Quantum Theory of Individual Electron and Photon Interactions: Electromagnetic Time Dilation, the Hyper-Canonical Dirac Equation, and Magnetic Moment Anomalies without Renormalization

Jay R. Yablon
Einstein Centre for Local-Realistic Physics
15 Thackley End
Oxford OX2 6LB, United Kingdom
yablon@alum.mit.edu

February 12, 2018

Abstract: Dirac's seminal 1928 paper "The Quantum Theory of the Electron" is the foundation of how we presently understand the behavior of fermions in electromagnetic fields, including their magnetic moments. In sum, it is, as titled, a quantum theory of individual electrons, but in classical electromagnetic fields comprising innumerable photons. Based on the electrodynamic time dilations which the author has previously presented and which arise by geometrizing the Lorentz Force motion, there arises an even-richer "hyper-canonical" variant of the Dirac equation which reduces to the ordinary Dirac equation in the linear limits. This advanced Dirac theory naturally enables the magnetic moment anomaly to be entirely explained without resort to renormalization and other ad hoc add-ons, and it also permits a detailed, granular understanding of how individual fermions interact with individual photons strictly on the quantum level. In sum, it advances Dirac theory to a quantum theory of the electron and the photon and their one-on-one interactions. Seven distinct experimental tests are proposed.

Contents

PART I: GAUGE SYMMETRY, TIME DILATION, AND THE ENERGY CONTENT OF MATERIAL BODIES IN CLASSICAL ELECTRODYNAMICS

1. From Minkowski Spacetime to Electromagnetic Interactions using Weyl's Local U(1) Gauge Symmetry: A Compact Review of the Known Physics

The modern concept of spacetime originated when Hermann Minkowski in his seminal paper [1] based on the Special Theory of Relativity [2], famously proclaimed that "from now onwards space by itself and time by itself will recede completely to become mere shadows and only a type of union of the two will still stand independently on its own." Following the advent of General Theory in [3], the invariant interval $c^2t^2-x^2-y^2-z^2$ Minkowski discovered became expressed via an infinitesimal metric line element $c^2d\tau^2=\eta_{\mu\nu}dx^\mu dx^\nu$ with a metric tensor $\mathrm{diag}\left(\eta_{\mu\nu}\right)=\left(1,-1,-1,-1\right)$ named for him. Moreover, it became understood that gravitational fields reside in a curved spacetime metric tensor $g_{\mu\nu}$ to which $\eta_{\mu\nu}$ defines the tangent space at each spacetime event, with a line element $c^2d\tau^2=g_{\mu\nu}dx^\mu dx^\nu$ specified according to Riemannian geometry which one of Gauss' preeminent students had been developed half a century earlier.

The equation $c^2d\tau^2=g_{\mu\nu}dx^\mu dx^\nu$ for the proper time line element $d\tau$ is often written in a number of different, albeit mathematically equivalent ways. For example, if one divides through by $d\tau^2$ and defines ("$\equiv$") a four-velocity $u^\mu\equiv dx^\mu/d\tau$ this equation becomes $c^2=g_{\mu\nu}u^\mu u^\nu$. By absorbing the spacetime indices into these vectors and writing $c^2=u_\sigma u^\sigma$, we see that the squared four-velocity is equal to the squared speed of light. Further, if we postulate some material mass m and multiply the foregoing through by m^2, also defining an energy-momentum vector $p^\mu=mu^\mu=mdx^\mu/d\tau=\left(E/c,\mathbf{p}\right)$, we arrive at $m^2c^2=g_{\mu\nu}p^\mu p^\nu=p_\sigma p^\sigma$, well-known as the relativistic energy momentum relation.

A next step often taken is to write down a complex function $\phi=s\exp\left(-ip_\sigma x^\sigma/\hbar\right)$ where $s\left(p^\nu\right)$ is a function of energy-momentum and $\exp\left(-ip_\sigma x^\sigma\right)$ is the kernel used in Fourier transforms between momentum space and configuration space. Using being the spacetime gradient operator $\partial_\mu=\left(\partial/c\partial t,\partial/\partial\mathbf{x}\right)=\left(\partial_t/c,\nabla\right)$ it is easy to see that $i\hbar\partial_\mu\phi=p_\mu\phi$. As a result, starting with $m^2c^2=p_\sigma p^\sigma$ and multiplying through from the right by ϕ, it is straightforward to form the operator equation $0=\left(\hbar^2\partial_\sigma\partial^\sigma+m^2c^2\right)\phi$, better-known as the Klein-Gordon equation for a free (non-interacting) particle.

It is also easy to see that by taking a simple scalar square root one can obtain the linear energy-momentum relation $mc=\pm\sqrt{g_{\mu\nu}p^\mu p^\nu}$, or $mc=\pm\sqrt{\eta_{\mu\nu}p^\mu p^\nu}$ in flat spacetime. But Dirac found in [4] that there exists an operator equation in flat spacetime – essentially a square-root of the Klein-Gordon equation – that uses a set of 4x4 matrices γ^μ *defined such that* $\frac{1}{2}\{\gamma^\mu\gamma^\nu+\gamma^\nu\gamma^\mu\}\equiv\eta^{\mu\nu}$. First we write $m^2c^2=\eta^{\mu\nu}p_\mu p_\nu=\frac{1}{2}\{\gamma^\mu\gamma^\nu+\gamma^\nu\gamma^\mu\}p^\mu p^\nu$. Then we observe that $(\gamma^\mu p_\mu)^2=(\gamma^\mu p_\mu)(\gamma^\nu p_\nu)=(\gamma^\nu p_\nu)(\gamma^\mu p_\mu)=\frac{1}{2}\{\gamma^\mu\gamma^\nu+\gamma^\nu\gamma^\mu\}p_\mu p_\nu=\eta^{\mu\nu}p_\mu p_\nu$. Therefore, $\pm\sqrt{\eta^{\mu\nu}p_\mu p_\nu}=\gamma^\mu p_\mu$. However, in order to connect this with $mc=\pm\sqrt{\eta_{\mu\nu}p^\mu p^\nu}$ two adjustments are required. First, because $\gamma^\mu p_\mu$ is a 4x4 matrix, the mass term mc needs to be formed into mc times a 4x4 identity matrix $I_{4\times4}$, which is implicitly understood, not explicitly shown. Second, because $mcI_{4\times4}$ is a diagonal matrix while $\gamma^\mu p_\mu$ cannot be diagonalized, simply equating $\gamma^\mu p_\mu=mc$ is mathematically nonsensical. Instead, we form a four-component Dirac spinor $u(p^\nu)$ and multiply from the right to obtain $(\gamma^\mu p_\mu-mc)u=0$. This makes mathematical sense as an operator equation with eigenvectors and eigenvalues. Note also that the $\pm$ sign, which results whenever a square-root is taken, gets absorbed into the components of γ^μ, all of which are ±1 or $\pm i$ with an balanced number of positive and negative entries. Further, similar to Klein-Gordon equation above, we write down a four-component spinor function $\psi=u\exp(-ip_\sigma x^\sigma/\hbar)$, deduce that $i\hbar\partial_\mu\psi=p_\mu\psi$, and so may write $(i\hbar\gamma^\mu\partial_\mu-mc)\psi=0$ which is Dirac's equation for a non-interacting fermion, e.g. electron in a configuration space.

Dirac's equation as developed above applies within a flat spacetime. To generalize to curved spacetime, thus to gravitation, we first define a set of Γ^μ having a parallel definition $\frac{1}{2}\{\Gamma^\mu\Gamma^\nu+\Gamma^\nu\Gamma^\mu\}\equiv g^{\mu\nu}$. We also establish a vierbein, a.k.a. tetrad e^μ_a, with both a superscripted Greek "spacetime/world" index and an early-in-the-alphabet subscripted Latin "Lorentz/ Minkowski" index, and define the tetrad by the relation $e^\mu_a\gamma^a\equiv\Gamma^\mu$. Consequently we deduce that $g^{\mu\nu}=\frac{1}{2}\{\gamma^a\gamma^b+\gamma^b\gamma^a\}e^\mu_a e^\nu_b=\eta^{ab}e^\mu_a e^\nu_b$. It is readily seen that the flat spacetime $g^{\mu\nu}=\eta^{\mu\nu}$ and $\Gamma^\mu=\gamma^\mu$ are obtained when $e^\mu_a=1$ along the $\mu=a$ diagonal and zero otherwise, i.e., when e^μ_a is a 4x4 unit matrix. Then, starting with $mc=\pm\sqrt{g_{\mu\nu}p^\mu p^\nu}$ we follow the exact same steps as in the previous paragraph, ending up with $(\Gamma^\mu p_\mu-mc)u=0$ in momentum space and $(i\hbar\Gamma^\mu\partial_\mu-mc)\psi=0$ in configuration space.

However, in curved spacetime, in order to couple the spinor fields ψ to gravity in a generally-covariant manner, we must also advance ∂_μ to a spin-covariant derivative $\partial_\mu\mapsto\nabla_\mu\equiv\partial_\mu-\frac{i}{4}\omega^{ab}_\mu\sigma_{ab}$, where a spin connection $\omega^{ab}_\beta=-\omega^{ba}_\beta$ which is antisymmetric in the Lorentz indexes a, b is defined using the gravitational-covariant derivative of $e^{\nu b}$ by $\omega^{ab}_\mu\equiv e_\nu{}^a\partial_{;\mu}e^{\nu b}=e_\nu{}^a(\partial_\mu e^{\nu b}+\Gamma^\nu_{\sigma\mu}e^{\sigma b})$, and where $\sigma_{ab}\equiv\frac{i}{2}[\gamma_a\gamma_b-\gamma_b\gamma_a]$ are the bilinear covariants which in the form of $\bar{\psi}\sigma^{\mu\nu}\psi$ contain the fermion polarization and magnetization bivectors. The extra term $-\frac{i}{4}\omega^{ab}_\mu\sigma_{ab}$ also makes its way back into the momentum space Dirac equation which thereby becomes $\left(\Gamma^{\mu}\left(p_{\mu}-\frac{i}{4}\omega_{\mu}^{ab}\sigma_{ab}\right)-mc\right)u=0$. The foregoing may all be thought of as equivalent albeit progressively-richer and more-revealing ways of writing the spacetime geometry metric interval $c^2d\tau^2=g_{\mu\nu}dx^{\mu}dx^{\nu}$.

In §9 of [3], one of the most important findings was not only that gravitation could be reduced to pure geometry based on a spacetime metric, but, in a phrase later coined by Wheeler [5], that the resulting theory was a theory of "geometro*dynamics*." Specifically, for a finite proper time $\tau=\int_A^B d\tau$ between any two events *A* and *B*, the lines $0=\delta\int_A^B d\tau$ of minimized variation are the geodesics of motion. Moreover, this equation of motion has been shown for over a century without empirical contradiction to describe gravitational motion. This calculation again begins with $c^2d\tau^2=g_{\mu\nu}dx^{\mu}dx^{\nu}$, now divided through by $c^2d\tau^2$ and turned into the number:

$$1=g_{\mu\nu}\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}. \tag{1.1}$$

Next, taking the scalar square root of this "1" enables us to write the variational equation as:

$$0=\delta\int_A^B d\tau=\delta\int_A^B(1)\,d\tau=\delta\int_A^B d\tau\sqrt{g_{\mu\nu}\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}}, \tag{1.2}$$

where the $\pm$ sign which attends to taking a square root may be discarded because of the zero on the left-hand side above. Then, using a well-known calculation reviewed in Appendix A because we shall shortly derive the Lorentz Force motion of classical electrodynamics in a similar way, one is able to derive the equation of motion (A.14), reproduced below:

$$\frac{d^2x^{\beta}}{d\tau^2}=-\Gamma^{\beta}{}_{\mu\nu}\frac{dx^{\mu}}{d\tau}\frac{dx^{\nu}}{d\tau}. \tag{1.3}$$

Given that (1.3) is derived when (1.2) is applied to the spacetime metric $c^2d\tau^2=g_{\mu\nu}dx^{\mu}dx^{\nu}$ merely divided through by $c^2d\tau^2$ in the form of (1.1), it is not uncommon to regard $c^2d\tau^2=g_{\mu\nu}dx^{\mu}dx^{\nu}$ as the first integral of this equation of motion. So once again, we arrive at an even-richer understanding of the simple metric $c^2d\tau^2=g_{\mu\nu}dx^{\mu}dx^{\nu}$ for curved spacetime geometry. And this now brings us to electrodynamics.

During the course of just over a decade, Hermann Weyl in [6], [7], [8] convincingly demonstrated that electromagnetism is a gauge theory based on a *local* U(1) internal symmetry group. The underlying principle of gauge symmetry is that the equations of physics – such as the Dirac equation or the Klein-Gordon equation or their respective Lagrangian densities – must remain invariant under transformations in a complex phase space defined by $\exp(i\Lambda)=\cos\Lambda+i\sin\Lambda$ where $\Lambda(t,\mathbf{x})$ is a *locally-variable* phase angle. Specifically, we require any physics equations containing a generalized function φ to be symmetric under a local transformation $\varphi \to \varphi' \equiv \exp(i\Lambda)\varphi$ which changes the direction but not the magnitude of the function in the phase space. However, because $\partial_\mu \varphi \to \partial_\mu \varphi' = \exp(i\Lambda)(\partial_\mu + i\partial_\mu \Lambda)\varphi$ violates this symmetry, we are required to define a gauge-covariant derivative $\mathcal{D}_\mu$ which likewise transforms as $\mathcal{D}_\mu \to \mathcal{D}'_\mu \equiv \exp(i\Lambda)\mathcal{D}_\mu$. So we introduce a vector gauge field A_μ and a charge q fashioned into $\mathcal{D}_\mu \equiv \partial_\mu - iqA_\mu / \hbar c$. Now $\mathcal{D}_\mu \varphi \to \mathcal{D}_\mu \varphi' = \exp(i\Lambda)\left[\partial_\mu - i\left(qA_\mu / \hbar c - \partial_\mu \Lambda\right)\right]\varphi$. Along with this, if we define $qA_\mu \to qA'_\mu \equiv qA_\mu + \hbar c \partial_\mu \Lambda$ as the transformation for the gauge field, then the $\partial_\mu \Lambda$ terms will cancel, so $\mathcal{D}_\mu \varphi \to \mathcal{D}_\mu \varphi' = \exp(i\Lambda)\left[\partial_\mu - iqA_\mu / \hbar c\right]\varphi = \exp(i\Lambda)\mathcal{D}_\mu \varphi$ is also redirected in the phase space just like $\varphi \to \varphi' \equiv \exp(i\Lambda)\varphi$, exactly as required. Note, in the above we adopt a convention where q is a positive charge. So for an electron, for example, we would set $q = -e$.

Then, armed with $\mathcal{D}_\mu \equiv \partial_\mu - iqA_\mu / \hbar c$, we merely substitute $\partial_\mu \mapsto \mathcal{D}_\mu$ into any physics equation containing ∂_μ operating on a general function φ, and are assured this equation will have a local U(1) gauge symmetry. So for Dirac's equation operating on $\varphi = \psi$, in flat spacetime where $\partial_{;\mu} e^{\nu b} = 0$ thus $\omega_\mu^{ab} = 0$ and $\nabla_\mu = \partial_\mu$ we substitute $\partial_\mu \mapsto \mathcal{D}_\mu \equiv \partial_\mu - iqA_\mu / \hbar c$ for the spin-covariant derivative to obtain $0 = \left(i\hbar \gamma^\mu \mathcal{D}_\mu - mc\right)\psi = \left(\gamma^\mu \left(i\hbar \partial_\mu + qA_\mu / c\right) - mc\right)\psi$. For the Klein-Gordon equation we obtain $0 = \left(\hbar^2 \mathcal{D}_\sigma \mathcal{D}^\sigma + m^2 c^2\right)\phi = \left(\hbar^2 \left(\partial_\sigma - iqA_\sigma / \hbar c\right)\left(\partial^\sigma - iqA^\sigma / \hbar c\right) + m^2 c^2\right)\phi$ by doing the same with $\varphi = \phi$. Empirical evidence for almost a century has established these to be correct equations for interacting fermions and bosons, with q being a physical electric charge and A^μ being a physical electromagnetic vector potential. In fact, if we subject a generalized gauge potential G^μ with related charges g to a gauge transformation $G^\mu \to G'^\mu \equiv \exp(i\Lambda) G^\mu$ and likewise require invariance of the field strength $F^{\mu\nu} = \partial^\mu G^\nu - \partial^\nu G^\mu$ under this transformation, we can even obtain $F^{\mu\nu} = \mathcal{D}^\mu G^\nu - \mathcal{D}^\nu G^\mu = \partial^{[\mu} G^{\nu]} - ig\left[G^\mu, G^\nu\right] / \hbar c$ using this heuristic prescription $\partial_\mu \mapsto \mathcal{D}_\mu$ with $\mathcal{D}_\mu \equiv \partial_\mu - igG_\mu / \hbar c$. This application of local gauge symmetry to gauge fields themselves, will be recognized to now yield a non-Abelian Yang-Mills [9] field strength such as that of $SU(2)_L$ weak and $SU(3)_{QCD}$ strong interactions.

From here we backtrack from configuration to momentum space via the relation $i\hbar \partial_\mu \varphi = p_\mu \varphi$ for an ordinary derivative operating on a function φ containing the Fourier kernel $\exp\left(-ip_\sigma x^\sigma\right)$. Consequently, using $\psi = u \exp\left(-ip_\sigma x^\sigma / \hbar\right)$ then removing the kernel, Dirac's equation becomes $\left(\gamma^\mu \left(p_\mu + qA_\mu / c\right) - mc\right)u = 0$ in flat spacetime, we reveals the electron magnetic moment, see, e.g., section 2.6 of [10]. Likewise, using $\phi = s \exp\left(-ip_\sigma x^\sigma / \hbar\right)$ and then removing the kernel, the Klein-Gordon equation becomes $0 = \left(\left(p_\sigma + qA_\sigma / c\right)\left(p^\sigma + qA^\sigma / c\right) - m^2 c^2\right)s$. Here, however, because there are no γ^μ matrices,

$s\left(p^{\nu}\right)$ may be removed, and we end up with a mathematically perfectly sensible equation $m^2c^2=\left(p_\sigma+qA_\sigma/c\right)\left(p^\sigma+qA^\sigma/c\right)$. Defining a gauge-covariant or "canonical" momentum $\pi^\mu \equiv p^\mu+qA^\mu/c$, this is compactly written as $m^2c^2=\pi_\sigma\pi^\sigma$, and is simply the relativistic energy-momentum relation $m^2c^2=p_\sigma p^\sigma$ generalized via local U(1) gauge symmetry to encompass a test charge q with mass m within a vector potential A^σ. From this we see that in momentum space in flat spacetime, requiring local U(1) gauge symmetry leads to a prescription $p^\mu \mapsto \pi^\mu$, which is the momentum-space parallel to the configuration space prescription $\partial_\mu \mapsto \mathfrak{D}_\mu$. So in momentum space Dirac's flat spacetime equation becomes $\left(\gamma^\mu\pi_\mu-mc\right)u=0$ and the relativistic energy momentum relation underpinning the Klein-Gordon equation becomes $m^2c^2=\pi_\sigma\pi^\sigma$.

Taking a closer look at the relation $m^2c^2=\pi_\sigma\pi^\sigma$ with $\pi^\mu \equiv p^\mu+qA^\mu/c$, we may write:

$$m^2c^2=\pi_\sigma\pi^\sigma=\left(p_\sigma+\frac{qA_\sigma}{c}\right)\left(p^\sigma+\frac{qA^\sigma}{c}\right)=p_\sigma p^\sigma+\frac{q}{c}\left(A_\sigma p^\sigma+p_\sigma A^\sigma\right)+\frac{q^2}{c^2}A_\sigma A^\sigma . \tag{1.4}$$

In the above we have avoided commuting p^σ with A^σ to combine the mixed terms $A_\sigma p^\sigma+p_\sigma A^\sigma$ into $2A_\sigma p^\sigma$ or $2p_\sigma A^\sigma$. This is because $A^\sigma=(\phi,\mathbf{A})$ is a function of the spacetime coordinates $x^\mu=(ct,\mathbf{x})$ while $p^\sigma=(E/c,\mathbf{p})$ is an energy momentum vector. So when we treat position and momentum as Heisenberg operator matrices we cannot commute **x** and **p** without exercising care, because of the canonical relation $\left[x_i,p_j\right]=i\hbar\delta_{ij}$. Likewise, because the Hamiltonian operator H has energy eigenvalues $H\left|s\right\rangle=\left(E-mc^2\right)\left|s\right\rangle$ when operating on a state vector $\left|s\right\rangle$, the Heisenberg Equation of motion $\left[H,A^\nu\right]=-i\hbar d_t A^\nu+i\hbar\partial_t A^\nu$ (take careful note of the total versus partial derivatives) also requires us to exercise care when we commute $cp^0=E$ with $A^0=\phi$ whenever fixed-basis state vectors $\left|s\right\rangle$ and field operators ϕ are involved. So to combine terms in (1.4) to show, say, $2A_\sigma p^\sigma$ while not ignoring Heisenberg commutation, we may make use of the commutator $\left[p_\sigma,A^\sigma\right]=p_\sigma A^\sigma-A_\sigma p^\sigma$ to identically rewrite (1.4) as:

$$m^2c^2=\pi_\sigma\pi^\sigma=\left(p_\sigma+\frac{qA_\sigma}{c}\right)\left(p^\sigma+\frac{qA^\sigma}{c}\right)=p_\sigma p^\sigma+2\frac{q}{c}A_\sigma p^\sigma+\frac{q}{c}\left[p_\sigma,A^\sigma\right]+\frac{q^2}{c^2}A_\sigma A^\sigma . \tag{1.5}$$

Then, if we choose to approximate around these commutation issues and thereby set $\left[p_\sigma,A^\sigma\right]=0$ which amounts to taking a classical $\hbar\to 0$ limit, (1.5) easily reduces to:

$$m^2c^2=\pi_\sigma\pi^\sigma=\left(p_\sigma+\frac{qA_\sigma}{c}\right)\left(p^\sigma+\frac{qA^\sigma}{c}\right)=p_\sigma p^\sigma+2\frac{q}{c}A_\sigma p^\sigma+\frac{q^2}{c^2}A_\sigma A^\sigma . \tag{1.6}$$

All of the foregoing is well-known, well-established, empirically-validated physics. Now, however, continuing deductively from the above, we shall uncover some equally-valid new relations and new physics which do not appear to be known to date. At the outset we will work from the classical approximation (1.6) in which we have set $\left[p_\sigma, A^\sigma\right]=0$ and thus effectively set $\hbar=0$. Later, after sufficient development in section 7, we will shift over and work from (1.5) to fully account for the quantum mechanics of the commutation $\left[p_\sigma, A^\sigma\right]$, and thereby will be able to see precisely how quantum mechanics alters the classical results we shall obtain from (1.6).

2. Derivation of Geodesic Lorentz Force Motion from Local U(1) Gauge Symmetry

Starting with the classical $\hbar \to 0$ relation (1.6), let us use the definitions $p^\mu \equiv mu^\mu$ for the ordinary energy-momentum and $u^\mu \equiv dx^\mu / d\tau$ for the 4-velocity to write (1.6) as:

$$m^2c^2 = \pi_\sigma \pi^\sigma = m^2 u_\sigma u^\sigma + 2\frac{qm}{c} A_\sigma u^\sigma + \frac{q^2}{c^2} A_\sigma A^\sigma = m^2 \frac{dx_\sigma}{d\tau}\frac{dx^\sigma}{d\tau} + 2\frac{qm}{c} A_\sigma \frac{dx^\sigma}{d\tau} + \frac{q^2}{c^2} A_\sigma A^\sigma . \tag{2.1}$$

Then, continuing to backtrack, we divide the above through by m^2c^2 and also raise an index to show the metric tensor in the first term after the final equality. We thereby obtain:

$$1 = \frac{\pi_\sigma \pi^\sigma}{mc^2} = g_{\mu\nu} \frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau} + 2\frac{q}{mc^2} A_\sigma \frac{dx^\sigma}{cd\tau} + \frac{q^2}{m^2c^4} A_\sigma A^\sigma . \tag{2.2}$$

The above is identical to (1.1) unless both $q \neq 0$ and $A^\sigma \neq 0$. That is, unless we have *both* a test charge with a charge-to-mass ratio q/m, and *also* a potential A^σ with which that test charge is interacting, (2.2) is the same as (1.1). This using (2.2) with *either* $q=0$ *or* $A^\sigma=0$ in the variational equation (1.2) will produce the gravitational geodesic motion of (1.3).

This raises the question whether using (2.2) with both $q \neq 0$ and $A^\sigma \neq 0$ in the variation $0=\delta\int_A^B d\tau$ as in (1.2) might produce *the Lorentz Force motion of electrodynamics together with the gravitational motion.* In other words, (2.2) raises the question whether the combined classical gravitational and electromagnetic motions can *both* be derived as geodesic motions from a variation using (2.2), which, as is easily seen, is just $m^2c^2=\pi_\sigma\pi^\sigma$ from (1.6) divided through by through by m^2c^2. And (1.6) of course, is in turn merely the relativistic energy-momentum relation $m^2c^2 = p_\sigma p^\sigma$ following application of the $p^\mu \mapsto \pi^\mu$ prescription which comes about by requiring Weyl's local U(1) gauge symmetry. And $m^2c^2 = p_\sigma p^\sigma$ is in turn just another way of representing the metric $c^2 d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu$ once a rest mass m has been postulated and the metric multiplied through by $m^2 / d\tau^2$ while lowering an index. So all roads lead back to $c^2 d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu$.

To prove that the electrodynamic Lorentz Force motion can be understood as geodesic motion just like gravitational motion, as we did at (1.1) to (1.3), we first take the square root of the "1" in (2.2) and use it in the variational equation, to write the following, in contrast to (1.2):

$$0=\delta\int_A^B d\tau=\delta\int_A^B (1)\, d\tau=\delta\int_A^B d\tau\sqrt{g_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+2\frac{q}{mc^2}A_\sigma\frac{dx^\sigma}{cd\tau}+\frac{q^2}{m^2c^4}A_\sigma A^\sigma}\ . \tag{2.3}$$

We then apply δ to the integrand and use (2.2) to remove the denominator, obtaining:

$$0=\delta\int_A^B d\tau=\frac{1}{2}\int_A^B d\tau\delta\left(g_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+2\frac{q}{mc^2}A_\sigma\frac{dx^\sigma}{cd\tau}+\frac{q^2}{m^2c^4}A_\sigma A^\sigma\right). \tag{2.4}$$

The first of the three terms corresponds with (A.1) which leads to gravitational motion. So we segregate that term right away, then apply (A.12) which is directly derived from (A.1), to obtain:

$$\begin{aligned}0=\delta\int_A^B d\tau=&\int_A^B \delta x^\alpha d\tau\left(\frac{1}{2}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-g_{\alpha\nu}\frac{d^2x^\nu}{c^2d\tau^2}\right)\\&+\int_A^B d\tau\delta\left(\frac{q}{mc^2}A_\sigma\frac{dx^\sigma}{cd\tau}+\frac{q^2}{2m^2c^4}A_\sigma A^\sigma\right)\end{aligned}. \tag{2.5}$$

Because $-\Gamma^\beta{}_{\mu\nu}=\frac{1}{2}g^{\beta\alpha}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)$, we see that the gravitational motion (A.14) i.e. (1.3) is already contained in the top line above. So now let's develop the bottom line which contains the additional electrodynamic terms added by the U(1) gauge symmetry via the parallel configuration and momentum space rules $\partial_\mu \mapsto \mathfrak{D}_\mu$ and $p^\mu \mapsto \pi^\mu$ reviewed in section 1.

For the bottom line of (2.5) we first distribute δ using the product rule, and assume no variation in the charge-to-mass ratio i.e. that $\delta(q/m)=0$ over the path from *A* to *B*, thus finding:

$$\begin{aligned}0=\delta\int_A^B d\tau=&\int_A^B \delta x^\alpha d\tau\left(\frac{1}{2}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-g_{\alpha\nu}\frac{d^2x^\nu}{c^2d\tau^2}\right)\\&+\int_A^B d\tau\left(\frac{q}{mc^2}\delta A_\sigma\frac{dx^\sigma}{cd\tau}+\frac{q}{mc^2}A_\sigma\delta\frac{dx^\sigma}{cd\tau}+\frac{q^2}{2m^2c^4}\delta\left(A_\sigma A^\sigma\right)\right)\end{aligned}. \tag{2.6}$$

From (A.3) we may deduce that $\delta A_\sigma=\delta x^\alpha\partial_\alpha A_\sigma$ and $\delta\left(A_\sigma A^\sigma\right)=\delta x^\alpha\partial_\alpha\left(A_\sigma A^\sigma\right)$. We use these as well as $\delta d=d\delta$ employed for (A.2) to advance the above to:

$$0=\delta\int_A^B d\tau=\int_A^B \delta x^\alpha d\tau\left(\frac{1}{2}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-g_{\alpha\nu}\frac{d^2x^\nu}{c^2d\tau^2}\right)$$
$$+\int_A^B d\tau\left(\frac{q}{mc^2}\delta x^\alpha\partial_\alpha A_\sigma\frac{dx^\sigma}{cd\tau}+\frac{q}{mc^2}A_\sigma\frac{d\delta x^\sigma}{cd\tau}+\frac{q^2}{2m^2c^4}\delta x^\alpha\partial_\alpha\left(A_\sigma A^\sigma\right)\right). \quad (2.7)$$

We next use (A.10) to obtain $dA_\sigma / cd\tau=\partial_\alpha A_\sigma dx^\alpha / cd\tau$. Then, for the second term on the bottom line above, to set up an integration-by-parts, we use this with the product rule to form:

$$\frac{d}{cd\tau}\left(A_\sigma\delta x^\sigma\right)=\delta x^\sigma\frac{dA_\sigma}{cd\tau}+A_\sigma\frac{d\delta x^\sigma}{cd\tau}=\delta x^\sigma\partial_\alpha A_\sigma\frac{dx^\alpha}{cd\tau}+A_\sigma\frac{d\delta x^\sigma}{cd\tau}. \quad (2.8)$$

Using (2.8) in (2.7) then produces:

$$0=\delta\int_A^B d\tau=\int_A^B \delta x^\alpha d\tau\left(\frac{1}{2}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-g_{\alpha\nu}\frac{d^2x^\nu}{c^2d\tau^2}\right)$$
$$+\int_A^B d\tau\left(\frac{q}{mc^2}\delta x^\alpha\partial_\alpha A_\sigma\frac{dx^\sigma}{cd\tau}+\frac{q}{mc^2}\left(\frac{d}{cd\tau}\left(A_\sigma\delta x^\sigma\right)-\delta x^\sigma\partial_\alpha A_\sigma\frac{dx^\alpha}{cd\tau}\right)+\frac{q^2}{2m^2c^4}\delta x^\alpha\partial_\alpha\left(A_\sigma A^\sigma\right)\right). \quad (2.9)$$

The term containing total integral in the above is equal to zero because of the boundary conditions on the definite integral in the variation. Specifically, in the above:

$$\int_A^B d\tau\frac{d}{d\tau}\left(A_\sigma\delta x^\sigma\right)=\int_A^B d\left(A_\sigma\delta x^\sigma\right)=\left(A_\sigma\delta x^\sigma\right)\Big|_A^B=0, \quad (2.10)$$

This is zero for the same reasons that (A.7) is zero when calculating the gravitational geodesics. Consequently, using (2.10) in (2.9) and with a renaming of summed indexes so there is a δx^α with a common α index in all terms, then factoring this out, (2.9) becomes:

$$0=\delta\int_A^B d\tau=\int_A^B \delta x^\alpha d\tau\left(\begin{array}{l}\frac{1}{2}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-g_{\alpha\nu}\frac{d^2x^\nu}{c^2d\tau^2}\\ +\frac{q}{mc^2}\left(\partial_\alpha A_\sigma-\partial_\sigma A_\alpha\right)\frac{dx^\sigma}{cd\tau}+\frac{q^2}{2m^2c^4}\partial_\alpha\left(A_\sigma A^\sigma\right)\end{array}\right). \quad (2.11)$$

It is very important that the integration-by-parts produced both a sign reversal as well as an index reversal, because $F_{\alpha\sigma}=\partial_\alpha A_\sigma-\partial_\sigma A_\alpha$ is the covariant-indexed electromagnetic field strength.

Now we are at (A.12) for the gravitational geodesics, but with some new terms. For the same reasons as at (A.12), the expression inside the large parenthesis above must be zero. So setting this to zero, using $F_{\alpha\sigma}=\partial_\alpha A_\sigma-\partial_\sigma A_\alpha$, multiplying all terms by $g^{\beta\alpha}$ to raise an index, using $-\Gamma^\beta{}_{\mu\nu}=\frac{1}{2}g^{\beta\alpha}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)$, and segregating the acceleration, yields:

$$\frac{d^2x^\beta}{c^2d\tau^2}=-\Gamma^\beta{}_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+\frac{q}{mc^2}F^\beta{}_\sigma\frac{dx^\sigma}{cd\tau}+\frac{q^2}{2m^2c^4}\partial^\beta\left(A_\sigma A^\sigma\right). \tag{2.12}$$

So it is possible to derive (2.12) from the variation $0=\delta\int_A^B d\tau$ using $1=\pi_\sigma\pi^\sigma/mc^2$ from (2.2) which simply restates the locally U(1) gauge-symmetric relativistic energy-momentum relation $m^2c^2=\pi_\sigma\pi^\sigma$ of (1.6). Therefore the Lorentz Force motion which has been thoroughly validated empirically over the course of decades can indeed be understood as geodesic motion just like the gravitational motion. This does not appear to have previously been reported in the literature, and so warrants attention at least from viewpoint of at least *mathematical* physics.

However (2.12) also has an extra term $\left(q^2/2m^2c^4\right)\partial^\beta\left(A_\sigma A^\sigma\right)$ which warrants *physical* attention. As we shall later see, this term is naturally removed by a variant of the Lorenz gauge $\partial_\sigma A^\sigma=0$ when (1.5) is applied with the commutator $\left[p_\sigma,A^\sigma\right]\neq 0$ i.e. $\hbar\neq 0$ in accordance with quantum mechanics. In other words, this added term arises precisely because we have neglected quantum mechanics by using (1.6) rather than (1.5) in the variation (2.3), and disappears once quantum mechanics is taken into account and the commutator not approximated to zero.

3. The Canonical Relativistic Energy-Momentum Relation, and the Apparently "Peculiar" Quadratic Line Element with which it is Synonymous

At (2.1) we took the relation $m^2c^2=\pi_\sigma\pi^\sigma$ of (1.4) in the classical $\hbar\to 0$ limit and divided through by m^2c^2 to arrive at (2.2) which, when used in the variation (2.3), yielded the geodesic equation (2.12). This includes Lorentz Force motion plus an extra term containing $\partial^\beta\left(A_\sigma A^\sigma\right)$. Let us now take this same $m^2c^2=\pi_\sigma\pi^\sigma$ of (1.4), (1.5) and use $p^\sigma=mdx^\sigma/d\tau$ to obtain:

$$\begin{aligned}m^2c^2=\pi_\sigma\pi^\sigma&=\left(m\frac{dx_\sigma}{d\tau}+\frac{qA_\sigma}{c}\right)\left(m\frac{dx^\sigma}{d\tau}+\frac{qA^\sigma}{c}\right)\\&=m^2\frac{dx_\sigma}{d\tau}\frac{dx^\sigma}{d\tau}+2\frac{qm}{c}A_\sigma\frac{dx^\sigma}{d\tau}+\frac{q}{c}\left[m\frac{dx_\sigma}{d\tau},A^\sigma\right]+\frac{q^2}{c^2}A_\sigma A^\sigma\end{aligned}. \tag{3.1}$$

In the classical $\hbar\to 0$ limit of (1.6) where we neglect commutation by setting $\left[p_\sigma,A^\sigma\right]=0$, using the approximation sign "$\cong$" prior to the final expression as a reminder of this, we obtain:

$$m^2c^2=\pi_\sigma\pi^\sigma=\left(m\frac{dx_\sigma}{d\tau}+\frac{qA_\sigma}{c}\right)\left(m\frac{dx^\sigma}{d\tau}+\frac{qA^\sigma}{c}\right)\cong m^2\frac{dx_\sigma}{d\tau}\frac{dx^\sigma}{d\tau}+2\frac{qm}{c}A_\sigma\frac{dx^\sigma}{d\tau}+\frac{q^2}{c^2}A_\sigma A^\sigma. \tag{3.2}$$

Then, also defining a gauge-covariant coordinate element $\mathcal{D}x^\mu \equiv dx^\mu + \left(q/mc^2\right)A^\mu c d\tau$, we simply multiply through by $d\tau^2/m^2$ and raise some selected indices to obtain:

$$\begin{aligned} c^2 d\tau^2 &= \frac{d\tau^2}{m^2}\pi_\sigma \pi^\sigma = \left(dx_\sigma + \frac{q}{mc^2}A_\sigma c d\tau\right)\left(dx^\sigma + \frac{q}{mc^2}A^\sigma c d\tau\right) = g_{\mu\nu}\mathcal{D}x^\mu \mathcal{D}x^\nu \\ &\cong g_{\mu\nu}dx^\mu dx^\nu + 2\frac{q}{mc^2}A_\sigma dx^\sigma c d\tau + \frac{q^2}{m^2c^4}A_\sigma A^\sigma c^2 d\tau^2 \end{aligned} . \tag{3.3}$$

The above is simply the metric equation $c^2 d\tau^2 = g_{\mu\nu}dx^\mu dx^\nu$ supplemented by new terms which come about because of gauge symmetry. These new terms are non-zero whenever there is a test charge with $q/m \neq 0$ situated in a gauge potential $A^\sigma \neq 0$. They arise because of the local U(1) gauge symmetry, and in fact reveal that the momentum space prescription $p^\mu \mapsto \pi^\mu$ and the configuration space prescription $\partial_\mu \mapsto \mathfrak{D}_\mu$ previously reviewed also go hand-in-hand with a parallel prescription $dx^\mu \mapsto \mathcal{D}x^\mu$ for the infinitesimal coordinate interval.

However, this metric (3.3) is unusual because it is *quadratic* in the line element $ds = c d\tau$. This quadratic is seen if we rewrite the bottom line of (3.3) which contains the classical $\hbar \rightarrow 0$ line element, with the approximation sign removed, in the form:

$$0 = \left(1 - \frac{q^2}{m^2c^4}A_\sigma A^\sigma\right)c^2 d\tau^2 - 2\frac{q}{mc^2}A_\sigma dx^\sigma c d\tau - g_{\mu\nu}dx^\mu dx^\nu , \tag{3.4}$$

and then use this in the quadratic equation to obtain the solution:

$$c d\tau = \frac{\frac{q}{mc^2}A_\sigma dx^\sigma \pm \sqrt{\left[g_{\mu\nu}\left(1 - \frac{q^2}{m^2c^4}A_\sigma A^\sigma\right) + \frac{q^2}{m^2c^4}A_\mu A_\nu\right]dx^\mu dx^\nu}}{1 - \frac{q^2}{m^2c^4}A_\sigma A^\sigma} . \tag{3.5}$$

Now, on the one hand, the metric (3.3) is just another way of stating the well-established relation $m^2c^2 = \pi_\sigma \pi^\sigma$ which is merely the relativistic energy-momentum relation $m^2c^2 = p_\sigma p^\sigma$ after imposing local U(1) gauge symmetry which causes the momentum space replacement $p^\mu \mapsto \pi^\mu$. In (3.3) that relation is written as $c^2 d\tau^2 = \left(d\tau^2/m^2\right)\pi_\sigma \pi^\sigma$, which is just another variant of $1 = \pi_\sigma \pi^\sigma / mc^2$ which was used in (2.3) to obtain the geodesic motion in (2.12).

On the other hand, when couched in the form of (3.3), and especially after obtaining the quadratic solution (3.5), this metric (3.3) <u>*appears*</u> *to have some problems*, and certainly, as a quadratic in $d\tau$, it is an unusual line element. One might notice that the metric (3.3), (3.5) is a function $d\tau(q/m)$ of the q/m ratio of a test charge and suppose this to mean that the *invariant* line element $ds = cd\tau$ and the background fields A^μ and $g_{\mu\nu}$ are actually not invariant when q/m is changed, which would not be permitted by field theory. And, one may notice that the term $A_\sigma A^\sigma$ is not invariant under a local U(1) gauge transformation, giving the line element a gauge-dependency. One might even go so far as to believe that this is a "peculiar" or even "aberrant" line element that cannot be associated to a Riemannian geometry, and moreover, that geodesics calculated starting with this line element are strongly non-linear involving irrational functions of electromagnetic potential. And one might then conclude that any development based on (3.3) can lead to no more than a chain of allegations and mistakes.

At the same time, however, (3.3) is simply (2.2) multiplied through by $c^2 d\tau^2$. When (2.2) is used in the variation (2.3) the resulting geodesics are given by (2.12) which does contain both the gravitational motion *and the Lorentz Force motion*, differing only by the final $\partial^\beta \left(A_\sigma A^\sigma\right)$ term which is a non-linear function of the electromagnetic potential, and which we still need to attend to. So to dismiss (3.3) out of hand because of its unusual form or the foregoing conceptual challenges would be a mistake. This is because if $c^2 d\tau^2 = d\tau^2 \pi_\sigma \pi^\sigma / m^2$ in (3.3) is a wrong equation then so too is $m^2 c^2 = \pi_\sigma \pi^\sigma$ in (1.6), given that *these are the very same equation* obtained from one another by the elementary algebra of multiplying both sides of an equation by the same objects. And if $m^2 c^2 = \pi_\sigma \pi^\sigma$ is a wrong equation, this would precipitate an unwarranted crisis in gauge theory itself, because the prescription to go from $m^2 c^2 = p_\sigma p^\sigma$ to $m^2 c^2 = \pi_\sigma \pi^\sigma$ via $p^\mu \mapsto \pi^\mu$ would also be wrong, yet this prescription is fundamental to local gauge theory as reviewed between (1.3) and (1.4). Or, $m^2 c^2 = p_\sigma p^\sigma$ would have to be wrong, which would be in collision with all the relativistic physics we know. Therefore, we have little choice but to adopt the view that (3.3) though peculiar in appearance is actually just as correct as $m^2 c^2 = \pi_\sigma \pi^\sigma$ with which it is synonymous. And we now also know that the $1 = \pi_\sigma \pi^\sigma / mc^2$ variant of (3.3) which is (2.2) produces the well-established geodesic motion contain in (2.12), plus an extra term still to be studied. Consequently, taking (3.3) as a challenge not than a mistake, we must find out more about the heretofore undiscovered physics which arises when the metric (3.3) is carefully studied in depth to all it its logical conclusions. This study will now become the focus of the rest of this paper.

4. The Quadratic Line Element at Rest with no Gravitation

The metric (3.3) is unusual in appearance for the several reasons laid out above, and yet it is not incorrect unless $m^2 c^2 = \pi_\sigma \pi^\sigma$ is incorrect, which it is not. To make better sense of (3.3), it is helpful to place the vector potential and the test charge into a rest frame thus placing the test charge and the source of the potential at rest relative to one another, and to work in flat spacetime. To do so we take a classical vector potential $A^\mu = \left(\phi, \mathbf{A}\right)$ and transform this to a rest frame so that $A^\mu = \left(\phi_0, \mathbf{0}\right)$ where ϕ_0 is the proper scalar potential. Additionally, starting with the coordinate element $dx^\mu = \left(cdt, d\mathbf{x}\right)$ we set $dx^\mu = \left(cdt, \mathbf{0}\right)$ to place the test particle in the same rest frame. Then then set $g_{\mu\nu} = \eta_{\mu\nu}$ to work in flat spacetime. Thus, at rest without gravitation, the classical $\hbar \to 0$ metric (3.3) becomes:

$$d\tau^2 = dt^2 + 2\frac{q\phi_0}{mc^2}dtd\tau + \frac{q^2{\phi_0}^2}{m^2c^4}d\tau^2 . \tag{4.1}$$

It will be seen that this is quadratic in both $d\tau$ and dt, so we can solve this equation either way and obtain the same result. Choosing to write the quadratic in dt we have:

$$0 = dt^2 + 2\frac{q\phi_0}{mc^2}d\tau dt - \left(1 - \frac{q^2{\phi_0}^2}{m^2c^4}\right)d\tau^2 . \tag{4.2}$$

Via the quadratic equation this solves to:

$$dt = -\frac{q\phi_0}{mc^2}d\tau \pm \sqrt{\frac{q^2{\phi_0}^2}{m^2c^4}d\tau^2 + \left(1 - \frac{q^2{\phi_0}^2}{m^2c^4}\right)d\tau^2} = -\frac{q\phi_0}{mc^2}d\tau \pm d\tau = \left(\pm 1 - \frac{q\phi_0}{mc^2}\right)d\tau . \tag{4.3}$$

Then, imposing the condition that when $q=0$ or $\phi_0 = 0$ we must have $dt = d\tau$ so that in the absence of any electromagnetic interaction (or motion or gravitation) the coordinate time flows at the same rate as the proper time, we can discard the minus sign in (4.3), obtaining the simplified:

$$\frac{dt}{d\tau} = 1 - \frac{q\phi_0}{mc^2} . \tag{4.4}$$

With $d\tau$ segregated this is alternatively written as:

$$d\tau = \frac{1}{1 - \frac{q\phi_0}{mc^2}}dt . \tag{4.5}$$

The above (4.5) is the exact quadratic solution for the "peculiar" line element (3.5) at rest and absent gravitation. So (3.5) is the general case of (4.5), obtained by restoring motion via a Lorentz transform and gravitational fields by curving the spacetime. And (3.3) to which (4.4), (4.5) is the at rest solution absent gravitation, is just an algebraic variant of the well-established $m^2c^2 = \pi_\sigma \pi^\sigma$ which in turn is merely the relativistic relation $m^2c^2 = p_\sigma p^\sigma$ with local U(1) gauge symmetry.

Now, it is well-established from Special and General Relativity that when two clocks are in relative motion and / or are differently-situated in a gravitational potential, the ratio of the time coordinate element to the proper time element $dt/d\tau \neq 1$. This is time dilation, and when multiplied through by mc^2 to obtain $E = p^0 = mc^2 \cdot dt/d\tau$ this also gives us the total energy content of the material body with mass *m*. Yet (4.4) and (4.5) indicate that *even at rest and absent gravitation*, whenever there is a test charge with $q/m \neq 0$ in a proper scalar potential $\phi_0 \neq 0$ we continue to have $dt/d\tau \neq 1$. This result – which is brand new physics – teaches *that there are also time dilations which occur whenever there are electromagnetic interactions*. So we now must study these electromagnetic time dilations and come to understand their operational meaning and how they are observed in the natural world.

5. Derivation of Electromagnetic Interaction Time Dilations using an Inequivalence Principle

We observed earlier following (3.5) that one of the perplexing features of (3.3) and (3.5) is that they are functions $d\tau(q/m)$ of the q/m ratio of a test charge. But of course, the line element $ds = cd\tau$ cannot change when q/m changes, but must be invariant under such changes. So too, field theory mandates that the background fields A^μ and $g_{\mu\nu}$ also be invariant when q/m changes. So the question now arises, how do we ensure that (3.3) and (3.5) adhere to this mandate?

Ever since Galileo's legendary Pisa experiment it has been known that if two different masses m and $m' \neq m$ are dropped under the very same circumstances in the very same gravitational field, the motion will be exactly the same for each mass. This came to be understood as signifying an experimental equality between gravitational and inertial mass. By elevating this to the equivalence principle, Einstein was able to find a geometric way of formulating gravity. This is seen by the absence of the mass m in the gravitational motion that is part of (2.12). But for electromagnetism – in fundamental contrast to gravitation – two different test charges with q/m and $q'/m' \neq q/m$ do *not* exhibit identical motions even in identical electromagnetic fields under identical circumstances, as seen by the presence of this q/m ratio in the Lorenz Force motion of (2.12). This is understood to signify an experimental *inequality* between electrical mass a.k.a. charge and inertial mass. So now, we formally elevate this to an *inequivalence principle* which plays the same role in electrodynamics that the equivalence principle plays in gravitation, by taking the affirmative step of postulating a brand new symmetry principle which mandates as follows:

> **Charge-to-Mass Ratio Gauge Symmetry Postulate:** The metric interval $d\tau$ and background fields A^μ and $g_{\mu\nu}$, and by implication $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, must remain invariant under any and all transformations which re-scale, i.e. *re-gauge* the charge-to-mass ratio via a re-gauging transformation $q/m \rightarrow q'/m' \neq q/m$.

To implement this principle, we first inventory all of the physical numbers and objects appearing in the "peculiar" quadratic metric (3.3). These are the speed of light *c*, the line element $d\tau$, the metric tensor $g_{\mu\nu}$ containing the gravitational field, the gauge field A^μ which is the electromagnetic potential, the q/m ratio, and the coordinate elements dx^μ. So, under a re-gauging $q/m \rightarrow q'/m' \neq q/m$ of the charge-to-mass ratio, we of course require the speed of light to remain invariant, $c \rightarrow c' \equiv c$. But we also require, by the above symmetry principle, that $d\tau \rightarrow d\tau' \equiv d\tau$, $g_{\mu\nu} \rightarrow g'_{\mu\nu} \equiv g_{\mu\nu}$ and $A^\mu \rightarrow A'^\mu \equiv A^\mu$ also remain invariant. So the only objects remaining which may be transformed when we re-gauge $q/m \rightarrow q'/m' \neq q/m$ are the coordinate elements dx^μ. We know very well from the Special and General Theories of Relativity that the observed $dx^\mu \rightarrow dx'^\mu \neq dx^\mu$ do in fact change when two different observers are in relative motion or have different placements in a gravitational field. And (4.4), (4.5) already indicate that this is also true of at least the time element $dx^0 = cdt$ when there are electrodynamic interactions.

So now we work from (3.3) to *define* a coordinate transformation $dx^\mu \to dx'^\mu \neq dx^\mu$ which occurs whenever we transform $q/m \to q'/m' \neq q/m$ in accordance with these symmetries, via:

$$\begin{aligned} c^2 d\tau^2 &= \frac{d\tau^2}{m^2}\pi_\sigma \pi^\sigma = \left(dx_\sigma + \frac{q}{mc^2} A_\sigma c d\tau \right)\left(dx^\sigma + \frac{q}{mc^2} A^\sigma c d\tau \right) = g_{\mu\nu} \mathcal{D}x^\mu \mathcal{D}x^\nu \\ \to c^2 d\tau'^2 &= \frac{d\tau^2}{m^2}\pi'_\sigma \pi'^\sigma \equiv c^2 d\tau^2 = \left(dx'_\sigma + \frac{q'}{m'c^2} A_\sigma c d\tau \right)\left(dx'^\sigma + \frac{q'}{m'c^2} A^\sigma c d\tau \right) = g_{\mu\nu} \mathcal{D}x'^\mu \mathcal{D}x'^\nu \end{aligned} . \quad (5.1)$$

Note that $\mathcal{D}x^\mu = dx^\mu + \left(q/mc^2\right) A^\mu c d\tau \to \mathcal{D}x'^\mu = dx'^\mu + \left(q'/m'c^2\right) A^\mu c d\tau$ is the transformation for the gauge-covariant coordinate elements $\mathcal{D}x^\mu$. If we then apply the $\hbar = 0$ classical approximation from (1.6) which sets $\left[p_\sigma, A^\sigma \right] = 0$, the above transformation $dx^\mu \to dx'^\mu \neq dx^\mu$ becomes:

$$\begin{aligned} c^2 d\tau^2 &= g_{\mu\nu} dx^\mu dx^\nu + 2\frac{q}{mc^2} A_\sigma dx^\sigma c d\tau + \frac{q^2}{m^2 c^4} A_\sigma A^\sigma c^2 d\tau^2 \\ \to c^2 d\tau'^2 \equiv c^2 d\tau^2 &= g_{\mu\nu} dx'^\mu dx'^\nu + 2\frac{q'}{m'c^2} A_\sigma dx'^\sigma c d\tau + \frac{q'^2}{m'^2 c^4} A_\sigma A^\sigma c^2 d\tau^2 \end{aligned} . \quad (5.2)$$

Now we move to a rest frame and remove all gravitation to directly deduce what happens to the time coordinate when we re-gauge $q/m \to q'/m' \neq q/m$. This is the exact same calculation we did from (4.1) to (4.5), except now we have some transformed objects annotated with "primes." So with $A^\mu = \left(\phi_0, \mathbf{0}\right)$ and $dx^\mu = \left(cdt, \mathbf{0}\right)$ and $g_{\mu\nu} = \eta_{\mu\nu}$ the above becomes (contrast (4.1)):

$$d\tau^2 = dt^2 + 2\frac{q\phi_0}{mc^2} dt d\tau + \frac{q^2 \phi_0{}^2}{m^2 c^4} d\tau^2 = dt'^2 + 2\frac{q'\phi_0}{m'c^2} dt' d\tau + \frac{q'^2 \phi_0{}^2}{m'^2 c^4} d\tau^2 . \quad (5.3)$$

This contains a first quadratic for dt and a second quadratic for dt'. We already have the solution for dt, which is (4.4). So the solution for dt', shown together with (4.4) for dt, is:

$$\frac{dt}{d\tau} = 1 - \frac{q\phi_0}{mc^2}; \qquad \frac{dt'}{d\tau} = 1 - \frac{q'\phi_0}{m'c^2} . \quad (5.4)$$

Now, because of the above symmetry postulate, $d\tau$ is the same invariant object in each of $dt/d\tau$ and $dt'/d\tau$ above. Likewise, c and ϕ_0 are also the same. And we used the same $\eta_{\mu\nu}$ to derive each of (5.4). Therefore, with two different massive charged bodies both at rest *in the same proper potential* ϕ_0, one with q/m and the other with q'/m', we deduce from (5.4) that the ratio:

$$\frac{dt}{dt'} = \frac{1 - q\phi_0 / mc^2}{1 - q'\phi_0 / m'c^2} . \tag{5.5}$$

Because the above compares measurements of time, we should be more specific about what is meant by the rate at which time flows for various charged bodies. The meaning and construction of so-called "geometrodynamic clocks" has been widely developed in the literature, see, e.g. section 5.2 of Ohanian's [11]. What (5.5) tells us is that if we start with an electrically-neutral material body which qualifies as a true geometrodynamic clock (g-clock), for example, a cesium oscillator through which a second is defined in the International System of Units (SI) by the standard of 9,192,631,770 oscillation "ticks," then if that clock is charged and placed into an electromagnetic proper potential ϕ_0, the rate of time signaling will be altered based on (5.5). So suppose that we wish to measure the ratio (5.5). One experiment we might do is to start with two identical, electrically-neutral g-clocks. We leave the first g-clock neutral so it maintains $q = 0$. We then charge the second g-clock to $q' \neq 0$. We then use the neutral $q = 0$ g-clock as a laboratory clock to measure the laboratory time element dt, and compare this to the dt' element measured by oscillations of the second $q' \neq 0$ clock. So for this experiment, with $q = 0$ (5.5) becomes:

$$\frac{dt}{dt'} = \frac{1}{1 - \dfrac{q'\phi_0}{m'c^2}} . \tag{5.6}$$

In Relativity Theory the time dilation factors $\gamma_v \equiv dt / d\tau = 1/\sqrt{1 - v^2 / c^2}$ for motion and $\gamma_g \equiv dt / d\tau = 1/\sqrt{g_{00}}$ for gravitational interaction associate dt with the time ticked off by the laboratory clock of an observer at rest or outside a gravitational field, and $d\tau$ with the proper time ticked off by an observed clock in relative motion or inside the gravitational field. The derivations of these two relativistic relations are reviewed in Appendix B. So in (5.6), we make a parallel association of dt with the neutral laboratory clock resting with an observer. Then, absent any gravitation or motion we now equate dt' with $d\tau$ so that $d\tau \equiv dt'$ becomes the proper time ticked off by the charged q' / m' clock being observed. With this we have:

$$\frac{dt}{d\tau} = \frac{1}{1 - \dfrac{q'\phi_0}{m'c^2}} . \tag{5.7}$$

Finally, as a matter of notational convention, because (5.7) compares a neutral $q = 0$ laboratory g-clock with dt, to a charged $q' \neq 0$ g-clock with $d\tau$, the primes are no longer needed, so we re-denote q' to q and m' to m. We then use (5.7) so re-notated to define an electromagnetic time dilation factor γ_e comparing the ratio of time ticked off by the neutral g-clock of an observer to time ticked off by an observed charged g-clock, as follows:

$$\gamma_{em} \equiv \frac{dt}{d\tau} = \frac{1}{1-\dfrac{q\phi_0}{mc^2}} = 1 + \frac{q\phi_0}{mc^2} + \left(\frac{q\phi_0}{mc^2}\right)^2 + \left(\frac{q\phi_0}{mc^2}\right)^3 + \left(\frac{q\phi_0}{mc^2}\right)^4 + \ldots = \sum_{n=0}^{\infty}\left(\frac{q\phi_0}{mc^2}\right)^n . \quad (5.8)$$

Above, $q\phi_0 / mc^2$ is the key dimensionless ratio which determines the numerical size of γ_{em}. Because $E_e = q\phi_0$ is the energy of electromagnetic interaction between the test charge q and the source of the potential ϕ_0, we see that $q\phi_0 / mc^2 = E_e / E_0$ is the dimensionless ratio of this electromagnetic interaction energy to the rest energy $E_0 = mc^2$ of the test charge.

It is illustrative to examine (5.8) in the special case where a positive charge Q generates a Coulomb proper scalar potential $\phi_0 = k_e Q / r$, with $k_e = 1/4\pi\varepsilon_0 = \mu_0 c^2 / 4\pi = 10^{-7} c^2 \mathrm{N/A}^2$ being the Coulomb constant. For a test body with positive charge q and mass m at rest in the potential at a distance r from Q, the electromagnetic interaction energy $E_e = q\phi_0 = k_e Qq / r$ is repulsive because lower energy states are achieved by two like-charges moving farther apart. The ratio of this interaction energy to the test charge rest mass is $q\phi_0 / mc^2 = k_e Qq / mc^2 r$. Here, (5.8) becomes:

$$\gamma_{em} \equiv \frac{dt}{d\tau} = \frac{1}{1-\dfrac{k_e Qq}{mc^2 r}} = 1 + \frac{k_e Qq}{mc^2 r} + \left(\frac{k_e Qq}{mc^2 r}\right)^2 + \left(\frac{k_e Qq}{mc^2 r}\right)^3 + \left(\frac{k_e Qq}{mc^2 r}\right)^4 + \ldots = \sum_{n=0}^{\infty}\left(\frac{k_e Qq}{mc^2 r}\right)^n . \quad (5.9)$$

Because $dt / d\tau > 1$ when Q and q both have the same sign and are therefore repelling, the neutral laboratory g-clock will emit more "tick" signals during a given time than the observed charged g-clock being observed. So we learn that time dilates for a *repulsive* electromagnetic interactions between two like-charges, just as it dilates for the attractive gravitational interaction between what are always two like-masses. That is, time dilation occurs for interactions between *like charges*, which interactions for gravitation are attractive and for electromagnetism are repulsive, owing to the respective spin-2 gravitons and spin-1 photons that quantum-mediate these interactions. This also means that time contracts for attractive electromagnetic interactions between unlike charges.

As a numeric benchmark for classical interactions, consider that the two charges each have $Q = q = 1\mathrm{C}$, the test particle has a rest mass $m = 1\mathrm{kg}$, and the separation $r = 1\mathrm{m}$. Therefore, the dimensionless ratio of interaction to rest energy $q\phi_0 / mc^2 = k_e / c^2 = 10^{-7}$, and the time dilation is $\gamma_{em} \cong 1 + 10^{-7}$ (to parts per 10^{-14}, from the next-higher-order term in (5.9)). At the same time, this interaction energy $q\phi_0 = k_e = 10^{-7} c^2 \,\mathrm{J} = 8.897 \times 10^9 \,\mathrm{J}$ is exceedingly large. The release of this much energy per second would yield a power of approximately 8.897 GW, which roughly approximates seven or eight nuclear power plants, or four times the power of the Hoover Dam, or the power of about seventy five jet engines, or the power output of a single space shuttle launch, or of a single lightning bolt. So it takes tremendously large electromagnetic interactions to produce very small time dilations. For electromagnetic interactions encountered in daily experience, this dilation will be much smaller. For example, a kW-order interaction would dilate time to about one part in 10^{14}. For a cesium clock ticking every 1.09×10^{-10} seconds, the discrepancy for a kW-order interaction would be about 1 tick per ten thousand seconds – about 2.75 hours.

Knowing from (5.8) that time dilates for repulsive electromagnetic interactions, one can design an even-simpler experiment to test for these time dilations, at least qualitatively: take a first neutral g-clock, and synchronize it with a second neutral g-clock. Then charge the second g-clock and use the first g-clock as a control to measure its time oscillations. Because there will now be an internal repulsive self-interaction energy between and among the various elemental parts of the charged clock, the mere charging of the clock should cause the oscillatory period to dilate.

As we now also show, the well-known energy content of electromagnetically-interacting bodies provides direct empirical evidence time really does dilate in accordance with (5.8) and (5.9).

6. The Energy Content of Electromagnetically-Interacting, Moving and Gravitating Material Bodies

Einstein's pioneering paper [12] first used a time dilation factor γ_v in the simple calculation $E=mc^2\gamma_v=mc^2\cdot dt/d\tau=mc^2/\sqrt{1-v^2/c^2}\cong mc^2+\frac{1}{2}mv^2$ to uncover the rest energy relation now known as $E_0=mc^2$. In this calculation, the Newtonian kinetic energy $E_v=\frac{1}{2}mv^2$ is shown to be a comparatively tiny addition to the huge rest energy $E_0=mc^2$ of a mass *m*, for non-relativistic velocities $v/c\ll1$. Moreover, the kinetic energy in general is seen to be the *nonlinear* $E_v=mc^2\cdot(dt/d\tau-1)=mc^2\left(1/\sqrt{1-v^2/c^2}-1\right)$ in which the Newtonian $\frac{1}{2}mv^2$ is the lowest-order term in the McLaurin series $E_{\text{kin}}=\frac{1}{2}mv^2\Sigma_{n=0}^{\infty}\left((2n+1)!!/2^n(n+1)!\right)\left(v^2/c^2\right)^n$, with $\frac{1}{2}mv^2$ multiplied by higher order terms v^2/c^2, v^4/c^4, v^6/c^6, etc. times a series of numeric coefficients.

Einstein later showed in [3] that this carries over to gravitational energies, but now with a gravitational time dilation $\gamma_g=dt/d\tau=1/\sqrt{g_{00}}$ which leads to the energy content relation $E=mc^2\gamma_g=mc^2\cdot dt/d\tau=mc^2/\sqrt{g_{00}}$. For a Schwarzschild metric with $g_{00}=1-2GM/c^2r$ this produces $E=mc^2/\sqrt{1-2GM/c^2r}\cong mc^2+GMm/r$. Here, the negative* Newtonian gravitational interaction energy $-E_g=GMm/r$ is seen to be a comparatively tiny addition to the rest energy mc^2 for weak gravitational interactions in which the ratio of gravitational energy to rest energy $GM/c^2r=(GMm/r)/mc^2\ll1$. Here too, $-E_g=mc^2(dt/d\tau-1)=mc^2/\left(1/\sqrt{1-2GM/c^2r}-1\right)$ is a nonlinear energy, with a series $-E_g=(GMm/r)\Sigma_{n=0}^{\infty}\left((2n+1)!!/(n+1)!\right)\left(GM/c^2r\right)^n$. In this

* Even though the mass *m* gains energy in the gravitational field and thus increases its ability to do work, e.g., by falling toward *M*, the gravitational interaction energy must be negative. This is because gravitation is an attractive interaction so that lower energy states must correlate with the two masses moving closer.

situation, the Newtonian GMm/r is multiplied by a higher-order succession of terms GM/c^2r, $\left(GM/c^2r\right)^2$, $\left(GM/c^2r\right)^3$ etc. terms times a series of coefficients.

As it happens, the electromagnetic time dilation (5.8) when multiplied through by the rest energy mc^2 yields similar information about the energy content of electromagnetically-interacting bodies. Working from (5.8) in the same way as reviewed just above, it is readily calculated that:

$$E = mc^2\gamma_{em} = mc^2\frac{dt}{d\tau} = \frac{mc^2}{1-\frac{q\phi_0}{mc^2}} = mc^2 + q\phi_0\left(1+\frac{q\phi_0}{mc^2}+\left(\frac{q\phi_0}{mc^2}\right)^2+\ldots\right) = mc^2 + q\phi_0\sum\nolimits_{n=0}^{\infty}\left(\frac{q\phi_0}{mc^2}\right)^n . \quad (6.1)$$

Here, the known interaction energy $E_e = q\phi_0$ is seen to be a comparatively tiny addition to the rest energy mc^2 for interactions in which the dimensionless ratio of electromagnetic interaction energy $q\phi_0$ to rest energy mc^2 is very small, $q\phi_0/mc^2 \ll 1$. Here, when $q\phi_0/mc^2$ grows measurably larger – in a new result that does not appear to have been reported in the literature at least for classical electromagnetic interactions – *the electromagnetic interaction energy becomes non-linear just like special and general relativistic energies*. Now, in general, electromagnetic interaction energy is given by the non-linear series $E_e = q\phi_0\Sigma_{n=0}^{\infty}\left(q\phi_0/mc^2\right)^n$, and the higher order multipliers of the known energy $q\phi_0$ are $q\phi_0/mc^2$, $\left(q\phi_0/mc^2\right)^2$, $\left(q\phi_0/mc^2\right)^3$ etc. So for a Coulomb potential $\phi_0 = k_eQ/r$ (6.1) above becomes:

$$\begin{aligned} E = mc^2\gamma_{em} = mc^2\frac{dt}{d\tau} &= \frac{mc^2}{1-\frac{k_eQq}{mc^2r}} = mc^2 + \frac{k_eQq}{r}\left(1+\frac{k_eQq}{mc^2r}+\left(\frac{k_eQq}{mc^2r}\right)^2+\left(\frac{k_eQq}{mc^2r}\right)^3+\ldots\right) \\ &= mc^2 + \frac{k_eQq}{r}\sum\nolimits_{n=0}^{\infty}\left(\frac{k_eQq}{mc^2r}\right)^n \end{aligned} . \quad (6.2)$$

Just as with $E_v = \frac{1}{2}mv^2$ for motion and $-E_g = GMm/r$, the Coulomb interaction energy $E_e = k_eQq/r$ is likewise a tiny correction to the to the rest energy mc^2, precisely as is observed. But the complete energy $E_e = \left(k_eQq/r\right)\Sigma_{n=0}^{\infty}\left(k_eQq/mc^2r\right)^n$ is non-linear. For the classical benchmark $q\phi_0/mc^2 = k_e/c^2 = 10^{-7}$ given at the end of the last section, the interaction energy $q\phi_0 = k_e = 10^{-7}c^2\,\mathrm{J} = 8.897\times10^9\,\mathrm{J}$ is increased by a scant one part in 10^7 owing to the first correction term k_eQq/mc^2r in the series. Nonetheless, (6.2) gives a precise prediction of the magnitude of these newly-predicted non-linear corrections.

When there are both motion and gravitation, the special and general relativistic time dilations are compounded by multiplication, so the total time dilation $\gamma = dt/d\tau = \gamma_v\gamma_g$, with a total energy content $E = mc^2\gamma_v\gamma_g$. We may therefore expect that when there are electromagnetic interactions in addition to motion and gravitation, $\Gamma \equiv dt / d\tau = \gamma_v \gamma_g \gamma_{em}$ will be the complete time dilation, and the total energy content of the material body will be $E = mc^2\Gamma = mc^2 \gamma_v \gamma_g \gamma_{em}$. If we compute this using while also showing the linear limit, we obtain:

$$\begin{aligned} E &= mc^2\Gamma = mc^2 \frac{dt}{d\tau} = mc^2 \gamma_v \gamma_g \gamma_{em} = mc^2 \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \frac{1}{\sqrt{1-\frac{2GM}{c^2 r}}} \frac{1}{1-\frac{k_e Qq}{mc^2 r}} \\ &\cong mc^2 \left(1+\frac{1}{2}\frac{v^2}{c^2}\right)\left(1+\frac{GM}{c^2 r}\right)\left(1+\frac{k_e Qq}{mc^2 r}\right) \\ &= mc^2 + \frac{1}{2}mv^2 + \frac{GMm}{r} + \frac{1}{2}\frac{GMm}{c^2 r}v^2 + \frac{k_e Qq}{r} + \frac{1}{2}\frac{k_e Qq}{c^2 r}v^2 + \frac{GM}{r}\frac{k_e Qq}{c^2 r} + \frac{1}{2}\frac{GM}{c^2 r}\frac{k_e Qq}{c^2 r}v^2 \end{aligned} \quad . \tag{6.3}$$

What we see here, in succession, are 1) the rest energy mc^2, 2) the kinetic energy of the mass *m*, 3) the gravitational interaction energy of the mass, 4) the kinetic energy of the gravitational energy, 5) the Coulomb interaction energy of the charged mass, 6) the kinetic energy of the Coulomb energy, 7) the gravitational energy of the Coulomb energy and 8) the kinetic energy of the gravitational energy of the Coulomb energy. Numbers 1 through 4 above are standard results that are obtained when one applies the Special and General theories at the same time. Numbers 1 through 4 are well-established in relativity theory. Numbers 5 through 8 incorporate the new findings (5.8) and (5.8) of an electromagnetic time dilation. All of these accords entirely with empirical observations of the linear limits of these same energies.

Of course, $E = p^0 c$ in (6.3) is the time component of the energy-momentum four-vector $cp^\mu = (E, c\mathbf{p}) = mcdx^\mu / d\tau$. By the chain rule, the relativistic four velocity $dx^\mu / d\tau = (dx^\mu / dt)(dt / d\tau)$, and because $dx^\mu = (cdt, d\mathbf{x})$ the ordinary four-velocity $dx^\mu / dt = (c, d\mathbf{x} / dt) = (c, \mathbf{v}) \equiv v^\mu$. Because the composite time dilation $\Gamma \equiv dt / d\tau = \gamma_v \gamma_g \gamma_{em}$ is validated at least at lowest order by the energy content shown in (6.3), we may combine the foregoing to deduce that $dx^\mu / d\tau = \gamma_v \gamma_g \gamma_e v^\mu = \Gamma v^\mu$. Therefore, when all of motion and gravitation and electrodynamic interactions are present, the Lorentz four-vector p^μ in (1.6), of which (6.3) sits in the time component, is deduced to be:

$$cp^\mu = \begin{pmatrix} E & c\mathbf{p} \end{pmatrix} = mcdx^\mu / d\tau = mcv^\mu \gamma_v \gamma_g \gamma_{em} = mcv^\mu \Gamma . \tag{6.4}$$

Likewise, we may deduce that in the "peculiar" quadratic metric of (3.3), the coordinate elements with all of motion and gravitation and electrodynamics are $dx^\mu = \gamma_v \gamma_g \gamma_{em} v^\mu d\tau = \Gamma v^\mu d\tau$. This is a way to reintroduce motion and gravitation and Lorentz covariance into the quadratic solution (4.4) obtained at rest and absent gravitation, and into the consequent (5.8) for a neutral laboratory g-clock used to measure time signals from an identical g-clock which is charged.

Finally, because gauge symmetry results in replacing $p^\mu \to \pi^\mu = p^\mu + qA^\mu / c$, it is necessary for A^μ to transform in the same general covariant manner as p^μ. Absent gravitation, and absent being aware of electromagnetic time dilations, the four-potential is normally defined in relation to motion by $A^\mu = \phi_0 v^\mu (dt/d\tau)/c = \phi_0 \gamma_v v^\mu / c$, where $v^\mu = (c, \mathbf{v})$. But when all time dilations are considered, energy content is changed, and general covariance requires that this now be extended to:

$$A^\mu = (\phi \quad \mathbf{A}) = \phi_0 (v^\mu / c)(dt/d\tau) = \phi_0 v^\mu \gamma_v \gamma_g \gamma_{em} / c = \phi_0 v^\mu \Gamma / c. \tag{6.5}$$

We see that both (6.4) and (6.5) contain the same time dilation and motion kernel $v^\mu \Gamma$.

7. Energy-Momentum Gradients, and Heisenberg Rules for Momentum Commutation in view of Electromagnetic Time Dilations

When the energy-momentum of a particle depends only on its rest mass and its motion, then $\gamma_g = 1$ and $\gamma_{em} = 1$, so (6.4) of course becomes the special relativistic $p^\mu = mv^\mu \gamma_v$. Because $\partial_\alpha \gamma_v = 0$ and $\partial_\alpha v^\mu = 0$, this has no spacetime dependency, which may be expressed differentially via $\partial_\alpha p^\mu = m\partial_\alpha (\gamma_v v^\mu) = 0$. When there is a gravitational field, then $p^\mu = mv\gamma_v \gamma_g$ and there is a spacetime dependency, because $\gamma_g(t, \mathbf{x}) = 1/\sqrt{g_{00}(t, \mathbf{x})}$ is a function of space and time. Thus, the four-gradient $\partial_\alpha \gamma_g = -\frac{1}{2}\partial_\alpha g_{00} / (g_{00})^{1.5} \neq 0$ and $\partial_\alpha p^\mu = m\partial_\alpha (\gamma_v \gamma_g v^\mu) \neq 0$. However, because g_{00} is a component of the metric tensor with gravitational-covariant derivative $\partial_{;\alpha} g_{\mu\nu} = 0$, the gravitational covariant derivative $\partial_{;\alpha} p^\mu = m\partial_{;\alpha} (\gamma_v \gamma_g v^\mu) = 0$ of the energy-momentum is still zero.

For electromagnetic interactions, this is no longer the case. Now, the energy-momentum $p^\mu(t, \mathbf{x}) = mv^\mu \gamma_v \gamma_g \gamma_{em}(t, \mathbf{x})$ takes on an explicit spacetime dependency, because as deduced in (5.8), the electromagnetic time dilation is a function $\gamma_{em}(t, \mathbf{x}) = 1/(1 - q\phi_0(t, \mathbf{x})/mc^2)$ of spacetime, because the proper potential $\phi_0(t, \mathbf{x})$ is (or may be) a function of space and time. Expressed differentially, $\partial_{;\alpha} \gamma_{em} = \partial_\alpha \gamma_{em} \neq 0$, so that $\partial_\alpha p^\mu \neq 0$ and even $\partial_{;\alpha} p^\mu \neq 0$. This spacetime dependency of the energy momentum stemming from $\phi_0(t, \mathbf{x})$ has a number of useful and important properties that it now behooves us to explore.

Absent gravitation, with $g_{\mu\nu} = \eta_{\mu\nu}$ thus $\gamma_g = 1$, the complete time dilation $\Gamma = dx^\mu / d\tau = \gamma_v \gamma_{em}$. From the time component of (6.5) we find that $\phi = \phi_0 \gamma_v \gamma_{em}$ which we invert to $\phi_0 = \phi / \gamma_v \gamma_{em}$. We then use this to write the electromagnetic time dilation (5.8) as:

$$\gamma_{em} = \frac{1}{1 - q\phi_0 / mc^2} = \frac{1}{1 - q\phi / \gamma_v \gamma_{em} mc^2} = \frac{\gamma_v \gamma_{em}}{\gamma_v \gamma_{em} - q\phi / mc^2} . \tag{7.1}$$

Upon dividing through by γ_{em} then taking the reciprocal of both sides we obtain:

$$1 = \frac{\gamma_v \gamma_{em} - q\phi / mc^2}{\gamma_v} = \gamma_{em} - \frac{q\phi}{mc^2 \gamma_v} , \tag{7.2}$$

which easily restructures into an alternative expression for γ_{em}, namely:

$$\gamma_{em} = 1 + \frac{q\phi}{mc^2 \gamma_v} . \tag{7.3}$$

The time component of (6.4) contains the total energy $E = mc^2 \gamma_v \gamma_{em}$ with $\gamma_g = 1$, see also (6.3), which energy, in view of (7.3), may be written as:

$$E = mc^2 \gamma_v \left(1 + \frac{q\phi}{mc^2 \gamma_v} \right) = mc^2 \gamma_v + q\phi . \tag{7.4}$$

So the total energy $E = mc^2 \gamma_v \gamma_{em}$ is alternatively written as the rest-plus motion energy $mc^2 \gamma_v$, plus the electromagnetic interaction potential energy $q\phi$. So if we take the space-gradient $-\partial^\alpha = \nabla^\alpha = \nabla$ of the above, then apply the relation $\nabla \phi = -\left(\mathbf{E} + \dot{\mathbf{A}} / c \right)$ between the potential gradient and the electric field $\mathbf{E}$ and time derivative $\dot{\mathbf{A}} = \partial \mathbf{A} / \partial t$ of the three-potential, we obtain:

$$-\partial^j E = \nabla E = -q \partial^j \phi = q \nabla \phi = -q \left(E^j + \dot{A}^j / c \right) = -q \left(\mathbf{E} + \dot{\mathbf{A}} / c \right) . \tag{7.5}$$

It is also useful to separately take the gradient of the time dilation (7.3), namely:

$$\partial^j \gamma_{em} = \frac{q}{mc^2 \gamma_v} \partial^j \phi = \frac{1}{mc^2 \gamma_v} q \left(E^j + \dot{A}^j / c \right) = -\frac{q}{mc^2 \gamma_v} \nabla \phi = -\nabla \gamma_{em} = \frac{1}{mc^2 \gamma_v} q \left(\mathbf{E} + \dot{\mathbf{A}} / c \right) . \tag{7.6}$$

We will find it useful to include γ_v inside the gradient and write this as a gradient of the total time dilation $dt / d\tau = \gamma_v \gamma_{em} = E / mc^2$, as such:

$$\partial^j \left(\frac{dt}{d\tau} \right) = \partial^j \left(\gamma_v \gamma_{em} \right) = \partial^j \left(\frac{E}{mc^2} \right) = \frac{q}{mc^2} \partial^j \phi = \frac{q}{mc^2} \left(E^j + \dot{A}^j / c \right) . \tag{7.7}$$

Having deduced the time component of (6.4) in flat spacetime, now we turn to the space components $c\mathbf{p} = cp^i = mcv^i\gamma_v\gamma_{em}$. Using (7.6), we find that:

$$\nabla c\mathbf{p} = -\partial^j cp^i = -mcv^i\gamma_v\partial^j\gamma_{em} = -\frac{1}{mc^2\gamma_v}mcv^i\gamma_v q\partial^j\phi . \tag{7.8}$$

Multiply through by $-E = -mc^2\gamma_v\gamma_{em}$ then applying $cp^i = mc\gamma_{em}\gamma_v v^i$ and $\partial^j\phi = E^j + \dot{A}^j / c$ yields:

$$\begin{aligned} &mc^2\gamma_v\gamma_{em}\partial^j cp^i = mcv^i\gamma_v\gamma_{em}q\partial^j\phi \\ &= E\partial^j cp^i = q\partial^j\phi cp^i = q\left(E^j + \dot{A}^j / c\right)cp^i = -E\nabla c\mathbf{p} = -q\nabla\phi c\mathbf{p} = q\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)c\mathbf{p} \end{aligned} . \tag{7.9}$$

With $cp^\mu = \begin{pmatrix} E & c\mathbf{p} \end{pmatrix}$ we then assemble equations (7.5) and (7.9) into spacetime-covariant form:

$$E\partial^j cp^\mu = q\partial^j\phi cp^\mu = q\left(E^j + \dot{A}^j / c\right)cp^\mu = -E\nabla cp^\mu = -q\nabla\phi cp^\mu = q\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)cp^\mu . \tag{7.10}$$

This now reveals the heuristic rule $-E\nabla \mapsto -q\nabla\phi \mapsto q\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)$ for when a gradient is applied to the flat spacetime four-momentum $cp^\mu = mc\gamma_v\gamma_{em}v^\mu$ of (6.4) with electromagnetic time dilation.

Another important and useful consequence of the relation $cp^\mu = mcv^\mu\gamma_v\gamma_{em}$ is that the three-momentum $p^i = \mathbf{p}$ no longer commutes as between momenta oriented along orthogonal spacetime axes, that is, $\left[p^i, p^j\right] = p^i p^j - p^j p^i \neq 0$ when $i \neq j$. Written in vector notation, this means that the momentum self-cross product $\mathbf{p}\times\mathbf{p} \neq 0$. To derive this with specificity, we begin with Heisenberg's canonical commutation relation $[p_x, x] = -i\hbar$ which of course underlies the uncertainty principle. It is easily calculated that $[p_x, x^n] = -i\hbar n x^{n-1}$. Moreover, because elementary calculus teaches that $\partial_x x^n = nx^{n-1}$ we may combine the foregoing into $[p_x, x^n] = -i\hbar n x^{n-1} = -i\hbar\partial_x x^n$. Therefore, for any function $b(\mathbf{x})$ expansible as a Maclaurin series in $\mathbf{x}$, noting that $\partial^i = -\nabla = -\left(\partial_x, \partial_y, \partial_z\right)$, we may generalize this in well-known fashion to the well-established relation $\left[p^i, b\right] = i\hbar\partial^i b$. This is then generalizable to any vector, tensor, etc. object $O(\mathbf{x})$ whereby $\left[p^i, O\right] = i\hbar\partial^i O$ and $\left[O, p^j\right] = -i\hbar\partial^j O$. Thus, if we generalize to a vector $b \mapsto b^j(\mathbf{x})$, with the momentum to the left of the commutator, this becomes $\left[p^i, b^j\right] = i\hbar\partial^i b^j$. With momentum to the right of $b^j(\mathbf{x})$, and with renamed indexes, the right-side relation is $\left[b^i, p^j\right] = -i\hbar\partial^j b^i$.

Now, the space components of $p^i = mv^i\gamma_v\gamma_{em}$ include both the non-relativistic momentum $p^i_{NR} = mv^i$ and the time dilation multiplier $dt/d\tau = \gamma_v\gamma_{em}$, that is, $p^i = p^i_{NR}\gamma_v\gamma_{em}$. However, the Heisenberg operator (op) matrices for momentum correspond to the *non-relativistic momentum only*, $p^i_{NR} \Leftrightarrow p^i_{op}$. With this in mind, were we to assign $b^j \mapsto p^j$ then the left-side commutator relation would become $\left[p^i, p^j\right] = i\hbar\partial^i p^j$. But, were we to instead assign $b^i \mapsto p^i$ then the right-side commutator this would become $\left[p^i, p^j\right] = -i\hbar\partial^j p^i$. So both relations together would imply that $i\hbar\partial^i p^j = -i\hbar\partial^j p^i$ is an antisymmetric tensor, which is not necessarily so. However, this overlooks the fact that *unlike any other vector*, the space-component momentum vector $p^i = mv^i\gamma_v\gamma_{em}$ is a hybrid momentum and spacetime vector. This is because it includes both $mv^i = p^i_{NR} \Leftrightarrow p^i_{op}$ which is a pure momentum against which functions of the space coordinates x^j are commuted, and because it also includes $\gamma_{em}(t,\mathbf{x}) = 1/\left(1-q\phi_0/mc^2\right)$ which is a function of spacetime because the proper scalar potential $\phi_0(t,\mathbf{x})$ is a function of spacetime. Specifically, $p^i(t,\mathbf{x}) = p^i_{NR}\gamma_v\gamma_{em}(t,\mathbf{x})$. Consequently, there is a self-commutativity wherein when we commute $\left[p^i, p^j\right]$, we are commuting the space-dependent portion of p^j to the left past the p^i_{NR} portion of p^i, while *simultaneously* commuting the space-dependent portion of p^i to the right past the p^j_{NR} portion of p^j. It is only the pure non-relativistic momentum which is self-commuting along all orthogonal pairs of space coordinates, $\left[p^i_{NR}, p^j_{NR}\right] = 0$.

Given the foregoing, if we write the two commuting momentum vectors as $cp^i = mcv^i\gamma_v\gamma_{em} = cp^i_{NR}\left(\gamma_v\gamma_{em}\right)$ and $cp^j = mcv^j\gamma_v\gamma_{em} = cp^j_{NR}\left(\gamma_v\gamma_{em}\right)$, and if we write the left-momentum and right-momentum commutativity relations as $\left[cp^i_0, \gamma_v\gamma_{em}\right] = i\hbar c\partial^i\left(\gamma_v\gamma_{em}\right)$ and $\left[\gamma_v\gamma_{em}, cp^j_0\right] = -i\hbar c\partial^j\left(\gamma_v\gamma_{em}\right)$, then also making use of $\left[p^i_{NR}, p^j_{NR}\right] = 0$, we may calculate:

$$\begin{aligned}
cp^i cp^j &= cp^i_{NR}\gamma_v\gamma_{em}cp^j_{NR}\gamma_v\gamma_{em} = \gamma_v\gamma_{em}cp^j_{NR}cp^i_{NR}\gamma_v\gamma_{em} + i\hbar c\partial^i\left(\gamma_v\gamma_{em}\right)cp^j_{NR}\gamma_v\gamma_{em} \\
&= cp^j_{NR}\gamma_v\gamma_{em}cp^i_{NR}\gamma_v\gamma_{em} - i\hbar c\partial^j\left(\gamma_v\gamma_{em}\right)cp^i_{NR}\gamma_v\gamma_{em} + i\hbar c\partial^i\left(\gamma_v\gamma_{em}\right)cp^j_{NR}\gamma_v\gamma_{em} \\
&= cp^j cp^i - i\hbar c\partial^j\left(\gamma_v\gamma_{em}\right)cp^i + i\hbar c\partial^i\left(\gamma_v\gamma_{em}\right)cp^j
\end{aligned} \quad . \tag{7.11}$$

Recasting this as a commutator, and then using (7.7), this becomes:

$$\begin{aligned}&\left[cp^i, cp^j\right] = -i\hbar c\partial^j\left(\gamma_v\gamma_{em}\right)cp^i + i\hbar c\partial^i\left(\gamma_v\gamma_{em}\right)cp^j \\ &= -i\hbar c\frac{q}{mc^2}\partial^j\phi cp^i + i\hbar c\frac{q}{mc^2}\partial^i\phi cp^j = -i\frac{\hbar cq}{mc^2}\left(E^j + \dot{A}^j / c\right)cp^i + \frac{\hbar cq}{mc^2}\left(E^j + \dot{A}^j / c\right)cp^j \ . \\ &= i\hbar c\frac{q}{mc^2}\nabla\phi c\mathbf{p} - i\hbar c\frac{q}{mc^2}\nabla\phi c\mathbf{p} = -i\frac{\hbar cq}{mc^2}\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)c\mathbf{p} + \frac{\hbar cq}{mc^2}\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)c\mathbf{p}\end{aligned} \qquad (7.12)$$

So the momentum self-commutator actually contains the gradient $\nabla\phi$ of the scalar potential, and, via $\nabla\phi = -\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)$, the electric field $\mathbf{E}$ and time derivative $\dot{\mathbf{A}} = \partial\mathbf{A} / \partial t$ of the three-vector potential. Note that $\left[cp^i, cp^j\right] = 0$ when $i = j$, so that this commutation relation only becomes non-zero, $\left[cp^i, cp^j\right] \neq 0$, as to orthogonal pairs of space coordinates where $i \neq j$.

Now let us additionally consider the circumstance where p^i is the momentum of an individual charged lepton (i.e., the electron, or the mu or tau lepton), the mass m is lepton rest mass, and the charge $q = -e$ is the electric charge quantum. The Bohr magneton is $\mu_B = \hbar e / 2mc$, so that given the specific emergence of $\hbar cq / mc^2$ in (7.12), we may rewrite the commutator as:

$$\begin{aligned}\left[cp^i, cp^j\right] &= 2i\mu_B\left(\partial^j\phi cp^i - \partial^i\phi cp^j\right) = 2i\mu_B\left(\left(E^j + \dot{A}^j / c\right)cp^i - \left(E^j + \dot{A}^j / c\right)cp^j\right) \\ &= -2i\mu_B\left(\nabla\phi c\mathbf{p} - \nabla\phi c\mathbf{p}\right) = 2i\mu_B\left(\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)c\mathbf{p} - \left(\mathbf{E} + \dot{\mathbf{A}} / c\right)c\mathbf{p}\right)\end{aligned} \ . \qquad (7.13)$$

Now let's return to (7.9), which contains terms very similar to those in the commutator (7.13). This gradient may be specialized to a dot (inner) product by forming

$$E\partial^i cp^i = q\partial^i\phi cp^i = q\left(E^i + \dot{A}^i / c\right)cp^i = -E\nabla\cdot c\mathbf{p} = -q\nabla\phi\cdot c\mathbf{p} = q\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)\cdot c\mathbf{p} \ . \qquad (7.14)$$

But it is especially of interest to form the cross product:

$$\begin{aligned}&\varepsilon^{kji}E\partial^j cp^i = \varepsilon^{kji}q\partial^j\phi cp^i = \varepsilon^{kji}q\left(E^j + \dot{A}^j / c\right)cp^i \\ &= -E\left(\nabla\phi\times c\mathbf{p}\right)^k = -q\left(\nabla\phi\times c\mathbf{p}\right)^k = q\left(\left(\mathbf{E} + \dot{\mathbf{A}} / c\right)\times c\mathbf{p}\right)^k \ .\end{aligned} \qquad (7.15)$$

This is because (7.13) contains $\partial^j\phi cp^i - \partial^i\phi cp^j$ which is also a cross product. Specifically, given that $\varepsilon^{kij}\left[cp^i, cp^j\right] = \varepsilon^{kij}cp^i cp^j - \varepsilon^{kij}cp^j cp^i = 2\left(c\mathbf{p}\times c\mathbf{p}\right)^k$, we may turn (7.13) into an explicit cross product by multiplying through by ε^{kij} to form (recall $\partial^j = -\nabla^j$):

$$\begin{aligned}&\varepsilon^{kij}\left[cp^i,cp^j\right]=2i\varepsilon^{kij}\mu_B\left(\partial^j\phi cp^i-\partial^i\phi cp^j\right)=2i\varepsilon^{kij}\mu_B\left(\left(E^j+\dot{A}^j/c\right)cp^i-\left(E^j+\dot{A}^j/c\right)cp^j\right)\\&=2\left(c\mathbf{p}\times c\mathbf{p}\right)^k=2i\mu_B\left(\nabla\phi\times c\mathbf{p}\right)^k=-2i\mu_B\left(\left(\mathbf{E}+\dot{\mathbf{A}}/c\right)\times c\mathbf{p}\right)^k\end{aligned}\quad . \tag{7.16}$$

If we now set $q=-e$ in (7.15) and so apply this to the charged leptons, then multiply through by $i\hbar c/mc^2$ and use $\mu_B=\hbar e/2mc$, we separately obtain:

$$-\frac{E}{mc^2}\left(i\hbar c\nabla\phi\times c\mathbf{p}\right)^k=2i\mu_B\left(\nabla\phi\times c\mathbf{p}\right)^k=-2i\mu_B\left(\left(\mathbf{E}+\dot{\mathbf{A}}/c\right)\times c\mathbf{p}\right)^k . \tag{7.17}$$

Dropping the k superscript which is absorbed into $\times^k=\times$, we than combine (7.16) and (7.17) into:

$$\begin{aligned}2c\mathbf{p}\times c\mathbf{p}&=-\left(E/mc^2\right)i\hbar c\nabla\phi\times c\mathbf{p}=-\left(dt/d\tau\right)i\hbar c\nabla\phi\times c\mathbf{p}=-\gamma_v\gamma_{em}i\hbar c\nabla\phi\times c\mathbf{p}\\&=2i\mu_B\nabla\phi\times c\mathbf{p}=-2i\mu_B\left(\mathbf{E}+\dot{\mathbf{A}}/c\right)\times c\mathbf{p}\end{aligned}\quad . \tag{7.18}$$

So as previewed in the paragraph following (7.10), we see that indeed, the momentum self-cross product $\mathbf{p}\times\mathbf{p}\neq 0$. And we have learned that this arises because the electromagnetic time dilation $\gamma_{em}\left(t,\mathbf{x}\right)=1/\left(1-q\phi_0\left(t,\mathbf{x}\right)/mc^2\right)$ is included in flat spacetime within the total energy momentum $cp^\mu=mcv^\mu\gamma_v\gamma_{em}=p^\mu_{NR}\gamma_v\gamma_{em}$ of (6.4). This is the first of numerous quantum mechanical results that we shall now begin to explore as we turn from classical to quantum electrodynamics.

PART II: COVARIANT GAUGE FIXING TO REMOVE TWO DEGREES OF FREEDOM FROM THE GAUGE POTENTIAL, YIELDING A MASSLESS PHOTON WITH TWO HELICITY STATES

8. Heisenberg / Ehrenfest Equations of Time Evolution and Space Configuration

Thus far all the development has been based on (1.6), which is the relativistic energy-momentum relation $m^2c^2=p_\sigma p^\sigma$ turned into $m^2c^2=\pi_\sigma\pi^\sigma$ via the prescription $p^\sigma\mapsto\pi^\sigma$ which arises from imposing local U(1) gauge symmetry, taken in the classical $\hbar=0$ limit by regarding the commutator relation to be $\left[p_\sigma,A^\sigma\right]=0$. Now, we return to the commutator $\left[p_\sigma,A^\sigma\right]$ in (1.5) and no longer approximate this to zero, but instead treat this quantum mechanically.

It was reviewed early in section 1 how when operating on a Fourier kernel $\exp\left(-ip_\sigma x^\sigma/\hbar\right)$ with the spacetime gradient ∂_μ, we obtain $\partial_\mu\exp\left(-ip_\sigma x^\sigma/\hbar\right)=-\left(ip_\mu/\hbar\right)\exp\left(-ip_\sigma x^\sigma/\hbar\right)$, where we assume that $\partial_\mu p_\sigma=0$ i.e. that the components of energy momentum are not functions of spacetime. So when we form a function such as $\phi = s\exp\left(-ip_\sigma x^\sigma / \hbar\right)$ with $s\left(p^\nu\right)$ a function of momentum but, importantly, not of spacetime because $\partial_\mu s\left(p^\nu\right)=0$, or such as $\psi = u\exp\left(-ip_\sigma x^\sigma / \hbar\right)$ with $u\left(p^\nu\right)$ and $\partial_\mu u\left(p^\nu\right)=0$, then we obtain $i\hbar\partial_\mu\phi = p_\mu\phi$ in the former and $i\hbar\partial_\mu\psi = p_\mu\psi$ in the latter case. Then for the Klein-Gordon and Dirac equations respectively, these operations allow for toggling between momentum and configuration space via $i\hbar\partial_\mu \leftrightarrow p_\mu$.

We now add to this, that energies $W = E - mc^2 = cp^0 - mc^2$ are eigenstates $H\left|s\right\rangle = W\left|s\right\rangle$ of a Hamiltonian operator H operating on a ket $\left|s\right\rangle$. Therefore, we may similarly form a Hamiltonian-momentum four-vector defined as $H^\mu \equiv \left(H + mc^2, c\mathbf{p}\right)$ for which $H^\mu\left|s\right\rangle = cp^\mu\left|s\right\rangle$, then use this in a Fourier-type kernel $\exp\left(-iH_\sigma x^\sigma / \hbar c\right)$ with the derivative $\partial_\mu \exp\left(-iH_\sigma x^\sigma / \hbar c\right) = -\left(iH_\mu / \hbar c\right)\exp\left(-ip_\sigma x^\sigma / \hbar c\right)$, likewise assuming that $\partial_\mu H = 0$ i.e. that the Hamiltonian is spacetime-independent, whole of course $\partial_\mu\left(mc^2\right)=0$. This is of interest because $\exp\left(-iH_\sigma x^\sigma / \hbar c\right) = \exp\left(-i\left(H+mc^2\right)t/\hbar + i\mathbf{p}\cdot\mathbf{x}/\hbar\right) = \exp\left(-i\left(H+mc^2\right)t/\hbar\right)\exp\left(i\mathbf{p}\cdot\mathbf{x}/\hbar\right)$ and because $U\left(t\right) = \exp\left(-iHt/\hbar\right)$ is the time evolution operator used in both the Heisenberg and Schrödinger pictures of quantum mechanics. The separation of this exponential into time and space operators via $\exp\left(A+B\right) = \exp A \exp B = \exp B \exp A$ is allowed because each of the four terms in $H_\sigma x^\sigma / c = \left(H+mc^2\right)t - p_x x - p_y y - p_z z = Ht - \mathbf{p}\cdot\mathbf{x}$ commutes with all other three.

Now, we generalize all of the foregoing by defining a ket $\left|s\right\rangle \equiv \exp\left(-iH_\sigma x^\sigma / \hbar c\right)\left|s_0\right\rangle$. This ket is a generalized state object including both a Fourier-type kernel $\exp\left(-iH_\sigma x^\sigma / \hbar c\right)$ which contains the Hamiltonian $H^0 = H + mc^2$, and a fixed-state ket $\left|s_0\right\rangle$ *defined* to be independent of spacetime, $\partial_\mu\left|s_0\right\rangle \equiv 0$, as designated by the subscript 0. The definition $\partial_\mu\left|s_0\right\rangle \equiv 0$ is important, and is the generalization of how we use $\partial_\mu s\left(p^\nu\right)=0$ and $\partial_\mu u\left(p^\nu\right)=0$ with $\partial_\mu p_\sigma = 0$ to toggle between configuration and momentum space for the Klein-Gordon and Dirac equations, respectively. As a consequence of these definitions, we may deduce that $H_\mu\left|s\right\rangle = i\hbar c\partial_\mu\left|s\right\rangle$.

Given that $H = H^\dagger$ is a Hermitian operator, we may also obtain the Hermitian conjugate of $\left|s\right\rangle$ which is the bra $\left\langle s\right| = \left\langle s\right|_0 \exp\left(iH_\sigma x^\sigma / \hbar c\right)$. As is customary we normalize the bra and ket to $\left\langle s | s\right\rangle = 1$. We then start by forming the operator relation:

$$\left\langle A^\nu\right\rangle = \left\langle s\right|A^\nu\left|s\right\rangle = \left\langle s_0\right|\exp\left(iH_\sigma x^\sigma / \hbar c\right)A^\nu\exp\left(-iH_\sigma x^\sigma / \hbar c\right)\left|s_0\right\rangle. \tag{8.1}$$

This is the expectation value for the gauge field A^ν, given that $\langle A^\nu \rangle = \langle s | A^\nu | s \rangle$. Now, our goal is to deduce the time-dependency $d\langle A^\nu \rangle / dt$, and thereafter, the space-dependency $d\langle A^\nu \rangle / d\mathbf{x}$.

The first step is to separate $\exp(-iH_\sigma x^\sigma / \hbar c) = \exp(-i(H + mc^2)t/\hbar)\exp(-ip_k x^k / \hbar)$ into time and space components with $-p_k x^k = \mathbf{p} \cdot \mathbf{x}$, via the standard $\exp(A+B) = \exp A \exp B$ because the commutator $[Ht, \mathbf{p}\cdot\mathbf{x}] = 0$. So for the ket we obtain the relation $\exp(-iH_\sigma x^\sigma / \hbar c)|s_0\rangle = \exp(-i(H+mc^2)t/\hbar)\exp(i\mathbf{p}\cdot\mathbf{x}/\hbar)|s_0\rangle$, with a conjugate relation for the bra. Then, for convenient notation we define the bra $|s_{0,\mathbf{x}}\rangle \equiv \exp(i\mathbf{p}\cdot\mathbf{x}/\hbar)|s_0\rangle$. Because $\partial_t |s_0\rangle = 0$ by definition, it is easy to see that $\partial_t |s_{0,\mathbf{x}}\rangle = 0$, but that $\hbar\partial_k |s_{0,\mathbf{x}}\rangle = \hbar\nabla |s_{0,\mathbf{x}}\rangle = -ip_k |s_{0,\mathbf{x}}\rangle = i\mathbf{p}|s_{0,\mathbf{x}}\rangle \neq 0$, so that $|s_{0,\mathbf{x}}\rangle$ varies in space but not over time. The subscripts $0, \mathbf{x}$ thus mean that $x^\mu = (ct, \mathbf{x}) = (0, \mathbf{x})$. If we view all of physics as describing the evolution over time of configurations of matter in space, then because $\exp(-iHt/\hbar)$ is the time evolution operator, we may regard $\exp(i\mathbf{p}\cdot\mathbf{x}/\hbar)$ as a space configuration operator. Likewise, now we may write $|s\rangle = \exp(-iHt/\hbar)|s_{0,\mathbf{x}}\rangle$. Likewise, also because $[Ht, \mathbf{p}\cdot\mathbf{x}] = 0$, the bra $\langle s| = \langle s_{0,\mathbf{x}}|\exp(i(H+mc^2)t/\hbar)$. In this notation, we may then rewrite (8.1) as:

$$\langle A^\nu \rangle = \langle s | A^\nu | s \rangle = \langle s_{0,\mathbf{x}} | \exp(iHt/\hbar) A^\nu \exp(-iHt/\hbar) | s_{0,\mathbf{x}} \rangle, \tag{8.2}$$

with the rest mass term in $H + mc^2$ cancelling out because $\exp(imc^2 t/\hbar)\exp(-imc^2 t/\hbar) = 1$. The above will be recognized as the usual starting point for deriving the Heisenberg equation of motion.

Because $\partial_t |s_{0,\mathbf{x}}\rangle = 0$, the total derivative of (8.2) with respect to time is the following:

$$\begin{aligned}
&\frac{d}{dt}\langle A^\nu \rangle = \frac{d}{dt}\langle s | A^\nu | s \rangle = \frac{d}{dt}\langle s_{0,\mathbf{x}} | \left(\exp(iHt/\hbar) A^\nu \exp(-iHt/\hbar)\right) | s_{0,\mathbf{x}} \rangle \\
&= \langle s_{0,\mathbf{x}} | \exp(iHt/\hbar) \left(\frac{i}{\hbar}[H, A^\nu] + \frac{\partial A^\nu}{\partial t} \right) \exp(-iHt/\hbar) | s_{0,\mathbf{x}} \rangle \\
&= \langle s | \left(\frac{i}{\hbar}[H, A^\nu] + \frac{\partial A^\nu}{\partial t} \right) | s \rangle = \frac{i}{\hbar}\langle [H, A^\nu] \rangle + \left\langle \frac{\partial A^\nu}{\partial t} \right\rangle
\end{aligned} \tag{8.3}$$

This is recognizable as Ehrenfest's theorem, which is merely the expectation value of the Heisenberg equation of motion in the Heisenberg picture. Also applying the eigenvalue relations $H|s\rangle = E|s\rangle$ and $\langle s|H = \langle s|E$, we may rewrite this overall result, retaining bras and kets, as:

$$\langle s|\left[H,A^{\nu}\right]|s\rangle=\langle s|\left[E,A^{\nu}\right]|s\rangle=i\hbar\langle s|\frac{\partial A^{\nu}}{\partial t}|s\rangle-i\hbar\frac{d}{dt}\langle s|A^{\nu}|s\rangle. \tag{8.4}$$

Note that all of the above are also equal to $\langle s|\left[\left(H+mc^{2}\right),A^{\nu}\right]|s\rangle$, and it is really $\left(H+mc^{2}\right)|s\rangle=E|s\rangle$ which enables us to interchange $E\leftrightarrow H$ in this context. We then reintroduce spacetime indexes in flat spacetime, to rewrite the above using $p_0=E/c$ as:

$$\langle s|\left[p_{0},A^{\nu}\right]|s\rangle=i\hbar\langle s|\frac{\partial A^{\nu}}{\partial x^{0}}|s\rangle-i\hbar\frac{d}{dx^{0}}\langle s|A^{\nu}|s\rangle=i\hbar\langle s|\partial_{0}A^{\nu}|s\rangle-i\hbar d_{0}\langle s|A^{\nu}|s\rangle. \tag{8.5}$$

Because our interest is the commutator $\left[p_{\sigma},A^{\sigma}\right]=\left[p_{0},A^{0}\right]+\left[p_{k},A^{k}\right]$ in (1.5), we find that when sandwiched between a bra and a ket as defined above, the term $\langle s|\left[p_{0},A^{0}\right]|s\rangle$ is the $\nu=0$ component of (8.5) above.

Next, let us obtain the space-dependency $d\left\langle A^{\nu}\right\rangle/d\mathbf{x}$ for (8.1). We can sample, say, the *z* axis, then generalize to *x* and *y*. First, we segregate the z-axis term to the front of the kernel $\exp\left(-iH_{\sigma}x^{\sigma}/\hbar c\right)=\exp\left(-ip_{3}x^{3}/\hbar\right)\exp\left(-ip_{2,1}x^{2,1}/\hbar\right)\exp\left(-i\left(H_{0}+mc^{2}\right)x^{0}/\hbar c\right)$. Again, this is permitted because all four terms in $H_{\sigma}x^{\sigma}/c=\left(H+mc^{2}\right)t-p_{x}x-p_{y}y-p_{z}z$ mutually commute. Then, we define $\left|s_{t,x,y,0}\right\rangle\equiv\exp\left(-ip_{2,1}x^{2,1}/\hbar\right)\exp\left(-i\left(H_{0}+mc^{2}\right)x^{0}/\hbar c\right)\left|s_{0}\right\rangle$ to be another ket which varies over time and over *x* and *y* but not over *z*, thus $\partial_{z}\left|s_{t,x,y,0}\right\rangle=0$. Therefore, $|s\rangle=\exp\left(-ip_{3}x^{3}/\hbar\right)\left|s_{t,x,y,0}\right\rangle$. Given that $p_3=-p_z$ in flat spacetime, using this and its conjugate bra $\langle s|$ in (8.1) yields:

$$\left\langle A^{\nu}\right\rangle=\langle s|A^{\nu}|s\rangle=\left\langle s_{t,x,y,0}\right|\exp\left(-ip_{z}z/\hbar\right)A^{\nu}\exp\left(ip_{z}z/\hbar\right)\left|s_{t,x,y,0}\right\rangle. \tag{8.6}$$

Then, using $\partial_{z}\left|s_{t,x,y,0}\right\rangle=0$, we take the *z*-axis total derivative of (8.6) to obtain:

$$\begin{aligned}
&\frac{d}{dz}\left\langle A^{\nu}\right\rangle=\frac{d}{dz}\langle s|A^{\nu}|s\rangle=\frac{d}{dz}\left\langle s_{t,x,y,0}\right|\exp\left(-ip_{z}z/\hbar\right)A^{\nu}\exp\left(ip_{z}z/\hbar\right)\left|s_{t,x,y,0}\right\rangle\\
&=\left\langle s_{t,x,y,0}\right|\exp\left(-ip_{z}z/\hbar\right)\left(-\frac{i}{\hbar}\left[p_{z},A^{\nu}\right]+\frac{\partial A^{\nu}}{\partial z}\right)\exp\left(ip_{z}z/\hbar\right)\left|s_{t,x,y,0}\right\rangle\\
&=\langle s|\left(-\frac{i}{\hbar}\left[p_{z},A^{\nu}\right]+\frac{\partial A^{\nu}}{\partial z}\right)|s\rangle=-\frac{i}{\hbar}\left\langle\left[p_{z},A^{\nu}\right]\right\rangle+\left\langle\frac{\partial A^{\nu}}{\partial z}\right\rangle
\end{aligned} \tag{8.7}$$

This is an Ehrenfest-type equation for the z evolution. Then generalizing to the other two space dimensions and also using $p_k = -\mathbf{p}$, we rewrite this in the form of (8.5), as:

$$\langle s|\left[p_k, A^\nu\right]|s\rangle = i\hbar\langle s|\frac{\partial A^\nu}{\partial x^k}|s\rangle - i\hbar\frac{d}{dx^k}\langle s|A^\nu|s\rangle = i\hbar\langle s|\partial_k A^\nu|s\rangle - i\hbar d_k\langle s|A^\nu|s\rangle . \tag{8.8}$$

Comparing (8.5) with (8.8), we see that these are simply the time and space parts of a Lorentz-covariant relation, and so may be combined into a single relation:

$$\langle s|\left[p_\mu, A^\nu\right]|s\rangle = i\hbar\langle s|\partial_\mu A^\nu|s\rangle - i\hbar\partial_\mu\langle s|A^\nu|s\rangle = \left\langle\left[p_\mu, A^\nu\right]\right\rangle = i\hbar\left(\left\langle\partial_\mu A^\nu\right\rangle - \partial_\mu\left\langle A^\nu\right\rangle\right). \tag{8.9}$$

Above, we have also replaced what were originally the total derivatives into partial derivatives, $d \mapsto \partial$, because we now have combined the $d_\sigma = d/dx^\sigma$ taken in all four spacetime dimensions into one relation. Now, even with the same ∂_μ in both terms on the right hand side above, we see with clarity that the expected value of the commutator, $\left\langle\left[p_\mu, A^\nu\right]\right\rangle$, measures $i\hbar$ times *the difference between the expected value of the four-gradient,* $\left\langle\partial_\mu A^\nu\right\rangle$*, and the four-gradient of the expected value,* $\partial_\mu\left\langle A^\nu\right\rangle$. Summing indexes this becomes:

$$\langle s|\left[p_\sigma, A^\sigma\right]|s\rangle = i\hbar\langle s|\partial_\sigma A^\sigma|s\rangle - i\hbar\partial_\sigma\langle s|A^\sigma|s\rangle = \left\langle\left[p_\sigma, A^\sigma\right]\right\rangle = i\hbar\left(\left\langle\partial_\sigma A^\sigma\right\rangle - \partial_\sigma\left\langle A^\sigma\right\rangle\right). \tag{8.10}$$

Now we have derived the correct quantum mechanical treatment of the commutator $\left[p_\sigma, A^\sigma\right]$ in (1.5): When this commutator is sandwiched within $\langle s|\left[p_\sigma, A^\sigma\right]|s\rangle$ using $\langle s|$ and $|s\rangle$ developed above, it is evaluated according to the Ehrenfest-type equation (8.10) above, which contains the expected value of the Heisenberg-picture equation of motion in its time term, and three space-component terms containing expectation values for Heisenberg-picture equations of configuration. Combined in the summed form of (8.10), these terms Lorentz transform as a scalar. Although derived in flat spacetime, we can generalize to curved spacetime by simply writing the commutator term as $\left\langle\left[p_\sigma, A^\sigma\right]\right\rangle = \left\langle g_{\mu\nu}\left[p^\mu, A^\nu\right]\right\rangle$.

9. Arriving at a Massless Photon by Gauge-Covariant, Lorentz-Covariant Gauge Fixing of the Klein-Gordon Equation to Remove Two Degrees of Freedom from the Gauge Field

With the result (8.10), we return to (1.5) with $A_\sigma p^\sigma = A^\sigma p_\sigma$ and $p_\sigma p^\sigma = p^\sigma p_\sigma$, but now sandwich this between the bra $\langle s|$ and the ket $|s\rangle$ developed in the previous section, to write:

$$0 = \langle s|\left(\pi_\sigma\pi^\sigma - m^2c^2\right)|s\rangle = \langle s|\left(p^\sigma p_\sigma + 2\frac{q}{c}A^\sigma p_\sigma - m^2c^2 + \frac{q}{c}\left[p_\sigma, A^\sigma\right] + \frac{q^2}{c^2}A_\sigma A^\sigma\right)|s\rangle . \tag{9.1}$$

This is just the Klein-Gordon equation $0=\left(\hbar^2\left(\partial_\sigma - iqA_\sigma/\hbar c\right)\left(\partial^\sigma - iqA^\sigma/\hbar c\right)+m^2c^2\right)\phi$ restated in momentum space as $0=\left(\left(p_\sigma + qA_\sigma/c\right)\left(p^\sigma + qA^\sigma/c\right)-m^2c^2\right)s$ with the earlier *s* turned into a ket $|s\rangle$ and with a front-appended bra $\langle s|$. For two random variables *A* and *B*, the expectation value is linear, $\langle A+B\rangle=\langle A\rangle+\langle B\rangle$. So the commutator term in (9.1) may be separately treated as $(q/c)\langle s|\left[p_\sigma, A^\sigma\right]|s\rangle$, enabling us to directly substitute (8.10) into (9.1). The result is:

$$0=\langle s|\left(p^\sigma p_\sigma + 2\frac{q}{c}A^\sigma p_\sigma - m^2c^2 + \frac{q}{c}i\hbar\partial_\sigma A^\sigma + \frac{q^2}{c^2}A_\sigma A^\sigma\right)|s\rangle - \frac{q}{c}i\hbar\partial_\sigma\langle s|A^\sigma|s\rangle. \tag{9.2}$$

Again, this is still the Klein-Gordon equation, in momentum space, with a bra in front. In (9.2), $(q/c)i\hbar\partial_\sigma A^\sigma + (q/c)^2 A_\sigma A^\sigma = (q/c)\left(i\hbar\partial_\sigma + (q/c)A_\sigma\right)A^\sigma$, which contains the gauge-covariant derivative in the form $i\hbar\partial_\sigma + qA_\sigma/c = i\hbar\mathfrak{D}_\sigma$. Thus (9.2) becomes:

$$0=\langle s|\left(p^\sigma p_\sigma + 2\frac{q}{c}A^\sigma p_\sigma - m^2c^2 + i\hbar\frac{q}{c}\mathfrak{D}_\sigma A^\sigma\right)|s\rangle - i\hbar\frac{q}{c}\partial_\sigma\langle s|A^\sigma|s\rangle. \tag{9.3}$$

Now, it is very common practice in U(1) gauge theory to remove one degree of freedom by imposing the Lorenz gauge $\partial_\sigma A^\sigma = 0$. However, *a priori*, the gauge field A^σ has four independent components, while the photon which this represents in quantum theory is massless and so only has two transverse degrees of freedom. Because $\langle s|\mathfrak{D}_\sigma A^\sigma|s\rangle = \langle\mathfrak{D}_\sigma A^\sigma\rangle$ and $\partial_\sigma\langle s|A^\sigma|s\rangle = \partial_\sigma\langle A^\sigma\rangle$, (9.3) affords us the opportunity to remove two degrees of freedom. First, we may impose the Lorentz-covariant and gauge-covariant Lorenz gauge fixing condition:

$$\langle s|\mathfrak{D}_\sigma A^\sigma|s\rangle = \langle\mathfrak{D}_\sigma A^\sigma\rangle = \left\langle\partial_\sigma A^\sigma - i\frac{q}{\hbar c}A_\sigma A^\sigma\right\rangle = 0, \tag{9.4}$$

which sets the expected value $\langle\mathfrak{D}_\sigma A^\sigma\rangle$ of the gauge-covariant derivative $\mathfrak{D}_\sigma A^\sigma$ of the gauge field A^σ to zero. Second, we may impose the Lorentz-covariant gauge fixing condition:

$$\partial_\sigma\langle s|A^\sigma|s\rangle = \partial_\sigma\langle A^\sigma\rangle = 0 \tag{9.5}$$

which is the usual Lorenz gauge used to set the *expected value* $\langle A^\sigma\rangle$ of the gauge field A^σ to zero. If we impose both (9.4) and (9.5) on (9.3), then we can remove *two of the four degrees of freedom* from the gauge field, in a covariant manner, ensuring that A^σ will only retain two degrees of freedom which is precisely what is needed for this to represent massless photon quanta.

Therefore, we now proceed to impose both (9.4) and (9.5) on (9.3), to simplify this to:

$$0=\left\langle s\right|\left(p^{\sigma}p_{\sigma}+2\frac{q}{c}A^{\sigma}p_{\sigma}-m^{2}c^{2}\right)\left|s\right\rangle=\left\langle p^{\sigma}p_{\sigma}+2\frac{q}{c}A^{\sigma}p_{\sigma}-m^{2}c^{2}\right\rangle, \tag{9.6}$$

while the gauge field loses two of its four degrees of freedom. We may also again apply the heuristic rule $p_{\sigma}\mapsto i\hbar\partial_{\sigma}$ in the above to write this, with sign flip and the bra removed, as:

$$0=\left(\hbar^{2}\partial^{\sigma}\partial_{\sigma}-2i\frac{\hbar}{c}qA^{\sigma}\partial_{\sigma}+m^{2}c^{2}\right)\left|s\right\rangle. \tag{9.7}$$

We have removed the bra in the above and so *not* written this as an expected value equation, because when ∂_{σ} appears in the equation, it needs to operate on a ket to its right, as $\partial_{\sigma}\left|s\right\rangle$. This is now a gauge-fixed Klein-Gordon equation in configuration space, in which the gauge field A^{σ} contains two not four degrees of freedom, precisely as is required for a massless photon. By the Correspondence Principle, the classical equation obtained from (9.6) is:

$$m^{2}c^{2}=p^{\sigma}p_{\sigma}+2\frac{q}{c}A^{\sigma}p_{\sigma}. \tag{9.8}$$

This should be contrasted with (1.5) from which the final two $\left[p_{\sigma},A^{\sigma}\right]$ and $A_{\sigma}A^{\sigma}$ terms have been removed using $\left\langle s\right|$ and $\left|s\right\rangle$ to turn (1.5) from a classical into a quantum mechanical equation, and then imposing the gauge conditions (9.5) and (9.6). What we learn from all this is that quantum mechanics, combined with two covariant gauge fixing conditions removing two degrees of freedom from the gauge fields, has brought about a wholesale change to the classical equation (1.5) by removing two of its terms.

10. Classical and Quantum Mechanical Geodesic Equations of Gravitational and Electromagnetic Motion

Now, let s work from the expectation value equation in (9.6), apply $p^{\sigma}=mdx^{\sigma}/d\tau$ throughout, and raise an index in the first term, and move the term with $m^{2}c^{2}$ to the left, thus:

$$\left\langle m^{2}c^{2}\right\rangle=\left\langle m^{2}g_{\mu\nu}\frac{dx^{\mu}}{d\tau}\frac{dx^{\nu}}{d\tau}+2\frac{qm}{c}A_{\sigma}\frac{dx^{\sigma}}{d\tau}\right\rangle. \tag{10.1}$$

It will be seen that this is the parallel equation to (2.1), but that two things have now changed: First, the term with $A_{\sigma}A^{\sigma}$ is gone as a consequence of the gauge conditions (9.4) and (9.5). Second the entire equation is an expectation value equation. By the Correspondence Principle and Ehrenfest's theorem, we know that the classical equation implied by (10.1) is simply (10.1) with the expectation brackets removed, which is (2.1) without the $A_{\sigma}A^{\sigma}$ term. Therefore, it is easy to see that if start with the classical equation implied by (10.1) via Correspondence, and repeat all the same steps earlier taken from (2.1) through (2.12) starting with the variational equation $0=\delta\int_A^B d\tau$ of (2.3) for *geodesic motion*, we will end up with the classical equation of motion:

$$\frac{d^2x^\beta}{c^2d\tau^2}=-\Gamma^\beta{}_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+\frac{q}{mc^2}F^\beta{}_\sigma\frac{dx^\sigma}{cd\tau}. \tag{10.2}$$

This has the gauge-dependent $\partial^\beta\left(A_\sigma A^\sigma\right)$ term removed as a consequence of the gauge fixing in (9.4) and (9.5), it accords precisely with the known classical physical motions for gravitation and electrodynamics, and it is entirely geodesic motion because of its derivation from a variation.

Now, however, we can also obtain the *quantum mechanical equation of motion* based on (10.1). First, we note that the mass term may be written as $\left\langle m^2c^2\right\rangle=m^2c^2$ because *m* and *c* are numbers with definite values and zero variance, not statistical values. So too for *q*. Second, dx^μ are coordinate elements and $d\tau$ is a proper time element which also represent definite, not statistical measurement numbers *against which* we measure statistical spreads. That is, even when we graph a probability distribution, we still do so against definite measurement axes. The statistical objects in (10.1) are the gravitational fields in $\left\langle g_{\mu\nu}\right\rangle$, and the gravitational fields and electromagnetic potential in $\left\langle A_\sigma\right\rangle=\left\langle g_{\sigma\tau}A^\tau\right\rangle$, though for now it will be convenient to retain the lower-indexed form A_σ to absorb the gravitational field. As a result, we may refine (10.1) into:

$$m^2c^2=m^2\left\langle g_{\mu\nu}\right\rangle\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}+2\frac{qm}{c}\left\langle A_\sigma\right\rangle\frac{dx^\sigma}{d\tau}. \tag{10.3}$$

We then divide both sides through by m^2c^2 to write this as:

$$1=\left\langle g_{\mu\nu}\right\rangle\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+2\frac{q}{mc^2}\left\langle A_\sigma\right\rangle\frac{dx^\sigma}{cd\tau} \tag{10.4}$$

Contrasting to (2.2), the difference is that the $A_\sigma A^\sigma$ term is now gone, and the two fields $g_{\mu\nu}$ and A_σ are now expectation values $\left\langle g_{\mu\nu}\right\rangle$ and $\left\langle A_\sigma\right\rangle$.

So if we now employ this "1" in a minimized variation as in (2.3), it turns out that all the steps taken from (2.3) through (2.11) will be exactly the same, except that $\left\langle g_{\mu\nu}\right\rangle$ will end up wherever there was a $g_{\mu\nu}$, and $\left\langle A_\sigma\right\rangle$ wherever there was a A_σ, in (2.11). Therefore, the counterpart to (2.11) based on (10.4) now turns out to be:

$$0=\delta\int_A^B d\tau=\int_A^B \delta x^\alpha d\tau\begin{pmatrix}\frac{1}{2}\left(\partial_\alpha\left\langle g_{\mu\nu}\right\rangle-\partial_\mu\left\langle g_{\nu\alpha}\right\rangle-\partial_\nu\left\langle g_{\alpha\mu}\right\rangle\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-\left\langle g_{\alpha\nu}\right\rangle\frac{d^2x^\nu}{c^2d\tau^2}\\+\frac{q}{mc^2}\left(\partial_\alpha\left\langle A_\sigma\right\rangle-\partial_\sigma\left\langle A_\alpha\right\rangle\right)\frac{dx^\sigma}{cd\tau}\end{pmatrix}. \tag{10.5}$$

As before, and for the same reasons, the term inside the large parenthesis must be zero, so that:

$$\left\langle g_{\alpha\nu}\right\rangle\frac{d^2x^\nu}{c^2d\tau^2}=\frac{1}{2}\left(\partial_\alpha\left\langle g_{\mu\nu}\right\rangle-\partial_\mu\left\langle g_{\nu\alpha}\right\rangle-\partial_\nu\left\langle g_{\alpha\mu}\right\rangle\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+\frac{q}{mc^2}\left(\partial_\alpha\left\langle A_\sigma\right\rangle-\partial_\sigma\left\langle A_\alpha\right\rangle\right)\frac{dx^\sigma}{cd\tau}. \tag{10.6}$$

In contrast to its counterpart (2.12), the above needs to be treated with some care, because $-\left\langle \Gamma^\beta{}_{\mu\nu}\right\rangle=\frac{1}{2}\left\langle g^{\beta\alpha}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\right\rangle$ and $\left\langle F_{\alpha\sigma}\right\rangle=\left\langle \partial_\alpha A_\sigma-\partial_\sigma A_\alpha\right\rangle$ contain *expectation values of derivatives*, while (10.6) is distinguished by having the terms $\frac{1}{2}\left(\partial_\alpha\left\langle g_{\mu\nu}\right\rangle-\partial_\mu\left\langle g_{\nu\alpha}\right\rangle-\partial_\nu\left\langle g_{\alpha\mu}\right\rangle\right)$ and $\partial_\alpha\left\langle A_\sigma\right\rangle-\partial_\sigma\left\langle A_\alpha\right\rangle$ containing *derivatives of expectation values*. This is where the Heisenberg equation of time evolution and space configuration comes back into play, because this very same distinction is measured by the commutators of the fields with energy momentum. So, from (8.9):

$$\partial_\alpha\left\langle A_\sigma\right\rangle=\left\langle \partial_\alpha A_\sigma\right\rangle+i\left\langle\left[p_\alpha,A_\sigma\right]\right\rangle/\hbar . \tag{10.7}$$

And because this applies generally to field operators, not only to A_σ, for $g_{\mu\nu}$ we may also write:

$$\partial_\alpha\left\langle g_{\mu\nu}\right\rangle=\left\langle \partial_\alpha g_{\mu\nu}\right\rangle+i\left\langle\left[p_\alpha,g_{\mu\nu}\right]\right\rangle/\hbar . \tag{10.8}$$

Then, using (10.7) and (10.8) in (10.6) and rearranging somewhat yields:

$$\begin{aligned}\left\langle g_{\alpha\nu}\right\rangle\frac{d^2x^\nu}{c^2d\tau^2}&=\frac{1}{2}\left\langle\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right\rangle\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+\frac{q}{mc^2}\left\langle\partial_\alpha A_\sigma-\partial_\sigma A_\alpha\right\rangle\frac{dx^\sigma}{cd\tau}\\&+\frac{i}{2\hbar}\left\langle\left[p_\alpha,g_{\mu\nu}\right]-\left[p_\mu,g_{\nu\alpha}\right]-\left[p_\nu,g_{\alpha\mu}\right]\right\rangle\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+\frac{iq}{\hbar mc^2}\left\langle\left[p_\alpha,A_\sigma\right]-\left[p_\sigma,A_\alpha\right]\right\rangle\frac{dx^\sigma}{cd\tau}\end{aligned}. \tag{10.9}$$

Now, we have a term $\left\langle\partial_\alpha A_\sigma-\partial_\sigma A_\alpha\right\rangle=\left\langle F_{\alpha\sigma}\right\rangle$ placed inside expectation values. Moreover, with simple re-indexing, we also find $\frac{1}{2}\left\langle\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right\rangle=-\left\langle g_{\alpha\beta}\Gamma^\beta{}_{\mu\nu}\right\rangle$. So with these replacements (10.9) becomes:

$$\begin{aligned}\left\langle g_{\alpha\nu}\right\rangle\frac{d^2x^\nu}{c^2d\tau^2}&=-\left\langle g_{\alpha\beta}\Gamma^\beta{}_{\mu\nu}\right\rangle\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+\frac{q}{mc^2}\left\langle F_{\alpha\sigma}\right\rangle\frac{dx^\sigma}{cd\tau}\\&+\frac{i}{2\hbar}\left\langle\left[p_\alpha,g_{\mu\nu}\right]-\left[p_\mu,g_{\nu\alpha}\right]-\left[p_\nu,g_{\alpha\mu}\right]\right\rangle\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+\frac{iq}{\hbar mc^2}\left\langle\left[p_\alpha,A_\sigma\right]-\left[p_\sigma,A_\alpha\right]\right\rangle\frac{dx^\sigma}{cd\tau}\end{aligned}. \tag{10.10}$$

To further simplify, we raise the free index α inside the expectation brackets. Although raising an index, for example, via $X^{\mu}=g^{\mu\sigma}X_{\sigma}$ for some X_{σ} involves multiplying by $g^{\mu\sigma}$, we may still perform this entirely within the brackets because if $\langle X\rangle=\langle Y\rangle$ then $\langle g^{\mu\sigma}X\rangle=\langle g^{\mu\sigma}Y\rangle$ for any objects *X*, *Y*. When we raise an index for $\langle g_{\alpha\nu}\rangle$ we have $\langle \delta^{\alpha}{}_{\nu}\rangle=\delta^{\alpha}{}_{\nu}$ which removes the expectation value because the Kronecker delta $\delta^{\alpha}{}_{\nu}$ is just a 4x4 identity matrix; likewise for $\langle g_{\alpha\beta}\Gamma^{\beta}{}_{\mu\nu}\rangle$ we have $\langle \delta^{\alpha}{}_{\beta}\Gamma^{\beta}{}_{\mu\nu}\rangle=\delta^{\alpha}{}_{\beta}\langle\Gamma^{\beta}{}_{\mu\nu}\rangle$. And when we do this for e.g. $[p_{\nu},g_{\alpha\mu}]$ we obtain $[p_{\nu},\delta^{\alpha}{}_{\mu}]=0$. So this removes the two commutators $[p_{\mu},g_{\nu\alpha}]$ and $[p_{\nu},g_{\alpha\mu}]$ which have an index α in the metric tensor. With all of this, also raising the remaining indexes to explicitly show all appearances of the gravitational field, we arrive at our final result:

$$\boxed{\begin{aligned}\frac{d^2x^{\alpha}}{c^2d\tau^2}=&-\langle\Gamma^{\alpha}{}_{\mu\nu}\rangle\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}+\frac{q}{mc^2}\langle g_{\sigma\beta}F^{\alpha\beta}\rangle\frac{dx^{\sigma}}{cd\tau}\\&+\frac{i}{2\hbar}\langle[p^{\alpha},g_{\mu\nu}]\rangle\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}+\frac{iq}{\hbar mc^2}\langle g_{\sigma\beta}[p^{\alpha},A^{\beta}]-g_{\sigma\beta}[p^{\beta},A^{\alpha}]\rangle\frac{dx^{\sigma}}{cd\tau}\end{aligned}}. \tag{10.11}$$

For classical theory, where all the commutators become zero and the expectation values are removed via the Correspondence Principle, (10.11) becomes the well-settled classical equation (10.2). So – very importantly – using the gauge conditions (9.4) and (9.5) to remove two terms from (9.3) which descended from (1.5), has caused the gauge-dependent term $\partial^{\beta}(A_{\sigma}A^{\sigma})$ to vanish from (2.12) in favor of (10.2), which accords entirely with the robustly confirmed motions of particles in gravitational and electromagnetic fields, and which motions are now seen to *both* be geodesic motions. When a classical system approaches a scale where quantum commutation cannot be neglected, (10.11) applies. And in a fully-quantum setting, where the commutators are large enough so the classical terms with $\langle\Gamma^{\alpha}{}_{\mu\nu}\rangle$ and $\langle g_{\sigma\beta}F^{\alpha\beta}\rangle$ become negligible due to the very tiny $\hbar$ in the *denominator* of the commutator terms, (10.11) becomes a quantum motion equation:

$$\frac{d^2x^{\alpha}}{c^2d\tau^2}=\frac{i}{2\hbar}\langle[p^{\alpha},g_{\mu\nu}]\rangle\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}+\frac{iq}{\hbar mc^2}\langle g_{\sigma\beta}[p^{\alpha},A^{\beta}]-g_{\sigma\beta}[p^{\beta},A^{\alpha}]\rangle\frac{dx^{\sigma}}{cd\tau}. \tag{10.12}$$

Finally, in (10.12) we can make good use of the generalized uncertainty relation $\sigma(A)\sigma(B)\ge\frac{1}{2}\left|i\langle[A,B]\rangle\right|$ for any two objects which are non-commuting, where σ represents statistical standard deviation. By this relation, $\sigma(p^{\alpha})\sigma(g_{\mu\nu})/\hbar\ge(i/2\hbar)\left|\langle[p^{\alpha},g_{\mu\nu}]\rangle\right|$. Therefore, when we consider gravitation alone by setting $q=0$ or $A^{\alpha}=0$, (10.12) becomes:

$$\sigma(p^{\alpha})\sigma(g_{\mu\nu})\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}\ge\hbar\frac{d^2x^{\alpha}}{c^2d\tau^2}=\frac{i}{2}\langle[p^{\alpha},g_{\mu\nu}]\rangle\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}, \tag{10.13}$$

which is an uncertainty relation for quantum gravitational interactions. Conversely, when we consider electromagnetic interactions alone in flat spacetime, (10.12) becomes:

$$\frac{q}{mc^2}\eta_{\sigma\beta}\left(\sigma\left(p^\alpha\right)\sigma\left(A^\beta\right)-\sigma\left(p^\beta\right)\sigma\left(A^\alpha\right)\right)\frac{dx^\sigma}{cd\tau}\geq\frac{\hbar}{2}\frac{d^2x^\alpha}{c^2d\tau^2}=\frac{iq}{2mc^2}\eta_{\sigma\beta}\left\langle\left[p^\alpha,A^\beta\right]-\left[p^\beta,A^\alpha\right]\right\rangle\frac{dx^\sigma}{cd\tau}. \tag{10.14}$$

This is an uncertainty relation for quantum electromagnetic interactions. Both (10.13) and (10.14) are actually four independent equations, with the free index α. In both of these relations, the lower bound on the uncertainty spread is established by the four-acceleration $d^2x^\alpha / c^2d\tau^2$. For gravitation, the coefficient of the acceleration is $\hbar$. And for electromagnetism, it is noteworthy that the coefficient is $\hbar/2$ which is also the magnitude of fermion spins. And it is again worth noting that because of the gauge conditions (9.4) and (9.5) all unphysical gauge freedom has been removed from A^μ, so that there is no gauge ambiguity in $\sigma\left(p^\alpha\right)\sigma\left(A^\mu\right)-\sigma\left(p^\mu\right)\sigma\left(A^\alpha\right)$.

Finally, it is helpful to directly contrast the classical equations of motion with the quantum equations of *motion uncertainty*. For gravitation absent electromagnetism this contrast is:

$$-\Gamma^\alpha{}_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}=\frac{d^2x^\alpha}{c^2d\tau^2}\quad\text{versus}\quad\sigma\left(p^\alpha\right)\sigma\left(g_{\mu\nu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}\geq\hbar\frac{d^2x^\alpha}{c^2d\tau^2}. \tag{10.15}$$

For electromagnetism absent gravitation, mindful that $F^{\alpha\mu}=\partial^\alpha A^\mu-\partial^\mu A^\alpha$, this is:

$$\frac{q}{mc^2}\eta_{\mu\nu}F^{\alpha\mu}\frac{dx^\nu}{cd\tau}=\frac{d^2x^\alpha}{c^2d\tau^2}\quad\text{versus}\quad\frac{q}{mc^2}\eta_{\mu\nu}\left(\sigma\left(p^\alpha\right)\sigma\left(A^\mu\right)-\sigma\left(p^\mu\right)\sigma\left(A^\alpha\right)\right)\frac{dx^\nu}{cd\tau}\geq\frac{\hbar}{2}\frac{d^2x^\alpha}{c^2d\tau^2}. \tag{10.16}$$

In (10.15) we see that for a given acceleration, as the momentum uncertainty $\sigma\left(p^\alpha\right)$ for the mass in the gravitational field becomes smaller the field uncertainty $\sigma\left(g_{\mu\nu}\right)$ grows larger, and vice versa. In (10.16) we see a similar incompatibility between momentum uncertainty and electromagnetic potential uncertainty.

11. The Simplified Quadratic Line Element following Gauge Fixing

If we again start with (10.4) and multiply each side through by $c^2d\tau^2$ we obtain the metric:

$$c^2d\tau^2=\left\langle g_{\mu\nu}\right\rangle dx^\mu dx^\nu+2\frac{q}{mc^2}\left\langle A_\sigma\right\rangle dx^\sigma cd\tau \tag{11.1}$$

It will be seen that this is the "unusual" quadratic metric (3.3) from earlier, but with the same two changes reviewed after (10.4): the $A_\sigma A^\sigma$ is gone, and we now have expectation values $\left\langle g_{\mu\nu}\right\rangle$ and $\left\langle A_\sigma\right\rangle$. This remains quadratic in $c^2d\tau^2$, as is seen if we write this as (contrast (3.4)):

$$0 = c^2 d\tau^2 - 2\frac{q}{mc^2}\langle A_\sigma \rangle dx^\sigma c d\tau - \langle g_{\mu\nu}\rangle dx^\mu dx^\nu \tag{11.2}$$

But now the quadratic solution takes on a much simpler form that its counterpart (3.5), namely:

$$c d\tau = \frac{q}{mc^2}\langle A_\sigma \rangle dx^\sigma \pm \sqrt{\left(\langle g_{\mu\nu}\rangle + \frac{q^2}{m^2c^4}\langle A_\mu \rangle \langle A_\nu \rangle\right) dx^\mu dx^\nu} \ . \tag{11.3}$$

In particular, this no longer contains the ratio form of (3.5), and the term inside the square root is significantly simplified. In fact, if we make the two definitions:

$$G_{\mu\nu} \equiv \langle g_{\mu\nu}\rangle + \frac{q^2}{m^2c^4}\langle A_\mu \rangle \langle A_\nu \rangle; \qquad c^2 \mathrm{T}^2 \equiv G_{\mu\nu} dx^\mu dx^\nu , \tag{11.4}$$

then also employing $\langle A \rangle = \langle A_\sigma \rangle dx^\sigma$ which is the expected value of the differential one-form $A = A_\sigma dx^\sigma$ for the gauge field, we see that (11.3) can be written in the very simple form:

$$c d\tau = \frac{q}{mc^2}\langle A_\sigma \rangle dx^\sigma \pm \sqrt{G_{\mu\nu} dx^\mu dx^\nu} = \frac{q}{mc^2}\langle A \rangle \pm c d\mathrm{T} . \tag{11.5}$$

The above have several very interesting properties. First, the object $c^2 d\mathrm{T}^2 \equiv G_{\mu\nu} dx^\mu dx^\nu$ has a form very similar to the metric scalar $c^2 d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu$. Of course, $G_{\mu\nu}$ defined above cannot be formally regarded as a metric tensor because it does not have the metricity properties of $g_{\mu\nu}$ whereby $g_{\mu\nu;\sigma} = 0$ and $g^{\mu\sigma} g_{\sigma\nu} = \delta^\mu{}_\nu$. Nor is $d\mathrm{T}$ (necessarily) invariant; rather, the invariant is $c d\tau = q\langle A \rangle / mc^2 \pm c d\mathrm{T}$ with the possibility of some sub-relation between $q\langle A \rangle / mc^2$ and $c d\mathrm{T}$ which leaves $c d\tau$ unchanged. But what makes this of keen interest is that we may still think of of $G_{\mu\nu}$ as being a "quasi-geometric" object in the manner of $g_{\mu\nu}$ merely because $\pm\sqrt{G_{\mu\nu} dx^\mu dx^\nu}$ standing alone still does define a line element $c d\mathrm{T}$ (which differs from $c d\tau$ precisely by $q\langle A \rangle / mc^2$). Further, the $\pm\sqrt{G_{\mu\nu} dx^\mu dx^\nu}$ square root is very reminiscent of how Dirac's equation $(i\hbar \Gamma^\mu \partial_\mu - mc)\psi = 0$ is developed in flat spacetime from $\pm\sqrt{\eta_{\mu\nu} p^\mu p^\nu}$ and then generalized into curved spacetime using a tetrad $e^\mu_a \gamma^a \equiv \Gamma^\mu$, as earlier reviewed in section 1.

This point will be of keen interest here, because while Dirac's equation teaches about how individual electrons behave in an *electromagnetic field*, (11.5) will lead us to a variant of Dirac's equation which can be used to understand *how individual photons interact with individual electrons*. And in fact, (11.5) only has the form that it does (versus the earlier (3.5)), because at (9.4) and (9.5) we removed two of the four degrees of freedom from A^σ giving it precisely the properties expected of a massless photon. Indeed, the foregoing is why, following Dirac, part of the title of this paper is "Quantum Theory of the Electron *and the Photon*."

12. The Electromagnetic Time Dilation and Energy Content Relations, following Gauge Fixing

Before we proceed to this new variant of Dirac's equation, we first wish to determine the impact of the foregoing quantum development and gauge fixing on the electromagnetic time dilations (5.8) and (5.9). To do so, we develop the quadratic solution for the metric (11.1) when taken at rest in flat spacetime, just as we earlier did for the metric (3.3). To place (11.1) into flat spacetime, we need to set $\langle g_{\mu\nu}\rangle = \langle \eta_{\mu\nu}\rangle = \eta_{\mu\nu}$. So following the same steps that led to (4.1), it is easy to see that (11.1) will become:

$$d\tau^2 = dt^2 + 2\frac{q\langle\phi_0\rangle}{mc^2}dtd\tau \ . \tag{12.1}$$

Following the development from (4.1) to (4.4), again choosing to solve for dt, we see that in place of (4.4) we now have:

$$\frac{dt}{d\tau} = 1 - \frac{q\langle\phi_0\rangle}{mc^2} . \tag{12.2}$$

So the only difference is that now the scalar potential appears as an expectation value. Otherwise there is no change to the overall form of the equation. This is because when we had the earlier terms with $\phi_0{}^2$ that have now been eliminated because of the gauge fixing at (9.4) and (9.5), these terms nonetheless ended up cancelling inside the square root term in (4.3).

So if we repeat the development from (4.4) to (5.8), nothing else changes, and the earlier (5.8) and (6.1) for the time electromagnetic time dilation at rest in flat spacetime and its energy content via the relation $E = \gamma_{em} mc^2$, now becomes:

$$\gamma_{em} \equiv \frac{dt}{d\tau} = \frac{E}{mc^2} = \frac{1}{1 - \frac{q\langle\phi_0\rangle}{mc^2}} = 1 + \frac{q\langle\phi_0\rangle}{mc^2} + \left(\frac{q\langle\phi_0\rangle}{mc^2}\right)^2 + \left(\frac{q\langle\phi_0\rangle}{mc^2}\right)^3 + \left(\frac{q\langle\phi_0\rangle}{mc^2}\right)^4 + \ldots = \sum_{n=0}^{\infty}\left(\frac{q\langle\phi_0\rangle}{mc^2}\right)^n . \tag{12.3}$$

Now, the time dilation is based on the expected value of the scalar potential. When we employ a Coulomb potential, this will enter as $\langle\phi_0\rangle = k_e Q\langle 1/r\rangle$ where $\langle 1/r\rangle$ is the expectation value of the inverse separation between the two charges. Note, we have not used $1/\langle r\rangle$ because statistically, $\langle 1/r\rangle \neq 1/\langle r\rangle$. Rather, as is well known, $\langle 1/r\rangle \geq 1/\langle r\rangle$ for positive random variable *r*. The only distribution with $\langle 1/r\rangle = 1/\langle r\rangle$ is a Dirac delta $\delta(r)$. So the (5.9), (6.2) counterpart is:

$$\gamma_{em} \equiv \frac{dt}{d\tau} = \frac{E}{mc^2} = \frac{1}{1-\frac{k_e Qq}{mc^2}\left\langle \frac{1}{r} \right\rangle} = 1 + \frac{k_e Qq}{mc^2}\left\langle \frac{1}{r} \right\rangle + \left(\frac{k_e Qq}{mc^2}\left\langle \frac{1}{r} \right\rangle \right)^2 + \ldots = \sum_{n=0}^{\infty} \left(\frac{k_e Qq}{mc^2}\left\langle \frac{1}{r} \right\rangle \right)^n . \quad (12.4)$$

So we naturally find ourselves in a situation where must use an *expected separation* between Q and q, which is precisely where we do end up once we talk about interactions between electrons, protons, etc. which do not have positions with classical certainty. Thus, (12.4) naturally embeds the existence of Heisenberg position uncertainty via the appearance of $\langle 1/r \rangle$. In general, cf. (6.3), the energy content relation $E = \Gamma mc^2 = \gamma_v \gamma_g \gamma_{em} mc^2$ holds for both classical and quantum systems. The expectation values of quantum systems are embedded in the individual γ_v, γ_g, γ_{em}. The energy in excess of mc^2, is then $W = E - mc^2 = mc^2(\Gamma - 1)$. This means as well that the relation $p^\mu = m\Gamma v^\mu = m\gamma_v \gamma_g \gamma_{em} v^\mu$ obtained at (6.4) also continues to hold for a quantum system.

PART III: THE HYPER-CANONICAL DIRAC EQUATION FOR INDIVIDUAL ELECTRON AND PHOTON INTERACTIONS

13. Dirac's Equation with Electromagnetic Tetrads

Now we turn to Dirac's equation. As reviewed in section 1, to obtain Dirac's equation, we start with the entirely-classical relation $m^2c^2 = \eta^{\mu\nu} p_\mu p_\nu$ in flat spacetime, define a set of 4x4 γ^μ operator matrices $\frac{1}{2}\{\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu\} \equiv \eta^{\mu\nu}$, then use $(\gamma^\mu p_\mu)^2 = \frac{1}{2}\{\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu\} p_\mu p_\nu = \eta^{\mu\nu} p_\mu p_\nu$ to take the square root equation $mc = \pm\sqrt{\eta^{\mu\nu} p_\mu p_\nu} = \gamma^\mu p_\mu$ with the $\pm$ sign absorbed in the γ^μ definitions. Finally, because this result only makes sense if it operates on a spinor $u(p^\nu)$ which following the development in section 8 we represent as the ket $|u_0\rangle$ with $\partial_\mu |u_0\rangle = 0$, we are able to form $(\gamma^\mu p_\mu - mc)|u_0\rangle = 0$. If we then use the ket $|\psi\rangle \equiv \exp(-ip_\sigma x^\sigma)|u_0\rangle$ this readily becomes $(i\hbar\gamma^\mu \partial_\mu - mc)|\psi\rangle = 0$. We then introduce electromagnetic interactions by requiring local U(1) electromagnetic interactions which provides us with the gauge-covariant derivative $\partial_\mu \mapsto \mathfrak{D}_\mu \equiv \partial_\mu - iqA_\mu / \hbar c$. Finally, in curved spacetime, where the underlying equation is $mc = \pm\sqrt{g^{\mu\nu} p_\mu p_\nu}$, we also employ tetrads defined such that $g^{\mu\nu} = \eta^{ab} e_a^\mu e_b^\nu$. In this way, we turn $mc = \pm\sqrt{\eta^{\mu\nu} p_\mu p_\nu}$ or $mc = \pm\sqrt{g^{\mu\nu} p_\mu p_\nu}$ which is a classical equation, into the quintessentially quantum mechanical operator equation of Dirac.

As also reviewed in section 1, a similar process occurs with the Klein Gordon equation. Here we start with the same classical $m^2c^2 = \eta_{\mu\nu} p^\mu p^\nu$, have this operate on what we now write as the ket $|s_0\rangle$ with $\partial_\mu |s_0\rangle = 0$ in the form $(p_\sigma p^\sigma - m^2c^2)|s_0\rangle = 0$, then use $|s\rangle \equiv \exp(-ip_\sigma x^\sigma)|s_0\rangle$

to advance this to $0=\left(\hbar^2\partial_\sigma\partial^\sigma+m^2c^2\right)|s\rangle$, then use $\partial_\mu\mapsto\mathfrak{D}_\mu$ to add interactions. Here too, we turn a purely classical equation $m^2c^2=\eta_{\mu\nu}p^\mu p^\nu$ into a quantum mechanical equation. The key point of both these examples for the discussion to follow is this: the tried and true recipe of both Klein-Gordon and Dirac teaches us that we can start with a classical equation such as $mc=\pm\sqrt{\eta^{\mu\nu}p_\mu p_\nu}$ or $m^2c^2=\eta_{\mu\nu}p^\mu p^\nu$, use it to operate on a ket such as $|\psi\rangle$ or $|s\rangle$, and thereby produce a valid quantum mechanical equation.

With this in mind, we return to (11.3) which is the quadratic solution for the metric (11.1), which in turn descends from (9.6) which in turn is the Klein-Gordon equation in the form (1.5) sandwiched between a bra and a ket after applying the gauge conditions (9.4) and (9.5). By the Ehrenfest/Correspondence Principle, the classical equation we may extract from (11.3) by turning all expectation values into ordinary classical objects is:

$$cd\tau=\frac{q}{mc^2}A_\sigma dx^\sigma\pm\sqrt{\left(g_{\mu\nu}+\frac{q^2}{m^2c^4}A_\mu A_\nu\right)dx^\mu dx^\nu}=\frac{q}{mc^2}A_\sigma dx^\sigma\pm\sqrt{G_{\mu\nu}dx^\mu dx^\nu}\,, \tag{13.1}$$

Above, we also insert the classical value $G_{\mu\nu}=g_{\mu\nu}+\left(q^2/m^2c^4\right)A_\mu A_\nu$ from (11.4), so this is (11.5) in its classical limit. This is also the "peculiar" quadratic solution (3.5), once its Klein-Gordon counterpart is converted to a quantum operator equation and its gauge fixed using (9.4) and (9.5).

Because our present interest is in Dirac's equation, we multiply this classical result (13.1) through by $m/d\tau$ and swap upper and lower indexes, to obtain:

$$mc=\frac{q}{mc^2}A^\sigma p_\sigma\pm\sqrt{\left(g^{\mu\nu}+\frac{q^2}{m^2c^4}A^\mu A^\nu\right)p_\mu p_\mu}=\frac{q}{mc^2}A^\sigma p_\sigma\pm\sqrt{G^{\mu\nu}p_\mu p_\mu}\,, \tag{13.2}$$

so we have the square root in the exact same form as the classical curved spacetime equation $mc=\pm\sqrt{g^{\mu\nu}p_\mu p_\nu}$. Just as we do for Dirac's equation in curved spacetime, we now turn (13.2) above into an alternative form of Dirac's equation which applies specifically to the quantum interactions between individual electrons and individual photons, because the covariant removal of two degrees of freedom to produce a massless photon is *structurally embedded* in (13.2). Specifically, in the same way we generalize Dirac's equation into flat spacetime by defining a set of Γ^μ in terms of $g^{\mu\nu}$ by $\frac{1}{2}\left\{\Gamma^\mu\Gamma^\nu+\Gamma^\nu\Gamma^\mu\right\}\equiv g^{\mu\nu}$ and in terms of the tetrads e^μ_a by $e^\mu_a\gamma^a\equiv\Gamma^\mu$ so that $g^{\mu\nu}=\frac{1}{2}\left\{\gamma^a\gamma^b+\gamma^b\gamma^a\right\}e^\mu_a e^\nu_b=\eta^{ab}e^\mu_a e^\nu_b$, let us now use exactly the same approach to (13.2). From (11.4), we may extract classical equation:

$$G_{\mu\nu}=g_{\mu\nu}+\left(q^2/m^2c^4\right)A_\mu A_\nu \tag{13.3}$$

from the expectation value. To start we will work in flat spacetime so that $g_{\mu\nu}=\eta_{\mu\nu}$ and $G_{\mu\nu}=\eta_{\mu\nu}+\left(q^2/m^2c^4\right)A_\mu A_\nu$. Later, we will generalize back to curved spacetime.

Just as the gravitational tetrads e_a^{μ} contain both an upper Greek spacetime index and a lower early-in-the-alphabet Latin Lorentz index, we begin by defining a similar *electromagnetic tetrad* ε_y^{μ} (ε denoting electromagnetism) with an upper Greek spacetime index and a lower late-in-the alphabet Latin electromagnetic index. We also use these in flat spacetime to define a set of electromagnetic gamma matrices by the relation $\Gamma^{\mu}_{(\varepsilon)}\equiv\varepsilon_y^{\mu}\gamma^y$. Finally, we further define these $\Gamma^{\mu}_{(\varepsilon)}$ in terms of $G^{\mu\nu}$ by $\frac{1}{2}\left\{\Gamma^{\mu}_{(\varepsilon)}\Gamma^{\nu}_{(\varepsilon)}+\Gamma^{\nu}_{(\varepsilon)}\Gamma^{\mu}_{(\varepsilon)}\right\}\equiv G^{\mu\nu}$, then combine all these definitions by writing:

$$G^{\mu\nu}=\eta^{\mu\nu}+\frac{q^2}{m^2c^4}A^\mu A^\nu\equiv\tfrac{1}{2}\left\{\Gamma^{\mu}_{(\varepsilon)}\Gamma^{\nu}_{(\varepsilon)}+\Gamma^{\nu}_{(\varepsilon)}\Gamma^{\mu}_{(\varepsilon)}\right\}=\tfrac{1}{2}\left\{\gamma^y\gamma^z+\gamma^y\gamma^z\right\}\varepsilon_y^{\mu}\varepsilon_z^{\nu}=\eta^{yz}\varepsilon_y^{\mu}\varepsilon_z^{\nu}, \tag{13.4}$$

Just as $\left(\gamma^\mu p_\mu\right)^2=\eta^{\mu\nu}p_\mu p_\nu$ in flat spacetime and $\left(\Gamma^\mu p_\mu\right)^2=g^{\mu\nu}p_\mu p_\nu$ in curved spacetime, it is simple to deduce from the above definitions that $\left(\Gamma^{\mu}_{(\varepsilon)}p_\mu\right)^2=G^{\mu\nu}p_\mu p_\nu$. Then, the square root in (13.2) may be written as $\pm\sqrt{G^{\mu\nu}p_\mu p_\mu}=\Gamma^{\mu}_{(\varepsilon)}p_\mu$ which, as with $\Gamma^\mu p_\mu$ in Dirac's equation, is a 4x4 matrix. So this will now have to operate on a 4-component column vector.

For Dirac's momentum space flat spacetime equation $\left(\gamma^\mu\left(p_\mu+qA_\mu/c\right)-mc\right)\left|u_0\right\rangle=0$ we employ a Dirac spinor $u\left(p^\mu\right)$ that is independent of space and time which, in accord with the conventions developed in section 8, we denote as $\left|u_0\right\rangle$. Here, we use a similar four-component fixed-state ket $\left|U_0\right\rangle$ defined to be independent of spacetime, $\partial_\mu\left|U_0\right\rangle\equiv0$. Then, appending $\left|U_0\right\rangle$ to the right of (13.2), using $\pm\sqrt{G^{\mu\nu}p_\mu p_\mu}=\Gamma^{\sigma}_{(\varepsilon)}p_\sigma$ and setting everything to a zero (13.2) becomes:

$$\left(\left(\Gamma^{\sigma}_{(\varepsilon)}+\frac{q}{mc^2}A^\sigma\right)cp_\sigma-mc^2\right)\left|U_0\right\rangle=0. \tag{13.5}$$

This is to be contrasted (13.6) with Dirac's $\left(\gamma^\sigma\pi_\sigma-mc\right)\left|u_0\right\rangle=\left(\gamma^\sigma\left(p_\sigma+qA_\sigma/c\right)-mc\right)\left|u_0\right\rangle=0$. In the absence of electromagnetic fields, where either $q=0$ or $A^\sigma=0$, the tetrad ε_y^{μ} becomes a 4x4 unit matrix, and $\Gamma^{\sigma}_{(\varepsilon)}\equiv\varepsilon_y^{\sigma}\gamma^y=\gamma^\mu$, so that (13.5) this reduces to $\left(\gamma^\sigma p_\sigma-mc\right)\left|U_0\right\rangle=0$. Likewise, Dirac's momentum space equation reduces to $\left(\gamma^\sigma p_\sigma-mc\right)\left|u_0\right\rangle=0$. Because these two equations now have exactly the same operator $\gamma^\sigma p_\sigma-mc$, this also means that $\left|U_0\right\rangle\rightarrow\left|u_0\right\rangle$ when electromagnetic interactions vanish. Thus (13.5) becomes synonymous with Dirac's momentum space equation for free fermions. However, when there are electromagnetic interactions, (13.5) is a somewhat different equation from $\left(\gamma^{\sigma}\left(p_{\sigma}+qA_{\sigma}/c\right)-mc\right)\left|u_0\right\rangle=0$. Shortly, we shall study these differences. As a result of appending $\left|U_0\right\rangle$, the classical (13.2) is now a quantum mechanical equation (13.5).

If the further define a ket $\left|\Psi\right\rangle\equiv\exp\left(-iH_{\sigma}x^{\sigma}/\hbar c\right)\left|U_0\right\rangle$ which is a function of space and time due to the kernel $\exp\left(-iH_{\sigma}x^{\sigma}/\hbar c\right)$, then we may deduce $H_{\sigma}\left|\Psi\right\rangle=cp_{\sigma}\left|\Psi\right\rangle=i\hbar c\partial_{\sigma}\left|\Psi\right\rangle$ just as we did previously prior to (8.1) for $\left|s\right\rangle$. With this (13.5) can be turned into:

$$\left(i\hbar c\left(\Gamma^{\sigma}_{(\varepsilon)}+\frac{q}{mc^2}A^{\sigma}\right)\partial_{\sigma}-mc^2\right)\left|\Psi\right\rangle=0\,. \tag{13.6}$$

This is the new variant of Dirac's equation in configuration space in flat spacetime, which should be contrasted to the usual $0=\left(i\hbar\gamma^{\mu}\mathfrak{D}_{\mu}-mc\right)\left|\psi\right\rangle=\left(\gamma^{\mu}\left(i\hbar\partial_{\mu}+qA_{\mu}/c\right)-mc\right)\left|\psi\right\rangle$ for Dirac's configuration space equation in flat spacetime, as reviewed in section 1. As with (13.5), the two operators become identical when $q=0$ or $A^{\sigma}=0$ so that $\left|\Psi\right\rangle\rightarrow\left|\psi\right\rangle$, in which circumstance, (13.6) becomes synonymous with Dirac's configuration apace equation for free fermions.

Importantly, (13.5) and (13.6) also answer the question how to make sense of the "peculiar" line element in (3.3) and its equally perplexing solution (3.5): The quadratic solution (3.5) is in fact a new variant (13.5) of Dirac's equation in thick disguise, which is unmasked once we use the Heisenberg/Ehrenfest equations of motion and configuration, then remove two degrees of freedom from the gauge field A^{σ} via (9.4) and (9.5), thereby turning A^{σ} into a true massless photon. So as we shall also shortly see, (13.5) allows us to study interactions between *individual electrons and individual photons*. For pedagogic reference, given that the Dirac equation $0=\left(i\hbar\gamma^{\mu}\mathfrak{D}_{\mu}-mc\right)\left|\psi\right\rangle$ is the *canonical* result of applying local U(1) gauge symmetry to the ordinary $0=\left(i\hbar\gamma^{\mu}\partial_{\mu}-mc\right)\left|\psi\right\rangle$, we shall refer to equations (13.5) and (13.6) as Dirac's equation with electromagnetic tetrads embedded in $\Gamma^{\mu}_{(\varepsilon)}\equiv\varepsilon^{\mu}_{\ y}\gamma^{y}$ as the "hyper-canonical" Dirac equation.

14. The Electromagnetic Interaction Tetrad

Now we wish to derive the electromagnetic tetrad $\varepsilon^{\mu}_{\ y}$, in explicit component representation. The key relation for doing so is $\eta^{yz}\varepsilon^{\mu}_{\ y}\varepsilon^{\nu}_{\ z}\equiv\eta^{\mu\nu}+\left(q^2/m^2c^4\right)A^{\mu}A^{\nu}$ in (13.4). For compact notation we define the ratio $\rho\equiv q/mc^2$. Given that $\varepsilon^{\mu}_{\ y}=\delta^{\mu}_{\ y}$ is a 4x4 identify matrix when $q=0$ or A^{μ}, it will also help to define an $\varepsilon'^{\mu}_{\ y}$ according to $\varepsilon^{\mu}_{\ y}\equiv\delta^{\mu}_{\ y}+\varepsilon'^{\mu}_{\ y}$, to represent how $\varepsilon^{\mu}_{\ y}$ differs from the unit $\delta^{\mu}_{\ y}$. With these definitions we write the salient portion of (13.4) as:

$$\eta^{yz}\varepsilon^{\mu}_{\ y}\varepsilon^{\nu}_{\ z}=\eta^{yz}\left(\delta^{\mu}_{\ y}+\varepsilon'^{\mu}_{\ y}\right)\left(\delta^{\nu}_{\ z}+\varepsilon'^{\nu}_{\ z}\right)=\eta^{yz}\left(\delta^{\mu}_{\ y}\delta^{\nu}_{\ z}+\varepsilon'^{\mu}_{\ y}\delta^{\nu}_{\ z}+\delta^{\mu}_{\ y}\varepsilon'^{\nu}_{\ z}+\varepsilon'^{\mu}_{\ y}\varepsilon'^{\nu}_{\ z}\right)=\eta^{\mu\nu}+\rho A^{\mu}\rho A^{\nu}\,. \tag{14.1}$$

With $\eta^{yz}\delta^{\mu}_{y}\delta^{\nu}_{z}=\eta^{\mu\nu}$ and $\eta^{yz}\delta^{\nu}_{z}=\eta^{y\nu}$ and $\eta^{yz}\delta^{\mu}_{y}=\eta^{\mu z}$, and also subtracting $\eta^{\mu\nu}$ from each side, this easily simplifies to:

$$\eta^{y\nu}\varepsilon'^{\mu}_{y}+\eta^{\mu z}\varepsilon'^{\nu}_{z}+\eta^{yz}\varepsilon'^{\mu}_{y}\varepsilon'^{\nu}_{z}=\rho A^{\mu}\rho A^{\nu}\,. \tag{14.2}$$

The above contains sixteen (16) equations for each of $\mu=0,1,2,3$ and $\nu=0,1,2,3$. But, this is symmetric in μ and ν so in fact there are only ten (10) independent equations. Moreover, because A^{μ} has only four independent components, and also because we have already removed two degrees of freedom from A^{μ} via the gauge conditions (9.4) and (9.5), we anticipate that (14.2) will highlight this limited freedom by imposing definitive constraints on A^{μ}. Given that $\operatorname{diag}\left(\eta^{yz}\right)=\left(1,-1,-1,-1\right)$, the four $\mu=\nu$ "diagonal" equations in (14.2) produce the relations:

$$\begin{aligned}
&2\varepsilon'^{0}_{0}+\varepsilon'^{0}_{0}\varepsilon'^{0}_{0}-\varepsilon'^{0}_{1}\varepsilon'^{0}_{1}-\varepsilon'^{0}_{2}\varepsilon'^{0}_{2}-\varepsilon'^{0}_{3}\varepsilon'^{0}_{3}=\rho A^{0}\rho A^{0}\\
&-2\varepsilon'^{1}_{1}+\varepsilon'^{1}_{0}\varepsilon'^{1}_{0}-\varepsilon'^{1}_{1}\varepsilon'^{1}_{1}-\varepsilon'^{1}_{2}\varepsilon'^{1}_{2}-\varepsilon'^{1}_{3}\varepsilon'^{1}_{3}=\rho A^{1}\rho A^{1}\\
&-2\varepsilon'^{2}_{2}+\varepsilon'^{2}_{0}\varepsilon'^{2}_{0}-\varepsilon'^{2}_{1}\varepsilon'^{2}_{1}-\varepsilon'^{2}_{2}\varepsilon'^{2}_{2}-\varepsilon'^{2}_{3}\varepsilon'^{2}_{3}=\rho A^{2}\rho A^{2}\\
&-2\varepsilon'^{3}_{3}+\varepsilon'^{3}_{0}\varepsilon'^{3}_{0}-\varepsilon'^{3}_{1}\varepsilon'^{3}_{1}-\varepsilon'^{3}_{2}\varepsilon'^{3}_{2}-\varepsilon'^{3}_{3}\varepsilon'^{3}_{3}=\rho A^{3}\rho A^{3}
\end{aligned}\,. \tag{14.3a}$$

Likewise the three $\mu=0$, $\nu=1,2,3$ mixed time and space relations in (14.2) are:

$$\begin{aligned}
&-\varepsilon'^{0}_{1}+\varepsilon'^{1}_{0}+\varepsilon'^{0}_{0}\varepsilon'^{1}_{0}-\varepsilon'^{0}_{1}\varepsilon'^{1}_{1}-\varepsilon'^{0}_{2}\varepsilon'^{1}_{2}-\varepsilon'^{0}_{3}\varepsilon'^{1}_{3}=\rho A^{0}\rho A^{1}\\
&-\varepsilon'^{0}_{2}+\varepsilon'^{2}_{0}+\varepsilon'^{0}_{0}\varepsilon'^{2}_{0}-\varepsilon'^{0}_{1}\varepsilon'^{2}_{1}-\varepsilon'^{0}_{2}\varepsilon'^{2}_{2}-\varepsilon'^{0}_{3}\varepsilon'^{2}_{3}=\rho A^{0}\rho A^{2}\\
&-\varepsilon'^{0}_{3}+\varepsilon'^{3}_{0}+\varepsilon'^{0}_{0}\varepsilon'^{3}_{0}-\varepsilon'^{0}_{1}\varepsilon'^{3}_{1}-\varepsilon'^{0}_{2}\varepsilon'^{3}_{2}-\varepsilon'^{0}_{3}\varepsilon'^{3}_{3}=\rho A^{0}\rho A^{3}
\end{aligned}\,. \tag{14.3b}$$

Finally, the pure-space relations with $\mu,\nu=1,2$, $\mu,\nu=2,3$ and $\mu,\nu=3,1$ are:

$$\begin{aligned}
&-\varepsilon'^{1}_{2}-\varepsilon'^{2}_{1}+\varepsilon'^{1}_{0}\varepsilon'^{2}_{0}-\varepsilon'^{1}_{1}\varepsilon'^{2}_{1}-\varepsilon'^{1}_{2}\varepsilon'^{2}_{2}-\varepsilon'^{1}_{3}\varepsilon'^{2}_{3}=\rho A^{1}\rho A^{2}\\
&-\varepsilon'^{2}_{3}-\varepsilon'^{3}_{2}+\varepsilon'^{2}_{0}\varepsilon'^{3}_{0}-\varepsilon'^{2}_{1}\varepsilon'^{3}_{1}-\varepsilon'^{2}_{2}\varepsilon'^{3}_{2}-\varepsilon'^{2}_{3}\varepsilon'^{3}_{3}=\rho A^{2}\rho A^{3}\\
&-\varepsilon'^{3}_{1}-\varepsilon'^{1}_{3}+\varepsilon'^{3}_{0}\varepsilon'^{1}_{0}-\varepsilon'^{3}_{1}\varepsilon'^{1}_{1}-\varepsilon'^{3}_{2}\varepsilon'^{1}_{2}-\varepsilon'^{3}_{3}\varepsilon'^{1}_{3}=\rho A^{3}\rho A^{1}
\end{aligned}\,. \tag{14.3c}$$

Now, the right hand side of all ten of (14.3) have nonlinear products $\rho A^{\mu}\rho A^{\nu}$ of two field terms. On the left of each there is a mix of linear and nonlinear expressions containing the ε^{μ}_{y}. In (14.3a) the linear appearances are of ε'^{0}_{0}, ε'^{1}_{1}, ε'^{2}_{2} and ε'^{3}_{3} respectively. Given that the complete tetrad $\varepsilon^{\mu}_{y}\equiv\delta^{\mu}_{y}+\varepsilon'^{\mu}_{y}$, let us require that $\varepsilon^{\mu}_{y}=\delta^{\mu}_{y}$ for the four $\mu=y$ components, therefore, $\varepsilon'^{0}_{0}=\varepsilon'^{1}_{1}=\varepsilon'^{2}_{2}=\varepsilon'^{3}_{3}$ for $\mu=y$. This is consistent with $\varepsilon^{\mu}_{y}=\delta^{\mu}_{y}$ generally when $q=0$ or A^{μ}, and it means that the field components ρA^{μ} will all appear in off-diagonal components of ε^{μ}_{y}. In (14.3b), let us eliminate the linear terms by requiring $\varepsilon'^{0}_{1}=\varepsilon'^{1}_{0}$, $\varepsilon'^{0}_{2}=\varepsilon'^{2}_{0}$, and $\varepsilon'^{0}_{3}=\varepsilon'^{3}_{0}$, which is symmetric in μ and *y*. In (14.3c) we likewise remove the linear terms by requiring $\varepsilon_2^{\prime 1}=-\varepsilon_1^{\prime 2}$, $\varepsilon_3^{\prime 2}=-\varepsilon_2^{\prime 3}$ and $\varepsilon_1^{\prime 3}=-\varepsilon_3^{\prime 1}$ which is antisymmetric in μ and *y*. With all of this (14.3) reduce to:

$$\begin{aligned}&-\varepsilon_1^{\prime 0}\varepsilon_1^{\prime 0}-\varepsilon_2^{\prime 0}\varepsilon_2^{\prime 0}-\varepsilon_3^{\prime 0}\varepsilon_3^{\prime 0}=\rho A^0\rho A^0\\&+\varepsilon_0^{\prime 1}\varepsilon_0^{\prime 1}-\varepsilon_2^{\prime 1}\varepsilon_2^{\prime 1}-\varepsilon_3^{\prime 1}\varepsilon_3^{\prime 1}=\rho A^1\rho A^1\\&+\varepsilon_0^{\prime 2}\varepsilon_0^{\prime 2}-\varepsilon_1^{\prime 2}\varepsilon_1^{\prime 2}-\varepsilon_3^{\prime 2}\varepsilon_3^{\prime 2}=\rho A^2\rho A^2\\&+\varepsilon_0^{\prime 3}\varepsilon_0^{\prime 3}-\varepsilon_1^{\prime 3}\varepsilon_1^{\prime 3}-\varepsilon_2^{\prime 3}\varepsilon_2^{\prime 3}=\rho A^3\rho A^3\end{aligned}, \tag{14.4a}$$

$$\begin{aligned}&-\varepsilon_2^{\prime 0}\varepsilon_2^{\prime 1}-\varepsilon_3^{\prime 0}\varepsilon_3^{\prime 1}=\rho A^0\rho A^1\\&-\varepsilon_1^{\prime 0}\varepsilon_1^{\prime 2}-\varepsilon_3^{\prime 0}\varepsilon_3^{\prime 2}=\rho A^0\rho A^2\\&-\varepsilon_1^{\prime 0}\varepsilon_1^{\prime 3}-\varepsilon_2^{\prime 0}\varepsilon_2^{\prime 3}=\rho A^0\rho A^3\end{aligned}, \tag{14.4b}$$

$$\begin{aligned}&\varepsilon_0^{\prime 1}\varepsilon_0^{\prime 2}-\varepsilon_3^{\prime 1}\varepsilon_3^{\prime 2}=\rho A^1\rho A^2\\&\varepsilon_0^{\prime 2}\varepsilon_0^{\prime 3}-\varepsilon_1^{\prime 2}\varepsilon_1^{\prime 3}=\rho A^2\rho A^3\\&\varepsilon_0^{\prime 3}\varepsilon_0^{\prime 1}-\varepsilon_2^{\prime 3}\varepsilon_2^{\prime 1}=\rho A^3\rho A^1\end{aligned}. \tag{14.4c}$$

Next, for the space components of A^μ, we assign $\varepsilon_0^{\prime 1}=-\rho A^1$, $\varepsilon_0^{\prime 2}=-\rho A^2$ and $\varepsilon_0^{\prime 3}=-\rho A^3$ for the components of the tetrad which have a space world index and a time Lorentz index. By the earlier symmetric relations $\varepsilon_1^{\prime 0}=\varepsilon_0^{\prime 1}$, $\varepsilon_2^{\prime 0}=\varepsilon_0^{\prime 2}$, and $\varepsilon_3^{\prime 0}=\varepsilon_0^{\prime 3}$ this means $\varepsilon_1^{\prime 0}=-\rho A^1$, $\varepsilon_2^{\prime 0}=-\rho A^2$ and $\varepsilon_3^{\prime 0}=-\rho A^3$ as well. Substituting this in (14.4) and reducing then brings us to:

$$\begin{aligned}&-\rho A^1\rho A^1-\rho A^2\rho A^2-\rho A^3\rho A^3=\rho A^0\rho A^0\\&+\rho A^1\rho A^1-\varepsilon_2^{\prime 1}\varepsilon_2^{\prime 1}-\varepsilon_3^{\prime 1}\varepsilon_3^{\prime 1}=\rho A^1\rho A^1\\&+\rho A^2\rho A^2-\varepsilon_1^{\prime 2}\varepsilon_1^{\prime 2}-\varepsilon_3^{\prime 2}\varepsilon_3^{\prime 2}=\rho A^2\rho A^2\\&+\rho A^3\rho A^3-\varepsilon_1^{\prime 3}\varepsilon_1^{\prime 3}-\varepsilon_2^{\prime 3}\varepsilon_2^{\prime 3}=\rho A^3\rho A^3\end{aligned}, \tag{14.5a}$$

$$\begin{aligned}&-\rho A^2\varepsilon_2^{\prime 1}-\rho A^3\varepsilon_3^{\prime 1}=\rho A^0\rho A^1\\&-\rho A^1\varepsilon_1^{\prime 2}-\rho A^3\varepsilon_3^{\prime 2}=\rho A^0\rho A^2\\&-\rho A^1\varepsilon_1^{\prime 3}-\rho A^2\varepsilon_2^{\prime 3}=\rho A^0\rho A^3\end{aligned}, \tag{14.5b}$$

$$\begin{aligned}&-\varepsilon_3^{\prime 1}\varepsilon_3^{\prime 2}=0\\&-\varepsilon_1^{\prime 2}\varepsilon_1^{\prime 3}=0\\&-\varepsilon_2^{\prime 3}\varepsilon_2^{\prime 1}=0\end{aligned}. \tag{14.5c}$$

Because (14.4) all contain products of two tetrads it would be possible to make the oppositely-signed assignments $\varepsilon_0^{\prime 1}=+\rho A^1$, $\varepsilon_0^{\prime 2}=+\rho A^2$ and $\varepsilon_0^{\prime 3}=+\rho A^3$ without changing the results (14.5) at all, because as to this sign ambiguity, $(\pm 1)^2 = +1$. As we shall later see at (19.13) supra, we choose the minus sign because this is required to ensure that (13.5) produces solutions identical to Dirac's usual $\left(\gamma^\sigma \left(p_\sigma + qA_\sigma / c\right) - mc\right)\left|u_0\right\rangle = 0$ in the weak field linear limit.

Next, one way to satisfy the earlier relation $\varepsilon'^1_2 = -\varepsilon'^2_1$, $\varepsilon'^2_3 = -\varepsilon'^3_2$ and $\varepsilon'^3_1 = -\varepsilon'^1_3$ following (14.3) is to set all six of these to zero. This will satisfy all of (14.5c) identically, and will also satisfy the last three relations (14.5a) identically. We may also divide out ρ^2 from the first relation (14.5a), and all of (14.5b) may be combined into one, so now all we have left to satisfy are:

$$A^0 A^0 + A^1 A^1 + A^2 A^2 + A^3 A^3 = 0, \tag{14.6a}$$

$$0 = \rho A^0 \rho A^1 = \rho A^0 \rho A^2 = \rho A^0 \rho A^3. \tag{14.6b}$$

If we posit that at least one of the three $A^1 \neq 0$, $A^2 \neq 0$ and $A^3 \neq 0$, then we are required by (14.6b) to set $A^0 = 0$. The only relation we now have left to satisfy is (14.6a), which is the $\mu\nu = 00$ pure-time component of (14.2). Because of (14.6b), (14.6a) becomes:

$$A^1 A^1 + A^2 A^2 + A^3 A^3 = \mathbf{A} \cdot \mathbf{A} = \mathbf{A}^2 = 0. \tag{14.7}$$

Consolidating (14.6) and (14.7) into generally covariant form, we obtain:

$$A^0 = A_0 = 0; \quad -A^k A_k = \mathbf{A} \cdot \mathbf{A} = \mathbf{A}^2 = 0; \quad A^\sigma A_\sigma = 0. \tag{14.8}$$

Now, subject to (14.8) which we shall review in depth momentarily, we obtained each component of the tetrad ε^μ_y. Collecting all of the results from (14.3) through (14.8), reassembling the complete tetrad $\varepsilon^\mu_y \equiv \delta^\mu_y + \varepsilon'^\mu_y$, and restoring $\rho = q / mc^2$, what we have deduced is that the simultaneous equations in (14.1) are solved by:

$$\varepsilon^\mu_y = \begin{pmatrix} 1 & -\rho A^1 & -\rho A^2 & -\rho A^3 \\ -\rho A^1 & 1 & 0 & 0 \\ -\rho A^2 & 0 & 1 & 0 \\ -\rho A^3 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -qA^1 / mc^2 & -qA^2 / mc^2 & -qA^3 / mc^2 \\ -qA^1 / mc^2 & 1 & 0 & 0 \\ -qA^2 / mc^2 & 0 & 1 & 0 \\ -qA^3 / mc^2 & 0 & 0 & 1 \end{pmatrix}. \tag{14.9}$$

The A^1, A^2 and A^3 above subject to the further constraint (14.8) which means that only two of the three A^k in (14.9) are truly independent. Thus, there are indeed only two degrees of freedom in the original A^μ which again is a downstream result of the gauge conditions (11.4) and (11.5).

15. Massless Photons with Two Helicity States and Coulomb Gauge

Equation (14.7) also part of (14.8), which is the $\mu\nu=00$ pure-time component of (14.2) shown expressly in the top line of (14.3a), is consequential. First, because the Pythagorean sum in (14.7) is equal to zero, it is impossible for all three of A^1, A^2 and A^3 to simultaneously be non-zero and real. In fact, if any of these is real, then at least one other must be imaginary. This means that A^μ under conditions (14.8) no longer represents a classical field $A^\mu=(\phi,\mathbf{A})$ with four real components and inherent gauge ambiguity, but rather a *massless* photon quantum with two degrees of freedom and no gauge ambiguity. This all is confirmed by the fact that $A^0=0$, making it impossible for a massive gauge field travelling along the *z* axis (denoted $\hat{z}$) to keep a longitudinal polarization $\varepsilon_\mu(\hat{z})=(c|\mathbf{p}|,0,0,E)/Mc^2$, see, e.g., section 6.12 of [13] at [6.92]. In the discussion to follow, we shall use "γ," a customary photon notation, as a subscript to designate when particular fields are those of an individual photon. This will distinguish from classical fields external to the photon, for which shall use the subscript "c".

Second, if this photon propagates along the *z* axis and has energy $cq^0=E=h\nu$, then its energy-momentum four-vector is

$$cq^\mu(\hat{z})=(E,0,0,cq_z)=(h\nu,0,0,cq_z). \tag{15.1}$$

Also, because the photon is massless, we must have $0=m_\gamma^2c^4=c^2q_\sigma q^\sigma$ (we shall now use the subscript "γ" to denote photon). Together with the above, this implies that $c^2\mathbf{q}\cdot\mathbf{q}=(h\nu)^2$. With (15.1), the longitudinal orthogonal polarization component must be $\varepsilon^3=0$, thus $A^3=0$. Now (14.7) reduces given (15.1), to $A^1A^1+A^2A^2=0$ which in turn means $A^1=\pm iA^2$. Then, the relation $A^1=\pm iA^2$ is solved by the right- and left-polarization vectors:

$$\varepsilon^\mu_{R,L}(\hat{z})\equiv(0,\mp1,-i,0)/\sqrt{2}, \tag{15.2}$$

again, [13] at [6.92]. In general, this means that:

$$A^\mu_\gamma=A\varepsilon^\mu\exp(-iq_\sigma x^\sigma/\hbar), \tag{15.3}$$

with a dimensionless polarization vector ε^μ and an amplitude *A* having dimensions of energy-per-charge to keep balance because those are the dimensions of A^μ_γ. So not only have the two covariant gauge conditions (9.4) and (9.5) forced A^μ_γ to be massless photons, but they have also forced A^μ_γ to assume the known *right- and left-handed photon helicities*. This is what has become of our remaining two degrees of freedom, precisely in accord with known theory and observation. And, because of (15.2), the gauge potential now introduces imaginary terms into the Riemannian geometry above and beyond the Fourier kernels often used to transform between configuration and momentum space, thus producing a type of Kähler Geometry. Because the only part of (15.3) which is a function of spacetime is the Fourier kernel $\exp\left(-iq_\sigma x^\sigma / \hbar\right)$, this means that in general,

$$i\hbar c\partial_\sigma A_\gamma^\mu = cq_\sigma A_\gamma^\mu = \begin{pmatrix} i\hbar\partial_t & i\hbar c\nabla \end{pmatrix} A_\gamma^\mu = \begin{pmatrix} h\nu & -c\mathbf{q} \end{pmatrix} A_\gamma^\mu . \tag{15.4}$$

Third, it is clear from the above that $q_\sigma \varepsilon^\sigma = 0$, which is a form of the Lorentz gauge that emerges from the classical rendition of (9.5). But because $A_\gamma^0 = 0$ thus $\varepsilon^0 = 0$ we may also deduce that $\mathbf{q}\cdot\boldsymbol{\varepsilon} = \mathbf{0}$ which is the Coulomb gauge. Ordinarily this is a non-covariant gauge choice, see section 6.9 of [13] at [6.67]. Yet here, this is a *covariant* gauge, because it is a consequence of the covariant gauge conditions (9.4) and (9.5). Indeed, (9.4) and (9.5) are responsible for the very structure of (13.5), having caused the peculiar quadratic solution (3.5) to eventually turn into (13.5) via the definitions (13.4) that led among other results, to $A^0 = A_\gamma^0 = 0$ in (14.6). Assembling this with other immediate corollaries and (15.2), we find that:

$$q_\sigma \varepsilon^\sigma = 0;\quad q^k \varepsilon^k = \mathbf{q}\cdot\boldsymbol{\varepsilon} = 0;\quad q_\sigma A_\gamma^\sigma = 0;\quad q^k A_\gamma^k = \mathbf{q}\cdot\mathbf{A}_\gamma = 0;\quad \partial_\sigma A_\gamma^\sigma = 0;\quad \partial_k A_\gamma^k = \nabla\cdot\mathbf{A}_\gamma = \mathbf{0} . \tag{15.5}$$

The above, together with (14.8), will be used extensively to zero out many contracted terms in subsequent calculations.

Finally, let us consider the relation $\mathbf{B} = \nabla\times\mathbf{A}$ between the vector potential $\mathbf{A}$ and the magnetic field $\mathbf{B}$, which via (15.4) may be written as $\hbar\mathbf{B}_\gamma = i\left(\mathbf{q}\times\mathbf{A}_\gamma\right)$ for an individual photon. Referring to just prior to (15.3), it is helpful to note that $\mathbf{q}(\hat{z}) = (0,0,q_z)$ and $\boldsymbol{\varepsilon}_{R,L}(\hat{z}) = (\mp 1, -i, 0)/\sqrt{2}$. Thus, $\mathbf{q}\times\boldsymbol{\varepsilon} = \begin{pmatrix} -iq_z & \mp q_z & 0 \end{pmatrix}/\sqrt{2}$ and $\left(\mathbf{q}\times\boldsymbol{\varepsilon}\right)^2 = 0$. Because $A_\gamma^\mu = A\varepsilon^\mu \exp\left(-iq_\sigma x^\sigma / \hbar\right)$, this also means that $\left(\mathbf{q}\times\mathbf{A}_\gamma\right)^2 = -\hbar^2 {\mathbf{B}_\gamma}^2 = 0$. Recognizing that we can rotate to propagate any other direction without changing the invariant features of this result, this leads to several important observations:

First, this emphasizes how $\mathbf{A} = \mathbf{A}_\gamma$ has now been fully converted to the quantum potential for a photon, and *is not and can no longer be regarded* as a classical potential $\mathbf{A} = \mathbf{A}_c$. Second, as the mediator of electromagnetic interactions, the *z*-traversing photon must have a magnetic field which we now know has the components

$$\hbar\mathbf{B}_\gamma = i\left(\mathbf{q}\times\mathbf{A}_\gamma(\hat{z})\right) = A\begin{pmatrix} q_z & \mp iq_z & 0 \end{pmatrix}\exp\left(-iq_\sigma x^\sigma / \hbar\right)/\sqrt{2} . \tag{15.6}$$

Just as $\mathbf{A}_\gamma$ is orthogonal to $\mathbf{q}$, so too $\mathbf{B}_\gamma$ is orthogonal to $\mathbf{q}$. Third, although the photon magnetic field is nonzero, this $\mathbf{B}_\gamma$ has imaginary components just like $\mathbf{A}_\gamma$, over and above the complex kernel $\exp\left(-iq_\sigma x^\sigma / \hbar\right)$. Fourth, from (15.6) we may calculate that $\mathbf{B}_\gamma^2 = 0$, which stems directly from $\mathbf{A}^2 = \mathbf{A}_\gamma^2 = 0$ found in (14.8). Thus, just as a photon carries energy even though as a luminous boson it is massless, so too a photon has a non-zero magnetic field even though the *magnitude* of that magnetic field is zero, $\left|\mathbf{B}_\gamma\right| = 0$. Fifth, the fact that a classical magnetic field can have a non-zero magnitude $\left|\mathbf{B}_c\right| \neq 0$ is one clear indicator why $\mathbf{A} = \mathbf{A}_\gamma$ *must* be regarded as a photon rather than a classical potential. Sixth, the magnetic field still carries the kernel $\exp\left(-iq_\sigma x^\sigma / \hbar\right)$. As a result, also using (15.1), the four-gradient has the identical form to (15.4) for the photon:

$$i\hbar c \partial_\sigma \mathbf{B}_\gamma = c q_\sigma \mathbf{B}_\gamma = \begin{pmatrix} i\hbar\partial_t & i\hbar c \nabla \end{pmatrix} \mathbf{B}_\gamma = \begin{pmatrix} h\nu & -c\mathbf{q} \end{pmatrix} \mathbf{B}_\gamma \tag{15.7}$$

Seventh, although the classical $A_c^{\ \mu} = \left(\phi, \mathbf{A}_c\right)$ which is an amalgamation of countless individual photons can always be Lorentz transformed into a rest frame $A_c^{\ \mu} = \left(\phi_0, \mathbf{0}\right)$, the photon with $A_\gamma^{\ \mu} = A\varepsilon^\mu \exp\left(-iq_\sigma x^\sigma / \hbar\right)$ as a luminous particle can never be so-transformed. No matter what direction the photon travels, its time component $A_\gamma^0 = 0$ as we deduced at (14.8).

Eighth, although there is no *Lorentz* transformation $A_\gamma^{\ \mu} \to A_c^{\ \mu}$ that can place the photon into a rest frame, it is *formally* possible to use a *U(1) gauge* transformation $qA^\mu \to qA'^\mu \equiv qA^\mu + \hbar c \partial^\mu \Lambda$ to do transform a photon potential into a classical potential. . Specifically, we assign $A^\mu = A_\gamma^\mu$ and $A'^\mu = A_c^\mu$ and write the transformation as $qA_\gamma^\mu \to qA_c^\mu \equiv qA_\gamma^\mu + \hbar c \partial^\mu \Lambda$, whereby the arbitrary gauge parameter $\Lambda\left(t, \mathbf{x}\right)$ is defined by $\hbar c \partial^\mu \Lambda \equiv qA_c^\mu - qA_\gamma^\mu$. For the time component, because $A_\gamma^{\ 0} = 0$ and $A_c^{\ 0} = \phi$, rather simply, we obtain $\hbar c \partial^0 \Lambda \equiv q\phi$. For the space components, we find an interesting wrinkle, owing to the fact that (15.2) and (15.3) are complex, not real, because $\sqrt{2} A_\gamma^\mu = A\left(0, \mp 1, -i, 0\right) \exp\left(-iq_\sigma x^\sigma / \hbar\right)$ for a *z*-traversing photon is a complex vector. Therefore, together with $\hbar c \partial^0 \Lambda \equiv q\phi$ above, and mindful that $\partial^k = -\left(\partial_x, \partial_y, \partial_z\right) = -\nabla$ whereby $-\hbar c \nabla \Lambda \equiv q\mathbf{A}_c - q\mathbf{A}_\gamma$, we find:

$$\begin{aligned} &\hbar c \partial^0 \Lambda = +\hbar c \partial \Lambda / \partial t \equiv q\phi \\ &\hbar c \partial^1 \Lambda = -\hbar c \partial \Lambda / \partial x \equiv qA_c^1 \pm qA \exp\left(-iq_\sigma x^\sigma / \hbar\right) / \sqrt{2} \\ &\hbar c \partial^2 \Lambda = -\hbar c \partial \Lambda / \partial y \equiv qA_c^2 + iqA \exp\left(-iq_\sigma x^\sigma / \hbar\right) / \sqrt{2} \\ &\hbar c \partial^3 \Lambda = -\hbar c \partial \Lambda / \partial x \equiv qA_c^\mu \end{aligned} \tag{15.8}$$

Because the above sets $\partial^1 \Lambda$ and $\partial^2 \Lambda$ to complex numbers stemming from A_γ^μ being a complex vector, the gauge parameter $\Lambda\left(t, \mathbf{x}\right)$ used in this transformation must also be a *complex* number, $\Lambda = a + ib$, once again signifying a type of Kähler Geometry. This is also of interest because historically, Weyl spent over a decade [6], [7], [8] pursuing the ultimately incorrect view that equations of nature should be invariant under a true "gauge" transformation $\varphi \to \varphi' \equiv \exp\left(\Lambda\right)\varphi$ rather than what we know is the correct $\varphi \to \varphi' \equiv \exp\left(i\Lambda\right)\varphi$. Writing Weyl's original misconception as $\varphi \to \varphi' \equiv \exp\left(i\left(-i\Lambda\right)\right)\varphi$, we see that what we understand today to be a real gauge angle was in Weyl's original view equivalent to an imaginary gauge angle. What (15.8) shows is that when we wish to transform between $A_\gamma{}^\mu \leftrightarrow A_\mathrm{c}{}^\mu$, we actually require a hybrid of both Weyl's original view and his eventual result: a complex gauge parameter $\Lambda = a + ib$, with an underlying transformation $\varphi \to \varphi' \equiv \exp\left(i\Lambda\right)\varphi = \exp\left(i\left(a+ib\right)\right)\varphi$. And it is very luminosity of non-material, massless photons with complex components, which is the cause of this. The imaginary part of $\Lambda = a + ib$ only becomes non-zero for transformations between a material $A_\mathrm{c}{}^\mu$ and a luminous $A_\gamma{}^\mu$.

Ninth, although it is *formally* possible to transform between $A_\gamma{}^\mu \leftrightarrow A_\mathrm{c}{}^\mu$, *physically* we cannot do so: At (9.4) and (9.5) we removed two degrees of freedom from A^μ, thereby removing any remaining freedom to transform $A_\gamma{}^\mu \leftrightarrow A_\mathrm{c}{}^\mu$. Equivalently, once two degrees of freedom were covariantly removed from the unrestricted A^μ turning it into the A_γ^μ of (15.3) with the properties (14.8) and (15.5) of a photon which is restricted to the Lorenz and Coulomb gauges, we "broke" the gauge symmetry, and can no longer transform $A_\gamma{}^\mu$ back to $A_\mathrm{c}{}^\mu$. In other words, we cannot "unbreak" a broken gauge symmetry. But we can always trace back the breaking.

Tenth, and finally, although we cannot transform between $A_\gamma{}^\mu \leftrightarrow A_\mathrm{c}{}^\mu$, the electric and magnetic fields contained in the field strength $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ are invariant under a gauge transformation, as has long been well known, because the antisymmetry of $F^{\mu\nu}$ washes out any gauge transformation. Although gauge theory was not known in the late-19th century, in retrospect one reason that Heaviside reformulated Maxwell's equations to eliminate the potential, and went so far as to erroneously argue that physics ought not even bother with a potential and should only use electric and magnetic fields, was because of what we now understand to be the gauge symmetry of **E** and **B**. Therefore, the electric and magnetic fields are invariant under a gauge transformation between $A_\gamma{}^\mu \to A_\mathrm{c}{}^\mu$, and so the transformation from $\left|\mathbf{B}_\gamma\right| = 0$ to $\left|\mathbf{B}_\mathrm{c}\right| \neq 0$ is gauge invariant. However, knowing this, whenever we start with $\mathbf{A}_\gamma$ and end up with a quantity such as $\mathbf{B}_\gamma^2$ in an equation, it is best to leave this as is rather than set $\mathbf{B}_\gamma^2 = 0$, in order to preserve the ability to let conduce a gauge transformation $A_\gamma{}^\mu \to A_\mathrm{c}{}^\mu$ from which $\mathbf{B}_\gamma^2 = 0 \to \mathbf{B}_\mathrm{c}^2 \neq 0$.

This is important to keep in mind, because in the next several sections we will be developing a Dirac Hamiltonian using (14.8) and (15.1) through (15.7) to reduce terms containing **A**, and will set $A^0 = \phi = 0$ throughout using (14.8), (15.2) and (15.3), effectively removing the two degrees of gauge freedom from A^μ as a downstream consequence of the gauge fixing conditions (9.4) and (9.5). Again, this may be thought of as "breaking" the gauge symmetry. Once this is done, however, those terms in the Hamiltonian which contain **A** will *not* be invariant under the quantum-to-classical potential gauge transformation $A_\gamma{}^\mu \to A_\mathrm{c}{}^\mu$. So **A** must be interpreted as the gauge potential for a single individual photon. Conversely, **A** cannot be regarded as part of a classical external potential $A_{\mathrm{c}}^{\ \mu}=(\phi,\mathbf{A}_{\mathrm{c}})$, because the symmetry breaking conditions we will have imposed using (14.8) and (15.1) through (15.7) are conditions that are not followed by a classical potential which has a rest frame, but only by a luminous photon which can never be at rest. Again, this is why, contrasting Dirac's original theory, this paper is titled a "Quantum Theory of *Individual* Electron and Photon Interactions."

This is also important to keep in mind because although **A** cannot be interpreted as an external potential owing to how the gauge symmetry has been broken, in terms which will also arise containing the electric and magnetic fields **E** and **B**, these fields *can* in interpreted as either classical or quantum fields, precisely because **E** and **B** are invariant under gauge transformations. Thus, **E** and **B** will enter the Hamiltonian in exactly the same form whether the gauge potentials are classical $A_{\mathrm{c}}^{\ \mu}$ or quantum mechanical $A_{\gamma}^{\ \mu}$, cf. earlier reference to Heaviside. Put differently, the Hamiltonian terms containing **E** and **B** sans **A** are invariant under gauge transformations and so are invariant under transformations between classical and quantum potentials. Thus, the **E** and **B** which appear, regardless of how we interpret the A^{μ} from which they arise via $-\nabla\phi=\mathbf{E}+\partial\mathbf{A}/c\partial t$ or by $\mathbf{B}=\nabla\times\mathbf{A}$ (which of course have the generally-covariant formulation $F^{\mu\nu}=\partial^{\mu}A^{\nu}-\partial^{\nu}A^{\mu}$), can be interpreted and used either as the **E**, **B** fields of an individual photon, or as classical external **E**, **B** fields arising from the stationary linear amalgamation of a countless multitude of individual luminously-propagating photons.

Finally, it will be of use to examine the electric and magnetic fields associated with a single photon quantum. As with a classical field, the photon field strength is $F^{\mu\nu}=\partial^{\mu}A^{\nu}-\partial^{\nu}A^{\mu}$ which is of course gauge symmetric and thus invariant under a gauge transformation $A_{\gamma}^{\ \mu}\leftrightarrow A_{\mathrm{c}}^{\ \mu}$. Using (15.4) we write this as:

$$i\hbar cF_{\gamma}^{\ \mu\nu}=i\hbar c\partial^{\mu}A_{\gamma}^{\nu}-i\hbar c\partial^{\nu}A_{\gamma}^{\mu}=cq^{\mu}A_{\gamma}^{\nu}-cq^{\nu}A_{\gamma}^{\mu}. \tag{15.9}$$

Using (15.5) and $q_{\sigma}q^{\sigma}=0$ for a luminous photon, the photon current density four-vector is then:

$$\begin{aligned}-4\pi\hbar^{2}J^{\nu}&=-4\pi\hbar^{2}\left(c\rho_{em}\quad \mathbf{J}\right)=-\hbar^{2}c\partial_{\mu}F_{\gamma}^{\ \mu\nu}=i\hbar c\partial_{\mu}\left(i\hbar\partial^{\mu}A_{\gamma}^{\nu}-i\hbar\partial^{\nu}A_{\gamma}^{\mu}\right)\\ &=i\hbar c\partial_{\mu}\left(q^{\mu}A_{\gamma}^{\nu}-q^{\nu}A_{\gamma}^{\mu}\right)=cq_{\mu}q^{\mu}A_{\gamma}^{\nu}-cq_{\mu}q^{\nu}A_{\gamma}^{\mu}=-cq^{\nu}\left(q_{0}A_{\gamma}^{0}+q_{k}A_{\gamma}^{k}\right)=0\end{aligned}, \tag{15.10}$$

that is, $\left(\rho_{em}\quad \mathbf{J}\right)=0$, using the notation ρ_{em} to distinguish this charge density from the substitute variable $\rho=q/mc^{2}$. Note that this zero arises from $q_{\mu}q^{\mu}=0$ because of the massless photon, from $A_{\gamma}^{0}=0$ in (14.8), and from $\mathbf{q}\cdot\mathbf{A}_{\gamma}=0$ in (15.5). So as expected, the photon is not an electromagnetic *source*, but rather is the electromagnetic mediator. But the electric and magnetic fields are still not zero. Specifically, using (14.8) and (15.1) in (15.9) we find:

$$i\hbar c\mathbf{E}_{\gamma}=i\hbar cE_{\gamma}^{\ j}=i\hbar cF_{\gamma}^{\ j0}=i\hbar cF_{\gamma}^{\ j0}=i\hbar cF_{\gamma}^{\ j0}=cq^{j}A_{\gamma}^{\ 0}-cq^{0}A_{\gamma}^{\ j}=-h\nu A_{\gamma}^{\ j}=-h\nu\mathbf{A}_{\gamma}, \tag{15.11}$$

i.e., $h\nu\mathbf{A}_\gamma = -i\hbar c\mathbf{E}_\gamma$ which simplifies to $ic\mathbf{E}_\gamma = -2\pi\nu\mathbf{A}_\gamma$. However, also by (14.8), this means that the magnitude $\mathbf{E}_\gamma{}^2 = 0$. For the magnetic field, we likewise have

$$-i\hbar\mathbf{B}_\gamma = -i\hbar B^i_\gamma = \tfrac{1}{2}\varepsilon^{ijk} i\hbar F_\gamma{}^{jk} = \tfrac{1}{2}\left(\varepsilon^{ijk} q^j A^k_\gamma - \varepsilon^{ijk} q^k A^j_\gamma\right) = \varepsilon^{ijk} q^j A^k_\gamma = \mathbf{q}\times\mathbf{A}_\gamma = -i\hbar\nabla\times\mathbf{A}_\gamma\,, \tag{15.12}$$

which is the usual relation $\mathbf{B} = \nabla\times\mathbf{A}$. However, when we take the magnitude using (14.8) and (15.5) we likewise obtain what we already saw at (15.6) namely:

$$-\hbar^2\mathbf{B}_\gamma{}^2 = -\hbar^2 B^i_\gamma B^i_\gamma = \varepsilon^{ijk}\varepsilon^{ilm} q^j A^k_\gamma q^l A^m_\gamma = q^j A^k_\gamma q^j A^k_\gamma - q^j A^k_\gamma q^k A^j_\gamma = 0\,. \tag{15.13}$$

So, an individual photon has energy but no mass, and has electric and magnetic fields which are non-zero but have zero magnitude. Additionally, writing (15.11) as $\hbar c\mathbf{E}_\gamma = ih\nu\mathbf{A}_\gamma = 2\pi i\hbar\nu\mathbf{A}_\gamma$, then taking the spacetime gradient of each side and using (15.4), we obtain (contrast (15.7) for **B**):

$$\begin{aligned} i\hbar c\partial_\sigma\mathbf{E}_\gamma &= -h\nu\partial_\sigma\mathbf{A}_\gamma = -2\pi\hbar\nu\,\partial_\sigma\mathbf{A}_\gamma = 2\pi i\nu q_\sigma\mathbf{A}_\gamma = 2\pi i\nu q_\sigma\mathbf{A}_\gamma = cq_\sigma\mathbf{E}_\gamma \\ &= \begin{pmatrix} i\hbar\partial_t & i\hbar c\nabla \end{pmatrix}\mathbf{E}_\gamma = \begin{pmatrix} h\nu & -c\mathbf{q} \end{pmatrix}\mathbf{E}_\gamma \end{aligned}\,. \tag{15.14}$$

Again, what will be of particular interest is that while $\mathbf{A}_\gamma$ for the photon is not invariant under the quantum-to-classical gauge transformation $A_\gamma{}^\mu \rightarrow A_{\mathrm{c}}{}^\mu$, the electric and magnetic fields in $F^{\mu\nu}$ *are* invariant. Therefore, when we encounter composite terms such as $ih\nu\mathbf{A}_\gamma = \hbar c\mathbf{E}_\gamma$ where the gauge-dependent $\mathbf{A}_\gamma$ is multiplied by the photon energy $h\nu$, or such as $i\mathbf{q}\times\mathbf{A}_\gamma = \hbar\mathbf{B}$ where the gauge-dependent $\mathbf{A}_\gamma$ is crossed with the photon momentum **q**, these *composite* terms are invariant under gauge transformations. This means that these composite terms, and the field strength $F^{\mu\nu}$ generally, are gauge-invariant whether they represent the electric and magnetic fields of a single photon, or classical electric and magnetic fields externally-applied to a single photon. Thus, wherever **E** and **B** appear, whether obtained from a classical potential $A_{\mathrm{c}}{}^\mu$ or a single-photon $A_\gamma{}^\mu$, these **E** and **B** fields (but not the **A** alone) may be regarded at will (under some carefully-proscribed restraints) as either classical fields or as individual photon fields. We shall see all of this in detail over the next several sections.

16. Maxwell's Equations for Individual Photons

To illustrate the foregoing considerations about the relation between a classical potential $A_{\mathrm{c}}{}^\mu$ and the potential $A^\mu_\gamma = A\varepsilon^\mu \exp\left(-iq_\sigma x^\sigma / \hbar\right)$ for an individual photon, in this section we shall apply Maxwell's equations to individual photons. Not only is this study of independent value in its own right, but it will illustrate how to properly navigate between classical fields and those of a single photon quantum. This will be indispensable when we return to developing the hyper-canonical Dirac equation, starting in the next section.

The relation between a four-potential $\mathbf{A}$ and the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ is covariantly-formulated by $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, which separates into the component equations $\nabla\phi = -\mathbf{E} - \partial\mathbf{A}/c\partial t = -\mathbf{E} - \dot{\mathbf{A}}/c$ and $\nabla\times\mathbf{A} = \mathbf{B}$. In turn, Maxwell's equations are covariantly-formulated by $4\pi J^\mu = \partial_\alpha F^{\alpha\mu}$ and $\partial^\alpha F^{\mu\nu} + \partial^\mu F^{\nu\alpha} + \partial^\nu F^{\alpha\mu} = 0$, a.k.a. $\partial_\alpha {}^*F^{\alpha\mu} = 0$ using the dual fields ${}^*F_{\mu\nu} = \frac{1}{2!}\varepsilon_{\sigma\tau\mu\nu}F^{\sigma\tau}$. The former separates into the component equations $\nabla\cdot\mathbf{E} = 4\pi\rho_{em}$ and $\nabla\times\mathbf{B} = \left(4\pi\mathbf{J} + \partial\mathbf{E}/\partial t\right)/c = \left(4\pi\mathbf{J} + \dot{\mathbf{E}}\right)/c$ which are Gauss' and Ampere's Laws for electricity. The latter separates into $\nabla\cdot\mathbf{B} = 0$ and $\nabla\times\mathbf{E} = -\partial\mathbf{B}/c\partial t = -\dot{\mathbf{B}}/c$ which are Gauss' and Faraday's Laws for magnetism. The absence of a magnetic field divergence in $\nabla\cdot\mathbf{B} = 0$ – colloquially expressed as the non-existence and non-observation of magnetic monopoles – is a mathematical identity that results from inserting the antisymmetric $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ into the cyclic field combination $\partial^\alpha F^{\mu\nu} + \partial^\mu F^{\nu\alpha} + \partial^\nu F^{\alpha\mu} = 0$. Using vectors, this is expressed by the identity $\nabla\cdot\mathbf{B} = \nabla\cdot\left(\nabla\times\mathbf{A}\right) = 0$, namely, the divergence of the curl is zero. In the language of exterior calculus, this has the simplified form $ddA = 0$, and exemplifies the geometric rule $dd = 0$ that the exterior derivative of an exterior derivative, or the boundary of a boundary, is zero.

Most importantly for the present development, these relations apply invariantly, whether $A^\mu = A^\mu_{\text{c}}$ is a classical material potential or $A^\mu = A^\mu_\gamma$ is the quantum potential for a single photon. Thus, $F_{\text{c}}^{\mu\nu} = \partial^\mu A^\nu_{\text{c}} - \partial^\nu A^\mu_{\text{c}}$, $4\pi J^\mu = \partial_\alpha F_{\text{c}}^{\alpha\mu}$ and $\partial^\alpha F_{\text{c}}^{\mu\nu} + \partial^\mu F_{\text{c}}^{\nu\alpha} + \partial^\nu F_{\text{c}}^{\alpha\mu} = 0$ for a classical potential, and $F_\gamma^{\mu\nu} = \partial^\mu A^\nu_\gamma - \partial^\nu A^\mu_\gamma$, $4\pi J^\mu = \partial_\alpha F_\gamma^{\alpha\mu}$ and $\partial^\alpha F_\gamma^{\mu\nu} + \partial^\mu F_\gamma^{\nu\alpha} + \partial^\nu F_\gamma^{\alpha\mu} = 0$ for an individual quantum photon potential. First, let us study these equations as applied to an individual photon, with the energy-momentum four-vector of (15.1) and the photon potential (15.3), absent any external potentials or fields or charge densities.

For an individual photon, for which the scalar potential $A^0 = \phi = 0$ because of (14.8) as represented in (15.2) and (15.3), the curl equation $\mathbf{B} = \nabla\times\mathbf{A}$ remains the same, but the gradient equation reduces to $\nabla\phi = 0 = -\mathbf{E} - \dot{\mathbf{A}}/c$ or, more directly, $\mathbf{E} = -\dot{\mathbf{A}}/c$. Employing (15.4) which, using $h = 2\pi\hbar$ and the radian frequency $\omega = 2\pi\nu$, contains both $i\hbar\dot{\mathbf{A}} = h\nu\mathbf{A} = \hbar\omega\mathbf{A}$ and $i\hbar\nabla\mathbf{A} = -\mathbf{q}\mathbf{A}$, we multiply through by $i\hbar c$, then convert into momentum space, to obtain:

$$\begin{aligned} i\hbar c\mathbf{E}_\gamma &= -i\hbar\dot{\mathbf{A}}_\gamma = -\hbar\omega\mathbf{A}_\gamma \\ i\hbar c\mathbf{B}_\gamma &= i\hbar c\nabla\times\mathbf{A}_\gamma = -c\mathbf{q}\times\mathbf{A}_\gamma \end{aligned} . \tag{16.1}$$

This relates the fields $\mathbf{E}_\gamma$ and $\mathbf{B}_\gamma$ *of an individual photon*, to the photon three- potential $\mathbf{A}_\gamma$. We showed via $\mathbf{E}_\gamma{}^2 = 0$ at (15.11) and $\mathbf{B}_\gamma^2 = 0$ at (15.6) how these two fields, although nonzero, do have zero magnitudes, just as a photon has non-zero energy $h\nu = \hbar\omega$ but zero rest mass. Now let's turn to Maxwell's equations.

Still for an individual photon, (15.10) teaches that $\left(c\rho_{em} \quad \mathbf{J}\right)=0$, i.e., that the luminous photon does not act as an electromagnetic source but only as an interaction mediator. Therefore, in covariant form Maxwell's equations reduce to the source-free, duality-symmetric $\partial_\alpha F^{\alpha\mu}=0$, and $\partial^\alpha F^{\mu\nu}+\partial^\mu F^{\nu\alpha}+\partial^\nu F^{\alpha\mu}=0$ a.k.a. $\partial_\alpha * F^{\alpha\mu}=0$. In component form, this produces $\nabla\cdot\mathbf{E}_\gamma=0$ and $\nabla\cdot\mathbf{B}_\gamma=0$ for the divergence equations, and $c\nabla\times\mathbf{E}_\gamma=-\dot{\mathbf{B}}_\gamma$ and $c\nabla\times\mathbf{B}_\gamma=\dot{\mathbf{E}}_\gamma$ for the curl equations. To convert these into momentum space, we turn to (15.7) and (15.14), which again using ω, contain $i\hbar\dot{\mathbf{B}}_\gamma=\hbar\omega\mathbf{B}_\gamma$ and $i\hbar\nabla\mathbf{B}_\gamma=-\mathbf{q}\mathbf{B}_\gamma$, $i\hbar\dot{\mathbf{E}}_\gamma=\hbar\omega\mathbf{E}_\gamma$ and $i\hbar\nabla\mathbf{E}_\gamma=-\mathbf{q}\mathbf{E}_\gamma$. As a result, Maxwell's equations for an individual photon convert to momentum space as follows:

$$\begin{aligned}
&i\hbar c\nabla\cdot\mathbf{E}_\gamma=-c\mathbf{q}\cdot\mathbf{E}_\gamma=0\\
&i\hbar c\nabla\cdot\mathbf{B}_\gamma=-c\mathbf{q}\cdot\mathbf{B}_\gamma=0\\
&i\hbar c\nabla\times\mathbf{E}_\gamma=-c\mathbf{q}\times\mathbf{E}_\gamma=-i\hbar\dot{\mathbf{B}}_\gamma=-\hbar\omega\mathbf{B}_\gamma\\
&i\hbar c\nabla\times\mathbf{B}_\gamma=-c\mathbf{q}\times\mathbf{B}_\gamma=i\hbar\dot{\mathbf{E}}_\gamma=\hbar\omega\mathbf{E}_\gamma
\end{aligned}. \tag{16.2}$$

Equations (16.2) establish paired relations $c\mathbf{q}\times\mathbf{E}_\gamma=\hbar\omega\mathbf{B}_\gamma$ and $c\mathbf{q}\times\mathbf{B}_\gamma=-\hbar\omega\mathbf{E}_\gamma$ between the electric and magnetic fields $\mathbf{E}_\gamma$ and $\mathbf{B}_\gamma$ of a photon. Again, these are non-zero, but have zero magnitude.

It is further possible to use (16.1) to write (16.2) multiplied through by another $i\hbar c$ in terms of the photon three-potential $\mathbf{A}_\gamma$ as:

$$\begin{aligned}
&-\hbar^2c^2\nabla\cdot\mathbf{E}_\gamma=-i\hbar cc\mathbf{q}\cdot\mathbf{E}_\gamma=c\mathbf{q}\cdot\left(\hbar\omega\mathbf{A}_\gamma\right)=0\\
&-\hbar^2c^2\nabla\cdot\mathbf{B}_\gamma=-i\hbar cc\mathbf{q}\cdot\mathbf{B}_\gamma=c\mathbf{q}\cdot\left(c\mathbf{q}\times\mathbf{A}_\gamma\right)=0\\
&-\hbar^2c^2\nabla\times\mathbf{E}_\gamma=-i\hbar cc\mathbf{q}\times\mathbf{E}_\gamma=-i\hbar c\left(\hbar\omega\mathbf{B}_\gamma\right)=\hbar\omega\left(c\mathbf{q}\times\mathbf{A}_\gamma\right)\\
&-\hbar^2c^2\nabla\times\mathbf{B}_\gamma=-i\hbar cc\mathbf{q}\times\mathbf{B}_\gamma=i\hbar c\left(\hbar\omega\mathbf{E}_\gamma\right)=c\mathbf{q}\times\left(c\mathbf{q}\times\mathbf{A}_\gamma\right)=-\hbar\omega\left(\hbar\omega\mathbf{A}_\gamma\right)
\end{aligned}. \tag{16.3}$$

The first equation contains $\mathbf{q}\cdot\mathbf{A}_\gamma=0$, which is merely (15.5) for the Coulomb gauge which characterizes a photon. Again, it is of consequence that is was *covariantly* derived from (9.4) and (9.5). The second equation contains $\mathbf{q}\cdot\left(\mathbf{q}\times\mathbf{A}_\gamma\right)=0$, which is the momentum space formulation of the identity that the divergence of the curl is zero. From the latter two equations in (16.3) we may extract the momentum space relations:

$$\begin{aligned}
&c\mathbf{q}\times\mathbf{E}_\gamma=i\omega\left(\mathbf{q}\times\mathbf{A}_\gamma\right)\\
&c\mathbf{q}\times\mathbf{B}_\gamma=-i\omega\left(\hbar\omega\mathbf{A}_\gamma/c\right)
\end{aligned} \tag{16.4}$$

between the fields $\mathbf{E}_\gamma$ and $\mathbf{B}_\gamma$ of a photon, and the photon three-potential $\mathbf{A}_\gamma$. Taken together, these are the Maxwell's equations in momentum space, for an individual photon, absent any external potentials or fields or sources.

Second, starting with the individual photon studied in (16.1) through (16.4), we next introduce a classical external potential ϕ, which of course is the time component of the four-vector $A_{\mathrm{c}}^{\mu} = \begin{pmatrix} \phi & \mathbf{A}_{\mathrm{c}} \end{pmatrix}$. We also place an observer at rest in the potential so that $A_{\mathrm{c}}^{\mu} = \begin{pmatrix} \phi & \mathbf{A}_{\mathrm{c}} \end{pmatrix} = \begin{pmatrix} \phi_0 & \mathbf{0} \end{pmatrix}$ or that observer. We further introduce a classical external charge density having the four-vector $J^{\mu} = \begin{pmatrix} c\rho_{em} & \mathbf{J} \end{pmatrix}$. And finally, we shall have the have the photon traverse a region of spacetime in which these A_{c}^{μ} and J^{μ} are non-zero. Under these new circumstances, let us now repeat the calculations of equations (16.1) through (16.4).

As to (16.1), the photon magnetic field continues $\mathbf{B}_\gamma$ to bear the relation $\mathbf{B}_\gamma = \nabla \times \mathbf{A}_\gamma$ to the photon three-potential $\mathbf{A}_\gamma$. However, the photon electric field $\mathbf{E}_\gamma$ will now bear the relation $\mathbf{E}_\gamma = -\nabla\phi - \dot{\mathbf{A}}_\gamma / c$ to $\mathbf{A}_\gamma$, and specifically, this relation will now be modified by the new term $\nabla\phi$ which was zero in (16.1). As such, given the non-zero $\phi = \phi_0$, (16.1) now becomes:

$$\begin{aligned} i\hbar c\mathbf{E}_\gamma &= -i\hbar c\nabla\phi_0 - i\hbar\dot{\mathbf{A}}_\gamma = c\mathbf{q}\phi_0 - \hbar\omega\mathbf{A}_\gamma \\ i\hbar c\mathbf{B}_\gamma &= i\hbar c\nabla\times\mathbf{A}_\gamma = -c\mathbf{q}\times\mathbf{A}_\gamma \end{aligned} \tag{16.5}$$

Clearly, this will revert to (16.1) when $\phi_0 = 0$, as it must.

As to (16.2), because we are now allowing a non-zero source J^{μ}, we must use the complete Maxwell equations with sources, so that (16.2) now becomes:

$$\begin{aligned} i\hbar c\nabla\cdot\mathbf{E}_\gamma &= -c\mathbf{q}\cdot\mathbf{E}_\gamma = 4\pi i\hbar c\rho_{em} \\ i\hbar c\nabla\cdot\mathbf{B}_\gamma &= -c\mathbf{q}\cdot\mathbf{B}_\gamma = 0 \\ i\hbar c\nabla\times\mathbf{E}_\gamma &= -c\mathbf{q}\times\mathbf{E}_\gamma = -i\hbar\dot{\mathbf{B}}_\gamma = -\hbar\omega\mathbf{B}_\gamma \\ i\hbar c\nabla\times\mathbf{B}_\gamma &= -c\mathbf{q}\times\mathbf{B}_\gamma = i\hbar\left(4\pi\mathbf{J} + \dot{\mathbf{E}}_\gamma\right) = 4\pi i\hbar\mathbf{J} + \hbar\omega\mathbf{E}_\gamma \end{aligned} \quad . \tag{16.6}$$

It is easily seen that when $J^{\mu} = 0$, the above will revert to (16.2), as it must.

Now, as we did at (16.3), let us again multiply the above through by another $i\hbar c$ and then combine with (16.5). This produces:

$$\begin{aligned}
&-\hbar^2c^2\nabla\cdot\mathbf{E}_\gamma=-i\hbar c c\mathbf{q}\cdot\mathbf{E}_\gamma=-c\mathbf{q}\cdot c\mathbf{q}\phi_0+c\mathbf{q}\cdot\left(\hbar\omega\mathbf{A}_\gamma\right)=-\left(\hbar\omega\right)^2\phi_0=-4\pi\hbar^2c^2\rho_{em}\\
&-\hbar^2c^2\nabla\cdot\mathbf{B}_\gamma=-i\hbar c c\mathbf{q}\cdot\mathbf{B}_\gamma=c\mathbf{q}\cdot\left(c\mathbf{q}\times\mathbf{A}_\gamma\right)=0\\
&-\hbar^2c^2\nabla\times\mathbf{E}_\gamma=-i\hbar c c\mathbf{q}\times\mathbf{E}_\gamma=c\mathbf{q}\times\left(c\mathbf{q}\phi_0+\hbar\omega\mathbf{A}_\gamma\right)=-i\hbar c\left(\hbar\omega\mathbf{B}_\gamma\right)=\hbar\omega\left(c\mathbf{q}\times\mathbf{A}_\gamma\right)\\
&-\hbar^2c^2\nabla\times\mathbf{B}_\gamma=-i\hbar c c\mathbf{q}\times\mathbf{B}_\gamma=-4\pi\hbar^2c\mathbf{J}+\hbar\omega i\hbar c\mathbf{E}_\gamma=c\mathbf{q}\times\left(c\mathbf{q}\times\mathbf{A}_\gamma\right)\\
&\qquad=c\mathbf{q}\left(c\mathbf{q}\cdot\mathbf{A}_\gamma\right)-\mathbf{A}_\gamma\left(c\mathbf{q}\cdot c\mathbf{q}\right)=-\left(\hbar\omega\right)^2\mathbf{A}_\gamma=-\left(\hbar\omega\right)^2\mathbf{A}_\gamma+c\mathbf{q}\hbar\omega\phi_0-4\pi\hbar^2c\mathbf{J}
\end{aligned}\quad. \tag{16.7}$$

To reduce, we use $c\mathbf{q}\cdot c\mathbf{q}=\left(h\nu\right)^2=\left(\hbar\omega\right)^2$ from (15.1) and $\mathbf{q}\cdot\mathbf{A}_\gamma=0$ from (15.5) in the first line. We see the identity $c\mathbf{q}\times c\mathbf{q}\phi_0=0$ in the third line. In the final equation we use the triple cross identity $c\mathbf{q}\times\left(c\mathbf{q}\times\mathbf{A}_\gamma\right)=c\mathbf{q}\left(c\mathbf{q}\cdot\mathbf{A}_\gamma\right)-\mathbf{A}_\gamma\left(c\mathbf{q}\cdot c\mathbf{q}\right)$, then $c\mathbf{q}\cdot c\mathbf{q}=\left(\hbar\omega\right)^2$, again together with $\mathbf{q}\cdot\mathbf{A}_\gamma=0$ from (15.5). It is easily seen that (16.7) reverts to (16.3) when $J^\mu=0$ and $\phi_0=0$, as it must.

Finally, as we did at (16.4), we isolate the momentum space relations in the above. But first, we reorder the final equation in (16.7) above into the second position in (16.8) below, so as to group together the Maxwell's equations pairs which are generally covariant, also showing underlying the covariant equation. The result is:

$$\left.\begin{aligned}
&c\mathbf{q}\cdot\mathbf{E}_\gamma=-i\omega\hbar\omega\phi_0/c=-4\pi i\hbar c\rho_{em}\\
&c\mathbf{q}\times\mathbf{B}_\gamma=-i\omega\hbar\omega\mathbf{A}_\gamma/c=-i\omega\hbar\omega\mathbf{A}_\gamma/c+i\omega\mathbf{q}\phi_0-4\pi i\hbar\mathbf{J}
\end{aligned}\right\}\ \partial_\alpha F^{\alpha\mu}=4\pi J^\mu$$
$$\left.\begin{aligned}
&c\mathbf{q}\cdot\mathbf{B}_\gamma=ic\mathbf{q}\cdot\left(\mathbf{q}\times\mathbf{A}_\gamma\right)/\hbar=0\\
&c\mathbf{q}\times\mathbf{E}_\gamma=i\omega\left(\mathbf{q}\times\mathbf{A}_\gamma\right)
\end{aligned}\right\}\ \partial^\alpha F^{\mu\nu}+\partial^\mu F^{\nu\alpha}+\partial^\nu F^{\alpha\mu}=0\quad. \tag{16.8}$$

Following this reordering, we see that the second and fourth equations above are respectively identical to the lower and upper equations in (16.4).

Further, from the first and second $4\pi J^\mu=\partial_\alpha F^{\alpha\mu}$ equations, we may deduce:

$$\begin{aligned}
&4\pi\hbar^2c^2\rho_{em}=\hbar\omega\left(\hbar\omega\phi_0\right)\\
&4\pi\hbar^2c\mathbf{J}=c\mathbf{q}\left(\hbar\omega\phi_0\right)
\end{aligned}, \tag{16.9}$$

which combines with clarity into the covariant relation:

$$4\pi\hbar^2cJ^\mu=\hbar\omega\phi_0q^\mu=4\pi\hbar^2c\begin{pmatrix}c\rho_{em} & \mathbf{J}\end{pmatrix}=\hbar\omega\phi_0\begin{pmatrix}\hbar\omega & c\mathbf{q}\end{pmatrix}. \tag{16.10}$$

This, as it must be, is manifestly the same as (15.10) written as $4\pi\hbar^2 cJ^\nu = cq^\nu cq_0 A^0$ after the replacement $A_\gamma^0 = 0 \mapsto A_{\rm c}^0 = \phi = \phi_0$ of the photon potential with an external potential at rest. This occurred prior to (16.5) when we placed the photon in the external potential $A_{\rm c}^\mu = (\phi_0 \quad \mathbf{0})$ and also introduced a non-zero $J^\mu = (\rho_{em} \quad \mathbf{J})$. The offsetting terms $-i\omega\hbar\omega\mathbf{A}_\gamma / c$ in the second line of (16.8) which cancel in (16.9), stem from the relation $q_\mu q^\mu = 0$ for a luminous photon, which was applied in (15.10). We see from (16.10) that as soon as $\phi_0 = 0$, so too does $J^\mu = 0$. So, introducing the external scalar potential $\phi = \phi_0$ is synonymous with introducing J^μ. On reflection, this is obvious: Equations (16.1) through (16.4) describe source-free electromagnetic fields. But a scalar potential must have a material source. Equation (16.10) says exactly that.

Third, let us next Lorentz transform the external potential out of the rest frame and into some relative motion, so that $A_{\rm c}^\mu = (\phi_0 \quad \mathbf{0}) \to (\phi \quad \mathbf{A}_{\rm c})$. What happens to equations (16.7)? Because $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ and Maxwell's $4\pi J^\mu = \partial_\alpha F^{\alpha\mu}$ and $\partial^\alpha F^{\mu\nu} + \partial^\mu F^{\nu\alpha} + \partial^\nu F^{\alpha\mu} = 0$ have identical form whether applied to a classical material $A_{\rm c}^\mu$ potential or a quantum luminous potential A_γ^μ related thereto by the gauge transformation reviewed at (15.8), *nothing at all changes in the form of equations (16.7).* Because electromagnetism is a linear, abelian interaction, potentials are additive, so that the photon potential add to the external potential, yielding an overall potential $A^\mu = A_{\rm c}^\mu + A_\gamma^\mu$. (For practical purposes, $A^\mu = A_{\rm c}^\mu + A_\gamma^\mu \cong A_{\rm c}^\mu$ because the individual photon potential is swamped by the external potential.) So in (16.7), all we need to do *as to form*, is replace $\phi_0 \mapsto \phi$ (to take this out of rest) and replace $\mathbf{A}_\gamma \mapsto \mathbf{A}$ (to add the motion components $\mathbf{A}_{\rm c}$ to the photon components $\mathbf{A}_\gamma$). As to substance, however, there is an important change: Whereas $\mathbf{q} \cdot \mathbf{A}_\gamma = 0$ for an individual photon from (15.5), this is generally not true for an external classical three-potential. Rather, we must regard $\mathbf{q} \cdot \mathbf{A}_{\rm c} \neq 0$ to be nonzero. So, we no longer remove this term. As a result, (16.7), rearranged to have the covariant orderings of (16.8), now becomes:

$$\begin{aligned}
&-\hbar^2 c^2 \nabla \cdot \mathbf{E}_\gamma = -i\hbar c c\mathbf{q} \cdot \mathbf{E}_\gamma = -(\hbar\omega)^2 \phi + \hbar c\omega \mathbf{q} \cdot \mathbf{A} = -4\pi\hbar^2 c^2 \rho_{em} \\
&-\hbar^2 c^2 \nabla \times \mathbf{B}_\gamma = -i\hbar c c\mathbf{q} \times \mathbf{B}_\gamma = -(\hbar\omega)^2 \mathbf{A} + \hbar c\omega \mathbf{q}\phi - 4\pi\hbar^2 c\mathbf{J} = -(\hbar\omega)^2 \mathbf{A} + c\mathbf{q}(c\mathbf{q} \cdot \mathbf{A}) \\
&-\hbar^2 c^2 \nabla \cdot \mathbf{B}_\gamma = -i\hbar c c\mathbf{q} \cdot \mathbf{B}_\gamma = c\mathbf{q} \cdot (c\mathbf{q} \times \mathbf{A}) = 0 \\
&-\hbar^2 c^2 \nabla \times \mathbf{E}_\gamma = -i\hbar c c\mathbf{q} \times \mathbf{E}_\gamma = \hbar c\omega (c\mathbf{q} \times \mathbf{A})
\end{aligned} \quad . \qquad (16.11)$$

The source equations contained in the first two lines above are now:

$$\begin{aligned} 4\pi\hbar^2 c^2 \rho_{em} &= \hbar\omega\left(\hbar\omega\phi - c\mathbf{q}\cdot\mathbf{A}\right) \\ 4\pi\hbar^2 c\mathbf{J} &= c\mathbf{q}\left(\hbar\omega\phi - c\mathbf{q}\cdot\mathbf{A}\right) \end{aligned}, \tag{16.12}$$

which, as they must, reduce to (16.9) when $\mathbf{q}\cdot\mathbf{A}=0$ and the external is Lorentz transformed to a rest frame, $\phi \to \phi_0$. This are, and also must be, the same as $4\pi\hbar^2 cJ^\nu = cq^\nu\left(cq_0A^0 + cq_kA^k\right)$ from (15.10), with $A_\gamma^\mu \mapsto A^\mu = A_\mathrm{c}^\mu + A_\gamma^\mu$.

Fourth and finally, the electric and magnetic fields $\mathbf{E}_\gamma$ and $\mathbf{B}_\gamma$ in (16.11) are still those of an individual photon. Now, let us introduce classical, external electric and magnetic fields $\mathbf{E}_\mathrm{c}$ and $\mathbf{B}_\mathrm{c}$, and ask: what now happens to (16.11)? Again, because electromagnetism is abelian, these fields are additive to those of the photon, so that the total $F^{\mu\nu} = F_\mathrm{c}^{\mu\nu} + F_\gamma^{\mu\nu} \cong F_\mathrm{c}^{\mu\nu}$, again with the photon's $F_\gamma^{\mu\nu}$ swamped by the external classical $F_\mathrm{c}^{\mu\nu}$. Moreover, although the photon potential $A_\gamma^\mu = A\varepsilon^\mu \exp\left(-iq_\sigma x^\sigma/\hbar\right)$ of (15.3) has broken gauge symmetry and is not invariant under the transformation $A_\gamma{}^\mu \to A_\mathrm{c}{}^\mu$, the electric and magnetic fields **E** and **B** *are gauge-invariant fields.* Again, although unknown in the late 19th century, this was central though unbeknownst to Heaviside's reformulation of Maxwell's Treatise to contain only electric and magnetic fields without potentials. So, the introduction of external $\mathbf{E}_\mathrm{c}$ and $\mathbf{B}_\mathrm{c}$ does not in any way change the *form* of equations (16.11). All we need do is replace $\mathbf{E}_\gamma \mapsto \mathbf{E} = \mathbf{E}_\mathrm{c} + \mathbf{E}_\gamma$ and $\mathbf{B}_\gamma \mapsto \mathbf{B} = \mathbf{B}_\mathrm{c} + \mathbf{B}_\gamma$ throughout. Therefore, with these external electromagnetic fields, (16.11) finally becomes:

$$\begin{aligned} -\hbar^2c^2\nabla\cdot\mathbf{E} &= -i\hbar c c\mathbf{q}\cdot\mathbf{E} = -\left(\hbar\omega\right)^2\phi + \hbar c\omega\mathbf{q}\cdot\mathbf{A} = -4\pi\hbar^2c^2\rho_{em} \\ -\hbar^2c^2\nabla\times\mathbf{B} &= -i\hbar c c\mathbf{q}\times\mathbf{B} = -\left(\hbar\omega\right)^2\mathbf{A} + \hbar c\omega\mathbf{q}\phi - 4\pi\hbar^2 c\mathbf{J} = -\left(\hbar\omega\right)^2\mathbf{A} + c\mathbf{q}\left(c\mathbf{q}\cdot\mathbf{A}\right) \\ -\hbar^2c^2\nabla\cdot\mathbf{B} &= -i\hbar c c\mathbf{q}\cdot\mathbf{B} = c\mathbf{q}\cdot\left(c\mathbf{q}\times\mathbf{A}\right) = 0 \\ -\hbar^2c^2\nabla\times\mathbf{E} &= -i\hbar c c\mathbf{q}\times\mathbf{E} = \hbar c\omega\left(c\mathbf{q}\times\mathbf{A}\right) \end{aligned}. \tag{16.13}$$

In (16.13), *all* of the fields and potentials are now classical and external, added to and swamping those of the individual photon. All that remains to represent the individual photon is its energy-momentum vector cq^μ which, via (15.10), is $cq^\mu\left(\hat{z}\right) = \left(h\nu, 0, 0, cq_z\right) = \left(\hbar\omega, 0, 0, \hbar\omega\right)$ for propagation along the positive *z* axis. So (16.13) now characterizes the behavior of the luminous photon energy-momentum cq^μ propagating through external potentials **A** and fields **E** and **B**, and even through spacetime regions with non-zero charge densities ρ_{em} and currents **J**.

Before we conclude, there are a few other lessons we may learn from the foregoing development which will be important as we momentarily return to the development of the hyper-canonical Dirac equation. First, and of great usefulness, the relations $i\hbar\partial_\sigma A_\gamma^\mu = q_\sigma A_\gamma^\mu$ in (15.4),

$i\hbar\partial_\sigma \mathbf{B}_\gamma = q_\sigma \mathbf{B}_\gamma$ in (15.7) and $i\hbar\partial_\sigma \mathbf{E}_\gamma = q_\sigma \mathbf{E}_\gamma$ in (15.14) all allow the heuristic replacement $i\hbar\partial_\sigma \mapsto q_\sigma$ whenever the spacetime gradient ∂_σ operates on *any* of ϕ, $\mathbf{A}_\gamma$, $\mathbf{B}_\gamma$ or $\mathbf{E}_\gamma$. But this is not a general replacement that can be used indiscriminately; its use depends integrally on the operand of ∂_σ. For counterexamples, consider $i\hbar\partial_\sigma |\Psi\rangle = p_\sigma |\Psi\rangle$ thus $i\hbar\partial_\sigma \mapsto p_\sigma$ used at (13.6) when the operand is a fermion wavefunction, and $\nabla p^\mu = -\left(q\left(\mathbf{E}+\dot{\mathbf{A}}/c\right)/E\right)p^\mu$ from (7.10) when the operand is a material energy-momentum p^μ. This leads to the question: can we still apply $i\hbar\partial_\sigma \mapsto q_\sigma$ when the operand is an external classical field $\mathbf{A}_\mathrm{c}$, $\mathbf{E}_\mathrm{c}$ or $\mathbf{B}_\mathrm{c}$?

We need look no further than (16.13) above to directly see that $i\hbar\nabla\mathbf{E} = -\mathbf{qE}$ and $i\hbar\nabla\mathbf{B} = -\mathbf{qB}$ from (15.14) and (15.7) remain fully intact. Likewise, because $\mathbf{B} = \nabla\times\mathbf{A}$, we may discern from the $c\mathbf{q}\times\mathbf{A}$ terms that so too does $i\hbar\nabla\mathbf{A} = -\mathbf{qA}$ from (15.4). Thus, although the classical fields do not contain a Fourier kernel $\exp\left(-iq_\sigma x^\sigma/\hbar\right)$, the symmetry relations applied above to go from (16.2) to (16.13) lead us to conclude that that (15.4), (15.7) and (15.14) do generalize to external classical fields, without the γ designation. Therefore, generally:

$$i\hbar\partial_\sigma A^\mu = q_\sigma A^\mu; \quad i\hbar\partial_\sigma \mathbf{B} = q_\sigma \mathbf{B}; \quad i\hbar c\partial_\sigma \mathbf{E} = cq_\sigma \mathbf{E}. \tag{16.14}$$

In sum, from the development that led from (16.2) to (16.13), we may conclude that $i\hbar\partial_\sigma \mapsto q_\sigma$ can still be used as a heuristic rule whenever the operand is a classical $\mathbf{A}_\mathrm{c}$, $\mathbf{E}_\mathrm{c}$ or $\mathbf{B}_\mathrm{c}$.

Second, while the relations $A_\gamma^0 = \phi = 0$ and $\mathbf{A}_\gamma^2 = 0$ from (14.8), and $\mathbf{q}\cdot\mathbf{A}_\gamma = 0$ and $\nabla\cdot\mathbf{A}_\gamma = 0$ from (15.5) apply to individual photons and will be very helpful to reduce many terms from the Dirac equation when we are considering individual photon behavior, these relations all do *not* apply for a classical potential. Specifically, $A_\mathrm{c}^0 = \phi \neq 0$ and $\mathbf{A}_\mathrm{c}^2 \neq 0$ and $\nabla\cdot\mathbf{A}_\mathrm{c} \neq 0$ for a classical potential, and for a photon in a classical potential, $\mathbf{q}\cdot\mathbf{A}_\mathrm{c} \neq 0$. Therefore, although the symmetry relations reviewed and used to go from (16.2) to (16.13) do enable the heuristic replacement $i\hbar\partial_\sigma \mapsto q_\sigma$ to be inherited by the classical $\mathbf{A}_\mathrm{c}$, $\mathbf{E}_\mathrm{c}$ or $\mathbf{B}_\mathrm{c}$ whenever they are operands of ∂_σ as generalized in (16.14), we will wish to leave $\mathbf{q}\cdot\mathbf{A}_\gamma = 0$ as is without zeroing it out, in those situations where we anticipate later wishing to generalize to a classical potential for which $\mathbf{q}\cdot\mathbf{A}_\mathrm{c} \neq 0$. For example, consider $-\hbar^2c^2\nabla\cdot\mathbf{E} = -i\hbar c c\mathbf{q}\cdot\mathbf{E} = -\left(\hbar\omega\right)^2\phi + \hbar c\omega\mathbf{q}\cdot\mathbf{A} = -4\pi\hbar^2c^2\rho_{em}$ from the top line of (16.13), which includes $i\hbar\nabla\cdot\mathbf{E} = -\mathbf{q}\cdot\mathbf{E}$. For an individual photon, $\phi = 0$ and $\mathbf{q}\cdot\mathbf{A} = 0$, which would imply that $\rho_{em} = 0$, which is the time component of (15.10). But if we encounter a $\nabla\cdot\mathbf{E}_\gamma$ such as in (16.2) but anticipate wanting to examine $\nabla\cdot\mathbf{E}$ generally, we will refrain from setting $\phi = 0$ and $\mathbf{q}\cdot\mathbf{A} = 0$ even when these are zero. Simply put, it is easier to set $\phi = 0$ and $\mathbf{q}\cdot\mathbf{A} = 0$ in (16.13) and revert to (16.2), than to start with (16.2) and generalize to

(16.13) (as we have done here to illustrate this very point). We shall keep this in mind as we now return to the hyper-canonical Dirac equation (13.6) and seek to develop this in the most general form so we can study the interactions of individual fermions and photons in external classical fields and with external sources.

17. The Hyper-Canonical Dirac Equation Generalized to Curved Spacetime

As it stands, while hyper-canonical Dirac equation (13.6) is modeled after Dirac's equation in curved spacetime because of its use of the tetrad ε_y^μ with components deduced in (14.9), it does not yet apply to gravitation. To advance (13.6) to gravitation, let us consider the electromagnetic ε_y^μ alongside the ordinary gravitational tetrad e_a^μ, as well as the electromagnetic gamma matrices $\Gamma_{(\varepsilon)}^\mu \equiv \varepsilon_y^\mu \gamma^y$ alongside the ordinary gravitational $\Gamma_{(g)}{}^\mu \equiv e_a^\mu \gamma^a$. Now, when we have both electromagnetism *and* gravitation, we are required to define a set of complete Γ^μ containing both electromagnetism and gravitation which generalize (13.4) from $\eta^{\mu\nu} \mapsto g^{\mu\nu}$ and also satisfy (13.3), and so are defined *such that* $G^{\mu\nu} \equiv g^{\mu\nu} + \left(q^2 / m^2c^4\right) A^\mu A^\nu \equiv \tfrac{1}{2}\left\{\Gamma^\mu\Gamma^\nu + \Gamma^\nu\Gamma^\mu\right\}$.

Given the two separate tetrad definitions $\Gamma_{(\varepsilon)}^\mu \equiv \varepsilon_y^\mu \gamma^y$ and $\Gamma_{(g)}{}^\mu \equiv e_a^\mu \gamma^a$, there are two possible choices for constructing the complete Γ^μ. The first is to start with the electromagnetic $\Gamma_{(\varepsilon)}^a \equiv \varepsilon_y^a \gamma^y$ developed above (note index switch from μ to *a*), then compound this with gravitation by defining $\Gamma^\mu \equiv e_a^\mu \Gamma_{(\varepsilon)}^a = e_a^\mu \varepsilon_y^a \gamma^y$. The second is to start with the usual gravitational $\Gamma_{(g)}{}^y \equiv e_a^y \gamma^a$ (note index switch from μ to *y*), then compound this with electromagnetism by defining $\Gamma^\mu \equiv \varepsilon_y^\mu \Gamma_{(g)}{}^y = \varepsilon_y^\mu e_a^y \gamma^a$. If we place no restrictions on the ordinary metric tensor $g^{\mu\nu}$ (other than its usual $\mu \leftrightarrow \nu$ symmetry), then these two choices are *not* the same, because (again with some index renaming) $e_a^\mu \varepsilon_y^a \gamma^y \neq \varepsilon_a^\mu e_y^a \gamma^y$ i.e. $\left[e_a^\mu \varepsilon_y^a - \varepsilon_a^\mu e_y^a\right]\gamma^y \neq 0$. Formally stated: the electromagnetic and gravitational tetrads operating on the Dirac gamma do not commute, $[e,\varepsilon]\gamma \neq 0$. Generally, two objects not commuting means they are not independent; presently, $[e,\varepsilon]\gamma \neq 0$ tells us that the electromagnetic interaction energies contained in ε gravitate thus changing the gravitational *e*, as they should. Now, let us examine these two possible choices.

Choosing $\Gamma^\mu \equiv \varepsilon_y^\mu e_a^y \gamma^a$ would yield $G^{\mu\nu} = \varepsilon_y^\mu \varepsilon_z^\nu e_a^y e_b^z \eta^{ab} = \varepsilon_y^\mu \varepsilon_z^\nu g^{yz}$. A simple calculation shows that this is the *incorrect* choice: Sample $G^{00} = \varepsilon_y^0 \varepsilon_z^0 g^{yz}$ and insert the tetrad (14.9) for *z* axis photon propagation, thus $A^3 = 0$. Then set $g^{\mu\nu} = \eta^{\mu\nu}$. Because $G^{\mu\nu} = g^{\mu\nu} + \left(q^2 / m^2c^4\right) A^\mu A^\nu$ and using $A^0 = 0$ from (14.8), we must have $G^{\mu\nu} = \eta^{\mu\nu}$. But in fact this ordering of the tetrads produces the contradictory $G^{00} = \eta^{00} - \rho A^1 \rho A^1 - \rho A^2 \rho A^2$. So this is wrong.

The correct choice is rather to define a complete tetrad

$$\mathrm{E}_y^{\mu} \equiv e_a^{\mu}\varepsilon_y^{a}, \tag{17.1}$$

and likewise to define the complete Γ^{μ} for electromagnetism and gravitation are by:

$$\Gamma^{\mu} \equiv e_a^{\mu}\Gamma_{(\varepsilon)}^{a} = e_a^{\mu}\varepsilon_y^{a}\gamma^{y} = \mathrm{E}_y^{\mu}\gamma^{y}. \tag{17.2}$$

Importantly, because $\eta^{\mu\nu} = \tfrac{1}{2}\left\{\gamma^{\mu},\gamma^{\nu}\right\}$, it is the electromagnetic tetrad which directly couples to flat Minkowski spacetime via $\Gamma_{(\varepsilon)}^{a} = \varepsilon_y^{a}\gamma^{y}$. This is in turn coupled to curved spacetime thus gravitation by the subsequent $\Gamma^{\mu} \equiv e_a^{\mu}\Gamma_{(\varepsilon)}^{a}$.

Combining these definitions and the $G^{\mu\nu} \equiv g^{\mu\nu} + \left(q^2/m^2c^4\right)A^{\mu}A^{\nu} \equiv \tfrac{1}{2}\left\{\Gamma^{\mu}\Gamma^{\nu}+\Gamma^{\nu}\Gamma^{\mu}\right\}$ requirement, we obtain:

$$\begin{aligned} G^{\mu\nu} &\equiv \tfrac{1}{2}\left\{\Gamma^{\mu}\Gamma^{\nu}+\Gamma^{\nu}\Gamma^{\mu}\right\} = e_a^{\mu}e_b^{\nu}\varepsilon_y^{a}\varepsilon_z^{b}\,\tfrac{1}{2}\left\{\gamma^{y}\gamma^{z}+\gamma^{z}\gamma^{y}\right\} = e_a^{\mu}e_b^{\nu}\varepsilon_y^{a}\varepsilon_z^{b}\eta^{yz} = \mathrm{E}_y^{\mu}\mathrm{E}_z^{\nu}\eta^{yz} \\ &= e_a^{\mu}e_b^{\nu}\left(\eta^{ab}+\rho A^{a}\rho A^{b}\right) = e_a^{\mu}e_b^{\nu}\eta^{ab} + e_a^{\mu}e_b^{\nu}\rho A^{a}\rho A^{b} = g^{\mu\nu} + e_a^{\mu}e_b^{\nu}\rho A^{a}\rho A^{b} \equiv g^{\mu\nu} + \rho A^{\mu}\rho A^{\nu} \end{aligned}. \tag{17.3}$$

For this to all be correct, it is necessary via the final definition that:

$$A^{\mu}A^{\nu} = e_a^{\mu}e_b^{\nu}A^{a}A^{b}. \tag{17.4}$$

be true. It will be seen making use of (14.9) for each of the sixteen pairwise μ,ν combinations, as well as $A^{0}=0$ from (14.8), that (17.4) is indeed true; thus so is (17.3).

We may then use Γ^{μ} from (17.2) in place of $\Gamma_{(\varepsilon)}^{\sigma}$ to advance (13.6) to:

$$\left(i\hbar c\left(\Gamma^{\sigma}+\frac{q}{mc^2}A^{\sigma}\right)\partial_{\sigma} - mc^2\right)\left|\Psi\right\rangle = 0. \tag{17.5}$$

Clearly, when $g_{\mu\nu} = \eta_{\mu\nu}$ thus $e_a^{\mu} = \delta_a^{\mu}$ the above will revert to (13.6). This is the hyper-canonical Dirac equation encompassing electromagnetism and gravitation, via the successive couplings of the electromagnetic and gravitational tetrads in $\Gamma^{\mu} = e_a^{\mu}\varepsilon_y^{a}\gamma^{y}$, but without yet a proper spin connection. We now review why this is needed and how it is introduced.

18. The Hyper-Canonical Spin Connection

As reviewed in section 1, in curved spacetime, in order to couple the spinor fields ψ to gravitation, we must advance ∂_{μ} in the Dirac equation $\left(i\hbar\Gamma^{\mu}\partial_{\mu} - mc\right)\psi = 0$ to a spin-covariant derivative $\partial_{\mu} \mapsto \nabla_{\mu} \equiv \partial_{\mu} - \tfrac{i}{4}\omega_{\mu}^{ab}\sigma_{ab}$ using a spin connection $\omega_{\mu}^{ab} \equiv e_{\nu}{}^{a}\partial_{;\mu}e^{\nu b}$, whereby Dirac's equation becomes $\left(i\hbar\Gamma^{\mu}\nabla_{\mu}-mc\right)\psi=0$, and where $\partial_{;\mu}e^{\nu b}=\partial_{\mu}e^{\nu b}+\Gamma^{\nu}_{\sigma\mu}e^{\sigma b}$ is the gravitational covariant derivative of $e^{\nu b}$. Put more formally, in the usual Dirac equation in curved spacetime, ∇_{μ} in $\nabla_{\mu}\psi$ does correctly operate as a generally-covariant vector because it contains the covariant $\partial_{;\mu}$, while ∂_{σ} in $\partial_{\sigma}\psi$ does not. This is why the spin connection is required. The same considerations must now be applied to ∂_{μ} in the Dirac-type equation (17.5) with gravitation.

To guide us on how to construct the required spin connection for (17.5), let us briefly review in more detail how this is ordinarily done for Dirac's equation. First, we note from the product rule that $\partial_{;\mu}\left(e_{\nu}{}^{a}e^{\nu b}\right)=\partial_{;\mu}e_{\nu}{}^{a}e^{\nu b}+e_{\nu}{}^{a}\partial_{;\mu}e^{\nu b}=\partial_{;\mu}e_{\nu}{}^{a}e^{\nu b}+\omega_{\mu}^{ab}$. So ω_{μ}^{ab} is actually one of the two terms in the covariant derivative $\partial_{;\mu}$ of $e_{\nu}{}^{a}e^{\nu b}=g_{\mu\nu}e^{\mu a}e^{\nu b}$. With this in mind, we start with $g^{\mu\nu}=e^{\mu}_{a}e^{\nu}_{b}\eta^{ab}$ and calculate that $g^{\mu\nu}=e^{\mu}_{a}e^{\nu}_{b}\eta^{ab}=e^{\mu}_{a}e^{\nu b}\delta^{a}{}_{b}=e^{\mu}_{a}e^{\nu a}$. This true because η^{ab} is used to raise and lower the flat spacetime Lorentz indexes. We may now lower a world index to obtain $\delta^{\mu}{}_{\nu}=e^{\mu}_{a}e^{a}_{\nu}$. Then, we again start with $g^{\mu\nu}=e^{\mu}_{a}e^{\nu}_{b}\eta^{ab}$ but this time we form the identity $g^{\mu\nu}=\delta^{\mu}{}_{\sigma}\delta^{\nu}{}_{\tau}g^{\sigma\tau}=e^{\mu}_{a}e^{\nu}_{b}\eta^{ab}$. We then use $\delta^{\mu}{}_{\nu}=e^{\mu}_{a}e^{a}_{\nu}$ with renamed indexes to write this identity as $g^{\mu\nu}=e^{\mu}_{a}e^{a}_{\sigma}e^{\nu}_{b}e^{b}_{\tau}g^{\sigma\tau}=e^{\mu}_{a}e^{\nu}_{b}\eta^{ab}$. We then divide out the two tetrads common to each side to deduce $\eta^{ab}=e^{a}_{\sigma}e^{b}_{\tau}g^{\sigma\tau}$, which is the inverse of $g^{\mu\nu}=e^{\mu}_{a}e^{\nu}_{b}\eta^{ab}$. By simple rearrangement of indexes in this inverse we obtain $\eta^{ab}=e^{a}_{\nu}e^{\nu b}$ which is the term for which the derivative is taken at the start of this paragraph using the product.

Therefore, because η^{ab} is a constant, $0=\partial_{;\mu}\eta^{ab}=\partial_{;\mu}e^{a}_{\nu}e^{\nu b}+e^{a}_{\nu}\partial_{;\mu}e^{\nu b}$. This in turn leads us with some further rearrangement to deduce that $\omega_{\mu}^{ab}=e^{a}_{\nu}\partial_{;\mu}e^{\nu b}=-e^{b}_{\nu}\partial_{;\mu}e^{\nu a}=-\omega_{\mu}^{ba}$, from which we learn that ω_{μ}^{ab} is antisymmetric in its *a*, *b* indexes. This is important, because antisymmetric tensors $A^{[\mu\nu]}$ can always be constructed by $A^{[\mu\nu]}\equiv\frac{1}{2}\left[A^{\mu\nu}-A^{\nu\mu}\right]$ from arbitrary tensors $A^{\mu\nu}$, which in the present context would have us defining $\omega_{\mu}^{ab}\equiv\frac{1}{2}\left[e_{\nu}{}^{a}\partial_{;\mu}e^{\nu b}-e^{b}_{\nu}\partial_{;\mu}e^{\nu a}\right]$. However, this construction is unnecessary here, because $\omega_{\mu}^{ab}=e^{a}_{\nu}\partial_{;\mu}e^{\nu b}=-e^{b}_{\nu}\partial_{;\mu}e^{\nu a}=-\omega_{\mu}^{ba}$ shows how $\omega_{\mu}^{ab}\equiv e_{\nu}{}^{a}\partial_{;\mu}e^{\nu b}$ is naturally antisymmetric without special construction. Therefore, we also deduce that $\partial_{;\mu}\left(e_{\nu}{}^{a}e^{\nu b}\right)=\partial_{;\mu}e_{\nu}{}^{a}e^{\nu b}+e_{\nu}{}^{a}\partial_{;\mu}e^{\nu b}=\omega_{\mu}^{ba}+\omega_{\mu}^{ab}=0$.

Now we turn to the spin-covariant derivative $\nabla_{\mu}=\partial_{\mu}-\frac{i}{4}\omega_{\mu}^{ab}\sigma_{ab}$. Starting with $g^{\mu\nu}=\frac{1}{2}\left\{\Gamma^{\mu}_{(g)}\Gamma^{\nu}_{(g)}+\Gamma^{\nu}_{(g)}\Gamma^{\mu}_{(g)}\right\}$ with $\Gamma^{\mu}_{(g)}=e^{\mu}_{a}\gamma^{\alpha}$ we see that $\Gamma_{(g)\nu}\Gamma^{\nu}_{(g)}=\delta_{\nu}{}^{\nu}=4$ and via the product rule that $\partial_{;\mu}\left(\Gamma_{(g)\nu}\Gamma^{\nu}_{(g)}\right)=\partial_{;\mu}\Gamma_{(g)\nu}\Gamma^{\nu}_{(g)}+\Gamma_{(g)\nu}\partial_{;\mu}\Gamma^{\nu}_{(g)}=0$ thus $-\partial_{;\mu}\Gamma_{(g)\nu}\Gamma^{\nu}_{(g)}=\Gamma_{(g)\nu}\partial_{;\mu}\Gamma^{\nu}_{(g)}$. So writing the derivative out fully and using $\Gamma_{(g)\nu}=e_{\nu}{}^{a}\gamma_{a}$ and $\partial_{;\mu}\Gamma^{\nu}_{(g)}=\partial_{;\mu}e^{\nu b}\gamma_{b}$, and with the Dirac covariants $\sigma_{ab}=\frac{i}{2}\left[\gamma_{a}\gamma_{b}-\gamma_{b}\gamma_{a}\right]$, we obtain:

$$\begin{aligned}\nabla_\mu &= \partial_\mu - \tfrac{i}{4}\omega_\mu^{ab}\sigma_{ab} = \partial_\mu + \tfrac{1}{8}\left(e_\nu{}^a \partial_{;\mu} e^{\nu b}\right)\left[\gamma_a\gamma_b - \gamma_b\gamma_a\right] = \partial_\mu + \tfrac{1}{8}\left[\Gamma_{(g)\nu}\partial_{;\mu}\Gamma^\nu_{(g)} - \partial_{;\mu}\Gamma^\nu_{(g)}\Gamma_{(g)\nu}\right] \\ &= \partial_\mu + \tfrac{1}{4}\Gamma_{(g)\nu}\partial_{;\mu}\Gamma^\nu_{(g)} = \partial_\mu + \tfrac{1}{4}\Gamma_{(g)\nu}\partial_\mu\Gamma^\nu_{(g)} + \tfrac{1}{4}\Gamma^\nu_{\sigma\mu}\Gamma_{(g)\nu}\Gamma^\sigma_{(g)}\end{aligned}, \tag{18.1}$$

where $\Gamma^\nu{}_{\sigma\mu} = \frac{1}{2}g^{\nu\beta}\left(\partial_\sigma g_{\mu\beta} + \partial_\mu g_{\beta\sigma} - \partial_\beta g_{\sigma\mu}\right)$ are the Christoffel connections. With the derivative written in this way, Dirac's free-fermion equation in curved spacetime now becomes:

$$\left(i\hbar c\Gamma^\sigma_{(g)}\nabla_\sigma - mc^2\right)\psi = \left(i\hbar c\Gamma^\sigma_{(g)}\partial_\sigma + \tfrac{1}{4}i\hbar c\Gamma^\sigma_{(g)}\Gamma_{(g)\nu}\partial_{;\sigma}\Gamma^\nu_{(g)} - mc^2\right)\psi = 0\,. \tag{18.2}$$

It is easier to calculate using this form of ∇_μ because the Lorentz indexes a, b are entirely hidden. Furthermore, we see that the factor of ¼ in (18.1) simply normalizes the $\frac{1}{4}\Gamma_{(g)\nu}\partial_{;\mu}\Gamma^\nu_{(g)}$ term to $\frac{1}{4}\Gamma_{(g)\nu}\Gamma^\nu_{(g)} = 1$ deduced just above. And, we see clearly why the spin connection containing $\partial_{;\sigma}\Gamma^\nu_{(g)} = \partial_\sigma\Gamma^\nu_{(g)} + \Gamma^\nu_{\sigma\tau}\Gamma^\tau_{(g)}$ rather than merely $\partial_\sigma\Gamma^\nu_{(g)}$ is needed to ensure that the derivative ∇_μ operates covariantly on ψ. Then, when we reduce to flat spacetime, $\partial_{;\sigma}\Gamma^\nu_{(g)}$ of course becomes $\partial_\sigma\Gamma^\nu_{(g)}$. But also, $\partial_\sigma\Gamma^\nu_{(g)} \to \partial_\sigma\gamma^\nu = 0$ because $e_b^\nu \to \delta_b^\nu$. As a consequence, the entire $\nabla_\sigma \to \partial_\sigma$ and Dirac's free-fermion equation reduces to the usual familiar $\left(i\hbar c\gamma^\sigma\partial_\sigma - mc^2\right)\psi = 0$.

With the foregoing in mind, we return to (17.3) and specifically the relation $G^{\mu\nu} = \mathrm{E}_y^\mu \mathrm{E}_z^\nu \eta^{yz}$ which is analogous to $g^{\mu\nu} = e_a^\mu e_b^\nu \eta^{ab}$ which was central to the review just concluded. Now, our objective is to obtain a spin connection analogous to that in $\partial_{;\mu}\left(e_\nu{}^a e^{\nu b}\right) = \partial_{;\mu}e_\nu{}^a e^{\nu b} + e_\nu{}^a\partial_{;\mu}e^{\nu b} = \partial_{;\mu}e_\nu{}^a e^{\nu b} + \omega_\mu^{ab}$. To focus for the moment simply on the electromagnetic tetrad, we set $g_{\mu\nu} = \eta_{\mu\nu}$ thus $e_a^\mu = \delta_a^\mu$ thus $G^{\mu\nu} = \eta^{\mu\nu} + \rho A^\mu \rho A^\nu = \varepsilon_y^\mu \varepsilon_z^\nu \eta^{yz}$. Similarly to the above, the Lorentz indexes are raised and lowered with η^{yz}, so we may write $G^{\mu\nu} = \eta^{\mu\nu} + \rho A^\mu \rho A^\nu = \varepsilon_y^\mu \varepsilon^{\nu y}$, i.e., $\varepsilon_y^\mu \varepsilon^{\nu y} = \eta^{\mu\nu} + \rho A^\mu \rho A^\nu$. This differs from the form of the earlier $e_a^\mu e^{\nu a} = g^{\mu\nu}$ because of the extra $\rho A^\mu \rho A^\nu$ term, so we cannot calculate the inverse in the same way. Instead, because $e_\nu{}^a e^{\nu b}$ is the starting point for the relation $\partial_{;\mu}\left(e_\nu{}^a e^{\nu b}\right) = \partial_{;\mu}e_\nu{}^a e^{\nu b} + e_\nu{}^a\partial_{;\mu}e^{\nu b} = \omega_\mu^{ba} + \omega_\mu^{ab} = 0$, let us simply start by using (14.9) to explicitly construct $\varepsilon_\nu^y \varepsilon^{\nu z}$. As in the above we may use η_{yz} to operate on the Lorentz indexes and $g_{\mu\nu} = \eta_{\mu\nu}$ to operate on the spacetime indexes, so that $\varepsilon_\nu^y = \eta_{\mu\nu}\eta^{wy}\varepsilon_w^\mu$ and $\varepsilon^{\nu z} = \eta^{xz}\varepsilon_x^\nu$, thus, $\varepsilon_\nu^y \varepsilon^{\nu z} = \eta_{\mu\nu}\eta^{wy}\eta^{xz}\varepsilon_w^\mu \varepsilon_x^\nu$. After we do this explicit construction, we obtain $\varepsilon_\nu^y \varepsilon^{\nu z} = \eta^{yz} + \rho^2 A^y A^z$, which is the inverse of $\varepsilon_y^\mu \varepsilon^{\nu y} = \eta^{\mu\nu} + \rho A^\mu \rho A^\nu$. So the inverse again swaps Lorentz and spacetime indexes, but there is now an extra term with $\rho^2 A^y A^z$.

As a consequence, using the *ordinary* derivative because we are presently considering $g_{\mu\nu} = \eta_{\mu\nu}$ thus $\partial_{;\sigma} = \partial_\sigma$, we obtain:

$$\partial_\sigma\left(\varepsilon^y_\nu\varepsilon^{\nu z}\right)=\partial_\sigma\varepsilon^y_\nu\varepsilon^{\nu z}+\varepsilon^y_\nu\partial_\sigma\varepsilon^{\nu z}=\rho^2\partial_\sigma A^y A^z+\rho^2 A^y\partial_\sigma A^z . \tag{18.3}$$

We may rewrite this to define an electromagnetic tetrad spin connection by:

$$\Omega^{yz}_\sigma\equiv\varepsilon^y_\nu\partial_\sigma\varepsilon^{\nu z}-\rho^2 A^y\partial_\sigma A^z=\varepsilon^z_\nu\partial_\sigma\varepsilon^{\nu y}+\rho^2 A^z\partial_\sigma A^y=-\Omega^{zy}_\sigma . \tag{18.4}$$

This is naturally antisymmetric in the Lorentz indexes, $\Omega^{yz}_\sigma=-\Omega^{zy}_\sigma$, but only with the extra term $\rho^2 A^y\partial_\sigma A^z$ included.

Contrasting (18.1), we then define a spin-covariant derivative for the electromagnetic tetrads by $\nabla_\sigma\equiv\partial_\sigma-\frac{i}{4}\Omega^{yz}_\sigma\sigma_{yz}$. Using $\Gamma_{(e)\nu}=\varepsilon^y_\nu\gamma_y$ and $\sigma_{ab}=\frac{i}{2}\left[\gamma_a,\gamma_b\right]$, this is rewritten as:

$$\begin{aligned}\nabla_\sigma&\equiv\partial_\sigma-\tfrac{i}{4}\Omega^{yz}_\sigma\sigma_{yz}=\partial_\sigma+\tfrac{1}{8}\left(\varepsilon^y_\nu\partial_\sigma\varepsilon^{\nu z}-\rho^2A^y\partial_\sigma A^z\right)\left[\gamma_y\gamma_z-\gamma_z\gamma_y\right]\\&=\partial_\sigma+\tfrac{1}{8}\left[\varepsilon^y_\nu\gamma_y\partial_\sigma\varepsilon^{\nu z}\gamma_z-\partial_\sigma\varepsilon^{\nu z}\gamma_z\varepsilon^y_\nu\gamma_y\right]-\tfrac{1}{8}\rho^2\gamma_y\gamma_z\left[A^y\partial_\sigma A^z-A^z\partial_\sigma A^y\right]\\&=\partial_\sigma+\tfrac{1}{8}\left[\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}-\partial_\sigma\Gamma^\nu_{(e)}\Gamma_{(e)\nu}\right]-\tfrac{1}{8}\rho^2\gamma_y\gamma_z\left[-\partial_\sigma A^yA^z+\partial_\sigma A^zA^y\right]\\&=\partial_\sigma+\tfrac{1}{4}\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}-\tfrac{1}{8}i\rho^2\gamma_y\gamma_z q_\sigma\left[A^y,A^z\right]/\hbar=\partial_\sigma+\tfrac{1}{4}\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}\end{aligned} . \tag{18.5}$$

The final result, $\nabla_\sigma=\partial_\sigma+\frac{1}{4}\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}$, has identical form to (18.1), except that $\frac{1}{4}\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}$ contains an ordinary derivative because at the moment we are using $g_{\mu\nu}=\eta_{\mu\nu}$. To reduce in the above, just as before (18.1) we may deduce that $G_\nu^{\ \nu}=\Gamma_{(e)\nu}\Gamma^\nu_{(e)}=\delta_\nu^{\ \nu}+\rho A_\nu\rho A^\nu=4$ using (17.3) with $g_{\mu\nu}=\eta_{\mu\nu}$ and applying $A^\sigma A_\sigma=0$ from (14.8). Therefore $-\partial_\sigma\Gamma^\nu_{(e)}\Gamma_{(e)\nu}=\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}$. Additionally we apply $\partial_\sigma\left(A^yA^z\right)=\partial_\sigma A^yA^z+A^y\partial_\sigma A^z$ and the fact that photon vectors are commuting $\left[A^y,A^z\right]=0$ because electrodynamics is an abelian gauge theory. So although Ω^{yz}_σ in (18.4) has an extra term $-\rho^2A^y\partial_\sigma A^z$ not contained in the gravitational $\omega^{ab}_\mu=e^a_\nu\partial_{;\mu}e^{\nu b}$, this term washes out from ∇_σ in (18.5) owing – not to any of the zero relations in (15.5) – but to electrodynamics being an abelian gauge theory.

With this we return first to (13.6) because that applies to electromagnetism absent gravitation, and we advance $\partial_\sigma\mapsto\nabla_\sigma$ using (18.5), thus obtaining:

$$\left(\left(\Gamma^\sigma_{(\varepsilon)}+\frac{q}{mc^2}A^\sigma\right)i\hbar c\nabla_\sigma-mc^2\right)\left|\Psi\right\rangle=\left(\left(\Gamma^\sigma_{(\varepsilon)}+\frac{q}{mc^2}A^\sigma\right)i\hbar c\left(\partial_\sigma+\frac{1}{4}\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}\right)-mc^2\right)\left|\Psi\right\rangle=0 . \tag{18.6}$$

To then broaden this to apply to gravitation, we merely generalize $\eta_{\mu\nu}\mapsto g_{\mu\nu}$. Via the minimal coupling principle we simultaneously generalize $\Gamma^a_{(\varepsilon)}=\varepsilon^a_y\gamma^y$ back to $\Gamma^\mu=\mathrm{E}^\mu_y\gamma^y=e^\mu_a\varepsilon^a_y\gamma^y$ using

(17.2) and $G^{\mu\nu}=\eta^{\mu\nu}+\rho A^{\mu}\rho A^{\nu}=\varepsilon^{\mu}_{y}\varepsilon^{\nu}_{z}\eta^{yz}$ back to $G^{\mu\nu}=g^{\mu\nu}+\rho A^{\mu}\rho A^{\nu}=\mathrm{E}^{\mu}_{y}\mathrm{E}^{\nu}_{z}\eta^{yz}$ using (17.3). And finally, we turn the ordinary derivative in $\partial_{\sigma}\Gamma^{\nu}_{(e)}$ into a covariant derivative of the form $\partial_{\sigma}\Gamma^{\nu}_{(e)}\mapsto\partial_{;\sigma}\Gamma^{\nu}=\partial_{\sigma}\Gamma^{\nu}+\Gamma^{\nu}_{\mu\sigma}\Gamma^{\mu}$ because the now spacetime is curved. Therefore, the spin-covariant derivative (18.5) becomes:

$$\nabla_{\sigma}=\partial_{\sigma}+\tfrac{1}{4}\Gamma_{\nu}\partial_{;\sigma}\Gamma^{\nu}=\partial_{\sigma}+\tfrac{1}{4}\Gamma_{\nu}\partial_{\sigma}\Gamma^{\nu}+\tfrac{1}{4}\Gamma^{\nu}_{\mu\sigma}\Gamma_{\nu}\Gamma^{\mu}\,. \tag{18.7}$$

Finally, the complete hyper-canonical Dirac equation with gravitation is:

$$\boxed{\left(i\hbar c\left(\Gamma^{\sigma}+\frac{q}{mc^{2}}A^{\sigma}\right)\nabla_{\sigma}-mc^{2}\right)\left|\Psi\right\rangle=\left(i\hbar c\left(\Gamma^{\sigma}+\frac{q}{mc^{2}}A^{\sigma}\right)\left(\partial_{\sigma}+\frac{1}{4}\Gamma_{\nu}\partial_{;\sigma}\Gamma^{\nu}\right)-mc^{2}\right)\left|\Psi\right\rangle=0}\,. \tag{18.8}$$

Using the relations $\left|\Psi\right\rangle\equiv\exp\left(-iH_{\sigma}x^{\sigma}/\hbar c\right)\left|U_{0}\right\rangle$ and $i\hbar c\partial_{\sigma}\left|\Psi\right\rangle=cp_{\sigma}\left|\Psi\right\rangle$ from prior to (13.6), to parallel (18.7) we may also define a spin-covariant momentum:

$$c\Pi_{\sigma}\equiv cp_{\sigma}+\tfrac{1}{4}i\hbar c\Gamma_{\nu}\partial_{;\sigma}\Gamma^{\nu}=cp_{\sigma}+\tfrac{1}{4}i\hbar c\Gamma_{\nu}\partial_{\sigma}\Gamma^{\nu}+\tfrac{1}{4}i\hbar c\Gamma^{\nu}_{\mu\sigma}\Gamma_{\nu}\Gamma^{\mu}\,, \tag{18.9}$$

and then convert (18.8) into momentum space to arrive at:

$$\boxed{\left(\left(\Gamma^{\sigma}+\frac{q}{mc^{2}}A^{\sigma}\right)c\Pi_{\sigma}-mc^{2}\right)\left|U_{0}\right\rangle=\left(\left(\Gamma^{\sigma}+\frac{q}{mc^{2}}A^{\sigma}\right)\left(cp_{\sigma}+\frac{1}{4}i\hbar c\Gamma_{\nu}\partial_{;\sigma}\Gamma^{\nu}\right)-mc^{2}\right)\left|U_{0}\right\rangle=0}\,. \tag{18.10}$$

Respectively, (18.8) and (18.10) are the hyper-canonical Dirac equations in configuration and momentum space, including electrodynamics and gravitation. All Lorentz indexes are hidden in $\Gamma_{\nu}=\mathrm{E}^{y}_{\nu}\gamma_{y}$. If we contrast (18.8) to the usual Dirac equation $\left(i\hbar c\Gamma^{\sigma}_{(g)}\left(\partial_{\sigma}+\tfrac{1}{4}\Gamma_{(g)\nu}\partial_{;\sigma}\Gamma^{\nu}_{(g)}\right)-mc^{2}\right)\psi=0$ with gravitation, there are two main differences: First, in the usual Dirac equation $\Gamma^{\nu}_{(g)}=e^{\nu}_{a}\gamma^{a}$ couples only to the gravitational e^{ν}_{a}, whereas $\Gamma^{\nu}=\mathrm{E}^{\nu}_{y}\gamma^{y}=e^{\nu}_{a}\varepsilon^{a}_{y}\gamma^{y}$ in (18.8) contains both the electromagnetic ε^{a}_{y} derived in (14.9) and the gravitational e^{ν}_{a}. Thus, we replace $\Gamma^{\nu}_{(g)}\mapsto\Gamma^{\nu}$. Second, in the usual Dirac equation $\Gamma^{\sigma}_{(g)}$ stands alone contracting with $\nabla_{\sigma}=\partial_{\sigma}+\tfrac{1}{4}\Gamma_{(g)\nu}\partial_{;\sigma}\Gamma^{\nu}_{(g)}$, while in the hyper-canonical Dirac equation we find that $\Gamma^{\sigma}_{(g)}\mapsto\Gamma^{\sigma}+qA^{\sigma}/mc^{2}$, adding an extra qA^{σ}/mc^{2} term. This is reminiscent (and in fact yet another downstream consequence) of how $\partial_{\mu}\mapsto\mathfrak{D}_{\mu}=\partial_{\mu}-iqA_{\mu}/\hbar c$ and $p^{\mu}\mapsto\pi^{\mu}=p^{\mu}+qA^{\mu}/c$ as a result of Weyl's Local U(1) Gauge Symmetry, as reviewed in section 1. These (18.8) and (18.10) are now in a form enabling Hamiltonian calculations to be carried out as simply as possible, which is our next undertaking.

PART IV: THE HYPER-CANONCIAL DIRAC HAMILTONIAN, WITH INHERENT MAGNETIC MOMENT ANOMALY AND NO RENORMALIZATION

19. Preparing the Dirac Equation for Calculating the Hamiltonian

To obtain the Dirac Hamiltonian, we start with (18.10) in momentum space. Because our interest is in the electrodynamic Hamiltonian and particularly showing how (18.10) naturally contains the magnetic moment anomaly obviating any need for renormalization, we shall eliminate gravitation and work in flat spacetime by setting $g_{\mu\nu}=\eta_{\mu\nu}$. This also means that we replace $\Gamma^{\nu}=\mathrm{E}^{\nu}_{y}\gamma^{y}=e^{\nu}_{a}\varepsilon^{a}_{y}\gamma^{y}$ with $\Gamma^{\nu}_{(e)}=\varepsilon^{\nu}_{y}\gamma^{y}$ because $e^{\nu}_{a}=\delta^{\nu}_{a}$, and that we replace $\partial_{;\sigma}\Gamma^{\nu}$ with the ordinary derivative $\partial_{\sigma}\Gamma^{\nu}$ because in flat spacetime the connections $\Gamma^{\nu}_{\mu\sigma}=0$. With these changes, and using $\rho=q/mc^{2}$ for compactness, (18.10) becomes:

$$\left(\left(\Gamma^{\sigma}_{(e)}+\rho A^{\sigma}\right)c\Pi_{(e)\sigma}-mc^{2}\right)\left|U_{0}\right\rangle=\left(\left(\Gamma^{\sigma}_{(e)}+\rho A^{\sigma}\right)\left(cp_{\sigma}+\tfrac{1}{4}i\hbar c\Gamma_{(e)\nu}\partial_{\sigma}\Gamma^{\nu}_{(e)}\right)-mc^{2}\right)\left|U_{0}\right\rangle=0\,. \tag{19.1}$$

This will be our starting point for extracting the Hamiltonian, and it includes the spin connection term $\tfrac{1}{4}i\hbar c\Gamma_{(e)\nu}\partial_{\sigma}\Gamma^{\nu}_{(e)}$ which descends from $\tfrac{1}{4}i\hbar c\Gamma_{\nu}\partial_{;\sigma}\Gamma^{\nu}$ in (18.10) and does not disappear as reviewed after (18.2) where $\partial_{\sigma}\Gamma^{\nu}_{(g)}\rightarrow\partial_{\sigma}\gamma^{\nu}=0$. This is because here, $\partial_{;\sigma}\Gamma^{\nu}\rightarrow\partial_{\sigma}\Gamma^{\nu}_{(e)}\neq 0$. As we shall soon see, this spin connection term is central to how (19.1) obviates renormalization.

First, we obtain the four components of $\Gamma^{\sigma}_{(\varepsilon)}=\varepsilon^{\sigma}_{y}\gamma^{y}$ using ε^{σ}_{y} derived in (14.9), as such:

$$\begin{aligned}
\Gamma^{0}_{(\varepsilon)}&=\varepsilon^{0}_{y}\gamma^{y}=\varepsilon^{0}_{0}\gamma^{0}+\varepsilon^{0}_{1}\gamma^{1}+\varepsilon^{0}_{2}\gamma^{2}+\varepsilon^{0}_{3}\gamma^{3}=\gamma^{0}-\rho A^{1}\gamma^{1}-\rho A^{2}\gamma^{2}-\rho A^{3}\gamma^{3}=\gamma^{0}-\rho A^{k}\gamma^{k}\\
\Gamma^{1}_{(\varepsilon)}&=\varepsilon^{1}_{y}\gamma^{y}=\varepsilon^{1}_{0}\gamma^{0}+\varepsilon^{1}_{1}\gamma^{1}=\gamma^{1}-\rho A^{1}\gamma^{0}\\
\Gamma^{2}_{(\varepsilon)}&=\varepsilon^{2}_{y}\gamma^{y}=\varepsilon^{2}_{0}\gamma^{0}+\varepsilon^{2}_{2}\gamma^{2}=\gamma^{2}-\rho A^{2}\gamma^{0}\\
\Gamma^{3}_{(\varepsilon)}&=\varepsilon^{3}_{y}\gamma^{y}=\varepsilon^{3}_{0}\gamma^{0}+\varepsilon^{3}_{3}\gamma^{3}=\gamma^{3}-\rho A^{3}\gamma^{0}
\end{aligned}\,. \tag{19.2}$$

This may be consolidated into:

$$\Gamma^{\sigma}_{(\varepsilon)}=\begin{pmatrix}\Gamma^{0}_{(\varepsilon)} & \Gamma^{k}{}_{(\varepsilon)}\end{pmatrix}=\begin{pmatrix}\gamma^{0}-\rho A^{k}\gamma^{k} & \gamma^{k}-\rho A^{k}\gamma^{0}\end{pmatrix}=\begin{pmatrix}\Gamma^{0}_{(\varepsilon)} & \boldsymbol{\Gamma}_{(\varepsilon)}\end{pmatrix}=\begin{pmatrix}\gamma^{0}-\rho\mathbf{A}\cdot\boldsymbol{\gamma} & \boldsymbol{\gamma}-\rho\mathbf{A}\gamma^{0}\end{pmatrix}. \tag{19.3}$$

Let us also write down the γ^{μ} in the in the Dirac representation, which are:

$$\gamma^{0}=\begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix};\quad \gamma^{k}=\boldsymbol{\gamma}=\begin{pmatrix}0 & \sigma^{k}\\ -\sigma^{k} & 0\end{pmatrix}=\begin{pmatrix}0 & \boldsymbol{\sigma}\\ -\boldsymbol{\sigma} & 0\end{pmatrix} \tag{19.4}$$

It will be helpful at various times in the upcoming calculations to make use of the Dirac relation $\gamma^j b^j \gamma^k c^k = (\boldsymbol{\gamma}\cdot\mathbf{b})(\boldsymbol{\gamma}\cdot\mathbf{c}) = -I_{2\times 2}(\boldsymbol{\sigma}\cdot\mathbf{b})(\boldsymbol{\sigma}\cdot\mathbf{c})$ easily apparent from (19.4) , where *I* is a unit matrix. Furthermore, the Pauli matrices satisfy the identity $(\boldsymbol{\sigma}\cdot\mathbf{b})(\boldsymbol{\sigma}\cdot\mathbf{c}) = \mathbf{b}\cdot\mathbf{c} + i\boldsymbol{\sigma}\cdot(\mathbf{b}\times\mathbf{c})$ so that this Dirac relation becomes $(\boldsymbol{\gamma}\cdot\mathbf{b})(\boldsymbol{\gamma}\cdot\mathbf{c}) = -I_{2\times 2}\left(I_{2\times 2}\mathbf{b}\cdot\mathbf{c} + i\boldsymbol{\sigma}\cdot(\mathbf{b}\times\mathbf{c})\right)$. In the special case where $\mathbf{b}=\mathbf{c}$ are the same vector and *not* sums of independent vectors (so that $\mathbf{b}\times\mathbf{b}=0$), this means that $(\boldsymbol{\gamma}\cdot\mathbf{b})^2(\boldsymbol{\gamma}\cdot\mathbf{b}) = -I_{4\times 4}\mathbf{b}^2$. And in the further special case where $\mathbf{b}=\mathbf{A}$ and $\mathbf{A}$ is the photon three-vector potential for which $\mathbf{A}^2=0$ via (14.8), we find that $(\boldsymbol{\gamma}\cdot\mathbf{A})^2=0$.

With this in mind we use (19.3) to calculate the spin connection term in (19.1), to find:

$$\begin{aligned}
&\tfrac{1}{4}i\hbar c\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)} = \tfrac{1}{4}i\hbar c\eta_{\mu\nu}\Gamma^\mu_{(e)}\partial_\sigma\Gamma^\nu_{(e)} = \tfrac{1}{4}i\hbar c\left(\Gamma^0_{(e)}\partial_\sigma\Gamma^0_{(e)} - \Gamma^k_{(e)}\partial_\sigma\Gamma^k_{(e)}\right)\\
&= \tfrac{1}{4}i\hbar c\left(\left(\gamma^0 - \rho A^k\gamma^k\right)\partial_\sigma\left(\gamma^0 - \rho A^l\gamma^l\right) - \left(\gamma^k - \rho A^k\gamma^0\right)\partial_\sigma\left(\gamma^k - \rho A^k\gamma^0\right)\right).\\
&= \tfrac{1}{4}i\hbar c\left(-\gamma^0\partial_\sigma\rho A^l\gamma^l + \gamma^k\partial_\sigma\rho A^k\gamma^0\right) = \tfrac{1}{2}\gamma^k\gamma^0 cq_\sigma\rho A^k
\end{aligned}\tag{19.5}$$

In the above, we have reduced using $\partial_\sigma\gamma^\mu=0$, $\mathbf{A}^2=0$, $(\boldsymbol{\gamma}\cdot\mathbf{A})^2=0$, $i\hbar\partial_\sigma A^\mu = q_\sigma A^\mu$ from (15.4), and $\gamma^k\gamma^0 = -\gamma^0\gamma^k$.

The hyper-canonical Dirac equation (19.1) contains (19.5) in the form of $\left(\Gamma^\sigma_{(e)} + \rho A^\sigma\right)\tfrac{1}{4}i\hbar c\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}$. However, $\rho A^\sigma\tfrac{1}{4}i\hbar c\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)} = \rho A^\sigma\tfrac{1}{2}\gamma^k\gamma^0 cq_\sigma\rho A^k = 0$ using (19.5) combined with $q_\sigma A^\sigma=0$ in (15.5). Therefore, from (19.5) and again using (19.3) we may calculate:

$$\begin{aligned}
&\left(\Gamma^\sigma_{(e)} + \rho A^\sigma\right)\tfrac{1}{4}i\hbar c\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)} = \tfrac{1}{2}\Gamma^\sigma_{(e)}\gamma^k\gamma^0 cq_\sigma\rho A^k = \tfrac{1}{2}\eta_{\mu\nu}\Gamma^\mu_{(e)}\gamma^k\gamma^0 cq^\nu\rho A^k\\
&= \tfrac{1}{2}\Gamma^0_{(e)}\gamma^k\gamma^0 cq^0\rho A^k - \tfrac{1}{2}\Gamma^j_{(e)}\gamma^k\gamma^0 cq^j\rho A^k\\
&= \tfrac{1}{2}\left(\gamma^0 - \rho A^j\gamma^j\right)\gamma^k\gamma^0 cq^0\rho A^k - \tfrac{1}{2}\left(\gamma^j - \rho A^j\gamma^0\right)\gamma^k\gamma^0 cq^j\rho A^k\\
&= \tfrac{1}{2}\left(-\gamma^k cq^0\rho A^k - \gamma^0\gamma^j\gamma^k cq^j\rho A^k\right)\\
&= -\tfrac{1}{2}\left(cq^0(\boldsymbol{\gamma}\cdot\rho\mathbf{A}) + \gamma^0(\boldsymbol{\gamma}\cdot c\mathbf{q})(\boldsymbol{\gamma}\cdot\rho\mathbf{A})\right)
\end{aligned}\quad .\tag{19.6}$$

To reduce the above, we use $\gamma^k\gamma^0 = -\gamma^0\gamma^k$, $\gamma^0\gamma^0=1$, $(\boldsymbol{\gamma}\cdot\mathbf{A})^2=0$, and $\mathbf{q}\cdot\mathbf{A}=0$ from (15.5).

We therefore see that the spin connection contributes two additional terms that would be absent from (19.1) if it did not include the spin connection and had been left as is at (17.5). Given that the photon energy momentum vector $cq^\mu = (h\nu, c\mathbf{q})$ where $cq^0 = h\nu$ is the photon energy, the first spin term $-\tfrac{1}{2}cq^0(\boldsymbol{\gamma}\cdot\rho\mathbf{A}) = -\tfrac{1}{2}h\nu(\boldsymbol{\gamma}\cdot\rho\mathbf{A})$ places the photon energy directly into the momentum space Dirac equation. As to the second term, we may use the identity $(\gamma\cdot\mathbf{b})(\gamma\cdot\mathbf{c})=-(\mathbf{b}\cdot\mathbf{c}+i\boldsymbol{\sigma}\cdot(\mathbf{b}\times\mathbf{c}))$ (with I matrices implicit) as well as $i\hbar\partial_\sigma A^\mu = q_\sigma A^\mu$ from (15.4) and $\mathbf{q}\cdot\mathbf{A}=0$ from (15.5) and the field strength $F^{ij}=\partial^i A^j-\partial^j A^i$ to deduce that:

$$\begin{aligned}
&-\tfrac{1}{2}\gamma^0(\gamma\cdot c\mathbf{q})(\gamma\cdot\rho\mathbf{A})=\tfrac{1}{2}i\gamma^0\boldsymbol{\sigma}\cdot(c\mathbf{q}\times\rho\mathbf{A})=\tfrac{1}{2}i\gamma^0\varepsilon^{ijk}\sigma^i cq^j\times\rho A^k=-\tfrac{1}{2}\hbar\gamma^0\varepsilon^{ijk}\sigma^i c\partial^j\rho A^k\\
&=-\tfrac{1}{2}\gamma^0\left(\hbar\sigma^1\left(c\partial^2\rho A^3-c\partial^3\rho A^2\right)+\hbar\sigma^2\left(c\partial^3\rho A^1-c\partial^1\rho A^3\right)+\hbar\sigma^3\left(c\partial^1\rho A^2-c\partial^2\rho A^1\right)\right)\quad.\\
&=-\tfrac{1}{2}\gamma^0\hbar c\rho\left(\sigma^1F^{23}+\sigma^2F^{31}+\sigma^3F^{12}\right)=\tfrac{1}{2}\gamma^0\hbar c\rho\left(\sigma^1B^1+\sigma^2B^2+\sigma^3B^3\right)=\tfrac{1}{2}\gamma^0\hbar c\rho\boldsymbol{\sigma}\cdot\mathbf{B}
\end{aligned}\tag{19.7}$$

The above also embeds $i\boldsymbol{\sigma}\cdot(\mathbf{q}\times\mathbf{A})=\hbar\boldsymbol{\sigma}\cdot(\nabla\times\mathbf{A})=\hbar\boldsymbol{\sigma}\cdot\mathbf{B}$, i.e., $\mathbf{B}=\nabla\times\mathbf{A}$ or $i(\mathbf{q}\times\mathbf{A})=\hbar\mathbf{B}$. Now, we now replace the substitute variable $\rho=q/mc^2$ with the actual charge $q=-e$ of the charged leptons. Then we define a triplet of spin matrices $\mathbf{S}\equiv\tfrac{1}{2}\hbar\boldsymbol{\sigma}$ and the Dirac g-factor $g_D=2$ in the usual way, so that

$$\boldsymbol{\mu}_D=-\frac{\hbar e}{2mc}\boldsymbol{\sigma}=-\mu_B\boldsymbol{\sigma}=-2\frac{e}{2mc}\mathbf{S}=-g_D\frac{e}{2mc}\mathbf{S}=-\frac{g_D}{2}\frac{e\hbar}{2mc}\boldsymbol{\sigma}=-\frac{g_D}{2}\mu_B\boldsymbol{\sigma}\tag{19.8}$$

is the $g_D=2$ Dirac g-factor and $\mu_B=\hbar e/2mc$ is the Bohr magneton. Now (19.7) becomes:

$$-\tfrac{1}{2}\gamma^0(\gamma\cdot c\mathbf{q})(\gamma\cdot\rho\mathbf{A})=\tfrac{1}{2}\gamma^0\hbar c\rho\boldsymbol{\sigma}\cdot\mathbf{B}=-\gamma^0\frac{\hbar ce}{2mc^2}\boldsymbol{\sigma}\cdot\mathbf{B}=-\gamma^0\mu_B\boldsymbol{\sigma}\cdot\mathbf{B}\,.\tag{19.9}$$

We then use all of this in (19.6) to finally obtain:

$$\left(\Gamma^\sigma_{(e)}+\rho A^\sigma\right)\tfrac{1}{4}i\hbar c\Gamma_{(e)\nu}\partial_\sigma\Gamma^\nu_{(e)}=-\gamma^0\mu_B\boldsymbol{\sigma}\cdot\mathbf{B}-\tfrac{1}{2}h\nu(\gamma\cdot\rho\mathbf{A})=\gamma^\mu\Sigma_\mu=\eta_{\mu\nu}\gamma^\mu\Sigma^\nu\,,\tag{19.10}$$

where, using $\rho=q/mc^2=-e/mc^2$ and $\mu_B=\hbar e/2mc$ to write (15.11) as $\tfrac{1}{2}h\nu\rho\mathbf{A}=i\mu_B\mathbf{E}$, we define a spin-connection four-vector:

$$\Sigma^\mu\equiv\begin{pmatrix}-\mu_B\boldsymbol{\sigma}\cdot\mathbf{B} & \tfrac{1}{2}h\nu\rho\mathbf{A}\end{pmatrix}=\mu_B\begin{pmatrix}-\boldsymbol{\sigma}\cdot\mathbf{B} & i\mathbf{E}\end{pmatrix}.\tag{19.11}$$

It will become of importance that the magnetic moment $-\mu_B\boldsymbol{\sigma}\cdot\mathbf{B}$ and the electric field $i\mu_B\mathbf{E}$ transform as the time and space components of a four-vector. Because of $\boldsymbol{\sigma}\cdot\mathbf{B}$, this implicitly contains a 2x2 matrix, and $\gamma^\mu\Sigma_\mu=\gamma^\mu I_{2\times2}\Sigma_\mu$.

Returning to (19.1) and inserting (19.10) now produces the rather simplified:

$$0=\left(\left(\Gamma^\sigma_{(e)}+\rho A^\sigma\right)cp_\sigma+\gamma^\sigma\Sigma_\sigma-mc^2\right)|U_0\rangle\,.\tag{19.12}$$

As the final step prior extracting the Hamiltonian, we raise an index using $g_{\mu\nu}=\eta_{\mu\nu}$ and substitute (19.3) into the above. We also reduce using $A^0=0$ from (14.8) and insert the fermion energy-momentum $cp^\mu=(E,c\mathbf{p})$ and the spin connection vector (19.11), so the above becomes:

$$\begin{aligned}0&=\left(\Gamma^0_{(e)}cp^0-\Gamma^k_{(e)}cp^k-\rho A^k cp^k+\gamma^0\Sigma^0-\gamma^k\Sigma^k-mc^2\right)|U_0\rangle\\&=\left(\gamma^0 cp^0-\gamma^k cp^k-\gamma^k\rho A^k cp^0+\left(\gamma^0-1\right)\rho A^k cp^k+\gamma^0\Sigma^0-\gamma^k\Sigma^k-mc^2\right)|U_0\rangle\\&=\left(\gamma^0\left(cp^0+\Sigma^0\right)-mc^2-\gamma^k\left(cp^k+\rho A^k cp^0+\Sigma^k\right)+\left(\gamma^0-1\right)\rho A^k cp^k\right)|U_0\rangle\\&=\left(\gamma^0\left(E+\rho\mathbf{A}\cdot c\mathbf{p}-\mu_B\boldsymbol{\sigma}\cdot\mathbf{B}\right)-\boldsymbol{\gamma}\cdot\left(c\mathbf{p}+E\rho\mathbf{A}+i\mu_B\mathbf{E}\right)-\left(mc^2+\rho\mathbf{A}\cdot c\mathbf{p}\right)\right)|U_0\rangle\end{aligned}\quad . \tag{19.13}$$

The term $c\mathbf{p}+\rho\mathbf{A}E$ within the above reveals why we chose to use a minus sign rather than a plus sign back at (14.5) though either choice seemed permissible: Restoring $\rho=q/mc^2$, this term becomes $c\mathbf{p}+\rho\mathbf{A}E=c\mathbf{p}+\left(E/mc^2\right)q\mathbf{A}$. Now, from prior to (1.4), the canonical momentum $\pi^\mu=p^\mu+qA^\mu/c$ has space components $c\boldsymbol{\pi}=c\mathbf{p}+q\mathbf{A}$. So in the limiting case where $E/mc^2=1$ we have $c\mathbf{p}+\rho\mathbf{A}E=c\mathbf{p}+q\mathbf{A}=c\boldsymbol{\pi}$.

So, to further simplify (19.13) and highlight the canonical momentum $c\boldsymbol{\pi}=c\mathbf{p}+q\mathbf{A}$, for the terms having a scaler product with $\boldsymbol{\gamma}$ we define a "hyper-momentum" vector:

$$c\mathbf{\Pi}=c\Pi^k\equiv c\mathbf{p}+E\rho\mathbf{A}+i\mu_B\mathbf{E}=cp+E\rho A^k+i\mu_B E^k\,. \tag{19.14}$$

In the limit $E\to mc^2$ and with $\mathbf{E}=0$ this reduces $\mathbf{\Pi}\to\boldsymbol{\pi}$ to the U(1) canonical momentum $\boldsymbol{\pi}$. Additionally, for the set of terms multiplied by γ^0 in (19.13) we define a "hyper-energy":

$$\mathrm{E}_{(2\times2)}\equiv E+\rho\mathbf{A}\cdot c\mathbf{p}-\mu_B\boldsymbol{\sigma}\cdot\mathbf{B} \tag{19.15}$$

which because of $\boldsymbol{\sigma}\cdot\mathbf{B}$ is also a 2x2 matrix. And, for the terms which have no γ^0 at all, we define a "hyper-rest energy":

$$\mathrm{M}c^2\equiv mc^2+\rho\mathbf{A}\cdot c\mathbf{p}\,. \tag{19.16}$$

It will be appreciated from (19.13) that (19.15) and (19.16) transform respectively as the time and space components of a four-vector in spacetime. Therefore, we additionally define a "hyper-energy-momentum" vector and also employ (19.11) as follows:

$$\begin{aligned}&c\mathrm{P}^\mu\equiv\left(\mathrm{E}\quad c\Pi_x\quad c\Pi_y\quad c\Pi_z\right)=\left(\mathrm{E}\quad c\mathbf{\Pi}\right)\\&\equiv\left(E+\rho\mathbf{A}\cdot c\mathbf{p}-\mu_B\boldsymbol{\sigma}\cdot\mathbf{B}\quad c\mathbf{p}+E\rho\mathbf{A}+i\mu_B\mathbf{E}\right)=\left(E+\rho\mathbf{A}\cdot c\mathbf{p}\quad c\mathbf{p}+E\rho\mathbf{A}\right)+\Sigma^\mu\end{aligned}\quad . \tag{19.17}$$

Then making use of all of (19.14) through (19.17) with $g_{\mu\nu} = \eta_{\mu\nu}$, we compact (19.13) to:

$$\begin{aligned} 0 &= \left(\gamma^0\left(E+\rho\mathbf{A}\cdot c\mathbf{p}-\mu_B\boldsymbol{\sigma}\cdot\mathbf{B}\right)-\boldsymbol{\gamma}\cdot\left(c\mathbf{p}+\rho\mathbf{A}\left(E+\tfrac{1}{2}h\nu\right)\right)-\left(mc^2+\rho\mathbf{A}\cdot c\mathbf{p}\right)\right)\left|U_0\right\rangle \\ &= \left(\gamma^0\mathrm{E}-\boldsymbol{\gamma}\cdot c\boldsymbol{\Pi}-\mathrm{M}c^2\right)\left|U_0\right\rangle = \left(\gamma^\sigma c\mathrm{P}_\sigma-\mathrm{M}c^2\right)\left|U_0\right\rangle \end{aligned}. \tag{19.18}$$

This final result, $0=\left(\gamma^\sigma c\mathrm{P}_\sigma-\mathrm{M}c^2\right)\left|U_0\right\rangle$, has exactly the same form as the free-particle Dirac equation $\left(\gamma^\sigma cp_\sigma-mc^2\right)u_0=0$, with the "hyper-canonical" substitutions $p^\sigma \mapsto \mathrm{P}^\sigma$, $m\mapsto\mathrm{M}$ and $u_0\mapsto\left|U_0\right\rangle$. With this, we are ready to calculate the Dirac "hyper-canonical Hamiltonian."

To Be Continued…

Appendix A: Review of Derivation of the Gravitational Geodesic Motion from a Variation

To derive (1.3) from (1.2) we first apply δ to the (1.2) integrand and then use (1.1) to clear the denominator but keep the factor .5 arising from differentiating the square root, yielding:

$$0=\delta\int_A^B d\tau=\frac{1}{2}\int_A^B d\tau\delta\left(g_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}\right). \tag{A.1}$$

The variation symbol δ commutes with the derivative symbol d such that $\delta d = d\delta$, and operates in the same way as d and so distributes via the product rule according to:

$$0=\delta\int_A^B d\tau=\frac{1}{2}\int_A^B d\tau\left(\delta g_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+g_{\mu\nu}\frac{d\delta x^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+g_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{d\delta x^\nu}{cd\tau}\right). \tag{A.2}$$

Now, one can use the chain rule in the small variation $\delta\to\partial$ limit to show that $\delta g_{\mu\nu}=\delta x^\alpha\partial_\alpha g_{\mu\nu}$. Indeed, the generic calculation for any field ϕ (taking $\delta\cong\partial$), is:

$$\delta x^\alpha\partial_\alpha\phi=\delta x^\alpha\frac{\partial\phi}{\partial x^\alpha}\cong\partial x^\alpha\frac{\delta\phi}{\partial x^\alpha}=\frac{\partial x^\alpha}{\partial x^\alpha}\delta\phi=\delta\phi. \tag{A.3}$$

Additionally, we may use the symmetry of $g_{\mu\nu}$ to combine the second and third term inside the parenthesis in (A.2). Thus, (A.2) becomes:

$$0=\delta\int_A^B d\tau=\frac{1}{2}\int_A^B d\tau\left(\delta x^\alpha\partial_\alpha g_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}+2g_{\mu\nu}\frac{d\delta x^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}\right). \tag{A.4}$$

The next step is to integrate by parts. From the product rule, we may obtain:

$$\frac{d}{cd\tau}\left(\delta x^{\mu} g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\right)=g_{\mu\nu}\frac{d\delta x^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}+\delta x^{\mu}\frac{d}{cd\tau}\left(g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\right). \tag{A.5}$$

It will be recognized that the first term after the equality in (A.5) is the same as the final term in (A.4) up to the factor of 2. So we use (A.5) in (A.4) to write:

$$0=\delta\int_{A}^{B}d\tau=\frac{1}{2}\int_{A}^{B}d\tau\left(\delta x^{\alpha}\partial_{\alpha}g_{\mu\nu}\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}+2\frac{d}{cd\tau}\left(\delta x^{\mu}g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\right)-2\delta x^{\mu}\frac{d}{cd\tau}\left(g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\right)\right). \tag{A.6}$$

The middle term in the above, which is a total integral, is equal to zero because of the boundary conditions on the variation. Specifically, this middle term is:

$$\int_{A}^{B}d\tau\frac{d}{cd\tau}\left(\delta x^{\mu}g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\right)=\frac{1}{c}\int_{A}^{B}d\left(\delta x^{\mu}g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\right)=\frac{1}{c}g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\delta x^{\mu}\Big|_{A}^{B}=0. \tag{A.7}$$

This definite integral is zero because the two worldlines intersect at the boundary events *A* and *B* but have a slight variational difference between *A* and *B* otherwise, so that $\delta x^{\sigma}(A)=\delta x^{\sigma}(B)=0$ while $\delta x^{\sigma}\neq 0$ elsewhere. Therefore we may zero out the middle term and rewrite (A.6) as:

$$0=\delta\int_{A}^{B}d\tau=\frac{1}{2}\int_{A}^{B}d\tau\left(\delta x^{\alpha}\partial_{\alpha}g_{\mu\nu}\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}-2\delta x^{\mu}\frac{d}{cd\tau}\left(g_{\mu\nu}\frac{dx^{\nu}}{cd\tau}\right)\right). \tag{A.8}$$

Next, in the final term above, we distribute the $d/cd\tau$ via the product rule to each of $g_{\mu\nu}$ and $dx^{\nu}/cd\tau$, so that this becomes:

$$0=\delta\int_{A}^{B}d\tau=\frac{1}{2}\int_{A}^{B}d\tau\left(\delta x^{\alpha}\partial_{\alpha}g_{\mu\nu}\frac{dx^{\mu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}-2\delta x^{\mu}\frac{dg_{\mu\nu}}{cd\tau}\frac{dx^{\nu}}{cd\tau}-2\delta x^{\mu}g_{\mu\nu}\frac{d^{2}x^{\nu}}{c^{2}d\tau^{2}}\right). \tag{A.9}$$

For the first time, we see an acceleration $d^{2}x^{\nu}/d\tau^{2}$. It is then straightforward to apply the chain rule to deduce $dg_{\mu\nu}/cd\tau=\partial_{\alpha}g_{\mu\nu}\left(dx^{\alpha}/cd\tau\right)$, which is a special case of the generic relation for any field ϕ given by:

$$\frac{d\phi}{cd\tau}=\frac{\partial\phi}{\partial x^{\alpha}}\frac{dx^{\alpha}}{cd\tau}=\partial_{\alpha}\phi\frac{dx^{\alpha}}{cd\tau}. \tag{A.10}$$

As a result, (A.9) now becomes:

$$0=\delta\int_A^B d\tau=\frac{1}{2}\int_A^B d\tau\left(\delta x^\alpha\partial_\alpha g_{\mu\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-2\delta x^\mu\partial_\alpha g_{\mu\nu}\frac{dx^\alpha}{cd\tau}\frac{dx^\nu}{cd\tau}-2\delta x^\mu g_{\mu\nu}\frac{d^2x^\nu}{c^2d\tau^2}\right). \quad (A.11)$$

At this point we have a coordinate variation in front of all terms, but the indexes are not the same. So we need to re-index to be able to factor out the same coordinate variation from all terms. We thus rename the summed indexes $\mu\leftrightarrow\alpha$ in the second and third terms and factor out the resulting δx^α from all three terms. And we also use the symmetry of $g_{\mu\nu}$ to split the middle term into two, then cycle all indexes, then factor out all the terms containing derivatives of $g_{\mu\nu}$. The result of all this re-indexing, also moving the outside coefficient of ½ into the integrand, is:

$$0=\delta\int_A^B d\tau=\int_A^B \delta x^\alpha d\tau\left(\frac{1}{2}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-g_{\alpha\nu}\frac{d^2x^\nu}{c^2d\tau^2}\right). \quad (A.12)$$

Now we are ready for the final steps. Because the worldlines under consideration are for material particles, the proper time $d\tau\neq 0$. Likewise, while $\delta x^\sigma(A)=\delta x^\sigma(B)=0$ at the boundaries, between these boundaries where the variation occurs, $\delta x^\sigma\neq 0$. Therefore, for the overall expression (A.12) to be equal to zero, the expression inside the large parenthesis must be zero. Consequently:

$$0=\frac{1}{2}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}-g_{\alpha\nu}\frac{d^2x^\nu}{c^2d\tau^2}. \quad (A.13)$$

From here, we multiply through by $g^{\beta\alpha}$, apply $-\Gamma^\beta{}_{\mu\nu}=\frac{1}{2}g^{\beta\alpha}\left(\partial_\alpha g_{\mu\nu}-\partial_\mu g_{\nu\alpha}-\partial_\nu g_{\alpha\mu}\right)$ for the Christoffel symbols, flip the sign, and segregate the acceleration term to obtain the final result:

$$\frac{d^2x^\beta}{d\tau^2}=-\Gamma^\beta{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}. \quad (A.14)$$

Appendix B: Review of Derivation of Time Dilations in Special and General Relativity

To derive time dilations in the Special Theory of Relativity, we begin with the flat spacetime metric $c^2d\tau^2=\eta_{\mu\nu}dx^\mu dx^\nu$ which using a squared velocity $v^2=\left(dx^k/dt\right)\left(dx^k/dt\right)$ is easily restructured with the chain rule into $1=\left(dt/d\tau\right)^2\left(1-v^2/c^2\right)$, then into the familiar $\gamma_v\equiv dt/d\tau=1/\sqrt{1-v^2/c^2}$, with γ_v defined as the motion-induced time dilation. In the General Theory we start with the line element $c^2d\tau^2=g_{\mu\nu}dx^\mu dx^\nu$ in which the metric tensor $g_{\mu\nu}$ contains the gravitational field. We isolate gravitation from motion by setting $dx^k=0$ to place the clock at rest in the gravitational field. This is just as we did to place the test charge at rest in the electromagnetic potential to reach (4.1) and (5.3) above, isolating electromagnetic effects from motion effects. The line element then becomes $c^2 d\tau^2 = g_{00} dx^0 dx^0 = g_{00} c^2 dt^2$ which rearranges to $dt^2 / d\tau^2 = 1/g_{00}$. We then take the positive square root $\gamma_g \equiv dt/d\tau = 1/\sqrt{g_{00}}$ so that this approaches 1 in the flat spacetime $g_{00} = \eta_{00} = 1$ limit, with γ_g defined as the gravitationally-induced time dilation. For motion dt is the coordinate time element in the rest frame of the observer and $d\tau$ is the proper time element ticked off by a g-clock in motion relative to the observer. For gravitation dt is the coordinate time element in the frame of an observer outside the gravitational field and $d\tau$ is the proper time element ticked off by a g-clock inside the gravitational field.

References

[1] H. Minkowski, Raum und Zeit, Physikalische Zeitschrift 10, 104-111 (1909)
[2] A. Einstein, *On the Electrodynamics of Moving Bodies*, Annalen der Physik 17 (10): 891–921 (1905)
[3] A. Einstein, *The Foundation of the Generalised Theory of Relativity*, Annalen der Physik 354 (7), 769-822 (1916)
[4] Dirac, P. A. M., *The Quantum Theory of the Electron*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 117 (778): 610 (1928)
[5] J. A. Wheeler, Geometrodynamics, New York: Academic Press (1963)
[6] H. Weyl, *Gravitation and Electricity*, Sitzungsber. Preuss. Akad.Wiss., 465-480. (1918).
[7] H. Weyl, *Space-Time-Matter* (1918)
[8] H. Weyl, *Electron und Gravitation*, Zeit. f. Physik, 56, 330 (1929)
[9] C. N. Yang and R. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Physical Review. **96** (1): 191–195. (1954)
[10] L. H. Ryder, *Quantum Field Theory*, Second edition, Cambridge (1996)
[11] H. C. Ohanian, *Gravitation and Spacetime*, Norton (1976)
[12] A. Einstein, *Does the inertia of a body depend upon its energy-content?*, Annalen der Physik, **18**:639 (1905)
[13] F. Halzen and A.D. Martin, *Quarks and Leptons: An introductory Course in Modern Particle Physics*, Wiley (1984)