potential for the Higgs field is chosen right, the original symmetry described by G_{GUT} will be broken to the Standard model gauge group at low temperatures and eventually to $\mathrm{SU}(3)_{\text{c}} \times \mathrm{U}(1)_{\text{em}}$ [12, 13, 14]. The original GUT symmetry is expected to be restored at temperatures around the GUT scale, about 10^{16} GeV.

The appeal of Grand Unified Theories lies not only in the fact that a single simple symmetry group is more aesthetically pleasing than the product of Lie groups that the Standard model gives. Unification of the three fundamental forces into a single symmetry group means that only a single coupling constant is needed to describe the coupling of fields to all the gauge groups, which after *symmetry breaking* reduce to the separate coupling constants for $\mathrm{SU}(3)_{\text{c}}$, $\mathrm{SU}(2)_{I_{\text{w}}}$ and $\mathrm{U}(1)_{\text{Y}}$. Embedding the electromagnetic gauge group $\mathrm{U}(1)_{\text{em}}$ in a compact simple Lie group would furthermore make the quantisation of electric charge, which is not a strict requirement of quantum electrodynamics, necessary [13].

In section 2.3 we have analysed the conditions for such linear gauge theories to admit monopole solutions and the conclusion was that monopole solutions exist if and only if the vacuum manifold $\mathcal{M}$ of the potential contains non-contractible 2-spheres, i.e. if $\pi_2(\mathcal{M}) \neq \{1\}$. Under the reasonable assumption that the Grand Unified Gauge group acts transitively[1] on $\mathcal{M}$, the vacuum manifold will be diffeomorphic to the coset space $\mathcal{M} \cong G_{\text{GUT}}/G_{\text{SM}}$. Since G_{GUT} is a simple Lie group, it can be replaced by a compact covering group which has both a trivial second *and* first homotopy group[2], so the homotopy theorems mentioned in section 1.2 tell us that

$$\pi_2(\mathcal{M}) \simeq \pi_1(\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)) = \pi_1(\mathrm{U}(1)) \simeq \mathbb{Z}, \tag{3.6}$$

which is non-trivial. The existence of monopole solutions is therefore a general prediction of Grand Unified Theories. The mass of such a monopole is the result of an integral of the energy density from equations (2.2) and (2.3) and is typically of the order 10^{17} GeV [6, 8, 12]. Between the Grand Unified phase transition and now there has also been an electroweak symmetry breaking in which the standard model gauge group breaks further, to $\mathrm{SU}(3)_{\mathrm{c}} \times \mathrm{U}(1)_{\mathrm{em}}$. This does not change a lot about the monopoles already created since $\pi_1(\mathrm{SU}(3) \times \mathrm{U}(1)) = \pi_1(G_{\text{SM}}) = \mathbb{Z}$, .

The magnetic behaviour of these monopoles is not immediately obvious at first sight since we have not even defined what we mean by electromagnetism. The value of the Higgs field at some point in space-time far away from the monopole corresponds to an element $g\,G_{\text{SM}} \simeq G_{\text{SM}}$ of $G_{\text{GUT}}/G_{\text{SM}}$. This element is the symmetry group to which the original GUT symmetry breaks at that point, so around a monopole the definition of electromagnetism itselfrotates in some sense, which explains the magnetic nature of the monopole. This can be seen explicitly in the example discussed in section 3.4.

Although it would be preferable to assume that the gauge group G_{GUT} is simple, semisimple Lie groups are also often considered. Grand Unified Theories with semisimple Lie groups also admit monopole solutions by exactly the same arguments. If also the assumption that G_{GUT} is semisimple is dropped, the existence of monopole solutions is no longer guaranteed and a more detailed analysis of the groups involved and the way in which symmetry is broken is required.

3.3 Monopole formation

We have argued in section 3.2 why Grand Unified Theories with semisimple symmetry groups generally predict the existence of magnetic monopole solutions after symmetry breaking, but this does not explain why we might expect them to appear in nature and it certainly does not give any predictions about how many we might expect to find in the visible Universe. For now, we will assume that there is only a single *phase transition* from the original GUT symmetry group to the standard model gauge group at the GUT scale, so at a temperature of around $T_{\text{GUT}} \sim 10^{16}$ GeV and that there is no inflation.

Above this critical temperature the original GUT symmetry was restored and the Higgs field Φ had a vanishing expectation value. As the temperature drops below T_{GUT}, it becomes favourable for the Higgs field to assume a non-zero expectation value in the

[1] Recall that this means that for any two element $x, y \in \mathcal{M}$ there exists a $g \in G_{\text{GUT}}$ such that $y = g\,x$. This assumption is reasonable because any deviation from it would be completely coincidental and highly unlikely once thermal contributions to the potential are considered (you would need to choose the temperature just right).

[2] This is possible because any representation of G_{GUT} is also a representation of its covering group, which has a trivial first and second homotopy group, as mentioned in section 1.2

vacuum manifold of the effectively potential and thereby break the symmetry. What value is chosen for the expectation value $\langle\Phi\rangle$ depends on random fluctuations, so we would expect different choices to be made in different regions of space. At the time of monopole formation, there is some *correlation length*, denoted by ξ, which is the typical range beyond which fluctuations in Φ are uncorrelated. The regions in which Φ assumes a different expectation value are determined by such fluctuations and will thus have a size of roughly order ξ^3 [6, 8, 12, 15].

The value of this correlation length depends on the details of phase transition (it generally diverges at the critical temperature), but we can write down an upper bound for it. Since regions that are not in causal contact cannot possibly be correlated in any way, the correlation length should satisfy $\xi < \ell_{\text{GUT}}$, where ℓ_{GUT} is the *causal horizon* at the time of the GUT phase transition[3], which is roughly of the order of 10^{-27} cm. The discussion above assumed that the phase transition was of second order, but it also applies to the case with a first order phase transition. In this case the relevant time and temperature become those at which the nucleation of bubbles becomes probable. The parameter ξ should then be replaced by the typical bubble size, which is still bounded in the same way by causality [6].

Where different regions meet, they will generally try to align to minimise their energy, but this is not always possible if the vacuum manifold has a non-trivial homotopy group. In the case of monopoles $\pi_2(\mathcal{M})$ is the relevant homotopy group, so if this homotopy group is non-trivial (as is the case for any GUT with a semisimple symmetry group) then there is a probability p that the orientation of the Higgs field around this point is topologically non-trivial and a monopole will form. This is called the *Kibble mechanism*. The exact value of p will not be much smaller than 1 and depends on the shape of the vacuum manifold and can be estimated by discretising the vacuum manifold and looking at how of the possible configuration around a point result in monopoles. We will assume p to be of order 0.1, which is the value that would be obtained if the vacuum manifold is the 2-sphere. If we take the GUT scale to be roughly 10^{16} GeV, the horizon to be $\ell_{\text{GUT}} \sim 8 \times 10^{-28}(T_{\text{GUT}}/10^{16}\text{GeV})^{-2}$ cm and $p \sim 0.1$, then right after the phase transition the number density of monopoles is at least [6]

$$n_{\text{MM,GUT}} \sim p\,\xi^{-3} \gtrsim p\,\ell_{\text{GUT}}^{-3} \sim 2 \times 10^{80}(T_{\text{GUT}}/10^{16}\text{ GeV})^6\ \text{cm}^{-3}. \tag{3.7}$$

Another way to express this is through the dimensionless ratio of the monopole number density and the temperature, $n_{\text{MM,GUT}}/T_{\text{GUT}}^3 \sim 4 \times 10^{-7}(T_{\text{GUT}}/10^{16}\text{ GeV})^3$, at the time of the phase transition.

3.3.1 The monopole problem

A monopole and an *anti-monopole*[4] can *annihilate*, a process which preserves the total topological charge, releasing their total mass as energy (in the form of parti-

[3] The horizon is approximately given by $\ell \simeq C\, m_{\text{P}}\, T_{\text{GUT}}^{-2}$ where $m_{\text{P}} \simeq 10^{19}$GeV is the Planck mass and $C = 0.6\, g_*^{-1/2} \sim 1/20$ with g_* the effective number of spin degrees of freedom [6]. It is roughly equal to the Hubble radius, $1/H_{\text{GUT}}$, and to one over the age of the universe at the time of the GUT phase transition, $1/t_{\text{GUT}}$.

[4] An anti-monopole is a monopole with an opposite charge. A monopole and an anti-monopole can annihilate because their combination is actually a topologically trivial solution, but they need to be brought together first.

cles). Because of their topological nature, monopoles cannot decay into other particles and pair annihilation is therefore the *only* possible way in which the number of monopoles in a comoving volume can be reduced. The attractive force between monopoles of opposite charge should of the order of the Coulomb force, so the rate at which monopoles and anti-monopoles capture each other and subsequently annihilate can be estimated [16, 17]. Pair annihilation is a relatively slow and inefficient process due to the high monopole mass and it turns out to be unable to keep up with the expansion of the universe when the ratio n_{MM}/T^3 falls below $10^{-8}(m_{\mathrm{MM}}/10^{17}\ \mathrm{GeV})$, where T is the temperature and m_{MM} is the monopole mass [6].

Since monopoles cannot decay, their number density must scale as a^{-3} once monopole annihilation has ceased to be effective, where a is the scale factor. Under the assumption that the universe has cooled down adiabatically afterwards, the total *entropy* is preserved and the entropy density s therefore also scales as a^{-3}, which tells us that the ratio n_{MM}/s is preserved [12]. At a temperature T the entropy density is given by $s = g_* \frac{2\pi^2}{45} T^3$, with g_* is the effective number of spin degrees of freedom [18, 19].

If the initial ratio, $n_{\mathrm{MM,GUT}}/T^3_{\mathrm{GUT}} \sim 4 \times 10^{-7}(T_{\mathrm{GUT}}/10^{16}\ \mathrm{GeV})^3$, is smaller than the threshold $10^{-8}(m_{\mathrm{MM}}/10^{17}\ \mathrm{GeV})$, monopole annihilation plays no role and the ratio between the monopole density and the entropy at a later time will be the same as at the GUT scale. If we now take $g_* \sim 100$ then

$$n_{\mathrm{MM}}/s \sim 4 \times 10^{-9}(T_{\mathrm{GUT}}/10^{16}\ \mathrm{GeV})^3, \tag{3.8}$$

which corresponds to $n_{\mathrm{MM}} \simeq 2 \times 10^{-7}(T_{\mathrm{GUT}}/10^{16}\ \mathrm{GeV})^3\ \mathrm{cm}^{-3}$ today ($T = 2.725$ K and $g_* \sim 3.91$ [19]).

If on the other hand the initial ratio is greater, then pair annihilation reduces the monopole density to this threshold, after which the number of monopoles per comoving volume freezes [17]. If initially $g_* \sim 100$, then at a later time this gives a monopole density of

$$n_{\mathrm{MM}}/s \sim 10^{-10}(m_{\mathrm{MM}}/10^{17}\ \mathrm{GeV}), \tag{3.9}$$

which gives $n_{\mathrm{MM}} \simeq 4 \times 10^{-9}(m_{\mathrm{MM}}/10^{17}\ \mathrm{GeV})\ \mathrm{cm}^{-3}$ today. For $T_{\mathrm{GUT}} \sim 10^{16}$ GeV and $m_{\mathrm{MM}} \sim 10^{17}$ GeV, the latter scenario applies and the total monopole energy density today would thus be $\rho_{\mathrm{MM}} \sim 4 \times 10^8\ \mathrm{GeV/cm}^3$.

The total energy density of baryonic matter is around $2 \times 10^{-7}\ \mathrm{GeV/cm}^3$, which corresponds to about 4% of the total energy density of our Universe [18, 19]. This would mean that the energy density of monopoles would exceed that of baryonic matter by around 15 orders of magnitude and would therefore completely dominate the energy density of the universe today. If the abundance of monopoles is this great, it should be possible to observe them, so the fact that no evidence of the existence monopoles has been found conflicts with this prediction (see chapter 4 for more on detection methods and upper bounds on the monopole density). This defines the *monopole problem*.

The easiest way to solve the monopole problem would of course be give up the idea of grand unification or to loosen the demands for the grand unified gauge group. If the electromagnetic gauge group $\mathrm{U}(1)_{\mathrm{em}}$ is not embedded in a semisimple Lie group, but for instance in a group of the form $H \times \mathrm{U}(1)$, then there is no need for monopoles to appear [12, 17]. Allowing for this would however more or less undermine the whole idea behind Grand Unified Theories. Alternatively, it may be possible that reality is described by a Grand Unified Theory, but that the unification temperature was never reached in the early universe [8].

An interesting possibility was suggested by Langacker and Pi [14] in 1980, who realised that when the GUT symmetry group is broken down to the standard model gauge group in several stages the monopole density could be reduced through the formation of *cosmic strings* (for more on those, see [8,12,20] or the paper by Maarten Verdult). The GUT symmetry could for instance be broken in the pattern

$$G_{\text{GUT}} \to \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \to \mathrm{SU}(3) \to \mathrm{SU}(3) \times \mathrm{U}(1)_{\text{em}}. \tag{3.10}$$

After the second phase transition, the monopoles and anti-monopoles that were created during the first phase transition would get connected by strings. This causes the monopoles and anti-monopoles to be pulled together by the string's tension, making the pair-annihilation process a lot more efficient and reducing the monopole density to about one monopole per horizon. During the third phase transition the electromagnetic $\mathrm{U}(1)$ symmetry group is restored and any remaining strings disappear.

By far the most satisfying solution to the monopole problem is inflation, which was in fact in part motivated by the monopole problem (along with the horizon and flatness problem) [8, 12, 21]. If monopoles are produced before or in the early stages of inflation they are diluted along with all other matter up to the point where they become entirely unobservable and there might not even be a single monopole in the currently observable universe. If the temperature after inflation due to preheating is high enough, monopoles could still be produced through thermal fluctuations. This would typically result in a very low monopole density, but according to some models it could be high enough to make them potentially detectable [12, 22].

3.4 An example of a 't Hooft-Polyakov monopole

In the previous section we have argued why any Grand Unified Theory that embeds a gauge group containing $\mathrm{U}(1)_{\text{em}}$ in a semisimple compact gauge group allows for the existence of magnetic monopoles. This was actually first realised in 1974 by 't Hooft [23] and Polyakov [24], which is why magnetic monopoles that occur through this argument are often called *'t Hooft-Polyakov monopoles*. These monopoles are fundamentally different from the Dirac monopole discussed before because they are non-singular and all fields involved are globally defined. Instead of a singularity they have a core in which the broken symmetry is restored, but outside this core the 't Hooft-Polyakov and the Dirac monopole turn out to look very similar. In this section we will discuss an example of a 't Hooft-Polyakov monopole and we will explicitly see why it is magnetic.

At the time when 't Hooft and Polyakov published their findings, it was not yet established that the $\mathrm{SU}(2) \times \mathrm{U}(1)$ mechanism for the electroweak symmetry breaking, which is incorporated in the standard model, was the right way to go. One appealing competitor was was the Georgi-Glashow $\mathrm{SU}(2)$ model. This model [13] embeds the electromagnetic gauge group $\mathrm{U}(1)_{\text{em}}$ in the compact, simple Lie group $\mathrm{SU}(2)$ (instead of $\mathrm{SU}(2)_{I_{\text{w}}} \times \mathrm{U}(1)_{\text{em}}$). It has a three-dimensional Higgs field that transforms via the adjoint representation of $\mathrm{SU}(2)$ (instead of a two dimensional complex Higgs field) and after symmetry breaking only a photon field and two massive, electrically charged, gauge bosons appear (contradicting the later observation of weak neutral current interactions). Since $\mathrm{SU}(2)$ is a simple compact Lie group, it has a trivial first and second homotopy group and magnetic monopoles are therefore expected to appear if this original $\mathrm{SU}(2)$ symmetry is broken to $\mathrm{U}(1)$.

The Lie algebra corresponding to the Lie group $\mathrm{SU}(2)$ is $\mathfrak{su}(2)$. This Lie algebra is generated by the complex 2×2-matrices $t^a = \tau^a/2$ for $a = 1, 2, 3$, where τ^1, τ^2, τ^3 are the three Pauli matrices, so we can write any element $g \in \mathrm{SU}(2)$ as $g = \exp(ig^a t^a)$ for some coefficients g^a. This choice of generators means that the commutator of two of these generators is given by $[t^a, t^b] = i\,\varepsilon_{abc} t^c$. We furthermore have $2\ \mathrm{Tr}(t^a t^b) = 2\ \mathrm{Tr}(\tau^a \tau^b) = \delta_{ab}$, so twice the trace defines an inner product on $\mathfrak{su}(2)$ with respect to which the generators t^i are orthonormal. It is furthermore useful to note that each of the generators t^a separately generate a subgroup of $\mathrm{SU}(2)$ that is isomorphic to $\mathrm{U}(1)$.

Apart from this group, the model has the following ingredients:

- A 3 dimensional scalar Higgs field $\Phi = \Phi^a t^a$ that takes values in the Lie algebra $\mathfrak{su}(2)$. The Higgs field Φ transforms via the adjoint representation of $\mathrm{SU}(2)$, so a group element $g(x^\mu) \in \mathrm{SU}(2)$ sends $\Phi(x^\mu)$ to $g(x^\mu)\cdot\Phi(x^\mu)\cdot g(x^\mu)^{-1}$, where the multiplication is just matrix multiplication. For an infinitesimal transformation by $i\xi(x^\mu) = i\xi^a(x^\mu)t^a$ this reads $\Phi(x^\mu) \mapsto \Phi(x^\mu) + [i\xi(x^\mu), \Phi(x^\mu)]$.
  We will write $|\Phi|^2 \equiv 2\,\mathrm{Tr}(\Phi^2) = (\Phi^1)^2 + (\Phi^2)^2 + (\Phi^3)^2$, which can easily be verified to be gauge invariant.
- A gauge field $A_\mu = A^a_\mu t^a$ that also takes values in $\mathfrak{su}(2)$ and couples to the Higgs field via the covariant derivative $D_\mu\Phi = \partial_\mu\Phi + i\,e\,[A, \Phi]$. This field transforms under gauge transformations in such a way that $D_\nu\Phi$ transforms the same way as Φ does. This tells us that $|D_\mu\Phi|^2 \equiv 2\,\mathrm{Tr}(D_\mu\Phi D^\mu\Phi) = D_\mu\Phi^a D^\mu\Phi^a$ is also gauge invariant.
- This defines the field-strength tensor $F_{\mu\nu} = [D_\mu, D_\nu]/(i\,e) = \partial_\mu A_\nu - \partial_\nu A_\mu + i\,e\,[A_\mu, A_\nu]$, which is an element of $\mathfrak{su}(2)$ and can be written as $F_{\mu\nu} = F^a_{\mu\nu} t^a$ as with A_μ. Due to the orthonormality of t^a with respect to twice the trace, we have $2\ \mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) = F^a_{\mu\nu}F^{a\mu\nu}$, which turns out to be a gaugeinvariant expression.
- A (Mexican hat) potential $V(\Phi)$ given by $V(\Phi) = \frac{1}{4}\lambda(\eta^2 - |\Phi|^2)$, that takes the minimal value zero on the vacuum manifold $\mathcal{M} = S^2 \subseteq \mathfrak{su}(2)$.

In addition to these fields, we would also need some number of fermionic fields to describe our world, but we will ignore these for the moment as they are not important for the symmetry breaking. With the fields that we have, we can write down the following gauge invariant Lagrangian density, which determines the dynamics of the Georgi-Glashow $\mathrm{SU}(2)$ model

$$\mathcal{L} = -\,\mathrm{Tr}(D_\mu\Phi^a D^\mu\Phi^a) - \tfrac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} - \tfrac{1}{4}\lambda(\eta^2 - |\Phi|^2)^2, \tag{3.11}$$

where $|\Phi|^2 := 2\,\mathrm{Tr}(\Phi^2) = (\Phi^1)^2 + (\Phi^2)^2 + (\Phi^3)^2$. This gives us the following field equations [9]

$$D_\mu D^\mu\Phi = -\lambda\,(\eta^2 - |\Phi|^2)\,\Phi \qquad \text{and} \qquad D_\mu F^{\mu\nu} = -i\,e\,[D^\nu\Phi, \Phi]. \tag{3.12}$$

By expanding the Higgs field $\Phi = (\eta + \phi)\,t^3$ around the vacuum solution $\Phi = \eta\,t^3$ and writing $A_\mu = W^1_\mu t^1 + W^2_\mu t^2 + a_\mu t^3$, it is possible to read off the masses of all the particles that appear. Since this is not what we are interested in, I will not do this here, but just state that the Higgs particle attains a mass $\sqrt{2\,\lambda}\,\eta$, that there are two W bosons particles (there are no Z-bosons in this theory) with mass $2\,\eta$ and a massless photon [9, 13].

Since the original symmetry was described by $\mathrm{SU}(2)$, which is a (semi)simple Lie group, and the remaining symmetry after symmetry breaking is $\mathrm{U}(1)$, we know from sections 1.1 and 3.2 that $\pi_2(\mathcal{M}) = \pi_1(\mathrm{U}(1)) = \mathbb{Z}$. This is indeed the case as $\mathcal{M}$ is a 2-sphere in $\mathfrak{su}(2)$ and $\pi_2(S^2) = \mathbb{Z}$, so monopole solutions should indeed exist.

3.4.1 The static monopole solution

A static solution representing a pure magnetic monopole was found by 't Hooft [23] and Polyakov [24]. The solution is described by fields Φ and A_μ (which are time-independent) of the form

$$\Phi = \eta\, h(r)\, \frac{x^a}{r}\, t^a, \qquad A_0 = 0 \qquad \text{and} \qquad A_i = \tfrac{-1}{e}(1 - k(r))\, \varepsilon_{ija} \frac{x^j}{r^2} t^a, \tag{3.13}$$

where h and k are functions of the distance $r = \sqrt{X^i x^i}$ to the origin [6, 9]. These fields are rotationally symmetric in the sense that a rotation has the same effect as a global (i.e. coordinate independent) gauge transformation [9].

We of course need $h(0) = 0$ and $k(0) = 1$ to avoid getting a singularity at the origin. To make sure that the total energy is finite (see section 2.2) we furthermore demand that $\lim_{r\to\infty} h(r) = 1$ and $\lim_{r\to\infty} k(r) = 0$. An important question is of course whether such a solution defines a monopole and this turns out to be the case. To see this we note that Φ points outwards radially and that $|\Phi| \to \eta$ as $r \to \infty$, so the asymptotic function $\Phi_\infty\colon S^2 \to \mathcal{M}$ (as defined in section 2.2) sends a point $\hat{\mathbf{x}}$ on the unit sphere to $\eta\,\hat{\mathbf{x}} \in \mathcal{M} = \eta\, S^2$. Since the 2-sphere is not contractible, it is easily seen that there exists no continuous deformation of Φ_∞ that maps S^2 into a single point, so Φ defines a monopole solution.

Unfortunately the field equations cannot be solved analytically, even for solutions of this form, so numerical methods need to be used to find the functions h and k. In terms of the dimensionless parameter $s = \frac{1}{2}\eta\, e\, r$, the field equations (3.12) for solutions of this form reduce to (N.B. in our units $e^2 = 4\,\pi\,\alpha$, where $\alpha \simeq \frac{1}{137}$)

$$\frac{\mathrm{d}^2 h}{\mathrm{d}s^2} + \frac{2}{s}\,\frac{\mathrm{d}h}{\mathrm{d}s} = \frac{2}{s^2}\, k^2 h - 4\, e^{-2}\lambda\,(1 - h^2)\, h, \tag{3.14}$$

$$\frac{\mathrm{d}^2 k}{\mathrm{d}s^2} = \frac{1}{s^2}(k^2 - 1)\, k + h^2 k. \tag{3.15}$$

A few solutions to these equations for different values of $e^{-2}\lambda$ have been plotted in figure 3.2. What is most important to note about these solutions is that both h and k converge exponentially to their asymptotic values (1 and 0 respectively) as $r \to \infty$. For $r \gg (\eta\, e)^{-1}$ the fields Φ and A_i therefore look approximately like

$$\Phi \simeq \eta\, \frac{x^a}{r}\, t^a, \qquad A_0 = 0 \qquad \text{and} \qquad A_i \simeq -\varepsilon_{ijk} \frac{x^j}{e\, r^2} t^a, \tag{3.16}$$

but for $r \lesssim (\eta\, e)^{-1}$ their behaviour is more complicated. The region where $r \lesssim (\eta\, e)^{-1}$ is called the *core* of the monopole.

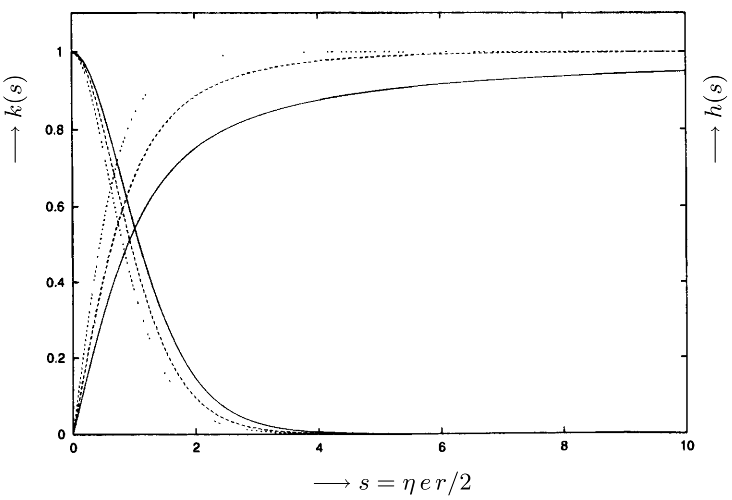


Figure 3.2: Solutions $h(s)$ (going up) and $k(s)$ (going down) to the field equations (3.14) for $\lambda \to 0$ (solid), $\lambda = e^2/40$ (dashed) and $\lambda = e^2/4$ (dotted). This figure originally appeared in [9].

3.4.2 Properties of the solution

For these solutions the mass (total energy with respect to the groundstate) of the monopole solution can be calculated by integration of the energy density obtained from the Lagrangian. The resulting mass for several values of $e^{-2}\lambda$ has been plotted in figure 3.3. We see that for very large values of λ the energy asymptotically goes to about $1.787 \times 8\,\pi\,\eta/e^2$ and that it goes to $8\,\pi\,\eta/e^2$ for very small values of λ [9, 25]. These two extremes are not very far apart, so we can safely say that it is of order $8\,\pi\,\eta/e^2$ for any value of λ.

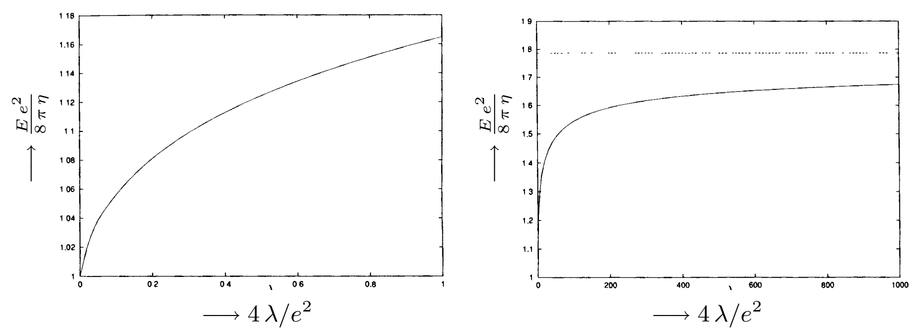


Figure 3.3: The dependence of the mass of the monopole solution on λ. The parameter $4\,\lambda/e^2$ has been plotted on the horizontal axis and on the vertical axis we see $E\,e^2/(8\,\pi\,\eta)$. This figure originally appeared in [9].

Although it does seem likely, it has not been proven that the energy of the solution described above is a minimum for all solutions in its class except in the limit for $\lambda \to 0$ [9]. It has however been shown numerically that small deformations of this solution increase its energy [26], so in that sense we are certain that we are dealing with a stable solution.

It is not immediately obvious why this solution should describe a *magnetic* monopole, because for that we first need to define exactly what we even *mean* by electromagnetism in this situation. Luckily, there is a relatively easy way to do this. At any point in space-time, there is a preferential direction for the fields taking values in the Lie algebra $\mathfrak{su}(2)$ which is set by the direction in which Φ points, so we can define the new field-strength tensor $f_{\mu\nu}$ to be the part of $F_{\mu\nu}$ pointing in this direction, so

$$f_{\mu\nu} \equiv \frac{2}{|\Phi|} \operatorname{Tr}(F_{\mu\nu}\Phi) = F^a_{\mu\nu} \frac{\Phi^a}{|\Phi|}. \tag{3.17}$$

Under a gauge transformation defined through $g(x^\mu)$, Φ transforms to $g\Phi g^{-1}$ and $F_{\mu\nu}$ transforms to $gF_{\mu\nu}g^{-1}$, so both $\operatorname{Tr}(F_{\mu\nu}\Phi)$ and $|\Phi| = \sqrt{2\operatorname{Tr}(\Phi^2)}$ are invariant and $f_{\mu\nu}$ is therefore gauge-invariant as well.

It is not possible to globally impose the *unitary gauge*[5] because the configuration of Φ is non-trivial. We can however *locally* choose the unitary gauge on any contractible region that does not contain the origin, so we can do it simultaneously do this everywhere except in the origin and a line (string) from the origin to infinity. If we write $\Phi = (\eta + \phi)t^3$ (possible by definition of the unitary gauge) and $A_\mu = W^1_\mu t^1 + W^2_\mu t^2 + a_\nu t^3$, then we see that outside the core (for $r \gg (\eta\, e)$)

$$F_{\mu\nu} \simeq f_{\mu\nu} t^3, \qquad A_{\mu\nu} \simeq a_\mu t^3 \qquad \text{and} \qquad f_{\mu\nu} \simeq \partial_\mu a_\nu - \partial_\nu a_\mu \tag{3.18}$$

and that the field equations for $F_{\mu\nu}$ furthermore reduce to $\partial_\mu f^{\mu\nu} \simeq 0$. Outside the core we thus get the usual field strength tensor in terms of the photon field a_μ, satisfying the usual field equations.

We can define the magnetic field B_i as we would usually do, but now in terms of this new field-strength tensor, to be given by $B_i = -\frac{1}{2}\varepsilon_{ijk} f_{jk}$. If we do this for the monopole solution we see that outside the core of the monopole (so for $r \gg (\eta\, e)^{-1}$) the magnetic field is approximately given by

$$B_i = -\tfrac{1}{2}\varepsilon_{ijk} f_{jk} \simeq -\frac{1}{e}\frac{x^i}{r^3}, \qquad \text{so} \qquad \mathbf{B} \simeq -\frac{\hat{\mathbf{r}}}{e\, r^2}, \tag{3.19}$$

which we recognise as the magnetic field of a Dirac monopole with a magnetic charge of $-4\,\pi/e$, which is twice the Dirac charge (up to a sign).

A monopole with the opposite charge can be obtained by reflecting this solution through the origin ($x^i \mapsto -x^i$) or by changing the sign in front of the Higgs field ($\Phi \mapsto -\Phi$). This solution also has the opposite topological charge (it corresponds to the inverse element of $\pi_2(\mathcal{M})$) and thus describes the so-called anti-monopole. In addition to this stable (stable against small perturbations) monopole solutions have been found that carry an electric charge in addition to a magnetic charge. These states are generally periodic instead of static and they have a higher energy, so in that sense they are excitations of the ordinary monopole state. Monopoles with an additional electric charge are called *Dyons* [9].

[5] The unitary gauge is the gauge in which Φ has been forced into the form $\Phi = (\eta + \phi)t^3$, which is what we did before when we expanded around the vacuum.

Since f_{ij} is defined independently from any gauge, the magnetic field is also gauge-independent and it is well-defined everywhere (except for the origin). The photon field a_μ on the other hand is gauge dependent and cannot be simultaneously defined on the entire space, exactly like in the case of the Dirac monopole. We now see that at long range this solution exactly describes a Dirac monopole, but with the singularity now replaced by a non-singular core.

The field-strength tensor $f_{\mu\nu}$ determines the magnetic field, but the choice we made for it was not entirely unique. Instead of the choice we made in equation (3.17), 't Hooft [23] suggested the following definition for the electromagnetic gauge field

$$f_{\mu\nu} = \frac{1}{|\Phi|}\Phi^1 F^a_{\mu\nu} + \frac{1}{e\,|\Phi|^3}\varepsilon_{abc}\Phi^a(D_\mu\Phi^b)(D_\nu\Phi^c). \tag{3.20}$$

If we now choose the unitary gauge (locally), write $\Phi = (\eta + \phi)t^3$ and $A_\mu = W^1_\mu t^1 + W^2_\mu t^2 + a_\mu t^3$, then the equations

$$f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu \qquad \text{and} \qquad \partial_\mu f^{\mu\nu} = 0 \tag{3.21}$$

become exact and hold everywhere except in the origin [23], making this an attractive choice. Inside the core the magnetic field would look completely different, but outside the core, the magnetic field for the Dirac monopole is again obtained. We see that different choices are possible for the electromagnetic field strength tensor that give rise to different electromagnetic fields in the core, but outside the core the Dirac monopole is always retrieved [7, 12]. The total magnetic charge of the monopole is therefore well-defined and we know that this charge is necessarily located somewhere in the core of the monopole, but exactly how it is distributed depends on the choice made for the field strength $f_{\mu\nu}$.

4. Observational bounds

Many people feel that magnetic monopoles should exist and an enormous amount of theoretical work has been done to derive their behaviour and their properties, but the fact of the matter is that not a single shred of evidence of their existence has been to confirm or hint at their existence. Even long before it was realised that GUT theories predicted the existence of monopoles, people have been searching for them in earth-bound materials, moon rocks and meteorites, as well as among the by-products of high-energy collisions in particle accelerators, with only negative results [27, 28].

Upper bounds for the possible monopole abundance can however be obtained by cosmological and astrophysical considerations. Furthermore, since many people have searched for monopoles in *cosmic radiation* in earth-bound experiments and none have yet been found, further bounds can be put on the scarcity of monopole, which will be the subject of this chapter.

4.1 The overclosure bound

A simple bound on the magnetic monopole density follows from the fact that we require the current magnetic monopole mass density to be smaller than the (observed) *Critical density* of the universe [12, 29, 30]. This simply follows from the Friedmann equation [18, 19]

$$H_0^2 = \tfrac{8}{3}\,\pi\, G_{\mathrm{N}}\rho - \frac{k}{a^2}, \tag{4.1}$$

which holds if we assume energy to be distributed roughly uniformly throughout the Universe. Here $H_0 \sim 73$ km/s/Mpc is the value of the Hubble parameter today [18] and G_{N} is Newton's constant. The parameters ρ and k are the total energy density of the Universe and the observed curvature of the universe. This can be rewritten as

$$\rho = \rho_{\mathrm{cr}} + k/H_0^2, \qquad \text{with} \qquad \rho_{\mathrm{cr}} = \frac{3\,H_0^2}{8\,\pi\,G_{\mathrm{N}}} \simeq 5.6 \times 10^{-6}\ \mathrm{GeV/cm^3}, \tag{4.2}$$

where ρ_{cr} is the aforementioned critical density [12, 30, 19]. Since the cosmological constant is expected to give a positive contribution [31] and the universe is observed to be approximately flat [18, 21] this gives us an upper bound for the contribution of magnetic monopoles to the total energy density of the Universe. This is called the overclosure bound a larger energy density would necessarily imply that the curvature parameter k is large, which would result in a ("very") closed universe [18, 19].

Because magnetic monopoles are expected to be extremely heavy (masses of the order of 10^{17} GeV) they will move very slowly, with velocities of the order of $10^{-20}\,c$ [32] if left alone. Nearby monopoles will however have been accelerated to velocities of around $10^{-3}c$ by the gravitational pull of the Galaxy (this is the escape velocity of our Galaxy) [12, 33, 34] and relatively light monopoles can be accelerated further by the Galaxy's magnetic field (see section 4.2 for details). This means that magnetic monopoles are generally non-relativistic, so we obtain an upper bound for their number density, which is [12, 30]

$$n_{\mathrm{MM}} \lesssim 10^{-22}(m_{\mathrm{MM}}/10^{17}\ \mathrm{GeV})^{-1}\ \mathrm{cm}^{-3}, \tag{4.3}$$

where m_{MM} is the monopole mass (in GeV). This density is 13 orders of magnitude below the number density calculated in section 3.3.1, which further explains why that prediction was unacceptable.

We would like to give this bound as a bound on the flux of monopoles instead of a density because the flux is something we can actually hope to measure. The flux of monopoles is given by $F \simeq \beta\, n_{\mathrm{MM}}$, where β is the monopole speed as a fraction of the speed of light. With our previous estimate that $\beta \sim 10^{-3}c$ we obtain the bound [28, 12]

$$F < F_{\text{uniform}} \sim 10^{-12}(m/10^{17}\ \mathrm{GeV})\ \mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}. \tag{4.4}$$

This upper bound was obtained by assuming a uniform distribution of monopoles throughout the Universe, which is why it is labelled the uniform bound in figure 4.3 on page 34.

4.2 The Parker bound

Another bound for the magnetic monopole flux that is widely accepted is the so-called *Parker bound* [12, 33, 34, 35]. Our Galaxy has a magnetic field of about about $\sim 3\ \mu\mathrm{G} = 1.2 \times 10^{-9}$ T due to a *Dynamo effect* by which the rotation of the galaxy induces tiny currents. Unlike the Earth's magnetic field, this field is distributed more or less randomly and varies over distance of the order of 10^{21} cm ~ 300 pc [12, 29]. The dynamo effect generating these fields is not completely understood, but it is believed to be able to renew these currents over a typical timescale of about 10^{8} yrs, which is about the galactic rotation period.

A magnetic monopole with negligible velocity travelling through a domain in which the magnetic fields points in some direction (over a distance of ~ 300 pc) is accelerated to a velocity of roughly $v_{\text{mag}} \simeq 10^{-3}(m/10^{17}\ \mathrm{GeV})^{-1/2}$ [32, 12], which corresponds to an energy gain of $\Delta E \simeq 10^{11}$GeV [32]. For monopoles with a mass less than 10^{17} GeV this velocity is larger than their initial velocity, so the monopoles are more or less swept along by the magnetic fields, either gaining or losing this energy when they pass through a domain. Since their kinetic energy cannot become negative this looks like a random walk and after passing through N domains the kinetic energy will be roughly $\sqrt{N}\,\Delta E$ [32].

A more detailed analysis [33, 32] for monopoles passing through the Galaxy gives us an estimate for the rate at which the Galactic magnetic field is dissipated. Since this should be smaller than the renewal rate of 10^8 yrs the following upper bound for the monopole flux is obtained [32, 33, 12]

$$F < F_{\text{Parker},<\,10^{17}\text{ GeV}} \sim 10^{-15}\ \text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}. \tag{4.5}$$

For heavy monopoles with a mass greater than 10^{17} GeV the analysis becomes completely different since the monopole velocity will then not be significantly influenced by the magnetic fields and they will mainly get deflected. The monopoles will still gain energy overall while travelling through the Galaxy because they are accelerated in directions transverse to their original velocity while travelling through a domain [12, 32]. This gives the higher, mass dependent limit for the monopole flux [12, 32, 34]

$$F < F_{\text{Parker},>\,10^{17}\text{ GeV}} \sim 10^{-15}(m/10^{17}\ \text{GeV})\ \text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}. \tag{4.6}$$

This bound was later improved and the new bound by taking many new factors into account [35, 36]. This new bound is called the *Extended Parker bound* and is given by

$$F < F_{\text{Parker, Extended}} \sim 10^{-16}(m/10^{17}\ \text{GeV})\ \text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}. \tag{4.7}$$

Both the normal Parker bound and this extended Parker bound are shown in figure 4.3 on page 34 from which we see that the Parker bounds are an improvement on the overclosure bound only for masses below 10^{17} GeV.

4.3 Direct monopole detection

Detection methods for magnetic monopoles can be classified into three categories: *Induction*, *Energy loss* and *Catalysis techniques*. In the remainder of this section I will describe each these techniques in some detail and discuss their (negative) results.

4.3.1 Induction techniques

Whenever a magnetic monopole passes through a closed loop of superconducting wire, something special happens. To see exactly what happens we can use Maxwell's equations, extended to include the magnetic charge and current. The relevant extended Maxwell equation reads

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} - \mathbf{j}_\text{m}, \quad \text{or equivalently} \quad \int_{\partial S} \mathbf{E} \cdot \text{d}\mathbf{r} = -\frac{\partial \Phi_\text{m}}{\partial t} - I_\text{m}, \tag{4.8}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic field respectively and $\mathbf{j}_\text{m}$ is the magnetic current density. In the integrated version of the equation the integral is over the boundary ∂S of an arbitrary some surface S and Φ_m and I_m are the magnetic flux and total magnetic current through this surface.

Inside the superconducting wire $\mathbf{E} = 0$, so the left-hand side of the integrated Maxwell equation vanishes completely if we choose the surface S in such a way that ∂S is entirely contained in the wire. The resulting equation, $\frac{\partial}{\partial t}\Phi_\text{m} = -I_\text{m}$, tells us that when a magnetic monopole with magnetic charge g passes through the loop, the total magnetic flux through the loop also changes, by $\Delta\Phi = -g$. For a monopole with the Dirac charge $g = g_{\rm D} = 2\,\pi/e$ the total flux change is $\Delta\Phi = -2\,\Phi_{\rm m,o}$, where $\Phi_{m,o} = \pi/e$ is the magnetic flux quantum.

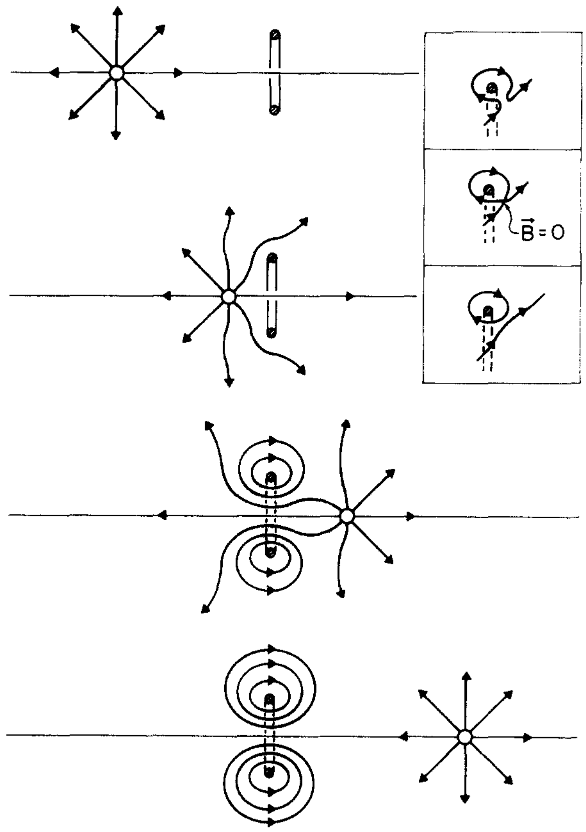


Figure 4.1: Bending and breaking of magnetic lines as a monopole passes through a loop of superconducting wire. This figure originally appeared in [32].

What happens has been depicted in figure 4.1: Because the magnetic field lines of a monopole cannot pass through the superconducting wire, they will bend around the wire loop as the monopole approaches it. Once the monopole has gone through, its total magnetic flux will therefore pass back through the loop, causing the aforementioned magnetic flux. The field lines will break where the magnetic field vanishes and form loops around the wire as the monopole moves away from it [32, 6].

An electric supercurrent, $I_{\rm ind} = g/L$,where L is the inductance of the loop, is in fact induced in this process through the Meissner-effect [37]. If the superconducting loop is shielded from external magnetic fields, then the jump in the magnetic flux through the loop (and the supercurrent) whenever a monopole passes through it should be measurable using a SQUID (superconducting quantum interference device) magnetometer.

Experiments with pseudopoles, which are the poles of very long magnetic dipoles (such as electromagnets), show that it should be possible to detect the current induced by a monopole with the Dirac charge with such a set-up. In 1982 a candidate events of a magnetic monopole passing through such a superconducting loop was discovered [38], but this result is controversial as no one has been able to reproduce it [37].

An advantage of this approach is that the measurement of a magnetic monopole (or dyon) depends only on its magnetic charge and not on its mass, velocity, its possible electric charge or any other properties. Unfortunately however, building large arrays of such detectors and shielding them properly would be very difficult and expensive and the combination of all experiments based on induction techniques have only been able to provide us with an upper bound of about $2 \times 10^{-14}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ for the total monopole flux [29][1] at the 90% confidence level [28, 27], which is well above (i.e. weaker than) the original Parker bound of $\sim 10^{-15}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ and the overclosure bound for any monopole mass.

4.3.2 Energy loss techniques

Since magnetic monopoles carry a magnetic charge, it makes sense to expect them to interact with matter through electromagnetism and slow down. Monopoles travelling through matter are indeed expected to lose energy through interactions and this can happen through a variety of mechanisms. Which mechanism is most dominant depends on the nature of the matter they travel through and is also strongly dependent on the monopole velocity [32, 39]. The exact mechanisms for this energy loss is quite complicated, but has been thoroughly reviewed in [32, 29]. Although the energy loss of GUT monopoles is very significant (of the order of $\sim 100\ \mathrm{MeV/cm}$), the kinetic energy of a magnetic monopole is large enough (even at non-relativistic velocities) to nevertheless make them very penetrating. Larger and more dense astrophysical objects on the other hand can capture a significant number of magnetic monopoles [29, 32, 36].

A number of results that have been obtained through energy loss techniques have been shown in figure 4.2. Most notably among these are the results from the MACRO detector, which is a large multipurpose detector that was mainly designed to detect (the absence of) magnetic monopoles travelling through the earth both through the energy loss mechanisms described here and through catalysed nucleon decay, which is described in section 4.3.3 (for more details see [32, 40, 41]). None of these experiments have resulted in the detection of even a single monopole, which tells us that the flux of magnetic monopoles with a velocity above $4 \times 10^{-5} c$ is bounded by [28, 41]

$$F < 1.4 \times 10^{-16}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}, \tag{4.9}$$

which is well below the original Parker bound, as can be seen in figure 4.2.

4.3.3 Catalysis techniques

Inside the core a Grand Unified monopole, the original GUT symmetry is restored, which means that in the vicinity of a monopole processes can occur that are otherwise impossible in the standard model. We could say that the monopole catalyses these

[1] sr stands for steradian and is a unit of solid angle. 4π sr corresponds to an entire sphere.

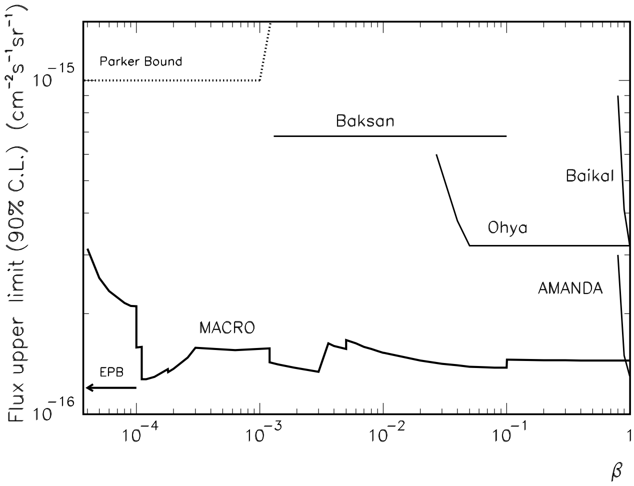


Figure 4.2: The result of the MACRO energy loss experiments and several others. The Parker bound has been indicated, as well as the extended Parker bound (EPB) for monopoles with a mass of 10^{17} GeV. Bounds on the monopole flux as a function of the monopole mass for white dwarfs and neutron stars. The curved lines are the corrected results for $g = g_{\rm D}(solid)$ and $g = 2\,g_{\rm D}$ (dotted). The Parker bound and the extended Parker bound have been shown, as well as the (uniform) overclosure bound. This figure originally appeared in [36].

processes. A process of particular interest is the *Catalysed decay of nucleons* via the *Rubakov-Callan mechanism*, where a nucleon decays into a positively charged lepton and a meson. The most relevant contributions can be written schematically as [30,40]

$$\begin{aligned} &M + p \to M + e^+\pi^0, & & M + p \to M + \mu^+ K^0, \\ &M + n \to M + e^+\pi^-, &\text{and}\quad & M + n \to M + \mu^+ K^-. \end{aligned}$$

The cross section for such interactions was initially believed to be of approximately the order of 10^{-56} cm^2, but Rubakov [42] and Callan [43] have shown that the cross section can in fact be much greater, namely comparable to the size of the nucleons.

The catalysis cross section is believed to roughly be of the form $\sigma = \sigma_0/\beta$, with σ_0 some parameter of the order $\sim 10^{-28}$ cm^2 and β the monopole velocity as a fraction of the speed of light [30, 40, 42, 43].The exact cross section is not known, only its order of magnitude, but the possibility that it is much smaller than this is not excluded. Under the assumption that these cross sections are indeed “large” (of the order $\sim 10^{-27}$ cm^2 or larger), it becomes possible to detect a monopole travelling through some materials. A monopole would cause one nucleon to decay for every few meters or even every few centimeters it travels through ordinary matter depending on the monopole velocity, the exact cross section and the material used.

Since a lot of energy is released in these processes (for decay into positrons around the rest mass of the nucleon, so ~ 1 GeV) a magnetic monopole travelling through matter should be observable through the nucelon decay it induces under the assumption that the aforementioned cross section is indeed large. There have therefore been

Experiment	Technique	Flux limit	β range
Soudan1	Proportional tubes	$F < 8.8 \times 10^{-14}$	$10^{-2} \lesssim \beta \lesssim 1$
IMB	Water Cherenkov	$F < (1-3) \times 10^{-15}$	$10^{-5} \lesssim \beta \lesssim 10^{-1}$
Kamiokande	Water Cherenkov	$F < 2.5 \times 10^{-15}$	$5 \times 10^{-5} \lesssim \beta \lesssim 10^{-3}$
Baikal	Underwater detector	$F < 6 \times 10^{-17}$	$v \simeq 10^{-5}$
MACRO	Streamer tubes	$F < 3 \times 10^{-16}$	$1.1 \times 10^{-4} \lesssim \beta \lesssim 5 \times 10^{-5}$

Table 4.1: Flux limits obtained by studying catalysed nucleon decay. The flux limits are in $\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$. This table originally appeared in [40]

numerous attempts to detect magnetic monopoles by looking at catalysed nucleon decay with detectors that specialise on different velocity ranges, but no catalysed nucleon decay has yet been observed [28]. Table 4.1 shows a number of flux limits that have been obtained in this way [40], although it should again be noted that they are only valid under the assumption that the aforementioned cross section is not negligible. All these bounds are roughly of the same order as the bound obtained from energy loss experiments and the extended Parker bound.

4.4 Observations from astrophysical objects

It is also possible to obtain upper bounds for the monopole flux, under the assumption of a large nucleon decay catalysis cross section, by looking at astrophysical objects (such as white dwarfs, neutron stars or large stars) [28, 36, 44]. The idea is that monopoles travelling through such objects can lose so much energy that they are captured by the object and will start to continually catalyse the decay of nucleons. Since energy is released in the processes described above, this should heat up the objects and thereby increase their luminosity. As more and more monopoles are captured as time progresses, monopoles could eventually become the main source of luminosity [36]. By comparing the observed luminosity of such objects with the luminosity they would have if they had been exposed to a certain monopole flux throughout their existence, further bounds can be put on the total monopole flux (monopoles are not *needed* to explain the luminosity of astrophysical objects, so only an upper bound is obtained).

The strongest bounds are obtained for objects that by themselves have a low luminosity, like neutron stars and white dwarfs. It has been shown that a white dwarfs and neutron stars capture almost all monopoles (with a mass below 10^{20} GeV) that hit them. The total number of monopoles in such an object will therefore be approximately given by $4\,\pi\,F\,A\,\tau$, where $4\,\pi\,F$ is the total monopole flux (from all directions), $A = 4\,\pi\,R^2(1 + (2\,G\,M)/R\,\beta^2))$ is the capture area (G is Newton's constant, and M and R are the mass and radius of the object) of the object and τ is the age of the object. It should be noted that annihilation of monopole anti-monopole pairs has not been taken into account because even after monopoles sink to the core of the object their density will not become high enough for annihilation to become effective [36].

These indirect searches based on astrophysical objects provide the lowest bounds for the monopole flux, but as with the earth-bound experiments these bounds are only meaningful under the assumption that the catalysis cross section is indeed large.

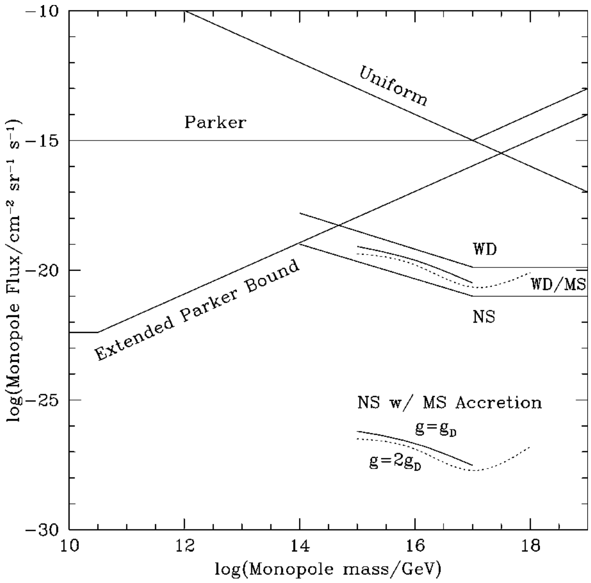


Figure 4.3: Bounds on the monopole flux as a function of the monopole mass for white dwarfs (WD) and neutron stars (NS). The curved lines are the improved results for $g = g_\mathrm{D}$ (solid) and $g = 2\,g_\mathrm{D}$ (dotted). The Parker bound and the extended Parker bound have been shown, as well as the (uniform) closure bound. This figure originally appeared in [36].

The lowest bound that was found by considering white dwarfs was from WD1136-286 [36]

$$F < 1.3 \times 10^{-20}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}/(\sigma_0/10^{-28}\ \mathrm{cm}^2), \tag{4.10}$$

where σ is the cross section discussed above and β is the magnetic monopole velocity. An even stronger bound that was obtained by looking at the pulsar PSR1929+10 [44], which limits the monopole flux to

$$F < 7 \times 10^{-22}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}/(\sigma_0/10^{-28}\ \mathrm{cm}^2)^{-1}, \tag{4.11}$$

where σ_0 is the quantity from section 4.3.3 such that the catalysis cross section is roughly $\sigma \simeq \sigma_0/\beta$. These flux limits are for particles with a velocity of $10^{-3}c$, but monopoles with a mass below 10^{17} GeV are expected to travel faster than $\beta \sim 10^{-3}$ due to acceleration by the Galaxy's magnetic field (see section 4.2). For such monopoles the capture radius of the objects is smaller, so an additional velocity (or mass) dependence should be added.

Both these results can be improved by also considering the monopoles that were already captured by these objects *before* they became white dwarfs or neutron stars (so when they were ordinary stars). Both the original results for white dwarfs and neutron stars, as well as the improved results have been collected in figure 4.3. We see that especially the improved bounds that follow from observations of the pulsar PSR1929+10 significantly exceeds the Parker bounds and the (uniform) overclosure bound by a few orders of magnitude [36]. It is important to once again note that these results are based on the assumption that the nucleon decay catalysis cross section is at least of the order $\sim 10^{-27}\ \mathrm{cm}^2$, which has not been ascertained.

5. Conclusion

The idea of particles carrying magnetic charge is already very old [32], but it was not until Dirac [11] showed in 1931 that their existence could explain the observed (yet unexplained) quantisation of electric charge that a widespread interest in it arose. Dirac's approach turned out to be a dead end, but when Unified Theories started to emergy and 't Hooft [23] and Polyakov [24] independently showed in 1974 that any such theory involving a semisimple gauge group not only allows their existence, but predicts it, magnetic monopoles once again became a hot topic.

Many years and numerous attempts to detect them later, one of the main problems of the predicted existence of monopoles remains that no evidence of their existence has been found [29, 37, 27]. This lack of observational evidence does not however exclude the possibility that magnetic monopoles do exist, since simple explanations for their scarcity are available. The question of whether or not magnetic monopoles exist at all will likely remain unanswered until either an all-encompassing theory of everything is found or until one is actually measured (both of which seem unlikely to happen any time soon). While their existence remains a mystery, we can at least be certain that if they do exist, they are a very rare phenomenon in our world.

Index

Bibliography

[1] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002, Available from: http://www.math.cornell.edu/~hatcher/AT/ATpage.html.

[2] M. W. Hirsch, *Differential Topology*, Springer-Veslag, 1994.

[3] E. P. van den Ban, Lie Groups, Lecture notes, 2003, Available from: http://www.math.uu.nl/people/ban/lecnotes/lie2005.pdf.

[4] H. Geiges, *An Introduction to Contact Topology*, Cambridge University Press, 2008.

[5] A. W. Knapp, *Lie Groups Beyond an Introduction*, Birkhäuser, 1996.

[6] J. P. Preskill, *Annual Review of Nuclear and Particle Science* **34**, 461 (1984).

[7] S. Coleman, *Aspects of Symmetry*, Cambridge University Press, 1985.

[8] R. Durrer, *New Astronomy Reviews* **43**, 111 (1999).

[9] N. Manton and P. Sutcliffe, *Topological Solitons*, Cambridge University Press, 2004.

[10] D. J. Griffiths, *Introduction to Electrodynamics*, Prentice-Hall, 3 edition, 1999.

[11] P. A. M. Dirac, *Proceedings of the Royal Society of London A* **133**, 60 (1931).

[12] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and other Topological Defects*, Cambridge University Press, 1994.

[13] H. Georgi and S. L. Glashow, *Physical Review Letters* **32**, 438 (1974).

[14] P. Langacker and S. Pi, *Physical Review Letters* **45**, 1 (1980).

[15] T. W. B. Kibble, *Physics Reports* **67**, 183 (1980).

[16] Y. B. Zeldovich and M. Y. Khoplov, *Physics Letters B* **79**, 239 (1978).

[17] J. P. Preskill, *Physical Review Letters* **43**, 1365 (1979).

[18] T. Prokopec, Lecture notes on Cosmology, Lecture notes, 2008, Available from: http://www.phys.uu.nl/~prokopec/.

[19] J. Rich, *Fundamentals of Cosmology*, Springer-Veslag, 2001.

[20] A. C. Davis and T. W. B. Kibble, *Contemporary Physics* **46**, 313 (2008), arXiv:hep-th/0505050.

[21] A. H. Guth, *Physical Review D* **23**, 347 (1981).

[22] J. P. Preskill and S. W. Hawking, *Monopoles in the Early Universe*, The Very Early Universe, Cambridge University Press, 1983.

[23] G. 't Hooft, *Nuclear Physics B* **79**, 6584 (1974).

[24] A. Polyakov, *JETP Letters (English)* **20**, 194 (1974).

[25] T. W. Kirkman and C. K. Zachos, *Physical Review D* **24**, 999 (1981).

[26] J. Baacke, *Zeitschrift für Physik C* **53**, 339 (1992).

[27] D. E. Groom, *The European Physical Journal C* **15**, 815 (1997).

[28] I. De Mitri, (2002), arXiv:hep-ex/0207051v1.

[29] G. Giacomelli and L. Patrizii, (2003), arXiv:hep-ex/0302011v2.

[30] F. A. Bais, J. Ellis, D. V. Nanopoulos, and K. A. Olive, *Nuclear Physics B* **219**, 189 (1983).

[31] T. Padmanabhan, *Physics Reports* **380**, 235 (2003).

[32] D. E. Groom, *Physics Reports* **149**, 323 (1986).

[33] E. N. Parker, *Astrophysics Journal* **160**, 383 (1970).

[34] M. S. Turner, E. N. Parker, and T. J. Bogdan, *Physical Review D* **26**, 1296 (1982).

[35] F. C. Adams, M. Fatuzzo, K. Freese, G. Tarlé, R. Watkins, and M. S. Turner, *Physical Review Letters* **70**, 2511 (1993).

[36] K. Freese and E. Krasteva, *Physical Review D* **59**, 063007 (1998).

[37] K. A. Milton, *Reports on Progress in Physics* **69**, 1637 (2006).

[38] B. Cabrera, *Physical Review Letters* **48**, 1378 (1982).

[39] S. P. Ahlen, *Physical Review D* **17**, 229 (1978).

[40] The MACRO Collaboration, (2002), arXiv:hep-ex/0207024v1.

[41] The MACRO Collaboration, (2002), arXiv:hep-ex/0207020v2.

[42] V. A. Rubakov, *Nuclear Physics B* **218**, 240 (1983).

[43] C. G. Callan, *Nuclear Physics B* **212**, 391 (1983).

[44] K. Freese, M. S. Turner, and D. N. Schramm, *Physical Review Letters* **51**, 1625 (1983).