CHAPTER I

INTRODUCTION

The purpose of this thesis is to establish the fundamental field equations through a logical process conventionally called the deduction approach to science. Unfortunately, this kind of process may not be accustomed to most physicists thus it is worth giving some short discussion at the outset. After that, the preliminary concepts of fields and the theories of electromagnetic fields and quantum gauge fields will be presented respectively in order to complete this chapter.

The Fundamental Approaches to Science (Lightman,1992)

In order to discover the laws of nature, a scientist always uses two logical approaches: the *inductive* and the *deductive*. In deductive science, the scientist performs a number of observations of nature, finds a pattern and then formulates the law or organizing principle. Ultimately, the law will be tested against future experiments. The construction of the Kepler's law for motion of the planets is an example of an inductive science. Kepler has used the heliocentric model of Copernicus and the positional observations of Tycho Brahe painstakingly accumulated over 20 years to deduce his famous three empirical laws of planetary motion (Zeilik and Smith,1987).

In deductive science, the scientist begins by postulating certain truths of nature, with only little guidance from outside experiments, and deduces the consequences of the postulates. The consequences are cast into predictions, which can then be pitted against observational tests. We see that inductive reasoning works from the bottom up, deductive from top down. Of course, all scientific theories of nature, whether they be arrived at by induction or deduction, must be abandoned if they are found to be wrong.

Although the deductive is rare, many twentieth-century physicists have used it successfully. Albert Einstein has founded his ingenious theory, the special theory of relativity, upon only two basic postulates, the form invariance of all physical laws and the constancy of the speed of light in empty space, which are applicable in all inertial frames of reference. Similarly, the modern theory of electrons, framed by Paul Dirac in the 1920s, was founded upon Dirac's intuition and love of mathematical beauty, not on observation of how electrons behaved. The unified theory of the electromagnetic force and the weak nuclear force, formulated by Sheldon Glashow, Abdus Salam, and Steven Weinberg in the 1960s, was built upon a sense of unity of nature, not upon the detailed trajectories of particles in atom smashers (Zee,1986).

The Preliminary Concepts of Fields

The concept of a *field* was originally developed by Euler for hydrodynamics but it was first introduced to physics by Michael Faraday about 140 years ago to give an explanation for electric and magnetic phenomena. Faraday had rejected a concept of action-at-a-distance and proposed instead that the interaction between charged particles should be mediated by a field which is the continuous function of spatial coordinates (Holliday and Resnick,1988). His account was pictorial and intuitive as we show in Fig.1. for the case of the electrostatic fields. After the first introduction by Faraday, the concepts of field have been widely accepted. They have also been generalized from being purely spatial to being spatio-temporal and are used not only for theory of electromagnetism but throughout physics, and in particular to account for gravitation in the General Theory of Relativity.

In the language of physics, fields can be discriminated into two kinds: the *classical fields* and the *quantum fields.* The classical field is a kind of tension or stress which can exist in empty space in the absent of matter. It reveals itself by producing *forces*, which act on any material object that happens to lie in the space of the field occupied (Dyson,1953). The two important classical fields are the electromagnetic field and gravitational field. In difference, the quantum field is more difficult to visualize than the classical field because, in quantum field theory, a field is assigned quantum mechanical properties and seen to be associated with a particular type of particle. It is always used to describe the properties of particle interactions. Actually, we are convinced, by quantum gauge field theories, to believe that all fields in nature are quantum fields. A classical field is just a special large scale manifestation of a quantum field. At this stage, most physicists believe that all physical fields may be described by the same principle called the unified field theory (Salam,1980).

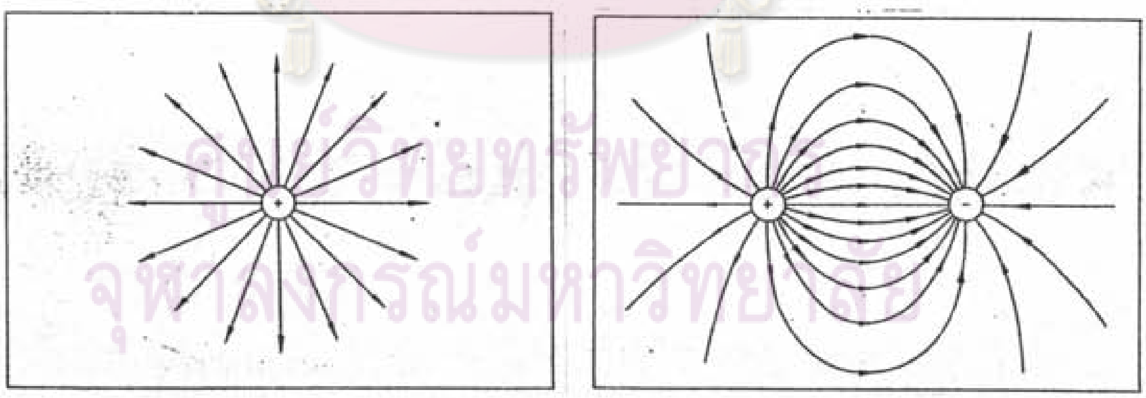

Fig.1 Field lines of electrostatic fields

In the present study, field will be treated mathematically as a differentiable function from a space or space-time manifold into a set of abstract entities, which may be the set of non-negative real numbers or the set of all real numbers. There are *scalar fields*. The abstract entities may also be *vectors*, which in an n-dimensional space can be represented by a set of n real numbers. *Tensors* are a further generalization of vectors, and can, for a natural number m, greater than 1, be represented for an n-dimensional space, by an n^m array of real numbers. We can see that a field is a function on a continuous space,

$$f : R^n \rightarrow R^{(nm)}$$

for some non-negative integer m. If m=0, it is a scalar field; if m=1, it is a vector field; if $m \geq 2$; it is a tensor field. (Lucas,1984)

Theory of Electromagnetic Fields

Electromagnetic field is the most well-known field in nature. Its details have completely elaborated since the end of nineteenth-century after the development for over a hundred years. The historical approach to the field can be shown as follows (Jackson,1975).

At the early time, electric and magnetic theories were investigated individually. The first quantitative studies of the electric phenomena were made by Charles A. Coulomb, on electrostatics. In 1784 Coulomb has established the force law for stationary charged particles by using a torsion balance (Serway,1990). He showed that the force **F** exerted between two small charged bodies q_1, located at point $\mathbf{x}_1$, and q_2 , located at point $\mathbf{x}_2$, can be written as

$$\mathbf{F} = Kq_1q_2 \frac{(\mathbf{x}_1-\mathbf{x}_2)}{|\mathbf{x}_1-\mathbf{x}_2|^3} \tag{1.1}$$

This equation is generally called *Coulomb's force law* and q_1 and q_2 are algebraic quantities which can be positive or negative. The constant of proportionality K depends on the system of units we used.

At this moment the *electric field* can be defined as the force per unit charge acting at a given point. It is a vector function of position, denoted by **E**. Therefore, from Eq.(1.1), we can rewrite as

$$\mathbf{F} = q\mathbf{E}, \tag{1.2}$$

where

$$\mathbf{E}(\mathbf{x}) = Kq_1 \frac{(\mathbf{x}-\mathbf{x}_1)}{|\mathbf{x}-\mathbf{x}_1|^3}. \tag{1.3}$$

We interpret this as the charge q_1, located at $\mathbf{x}_1$, generates the vector field **E** of which value at any point **x** is defined by Eq.(1.3). If the charges are so small and so numerous that they can be described by a charge density $\rho(\mathbf{x})$, then electric field $\mathbf{E}(\mathbf{x})$ can be written in the integral form,

$$\mathbf{E}(\mathbf{x}) = \int_V \frac{\rho(\mathbf{x}')(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^3}\, d^3x', \tag{1.4}$$

where $d^3x' = dx'dy'dz'$ is the three volume element at $\mathbf{x}'$ and the constant K is omitted according to the unit we preferred. This equation is the integral form of the Coulomb's law in Eq.(1.1). It is more suitable form for the evaluation of electric fields by using the so-called Gauss's law in three-space which states

$$\oint_s \mathbf{E}\cdot\mathbf{n}\, da = 4\pi \int_v \rho(\mathbf{x})\, d^3x, \qquad (1.5)$$

and the divergence theorem,

$$\oint_s \mathbf{v}\cdot\mathbf{n}\, da = \int_v \nabla\cdot\mathbf{v}\, d^3x.$$

Then we can readily prove from Eq.(1.4) that

$$\nabla\cdot\mathbf{E}(\mathbf{x}) = 4\pi\rho(\mathbf{x}), \qquad (1.6a)$$

which is the differential form of Gauss's law of electrostatics. If we take the curl of Eq.(1.4) we will find that

$$\nabla\times\mathbf{E}(\mathbf{x}) = 0. \qquad (1.6b)$$

The two equations in Eq.(1.6) show the properties of electric field for static case.

Similar to the studies of electric phenomena, the knowledge of magnetic phenomena was developed from the empirical facts. However, the distinction is that there is no evidence for the existence of magnetic charge to be source of magnetic field as electric field has. The studies of magnetic fields began shortly after Oersted's discovery in 1819 that a compass needle is deflected by a current-carrying conductor, Jean Baptiste Biot and Felix Savart had reported that a conductor carrying a steady current produces a force on magnet. From their experimental results, Biot and Savart were able to arrive at an expression for the *magnetic field* **B** at point **x** produced by electric current density **j** at point $\mathbf{x}'$ in the following convenient form

$$\mathbf{B}(\mathbf{x}) = \frac{1}{c}\int_v \mathbf{j}(\mathbf{x}')\times\frac{(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^3}\, d^3x' \qquad (1.7)$$

where c is the speed of light in vacuum. This equation is generally known as the *Biot-Savart law* of magnetostatics. By taking the divergence and curl of Eq.(1.7), we will obtain the laws for magnetostatics (Griffths,1989),

$$\nabla \cdot \mathbf{B}(\mathbf{x}) = 0, \tag{1.8a}$$

$$\nabla \times \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c}\mathbf{j}(\mathbf{x}). \tag{1.8b}$$

The combination of electric and magnetic fields was arisen from the study on time-dependent electric and magnetic fields made by Michael Faraday in 1831. Faraday discovered that these two fields also exert effects on each other. He found that a changing magnetic field produces electric forces, the effect is now known as *induction*, which can be expressed as

$$\nabla \times \mathbf{E}(\mathbf{x},t) + \frac{1}{c}\frac{\partial \mathbf{B}(\mathbf{x},t)}{\partial t} = 0. \tag{1.9}$$

This equation is called the differential form of *Faraday's law.* We also note that this is the time-dependent generalization of the statement, $\nabla \times \mathbf{E}(\mathbf{x}) = 0$, for the electrostatic fields. At this stage, the fundamental equations of electromagnetism can be summarized as follows,

Coulomb's law: $$\nabla \cdot \mathbf{E}(\mathbf{x}) = 4\pi\rho(\mathbf{x}), \tag{1.10a}$$

Ampere's law: $$\nabla \times \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c}\mathbf{j}(\mathbf{x}), \tag{1.10b}$$

Faraday's law: $$\nabla \times \mathbf{E}(\mathbf{x},t) + \frac{1}{c}\frac{\partial \mathbf{B}(\mathbf{x},t)}{\partial t} = 0, \tag{1.10c}$$

Absent of magnetic charge: $$\nabla \cdot \mathbf{B}(\mathbf{x}) = 0. \tag{1.10d}$$

These equations are written in macroscopic form and in Gaussian units. The continuity equation guaranteeing the conservation of electrical charge is

$$\frac{\partial \rho(\mathbf{x},t)}{\partial t} + \nabla\cdot\mathbf{j}(\mathbf{x},t) = 0 \qquad (1.11)$$

which requires that the charge variation in some arbitrary volume is caused by the flow of current through the surface of the volume.

The field equations in Eq.(1.10) still contain some logical inconsistency appeared in Ampere's law. This because we know from vector calculus that the divergence of curl of any vector $\mathbf{V}$ vanishes or, $\nabla\cdot(\nabla\times\mathbf{V}) = 0$, then, it follows from Ampere's law for steady currents that,

$$\nabla\cdot\mathbf{j} = \nabla\cdot(\nabla\times\mathbf{B}) = 0. \qquad (1.12)$$

But form the continuity equation Eq.(1.11), $\nabla\cdot\mathbf{j} = 0$ holds only if the charge density ρ is constant in time. It was, from a logical point of view, no a priori reason to expect that the static equations hold unchanged for time-dependent fields. However, it was James C. Maxwell who abolished this conflict. In 1864, Maxwell had modified Ampere's law to accommodate a time dependence by introducing a term called the *displacement current* to Eq.(1.10b) as follows:

$$\nabla\times\mathbf{B}(\mathbf{x}) = \frac{4\pi}{c}\mathbf{j}(\mathbf{x}) \longrightarrow \nabla\times\mathbf{B}(\mathbf{x},t) = \frac{4\pi}{c}\mathbf{j}(\mathbf{x},t) + \frac{1}{c}\frac{\partial \mathbf{E}(\mathbf{x},t)}{\partial t}. \qquad (1.13)$$

Then, the complete set of the electromagnetic field equations in Eq.(1.10), after being modified, become

Coulomb's law: $$\nabla\cdot\mathbf{E}(\mathbf{x},t) = 4\pi\rho(\mathbf{x},t), \qquad (1.14a)$$

Ampere-Maxwell's law: $$\nabla\times\mathbf{B}(\mathbf{x},t) - \frac{1}{c}\frac{\partial\mathbf{E}(\mathbf{x},t)}{\partial t} = \frac{4\pi}{c}\mathbf{j}(\mathbf{x},t), \qquad (1.14b)$$

Faraday's law: $$\nabla\times\mathbf{E}(\mathbf{x},t) + \frac{1}{c}\frac{\partial\mathbf{B}(\mathbf{x},t)}{\partial t} = 0, \qquad (1.14c)$$

Absent of magnetic charge: $$\nabla\cdot\mathbf{B}(\mathbf{x},t) = 0. \qquad (1.14d)$$

These four field equations are formed a complete set of *Maxwell equations* for time-dependent electromagnetic fields. Notice that the first two equations are now imply directly the existence of the continuity equation in Eq.(1.11). Maxwell equations together with electromagnetic force equation, or *Lorentz force law* for electrical charge q,

$$\mathbf{F} = q\mathbf{E} + \frac{q}{c}(\mathbf{v}\times\mathbf{B}), \qquad (1.15)$$

gives a complete description of all classical electromagnetic interactions.

It is often useful to eliminate the fields **E** and **B** appearing in the Maxwell equations in favor of a *vector potential* **A** and *scalar potential* Φ through the relations

$$\mathbf{B} \equiv \nabla\times\mathbf{A}, \qquad (1.16a)$$

$$\mathbf{E} \equiv -\nabla\Phi - \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t} \qquad (1.16b)$$

These definitions automatically satisfy the third and fourth of Maxwell's equations, as can be verified with the identities $\nabla\cdot(\nabla\times\mathbf{A}) = 0$ and $\nabla\times\nabla\Phi = 0$. Thus the four coupled first-order partial differential equations in Eq.(1.14) are converted to two inhomogeneous equations that are satisfied identically,

$$\nabla^2\Phi + \frac{1}{c}\frac{\partial(\nabla\cdot\mathbf{A})}{\partial t} = -4\pi\rho \tag{1.17a}$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} - \nabla\left(\nabla\cdot\mathbf{A} + \frac{1}{c}\frac{\partial\Phi}{\partial t}\right) = -\frac{4\pi\mathbf{j}}{c} \tag{1.17b}$$

where the vector identity $\nabla\times(\nabla\times\mathbf{A}) = \nabla(\nabla\cdot\mathbf{A}) - \nabla^2\mathbf{A}$ has been used. Therefore, the solution of the Maxwell equations is reduced to solving the coupled equations in Eq.(1.17) for potentials **A** and Φ. We now exploit a fundamental property of classical electrodynamics, *gauge invariance,* to decoupled these equations. The potential **A** and Φ as defined above are, indeed, not unique. The transformation of **A** and Φ that preserve the Maxwell equations are called *gauge transformation.* Because of the identity $\nabla\times\nabla\Phi = 0$, the transformation

$$\mathbf{A} \rightarrow \mathbf{A}' \equiv \mathbf{A} + \nabla\chi \tag{1.18a}$$

where χ is an arbitrary scalar function, will have no effect on **B**; to preserve **E,** in Eq.(1.16b), the scalar potential Φ must simultaneously be changed by

$$\Phi \rightarrow \Phi' \equiv \Phi - \frac{1}{c}\frac{\partial\chi}{\partial t}. \tag{1.18b}$$

These gauge transformmations, by construction, yield the same **E** and **B** fields as the potentials before the transformation, so they leave the Maxwell equations invariantly. We may use this invariance to choose a set of potential $(\Phi, \mathbf{A})$ that satisfy

$$\nabla.\mathbf{A} + \frac{1}{c}\frac{\partial\Phi}{\partial t} = 0. \tag{1.19}$$

Then, the second-order equations in Eq.(1.17) decouple to

$$\nabla^2\Phi - \frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2} = -4\pi\rho \tag{1.20a}$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\frac{4\pi\mathbf{j}}{c} \tag{1.20b}$$

and these equations may be solved independently for Φ and $\mathbf{A}$.

A procedure such as this is termed *fixing the gauge.* The particular choice of gauge defined by Eq.(1.19) is called the *Lorentz gauge* and Eq.(1.20) is termed the *Lorentz condition.* The Lorentz gauge is widely used because it leads to the two decoupled potential wave equations derived above, and because equations in the Lorentz gauge are easily written in covariant form. Another useful gauge condition is the *Coulomb gauge* where $\nabla\cdot\mathbf{A} = 0$, then the scalar potential satisfies the Poisson equation,

$$\nabla^2\Phi = -4\pi\rho$$

with solution,

$$\Phi(\mathbf{x},t) = \int_V \frac{\rho(\mathbf{x}',t)\, d^3x'}{|\mathbf{x}-\mathbf{x}'|}.$$

The scalar potential is just the *instantaneous* Coulomb potential due to the charge density $\rho(\mathbf{x},t)$. This is the origin of the name "Coulomb gauge."

Because of the logical addition of the Maxwell displacement current term in Eq.(1.10b), its role in empty space is still made some confusion. Rosser (Rosser,1976,1983) and Shadowitz (Shadowitz,1975) have pointed out that the Ampere-Maxwell's law in empty space,

$$\nabla \times \mathbf{B}(\mathbf{x},t) - \frac{1}{c}\frac{\partial \mathbf{E}(\mathbf{x},t)}{\partial t} = 0,$$

is *not* a relation between the magnetic field and one of its sources but is a relation between the two resultant field vectors **E** and **B**, valid at any field point in empty space. The Maxwell term cannot regard as the source of the magnetic field since the displacement current in empty space does not appear in the differential equations for the potentials Φ and **A** in either the Coulomb or Lorentz gauge. The only sources in Maxwell equations are the charge density ρ and the current density **j**. Therefore, it is pointless to make a statements such as "a changing electric field produces a magnetic field" as the old fashion does. This confusion will be dispersed if we consider the sources and fields in four-space rather than in the conventional three-space, as usual.

Maxwell equations have survived unscathed for over a century, particularly through a period when many of the traditional concepts and theories of physics suffered drastic changes from the invention of the special theory of relativity by Albert Einstein in 1905. Einstein founded his special theory of relativity on the basis of two postulates :

I. The form of physical laws is the same for observers in all inertial frames.
II. The speed of light is the same for observers in all inertial frames.

From these two postulates, he can derive the so-called *Lorentz transformation*, for two inertial frames moving along xx' axis (Rindler,1991),

$$x' = \gamma(x-vt), \quad y' = y, \quad z' = z, \quad t' = \gamma(t-vx/c^2) \qquad (1.22)$$

where $\gamma = 1/(1-v^2/c^2)^{1/2}$ is a constant of the transformation. This transformation was first shown by H.A. Lorentz in 1904 to be the transformation under which Maxwell equations are form invariance. Einstein also showed that Maxwell equations are form invariant under the Lorentz transformation. He also noticed that Newton's second law for an electron is not form invariant, he then proceeded to modify Newton's second law so that it is consistent with the postulates of relativity (Einstein,1905). Because of the form invariance, we could suggest that Maxwell equations should be able to written in form of tensor equations. This because the linear homogeneous Lorentz transformation in Eq.(1.22) is allowable for being a tensor transformations. (See for detail of tensor theory in Chapter II.) The manifestation of form invariance of Maxwell equations is shown in many standard texts on electromagnetism or electrodynamics, for examples, Jackson(1975) in Chapter 11, or Schwatz(1972) in Chapter 3.

The postulate of form invariance in the special theory of relativity can be applied to electrostatics, which is based on Coulomb's law Eq.(1.4), and the principle of superposition. Then, it could be generalized from the static case alone to obtain the complete set of Maxwell equations in Eq.(1.14) if the postulate that electric charge is a conserved scalar is hold true(Kobe,1986). Instead of Coulomb's law, the Biot-Savart law in Eq.(1.6) can also be used for being the initial equation to derive for the same result (Neuenschwander and Turner,1992). Similarly, Zeleny(1991) was derived Maxwell equations and Lorentz force law by using the postulate of symmetries of gauge invariance, Eq.(1.18), and the Lorentz invariance, Eq.(1.22).

Theory of Quantum Gauge Fields

Though we have learned from the special theory of relativity that all physical laws should be Lorentz invariance, which is a kinematics or geometric symmetry. Such symmetries do not, by themselves, *determine* the structure of the interactions of fields, but do help us to *map* those interaction from one space to another, e.g., from three-space to spacetime. In contrast, dynamical symmetries *do* determine interactions. Celebrated examples include general covariance and quantum gauge invariance. The former determines the gravitational field from the Lorentz-invariant theory, and the latter fixes the structure of the fundamental interactions from a knowledge of the free-field Lagrangians and symmetry groups. Now we will show that how the quantum gauge invariance works.

The *gauge principle*, which might also be described as a principle of local symmetry, is a statement about the *invariant* properties of physical laws. It requires that every continuous symmetry be a *local symmetry*, a phrase that is explained and discussed in detail later. The story of gauge principle began in 1918 when Emmy Noether has found the deeply significant relation of conservation law and continuous transformation which could be simply stated that *for every continuous symmetry of nature there is a corresponding conservation law and for every conservation law there is a symmetry.* (There is shown in Appendix A. for detail of Noether's theorem.) In 1919, Hermann Weyl used this idea to establish the unified theory of gravitation and electromagnetism. He thought that the electromagnetic theory should be also invariant under some kind of local symmetry, similar to the local symmetry that characterizes the General Theory of Relativity. The invariance that Weyl hoped to exploit was an invariance with respect to change of scale. The *scale* invariance is the requirement that physical laws be the same if the scale of all length measurements is changed by the same over all factors. Weyl wanted to require a *local gauge invariance* in which a scale change are allowed to be different at different points in space and time, analogous to the curvilinear coordinate transformations of the general relativity. The associated conservation law, corresponding to Noether's theorem, was to be the conservation of electric charge. Unfortunately, his theory did not success and Weyl later abandoned it (Weyl,1931).

It was until 1927 that Fritz London had pointed out that the symmetry associated with electric charge conservation was not a scale invariance, but a *phase* invariance, i.e., the invariance of quantum theory under an arbitrary change in the complex phase of the wave function. The invariance under a *global* phase change-multiplication of the wave function by a constant phase factor $e^{i\chi}$ -was trivial in fact; the nontrivial fact was that the existence of electromagnetic field allows a much broader kind of invariance, invariance under a *local* phase change, in which the phase factor varies arbitrary from one point to another in spacetime. It becomes an arbitrary function of x, y, z, and t, the coordinates of space-time. Weyl also played a part in this modification of his idea and continued to use the name "gauge symmetry" to described it although it was now a misnomer, since the word "gauge" historically refers to a choice of length scale, rather than to the assignment of complex phases. Weyl suggested that a principle of gauge invariance could develop a unified theory of electromagnetism and matter (Weyl, 1931), in a manner similar to his earlier unified theory of electromagnetism and gravitation (Weyl,1922).

Broadly speaking, all symmetry transformations of a theory can be classified into two categories; those that depend continuously on a set of parameters and those that correspond to some kind of reflection. Accordingly, they are known respectively as *continuous* and *discrete* transformations. Gauge theories are depended only on the continuous transformations, but continuous symmetries, whether spacetime or internal, can again be of two kinds. First, the *global transformation* of which the parameters of transformation, the global parameters, are constant. This implies that the transformation is the same at all spacetime points. In contrast, if the parameters of transformation depend on spacetime coordinates then the symmetry transformation is known as a *local transformation.* In such case *real* physical forces must be introduced to maintain the symmetry.

As an example, let us consider the time -dependent Schrödinger equation, associated with particle of charge q,

$$H\Psi(\mathbf{x},t) = \left[\frac{-\hbar^2}{2m}\nabla^2 + V(\mathbf{x})\right]\Psi(\mathbf{x},t) = E\Psi(\mathbf{x},t). \qquad (1.23)$$

Clearly, if $\Psi(\mathbf{x},t)$ is a solution of this equation, then so is $e^{iq\chi}\Psi(\mathbf{x},t)$, where χ is a constant parameter. In other words, any quantum mechanical wave function can only be defined up to a constant phase. A constant phase transformation is therefore, a symmetry of a quantum mechanical system. This kind of transformation is conserved the probability density, $\rho = \Psi^*\Psi$, of a quantum mechanical state, and, in fact, conservation of electric charge can be associated with just such a global phase transformation. The conservation of probability comes from the fact that it involves the factors Ψ^* and Ψ , so that the phase factors cancel.

Consider next a local phase transformation,

$$\Psi(\mathbf{x},t) \rightarrow e^{iq\chi(\mathbf{x},t)/c\hbar}\Psi(\mathbf{x},t), \qquad (1.24)$$

now $\chi(\mathbf{x},t)$ becomes a scalar function of spacetime coordinates, namely every point in four-space has a different phase. This transformation still gives conserved probability but it is no longer giving correct solution for time-dependent Schrödinger equation,

$$\left[\frac{-\hbar^2}{2m}\nabla^2 + V(\mathbf{x})\right]\Psi(\mathbf{x},t) = i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{x},t). \qquad (1.25)$$

This because the gradient and time derivative introduce the inhomogeneous terms

$$\begin{aligned}\nabla\left[e^{iq\chi(\mathbf{x},t)/c\hbar}\Psi(\mathbf{x},t)\right] &= e^{iq\chi(\mathbf{x},t)/c\hbar}\left[i(q/c\hbar)\left(\nabla\chi(\mathbf{x},t)\right)\Psi(\mathbf{x},t) + \nabla\Psi(\mathbf{x},t)\right)\right] \\ &\neq e^{iq\chi(\mathbf{x},t)/c\hbar}\,\nabla\Psi(\mathbf{x},t), \qquad (1.26a)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial t}\left[e^{iq\chi(\mathbf{x},t)/c\hbar}\Psi(\mathbf{x},t)\right] &= e^{iq\chi(\mathbf{x},t)/c\hbar}\left[i(q/c\hbar)\,\frac{\partial\chi(\mathbf{x},t)}{\partial t}\,\Psi(\mathbf{x},t) + \frac{\partial\Psi(\mathbf{x},t)}{\partial t}\right] \\ &\neq e^{iq\chi(\mathbf{x},t)/c\hbar}\,\frac{\partial}{\partial t}\Psi(\mathbf{x},t). \qquad (1.26b)\end{aligned}$$

Because of the appearance of inhomogeneous terms in Eq.(1.26), we see that the Schrödinger equation *cannot* be invariant under a local phase transformation.

The transformation of Eq.(1.24) can, however, be made a symmetry of Schrödinger equation if we arbitrarily introduce a modified gradient and modified derivative of time that contain a vector potential and scalar potential as follows:

$$\nabla \rightarrow \mathbf{D} \equiv \nabla - i(q/c\hbar)\mathbf{A}, \quad \frac{\partial}{\partial t} \rightarrow D^0 \equiv \frac{\partial}{\partial t} + i(q/c\hbar)\Phi \qquad (1.27)$$

In these notations, the Schrodinger equation is become

$$\left[\frac{-\hbar^2}{2m}\mathbf{D}^2 + V(\mathbf{x})\right]\Psi(\mathbf{x},t) = i\hbar\, D^0\,\Psi(\mathbf{x},t). \qquad (1.28)$$

To determine the invariant form, we also require that the vector and scalar potentials change under the transformation of Eq.(1.24) as

$$\mathbf{A} \rightarrow \mathbf{A}' \equiv \mathbf{A} + \nabla\chi(\mathbf{x},t) \tag{1.29a}$$

$$\Phi \rightarrow \Phi' \equiv \Phi - \frac{\partial \chi(\mathbf{x},t)}{\partial t} \tag{1.29b}$$

Now, the modified gradiant operator will transform under the combined change as desired, namely,

$$\begin{aligned}\left[\nabla - i(q/c\hbar)\mathbf{A}(\mathbf{x},t)\right]\Psi(\mathbf{x},t) &\rightarrow \left[\nabla - i(q/c\hbar)\mathbf{A}(\mathbf{x},t) - i(q/c\hbar)\nabla\chi(\mathbf{x},t)\right]\left(e^{iq\chi(\mathbf{x},t)/c\hbar}\,\Psi(\mathbf{x},t)\right)\\ &= e^{iq\chi(\mathbf{x},t)/c\hbar}\left[\nabla - i(q/c\hbar)\mathbf{A}(\mathbf{x},t)\right]\Psi(\mathbf{x},t)\end{aligned} \tag{1.30a}$$

and the modified time derivative becomes

$$\begin{aligned}\left[\frac{\partial}{\partial t} + i(q/c\hbar)\Phi\right]\Psi(\mathbf{x},t) &\rightarrow \left[\frac{\partial}{\partial t} + i(q/c\hbar)\Phi(\mathbf{x},t) - i(q/c\hbar)\frac{\partial \chi(\mathbf{x},t)}{\partial t}\right]\left(e^{iq\chi(\mathbf{x},t)/c\hbar}\,\Psi(\mathbf{x},t)\right)\\ &= e^{iq\chi(\mathbf{x},t)/c\hbar}\left[\frac{\partial}{\partial t} + i(q/c\hbar)\Phi(\mathbf{x},t)\right]\Psi(\mathbf{x},t).\end{aligned} \tag{1.30b}$$

These mean that local phase transformation in Eq.(1.24) will be a symmetry of the modified time-dependent Schrödinger equation

$$\left[\frac{-\hbar^2}{2m}\left[\nabla - i(q/c\hbar)\mathbf{A}(\mathbf{x},t)\right]^2 + V(\mathbf{x})\right]\Psi(\mathbf{x},t) = \left[i\hbar\frac{\partial}{\partial t} + i(q/c\hbar)\Phi(\mathbf{x},t)\right]\Psi(\mathbf{x},t), \tag{1.31}$$

provided we require the added vector and scalar potentials to transform as given in Eq.(1.29). We recognize Eq.(1.29) as a gauge transformation similar to that found in Maxwell equations Eq.(1.18). To preserve the invariance under a local phase transformation requires the introduction of additional fields. These fields are known as *gauge fields*, and lead to the introduction of definite physical forces. In the present case, $\mathbf{A}(\mathbf{x},t)$ and $\Phi(\mathbf{x},t)$ can be interpreted as the electromagnetic vector and scalar potentials respectively (Das and Ferbel, 1994). The complete treatment on this line to obtain the Maxwell equations Eq.(1.14) was shown by Kobe (Kobe,1978). He also showed that Maxwell equations could be derived by using the postulate of gauge invariance in Lagrangian and Hamiltonian mechanics as well (Kobe,1980;1981).

We find that to understand the principle of gauge principle in every case there is a characteristic logical pattern that emerges, represented graphically in Fig. 2, connecting conserved quantities, symmetries of nature, and gauge fields.

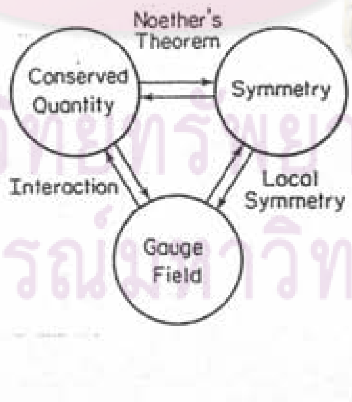

Fig. 2. The logical pattern of gauge theory (Mills,1989)

First, as we have seen, there is Noether's theorem, which states that for every conservation law there is an associated symmetry and vice versa; second, there is the fact, mentioned above, that the requirement of local symmetry leads to a gauge field theory of particular well-determined character; and third, we find that the gauge field theory determined in this way necessarily includes interactions between the gauge field and the conserved quantity with which we started. Thus we have the astonishing fact that *for every true conservation law there is a complete theory of a gauge field for which the given conserved quantity is the source.* The only restriction is that the conservation law be associated with a continuous symmetry (this would exclude, for example, parity, which is associated with reflection symmetry). The resulting theory has just one free parameter, the interaction strength. The explicit explanation on this logical pattern was given by Robert Mills (Mills,1989).

For some 25 years the idea of gauge invariance (almost always thought of in terms of local gauge invariance) was seen as a specific characteristic of electromagnetic theory but not of more fundamental significance. The idea that local gauge invariance might have a more universal significance in physics began to be considered in the early 1950s, particularly by the brilliant paper of C.N. Yang and R. Mills (Yang and Mills,1954). In 1954 Yang and Mills introduced the field that now bears their names by considering the form invariance under local rotations in isospin space of a derivative acting on a wave function. They proceeded in complete analogy with the derivation of Maxwell equations for the electromagnetic field from the principle of form invariance under local gauge transformations. Their work was extended to general gauge groups by R. Utiyama (Utiyama,1956).

With extensive studies, the principle of gauge invariance is now considered as the principle which underlies all fields of nature (Weinberg,1974). With this principle in mind, particle physicists are again actively pursuing the long sough goal of a unified field theory which would unify the strong, electromagnetic, weak, and gravitational interactions. There has been substantial progress in unifying the weak and electromagnetic interactions in name of electroweak theory (Weinberg,1980), and the strong interaction can also be brought in the Standard model theory (Georgi and Glashow,1980). Gravitational interactions still remain isolated, but the gravitational field equations have been recently derived from a gauge theory (Cho,1976). In all of those approaches the electromagnetic field is the most excellent example of a gauge theory, which serves as a paradigm for all other such theories.

Scope of Thesis

The purpose of this thesis is to establish the fundamental field equations from the continuity equation. In Chapter III, we begin by the assumption that there exists some conserved quantity, which is denoted by Q, in nature. If Q is conserved globally, then it is also conserved locally as well, the local conservation of Q implies that there exists the continuity equation for the conservation of Q, similar to Eq.(1.11) for electric charge. Because Q is a conserved scalar, it be conserved in all reference frames, through the continuity equation should have the invariant form (covariance), then it can be written in form of tensor equation. The *universal constant velocity* will be arisen naturally to preserve the form invariance of continuity equation in four-space. We then proceed on to develop the transformation under which the continuity equation is covariant. We will call this transformation *the inertial transformation* because it involves deeply on the relative motion of inertial frames. In Chapter IV, we propose that the sources will have interaction among themselves mediating by the second-rank tensor field. This tensor field is then proved, by using continuity equation, to be an antisymmetric tensor in four-space. It gives rise two real fields in three-space called the *polar field* and the *axial field*. The characteristics of these two fields can be manifested by four equations which we call the *fundamental field equations* for the conserved quantity Q. We also shown for example, if we take the speed of light in empty space to be the universal constant velocity of nature, as Einstein did in his special theory of relativity, we can readily prove that the fundamental field equations for the conservation of electric charge are exactly the well-known Maxwell equations we have already mentioned in Eq.(1.14). Finally, the conclusions and discussions are given in Chapter V. The sufficient details of tensor theory in four-space, which are being employed in Chapters III and IV, are mentioned in Chapter II.