M P Silverman 1998 Essay on Gauge Theory invited for The Physics Teacher by Editor Clifford Swartz

An essay for the 2025 International Year of Quantum Science and Technology (IYQ)

Symmetries for the Mind's Eye:
The Far-Reaching Significance of Lorentz and Gauge Invariance

M. P. Silverman

Department of Physics & Astronomy

Trinity College - 300 Summit Street

Hartford CT 06106 U.S.A.

Abstract

The principle of relativity requires that the laws of physics be invariant under a transformation from one inertial reference frame to another. Similarly, the gauge principle asserts that physical laws not change under a transformation of certain internal degrees of freedom. Implemented together, the two invariance principles are so restrictive as to determine almost uniquely the fundamental interactions between particles.

> "...Nature seems to take advantage of the simple mathematical representations of the symmetry laws. When one pauses to consider the elegance and the beautiful perfection of the mathematical reasoning involved and contrast it with the complex and far-reaching physical consequences, a deep sense of respect for the power of the symmetry laws never fails to develop."
>
> C. N. Yang [Nobel Lecture][1]

1. Introduction: Symmetry and Invariance

To those who appreciate the beauty of symmetry, nature abounds in shapes and patterns to delight the eye. Crystals, for example, afford one of the richest sources of visual symmetry in science. Who could fail to marvel at comely golden cubes of iron pyrite, vitreous rhombohedrons of calcite with their intriguing image-doubling properties, hexagonal prisms of quartz, or the infinite variety of six-sided dendritic crystals of ice that come to us gratuitously as snowflakes on a cold wintry day?

It is at the atomic or molecular level however—beyond the gross morphology accessible to our senses—that the symmetries of crystals reveal fundamental aspects of the interactions that hold matter together. There the bewildering variety of natural structures reduce to manageably small sets of symmetry elements, classifiable into different symmetry groups. Space groups, for example, characterize the various ways the lattice of an ideal crystal—one without edges or defects—can be displaced without altering the original pattern. In general, a symmetry element is a transformation that leaves a pattern invariant. To make this abstraction concrete, take a look, if possible, at the lovely Hungarian folk needlework of Reference 2 for an illustration of all the one- and two-dimensional space-group symmetries.[2]

The "lesson", if I may construe it so, of translational symmetry in crystallography is that there is no special or preferred point of origin in a crystal; the full lattice can be reconstructed from knowledge of a sufficient neighbourhood of *any* point within the crystal. This may seem an obvious point—and perhaps in the case of crystals it is. However, besides those geometric symmetries that please the eye there are more abstract mathematical symmetries in nature discernable only by the "mind's eye" (to borrow Richard Feynman's colourful imagery). The objects of interest are not physical objects like crystals, but physical laws like equations of motion. In these cases invariance under transformations analogous to translations or rotations are not trivial, but have profound implications.

Lorentz invariance and gauge invariance are two of the most consequential symmetries in physics. Together, they pose severe constraints on any candidate in the physicist's search for an ultimate "theory of everything".

2. Symmetry in Motion: The Principle of Relativity

Seated on a Boeing 747 cruising seven miles above ground at 900 km/hr, you are observing a cup of tea on the tray before you. What is it doing? In the absence of turbulence (and discounting small surface ripples produced by engine vibration), nothing at all. Like Sherlock Holmes's reference to the dog in the night[3], it is the *non*occurrence of any event that is remarkable. That cup of tea could just as well have been resting on your kitchen table at home.

According to Newton's second law, the net force on a system is proportional to its total mass and acceleration:

$$\mathbf{F} = m\mathbf{a} = m\frac{d\mathbf{v}}{dt} = m\frac{d^2\mathbf{r}}{dt^2} \tag{1}$$

Eq. (1) is a differential equation; once a mathematical expression for the force **F** is supplied and boundary conditions stipulated, the equation can (at least in principle) be solved to yield the velocity **v** or coordinate vector **r** as a function of time. Now supposing that Eq. (1) applies to a marble rolling on your kitchen table, then a transformation (called a "boost") of that table and everything on it to the constant velocity **V** of the Boeing 747 would transform the force on the marble as follows

$$\mathbf{F}' = m\mathbf{a}' = m\frac{d\mathbf{v}'}{dt} = m\frac{d(\mathbf{v}+\mathbf{V})}{dt} = m\frac{d\mathbf{v}}{dt} = \mathbf{F} \tag{2}$$

There is no change. The invariance (termed Galilean invariance) of the Newtonian force law under a boost is illustrative of a deep principle of physics: There is no single preferred frame (e.g. that of the "aether" at rest) in which to express the laws of classical mechanics. In fact, *all* physical laws are valid in all inertial (i.e. nonaccelerating) reference frames; this is the "principle of relativity."

Although the preceding principle is correct, the velocity addition [$\mathbf{v}' = \mathbf{v} + \mathbf{V}$] of Eq. (2), with which I adduced it, is not valid for velocities approaching that of light. The correct relation (called a Lorentz transformation) for comparing the spatial coordinates and time of two inertial frames was derived by Einstein in his special theory of relativity[4]

$$\begin{aligned} x' &= \gamma(x - Vt) \\ y' &= y; \quad z' = z \\ t' &= \gamma\left(t - \frac{Vx}{c^2}\right) \end{aligned} \tag{3}$$

Here the "primed" frame is moving at speed V along the +x-axis of the "unprimed" frame; c is the speed of light, and $\gamma = 1/\sqrt{1-(V/c)^2}$ is a factor that appears ubiquitously in relativistic formulas. Invariance under the space-time transformation (3) is called Lorentz invariance.

It is worth noting that Einstein did not create special relativity in response to any perceived experimental violation of Newton's laws; none was known at the time. Rather, he sought a description of motion (kinematics) consistent with the laws of electromagnetism deduced by James Clerk Maxwell in the 1840's. From these laws new symmetries for the mind's eye emerge.

3. Fields, Potentials, and Gauge

Formulated in words, the laws of electromagnetism comprise four disconnected empirical statements:

(1) The electric force between two point charges is proportional to each charge and falls off as the square of their separation [Coulomb's law].

(2) All magnets have both "north" and "south" ends [nonexistence of magnetic monopoles].

(3) A time-varying magnetic flux gives rise to an electromotive force [Faraday's law].

(4) Magnetism is generated by electric currents, both conduction currents (flow of electric charge) and displacement currents (time-varying electric flux) [Ampere-Maxwell law].

It is only when these statements are expressed mathematically in terms of electric (**E**) and magnetic (**B**) fields that they cease to be disconnected and lead to a self-consistent theory of all classical electromagnetic and optical phenomena.

The geometrical embodiment of the first statement (Coulomb's law) is that electric lines of force diverge radially and isotropically from each point electric charge. Since the second statement denies the existence of isolated north or south poles (point magnetic charges), there can be no corresponding lines of magnetic force diverging radially and isotropically from any region of space. Instead, magnetic lines of force form loops about their current sources. The algebraic consequence of the second statement is that any magnetic field—whatever its configuration—can be expressed as a particular spatial variation (the "curl") of a function designated the vector potential **A**.

$$\mathbf{B} = \nabla \times \mathbf{A} \quad \Rightarrow \quad \begin{cases} B_x = \dfrac{\partial A_z}{\partial y} - \dfrac{\partial A_y}{\partial z} \\ B_y = \dfrac{\partial A_x}{\partial z} - \dfrac{\partial A_z}{\partial x} \\ B_z = \dfrac{\partial A_y}{\partial x} - \dfrac{\partial A_x}{\partial y} \end{cases} \tag{4}$$

In terms of the coordinates (x, y, z) of a cartesian coordinate system the symbol ∇ takes the form $\nabla = \mathbf{x}(\partial / \partial x) + \mathbf{y}(\partial / \partial y) + \mathbf{z}(\partial / \partial z)$, where **x**, **y**, **z** are unit vectors along the corresponding axes, and can be used in "dot" and "cross" products with other vectors—as exemplified by the components of **B** in Eq. (4). The "dot" or scalar product, e.g. $\nabla \cdot \mathbf{B} = \dfrac{\partial B_x}{\partial x} + \dfrac{\partial B_y}{\partial y} + \dfrac{\partial B_z}{\partial z}$, is precisely what is meant mathematically by "divergence".

When the form (4) for **B** is substituted into the mathematical expression for statement 3 (Faraday's law), there results a general expression for **E** in terms of both **A** and another function called the scalar potential ϕ.

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \Rightarrow \begin{cases} E_x = -\frac{\partial\phi}{\partial x} - \frac{\partial A_x}{\partial t} \\ E_y = -\frac{\partial\phi}{\partial y} - \frac{\partial A_y}{\partial t} \\ E_z = -\frac{\partial\phi}{\partial z} - \frac{\partial A_z}{\partial t} \end{cases} \tag{5}$$

Note that the first term on the right side of (5) is not a dot product (since ϕ is not a vector); it is referred to as the "gradient" of ϕ and gives the slope of the function (along any direction) at each point, just as the name "gradient" suggests.

To an astute reader with the principle of relativity in mind, the introduction of magnetism may present a disturbing paradox. Consider the passage of a charged particle through a magnetic field. This is in fact easy to arrange experimentally. Adjust the time base of an oscilloscope so that the beam spot moves slowly across the screen; turn down the intensity control so that the spot is a small "particle"; and hold the poles of a horseshoe magnet up to the middle of the screen. The spot, advancing horizontally at uniform speed, deflects first up and then down (or vice versa, depending on pole sequence) as it moves past the magnet . Now move the magnet horizontally across the screen at the same leisurely pace as the spot. There is no deflection; the spot remains at rest relative to the magnet. A charged particle experiences a magnetic force (Lorentz force) only if it is moving.

Suppose, however, the magnet is at rest in the laboratory (together with the oscilloscope), but *you* are moving at constant velocity alongside the particle so that in your reference frame the particle is initially at rest. Theoretically there can be no Lorentz force. Would the particle deflect when the magnet moves past it? Can the existence or nonexistence of a physical force depend on the observer's reference frame? I have put these questions numerous times to physics students and teachers alike only to find them wide-eyed with surprise at this apparent "failure" of Nature to follow the simple proprieties of the principle of relativity.

The resolution of the paradox is that the charged particle does indeed swerve—but because of an *electric* force, not a magnetic force. The Lorentz transformation (3) that relates the coordinates and time of two inertial frames also intermixes electric and magnetic fields as follows

$$\begin{aligned} \mathbf{E}' &= \gamma(\mathbf{E} + \mathbf{V} \times \mathbf{B}) \\ \mathbf{B}' &= \gamma(\mathbf{B} - \mathbf{V} \times \mathbf{E}) \end{aligned} \quad . \tag{6}$$

V is the velocity of the "primed" frame relative to the "unprimed" frame in which the fields are **E** and **B**. In the preceding example of the magnet and oscilloscope, **E** is null in the laboratory, but there is a nonvanishing field **E'** in the rest frame of the observer and consequently an electric force on the charged particle. The laws of electromagnetism—explicitly, Maxwell's equations and the combined electric and magnetic force law[5]—are invariant in form under the Lorentz transformations (3), (6) of both coordinates and fields. But this is not the only invariance of electromagnetism.

Since the electromagnetic forces exerted on matter are directly proportional to the electric and magnetic fields, it should be no surprise that these fields must be (and are) uniquely determined in any well-posed system of charges and currents. The potentials **A** and ϕ, however, are not unique. If **A** and ϕ are the potentials that generate particular fields **E** and **B**, then the transformed potentials

$$\begin{aligned} \mathbf{A}' &= \mathbf{A} - \nabla\Lambda \\ \phi' &= \phi + \frac{\partial\Lambda}{\partial t} \end{aligned} \quad , \tag{7}$$

where $\Lambda(\mathbf{r})$ is an arbitrary function, also generate the same fields, as substitution into Eqs. (4) and (5) will readily confirm. Eq. (7) constitutes what is termed a gauge transformation, the word "gauge" ("Eich" in the original German designation by mathematician Hermann Weyl) conveying the notion of size or scale, just as in the gauge of a railroad track. The nonuniqueness of the potentials should likewise be no surprise; after all, the scalar potential corresponds to the "voltage" in an electrical circuit, and only differences in voltage are physically meaningful and measurable.

The invariance of Maxwell's equations under Lorentz and gauge transformations was initially regarded in each case as an interesting, but unimportant, curiosity. I am not sure whether Maxwell himself was even aware of these symmetries. In light of special relativity

(1905), Lorentz invariance was soon recognized as a criterion to be met by any admissible physical theory. But until the 1960's (with debates extending even into the 1980's), electromagnetic fields were widely regarded as fundamental, and potentials little more than auxiliary aids to calculation. Though not all physicists realized it, quantum mechanics had wrought a radical change in the understanding of classical electromagnetism.

4. Gauge Invariance and Quantum Mechanics

In contrast to classical mechanics, it is not the particle location **r** or velocity **v** that quantum theory provides, but rather an abstract complex-valued wave function $\psi(\mathbf{r},t)$ from which the probability of finding the particle within a specified region of space can be calculated. The quantum "wave", unlike waves of sound or light, does not represent the undulation of any physical medium, either material (like air) or immaterial (like an electromagnetic field). Rather, it is a purely mathematical function conveying statistical information about particles. Astonishingly, under experimental conditions such that the paths taken from source to detector are undetermined, the particles distribute themselves in patterns corresponding to the diffraction and interference of the wave function—even when the experiment is performed *one particle at a time*.[6] It is the inexplicable strangeness of apparent wavelike behaviour of quantum particles in the absence of any cooperative particle interaction that sustains a degree of mystery in a branch of physics whose fundamental principles were codified three quarters of a century ago.

The equation of motion of a free particle of mass m moving at a speed low compared to the speed of light is the Schrödinger equation

$$\left(\frac{p^2}{2m}\right)\psi = i\frac{d\psi}{dt} \tag{8}$$

Note that in quantum mechanics the linear momentum **p** and its "square" $p^2 = \mathbf{p}\cdot\mathbf{p}$ are not functions, but "operators" that take derivatives of ψ: $\mathbf{p} = -i\nabla$ and $p^2 = -\nabla^2 = -\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)$.

The fact that the state of a quantum particle is represented by a wave function rather than a trajectory leads to a symmetry requirement on the equation of motion that may at first glance seem useless, but which turns out to have extensive ramifications. Like the potentials of electromagnetic theory, the wave ψ is not unique, but can be multiplied by a factor $e^{-iq\Lambda}$, signifying an initial phase adjustment by $q\Lambda$ radians. (Why the particle charge q is included will become apparent shortly.) Only relative phase (like relative voltage) is physically meaningful; thus, in their spatial and temporal evolution both ψ and $\psi e^{-iq\Lambda}$ contain the same information about a physical system.

For constant Λ the preceding statement is obviously true, since $e^{-iq\Lambda}$ can be dropped from both sides of the Schrödinger equation (8). However, if an admissible quantum equation of motion is required to be invariant under a *local* —and not merely global—phase transformation (the gauge principle), the phase of the wave function can be reset at each and every point of space through which the particle moves. Inspection of Eq. (8) shows immediately that the spatially varying function $e^{-iq\Lambda(r)}$ does *not* drop out, the operator p^2 giving rise to first and second derivatives of Λ on the left side with no compensating terms on the right. The gauge principle can not be implemented for free particles.

It is at this point that the conceptually separate paths of quantum physics and classical electromagnetism fatefully merge. If the electromagnetic potentials are inserted into Eq. (8) according to a well-established procedure (known as "minimal coupling")

$$\left[\frac{(\mathbf{p} - q\mathbf{A})^2}{2m} + q\phi\right]\psi = i\frac{\partial\psi}{\partial t}, \tag{9}$$

then it is not difficult to confirm that the "new" Schrödinger equation (9) is unchanged in form under a gauge transformation (7) of the potentials *and* a phase transformation $\psi' = e^{-iq\Lambda(r)}\psi$ of the wave function, both transformations effected by the same arbitrary function Λ. From now on it is to be understood that the two transformations together comprise a gauge transformation. [Originally, Weyl believed (incorrectly) that Einstein's theory of gravity (general relativity) should be invariant under a *scale* change e^{Λ}—hence his adoption of the word "gauge". In quantum mechanics the appropriate invariance involves phase, not scale.]

From Eq. (9) follow all the interactions of a single nonrelativistic charged particle with electric and magnetic fields, including the Zeeman effect, Stark effect, paramagnetism, diamagnetism, emission of radiation, and radiation scattering, to cite but some of the most familiar. Although Eq. (9) was known to physicists well before gauge invariance was appreciated, its derivation on the basis of a symmetry requirement provides new insight into the nature of the electromagnetic potentials. In the context of gauge theory the potentials $\mathbf{A}$ and ϕ are no longer conceptually secondary functions, but primary functions necessary for the maintenance of a fundamental symmetry. Moreover, this fundamentality is supported by phenomena like the Aharonov-Bohm effect in which the spatial distribution of a beam of charged particles can be markedly altered by passage through a region of space where classical electromagnetic fields are *null*, but the potentials nonvanishing.[7]

5. Beyond Electromagnetism

In the historical evolution of electromagnetism, the force between elementary charges (Coulomb's law) was first inferred from experiment and eventually led (in conjunction with other empirically determined laws) to recognition of light as an electromagnetic wave. The imposition of gauge invariance and Lorentz invariance is so restrictive, however, that the chain of reasoning could have preceded in the opposite direction: insertion of "light" (i.e. $\mathbf{A}$ and ϕ) to make the quantum equation of motion gauge invariant—with subsequent *prediction* of Coulomb's law and Maxwell's equations. In gauge theory, physicists have for the first time the possibility of deducing fundamental interactions from considerations of mathematical symmetry.[8]

At the outset, the model builder seeking to account for the interactions of certain particles specifies the "internal degrees of freedom" of the particles and the group of symmetry elements that locally reset internal variables as the particles move from one space-time point to another. In electromagnetism, for example, apart from mass a particle is characterized by the sole property of electrical charge, corresponding to which is the single local variable of wave-function phase. But there are other possibilities. In the first extension of gauge theory beyond electromagnetism, C. N. Yang and R. L. Mills[9], attempting to model the strong nuclear interactions, constructed a gauge theory of the "nucleon" whose internal space had two degrees of freedom (isospin), a kind of "up" and "down" directionality (like the two components of electron spin) wherein the "up" state corresponded to the proton and the "down" state to the neutron. In this case gauge invariance was invoked to insure that the equations of motion remained unchanged in form under an arbitrary rotation of the isospin direction, a transformation with three adjustable variables.

In general, the construction of a gauge theory works more-or-less in the following way:

1. Write down in a manifestly Lorentz-invariant form a function (called the Lagrangian) that expresses the difference in kinetic and potential energies of the pertinent particles. In this form the three spatial coordinates (represented by an index $\mu = 1,2,3$ for x, y, z) and single time coordinate ($\mu = 0$ for t) are all treated on equal footing—a four-component vector x_μ—since the Lorentz transformation mixes space and time coordinates; the kinetic energy expression contains all four derivatives $\frac{\partial}{\partial x_\mu}$.

2. Generate the interactions of the particles by replacing each ordinary derivative $\frac{\partial}{\partial x_\mu}$ in the Lagrangian with a "gauge-covariant" derivative $\frac{D}{Dx_\mu}$ constructed in a prescribed manner from the gauge potentials. This replacement insures that the Lagrangian (and the equations of motion to which it gives rise) are invariant under a local gauge transformation. Eq. (9) shows that the gauge-covariant derivatives in electromagnetism are $\frac{D}{Dx} = \frac{\partial}{\partial x} - iqA_x$ (with corresponding forms for y and z) and $\frac{D}{Dt} = \frac{\partial}{\partial t} + iq\phi$. The Lagrangian will contain a set of four potentials A_μ (corresponding to ϕ, A_x, A_y, A_z) for each local variable that can be arbitrarily reset—e.g. one such variable (phase) in electromagnetism and three variables (rotation angles) in Yang-Mills theory.

3. Construct the fields from the gauge potentials by an operation that generalizes the curl expression of Eq. (4), and add to the Lagrangian a term representing the energy stored in the fields. In electromagnetism the components of **E** and **B** become the elements of a single mathematical structure, an "antisymmetric tensor" $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\nu} = -F_{\nu\mu}$, where the indices each take values 0,1,2,3, and $F_{01} = -E_x$, $F_{23} = B_x$, etc. Actually, it is the gauge-covariant derivatives that must be used to construct $F_{\mu\nu}$. In electromagnetism, both derivatives produce the same $F_{\mu\nu}$, but in Yang-Mills theory, the use of $\frac{D}{Dx_\mu}$ leads to additional terms with no electromagnetic counterparts.

4. Finally, deduce from the Lagrangian (by the methods of the calculus of variations) a set of relations (known as Euler-Lagrange equations) that prescribe the interaction of the particles with the potentials and fields, as well as the generation of the fields by particle currents. From the antisymmetry of the field tensor follow other relations among the fields themselves.

Although the procedure, as just outlined, may read like a cookie recipe and is not too difficult for a mathematical physicist to implement, the resulting equations are generally of daunting complexity. For example, in stark contrast to electromagnetism in which the fields themselves are uncharged although they are produced by charged particles, the components of Yang-Mills fields are subject to the same force as the particles and therefore interact with one another. As a consequence, the principle of superposition—so vital to the solution of electromagnetic problems—does not ordinarily apply. Also, the Yang-Mills analog to Maxwell's magnetic field **B** need not have a vanishing divergence; thus the theory leads to particles characteristic of magnetic monopoles (although none has yet been found). Clearly, this is a richer but far more difficult theory to work with than is electromagnetism.

6. Predictive Power of Gauge Invariance

Complex though it may be, gauge theory has provided the means of effecting the greatest syntheses of physical law since Maxwell. In the 1960's Weinberg and Salam, independently building on earlier work of Glashow, modeled electromagnetic interactions together with weak nuclear interactions (like beta decay of the neutron) under a single gauge-invariant mathematical framework.[10] The theory predicted new particles (carriers of the weak force) and hitherto unobserved interactions (like electron-neutrino scattering), all of which have since been confirmed in the laboratory. The three physicists received the Nobel Prize in Physics for 1979.

Further application of gauge-invariance to the hypothetical elementary constituents of the nucleus (quarks), has led to a theory of the strong nuclear interaction—the interaction that holds an atomic nucleus together despite mutual repulsion of positively charged protons. Termed "quantum chromodynamics" (in whimsical reference to quark "colour" as an attribute analogous to electrical charge), the resulting nonlinear equations of motion are exceptionally difficult to work with, although advanced analytical methods and the use of high-speed computers have led to much progress. One particular accomplishment is the prediction of a feature known as "asymptotic freedom", the *increase* in attraction between quarks as their separation increases. This would seem to account for why free quarks have not been observed—and are not expected to be observed—in deeply penetrating inelastic scattering of electrons by protons.

When, years ago, theoretical attempts to account satisfactorily for all but the electromagnetic interactions met with repeated failure, even highly imaginative physicists despairingly believed that "...in a few years the concepts of field theory will drop totally out of the vocabulary of day-to-day work in high energy physics"[11]. This never occurred. Figuratively and literally, gauge invariance provided shafts of light for exploring new and fruitful paths through the darkness.

References

[1]C. N. Yang, "The Law of Parity Conservation and Other Symmetry Laws of Physics", *Les Prix Nobel* (The Nobel Foundation, Stockholm, 1957) 393-403

[2]I. Hargittal and G. Lengyel, "The Seven One-Dimensional Space-Group Symmetries Illustrated by Hungarian Folk Needlework," *J. Chem. Educ.* **61** (1984) 1033-1034; "The Seventeen Two-Dimensional Space-Group Symmetries in Hungarian Needlework," **62** (1985) 35-36. There are 230 three-dimensional space-group symmetries (which regrettably cannot be represented by needlework).

[3]A. Conan Doyle, "The Silver Blaze" in *The Complete Sherlock Hiolmes* (Doubleday, NY, 1930) 383-401: *Inspector Gregory:* "Is there any other point to which you would wish to draw my attention?"; *Holmes:* "To the curious incident of the dog in the nighttime."; *Gregory:* "The dog did nothing in the nighttime."; *Holmes:* "That was the curious incident." [page 397]

[4]A. Einstein, "On the Electrodynamics of Moving Media", *Annalen der Physik* **17** (1905); see Einstein's readable account *Relativitty: The Special and General Theory* (Crown, New York, 1961)

[5]The four empirical laws of electromagnetism take the form: (1) $\nabla \cdot \mathbf{E} = \rho$, (2) $\nabla \cdot \mathbf{B} = 0$, (3) $\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t$, (4) $\nabla \times \mathbf{B} = \mathbf{j} + \partial \mathbf{E} / \partial t$ (in the Lorentz-Heaviside system of units); ρ is the density of electric charge, and $\mathbf{j}$ is the conduction current density. The effect of the fields on a point charge q moving at velocity $\mathbf{v}$ is given by the force $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$.

[6]See M. P. Silverman, *And Yet It Moves: Strange Systems and Subtle Questions in Physics* (Cambridge, 1993) Chaps 1-2 for discussion of electron interference.

[7]See M. P. Silverman, *More Than One Mystery: Explorations of Quantum Interference* (Springer, 1995) Chaps 1-3 for quantum effects of electromagnetic potentials.

[8]There are many technical expositions of gauge theory. A particularly readable one (at the graduate-student level) is K. Moriyasu, *An Elementary Primer For Gauge Theory,* (World Scientific, Singapore, 1983).

[9]C. N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance", *Phys. Rev.* **96** (1954) 191-195

[10]See P. C. W. Davies, *The Forces of Nature* (Cambridge, 1979) for discussion of both the weak and strong interactions.

[11]Freeman Dyson, quoted in Ref. 8, page 73.

BACKGROUND to the essay

In the 1980's and 90's I was an active contributor to the AAPT journal The Physics Teacher as an author, member of the editorial board, and for a while something of a "sounding board" to Clifford Swartz, the editor of TPT for nearly 30 years. During one phone call between us at some point in 1998, he said he would like to have an article explaining the basic idea of gauge theory and how this concept unifies physics. I suggested some notable physicists to contact, but he said that all the articles he saw so far were beyond the level of TPT. He asked if I would write one. My first book of personal researches, And Yet It Moves: Strange Systems and Subtle Questions in Physics, had been published by Cambridge a few years earlier (1993) and was receiving favourable comments for its clear discussion of complex issues, including the Aharonov-Bohm effect, which is illustrative of gauge theory in the realm of quantum physics and electromagnetism. I agreed and produced the preceding article.

The reviewers to whom the paper was sent liked the paper, but concluded that it was more suitable for another journal than TPT. The collective arguments of the reviewers were that the subject matter was too abstract, a high school teacher could not follow all of it, and a high school teacher could not use the contents in a classroom. I pointed out in response that the subject was indeed abstract, but so are all the laws of physics to one degree of another. And in any event, you knew the material is abstract; that is why you asked me to write the article. Moreover, an article of this kind does not have to be "used" in class at all. Surely there are TPT readers who have heard of the importance of gauge symmetry and simply want to know more about it. Self-instruction is an important part of education for both teachers and students. Furthermore, those who do want to know more should be able to follow the principle ideas of the paper and come away with a better understanding. After all, it is not necessary to understand everything in order to learn something. But in the end, Dr Swartz regretted that he would not be able to publish the piece.

Now—more than a quarter century later—I still believe that readers, especially those who have studied calculus, can benefit from the discussion of gauge symmetry presented in the preceding article. And since 2025 is the centennial year of the creation of quantum mechanics, I make it available as part of the activities of the International Year of Quantum Science and Technology (IYQ).

M P Silverman, G A Jarvis Prof Emeritus
Personal Website: https://mpsilverman.com
Author Website: https://www.amazon.com/stores/author/B001HMOC5O