J. Undergrad. Sci. 3: 47-55 (Summer 1996)

Theory of Magnetic Monopoles and Electric-Magnetic Duality: A Prelude to *S*-Duality

JUN S. SONG

We present a self-contained, elementary description of magnetic monopoles in classical physics. The electric-magnetic duality is discussed both in non-relativistic particle mechanics and in relativistic classical field theory. In the process, we will see that magnetic monopoles appear as soliton solutions in certain field theories. The paper concludes with brief comments on *S*-duality. The ultimate goal is to make the abstract understandable to the general public, thus making clear the *possibility* of the existence of magnetic monopoles.

Introduction

I do not think that I know what I do not know.

—Plato, *Apology*

Anyone who is familiar with elementary physics or chemistry would agree that an electron is a point-like particle with electric charge 1.602 x 10^{-19} Coulombs. It is equally well known that opposite electric charges attract and like ones repel. But what is charge? Qualitatively, there are two types of electric charge, + and - as commonly called, and the amount of charge determines the strength of the force between two charged objects. The more the charge, the stronger the force. There now arises another question: How is the force created, or, more appropriately, how do charged particles interact? To answer this question, we first note the remarkable fact that the electric force depends only on charge and not on the particular nature of particles. For example, the *electric* force between two muons, which are particles with the same charge as the electron but 200 times heavier, is equal to that between two electrons in the same external conditions. Hence, it seems that the fundamental concept that we must understand before we delve into more difficult questions is that of "charge."

In order to describe electromagnetism, Faraday introduced the concept of fields as mediators of interaction. Since our physical world has three spatial dimensions, electric and magnetic fields are 3-component vectors[1] defined at every point in space. See Figure 1. Equivalently, the fields are vector-valued functions of space and are sometimes called the vector fields. In classical electrodynamics, electric charges are sources of electric fields. A point-like particle with positive (negative) electric charge has an electric field radially pointing outwards (inwards), as in Figure 2. Hence, we define:

Definition 1. *Charges are sources of the fields. The stronger the charge, the stronger the fields.*

How, then, do charged particles interact? Electric charges are also responders to the electric field, so that a positively charged particle at position *x* responds to an external field at *x* by moving in the direction of the field. On the other hand, an electron and neutron do not interact *electrically* with each other because a neutron, carrying no electric charge, does not respond to the electric field created by the electron. Equivalently, the neutron does not create electric fields to which the electron can respond. Thus we also have:

Definition 2. *Charges[2] are also responders to the fields. The stronger the charge, the stronger the response.*

Charges are often called the coupling constants because their strength determines the coupling, or interaction, strength with the corresponding fields.

As we shall see shortly in greater detail, there is a striking symmetry between electric and magnetic fields in our description of electromagnetism. That is, the equations governing the dynamics of the electric and magnetic fields are unchanged when the fields are exchanged. For example, the energy density of electromagnetic fields[3] is

$$\frac{1}{8\pi}|\mathbf{E}|^2 + \frac{1}{8\pi}|\mathbf{B}|^2$$

which is manifestly symmetric in $\mathbf{E}$ and $\mathbf{B}$. Thus, analogous to the formulation of electric charges, it is certainly conceivable that there exist magnetic charges which are sources and responders to the magnetic fields. See Figure 3. A particle carrying magnetic charge is called a magnetic monopole.

The first comment that can be made about a magnetic monopole is that it has not been observed experimentally. Nevertheless, as Ed Witten once asserted in his Loeb Lecture at Harvard, almost all theoretical physicists believe in the existence of magnetic monopoles, or at least hope that there is one. There was an upsurge of interest in the subject in 1970s and 1980s for several compelling reasons.The study of magnetic monopoles has brought together many seemingly unrelated concepts in physics through the fascinating notion of duality. Duality is a symmetry that relates two distinct theories in such a way that they describe the same physics. Descriptions of magnetic monopoles in the modern physics lead to the strong/weak coupling duality that relates a theory that describes a strong force to another

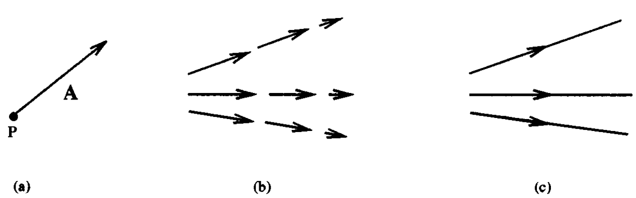

Figure 1. Vector fields. (a) The length of the arrow represents the magnitude of the vector $\mathbf{A}$ at point *P* and the arrow points in the direction of $\mathbf{A}$; **(b)** vector field evaluated at several points in space; **(c)** we usually connect the arrows into smooth lines such that the direction of the line represents the direction of the field and the concentration of the lines represents the magnitude.

JUN S. SONG is a recent graduate of Harvard College, having majored in physics. He independently authored this article, which has received the Bowdoin Prize for exemplary writing in the natural sciences. Mr. Song is also the recipient of the Thomas T. Hoopes Prize and the Herchel Smith Scholarship. He will be attending Cambridge University for the next year before pursuing a Ph.D. at the Massachusetts Institute of Technology.

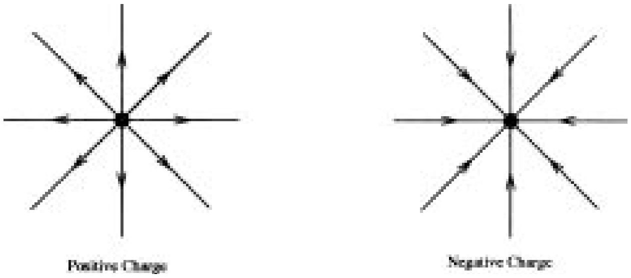

Figure 2. Electric field configurations of point particles.

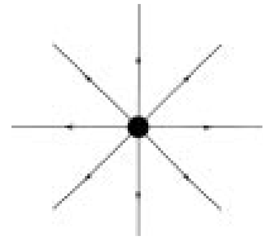

Figure 3. Magnetic field configurations of a magnetic monopole.

theory that describes a weak force. More precisely, calculations involving strong forces in one theory can be obtained from calculations involving weak forces in another theory related by the symmetry. Hence, the duality could play a prominent role in understanding the strong and weak interactions in nature.

On the other hand, the lack of experimental evidence of magnetic monopoles has led to their blind rejection by many people who are not really familiar with the underlying physics and mathematics. Outside the realm of theoretical physics, the absence of evidence has been mistranslated as an evidence of absence, and the current educational system reflects this unfortunate fact. For example, it is typically taught in high school or introductory physics classes that magnetic monopoles do not exist. Not surprisingly, we have discovered through a short survey that the response of the general public, when asked on the subject of magnetic monopole, dominantly falls into two categories: "What is a magnetic monopole?" and "It doesn't exist." It thus seems that many people have been wrongly taught what to think by the absolute prejudgment of the skeptics on the matter and have been misled into thinking that they know what they do not know. It is clearly illogical to argue absolutely that magnetic monopoles do not exist merely based on the absence of evidence. After all, most profound aspects of nature are not manifest, and open-mindedness and unceasing curiosity are what have allowed the astonishing progress in the 20th century physics.

Consequently, the main aim of this paper is two-fold. First, we will further explain what magnetic monopoles are in order to establish a common ground of understanding for possible debates. Then, we will address the question of how magnetic monopoles, *if* they exist, can be described in theoretical physics, leading us to the subject of electric-magnetic duality. At the same time, we will try to convey the theoretical reasons in favor of their existence. It should be noted that the ultimate goal of our exposition is not to convince the reader of the existence of magnetic monopoles but to make him or her aware of the *possibility* of their existence.

This paper is organized as follows. The next section discusses the electric-magnetic duality in classical electrodynamics. We then explain Dirac's quantum mechanical construction of monopole states and the quantization condition. In following sections, we discuss magnetic monopoles as classical soliton solutions in relativistic field theory and explain the Montonen-Olive duality conjecture. We conclude with brief comments on *S*-duality.

Since we do not assume much background in physics, many explicit calculations will have to be replaced by vague words. Even though our treatment would be inevitably incomplete and somewhat cursory, we hope that this paper would serve as an elementary guide to the curious minds who are not conversant in theoretical physics but are open to new suggestions.

Classical Electrodynamics

> *Scientists tend to overcompress, to make their arguments difficult to follow by leaving out too many steps. They do this because they have a hard time writing and they would like to get it over with as soon as possible.... Six weeks of work are subsumed into the word "obviously."*
>
> —Sidney Coleman

It is always difficult to explain a specific phenomenon that arises in a particular theory to a person who is unfamiliar with the theory itself. Hence, in order to preclude any unnecessary confusion, we begin each section with a general survey on various branches of modern physics.

Each description of our world is characterized by certain parameters.The moon revolving around the earth is characterized by a long distance scale since both the size of the moon and the earth-moon distance are large. On the other hand, the electron-proton system of a hydrogen atom is characterized by short distance. The two most pertinent parameters in the development of physics have been speed and length. It seems that there is no one branch of physics that is useful at all scales of speed and length. Broadly speaking, we can divide our theories into four categories, each being applicable to physics at different scales: non-relativistic classical mechanics (NRCM), relativistic classical mechanics (RCM), non-relativistic quantum mechanics (NRQM), and quantum field theory (QFT). Their relevant scales are summarized in Figure 4. We will see how magnetic monopoles arise in each theory.

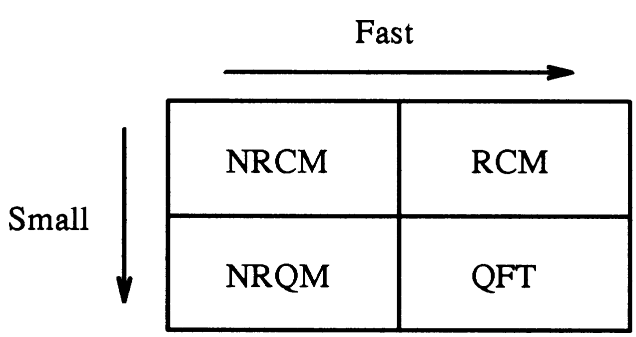

Figure 4. Four regimes of physics.

Most readers will be familiar with non-relativistic classical mechanics from their everyday experience. Examples of NRCM, which is applicable at low speed and long distance scales, are Newtonian mechanics and elementary electromagnetism. By low speed, we mean low compared to the speed of light which is c = 3 x 10^8 meters/second, and by long, we mean long compared to the atomic scale which is ordinarily about 10^{-10} meters. To describe the trajectory of a baseball thrown upwards against the gravity, Newton's 2nd law is sufficient.

Electric-Magnetic Duality: Marriage of Electricity and Magnetism. In classical electrodynamics, the fundamental quantities are the electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$. Electric charges both create and respond to electric fields, so that two charges interact because one charge responds to the field created by the other, and vice versa. Basically, all electromagnetic problems can be reduced to finding the electric and magnetic fields for given sources and boundary conditions. Even more fundamentally, all electromagnetic effects can be derived from a set of eight differential equations known as the Maxwell equations, which are over 100 years old. The following Maxwell equations for a vacuum without sources possess an interesting symmetry:

$$\begin{aligned} \nabla\cdot\mathbf{E} &= 0, \\ \nabla\cdot\mathbf{B} &= 0, \\ \nabla\times\mathbf{E}+\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} &= 0, \\ \nabla\times\mathbf{B}-\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} &= 0. \end{aligned} \tag{1}$$

Anyone[4] staring at the above equations can probably see that the above equations are symmetric under the exchange of $\mathbf{E}$ and $\mathbf{B}$. More precisely, they are invariant under

$$\mathbf{E}\rightarrow\mathbf{B} \quad \text{and} \quad \mathbf{B}\rightarrow-\mathbf{E} \tag{2}$$

This symmetry is called the electric-magnetic duality, and the exchange of electric and magnetic fields in Equation 2 is known as the duality transformation. The duality has the following physical interpretation: In classical physics, a vacuum is an empty space without any particles. The electric-magnetic duality simply implies that a theory that describes a vacuum consisting only of the electric and magnetic fields, $\mathbf{E}_1$ and $\mathbf{B}_1$ respectively, has the same physical interpretation as another theory that describes a vacuum with the electric field $\mathbf{E}_2 = \mathbf{B}_1$ and the magnetic field $\mathbf{B}_2 = -\mathbf{E}_1$. In particular, the energy densities are the same, i.e.,

$$\frac{1}{8\pi}\left|\mathbf{E}_1\right|^2+\frac{1}{8\pi}\left|\mathbf{B}_1\right|^2=\frac{1}{8\pi}\left|\mathbf{E}_2\right|^2+\frac{1}{8\pi}\left|\mathbf{B}_2\right|^2,$$

and the electromagnetic waves propagating in the two vacuo are identical. As we shall see, generalization of this dual description of the same physics by two distinct theories has a profound consequence in modern theoretical physics.

Unfortunately, the above symmetry *seems* to be spoiled in nature by the fact that we clearly have electric charges but have not yet observed any magnetic charges. To understand the statement, assume that there are electric charge density ρ_e and current density $\mathbf{J}_e$ but no corresponding magnetic counterparts. The Maxwell equations then become:

$$\begin{aligned} \nabla\cdot\mathbf{E} &= 4\pi\rho_e, \\ \nabla\cdot\mathbf{B} &= 0, \\ \nabla\times\mathbf{E}+\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} &= 0, \\ \nabla\times\mathbf{B}-\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} &= \frac{4\pi}{c}\mathbf{J}_e. \end{aligned} \tag{3)-(6}$$

The equations are no longer symmetric under the duality transformation. Equations 4 and 5 seem to be missing something on their right hand sides. To see exactly what they are missing, we need to explain the meaning of $\nabla\cdot\mathbf{B}$, also called the divergence of $\mathbf{B}$ or simply div$\mathbf{B}$. Let V be a volume enclosed by a surface S in space. $\nabla\cdot\mathbf{B}$ integrated over the volume V gives 4π times the total amount of magnetic charge g contained in V.[5] Similarly, $\nabla\cdot\mathbf{B}$ evaluated at point x gives 4π times the magnetic charge density at x. Hence, Equation 4 says that there is no magnetic charge at any point in space. Roughly speaking, moving charges are equivalent to currents. But because the above Maxwell equations assume that there is no magnetic charge, there is no magnetic current $\mathbf{J}_m$ on the right hand side of Equation 5. Hence, the absence of magnetic charge ruins the duality. In physics jargon, we say that the absence of magnetic charge breaks the symmetry.

In order to maintain the electric-magnetic duality, we need to weaken our assumption. That is, we now assume that magnetic monopoles may exist, but we just have not been able to observe them experimentally. We thus modify the previous Maxwell equations by putting in the magnetic charge and current densities, ρ_m and $\mathbf{J}_m$:

$$\begin{aligned} \nabla\cdot\mathbf{E} &= 4\pi\rho_e, \\ \nabla\cdot\mathbf{B} &= 4\pi\rho_m, \\ \nabla\times\mathbf{E}+\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} &= \frac{4\pi}{c}\mathbf{J}_m, \\ \nabla\times\mathbf{B}-\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} &= \frac{4\pi}{c}\mathbf{J}_e. \end{aligned} \tag{7)-(10}$$

The above equations now look more symmetric, but the symmetry is not entirely apparent. With a moment of thought, we see that our new Maxwell equations are left unchanged under the following duality transformations

$$\mathbf{E}\rightarrow\mathbf{B} \quad ; \quad \mathbf{B}\rightarrow-\mathbf{E} \tag{11}$$

$$\rho_e,\mathbf{J}_e\rightarrow\rho_m,\mathbf{J}_m \quad ; \quad \rho_m,\mathbf{J}_m\rightarrow-\rho_e,-\mathbf{J}_e \tag{12}$$

Physically, the duality transformation exchanges the roles of the electric and magnetic fields. Since charges are the sources and responders to the fields, we also need to exchange the electric and magnetic charge and current densities in order to leave the theory invariant. This duality symmetry can be understood by using a toy model. Suppose that there are two worlds—one much like our world, so call it Reality, and the other called Wonder Land. Call the electric and magnetic fields in Reality $\mathbf{E}$ and $\mathbf{B}$, respectively, and their counterparts in Wonder Land $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$. Similarly, the electric and magnetic charges in Reality are ρ_e and ρ_m and those in World Land $\tilde{\rho}_e$ and $\tilde{\rho}_m$, respectively. For simplicity, we assume that all charges are stationary so that there are no currents. Assume that Reality and Wonder Land are

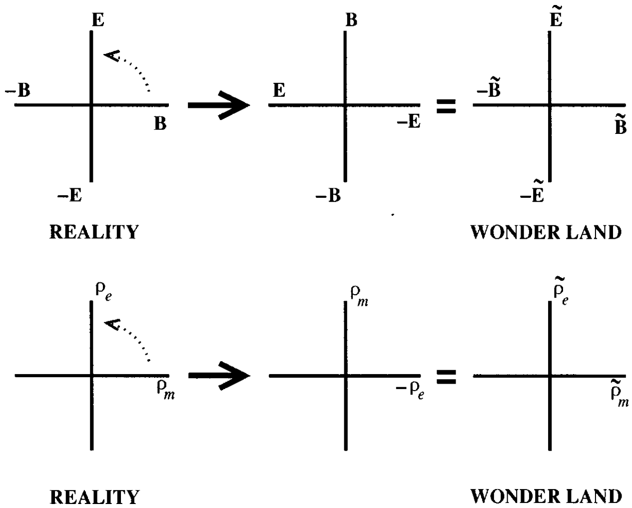

Figure 5. Picture of the dual worlds. The duality transformation amounts to a rotation by 90 degrees in the abstract field space.

related by the electric-magnetic duality. This assumption implies that $\tilde{\mathbf{E}} = \mathbf{B}$ and $\tilde{\mathbf{B}} = -\mathbf{E}$. Since charges are sources of the fields, it also implies that $\tilde{\rho}_e = \rho_m$ and $\tilde{\rho}_m = -\rho_e$. Furthermore, Reality has abundant electric charges, but magnetic charges are very rare. Wonder Land, on the other hand, has a lot of magnetic charges but very few electric charges. But remember that magnetic (electric) charges in Wonder Land become electric (magnetic) charges in Reality when we make the electric-magnetic duality transformations. Now suppose that a physicist in Reality wants to calculate the force of electric field **E** on an electron with electric charge *e*. After many sleepless nights, the physicist discovers that the electromagnetism in Reality admits a dual description. This dual description is none other than the corresponding electromagnetism in Wonder Land. Hence, to study the behavior of an electron in Reality, the physicist decides to do the calculation using Wonder Land's theory of electromagnetism. Throughout his calculation, the physicist must keep in mind that the roles of electricity and magnetism are exchanged in Reality and Wonder Land. Consequently, in Wonder Land, he calculates the force of the external *magnetic* field $\tilde{\mathbf{B}} = -\mathbf{E}$ on a magnetic monopole with *magnetic* charge $g = -e$. The electric-magnetic duality guarantees that the force he has just computed is equal to the force on the electron in Reality. Hence, the two theories of electromagnetism in Reality and Wonder Land provide dual descriptions of the same physics. One can use either theory to get the same answer as long as he remembers that the electricity and magnetism exchange roles. Figure 5 compactly summarizes the relevant points.

This extremely fascinating phenomenon of duality has a highly non-trivial and unexpected generalization in field theory. To prepare the reader for the upcoming discussion,we now digress to introduce the theory of special relativity.

Special Theory of Relativity: A Lightning Introduction. When the objects in which we are interested move at a speed comparable to the speed of light, simple Newtonian mechanics cannot be applied. Classical mechanics must be modified by the special theory of relativity when the relevant particles move at high speeds, and the resulting theory is the relativistic classical mechanics, the subject of the second box in Figure 4. The theory of special relativity was first proposed by Einstein. One of the most popularized equation in physics, namely $E = mc^2$, is a consequence of this theory. Special relativity is formulated based on Einstein's two postulates:

Postulate 1. *The laws of physics are the same to all inertial observers.*

Postulate 2. *The vacuum speed of light is the same to all inertial observers.*

An observer traveling at constant speed is said to be an inertial observer. The first postulate says that two inertial observers traveling at different speeds agree on the physical laws that describe any given phenomenon. The second postulate appears to contradict the human intuition. It states that the speed of light in vacuum is the same for both a stationary observer and an inertial observer who is traveling at *any* given constant speed. Suppose that I ride on an idealized Harvard Shuttle Bus traveling at very high, but constant, speed in the direction of light and that my twin brother[6] has missed the bus and stands still on the ground. We both measure the speed of light, which is in the direction of the Shuttle Bus. According to normal human intuition, the speed that I measure should be slower than that measured by my brother. Nevertheless, the second postulate correctly asserts that the speed must be the same for both of us.[7] From the two postulates, many consequences can be derived. For example, moving objects appear to be contracted and moving clocks seem to run slower with respect to a stationary clock.

Special relativity is the theory that has been most accurately tested by experiments, and no deviation from the predictions of the theory is known to date. A physical theory that incorporates special relativity in a consistent manner is said to be Lorentz invariant. The special theory of relativity is also the most treasured and upheld principle in physics, and the requirement of Lorentz invariance imposes severe restrictions on relativistic mechanics. In modern physics, we reject all relativistic theories that are not Lorentz invariant. It is truly amazing that the Maxwell equations, which were written down before the development of special relativity, are Lorentz invariant.

This concludes our brief digression on the subject. In the next section, we will concentrate on the non-relativistic quantum mechanical description of magnetic monopoles.

Quantum Mechanics and Dirac Monopole

Non-relativistic quantum mechanics describes physics at low speed and a short distance scale. In quantum mechanics, the distinction between particle and wave becomes blurry. Waves, which were classically thought to be distinct from particles, gain particle-like interpretations, and in turn, particles begin to behave like waves. So is an electron a particle or wave? The answer is both; electrons behave sometimes like particles and sometimes like waves. The same statement holds true for other particles and waves, and this mysterious phenomenon is referred to as particle-wave duality. The electromagnetic waves also gain a particle-like interpretation, and the associated particle is known as a photon.

Because of the limitation of space, we end our introduction to quantum mechanics with a short list of important features that characterize the theory. We hope that this survey will make the reader feel less intimidated by our subsequent discussions. We omit many important aspects which we consider unnecessary for the purpose of understanding this paper. A few defining characteristics of quantum mechanics are:

(1) *Everything* that we want to know about the behavior of a particular particle can be obtained from a single function associated with the particle in given external conditions. The function is called a wave-function, and changing the external conditions changes the wave-function. Despite the deceptive simplicity of the idea, finding the exact wave-function is impossible in most cases.

(2) The world looks discrete quantum mechanically. That is, when the world is viewed at the atomic scale, many quantities that we observe become discrete. A familiar example is the atomic spectra of hydrogen atoms, displaying the discrete energy levels of the atoms.

(3) The physical quantities that we can observe in experiments are called observables. Roughly speaking, each observable becomes an operator that acts on the wave-function. An operator acts on a function to get another function. By quantization of classical systems we mean this correspondence between observables and operators. This operator nature of observables accounts for the discreteness of the world in atomic scale.

Dirac Monopole and Charge Quantization. There is one fundamental problem in describing magnetic monopoles in quantum mechanics. In the last section, we noted that the fundamental quantities in classical electrodynamics are the electric and magnetic fields. In quantum mechanics, on the other hand, the electric and magnetic fields do not provide a complete description of the electromagnetic effects on the wave-functions of the charged particles. Hence, the fundamental quantities in quantum mechanical formulation of electrodynamics are *not* the electric and magnetic fields but another vector field called vector potential $\mathbf{A}$ and a function called scalar potential ϕ. One of the *identities* in vector calculus tells us that

Identity 1. *For all vector fields* $\mathbf{A}$, $\nabla \cdot (\nabla \times \mathbf{A}) \equiv 0$.

We emphasize that Identity 1 is true for *all* $\mathbf{A}$. Thus, if write $\mathbf{B}$ as

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{13}$$

then Equation 4 in the previous section is automatically satisfied. $\mathbf{A}$ in Equation 13 is the vector potential that plays the fundamental role in quantum mechanics. On the other hand, recall that Equation 8 states that

$$\nabla \cdot \mathbf{B} = 4\pi\rho_m \tag{8}$$

where ρ_m is the magnetic charge density. This equation tells us that $\mathbf{B}$ *cannot* be written as $\nabla \times \mathbf{A}$, because $\mathbf{B} = \nabla \times \mathbf{A}$ implies that $\nabla \cdot \mathbf{B} = (\nabla \times \mathbf{A}) = 0$ by Identity 1, producing a contradiction:

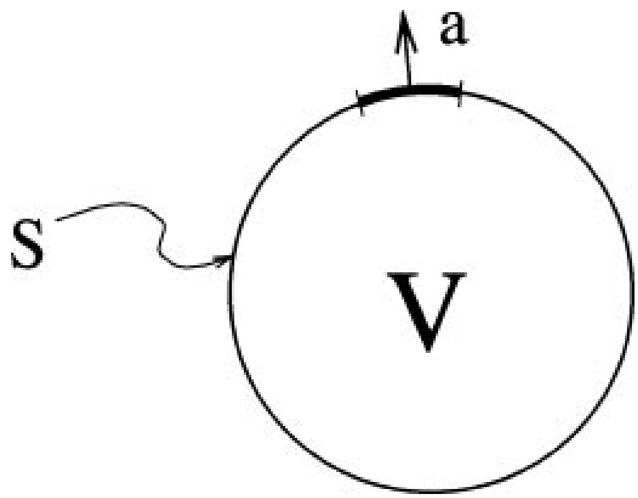

Figure 6. Cross section of a three-dimensional volume *V* bounded by a closed surface *S*.

$$\text{We want } \nabla \cdot \mathbf{B} = 4\pi\rho_m, \text{ but}$$
$$\mathbf{B} = \nabla \times \mathbf{A} \text{ implies } \nabla \cdot \mathbf{B} \equiv 0 \neq 4\pi\rho_m \Rightarrow \text{CONTRADICTION}$$

Hence it appears that Equation 14, which implies the presence of magnetic charge, forbids us from using Equation 13, which gives us the vector potential $\mathbf{A}$. But as we just noted, we *need* $\mathbf{A}$ in quantum physics. One of the reasons that we need the vector potential is that without it, we cannot describe an electron in a magnetic field. It thus appears that we must reject magnetic monopoles in quantum mechanics in order to maintain the vector potential, which must appear in the form of Equation 13.

Nevertheless, in 1931, P. A. M. Dirac showed that it is indeed possible to have both magnetic charge and the vector potential in quantum mechanics, and he derived an unexpectedly pleasant result. Let us first explain how he circumvented the problem for a point-like magnetic monopole, as in Figure 3. From the last section, recall that $\nabla \cdot \mathbf{B}$ integrated over a volume V gives 4π times the amount of magnetic charge enclosed in the volume. Hence, for a point-like magnetic monopole and *any* three-dimensional volume V enclosing the monopole, Equation 14 is equivalent to:

$$\int_V \nabla \cdot \mathbf{B} dV = 4\pi g, \tag{15}$$

where g is the magnetic charge of the monopole. Before we go on, we state a theorem from vector calculus.

Theorem 1 (Divergence Theorem). *Let V be a three-dimensional volume bounded by a closed surface S. Then, for a vector function* $\mathbf{F}$,

$$\int_V \nabla \cdot \mathbf{F} dV = \int_S \mathbf{F} \cdot d\mathbf{a},$$

where $d\mathbf{a}$ *is an infinitesimal area element pointing out of the surface S. See Figure 6.*

The theorem states that we can transform the original integral over the volume V into a new integral over the surface S bounding V. This is possible if we have that funny symbol[8] $\nabla \cdot$ in front of the vector function $\mathbf{F}$.

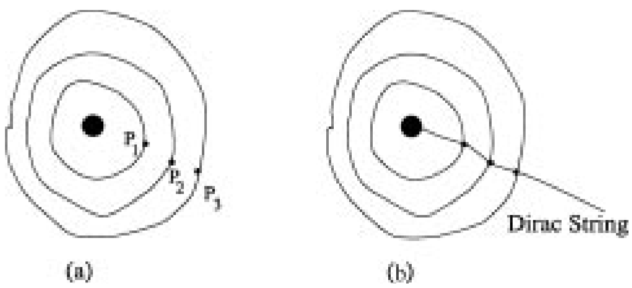

Figure 7. Dirac String. (a) *SOMETHING* has to be infinite at one point P_i on any surface bounding an arbitrary volume that contains the monopole. **(b)** Since the magnetic field **B** is a continuous function of space, we can join the singular points into a line connecting the monopole to infinity.

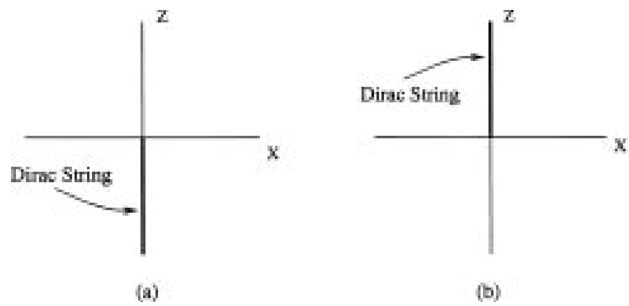

Figure 8. Two vector potentials. Since the vector potential is singular along the Dirac String, we need two vector potentials to describe electrons in the magnetic field of the magnetic monopole. **(a)** $\mathbf{A}_1$, which is singular along the negative z-axis, is used when the electron is away from the negative z-axis. **(b)** $\mathbf{A}_2$, which is singular along the positive z-axis, is used when the electron is near the negative z-axis.

Our task is to find the vector potential **A** such that the magnetic field **B** can be written in the form of Equation 13 as much as possible and, at the same time, Equation 15 is satisfied. Hence, in the spirit of Equation 13, we can decompose **B** into two parts:

$$\mathbf{B} = \nabla \times \mathbf{A} + SOMETHING. \qquad (16)$$

By Identity 1, the first term in the above expression does not contribute when it is substituted into the integral in Equation 15:

$$\int_V \nabla \cdot \mathbf{B} dV = \int_V \nabla \cdot (SOMETHING) dV = 4\pi g. \qquad (17)$$

Using the Divergence Theorem, we can rewrite it as

$$\int_S \nabla \cdot (SOMETHING) d\mathbf{a} = 4\pi g. \qquad (18)$$

Since our goal is to write out the magnetic field **B** in terms of $\nabla \times \mathbf{A}$ as closely as possible, we must let *SOMETHING* vanish in most places such that $\mathbf{B} = \nabla \times \mathbf{A}$ almost everywhere. (Remember that *SOMETHING* cannot vanish everywhere since Equation 18 says the integral over the surface *S* must not be zero.) Dirac argued that we can judiciously chose *SOMETHING* such that it vanishes everywhere on the surface *S* expect at one point *P* where it is infinite. *SOMETHING* must be infinite at *P* for the following reason. Suppose that *SOMETHING* is zero everywhere on the surface *S* except at one point where it is finite. It can be proven mathematically that the integral of such a function over the surface is zero, which contradicts Equation 18. Hence, on the surface *S*, we have

$$\mathbf{B} = \nabla \times \mathbf{A} + (SOMETHING \text{ which is infinite at point } P \text{ and zero elsewhere}). \qquad (19)$$

Now, recall that the volume *V* was arbitrarily chosen. That is, Equation 15 holds for any *V* that contains the magnetic monopole. Then, in the above argument, *SOMETHING* must be infinite at one point on each surface bounding an arbitrary *V*. This implies that *SOMETHING* must be infinite on a line connecting the monopole to infinity, as can be clearly seen in Figure 7. This line of singularity is called the Dirac String. Since the magnetic field is a physical quantity that we can measure in a laboratory, it should not be infinite at any point. The regularity of the magnetic field in Equation 19 implies that the vector potential **A** has to be infinite or singular along the Dirac String to cancel the infinity of *SOMETHING*. Hence, in the presence of a magnetic monopole, the vector potential cannot be defined everywhere. Thus, to describe the physics of magnetic monopoles, we must use two vector potentials $\mathbf{A}_1$ and $\mathbf{A}_2$ which are related by a transformation. See Figure 8. Our approach is less elegant than Dirac's original than Dirac's original formulation, but the result is the same. This monopole with a line of singularity is called the Dirac monopole.

We now state a remarkable consequence of this construction without proof. Dirac has shown that the existence of magnetic monopoles explains the quantization of electric charge. In nature, all electric charges seem to appear as integral multiples of the electron's charge. If we call electron's charge *e*, then all electric charges that we find in nature can be written as *en*, for some integer *n*. This peculiar property of the charges is known as the quantization of electric charge. Prior to Dirac, no one could explain this phenomenon. Using magnetic monopoles, Dirac unexpectedly found a possible explanation for the charge quantization. More precisely, the Dirac quantization condition says that in the presence of magnetic monopoles, the product of electric and magnetic charges must be an integral multiple of 1/2. In equations, all electric and magnetic charges in nature, e_i and g_j respectively, must satisfy:

Theorem 2 (Dirac Quantization).

$$e_i g_j = (1/2) n_{ij}, \text{ for some integer } n_{ij}. \qquad (20)$$

We emphasize that the Dirac quantization condition must hold for *all* magnetic and electric charges in nature. This is possible only if there exist basic units of electric and magnetic charges, e_0 and g_0 respectively, such that all charges are integral multiples of them:

$$\text{For all } e_i \text{ in nature, } e_i = n_i e_0 \quad (n_i \text{ is an integer}). \qquad (21)$$

$$\text{For all } g_j \text{ in nature, } g_j = n_j g_0 \quad (n_j \text{ is an integer}). \qquad (22)$$

Furthermore, e_0 and g_0 are unique up to sign and they satisfy the Dirac quantization condition themselves.

$$e_0 g_0 = (1/2) n_0, \text{ for some integer } n_0. \qquad (23)$$

Thus, the existence of magnetic monopoles provides an explanation for the fact that electric charges in nature are integral multiples of the electron's charge. Ignoring the quarks for the moment, we can let e_0 be equal to the charge of the electron.

Strong-Weak Coupling Duality: First Child. We now discuss the profound implication of electric-magnetic duality in conjunction with the Dirac quantization condition. Recall that the duality transformation exchanges the electric and magnetic fields by exchanging the corresponding charges:

Electric charge e —> Magnetic charge g
Magnetic charge g —> -(Electric charge e) (24)

Now, using the Dirac quantization condition, we get

Electric charge e —> Magnetic charge $g = (1 / 2e)$
Magnetic charge g —> -(Electric charge e) = $-(1 / 2g)$, (25)

where we let $n = 1$ for simplicity. But from the Dirac quantization condition we see that as the quantity e approaches zero, $g = (1 / 2e)$ approaches infinity, and vice versa. Hence, since the strength of charge determines the strength of the force involved, the electric-magnetic duality together with the Dirac quantization condition implies that there is a symmetry that exchanges the strong and weak forces. In physics jargon, we say that the duality reverses the strong and weak coupling constants. In nature, there exists a nuclear force which is strong and a weak force by which particles decay. If a strong-weak coupling duality similar to the electric-magnetic duality exists in nature, then the strong and weak forces could be merely different manifestations of the same thing. In that case, the human intellect would reveal one of nature's hidden beauties, giving us the understanding of one of many mysteries that underlie our existence. We devote the next section to generalizing the strong-weak coupling duality to relativistic field theory.

Classical Field Theory and Solitons

The very concept of law of nature reflects a terminology which appears to be the heritage of a normative metaphor rooted in the ancestral image of a universe ruled by God. Maybe, a deeper concept which underlies more basically our way of thinking in physics is that of symmetry, which moves us from the normative or legal metaphor to the belief that beauty is the closest to truth.

—César Gómez

The beauty of physics lies in our description of nature as well as in nature itself. Beauty often reveals itself as a symmetry or duality in our theories, serving as a guiding principle for physicists. Nowhere in theoretical physics is the concept of symmetry more pronounced than in classical and quantum field theory. Quantum field theory (QFT) is an incorporation of special relativity into quantum mechanics and forms the fourth box in Figure 4. As the figure suggests, QFT describes the atomic world at high speed. It should be noted that special relativity itself is a space-time symmetry. Recall that in quantum mechanics, we quantized the classical systems by turning the observables into operators and assigning wave-functions to particles. In QFT, we quantize the wave-functions, i.e., we turn the wave-functions into operators. These operators are called quantum fields because they satisfy the defining characteristics of fields discussed in the second section. Namely, they are defined at every point in space-time and are often vectors in an abstract vector space. There is a one-to-many correspondence between particles and fields. That is, for each particle, we can find many fields describing the particle, but we usually choose the one that is most convenient. Intuitively, it is clear that particles traveling at high speed have high energies. Hence, QFT provides a quantum mechanical description of physics at very high energy. When energy gets high, many surprising things can happen. A photon, which is the particle associated with the electromagnetic fields, can transform into an electron-positron pair. The positron is the anti-particle of the electron, with the same mass but opposite quantum numbers.[9] It thus seems possible that photons can create a magnetic monopole-anti-monopole pair. There are several reasons for not observing such a creation in the laboratory, two of which are:

(1) The calculated attractive force between a magnetic monopole and its anti-particle is much greater than that between an electron and a positron. It is consequently much more difficult to create the monopole-anti-monopole pair in the laboratory.

(2) Calculations predict that magnetic monopoles are superheavy. Heavy particles require high energy to be produced, and magnetic monopoles may be too heavy to be produced in current high energy accelerators.

Classical field theory (CFT) is the long distance limit of QFT, and thus, it is the limit in which operators become ordinary functions and numbers. Hence, the quantum fields become ordinary fields much like the electric and magnetic fields in classical electrodynamics, and we do not have to worry about subtleties, such as removing infinities from physical quantities, that arise in quantum physics. We say that there is a symmetry[10] when the Lagrangian[11] of our system has the same form under certain transformations on the fields. In field theory, symmetry *governs* the dynamics by restricting the form of the Lagrangian from which all relevant equations and interactions are derived. An example of symmetry transformations is multiplication of the fields by a complex number. When we make different transformations at different space-time points, we must introduce a new field in order to maintain the symmetry. This new field is called the gauge field and is responsible for the interactions among various particles. For example, when we multiply the fields by different complex numbers at different points in space-time and demand that the Lagrangian is left unchanged under the transformation, then we must introduce a new gauge field, and this field is nothing but the photon field, which is a combination of the vector potential $\mathbf{A}$ and the scalar potential ϕ. Furthermore, from the Lagrangian, we can derive the equations of motion for the fields, and in the example just mentioned, the equations are indeed the Maxwell equations. We now state a beautiful theorem without proof:

Theorem 3 (Noether's Theorem). *For every symmetry, there is a conserved quantity.*

What is the conserved quantity in the above example of multiplication by complex numbers? It is the electric charge! In general, the conserved quantities are the charges that

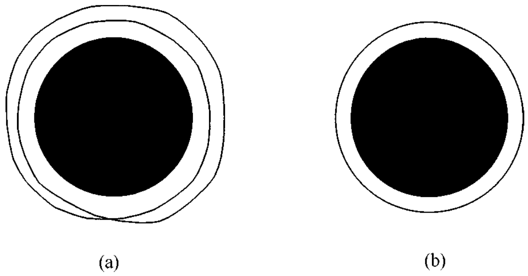

Figure 9. Topological charge. The number of times of winding around the disk is conserved in *two* dimensions. (a) cannot be deformed into (b) without breaking the curve or going to higher spatial dimensions.

correspond to different forces in nature. We sometimes call such charges Noetherian charges.

There is another type of conserved charge called the topological charge. Topological charge is not a consequence of some continuous symmetry, but rather it arises from topological considerations. For example, consider the map from a circle onto another circle in two dimensions. The number of times that the first circle winds around the second circle is conserved because we cannot change the number unless we break the circle in some way. Equivalently, consider wrapping around a solid disk with a rubber band in two dimensions. The number of times that the rubber band winds around the disk is conserved because the number cannot change unless we tear the rubber band apart or go to higher spatial dimensions. This argument is clear from Figure 9. In general, topological charges are conserved quantities arising from the fact that we cannot deform two maps, or field configurations, continuously into each other.

Montonen-Olive Duality Conjecture: Second Child. We are now ready to discuss the electric-magnetic duality conjecture by Montonen and Olive. In certain field theories, magnetic monopoles arise as solitons. Classically, solitons are extended objects with smeared out energy or mass densities, whereas particles are point-like objects with localized point-like mass densities. A soliton is interpreted as a composite particle which is a bound state of two or more elementary particles. Hence, in some field theories, we can have magnetic monopoles which are solitons composed of more elementary particles. Magnetic monopoles as solitons were first constructed by Polyakov and 't Hooft. Surprisingly, magnetic and electric charges have different origins in field theory. As previously mentioned, electric charges are the conserved quantities under a continuous symmetry. On the other hand, 't Hooft and Polyakov have shown that magnetic charges appear as topological charges. Furthermore, the electric and magnetic charges satisfy the Dirac quantization condition.

In 1977, Montonen and Olive conjectured that the field theory which was considered by 't Hooft and Polyakov contains the electric-magnetic duality in the following sense. Under the electric-magnetic duality transformation, the original field theory becomes a dual theory, with possibly different forms of Lagrangian and interactions, such that:

(1) Electric and magnetic charges exchange roles. That is, the magnetic charge, which is a topological charge in the original field theory, becomes a Noetherian charge in the dual theory.

(2) In the original field theory, the magnetic monopoles are solitons, and the gauge fields are elementary particles. In the dual theory, magnetic monopoles become the elementary gauge particles, and solitons carry electric charge which is now topological.

(3) Since the electric and magnetic charges satisfy the Dirac quantization condition, the duality transformation exchanges the strong and weak couplings.

Montonen and Olive conjectured that this "dual" theory is in fact exactly the original theory with just the charges and the fields relabeled. Even though they have not been able to prove the conjecture, their work has provided a significant basis for the subject of *S*-duality, which is one of the most active research areas in today's high energy particle theory. We will now make a few remarks on *S*-duality.

S-*Duality: The Next Generation.* *S*-duality is a generalization of the electric-magnetic duality to include more symmetries. Recall that the electric-magnetic duality has only two symmetry transformations—namely, the identity transformation, which does absolutely nothing to the theory, and the duality transformation, which exchanges the charges. In mathematical notations, the electric-magnetic duality acts on electric charge e as

$$e \to e \qquad \text{or} \qquad e \to \frac{1}{2e}.$$

S-duality generalizes these two transformation to a group of transformations consisting of infinite elements. More precisely, *S*-duality asserts that certain field theories are unchanged under all transformations for the form

$$\tau \to \frac{a\tau + b}{c\tau + d} \tag{27}$$

a, b, c, d are integers and ad - bc = 1,

where τ is a generalized charge. The set of all such transformations is called the modular or $SL(2,\mathbf{Z})$ group. If *S*-duality is a true symmetry of nature, then we have an infinite number of theories that are equivalent to each other and which are all related by the modular group. Further investigation of this subject would be beyond the scope of our presentation.

Conclusion

In this paper, we have seen that magnetic monopoles are certainly possible in physics. The existence of magnetic charge in quantum mechanics provides an explanation for charge quantization in nature. Generalization of the electric-magnetic duality to classical field theory leads to many surprising results, combining seemingly unrelated topics such as solitons, strong and weak forces, and charge quantization into one unifying theory. Despite the many attractive features, the present theory does not require magnetic monopoles to exist. Likewise, it does not forbid them to exist. In fact, it is very hard to believe that magnetic monopoles which lead to extremely profound theories such as *S*-duality are mere accidents in our attempts to understand nature.

Symmetries in physics are almost always hidden and often defy what is apparent. In this paper, we have tried to show the reader the hidden symmetry of nature that the physics of magnetic monopoles may reveal someday. The development of the theory of *S*-duality is still in its rudimentary stage, and a lot of work needs to be done to unravel the mysteries of the physical world.

The author regrets that he couldn't discuss the last section to his heart's content due to the lack of time and space. An extensive study of gauge field theory and *S*-duality using the language of algebraic topology will be treated elsewhere by the author.[12]

Acknowledgments

I would like to thank Professor Arthur Jaffe for getting me interested in the subject of *S*-duality and for useful discussions.

No pretense of originality is made in this paper. Most things that we present can be found in the literature, but probably in a form that the reader might not understand.

I also would like to thank C. Ahn for proofreading the preliminary version and for enlightening comments.

And finally, I heartily thank Emanuela and Jurij Striedter, the Co-Masters of Cabot House, for their hospitality during my stay at Cabot, and I would like to dedicate this work in honor of Jurij Striedter's 70th birthday. I regret the fact that my feeble power allows me to return their kindness only in this form. I wish them a happy and peaceful retirement.

References

(1) A vector has both a magnitude and a direction in space.

(2) We avoid using the term "electric charge" because the two definitions hold true for other forces of fundamental interaction. For example, mass is a source for the gravitational field and also responds to the field. The heavier the mass, the stronger the gravitational field and force. Hence, mass can be viewed as a gravitational charge.

(3) Throughout this paper, $\mathbf{E}$ denotes the electric field, $\mathbf{B}$ the magnetic field. Bold-faced letters represents vectors and $|\ \ |$ represents the magnitude of a vector.

(4) We feel that knowledge of vector calculus is totally unnecessary for the purpose of understanding the underlying symmetry. A reader who is not familiar with vector calculus may think of such marks as $\boldsymbol{\nabla}$ as $\nabla\cdot$ as being abstract symbols, or as different ways of writing derivatives.

(5) This is true in general for any vector field $\mathbf{A}$ that satisfies the inverse square law, in particular for the gravitational field.

(6) The author is not lying.

(7) This is true if we assume that the air is replaced by vacuum and that my brother and I are wearing really expensive space-suits.

(8) We have been intentionally avoiding the actual discussion of vector calculus for the same reason that we don't think that a person has to know the meaning and name of individual playing cards in order to sort a deck of cards into piles of the same suit. The Divergence Theorem is merely a useful device that allows us to throw away the $\nabla\cdot$ notation without the need to worry what it means.

(9) Roughly, quantum numbers are conserved quantities that characterize each particle. An example would be electric charge.

(10) By symmetry, we really mean continuous symmetries that can vary smoothly. For example, if the symmetry is multiplication by a number, we should be able to vary that number continuously to other numbers and the symmetry should still hold.

(11) Lagrangian is equal to the kinetic energy *minus* the potential energy of the system and is expressed in terms of the fields.

(12) Song, J. S. 1996. *Geometry of* S*-Duality and Gauge Theories* (A. B. Thesis, Harvard University).

Uncited References

Cheng, T. P., and L. F. Li. 1992. *Gauge Theory of Elementary Particle Physics* (Oxford University Press).

Coleman, S. 1975. "Quantum sine-Gordon equation as the massive Thirring model." *Phys. Rev.* D11: 2088.

Dirac, P. A. M. 1931. "Quantised singularities in the electromagnetic field." *Proc. Roy. Soc.* A133: 60.

Goddard, P., and D. Olive. 1978. "Magnetic monopolies in gauge field theories." *Rep. Prog. Phys.* 41: 1357.

Goddard, P., et al. 1977. "Gauge theories and magnetic charge." *Nucl. Phys.* B125: 1.

Jackson, J. D. 1975. *Classical Electrodynamics* (John Wiley & Sons).

Montonen, C., and D. Olive. 1977. "Magnetic monopoles as gauge particles?" *Phys. Lett.* 72B: 117.

Song, J. S. 1995. *Gauge Field Theories, Supersymmetry, and Self-Dual Einstein-Yang-Mills Equations* (Harvard University Archive).

Witten, E., and D. Olive. 1978. "Supersymmetry algebras that include topological charges." *Phys. Lett.* 78B: 97.

Witten, E. *On* S*-Duality in Abelian Gauge Theory* (hep-th/9505186).

Wu, T. T., and C. N. Yang. "Concept of nonintegrable phase factors and global formulation of gauge fields." *Phys. Rev*. D12: 3845.