

Introduction to magnetic monopoles

Arttu Rajantie

To cite this article: Arttu Rajantie (2012) Introduction to magnetic monopoles, Contemporary Physics, 53:3, 195-211, DOI: [10.1080/00107514.2012.685693](https://doi.org/10.1080/00107514.2012.685693)

To link to this article: <https://doi.org/10.1080/00107514.2012.685693>



Published online: 19 Jun 2012.



Submit your article to this journal [↗](#)



Article views: 13113



View related articles [↗](#)



Citing articles: 6 View citing articles [↗](#)

Introduction to magnetic monopoles

Arttu Rajantie*

Department of Physics, Imperial College London, London SW7 2AZ, UK

(Received 12 March 2012; final version received 13 April 2012)

One of the most basic properties of magnetism is that a magnet always has two poles, north and south, which cannot be separated into isolated poles, i.e. magnetic monopoles. There are strong theoretical arguments why magnetic monopoles should exist, but in spite of extensive searches they have never been found. This article gives an introduction to theoretical and experimental aspects of magnetic monopoles and discusses their role in our understanding of fundamental physics, both historically and today.

Keywords: electromagnetism; magnetic monopole; particle physics; quantum field theory

1. History of magnetic monopoles

Magnetism has a very long history (for more detailed discussion, see [1,2]), and for most of that time it has been seen as something mysterious. Even today, we still do not fully understand one very elementary property of magnets, which we all learn very early on in school: Why does a magnet always have two poles, north and south? Or, in other words, why cannot magnetic field lines end? Electric field lines end in electric charges, but it appears that there are no magnetic charges. The purpose of this paper is to give a brief summary of the physics of magnetic charges, or magnetic monopoles, and reasons why they may exist.

The earliest description of magnetism is attributed to Thales of Miletus, northern Greece, who reported that pieces of rock from Magnesia (naturally magnetic form of magnetite known as lodestone) had strange properties. He had also noted that amber could attract light objects such as feathers or hair after it had been rubbed, which is now understood as an effect of static electricity. In a somewhat similar way, lodestone could attract iron, but no other material, even other metals. Even more mysteriously, when hung from a piece of string, a piece of lodestone would always orient itself in the same way. Unlike amber, lodestone did not need rubbing or any other preparation to acquire these properties, but when an iron needle was rubbed with lodestone, these properties were transferred to it.

Lodestone was also known elsewhere, for example in China. While the Greeks thought that the rocks were attracted to the place where they had been mined, the Chinese noticed that a piece of lodestone actually always pointed north or south. At the start of the last

millennium, they already knew how to use this effect as a navigational tool in the form of a compass, and by 1187 this invention had found its way to Europe.

The attempts to explain and understand these powers of magnets were based on pure philosophical speculation until 1269 when Petrus Peregrinus of Maricourt wrote a letter called *Epistola de magnete* which described experiments he had carried out on magnets and the conclusions he had drawn from them. He showed that a magnet has two opposite poles (and introduced the term ‘polus’ to describe them), and that when a magnet is cut in two halves, each of them still has two opposite poles. He found that two like poles repel each other, but he failed to attribute the north–south orientation of the compass needle to the Earth’s magnetism. Instead, he accepted the prevailing view that it was due to its attraction to celestial poles.

It took three more centuries before further progress was made. It had been assumed that whatever force makes the compass needle turn north, it acts in the horizontal direction, but in 1576 Robert Norman found that in fact the needle tends to tilt downwards. At first he thought that he had simply failed to balance his needle properly, but further experiments confirmed his finding. A needle that was perfectly balanced would, after being magnetised, turn downwards. This was a clear hint that the force causing the orientation of the needle is due to the Earth rather than the heavens, although Norman failed to draw this conclusion from his findings. More importantly for our purposes, he found that there was actually no overall attraction at all, only a change in orientation. This is a direct sign that there are no magnetic charges because

*Email: a.rajantie@imperial.ac.uk

they would experience a force in the direction of the magnetic field.

Relatively shortly after this, in 1600, William Gilbert published a book titled *De Magnete* [3] in which he tried to construct the first consistent theory of magnetism, based on careful and systematic experiments. In particular, he was the first to understand that the orientation of the compass needle was caused by the magnetism of the Earth. This book is regarded as the start of the scientific study of magnetism.

During the following centuries, many aspects of magnetism and its connection with electricity were discovered, and theories describing it became more specific and more quantitative. Nevertheless, it was widely believed that magnetism was caused by two oppositely charged magnetic fluids, which consist of magnetic molecules, or magnetic monopoles as we would call them. This idea persisted well into the nineteenth century, until Andre-Marie Ampère showed that magnetic fields are generated by electric currents and eventually Michael Faraday demonstrated that the magnetic fluids do not exist [4].

Building on Faraday's findings and physical intuition about electric and magnetic fields, James Clerk Maxwell developed his theory of electrodynamics, which he published in 1864 [5]. This theory was a massive breakthrough in physics, because it not only unified magnetism and electricity in one simple theory but also explained the properties of light, which it showed to be an electromagnetic wave travelling through space. Furthermore, the theory showed that the speed of light had to be constant, and therefore paved the way for the theory of relativity.

In his theory, Maxwell assumed that no magnetic charges exist, but the theory did not explain why, and did not forbid them. Magnetic charges were sometimes discussed to illustrate some aspect of physics or as a convenient mathematical tool, but they were not considered to be real, physical objects. One exception to this was Pierre Curie [6], who speculated on the possibility of free magnetic charges in 1894.

When quantum mechanics was developed at the start of the twentieth century, it first appeared to forbid isolated magnetic charges. However, in 1931 Paul Dirac showed that in fact, quantum mechanics allows certain quantised magnetic charges [7]. Furthermore, the existence of magnetic charges would explain the observed quantisation of electric charge, which until then had been a mystery. Finally, in 1974, Gerard 't Hooft and Alexander Polyakov showed that magnetic monopoles are predicted by many particle physics models [8,9], in particular by all Grand Unified Theories (GUTs), which aim to describe electromagnetic interactions and the weak and strong nuclear forces by a single unified theory [10]. The same

conclusion also applies generally to more modern 'theories of everything' (TOEs) such as superstring theory, which also include gravitation [11].

From a theoretical point of view, we would therefore expect magnetic monopoles to exist. Finding one would be an incredible breakthrough for high energy physics. They derive their properties from the laws of physics at very high energies, which are far out of reach of any particle accelerator experiments. They are absolutely stable, so they would not decay to other particles, unlike most other particles physicists are hoping to find, and they interact relatively strongly through the electromagnetic field, which means that they would be easy to study experimentally. Their discovery would therefore allow physicists to design experiments to test GUTs and TOEs directly. Unfortunately, all attempts to find them (see, e.g. [12,13]) have failed so far, but they have still played a key role in the development of particle physics and cosmology in recent decades.

The purpose of this article is to give a brief introduction to the physics of magnetic monopoles. For a more detailed review, see, e.g. [14,15]. A comprehensive and up-to-date magnetic monopole bibliography is also available [16].

2. Monopole solutions

2.1. Monopoles in classical electrodynamics

Classical electrodynamics is described elegantly by four Maxwell's equations. In the absence of any electric and magnetic charges, they read,¹

$$\vec{\nabla} \cdot \vec{E} = 0, \quad (1)$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad (2)$$

$$\vec{\nabla} \cdot \vec{B} = 0, \quad (3)$$

$$\vec{\nabla} \times \vec{B} = \frac{\partial \vec{E}}{\partial t}, \quad (4)$$

where \vec{E} and \vec{B} are vectors representing the electric and magnetic fields, respectively. In these equations the symbol $\vec{\nabla} \cdot$ is known as the *divergence* of the vector field, and $\vec{\nabla} \times$ is known as the *curl*. If one represents the electric and magnetic fields by field lines, non-zero divergence would indicate that field lines are ending, whereas the curl describes their curvature. Therefore, Equations (1) and (3) show that neither electric nor magnetic field lines have start or end points or, in other words, that they are closed. Equations (2) and (4) show that time-dependent magnetic fields generate electric fields, and vice versa.

These equations have a beautiful symmetry known as electric-magnetic duality: if we replace the electric and magnetic fields as

$$\vec{E} \rightarrow \vec{B}, \quad \text{and} \quad \vec{B} \rightarrow -\vec{E}, \quad (5)$$

the equations remain unchanged. This means that the electric and magnetic fields themselves behave in exactly the same way.

Of course, electric charges exist, and electric field lines can start and end at electric charges. More precisely, the electric field around the electric charge q has the form

$$\vec{E}(\vec{r}) = \frac{q}{4\pi r^3} \vec{r} \quad (6)$$

and a charge moving in an electromagnetic field at velocity \vec{v} experiences the Lorentz force

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}). \quad (7)$$

The duality (5) shows that if magnetic charges existed, the magnetic field around a magnetic charge g would be the dual of Equation (6),

$$\vec{B}(\vec{r}) = \frac{g}{4\pi r^3} \vec{r}, \quad (8)$$

and a magnetic charge would experience the force

$$\vec{F} = g(\vec{B} - \vec{v} \times \vec{E}). \quad (9)$$

However, because no magnetic charges have been found, the duality symmetry appears to be broken. This means that, in fact, the essential difference between electricity and magnetism is that electric charges exist but magnetic charges do not.

This raises the question why nature has this asymmetry. From the point of view of classical electrodynamics, there is no reason why there could not be magnetic charges, and if they did, the duality symmetry would be intact. In other words, classical electrodynamics is perfectly compatible with the notion of magnetic monopoles, and from the aesthetic point of view it is strange that magnetic monopoles do not seem to exist, because their existence would make the theory more symmetric.

2.2. Monopoles in quantum theory

In quantum theory, the question about the possible existence of magnetic monopoles becomes more complex but also more intriguing. This is because it turns out that in quantum mechanics, electromagnetic

forces have to be described in terms of scalar and vector potentials ϕ and \vec{A} [17], instead of the electric and magnetic fields \vec{E} and \vec{B} . These potentials appear in the theory as position-dependent fields, whose variations give rise to electric and magnetic fields through the relations

$$\begin{aligned} \vec{E} &= -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla} \phi, \\ \vec{B} &= \vec{\nabla} \times \vec{A}. \end{aligned} \quad (10)$$

In classical electrodynamics these potentials are widely used as convenient calculation tools, and the scalar potential is, in fact, simply the familiar electric potential which we learn about in school, but their use is optional and one can perfectly well do calculations in terms of the electric and magnetic fields, too.

The way the scalar and vector potentials are introduced through Equation (10) does not respect the duality symmetry between \vec{E} and \vec{B} . However, the potentials themselves are not physical, observable quantities, and they describe exactly the same physics as the original Maxwell equations (1)–(4). In fact, there are an infinite number of different potentials that give rise to the same electric and magnetic fields, because they remain unchanged if one changes ϕ and \vec{A} by

$$\phi \rightarrow \phi - \frac{\partial \lambda}{\partial t}, \quad \vec{A} \rightarrow \vec{A} + \vec{\nabla} \lambda, \quad (11)$$

where λ is any function. Changing from one of these physically equivalent potential configurations to another by Equation (11) is known as a *gauge transformation*, and because physical quantities do not change under such gauge transformations we say that the theory has a *gauge symmetry*. The gauge symmetry of Maxwell's equations is known mathematically as U(1). Similar, but somewhat more complex gauge symmetries play a central role in our current theories of particle physics. They are seen as the fundamental principle that determines the nature of the elementary particle interactions, and we will come back to them shortly.

Nevertheless, the potentials (10) appear to forbid magnetic charges. This is because one of the basic results in vector analysis is that the divergence of the curl of a vector field always vanishes, so that one finds

$$\vec{\nabla} \cdot \vec{B} = \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = 0. \quad (12)$$

Therefore, if the magnetic field is represented by the vector potential through Equation (10), then its field

lines can never have end points. This appears to show that one cannot describe magnetic monopoles using the vector potential.

This restriction becomes important in quantum mechanics. Remember that in quantum mechanics a particle is described in terms of its wave function ψ , which is a complex number. In general, the particle does not have a well-defined position, but the probability of finding the particle at a particular position is given by the square of the absolute value of the wave function, $|\psi|^2$. The complex nature of the wave function becomes apparent in interference experiments. For example, in the famous *double-slit experiment* (see Figure 1), particles can travel through either of two parallel slits, and the final wave function $\psi(\vec{x})$ on the screen behind them is a linear superposition of the wave functions $\psi_1(\vec{x})$ and $\psi_2(\vec{x})$ corresponding to each of the two routes, $\psi(\vec{x}) = \psi_1(\vec{x}) + \psi_2(\vec{x})$. At places where the complex phases of the two wave functions coincide, one has *constructive interference* and the probability is enhanced, and at places where the complex phases are opposite, the two wave functions cancel each other leading to *destructive interference* and reduced probability.

When one describes an electrically charged particle in quantum mechanics, the complex phase θ of the wave function ψ depends on the vector potential \vec{A} . More precisely, when moving in space, the complex phase changes by the rate given by the component of the vector potential parallel to the motion, which can be expressed as

$$\vec{\nabla}\theta = q\vec{A}. \quad (13)$$

This is compatible with the gauge symmetry, because neither the complex phase nor the vector potential are directly observable quantities. In fact, the gauge transformation (11) corresponds to changing the complex phase by a position-dependent amount,

$$\theta(\vec{x}) \rightarrow \theta(\vec{x}) + \lambda(\vec{x}). \quad (14)$$

Nevertheless, this connection between the wave function and the vector potential shows that one needs to use the vector potential instead of the magnetic field in quantum mechanics. Because the vector potential cannot describe a magnetic monopole, does this mean that quantum mechanics prohibits the existence of magnetic charges?

In 1931, Paul Dirac showed that it does not [7]. First, he was able to find a vector potential that describes a magnetic monopole. He did this in essentially the same way Faraday had constructed a ‘magnetic monopole’ for his experiments 110 years earlier [18]: in order to investigate the motion of a magnetic monopole in the presence of an electric

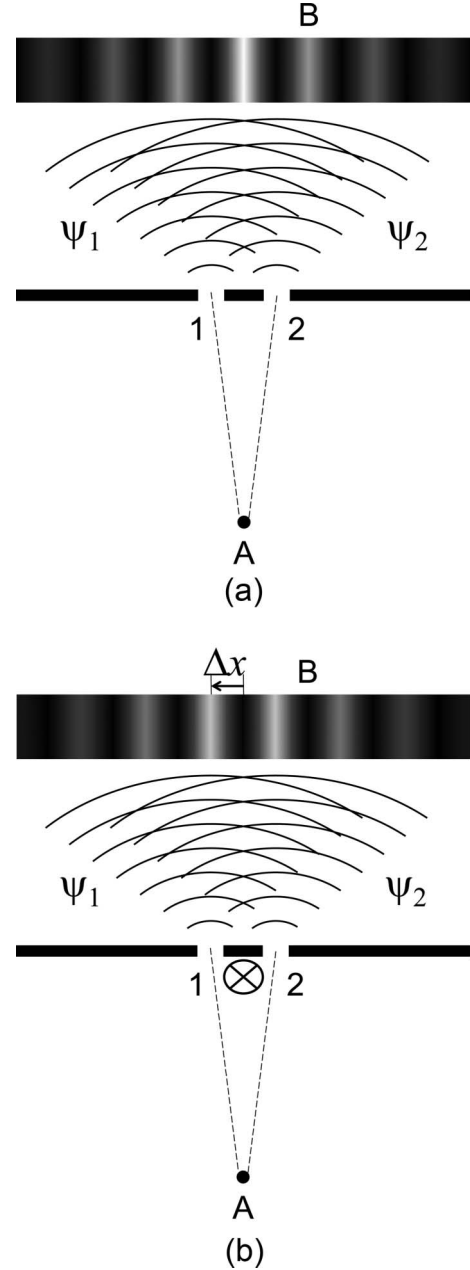


Figure 1. (a) In the double-slit experiment, charged particles from the source A travel through one of two slits 1 and 2, and are detected on a screen B behind them. The wave functions ψ_1 and ψ_2 interfere depending on their complex phases, producing a pattern consisting of peaks and troughs on the screen. (b) When a solenoid \otimes (or a Dirac string) carrying magnetic flux is placed between the slits, it changes the relative complex phases of the two wave functions ψ_1 and ψ_2 . This shifts the interference pattern by a distance Δx which is proportional to the flux Φ . For $\Phi = 2\pi/q$, this distance is equal to the distance between peaks, and therefore the pattern is unchanged and the Dirac string is unobservable.

current, Faraday used a long magnet in a vessel filled with mercury, in such a way that one of the poles was under the surface and the other above it. That way the

upper pole behaved essentially like a magnetic monopole.

To be more precise, let us consider a long and thin solenoid, i.e. a long straight coil of wire. When an electric current runs through the solenoid, it produces a magnetic field inside it. Because the field lines cannot end, this field spreads out in all directions from the end of the solenoid. It is useful to describe this in terms of the magnetic flux Φ , which is the magnetic field integrated over a cross-sectional surface,

$$\Phi = \int d\vec{S} \cdot \vec{B}. \quad (15)$$

Because field lines cannot end, the flux coming out of the end of the solenoid has to be equal to the flux Φ inside it.

If the length of the solenoid is much greater than its width, the shape of the magnetic field around the end of the solenoid looks exactly like that of a magnetic monopole with magnetic charge g equal to the magnetic flux, so that $g = \Phi$. Along the length of the solenoid, the magnetic field is confined inside it, apart from the monopole fields at the two ends. Because there are no magnetic charges in this system, one can write down the vector potential for it without any difficulty. Mathematically, one can imagine first making the solenoid infinitesimally thin and long so that the other endpoint is infinitely far away, and then forget about the solenoid itself and only keep the vector potential, which can be written as

$$\vec{A}(\vec{r}) = \frac{g}{4\pi|\vec{r}|} \frac{\vec{r} \times \hat{k}}{|\vec{r}| - \vec{r} \cdot \hat{k}}. \quad (16)$$

Here \hat{k} is a unit vector pointing in the direction of the solenoid. This vector potential describes a magnetic monopole connected to an infinitesimally thin line (known as a *Dirac string*) carrying the magnetic flux to it. Because in classical electrodynamics magnetic fields only affect particles that are inside the field, the Dirac string is unobservable if its thickness is zero, and therefore the vector potential really describes an isolated magnetic monopole. The price one has to pay is that the vector potential is singular along the Dirac string.

However, the situation is more complicated in quantum mechanics. As we saw earlier, the vector potential \vec{A} affects the complex phase of the particle's wave function, and therefore the Dirac string can give rise to an interference effect even if the particles are not actually touching it. If we consider again the double slit experiment, and imagine having a Dirac string placed between the two slits, the wave functions corresponding to the two slits would pick up a different

complex phase factor because they are travelling through a different vector potential. This would shift the interference pattern on the screen left or right depending on the magnetic flux [17].

It is fairly easy to calculate the extra complex phase $\Delta\theta$ a particle picks up. According to Equation (13), it is simply given by integrating the vector potential \vec{A} along the trajectory and multiplying this by the electric charge q of the particle. We can write this as

$$\Delta\theta_X = q \int_{A \rightarrow X \rightarrow B} d\vec{r} \cdot \vec{A}, \quad (17)$$

where $X = 1$ or 2 , depending on which slit the particle goes through. The interference pattern on the screen will only be affected by the Dirac string if the particles passing through different slits pick up a different complex phase. We can write this difference as

$$\Delta\theta = \Delta\theta_1 - \Delta\theta_2 = q \oint_C d\vec{r} \cdot \vec{A}, \quad (18)$$

which represents an integral along a closed curve C from A to 1 to B to 2 back to A . A general result from vector analysis known as Stokes's theorem shows that the integral is equal to the magnetic flux Φ through the area enclosed by the curve C . Combining this with the observation above that the flux is equal to the magnetic charge g of the monopole, we find the relation

$$\Delta\theta = q \oint_C d\vec{r} \cdot \vec{A} = q \int d\vec{S} \cdot \vec{B} = q\Phi = qg, \quad (19)$$

which shows that the complex phase difference between the two slits $\Delta\theta$ is equal to the product of the electric charge q and the magnetic charge g . The Dirac string of any magnetic monopole with non-zero magnetic charge g would therefore generate a non-zero difference in the complex phase. Now, it is important to realise that the complex phase is only defined modulo 2π . Two complex numbers whose phases differ by an integer multiple of 2π are equal. Therefore, the Dirac string is only observable if the phase difference $\Delta\theta$ is not an integer multiple of 2π . Conversely, if the electric and the magnetic charges q and g satisfy [7]

$$\frac{qg}{2\pi} \in \mathbb{Z}, \quad (20)$$

then the Dirac string is unobservable and the electrically charged particle only feels the magnetic field of the monopole. This is known as the *Dirac quantisation condition*.

If there are several different types of electrically charged particles, then in order for the Dirac string to be completely unobservable, the electric charge of each of the particle species has to satisfy Equation (20). Dirac realised that this is only possible if the electric charge is quantised, meaning that the electric charge of any particle is an integer multiple of some elementary charge, which itself satisfies Equation (20). This is a remarkable finding because electric charges are indeed quantised in nature: they are all integer multiples of the electron charge e (or one third of it, $e/3$, if one includes quarks but they are not observed as individual particles). Therefore, Dirac had found that this could be explained if magnetic monopoles exist, and, perhaps emboldened by his earlier successful prediction of the positron, he concluded that it would be surprising if magnetic monopoles did not exist [7,19].

Even though we now have a much better understanding of elementary particles and their properties than in Dirac's time, and even though there are other ways one can try to understand the quantisation of electric charge, they all seem to predict magnetic monopoles at the same time, so it seems that Dirac's insight may well have been valid. For example, string theorist Joseph Polchinski [20] has argued that any theory that explains charge quantisation would inevitably predict the existence of magnetic monopoles.

Although, as Dirac showed, magnetic monopoles can be consistently described in quantum theory, they do not appear automatically unlike the positron. Quite the contrary, it is actually difficult to add magnetic monopoles to quantum electrodynamics, the quantum version of Maxwell's theory [21]. This can be understood as the flip side of Dirac's quantisation argument: in quantum electrodynamics electric charges are not naturally quantised, and therefore including magnetic monopoles has to change the structure of the theory significantly to force the quantisation condition (20).

2.3. 't Hooft–Polyakov monopoles

By the 1970s, physicists had moved on from quantum electrodynamics to develop a theoretical description of strong and weak nuclear forces, which would eventually become our current Standard Model of particle physics. The guiding principle was the concept of gauge symmetry that we encountered in the context of Maxwell's theory of electrodynamics, and although the gauge transformations involved are more complicated, the structure of these theories is very similar to Maxwell's theory.

In particular, progress was being made to unify electromagnetism and the weak nuclear force into one *electroweak* force (see Figure 2). The theory that achieved this [23–25] has two different gauge

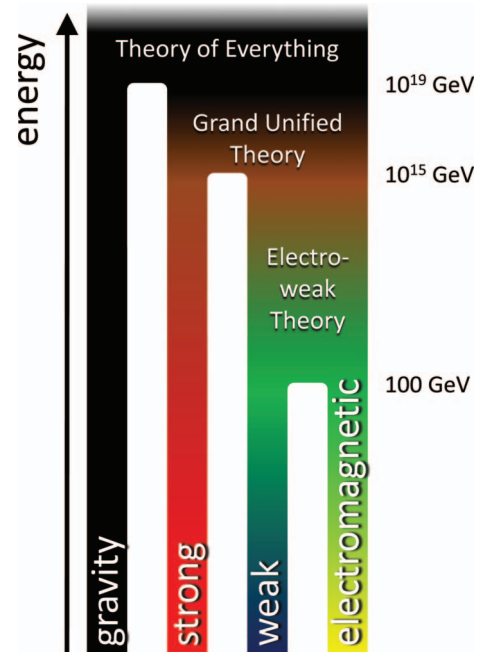


Figure 2. Unification of fundamental forces. There is strong experimental evidence that at energies above 100 GeV the electromagnetic and weak nuclear forces are described by a unified electroweak theory. It is possible that this is unified with strong nuclear forces to a Grand Unified Theory (GUT) above 10^{15} GeV. Finally, it is known that at energies above the *Planck scale*, 10^{19} GeV, the gravitational force becomes so strong that it needs to be included in the theory, and it is believed that it will be unified with all the elementary particle forces into a Theory of Everything, which may be a superstring theory or the M-theory [22].

symmetries known as $U(1)$ and $SU(2)$. The $U(1)$ gauge symmetry has the same structure as the gauge symmetry of Maxwell's equations, but the $SU(2)$ symmetry is somewhat more complicated. In particular, in $U(1)$ the order in which one makes two gauge transformations does not matter, but in $SU(2)$ it does. Because of this, the former is known as an Abelian symmetry, and the latter as a non-Abelian symmetry.

At low energies, this $SU(2) \times U(1)$ symmetry is 'broken' because of a hypothetical *Higgs field* ϕ [26–28], which has a non-zero value in the vacuum and behaves essentially like a quantum wave function spread over the whole space. Just like the complex phase of the wave function, the Higgs field is sensitive to the values of the vector potentials corresponding to the two gauge symmetries. Although the theory itself still has the gauge symmetry, the vacuum state with the non-zero Higgs field is no longer symmetric, and therefore only certain gauge transformations are allowed. These turn out to correspond exactly to the $U(1)$ gauge symmetry of Maxwell's equations, and therefore they give rise to the electromagnetic forces.

As a by-product the theory predicted the existence of some new particles, massive W and Z bosons which carry the weak interactions and were discovered in 1983, and the Higgs boson which the Large Hadron Collider (LHC) at CERN is currently trying to find.

In 1974 Gerard 't Hooft [8] and Alexander Polyakov [9] considered a slightly different variant of this theory, which had been proposed earlier by Georgi and Glashow [29]. Remarkably, 't Hooft and Polyakov found that this theory would necessarily contain both electric and magnetic charges.

The Georgi–Glashow model is somewhat simpler than the $SU(2) \times U(1)$ electroweak theory. It has only one gauge symmetry, a non-Abelian symmetry known as $SO(3)$. The Higgs field consists of three real-valued components ϕ^a , with $a \in \{1,2,3\}$. We can therefore think of it as a three-component vector $\boldsymbol{\phi}$, although it is important to keep in mind that it is a vector in the abstract ‘internal’ space rather than real space. The Higgs field carries potential energy whose density depends on its value as

$$V(\boldsymbol{\phi}) = \frac{\lambda}{4} (|\boldsymbol{\phi}|^2 - v^2)^2, \quad (21)$$

where λ and v are parameters that would have to be determined from experiments.

Because the potential $V(\boldsymbol{\phi})$ depends only on the length $|\boldsymbol{\phi}|$ of the Higgs field vector, it is clearly spherically symmetric in the internal space. The gauge transformations correspond to rotating this internal space, and the gauge symmetry means that such a rotation does not change the laws of physics. As in electrodynamics [see Equation (14)], this rotation can be position-dependent, and this requires a three-component *gauge field* \vec{A}^a , which is the analogue of the electromagnetic potentials.

In the vacuum state, the potential energy (21) is minimised and therefore the length of the Higgs field vector is non-zero, $|\boldsymbol{\phi}| = v$. However, it can point in any direction, so there is a sphere of possible vacuum states (see Figure 3). In analogy with Equation (13), the Higgs and gauge fields are related by the requirement that the covariant derivative, defined as

$$(\vec{D}\boldsymbol{\phi})^a = \vec{\nabla}\boldsymbol{\phi}^a + q\epsilon^{abc}\vec{A}^b\boldsymbol{\phi}^c, \quad (22)$$

has to vanish, $(\vec{D}\boldsymbol{\phi})^a = 0$ for all a . Here ϵ^{abc} is the Levi–Civita tensor, and we follow the usual convention that repeated indices b and c are summed over. Once the direction of the Higgs field is fixed, only rotations around that axis are allowed, so the original $SO(3)$ symmetry is said to be spontaneously broken. The remaining rotations correspond to the $U(1)$ gauge transformations (11) of Maxwell’s equations, and

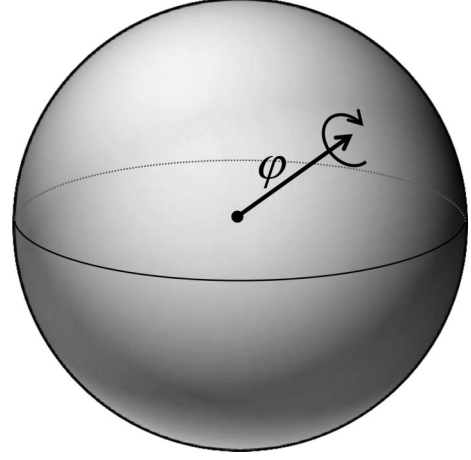


Figure 3. Spontaneous symmetry breaking. The Georgi–Glashow model of electroweak unification has an $SO(3)$ gauge symmetry which correspond to rotations in the three-dimensional internal space. However, the Higgs field ϕ has a non-zero length, and therefore its possible values lie on a sphere in the internal space, corresponding to different possible vacuum states. Because of the $SO(3)$ symmetry, all vacuum states are identical, but once the vacuum state has been chosen, only rotations around the Higgs field axis are possible. Therefore the full $SO(3)$ symmetry is spontaneously broken to a smaller $U(1)$ symmetry, which gives rise to electromagnetism.

therefore they give rise to electromagnetism. Furthermore, the requirement that a charged particle would have to behave consistently under any rotation of the whole three-dimensional internal space, not just those around the Higgs field vector axis, gives rise to a condition that the electric charge has to be quantised in units of some elementary charge q .

't Hooft and Polyakov considered what happens if the direction of the Higgs field is not the same everywhere. In most cases, the field would quickly find its way to the vacuum state. However, they found that this would not happen to the so-called hedgehog configuration, which one can visualise by imagining that the directions in the internal space correspond to directions in real space. The hedgehog configuration would then be the case in which the Higgs field vector points away from the origin everywhere (see Figure 4). We can write this field configuration as

$$\phi^a(\vec{r}) = vf(r)\frac{r_a}{r}, \quad (23)$$

where r_a is component a of the position vector \vec{r} , and $r = |\vec{r}|$. The function $f(r)$, which one usually has to find numerically, describes the length of the Higgs field vector and is a function of radius r only, so that the configuration is spherically symmetric. It follows from

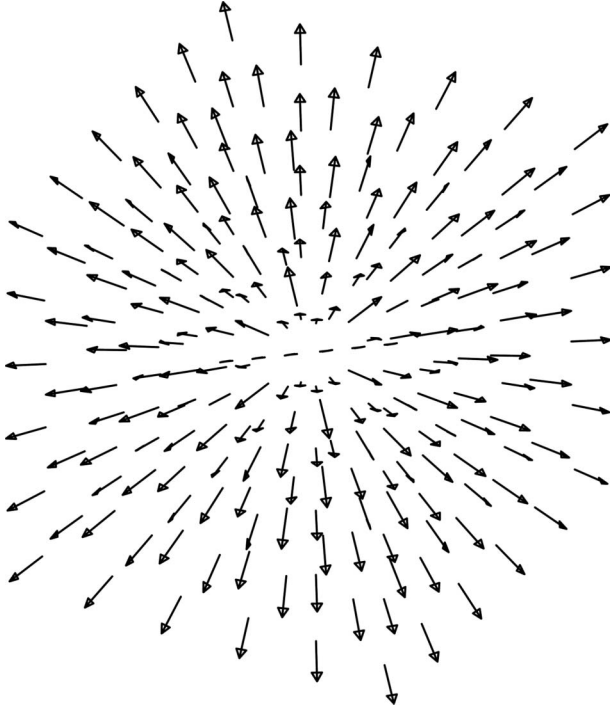


Figure 4. The hedgehog configuration. In the 't Hooft–Polyakov monopole solution, the Higgs field vector (indicated by the arrows) points away from the origin everywhere and its length approaches the vacuum value v far from the origin. This configuration cannot be turned continuously into the uniform vacuum state, so it is topologically stable. In order for the field to be continuous, it cannot be in the vacuum state at the origin, and therefore there is a localised lump of energy (in other words, a particle) at the origin.

Equation (22), that far from the origin the gauge field is

$$A_i^a(\vec{r}) = \frac{1}{q} \epsilon_{aij} \frac{r_j}{r^2}, \quad (24)$$

where i and j correspond to directions in real space and a in the internal space. This solution is an example of a topological defect or a topological soliton [30]: it is stable because it is impossible to turn it continuously into the uniform vacuum state.

Far from the origin, the Higgs field approaches its vacuum length, which means that $f(r) \rightarrow 1$, and its direction is almost uniform. However, in order to be continuous, the field has to go to zero at the origin, $f(0) = 0$, and its direction has to vary rapidly around it, which means that it will have non-zero energy. Therefore, the hedgehog configuration appears physically as a lump of energy localised in a small volume near the origin. Because according to the theory of relativity, mass is energy, this means that it is

essentially a massive particle. Furthermore, because the Higgs field vanishes at the origin, the full $\text{SO}(3)$ gauge symmetry is unbroken inside the particle.

However, the most striking result is that the particle has a magnetic charge. Because electromagnetism is generated by rotations around the Higgs field vector, one can calculate the magnetic field in the hedgehog configuration. 't Hooft found that it is given by the expression [8]

$$B_i = \epsilon_{ijk} \hat{\phi}^a F_{jk}^a - \frac{1}{q} \epsilon_{ijk} \epsilon^{abc} \hat{\phi}^a (D_j \hat{\phi})^b (D_k \hat{\phi})^c, \quad (25)$$

where $\hat{\phi}^a = \phi^a / |\phi|$ and

$$F_{jk}^a = \partial_j A_k^a - \partial_k A_j^a + q \epsilon^{abc} A_j^b A_k^c. \quad (26)$$

Substituting the solutions (23) and (24), one finds that this is just the magnetic field (8) of a monopole with magnetic charge

$$g = \frac{4\pi}{q}, \quad (27)$$

which is twice the minimum allowed by the Dirac quantisation condition (20). In fact, Equation (25) has the general topological property that when integrated over any closed surface, it always gives an integer times $4\pi/q$. According to Equation (15) such an integral is equal to the magnetic charge inside the surface, and therefore this means that the magnetic charge is quantised in any field configuration, not just in the 't Hooft–Polyakov hedgehog solution.

The mass of these 't Hooft–Polyakov monopoles would be roughly 100 GeV, determined by the energy scale associated with weak nuclear forces. However, experimental results quickly ruled out the Georgi–Glashow model as a theory of electroweak unification, and the successful $\text{SU}(2) \times \text{U}(1)$ theory does not allow similar magnetic monopole solutions.

However, Georgi and Glashow noticed very soon that their theory could be modified to do something even more ambitious, namely to unify the electroweak theory with strong nuclear forces into one Grand Unified Theory (GUT) [31], which would describe all known elementary particle forces except gravity (see Figure 2). They achieved this very elegantly using only one non-Abelian gauge symmetry known as $\text{SU}(5)$, and two Higgs fields. At very high energies of around 10^{15} GeV, one of the Higgs fields would break the GUT symmetry into three pieces, the $\text{SU}(2)$ and $\text{U}(1)$ symmetries of the electroweak theory and a further $\text{SU}(3)$ symmetry that describes strong nuclear forces, and the second Higgs would then play exactly the same role as in the electroweak theory.

Because the structure of the SU(5) GUT was so similar to the Georgi–Glashow electroweak theory, it also predicted 't Hooft–Polyakov monopoles, but their mass would now be much higher, around 10^{16} GeV (or 10 ng in SI units). They also have the interesting property that they can catalyse nucleon decay [32–34]. According to the Standard Model of particle physics, the proton is a stable particle and cannot decay. In the grand unified theory this is no longer true, but the proton still has a very long lifetime because the decay requires a quantum tunnelling process through a very high potential barrier. However, as we saw earlier, the full SU(5) symmetry is unbroken inside the monopole and therefore the potential barrier disappears. This means that when a proton or a neutron gets in contact with a GUT monopole, they can decay very fast.

In fact, the existence of magnetic monopoles is not specific to the SU(5) GUT. Instead, it is an unavoidable consequence of grand unification: any theory that would describe all strong and electroweak forces by one grand unified interaction would have 't Hooft–Polyakov monopoles [35]. The reason for this lies in a branch of mathematics known as topology, which studies whether one function can be continuously deformed to another.

The way to demonstrate that the 't Hooft–Polyakov hedgehog configuration is stable is to consider a sphere around the origin. In the Georgi–Glashow model, the Higgs field is a three-component vector, and because its length is fixed, the set of possible values it can take also the form a sphere in the internal space. In the hedgehog solution, the field values on a sphere in real space wrap once around the sphere in the internal space, and this cannot be continuously deformed into the vacuum state, in which the field values do not wrap around the sphere, and this makes the configuration stable. In topology, homotopy theory tells precisely when stable configurations like this exist, and one can show that whenever a gauge symmetry is broken by a Higgs field in such a way that a U(1) gauge symmetry remains, stable hedgehog configurations exist [35,36]. Because a U(1) gauge symmetry is needed for electromagnetism, this implies that any GUT that unifies electromagnetism with other forces into one gauge symmetry would always have stable hedgehogs, i.e. 't Hooft–Polyakov monopoles.

This was a remarkable finding, because there were (and still are) many theoretical reasons to expect grand unification. First of all, the Standard Model of particle physics has a fairly complex structure consisting of three different gauge symmetries and a wide range of different matter particles each with their own properties, which gets a simple explanation in grand unified theories. The strengths of the three interactions are

also such that they coincide when extrapolated to the energy of 10^{15} GeV. However, even if grand unification is valid as a general concept, we cannot be sure that the Georgi–Glashow SU(5) GUT is the correct theory. It is therefore important that the prediction of magnetic monopoles is general, and not specific to this particular theory.

It is obviously almost impossible to test grand unified theories with accelerator experiments, because that would require energies around 10^{15} GeV, which is one trillion times higher than what the LHC can reach. It is therefore exciting that magnetic monopoles can give a potential way to study them experimentally. Another generic prediction of grand unified theories is that the proton can decay. This means that hydrogen atoms, which have a single proton as their nucleus should have finite lifetime. Because every water molecule contains two hydrogen atoms, one should be able to detect this by observing a large enough tank of water. At some point a proton should decay in the tank, creating a tiny flash of light, which can be detected. However, no proton decays have ever been seen in these experiments. This rules out the original Georgi–Glashow SU(5) GUT, but there are other GUTs which predict a much longer proton lifetime and which are therefore still compatible with experiments.

These days, the focus in high energy physics has shifted from grand unified theories to even more ambitious 'theories of everything' such as superstring theory, which aim to include also gravity. The structure of these theories is different because all four forces are unified at the same energy and because they are not based on fields but on other types of objects such as strings or membranes. Therefore the argument for the existence of 't Hooft–Polyakov monopoles does not apply in the same form. However, they still generally predict the existence of magnetic monopoles [11].

3. Monopoles in cosmology

If the fundamental laws of physics allow the existence of magnetic monopole particles, then it is likely that they would have been produced in the Big Bang or shortly afterwards in the early universe [10,37]. Therefore, with a good enough understanding of the production mechanism and the early universe, one should be able to calculate how many monopoles there should be in the universe today. However, even very simple estimates show that the prediction is far too high to be compatible with observations [10]. This is known as the *monopole problem*.

To see this, let us start by looking at how the monopoles may have been produced in the early

universe. The basic idea is known as the *Kibble mechanism* after Tom Kibble who discovered it in 1976 [36], and although one can argue that it is not strictly speaking applicable to grand unified theories [38], the main conclusions are still valid [39,40].

For simplicity, let us consider the $SO(3)$ Georgi–Glashow model in which the Higgs field can be thought of as a three-component vector. The early universe was very hot, and if it was hot enough the GUT symmetry would have been initially unbroken, and the Higgs field was zero. When the universe expanded and cooled down, it went through a phase transition to the phase with the broken gauge symmetry when the universe was very young, only 10^{-35} s old. In this phase transition, the Higgs field became non-zero, and because it is a vector it has to have a direction. In order to minimise the energy, the direction would have to be the same everywhere.

However, the transition took place in a finite time. Remember that according to relativity, information cannot travel faster than the speed of light. If you consider two points in space that are so far apart that information from one cannot reach the other before the transition is completed, then those two points have to choose the direction of the Higgs field independently of each other. Because of the symmetry, every direction is equally likely. Therefore it follows that after the transition the Higgs field will be uniform on short scales, up to the distance ξ that light was able to travel during the transition, but random and uncorrelated at longer distances. Roughly speaking, we can imagine that the universe consists of domains of size ξ , in each of which the direction of the Higgs field is uniform but random. In the early universe, this correlation distance cannot have been longer than the size of the particle horizon, which is roughly given by the age of the universe or the inverse of the expansion rate (or Hubble rate) H at the time.

On the other hand, the Higgs field has to be continuous, so where two domains meet, it will interpolate smoothly between them. However, consider a point between four domains. It is possible that, when the field interpolates between each pair of them, it ends up wrapping around the sphere in the internal space. In that case, the field cannot continuously interpolate between all four domains without vanishing in the middle, and a monopole or an anti-monopole (a monopole with negative magnetic charge) is formed (see Figure 5). The probability that the directions in the four domains are such that this happens is independent of their size or any other parameter, so it is simply some constant number which can be calculated in principle but which is clearly going to be not much less than one. Therefore one concludes that roughly one monopole or anti-monopole is produced

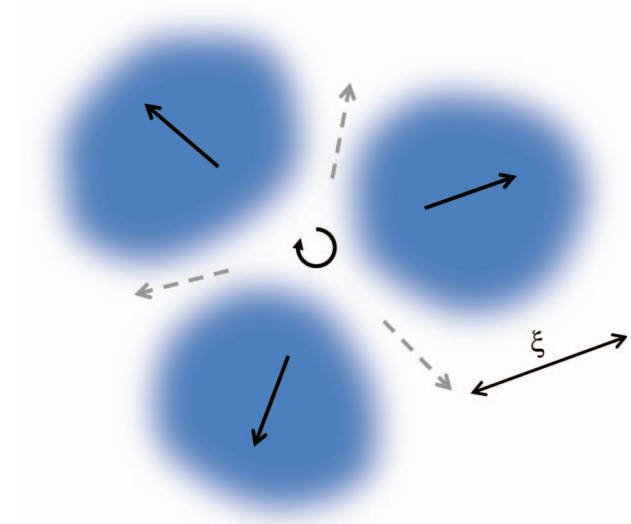


Figure 5. A two-dimensional example of the Kibble mechanism. The Higgs field is random at distances longer than the correlation length ξ . Therefore we can imagine that the universe consists of domains of size ξ , each with a random but more or less uniform field direction (solid arrows). The field interpolates continuously between any pair of two domains (dashed arrows), but where three domains meet, it may not be able to do that without vanishing in the middle. In that case a topological defect (a vortex) is formed. Similarly, in three dimensions, monopoles can be formed where four domains meet.

per domain, or per particle horizon [36]. Because the particle horizon was very short in the very early universe, when the GUT phase transition would have taken place, this leads to a very high density of monopoles and anti-monopoles.

Nevertheless, the magnetic monopoles carry initially only a very small fraction of the total energy, most of which is in the form of radiation. Because both monopoles and anti-monopoles are produced in equal numbers, they can initially meet and destroy each other reducing their number density. However, after a while those processes stop and then the number density of monopoles decreases only because of the expansion of the universe. As a result, the number density of magnetic monopoles in the universe today would be comparable to the number density of protons and neutrons [10,35]. The problem is that the mass of a monopole, at roughly 10^{16} GeV, is sixteen orders of magnitude higher than the mass of a proton or neutron. It is therefore absolutely clear that the prediction cannot be valid.

One possible solution to this monopole problem is that magnetic monopoles simply do not exist. However, in 1980 Alan Guth proposed an alternative explanation in the form of a theory called inflation

[41]. According to this theory, the universe expanded at an accelerating rate in its very early stages. If this happened after the GUT phase transition (or during it as Guth originally thought), it could potentially dilute the monopole density to an acceptable level. With a sufficiently long period of acceleration, inflation would also solve two other major puzzles in Big Bang cosmology, known as the horizon and flatness problems. Furthermore, a more detailed calculation shows that inflation also explains the origin of primordial density perturbations which acted as the initial seeds for the formation of structures such as galaxies and galaxy clusters. These same perturbations can be observed more directly as temperature fluctuations in the cosmic microwave background radiation. Measurements of these temperature fluctuations, by the WMAP satellite [42] and other experiments, have shown that they agree astonishingly well with the predictions of the inflationary theory.

In spite of its great successes, inflation is more a general idea than a specific theory. There are a large number of theoretical models that give rise to enough inflation to solve the monopole, horizon and flatness problems and can produce perturbations that are compatible with observations. In most of these models, inflation takes place at energies well below the GUT scale. Therefore, if monopoles were formed at all, this would have happened before inflation. Inflation would then have solved the monopole problem by diluting their number density down to a small enough level.

However, there are some models in which GUT or intermediate mass monopoles are formed in a phase transition at the end of inflation or shortly afterwards [40]. In those cases it will be important to understand the precise mechanism by which the monopoles form and the way they evolve afterwards.

4. Monopole searches

Following the theoretical prediction by 't Hooft and Polyakov, there have been many attempts to detect magnetic monopoles in experiments. These fall into three categories: one can try to produce them in accelerator experiments such as the LHC, one can try to find existing magnetic monopoles either in cosmic rays or trapped in materials, or one can look for indirect signs of magnetic monopoles in astronomical observations.

For direct searches there are three primary methods: the first is to use a superconducting ring. If a magnetic monopole passes through the ring, the changing magnetic field induces a current, which can be measured. The change in the current also gives the strength of the magnetic charge very accurately. The second method relies on the fact that magnetic

monopoles are very highly ionising: because of their strong magnetic charge, a magnetic monopole travelling through matter would easily strip electrons off atoms, leaving a clear track in detector plates. Because of the form of the Lorentz force (7), this effect is velocity dependent, which allows one to distinguish a monopole from a strong electric charge. The third approach is to look for signs of nucleon decay catalysed by monopoles travelling through matter [33,43].

The following is a brief summary of different experimental searches. More details can be found in [12,13].

4.1. Accelerator searches

In principle, magnetic monopoles should be produced in particle accelerator experiments if the collision energy is sufficiently high, higher than $2Mc^2$. For GUT monopoles the required energy is at least one trillion times higher than the energies available at the Large Hadron Collider (LHC). Therefore, it is unrealistic to expect that they could be produced in any foreseeable particle accelerators.

However, we really only know the laws of physics up to the electroweak scale, which is at around 100 GeV, and therefore it is perfectly possible that there are magnetic monopoles that are much lighter than GUTs would predict. These are called intermediate-mass monopoles, and if they are light enough, they could very well be produced at the LHC, although there are huge theoretical uncertainties in their predicted production rate [44] (see Section 5 for more discussion).

Looking for magnetic monopoles is very different from looking for most other hypothetical particles. Exotic particles are usually very difficult to detect because they have a very short lifetime. The particle decays very quickly into known particles, and therefore one has to analyse the particles produced in the collision very carefully to see if some of them might have been created by the hypothetical new particle. This is very difficult, because it is much more likely that the same known particles were produced in other processes.

In contrast, magnetic monopoles are stable. A monopole can only be destroyed when it meets an anti-monopole. Therefore, once created in an experiment, the monopole will not decay. Because of its magnetic charge, the monopole will also interact relatively strongly with the electromagnetic field. This interaction is well understood, and also very different from any known particles, and therefore there would be little risk of confusion.

Because of these differences, the main LHC experiments are not well suited for looking for

magnetic monopoles. Instead, a dedicated experiment called MoEDAL [45] is being constructed for that purpose. Compared with the other LHC experiments, MoEDAL is very simple: it consists of plastic nuclear track detector sheets placed around the LHCb experiment. If monopoles are produced in collisions, they will fly through the sheets, leaving tell-tale marks on them which can be detected when the sheets are removed and analysed. Monopoles have also been searched for in different ways in other accelerators in the past, such as Tevatron, LEP and HERA, but with no success. This excludes monopoles with mass less than roughly 1 TeV [46].

In principle, quantum field theory also allows the possibility that a ‘virtual’ pair or two oppositely charged monopoles are produced and destroy each other after a very short time. This process is possible even when the collision energy is not quite high enough to produce two real monopoles. It is much harder to look for virtual monopoles, but there have been some attempts [47,48]

4.2. Direct searches

Instead of trying to produce monopoles in an experiment, one can also try to look for monopoles that already exist in the universe. Because monopoles are stable, even monopoles created in the very early universe should still be around. Obviously, how easy it is to find them depends on how many of them there are, or more precisely on their flux F . By this we mean the number of monopoles that would hit a unit area per unit time per solid angle. The flux is conventionally measured in units of $\text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$, where sr stand for *steradian* and is the unit of solid angle. The solid angle of a whole sphere is 4π steradians. Therefore the number of monopoles hitting a sphere of radius R in time t is equal to $(4\pi \text{ steradians}) \times FRt$. This clearly depends not only on the number of monopoles but also on their velocity.

There have been many attempts to look for magnetic monopoles in cosmic rays, which are particles hitting the Earth from space. The higher the flux, the more often a monopole cosmic ray would hit a detector. Some early experiments appeared to show promising evidence for them. In 1973, a cosmic ray balloon experiment showed a particle track that appeared to be consistent with a magnetic monopole [50], but it was soon accepted that it was more likely to have been a platinum nucleus [51]. On Valentine’s day in 1982, a laboratory experiment using a superconducting ring showed a sudden jump in the current by exactly the amount that a magnetic monopole passing through the ring would generate [52], and another event was seen in a similar experiment at

Imperial College in 1985 [53]. However, because later experiments have not been able to reproduce them, they are believed to have been caused by other effects. Nevertheless, they motivated physicists to design larger experiments which were able to put stronger limits on the flux of the magnetic monopoles. Furthermore, these experiments were also able to look for other types of particles in cosmic rays and they paved the way for current dark matter searches.

The strongest bounds for GUT monopoles come from an underground experiment called MACRO (Monopole, Astrophysics and Cosmic Ray Observatory) which was based in Gran Sasso, Italy and operated from 1989 to 2000. It was specially designed to detect magnetic monopoles, and contained different types of detectors using both ionisation and nucleon decay methods, with a total effective surface area of roughly $10,000 \text{ m}^2$. It did not detect any monopoles, and from this one can conclude that the monopole flux cannot be higher than roughly $F \lesssim 10^{-16} \text{ cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$ [54] [see Figure 6(a)]. Several other experiments such as AMANDA [55] and Baikal [56] have produced similar bounds later. For lighter, intermediate-mass monopoles the RICE experiment (Radio Ice Cherenkov Experiment) at the South Pole has given an even stronger bound of roughly $F \lesssim 10^{-16} \text{ cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$ [49] [see Figure 6(b)].

Because the Earth has been bombarded by cosmic rays for billions of years, monopoles could have well been trapped in matter on Earth. There have been attempts to find magnetic monopoles trapped in materials like moon rock, meteorites and sea water, but with no success [57].

4.3. Astrophysical bounds

If magnetic monopoles exist, they would produce astrophysical effects, which we can hope to detect in astronomical observations. Perhaps the most significant effect is that magnetic fields would accelerate them, generating magnetic currents which would drain energy from the field. The rate at which the energy of the magnetic field would decrease depends on the monopole flux. If the flux was very high, we would not expect to find any large-scale magnetic fields in space because they would have dissipated away. However, observations show that there is a magnetic field of roughly $3\mu\text{G}$ in our galaxy, from which we can conclude that the monopole flux cannot be very high. This is known as the *Parker bound* [58,59]. The strength of this effect depends also on the mass of the monopoles. If they are very heavy, around 10^{17} GeV , their motion is dominated by gravitational forces rather than magnetic fields, and therefore the

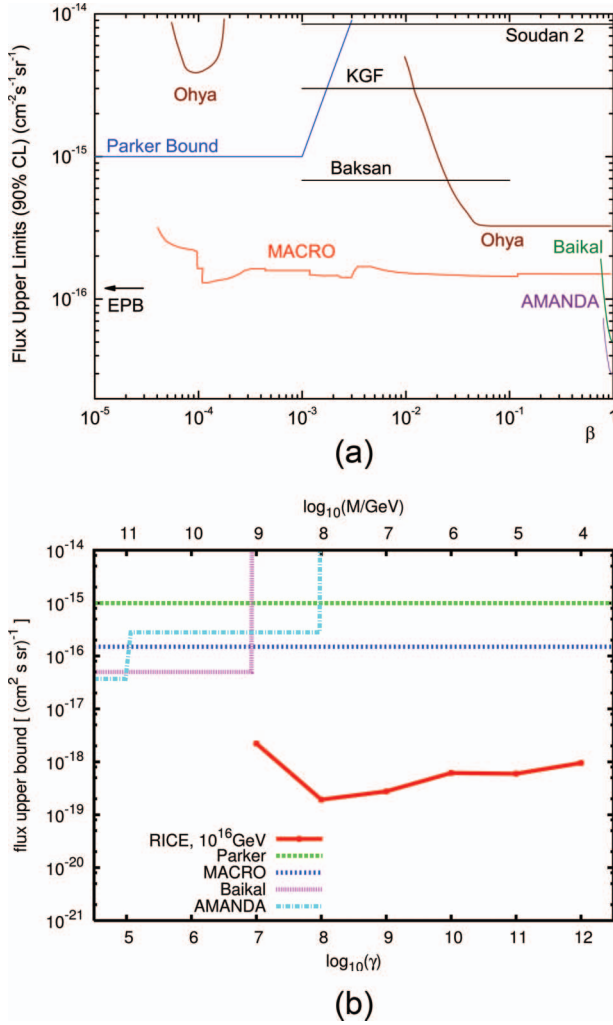


Figure 6. Upper bounds for the flux of (a) GUT monopoles (from [13]), and (b) intermediate mass magnetic monopoles (from [49]), from several experiments as well as the astrophysical Parker bound. In (a) the x axis β is the velocity of the monopoles in units of the speed of light, $\beta = v/c$. Reprinted figure with permission from D. Hogan, D. Besson, J. Ralston, I. Kravchenko, and D. Seckel, Phys. Rev. D 78, 075031 (2008). Copyright (2008) by the American Physical Society.

bound is weak (see Figure 6). On the other hand, for lighter monopoles one finds a stronger bound

$$F \lesssim 10^{-15} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}. \quad (28)$$

An even simpler bound follows from the total mass of the magnetic monopoles in the universe. From astronomical observations [42], we know that matter particles make up approximately 23% of the total energy in the universe, and roughly one third of this can be accounted for by known particles. In principle, some or all of the remaining dark matter could be magnetic monopoles. If they are light enough, less than

10^{17} GeV , they would not be bound to galaxies by gravity, and therefore their density would be more or less uniform. From the observed dark matter density, one obtains an upper bound

$$F \lesssim 10^{-16} (10^{16} \text{ GeV}/M) (v/10^{-3} c) \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}, \quad (29)$$

where M is the mass of a monopole, v is their average speed, and c is the speed of light. Because of the mass dependence, this bound is only relevant for heavy monopoles, such as GUT ones, but for them it is comparable to other bounds.

5. Monopoles and quantum field theory

Even though magnetic monopoles have never been discovered, and therefore we do not know if they really exist, they have still been extremely useful for theoretical physics. We already saw in Section 3 how they were the motivation for inflation, which has later been confirmed by other observations. They have also played an equally important role in understanding quantum field theories.

Attempts to include magnetic monopoles have forced physicists to look at quantum field theories from a new perspective and to develop new calculation techniques. Among the early examples is Schwinger's formulation of quantum electrodynamics with magnetic charges in 1966 [21], but this direction became much more important with the theoretical discovery of 't Hooft–Polyakov monopoles in 1974. The conventional interpretation of quantum field theories was that particles correspond to field quanta, or quantised waves, so that each particle species would correspond to one particular field. For example, photons are the quanta of the electromagnetic field. However, 't Hooft–Polyakov monopoles are localised field configurations, which appear in the theory in addition to all the field quanta. These days analogous topological solitons [30] play a central role especially in superstring theory.

Describing solitons such as 't Hooft–Polyakov monopoles in quantum field theory gives rise to new challenges. The conventional way to do calculations in quantum field theory, known as perturbation theory, is designed for field quanta, and it cannot be used to study topological solitons. Furthermore, it assumes that the quanta interact with each other weakly. For example, the strength of the interactions between electrons and photons is given by the electron charge e , or perhaps more precisely by the fine structure constant $\alpha = e^2/4\pi \approx 1/137$. This is a small number, and therefore the interaction is weak, and perturbation theory works. The problem is that the Dirac quantisation condition (20) tells us that the magnetic charge of

the magnetic monopole is very large. The analogous magnetic fine structure constant would be $\alpha_M = g^2/4\pi = 1/\alpha \approx 137$, which is not a small number. Therefore the magnetic interaction between two monopoles is actually very strong, and perturbation theory is generally not applicable.

The main alternative to perturbation theory is to do numerical lattice field theory simulations. These simulations are widely used to study quantum chromodynamics (QCD), the theory of the strong nuclear force. As the name suggests, this interaction is not weak either. However, because magnetic monopoles and other solitons are very different in nature from the particles involved in QCD, the techniques used in lattice QCD are not always directly applicable. Instead, one needs to develop new techniques to calculate the properties of magnetic monopoles, such as their mass and interactions with other particles. This is particularly important because there are experimentally important quantities which one cannot calculate reliably using other techniques, for example the effects of virtual monopole–anti-monopole pairs. Most importantly, semiclassical calculations suggest that the pair production rate of magnetic monopoles in particle collisions is suppressed by a massive factor which would make it practically impossible to find them in accelerator experiments such as the LHC [44], and it is crucial to check whether this is also true in quantum theory. There has been significant progress recently in developing these methods, and they have been used to

calculate the mass and the interaction form factors of magnetic monopoles [60–62]. Figure 7 shows a visualisation of a 't Hooft–Polyakov monopole in such a lattice field theory simulation.

Magnetic monopoles may also be able to help us understand the physics of QCD better. The theory is defined in terms of two types of particles: quarks which are matter particles, and gluons which mediate the strong force. However, neither quarks nor gluons have been seen as free particles in experiments. Instead, we know that they are confined into protons, neutrons and other hadron particles, which are all made up of several quarks. If we tried to pull the quarks apart, we would feel a constant force keeping them together as if they were connected by a string. This phenomenon is called *confinement*, and we know from lattice field theory simulations that QCD can describe it correctly, but there is no rigorous proof because we do not have any reliable ways of doing analytic calculations with the theory.

In 1976, Gerard 't Hooft [65] and Stanley Mandelstam [66] suggested that confinement could be explained by magnetic monopoles. Note first that under certain circumstances (typically low temperature) some materials become superconducting. This means that electric currents can flow without any resistance, but also that magnetic fields are repelled. If one could put a magnetic monopole inside a superconductor, its magnetic field would be confined into a string known as an Abrikosov flux tube. A monopole and an anti-monopole would therefore be connected by a string.

't Hooft and Mandelstam realised that confinement looks very much like a dual version of this in the sense of the electric-magnetic duality (5), i.e. when we swap electric and magnetic fields and charges. Because superconductivity is caused by the behaviour of electrically charged particles, confinement could therefore be explained by magnetic monopoles behaving in a similar way. These monopoles would not be the 'real' GUT monopoles, but analogous mathematical solutions in QCD. It is difficult to make this idea precise in QCD, but in some simpler theories, people have been able to show rigorously that it is correct and that confinement is indeed caused by magnetic monopoles [67,68].

In particular, Claus Montonen and David Olive argued in 1977 that the electric-magnetic duality (5) should be valid not only in electrodynamics, but also in the $SO(3)$ Georgi–Glashow model with 't Hooft–Polyakov monopoles [69]. Their arguments were based on the semiclassical approximation, and it is difficult to test their conjectured duality in the quantum theory, because of the problems associated with describing magnetic monopoles in quantum field theory [61].

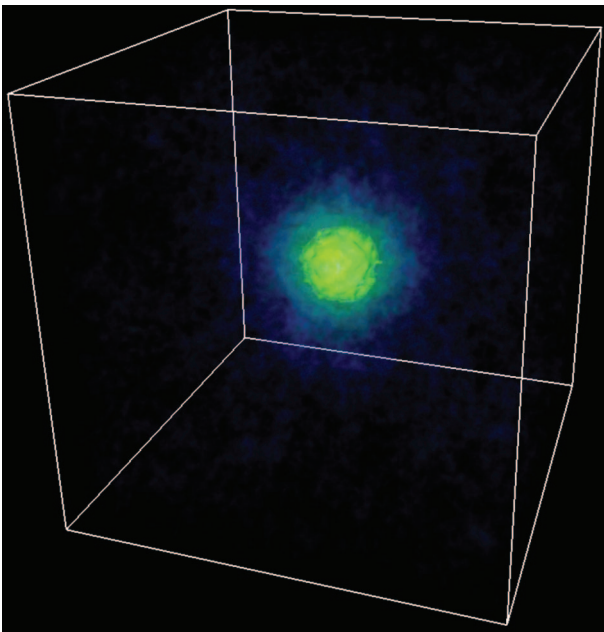


Figure 7. A 't Hooft–Polyakov monopole configuration in a lattice field theory simulation. This figure was produced using VAPOR [63,64].

However, it has been shown to be true in supersymmetric versions of the theory [68], in which calculations are simpler. This Montonen–Olive duality is very useful for many calculations, because it turns the strong magnetic coupling α_M into a weak electric coupling. Therefore one can use perturbation theory to do calculations even when interactions are strong. Similar dualities have also turned out to be very important in superstring theory [70] where they can be used to relate different variants of the theory to each other. This suggests that there is actually only one fundamental theory, known as M-theory [22], and different superstring theories are simply different ways of describing the same physics.

6. Conclusions

One of the most fundamental properties of magnetism is that there are no magnetic charges, or magnetic monopoles. Yet, as we have seen, there are strong theoretical reasons to expect that they should exist. Magnetic monopoles are considered to be ‘one of the safest bets one can make about physics not yet seen’ [20]. Theorists believe that they are simply so rare that we have not detected them yet, and so heavy that we have not been able to produce them in particle accelerators. However, as Petrus Peregrinus wrote [2], ‘experience rather than argument is the basis of certainty in science’, and therefore the search has to continue. The discovery of magnetic monopoles would not only confirm that we are on the right track but also give us a unique way to probe physics at extremely high energies, far beyond the reach of any foreseeable particle accelerators.

Nevertheless, even without such a discovery, magnetic monopoles have been driving our understanding of physics towards more and more fundamental levels for centuries, and this progress is still continuing today in quantum field theory, where monopoles and the electric-magnetic duality give us access to phenomena we would not be able to study and understand in other ways. The absence of magnetic monopoles in the universe also led to the theory of inflation, which is the cornerstone of modern cosmology. Therefore, whether they exist or not, magnetic monopoles have a central place in our understanding of the universe.

Acknowledgements

The author would like to thank David Weir for extensive discussions, and acknowledge funding by the STFC grant ST/J000353/1 and Royal Society International Joint Project JP100273. The calculations for Figure 7 were performed on the COSMOS Consortium supercomputer within the DiRAC Facility jointly funded by STFC and the Large Facilities Capital Fund of BIS.

Note

1. In SI units, these equations also involve the permittivity ϵ_0 and permeability μ_0 of the vacuum, but for simplicity I am using *natural units* in which these are both set equal to 1, which also implies that the speed of light is equal to one, $c = 1$. Furthermore, one also sets Planck’s constant to one, $\hbar = 1$. This has the consequence that everything can be expressed in units of energy. The unit of energy conventionally used in particle physics is gigaelectronvolt $1 \text{ GeV} \approx 1.6 \times 10^{-10} \text{ J}$.

Notes on contributor



Arttu Rajantie is Reader in Theoretical Physics at Imperial College London. He obtained his PhD from the University of Helsinki in 1997. After that he worked for two years at the University of Sussex, and then for five years in the Department of Applied Mathematics and Theoretical Physics (DAMTP) at the University of Cambridge, where he was also a fellow of Churchill College, until he joined Imperial College in 2005. He is interested in non-linear phenomena in quantum field theory, and their implications for particle physics and cosmology. In particular, he has developed ways to study magnetic monopoles and other topological solitons using lattice Monte Carlo simulations.

References

- [1] B. Baigrie, *Electricity and Magnetism – A Historical Perspective*, Greenwood Press, Westport, CT, 2007.
- [2] G.L. Verschuur, *Hidden Attraction – The Mystery and History of Magnetism*, Oxford University Press, Oxford, 1996.
- [3] W. Gilbert, *De Magnete (Translated by P. Fleury Mottelay)*, Dover Publications, New York, 1991.
- [4] M. Faraday, *Experimental Researches in Electricity, Vol. III*, Taylor and Francis, London, 1855.
- [5] J.C. Maxwell, *A dynamical theory of the electromagnetic field*, Philos. Trans. R. Soc. London 155 (1865), pp. 459–512.
- [6] P. Curie, *Sur la possibilité d’existence de la conductibilité magnétique et du magnétisme libre*, Séances de la Société française de physique, Paris, 1894.
- [7] P. Dirac, *Quantized Singularities in the Electromagnetic Field*, Proc. R. Soc. London A 133 (1931), pp. 60–72.
- [8] G. ’t Hooft, *Magnetic Monopoles in Unified Gauge Theories*, Nucl. Phys. B 79 (1974), pp. 276–284.
- [9] A.M. Polyakov, *Particle Spectrum in the Quantum Field Theory*, JETP Lett. 20 (1974), pp. 194–195.
- [10] J. Preskill, *Cosmological Production of Superheavy Magnetic Monopoles*, Phys. Rev. Lett. 43 (1979), p. 1365.
- [11] M. Duff, R.R. Khuri, and J. Lu, *String solitons*, Phys. Rep. 259 (1995), pp. 213–326.
- [12] K. Nakamura, *et al.*, *Review of particle physics*, J. Phys. G G37 (2010), 075021.
- [13] G. Giacomelli, L. Patrizii, and Z. Sahnoun, *Searches for Magnetic Monopoles and... beyond*, preprint (2011). Available at arXiv:1105.2724.

- [14] K.A. Milton, *Theoretical and experimental status of magnetic monopoles*, Rep. Prog. Phys. 69 (2006), pp. 1637–1712.
- [15] Y.M. Shnir, *Magnetic Monopoles*, Springer, Berlin, 2005.
- [16] S. Balestra, G. Giacomelli, M. Giorgini, L. Patrizii, V. Popa, Z. Sahnoun, V. Togo, *Magnetic monopole bibliography-II, preprint* (2011). Available at arXiv:1105.5587.
- [17] Y. Aharonov and D. Bohm, *Significance of electromagnetic potentials in the quantum theory*, Phys. Rev. 115 (1959), pp. 485–491.
- [18] M. Faraday, *Experimental Researches in Electricity, Vol. II* Taylor and Francis, London, 1844.
- [19] P. Dirac, *The theory of magnetic poles*, Phys. Rev. 74 (1948), pp. 817–830.
- [20] J. Polchinski, *Monopoles, duality, and string theory*, Int. J. Mod. Phys. A19S1 (2004), pp. 145–156.
- [21] J.S. Schwinger, *Magnetic charge and quantum field theory*, Phys. Rev. 144 (1966), pp. 1087–1093.
- [22] M. Duff, *M theory (the theory formerly known as strings)*, Int. J. Mod. Phys. A 11 (1996), pp. 5623–5642.
- [23] S. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. 22 (1961), pp. 579–588.
- [24] A. Salam and J.C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. 13 (1964), pp. 168–171.
- [25] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. 19 (1967), pp. 1264–1266.
- [26] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. 13 (1964), pp. 321–323.
- [27] P.W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. 13 (1964), pp. 508–509.
- [28] G. Guralnik, C. Hagen, and T. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. 13 (1964), pp. 585–587.
- [29] H. Georgi and S.L. Glashow, *Unified weak and electromagnetic interactions without neutral currents*, Phys. Rev. Lett. 28 (1972), p. 1494.
- [30] N. Manton and P. Sutcliffe, *Topological Solitons*, Cambridge University Press, Cambridge, 2004.
- [31] H. Georgi and S. Glashow, *Unity of all elementary particle forces*, Phys. Rev. Lett. 32 (1974), pp. 438–441.
- [32] J. Callan Curtis G, *Monopole catalysis of baryon decay*, Nucl. Phys. B 212 (1983), p. 391.
- [33] V. Rubakov, *Adler–Bell–Jackiw anomaly and fermion number breaking in the presence of a magnetic monopole*, Nucl. Phys. B 203 (1982), pp. 311–348.
- [34] V. Rubakov, *Monopole catalysis of proton decay*, Rep. Prog. Phys. 51 (1988), pp. 189–241.
- [35] J. Preskill, *Magnetic monopoles*, Ann. Rev. Nucl. Part. Sci. 34 (1984), pp. 461–530.
- [36] T. Kibble, *Topology of cosmic domains and strings*, J. Phys. A 9 (1976), pp. 1387–1398.
- [37] Y. Zeldovich and M. Khlopov, *On the concentration of relic magnetic monopoles in the universe*, Phys. Lett. B 79 (1978), pp. 239–241.
- [38] A. Rajantie, *Magnetic monopoles from gauge theory phase transitions*, Phys. Rev. D 68 (2003), p. 021301.
- [39] A. Rajantie, *Formation of topological defects in gauge field theories*, Int. J. Mod. Phys. A 17 (2002), pp. 1–44.
- [40] A. Rajantie, *Defect formation in the early universe*, Contemp. Phys. 44 (2003), pp. 485–502.
- [41] A.H. Guth, *The inflationary universe: a possible solution to the horizon and flatness problems*, Phys. Rev. D 23 (1981), pp. 347–356.
- [42] E. Komatsu, K.M. Smith, J. Dunkley, C.L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M.R. Nolte, L. Page, D.N. Spergel, M. Halpern, R.S. Hill, A. Kogut, M. Limon, S.S. Meyer, N. Odegard, G.S. Tucker, J.L. Weiland, E. Wollack and E.L. Wright, *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation*, Astrophys. J. Suppl. 192 (2011), p. 18.
- [43] S. Errede, et al., *Experimental limits on magnetic monopole catalysis of nucleon decay*, Phys. Rev. Lett. 51 (1983), p. 245.
- [44] A. Drukier and S. Nussinov, *Monopole pair creation in energetic collisions: Is it possible?*, Phys. Rev. Lett. 49 (1982), p. 102.
- [45] J. Pinfold, *MoEDAL becomes the LHC’s magnificent seventh*, CERN Cour. 50N4 (2010), pp. 19–20.
- [46] M. Fairbairn, A. Kraan, D. Milstead, T. Sjostrand, P.Z. Skands, et al., *Stable massive particles at colliders*, Phys. Rep. 438 (2007), pp. 1–63.
- [47] M. Acciarri, et al., *Search for anomalous Z to gamma gamma events at LEP*, Phys. Lett. B 345 (1995), pp. 609–616.
- [48] B. Abbott, et al., *A Search for heavy point – like Dirac monopoles*, Phys. Rev. Lett. 81 (1998), pp. 524–529.
- [49] D. Hogan, D. Besson, J. Ralston, I. Kravchenko, and D. Seckel, *Relativistic magnetic monopole flux constraints from RICE*, Phys. Rev. D 78 (2008), p. 075031.
- [50] P. Price, E. Shirk, W. Osborne, and L.S. Pinsky, *Evidence for detection of a moving magnetic monopole*, Phys. Rev. Lett. 35 (1975), pp. 487–490.
- [51] L.W. Alvarez, *Analysis of a reported magnetic monopole*, in Proceedings of the 1975 International Symposium on Lepton and Photon Interactions at High Energy, Stanford, 1975.
- [52] B. Cabrera, *First results from a superconductive detector for moving magnetic monopoles*, Phys. Rev. Lett. 48 (1982), pp. 1378–1380.
- [53] A. Caplin, M. Hardiman, M. Koratzinos, and J. Schouten, *Observation of an unexplained event from a magnetic monopole detector*, Nature 321 (1986), pp. 402–406.
- [54] M. Ambrosio, et al., *Final results of magnetic monopole searches with the MACRO experiment*, Eur. Phys. J. C 25 (2002), pp. 511–522.
- [55] R. Abbasi, Y. Abdou, T. Abu-Zayyad, J. Adams, J. Aguilar, et al., *Search for relativistic magnetic monopoles with the AMANDA-II neutrino telescope*, Eur. Phys. J. C 69 (2010), pp. 361–378.
- [56] K. Antipin, et al., *Search for relativistic magnetic monopoles with the Baikal Neutrino Telescope*, Astropart. Phys. 29 (2008), pp. 366–372.
- [57] H. Jeon and M.J. Longo, *Search for magnetic monopoles trapped in matter*, Phys. Rev. Lett. 75 (1995), pp. 1443–1446.
- [58] E.N. Parker, *The origin of magnetic fields*, Astrophys. J. 160 (1970), p. 383.
- [59] M.S. Turner, E.N. Parker, and T. Bogdan, *Magnetic monopoles and the survival of galactic magnetic fields*, Phys. Rev. D 26 (1982), p. 1296.
- [60] A. Davis, T. Kibble, A. Rajantie, and H. Shanahan, *Topological defects in lattice gauge theories*, J. High Energy Phys. (online) 0011 (2000), 010.
- [61] A. Rajantie, *Mass of a quantum ‘t Hooft–Polyakov monopole*, J. High Energy Phys. (online) 0601 (2006), 088.
- [62] A. Rajantie and D.J. Weir, *Nonperturbative study of the ‘t Hooft–Polyakov monopole form factors*, Phys. Rev. D 85 (2012), 025003.

- [63] J. Clyne and M. Rast, *A prototype discovery environment for analyzing and visualizing terascale turbulent fluid flow simulations*, in *Visualization and Data Analysis 2005 (Proceedings Volume)*, R.F. Erbacher, J.C. Roberts, M.T. Grohn and K. Borner, eds., Proceedings of SPIE Vol. 5669, SPIE, Bellingham, WA, 2005.
- [64] J. Clyne, P. Mininni, A. Norton, and M. Rast, *Interactive desktop analysis of high resolution simulations: application to turbulent plume dynamics and current sheet formation*, New J. Phys 9 (2007).
- [65] G. 't Hooft, *Gauge theories with unified weak, electromagnetic and strong interactions*, in EPS International Conference on High Energy Physics, Palermo, 1975, pp. 1225–1249.
- [66] S. Mandelstam, *Vortices and quark confinement in nonabelian gauge theories*, Phys. Rep. 23 (1976), pp. 245–249.
- [67] A.M. Polyakov, *Quark confinement and topology of gauge groups*, Nucl. Phys. B 120 (1977), pp. 429–458.
- [68] N. Seiberg and E. Witten, *Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory*, Nucl. Phys. B 426 (1994), pp. 19–52.
- [69] C. Montonen and D.I. Olive, *Magnetic monopoles as gauge particles?*, Phys. Lett. B 72 (1977), p. 117.
- [70] J. Polchinski, *String duality: a colloquium*, Rev. Mod. Phys. 68 (1996), pp. 1245–1258.