An Elementary PRIMER For GAUGE THEORY

K. Moriyasu

An Elementary Primer For GAUGE THEORY

K. Moriyasu

Senior Research Physicist
University of Washington, Seattle

World Scientific

World Scientific Publishing Co Pte Ltd
P O Box 128
Farrer Road
Singapore 9128

ISBN 9971-950-83-9
9971-950-94-4 pbk

Printed in Singapore by Richard Clay (S. E. Asia) Pte. Ltd.

PREFACE

The understanding of nuclear and elementary particle physics has now reached a historical turning point. During the last decade, a revolution has quietly occurred – a revolution called "Gauge Theory". For the first time in 50 years, since the birth of modern nuclear physics, gauge theory allows us to understand how the fundamental forces of nature may be unified within a single coherent theory. The discovery of gauge theory rivals in importance the development of both relativity and quantum mechanics. In contrast to the situation less than 10 years ago, gauge theory now dominates nearly all phases of elementary particle physics today. Even the reasons for performing new experiments are now judged by their relevance for testing the predictions of gauge theory.

Clearly, such an exciting development should be widely accessible and understandable not only to theoreticians but also to experimental physicists, students and the "intelligent layman" as well. Like politics and war, gauge theory has become too important to be left only to the experts. Unfortunately, for the reader who wishes to first understand the basic physical ideas behind gauge theory, the published literature can present a daunting challenge. The reason for the difficulty is that gauge theory represents a totally new synthesis of quantum mechanics and symmetry ideas which have been applied to the entire field of elementary particle physics.

I believe that gauge theory can be appreciated by the non-expert; that is the raison d'etre for this primer. In order to emphasize the physics of gauge theory rather than the mathematical formalism, I have used a new intuitive approach and designed the text primarily for the reader with only a background in quantum mechanics. My goal in this primer is to hopefully leave the reader with an appreciation of the elegance and beauty of gauge theory.

This book was motivated by my own desire as a "non-expert" to learn something about gauge theory. Over a period of 4-5 years, I wrote a series of short pedagogical articles on gauge theory topics for the American and European Journals of Physics. These articles allowed me to test the ideas and the writing style for this primer. I also found that trying to satisfy the high standards of the referees for these journals encouraged me to develop much clearer explanations for many gauge theory topics. I am indebted to these referees who do their work in anonymity.

K. Moriyasu

Seattle,
July, 1983

CONTENTS

CHAPTER I

INTRODUCTION

> *. . . the best reason for believing in a renormalizable gauge theory of the weak and electromagnetic interactions is that it fits our preconceptions of what a fundamental field theory should be like.*
>
> S. Weinberg, 1974[1]

Modern gauge theory has emerged as one of the most significant and far-reaching developments of physics in this century. It has allowed us for the first time to realize at least a part of the age old dream of unifying the fundamental forces of nature. We now believe that electromagnetism, that most useful of all forces, has been successfully unified with the nuclear weak interaction, the force which is responsible for radioactive decay. What is most remarkable about this unification is that these two forces differ in strength by a factor of nearly 100000. This brilliant accomplishment by the Weinberg-Salam gauge theory, and the insight gained from it, have encouraged the hope that all of the fundamental forces may be unified within a gauge theory framework. At the same time, it has been realized that the potential areas of application for gauge theory extend far beyond elementary particle physics. Although much of the impetus for gauge theory came from new discoveries in particle physics, the basic ideas behind gauge symmetry have also appeared in other areas as seemingly unrelated as condensed matter physics, non-linear wave phenomena and even pure mathematics. This diversity of interest in gauge theory indicates that it is in fact a very general area of study and not exclusively limited to elementary particles.

[1] S. Weinberg, *Rev. Mod. Phys.* 46, 255 (1974).

In this primer for gauge theory, our purpose is to present an elementary introduction which will provide an adequate background for appreciating both the new theoretical developments and the experimental investigations into gauge theory. We have therefore adopted a very general pedagogical approach which should be useful for very different areas of physics. Like any new topic in physics, the study of gauge theory requires some familiarity with background material from other areas of physics and mathematics. Gauge theory represents a new synthesis of quantum mechanics and symmetry. At the same time, it is also a direct descendent of quantum electrodynamics; thus much of the published literature on modern gauge theory is written in the language of renormalizable quantum field theory which has proven so useful in electrodynamics.

In this primer, we have adopted the point of view that it is possible to learn the fundamentals of gauge theory by using a much simpler semiclassical approach. By semiclassical, we mean Maxwell's electromagnetism and old-fashioned Schrödinger quantum theory where the electromagnetic field is not second quantized. By using such an approach, we can emphasize the new physics of gauge theory without the added technical complexities of quantum field theory. A limitation of our approach is that we cannot discuss the problems associated with the quantization of gauge theory in any rigorous fashion. However, since these problems are among the most subtle and difficult in gauge theory, we feel that they can best be studied separately in more advanced treatments such as the excellent review of Abers and Lee.[2]

One essential requisite for the study of gauge theory is at least a nodding acquaintance with some of the terminology of group theory. The heart of any gauge theory is the gauge symmetry group and the crucial role that it plays in determining the dynamics of the theory. Fortunately, much of the necessary group theory is already familiar to physics students from the treatment of angular momentum operators in quantum mechanics. The essential difference in

[2] E. Abers and B. Lee, *Phys. Rep.* 9C, 2 (1973).

gauge theory is that the symmetry group is not associated with any physical coordinate transformation in space-time. Gauge theory is based on an "internal" symmetry. Therefore, one cannot speak of angular momentum operators, but must replace them with the more abstract concept of group generators. This is more than a mere change of labels because the generators have mathematical properties which were previously ignored in quantum mechanics but are very useful in gauge theory. In particular, we will see that the proper understanding of gauge invariance leads naturally to a geometrical description of gauge theory that is both highly intuitive and strongly resembles the familiar geometrical picture of general relativity. By exploiting this geometrical feature of gauge theory, we can often find much simpler interpretations of complicated physical phenomena such as gauge symmetry breaking, which is one of the most important ingredients of the Weinberg-Salam theory.

This primer is generally organized into three sections. The first section consisting of Chapters II through V introduces the concept of gauge invariance and describes the essential ingredients and physical assumptions which go into the building of a general gauge theory. We begin in Chapter II with the original inspiration of Hermann Weyl and briefly review why gauge theory was re-discovered three times in different physical contexts before the correct interpretation of gauge invariance was finally understood. In Chapter III, the geometrical interpretation of gauge symmetry is discussed and simple arguments are used to motivate and derive the essential mathematical building blocks of gauge theory. In Chapter IV, the familiar case of electromagnetism is used as a pedagogical guide for the construction of the Yang-Mills theory. The canonical Lagrangian formalism is introduced and the equations of motion are derived and discussed. The non-Abelian versions of Maxwell's equations are presented in Chapter V and compared with electromagnetism. The unique problems caused by non-linearity and lack of superposition in Yang-Mills gauge theories are discussed.

The second section of this primer deals with the new description of the electromagnetic, weak and strong forces as gauge theories.

We begin in chapter VI by briefly reviewing the salient experimental and theoretical features of the weak nuclear interaction which lead to the idea of a gauge theory. In chapter VII, we introduce the general formalism for understanding how gauge symmetry is broken by the Higgs mechanism. Several physical examples are discussed in detail to illustrate the dynamical mechanism responsible for symmetry breaking in different applications. In Chapter VIII, we present a simple introduction to the Weinberg-Salam theory of the unified weak and electromagnetic interactions. The procedure for unifying the weak and electromagnetic gauge symmetry groups is discussed in detail. Symmetry breaking of the weak interaction is introduced and the masses of the gauge vector are derived. In chapter IX, we present a brief introduction to the basic physical ideas in the color gauge theory of strong interactions. A simple intuitive argument is given for the new phenomena of "asymptotic freedom". By using an analogy between the vacuum of color gauge theory and a dielectric medium, it is shown how the so-called "running coupling constant" can be obtained.

The third section of this primer provides an introduction to some of the new "non-perturbative" features of gauge theory. In chapter X, we present a simple study of monopoles and vortices and explain how their properties can be understood as topological features of gauge theory.

In the appendix, we briefly summarize some of the key group theory terminology used in this primer.

CHAPTER II

THE REDISCOVERY OF GAUGE SYMMETRY

. . gauge invariance has no physical meaning, but must be satisfied for all observable quantities in order to ensure that the arbitrariness of A *and* ϕ *does not affect the field strength.*

Röhrlich, 1965[1]

2.1 Introduction

Gauge invariance was recognized only recently as the physical principle governing the fundamental forces between the elementary particles. Yet the idea of gauge invariance was first proposed by Hermann Weyl[2] in 1919 when the only known elementary particles were the electron and proton. It required nearly 50 years for gauge invariance to be "rediscovered" and its significance to be understood. The reason for this long hiatus was that Weyl's physical interpretation of gauge invariance was shown to be incorrect soon after he had proposed the theory. Gauge invariance only managed to survive because it was known to be a symmetry of Maxwell's equations and thus became a useful mathematical device for simplifying many calculations in electrodynamics. In view of the present success of gauge theory, we can say that gauge invariance was a classical case of a good idea which was discovered long before its time.

In this chapter, we present a brief historical introduction to the discovery and evolution of gauge theory. The early history of gauge theory can be divided naturally into old and new periods where the dividing line occurs in the 1950's. In the old period, we will return to Weyl's original gauge theory to gain insight into

[1] F. Röhrlich, *Classical Charged Particles* (Addison Wesley, Reading, Mass., 1965).
[2] H. Weyl, *Ann. Physik* 59, 101 (1919).

several key questions. The most important question is what motivated Weyl to propose the idea of gauge invariance as a physical symmetry? How did he manage to express it in a mathematical form that has remained almost the same today although the physical interpretation has altered radically? And, how did the development of quantum mechanics lead Weyl himself to a rebirth of gauge theory?

The new period of gauge theory begins with the pioneering attempt of Yang and Mills[3] to extend gauge symmetry beyond the narrow limits of electromagnetism. Here we will review the radically new interpretation of gauge invariance required by the Yang-Mills theory and the reasons for the failure of the original theory. By comparing the new theory with that of Weyl, we can see that many of the original ideas of Weyl have been rediscovered and incorporated into the modern theory.

2.2 The Einstein Connection

In 1919, only two fundamental forces of nature were thought to exist – electromagnetism and gravitation. In that same year, a group of scientists also made the first experimental observation of starlight bending in the gravitational field of the sun during a total eclipse[4]. The brilliant confirmation of Einstein's General Theory of Relativity inspired Hermann Weyl to propose his own revolutionary idea of gauge invariance in 1919. To see how this came about, let us first briefly recall some basic ideas involved in relativity.

The fundamental concept underlying both special and general relativity is that there are no absolute frames of reference in the universe. The physical motion of any system must be described relative to some arbitrary coordinate frame specified by an observer, and the laws of physics must be independent of the choice of frame.

In special relativity, one usually defines convenient reference frames which are called "inertial", i.e. moving with uniform velocity.

[3]C. N. Yang and R. L. Mills, *Phys. Rev.* 96, 191 (1954).
[4]H. von Kluber, *Vistas in Astronomy* 3, 47 (1960).

For example, consider a particle which is moving with constant velocity ν with respect to an observer. Let S be the rest frame of of the observer and S′ be an inertial frame which is moving at the same velocity as the particle. The observer can either state that the particle is moving with velocity ν in S or that it is at rest in S′. The important point to be noted from this trivial example is that the inertial frame S′ can always be related by a simple Lorentz transformation to the the observer's frame S. The transformation depends only on relative velocity between particle and observer, not on their positions in space-time. The particle and observer can be infinitesimally close together or at opposite ends of the universe; the Lorentz transformation is still the same. Thus the Lorentz transformation, or rather the Lorentz symmetry group of special relativity, is an example of "global" symmetry.

In general relativity, the description of relative motion is much more complicated because one is dealing with the motion of a system in a gravitational field. For the sake of illustration, let us consider the following "gedanken" exercise for measuring the motion of a test particle which is moving through a gravitational field. The measurement is to be performed by a physicist in an elevator. The elevator cable has broken so that the elevator and physicist are falling freely[5]. As the particle moves through the field, the physicist determines its motion with respect to the elevator. Since both particle and elevator are falling in the same field, the physicist can describe the particle's motion as if there were no gravitational field. The acceleration of the elevator cancels out the acceleration of the particle due to gravity. This is a simple example of the principle of equivalence, which follows from the well-known fact that all bodies accelerate at the same rate in a given gravitational field (e.g. 9.8 m/sec^2 on the surface of the earth).

Let us now compare the physicist in the falling elevator with the observer in the inertial frame in special relativity. It might appear that the elevator corresponds to an accelerating or "non-inertial" frame that is analogous to the frame S′ in which the particle

[5]P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hill, New York, 1946).

appeared to be at rest. However, this is not true because a real gravitational field does not produce the same acceleration at every point in space. As one moves infinitely far away from the source, the gravitational field will eventually vanish. Thus, the falling elevator can only be used to define a reference frame within an infinitesimally small region where the gravitational field can be considered to be uniform. Over a finite region, the variation of the field may be sufficiently large for the acceleration of the particle not to be completely cancelled.[a]

We see that an essential difference between special and general relativity is that a reference frame can only be defined "locally" or at a single point in a gravitational field. This creates a fundamental problem. To illustrate the difficulty, let us now suppose that there are many more physicists in nearby falling elevators. Each physicist makes an independent measurement so that the path of the particle in the gravitational field can be determined. How are the individual measurements to be related to each other? The measurements were made in separate elevators at different locations in the field. Clearly, one cannot perform an ordinary Lorentz transformation between the elevators. If the different elevators were related only by a Lorentz transformation, the acceleration would have to be independent of position and the gravitational field could not decrease with distance from the source.

Einstein solved the problem of relating nearby falling frames by defining a new mathematical relation known as a "connection". To understand the meaning of a connection, let us consider a 4-vector A_μ which represents some physically measured quantity. Now suppose that the physicist in the elevator located at x observes that A_μ changes by an amount dA_μ and a second physicist in a different elevator at x' observes a change dA'_μ. How do we relate the changes dA_μ and dA'_ν? In special relativity, the differential dA_μ is also a vector like A_μ itself. Thus, the differential dA'_ν in the

[a] A strongly varying gravitational field gives rise to "tidal" forces which can produce some unusual effect For example, see the science fiction story *Neutron Star* by L. Niven (Ballantyne, New York, 1968).

the elevator at x' is given by the familiar relation

$$\mathrm{d}A'_\nu = \frac{\partial x^\mu}{\partial x'^\nu}\,\mathrm{d}A_\mu \qquad \text{(II - 1)}$$

where, according to the usual convention[b], the repeated index μ is summed over the values $= 0, 1, 2, 3$. The simple relation (II - 1) follows from the fact that the Lorentz transformation between x and x' is a linear transformation. What happens in general relativity? We can no longer assume that the transformation from x to x' is linear. Thus, we must write for $\mathrm{d}A'_\nu$ the general expression

$$\begin{aligned}\mathrm{d}A'_\nu &= \frac{\partial x^\mu}{\partial x'^\nu}\,\mathrm{d}A_\mu + A_\mu\,\mathrm{d}\left(\frac{\partial x^\mu}{\partial x'^\nu}\right) \\ &= \frac{\partial x^\mu}{\partial x'^\nu}\,\mathrm{d}A_\mu + A_\mu\,\frac{\partial^2 x^\mu}{\partial x'^\nu\,\partial x'^\lambda}\,\mathrm{d}x'^\lambda \quad . \qquad \text{(II - 2)}\end{aligned}$$

Clearly, the second derivatives $\partial^2 x^\mu/\partial x'^\nu \partial x'^\lambda$ will vanish if the x^μ are linear functions of the x'^ν.

How do we interpret the physical meaning of the extra term in (II - 2)? Such terms are actually quite familiar in physics. They occur in "curvilinear" coordinate systems. For example, suppose that two physicists are located on a circular path at the positions x, y and $x' = x + \mathrm{d}x$, $y' = y + \mathrm{d}y$ as shown in Fig. (2-1). The curved path could be the equator of the earth. Using the familiar curvilinear coordinates:

$$x = R\cos\phi\,, \qquad y = R\sin\phi \quad , \qquad \text{(II - 3)}$$

it can be easily seen that the differentials $\mathrm{d}x$ and $\mathrm{d}y$ depend on the coordinates x and y. Now suppose that the physicist at x, y measures

[b] The components of the 4-vector $A^\mu = (A^0, \mathbf{A})$ and $A_\mu = (A_0, \mathbf{A})$ with $A^0 = -A_0$. Vector components with upper and lower indices are related by $x_\mu = g_{\mu\nu}\, x^\nu$, where $g_{\mu\nu}$ is the metric tensor which appears in the definition of the invariant space-time interval $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu$. The components of $g_{\mu\nu}$ are $g_{11} = g_{22} = g_{33} = 1$, $g_{00} = -1$, and all other components are zero.