Jay R. Yablon, November 7, 2018

Theory of Fermion Masses, Mixing Angles, Lagrangian Potentials and Beta Decays, based on Higgs Bosons arising from the Scaler Fields of a Kaluza Klein Theory given Five-Dimensional General Covariance by Dirac's Quantum Theory of the Electron

Jay R. Yablon, November 7, 2018

Abstract: Why the twelve elementary fermions have the masses they have, (and what the neutrino masses actually are) is one of the deepest unsolved mysteries of modern physics. We crack this puzzle using a theory of fermion masses which succeeds in reparameterizing all twelve fermion masses in terms of other known parameters for which the theoretical interconnection to these masses have not heretofore been understood. The first step is to "repair" long-recognized perplexities of Kaluza-Klein theory using Dirac's quantum theory of the electron to enforce general covariance across all five dimensions. One consequence of this is the emergence of a modified Dirac equation for fermions which naturally contains the Kaluza-Klein scaler. After establishing a connection between this Kaluza-Klein scaler and the standard model Higgs scaler, we use the latter to connect the known masses of all the quarks and charged leptons to the CKM and PMNS mixing angles and several other parameters which have heretofore not been theoretically connected to these masses. Then, after using the Newton gravitational constant and the Fermi vacuum to establish a sum of neutrino masses in the exact range expected from experiments, it also becomes possible to predict the rest masses of the three neutrinos. Also predicted are the existence and rest mass of a second leptonic Higgs boson, and tighter values for several other known parameters including the mass of the established Higgs boson. Also uncovered is a new, deep role for the cosmological neutrino background (CvB) and the Higgs boson in triggering and facilitating weak interaction beta decay events.

Contents

Preface, and Guide for Efficient Reading

Preface

This manuscript is two papers in one. One is about Kaluza-Klein Theory. The other is about particles physics and the rest masses and weak beta decays of the elementary fermions. This began as an effort to "repair" five-dimensional Kaluza-Klein theory in advance of its 2019 centenary, by using Dirac's Quantum Theory of the Electron as the basis for requiring Kaluza-Klein theory to be generally-covariant across all five of its dimensions. Unexpectedly, this turned into a theory through which it became possible to explain all twelve of the observed elementary fermion masses in relation to other heretofore independent parameters including the CKM and PMNS mixing angles. Of course, Kaluza-Klein theory started in 1919 as a classical theory to unify Maxwell's electrodynamics with general relativistic gravitation, before we even had modern gauge theory or Dirac theory or much of modern quantum theory. Because one would not normally expect to be talking about Kaluza-Klein theory and the elementary fermion masses of modern particle physics in the same breath – much less claim that a detailed study of Kaluza-Klein theory can lead to a deep understanding of these fermion masses – it is important to overview the trail that lead from one to the other, and why it is that this is all best-presented in one complete paper.

As this manuscript will demonstrate, if we start with the Klauza-Klein metric tensor denoted G_{MN} and then follow Dirac by finding set of Dirac-type matrices Γ_{M}, $\text{M} = 0,1,2,3,5$ defined such that the anticommutator $\frac{1}{2}\{\Gamma_{\text{M}},\Gamma_{\text{N}}\} \equiv G_{\text{MN}}$, then not only are the most perplexing century-old problems of Kaluza-Klein theory repaired, but the Γ_{M} matrices so-defined can be used to formulate a modified Dirac equation $\left(i\hbar c\Gamma^{\text{M}}\partial_{\text{M}} - mc^2\right)\Psi = 0$ in five dimensions, where Ψ is a fermion wavefunction. Of course, at its most fundamental level, Dirac's equation governs the behavior of fermions, including the six quarks and six leptons which we presently understand to be the fundamental constituents of matter. So, it is natural to inquire whether this modified Dirac equation can go so far as to help explain the observed pattern of fermion masses.

In the modern era, it is well-understood that Higgs field and especially the scalar Higgs bosons are at the heart of how all massive particles acquire their observed rest masses without a violation of gauge symmetry. This mechanism is explicitly understood and has been empirically confirmed for spin-1 bosons via the massive *W* and *Z* bosons of electroweak theory, and been further-established by the empirical observation of the Higgs scalar and its mass in the vicinity of 125 GeV. For spin-½ fermions, it is generally assumed that the Higgs scalars are also the mainspring of gauge-invariant mass acquisition, but the specifics of how this occurs are not yet well-understood.

Where Kaluza-Klein theory despite its very-early genesis is perhaps most prescient, is that in addition to its metric tensor G_{MN} containing the purely-gravitational metric tensor $g_{\mu\nu}$ for what in the quantum world are spin-2 gravitons and the gauge potential four-vector A_μ for spin-1

photons, it also contains a scalar field ϕ for a spin-0 scalar boson. So, when we use the five-covariant Γ_{M} in a five-dimensional Dirac equation $\left(i\hbar c\Gamma^{\text{M}}\partial_{\text{M}} - mc^2\right)\Psi = 0$, a scalar field is implicitly contained in this equation. So, it is also natural to seek out a possible connection between the Kaluza-Klein scalar, and the modern-era scalar that is well-known as the Higgs boson.

Following this approach, it turns out that the Kaluza-Klein scalar can in fact be connected to the modern Higgs scalar, and that once this is done, the fermion masses naturally follow. The easiest masses to deduce are those of the top, charm and up masses for isospin-up quarks, because of the uniquely-large mass of the top quark. These three quarks reveal the pattern for understanding fermion masses. This pattern is confirmed by being successfully extended to the bottom, strange and down quark masses. The lepton masses do not follow as simply as do the quark masses. In order to fit the charge leptons to this pattern, a specific amount of "extra" energy must be added to the sum of the electron, muon and tauon masses. Initially merely a new parameter, this extra energy turns out to be directly related to the sum of the three neutrino masses, with the ratio of Newton's gravitational constant G to the Fermi constant G_F cementing the relation. Once this connection is understood, not only does the quark mass pattern become extended to the known tau lepton masses and mixing angles, but so too, it becomes possible for the first time to predict not only the sum of the neutrino masses which turns out to be $m_{\nu e} + m_{\nu\mu} + m_{\nu\tau} = 0.133 \text{ eV}/c^2$, but also to predict the individual neutrino masses. Also predicted is a second Higgs scalar for leptons, with a rest mass which, independently, turns out to be only a few MeV above the free proton and neutron rest masses.

When all of this is completed, the twelve quark and lepton masses come to be understood entirely in terms of eleven heretofore-independent parameters, specifically: the three real CKM mixing angles, the three real PMNS mixing angles, the mass of the Higgs boson, the Newton and Fermi constants, the value $\alpha\left(M_W{}^2\right)$ of the electromagnetic running coupling at a probe energy equal to the W boson mass, and the weak (Weinberg) mixing angle. Only one of the twelve elementary fermion masses (or one fermion mass sum) has to be retained as a parameter in its own right, and for this purpose we utilize the neutrino mass sum. In the process of reparameterizing the fermion masses in the foregoing fashion, we acquire a much deeper understanding of the role that Higgs bosons and Higgs fields play in weak beta-decays, and discover that leptonic beta decays are accompanied not only by Z boson masses of about 80 GeV, but also by heretofore unknown energy exchanges in the zone of 100 TeV.

The opening to be able to uncover these findings, however, originates in using Dirac theory to render Kaluza-Klein theory generally covariant in five dimensions, then tracking down how the Kaluza-Klein scalar in the five-dimensional Dirac equation becomes connected to the Higgs boson. Once that connection is established, the path is cleared to understand the theoretical basis for why the elementary fermions have the particular pattern of masses they are observed to have, and crack one of the deepest puzzles that modern physics has to offer.

Reader Guide

In its original formulation to "repair" Kaluza-Klein theory, this manuscript was about forty pages long. With the addition of the Theory of Fermion Masses, the length has more than tripled. Accordingly, consideration was given to separating Part I and Part II of the present manuscript into two companion papers. While the numeric connections used to reparametrize the fermion masses could have been separated from its physical origins and presented in this way, such a separation would have largely obscured the physical grounding of these connections in the Higgs bosons and fields. Especially, the premier role of Higgs theory as to how particle masses are acquired but also as to the underlying mechanism of weak beta decay would have been obscured with such a separation. So, it was decided to keep both parts in one manuscript and instead present this brief "guide" for efficient reading.

Most readers of scientific papers – especially lengthy papers such as this one – are not only looking for an efficient way to study a paper, but also, want to be able to decide fairly quickly whether to even devote any time at all to studying a paper. And this decision is based on the reader's sense about whether the paper contains sound, new science. For a theoretical physics paper, having the theory presented make convincing points of contact with empirical data – especially previously-unexplained data such as the elementary fermion masses – is very important, and is likely a primary screening criterion for most serious readers. Therefore, while the reader can certainly study this manuscript from start to finish in a linear way, it is suggested that the reader might instead wish to dive directly into the connections between fermion mass and other parameters such as the CKM and PMNS mixing angles, convince him or herself that these connections are properly-established and not previously-known, and then work outward to assimilate the surrounding theory which both leads to and further supports these connections.

In the event the reader decides to adopt this suggestion, the place to start is in section 14, for the up, charm and top quarks. Recognizing that the Fermi vev $v \cong 246.22\,\text{GeV}$ and that the energy $\frac{1}{\sqrt{2}}v \cong 174.10\,\text{GeV}$ cut by a $\sqrt{2}$ factor appears widely in Higgs field theory, the reader should first be convinced that coupling and mass sums (12.2) and (12.3) are indeed a true empirical relations within experimental error bars, and that (12.4) is therefore a warranted refinement for the top quark mass. The reader should next review the bi-unitary mass matrix transformations (12.7) to (12.9) and become convinced that the connections in (12.12) between two of the mass mixing angles and two of the three real CKM angles are also true within experimental errors.

If the reader clears section 14, he or she should next review section 15 for the down, strange and bottom masses. The reader should first become convinced that the same type of bi-unitary transformation when used with a mass sum $\frac{1}{\sqrt{2}}v_{\Downarrow} \equiv m_d c^2 + m_s c^2 + m_b c^2$ of (13.2) defining a second (local) vev $v_{\Downarrow}$ for isospin-down quarks, now produces the first of the four relations in (13.6) where the third real CKM angle is also related to a third mass mixing angle within experimental errors.

At this point, section 16 should be skimmed just long enough for the reader to become convinced that the theoretical relation $m_h c^2 \equiv \frac{1}{2}\left(v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}\right)$ (with $v_{\Uparrow} \equiv v$) precisely specifying the observed Higgs boson mass m_h in relation to the isospin-up and isospin-down vevs is also true within experimental errors. This means that $v_{\Downarrow}$ in (15.2) is not a new parameter, but rather, is a function $v_{\Downarrow}\left(v, m_h\right)$ of the Fermi vev and the Higgs mass. The reader should also be convinced that this now gives the parameter λ from the Lagrangian density – which has for decades been a theoretically-disconnected parameter – the theoretical valuation of (16.6) in terms of the two vacuum minima, namely, the fermi vev $v_{\Uparrow} \equiv v$ which is a *global* minimum, and the isospin-down quark vev $v_{\Downarrow}$ which is a *local* minimum. The reader should also be convinced that as a result of all this, the six quark masses have been effectively reparametrized in terms of five heretofore-disconnected parameters, namely, the three CKM mixing angles θ_{C21}, θ_{C23}, θ_{C31}, the Fermi vev, and the Higgs mass, as reviewed prior to (16.4). Moreover, if the reader is willing to credit, at least on a preliminary basis, that the relation $3\left(m_d - m_u\right)/\left(2\pi\right)^{1.5} = m_e$ discussed at (15.7) between the mass of the electron and the mass difference between the down and up quarks is true within known error bars and may in fact be generally true, then the electron rest mass itself becomes a sixth parameter, whereby all six quark masses are fully reparameterized according to $m_u, m_c, m_t, m_d, m_s, m_b = F\left(v, \theta_{C31}, \theta_{C23}, \theta_{C21}, m_h, m_e\right)$ in (17.1), into other previously-disconnected parameters. At this point, further reparameterization moves into the lepton masses.

Next, the reader should turn to section 18 to see how the charged lepton masses may be similarly reparameterized using bi-unitary mass matrix transformations in terms of two of the three real PMNS angles, namely θ_{P12} and θ_{P13}, but only by postulating an extra energy $\delta_{\Downarrow}$ which is added to the rest energy sum $m_\tau c^2 + m_\mu c^2 + m_e c^2$ when it is defined at (18.11). This means that we start with the masses m_τ, m_μ, m_e, supplement these with a new unknown parameter $\delta_{\Downarrow}$. But we wash this out when we find at (18.16) that this sum with the extra energy can be related to the Fermi vev via $m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_{\Downarrow} = \alpha\left(M_W{}^2\right) v_{\Uparrow}$ within experimental errors using the strength $\alpha\left(M_W{}^2\right)$ of the electromagnetic running coupling at a probe energy $Q^2 = M_W^2 c^4$ equivalent to the rest mass of the *W* boson, which boson must *always* be present at the vertex of any beta decay between a charged lepton and a neutrino. The fermi vev is already a parameter used for quarks, so we effectively reparameterize $m_\tau, m_\mu, m_e, \delta_{\Downarrow} = F\left(\theta_{P12}, \theta_{P13}, \alpha\left(M_W{}^2\right), \delta_{\Downarrow}\right)$. The new parameter $\delta_{\Downarrow}$ remains independent for now, and its study takes place when we turn to neutrino masses.

At this point the reader should review section 19 for the neutrino masses, where we seek to reparameterize the remaining set of mass/energy numbers $\left\{m_{\nu e}, m_{\nu\mu}, m_{\nu\tau}, \delta_{\Downarrow}\right\}$. The reader should first confirm that (19.1) is a correct numeric calculation, and then become convinced that the physical connection (19.2b) between the extra energy $\delta_{\Downarrow}$ and the sum of the neutrino rest

masses is warranted within experimental errors for the upper limit of the neutrino mass sum. Because the ratio of the Fermi vev to the Planck energy is at the center of (19.2b), and because the Planck energy is merely a restatement of the Newton gravitational constant *G*, this means that in a single stroke, we eliminate $\delta_{\Downarrow}$ as an unknown parameter and trade it for the known parameter *G* and as a result, introduce *gravitation* into particle physics by way of the very tiny neutrino masses. This is because, as seen in (19.2c), the extra energy $\delta_{\Downarrow}$ is directly equal to the sum $\left(m_{\nu e}+m_{\nu\mu}+m_{\nu\tau}\right)c^2$ of the neutrino rest energies, times an amplification factor $\sqrt{2M_Pc^2/v}$ which is at bottom based on the ratio of the Fermi constant $G_F=1.1663787(6)\times10^{-5}\ \text{GeV}^{-2}$ to the Newton constant $G=\ 6.708\ 61(31)\times10^{-39}\ \text{GeV}^{-2}$, in natural units. The reader should then confirm that this is used together with the known data (19.3) to predict neutrino masses at (19.4) which are well within the ranges that they are expected to be. Finally, the reader should become convinced that the connection to the neutrino masses of the final PMNS angle θ_{P23} in (19.7), using the same type of bi-unitary transformations previously employed for other masses, is correctly carried out.

If the reader is convinced that the foregoing does represent a true reparameterization of the fermion rest masses, then it should be clear the net result of all this that all twelve fermion rest masses will have been reparameterized according to (20.5), and that all told, twenty-two physics parameters which are heretofore been regarded as independent, will have been reduced down to eleven parameters, removing eleven independent unknowns from our understanding of the natural world. This alone should then provide motivation for the reader to study the balance of the paper to see how its was possible to obtain these results starting with connecting the Higgs fields of particle physics to scaler fields of a Kaluza-Klein theory given five-dimensional general covariance by Dirac's quantum theory of the electron.

1. Introduction – The Incompatibility of Kaluza-Klein and Dirac Theories

About a century ago with the 1920s approaching, much of the physics community was trying to understand the quantum reality that Planck had first uncovered almost two decades prior [1]. But with the General Theory of Relativity [2] having recently placed gravitation and the dynamical behavior of gravitating objects onto an entirely geometric and geodesic foundation (which several decades later Wheeler would dub "geometrodynamics" [3]), a few scientists were trying to scale the next logical hill, which – with weak and strong interactions not yet known – was to obtain a geometrodynamic theory of electromagnetism. Besides Einstein's own work on this which continued for the rest of his life [4], the two most notable efforts were those of Hermann Weyl [5], [6] who was just starting to develop his U(1) gauge theory in four dimensions (which turned out to be a theory of "phase" invariance [7] that still retains the original moniker "gauge"), and Kaluza [8] then Klein [9], [10] who quite successfully used a fifth dimension to geometrize the Lorentz Force motion and the Maxwell Stress-Energy tensor (see, e.g., [11] and [12]). This is a very attractive aspect of Kaluza-Klein theory, and it remains so because even today, despite almost a century of efforts to do so, U(1) gauge theory has not yet successfully been able to place the Lorentz Force dynamics and the Maxwell Stress Energy on an entirely-geometrodynamic foundation. And as will be appreciated by anyone who has studied this problem seriously, it is the

inequivalence of electrical mass (a.k.a. charge) and inertial mass which has been the prime hindrance to being able to do so.

Notwithstanding these Kaluza "miracles" of geometrizing the Lorentz Force motion and the Maxwell Stress-Energy, this fifth dimension and an associated scalar field known as the graviscalar or radion or dilaton raised its own new challenges, many of which will be reviewed here. These have been a legitimate hurdle to the widespread acceptance of Kaluza-Klein theory as a theory of what is observed in the natural world. It is important to keep this historical sequencing in mind, because Kaluza's work in particular predated what we now know to be modern gauge theory and so was the "first" geometrodynamic theory of electrodynamics. And it of course predated any substantial knowledge about the weak and strong interactions. Of special interest in this paper, Kaluza-Klein also preceded Dirac's seminal Quantum Theory of the Electron [13] which today is the foundation of how we understand fermion behavior.

Now in Kaluza-Klein theory, the metric tensor which we denote by G_{MN} and its inverse G^{MN} obtained by $G^{\text{MA}}G_{\text{AN}} = \delta^{\text{M}}{}_{\text{N}}$ are specified in five dimensions with an index $\text{M} = 0,1,2,3,5$, and may be represented in the 2x2 matrix format:

$$G_{\text{MN}} = \begin{pmatrix} g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu & \phi^2 k A_\mu \\ \phi^2 k A_\nu & \phi^2 \end{pmatrix}; \quad G^{\text{MN}} = \begin{pmatrix} g^{\mu\nu} & -A^\mu \\ -A^\nu & g_{\alpha\beta}A^\alpha A^\beta + 1/\phi^2 \end{pmatrix}. \tag{1.1}$$

In the above $g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$ transforms as a 4x4 tensor symmetric in spacetime. This is because $g_{\mu\nu} = g_{\nu\mu}$ is a symmetric tensor, and because electrodynamics is an abelian gauge theory with a commutator $\left[A_\mu, A_\nu \right] = 0$. The components $G_{\mu 5} = \phi^2 k A_\mu$ and $G_{5\nu} = \phi^2 k A_\nu$ transform as covariant (lower-indexed) vectors in spacetime. And the component $G_{55} = \phi^2$ transforms as a scalar in spacetime. If we regard ϕ to be a dimensionless scalar, then the constant *k* must have dimensions of charge/energy because the metric tensor is dimensionless and because the gauge field A_μ has dimensions of energy/charge.

It is very important to understand that when we turn off all electromagnetism by setting $A_\mu = 0$ and $\phi = 0$, G^{MN} in (1.1) becomes singular. This is indicated from the fact that in this situation $\text{diag}\left(G_{\text{MN}}\right) = \left(g_{00}, g_{11}, g_{22}, g_{33}, 0\right)$ with a determinant $\left|G_{\text{MN}}\right| = 0$, and is seen directly from the fact that $G^{55} = g_{\alpha\beta}A^\alpha A^\beta + 1/\phi^2 = 0 + \infty$. Therefore, (1.1) relies upon ϕ being non-zero to avoid the degeneracy of a metric inverse singularity when $\phi = 0$.

We also note that following identifying the Maxwell tensor in the Kaluza-Klein fields via a five-dimensional the Einstein field equation, again with ϕ taken to be dimensionless, the constant *k* is found to be:

$$\frac{k^2}{2} \equiv \frac{2G}{c^4} 4\pi\epsilon_0 = \frac{2}{c^4}\frac{G}{k_e} \quad \text{i.e.,} \quad k = \frac{2}{c^2}\sqrt{\frac{G}{k_e}}, \tag{1.2}$$

where $k_e = 1/4\pi\varepsilon_0 = \mu_0 c^2/4\pi$ is Coulomb's constant and G is Newton's gravitational constant.

Now, as noted above, Kaluza-Klein theory predated Dirac's Quantum Theory of the Electron [13]. Dirac's later theory begins with taking an operator square root of the Minkowski metric tensor $\text{diag}\left(\eta^{\mu\nu}\right) = \left(+1,-1,-1,-1\right)$ by defining ("≡") a set of four operator matrices γ^μ according to the anticommutator relation $\frac{1}{2}\left\{\gamma^\mu,\gamma^\nu\right\} = \frac{1}{2}\left\{\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu\right\} \equiv \eta^{\mu\nu}$. The lower-indexed gamma operators are likewise defined such that $\frac{1}{2}\left\{\gamma_\mu,\gamma_\nu\right\} \equiv \eta_{\mu\nu}$. To generalize to curved spacetime thus to gravitation which employs the metric tensor $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$ defined such that $g^{\mu\alpha}g_{\alpha\nu} \equiv \delta^\mu{}_\nu$ and we define a set of Γ^μ with a parallel definition $\frac{1}{2}\left\{\Gamma^\mu,\Gamma^\nu\right\} \equiv g^{\mu\nu}$. We simultaneously define a vierbein a.k.a. tetrad e_a^μ with both a superscripted Greek "spacetime / world" index and a subscripted Latin "local / Lorentz / Minkowski" index using the relation $e_a^\mu\gamma^a \equiv \Gamma^\mu$. Thus, we deduce that $g^{\mu\nu} = \frac{1}{2}\left\{\Gamma^\mu,\Gamma^\nu\right\} = \frac{1}{2}\left\{\gamma^a\gamma^b + \gamma^b\gamma^a\right\}e_a^\mu e_b^\nu = \eta^{ab}e_a^\mu e_b^\nu$. So just as the metric tensor $g^{\mu\nu}$ transforms in four-dimensional spacetime as a contravariant (upper-indexed) tensor, these Γ^μ operators likewise transform in spacetime as a contravariant four-vector.

One might presume in view of Dirac theory that the five-dimensional G_{MN} and G^{MN} in the Kaluza-Klein metric tensor (1.1) can be likewise deconstructed into square root operators defined using the anticommutator relations:

$$\tfrac{1}{2}\left\{\Gamma_{\text{M}},\Gamma_{\text{N}}\right\} = \tfrac{1}{2}\left\{\Gamma_{\text{M}}\Gamma_{\text{N}} + \Gamma_{\text{N}}\Gamma_{\text{M}}\right\} \equiv G_{\text{MN}}; \quad \tfrac{1}{2}\left\{\Gamma^{\text{M}},\Gamma^{\text{N}}\right\} = \tfrac{1}{2}\left\{\Gamma^{\text{M}}\Gamma^{\text{N}} + \Gamma^{\text{N}}\Gamma^{\text{M}}\right\} \equiv G^{\text{MN}}, \tag{1.3}$$

where Γ_{M} and Γ^{M} transform as *five-dimensional vectors* in five-dimensional spacetime. This would presumably include a five-dimensional definition $\varepsilon_A^{\text{M}}\gamma^A \equiv \Gamma^{\text{M}}$ for a tetrad ε_A^{M}, where $\text{M} = 0,1,2,3,5$ is a world index and $A = 0,1,2,3,5$ is a local index, and where γ^5 is a fifth operator matrix which may or may not be associated with Dirac's $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$, depending upon the detailed mathematical calculations which determine this γ^5.

However, as we shall now demonstrate, the Kaluza-Klein metric tensors in (1.1) *cannot* be deconstructed into Γ_{M} and Γ^{M} in the manner of (1.3) without modification to their $G_{05} = G_{50}$ and G_{55} components, and without imposing certain constraints on the gauge fields A^μ which remove two degrees of freedom and fix the gauge of these fields to that of a photon. We represent these latter constraints by $A^\mu = A_\gamma{}^\mu$, with *a subscripted γ which denotes a photon and which is not a spacetime index*. This means that in fact, in view of Dirac theory which was developed afterwards, the Kaluza-Klein metric tensors (1.3) are really *not* generally-covariant in five dimensions. Rather,

they only have a four-dimensional spacetime covariance represented in the components of $G_{\mu\nu} = g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$ and $G^{\mu\nu} = g^{\mu\nu}$, and of $G_{\mu 5} = \phi^2 k A_\mu$ and $G^{\mu 5} = -A^\mu$, which are all patched together with fifth-dimensional components with which they are not generally-covariant. Moreover, even the spacetime components of (1.1) alone are not generally covariant even in the four spacetime dimensions alone, unless the gauge symmetry of the gauge field A_μ is broken to remove two degrees of freedom and fixed to that of a photon, $A^\mu = A_\gamma{}^\mu$.

In today's era when the General Theory of Relativity [2] is now a few years past its centenary, and when at least in classical field theory general covariance is firmly-established as a required principle for the laws of nature, it would seem essential that any theory of nature which purports to operate in five dimensions that include the four dimensions of spacetime, ought to manifest general covariance *across all five dimensions*, and ought to be wholly consistent at the "operator square root" level with Dirac theory. Accordingly, it is necessary to "repair" Kaluza-Klein theory to make certain that it adheres to such five-dimensional covariance. In so doing, many of the most-nagging, century-old difficulties of Kaluza-Klein theory are immediately resolved, including those related to the scalar field in $G_{55} = \phi^2$ and the degeneracy of the metric tensor when this field is zeroed out, as well as the large-magnitude terms which arise when the scalar field has a non-zero gradient. Moreover, the fourth spacelike dimension of Kaluza-Klein is instead revealed to be a second timelike dimension. And of extreme importance, this Kaluza-Klein fifth dimension which has spent a century looking for direct observational grounding, may be tied directly to the clear observational physics built around the Dirac γ^5, and the multitude of observed chiral and pseudoscalar and axial vector particle states that are centered about this γ^5. Finally, importantly, all of this happens without sacrificing the Kaluza "miracle" of placing electrodynamics onto a geometrodynamic footing. This is what will now be demonstrated.

PART I: THE MARRIAGE BETWEEN FIVE DIMENSIONAL KALUZA-KLEIN THEORY AND DIRAC'S QUANTUM THEORY OF THE ELECTRON

2. The Kaluza-Klein Tetrad and Dirac Operators in Four Dimensional Spacetime, and the Covariant Fixing of Gauge Fields to the Photon

The first step to ensure that Kaluza-Klein theory is covariant in five dimensions using the operator deconstruction (1.3), is to obtain the four-dimensional spacetime deconstruction:

$$\tfrac{1}{2}\left\{\Gamma_\mu, \Gamma_\nu\right\} = \tfrac{1}{2}\left\{\Gamma_\mu\Gamma_\nu + \Gamma_\nu\Gamma_\mu\right\} = \tfrac{1}{2}\varepsilon_{\mu a}\varepsilon_{\nu b}\left\{\gamma^a\gamma^b + \gamma^b\gamma^a\right\} = \eta^{ab}\varepsilon_{\mu a}\varepsilon_{\nu b} \equiv G_{\mu\nu} = g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu \qquad (2.1)$$

using a four-dimensional tetrad $\varepsilon_{\mu a}$ defined by $\varepsilon_{\mu a}\gamma^a \equiv \Gamma_\mu$, where $\mu = 0,1,2,3$ is a spacetime world index raised and lowered with $G^{\mu\nu}$ and $G_{\mu\nu}$, and $a = 0,1,2,3$ is a local Lorentz / Minkowski

tangent spacetime index raised and lowered with η^{ab} and η_{ab}. To simplify calculation, we set $g_{\mu\nu}=\eta_{\mu\nu}$ thus $G_{\mu\nu}=\eta_{\mu\nu}+\phi^2k^2A_\mu A_\nu$. Later on, we will use the minimal-coupling principle to generalize back from $\eta_{\mu\nu}\mapsto g_{\mu\nu}$. In this circumstance, the spacetime is "flat" *except for* the curvature in $G_{\mu\nu}$ brought about by the electrodynamic terms $\phi^2k^2A_\mu A_\nu$. We can further simplify calculation by defining an $\varepsilon'_{\mu a}$ such that $\delta_{\mu a}+\varepsilon'_{\mu a}\equiv\varepsilon_{\mu a}$, which represents the degree to which $\varepsilon_{\mu a}$ differs from the unit matrix $\delta_{\mu a}$. We may then write the salient portion of (2.1) as:

$$\begin{aligned}&\eta^{ab}\varepsilon_{\mu a}\varepsilon_{\nu b}=\eta^{ab}\left(\delta_{\mu a}+\varepsilon'_{\mu a}\right)\left(\delta_{\nu b}+\varepsilon'_{\nu b}\right)=\eta^{ab}\delta_{\mu a}\delta_{\nu b}+\delta_{\nu b}\eta^{ab}\varepsilon'_{\mu a}+\delta_{\mu a}\eta^{ab}\varepsilon'_{\nu b}+\eta^{ab}\varepsilon'_{\mu a}\varepsilon'_{\nu b}\\&=\eta_{\mu\nu}+\eta_{a\nu}\varepsilon'^{\,a}_{\mu}+\eta_{\mu b}\varepsilon'^{\,b}_{\nu}+\eta_{ab}\varepsilon'^{\,a}_{\mu}\varepsilon'^{\,b}_{\nu}=\eta_{\mu\nu}+\phi^2k^2A_\mu A_\nu\end{aligned}\quad.\tag{2.2}$$

Note that when electrodynamics is "turned off" by setting A_μ and / or by setting $\phi=0$ this reduces to $\eta^{ab}\varepsilon_{\mu a}\varepsilon_{\nu b}=\eta_{\mu\nu}$ which is solved by the tetrad being a unit matrix, $\varepsilon_{\mu a}=\delta_{\mu a}$. Subtracting $\eta_{\mu\nu}$ from each side of (2.2) we now need to solve:

$$\eta_{a\nu}\varepsilon'^{\,a}_{\mu}+\eta_{\mu b}\varepsilon'^{\,b}_{\nu}+\eta_{ab}\varepsilon'^{\,a}_{\mu}\varepsilon'^{\,b}_{\nu}=\phi^2k^2A_\mu A_\nu\,.\tag{2.3}$$

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. Given that $\mathrm{diag}\left(\eta_{ab}\right)=\left(1,-1,-1,-1\right)$, the four $\mu=\nu$ "diagonal" equations in (2.3) produce the relations:

$$\begin{aligned}&\eta_{a0}\varepsilon'^{\,a}_{0}+\eta_{0b}\varepsilon'^{\,b}_{0}+\eta_{ab}\varepsilon'^{\,a}_{0}\varepsilon'^{\,b}_{0}=2\varepsilon'^{\,0}_{0}+\varepsilon'^{\,0}_{0}\varepsilon'^{\,0}_{0}-\varepsilon'^{\,1}_{0}\varepsilon'^{\,1}_{0}-\varepsilon'^{\,2}_{0}\varepsilon'^{\,2}_{0}-\varepsilon'^{\,3}_{0}\varepsilon'^{\,3}_{0}=\phi^2k^2A_0A_0\\&\eta_{a1}\varepsilon'^{\,a}_{1}+\eta_{1b}\varepsilon'^{\,b}_{1}+\eta_{ab}\varepsilon'^{\,a}_{1}\varepsilon'^{\,b}_{1}=-2\varepsilon'^{\,1}_{1}+\varepsilon'^{\,0}_{1}\varepsilon'^{\,0}_{1}-\varepsilon'^{\,1}_{1}\varepsilon'^{\,1}_{1}-\varepsilon'^{\,2}_{1}\varepsilon'^{\,2}_{1}-\varepsilon'^{\,3}_{1}\varepsilon'^{\,3}_{1}=\phi^2k^2A_1A_1\\&\eta_{a2}\varepsilon'^{\,a}_{2}+\eta_{2b}\varepsilon'^{\,b}_{2}+\eta_{ab}\varepsilon'^{\,a}_{2}\varepsilon'^{\,b}_{2}=-2\varepsilon'^{\,2}_{2}+\varepsilon'^{\,0}_{2}\varepsilon'^{\,0}_{2}-\varepsilon'^{\,1}_{2}\varepsilon'^{\,1}_{2}-\varepsilon'^{\,2}_{2}\varepsilon'^{\,2}_{2}-\varepsilon'^{\,3}_{2}\varepsilon'^{\,3}_{2}=\phi^2k^2A_2A_2\\&\eta_{a3}\varepsilon'^{\,a}_{3}+\eta_{3b}\varepsilon'^{\,b}_{3}+\eta_{ab}\varepsilon'^{\,a}_{3}\varepsilon'^{\,b}_{3}=-2\varepsilon'^{\,3}_{3}+\varepsilon'^{\,0}_{3}\varepsilon'^{\,0}_{3}-\varepsilon'^{\,1}_{3}\varepsilon'^{\,1}_{3}-\varepsilon'^{\,2}_{3}\varepsilon'^{\,2}_{3}-\varepsilon'^{\,3}_{3}\varepsilon'^{\,3}_{3}=\phi^2k^2A_3A_3\end{aligned}\quad.\tag{2.4a}$$

Likewise, the three $\mu=0$, $\nu=1,2,3$ mixed time and space relations in (2.3) are:

$$\begin{aligned}&\eta_{a1}\varepsilon'^{\,a}_{0}+\eta_{0b}\varepsilon'^{\,b}_{1}+\eta_{ab}\varepsilon'^{\,a}_{0}\varepsilon'^{\,b}_{1}=-\varepsilon'^{\,1}_{0}+\varepsilon'^{\,0}_{1}+\varepsilon'^{\,0}_{0}\varepsilon'^{\,0}_{1}-\varepsilon'^{\,1}_{0}\varepsilon'^{\,1}_{1}-\varepsilon'^{\,2}_{0}\varepsilon'^{\,2}_{1}-\varepsilon'^{\,3}_{0}\varepsilon'^{\,3}_{1}=\phi^2k^2A_0A_1\\&\eta_{a2}\varepsilon'^{\,a}_{0}+\eta_{0b}\varepsilon'^{\,b}_{2}+\eta_{ab}\varepsilon'^{\,a}_{0}\varepsilon'^{\,b}_{2}=-\varepsilon'^{\,2}_{0}+\varepsilon'^{\,0}_{2}+\varepsilon'^{\,0}_{0}\varepsilon'^{\,0}_{2}-\varepsilon'^{\,1}_{0}\varepsilon'^{\,1}_{2}-\varepsilon'^{\,2}_{0}\varepsilon'^{\,2}_{2}-\varepsilon'^{\,3}_{0}\varepsilon'^{\,3}_{2}=\phi^2k^2A_0A_2\\&\eta_{a3}\varepsilon'^{\,a}_{0}+\eta_{0b}\varepsilon'^{\,b}_{3}+\eta_{ab}\varepsilon'^{\,a}_{0}\varepsilon'^{\,b}_{3}=-\varepsilon'^{\,3}_{0}+\varepsilon'^{\,0}_{3}+\varepsilon'^{\,0}_{0}\varepsilon'^{\,0}_{3}-\varepsilon'^{\,1}_{0}\varepsilon'^{\,1}_{3}-\varepsilon'^{\,2}_{0}\varepsilon'^{\,2}_{3}-\varepsilon'^{\,3}_{0}\varepsilon'^{\,3}_{3}=\phi^2k^2A_0A_3\end{aligned}\,.\tag{2.4b}$$

Finally, the pure-space relations in (2.3) are:

$$\begin{aligned}
&\eta_{a2}\varepsilon_1'^a+\eta_{1b}\varepsilon_2'^b+\eta_{ab}\varepsilon_1'^a\varepsilon_2'^b=-\varepsilon_1'^2-\varepsilon_2'^1+\varepsilon_1'^0\varepsilon_2'^0-\varepsilon_1'^1\varepsilon_2'^1-\varepsilon_1'^2\varepsilon_2'^2-\varepsilon_1'^3\varepsilon_2'^3=\phi^2k^2A_1A_2\\
&\eta_{a3}\varepsilon_2'^a+\eta_{2b}\varepsilon_3'^b+\eta_{ab}\varepsilon_2'^a\varepsilon_3'^b=-\varepsilon_2'^3-\varepsilon_3'^2+\varepsilon_2'^0\varepsilon_3'^0-\varepsilon_2'^1\varepsilon_3'^1-\varepsilon_2{}^2\varepsilon_3'^2-\varepsilon_2'^3\varepsilon_3'^3=\phi^2k^2A_2A_3\,.\\
&\eta_{a1}\varepsilon_3'^a+\eta_{3b}\varepsilon_1'^b+\eta_{ab}\varepsilon_3'^a\varepsilon_1'^b=-\varepsilon_3'^1-\varepsilon_1'^3+\varepsilon_3'^0\varepsilon_1'^0-\varepsilon_3'^1\varepsilon_1'^1-\varepsilon_3'^2\varepsilon_1'^2-\varepsilon_3'^3\varepsilon_1'^3=\phi^2k^2A_3A_1
\end{aligned}\tag{2.4c}$$

Now, we notice that the right-hand side of all ten of (2.4) have nonlinear second-order products $\phi^2k^2A_\mu A_\nu$ of field terms, while on the left of each there is a mix of linear first-order and nonlinear second-order expressions containing the ε'^a_μ. Our goal at the moment, therefore, is to eliminate all of the first order expressions from the left-hand sides of (2.4) to create a structural match whereby a sum of second order terms on the left is equal to a second order term on the right.

In (2.3a) the linear appearances are of $\varepsilon_0'^0$, $\varepsilon_1'^1$, $\varepsilon_2'^2$ and $\varepsilon_3'^3$ respectively. Noting that the complete tetrad $\varepsilon_\mu{}^a=\delta_\mu{}^a+\varepsilon'^a_\mu$ and that $\varepsilon_\mu{}^a=\delta_\mu{}^a$ when electrodynamics is turned off, we first require that $\varepsilon_\mu{}^a=\delta_\mu{}^a$ for the four $\mu=a$ diagonal components, and therefore, that $\varepsilon_0'^0=\varepsilon_1'^1=\varepsilon_2'^2=\varepsilon_3'^3=0$. As a result, the fields in $\phi^2k^2A_\mu A_\nu$ will all appear in off-diagonal components of the tetrad. With this, (2.4a) reduce to:

$$\begin{aligned}
&-\varepsilon_0'^1\varepsilon_0'^1-\varepsilon_0'^2\varepsilon_0'^2-\varepsilon_0'^3\varepsilon_0'^3=\phi^2k^2A_0A_0\\
&\varepsilon_1'^0\varepsilon_1'^0-\varepsilon_1'^2\varepsilon_1'^2-\varepsilon_1'^3\varepsilon_1'^3=\phi^2k^2A_1A_1\\
&\varepsilon_2'^0\varepsilon_2'^0-\varepsilon_2'^1\varepsilon_2'^1-\varepsilon_2'^3\varepsilon_2'^3=\phi^2k^2A_2A_2\\
&\varepsilon_3'^0\varepsilon_3'^0-\varepsilon_3'^1\varepsilon_3'^1-\varepsilon_3'^2\varepsilon_3'^2=\phi^2k^2A_3A_3
\end{aligned}\quad.\tag{2.5a}$$

In (2.4b) we achieve structural match using $\varepsilon_1'^1=\varepsilon_2'^2=\varepsilon_3'^3=0$ from above, and also by setting $\varepsilon_0'^1=\varepsilon_1'^0$, $\varepsilon_0'^2=\varepsilon_2'^0$, $\varepsilon_0'^3=\varepsilon_3'^0$, which is symmetric under $0\leftrightarrow a=1,2,3$ interchange. Therefore:

$$\begin{aligned}
&-\varepsilon_0'^2\varepsilon_1'^2-\varepsilon_0'^3\varepsilon_1'^3=\phi^2k^2A_0A_1\\
&-\varepsilon_0'^1\varepsilon_2'^1-\varepsilon_0'^3\varepsilon_2'^3=\phi^2k^2A_0A_2\,.\\
&-\varepsilon_0'^1\varepsilon_3'^1-\varepsilon_0'^2\varepsilon_3'^2=\phi^2k^2A_0A_3
\end{aligned}\tag{2.5b}$$

In (2.4c) we use $\varepsilon_1'^1=\varepsilon_2'^2=\varepsilon_3'^3=0$ from above and also set $\varepsilon_1'^2=-\varepsilon_2'^1$, $\varepsilon_2'^3=-\varepsilon_3'^2$, $\varepsilon_3'^1=-\varepsilon_1'^3$ which are antisymmetric under interchange of different space indexes. Therefore, we now have:

$$\begin{aligned}
&\varepsilon_1'^0\varepsilon_2'^0-\varepsilon_1'^3\varepsilon_2'^3=\phi^2k^2A_1A_2\\
&\varepsilon_2'^0\varepsilon_3'^0-\varepsilon_2'^1\varepsilon_3'^1=\phi^2k^2A_2A_3\,.\\
&\varepsilon_3'^0\varepsilon_1'^0-\varepsilon_3'^2\varepsilon_1'^2=\phi^2k^2A_3A_1
\end{aligned}\tag{2.5c}$$

In all of (2.5), we now only have matching-structure second-order terms on both sides.

For the next step, closely studying the space indexes in all of (2.5) above, we now make an educated guess at an assignment for the fields in $\phi^2 k^2 A_i A_j$. Specifically, also using the symmetric-interchange $\varepsilon_0'^1 = \varepsilon_1'^0$, $\varepsilon_0'^2 = \varepsilon_2'^0$, $\varepsilon_0'^3 = \varepsilon_3'^0$ from earlier, we now guess an assignment:

$$\varepsilon_0'^1 = \varepsilon_1'^0 = \phi k A_1; \quad \varepsilon_0'^2 = \varepsilon_2'^0 = \phi k A_2; \quad \varepsilon_0'^3 = \varepsilon_3'^0 = \phi k A_3 . \tag{2.6}$$

Because all space-indexed expressions in (2.5) contain second-order products of the above, it is possible to have also tried using a minus sign in all of (2.5) whereby $\varepsilon_0'^1 = \varepsilon_1'^0 = -\phi k A_1$, $\varepsilon_0'^2 = \varepsilon_2'^0 = -\phi k A_2$ and $\varepsilon_0'^3 = \varepsilon_3'^0 = -\phi k A_3$. But absent motivation to the contrary, we employ a plus sign which is implicit in the above. Substituting (2.6) into all of (2.5) and reducing now yields:

$$\begin{aligned}
&-A_1 A_1 - A_2 A_2 - A_3 A_3 = A_0 A_0 \\
&-\varepsilon_1'^2 \varepsilon_1'^2 - \varepsilon_1'^3 \varepsilon_1'^3 = 0 \\
&-\varepsilon_2'^1 \varepsilon_2'^1 - \varepsilon_2'^3 \varepsilon_2'^3 = 0 \\
&-\varepsilon_3'^1 \varepsilon_3'^1 - \varepsilon_3'^2 \varepsilon_3'^2 = 0
\end{aligned} \quad , \tag{2.7a}$$

$$\begin{aligned}
&-\phi k A_2 \varepsilon_1'^2 - \phi k A_3 \varepsilon_1'^3 = \phi^2 k^2 A_0 A_1 \\
&-\phi k A_1 \varepsilon_2'^1 - \phi k A_3 \varepsilon_2'^3 = \phi^2 k^2 A_0 A_2 \\
&-\phi k A_1 \varepsilon_3'^1 - \phi k A_2 \varepsilon_3'^2 = \phi^2 k^2 A_0 A_3
\end{aligned} , \tag{2.7b}$$

$$-\varepsilon_1'^3 \varepsilon_2'^3 = -\varepsilon_2'^1 \varepsilon_3'^1 = -\varepsilon_3'^2 \varepsilon_1'^2 = 0 . \tag{2.7c}$$

Now, one way to satisfy the earlier relations $\varepsilon_1'^2 = -\varepsilon_2'^1$, $\varepsilon_2'^3 = -\varepsilon_3'^2$, $\varepsilon_3'^1 = -\varepsilon_1'^3$ used in (2.5c) as well as to satisfy (2.7c), is to set all of the pure-space components:

$$\varepsilon_1'^2 = \varepsilon_2'^1 = \varepsilon_2'^3 = \varepsilon_3'^2 = \varepsilon_3'^1 = \varepsilon_1'^3 = 0 . \tag{2.8}$$

This disposes of (2.7c) and last three relations in (2.7a), leaving only the two constraints:

$$-A_1 A_1 - A_2 A_2 - A_3 A_3 = A_0 A_0 , \tag{2.9a}$$

$$0 = \phi^2 k^2 A_0 A_1 = \phi^2 k^2 A_0 A_2 = \phi^2 k^2 A_0 A_3 . \tag{2.9b}$$

These above relations (2.9) are extremely important. In (2.9b), if *any one* of A_1, A_2 or A_3 is *not* equal to zero, then we *must* have $A_0 = 0$. So, we take as a given that at least one of A_1, A_2 or A_3 is non-zero, whereby (2.9a) and (2.9b) together become:

$$A_0 = 0; \quad A_1 A_1 + A_2 A_2 + A_3 A_3 = 0\,, \tag{2.10}$$

These two constraints have removed two redundant degrees of freedom from the gauge field A_μ*, in a generally-covariant manner*. Moreover, for the latter constraint in $A_1 A_1 + A_2 A_2 + A_3 A_3 = 0$ to be satisfied, it is necessary that at least one of the space components of A_j be *imaginary*. For example, if $A_3 = 0$, then one way to solve the entirety of (2.10) is to have:

$$A_\mu = A\varepsilon_\mu \exp\left(-iq_\sigma x^\sigma / \hbar\right), \tag{2.11a}$$

with a polarization vector

$$\varepsilon_{R,L\mu}\left(\hat{z}\right) \equiv \begin{pmatrix} 0 & \pm 1 & +i & 0 \end{pmatrix} / \sqrt{2}\,, \tag{2.11b}$$

where *A* has dimensions of charge / energy to provide dimensional balance given the dimensionless $\varepsilon_{R,L\mu}$. But the foregoing is instantly-recognizable as the gauge potential $A_\mu = A_{\gamma\mu}$ for an individual photon (denoted with γ) with two helicity states propagating along the *z* axis, having an energy-momentum vector

$$cq^\mu\left(\hat{z}\right) = \begin{pmatrix} E & 0 & 0 & cq_z \end{pmatrix} = \begin{pmatrix} h\nu & 0 & 0 & h\nu \end{pmatrix}. \tag{2.11c}$$

This satisfies $q_\mu q^\mu = m_\gamma{}^2 c^2 = 0$, which makes this a massless, luminous field quantum. Additionally, we see from all of (2.11) that $A_\mu q^\mu = 0$, and $A_j q^j = 0$ as is also true for a photon. The latter $A_j q^j = 0$ is the so-called Coulomb gauge which is ordinarily imposed as a non-covariant gauge condition. But here, it has emerged in an entirely covariant fashion.

In short, what we have ascertained in (2.10) and (2.11) is that if the spacetime components $G_{\mu\nu} = g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$ of the Kaluza-Klein metric tensor with $g_{\mu\nu} = \eta_{\mu\nu}$ are to produce a set of Γ_μ satisfying the Dirac anticommutator relation $\frac{1}{2}\left\{\Gamma_\mu, \Gamma_\nu\right\} \equiv G_{\mu\nu}$, *the gauge symmetry of* A_μ *must be broken to correspond with that of the photon,* $A_\mu = A_{\gamma\mu}$. *The very act of deconstructing* $G_{\mu\nu}$ *into square root Dirac operators covariantly removes two degrees of freedom from the gauge field and forces it to become a photon field quantum.* Moreover, (2.11a) implies that $i\hbar\partial_\alpha A_\mu = q_\alpha A_\mu$ while (2.11c) contains the energy $E = h\nu$ of a single photon. So, *starting with an entirely-classical* $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$ *and merely requiring the formation of a set of* Γ_μ *transforming covariantly in spacetime with the anticommutator* $\frac{1}{2}\left\{\Gamma_\mu, \Gamma_\nu\right\} \equiv G_{\mu\nu}$, *we covariantly end up with some of the core relations of quantum mechanics*.

Even outside of the context of Kaluza-Klein theory, entirely in four-dimensional spacetime, the foregoing calculation solves the long-perplexing problem of how to covariantly eliminate the redundancy inherent in using a four-component Lorentz vector A_μ *to describe a classical*

electromagnetic wave or a quantum photon field with only two transverse degrees of physical freedom: If we posit a metric tensor given by $G_{\mu\nu} = g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$, and if we require the existence of a set of Dirac operators Γ_μ transforming as a covariant vector in spacetime and connected to the metric tensor such that $\frac{1}{2}\{\Gamma_\mu, \Gamma_\nu\} \equiv G_{\mu\nu}$, then we are given *no choice* but to have $A_\mu = A_{\gamma\mu}$ be the quantum field of a photon with two degrees of freedom covariantly-removed and only two degrees of freedom remaining.

Moreover, we have also deduced all of the components of the tetrad $\varepsilon_\mu{}^a = \delta_\mu{}^a + \varepsilon'_\mu{}^a$. Pulling together all of $\varepsilon'^0_0 = \varepsilon'^1_1 = \varepsilon'^2_2 = \varepsilon'^3_3 = 0$ together with (2.6) and (2.8), and setting $A_\mu = A_{\gamma\mu}$ to incorporate the pivotal finding in (2.10), (2.11) that the gauge-field must be covariantly fixed to the gauge field of a photon (again, γ is a subscript, *not a spacetime index*), this tetrad is:

$$\varepsilon_\mu{}^a = \delta_\mu{}^a + \varepsilon'_\mu{}^a = \begin{pmatrix} 1 & \phi k A_{\gamma 1} & \phi k A_{\gamma 2} & \phi k A_{\gamma 3} \\ \phi k A_{\gamma 1} & 1 & 0 & 0 \\ \phi k A_{\gamma 2} & 0 & 1 & 0 \\ \phi k A_{\gamma 3} & 0 & 0 & 1 \end{pmatrix}. \tag{2.12}$$

Finally, because $\varepsilon_{\mu a}\gamma^a = \varepsilon_\mu{}^\alpha \gamma_\alpha \equiv \Gamma_\mu$, we may use (2.12) to deduce that the Dirac operators:

$$\begin{aligned} \Gamma_0 &= \varepsilon_0{}^\alpha \gamma_\alpha = \varepsilon_0{}^0\gamma_0 + \varepsilon_0{}^1\gamma_1 + \varepsilon_0{}^2\gamma_2 + \varepsilon_0{}^3\gamma_3 = \gamma_0 + \phi k A_{\gamma j}\gamma_j \\ \Gamma_1 &= \varepsilon_1{}^\alpha \gamma_\alpha = \varepsilon_1{}^0\gamma_0 + \varepsilon_1{}^1\gamma_1 = \gamma_1 + \phi k A_{\gamma 1}\gamma_0 \\ \Gamma_2 &= \varepsilon_2{}^\alpha \gamma_\alpha = \varepsilon_2{}^0\gamma_0 + \varepsilon_2{}^2\gamma_2 = \gamma_2 + \phi k A_{\gamma 2}\gamma_0 \\ \Gamma_3 &= \varepsilon_3{}^\alpha \gamma_\alpha = \varepsilon_3{}^0\gamma_0 + \varepsilon_3{}^3\gamma_3 = \gamma_3 + \phi k A_{\gamma 3}\gamma_0 \end{aligned}, \tag{2.13}$$

which consolidate into a set of Γ_μ transforming as a four-vector in spacetime, namely:

$$\Gamma_\mu = \begin{pmatrix} \gamma_0 + \phi k A_{\gamma j}\gamma_j & \gamma_j + \phi k A_{\gamma j}\gamma_0 \end{pmatrix}. \tag{2.14}$$

It is a useful exercise to confirm that (2.14) above, inserted into (2.1), will produce $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$, which may then be generalized from $\eta_{\mu\nu} \mapsto g_{\mu\nu}$ in the usual way by applying the minimal coupling principle. As a result, we return to the Kaluza-Klein metric tensors in (1.1), but apply the foregoing to now rewrite these as:

$$G_{\text{MN}} = \begin{pmatrix} g_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu} & \phi^2 k A_{\gamma\mu} \\ \phi^2 k A_{\gamma\nu} & \phi^2 \end{pmatrix}; \quad G^{\text{MN}} = \begin{pmatrix} g^{\mu\nu} & -A_\gamma{}^\mu \\ -A_\gamma{}^\nu & g_{\alpha\beta} A_\gamma{}^\alpha A_\gamma{}^\beta + 1/\phi^2 \end{pmatrix}. \tag{2.15}$$

The only change we have made is to replace $A_\mu \mapsto A_{\gamma\mu}$, which is to represent the remarkable result that *even in four spacetime dimensions alone*, it is not possible to deconstruct $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$ into a set of Dirac Γ_μ defined using (2.1) without fixing the gauge field A_μ to that of a photon $A_{\gamma\mu}$. Now, we extend this general covariance to the fifth dimension.

3. Derivation of the "Dirac-Kaluza-Klein" (DKK) Five-Dimensional Metric Tensor

To ensure general covariance at the Dirac level in five-dimensions, it is necessary to first extend (2.1) into all five dimensions. For this we use the lower-indexed (1.3), namely:

$$\tfrac{1}{2}\{\Gamma_{\rm M},\Gamma_{\rm N}\} = \tfrac{1}{2}\{\Gamma_{\rm M}\Gamma_{\rm N} + \Gamma_{\rm N}\Gamma_{\rm M}\} \equiv G_{\rm MN}\,. \tag{3.1}$$

As just shown, the spacetime components of (3.1) with $g_{\mu\nu} = \eta_{\mu\nu}$ and using (2.14) will already reproduce $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$ in (2.15). Now we turn to the fifth-dimensional components.

We first find it helpful to separate the time and space components of $G_{\rm MN}$ in (2.15) and so rewrite this as:

$$G_{\rm MN} = \begin{pmatrix} G_{00} & G_{0k} & G_{05} \\ G_{j0} & G_{jk} & G_{j5} \\ G_{50} & G_{5k} & G_{55} \end{pmatrix} = \begin{pmatrix} g_{00} + \phi^2 k^2 A_{\gamma 0} A_{\gamma 0} & g_{0k} + \phi^2 k^2 A_{\gamma 0} A_{\gamma k} & \phi^2 k A_{\gamma 0} \\ g_{j0} + \phi^2 k^2 A_{\gamma j} A_{\gamma 0} & g_{jk} + \phi^2 k^2 A_{\gamma j} A_{\gamma k} & \phi^2 k A_{\gamma j} \\ \phi^2 k A_{\gamma 0} & \phi^2 k A_{\gamma k} & \phi^2 \end{pmatrix}. \tag{3.2}$$

We know of course that $A_{\gamma 0} = 0$, which is the constraint that first arose from (2.10). So, if we again work with $g_{\mu\nu} = \eta_{\mu\nu}$ and set $A_{\gamma 0} = 0$, the above simplifies to:

$$G_{\rm MN} = \begin{pmatrix} G_{00} & G_{0k} & G_{05} \\ G_{j0} & G_{jk} & G_{j5} \\ G_{50} & G_{5k} & G_{55} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \eta_{jk} + \phi^2 k^2 A_{\gamma j} A_{\gamma k} & \phi^2 k A_{\gamma j} \\ 0 & \phi^2 k A_{\gamma k} & \phi^2 \end{pmatrix}. \tag{3.3}$$

Next, let us *define* a Γ_5 to go along with the remaining Γ_μ in (2.14) in such a way as to *require* that the symmetric components $G_{j5} = G_{5j} = \phi^2 k A_{\gamma j}$ in (3.3) remain fully intact without any change. This is important, because these components in particular are largely responsible for the Kaluza "miracles" which reproduce Maxwell's equations together with the Lorentz Force motion and the Maxwell Stress-Energy Tensor. At the same time, because $A_{\gamma 0} = 0$ as uncovered at (2.10), we can always maintain covariance between the space components $G_{j5} = G_{5j} = \phi^2 k A_{\gamma j}$ and the time components $G_{05} = G_{50}$ in the manner of (1.1) by adding $\phi^2 k A_{\gamma 0} = 0$ to anything else we

deduce for $G_{05} = G_{50}$, so we lay the foundation for the Kaluza miracles to remain intact. We impose this requirement though (3.1) by writing the Γ_5 *definition* as:

$$\tfrac{1}{2}\{\Gamma_j, \Gamma_5\} = \tfrac{1}{2}\{\Gamma_j\Gamma_5 + \Gamma_5\Gamma_j\} \equiv G_{j5} = G_{j5} = \phi^2 k A_{\gamma j}\,. \tag{3.4}$$

Using $\Gamma_j = \gamma_j + \phi k A_{\gamma j}\gamma_0$ from (2.14) and adding in a zero, the above now becomes:

$$0 + \phi^2 k A_{\gamma j} \equiv \tfrac{1}{2}\{\Gamma_j\Gamma_5 + \Gamma_5\Gamma_j\} = \tfrac{1}{2}\{\gamma_j, \Gamma_5\} + \tfrac{1}{2}\phi k A_{\gamma j}\{\gamma_0, \Gamma_5\}\,, \tag{3.5}$$

which reduces down to a pair of anticommutation constraints on Γ_5, namely:

$$\begin{aligned} 0 &= \tfrac{1}{2}\{\gamma_j, \Gamma_5\} \\ \phi &= \tfrac{1}{2}\{\gamma_0, \Gamma_5\} \end{aligned}\,. \tag{3.6}$$

Now let's examine possible options for Γ_5.

Given that $\Gamma_0 = \gamma_0 + \phi k A_{\gamma j}\gamma_j$ and $\Gamma_j = \gamma_j + \phi k A_{\gamma j}\gamma_0$ in (2.14), we anticipate the general form for Γ_5 to be $\Gamma_5 \equiv \gamma_X + Y$ in which we define two unknowns to be determined using (3.6). First, *X* is one of the indexes 0, 1, 2, 3 or 5 of a Dirac matrix. Second, *Y* is a complete unknown which we anticipate will also contain a Dirac matrix as do the operators in (2.14). Using $\Gamma_5 \equiv \gamma_X + Y$ in (3.6) we first deduce:

$$\begin{aligned} 0 &= \tfrac{1}{2}\{\gamma_j\Gamma_5 + \Gamma_5\gamma_j\} = \tfrac{1}{2}\{\gamma_j\gamma_X + \gamma_j Y + \gamma_X\gamma_j + Y\gamma_j\} = \tfrac{1}{2}\{\gamma_j, \gamma_X\} + \tfrac{1}{2}\{\gamma_j, Y\} \\ 0 + \phi &= \tfrac{1}{2}\{\gamma_0\Gamma_5 + \Gamma_5\gamma_0\} = \tfrac{1}{2}\{\gamma_0\gamma_X + \gamma_0 Y + \gamma_X\gamma_0 + Y\gamma_0\} = \tfrac{1}{2}\{\gamma_0, \gamma_X\} + \tfrac{1}{2}\{\gamma_0, Y\} \end{aligned}\,. \tag{3.7}$$

From the top line, so long as $\gamma_X \neq -Y$ which means so long as $\Gamma_5 \neq 0$, we must have both the anticommutators $\{\gamma_j, \gamma_X\} = 0$ and $\{\gamma_j, Y\} = 0$. The former $\{\gamma_j, \gamma_X\} = 0$ excludes *X* being a space index 1, 2 or 3 leaving only $\gamma_X = \gamma_0$ or $\gamma_X = \gamma_5$. The latter $\{\gamma_j, Y\} = 0$ makes clear that whatever Dirac operator is part of *Y* must likewise be either γ_0 or γ_5. From the bottom line, however, we must also have the anticommutators $\{\gamma_0, \gamma_X\} = 0$ and $\tfrac{1}{2}\{\gamma_0, Y\} = \phi$. The former means that the only remaining choice is $\gamma_X = \gamma_5$, while given $\gamma_0\gamma_0 = 1$ and $\{\gamma_0, \gamma_5\} = 0$ the latter means that $Y = \phi\gamma_0$. Therefore, we conclude that $\Gamma_5 = \gamma_5 + \phi\gamma_0$. Thus, including this in (2.14) now gives:

$$\Gamma_{\mathrm{M}} = \begin{pmatrix} \gamma_0 + \phi k A_{\gamma k}\gamma_k & \gamma_j + \phi k A_{\gamma j}\gamma_0 & \gamma_5 + \phi\gamma_0 \end{pmatrix}. \tag{3.8}$$

With this final operator $\Gamma_5 \equiv \gamma_5 + \phi\gamma_0$, we can use all of (3.8) above in (3.1) to precisely reproduce $G_{j5} = \phi^2 k A_{\gamma j}$ and $G_{5k} = \phi^2 k A_{\gamma k}$ in (3.3), as well as $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$ given $A_{\gamma 0} = 0$. This leaves the remaining components $G_{05} = G_{50}$ and G_{55} to which we now turn.

If we use $\Gamma_0 = \gamma_0 + \phi k A_{\gamma j}\gamma_j$ and $\Gamma_5 = \gamma_5 + \phi\gamma_0$ in (3.1) to ensure that these remaining components are also fully covariant over all five dimensions, then we determine that:

$$\begin{aligned} G_{05} = G_{50} &= \tfrac{1}{2}\{\Gamma_0\Gamma_5 + \Gamma_5\Gamma_0\} = \tfrac{1}{2}\Big\{\big(\gamma_0 + \phi k A_{\gamma j}\gamma_j\big)\big(\gamma_5 + \phi\gamma_0\big) + \big(\gamma_5 + \phi\gamma_0\big)\big(\gamma_0 + \phi k A_{\gamma j}\gamma_j\big)\Big\} \\ &= \phi\gamma_0\gamma_0 + \tfrac{1}{2}\{\gamma_0, \gamma_5\} + \tfrac{1}{2}\phi k A_{\gamma j}\{\gamma_j, \gamma_5\} + \tfrac{1}{2}\phi^2 k A_{\gamma j}\{\gamma_j, \gamma_0\} = \phi \end{aligned}, \tag{3.9}$$

$$G_{55} = \Gamma_5\Gamma_5 = \big(\gamma_5 + \phi\gamma_0\big)\big(\gamma_5 + \phi\gamma_0\big) = \gamma_5\gamma_5 + \phi^2\gamma_0\gamma_0 + \phi\{\gamma_5\gamma_0 + \gamma_0\gamma_5\} = 1 + \phi^2 . \tag{3.10}$$

These two components are now different from those in (3.3). However, in view of this Dirac operator deconstruction these are required to be different to ensure that the metric tensor is completely generally-covariant across all five dimensions, just as we were required at (2.15) to set $A_j = A_{\gamma j}$ at (2.12) to ensure even basic covariance in four spacetime dimensions.

Consequently, changing (3.3) to incorporate (3.9) and (3.10), we now have:

$$G_{\mathrm{MN}} = \begin{pmatrix} G_{00} & G_{0k} & G_{05} \\ G_{j0} & G_{jk} & G_{j5} \\ G_{50} & G_{5k} & G_{55} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \phi \\ 0 & \eta_{jk} + \phi^2 k^2 A_{\gamma j} A_{\gamma k} & \phi^2 k A_{\gamma j} \\ \phi & \phi^2 k A_{\gamma k} & 1 + \phi^2 \end{pmatrix}. \tag{3.11}$$

This metric tensor is fully covariant across all five dimensions, and because it is rooted in the Dirac operators (3.8), we expect that this can be made fully compatible with Dirac's theory of the multitude of fermions observed in the natural world, as we shall examine further in section 5. Moreover, in the context of Kaluza-Klein theory, Dirac's Quantum Theory of the Electron [13] has also forced us to set $A_j = A_{\gamma j}$ in the metric tensor, and thereby also served up a *quantum theory of the photon*. Because of its origins in requiring Kaluza-Klein theory to be compatible with Dirac theory, we shall refer to the above as the "Dirac-Kaluza-Klein" (DKK) metric tensor, and shall give the same name to the overall theory based on this.

Importantly, when electrodynamics is turned off by setting $A_{\gamma j} = 0$ and $\phi = 0$ the signature of (3.11) becomes $\mathrm{diag}\big(G_{\mathrm{MN}}\big) = \big(+1,-1,-1,-1,+1\big)$ with a determinant $\big|G_{\mathrm{MN}}\big| = -1$, versus $\big|G_{\mathrm{MN}}\big| = 0$ in (1.1) as reviewed earlier. This means that the inverse obtained via $G^{\mathrm{MA}}G_{\mathrm{AN}} = \delta^{\mathrm{M}}{}_{\mathrm{N}}$ will be non-singular as opposed to that in (1.1), and that there is no reliance whatsoever on having $\phi \neq 0$ in order to avoid singularity. This in turn frees G_{55} from the energy requirements of ϕ which cause the fifth dimension in (1.1) to have a spacelike signature. And in fact, we see that as a result of this signature, *the fifth dimension in (3.11) is a second timelike, not fourth spacelike,*

dimension. In turn, because (3.10) shows that $G_{55} = 1 + \phi^2 = \gamma_5\gamma_5 + \phi^2$ obtains its signature from $\gamma_5\gamma_5 = 1$, it now becomes possible to fully associate the Kaluza-Klein fifth dimension with the γ_5 of Dirac theory. This is not possible when a theory based on (1.1) causes G_{55} to be spacelike even though $\gamma_5\gamma_5 = 1$ is timelike, because of this conflict between timelike and spacelike signatures. Moreover, having only $G_{55} = \phi^2$ causes G_{55} to shrink or expand or even zero out entirely, based on the magnitude of ϕ. In (3.11), there is no such problem. We shall review the physics consequences of all these matters more deeply in section 11 following other development. At the moment, we wish to consolidate (3.11) into the 2x2 matrix format akin to (1.1), which consolidates all spacetime components into a single expression with manifest four-dimensional covariance.

In general, as already hinted, it will sometimes simplify calculation to set $A_{\gamma 0} = 0$ simply because this puts some zeros in the equations we are working with; while at other times it will be better to explicitly include $A_{\gamma 0}$ knowing this is zero in order to take advantage of the consolidations enabled by general covariance. To consolidate (3.11) to 2x2 format, we do the latter, by restoring the zeroed $A_{\gamma 0} = 0$ to the spacetime components of (3.11) and consolidating them to $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$. This is exactly what is in the Kaluza-Klein metric tensor (1.1) when $g_{\mu\nu} = \eta_{\mu\nu}$, but for the fact that the gauge symmetry has been broken to force $A_\mu = A_{\gamma\mu}$. But we also know that $G_{05} = G_{50}$ and $G_{j5} = G_{5j}$ have been constructed at (3.9) and (3.4) to form a four-vector in spacetime. Therefore, referring to these components in (3.11), we now *define* a new covariant (lower-indexed) four-vector:

$$\Phi_\mu \equiv \begin{pmatrix} \phi & \phi^2 k A_{\gamma j} \end{pmatrix}. \tag{3.12}$$

Moreover, $G_{55} = \gamma_5\gamma_5 + \phi^2\gamma_0\gamma_0$ in (3.10) teaches that the underlying timelike signature (and the metric non-singularity) is rooted in $\gamma_5\gamma_5 = 1$, and via $\phi^2\gamma_0\gamma_0 = \phi^2$ that the square of the scalar field is rooted in $\gamma_0\gamma_0 = 1$ which has two time indexes. So, we may now formally assign $\eta_{55} = 1$ to the fifth component of the Minkowski metric signature, and we may assign $\phi^2 = \Phi_0\Phi_0$ to the fields in $G_{\mu\nu}$ and G_{55}. With all of this, and using minimal coupling to generalize $\eta_{\text{MN}} \mapsto g_{\text{MN}}$ which also means accounting for non-zero $g_{\mu 5}$, $g_{5\nu}$, (3.11) may now be compacted via (3.12) to the 2x2 form:

$$\boxed{G_{\text{MN}} = \begin{pmatrix} G_{\mu\nu} & G_{\mu 5} \\ G_{5\nu} & G_{55} \end{pmatrix} = \begin{pmatrix} g_{\mu\nu} + \Phi_0\Phi_0 k^2 A_{\gamma\mu} A_{\gamma\nu} & g_{\mu 5} + \Phi_\mu \\ g_{5\nu} + \Phi_\nu & g_{55} + \Phi_0\Phi_0 \end{pmatrix}}. \tag{3.13}$$

This is the Dirac-Kaluza-Klein metric tensor which will form the basis for all continued development from here, and it should be closely contrasted with (1.1). The next step is to calculate the inverse G^{MN} of (3.13) above.

4. Calculation of the Inverse Dirac-Kaluza-Klein Metric Tensor

As already mentioned, the modified Kaluza-Klein metric tensor (3.13) has a *non-singular* inverse G^{MN} specified in the usual way by $G^{\text{MA}}G_{\text{AN}} = \delta^{\text{M}}{}_{\text{N}}$. We already know this because when all electromagnetic fields are turned off and $g_{\text{MN}} = \eta_{\text{MN}}$, we have a determinant $\left|G_{\text{MN}}\right| = -1$ which is one of the litmus tests that can be used to demonstrate non-singularity. But because this inverse is essential to being able to calculate connections, equations of motion, and the Einstein field equation and related energy tensors, the next important step – which is entirely mathematical – is to explicitly calculate the inverse of (3.13). We shall now do so.

Calculating the inverse of a 5x5 matrix is a very cumbersome task if one employs a brute force approach. But we can take great advantage of the fact that the tangent space Minkowski tensor $\text{diag}\left(\eta_{\text{MN}}\right) = \left(+1,-1,-1,-1,+1\right)$ has two timelike and three spacelike dimensions when we set $A_{\gamma j} = 0$ and $\phi = 0$ to turn off the electrodynamic fields, by using the analytic blockwise inversion method detailed, e.g., in [14]. Specifically, we split the 5x5 matrix into 2x2 and 3x3 matrices along the "diagonal", and into 2x3 and 3x2 matrices off the "diagonal." It is best to work from (3.11) which does not show the time component $A_{\gamma 0} = 0$ because this is equal to zero for a photon, and which employs $g_{\mu\nu} = \eta_{\mu\nu}$. We expand this to show the entire 5x5 matrix, and we move the rows and columns so the ordering of the indexes is not $\text{M} = 0,1,2,3,5$, but rather is $\text{M} = 0,5,1,2,3$. With all of this, (3.11) may be rewritten as:

$$G_{\text{MN}} = \begin{pmatrix} G_{00} & G_{05} & G_{01} & G_{02} & G_{03} \\ G_{50} & G_{55} & G_{51} & G_{52} & G_{53} \\ G_{10} & G_{15} & G_{11} & G_{12} & G_{13} \\ G_{20} & G_{25} & G_{21} & G_{22} & G_{23} \\ G_{30} & G_{35} & G_{31} & G_{32} & G_{33} \end{pmatrix} = \begin{pmatrix} 1 & \phi & 0 & 0 & 0 \\ \phi & 1+\phi^2 & \phi^2 kA_{\gamma 1} & \phi^2 kA_{\gamma 2} & \phi^2 kA_{\gamma 3} \\ 0 & \phi^2 kA_{\gamma 1} & -1+\phi^2 k^2 A_{\gamma 1}A_{\gamma 1} & \phi^2 k^2 A_{\gamma 1}A_{\gamma 2} & \phi^2 k^2 A_{\gamma 1}A_{\gamma 3} \\ 0 & \phi^2 kA_{\gamma 2} & \phi^2 k^2 A_{\gamma 2}A_{\gamma 1} & -1+\phi^2 k^2 A_{\gamma 2}A_{\gamma 2} & \phi^2 k^2 A_{\gamma 2}A_{\gamma 3} \\ 0 & \phi^2 kA_{\gamma 3} & \phi^2 k^2 A_{\gamma 3}A_{\gamma 1} & \phi^2 k^2 A_{\gamma 3}A_{\gamma 2} & -1+\phi^2 k^2 A_{\gamma 3}A_{\gamma 3} \end{pmatrix}. \quad (4.1)$$

Then, we find the inverse using the blockwise inversion relation:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \\ -\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & \left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{pmatrix} \quad (4.2)$$

with the matrix block assignments:

$$\mathbf{A} = \begin{pmatrix} 1 & \phi \\ \phi & 1+\phi^2 \end{pmatrix}; \quad \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ \phi^2 kA_{\gamma 1} & \phi^2 kA_{\gamma 2} & \phi^2 kA_{\gamma 3} \end{pmatrix};$$
$$\mathbf{C} = \begin{pmatrix} 0 & \phi^2 kA_{\gamma 1} \\ 0 & \phi^2 kA_{\gamma 2} \\ 0 & \phi^2 kA_{\gamma 3} \end{pmatrix}; \quad \mathbf{D} = \begin{pmatrix} -1+\phi^2 k^2 A_{\gamma 1}A_{\gamma 1} & \phi^2 k^2 A_{\gamma 1}A_{\gamma 2} & \phi^2 k^2 A_{\gamma 1}A_{\gamma 3} \\ \phi^2 k^2 A_{\gamma 2}A_{\gamma 1} & -1+\phi^2 k^2 A_{\gamma 2}A_{\gamma 2} & \phi^2 k^2 A_{\gamma 2}A_{\gamma 3} \\ \phi^2 k^2 A_{\gamma 3}A_{\gamma 1} & \phi^2 k^2 A_{\gamma 3}A_{\gamma 2} & -1+\phi^2 k^2 A_{\gamma 3}A_{\gamma 3} \end{pmatrix}. \quad (4.3)$$

The two inverses we must calculate are $\mathbf{A}^{-1}$ and $\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}$. The former is a 2x2 matrix easily inverted, see, e.g. [15]. Its determinant $\left|\mathbf{A}\right|=1+\phi^2-\phi^2=1$, so its inverse is:

$$\mathbf{A}^{-1}=\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}. \tag{4.4}$$

Next, we need to calculate $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$, then invert this. We first calculate:

$$\begin{aligned}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}&=-\begin{pmatrix}0 & \phi^2kA_{\gamma 1}\\ 0 & \phi^2kA_{\gamma 2}\\ 0 & \phi^2kA_{\gamma 3}\end{pmatrix}\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}\begin{pmatrix}0 & 0 & 0\\ \phi^2kA_{\gamma 1} & \phi^2kA_{\gamma 2} & \phi^2kA_{\gamma 3}\end{pmatrix}\\ &=-\begin{pmatrix}0 & \phi^2kA_{\gamma 1}\\ 0 & \phi^2kA_{\gamma 2}\\ 0 & \phi^2kA_{\gamma 3}\end{pmatrix}\begin{pmatrix}-\phi^3kA_{\gamma 1} & -\phi^3kA_{\gamma 2} & -\phi^3kA_{\gamma 3}\\ \phi^2kA_{\gamma 1} & \phi^2kA_{\gamma 2} & \phi^2kA_{\gamma 3}\end{pmatrix}=-\begin{pmatrix}\phi^4k^2A_{\gamma 1}A_{\gamma 1} & \phi^4k^2A_{\gamma 1}A_{\gamma 2} & \phi^4k^2A_{\gamma 1}A_{\gamma 3}\\ \phi^4k^2A_{\gamma 2}A_{\gamma 1} & \phi^4k^2A_{\gamma 2}A_{\gamma 2} & \phi^4k^2A_{\gamma 2}A_{\gamma 3}\\ \phi^4k^2A_{\gamma 3}A_{\gamma 1} & \phi^4k^2A_{\gamma 3}A_{\gamma 2} & \phi^4k^2A_{\gamma 3}A_{\gamma 3}\end{pmatrix}\end{aligned}. \tag{4.5}$$

Therefore:

$$\begin{aligned}\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}&=\begin{pmatrix}-1+\left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 1} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 2} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 3}\\ \left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 1} & -1+\left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 2} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 3}\\ \left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 1} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 2} & -1+\left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 3}\end{pmatrix}.\\ &=\eta_{jk}+\left(\phi^2-\phi^4\right)k^2A_{\gamma j}A_{\gamma k}\end{aligned} \tag{4.6}$$

We can easily invert this using the skeletal mathematical relation $\left(1+x\right)\left(1-x\right)=1-x^2$. Specifically, using the result in (4.6) we may write:

$$\begin{aligned}&\left(\eta_{jk}+\left(\phi^2-\phi^4\right)k^2A_{\gamma j}A_{\gamma k}\right)\left(\eta_{kl}-\left(\phi^2-\phi^4\right)k^2A_{\gamma k}A_{\gamma l}\right)\\ &=\eta_{jk}\eta_{kl}+\left(\phi^2-\phi^4\right)k^2\left(\eta_{kl}A_{\gamma j}A_{\gamma k}-\eta_{jk}A_{\gamma k}A_{\gamma l}\right)-\left(\phi^2-\phi^4\right)^2k^4A_{\gamma j}A_{\gamma k}A_{\gamma k}A_{\gamma l}=\delta_{jl}\end{aligned}. \tag{4.7}$$

The $A_{\gamma j}A_{\gamma k}A_{\gamma k}A_{\gamma l}$ term zeros out because $A_{\gamma k}A_{\gamma k}=0$ for the photon field. Sampling the diagonal $j=l=1$ term, $\eta_{k1}A_{\gamma 1}A_{\gamma k}-\eta_{1k}A_{\gamma k}A_{\gamma 1}=-A_{\gamma 1}A_{\gamma 1}+A_{\gamma 1}A_{\gamma 1}=0$. Sampling the off-diagonal $j=1$, $l=2$ term, $\eta_{k1}A_{\gamma 2}A_{\gamma k}-\eta_{2k}A_{\gamma k}A_{\gamma 1}=-A_{\gamma 2}A_{\gamma 1}+A_{\gamma 2}A_{\gamma 1}=0$. By rotational symmetry, all other terms zero as well. And of course, $\eta_{jk}\eta_{kl}=\delta_{jl}$. So (4.7) taken with (4.6) informs us that:

$$\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}=\eta_{jk}-\left(\phi^2-\phi^4\right)k^2A_{\gamma j}A_{\gamma k}$$
$$=\begin{pmatrix}-1-\left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 1} & -\left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 3}\\ -\left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 1} & -1-\left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 3}\\ -\left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 1} & -\left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 2} & -1-\left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 3}\end{pmatrix}. \tag{4.8}$$

We now have all the inverses we need; the balance of the calculation is matrix multiplication.

From the lower-left block in (4.2) we use **C** in (4.3), with (4.4) and (4.8), to calculate:

$$-\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1}$$
$$=\begin{pmatrix}1+\left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 1} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 2} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 1}A_{\gamma 3}\\ \left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 1} & 1+\left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 2} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 2}A_{\gamma 3}\\ \left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 1} & \left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 2} & 1+\left(\phi^2-\phi^4\right)k^2A_{\gamma 3}A_{\gamma 3}\end{pmatrix}\begin{pmatrix}0 & \phi^2kA_{\gamma 1}\\ 0 & \phi^2kA_{\gamma 2}\\ 0 & \phi^2kA_{\gamma 3}\end{pmatrix}\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}, \tag{4.9}$$
$$=\begin{pmatrix}-\phi^3kA_{\gamma 1}-\left(\phi^2-\phi^4\right)\phi^3k^3A_{\gamma 1}A_{\gamma k}A_{\gamma k} & \phi^2kA_{\gamma 1}+\left(\phi^2-\phi^4\right)\phi^2k^3A_{\gamma 1}A_{\gamma k}A_{\gamma k}\\ -\phi^3kA_{\gamma 2}-\left(\phi^2-\phi^4\right)\phi^3k^3A_{\gamma 2}A_{\gamma k}A_{\gamma k} & \phi^2kA_{\gamma 2}+\left(\phi^2-\phi^4\right)\phi^2k^3A_{\gamma 2}A_{\gamma k}A_{\gamma k}\\ -\phi^3kA_{\gamma 3}-\left(\phi^2-\phi^4\right)\phi^3k^3A_{\gamma 3}A_{\gamma k}A_{\gamma k} & \phi^2kA_{\gamma 3}+\left(\phi^2-\phi^4\right)\phi^2k^3A_{\gamma 3}A_{\gamma k}A_{\gamma k}\end{pmatrix}=\begin{pmatrix}-\phi^3kA_{\gamma 1} & \phi^2kA_{\gamma 1}\\ -\phi^3kA_{\gamma 2} & \phi^2kA_{\gamma 2}\\ -\phi^3kA_{\gamma 3} & \phi^2kA_{\gamma 3}\end{pmatrix}$$

again using $A_{\gamma k}A_{\gamma k}=0$. We can likewise calculate $-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}$ in the upper-right block in (4.2), but it is easier and entirely equivalent to simply use the transposition symmetry $G_{\mathrm{MN}}=G_{\mathrm{NM}}$ of the metric tensor and the result in (4.9) to deduce:

$$-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}=\begin{pmatrix}-\phi^3kA_{\gamma 1} & -\phi^3kA_{\gamma 2} & -\phi^3kA_{\gamma 3}\\ \phi^2kA_{\gamma 1} & \phi^2kA_{\gamma 2} & \phi^2kA_{\gamma 3}\end{pmatrix}, \tag{4.10}$$

For the upper left block in (4.2) we use **B** in (4.3), with (4.4) and (4.9) to calculate:

$$\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1}$$
$$=\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}+\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}\begin{pmatrix}0 & 0 & 0\\ \phi^2kA_{\gamma 1} & \phi^2kA_{\gamma 2} & \phi^2kA_{\gamma 3}\end{pmatrix}\begin{pmatrix}\phi^3kA_{\gamma 1} & -\phi^2kA_{\gamma 1}\\ \phi^3kA_{\gamma 2} & -\phi^2kA_{\gamma 2}\\ \phi^3kA_{\gamma 3} & -\phi^2kA_{\gamma 3}\end{pmatrix}, \tag{4.11}$$
$$=\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}+\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}\begin{pmatrix}0 & 0\\ \phi^5k^2A_{\gamma k}A_{\gamma k} & -\phi^4k^2A_{\gamma k}A_{\gamma k}\end{pmatrix}=\begin{pmatrix}1+\phi^2 & -\phi\\ -\phi & 1\end{pmatrix}$$

again using $A_{\gamma k}A_{\gamma k}=0$. And (4.8) already contains the complete lower-right block in (4.2).

So, we now reassemble (4.8) through (4.11) into (4.2) to obtain the complete inverse:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} 1+\phi^2 & -\phi & -\phi^3 kA_{\gamma 1} & -\phi^3 kA_{\gamma 2} & -\phi^3 kA_{\gamma 3} \\ -\phi & 1 & \phi^2 kA_{\gamma 1} & \phi^2 kA_{\gamma 2} & \phi^2 kA_{\gamma 3} \\ -\phi^3 kA_{\gamma 1} & \phi^2 kA_{\gamma 1} & -1-\left(\phi^2-\phi^4\right)k^2 A_{\gamma 1}A_{\gamma 1} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 1}A_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 1}A_{\gamma 3} \\ -\phi^3 kA_{\gamma 2} & \phi^2 kA_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 2}A_{\gamma 1} & -1-\left(\phi^2-\phi^4\right)k^2 A_{\gamma 2}A_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 2}A_{\gamma 3} \\ -\phi^3 kA_{\gamma 3} & \phi^2 kA_{\gamma 3} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 3}A_{\gamma 1} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 3}A_{\gamma 2} & -1-\left(\phi^2-\phi^4\right)k^2 A_{\gamma 3}A_{\gamma 3} \end{pmatrix} \tag{4.12}$$

Then we reorder rows and columns back to the $\mathrm{M} = 0,1,2,3,5$ sequence and connect this to the contravariant (inverse) metric tensor G^{MN} to write:

$$G^{\mathrm{MN}} = \begin{pmatrix} 1+\phi^2 & -\phi^3 kA_{\gamma 1} & -\phi^3 kA_{\gamma 2} & -\phi^3 kA_{\gamma 3} & -\phi \\ -\phi^3 kA_{\gamma 1} & -1-\left(\phi^2-\phi^4\right)k^2 A_{\gamma 1}A_{\gamma 1} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 1}A_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 1}A_{\gamma 3} & \phi^2 kA_{\gamma 1} \\ -\phi^3 kA_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 2}A_{\gamma 1} & -1-\left(\phi^2-\phi^4\right)k^2 A_{\gamma 2}A_{\gamma 2} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 2}A_{\gamma 3} & \phi^2 kA_{\gamma 2} \\ -\phi^3 kA_{\gamma 3} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 3}A_{\gamma 1} & -\left(\phi^2-\phi^4\right)k^2 A_{\gamma 3}A_{\gamma 2} & -1-\left(\phi^2-\phi^4\right)k^2 A_{\gamma 3}A_{\gamma 3} & \phi^2 kA_{\gamma 3} \\ -\phi & \phi^2 kA_{\gamma 1} & \phi^2 kA_{\gamma 2} & \phi^2 kA_{\gamma 3} & 1 \end{pmatrix}. \tag{4.13}$$

In a vitally-important contrast to the usual Kaluza-Klein G^{MN} in (1.1), this is manifestly *not singular*. This reverts to $\mathrm{diag}\left(G^{\mathrm{MN}}\right) = \mathrm{diag}\left(\eta^{\mathrm{MN}}\right) = (+1,-1,-1,-1,+1)$ when $A_{\gamma\mu} = 0$ and $\phi = 0$ which is exactly the same signature as G_{MN} in (3.11). Then we consolidate to the 3x3 form:

$$G^{\mathrm{MN}} = \begin{pmatrix} G^{00} & G^{0k} & G^{05} \\ G^{j0} & G^{jk} & G^{j5} \\ G^{50} & G^{5k} & G^{55} \end{pmatrix} = \begin{pmatrix} 1+\phi^2 & -\phi^3 kA_{\gamma k} & -\phi \\ -\phi^3 kA_{\gamma j} & \eta^{jk} - \left(\phi^2-\phi^4\right)k^2 A_{\gamma j}A_{\gamma k} & \phi^2 kA_{\gamma j} \\ -\phi & \phi^2 kA_{\gamma k} & 1 \end{pmatrix}. \tag{4.14}$$

Now, the photon gauge vectors $A_{\gamma j}$ in (4.14) still have lower indexes, and with good reason: We cannot simply raise these indexes of components *inside the metric tensor* at will as we might for any other tensor. Rather, we must use the metric tensor (4.14) itself to raise and lower indexes, by calculating $A_\gamma{}^{\mathrm{M}} = G^{\mathrm{MN}}A_{\gamma\mathrm{N}}$. Nonetheless, it would be desirable to rewrite the components of (4.14) with all upper indexes, which will simplify downstream calculations. Given that $A_{\gamma 0} = 0$ for the photon and taking $A_{\gamma 5} = 0$, and raising indexes for $A_\gamma{}^0$ and $A_\gamma{}^5$ while sampling $A_\gamma{}^1$ and once again employing $A_{\gamma k}A_{\gamma k} = 0$, we may calculate:

$$\begin{aligned} A_\gamma{}^0 &= G^{0\mathrm{N}}A_{\gamma\mathrm{N}} = G^{01}A_{\gamma 1} + G^{02}A_{\gamma 2} + G^{03}A_{\gamma 3} = -\phi^3 kA_{\gamma k}A_{\gamma k} = 0 \\ A_\gamma{}^1 &= G^{1\mathrm{N}}A_{\gamma\mathrm{N}} = G^{11}A_{\gamma 1} + G^{12}A_{\gamma 2} + G^{13}A_{\gamma 3} = -A_{\gamma 1} - \left(\phi^2 + \phi^4\right)k^2 A_{\gamma 1}A_{\gamma k}A_{\gamma k} = -A_{\gamma 1}\,, \\ A_\gamma{}^5 &= G^{5\mathrm{N}}A_{\gamma\mathrm{N}} = G^{51}A_{\gamma 1} + G^{52}A_{\gamma 2} + G^{53}A_{\gamma 3} = -\phi^2 kA_{\gamma k}A_{\gamma k} = 0 \end{aligned} \tag{4.15}$$

The middle result applies by rotational symmetry to other space indexes, so that:

$$A_\gamma{}^\mu = G^{\mu\nu}A_{\gamma\nu} = \eta^{\mu\nu}A_{\gamma\nu} \mapsto A_\gamma{}^\mu = g^{\mu\nu}A_{\gamma\nu}\,, \tag{4.16}$$

which is the usual way of raising indexes in flat spacetime, generalized to $g^{\mu\nu}$ with minimal coupling. As a result, with $g^{\mu\nu} = \eta^{\mu\nu}$ we may raise the index in (3.12) to obtain:

$$\Phi^\mu = \begin{pmatrix} \phi & \phi^2 kA_\gamma{}^j \end{pmatrix} = \begin{pmatrix} \phi & -\phi^2 kA_{\gamma j} \end{pmatrix}. \tag{4.17}$$

We then use (4.17) to write (4.14) as:

$$G^{\mathrm{MN}} = \begin{pmatrix} G^{00} & G^{0k} & G^{05} \\ G^{j0} & G^{jk} & G^{j5} \\ G^{50} & G^{5k} & G^{55} \end{pmatrix} = \begin{pmatrix} 1+\phi^2 & -\phi^3 kA_{\gamma k} & -\Phi^0 \\ -\phi^3 kA_{\gamma j} & \eta^{jk} - \left(\phi^2 - \phi^4\right)k^2 A_{\gamma j}A_{\gamma k} & -\Phi^j \\ -\Phi^0 & -\Phi^k & 1 \end{pmatrix}. \tag{4.18}$$

Now we focus on the middle term, expanded to $\eta^{jk} - \phi^2 k^2 A_{\gamma j}A_{\gamma k} + \phi^4 k^2 A_{\gamma j}A_{\gamma k}$. Working from (4.17) we now calculate:

$$\Phi^0\Phi^0 = \phi^2;\quad \Phi^0\Phi^k = -\phi^3 kA_{\gamma k};\quad \Phi^j\Phi^0 = -\phi^3 kA_{\gamma j};\quad \Phi^j\Phi^k = \phi^4 k^2 A_{\gamma j}A_{\gamma k}\,. \tag{4.19}$$

So, we use (4.19) in (4.18), and raise the indexes using $A_{\gamma j}A_{\gamma k} = A_\gamma{}^j A_\gamma{}^k$ from (2.16), to write:

$$G^{\mathrm{MN}} = \begin{pmatrix} G^{00} & G^{0k} & G^{05} \\ G^{j0} & G^{jk} & G^{j5} \\ G^{50} & G^{5k} & G^{55} \end{pmatrix} = \begin{pmatrix} 1+\Phi^0\Phi^0 & \Phi^0\Phi^k & -\Phi^0 \\ \Phi^j\Phi^0 & \eta^{jk} - \phi^2 k^2 A_\gamma{}^j A_\gamma{}^k + \Phi^j\Phi^k & -\Phi^j \\ -\Phi^0 & -\Phi^k & 1 \end{pmatrix}. \tag{4.20}$$

Then, again taking advantage of the fact that $A_{\gamma 0} = 0$, while using $1 = \eta_{00} = \eta^{00}$ and $1 = \eta_{55} = \eta^{55}$ we may consolidate this into the 2x2 format:

$$G^{\mathrm{MN}} = \begin{pmatrix} G^{\mu\nu} & G^{\mu 5} \\ G^{5\nu} & G^{55} \end{pmatrix} = \begin{pmatrix} \eta^{\mu\nu} - \Phi^0\Phi^0 k^2 A_\gamma{}^\mu A_\gamma{}^\nu + \Phi^\mu\Phi^\nu & -\Phi^\mu \\ -\Phi^\nu & \eta^{55} \end{pmatrix}. \tag{4.21}$$

This is the inverse of (3.13) with $g_{\mu\nu} = \eta_{\mu\nu}$, and it is a good exercise to check and confirm that in fact, $G^{\mathrm{MA}} G_{\mathrm{AN}} = \delta^{\mathrm{M}}{}_{\mathrm{N}}$.

The final step is to apply minimal coupling to generalize $\eta^{\mathrm{MN}} \mapsto g^{\mathrm{MN}}$, with possible non-zero $g_{\mu 5}$, $g_{5\nu}$, $g^{\mu 5}$ and $g^{5\nu}$. With this last step, (4.21) now becomes:

$$\boxed{G^{\mathrm{MN}} = \begin{pmatrix} G^{\mu\nu} & G^{\mu 5} \\ G^{5\nu} & G^{55} \end{pmatrix} = \begin{pmatrix} g^{\mu\nu} - \Phi^0 \Phi^0 k^2 A_\gamma{}^\mu A_\gamma{}^\nu + \Phi^\mu \Phi^\nu & g^{\mu 5} - \Phi^\mu \\ g^{5\nu} - \Phi^\nu & g^{55} \end{pmatrix}}. \tag{4.22}$$

The above along with (3.13) are the direct counterparts to the Kaluza-Klein metric tensors (1.1). This inverse, in contrast to that of (1.1), is manifestly non-singular.

Finally, we commented after (2.6) that it would have been possible to choose minus rather than plus signs in the tetrad / field assignments. We make a note that had we done so, this would have carried through to a sign flip in all the $\varepsilon_k{}^0$ and $\varepsilon_0{}^k$ tetrad components in (2.12), it would have changed (2.14) to $\Gamma_\mu = \begin{pmatrix} \gamma_0 - \phi k A_{\gamma j} \gamma_j & \gamma_j - \phi k A_{\gamma j} \gamma_0 \end{pmatrix}$, and it would have changed (3.8) to include $\Gamma_5 = \gamma_5 - \phi \gamma_0$. Finally, for the metric tensors (4.22), all would be exactly the same, except that we would have had $G_{\mu 5} = G_{5\mu} = g_{\mu 5} - \Phi_\mu$ and $G^{\mu 5} = G^{5\mu} = g^{\mu 5} + \Phi^\mu$, with the vectors in (3.12) and (4.17) instead given by $\Phi_\mu = \begin{pmatrix} \phi & -\phi^2 k A_{\gamma j} \end{pmatrix}$ and $\Phi^\mu = \begin{pmatrix} \phi & -\phi^2 k A_\gamma{}^j \end{pmatrix}$. We note this because in a related preprint by the author at [16], this latter sign choice was required at [14.5] in a similar circumstance to ensure limiting-case solutions identical to those of Dirac's equation, as reviewed following [19.13] therein. Whether a similar choice may be required here cannot be known for certain without calculating detailed correspondences with Dirac theory based on the Γ_{M} in (3.8). In the next section, we will lay out the Dirac theory based on the Kaluza-Klein metric tensors having now been made generally-covariant in five dimensions.

5. The Dirac Equation with Five-Dimensional General Covariance

Now that we have obtained a Dirac-Kaluza-Klein metric tensor G_{MN} in (3.13) and its non-singular inverse G^{MN} in (4.22) which are fully covariant across all five dimensions and which are connected to a set of Dirac operators Γ_{M} deduced in (3.8) through the anticommutators (3.1), there are several additional calculations we shall perform which lay the foundation for deeper development. The first calculation, which vastly simplifies downstream calculation and provides the basis for a Dirac-type quantum theory of the electron and the photon based on Kaluza-Klein, is to obtain the contravariant (upper indexed) operators $\Gamma^{\mathrm{M}} = G^{\mathrm{MN}} \Gamma_{\mathrm{N}}$ in two component form which consolidates the four spacetime operators Γ^μ into a single four-covariant expression, then to do the same for the original Γ_{M} in (3.8).

As just noted, we may raise the indexes in the $\Gamma_{\rm M}$ of (3.8) by calculating $\Gamma^{\rm M} = G^{\rm MN}\Gamma_{\rm N}$. It is easiest to work from (3.8) together with the 3x3 form (4.20), then afterward consolidate to 2x2 form. So, we first calculate each of Γ^0, Γ^j and Γ^5 as such:

$$\begin{aligned}\Gamma^0 &= G^{0\rm N}\Gamma_{\rm N} = G^{00}\Gamma_0 + G^{0k}\Gamma_k + G^{05}\Gamma_5 \\ &= \left(1+\Phi^0\Phi^0\right)\left(\gamma_0 + \phi k A_{\gamma j}\gamma_j\right) + \Phi^0\Phi^k\left(\gamma_k + \phi k A_{\gamma k}\gamma_0\right) - \Phi^0\left(\gamma_5 + \phi\gamma_0\right), \\ &= \gamma^0 + \Phi^0 k A_\gamma{}^k\gamma^k + k A_\gamma{}^0\Phi^0\gamma^0 - \Phi^0\gamma^5\end{aligned} \tag{5.1a}$$

$$\begin{aligned}\Gamma^j &= G^{j\rm N}\Gamma_{\rm N} = G^{j0}\Gamma_0 + G^{jk}\Gamma_k + G^{j5}\Gamma_5 \\ &= \Phi^j\Phi^0\left(\gamma_0 + \phi k A_{\gamma k}\gamma_k\right) + \left(\eta^{jk} - \phi^2 k^2 A_\gamma{}^j A_\gamma{}^k + \Phi^j\Phi^k\right)\left(\gamma_k + \phi k A_{\gamma k}\gamma_0\right) - \Phi^j\left(\gamma_5 + \phi\gamma_0\right), \\ &= \gamma^j + \Phi^j k A_\gamma{}^k\gamma^k + k A_\gamma{}^j\Phi^0\gamma^0 - \Phi^j\gamma^5\end{aligned} \tag{5.1b}$$

$$\begin{aligned}\Gamma^5 &= G^{5\rm N}\Gamma_{\rm N} = G^{50}\Gamma_0 + G^{5k}\Gamma_k + G^{55}\Gamma_5 \\ &= -\Phi^0\left(\gamma_0 + \phi k A_{\gamma j}\gamma_j\right) - \Phi^k\left(\gamma_k + \phi k A_{\gamma k}\gamma_0\right) + \left(\gamma_5 + \phi\gamma_0\right) = \gamma^5.\end{aligned} \tag{5.1c}$$

To reduce the above, we have employed $\Phi^\mu = \begin{pmatrix}\phi & \phi^2 k A_\gamma{}^j\end{pmatrix}$ from (4.17) which implies that $\Phi^k A_{\gamma k} = 0$ via $A_k A_k = 0$ from (2.10). We have also used $A_\gamma{}^j = \eta^{jk}A_{\gamma k} = -A_{\gamma j}$ from (2.16), and the basic Dirac identities $\gamma^0 = \gamma_0$, $\gamma^k = \eta^{jk}\gamma_k = -\gamma_k$ and $\gamma^5 = \gamma_5$. We also include a term $k A_\gamma{}^0\Phi^0\gamma^0 = 0$ in (5.1a) to highlight the four-dimensional spacetime covariance with (5.1b), notwithstanding that this term is a zero because the gauge symmetry has been broken to that of a photon. Making use of this, we consolidate all of (5.1) above into the two-part:

$$\boxed{\Gamma^{\rm M} = \begin{pmatrix}\gamma^\mu + \Phi^\mu k A_\gamma{}^k\gamma^k + k A_\gamma{}^\mu\Phi^0\gamma^0 - \Phi^\mu\gamma^5 & \gamma^5\end{pmatrix}}. \tag{5.2}$$

As a final step to consolidate the Dirac matrices, we use the 2x2 consolidation of the metric tensor $G_{\rm MN}$ in (3.13), with $g_{\mu\nu} = \eta_{\mu\nu}$, to lower the indexes in (5.2) and obtain a two-part $\Gamma_{\rm M} = G_{\rm MN}\Gamma^{\rm N}$. Doing so we calculate:

$$\begin{aligned}\Gamma_\mu &= G_{\mu\rm N}\Gamma^{\rm N} = G_{\mu\nu}\Gamma^\nu + G_{\mu 5}\Gamma^5 \\ &= \left(\eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu}A_{\gamma\nu}\right)\left(\gamma^\nu + \Phi^\nu k A_\gamma{}^k\gamma^k + k A_\gamma{}^\nu\Phi^0\gamma^0 - \Phi^\nu\gamma^5\right) + \Phi_\mu\gamma^5, \\ &= \gamma_\mu + \Phi_\mu k A_{\gamma k}\gamma_k - \Phi_0\Phi_0 k^2 A_{\gamma\mu}A_{\gamma k}\gamma_k + k A_{\gamma\mu}\Phi_0\gamma_0\end{aligned} \tag{5.3a}$$

$$\begin{aligned}\Gamma_5 &= G_{5\rm N}\Gamma^{\rm N} = G_{5\nu}\Gamma^\nu + G_{55}\Gamma^5 \\ &= \Phi_\nu\left(\gamma^\nu + \Phi^\nu k A_\gamma{}^k\gamma^k + k A_\gamma{}^\nu\Phi^0\gamma^0 - \Phi^\nu\gamma^5\right) + \left(1+\Phi_0\Phi_0\right)\gamma^5. \\ &= \gamma_5 + \Phi_0\gamma_0\end{aligned} \tag{5.3b}$$

Above, we use the same reductions employed in (5.1), as well as $A_{\gamma\nu}A_{\gamma}{}^{\nu}=0$, $A_{\gamma\nu}\Phi^{\nu}=0$ and $\Phi_{\nu}\Phi^{\nu}=\phi^2$. We then consolidate this into the two-part:

$$\boxed{\Gamma_{\text{M}}=\left(\gamma_{\mu}+\left(\Phi_{\mu}-\Phi_0\Phi_0 kA_{\gamma\mu}\right)kA_{\gamma k}\gamma_k+kA_{\gamma\mu}\Phi_0\gamma_0\quad\gamma_5+\Phi_0\gamma_0\right)}. \tag{5.4}$$

Making use of $\Phi_{\mu}\equiv\left(\phi\quad\phi^2 kA_{\gamma j}\right)$ in (3.12), again mindful that $A_{\gamma\mu}=0$, and noting that $\Phi_{\mu}-\Phi_0\Phi_0 kA_{\gamma\mu}=\Phi_0=\phi$ for the $\mu=0$ time component and $\Phi_{\mu}-\Phi_0\Phi_0 kA_{\gamma\mu}=\Phi_k-\phi^2 kA_{\gamma k}=0$ for the $\mu=k$ space components, it is a good exercise to confirm that (5.4) does reduce precisely to $\Gamma_{\text{M}}=\left(\gamma_0+\phi kA_{\gamma k}\gamma_k\quad\gamma_j+\phi kA_{\gamma j}\gamma_0\quad\gamma_5+\phi\gamma_0\right)$ obtained in (3.8). Using (5.2) and (5.4) and reducing with $\Phi^{\mu}=\left(\phi\quad\phi^2 kA_{\gamma}{}^{j}\right)$, $\gamma^k\gamma_0=-\gamma_0\gamma^k$, $A_{\gamma j}\gamma_j A_{\gamma}{}^{k}\gamma^k=0$, $\Phi_{\mu}A_{\gamma}{}^{\mu}=0$ and $A_{\gamma\mu}A_{\gamma}{}^{\mu}=0$, it is also a good exercise to confirm that:

$$\Gamma_{\text{M}}\Gamma^{\text{M}}=\gamma_{\text{M}}\gamma^{\text{M}}=5\,. \tag{5.5}$$

And, it is a good exercise to confirm that (5.4) and (5.2) used in (1.3), see also (3.1), respectively reproduce the covariant and contravariant metric tensors (3.13) and (4.22).

Finally, having the upper-indexed (5.2) enables us to extend the Dirac equation governing fermion behavior into all five of the Kaluza-Klein dimensions, in the form of:

$$\boxed{\left(i\hbar c\Gamma^{\text{M}}\partial_{\text{M}}-mc^2\right)\Psi=0}. \tag{5.6}$$

If we then define a five-dimensional energy-momentum vector $cp^{\text{M}}=\left(cp^{\mu}\quad cp^5\right)$ containing the usual four-dimensional $cp^{\mu}=\left(E\quad c\mathbf{p}\right)$, and given that (3.13) and (4.22) provide the means to lower and raise indexes at will, we may further define the wavefunction $\Psi\equiv U_0\left(p^{\Sigma}\right)\exp\left(-ip_{\Sigma}x^{\Sigma}/\hbar\right)$ to include a Fourier kernel $\exp\left(-ip_{\Sigma}x^{\Sigma}/\hbar\right)$ over all five dimensions $x^{\Sigma}=\left(ct^0\quad\mathbf{x}\quad ct^5\right)$. These coordinates now include a *timelike* $x^5=ct^5$ which is heretofore distinguished from the ordinary time dimension $x^0=ct^0$ because as earlier reviewed, (3.13) has the tangent-space signature $\text{diag}\left(G_{\text{MN}}\right)=\left(+1,-1,-1,-1,+1\right)$. And $U_0\left(p^{\Sigma}\right)$ is a Dirac spinor which is now a function of all five components of p^{Σ} but independent of the coordinates x^{Σ}. In other words, $\partial_{\text{M}}U_0\left(p^{\Sigma}\right)=0$, which is why we include the 0 subscript. With all of this, we can convert (5.6) from configuration space to momentum space in the usual way, to obtain:

$$\boxed{\left(\Gamma^{\text{M}}cp_{\text{M}}-mc^2\right)U_0\left(p^{\Sigma}\right)=0}. \tag{5.7}$$

It is important to note that it is *not possible* to obtain the Dirac-type equations (5.6) and (5.7) from the usual Kaluza-Klein metric tensor and inverse (1.1), precisely because this metric

tensor is *not generally-covariant* across all five dimensions. And in fact, as we first deduced at (2.10), the Kaluza-Klein (1.1) are not even truly-covariant in the four spacetime dimensions alone unless we set the gauge field $A_\mu \mapsto A_{\gamma\mu}$ to that of a photon with only two transverse degrees of freedom. Of course, we do not at this juncture know precisely how to understand the fifth component cp^5 of the energy momentum or the second time dimension $x^5 = ct^5$. But it is the detailed development and study of the Dirac-Kaluza-Klein (DKK) equations (5.6) and (5.7) which may provide one set of avenues for understanding precisely how the energy cp^5 and the time t^5 are manifest in the natural world.

6. The Dirac-Kaluza-Klein Metric Tensor Determinant and Inverse Determinant

It is also helpful to calculate the metric tensor determinants. These are needed in a variety of settings, for example, to calculate the five-dimensional Einstein-Hilbert action, see e.g. [17], which expressly contains the determinant as part of the volume element $\sqrt{-g}\,\mathrm{d}^4x$ in four dimensions and which we anticipate will appear as $\sqrt{-G}\,\mathrm{d}^5x$ in five dimensions. As we shall later elaborate in section 23, the Einstein-Hilbert action provides what is perhaps the most direct path for understanding the fifth dimension as a "matter" dimension along the lines long-advocated by the 5D Space-Time-Matter Consortium [18]. Moreover, the Einstein-Hilbert action, from which the Einstein equation is also derived as reviewed in [17], is also essential for calculating quantum mechanical path integrals which would effectively provide a quantum field theory of gravitation in five-dimensions. For all these reasons, it is helpful to have obtained this determinant.

To calculate the determinant, we employ the block calculation method reviewed, e.g., at [19]. Specifically, for an invertible matrix which we have shown G_{MN} to be via G^{MN} in (4.22), the determinant is calculated with:

$$\left|G_{\mathrm{MN}}\right| = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \left|\mathbf{A}\right|\left|\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right|, \tag{6.1}$$

using the exact same blocks specified in (4.3) to calculate (4.2). Keep in mind that the blocks in (4.3) are based on having used what we now understand to be the tangent Minkowski-space $g_{\mathrm{MN}} = \eta_{\mathrm{MN}}$. As we found following (4.3), $\left|\mathbf{A}\right| = 1+\phi^2-\phi^2 = 1$, so (6.1) simplifies to $\left|G_{\mathrm{MN}}\right| = \left|\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right|$. Moreover, we already found $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ in (4.6). So, all that we need do is calculate the determinant of this 3x3 matrix, and we will have obtained $\left|G_{\mathrm{MN}}\right|$.

From (4.6) which we denote as the matrix $m_{ij} \equiv \mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$, we write out the full determinant, substitute (4.6), then reduce to obtain:

$$\begin{aligned} \left|m_{ij}\right| &= m_{11}m_{22}m_{33} + m_{12}m_{23}m_{31} + m_{13}m_{21}m_{32} - m_{13}m_{22}m_{31} - m_{12}m_{21}m_{33} - m_{11}m_{23}m_{32} \\ &= -1 + \left(\phi^2-\phi^4\right)k^2\left(A_{\gamma 1}A_{\gamma 1} + A_{\gamma 2}A_{\gamma 2} + A_{\gamma 3}A_{\gamma 3}\right) = -1 \end{aligned}. \tag{6.2}$$

Most of the terms cancel identically because of the equal number of + and – signs in the top line of (6.2). The only remaining term besides –1 itself, contains $A_{\gamma j}A_{\gamma j}=0$, which is zero because of (2.10) which removed two degrees of freedom from the gauge field and turned it into $A_\mu = A_{\gamma\mu}$ for a massless, luminous photon. So, we conclude, neatly, that $\left|\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right|=-1$, and because $\left|\mathbf{A}\right|=1$, that $\left|G_{\rm MN}\right|=-1=\left|\eta_{\rm MN}\right|$. Moreover, because $\left|M^{-1}\right|=\left|M\right|^{-1}$ for any square matrix, we likewise conclude that $\left|G^{\rm MN}\right|=-1=\left|\eta^{\rm MN}\right|$. Then, because the blocks in (4.3) are based on having used $g_{\rm MN}=\eta_{\rm MN}$, we may employ minimal coupling to generalize from $\eta_{\rm MN}\mapsto g_{\rm MN}$, so that the complete five-dimensional determinant and its inverse are:

$$\boxed{G\equiv\left|G_{\rm MN}\right|=\left|g_{\rm MN}\right|\equiv g;\quad G^{-1}\equiv\left|G^{\rm MN}\right|=\left|g^{\rm MN}\right|\equiv g^{-1}}. \tag{6.3}$$

In the above, the massless, luminous $A_\mu = A_{\gamma\mu}$ and the scalar field ϕ wash entirely out of the determinant, leaving the determinants entirely dependent upon $g_{\rm MN}$ which accounts for all curvatures *other than* those produced by $A_{\gamma\mu}$ and ϕ.

For the determinant of the four-dimensional spacetime components $G_{\mu\nu}$ alone, we employ the exact same calculation used in (6.1), but now we split $G_{\mu\nu}$ into a 1x1 time "block" with $\mathbf{A}=\left|\mathbf{A}\right|=1$, a 3x3 space block with the same $\mathbf{D}=\eta_{\mu\nu}+\phi^2k^2A_{\gamma\mu}A_{\gamma\nu}$, and the 1x3 and 3x1 blocks $\mathbf{B}=0$ and $\mathbf{C}=0$. So (6.1) becomes $\left|G_{\mu\nu}\right|=\left|\mathbf{A}\right|\left|\mathbf{D}\right|=\left|\mathbf{D}\right|$. We next note that $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ in (4.6) differs from $\mathbf{D}$ in (4.3) merely by the term $-\phi^4k^2A_{\gamma\mu}A_{\gamma\nu}$, which tells us that the calculation of $\left|\mathbf{D}\right|$ will produce the exact same result as (6.2) leading to $\left|G_{\mu\nu}\right|=-1=\left|\eta_{\mu\nu}\right|$, with the inverse following suit. Consequently, after generalizing $\eta_{\mu\nu}\mapsto g_{\mu\nu}$ via minimal coupling, we find that for the four dimensions of spacetime alone:

$$\left|G_{\mu\nu}\right|=\left|g_{\mu\nu}\right|;\quad\left|G^{\mu\nu}\right|=\left|g^{\mu\nu}\right|. \tag{6.4}$$

Here too, the massless, luminous $A_\mu = A_{\gamma\mu}$ with two degrees of freedom and the scalar ϕ are washed out entirely. Note, comparing (6.3) and (6.4), that we have reserved the notational definitions $G\equiv\left|G_{\rm MN}\right|$ and $g=\left|g_{\rm MN}\right|$ for the *five-dimensional* determinants. In four dimensions, we simply use the spacetime indexes to designate that (6.4) represents the four-dimensional spacetime subset of the five-dimensional metric tensor determinant and inverse.

7. The Dirac-Kaluza-Klein Lorentz Force Motion

Kaluza-Klein theory which will celebrate its centennial next year, has commanded attention for the past century for the very simple reason that despite all of its difficulties (most of which as will be reviewed in section 11 arise directly or indirectly from the degeneracy of the

metric tensor (1.1) and its lack of five-dimensional covariance at the Dirac level) because it successfully explains Maxwell's equations, the Lorentz Force motion and the Maxwell stress-energy tensor on an entirely geometrodynamic foundation. This successful geometrodynamic representation of Maxwell's electrodynamics – popularly known as the "Kaluza miracle" – arises particularly from the components $G_{\mu 5} = G_{5\mu} = \phi^2 k A_\mu$ of the metric tensor (1.1), because the electromagnetic field strength $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is among the objects which appear in the five-dimensional Christoffel connections $\tilde{\Gamma}^{\rm M}_{\rm AB}$ (particularly in $\tilde{\Gamma}^{\mu}_{\alpha 5}$ as we shall now detail), and because these $F^{\mu\nu}$ then make their way into the geodesic equation of motion in a form that can be readily connected to the Lorentz Force motion, and because they also enter the Einstein field equation in a form that can be likewise connected to the Maxwell stress-energy tensor. Therefore, it is important to be assured that in the process of remediating the various difficulties of Kaluza-Klein's metric tensor (1.1), the 5-covariant metric tensor (3.13) does not sacrifice any of the Kaluza miracle in the process.

In (3.13), $G_{\mu 5} = G_{5\mu} = \phi^2 k A_\mu$ from (1.1) which are responsible for the Kaluza miracle are replaced by $G_{\mu 5} = G_{5\mu} = g_{\mu 5} + \Phi_\mu$. For a flat Minkowski tangent space $g_{\rm MN} = \eta_{\rm MN}$ these reduce to $G_{\mu 5} = G_{5\mu} = \Phi_\mu$. At (3.4) we *required* $G_{j5} = G_{5j} = \phi^2 k A_{\gamma j}$ to precisely match $G_{\rm MN}$ from the Kaluza-Klein metric (1.1), maintaining the same spacetime covariance as $G_{\mu 5} = G_{5\mu}$ in (1.1) because $\phi^2 k A_{\gamma 0} = 0$, to keep the "miracle" intact. So, for a five-dimensional metric defined by:

$$c^2 d\mathrm{T}^2 \equiv G_{\rm MN} dx^{\rm M} dx^{\rm N} \tag{7.1}$$

the equation of motion obtained by minimizing the geodesic variation is:

$$\frac{d^2 x^{\rm M}}{c^2 d\mathrm{T}^2} = -\tilde{\Gamma}^{\rm M}_{\rm AB} \frac{dx^{\rm A}}{cd\mathrm{T}} \frac{dx^{\rm B}}{cd\mathrm{T}} = -\tilde{\Gamma}^{\rm M}_{\alpha\beta} \frac{dx^{\alpha}}{cd\mathrm{T}} \frac{dx^{\beta}}{cd\mathrm{T}} - 2\tilde{\Gamma}^{\rm M}_{\alpha 5} \frac{dx^{\alpha}}{cd\mathrm{T}} \frac{dx^{5}}{cd\mathrm{T}} - \tilde{\Gamma}^{\rm M}_{55} \frac{dx^{5}}{cd\mathrm{T}} \frac{dx^{5}}{cd\mathrm{T}} \tag{7.2}$$

just as in Kaluza-Klein theory, with connections of the "first" and "second" kinds specified by:

$$\begin{aligned} \tilde{\Gamma}_{\Sigma\,\rm AB} &= \tfrac{1}{2}\left(\partial_{\rm B} G_{\Sigma\rm A} + \partial_{\rm A} G_{\rm B\Sigma} - \partial_{\Sigma} G_{\rm AB}\right); \\ \tilde{\Gamma}^{\rm M}_{\rm AB} &= \tfrac{1}{2} G^{\rm M\Sigma}\left(\partial_{\rm B} G_{\Sigma\rm A} + \partial_{\rm A} G_{\rm B\Sigma} - \partial_{\Sigma} G_{\rm AB}\right) = G^{\rm M\Sigma}\tilde{\Gamma}_{\Sigma\,\rm AB} \end{aligned}, \tag{7.3}$$

likewise, just as in Kaluza-Klein theory. One may multiply (7.2) through by $d\mathrm{T}^2 / d\tau^2$ to obtain:

$$\frac{d^2 x^{\rm M}}{c^2 d\tau^2} = -\tilde{\Gamma}^{\rm M}_{\rm AB} \frac{dx^{\rm A}}{cd\tau} \frac{dx^{\rm B}}{cd\tau} = -\tilde{\Gamma}^{\rm M}_{\alpha\beta} \frac{dx^{\alpha}}{cd\tau} \frac{dx^{\beta}}{cd\tau} - 2\tilde{\Gamma}^{\rm M}_{\alpha 5} \frac{dx^{\alpha}}{cd\tau} \frac{dx^{5}}{cd\tau} - \tilde{\Gamma}^{\rm M}_{55} \frac{dx^{5}}{cd\tau} \frac{dx^{5}}{cd\tau} \tag{7.4}$$

which is the equation of motion with regard to the ordinary invariant spacetime metric line element $d\tau$, in which this four-dimensional proper time is defined by:

$$c^2 d\tau^2 \equiv G_{\mu\nu} dx^\mu dx^\nu = g_{\mu\nu} dx^\mu dx^\nu + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu} dx^\mu dx^\nu . \tag{7.5}$$

The space acceleration with regard to proper time τ is then given by $d^2x^j / d\tau^2$ with $\mathrm{M} = j = 1,2,3$ in (7.4). And if we then multiply this through by $d\tau^2 / dt^{0\ 2}$ (mindful again that we now need to distinguish dt^0 from the second time dimension dt^5), we obtain the space acceleration $d^2x^j / dt^{0\ 2}$ with regard to the ordinary time coordinate.

The above (7.1) through (7.5) are exactly the same as their counterparts in Kaluza-Klein theory, and they are exactly the same as what is used in the General Theory of Relativity in four spacetime dimensions alone, aside from minor notational changes intended to distinguish four- from five-dimensional objects. The only difference is that Kaluza-Klein theory uses the metric tensor (1.1) which has a spacelike fifth dimension, while the present DKK theory uses the metric tensor (3.13) which as a timelike fifth dimension. But the main reasons we are reviewing the equation of five-dimensional motion (7.4) is to be assured that the Kaluza miracle is not compromised by using the different metric tensor (3.13) rather than the usual (1.1).

As noted above, the connections $\tilde{\Gamma}^{\mathrm{M}}_{\alpha 5}$ are the particular ones responsible for the Kaluza-Klein representation of electrodynamics, whereby $\tilde{\Gamma}^{\mu}_{\alpha 5}$ governs accelerations in the four spacetime dimensions and $\tilde{\Gamma}^{5}_{\alpha 5}$ governs the fifth-dimensional acceleration. So, let's examine $\tilde{\Gamma}^{\mu}_{\alpha 5}$ more closely. Using (3.13) and (4.22) in (7.3) along with the symmetric $G_{\mathrm{MN}} = G_{\mathrm{NM}}$ we obtain:

$$\begin{aligned}
&\tilde{\Gamma}^{\mu}_{\alpha 5} = \tfrac{1}{2} G^{\mu\Sigma} \left(\partial_5 G_{\Sigma\alpha} + \partial_\alpha G_{5\Sigma} - \partial_\Sigma G_{\alpha 5} \right) \\
&= \tfrac{1}{2} G^{\mu\sigma} \left(\partial_5 G_{\sigma\alpha} + \partial_\alpha G_{5\sigma} - \partial_\sigma G_{\alpha 5} \right) + \tfrac{1}{2} G^{\mu 5} \partial_\alpha G_{55} \\
&= \tfrac{1}{2} \left(g^{\mu\sigma} - \Phi^0 \Phi^0 k^2 A_\gamma{}^\mu A_\gamma{}^\sigma + \Phi^\mu \Phi^\sigma \right) \left(\partial_5 \left(g_{\sigma\alpha} + \Phi_0 \Phi_0 k^2 A_{\gamma\sigma} A_{\gamma\alpha} \right) + \partial_\alpha \left(g_{5\sigma} + \Phi_\sigma \right) - \partial_\sigma \left(g_{\alpha 5} + \Phi_\alpha \right) \right) \\
&+ \tfrac{1}{2} \left(g^{\mu 5} - \Phi^\mu \right) \partial_\alpha \left(g_{55} + \Phi_0 \Phi_0 \right)
\end{aligned} . \tag{7.6}$$

For a flat tangent space $G_{\mathrm{MN}} = \eta_{\mathrm{MN}}$ with $\mathrm{diag}\left(\eta_{\mathrm{MN}}\right) = \left(+1,-1,-1,-1,+1\right)$ thus $\partial_\alpha G_{\mathrm{MN}} = 0$ this simplifies to:

$$\tilde{\Gamma}^{\mu}_{\alpha 5} = \tfrac{1}{2} \left(\eta^{\mu\sigma} - \Phi^0 \Phi^0 k^2 A_\gamma{}^\mu A_\gamma{}^\sigma + \Phi^\mu \Phi^\sigma \right) \left(\partial_5 \left(\Phi_0 \Phi_0 k^2 A_{\gamma\sigma} A_{\gamma\alpha} \right) + \partial_\alpha \Phi_\sigma - \partial_\sigma \Phi_\alpha \right) - \tfrac{1}{2} \Phi^\mu \partial_\alpha \left(\Phi_0 \Phi_0 \right) . \tag{7.7}$$

What is of special interest in (7.7) is the antisymmetric tensor term $\partial_\alpha \Phi_\sigma - \partial_\sigma \Phi_\alpha$, because this is responsible for an electromagnetic field strength $F_{\gamma\mu\nu} = \partial_\mu A_{\gamma\nu} - \partial_\nu A_{\gamma\mu}$. To see this, we rewrite (3.12) as:

$$\Phi_\mu = \left(\phi + \phi^2 k A_{\gamma 0} \quad \phi^2 k A_{\gamma j} \right) , \tag{7.8}$$

again taking advantage of $A_{\gamma 0} = 0$ to display the spacetime covariance of $A_{\gamma\mu}$. We then calculate the antisymmetric tensor in (7.7) in two separate bivector parts, as follows:

$$\begin{aligned}\partial_0\Phi_k-\partial_k\Phi_0&=\partial_0\left(\phi^2kA_{\gamma k}\right)-\partial_k\left(\phi+\phi^2kA_{\gamma 0}\right)\\&=\phi^2k\left(\partial_0A_{\gamma k}-\partial_kA_{\gamma 0}\right)+2\phi k\left(A_{\gamma k}\partial_0-A_{\gamma 0}\partial_k\right)\phi-\partial_k\phi\,,\\&=\phi^2kF_{\gamma 0k}-2\phi k\left(A_{\gamma 0}\partial_k-A_{\gamma k}\partial_0\right)\phi-\partial_k\phi\end{aligned}\tag{7.9a}$$

$$\begin{aligned}\partial_j\Phi_k-\partial_k\Phi_j&=\partial_j\left(\phi^2kA_{\gamma k}\right)-\partial_k\left(\phi^2kA_{\gamma j}\right)\\&=\phi^2k\left(\partial_jA_{\gamma k}-\partial_kA_{\gamma j}\right)+2\phi k\left(A_{\gamma k}\partial_j-A_{\gamma j}\partial_k\right)\phi\,.\\&=\phi^2kF_{\gamma jk}-2\phi k\left(A_{\gamma j}\partial_k-A_{\gamma k}\partial_j\right)\phi\end{aligned}\tag{7.9b}$$

We see the emergence of the field strength tensor $F_{\gamma\mu\nu}=\partial_\mu A_{\gamma\nu}-\partial_\nu A_{\gamma\mu}$ in its usual Kaluza-Klein form $\phi^2kF_{\gamma\mu\nu}$, modified to indicate that this arises from taking $F_\gamma^{\ \mu\nu}$ for a photon $A_\gamma^{\ \nu}$, which is a point to which we shall return momentarily. The only term which bars immediately merging both of (7.9) in a generally-covariant manner is the gradient $-\partial_k\phi$ in the 0*k* components of (7.9a). For this, noting that with reversed indexes $\partial_j\Phi_0-\partial_0\Phi_j$ (7.9a) will produce a gradient $+\partial_j\phi$ in the *j*0 components, we define a four-component $I_\mu\equiv\begin{pmatrix}1 & \mathbf{0}\end{pmatrix}$ and use this to form:

$$\begin{pmatrix}0 & -\partial_k\phi\\ \partial_j\phi & \mathbf{0}\end{pmatrix}=\begin{pmatrix}0\\ \partial_j\phi\end{pmatrix}\begin{pmatrix}1 & \mathbf{0}\end{pmatrix}-\begin{pmatrix}1\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}0 & \partial_k\phi\end{pmatrix}=\partial_\mu\phi I_\nu-I_\mu\partial_\nu\phi=-\left(I_\mu\partial_\nu-I_\nu\partial_\mu\right)\phi\,.\tag{7.10}$$

We then use this to covariantly combine both of (7.9) into:

$$\begin{aligned}\partial_\mu\Phi_\nu-\partial_\nu\Phi_\mu&=\phi^2kF_{\gamma\mu\nu}-2\phi k\left(A_{\gamma\mu}\partial_\nu-A_{\gamma\nu}\partial_\mu\right)\phi-\left(I_\mu\partial_\nu-I_\nu\partial_\mu\right)\phi\\&=\phi^2kF_{\gamma\mu\nu}-\left(\left(I_\mu+2\phi kA_{\gamma\mu}\right)\partial_\nu-\left(I_\nu+2\phi kA_{\gamma\nu}\right)\partial_\mu\right)\phi\end{aligned}\tag{7.11}$$

The newly-appearing vector $I_\mu+2\phi kA_{\gamma\mu}=\begin{pmatrix}1 & 2\phi kA_{\gamma j}\end{pmatrix}$ which we represent by now removing $A_{\gamma 0}=0$, is itself of interest, because the breaking of the gauge symmetry in section 2 caused $A_{\gamma 0}=0$ to come out of the photon gauge vector which only has two transverse degrees of freedom. But in this new vector $\begin{pmatrix}1 & 2\phi kA_{\gamma j}\end{pmatrix}$, the removed $A_{\gamma 0}=0$ is naturally replaced by the number 1, which is then included along with the remaining photon components $A_{\gamma j}$ multiplied by $2\phi k$. Again, the very small constant *k* which Kaluza-Klein theory fixes to (1.2) has dimensions of charge/energy, ϕ is taken to be dimensionless, and so $2\phi kA_{\gamma j}$ is dimensionless as well. Compare also $\Phi_\mu=\begin{pmatrix}\phi & \phi^2kA_{\gamma j}\end{pmatrix}$, then observe that $\Phi_\mu+\phi^2kA_{\gamma\mu}=\phi\left(I_\mu+2\phi kA_{\gamma\mu}\right)$.

Most importantly, we now see in (7.11) that the field strength $F_{\gamma\mu\nu}$ which is needed for the Lorentz Force motion and the Maxwell tensor, does indeed emerge inside of $\tilde{\Gamma}^\mu_{\alpha 5}$ as seen in (7.7)

just as it does from the usual Kaluza-Klein metric tensor (1.1), with the identical coefficients. But there is one wrinkle: $F_\gamma^{\mu\nu}$ is the field strength of a *single photon*, not a general classical $F^{\mu\nu}$ sourced by a material current density $J^\nu = (\rho \quad \mathbf{J})$ with a gauge potential $A^\mu = (\phi \quad \mathbf{A})$ which can always be Lorentz-transformed into a rest frame with $A^\mu = (\phi_0 \quad \mathbf{0})$ with ϕ_0 being the proper potential (note: this is a different ϕ from the Kaluza-Klein ϕ). In contrast, the photon $A_\gamma^{\ \mu}$ in (2.11) can never be placed at rest because the photon is a luminous, massless field quantum.

However, this can be surmounted using gauge symmetry, while making note of Heaviside's intuitions half a century before gauge theory which led him to formulate Maxwell's original theory without what would later be understood as a gauge potential. Specifically, even though the gauge symmetry is broken for $A_\gamma^{\ \mu}$ and it is therefore impossible to *Lorentz* transform the luminous $A_\gamma^{\ \mu}$ into a classical potential $A^\mu = (\phi \quad \mathbf{A})$ which can be placed at rest, or even to gauge transform $A_\gamma^{\ \mu} \rightarrow A^\mu$ from a luminous to a material potential because its gauge has already been fixed, the same impossibility *does not apply to gauge transformations of* $F_\gamma^{\ \mu\nu} = \partial^\mu A_\gamma^{\ \nu} - \partial^\nu A_\gamma^{\ \mu}$ obtained from this $A_\gamma^{\ \mu}$. This is because $F_{\gamma\mu\nu} = \partial_\mu A_{\gamma\nu} - \partial_\nu A_{\gamma\mu}$ is an antisymmetric tensor which, as is well-known, is *invariant under gauge transformations* $qA_\mu \rightarrow qA'_\mu \equiv qA_\mu + \hbar c \partial_\mu \Lambda$, where q is an electric charge and $\Lambda(t, \mathbf{x})$ is an unobservable scalar gauge parameter. To review, if we gauge transform some $qF_{\mu\nu} = q\partial_{;[\mu} A_{\nu]} \rightarrow qF'_{\mu\nu} = q\partial_{;[\mu} A_{\nu]} + \hbar c \left[\partial_{;\mu}, \partial_{;\nu}\right]\Lambda = qF_{\mu\nu}$, the gauge transformation washes out because the commutator $\left[\partial_{;\mu}, \partial_{;\nu}\right]\Lambda = 0$ even in curved spacetime. This is because the covariant derivative of a scalar is the same as its ordinary derivative, so that the covariant derivative $\partial_{;\mu}\partial_{;\nu}\Lambda = \partial_{;\mu}\partial_\nu\Lambda = \partial_\mu\partial_\nu\Lambda - \Gamma^\sigma_{\mu\nu}\partial_\sigma\Lambda$, with a similar expression under $\mu \leftrightarrow \nu$ interchange, and because $\Gamma^\sigma_{\mu\nu} = \Gamma^\sigma_{\nu\mu}$ is symmetric under such interchange.

So even though we cannot Lorentz transform $A_\gamma^{\ \mu}$ into A^μ, and even though the gauge of $A_\gamma^{\ \mu}$ is fixed so we cannot even gauge transform $A_\gamma^{\ \mu}$ into A^μ, we may perform a gauge transformation $F_{\gamma\mu\nu} \rightarrow F_{\mu\nu}$ precisely because the field strength (which was central to Heaviside's formulation of Maxwell in terms of its bivectors **E** and **B**) is invariant with respect to the gauge that was fixed to the photon in (2.11) as a result of (2.10). Another way of saying this is that $F_{\gamma\mu\nu} = \partial_\mu A_{\gamma\nu} - \partial_\nu A_{\gamma\mu}$ for a photon has the exact same form as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ for a materially-sourced potential which can be placed at rest, and that $F_{\gamma\mu\nu}$ enters into Maxwell's equations in exactly the same form as $F_{\mu\nu}$. The difference is that $F_{\gamma\mu\nu}$ emerges in source-free electrodynamics where the source current $J^\nu = 0$ while $F_{\mu\nu}$ emerges when there is a non-zero $J^\nu \neq 0$.

So irrespective of this $A^\mu = A_\gamma^{\ \mu}$ symmetry breaking which arose from (2.10) to ensure Dirac-level covariance of the Kaluza-Klein metric tensor, the luminous photon fields $F_{\gamma\mu\nu}$ emerging in (7.7) via (7.11) can always be gauge-transformed using $F_\gamma^{\ \mu\nu} \rightarrow F^{\mu\nu}$ into the classical

field strength of a classical materially-sourced potential $A^\mu = \begin{pmatrix} \phi & \mathbf{A} \end{pmatrix}$. Moreover, once we gauge transform $F_\gamma^{\ \mu\nu} \rightarrow F^{\mu\nu}$, the classical field strength $F^{\mu\nu}$ will contain innumerably-large numbers of photons mediating electromagnetic interactions, and so will entirely swamp out the individual $A_\gamma^{\ \mu}$ which represent individual photons. This transformation of $F_\gamma^{\ \mu\nu} \rightarrow F^{\mu\nu}$ by taking advantage of gauge symmetry, following by drowning out the impacts of individual photons as against classical fields, is exactly what the author did in Sections 21 and 23 of [16] to obtain the empirically-observed lepton magnetic moments at [23.5] and [23.6] of that same paper.

So, we now substitute (7.11) with a gauge-transformed $F_{\gamma\mu\nu} \rightarrow F_{\mu\nu}$ into (7.7), to find that:

$$\begin{aligned}
\tilde{\Gamma}^\mu_{\alpha 5} &= \tfrac{1}{2}\left(\eta^{\mu\sigma} - \Phi^0\Phi^0 k^2 A_\gamma^{\ \mu} A_\gamma^{\ \sigma} + \Phi^\mu\Phi^\sigma\right)\phi^2 k F_{\alpha\sigma} \\
&+ \tfrac{1}{2}\left(\eta^{\mu\sigma} - \Phi^0\Phi^0 k^2 A_\gamma^{\ \mu} A_\gamma^{\ \sigma} + \Phi^\mu\Phi^\sigma\right)\partial_5\left(\Phi_0\Phi_0 k^2 A_{\gamma\sigma}A_{\gamma\alpha}\right) \\
&- \tfrac{1}{2}\left(\eta^{\mu\sigma} - \Phi^0\Phi^0 k^2 A_\gamma^{\ \mu} A_\gamma^{\ \sigma} + \Phi^\mu\Phi^\sigma\right)\left(\left(I_\alpha + 2\phi k A_{\gamma\alpha}\right)\partial_\sigma - \left(I_\sigma + 2\phi k A_{\gamma\sigma}\right)\partial_\alpha\right)\phi \\
&- \tfrac{1}{2}\Phi^\mu\partial_\alpha\left(\Phi_0\Phi_0\right)
\end{aligned} \qquad (7.12)$$

From here, further mathematical reductions are possible. First, we noted earlier that $i\hbar\partial_\alpha A_{\gamma\mu} = q_\alpha A_{\gamma\mu}$ for the photon field in (2.11), which we extend to five dimensions as $i\hbar\partial_{\mathrm{A}} A_{\gamma\mu} = q_{\mathrm{A}} A_{\gamma\mu}$ by appending a fifth dimension in the Fourier kernel in (2.11a) just as we did for the fermion wavefunction following (5.6). Thus, we find $i\hbar A_\gamma^{\ \sigma}\partial_5 A_{\gamma\sigma} = A_\gamma^{\ \sigma} q_5 A_{\gamma\sigma} = 0$ and so may set $A_\gamma^{\ \sigma}\partial_5 A_{\gamma\sigma} = 0$. For similar reasons, see (4.17) and recall that $A_{\gamma 0} = 0$, we set $\Phi^\sigma\partial_5 A_{\gamma\sigma} = 0$. We also clear any remaining $A_\gamma^{\ \sigma} A_{\gamma\sigma} = 0$ and $\Phi^\sigma A_{\gamma\sigma} = 0$, and use $A_\gamma^{\ \sigma} I_\sigma = 0$ because $A_{\gamma 0} = 0$. Next, because $A_{\gamma 0} = 0$, wherever there is a remaining $A_{\gamma\sigma}$ summed with an object with an upper σ index, we set $\sigma = k = 1,2,3$ to the space indexes only. We also use $\eta^{\mu\sigma} I_\sigma = \eta^{00} I_0 = 1$. And we substitute $\Phi^0 = \Phi_0 = \phi$ throughout. Again mindful that $i\hbar\partial_{\mathrm{A}} A_{\gamma\mu} = q_{\mathrm{A}} A_{\gamma\mu}$, we also use $A_\gamma^{\ j} = \eta^{jk} A_{\gamma k}$ from (4.16) to raise some indexes. Finally, we apply all remaining derivatives, separate out time and space components for any summed indexes still left except for in $F_{\alpha\sigma}$, and reconsolidate. The result is that strictly mathematically, (7.12) reduces to:

$$\begin{aligned}
\tilde{\Gamma}^\mu_{\alpha 5} &= \tfrac{1}{2}\left(\eta^{\mu\sigma} - \phi^2 k^2 A_\gamma^{\ \mu} A_\gamma^{\ \sigma} + \Phi^\mu\Phi^\sigma\right)\phi^2 k F_{\alpha\sigma} \\
&+ \tfrac{1}{2}\left(\eta^{\mu 0} + 2\eta^{\mu k} k A_{\gamma k}\phi - \Phi^\mu\phi\right)\partial_\alpha\phi \\
&- \tfrac{1}{2}\left(\eta^{\mu 0} + \Phi^\mu\phi\right)\left(I_\alpha + 2\phi k A_{\gamma\alpha}\right)\partial_0\phi \\
&- \tfrac{1}{2}\left(\eta^{\mu k} - \phi^2 k^2 A_\gamma^{\ \mu} A_\gamma^{\ k} + \Phi^\mu\phi^2 k A_\gamma^{\ k}\right)\left(I_\alpha + 2\phi k A_{\gamma\alpha}\right)\partial_k\phi \\
&+ \phi\partial_5\phi k^2 A_\gamma^{\ \mu} A_{\gamma\alpha} + \tfrac{1}{2}\phi^2 k^2\partial_5 A_\gamma^{\ \mu} A_{\gamma\alpha} + \tfrac{1}{2}\phi^2 k^2 A_\gamma^{\ \mu}\partial_5 A_{\gamma\alpha}
\end{aligned} \qquad (7.13)$$

Now, it is the upper μ index in $\tilde{\Gamma}^{\mu}_{\alpha 5}$ which, when used in the equation of motion (7.4), will determine the coordinate against which the acceleration is specified in relation to the proper time interval $d\tau$. So, we now separate (7.13) into its time and space components, as such:

$$\begin{aligned}\tilde{\Gamma}^{0}_{\alpha 5} &= \tfrac{1}{2}\left(\eta^{0\sigma} + \phi\Phi^{\sigma}\right)\phi^{2}kF_{\alpha\sigma} \\ &+ \tfrac{1}{2}\left(1-\phi^{2}\right)\partial_{\alpha}\phi - \tfrac{1}{2}\left(1+\phi^{2}\right)\left(I_{\alpha} + 2\phi kA_{\gamma\alpha}\right)\partial_{0}\phi - \tfrac{1}{2}\phi^{3}kA_{\gamma}{}^{k}\left(I_{\alpha} + 2\phi kA_{\gamma\alpha}\right)\partial_{k}\phi\end{aligned} \quad , \tag{7.14a}$$

$$\begin{aligned}\tilde{\Gamma}^{j}_{\alpha 5} &= \tfrac{1}{2}\left(\eta^{j\sigma} - \phi^{2}k^{2}A_{\gamma}{}^{j}A_{\gamma}{}^{\sigma} + \phi^{2}kA_{\gamma}{}^{j}\Phi^{\sigma}\right)\phi^{2}kF_{\alpha\sigma} \\ &+ \left(1-\tfrac{1}{2}\phi^{2}\right)kA_{\gamma}{}^{j}\phi\partial_{\alpha}\phi \\ &- \tfrac{1}{2}\phi^{2}kA_{\gamma}{}^{j}\left(I_{\alpha} + 2\phi kA_{\gamma\alpha}\right)\phi\partial_{0}\phi \\ &- \tfrac{1}{2}\left(\eta^{jk} - \left(\phi^{2}-\phi^{4}\right)k^{2}A_{\gamma}{}^{j}A_{\gamma}{}^{k}\right)\left(I_{\alpha} + 2\phi kA_{\gamma\alpha}\right)\partial_{k}\phi \\ &+ k^{2}A_{\gamma}{}^{j}A_{\gamma\alpha}\phi\partial_{5}\phi + \tfrac{1}{2}\phi^{2}k^{2}A_{\gamma\alpha}\partial_{5}A_{\gamma}{}^{j} + \tfrac{1}{2}\phi^{2}k^{2}A_{\gamma}{}^{j}\partial_{5}A_{\gamma\alpha}\end{aligned} \quad . \tag{7.14b}$$

It is noteworthy that all terms in (7.13) containing the fifth dimensional derivative $\partial_{5} = \partial/\partial x^{5} = \partial/c\partial t^{5}$ also contain $A_{\gamma}{}^{\mu}$ and so drop out entirely from (7.14a) because $A_{\gamma}{}^{0} = 0$.

Now, as previewed prior to (7.12), $A_{\gamma\alpha}$ is the field for a single photon, which is inconsequential in physical effect compared to $F_{\alpha\sigma}$ which has now been gauge-transformed to a classical electric and magnetic field bivector consisting of innumerable photons. This is to say, if there is some interaction occurring in a classical electromagnetic field, a single photon more, or a single photon less, will be entirely undetectable for that interaction, akin to a single drop of water in an ocean. Moreover, the constant k is very small, so that the dimensionless $kA_{\gamma\alpha}$ will be very small in relation to the numbers ± 1 contained in $\eta^{\mu\nu}$. With this in mind, we may set $A_{\gamma\alpha} \cong 0$ as an extraordinarily-close approximation to zero all terms which contain $A_{\gamma\alpha}$ in (7.14). This includes for (7.14a), only retaining $\Phi^{0} = \phi$ in $\phi\Phi^{\sigma}\phi^{2}kF_{\alpha\sigma} = \phi\Phi^{0}\phi^{2}kF_{\alpha 0}$. And in (7.14b) we further use $\eta^{jk}\partial_{k} = -\partial_{j}$. So now, both of (7.14) reduce to:

$$\tilde{\Gamma}^{0}_{\alpha 5} = \tfrac{1}{2}\left(1+\phi^{2}\right)\phi^{2}kF_{\alpha 0} + \tfrac{1}{2}\left(1-\phi^{2}\right)\partial_{\alpha}\phi - \tfrac{1}{2}\left(1+\phi^{2}\right)I_{\alpha}\partial_{0}\phi, \tag{7.15a}$$

$$\tilde{\Gamma}^{j}_{\alpha 5} = \tfrac{1}{2}\phi^{2}kF_{\alpha}{}^{j} + \tfrac{1}{2}I_{\alpha}\partial_{j}\phi. \tag{7.15b}$$

Contrasting, we see that the former contains $F_{\alpha 0}$ while the latter contains $F_{\alpha}{}^{j}$ with a raised index. To properly compare we need to carefully raise the time index in (7.15a). To do this, we recall from after (2.11) that $i\hbar\partial_{\alpha}A_{\mu} = q_{\alpha}A_{\mu}$, $A_{\gamma}{}^{\alpha}q_{\alpha} = 0$, and $A^{j}q_{j} = 0$, which also means that $A_{\gamma}{}^{\alpha}\partial_{\alpha} = 0$ and $A_{\gamma}{}^{j}\partial_{j} = 0$, thus $\Phi^{j}\partial_{j} = 0$ when ∂_{α} operates on $A_{\gamma\mu}$. Recall as well that

$A_\gamma{}^\sigma A_{\gamma\sigma} = 0$ and $\Phi^\sigma A_{\gamma\sigma} = 0$. So, working from $F_{\gamma\sigma\nu} = \partial_\sigma A_{\gamma\nu} - \partial_\nu A_{\gamma\sigma}$ for an individual photon and using (4.22) with $g^{\mu\nu} = \eta^{\mu\nu}$, we first obtain, without yet fully reducing:

$$F_\gamma{}^\mu{}_\nu = G^{\mu\sigma} F_{\gamma\sigma\nu} = G^{\mu\sigma}\partial_\sigma A_{\gamma\nu} - G^{\mu\sigma}\partial_\nu A_{\gamma\sigma} = \left(\eta^{\mu\sigma} + \Phi^\mu\Phi^\sigma\right)\partial_\sigma A_{\gamma\nu} - \left(\eta^{\mu\sigma} + \Phi^\mu\Phi^\sigma\right)\partial_\nu A_{\gamma\sigma}. \tag{7.16}$$

Then, extracting the electric field bivector we obtain the field strength with a raised time index:

$$\begin{aligned} F_\gamma{}^0{}_\nu &= \left(\eta^{0\sigma}\partial_\sigma A_{\gamma\nu} + \Phi^0\Phi^\sigma\partial_\sigma A_{\gamma\nu}\right) - \left(\eta^{0\sigma}\partial_\nu A_{\gamma\sigma} + \Phi^0\Phi^\sigma\partial_\nu A_{\gamma\sigma}\right) \\ &= \left(\partial_0 A_{\gamma\nu} + \Phi^0\Phi^0\partial_0 A_{\gamma\nu}\right) - \left(\partial_\nu A_{\gamma 0} + \Phi^0\Phi^0\partial_\nu A_{\gamma 0}\right) \\ &= \left(1+\phi^2\right)\left(\partial_0 A_{\gamma\nu} - \partial_\nu A_{\gamma 0}\right) = \left(1+\phi^2\right) F_{\gamma 0\nu} \end{aligned} \quad . \tag{7.17}$$

Using the gauge transformation $F_{\gamma\mu\nu} \to F_{\mu\nu}$ discussed prior to (7.12) to write this as $F_\alpha{}^0 = \left(1+\phi^2\right) F_{\alpha 0}$, then using this in (7.15a), now reduces the equation pair (7.15) to:

$$\tilde{\Gamma}^0_{\alpha 5} = \tfrac{1}{2}\phi^2 k F_\alpha{}^0 + \tfrac{1}{2}\left(1-\phi^2\right)\partial_\alpha\phi - \tfrac{1}{2}\left(1+\phi^2\right) I_\alpha \partial_0\phi, \tag{7.18a}$$

$$\tilde{\Gamma}^j_{\alpha 5} = \tfrac{1}{2}\phi^2 k F_\alpha{}^j + \tfrac{1}{2} I_\alpha \partial_j\phi. \tag{7.18b}$$

These clearly manifest general spacetime covariance between the $\frac{1}{2}\phi^2 k F_\alpha{}^0$ and $\frac{1}{2}\phi^2 k F_\alpha{}^j$ terms.

At this point we are ready to use the above in the equation of motion (7.4). Focusing on the motion contribution from the $\tilde{\Gamma}^{\mathrm{M}}_{\alpha 5}$ term, we first write (7.4) as:

$$\frac{d^2 x^{\mathrm{M}}}{c^2 d\tau^2} = -2\tilde{\Gamma}^{\mathrm{M}}_{\alpha 5}\frac{dx^\alpha}{cd\tau}\frac{dx^5}{cd\tau} + \ldots \tag{7.19}$$

with a reminder that we are focusing on this particular term out of the three terms in (7.4). We then separate this into time and space components and use (7.18) with $F_\alpha{}^\mu = -F^\mu{}_\alpha$ and $I_\alpha = \begin{pmatrix} 1 & \mathbf{0} \end{pmatrix}$. Importantly, we also use the differential chain rule on the ϕ terms. We thus obtain:

$$\begin{aligned} \frac{d^2 x^0}{c^2 d\tau^2} &= -2\tilde{\Gamma}^0_{\alpha 5}\frac{dx^\alpha}{cd\tau}\frac{dx^5}{cd\tau} + \ldots = -\left(\phi^2 k F_\alpha{}^0 + \left(1-\phi^2\right)\partial_\alpha\phi - \left(1+\phi^2\right) I_\alpha\partial_0\phi\right)\frac{dx^\alpha}{cd\tau}\frac{dx^5}{cd\tau} + \ldots \\ &= \phi^2 k\frac{dx^5}{cd\tau}F^0{}_\alpha\frac{dx^\alpha}{cd\tau} + 2\frac{dx^5}{cd\tau}\phi^2\frac{d\phi}{cd\tau} + \ldots \end{aligned} \tag{7.20a}$$

$$\frac{d^2x^j}{c^2d\tau^2} = -2\tilde{\Gamma}^j_{\alpha 5}\frac{dx^\alpha}{cd\tau}\frac{dx^5}{cd\tau} + ... = -\left(\phi^2 kF_\alpha{}^j + I_\alpha \partial_j \phi\right)\frac{dx^\alpha}{cd\tau}\frac{dx^5}{cd\tau}$$
$$= \phi^2 k\frac{dx^5}{cd\tau}F^j{}_\alpha\frac{dx^\alpha}{cd\tau} - \frac{dx^5}{cd\tau}\frac{dx^0}{dx^j}\frac{d\phi}{cd\tau} + ... \tag{7.20b}$$

In both of the above, for the scalar we find a *derivative along the curve*, $d\phi/cd\tau$. Note further that in (7.20b) this is multiplied by the inverse of $dx^j/dx^0 = v^j/c$ where $v^j = dx^j/dt^0$ is an ordinary space velocity with reference to the ordinary time t^0 (versus the fifth-dimensional t^5). In contrast, in (7.20a) the objects covariant with this velocity term simply turned into the number 1 via the chain rule. Given its context, we understand v^j to be the space velocity of the scalar ϕ.

This raises an important question and gives us our first piece of solid information about the *physical nature* of the Kaluza-Klein scalar ϕ: Without the $d\phi/cd\tau$ term (7.20) consolidate into $d^2x^\mu/c^2d\tau^2 = \left(\phi^2 kdx^5/cd\tau\right)F^\mu{}_\alpha dx^\alpha cd\tau$ following which we can make the usual "Kaluza miracle" association with the Lorentz Force law. However, with this term, if ϕ is a *material* field or particle which can be Lorentz transformed to a rest frame with $v^j = 0$, then we have a problem, because the latter term in (7.20b) will become infinite, causing the space acceleration to likewise become infinite. The only way to avoid this problem, is to understand the scalar ϕ as a *luminous* entity which travels at the speed of light and which can never be Lorentz transformed to a rest frame, just like the photon. More to the point in terms of scientific method: we know from observation that the Lorentz force does not become infinite nor does it exhibit any observable deviations from the form $d^2x^\mu/c^2d\tau^2 = \left(\phi^2 kdx^5/cd\tau\right)F^\mu{}_\alpha dx^\alpha cd\tau$. Therefore, we use this observational evidence in view of (7.20b) to deduce that ϕ must be luminous.

To implement this luminosity, we first write the four-dimensional spacetime metric for a luminous particle such as the photon, and now also the scalar ϕ, using mixed indexes, as $0 = d\tau^2 = dx^0 dx_0 + dx^j dx_j$. This easily is rewritten as $dx^0 dx_0 = -dx^j dx_j$ and then again as:

$$\frac{dx^0}{dx^j} = -\frac{dx_j}{dx_0}. \tag{7.21}$$

This is the term of interest in (7.20b). Now, we want to raise indexes on the right side of (7.21) but must do so with (3.13). Using $\Phi_0 = \phi$ and $g_{\mu\nu} = \eta_{\mu\nu}$ as well as $A_{\gamma 0} = 0$ and $A_\gamma{}^\mu = \eta^{\mu\nu}A_{\gamma\nu}$ from (4.16), we find:

$$dx_0 = G_{0\nu}dx^\nu = \left(\eta_{0\nu} + \phi^2 k^2 A_{\gamma 0}A_{\gamma\nu}\right)dx^\nu = \eta_{0\nu}dx^\nu = dx^0$$
$$dx_j = G_{j\nu}dx^\nu = \left(\eta_{j\nu} + \phi^2 k^2 A_{\gamma j}A_{\gamma\nu}\right)dx^\nu = -dx^j + \phi^2 k^2 A_\gamma{}^j A_\gamma{}^k dx^k \tag{7.22}$$

Using the above in (7.21) then yields the *luminous* particle relation:

$$\frac{dx^0}{dx^j} = \frac{dx^j}{dx^0} - \phi^2 k^2 A_\gamma{}^j A_\gamma{}^k \frac{dx^k}{dx^0} = \hat{u}^j - \phi^2 k^2 A_\gamma{}^j A_\gamma{}^k \hat{u}^k \,. \tag{7.23}$$

Above, we also introduce a unit vector $\hat{u}^j = dx^j / dx^0$ with $\hat{u}^j \hat{u}^j = 1$ pointing in the direction of the luminous propagation of ϕ.

Inserting (7.23) for a luminous scalar into (7.20b) then produces:

$$\frac{d^2 x^j}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F^j{}_\alpha \frac{dx^\alpha}{cd\tau} - \hat{u}^j \frac{dx^5}{cd\tau}\frac{d\phi}{cd\tau} + \frac{dx^5}{cd\tau}\phi^2 k^2 A_\gamma{}^j A_\gamma{}^k \hat{u}^k \frac{d\phi}{cd\tau} \tag{7.24}$$

As we did starting at (7.15) we then set $A_{\gamma\alpha} \cong 0$ because the gauge vector for a single photon will be swamped by the innumerable photons contained in the classical field strength $F^j{}_\alpha$. As a result, using (7.24), we find that (7.20) together now become:

$$\frac{d^2 x^0}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F^0{}_\alpha \frac{dx^\alpha}{cd\tau} + 2\frac{dx^5}{cd\tau}\phi^2 \frac{d\phi}{cd\tau} \tag{7.25a}$$

$$\frac{d^2 x^j}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F^j{}_\alpha \frac{dx^\alpha}{cd\tau} - \hat{u}^j \frac{dx^5}{cd\tau}\frac{d\phi}{cd\tau} \tag{7.25b}$$

In (7.25b), ϕ has now been made luminous.

Finally, we are ready to connect this to the Lorentz Force motion, which we write as:

$$\frac{d^2 x^\mu}{c^2 d\tau^2} = \frac{q}{mc^2} F^\mu{}_\alpha \frac{dx^\alpha}{cdt} \,. \tag{7.26}$$

We start with the space components in (7.25b) combined with $\mu = j$ in (7.26) and use these to *define* the association:

$$\frac{d^2 x^j}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F^j{}_\alpha \frac{dx^\alpha}{cd\tau} - \hat{u}^j \frac{dx^5}{cd\tau}\frac{d\phi}{cd\tau} \equiv \frac{q}{mc^2} F^j{}_\alpha \frac{dx^\alpha}{cdt} \,. \tag{7.27}$$

For the moment, let us ignore the term $d\phi / d\tau$ to which we shall shortly return, and focus on the term with $F^j{}_\alpha$. If this is to represent Lorentz motion insofar as the $F^j{}_\alpha$ terms, then factoring out common terms from both sides, we obtain the following relation and its inverse:

$$\boxed{\phi^2 k \frac{dx^5}{cd\tau} = \phi^2 k \frac{dt^5}{d\tau} = \frac{q}{mc^2}; \quad \frac{dx^5}{cd\tau} = \frac{dt^5}{d\tau} = \frac{q}{\phi^2 kmc^2}} \,. \tag{7.28}$$

This is why electric charge – and to be precise, the charge-to-mass ratio – is interpreted as "motion" through the fifth dimension. However, because of the timelike fifth dimension in the metric tensor (3.13), the charge-to-energy ratio of a charged material body is no longer interpreted as *spatial motion through an unseen fourth space dimension*. Rather, it is understood as a *rate of time flow in a second time dimension*.

Next, we substitute the above for $dx^5 / cd\tau$ in each of (7.25) and reduce to obtain:

$$\frac{d^2x^0}{c^2d\tau^2} = \frac{q}{mc^2}F^0{}_\alpha \frac{dx^\alpha}{cd\tau} + 2\frac{q}{kmc^2}\frac{d\phi}{cd\tau} \tag{7.29a}$$

$$\frac{d^2x^j}{c^2d\tau^2} = \frac{q}{mc^2}F^j{}_\alpha \frac{dx^\alpha}{cd\tau} - \frac{\hat{u}^j}{\phi^2}\frac{q}{kmc^2}\frac{d\phi}{cd\tau} \tag{7.29b}$$

This does indeed reproduce the Lorentz motion, except for the $d\phi / d\tau$ term in each. Now, because there is no observed deviation for the Lorentz motion, in order to minimize the physical impact of these final terms, one might suppose that the luminous ϕ is an extremely small field $\phi \cong 0$ with $d\phi / d\tau \cong 0$, but this is problematic for two reasons: First, if *k* turns out to be the extremely small ratio $k = \left(2 / c^2\right)\sqrt{G / k_e}$ given by (1.2) as it is in Kaluza-Klein theory – and there is no reason to believe that *k* will turn out otherwise here – then the 1/*k* in both of (7.29) is an extremely large coefficient, which means that $d\phi / d\tau$ would have to be even more extraordinarily small. Second, even if $d\phi / d\tau \cong 0$ in part because we make ϕ extremely small, the presence of $1/\phi^2$ in (7.29b) still causes a problem, because an extremely small $\phi \to 0$ implies an extremely large $1/\phi^2 \to \infty$. Ironically, the $1/\phi^2$ which causes $G^{\mathrm{MN}} \to \infty$ in the usual Kaluza-Klein metric tensor (1.1) – which problem was solved by the non-singular (4.22) – nevertheless still persists, because of its appearance in (7.29b). And it persists in the form of a very large yet unobserved impact on the physical, observable Lorentz motion. The only apparent way to resolve this, is to *require* that $d\phi / d\tau = 0$. If that is the case, then (7.29) both condense precisely into the Lorentz Force motion.

Now, on first appearance, the thought that $d\phi / d\tau = 0$ seems to suggest that ϕ must be a constant field with no gradient, which as pointed out in [11] imposes unwarranted constraints on the electromagnetic field, and which also defeats the purpose of a "field" if that field has to be constant. But in (7.29), $d\phi / d\tau$ is *not* a gradient nor is it a time derivative. Rather, it is a *derivative along the curve* with curvature specified by the metric tensor (2.15), and it is related to the four-gradient $\partial_\mu\phi$ by the chain rule $d\phi / d\tau = \left(\partial\phi / \partial x^\mu\right)\left(dx^\mu / d\tau\right) = \partial_\mu\phi u^\mu$ with $u^\mu \equiv dx^\mu / d\tau$. Moreover, we have now learned at (7.20) that ϕ must be a luminous field, which requirement has been embedded in (7.29b). So, this derivative along the curve will be taken in *frames of reference which travel with the luminous field*, which luminous reference frames cannot ever be transformed into the rest frame – or even into a relatively-moving frame – of a *material* observer. As a result, it is indeed possible to have a zero $d\phi / d\tau$ in the luminous reference frame "along the curve" simultaneously with a non-zero gradient $\partial_\mu\phi \neq 0$ taken with reference to coordinates defined by a

material observer. As we now shall elaborate, this solves the "constant field / zero gradient" problems which have long plagued Kaluza-Klein theory, and teaches a great deal of new intriguing information about the physical properties of the scalar field ϕ.

8. Luminosity and Internal Second-Rank Dirac Symmetry of the Dirac-Kaluza-Klein Scalar

Let us take the final step of connecting (7.29) to the observed Lorentz Force motion with nothing else in the way, by formally setting the derivative along the curve for ϕ to zero, thus:

$$\frac{d\phi}{cd\tau} = \frac{\partial\phi}{\partial x^\mu}\frac{dx^\mu}{cd\tau} = 0 . \tag{8.1}$$

With this, both of (7.29) immediately become synonymous with the Lorentz Force motion (7.26). From the standpoint of scientific method, we can take (7.29) together with (7.26) as empirical evidence that (8.1) must be true. Now, let's explore what (8.1) – if it really is true – teaches us about the physical properties of ϕ.

To start, let us square (8.1) and so write this as:

$$\left(\frac{d\phi}{cd\tau}\right)^2 = \frac{\partial\phi}{\partial x^\mu}\frac{\partial\phi}{\partial x^\nu}\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau} = \partial_\mu\phi\partial_\nu\phi\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau} = 0 . \tag{8.2}$$

Next, let's write the four-dimensional spacetime metric (7.5) for a luminous particle using (3.13) with $g_{\mu\nu} = \eta_{\mu\nu}$ and $\Phi_0 = \phi$ as:

$$0 = c^2 d\tau^2 = G_{\mu\nu}dx^\mu dx^\nu = \eta_{\mu\nu}dx^\mu dx^\nu + \phi^2 k^2 A_{\gamma\mu}A_{\gamma\nu}dx^\mu dx^\nu . \tag{8.3}$$

We already used a variant of this to obtain (7.23). Then, also appending a ϕ^2 and using an overall minus sign which will become useful momentarily, we restructure this to:

$$-\left(\eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu}A_{\gamma\nu}\right)\frac{dx^\mu}{cd\tau}\frac{dx^\nu}{cd\tau}\phi^2 = 0 . \tag{8.4}$$

The above (8.4) describes a luminous particle in a five-dimensional spacetime with the metric tensor (3.13). So, we can use this luminosity to supply the zero for the squared derivative along the curve in (8.2) if, comparing (8.2) and (8.4), we define the relation:

$$\partial_\mu\phi\partial_\nu\phi \equiv -\left(\eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu}A_{\gamma\nu}\right)\phi^2 / \lambda\!\!\!^{-2} \neq 0 , \tag{8.5}$$

where $\lambda\!\!\!^{-} \equiv \lambda / 2\pi$ is a reduced wavelength of the scalar, needed and therefore introduced to balance the $1/\text{length}^2$ dimension of $\partial_\mu \phi \partial_\nu \phi$ with the dimensionless $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$. Now, all we need to do is determine the first-order $\partial_\mu \phi$ which satisfies (8.5).

What becomes apparent on close study of (8.5) is that there is no way to isolate a first-order $\partial_\mu \phi$ unless we make use of the Dirac gamma operators in a manner very similar to what Dirac originally used in [13] to take the operator "square root" of the Klein-Gordon equation. And in fact, the operator square root we need to take to separate out a linear $\partial_\mu \phi$ from (8.5) is precisely the $\Gamma_\mu = \begin{pmatrix} \gamma_0 + kA_{\gamma j}\gamma_j & \gamma_j + kA_{\gamma j}\gamma_0 \end{pmatrix}$ we found in (2.14) which satisfy (2.1) with $g_{\mu\nu} = \eta_{\mu\nu}$, that is, which satisfy $\frac{1}{2}\{\Gamma_\mu, \Gamma_\nu\} = \eta_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$. Therefore, we may now use these Γ_μ to take the square root of (8.5), where we also use $-i = \sqrt{-1}$ choosing $-i$ rather than $+i$ for reasons which will become apparent at (8.10), to obtain:

$$\lambda\!\!\!^{-} \partial_\mu \phi = -i\Gamma_\mu \phi . \tag{8.6}$$

Now, just as the photon gauge field (2.11a) contains a Fourier kernel $\exp\left(-iq_\sigma x^\sigma / \hbar\right)$ where q^μ is the photon energy-momentum, and the fermion wavefunction used in (5.6) contains a Fourier kernel $\exp\left(-ip_\Sigma x^\Sigma / \hbar\right)$ with a fermion five-momentum p^{M} (and we anticipate (5.6) will be used to inform us regarding p^5), let us specify a Fourier kernel $\exp\left(-is_\Sigma x^\Sigma / \hbar\right)$ with a five-dimensional s^{M} which we regard as the five-momentum of the luminous scalar ϕ. Moreover, because ϕ is dimensionless and so too is $\exp\left(-is_\Sigma x^\Sigma / \hbar\right)$, let us simply define:

$$\phi \equiv \tfrac{1}{\sqrt{2}}\left(\phi_1 + i\phi_2\right)\exp\left(-is_\Sigma x^\Sigma / \hbar\right). \tag{8.7}$$

Above, $\exp\left(-is_\Sigma x^\Sigma / \hbar\right)$ is a Fourier kernel in five dimensions, while $\phi_1 + i\phi_2$ is a dimensionless, complex-valued amplitude. This complex amplitude, albeit dimensionless, is chosen to be analogous to the energy-dimensioned scalar field is used to break symmetry via the standard model Higgs mechanism, which we denote by $\phi_h \equiv \frac{1}{\sqrt{2}}\left(\phi_{1h} + i\phi_{2h}\right)$. Specifically, ϕ_1 and ϕ_2 introduce two degrees of freedom which can be used to give mass to otherwise massless objects. Because $\left(\phi_1 + i\phi_2\right)^*\left(\phi_1 + i\phi_2\right) = \phi_1^2 + \phi_2^2$, the symmetry of the "circle" in the complex Euler plane of ϕ_1 and ϕ_2 can always be broken by choosing the $\phi_2 = 0$ orientation, see Figure 14.5 in [20]. In the standard model, once the symmetry is broken, the scalar field is expanded about the vacuum having an expectation value v via $\phi_h\left(x^\mu\right) = \frac{1}{\sqrt{2}}\left(v + h\left(x^\mu\right)\right)$, with fluctuations provided by the Higgs field $h\left(x^\mu\right)$. In the standard model, the vev is taken to be $v = 246.2196508\ \text{GeV}$ of the standard model, namely, the Fermi vacuum expectation associated with the Fermi coupling via $1/2v^2 = G_F / \sqrt{2}\left(\hbar c\right)^3$ based on the latest PDG data [21]. In (8.7), *which we will connect directly*

to the standard model Higgs mechanism in the section 11, the kernel $\exp\left(-is_{\Sigma}x^{\Sigma}/\hbar\right)$ provides a third degree of freedom based on the orientation of the angle $\theta = s_{\Sigma}x^{\Sigma}/\hbar$.

If we allow $\phi_1\left(x^{\mathrm{M}}\right)$ and $\phi_2\left(x^{\mathrm{M}}\right)$ to be functions of five-dimensional spacetime so they can be expanded about a minimum *v* in familiar form $\phi(x) = \frac{1}{\sqrt{2}}\left(v + h\left(x^{\Sigma}\right)\right)$ after choosing an $\phi_2 = 0$ orientation, then the five-gradient of (8.7) is straightforwardly calculated to be:

$$\partial_{\mathrm{M}}\phi = \left(\frac{\partial_{\mathrm{M}}\phi_1 + i\partial_{\mathrm{M}}\phi_2}{\phi_1 + i\phi_2} - i\frac{s_{\mathrm{M}}}{\hbar}\right)\phi\,. \tag{8.8}$$

If the then covariantly extend (8.6) into the fifth dimension in the form of $\lambda\!\!\!^{-}\partial_{\mathrm{M}}\phi = -i\Gamma_{\mathrm{M}}\phi$ and then apply (8.8) we find:

$$\lambda\!\!\!^{-}\partial_{\mathrm{M}}\phi = \lambda\!\!\!^{-}\left(\frac{\partial_{\mathrm{M}}\phi_1 + i\partial_{\mathrm{M}}\phi_2}{\phi_1 + i\phi_2} - i\frac{s_{\mathrm{M}}}{\hbar}\right)\phi = -i\Gamma_{\mathrm{M}}\phi\,. \tag{8.9}$$

Stripping off ϕ, following some algebraic rearrangement including multiplying through by *c*, then using $E = \hbar c/\lambda\!\!\!^{-} = \hbar\omega = hf$ for the energy magnitude of the scalar, we arrive at:

$$cs_{\mathrm{M}} = \hbar\omega\Gamma_{\mathrm{M}} - i\hbar c\frac{\partial_{\mathrm{M}}\phi_1 + i\partial_{\mathrm{M}}\phi_2}{\phi_1 + i\phi_2}\,. \tag{8.10}$$

The time component of $\hbar\omega\Gamma_0 = \hbar\omega\left(\gamma_0 + kA_{\gamma j}\gamma_j\right)$ within the energy component cs_0 above is positive for the upper (particle) components of $\mathrm{diag}\left(\gamma_0\right) = \left(+I, -I\right)$ in the Dirac representation, and negative for the lower (antiparticle) components, which we interpret using Feynman–Stueckelberg. Having these upper components be positive is the reason we used $-i = \sqrt{-1}$ at (8.6).

Finally, we insert (8.10) into (8.7) for the luminous scalar and reduce, to obtain:

$$\begin{aligned}\phi &= \frac{1}{\sqrt{2}}\left(\phi_1 + i\phi_2\right)\exp\left(-i\frac{\omega}{c}\Gamma_{\Sigma}x^{\Sigma} - \frac{\partial_{\Sigma}\phi_1 + i\partial_{\Sigma}\phi_2}{\phi_1 + i\phi_2}x^{\Sigma}\right)\\ &= \frac{1}{\sqrt{2}}\left(\phi_1 + i\phi_2\right)\exp\left(-i\frac{\omega}{c}\Gamma_{\Sigma}x^{\Sigma}\right)\exp\left(-\frac{\partial_{\Sigma}\phi_1 + i\partial_{\Sigma}\phi_2}{\phi_1 + i\phi_2}x^{\Sigma}\right)\end{aligned}\,. \tag{8.11}$$

The product separation of exponentials in the lower line is possible in view of the Zassenhaus-Baker-Campbell-Hausdorff relation $\exp\left(A + B\right) = \exp A\exp B\exp\left(-\left[A, B\right]/2\right)\ldots$ because although $\Gamma_{\Sigma}x^{\Sigma}$ is a 4x4 matrix operator, the second additive term in the top line is a 4x4 diagonal matrix which does commute with the first term, i.e., $\left[A, B\right] = 0$. Because (8.10) contains an

energy $E = \hbar\omega = hf$, we now must interpret ϕ as single luminous field quantum just as at (2.11) we were required to regard $A_\mu = A_{\gamma\mu}$ as an individual photon quantum. Significantly, both the energy-momentum five-vector cs_{M} in (8.10) for the scalar, and the scalar itself in (8.11), are actually 4x4 operator matrices owing to the presence of Γ_Σ in each. Thus, these both have an implied second rank index pair *AB* with Dirac spinor indexes $A = 1,2,3,4$ and $B = 1,2,3,4$.

To make use of the luminous scalar operator (8.11) in later calculations, it is helpful to separate the kernel $\exp\left(-i\omega\Gamma_\Sigma x^\Sigma / c\right)$ into sine and cosine terms using the Maclaurin series $\exp(-ix) = 1 - ix - \frac{1}{2!}x^2 + i\frac{1}{3!}x^3 + \frac{1}{4!}x^4 - \frac{1}{5!}ix^5 - ... = \cos x - i\sin x$. To do so, we first use the anticommutator (3.1) to calculate the square:

$$\left(\Gamma_\Sigma x^\Sigma\right)^2 = \tfrac{1}{2}\left\{\Gamma_{\mathrm{M}}\Gamma_{\mathrm{N}} + \Gamma_{\mathrm{N}}\Gamma_{\mathrm{M}}\right\}x^{\mathrm{M}}x^{\mathrm{N}} = G_{\mathrm{MN}}x^{\mathrm{M}}x^{\mathrm{N}} \equiv c^2\mathrm{T}^2, \tag{8.12}$$

where $\mathrm{S}^2 \equiv c^2\mathrm{T}^2 \equiv G_{\mathrm{MN}}x^{\mathrm{M}}x^{\mathrm{N}}$ is a *finite* invariant proper length / time in the five-dimensional geometry. Thus $\left(\omega\Gamma_\Sigma x^\Sigma / c\right)^2 \equiv \omega^2\mathrm{T}^2$. Then, we insert this into the series to obtain:

$$\begin{aligned}\exp\left(-i\frac{\omega}{c}\Gamma_\Sigma x^\Sigma\right) &= 1 - \frac{1}{2!}\omega^2\mathrm{T}^2 + \frac{1}{4!}\omega^4\mathrm{T}^4 - i\frac{\omega}{c}\Gamma_\Sigma x^\Sigma\left(1 - \frac{1}{3!}\omega^2\mathrm{T}^2 + \frac{1}{5!}\omega^4\mathrm{T}^4\right) + ... \\ &= \cos(\omega\mathrm{T}) - i\frac{\Gamma_\Sigma x^\Sigma}{c\mathrm{T}}\left(\omega\mathrm{T} - \frac{1}{3!}\omega^3\mathrm{T}^3 + \frac{1}{5!}\omega^5\mathrm{T}^5\right) + ... = \cos(\omega\mathrm{T}) - i\frac{\Gamma_\Sigma x^\Sigma}{c\mathrm{T}}\sin(\omega\mathrm{T})\end{aligned}, \tag{8.13}$$

To get to the sin term in the bottom line, we multiplied through by $1 = \omega\mathrm{T} / \omega\mathrm{T}$. Inserting this into (8.11) gives us the final expression for the luminous, dimensionless. massless scalar:

$$\boxed{\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)\left(\cos(\omega\mathrm{T}) - i\frac{\Gamma_\Sigma x^\Sigma}{c\mathrm{T}}\sin(\omega\mathrm{T})\right)\exp\left(-\frac{\partial_\Sigma\phi_1 + i\partial_\Sigma\phi_2}{\phi_1 + i\phi_2}x^\Sigma\right)}. \tag{8.14}$$

The Dirac operator characteristics of ϕ are now seen to be isolated in and stem from the $\Gamma_\Sigma x^\Sigma$ matrix which multiplies the $\sin(\omega\mathrm{T})$ term. In view of (8.12) it is clear that $\phi^*\phi = \frac{1}{2}\left(\phi_1^{\,2} + \phi_2^{\,2}\right)$, which is precisely the property this luminous scalar it should have in order to be able to break a local U(1) gauge symmetry, see, e.g., sections 14.7 and 14.8 in [20]. But as noted after (8.7), what is now the more-recognizable angle $\theta = \omega\mathrm{T}$ provides a third degree of freedom. In section 11 we shall see that (8.14) above further simplifies when we geometrize the fermion rest masses and break the symmetry such that two degrees of freedom give mass to fermions and the third degree of freedom gives mass to the massless scalar ϕ and produces the massive Higgs.

The luminous massless scalar operator (8.14) with second-rank Dirac internal symmetries solves the Kaluza-Klein problem of how to make the scalar field "constant" to remove what are otherwise some very large terms, while not unduly constraining the electromagnetic fields: The

gradient can be non-zero, while the derivative along the curve can be zero, $d\phi/cd\tau=0$, so long as the scalar is a luminous particle which also has a second rank Dirac structure. In turn, if we then return to the metric tensor G_{MN} in the form of, say, (3.11), we find that this too must also have implied Dirac indexes, that is, $G_{\text{MN}}=G_{\text{MN}AB}$ owing to the structure (8.14) of the scalar fields which sit in its fifth dimensional components. So (8.14) gives a second rank Dirac structure to the metric tensor, alongside of its already second-rank, five dimensional spacetime structure. And of course, with (8.14) being derived to obey $d\phi/d\tau=0$, (7.29) become synonymous with the electrodynamic Lorentz force motion, which is one of the key touchstones of Kaluza-Klein theory.

And so, the Kaluza-Klein fifth dimension, taken together with using Dirac theory to enforce general covariance across all five dimensions, has turned a metric tensor (1.1) with an entirely classical character, into a quantum field theory metric tensor with luminous photons and luminous scalar field quanta. If this is all in accord with physical reality, this means that nature actually has *three spin types of massless, luminous field quanta*: spin-2 gravitons, spin-1 photons and gluons, and spin-0 scalars with an internal second rank Dirac-tensor symmetry. This also means that the massless, luminous Kaluza-Klein scalar in (8.14) is not the same scalar as the usual Higgs, because the latter is massive and material. However, the scalar (8.14) has properties similar to the Higgs, and as we shall now see, it can be used to spontaneously break symmetry, whereby the two degrees of freedom in the amplitude $\left(\phi_1+i\phi_2\right)/\sqrt{2}$ give the fermions their rest mass, and the third degree of freedom $\theta=\omega\text{T}$ does produce the Higgs in its known *massive* form. And, there is a direct relation which we shall see between the DKK scalar field and the Higgs field which will enable us to develop a theory of fermion masses and mixing angles which fits the experimental data. Moreover, the type of Higgs production paired with top quarks reported out of CERN only in the past several weeks [22], [23], [24] is better-understood by using (8.14) as the scalar for spontaneous symmetry breaking leading to fermion rest masses.

9. How the Dirac-Kaluza-Klein Metric Tensor Resolves the Challenges faced by Kaluza-Klein Theory without Diminishing the Kaluza "Miracle," and Grounds the Now-Timelike Fifth Dimension in Manifestly-Observed Physical Reality

Now let's review the physics implications of everything that has been developed here so far. As has been previously pointed out, in the circumstance where all electrodynamic interactions are turned off by setting $A_{\gamma j}=0$ and what is now $\Phi_\mu=0$, then (3.13) reduces when $g_{\mu\nu}=\eta_{\mu\nu}$ to $\text{diag}\left(G_{\text{MN}}\right)=\left(+1,-1,-1,-1,+1\right)$ with $\left|G_{\text{MN}}\right|=-1$. And we saw at (6.3) that this result does not change at all, even when $A_{\gamma j}\neq 0$ and $\Phi_\mu\neq 0$. But in the same situation the usual Kaluza-Klein metric tensor (1.1) reduces to $\text{diag}\left(G_{\text{MN}}\right)=\left(+1,-1,-1,-1,0\right)$ with a determinant $\left|G_{\text{MN}}\right|=0$. This of course means the Kaluza-Klein metric tensor is not-invertible and therefore becomes singular when electrodynamic interactions are turned off. Again, this may be seen directly from the fact that when we set $A_{\gamma j}=0$ and $\phi=0$, in (1.1) we get $G^{55}=g_{\alpha\beta}A^\alpha A^\beta+1/\phi^2=0+\infty$. This degeneracy leads to a number of interrelated ills which have hobbled Kaluza-Klein as a viable theory of the natural world for a year shy of a century.

First, the scalar field ϕ carries a much heavier burden than it should, because Kaluza-Klein theory relies upon this field being non-zero to ensure that the five-dimensional spacetime geometry is non-singular. This imposes constraints upon ϕ which would not exist if it was not doing "double duty" as both a scalar field and as a structural element required to maintain the non-degeneracy of Minkowski spacetime extended to five dimensions.

Second, this makes it next-to-impossible to account for the fifth dimension in the observed physical world. After all, the space and time of real physical experience have a flat spacetime signature $\text{diag}\left(\eta_{\mu\nu}\right)=\left(+1,-1,-1,-1\right)$ which is structurally sound even in the absence of any fields whatsoever. But what is one to make of a signature which, when $g_{\mu\nu}=\eta_{\mu\nu}$ and $A_{\gamma k}=0$, is given by $\text{diag}\left(\eta_{\text{MN}}\right)=\left(+1,-1,-1,-1,\phi^2\right)$ with $\left|\eta_{\text{MN}}\right|=-\phi^2$? How is one to explain the physicality of a $G_{55}=\phi^2$ in the Minkowski signature which is based upon a field, rather than being either a timelike +1 or a spacelike –1 Pythagorean metric component? The Minkowski signature defines the *flat tangent* spacetime at each event, absent curvature. How can a tangent space which by definition should not be curved, be dependent upon a field ϕ which if it has even the slightest modicum of energy will cause curvature? This is an internal logical contradiction of the Kaluza-Klein metric tensor (1.1) that had persisted for a full century, and it leads to such hard-to-justify oddities as a fifth dimensional metric component $G_{55}=\phi^2$ and determinant $\left|\eta_{\text{MN}}\right|=-\phi^2$ which dilates or contracts (hence the sometime-used name "dilaton") in accordance with the behavior of ϕ^2.

Third, the DKK metric tensor (3.13) is obtained by *requiring* that it be possible to deconstruct the Kaluza-Klein metric tensor into a set of Dirac matrices obeying (3.1), with the symmetry of full five-dimensional general covariance. What we have found is that it is *not possible* to have 5-dimensional general covariance if $G_{05}=G_{50}=0$ and $G_{55}=\phi^2$ as in (1.1). Rather, general 5-dimensional covariance *requires* that $G_{05}=G_{50}=\phi$ and $G_{55}=1+\phi^2$ in (3.13). Further, even to have spacetime covariance in *four dimensions alone*, we are *required* to gauge the electromagnetic potential to that of the photon. Without these changes to the metric tensor components, it is simply not possible to make Kaluza-Klein theory compatible with Dirac theory and to have 5-dimensional general covariance. This means that there is no consistent way of using the usual (1.1) to account for the fermions which are at the heart of observed matter in the material universe. Such an omission – even without any of its other known ills – most-assuredly renders the KK metric (1.1) "unphysical."

Finally, there is the century-old demand which remains unmet to this date: "show me the fifth dimension!" There is no observational evidence at all to support the fifth dimension, at least in the form specified by (1.1), or in the efforts undertaken to date to remedy these problems.

But the metric tensors (3.13) and (4.22) lead to a whole other picture. First, by definition, a 5-covariant Dirac equation (5.6) can be formed, so there is no problem of incompatibility with Dirac theory. Thus, all aspects of fermion physics may be fully accounted for. Second, it should be obvious to anyone familiar with the γ_μ and $\gamma_5\equiv -i\gamma_0\gamma_1\gamma_2\gamma_3$ that one may easily use an

anticommutator $\eta_{MN} \equiv \frac{1}{2}\{\gamma_M, \gamma_N\}$ to form a five-dimensional Minkowski tensor with $\text{diag}(\eta_{MN}) = (+1,-1,-1,-1,+1)$, which has a Minkowski signature with two timelike and three spacelike dimensions. But it is not at all obvious how one might proceed to regard γ_5 as the generator of *a truly-physical fifth dimension* which is on an absolute par with the generators γ_μ of the four truly-physical dimensions which are time and space. This is true, notwithstanding the clear observational evidence that γ_5 has a multitude of observable physical impacts. The reality of γ_5 is most notable in the elementary fermions that contain the factor $\frac{1}{2}(1 \pm \gamma_5)$ for right- and left-chirality; in the one particle and interaction namely neutrinos acting weakly that are always left-chiral; and in the many observed pseudo-scalar mesons ($J^{PC} = 0^{-+}$) and pseudo-vector mesons ($J^{PC} = 1^{++}$ and $J^{PC} = 1^{+-}$) laid out in [25], all of which require the use of γ_5 to underpin their theoretical origins. So γ_5 is real and physical, as would therefore be any fifth dimension which can be *properly-connected* with γ_5.

But the immediate problem as pointed out in toward the end of [11], is that because $G_{55} = \phi^2$ in the Kaluza-Klein metric tensor (1.1), if we require electromagnetic energy densities to be positive, the fifth-dimension must have a *spacelike* signature. And this directly contradicts making γ_5 the generator of the fifth dimension because $\gamma_5\gamma_5 = 1$ produces a *timelike* signature. So, as physically-real and pervasive as are the observable consequences of the γ_5 matrix, the Kaluza-Klein metric tensor (1.1) does *not* furnish a theoretical basis for associating γ_5 with a fifth dimension, at the very least because of this timelike-versus-spacelike contradiction. This is yet another problem stemming from having ϕ carry the burden of maintaining the fifth-dimensional signature and the fundamental Pythagorean character of the Minkowski tangent space.

So, to summarize, on the one hand, Kaluza-Klein theory has a fifth physical dimension on a par with space and time, but it has been impossible to connect that dimension with actual observations in the material, physical universe, or to make credible sense of the dilation and contraction of that dimension based on the behavior of a scalar field. On the other hand, Dirac theory has an eminently-physical γ_5 with pervasive observational manifestations on an equal footing with γ_μ, but it has been impossible to connect this γ_5 with a true physical fifth dimension (or at least, with the Kaluza-Klein metric tensor (1.1) in five dimensions). At minimum this is because the metric tensor signatures conflict. *Kaluza-Klein has a fifth-dimension unable to connect to physical reality, while Dirac theory has a physically-real γ_5 unable to connect to a fifth dimension.* And the origin of this disconnect on both hands, is that the Kaluza-Klein metric tensor (1.1) cannot be deconstructed into Dirac-type matrices while maintaining five-dimensional general covariance according to (3.1). *To maintain general covariance and achieve a Dirac-type square root operator deconstruction of the metric tensor, (1.1) must be replaced by (3.13) and (4.22).*

Once we use (3.13) and (4.22) all these problems evaporate. Kaluza-Klein theory becomes fully capable of describing fermions as shown in (5.6). With $G_{55} = 1 + \phi^2$ the metric signature is decoupled from the energy requirements for ϕ, and with $|G_{MN}| = |g_{MN}|$ from (6.3) the metric tensor

determinant is entirely independent of both $A_{\gamma\mu}$ and ϕ. Most importantly, when $A_{\gamma j}=0$ and $\phi=0$ and $g_{\mathrm{MN}}=\eta_{\mathrm{MN}}$, because $\mathrm{diag}\left(G_{\mathrm{MN}}\right)=\left(+1,-1,-1,-1,+1\right)=\mathrm{diag}\left(\frac{1}{2}\{\gamma_{\mathrm{M}},\gamma_{\mathrm{N}}\}\right)=\mathrm{diag}\left(\eta_{\mathrm{MN}}\right)$, and because of this decoupling of ϕ from the metric signature, we now have a timelike $\eta_{55}=\gamma_5\gamma_5=+1$ which is directly generated by γ_5. As a consequence, the fifth dimension of Kaluza-Klein theory which has heretofore been disconnected from physical reality, can now be identified with a true physical dimension that has γ_5 as its generator, just as γ_0 is the generator of a truly-physical time dimension and γ_j are the generators of a truly-physical space dimensions. And again, γ_5 has a wealth of empirical evidence to support its reality.

Further, with a tangent space $\mathrm{diag}\left(\eta_{\mathrm{MN}}\right)=\left(+1,-1,-1,-1,+1\right)$ we now have two timelike and three spacelike dimensions, with matching tangent-space signatures between Dirac theory and the Dirac-Kaluza-Klein theory. With the fifth-dimension now being timelike not spacelike, the notion of "curling up" the fifth dimension into a tiny "cylinder" comes off the table completely, while the Feynman-Wheeler concept of "many-fingered time" returns to the table, providing a possible avenue to study future probabilities which congeal into past certainties as the arrow of time progresses forward with entropic increases. And because γ_5 is connected to a multitude of confirmed observational phenomena in the physical universe, the physical reality of the fifth dimension in the metric tensors (3.13) and (4.22) is now supported by every single observation ever made of the reality of γ_5 in particle physics, regardless of any other epistemological interpretations one may also arrive at for this fifth dimension.

Moreover, although the field equations obtained from (3.13) and (4.22) rather than (1.1) will change somewhat because now $G_{05}=G_{50}=\phi$ and $G_{55}=1+\phi^2$ and the gauge fields are fixed to the photon $A_\mu=A_{\gamma\mu}$ with only two degrees of freedom, there is no reason to suspect that the many good benefits of Kaluza-Klein theory will be sacrificed because of these changes which eliminate the foregoing problems. Indeed, we have already seen in sections 7 and 8 how the Lorentz force motion is faithfully reproduced. Rather, we simply expect some extra terms (and so expect some additional phenomenology) to emerge in the equations of motion and the field equations because of these modifications. But the Kaluza-Klein benefits having of Maxwell's equations, the Lorentz Force motion and the Maxwell-stress energy embedded, should remain fully intact when using (3.13) and (4.22) in lieu of (1.1), as illustrated in sections 7 and 8.

Finally, given all of the foregoing, beyond the manifold observed impacts of γ_5 in particle physics, there is every reason to believe that using the five-dimensional Einstein equation with the DKK metric tensors will fully enable us to understand this fifth dimension, at bottom, as a *matter dimension*, along the lines long-advocated by the 5D Space-Time-Matter Consortium [18]. This will be further examined in next section, and may thereby bring us ever-closer to uncovering the truly-geometrodynamic theoretical foundation at the heart of all of nature.

10. Pathways for Continued Exploration: The Einstein Equation, the "Matter Dimension," Quantum Field Path Integration, Epistemology of a Second Time Dimension, and All-Interaction Unification

Starting at (7.6) we obtained the connection $\tilde{\Gamma}^{\mathrm{M}}_{\alpha 5}$ in order to study the $\tilde{\Gamma}^{\mathrm{M}}_{\alpha 5}$ term in the equation of motion (7.4), because this is the term which provides the Lorentz Force motion which becomes (7.29) once ϕ is understood to be a luminous field with $d\phi / d\tau = 0$ as in (8.1). The reason this was developed in detail here, is to demonstrate that the DKK metric tensors (3.13) and (4.22) in lieu of the usual (1.1) of Kaluza-Klein do not in any way forego the Kaluza miracle, at least as regards the Lorentz Force equation of electrodynamic motion. But there are a number of further steps which can and should be taken to further develop the downstream implications of using the DKK metric tensors (3.13) and (4.22) in lieu of the usual (1.1) of Kaluza-Klein.

First, it is necessary to calculate all of the other connections $\tilde{\Gamma}^{\mathrm{M}}_{\mathrm{AB}}$ using (7.3) and the metric tensors (3.13) and (4.22) similarly to what was done in section 7, then to fully develop the remaining terms in the equations of motion (7.2), (7.4) which have not yet been elaborated here, and also to obtain the five-dimensional Riemann and Ricci tensors, and the Ricci scalar:

$$\begin{aligned}
&\hat{R}^{\mathrm{A}}_{\mathrm{BMN}} = \partial_{\mathrm{M}}\Gamma^{\mathrm{A}}{}_{\mathrm{NB}} - \partial_{\mathrm{N}}\Gamma^{\mathrm{A}}{}_{\mathrm{MB}} + \Gamma^{\mathrm{A}}{}_{\mathrm{M}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{NB}} - \Gamma^{\mathrm{A}}{}_{\mathrm{N}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{MB}} \\
&\hat{R}_{\mathrm{BM}} = \hat{R}^{\mathrm{T}}_{\mathrm{BMT}} = \partial_{\mathrm{M}}\Gamma^{\mathrm{T}}{}_{\mathrm{TB}} - \partial_{\mathrm{T}}\Gamma^{\mathrm{T}}{}_{\mathrm{MB}} + \Gamma^{\mathrm{T}}{}_{\mathrm{M}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{TB}} - \Gamma^{\mathrm{T}}{}_{\mathrm{T}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{MB}} \\
&\hat{R} = \hat{R}^{\Sigma}{}_{\Sigma} = G^{\mathrm{BM}}\hat{R}_{\mathrm{BM}} = G^{\mathrm{BM}}\partial_{\mathrm{M}}\Gamma^{\mathrm{T}}{}_{\mathrm{TB}} - G^{\mathrm{BM}}\partial_{\mathrm{T}}\Gamma^{\mathrm{T}}{}_{\mathrm{MB}} + G^{\mathrm{BM}}\Gamma^{\mathrm{T}}{}_{\mathrm{M}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{TB}} - G^{\mathrm{BM}}\Gamma^{\mathrm{T}}{}_{\mathrm{T}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{MB}}
\end{aligned} \quad . \tag{10.1}$$

Once these are obtained, these may then be placed into a fifth-dimensional Einstein field equation:

$$-\mathrm{K}\hat{T}_{\mathrm{MN}} = \hat{R}_{\mathrm{MN}} - \tfrac{1}{2}G_{\mathrm{MN}}\hat{R} \tag{10.2}$$

with a suitably-dimensioned constant K related to the usual κ to be discussed momentarily. This provides the basis for studying the field dynamics and energy tensors of the DKK geometry.

The development already presented here, should make plain that the Kaluza miracle will also be undiminished when the DKK metric tensors (3.13) and (4.22) are used in (10.2) in lieu of the usual Kaluza-Klein (1.1). Because $\tilde{\Gamma}^{\mu}_{\alpha 5}$ which we write as $\tilde{\Gamma}^{\mu}_{\alpha 5} = \frac{1}{2}\eta^{\mu\sigma}\phi^2 kF_{\alpha\sigma} + ...$ contains the electromagnetic field strength as first established at (7.13), we may be comfortable that the terms needed in the Maxwell tensor will be embedded in the (10.1) terms housed originally in $\Gamma^{\mathrm{A}}{}_{\mathrm{M}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{NB}} - \Gamma^{\mathrm{A}}{}_{\mathrm{N}\Sigma}\Gamma^{\Sigma}{}_{\mathrm{MB}}$. Moreover, because the electromagnetic source current density $\mu_0 J^{\mu} = \partial_{\sigma}F^{\sigma\mu}$, we may also be comfortable that Maxwell's source equation will be embedded in the terms housed originally in $\partial_{\mathrm{M}}\Gamma^{\mathrm{A}}{}_{\mathrm{NB}} - \partial_{\mathrm{N}}\Gamma^{\mathrm{A}}{}_{\mathrm{MB}}$. Moreover, because $\left(\hat{R}^{\mathrm{MN}} - \frac{1}{2}G^{\mathrm{MN}}\hat{R}\right)_{;\mathrm{M}} = 0$ which via (10.2) ensures a locally-conserved energy $\hat{T}^{\mathrm{MN}}{}_{;\mathrm{M}} = 0$ is contracted from the second Bianchi identity $\hat{R}^{\mathrm{A}}_{\mathrm{BMN;P}} + \hat{R}^{\mathrm{A}}_{\mathrm{BNP;M}} + \hat{R}^{\mathrm{A}}_{\mathrm{BPM;N}} = 0$, we may also be comfortable that Maxwell's magnetic charge equation $\partial_{;\alpha}F_{\mu\nu} + \partial_{;\mu}F_{\nu\alpha} + \partial_{;\nu}F_{\alpha\mu} = 0$ will likewise be embedded. In short, we may be comfortable based on what has already been developed here, that the Kaluza miracle will

remain intact once field equations are calculated. But we should expect some additional terms and information emerging from the field equation which do not appear when we use the usual (1.1).

Second, the Ricci scalar $\hat{R}$ is especially important because of the role it plays in the Einstein-Hilbert Action. This action provides a very direct understanding of the view that the fifth dimension is a matter dimension [18], and because this action can be used to calculate five-dimensional gravitational path integrals which may be of assistance in better understanding the nature of the second time dimension t^5. Let us briefly preview these development paths.

The Einstein-Hilbert action reviewed for example in [26], in four dimensions, is given by:

$$S = \int \left((1/2\kappa) R + \mathcal{L}_{\mathrm{M}} \right) \sqrt{-g}\, \mathrm{d}^4 x \ . \tag{10.3}$$

The derivation of the four-dimensional (10.2) from this is well-known, where $R = R^{\sigma}{}_{\sigma}$. So, in five dimensions, we immediately expect that (10.2) will emerge from extending (10.3) to:

$$\hat{S} = \int \left((1/2\mathrm{K}) \hat{R} + \hat{\mathcal{L}}_{\mathrm{M}} \right) \sqrt{-G}\, \mathrm{d}^5 x = \int \left((1/2\mathrm{K}) \hat{R}^{\sigma}{}_{\sigma} + (1/2\mathrm{K}) \hat{R}^{5}{}_{5} + \hat{\mathcal{L}}_{\mathrm{M}} \right) \sqrt{-G}\, \mathrm{d}^5 x \ , \tag{10.4}$$

using $\hat{R} = \hat{R}^{\sigma}{}_{\sigma} + \hat{R}^{5}{}_{5}$ from (10.1) and the G already obtained in (6.3), and where $\mathrm{K} \equiv \lambda\kappa$ contains some suitable length λ to balance the extra space dimensionality in $\mathrm{d}^5 x$ versus $\mathrm{d}^4 x$. In Kaluza-Klein theory based on (1.1) λ is normally the radius of the compactified fourth space dimension and is very small. Here, because there is a second time dimension, this should become associated with some equally-suitable period of time=length/c, but it may not necessarily be small if it is associated, for example, with the reduced wavelength $\lambda\!\!\!^{-} = c/\omega$ of the scalar deduced in (8.14), and if that wavelength is fairly large which is likely because these scalars ϕ have not exactly overwhelmed the detectors in anybody's particle accelerators or cosmological observatories.

However, the energy tensor $T^{\mu\nu}$ in four dimensions is placed into the Einstein equation by hand. This is why Einstein characterized the $R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R$ side of his field equation as "marble" and the $-\kappa T^{\mu\nu}$ side as "wood." And this $T^{\mu\nu}$ is defined from the Lagrangian density of matter by:

$$T_{\mu\nu} \equiv -2 \frac{\delta \mathcal{L}_{\mathrm{M}}}{\delta g^{\mu\nu}} + g_{\mu\nu} \mathcal{L}_{\mathrm{M}} \ . \tag{10.5}$$

Therefore, in the Five-Dimensional Space-Time-Matter view of [18], and referring to (10.4), the "wood" of $\hat{\mathcal{L}}_{\mathrm{M}}$ is discarded entirely, and rather, we associate

$$\hat{\mathcal{L}}_{\mathrm{M}} \equiv (1/2\mathrm{K}) \hat{R}^{5}{}_{5} \tag{10.6}$$

with the matter Lagrangian density. As a result, this is now also made of "marble."

Then (10.4) may be simplified to the 5-dimensional "vacuum" equation (see [27] at 428 and 429):

$$\hat{S} = \int (1/2\mathrm{K})\,\hat{R}\sqrt{-G}\,\mathrm{d}^5 x\,, \tag{10.7}$$

and the field equation (10.2) derived from varying (10.7) becomes the vacuum equation:

$$0 = \hat{R}_{\mathrm{MN}} - \tfrac{1}{2} G_{\mathrm{MN}} \hat{R}\,. \tag{10.8}$$

And we anticipate that the variation itself will produce the usual relation:

$$\hat{R}_{\mathrm{MN}} = \frac{\delta \hat{R}}{\delta g^{\mathrm{MN}}} = \frac{\delta \hat{R}^{\sigma}{}_{\sigma}}{\delta g^{\mathrm{MN}}} + \frac{\delta \hat{R}^{5}{}_{5}}{\delta g^{\mathrm{MN}}} \tag{10.9}$$

for the Ricci tensor, but now in five dimensions.

So, in view of (10.5) and (10.6), what we ordinarily think of as the energy tensor – which is now made of entirely geometric "marble," – is contained in those *components of (10.8)* which, also in view of (10.9) and $\hat{R} = \hat{R}^{\sigma}{}_{\sigma} + \hat{R}^{5}{}_{5}$, and given the zero of the vacuum in (10.8), are in:

$$-\mathrm{K}\hat{T}_{\mathrm{MN}} = \frac{\delta \hat{R}^{5}{}_{5}}{\delta G^{\mathrm{MN}}} - \frac{1}{2} G_{\mathrm{MN}} \hat{R}^{5}{}_{5} = \hat{R}_{\mathrm{MN}} - \frac{1}{2} G_{\mathrm{MN}} \hat{R} - \frac{\delta \hat{R}^{\sigma}{}_{\sigma}}{\delta g^{\mathrm{MN}}} + \frac{1}{2} G_{\mathrm{MN}} \hat{R}^{\sigma}{}_{\sigma} = -\frac{\delta \hat{R}^{\sigma}{}_{\sigma}}{\delta g^{\mathrm{MN}}} + \frac{1}{2} G_{\mathrm{MN}} \hat{R}^{\sigma}{}_{\sigma}\,. \tag{10.10}$$

In four dimensions, the salient part of the above now becomes (note sign flip):

$$\mathrm{K}\hat{T}_{\mu\nu} = \frac{\delta \hat{R}^{\sigma}{}_{\sigma}}{\delta g^{\mu\nu}} - \frac{1}{2} G_{\mu\nu} \hat{R}^{\sigma}{}_{\sigma}\,. \tag{10.11}$$

We then look for geometrically-rooted energy tensors that emerge in (10.8) and (10.11) using (10.1) which contain field configurations which up to multiplicative coefficients, resemble the Maxwell tensor, the tensors for dust, perfect fluids, and the like, which is all part of the Kaluza miracle. And because T^{00} is an energy-density, and because the integral of this over a three-dimensional space volume is an energy which divided by c^2 is a mass, from this view we see how the fifth dimension really is responsible for creating matter out of geometric "marble" rather than hand-introduced "wood."

In a similar regard, one of the most important outstanding problems in particle physics, is how to introduce fermion rest masses theoretically rather than by hand, and hopefully thereby explain why the fermions have the observed masses that they do. Here, just as the five spacetime dimensions introduce a "marble" energy tensor (10.11), we may anticipate that when the five-dimensional Dirac equation (5.6) is fully developed, there will appear amidst its Lagrangian density terms a fermion rest energy term $m'c^2\overline{\Psi}\Psi$ in which the mc^2 in (5.6) is occupied, not by a hand-added "wood" mass, but by some energy-dimensioned scalar number which emerges entirely

from the five dimensional geometry. In this event, just as we discarded $\hat{\mathcal{L}}_{\mathrm{M}}$ in (10.4) and replaced it with $(1/2\mathrm{K})\hat{R}^{5}{}_{5}$ at (10.6) to arrive at (10.7) and (10.8), we would discard the mc^2 in (5.6), change (5.6) to $i\hbar c\Gamma^{\mathrm{M}}\partial_{\mathrm{M}}\Psi = 0$ without any hand-added "wood" mass, and in its place use the $m'c^2$ emergent from the geometry in the $m'c^2\overline{\Psi}\Psi$ terms.

Third, the action $\hat{S} = \int (1/2\kappa)\hat{R}\sqrt{-G}\mathrm{d}^5x$, like any action, is directly used in the quantum field path integral, which using (10.7) is:

$$Z = \int DG_{\mathrm{MN}} \exp\left(i\hat{S}/\hbar\right) = \int DG_{\mathrm{MN}} \exp\left((i/\hbar)\int (1/2\mathrm{K})\hat{R}\sqrt{-G}\mathrm{d}^5x\right) . \tag{10.12}$$

Here, the only field over which the integration needs to take place is G_{MN}, because this contains not only the usual $g_{\mu\nu}$, but also the photon $A_{\gamma\mu}$ and the scalar ϕ. But aside from the direct value of (10.12) in finally quantizing gravity, one of the deeply-interesting epistemological issues raised by path integration, relates to the meaning of the fifth time dimension – not only as the matter dimension just reviewed – but also as an actual second dimension of time.

For example, Feynman's original formulation of path integration considers the multiple paths that an individual field quantum might take to get from a source point A to a detection point B, *in a given time*. And starting with Feynman-Stueckelberg it became understood that negative energy particles traversing forward in time may be interpreted as positive energy antiparticles moving backward through time. But with a second time dimension t^5, the path integral must now take into account *all of the possible paths through time* that the particle may have taken, which are no longer just forward and backward, but also *sideways* through what is now a *time plane*. Now, the time t^0 that we actually observe may well become associated with *the actual path taken through time* from amidst multiple time travel possibilities each with their own probability amplitudes, and t^5 may become associated with alternative paths not taken. If one has a deterministic view of nature, then of course the only reality rests with events which did occur, while events which may have occurred but did not have no meaning. But if one has a non-deterministic view of nature, then having a second time dimension to account for all the paths through time *which were not taken* makes eminent sense, and certainly makes much more intuitive and experiential sense than curling up a space dimension into a tiny loop. And if path integral calculations should end up providing a scientific foundation for the physical reality of time paths which could have occurred but never did, this could deeply affect human viewpoints of life and nature. So, while the thoughts just stated are highly preliminary, one would anticipate that a detailed analysis of path integration when there is a second time dimension may help us gain further insight into the physical nature of the fifth dimension as a time dimension, in addition to how this dimension may be utilized to turn the energy tensor from "wood" into "marble."

Finally, Kaluza-Klein theory only unifies gravitation and electromagnetism. As noted in the introduction, weak and strong interactions, and electroweak unification, were barely a glimmer a century ago when Kaluza first passed his new theory along to Einstein in 1919. This raises the question whether Kaluza-Klein theory "repaired" to be compatible with Dirac theory using the DKK metric tensor (3.13) and its inverse (4.22) might also provide the foundation for all-

interaction unification to include the weak and strong interactions in addition to gravitation and electromagnetism.

In ordinary four-dimensional gravitational theory, the metric tensor only contains gravitational fields $g_{\mu\nu}$. The addition of a Kaluza Klein fifth dimension adds a spin one vector gauge potential A_μ as well as a spin 0 scaler ϕ to the metric tensor as seen in (1.1). The former becomes the luminous $A_{\gamma\mu}$ of (2.11) and the latter becomes the luminous ϕ_{AB} of (8.14) for the DKK metric tensor (3.13) and inverse (4.22). So, it may be thought that if adding an extra dimension can unify gravitation with electromagnetism, adding additional dimensions beyond the fifth might bring in the other interactions as well. This has been one of the motivations for string theory in higher dimensions, which are then compactified down to the observed four space dimensions. But these higher-dimensional theories invariably regard the extra dimensions to be *spacelike* dimensions curled up into tiny loops just like the spacelike fifth dimension in Kaluza Klein. And as we have shown here, the spacelike character of this fifth dimension is needed to compensate for the singularity of the metric tensor when $\phi \to 0$ which is one of the most serious KK problems repaired by DKK. Specifically, when Kaluza-Klein is repaired by being made compatible with Dirac theory, the fifth dimension instead becomes a second *timelike* rather than a fourth spacelike dimension. So, if the curled-up spacelike dimension is actually a flaw in the original Kaluza-Klein theory because it is based on a metric degeneracy which can be and is cured by enforcing compatibility with Dirac theory over all five dimensions, it appears to make little sense to replicate this flaw into additional spacelike dimensions.

Perhaps the more fruitful path is to recognize, as is well-established, that weak and strong interactions are very similar to electromagnetic interactions insofar as all three are all mediated by spin-1 bosons in contrast to gravitation which is mediated by spin-2 gravitons. The only salient difference among the three spin-1 mediated interactions is that weak and strong interactions employ SU(2) and SU(3) Yang-Mills [28] internal symmetry gauge groups in which the gauge fields are non-commuting and may gain an extra degree of freedom and thus a rest mass by symmetry breaking, versus the commuting U(1) group of electromagnetism. Moreover, Yang-Mills theories have been extraordinarily successful describing observed particle and interaction phenomenology. So, it would appear more likely than not that once we have a U(1) gauge field with only the two photon degrees of freedom integrated into the metric tensor in five dimensions as is the case for the DKK metric tensors (3.13) and inverse (4.22), it is unnecessary to add any additional dimensions in order to pick up the phenomenology of weak and strong interactions. Rather, one simply generalizes abelian electromagnetic gauge theory to non-abelian Yang-Mills gauge theory in the usual way, all within the context of the DKK metric tensors (3.13) and inverse (4.22) and the geodesic equation of motion and Einstein equation machinery that goes along with them. Then the trick is to pick the right gauge group, the right particle representations, and the right method of symmetry breaking.

So from this line of approach, it seems as though one would first regard the U(1) gauge fields $A_{\gamma\mu}$ which are already part of the five dimensional DKK metric tensor (3.13), as non-abelian SU(N) gauge fields $G_\mu = T^i G^i{}_\mu$ with internal symmetry established by the group generators T^i which have a commutation relation $\left[T^i, T^j\right] = if^{ijk}T^k$ with group structure constants f^{ijk}. Prior

to any symmetry breaking each gauge field would have only two degrees of freedom and so be massless and luminous just like the photon because this constraint naturally emerges from (2.10). Then, starting with the metric tensor (3.13), one would replace $A_{\gamma\mu} \mapsto G_{\gamma\mu} = T^i G^i{}_{\gamma\mu}$ everywhere this field appears (with γ now understood to denote, not a photon, but another luminous field quantum), then re-symmetrize the metric tensor by replacing $G_{\gamma\mu}G_{\gamma\nu} \mapsto \frac{1}{2}\{G_{\gamma\mu}, G_{\gamma\nu}\}$ because these fields are now non-commuting. Then – at the risk of understating what is still a highly nontrivial problem – all we need do is discover the correct Yang-Mills GUT gauge group to use for these $G_{\gamma\mu}$, discover what particles are associated with various representations of this group, discover the particular way or ways in which the symmetry of this GUT group is broken and at what energy stages including how to add an extra degree of freedom to some of these $G_{\gamma\mu}$ or combinations of them to give them a mass such as is required for the weak W and Z bosons, discover the origin of the chiral asymmetries observed in nature such as those of the weak interactions, discover how the observed fermion phenomenology becomes replicated into three fermion generations, discover how to produce the observed $G \supset SU(3)_C \times SU(2)_W \times U(1)_{em}$ phenomenology observed at low energies, and discover the emergence during symmetry breaking of the observed baryons and mesons of hadronic physics, including protons and neutrons with three confined quarks. How do we do this?

There have been many GUT theories proposed since 1954 when Yang-Mills theory was first developed, and the correct choice amongst these theories is still on open question. As an example, in an earlier paper [29] the author did address these questions using a $G = SU(8)$ GUT group in which the up and down quarks with three colors each and the electron and neutrino leptons form the 8 components of an octuplet $\left(\nu, (u_R, d_G, d_B), e, (d_R, u_G, u_B)\right)$ in the fundamental representation of SU(8), with (u_R, d_G, d_B) having the quark content of a neutron and (d_R, u_G, u_B) the quark content of a proton. Through three stages of symmetry breaking at the Planck energy, at a GUT energy, and at the Fermi vev energy, this was shown to settle into the observed $SU(3)_C \times SU(2)_W \times U(1)_{em}$ low-energy phenomenology including the condensing of the quark triplets into protons and neutrons, the replication of fermions into three generations, the chiral asymmetry of weak interactions, and the Cabibbo mixing of the left-chiral projections of those generations. As precursor to this SU(8) GUT group, in [48] and [30], based on [47], it was shown that the nuclear binding energies of fifteen distinct nuclides, namely ^{2}H, ^{3}H, ^{3}He, ^{4}He, ^{6}Li, ^{7}Li, ^{7}Be, ^{8}Be, ^{10}B, ^{9}Be, ^{10}Be, ^{11}B, ^{11}C, ^{12}C and ^{14}N, are genomic "fingerprints" which can be used to establish "current quark" masses for the up and down quarks to better than 1 part in 10^5 and in some cases 10^6 *for all fifteen nuclides*, entirely independently of the renormalization scheme that one might otherwise use to characterize current quark masses. This is because one does not need to probe the nucleus at all to ascertain quark masses, but merely needs to decode the mass defects, alternatively nuclide weights, which are well-known with great precision and are independent of observational methodology. Then, in [7.6] of [31], the quark masses so-established by decoding the fingerprints of the light nucleon mass defects, in turn, were used to retrodict the observed masses of the proton and neutron as a function of only these up and down quark masses and the Fermi vev and a determinant of the CKM mixing matrix, within all experimental errors for all of these input and output parameters, *based directly on the SU(8) GUT group and particle representation and symmetry breaking cascade of [29].* So if one were to utilize the author's

example of a GUT, the $A_{\gamma\mu} \mapsto G_{\gamma\mu} = T^i G^i{}_{\gamma\mu}$ in the DKK metric (3.13) would be regarded to have an SU(8) symmetry with the foregoing octuplet in its fundamental representation. Then one would work through the same symmetry breaking cascade, but now also having available the equation of motion (7.2) and the Einstein equation (10.8) so that the motion for all interactions is strictly geodesic motion and the field dynamics and energy tensors are at bottom strictly geometrodynamic and fully gravitational.

In 2019, the scientific community will celebrate the centennial of Kaluza-Klein theory. Throughout this entire century, Kaluza-Klein theory has been hotly debated and has had its staunch supporters and its highly-critical detractors. And both are entirely justified. The miracle of geometrizing Maxwell's electrodynamics and the Lorentz motion and the Maxwell stress-energy tensors in a theory which is unified with gravitation and turns Einstein's "wood" tensor into the "marble" of geometry is tremendously attractive. But a theory which is rooted in a degenerate metric tensor with a singular inverse and a scalar field which carries the entire new dimension on its shoulders and which contains an impossible-to-observe curled up fourth space dimension, not to mention a structural incompatibility with Dirac theory and thus no ability to account for fermion phenomenology, is deeply troubling.

By using Dirac theory itself to force five-dimensional general covariance upon Kaluza-Klein theory and cure all of these troubles while retaining all the Kaluza miracles and naturally and covariantly breaking the symmetry of the gauge fields by removing two degrees of freedom and thereby turning classical fields into quantum fields, to uncover additional new knowledge about our physical universe in the process, and to possibly lay the foundation for all-interaction unification, we deeply honor the work and aspirations of our physicist forebears toward a unified geometrodynamic understanding of nature as the Kaluza-Klein centennial approaches.

PART II: THE DIRAC-KALUZA-KLEIN SCALER, THE HIGGS FIELD, AND A THEORY OF FERMION MASSES, MIXING ANGLES AND BETA DECAYS WHICH FITS THE EXPERIMENTAL DATA

11. Spontaneous Symmetry Breaking of the Massless Luminous Dirac-Kaluza-Klein Scalar

Returing to where we left of at the end of section 8, let us now find out more about the new component cp^5 in the energy-momentum five-vector $cp^{\mathrm{M}} = \left(E \quad cp^j \quad cp^5\right)$ defined prior to the momentum space Dirac equation (5.7). Leaving p^{Σ} in $U_0\left(p^{\Sigma}\right)$ implicitly understood, we first swap upper and lower indexes in (5.7) and expand using the three-part metric tensor (3.8) as such:

$$\begin{aligned} 0 &= \left(\Gamma_{\mathrm{M}} cp^{\mathrm{M}} - mc^2\right)U_0 = \left(\Gamma_0 cp^0 + \Gamma_j cp^j + \Gamma_5 cp^5 - mc^2\right)U_0 \\ &= \left(\gamma_0 E + \phi k A_{\gamma k}\gamma_k E + \gamma_j cp^j + \phi k A_{\gamma j}\gamma_0 cp^j + \gamma_5 cp^5 + \phi\gamma_0 cp^5 - mc^2\right)U_0 \end{aligned} . \tag{11.1}$$

Now, mc^2 is the rest energy of the fermion. This is placed into the Dirac equation by hand, and it originates in the relativistic energy-momentum relation $m^2c^4 = g_{\mu\nu}p^\mu p^\nu$ into which it is also placed by hand. It would be very desirable to give this rest mass an interpretation purely in terms of the geometry and the cp^5 component of the five-momentum so it need not be entered by hand, and even more desirable if the scalar ϕ which we now know is luminous and massless can be used in a manner analogous to the Higgs mechanism to break symmetry and enable us to predict or at least better understand the observed pattern of fermion rest masses.

Toward these ends, suppose that starting with (11.1) we remove the hand-added mc^2 entirely and instead use the cp^5 terms, *by postulate*, to define the eigenvalue relation:

$$\Gamma_5 cp^5 U_0 = \left(\gamma_5 cp^5 + \phi\gamma_0 cp^5\right)U_0 \equiv -mc^2 U_0 . \tag{11.2}$$

By this postulate from which we shall now explore the implications, a fermion rest energy mc^2 represents the eigenvalues of the operator $-\Gamma_5 cp^5 = -\gamma_5 cp^5 - \phi\gamma_0 cp^5$. Not only are γ_0 and γ_5 4x4 Dirac operators as always, but from the result in (8.14), so too is the luminous scalar ϕ.

This highlights some very important points regarding using spontaneous symmetry breaking to arrive at a *fermion* rest masses which it is not presently known how to do in detail, as opposed to arriving at gauge boson rest masses which it is known how to do. Specifically, when a scalar field (also denoted ϕ) is used to break symmetry, for example, for a triplet of three weak interaction gauge fields $W^{a\mu}$ in the adjoint representation of a local SU(2) Yang-Mills [32] gauge group where $a = 1,2,3$ is an internal symmetry index associated with the SU(2) generators τ^a which have a commutator relation $\left[\tau^a, \tau^b\right] = \varepsilon^{abc}\tau^c$ (see section 14.9 of [20]), the scalars are placed into the fundamental representation of SU(2) whereby $\phi^T = \begin{pmatrix}\phi_\alpha & \phi_\alpha\end{pmatrix} = \begin{pmatrix}\phi_1 + i\phi_2 & \phi_3 + i\phi_4\end{pmatrix}$ is an SU(2) *doublet* of complex scalars providing four scalar degrees of freedom. This structural matching of the scalars in the fundamental representation of SU(2) with the gauge bosons in the adjoint representation of SU(2) enables the scalars to be coupled to the gauge fields the Lagrangian density term $g_W{}^2\phi^\dagger\tau^a W^a{}_\mu\tau^b W^{b\mu}\phi$, which coupling underlies the spontaneous symmetry breaking (note $\tau^a = \tau^{a\dagger}$ are Hermitian). So, if we restore into (11.2) the Fourier kernel and thus $\Psi \equiv U_0 \exp\left(-ip_\Sigma x^\Sigma / \hbar\right)$ specified prior to (5.7), and knowing the form of the Lagrangian density for the fermion rest masses, and explicitly showing the normally-implicit Dirac spinor indexes *A*, *B*, *C*, all ranging from 1 to 4, we see looking closely at (11.2) that the Lagrangian density will contain a term $\overline{\Psi}_A\phi_{AB}\gamma_{0BC}cp^5\Psi_C$. In other words, $\phi = \phi_{AB}$ in (8.14) couples perfectly to Dirac fermion wavefunctions, so symmetry can be broken and the fermions obtain rest masses.

This means that the seeming "oddity" of the luminous scalar having picked up a second rank Dirac structure in (8.14) in order to have $d\phi/d\tau = 0$ in (8.1) so that (7.29) can be covariantly collapsed to precisely reproduce the Lorentz force motion as geodesic motion in the Kaluza-Klein

geometry, actually makes perfect sense in view of (11.2): Gauge bosons have a Yang-Mills internal symmetry structure against which must be matched by the internal symmetries of the scalars used to spontaneously break symmetry and give mass to these gauge bosons via the Higgs mechanism, so that the scalars properly couple to the bosons. Likewise, *fermions* have a Dirac spinor structure (in addition to their Yang-Mills internal structure) against which we have to expect any scalars used to spontaneously break symmetry and give mass to the fermions will also have to have to be matched with a *Dirac structure*, so the scalars properly couple to the fermions. So, the luminous scalar (8.14) having a Dirac structure which couples with the Dirac structure of the fermions is in precisely the same league as the scalars used to break gauge boson symmetries having an internal symmetry structure to couple the internal symmetry of the gauge bosons. And it is in the same league, for example, as having to use a spin connection (see, e.g., [33]) for fermions to be able to covariantly couple to gravitation. So, notwithstanding the "oddity" of the scalar in (8.14) picking up Dirac structure, *this luminous massless scalar (8.14) turns out to be ready-made for generating fermion rest masses through spontaneous symmetry breaking using the Higgs mechanism*.

Finally, when we do the accounting for degrees of freedom, the luminous massless scalar (8.14) is also perfectly matched to generate fermion masses while also generating a massive Higgs scalar. By way of contrast, with a subscript *H* used to denote the standard Higgs mechanism, a scalar which we write as $\phi_h = \left(\phi_{1h} + i\phi_{2h}\right)/\sqrt{2}$ used to break a local U(1) gauge symmetry starts out with two scalar degrees of freedom provided by ϕ_{1H} and ϕ_{2H}. And of course, $\phi_h^{\dagger}\phi_h = \frac{1}{2}\left(\phi_{1h}{}^2 + \phi_{2h}{}^2\right)$ defines the "circle" for symmetry breaking. One of these degrees of freedom is "swallowed" by a gauge boson which starts out massless with two degrees of freedom (see, for example, (2.11b) for the photon polarization) and thereby becomes massive by acquiring a longitudinal polarization. The other degree of freedom is swallowed by a Higgs scalar $h(t,\mathbf{x})$ introduced by the expansion $\phi_h(t,\mathbf{x}) = v + h(t,\mathbf{x})$ about the vacuum vev *v*, thereby giving mass to that scalar. The empirical observation at CERN of the Higgs scalar [34] with the theoretically-anticipated mass is perhaps one of the most significant scientific events of the past few decades.

Here (8.14) contains the same form of expression $\left(\phi_1 + i\phi_2\right)/\sqrt{2}$ used in the Higgs mechanism. Likewise, $\phi^*\phi = \frac{1}{2}\left(\phi_1{}^2 + \phi_2{}^2\right)$ defines the circle for symmetry breaking. So, these fields ϕ_1 and ϕ_2 carry two degrees of freedom available to be "swallowed" by other particles during symmetry breaking via the Goldstone mechanism [35]. But there are two important differences which we shall now study: First, (8.14) has an additional the angle $\theta = \omega\mathrm{T}$ in the Fourier kernel, which can be oriented in any direction as an additional aspect of symmetry breaking and used to provide a third degree of freedom which can also be swallowed by other particles. Second, in contrast to the usual Higgs scalars, ϕ_1 and ϕ_2 in (8.14) are presently dimensionless, whereas in $\phi_h = \left(\phi_{1h} + i\phi_{2h}\right)/\sqrt{2}$ these have energy dimension and so can be connected with $\phi_h(t,\mathbf{x}) = v + h(t,\mathbf{x})$ following symmetry breaking to create a scalar field expansion about the Fermi vacuum. So, we will need to find a way to introduce an energy dimensionality the fields in to (8.14), which we shall do shortly.

Further, it is well-known that any hypothetical "massless" fermion would carry two degrees of freedom and be fully chiral: Consider that a generation ago, when neutrinos were thought to be massless before this was disproven by leptonic neutrino oscillations, the massless $v_L = \frac{1}{2}\left(1+\gamma^5\right)v$ would have had only two degrees of freedom, with right-chirality nonexistent. So, for (8.14) to generate a fermion mass, it is necessary that *both degrees of freedom* from ϕ_1 and ϕ_2 in (8.14) go into the fermion, so that the fermion can be bumped up from two to four degrees of freedom and acquire a mass. But this leaves nothing more for the scalar, so we cannot reveal a massive Higgs scalar unless there is a third degree of freedom. This, as we shall now see, is provided by the degree of freedom in $\theta = \omega\Gamma$, and is precisely the benefit of this third degree of freedom, because now masses can be acquired by both the fermions and the Higgs field.

With this overview, let's now proceed with some further calculations using (11.2). First, starting with the Dirac equation (11.1) we initially remove the hand-added mc^2 and so write this as the entirely geometric $0 = \Gamma_\Sigma cp^\Sigma U_0$. Then we reintroduce the mass term, but using (11.2), thus:

$$0 = \Gamma_\Sigma cp^\Sigma U_0 = \left(\Gamma_\sigma cp^\sigma + \Gamma_5 cp^5\right)U_0 = \left(\Gamma_\sigma cp^\sigma - mc^2\right)U_0 = \left(\Gamma_\sigma cp^\sigma + \gamma_5 cp^5 + \phi\gamma_0 cp^5\right)U_0 . \tag{11.3}$$

The fermion mass term is no-longer hand-added, but rather, originates in the fifth-dimensional operator $\Gamma_5 cp^5$, with its usual appearance $0 = \left(\Gamma_\sigma cp^\sigma - mc^2\right)U_0$ when the fifth-dimensional $\Gamma_5 cp^5$ is replaced by $-mc^2$. So, the momentum space Dirac equation (5.7) becomes $\Gamma_\Sigma cp^\Sigma U_0 = 0$, and the configuration space equation (5.6) is now simply $i\hbar c\Gamma^{\mathrm{M}}\partial_{\mathrm{M}}\Psi = 0$, without a hand-added mass.

Next, let us use the anticommutator (3.1) for three interdependent calculations, starting with $\Gamma_\Sigma cp^\Sigma U_0 = 0$ and $\Gamma_\sigma cp^\sigma U_0 = mc^2 U_0$ and $\Gamma_5 cp^5 U_0 = mc^2 U_0$ all of which are contained in (11.3), and the last of which is also (11.2). In all cases, we "square" the operators using an anticommutator, strip off the operand, and apply (3.1) to obtain, respectively:

$$\Gamma_{\mathrm{M}} cp^{\mathrm{M}}\Gamma_{\mathrm{N}} cp^{\mathrm{N}} = \tfrac{1}{2}\left\{\Gamma_{\mathrm{M}},\Gamma_{\mathrm{N}}\right\}cp^{\mathrm{M}}cp^{\mathrm{N}} = G_{\mathrm{MN}}cp^{\mathrm{M}}cp^{\mathrm{N}} = 0 , \tag{11.4a}$$

$$\Gamma_\mu cp^\mu\Gamma_\nu cp^\nu = \tfrac{1}{2}\left\{\Gamma_\mu,\Gamma_\nu\right\}cp^\mu cp^\nu = G_{\mu\nu}cp^\mu cp^\nu = m^2c^4 , \tag{11.4b}$$

$$\Gamma_5 cp^5\Gamma_5 cp^5 = G_{55}cp^5cp^5 = m^2c^4 . \tag{11.4c}$$

Note that $G_{\mu\nu}cp^\mu cp^\nu = m^2c^4$ in (11.4b) is just the usual form of the relativistic energy momentum relation prior to applying local gauge symmetry. Expanding (11.4a) in two-part form, we obtain:

$$0 = G_{\mu\nu}cp^\mu cp^\nu + 2G_{\mu 5}cp^\mu cp^5 + G_{55}cp^5cp^5 , \tag{11.5}$$

which we may then combine with (11.4b) and (11.4c) to write the chain of relations:

$$m^2c^4 = G_{\mu\nu}cp^\mu cp^\nu = G_{55}cp^5cp^5 = -G_{\mu 5}cp^\mu cp^5 . \tag{11.6}$$

As we shall see in section ???, (11.6) can be used to derive Weyl's local U(1) gauge theory [5], [6], [7] from Kaluza-Klein theory, but for the moment, we remain focused on spontaneous symmetry breaking to generate fermion rest masses.

Equation (11.4a) leads to a very interesting and important consequence for the five-dimensional metric line element $d\mathrm{S} = cd\mathrm{T}$ defined by:

$$c^2 d\mathrm{T}^2 \equiv G_{\mathrm{MN}} dx^{\mathrm{M}} dx^{\mathrm{N}} . \tag{11.7}$$

Specifically, if we further define the five-momentum in terms of mass and motion in the usual way by $cp^{\mathrm{M}} \equiv mc^2 dx^{\mathrm{M}} / cd\tau$ where $c^2 d\tau^2 \equiv G_{\mu\nu} dx^\mu dx^\nu$ is the four-dimensional line element, and if we then multiply (11.7) above through by $m^2c^4 / c^2 d\tau^2$, we obtain:

$$m^2c^4 \frac{d\mathrm{T}^2}{d\tau^2} = G_{\mathrm{MN}} mc^2 \frac{dx^{\mathrm{M}}}{cd\tau} mc^2 \frac{dx^{\mathrm{N}}}{cd\tau} = G_{\mathrm{MN}} cp^{\mathrm{M}} cp^{\mathrm{N}} . \tag{11.8}$$

Then, comparing (11.8) with (11.4a) which is equal to zero and identical to (11.8), and presuming non-zero $m \neq 0$ and $d\tau \neq 0$, the five-dimensional infinitesimal line element must also be zero:

$$\boxed{d\mathrm{S} = cd\mathrm{T} = 0}. \tag{11.9}$$

This is a very important and useful result, and it is one of the direct consequences of the postulated eigenvalue relation (11.2) for the fermion rest mass.

Our first use of this result, will be to break the symmetry of the sine and cosine terms in (8.14). In this regard, what we learn from (11.9) is that any *finite* five-dimensional proper metric interval $\mathrm{S} = c\mathrm{T} = \int d\mathrm{S} = \int cd\mathrm{T} = \mathrm{S}_0 = c\mathrm{T}_0$ obtained from (11.9) whether of length or time dimensionality can only be a constant of integration $\mathrm{S}_0 = c\mathrm{T}_0$. And this in turn means that T in (8.14) is zero up to a constant of integration, and specifically, that $\mathrm{T} = 0 + \mathrm{T}_0$. So, we now wish to use this finding in the most advantageous way possible.

Toward this end, starting with $\cos(\omega\mathrm{T})$ in (8.14), let us break the symmetry in the plane of the angle $\theta = \omega\mathrm{T}$ by imposing the symmetry-breaking constraint $\cos(\omega\mathrm{T}) \equiv 1$. This of course means that $\omega\mathrm{T} = \omega\mathrm{T}_0 = 2\pi n$ is quantized, with $n = 0, \pm(1, 2, 3...)$ being any integer. Using $c = \omega\lambda\!\!\!^{-}$ and $\lambda = 2\pi\lambda\!\!\!^{-}$ which we can do because ϕ in (8.14) is massless and luminous, this constraint $\omega\mathrm{T} = 2\pi n$ is alternatively formulated in terms of five-dimensional space-dimensioned proper line elements $\mathrm{S} = c\mathrm{T} = n\lambda$ which are essentially quantized units of five-dimensional length. As well, this means that $\sin(\omega\mathrm{T}) = 0$, but we need to be careful because there is also a $c\mathrm{T}$ in the denominator of the $\sin(\omega\mathrm{T})$ term in (8.14).

So, for the sin term, we insert the same $\omega \mathrm{T} = \omega \mathrm{T}_0 = 2\pi n$, then use $c = \omega \lambda\!\!\!^{-}$ and $\lambda = 2\pi \lambda\!\!\!^{-}$:

$$-i\frac{\Gamma_\Sigma x^\Sigma}{c\mathrm{T}}\sin\left(\omega \mathrm{T}\right) = -i\frac{\Gamma_\Sigma x^\Sigma \omega}{2\pi nc}\sin\left(2\pi n\right) = -i\frac{\Gamma_\Sigma x^\Sigma}{n\lambda}\sin\left(2\pi n\right). \tag{11.10}$$

If we select $n = 0$ which produces a $0/0$, then we deduce from the top line of (8.13) that (11.10) will be equal to $-i\Gamma_\Sigma x^\Sigma / \lambda\!\!\!^{-}$ and *not* be zero. But *for any other integer* $n \neq 0$, the above *will* be equal to zero. So, we break symmetry by restricting *n* to a *non-zero* integer $n = \pm\left(1, 2, 3...\right)$. With this final constraint (11.10) does become zero and (8.14) reduces to:

$$\phi = \frac{1}{\sqrt{2}}\left(\phi_1 + i\phi_2\right)\exp\left(-\frac{\partial_\Sigma \phi_1 + i\partial_\Sigma \phi_2}{\phi_1 + i\phi_2}x^\Sigma\right). \tag{11.11}$$

Having used $\mathrm{S} = c\mathrm{T} = n\lambda$ to break symmetry with $n = \pm\left(1, 2, 3...\right)$ being a positive or negative non-zero integer, we see that *finite* five-dimensional proper lengths are quantized integer multiples of the wavelength λ first specified in (8.5) for the now-luminous Kaluza-Klein scalar field ϕ. This follows a long tradition of quantization based on wavelength fitting which started with Bohr [36] and culminated with DeBroglie [37].

Importantly, with (11.11) we need no longer be concerned with the Dirac operator matrix Γ_Σ in ϕ, because we have broken symmetry so as to effectively diagonalize the operation of this operator. We do however need to be mindful that in breaking symmetry in this way, we have eliminated any overt appearance of the scalar frequency $f = \omega / 2\pi$ or wavelength $\lambda = 2\pi \lambda\!\!\!^{-}$ or energy $hf = \hbar\omega$ of the scalar ϕ, which were overt in (8.14), and particularly, the dimensionless ratio x^Σ / λ in (11.10). This does not mean that the scalar no longer has a frequency or wavelength or energy. Rather, it means that the symmetry breaking has *hidden* these attributes. Making note of all this, we shall find occasion shortly to reintroduce λ as the "initial value" for a constant of integration we will momentarily come across.

We complete the symmetry breaking in the usual way by again noting that $\phi^*\phi = \frac{1}{2}\left(\phi_1^{\,2} + \phi_2^{\,2}\right)$ defines a symmetry breaking circle, and by orienting the scalar in this circle by setting $\phi_2\left(x^\Sigma\right) = 0$. This further reduces (8.14) to its final symmetry-broken form:

$$\boxed{\phi = \frac{1}{\sqrt{2}}\phi_1\exp\left(-\frac{\partial_\Sigma \phi_1}{\phi_1}x^\Sigma\right)}. \tag{11.12}$$

Now let us return to (11.2) where we defined the fermion rest mass strictly in terms of the geometry of $\Gamma_5 = \gamma_5 + \phi\gamma_0$ and the fifth-dimensional component cp^5 of the energy-momentum vector. Into (11.2) we now insert the symmetry-broken (11.12) and restructure, to obtain:

$$0 = \left(\gamma_5 cp^5 + \frac{1}{\sqrt{2}} cp^5 \phi_1 \exp\left(-\frac{\partial_\Sigma \phi_1}{\phi_1} x^\Sigma \right) \gamma_0 + mc^2 \right) U_0 \,. \tag{11.13}$$

Of special interest in (11.13), is that whereas ϕ_1 has all along been physically dimensionless, now in (11.13) this is multiplied by cp^5 which has dimensions of energy. This means that $cp^5 \phi_1$ now has precisely the same characteristics as ϕ_{1h} in the scalar field $\phi_h = \left(\phi_{1h} + i\phi_{2h} \right) / \sqrt{2}$ employed in standard model Higgs field theory, and being an energy-dimensioned scalar field, may therefore present the opportunity for a connection with the standard model Higgs field *h*.

To pursue this possibility, we first use the Dirac representation of γ_M to write (11.13) as:

$$0 = \begin{pmatrix} mc^2 + \frac{1}{\sqrt{2}} cp^5 \phi_1 \exp\left(-\frac{\partial_\Sigma \phi_1}{\phi_1} x^\Sigma \right) & cp^5 \\ cp^5 & mc^2 - \frac{1}{\sqrt{2}} cp^5 \phi_1 \exp\left(-\frac{\partial_\Sigma \phi_1}{\phi_1} x^\Sigma \right) \end{pmatrix} \begin{pmatrix} U_{0A} \\ U_{0B} \end{pmatrix}. \tag{11.14}$$

The eigenvalues are obtained by setting the determinant of the above matrix to zero as such:

$$\left(mc^2 \right)^2 - \left(cp^5 \right)^2 - \left[\frac{1}{\sqrt{2}} cp^5 \phi_1 \exp\left(-\frac{\partial_\Sigma \phi_1}{\phi_1} x^\Sigma \right) \right]^2 = 0 \,. \tag{11.15}$$

Restructuring and taking both the $\pm$ square roots, we then obtain the eigenvalues:

$$\pm \sqrt{\left(mc^2 \right)^2 - \left(cp^5 \right)^2} = \frac{1}{\sqrt{2}} cp^5 \phi_1 \exp\left(-\frac{\partial_\Sigma \phi_1}{\phi_1} x^\Sigma \right). \tag{11.16}$$

The above now presents a differential equation for ϕ_1 as a function of the five-dimensional x^Σ.

To solve this equation, we first restructure a bit, then take the natural log of both sides, then use the identity $\ln\left(AB \right) = \ln A + \ln B$ to obtain:

$$\ln\left(\pm \frac{\sqrt{2}\sqrt{\left(mc^2 \right)^2 - \left(cp^5 \right)^2}}{cp^5} \right) = \ln\left(\phi_1 \exp\left(-\frac{\partial_\Sigma \phi_1}{\phi_1} x^\Sigma \right) \right) = \ln \phi_1 - \frac{\partial_\Sigma \phi_1}{\phi_1} x^\Sigma \,. \tag{11.17}$$

Then we isolate the rightmost term in the above, use $\partial_\Sigma = \partial / \partial x^\Sigma$ and $\ln\left(A / B \right) = \ln A - \ln B$, then further simplify, as such:

$$-\frac{1}{\phi_1}\frac{\partial\phi_1}{\partial x^\Sigma}x^\Sigma = \ln\left(\pm\frac{\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}}{cp^5}\right)-\ln\phi_1 = \ln\left(\pm\frac{\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}}{cp^5\phi_1}\right). \qquad (11.18)$$
$$= \ln\left(\pm\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}\right)-\ln\left(cp^5\phi_1\right)$$

Then we finally restructure this into:

$$\frac{1}{x^\Sigma}\partial x^\Sigma = -\frac{1}{\ln\left(\pm\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}\right)-\ln\left(cp^5\phi_1\right)}\frac{1}{\phi_1}\partial\phi_1 . \qquad (11.19)$$

Now, we set up an integral by converting $\partial \to d$ and placing an indefinite integral sign to operate on each side. And, to simplify the integration, we briefly define the substitute variables $y \equiv \phi_1$, $A \equiv \ln\left(\pm\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}\right)$ and $B \equiv cp^5$. Then we carry out the integration. Prior to the equal sign we employ an integration constant defined by $C \equiv \ln\left(1/L^5\right)$ with L^5 being a constant that has dimensions of length to the fifth power. Then we conclude by replacing the substitute variables, to obtain:

$$\int\frac{1}{x^\Sigma}dx^\Sigma = \int\left(\frac{dx^0}{x^0}+\frac{dx^1}{x^1}+\frac{dx^2}{x^2}+\frac{dx^3}{x^3}+\frac{dx^5}{x^5}\right) = \ln x^0+\ln x^1+\ln x^2+\ln x^3+\ln x^5+\ln\frac{1}{L^5}$$
$$= \ln\left(\frac{x^0x^1x^2x^3x^5}{L^5}\right) = \int -\frac{1}{A-\ln\left(By\right)}\frac{1}{y}dy = \ln\left(A-\ln\left(By\right)\right) \quad .$$
$$= \int -\frac{1}{\ln\left(\pm\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}\right)-\ln\left(cp^5\phi_1\right)}\frac{1}{\phi_1}d\phi_1 = \ln\left(\ln\left(\pm\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}\right)-\ln\left(cp^5\phi_1\right)\right)$$

(11.20)

The middle line includes using the generalized $\int -\left(1/\left(A-\ln\left(By\right)\right)y\right)dy = \ln\left(A-\ln\left(By\right)\right)+C$.

The upshot, now exponentiating each side and again using $\ln\left(A/B\right) = \ln A - \ln B$, is:

$$\frac{x^0x^1x^2x^3x^5}{L^5} = \ln\left(\pm\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}\right)-\ln\left(cp^5\phi_1\right) = \ln\left(\pm\frac{\sqrt{2}\sqrt{\left(mc^2\right)^2-\left(cp^5\right)^2}}{cp^5\phi_1}\right). \qquad (11.21)$$

Exponentiating one final time and isolating the energy dimensioned field $cp^5\phi_1$, and denoting the five dimensional coordinate set as $x^{\mathrm{M}} \equiv \mathbf{X}$, the final result is:

$$\boxed{cp^5\phi_1(\mathbf{X}) = \pm\sqrt{2}\sqrt{\left(mc^2\right)^2 - \left(cp^5\right)^2}\,\exp\left(-\frac{x^0x^1x^2x^3x^5}{L^5}\right)}. \tag{11.22}$$

The numerator inside the exponent, which we may write in consolidate form as $V_{(5)} \equiv x^0x^1x^2x^3x^5$, is a five-dimensional volume with dimensions of length to the fifth power. Because the argument of the exponential is required to be dimensionless, this means that the constant of integration is embodied in L^5 is likewise required to have dimensions of length to the fifth power. This is the first of several "initial conditions" we will utilize to determine this integration constant. Later, cognizant that the ratio x^Σ / λ in (11.10) and more generally the energy and wavelength of the scalar field ϕ of (8.14) became hidden when we broke symmetry, we will add to this initial condition an expectation L^5 relate in some way to the wavelength of a scalar field, at fifth order.

Now, to see how $cp^5\phi_1$ connects to Higgs fields, we turn our attention obtaining a direct expression for the fifth component of the energy momentum, $cp^5 = mc^2\left(dx^5 / cd\tau\right)$ which appears particularly in the radical above.

12. The Fifth-Dimensional Component of the Dirac-Kaluza-Klein Energy Momentum Vector

To directly study cp^5, recall that (7.28) connects electric charge to motion in the Kaluza-Klein fifth dimension. So, using $dx^5 / cd\tau = q / \phi^2 kmc^2$ in (7.28), and also borrowing *k* from (1.2), we obtain:

$$cp^5 = mc^2\frac{dx^5}{cd\tau} = \frac{q}{\phi^2 k} = \frac{qc^2}{2\phi^2}\sqrt{\frac{k_e}{G}}\,. \tag{12.1}$$

Formally speaking, we have not yet proved that (1.2) is the correct value of *k* for the DKK metric tensor (3.13), see also (3.12). Rather, we have "borrowed" the value for *k* which is determined using the ordinary Kaluza-Klein metric tensor (1.1) in the five-dimensional Einstein equation. When this calculation is carried out, included amidst the expressions obtained is the term combination $g^{\alpha\beta}F_{\mu\alpha}F_{\nu\beta} - \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}$ recognizable as the body of the Maxwell stress-energy tensor, see, e.g. [11]. Then, the definition (1.2) is required to match this body with its correct coefficients in the stress-energy. However, the DKK metric tensor does not *omit* any of the terms in (1.1). Rather, referring to (3.11) for $g_{\mathrm{MN}} = \eta_{\mathrm{MN}}$ in view of $A_{\gamma 0} = 0$, it merely adds terms while fixing the gauge field via $A_\mu \mapsto A_{\gamma\mu}$ to that of a photon field quantum. In particular, it adds a 1 to ϕ^2 in G_{55}, and it adds a ϕ to $A_{\gamma 0} = 0$ in G_{05}, while fixing $A_\mu \mapsto A_{\gamma\mu}$. Moreover, we proved in section 7 how the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ appears in the DKK equation of motion just as

it does in ordinary Kaluza-Klein theory following the gauge transformation $F_{\gamma\mu\nu} \rightarrow F_{\mu\nu}$ reviewed prior to (7.12), and at (7.28) how electric charge becomes connected to fifth-dimensional motion in the exact same way. There are additional terms in DKK, but no terms are lost. So, there is every reason to expect that the exact same stress energy body $g^{\alpha\beta}F_{\mu\alpha}F_{\nu\beta} - \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}$ will appear when the DKK metric tensor (3.13) is used in the five-dimensional Einstein equation, and that *k* will likewise turn out to be exactly the same as it is in (1.2). It is for this reason, in advance of a detailed calculation of the five-dimensional Einstein equation using the DKK (3.13) which will be the subject of a subsequent paper not this paper, that we "borrow" *k* from (1.2). But we shall also continue to show *k* in our calculations, in order to also obtain results without this borrowing.

We then combine (12.1) with $G_{55}cp^5cp^5 = m^2c^4$ from (11.4c) and $G_{55} = 1+\phi^2$ from (3.13) and (3.12) when $g_{\text{MN}} = \eta_{\text{MN}}$, and also borrow (1.2), to obtain:

$$G_{55}cp^5cp^5 = \left(1+\phi^2\right)\frac{q^2}{\phi^4k^2} = \left(1+\phi^2\right)\frac{k_eq^2c^4}{4G\phi^4} = m^2c^4. \tag{12.2}$$

This easily restructures into a quadratic for ϕ^2, which we write as:

$$0 = \frac{m^2c^4k^2}{q^2}\phi^4 - \phi^2 - 1 = 4\frac{Gm^2}{k_eq^2}\phi^4 - \phi^2 - 1, \tag{12.3}$$

which we see includes the very small dimensionless ratio Gm^2 / k_eq^2 of gravitational-to-electrical interaction strength for a charge *q* with mass *m*.

The next step is to solve the quadratic equation for (12.3). But first, because *q* and *m* in (12.3) are the charges and masses of individual fermions given the genesis of (12.3) in the DKK Dirac equation (11.3), it will be helpful to rewrite this ratio to facilitate downstream calculation. First, we observe that $q = Qe$ for any individual fermion, where *Q* is the electric charge generator for that fermion, and where the charge strength *e* is related to the electromagnetic running coupling by $\alpha = k_ee^2 / \hbar c$, with $\alpha = 1/137.035999139$ being the low-probe value of the running fine structure number as reported in [21]. The charge generator $Q = -1$ for the e,μ,τ leptons, $Q = +2/3$ for the u,c,t quarks and $Q = -1/3$ for the d,s,b quarks, and have reversed signs for the antiparticles. Also, we note that the Planck mass $M_P = 1.220910\times10^{19}$ GeV / c^2 using the value reported in [38], is *defined* as the mass for which the coupling strength $GM_P{}^2 \equiv \hbar c$. Therefore, we may calculate that the ratio Gm^2 / k_eq^2 in (12.3) may be rewritten as:

$$\frac{Gm^2}{k_eq^2} = \frac{Gm^2}{Q^2k_ee^2} = \frac{Gm^2}{Q^2\alpha\hbar c} = \frac{Gm^2}{Q^2\alpha GM_P{}^2} = \frac{m^2}{Q^2\alpha M_P{}^2}. \tag{12.4a}$$

For the square root of this, which will also be of interest, we write:

$$\sqrt{\frac{Gm^2}{k_e q^2}} = \sqrt{\frac{G}{k_e}}\frac{m}{q} = \sqrt{\frac{G}{k_e}}\frac{m}{Qe} = \frac{\sqrt{G}m}{Q\sqrt{\alpha}\sqrt{\hbar c}} = \frac{m}{Q\sqrt{\alpha}M_P} \tag{12.4b}$$

without the $\pm$ that regularly arises when taking a square root, because masses such as m and M_P are always taken to be positive numbers, because $\sqrt{\alpha}$ is always taken to be a positive dimensionless measure of charge strength, and because it is important to maintain the proper positive or negative sign for Q without washing it out with a $\pm$. The above enables us to readily use each fermion's m/M_P ratio, as well as to directly account for its positive or negative Q.

Solving (12.3) with the quadratic equation, and using (12.4), the positive and negative roots are found to be at:

$$\phi_\pm{}^2 = \frac{q^2}{2m^2c^4k^2}\left(1\pm\sqrt{1+4\frac{m^2c^4k^2}{q^2}}\right) = \frac{k_e q^2}{8Gm^2}\left(1\pm\sqrt{1+16\frac{Gm^2}{k_e q^2}}\right) = \frac{Q^2\alpha M_P{}^2}{8m^2}\left(1\pm\sqrt{1+16\frac{m^2}{Q^2\alpha M_P{}^2}}\right). \tag{12.5}$$

Placing this into ϕ^2 in (12.1) we arrive at two corresponding values for cp^5, namely:

$$cp_\pm{}^5 = \frac{2m^2c^4k}{q}\frac{1}{1\pm\sqrt{1+4\frac{m^2c^4k^2}{q^2}}} = 4\sqrt{\frac{Gm^2}{k_e q^2}}\frac{mc^2}{1\pm\sqrt{1+16\frac{Gm^2}{k_e q^2}}} = 4\frac{m}{Q\sqrt{\alpha}M_P}\frac{mc^2}{1\pm\sqrt{1+16\frac{m^2}{Q^2\alpha M_P{}^2}}} \tag{12.6}$$

Applying what we now write as $cp_\pm{}^5 = mc^2\left(dx_\pm{}^5/cd\tau\right)$, it is also helpful to obtain:

$$\frac{dx^5{}_\pm}{cd\tau} = \frac{2mc^2k}{q}\frac{1}{1\pm\sqrt{1+4\frac{m^2c^4k^2}{q^2}}} = 4\sqrt{\frac{Gm^2}{k_e q^2}}\frac{1}{1\pm\sqrt{1+16\frac{Gm^2}{k_e q^2}}} = 4\frac{m}{Q\sqrt{\alpha}M_P}\frac{1}{1\pm\sqrt{1+16\frac{m^2}{Q^2\alpha M_P{}^2}}} \tag{12.7}$$

for the "motion" $dx^5{}_\pm/cd\tau = dt_\pm{}^5/d\tau$, which is really a rate of time progression through the timelike fifth DKK dimension. Also, because the DKK metric tensor component $G_{55} = 1+\phi^2$ for $g_{\mathrm{MN}} = \eta_{\mathrm{MN}}$, see (3.11), which we now write as $G_{55\pm} = 1+\phi_\pm{}^2$, it is also useful to employ (12.5) to write these two solutions as:

$$\begin{aligned} G_{55\pm} = 1+\phi_\pm{}^2 &= 1+\frac{q^2}{2m^2c^4k^2}\left(1\pm\sqrt{1+4\frac{m^2c^4k^2}{q^2}}\right) = 1+\frac{k_e q^2}{8Gm^2}\left(1\pm\sqrt{1+16\frac{Gm^2}{k_e q^2}}\right) \\ &= 1+\frac{Q^2\alpha M_P{}^2}{8m^2}\left(1\pm\sqrt{1+16\frac{m^2}{Q^2\alpha M_P{}^2}}\right) \end{aligned}, \tag{12.8}$$

Now, the ratio $16Gm^2 / k_e q^2 = 16m^2 / Q^2 \alpha M_P{}^2 \ll 1$ inside the radicals above is a very small number for all of the elementary fermions with an electrical charge $Q \neq 0$, because the ratio $m^2 / M_P{}^2$ is on the order of 10^{-40} for all of the known fermions. Moreover, even if we had not "borrowed" from (1.2), we likewise expect $4m^2c^4k^2 / q^2 \ll 1$ to be a very small number. Therefore, it is helpful to use the first three terms of the series expansion $\sqrt{1+x^2} = 1 + \frac{1}{2}x^2 - \frac{1}{8}x^4 + \ldots$ in each of (12.5) through (12.7) to obtain:

$$\begin{aligned} \phi_\pm{}^2 &= \frac{q^2}{2m^2c^4k^2} \pm \left(\frac{q^2}{2m^2c^4k^2} + 1 - \frac{m^2c^4k^2}{q^2} + \ldots \right) = \frac{k_e q^2}{8Gm^2} \pm \left(\frac{k_e q^2}{8Gm^2} + 1 - 4\frac{Gm^2}{k_e q^2} + \ldots \right) \\ &= \frac{Q^2\alpha M_P{}^2}{m^2} \pm \left(\frac{Q^2\alpha M_P{}^2}{m^2} + 1 - 4\frac{m^2}{Q^2\alpha M_P{}^2} + \ldots \right) \end{aligned} , \tag{12.9}$$

$$\begin{aligned} cp_\pm{}^5 &= \frac{2m^2c^4k}{q} \frac{1}{1 \pm \left(1 + 2\frac{m^2c^4k^2}{q^2} - 2\left(\frac{m^2c^4k^2}{q^2} \right)^2 + \ldots \right)} = 4\sqrt{\frac{Gm^2}{k_e q^2}} \frac{mc^2}{1 \pm \left(1 + 8\frac{Gm^2}{k_e q^2} - 32\left(\frac{Gm^2}{k_e q^2} \right)^2 + \ldots \right)} \\ &= 4\frac{m}{Q\sqrt{\alpha} M_P} \frac{mc^2}{1 \pm \left(1 + 8\frac{m^2}{Q^2\alpha M_P{}^2} - 32\left(\frac{m^2}{Q^2\alpha M_P{}^2} \right)^2 + \ldots \right)} \end{aligned} , \tag{12.10}$$

$$\begin{aligned} \frac{dx_\pm{}^5}{cd\tau} &= \frac{2mc^2k}{q} \frac{1}{1 \pm \left(1 + 2\frac{m^2c^4k^2}{q^2} - 2\left(\frac{m^2c^4k^2}{q^2} \right)^2 + \ldots \right)} = 4\sqrt{\frac{Gm^2}{k_e q^2}} \frac{1}{1 \pm \left(1 + 8\frac{Gm^2}{k_e q^2} - 32\left(\frac{Gm^2}{k_e q^2} \right)^2 + \ldots \right)} \\ &= 4\frac{m}{Q\sqrt{\alpha} M_P} \frac{1}{1 \pm \left(1 + 8\frac{m^2}{Q^2\alpha M_P{}^2} - 32\left(\frac{m^2}{Q^2\alpha M_P{}^2} \right)^2 + \ldots \right)} \end{aligned} . \tag{12.11}$$

Now let's consider the separate $\pm$ solutions originating when we applied the quadratic equation to (12.3), as well as certain inequalities. Using $m^2 / M_P{}^2 \ll 1$ which is valid to a 1 part in 10^{40} approximation, (12.9) separates into:

$$\begin{aligned} \phi_+{}^2 &\cong \frac{q^2}{m^2c^4k^2} = \frac{k_e q^2}{4Gm^2} = \frac{Q^2\alpha M_P{}^2}{4m^2} \gg 1 \\ \phi_-{}^2 &\cong -1 + \frac{m^2c^4k^2}{q^2} = -1 + 4\frac{m^2}{Q^2\alpha M_P{}^2} \cong -1 \end{aligned} . \tag{12.12}$$

Likewise, also using $1/(1-x) \cong 1+x$ for $x \ll 1$, (12.10) separates into:

$$\begin{aligned} cp_+{}^5 &\cong \frac{m^2c^4k}{q} = 2\sqrt{\frac{Gm^2}{k_eq^2}}mc^2 = 2\frac{m}{Q\sqrt{\alpha}M_P}mc^2 = 2\frac{1}{Q\sqrt{\alpha}M_P}m^2c^2 \ll mc^2 \\ cp_-{}^5 &\cong -\frac{q}{mc^2k} = -\frac{1}{2}\sqrt{\frac{k_eq^2}{Gm^2}}mc^2 = -\frac{1}{2}\frac{Q\sqrt{\alpha}M_P}{m}mc^2 = -\frac{1}{2}Q\sqrt{\alpha}M_Pc^2 \gg mc^2 \end{aligned} \quad . \tag{12.13}$$

And for (12.11) we similarly obtain:

$$\begin{aligned} \frac{dx_+{}^5}{cd\tau} &\cong \frac{mc^2k}{q} = 2\sqrt{\frac{Gm^2}{k_eq^2}} = 2\frac{m}{Q\sqrt{\alpha}M_P} \ll 1 \\ \frac{dx_-{}^5}{cd\tau} &\cong -\frac{q}{m^2c^4k} = -\frac{1}{2}\sqrt{\frac{k_eq^2}{Gm^2}} = -\frac{1}{2}\frac{Q\sqrt{\alpha}M_P}{m} \gg 1 \end{aligned} \quad . \tag{12.14}$$

Also, because the DKK metric tensor component $G_{55} = 1+\phi^2$ for $g_{\text{MN}} = \eta_{\text{MN}}$, see (3.11), which we wrote as $G_{55\pm} = 1+\phi_\pm{}^2$ at (12.8), it is also useful to use (12.12) to write the two solutions as:

$$\begin{aligned} G_{55+} &= 1+\phi_+{}^2 \cong 1+\frac{q^2}{m^2c^4k^2} = 1+\frac{k_eq^2}{4Gm^2} = 1+\frac{Q^2\alpha M_P{}^2}{4m^2} \cong \frac{Q^2\alpha M_P{}^2}{4m^2} \gg 1 \\ 0 < G_{55-} &= 1+\phi_-{}^2 \cong \frac{m^2c^4k^2}{q^2} = 4\frac{Gm^2}{k_eq^2} = 4\frac{m^2}{Q^2\alpha M_P{}^2} \ll 1 \end{aligned} \quad . \tag{12.15}$$

Note also, referring to (3.11) through (3.13), that (12.12) can be used in the $G_{\mu 5} = G_{5\mu}$ metric tensor components. With $M_P/m \cong 10^{20}$ roughly, $dx_-{}^5/cd\tau \cong 10^{20}$ in (12.14) reproduces the usual result from ordinary Kaluza-Klein theory, see toward the end of [12]. However, G_{55-}, albeit very small, still retains a timelike rather than a spacelike signature, so that $dx^5{}_-/cd\tau = dt_-{}^5/d\tau$ is a very rapid rate of fifth dimensional time flow, and not a space velocity on the order of $10^{20}c$.

Finally, we return to (11.22) and use the next-to-last expression in the two solutions (12.13) to likewise split this into:

$$\begin{aligned} cp_+{}^5\phi_1 &= \pm\sqrt{2}mc^2\sqrt{1-4\left(\frac{m}{Q\sqrt{\alpha}M_P}\right)^2}\exp\left(-\frac{x^0x^1x^2x^3x^5}{L^5}\right) \cong \pm\sqrt{2}mc^2\exp\left(-\frac{x^0x^1x^2x^3x^5}{L^5}\right) \\ cp_-{}^5\phi_1 &= \pm\sqrt{2}mc^2\sqrt{1-\frac{1}{4}\left(\frac{Q\sqrt{\alpha}M_P}{m}\right)^2}\exp\left(-\frac{x^0x^1x^2x^3x^5}{L^5}\right) \cong \pm\frac{i}{\sqrt{2}}M_Pc^2Q\sqrt{\alpha}\exp\left(-\frac{x^0x^1x^2x^3x^5}{L^5}\right) \end{aligned} \quad . \tag{12.16}$$

Recall again that these two solutions for $cp_+{}^5\phi_1$ and $cp_-{}^5\phi_1$, represent the two roots solutions to the quadratic (12.3) which we obtained starting at (12.5). We note that up to parts per 10^{40} the former

is independent of the fermion charge generator Q and $\sqrt{\alpha}$, while the latter is not. As we shall see in the next section, the former solution applies in the *Fermi vacuum* with an energy vev $v = 246.2196508\,\text{GeV}$ rooted in the Fermi coupling via $\sqrt{2}G_F v^2 = (\hbar c)^3$ [21], while, the latter applies in the *Planck vacuum* in which the Planck energy $M_P c^2 = 1.220910\times 10^{19}\,\text{GeV}$ [39] is established from the Newton coupling via $GM_P{}^2 \equiv \hbar c$.

13. Connection between the Dirac-Kaluza-Klein Scalar and the Higgs Field

At the outset it should be noted that $cp_+{}^5\phi_1(\mathbf{X})$ and $cp_-{}^5\phi_1(\mathbf{X})$ in (12.16) are both energy-dimensioned scalar fields, as is the Higgs field $h(\mathbf{X})$. As to the ratio $V_{(5)}/L^5 = x^0x^1x^2x^3x^5/L^5$, recall that L^5 arose at (11.20) via the constant of integration $C \equiv \ln(1/L^5)$. A length-to-the-fifth power dimensionality was required as an "initial condition" to provide a proper argument to the exponential in view of $V_{(5)}$ also having a fifth-order length dimension. Another "initial condition" we impose on this integration constant is that the overall ratio $V_{(5)}/L^5$ must be invariant under rotations and boosts (Poincare invariance for six of the ten parameters of the Poincare symmetry group), which means that L^5 must rotate and Lorentz transform and rotate in the same manner as $V_{(5)}$. Thus, we expect that each of the five L in L^5 will be in the nature of one component of a five-dimensional wavevector k^{M}, and that $V_{(5)}/L^5$ will take the invariant form of a product $\Pi_{\Sigma=0,1,2,3,5}(k_\Sigma x^\Sigma)$. Again, these are simply "initial conditions" imposed on a constant of integration whereby we require a) proper physical dimensionality and b) rotational and Lorentz symmetry.

We also recall that in the standard model, we expect a fermion (*f*) rest energy $m_f c^2$ to be related to $v \cong 246.2196508\,\text{GeV}$ by the relation $m_f c^2 = G_f v/\sqrt{2}$ where G_f is an arbitrary coupling not provided by presently-known theory and only deducible by knowing the observed fermion mass, see, e.g., [15.32] in [20]. So for the moment, irrespective of what number L^5 actually is or what $cp_+{}^5\phi_1(\mathbf{X})$ and $cp_-{}^5\phi_1(\mathbf{X})$ physically represent, for a coordinate assignment $x^{\text{M}} = \mathbf{0} = (0,0,0,0,0)$ at an origin the solutions where $\exp(-V_{(5)}/L^5) = 0$ (12.16) reduce to:

$$\begin{array}{l} cp_+{}^5\phi_1(\mathbf{0}) = \pm\sqrt{2}m_f c^2 = \pm G_f v \\ cp_-{}^5\phi_1 = \pm\frac{1}{\sqrt{2}} iM_P c^2 Q\sqrt{\alpha} \end{array}, \tag{13.1}$$

where we have replaced the approximation sign $\cong$ in (12.16) with an equal sign by recognizing that the 1 part per 10^{40} discrepancy is exceptionally small and unlikely to be observable.

Conversely, again irrespective of L^5, for a coordinate assignment with $V_{(5)}/L^5 \gg 1$, the exponential will approach zero, and (12.16) will reduce to:

$$\begin{matrix} cp_+{}^5\phi_1\left(V_{(5)} / L_+{}^5 \gg 1\right) = 0 \\ cp_-{}^5\phi_1\left(V_{(5)} / L_-{}^5 \gg 1\right) = 0 \end{matrix}, \tag{13.2}$$

So, these energy-dimensioned fields $cp_+{}^5\phi_1(\mathbf{X})/\sqrt{2}$ and $\sqrt{2}cp_-{}^5\phi_2(\mathbf{X})$ are equal to zero far from the origin, while at the origin, they are equal to $\pm m_f c^2$ and $\pm M_P c^2 Q\sqrt{\alpha}$ respectively, where $m_f c^2$ is a fermion rest energy and $M_P c^2$ is the Planck energy.

Now, the Higgs field $h(\mathbf{X})$ is a scalar field with dimensions of energy. Therefore, as with any energy field, the physics transpiring in this field will favor states of lower energy and disfavor states of higher energy. Of course, Heisenberg uncertainty does not permit us to talk about the "position" of a fermion in more than a statistical way, and so we cannot technically say that a fermion is "at a given coordinate" x^{M} in the five-dimensional space. But we can say that if the Higgs field provides energy "wells" for the fermions from which the fermions also obtain their rest masses, then the fermions will find "nests" in which they are most likely to situate at energetically-minimized locations in the Higgs field. Additionally, in the standard model, the initial scalar field which we denote by ϕ_h to distinguish from the Kaluza-Klein field ϕ, is given the following assignments at the various steps of symmetry breaking:

$$\phi_h = \tfrac{1}{\sqrt{2}}\left(\phi_{1h} + i\phi_{2h}\right) = \tfrac{1}{\sqrt{2}}\phi_{1h} = \tfrac{1}{\sqrt{2}}\left(v + h\right). \tag{13.3}$$

That is, we first assign $\phi_h = \tfrac{1}{\sqrt{2}}\left(\phi_{1h} + i\phi_{2h}\right)$. Then we break symmetry in the $\phi_{1h}{}^2 + \phi_{2h}{}^2$ circle by setting $\phi_{2h} = 0$. Then, working from the leading terms of a potential $V = \tfrac{1}{2}\mu^2\phi_h{}^2 + \tfrac{1}{4}\lambda\phi_h{}^4$ for the scalar field, we find that this potential has minima at $\phi_h = \pm v = \pm\sqrt{-\mu^2/\lambda}$. Finally, we perturbatively expand around the minima at $\phi_h = v$ using the Higgs field h which represents quantum fluctuations about the minima. Note that V has physical dimensions of *energy to the fourth power*, because in the simplest setting this is part of a Lagrangian density $\mathcal{L} = T - V$ with $T = \tfrac{1}{2}\left(\partial_\mu\phi_h\right)\left(\partial^\mu\phi_h\right)$ and V as above. And in more advanced settings where boson masses are generated, $V = \mu^2\phi_h{}^\dagger\phi_h + \lambda\left(\phi_h{}^\dagger\phi_h\right)^2$ and $T = \left(D_\mu\phi_h\right)^\dagger\left(D^\mu\phi_h\right)$ with $D_\mu = \partial_\mu + iqA_\mu$, but V still reduces to the simpler form $V = \tfrac{1}{2}\mu^2\phi_h{}^2 + \tfrac{1}{4}\lambda\phi_h{}^4$ after symmetry is broken with (13.3) because of the $1/\sqrt{2}$ coefficient being squared or raised to the fourth power. This is all nicely reviewed in section 14.6 through 14.9 of [20].

Most importantly for the present discussion, because $\phi_h = \pm v$ are the minima of the potential V and because $v = 246.2196508\,\mathrm{GeV}$ is a constant energy number, the expectation value $\langle\phi_h\rangle = \tfrac{1}{\sqrt{2}}\left(\pm v + \langle h\rangle\right) = \pm\tfrac{1}{\sqrt{2}}v$, which means that the expectation value of the Higgs field $\langle h\rangle = 0$. This of course makes sense because the Higgs field is *defined* to represent quantum *fluctuations*

about the minima in the potential *V*. But by being very explicit about all of this, now we see how to assign $cp^5\phi$ to the respective Higgs fields in both the Fermi and the Planck vacuums.

Specifically, for both solutions (12.16), at $V_{(5)}/L^5=0$ the exponential $\exp\left(-V_{(5)}/L^5\right)=1>0$ is above zero. Further, where $V_{(5)}/L^5\gg 1$, the exponential $\exp\left(-V_{(5)}/L^5\right)=0$ *drops* to zero. So, if we want the origin at $x^{\mathrm{M}}=\mathbf{0}$ be the most energetically-favorable locale for a fermion to "nest" at, we must choose the – signs from the $\pm$ in (12.17) for both solutions. Then, following this choice of sign, we assign $cp_+{}^5\phi_1(\mathbf{X})\equiv h_+(\mathbf{X})$ and $cp_-{}^5\phi_1(\mathbf{X})\equiv \frac{i}{2}Q\sqrt{\alpha}h_-(\mathbf{X})$, with h_+ and h_- representing Higgs field associated with each respective solution. As a result, also showing $m_fc^2=G_fv/\sqrt{2}$, (12.16) now become:

$$\begin{aligned} h_+(\mathbf{X}) &= -\sqrt{2}m_fc^2\exp\left(-\frac{x^0x^1x^2x^3x^5}{L_+{}^5}\right)=-\sqrt{2}m_fc^2\exp\left(-\frac{V_{(5)}}{L_+{}^5}\right)=-G_fv\exp\left(-\frac{V_{(5)}}{L_+{}^5}\right) \\ h_-(\mathbf{X}) &= -\sqrt{2}M_Pc^2\exp\left(-\frac{x^0x^1x^2x^3x^5}{L_-{}^5}\right)=-\sqrt{2}M_Pc^2\exp\left(-\frac{V_{(5)}}{L_-{}^5}\right) \end{aligned}. \tag{13.4}$$

The latter assignment includes $Q\sqrt{\alpha}$, to make the background field $h_-(\mathbf{X})$ independent of the specific charge generator *Q* of any fermion which may be situated in this field, a factor of *i* to maintain a real relation between h_- and the energy-times-exponential term, and a factor of ½ to have the exact same form in both solutions with the sole difference being m_f in the former and M_P in the later. We have also added + and – subscripts to the *L* arising from the integration constant for each of the two solutions to the quadratic (12.3). Going forward, we also make the notational definition $h\equiv h_+$ and $H\equiv h_-$, with the capitalization of the latter indictive of the inordinately-higher energies signaled by the appearance of M_Pc^2 versus m_fc^2 in the bottom versus the top relation (13.4). Now let's examine the evidence in favor of these assignments.

With these assignments, the Higgs fields will have minima at the origin which means the origin will be the most energetically-favorable locale for fermions to nest at. So now, given the energy minima at the origin, we can think of a fermion being situated at the origin $x^{\mathrm{M}}=\mathbf{0}$ with the highest statistical likelihood. As to the handling of the $\frac{1}{\sqrt{2}}$ factor, we observe from (13.3) that following symmetry breaking, $\phi_h=\frac{1}{\sqrt{2}}(v+h)$ and that the energy in the potential $V=\frac{1}{2}\mu^2\phi_h{}^2+\frac{1}{4}\lambda\phi_h{}^4$ is obtained from ϕ_h in which $\frac{1}{\sqrt{2}}$ is the coefficient of $v+h$ and so "cuts" the energies of *v* and *h* by this same factor as regards the physical energies. Thus, at the origin where a fermion is most likely to nest, $\frac{1}{\sqrt{2}}h(\mathbf{0})=-m_fc^2$ and $\frac{1}{\sqrt{2}}H(\mathbf{0})=-M_Pc^2$, while far from the origin $h\left(V_{(5)}/L_+{}^5\gg 1\right)=0$ and $H\left(V_{(5)}/L_-{}^5\gg 1\right)=0$. For both solutions, far from a fermion the Higgs fields have zero energies which are consistent with their expected values $\langle h\rangle=0$ and $\langle H\rangle=0$,.

Conversely, at the origin where we expect fermions to nest, we have $\frac{1}{\sqrt{2}}h(\mathbf{0})+m_f c^2=0=\langle h\rangle$ and $\frac{1}{\sqrt{2}}H(\mathbf{0})+M_P c^2=0=\langle H\rangle$. The energy effect is that the Higgs field is perturbed below its expected value, *but in a fashion that precisely counterbalances the rest energy it has given to the fermion* for *h*, and that counterbalances the Planck energy for *H*. That is, for *h*, when and where a fermion "embeds" or "nests" itself in a Higgs field, it appropriates an energy $m_f c^2$ out from the Higgs field for its own rest energy, and as a result, it creates a perturbation which drops the Higgs field down to an energy $-m_f c^2$. This is energy conservation appearing in yet another guise.

As to the $H\equiv h_-$ solutions, it is the serendipitous appearance of the Planck energy $M_P c^2$ which leads to us conclude that *H* represents a Higgs field in the Wheeler / Planck vacuum [3], [40], where we expect that the rest masses of all fermions will converge and become synonymous with the Planck mass based on our limited understanding of Grand Unified Theories which include gravitation. In this *H* solution, each fermion similarly takes an energy $M_P c^2$ for itself from the Higgs, and the Higgs itself compensates by dropping to an energy $-M_P c^2$. Again, this is energy conservation in another guise. It should also be noted that the geometrodynamic Wheeler / Planck vacuum comes about when – on statistical average – an inordinate number of Planck-energy fluctuations are separated from one another by the Planck length, whereby the negative gravitational energy arising from the Newtonian gravitational interactions of these fluctuations precisely cancels the energies of the fluctuations themselves, creating this geometrodynamic "vacuum." Later, following Hawking's [41], it became understood that the black holes inherent in this vacuum have the statistical character of a blackbody spectrum. Now, the association in (13.4) and particularly $\frac{1}{\sqrt{2}}H(\mathbf{0})=-M_P c^2$ is suggestive of the latter Higgs field $H\equiv h_-$ being synonymous with the *quantum gravitational field.*

The aforementioned energy conservation is seen explicitly by integrating (13.4) from zero to infinity over the entire five-dimensional volume $V_{(5)}=x^0x^1x^2x^3x^5$ to ascertain the total energy taken out of the Higgs field to give mass to a fermion. Again, recognizing from $\phi_h=\frac{1}{\sqrt{2}}(v+h)$ and $V=\frac{1}{2}\mu^2\phi_h{}^2+\frac{1}{4}\lambda\phi_h{}^4$ that we calculate energies by cutting *h* by a factor of $\frac{1}{\sqrt{2}}$, we find that:

$$\begin{aligned}\frac{1}{\sqrt{2}}\frac{1}{L_+{}^5}\int_0^\infty h(\mathbf{X})\,dV_{(5)}&=-\frac{1}{L_+{}^5}m_f c^2\int_0^\infty\exp\left(-\frac{V_{(5)}}{L_+{}^5}\right)dV_{(5)}=m_f c^2\exp\left(-\frac{V_{(5)}}{L_+{}^5}\right)\Bigg|_0^\infty=-m_f c^2\\ \frac{1}{\sqrt{2}}\frac{1}{L_-{}^5}\int_0^\infty H(\mathbf{X})\,dV_{(5)}&=-\frac{1}{L_-{}^5}M_P c^2\int_0^\infty\exp\left(-\frac{V_{(5)}}{L_-{}^5}\right)dV_{(5)}=M_P c^2\exp\left(-\frac{V_{(5)}}{L_-{}^5}\right)\Bigg|_0^\infty=-M_P c^2\end{aligned}\quad . \qquad (13.5)$$

So indeed, with the association $cp_+{}^5\phi_1(\mathbf{X})\equiv h_+(\mathbf{X})$ made in (13.4), and with choosing the – signs in (12.17) to place the energy minima at the origin, we find that a fermion with a mass $m_f c^2$ gains this mass by extracting a total energy $m_f c^2$ out of the Higgs field. That is, integrated over the

entire expanse of $V_{(5)}$, the perturbations in the Higgs field all sum to $-m_f c^2$, *with the Higgs field thereby having given up precisely the amount of energy which the fermion acquired for its rest energy*. Likewise with $cp_-{}^5\phi_1(\mathbf{X}) \equiv \frac{i}{2}Q\sqrt{\alpha}h_-(\mathbf{X})$ for $M_P c^2$ and $-M_P c^2$ in the Planck vacuum. It is because of these results, that the evidence supports the identifications of $cp_+{}^5\phi_1(\mathbf{X}) \equiv h_+(\mathbf{X})$ and $cp_-{}^5\phi_1(\mathbf{X}) \equiv \frac{i}{2}Q\sqrt{\alpha}h_-(\mathbf{X})$ with the Higgs fields in (13.4).

Next, let us use $\phi_h = \frac{1}{\sqrt{2}}(v+h)$ from (13.3), and using uppercase notation let us form a similar $\Phi_h = \frac{1}{\sqrt{2}}(M_P c^2 + H)$ to represent Planck vacuum fluctuations. Using these in each of (13.4) we obtain:

$$\begin{aligned}\phi_{h+}(\mathbf{X}) &= \frac{1}{\sqrt{2}}(v+h) = \frac{1}{\sqrt{2}}\left(v - \sqrt{2}m_f c^2 \exp\left(-\frac{V_{(5)}}{L_+{}^5}\right)\right) = \frac{v}{\sqrt{2}}\left(1 - G_f \exp\left(-\frac{V_{(5)}}{L_+{}^5}\right)\right)\\ \phi_{h-}(\mathbf{X}) &= \frac{1}{\sqrt{2}}(M_P c^2 + H) = \frac{1}{\sqrt{2}}\left(M_P c^2 - \sqrt{2}M_P c^2 \exp\left(-\frac{V_{(5)}}{L_-{}^5}\right)\right) = \frac{M_P c^2}{\sqrt{2}}\left(1 - \sqrt{2}\exp\left(-\frac{V_{(5)}}{L_-{}^5}\right)\right)\end{aligned} \quad . (13.6)$$

Although the Planck-scale "minus" quadratic solutions should be closely studied in relation to the detailed physics of the Planck vacuum, from here forward, we shall focus on the "plus" solutions because these will lead to a detailed understanding of how the fermions acquire their rest masses from the Higgs field. Of particular interest now, is the expression for ϕ_{h+} written in terms of the coupling $m_f c^2 = \frac{1}{\sqrt{2}}vG_f$ of the fermion rest mass to the vacuum. This is the fermion counterpart of the coupling form $M_b c^2 = \frac{1}{2}vg$ by which massive vector bosons obtain their masses in the standard model. Setting $x = V_{(5)} / L_+{}^5$ and $y = \sqrt{2}\phi_{h+} / v$ this function takes the mathematical form of $y = 1 - G_f \exp(-x)$. For $x \gg 1$ this function has a flat line at $y = 1$. Near the $x = 0$ origin where a fermion is energetically most likely to situate this dips into an energy well with a minimum $y = 1 - G_f$ at the origin. For light fermions such as the electron, and in fact, for all fermions except the top quark, which all have empirical values $G_f \ll 1$, this energy well is only mildly depressed below $y = 1$. But for the top quark which has $G_t \cong 1$ slightly less than 1, in a result of significance, this energy well dips almost down to zero. Let us now take a closer look at this.

Using the empirical value $v/\sqrt{2} = 174.1035847$ GeV, and using empirical mass data from PDG's [42], the dimensionless couplings $G_f = m_f c^2 / \frac{1}{\sqrt{2}}v$ for the up, charm and top quarks are:

$$G_u = 0.000013^{+0.000003}_{-0.000002};\quad G_c = 0.00732^{+0.00014}_{-0.00020};\quad G_t = 0.99366 \pm 0.00230, \tag{13.7a}$$

while the down, strange and bottom quarks these are:

$$G_d = 0.000027^{+0.000003}_{-0.000002};\quad G_s = 0.000546^{+0.000052}_{-0.000017};\quad G_b = 0.02401^{+0.00023}_{-0.00017}. \tag{13.7b}$$

And using the PDG mass data from [43], for the three charged leptons we obtain:

$$G_e = 2.935028288\times10^{-6}\pm1.8\times10^{-14}; G_\mu = 6.06870758\times10^{-4}\pm1.4\times10^{-11}; G_\tau = 0.0102058\pm6.9\times10^{-7}. \quad (13.7c)$$

In (13.7a) we see clearly how G_t is just under 1, irrespective of the error bars. In other words, although G_t is close to 1, it is not possible for this to be equal to 1, because such a result would be *outside* the errors bars.

As we start speak about quark masses and their "errors," it must be noted that the error bars of the quark masses in [42] are not just ordinary experimental errors owing to limitations in the resolutions of observational equipment. Rather, as elaborated in [44], "Unlike the leptons, quarks are confined inside hadrons and are not observed as physical particles. Quark masses therefore cannot be measured directly, but must be determined indirectly through their influence on hadronic properties. Although one often speaks loosely of quark masses as one would of the mass of the electron or muon, any quantitative statement about the value of a quark mass must make careful reference to the particular theoretical framework that is used to define it. It is important to keep this *scheme dependence* in mind when using the quark mass values tabulated in the data listings (original emphasis)." For the moment, we will speak "loosely" about these error spreads, and later on, will discuss these spreads in more precise terms in relation to observational schemes.

It is illustrative to show the graph of $\phi_{h+}(\mathbf{X})$ in (13.6) for the top quark which is by far the most-massive elementary fermion, and by way of contrast, for the electron which is the lightest fermion. These two plots are shown below:

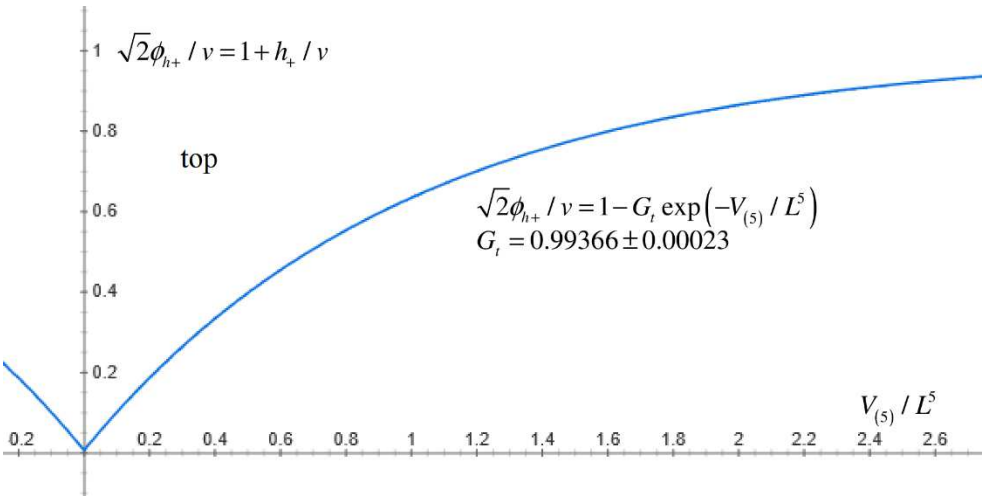

Figure 1: Higgs field extraction of rest energy from the Fermi vacuum, for the top quark

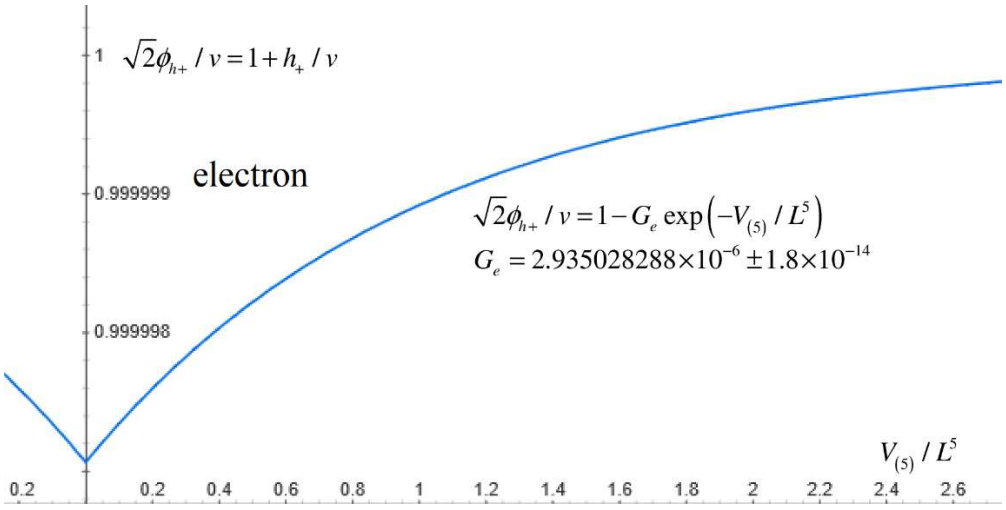

Figure 2: Higgs field extraction of rest energy from the Fermi vacuum, for the electron

We see in these two graphs both based upon the underlying mathematical relation $y = 1 - G_f \exp\left(-x\right)$, that the curves have exactly the same shape, culminating in an energy well minimum at the origin of the Higgs well where the fermion is energetically most likely to be nested. But the energy extracted to give the electron its rest mass only causes a very slight dip down to $1 + h_+ / v \cong .999997$ at the origin. In contrast, at the center of the top quark error-bar range the energy needed to give the top quark its mass depresses to $1 + h_+ / v = 1 - G_t \cong 0.00634$ at the origin, only slightly larger than zero. That is, at the origin of the Higgs well for the top quark, *almost all, but not all*, of the energy has been removed from the vacuum to give the top quark its rest mass. Moreover, in accordance with (13.5), all of the energy extracted from the vacuum in this way, integrated over $dV_{(5)}$ from the origin out to infinity, its precisely equal to the rest energy of each fermion, and likewise for all fermions. So, the fermions do acquire their rest masses by quite literally sucking out of the Fermi vacuum via the Higgs field, an energy exactly equal to their rest energies. We will later find that Figure 2 is not-quite correct, because in fact the Higgs vacuum possesses *four* vev minima, not one: a global minimum for isospin-up quarks, a local minimum for isospin-down quarks, a global minimum for charged (isospin-down) leptons, and a local minimum for (isospin-up) neutrinos. Therefore, the vev against which the "dip" in Figure 2 occurs, will be set by the charged lepton vev. And for each fermion type, its "dip" will occur as against its respective vev minimum. But this is not knowable at this point, and is revealed only by studying the fermion masses in detail and connecting them to the CKM and neutrino oscillation (PMNS) mixing angles, as we shall do in the next several sections.

Finally, the very recent observation at CERN of a clear affinity between the Higgs boson and the top quark [22], [23], [24] is very graphically understood on the basis of Figure 1 as a manifestation of how the top quark – uniquely amongst all fermions given the empirical data in (13.7) – draws almost all of the energy out of the Fermi vacuum, in its immediate vicinity where $V_{(5)} / L^5 \cong 0$. This insight about how the top quark removes almost all of the energy from the Fermi vacuum, while very interesting in its own right and illustrative of the observed $tH\overline{H}$ affinity,

points toward a deeper meaning that leads directly to a theory of why the fermions actually have the rest masses that they do. This is the subject of the next several sections.

14. Theory of Fermion Rest Masses and Mixing: Up, Charm and Top

It is highly intriguing in its own right that the mass coupling $G_t = 0.99366 \pm 0.00230$ for the top quark is very close to 1 but just under 1, and also, that this closeness to 1 is *outside* the error bars. In other worlds, there is no possibility that $G_t = 1$ and simply needs to be established as such by more accurate testing. This leads us to raise the question whether the *sum* of the up-plus-charm-plus-top rest energies might yield total energy for which the coupling is equal to 1 *within experimental and scheme-dependent errors*, and if so, whether this could be of theoretically significances as the "tip of the spear" toward developing a viable theory that solves the presently-unsolved mystery puzzle of why the fermions have the rest masses that they do.

It turns out that if we calculate this coupling $G_f = m_f c^2 / \frac{1}{\sqrt{2}} v$ for the *sum* of the three isospin-up quark masses, and account for the error bars in all three, we obtain:

$$G_{u+c+t} = \left(m_u + m_c + m_t \right) c^2 / \tfrac{1}{\sqrt{2}} v = 1.000997^{+0.0024439}_{-0.0025008}, \text{ i.e. } \quad 0.998496 < G_{u+c+t} < 1.003441 . \qquad (14.1)$$

So, given the errors, it *is possible* that the sum of these three quark masses *is exactly equal to* $\frac{1}{\sqrt{2}} v$ and that this equality is a true relation of physical significance. If this is so, then because $v = 246.2196508$ is known with greater precision than any of the up, charm or top quark masses, we first of all have an immediate resource for narrowing the range of error in the top quark mass, down to the error range of the charm quark. This provides immediacy in its ability to be confirmed or contradicted by more-precise experiments to measure the top quark mass. Secondly, if this equality is true, then it becomes possible to account for all three quark masses using bi-unitary CKM-style mixing rotations acting on a mass matrix, which possibility has been entertained on and off for four decades, see., e.g., [45]. Third, once these bi-unitary transformations are established for the isospin-up quark masses, similar transformations may be established for the isospin-down quarks, and for the charged leptons. Fourth and finally, once such transformations have been established, we are able to revisit the potentials $V = \frac{1}{2} \mu^2 {\phi_h}^2 + \frac{1}{4} \lambda {\phi_h}^4$ for the Higgs theory scalar, and reestablish these in a fashion that ties together all of the foregoing fermion masses with the very tiny masses of the neutrinos. Let us now take up each of these four matters in turn.

First, let us use the empirical data that $0.998496 < G_{u+c+t} < 1.003441$ to postulate that in fact, this coupling

$$G_{u+c+t} = \left(m_u + m_c + m_t \right) c^2 / \tfrac{1}{\sqrt{2}} v \equiv 1, \qquad (14.2)$$

based on this being true within experimental and scheme-dependent errors. Directly in terms of rest energies, this means:

$$m_u c^2 + m_c c^2 + m_t c^2 \equiv \tfrac{1}{\sqrt{2}} v = 174.1035847 \text{ GeV} . \tag{14.3}$$

The precision in (14.3) for v is far greater than the precision in either $m_u = .0022^{+.0005}_{-.0004}$ GeV or $m_c = 1.275^{+.025}_{-.035}$ GeV, as well as in $m_t = 173.0 \pm 0.4$ GeV, see [42]. So, we need not be concerned with the precision in v, but instead will account for the errors particularly in m_c. Combining (14.3) with the known up and charm mases we deduce that:

$$\boxed{\begin{array}{l} m_t c^2 = 174.1035847 \text{ GeV} - m_c c^2 - m_u c^2 = 172.8264^{+0.0354}_{-0.0255} \text{ GeV}; \\ \text{i.e.} \quad 172.8009 < m_t c^2 < 172.8618 \text{ GeV} \end{array}} . \tag{14.4}$$

which is more accurate than the currently-known range $172.6 < m_t < 173.4$ GeV by two-to-three orders of magnitude. *This result in (2.4) is a prediction which can and should be tested in experiments designed to obtain a more precise direct measurement of the top quark mass.* If (14.4) is true, then it is also convenient for the next step to collect all of these quark masses together:

$$m_u c^2 = .0022^{+.0005}_{-.0004} \text{ GeV}; \quad m_c c^2 = 1.275^{+.025}_{-.035} \text{ GeV}; \quad m_t c^2 = 172.8264^{+0.0354}_{-0.0255} \text{ GeV} . \tag{14.5}$$

We may also revise the isospin-up quark couplings in (13.7a) as such:

$$G_u = 0.000013^{+0.000003}_{-0.000002}; \quad G_c = 0.00732^{+0.00014}_{-0.00020}; \quad G_t = 0.99266^{+0.00020}_{-0.00015}, \tag{14.6}$$

Second, taking the foregoing to be true, and also given what we just learned in relation to Figure 1, let us now form the following hypothesis of how these three fermions obtain their mass: In Figure 1, at the origin of the Higgs field energy well where the top quark is energetically most likely to be seated, almost all of the energy, but not quite all of the energy, is drawn out of the Fermi vacuum and used to give the mass to the top quark, via the energy integration calculated in (13.5). But if there was to exist a single quark with the sum (14.3) of all three quark masses – or if the masses of all three quark masses could be transformed into the mass of a single quark – then that single quark would draw the entirety of the energy out of the Fermi vacuum at the origin of its Higgs field energy well. And in fact, the type bi-unitary mass matrix transformations discussed in [45] provide the precise vehicle for this to occur.

Specifically, we know there is a Fermi vacuum with an energy that has an expected value $v = 246.2196508 \text{ GeV}$, and that fermions acquire their masses by drawing energy out of this vacuum. So one means for the top, charm and up quarks to acquire their masses would be for all three quark to start out formally massless (i.e. with two degree of freedom), for the symmetry to be broken in the manner reviewed leading to (11.12) whereby the top quark gains a mass of $m_t c^2 \equiv \tfrac{1}{\sqrt{2}} v = 174.1035847 \text{ GeV}$ which depressed the vacuum down to a rock bottom 0 GeV at the origin of the Higgs well, and where some of this mass is then rotated over to the charm and up quarks via a bi-unitary transformation operating on a mass matrix with the rest energies $m_t c^2$, $m_c c^2$ and $m_u c^2$ on its diagonal. Specifically, let us begin with a mass matrix defined by:

$$M_{uct}c^2 \equiv \begin{pmatrix} m_t & \sqrt{m_t m_c} & \sqrt{m_t m_u} \\ \sqrt{m_t m_c} & m_c & \sqrt{m_c m_u} \\ \sqrt{m_t m_u} & \sqrt{m_c m_u} & m_u \end{pmatrix} c^2 = \begin{pmatrix} \frac{1}{\sqrt{2}}v & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 174.1035847\ \text{GeV} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (14.7)$$

Then, let us transform this into $M_{uct} \to M'_{uct} = U^\dagger M'_{uct} U$, where U is a unitary matrix $U^\dagger U = 1$. The important point to note, is that under a bi-unitary transformation the trace $\text{tr}\left(M_{uct}\right) = \text{tr}\left(M'_{uct}\right)$ is preserved so that $m_u c^2 + m_c c^2 + m_t c^2 \equiv \frac{1}{\sqrt{2}}v$ in (14.3) will remain true not matter what specific angles or phases are used in this transformation.

The next deliberation is what to use for the unitary matrix U. As a 3x3 matrix this could have up the three real angles θ_{21}, θ_{32}, θ_{31} and one imaginary phase δ in the same manner as the CKM mixing matrix used to characterize generation-changing weak interaction beta decays for both quark and leptons. But the up, charm and top masses represent three unknown mass parameters. The relation (14.3) reduces this down to two unknown parameters, plus the Fermi vev which is known. So, we ought not use more than two real angles without a phase (or more precisely, with $\delta = 0$ so $\exp i\delta = 1$) to re-parametrize these two unknown masses, so that we simply trade two mass unknowns for two angle unknowns. For this purpose, we may choose any two of θ_{21}, θ_{32}, θ_{31} and structure the matrices accordingly. Specifically, we may choose a first parametrization with θ_{32} and θ_{21}, whereby some of the mass in $m_t c^2 = \frac{1}{\sqrt{2}}v$ first is rotated into $m_c c^2$, then "downward cascades" into $m_u c^2$. The second parameterization is to use θ_{32} and θ_{31} where the top quark mass is "distributed" to both the charm and up quarks. The third alternative is to use θ_{31} and θ_{21} where the top mass rotates into the up quark, then "upward cascades" into the charm quark. For reasons that momentarily become apparent, it is fruitful to develop both the "downward cascade" and the "distribution" parameterizations, while the third parameterization turns out to be duplicative of the first but with a 90-degree rotation of one of the angles.

Using the "downward cascade" parameterization, this bi-unitary transformation is:

$$M_{uct}c^2 \to M'_{uct}c^2 = U^\dagger M_{uct}c^2 U$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{21} & -s_{21} \\ 0 & s_{21} & c_{21} \end{pmatrix} \begin{pmatrix} c_{32} & -s_{32} & 0 \\ s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}}v & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_{32} & s_{32} & 0 \\ -s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{21} & s_{21} \\ 0 & -s_{21} & c_{21} \end{pmatrix}. \quad (14.8a)$$

$$= \frac{1}{\sqrt{2}}v \begin{pmatrix} {c_{32}}^2 & c_{32}\,s_{32}\,c_{21} & c_{32}\,s_{32}\,s_{21} \\ c_{32}\,s_{32}\,c_{21} & {s_{32}}^2\,{c_{21}}^2 & {s_{32}}^2\,c_{21}\,s_{21} \\ c_{32}\,s_{32}\,s_{21} & {s_{32}}^2\,c_{21}\,s_{21} & {s_{32}}^2\,{s_{21}}^2 \end{pmatrix}_I = \begin{pmatrix} m'_t & \sqrt{m'_t m'_c} & \sqrt{m'_t m'_u} \\ \sqrt{m'_t m'_c} & m'_c & \sqrt{m'_c m'_u} \\ \sqrt{m'_t m'_u} & \sqrt{m'_c m'_u} & m'_u \end{pmatrix} c^2$$

So now the energy $\frac{1}{\sqrt{2}}v$ from the Fermi vacuum that started out all in the top quark has been rotated into and shared with the charm and up quarks. With the "distribution" parameterization we obtain:

$$M_{uct}c^2 \rightarrow M'_{uct}c^2 = U^\dagger M_{uct}c^2 U$$

$$= \begin{pmatrix} c_{31} & 0 & -s_{31} \\ 0 & 1 & 0 \\ s_{31} & 0 & c_{31} \end{pmatrix}\begin{pmatrix} c_{32} & -s_{32} & 0 \\ s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}}v & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} c_{32} & s_{32} & 0 \\ -s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} c_{31} & 0 & s_{31} \\ 0 & 1 & 0 \\ -s_{31} & 0 & c_{31} \end{pmatrix}. \qquad (14.8b)$$

$$= \frac{1}{\sqrt{2}}v\begin{pmatrix} c_{32}{}^2 c_{31}{}^2 & c_{32}\, s_{32}\, c_{31} & c_{32}{}^2 c_{31}\, s_{31} \\ c_{32}\, s_{32}\, c_{31} & s_{32}{}^2 & c_{32}\, s_{32}\, s_{31} \\ c_{32}{}^2 c_{31}\, s_{31} & c_{32}\, s_{32}\, s_{31} & c_{32}{}^2 s_{31}{}^2 \end{pmatrix}_{II} = \begin{pmatrix} m'_t & \sqrt{m'_t m'_c} & \sqrt{m'_t m'_u} \\ \sqrt{m'_t m'_c} & m'_c & \sqrt{m'_c m'_u} \\ \sqrt{m'_t m'_u} & \sqrt{m'_c m'_u} & m'_u \end{pmatrix} c^2$$

There are three mathematical points to note in (14.8). First, as already mentioned, the trace is preserved under (14.8), because $s_{32}{}^2 s_{21}{}^2 + s_{32}{}^2 c_{21}{}^2 + c_{32}{}^2 = 1$ in the former and $c_{32}{}^2 c_{31}{}^2 + c_{32}{}^2 s_{31}{}^2 + s_{32}{}^2 = 1$ in the latter. Thus, $m_u c^2 + m_c c^2 + m_t c^2 = m'_u c^2 + m'_c c^2 + m'_t c^2 = \frac{1}{\sqrt{2}}v$, so we also preserve (14.3) as required. Second, all of the square root relations in the off-diagonal positions are preserved, viz: $m'_t m'_c = c_{32}{}^2 s_{32}{}^2 c_{21}{}^2$, $m'_t m'_u = c_{32}{}^2 s_{32}{}^2 s_{21}{}^2$ and $m'_c m'_u = s_{32}{}^4 c_{21}{}^2 s_{21}{}^2$ in the former while $m'_t m'_c = c_{32}{}^2 s_{32}{}^2 c_{31}{}^2$, $m'_t m'_u = c_{32}{}^4 c_{31}{}^2 s_{31}{}^2$ and $m'_c m'_u = c_{32}{}^2 s_{32}{}^2 s_{31}{}^2$, whether calculated from the diagonal or the off-diagonal elements. Third, the masses and their associated couplings related by $\frac{1}{\sqrt{2}}vG_f = m_f c^2$ are the same no matter which parameterization scheme we use, but that *the angles are defined differently depending on the scheme*. For this reason we have denoted all of the angles on the final lines of (14.8) by the *I* and *II* subscripts outside the matrix containing the sines and cosines of these angles. Note also, if we use the mass-to-vacuum coupling relation $\frac{1}{\sqrt{2}}vG_f = m_f c^2$, then dropping the primes of the transformations in (14.8) from here on, we can explicitly identify these couplings to be:

$$G_{uct} = \begin{pmatrix} G_t & \sqrt{G_t G_c} & \sqrt{G_t G_u} \\ \sqrt{G_t G_c} & G_c & \sqrt{G_c G_u} \\ \sqrt{G_t G_u} & \sqrt{G_c G_u} & G_u \end{pmatrix} = \begin{pmatrix} c_{32}{}^2 & c_{32}\, s_{32}\, c_{21} & c_{32}\, s_{32}\, s_{21} \\ c_{32}\, s_{32}\, c_{21} & s_{32}{}^2 c_{21}{}^2 & s_{32}{}^2 c_{21}\, s_{21} \\ c_{32}\, s_{32}\, s_{21} & s_{32}{}^2 c_{21}\, s_{21} & s_{32}{}^2 s_{21}{}^2 \end{pmatrix}_I, \qquad (14.9a)$$

$$G_{uct} = \begin{pmatrix} G_t & \sqrt{G_t G_c} & \sqrt{G_t G_u} \\ \sqrt{G_t G_c} & G_c & \sqrt{G_c G_u} \\ \sqrt{G_t G_u} & \sqrt{G_c G_u} & G_u \end{pmatrix} = \begin{pmatrix} c_{32}{}^2 c_{31}{}^2 & c_{32}\, s_{32}\, c_{31} & c_{32}{}^2 c_{31}\, s_{31} \\ c_{32}\, s_{32}\, c_{31} & s_{32}{}^2 & c_{32}\, s_{32}\, s_{31} \\ c_{32}{}^2 c_{31}\, s_{31} & c_{32}\, s_{32}\, s_{31} & c_{32}{}^2 s_{31}{}^2 \end{pmatrix}_{II}. \qquad (14.9b)$$

Note, that for both of these, the trace $\mathrm{tr}G_{uct} = G_t + G_c + G_u = 1$. This is another refection of (14.3).

Now we turn to the empirical data and calculate these angles to see if they bear any relation to any other known empirical particle data. Specifically, we use the revised mass coupling data in (14.6) to calculate θ_{I32} and θ_{I21} in (14.9a), and θ_{II32} and θ_{II31} in (14.9b). From (14.9a) we first deduce $\mathrm{c}_{I32}{}^2 = G_t$, then $\mathrm{c}_{I21}{}^2 = G_c / \mathrm{s}_{I32}{}^2 = G_c / \left(1 - \mathrm{c}_{I32}{}^2\right)$, then ascertain the angles in both radians and degrees. From (14.9b) we likewise deduce $\mathrm{s}_{II32}{}^2 = G_c$ followed by $\mathrm{c}_{II31}{}^2 = G_t / \mathrm{c}_{II32}{}^2 = G_t / \left(1 - \mathrm{s}_{II32}{}^2\right)$ followed by the angles. In this way, we calculate that:

$$\begin{aligned}
\theta_{I32} &= 0.08575^{+0.00085}_{-0.00120}\ \text{rad} = 4.91338^{+0.04893}_{-0.06874}\ ^\circ \\
\theta_{II32} &= 0.08568^{+0.00084}_{-0.00119}\ \text{rad} = 4.90914^{+0.04801}_{-0.06801}\ ^\circ \\
\theta_{I21} &= 0.04152^{+0.00403}_{-0.00343}\ \text{rad} = 2.37864^{+0.23071}_{-0.19673}\ ^\circ \\
\theta_{II31} &= 0.00357^{+0.00038}_{-0.00034}\ \text{rad} = 0.20442^{+0.02206}_{-0.01953}\ ^\circ
\end{aligned} \quad . \tag{14.10}$$

The scheme-dependent θ_{32} differ but slightly as between these two parameterizations, and the angles of approximately $4.91°$ do not "ring any bells" with regard to other known empirical data. *But as to θ_{I21} and θ_{II31} one cannot help but notice based on the 2018 PDG data [46] that these are equal to two of the three CKM quark mixing angles within experimental errors*. Specifically, using the Wolfenstein parameterization reviewed in [46], it is possible in a known manner to deduce that for the empirically-observed standard parameterization CKM angles (subscript *C*):

$$\begin{aligned}
\theta_{C12} &= 0.2265 \pm 0.0005\ \text{rad} = 12.975 \pm 0.026\ ^\circ \\
\theta_{C13} &= 0.0036^{+0.0003}_{-0.0002}\ \text{rad} = 0.209^{+0.015}_{-0.013}\ ^\circ \\
\theta_{C23} &= 0.0422 \pm 0.0009\ \text{rad} = 2.415 \pm 0.053\ ^\circ \\
\delta_C &= 1.2391^{+0.0348}_{-0.0335}\ \text{rad} = 70.998^{+1.995}_{-1.917}\ ^\circ
\end{aligned} \quad . \tag{14.11}$$

Doing the comparisons, we see that $\theta_{I21} = 2.37864^{+0.23071}_{-0.19673}\ ^\circ$ versus $\theta_{C23} = 2.415 \pm 0.053\ ^\circ$ which overlap within the error bars, and that $\theta_{II31} = 0.20442^{+0.02206}_{-0.01953}\ ^\circ$ versus $\theta_{C13} = 0.209^{+0.015}_{-0.013}\ ^\circ$ which likewise overlap within the error bars. In fact, θ_{I21} which has a wider error bar has a central portion fitting entirely within the error range for θ_{C23}, and θ_{II31} with a wider spread also has a central region fitting entirely within the errors for θ_{C13}.

So, our goal was to see whether the mass mixing angles in the bi-unitary transformation $M_{uct} \to M'_{uct} = U^\dagger M'_{uct} U$ bore any relation to any known data. And in the comparison between (14.10) and (14.11) we found that we have two "hits" directly in the middle of the empirical data for two of the three real CKM mixing angles. (Shortly, we will likewise connect with the third real CKM angle using the isospin-down quark masses.) With two such hits not one, the statistical chances of this being a coincidence are extremely remote. Therefore, let us now conclude that this concurrence between (14.10) and (14.11) in fact is the discovery of two fundamental physical relations whereby we use the empirical data concurrence to define the physical relations:

$$\boxed{\begin{array}{l}\theta_{I21} \equiv \theta_{C23} = 2.415 \pm 0.053\,^\circ \\ \theta_{II31} \equiv \theta_{C13} = 0.209^{+0.015}_{-0.013}\,^\circ\end{array}}. \tag{14.12}$$

Note also that we have used the CKM angles to reestablish the empirical range of the mass-based θ_{I21} and θ_{II31} because the former have tighter (smaller) error ranges. Consequently, we may use (14.12) to more tightly tune the isospin-up quark masses using CKM data, rather than vice versa.

Once we have made the connections in (14.12), it becomes possible to express the isospin-up quark masses via their couplings $G_f = m_f c^2 / \frac{1}{\sqrt{2}} v$, directly in terms of the CKM mixing angles, and vice versa. From the relations embedded in (14.9) which were used to obtain (14.10), we may now use (14.12) to find that:

$$\begin{array}{l}\cos^2\theta_{C23} = \cos^2\theta_{I21} = \dfrac{G_c}{\sin^2\theta_{I32}} = \dfrac{G_c}{1-\cos^2\theta_{I32}} = \dfrac{G_c}{1-G_t} \\ \cos^2\theta_{C13} = \cos^2\theta_{II31} = \dfrac{G_t}{\cos^2\theta_{II32}} = \dfrac{G_t}{1-\sin^2\theta_{II32}} = \dfrac{G_t}{1-G_c}\end{array}. \tag{14.13}$$

Then, solving (14.13) as simultaneous equations in G_t and G_c, while also using $G_u = \mathrm{s}_{I32}{}^2\,\mathrm{s}_{I21}{}^2 = \mathrm{c}_{II32}{}^2\,\mathrm{s}_{II31}{}^2$ from (14.9) along with (14.12), we are able to deduce:

$$G_t = \frac{\sin^2\theta_{C23}\cos^2\theta_{C31}}{1-\cos^2\theta_{C23}\cos^2\theta_{C31}};\quad G_c = \frac{\cos^2\theta_{C23}\sin^2\theta_{C31}}{1-\cos^2\theta_{C23}\cos^2\theta_{C31}};\quad G_u = G_c\tan^2\theta_{C23} = G_t\tan^2\theta_{C31}. \tag{14.14}$$

This expresses the $G_f = m_f c^2 / \frac{1}{\sqrt{2}} v$ for the isospin-up quarks, entirely in terms of the CKM angles θ_{C31} and θ_{C32} which mix the third-generation quarks with the first and second generations. Only two of the three relations (14.14) are independent. But together with (14.4) which related the sum of the three isospin-up quarks directly to the Fermi vev, we have now expressed all of these three quark masses as functions $m_u, m_c, m_t = F\left(v, \theta_{C31}, \theta_{C23}\right)$ of other known parameters, namely, the Fermi G_F coupling and its related vev, and the two third-generation CKM mixing angles. In this way, what began at the start of this section as *twelve* unexplained fermion rest masses (six quarks flavors and six lepton flavors) have now been reduced down to only *nine* remaining unexplained masses. *Three of these twelve masses, for the isospin-up quarks, can now be expressed entirely in terms of other known physical parameters.*

In fact, there is a very simple geometric interpretation of the results in (14.13). From (14.9) we may use $G_t + G_c + G_u = 1$ then $\frac{1}{\sqrt{2}} v G_f = m_f c^2$ to rewrite (14.13) as:

$$\cos^2\theta_{C23}=\cos^2\theta_{I21}=\frac{G_c}{1-G_t}=\frac{G_c}{G_c+G_u}=\frac{\sqrt{G_c}^{\,2}}{\sqrt{G_c}^{\,2}+\sqrt{G_u}^{\,2}}=\frac{\sqrt{m_c}^{\,2}}{\sqrt{m_c}^{\,2}+\sqrt{m_u}^{\,2}}$$
$$\cos^2\theta_{C13}=\cos^2\theta_{II31}=\frac{G_t}{1-G_c}=\frac{G_t}{G_t+G_u}=\frac{\sqrt{G_t}^{\,2}}{\sqrt{G_t}^{\,2}+\sqrt{G_u}^{\,2}}=\frac{\sqrt{m_t}^{\,2}}{\sqrt{m_t}^{\,2}+\sqrt{m_u}^{\,2}} \quad . \tag{14.15}$$

If we now establish a three-dimensional rest mass space in which the square roots $\sqrt{m_c}$, $\sqrt{m_u}$ and $\sqrt{m_t}$ are respectively plotted against the *x*, *y*, and *z* axes, we see that $\theta_{C13}=\theta_{II31}=\theta$ is simply the polar angle θ of descent from the *z* axis and $\theta_{C23}=\theta_{I21}=\phi$ is the azimuthal axis of rotation through the *x* and *y* plane about the *z* axis, using spherical coordinates. This is graphically illustrated below, using the quarks mass values in (14.5):

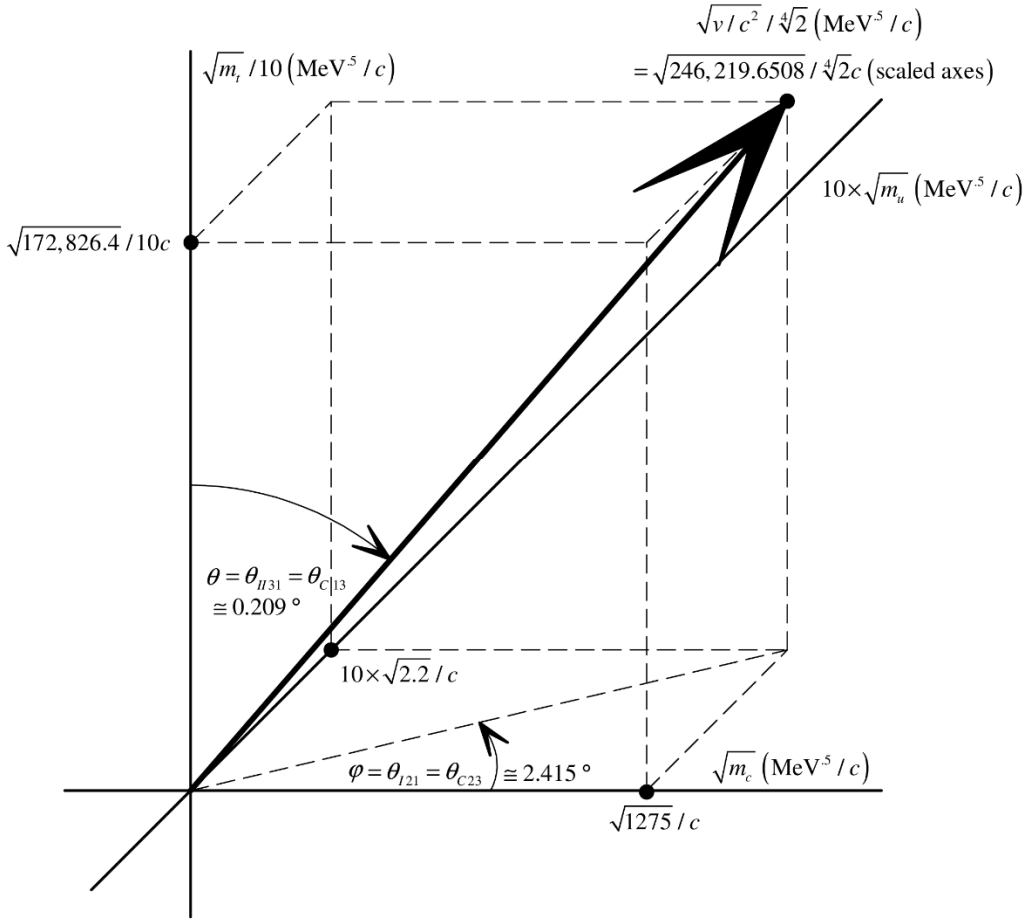

Figure 3: Isospin-Up Quark Mixing in Rest Mass Space

Because, comparatively speaking, the top quark mass is so huge and the up quark mass is so small, even after taking square roots the top-to-up ratio is about 280-to-1. So, any visual representation drawn to scale with be difficult to see. Therefore, in the above we have rescaled the axis for the top mass by dividing by 10 and rescaled the axis for the up mass by multiplying by 10. What is

remarkable is not only that the Fermi vev of about 246.22 GeV can be rotated in this square root space to produce the mass of each quark as illustrated, *but that the azimuthal and polar angles correspond also to two of the three CKM mixing angles*.

One final point is worth noting before we move on to examine the isospin-down quark masses. As between the first and second parameterizations, we also uncovered two other angles θ_{I32} and θ_{II32} in (14.10). And we did not develop an available third parameterization using what we shall denote as θ_{III21} and θ_{III31}. This is because as noted, the angles obtained from the bi-unitary transformations in (14.8) are parameter-dependent but the masses and their couplings are not nor can they be. So, once we attach a physical significance to $\cos^2\theta_{I21}$ and $\cos^2\theta_{II31}$ in (14.12) we have squeezed all of the *independent* information we can out the bi-unitary transformations. The remaining angles θ_{I32} and θ_{II32} furnish no further information, as they are not independent of the physical connections established in (14.12) but simply contain redundant information. Likewise, the third parameterization using what we shall denote as θ_{III21} and θ_{III31} produces a $90°-\theta_{III21}=\theta_{I21}$ and $\theta_{III31}=\theta_{I32}$ which effectively rotates of θ_{I21} by 90 degrees, a renames θ_{I32} to θ_{III31}, and in the process, also flip the signs of all the square roots which contain G_c. As such, this too is redundant and adds no new salient data.

15. Theory of Fermion Rest Masses and Mixing: Down, Strange and Bottom

If it is possible to express the three up quark masses as $m_u, m_c, m_t = F\left(v, \theta_{C31}, \theta_{C23}\right)$, and given that the two CKM angles parameterize generation changes during weak beta decays between isospin-up and isospin-down quarks, and because $\theta_{C12}=12.975\pm0.026°$ in (14.11) is still unaccounted for, it is natural to examine whether a carbon copy of the bi-unitary transformations in the last section can be used the characterize the down, strange and bottom quark masses in a similar fashion, while also relating $\theta_{C12}=12.975\pm0.026°$ to these masses. From here, to avoid notational confusion, we shall start to use the subscript $\Downarrow$ to denote various angles and objects associated with the isospin-down quarks when necessary to distinguish from the results of the last section, and will add the subscript $\Uparrow$ to the objects and angles of the last section when necessary to establish a clear distinction.

To cut right to the chase, let us replicate (14.9a) identically, but with the substitutions $u\mapsto d$, $c\mapsto s$ and $t\mapsto b$, as such:

$$G_{dsb}=\begin{pmatrix} G_b & \sqrt{G_bG_s} & \sqrt{G_bG_d} \\ \sqrt{G_bG_s} & G_s & \sqrt{G_sG_d} \\ \sqrt{G_bG_d} & \sqrt{G_sG_d} & G_d \end{pmatrix}=\begin{pmatrix} \mathrm{c}_{32}{}^2 & \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{c}_{21} & \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{s}_{21} \\ \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{c}_{21} & \mathrm{s}_{32}{}^2\,\mathrm{c}_{21}{}^2 & \mathrm{s}_{32}{}^2\,\mathrm{c}_{21}\,\mathrm{s}_{21} \\ \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{s}_{21} & \mathrm{s}_{32}{}^2\,\mathrm{c}_{21}\,\mathrm{s}_{21} & \mathrm{s}_{32}{}^2\,\mathrm{s}_{21}{}^2 \end{pmatrix}_{I\Downarrow}, \tag{15.1a}$$

$$G_{dsb}=\begin{pmatrix} G_b & \sqrt{G_bG_s} & \sqrt{G_bG_d} \\ \sqrt{G_bG_s} & G_s & \sqrt{G_sG_d} \\ \sqrt{G_bG_d} & \sqrt{G_sG_d} & G_d \end{pmatrix}=\begin{pmatrix} \mathrm{c}_{32}{}^2\,\mathrm{c}_{31}{}^2 & \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{c}_{31} & \mathrm{c}_{32}{}^2\,\mathrm{c}_{31}\,\mathrm{s}_{31} \\ \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{c}_{31} & \mathrm{s}_{32}{}^2 & \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{s}_{31} \\ \mathrm{c}_{32}{}^2\,\mathrm{c}_{31}\,\mathrm{s}_{31} & \mathrm{c}_{32}\,\mathrm{s}_{32}\,\mathrm{s}_{31} & \mathrm{c}_{32}{}^2\,\mathrm{s}_{31}{}^2 \end{pmatrix}_{II\Downarrow} . \tag{15.1b}$$

Here, it is clear that $G_b+G_s+G_d=1$. As with (14.8) and (14.9), these coupling matrices utilizing the first and second parameterizations arise following a bi-unitary transformation $M_{dsb}c^2\rightarrow M'_{dsb}c^2=U^{\dagger}{}_{\Downarrow}M'_{dsb}c^2U_{\Downarrow}$ in which before the transformation, M_{dsb} contains all of the rest mass in the bottom quark. Because the diagonals sum to 1, $m_dc^2+m_sc^2+m_bc^2$ is invariant under these unitary transformations. Of course, however, the empirical $m_dc^2+m_sc^2+m_bc^2\neq m_uc^2+m_cc^2+m_tc^2=\frac{1}{\sqrt{2}}v$. So, to mirror the development of the last section we shall need to postulate a new, second vev *defined* by:

$$\tfrac{1}{\sqrt{2}}v_{\Downarrow}\equiv m_dc^2+m_sc^2+m_bc^2=4.2797^{+0.284}_{-0.192}\ \text{GeV}\ , \tag{15.2}$$

while re-denoting $\frac{1}{\sqrt{2}}v_{\Uparrow}=174.1035847\ \text{GeV}$ to identify this as the vacuum which, cut by the same $\sqrt{2}$ factor, is now understood to be equal to the sum of the isospin-up quarks. The be clear: at the moment, the existence of this second vev this is a postulate, intended to see if we can account for the remaining CKM angle $\theta_{C12}=12.975\pm0.026\,^\circ$ in the same way we have already accounted for the other two CKM angles. If we can, then the postulate is validated. If, not, then it is not. Along with the definition above, we have used the empirical data in [42] to provide a numerical value for $\frac{1}{\sqrt{2}}v_{\Downarrow}=4.2797^{+0.284}_{-0.192}\ \text{GeV}$, where the upside error of $284\ \text{MeV}$ is based on the unlikely event of all three quarks having a mass at the top end of their error bars and the downside error of $192\ \text{MeV}$ conversely is based on all three quarks being at the low end.

Now, in (13.7b) we calculated couplings *G* for the down, strange and bottom quarks which were based on the relation $G_f=m_fc^2/\frac{1}{\sqrt{2}}v_{\Uparrow}$. But in introducing $\frac{1}{\sqrt{2}}v_{\Downarrow}$ with a much smaller energy, we are implicitly introducing the prospect that the potential $V(\phi_h)=\frac{1}{2}\mu^2\phi_h{}^2+\frac{1}{4}\lambda\phi_h{}^4+...$ not only has a first minimum at $v_{\Uparrow}=\sqrt{2}\cdot174.1035847\ \text{GeV}$, but has a *second minimum* at $v_{\Downarrow}=\sqrt{2}\cdot4.2797^{+0.284}_{-0.192}\ \text{GeV}$. This in turn requires us to no longer ignore the higher order terms in the potential which will be of order $\phi_h{}^6$, $\phi_h{}^8$, $\phi_h{}^{10}$ and so on, because we cannot have a second minimum (and perhaps a third for the charged lepton masses and a fourth right near zero for the neutrino masses) without these higher order terms. We will examine this more closely in the next section, but for the moment, let us simply posit that there is some $V(\phi_h)$, not yet known, which has a second minimum at $v_{\Downarrow}=\sqrt{2}\cdot4.2797^{+0.284}_{-0.192}\ \text{GeV}$, and indeed, which is ascertained *subject to the requirement* that it have this second minimum at this exact energy as well as the usual first minimum at the energy $v_{\Uparrow}=\sqrt{2}\cdot174.1035847\ \text{GeV}$.

Now, because the trace of the matrices in (15.1) sums to 1 by trigonometric identity and thus $G_d + G_s + G_b = 1$, the relation (15.2) requires us to recalibrate the coupling for each individual quark to $G_{d,s,b} = m_{d,s,b}c^2 / \frac{1}{\sqrt{2}} v_{\Downarrow}$, using the second minimum at $v_{\Downarrow} = \sqrt{2} \cdot 4.2797^{+0.284}_{-0.192}$ GeV rather than the first minimum at $v_{\Uparrow} = 246.2196508\,\text{GeV}$. Similarly to the procedure followed at (14.9), we use (15.1a) to calculate $\mathrm{c}_{I\Downarrow 32}{}^2 = G_b$ followed by $\mathrm{c}_{I\Downarrow 21}{}^2 = G_s / \mathrm{s}_{I\Downarrow 32}{}^2$ followed by the two angles, and (15.1b) to calculate $\mathrm{s}_{II\Downarrow 32}{}^2 = G_s$ followed by $\mathrm{c}_{II\Downarrow 31}{}^2 = G_b / \mathrm{c}_{II\Downarrow 32}{}^2$ followed by the two angles. However, unlike in the last section where $v_{\Uparrow}$ was independently-known because it is simply the vev energy magnitude associated with the Fermi coupling constant, the $\frac{1}{\sqrt{2}} v_{\Downarrow} = 4.2797^{+0.284}_{-0.192}$ GeV in (15.2) is itself a function of the down, strange and bottom masses and so is subject to their error bars. Moreover, the $G_{d,s,b}$ of the individual quarks are interdependent with and so subject to the error bars of the other two quarks. As a result, we shall review four different calculations each based on different assumptions about the error bars in the quark mass measurements and in the CKM mixing angle $\theta_{C12} = 12.975 \pm 0.026\,^\circ$ in (14.11).

Drawing again from PDG's [42], we start with the individual quark masses $m_d c^2 = .0047^{+0.0005}_{-0.0003}$ GeV, $m_s c^2 = .095^{+.009}_{-0.003}$ GeV and $m_b c^2 = 4.18^{+0.04}_{-0.03}$ GeV which, it will be noted, sum to the result in (15.2). In the first calculation we simply use the central value of the error bars in PDG's [42] for each of the three quark masses to calculate the four mass mixing angles as reviewed in the previous paragraph, as such:

$$\begin{aligned} \theta_{I\Downarrow 32} &= 0.153\,\text{rad} = 8.779^\circ \\ \theta_{II\Downarrow 32} &= 0.150\,\text{rad} = 8.568^\circ \\ \theta_{I\Downarrow 21} &= 0.219\,\text{rad} = 12.540^\circ \\ \theta_{II\Downarrow 31} &= 0.034\,\text{rad} = 1.921\,^\circ \end{aligned} \quad . \tag{15.3}$$

At the lower end of the empirical $\theta_{C12} = 12.975 \pm 0.026\,^\circ$, is the value $\theta_{C12} = 12.949^\circ$, which differs from $\theta_{I\Downarrow 21} = 12.540^\circ$ in (14.10) by a mere 0.409°. Coupled with having already connected two mass mixing angles to the real CKM angles in (14.12), this leads us to suspect that $\theta_{I\Downarrow 21}$ is in fact physically equivalent to θ_{C12}, i.e., that $\theta_{I\Downarrow 21} = \theta_{C12}$. So, the next step is to see if such a suspected connection falls within the error bars for the three quark masses that went into the calculation summarized in (14.10).

It turns out that $\theta_{I\Downarrow 21}$ which was calculated to be 12.540° in (14.10) is very-sensitive to variations in the down quark mass, is moderately-sensitive to variations in the strange quark mass, and is virtually unaffected by variations in the bottom quark mass. So for a second calculation, we leave the strange and the bottom masses alone at their centers by using $m_s c^2 = .095$ GeV and $m_b c^2 = 4.18$ GeV, and simply see whether there is some value for the down quark mass that will

enable $\theta_{I\Downarrow 21} = \theta_{C12}$ to in fact become a valid relation within the errors of $m_d c^2 = .0047^{+0.0005}_{-0.0003}$ GeV and $\theta_{C12} = 12.975 \pm 0.026\,°$. The combinations of results turn out to be:

$$\begin{aligned}
&\text{if } m_b c^2 = 4.18 \text{ GeV and } m_s c^2 = 95 \text{ MeV and } m_d c^2 = 5.064 \text{ MeV, then } \theta_{I\Downarrow 21} = 13.001° \\
&\text{if } m_b c^2 = 4.18 \text{ GeV and } m_s c^2 = 95 \text{ MeV and } m_d c^2 = 5.043 \text{ GeV, then } \theta_{I\Downarrow 21} = 12.975° \\
&\text{if } m_b c^2 = 4.18 \text{ GeV and } m_s c^2 = 95 \text{ MeV and } m_d c^2 = 5.022 \text{ GeV, then } \theta_{I\Downarrow 21} = 12.949°
\end{aligned} \qquad (15.4)$$

That is, now in MeV, with the bottom and strange quarks left at their centers, a down quark mass in the range $m_d c^2 = 5.043 \pm .021$ MeV corresponds to the range $\theta_{C12} = 12.975 \pm 0.026\,°$ of the first and second generation CKM mixing angle. Because the error bars for the down quark mass can be as high as $m_d c^2 = 5.2$ MeV, we have now established that $\theta_{I\Downarrow 21} = \theta_{C12}$ can indeed be a valid physical relationship within the known error bars for the isospin-down quark masses and the CKM mixing angles. It also turns out that for the down quark mass taken closer to its central value $m_d c^2 = 4.7$ MeV, it is necessary to reduce the strange quark mass somewhat to stay within the range of $\theta_{C12} = 12.975 \pm 0.026\,°$ for the CKM mixing angle.

So, in a third calculation, knowing that $\theta_{I\Downarrow 21}$ is most sensitive to the down mass which needs to be elevated above $m_d c^2 = 4.7$ MeV to hit the CKM target of $\theta_{C12} = 12.975 \pm 0.026\,°$, we start with a lower down quark mass assumed now to be $m_d c^2 = 4.9$ MeV. Then, we examine the ranges of acceptable values for the strange quark mass which achieve $\theta_{C12} = 12.975 \pm 0.026\,°$. The result of this calculation are as follows:

$$\begin{aligned}
&\text{if } m_b c^2 = 4.18 \text{ GeV and } m_d c^2 = 4.9 \text{ MeV and } m_s c^2 = 91.918 \text{ MeV, then } \theta_{I\Downarrow 21} = 13.001° \\
&\text{if } m_b c^2 = 4.18 \text{ GeV and } m_d c^2 = 4.9 \text{ MeV and } m_s c^2 = 92.299 \text{ MeV, then } \theta_{I\Downarrow 21} = 12.975°. \\
&\text{if } m_b c^2 = 4.18 \text{ GeV and } m_d c^2 = 4.9 \text{ MeV and } m_s c^2 = 92.683 \text{ MeV, then } \theta_{I\Downarrow 21} = 12.949°
\end{aligned} \qquad (15.5)$$

Noting again that $m_s c^2 = .095^{+.009}_{-0.003}$ GeV, we see that for the top line calculation the strange mass falls just below the error bar, while for the middle and bottom line calculations the strange mass ends up below its center but still within the PDG error range. Weighing all of the data, for the example of $m_d c^2 = 4.9$ MeV whereby the down and strange quarks "share" the variations with the down moved above center but not as high as in (15.5) and to compensate the strange is moved below center, and given that with a $^{+.009}_{-0.003}$ GeV variation the strange quark has less movement available on the low end than on the high end, it seems most reasonable to expect that θ_{C12} is likely on the low end of $\theta_{C12} = 12.975°$ than on the high end. That is, with $\theta_{I\Downarrow 21} = \theta_{C12}$ taken to be a correct physical relationship given that it is in fact true within the experimental and scheme-dependent error bars, we expect that a) the down quark mass is higher than the middle of $m_d c^2 = 4.7^{+0.5}_{-0.3}$ MeV, b) the strange quark mass is lower than the middle of $m_s c^2 = 95^{+9}_{-3}$ MeV, and c) the CKM angle is lower than the middle of $\theta_{C12} = 12.975 \pm 0.026\,°$.

But what is most important is that $\theta_{I\Downarrow 21} = \theta_{C12}$ *is in fact a correct physical relationship within the known experimental and scheme-dependent errors for the pertinent empirical data.* Once this relationship is taken to be a given, it then becomes possible to more finely tune the up and strange masses and the CKM angle θ_{C12}. Again, the bottom mass has negligible impact on any of this. So, we now take the step of establishing $\theta_{I\Downarrow 21} = \theta_{C12}$ as a true physical relationship, and adding this to (14.2) updated to differentiate isospin-up from isospin-down, whereby:

$$\boxed{\begin{array}{l} \theta_{I\Downarrow 21} \equiv \theta_{C12} = 12.975 \pm 0.026\,^\circ \\ \theta_{II\Downarrow 31} \quad\;\; = 1.921\,^\circ \\ \theta_{I\Uparrow 21} \equiv \theta_{C23} = 2.415 \pm 0.053\,^\circ \\ \theta_{II\Uparrow 31} \equiv \theta_{C13} = 0.209^{+0.015}_{-0.013}\,^\circ \end{array}}. \tag{15.6}$$

Now, *all three of the CKM mixing angles have been connected to mixing angles which are the direct result of bi-unitary transformations operating on quark mass matrices.* There is also a fourth "leftover" angle $\theta_{II\Downarrow 31} = 1.921\,^\circ$, also shown.

In a fourth and final calculation, which also necessitates a brief preface, we address the scheme-dependency of the quark masses about which to this point we have been speaking loosely. Although the quark masses deduced from hadronic scattering experiments are scheme-dependent as reviewed in [44], this does not mean we ought to conclude the quarks do not each have an objective mass that is scheme-independent, as do the leptons. In this regard, the key statement in [44] is that "quark masses therefore cannot be measured directly, but *must be determined indirectly through their influence on hadronic properties*." Ordinarily, these influences are observed in scattering experiments. However, in [11.22] of [47] and [10.1] of [48], the author demonstrated the existence of a pair of simultaneous equations

$$\begin{cases} 3\left(m_d - m_u\right)/\left(2\pi\right)^{\frac{3}{2}} = m_e \\ M_n - M_p = m_u - \left(3m_d + 2\sqrt{m_\mu m_d} - 3m_u\right)\Big/\left(2\pi\right)^{\frac{3}{2}} \end{cases} \tag{15.7}$$

through which the up and down quark masses may be deduced with extremely high precision based on the tightly-known, scheme-independent rest mass of the electron m_e and the tightly-known, scheme independent difference $M_n - M_p$ between the neutron mass and the proton mass. In this scheme, named the "Electron, Proton, Neutron (EPN) scheme," the electron, proton and neutron masses as well as nuclear binding energies and mass defects express *indirect* influences and manifestations of objective quark masses and are essentially "fingerprints" or parts of a "nuclear genome" from which the quark masses may be inferred. Using, (15.7), one may deduce very precise values for the up and down quark masses, which are:

$$\begin{array}{l} m_u = 0.002387339327 \text{ u} = 2.22379240 \text{ MeV} \\ m_d = 0.005267312526 \text{ u} = 4.90647034 \text{ MeV} \end{array} . \tag{15.8}$$

So, in the fourth and final calculation we use these very precise values of the up and down quark masses, which enables us to tighten up the error ranges for other quantities which are interconnected with these.

So now, with (15.8), we repeat the calculations of (15.4) and (15.5) to obtain:

$$\begin{array}{l} \text{if } m_b c^2 = 4.18 \text{ GeV and } m_d c^2 = 4.90647034 \text{ MeV and } m_s c^2 = 92.039 \text{ MeV, then } \theta_{I\Downarrow 21} = 13.001° \\ \text{if } m_b c^2 = 4.18 \text{ GeV and } m_d c^2 = 4.90647034 \text{ MeV and } m_s c^2 = 92.421 \text{ MeV, then } \theta_{I\Downarrow 21} = 12.975° \\ \text{if } m_b c^2 = 4.18 \text{ GeV and } m_d c^2 = 4.90647034 \text{ MeV and } m_s c^2 = 92.806 \text{ MeV, then } \theta_{I\Downarrow 21} = 12.949° \end{array} . \tag{15.9}$$

It should also be noted that keeping the down and strange masses as is, and using a bottom quark mass anywhere over the entire range given by $m_b c^2 = 4.18^{+0.04}_{-0.03}$ GeV, produces absolutely no change in the value of $\theta_{I\Downarrow 21}$. This is why we made the statement at (15.4) that this mass mixing angle and therefore the CKM angle $\theta_{C12} = \theta_{I\Downarrow 21}$ now related to this by (15.6) is virtually unaffected by variations in the bottom quark mass. Based on the match to the central empirical date $\theta_{C12} = 12.975 \pm 0.026°$ in (14.11), we shall henceforth use the middle line of (15.9) for the mid-range masses of the isospin-down quarks.

Because we have shown in (15.6) that $\theta_{I\Downarrow 21} = \theta_{C12} = 12.975 \pm 0.026°$ within experimental error bars, and because this is based on the postulate that the Higgs vacuum has a second vev $\frac{1}{\sqrt{2}} v_{\Downarrow} \equiv m_d c^2 + m_s c^2 + m_b c^2$ which represents another minimum of the Lagrangian potential, the connection established in (15.6) also is confirming evidence that this second vev postulated in (15.2) does in fact physically exist. The first minimum was of course independently-set by the fermi vev $v_{\Uparrow} = v = 246.2196508 \text{ GeV}$. But at the moment, all we know about $v_{\Downarrow}$ are the masses of the down, strange and bottom quarks of which this is the sum. Therefore, it is important to get the tightest error bar fit that we can for this second vev. For this purpose, we use the calculation in (15.9), and we use $\theta_{C12} = 12.975 \pm 0.026°$ to set the outer bounds on $v_{\Downarrow}$. As a result, recognizing that $m_b c^2 = 4.18^{+0.04}_{-0.03}$ GeV is the least-precise ingredient that goes into this vev, we calculate $\frac{1}{\sqrt{2}} v_{\Downarrow} = 4.2773 \pm 0.0004 \text{ GeV}$, which now replaces (15.2). Note, the high end of $v_{\Downarrow}$ corresponds to the low end of θ_{C12} and vice-versa. Any further precision in this number will depend entirely up ascertaining additional precision for the bottom mass. Rewritten without $\sqrt{2}$ to enable direct comparison to the Fermi vev including its error bars in [21], we have:

$$\begin{array}{l} v_{\Downarrow} = \sqrt{2}\left(m_d c^2 + m_s c^2 + m_b c^2\right) = 6.0491 \pm 0.0005 \text{ GeV} \\ v_{\Uparrow} = \sqrt{2}\left(m_u c^2 + m_c c^2 + m_t c^2\right) = 246.2196508 \pm 0.0000633 \text{ GeV} \end{array} . \tag{15.10}$$

Again, the outer bounds on $v_{\Downarrow}$ are now set, not by the masses, but by $\theta_{C12} = 12.975 \mp 0.026\,°$, which this sign flip correspondence explicit. The ratio $v_{\Uparrow} / v_{\Downarrow} \cong 40.7035$. The Higgs field rest energy extraction plots for the isospin-down quarks look identical to those of Figures 1 and 2 with the depth dependent upon the particle mass, with the exception that while $h_+ / v = h_+ / v_{\Uparrow}$ in Figures 1 and 2, for the isospin-down quarks it becomes $h_+ / v = h_+ / v_{\Downarrow}$. So, the energy drop begins in a vacuum with a magnitude that is smaller by a factor of just over 40.

It is also possible using $v_{\Uparrow}$ in (15.10) together with $m_u = 2.22379240$ MeV obtained in the EPN scheme via (15.7) and $m_c c^2 = 1.275^{+0.025}_{-0.035}$ GeV from [42] to tighten our knowledge of the top quark mass. This is presently known to be $m_t c^2 = 173 \pm 0.4$ GeV based on [42]. Now, the central value and error bar range are inherited from the charm quark, whereby:

$$\boxed{m_t c^2 = 172.826^{+0.025}_{-0.035}\ \text{GeV}}. \tag{15.11}$$

Moreover, because as noted the results in (15.9) are impervious to bottom mass swings over the whole range of $m_b c^2 = 4.18^{+0.04}_{-0.03}$ GeV, we can use the very precise down quark mass in (15.8) and the fairly tight $\theta_{C12} = \theta_{I\Downarrow 21}$ to calculate a more precise magnitude for the strange quark mass. This is presently known to be $m_s c^2 = 95^{+9}_{-3}$ MeV, and is now tightened to:

$$\boxed{m_s c^2 = 92.421 \pm^{+0.385}_{-0.382}\ \text{MeV}}. \tag{15.12}$$

These are both more than ten times as accurate as what is presently known for the top and strange quark masses, and constitute two additional empirical predictions of this theory which can and should be tested.

The connection in (15.6) whereby $\theta_{I\Downarrow 21} \equiv \theta_{C12} = 12.975 \pm 0.026\,°$ also means that there are some additional theoretical relations between the CKM mixing angles and the $\Downarrow$ quark masses as represented by their couplings $G_{d,s,b}\left(v_{\Downarrow}\right) = m_{d,s,b} c^2 / \frac{1}{\sqrt{2}} v_{\Downarrow}$. These relations, assembled with the earlier (14.13) updated to reflect that these are $\Uparrow$ quarks which use a different vev, are:

$$\boxed{\begin{aligned}
\cos^2\theta_{C21}=\cos^2\theta_{I\Downarrow 21}&=\frac{G_s\left(v_{\Downarrow}\right)}{\sin^2\theta_{I\Downarrow 32}{}^2}=\frac{G_s\left(v_{\Downarrow}\right)}{1-\cos^2\theta_{I\Downarrow 32}{}^2}=\frac{G_s\left(v_{\Downarrow}\right)}{1-G_b\left(v_{\Downarrow}\right)}\\
\cos^2\theta_{II\Downarrow 31}&=\frac{G_b\left(v_{\Downarrow}\right)}{\cos^2\theta_{II\Downarrow 32}}=\frac{G_b\left(v_{\Downarrow}\right)}{1-\sin^2\theta_{II\Downarrow 32}}=\frac{G_b\left(v_{\Downarrow}\right)}{1-G_s\left(v_{\Downarrow}\right)}\\
\cos^2\theta_{C23}=\cos^2\theta_{I\Uparrow 21}&=\frac{G_c\left(v_{\Uparrow}\right)}{\sin^2\theta_{I\Uparrow 32}}=\frac{G_c\left(v_{\Uparrow}\right)}{1-\cos^2\theta_{I\Uparrow 32}}=\frac{G_c\left(v_{\Uparrow}\right)}{1-G_t\left(v_{\Uparrow}\right)}\\
\cos^2\theta_{C13}=\cos^2\theta_{II\Uparrow 31}&=\frac{G_t\left(v_{\Uparrow}\right)}{\cos^2\theta_{II\Uparrow 32}}=\frac{G_t\left(v_{\Uparrow}\right)}{1-\sin^2\theta_{II\Uparrow 32}}=\frac{G_t\left(v_{\Uparrow}\right)}{1-G_c\left(v_{\Uparrow}\right)}
\end{aligned}}\quad . \tag{15.13}$$

We have also included the "leftover" angle $\theta_{II\Downarrow 31}=0.034\text{ rad}=1.921°$ which, at least for the moment, does not relate any other independently-known data, but which, like $\cos^2\theta_{C21}=\cos^2\theta_{I\Downarrow 21}$ above, is a function of the strange and bottom quark couplings.

Then, solving the top two (15.13) as simultaneous equations in G_b and G_s, while also using $G_d=\mathrm{s}_{I\Downarrow 32}{}^2\,\mathrm{s}_{I\Downarrow 21}{}^2=\mathrm{c}_{II\Downarrow 32}{}^2\,\mathrm{s}_{II\Downarrow 31}{}^2$ from (15.1) along with the results in (15.6), and assembling this with (14.14) also updated to reflect the $\Uparrow$ vacuum, we obtain:

$$\begin{aligned}
&G_b=\frac{\sin^2\theta_{C23}\cos^2\theta_{II\Downarrow 31}}{1-\cos^2\theta_{C23}\cos^2\theta_{II\Downarrow 31}};\ G_s=\frac{\cos^2\theta_{C23}\sin^2\theta_{II\Downarrow 31}}{1-\cos^2\theta_{C23}\cos^2\theta_{II\Downarrow 31}};\ G_d=G_s\tan^2\theta_{C12}=G_b\tan^2\theta_{II\Downarrow 31}\\
&G_t=\frac{\sin^2\theta_{C23}\cos^2\theta_{C31}}{1-\cos^2\theta_{C23}\cos^2\theta_{C31}};\quad G_c=\frac{\cos^2\theta_{C23}\sin^2\theta_{C31}}{1-\cos^2\theta_{C23}\cos^2\theta_{C31}};\quad G_u=G_c\tan^2\theta_{C23}=G_t\tan^2\theta_{C31}
\end{aligned}\quad . \tag{15.14}$$

In contrast to (14.14) where $m_u, m_c, m_t=F\left(v,\theta_{C31},\theta_{C23}\right)$ so that all three quark masses may be expressed as a function of three independently-know parameters, the three $G_{d,s,b}$ and associated quark masses are now reduced in "freedom" by only one independently-known parameter, namely, the third mixing angle $\cos^2\theta_{C21}$. So, we may write $m_d, m_s, m_b=F\left(m_d, m_s, \theta_{C21}\right)$ or alternatively $m_d, m_s, m_b=F\left(m_d, m_b, \theta_{C21}\right)$, because $\cos^2\theta_{C21}=G_s/\left(1-G_b\right)$ eliminates either G_s or G_b but not both as independent parameters. Thus, all told, we have now taken six previously-unexplained quark masses, and reduced this to two unexplained quark masses, plus the three CKM angles, plus the Fermi vev. So now we focus on the question of the remaining two quark masses.

Similarly to (14.15), using $G_b+G_s+G_d=1$ from (15.1) we may rewrite the upper two relations (15.13) as:

$$\cos^2\theta_{C21}=\cos^2\theta_{I\Downarrow 21}=\frac{G_s}{1-G_b}=\frac{G_s}{G_s+G_d}=\frac{\sqrt{G_s}^2}{\sqrt{G_s}^2+\sqrt{G_d}^2}=\frac{\sqrt{m_s}^2}{\sqrt{m_s}^2+\sqrt{m_d}^2}$$
$$\cos^2\theta_{II\Downarrow 31}=\frac{G_b}{1-G_s}=\frac{G_b}{G_b+G_d}=\frac{\sqrt{G_b}^2}{\sqrt{G_b}^2+\sqrt{G_d}^2}=\frac{\sqrt{m_b}^2}{\sqrt{m_b}^2+\sqrt{m_d}^2} \quad . \tag{15.15}$$

Then we may graph a similar geometric relationship in a three-dimensional rest mass space in which the square roots $\sqrt{m_s}$, $\sqrt{m_d}$ and $\sqrt{m_b}$ are plotted against the *x*, *y*, and *z* axes. Here, the masses are close enough once the square root is taken, that they may be drawn to scale, without re-scaling any axis. The result is shown below:

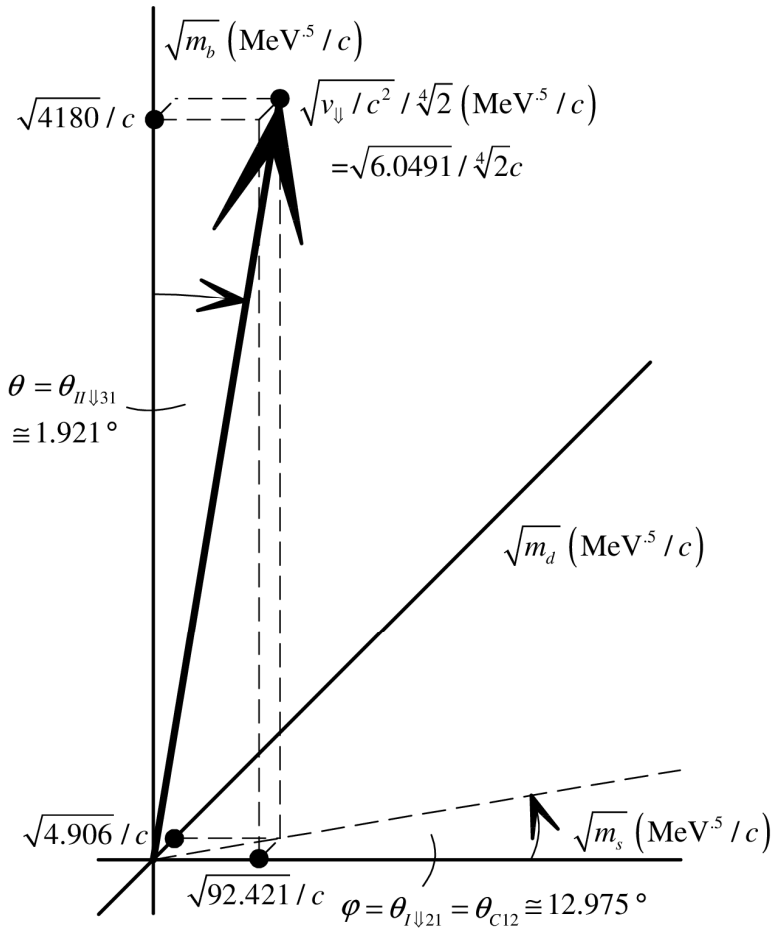

Figure 4: Isospin-Down Quark Mixing in Rest Mass Space

Here too, what is remarkable, taken together with Figure 3, is *that the azimuthal angle* φ *here corresponds also to the third of the three CKM mixing angles*. Each of the four angles in Figures 3 and 4 is needed to specify the projections of the vector associated with the vacuum into each of the individual masses, but only three of these angles are used also for CKM mixing.

Taking stock of where we are at the moment, there are two reasons, one for each, why there are two free parameters still remaining in the isospin-down portion of (15.14). First, there are only three real CKM mixing angles, not four. Two of those already went into $m_u, m_c, m_t = F\left(v, \theta_{C31}, \theta_{C23}\right)$. All that was left for the isospin-down masses was θ_{C21}. The "leftover" angle $\cos^2\theta_{II\Downarrow 31}$ previously-referenced at (15.6) and (15.13), if it had an independent basis, could squeeze out another degree of freedom from m_d, m_s, m_b. One possibility to consider is whether the "leftover" CP-violating phase angle δ_{13} associated with the 13 transition in the standard CKM parameterization bears some relation to "leftover" angle $\cos^2\theta_{II\Downarrow 31}$. But at the moment, whether this phase can provide an independent basis for the leftover mass mixing angle, or some other basis is required, is presently not clear. Second, for the isospin-up quarks, we had available $v_{\Downarrow} = v$ as an independent energy number supplied by the Fermi vev. On the other hand, at present, we have no *independent* information about $v_{\Downarrow}$ in (15.10). Rather, we only know about this from the $m_d c^2 + m_s c^2 + m_b c^2$ mass sum. So, to squeeze out another degree of freedom from the unknown numbers in the natural world, we would need to have *independent* knowledge of $v_{\Downarrow} = v$ separately from its value in (15.10) arrived at from the quark masses themselves. And this raises the final matter to be explored before we turn to the lepton masses.

16. The Two-Minimum, Two Maximum Lagrangian Potential for Quarks

When we first introduced the postulate of a second vev for the isospin-down quarks, this was speculative. But because this postulate led to the connection $\theta_{I\Downarrow 21} \equiv \theta_{C12} = 12.975 \pm 0.026\,^\circ$ with observed empirical data at (15.6), this connection is confirming evidence of this second vev. Normally, the Lagrangian potential $V = \frac{1}{2}\mu^2\phi_h^{\,2} + \frac{1}{4}\lambda\phi_h^{\,4} + ...$ with higher-order terms above $\phi_h^{\,4}$ neglected is used to establish the vev and the Higgs fields in a well-known manner, see, e.g., section 14.6 of [20] including Figure 14.3. But if is there is now to be a second minimum at $v_{\Downarrow}$, we can no longer neglect these higher order terms, because they will need to be responsible for providing this second minimum.

To review the standard model calculation so that we can consider the form of the required higher order terms, we start with *V* sans any higher order terms then calculate its first derivative $V' = dV / d\phi_h = \phi_h\left(\mu^2 + \lambda\phi_h^{\,2}\right)$. This will equal zero at either $\phi_h = 0$ or at $-\mu^2 / \lambda = \phi_h^{\,2}$, so these two points on the domain will be minima or maxima of the original *V* dependent on overall sign. We then assert the condition that the latter stationary point is to be a minimum at $\phi_h = v = v_{\Uparrow}$ (now distinguishing the Fermi vev $v_{\Uparrow}$ from $v_{\Downarrow}$) by defining $-\mu^2 / \lambda = \phi_h^{\,2} \equiv v_{\Uparrow}^{\,2}$. At the same time, the expansion $v_{\Uparrow} \mapsto v_{\Uparrow} + h\left(t, \mathbf{x}\right)$ about the vacuum using a Higgs field reveals a Higgs boson rest energy $m_h^{\,2}c^4 = -2\mu^2 = 2\lambda v_{\Uparrow}^{\,2}$ in the usual way. These two results may be combined to inform us that $\lambda = m_h^{\,2}c^4 / 2v_{\Uparrow}^{\,2}$, which says that the parameter λ is undetermined unless and until we know the mass of the Higgs boson. So, with these items of information we may return to the original *V* as well as $V' = dV / d\phi_h$ and rewrite these as standard model relations:

$$V = \lambda\left(-\tfrac{1}{2}v_{\Uparrow}{}^2\phi_h{}^2 + \tfrac{1}{4}\phi_h{}^4\right) = -\frac{1}{4}m_h{}^2c^4\phi_h{}^2 + \frac{1}{8}\frac{m_h{}^2c^4}{v_{\Uparrow}{}^2}\phi_h{}^4$$
$$V' = \lambda\phi_h\left(\phi_h{}^2 - v_{\Uparrow}{}^2\right) = \frac{m_h{}^2c^4}{2v_{\Uparrow}{}^2}\phi_h\left(\phi_h{}^2 - v_{\Uparrow}{}^2\right) \qquad (16.1a)$$

Now, we note from (15.10) that the two vevs which specify minima of the potential are stepped up from their respective fermion mass sums by a factor of $\sqrt{2}$. At the same time following symmetry breaking and defining a Higgs field to represent perturbations about the vev minimum, there is a similar $\sqrt{2}$ factor in the relations $v_{\Uparrow} + h = \phi_{1h} = \sqrt{2}\phi_h$ in (13.3), now using $v_{\Uparrow} = v$. So, at the minimum where h = 0, we have $v_{\Uparrow} = \phi_{1h} = \sqrt{2}\phi_h$, or $v_{\Uparrow}{}^2 = \phi_{1h}{}^2 = 2\phi_h{}^2$. In contrast, from V' in (16.1b) as written, the minimum will occur when $\phi_h{}^2 = v_{\Uparrow}{}^2$. Thus, in view of what we learned from (15.10) about the relation between mass sums and vevs, we see that (16.1b) needs to be recalibrated so that the minimum occurs when $\phi_{1h}{}^2 - v_{\Uparrow}{}^2 = 0$, not $\phi_h{}^2 - v_{\Uparrow}{}^2$. That is, (4.1) needs to be recalibrated by a factor of $\sqrt{2}$, which is most simply achieved by replacing $\phi_h \mapsto \phi_{1h} = \sqrt{2}\phi_h$ in (16.1b). Doing so, with $V' = dV / d\phi_{1h}$, this now becomes:

$$V = \lambda\left(-\tfrac{1}{2}v_{\Uparrow}{}^2\phi_{1h}{}^2 + \tfrac{1}{4}\phi_{1h}{}^4\right) = -\frac{1}{4}m_h{}^2c^4\phi_{1h}{}^2 + \frac{1}{8}\frac{m_h{}^2c^4}{v_{\Uparrow}{}^2}\phi_{1h}{}^4 = m_h{}^2c^4\left(-\frac{1}{4}\phi_{1h}{}^2 + \frac{1}{8}\frac{1}{v_{\Uparrow}{}^2}\phi_{1h}{}^4\right)$$
$$V' = \lambda\phi_{1h}\left(\phi_{1h}{}^2 - v_{\Uparrow}{}^2\right) = \frac{m_h{}^2c^4}{2v_{\Uparrow}{}^2}\phi_{1h}\left(\phi_{1h}{}^2 - v_{\Uparrow}{}^2\right) \qquad (16.1b)$$

Given that $\lambda = m_h{}^2c^4 / 2v_{\Uparrow}{}^2 > 0$, and examining V', we see that V will have a maximum at $\phi_{1h} = 0$ and a minimum at $\phi_{1h}{}^2 = v_{\Uparrow}{}^2$. When we break symmetry of $\phi_h = \frac{1}{\sqrt{2}}\left(\phi_{1h} + i\phi_{2h}\right)$ from (13.3) in the symmetry circle, in addition to choosing $\phi_{h2} = 0$, we also chose $\phi_{1h} = +v_{\Uparrow}$ as between the two possible choices $\phi_{1h} = \pm v_{\Uparrow}$. Empirically, $v_{\Uparrow} = v = 246.2196508 \pm 0.0000633$ GeV is obtained from the Fermi coupling constant G_F. We calculated $v_{\Downarrow} = 6.0491 \pm 0.0005$ GeV at (15.10) from the sum $\frac{1}{\sqrt{2}}v_{\Downarrow} = m_dc^2 + m_sc^2 + m_bc^2 = 4.2773 \pm 0.0004$ GeV of the isospin-down quarks. And while over four decades passed between when the Higgs boson was first postulated and when it was finally observed, today we have experimental data showing the Higgs boson to have a rest energy $m_hc^2 = 125.18 \pm 0.16$ GeV, see PDG's [49]. It is noteworthy that m_hc^2 is just a touch larger than half the Fermi vev, and to be precise, that $m_hc^2 - v_{\Uparrow} / 2 = 2.07 \pm 0.16$ GeV. Also, because we now know the Higgs mass empirically, we may deduce that the undetermined parameter $\lambda = m_h{}^2c^4 / 2v_{\Uparrow}{}^2 = 0.1292 \pm 0.0003$. Were the Higgs mass to be exactly equal to half the Fermi vev, we would have $\lambda = 1/8$. The consequences of this slight deviation from $\lambda = 1/8$ are important, and will drive many of the results now to be reviewed. Finally, at the minimum

where $\phi_{1h}{}^2 = v_{\Uparrow}{}^2$, using the center values of the data for m_h and $v_{\Uparrow}$, the upper (16.1b) yields $V\left(\phi_{1h}{}^2 = v_{\Uparrow}{}^2\right) = -\frac{1}{8} m_h{}^2 c^4 v_{\Uparrow}{}^2 = -\left(104.39 \text{ GeV}\right)^4$.

Turning to theory, and referring back to Figure 1 and 2, if $V\left(\phi_h\right)$ is to have a second (local) minimum at $\phi_{1h} = v_{\Downarrow}$ to provide a "nest" for isospin-down quarks along with its first (global) minimum at $\phi_{1h} = v_{\Uparrow}$ where isospin-up quarks are "nested," as well as its maximum at $\phi_{1h} = 0$, then it must now also have a *second maximum* at some definitive $v_{\Downarrow} < \phi_{1h} < v_{\Uparrow}$ in between the two minimum points. (Note that $\sqrt{2}\phi_{h+} \mapsto \sqrt{2}\phi_h = \phi_{1h}$ in the notation we are using presently.) This is not optional: mathematics demands that if a function as two minima, it inexorably must have a maximum somewhere between these two minima. So this raises an obvious question: where might this second *maximum* be? Just as $v_{\Uparrow}$ and $v_{\Downarrow}$ are physically meaningful numbers, we expect that the energy of ϕ_{1h} at this second maximum should have some physical meaning, for example, that it may be the rest mass of an elementary particle. The empirical rest masses of significance between $v_{\Downarrow}$ (about 6 GeV) and $v_{\Uparrow}$ (about 246 GeV) are the top quark mass, the masses M_W and M_Z of the electroweak vector bosons, and the Higgs mass. The top mass and the electroweak bosons are theoretically accounted for in other ways, so let's make an educated guess that it is the Higgs mass itself which establishes the second maximum. Specifically, as a preliminary hypothesis to again be tested against empirical data, let us hazard a guess (to which we shall momentarily make a slight adjustment) that:

$$\phi_{1h}\left(x^{\mathrm{M}}\right) \equiv m_h c^2 = \sqrt{2\lambda} v_{\Uparrow}. \tag{16.2}$$

In other words, this second maximum occurs precisely where ϕ_{1h} has an energy equal to the energy equivalent of the Higgs mass itself. Then let's again turn back to the empirical data to test this.

As noted just above, the Higgs rest mass $m_h c^2 = v_{\Uparrow}/2 + 2.07 \pm 0.16$ GeV is slightly above the halfway point between zero and the Fermi vev $v_{\Uparrow} = 246.2196508 \pm 0.0000633$ GeV. Another way to say this is that twice the Higgs mass is $2m_h c^2 = v_{\Uparrow} + 4.14 \pm 0.32$ GeV, which exceeds the Fermi vev by 4.14 ± 0.32 GeV. Comparing $\frac{1}{\sqrt{2}} v_{\Downarrow} = m_d c^2 + m_s c^2 + m_b c^2 = 4.2773 \pm 0.0004$ GeV from (3.10) we see that these two numbers match up *within experimental errors*. This means that within experimental errors, the Higgs mass is *exactly halfway* between $\frac{1}{\sqrt{2}} v_{\Downarrow} = 4.2773 \pm 0.0004$ GeV and $v_{\Uparrow} = 246.2196508 \pm 0.0000633$ GeV. Or, put differently, *if we now theoretically define the Higgs mass to be the average*:

$$\boxed{\begin{aligned} 125.18 \pm 0.16 \text{ GeV} = m_h c^2 &\equiv \frac{v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}}{2} = 125.2485 \pm 0.0002 \text{ GeV} \\ &= \frac{\sqrt{2}\left(m_u c^2 + m_c c^2 + m_t c^2\right) + m_d c^2 + m_s c^2 + m_b c^2}{2} \end{aligned}}. \tag{16.3}$$

of $v_{\Uparrow} = v = \sqrt{2}\left(m_u c^2 + m_c c^2 + m_t c^2\right)$ *and* $\frac{1}{\sqrt{2}} v_{\Downarrow} = m_d c^2 + m_s c^2 + m_b c^2$ *using the data from (15.10), we find that this relationship* $m_h c^2 = \left(v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}\right)/2$ *in (16.3) is true within experimental errors.*

The question now becomes whether $m_h c^2 = \left(v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}\right)/2$ above really is a relation of genuine physical significance, or is just a coincidence. There are good reasons why this is a real relation: First, a second maximum is required at some $v_{\Downarrow} < \phi_{1h} < v_{\Uparrow}$. Second, given this domain for the maximum, it makes particular sense in the present context for the maximum to be established right at the domain point where $\phi_{1h} = \frac{1}{\sqrt{2}}\left(v + h\right) = m_h c^2$, as in (16.2). Third, it makes sense for the maximum to be fairly close to the halfway point between $v_{\Downarrow}$ and $v_{\Uparrow}$, as $\phi_{1h} \equiv m_h c^2$ is. Fourth, $\phi_{1h} \equiv m_h c^2$ is in fact precisely halfway between $v_{\Uparrow}$ and $\frac{1}{\sqrt{2}} v_{\Downarrow}$ within experimental errors based on empirical data, which certainly qualifies as "fairly close" to the halfway point. Finally, the Higgs mass itself and the related parameter $\lambda = {m_h}^2 c^4 / 2{v_{\Uparrow}}^2$ have long been entirely unexplained as a theoretical matter. Given that we now have good empirical data for the Higgs mass, and that $m_h c^2 = \left(v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}\right)/2$ is confirmed by that data within experimental errors, we now regard (16.3) to be a new, correct theoretical relation of physical significance. Finally, because the empirical data on the right in (16.3) has a tighter error bound than the data on the left, we further use (16.3) as a *prediction* that as the Higgs mass becomes measured even more tightly than at present, it will be found to fit in the range $m_h c^2 = 125.2485 \pm 0.0002$ GeV. Now let's proceed forward regarding $m_h c^2 = \left(v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}\right)/2$ in (16.3) to be a true theoretical physical relation for the Higgs mass.

Following (15.14) we noted using $m_d, m_s, m_b = F\left(m_d, m_s, \theta_{C21}\right)$ that we had squeezed one degree of freedom from the isospin-down quark masses via the relation (15.13) for the CKM mixing angle θ_{C21} and these masses. With the discovery of (16.3), we now have a basis for expressing the previously-undetermined number $v_{\Downarrow}$ as a function $v_{\Downarrow} = F\left(v_{\Uparrow}, m_h\right)$. In other words, given the Higgs mass and the Fermi vev, we may deduce $v_{\Downarrow} = \sqrt{2}\left(m_d c^2 + m_s c^2 + m_b c^2\right)$. This means that if we choose to regard the Higgs mass as a "given" number, we have squeezed yet another unexplained energy number out of the parameters which drive the natural world. So, we can remove m_s from the prior relation and now write $m_d, m_s, m_b = F\left(m_d, m_h, \theta_{C21}\right)$. Together with $m_u, m_c, m_t = F\left(v, \theta_{C31}, \theta_{C23}\right)$, *we have now eliminated five (5) out of the six unexplained quark masses* and "explained" them insofar as they relate to θ_{C21}, θ_{C23}, θ_{C31}, v, and m_h. Of course, this does not "explain" why the five numbers θ_{C21}, θ_{C23}, θ_{C31}, v, and m_h have the empirical values that they have. But we have explained how these are related to the quark masses and so have rendered five of these six mass numbers into the status of "redundant" data. "Explaining" why these five numbers have their observed values, should an expected by-product of a successful GUT theory and its stages of symmetry breaking down to observable energies.

Again turning to the square roots of masses, if we write (16.3) as:

$$\left(\sqrt{v_{\Uparrow}}/c\right)^2+\left(\sqrt{v_{\Downarrow}}/\sqrt[4]{2}c\right)^2=\left(\sqrt{2m_h}\right)^2, \tag{16.4}$$

we see a Pythagorean relation amongst $\sqrt{v_{\Uparrow}}/c$, $\sqrt{v_{\Downarrow}}/\sqrt[4]{2}c$ and $\sqrt{2m_h}$, with the former two on the legs of a right triangle and the latter on the hypotenuse. This can be used to define an angle:

$$\sin\theta_v\equiv\frac{\sqrt{v_{\Downarrow}}/\sqrt[4]{2}c}{\sqrt{2m_h}};\quad \cos\theta_v=\frac{\sqrt{v_{\Uparrow}}/c}{\sqrt{2m_h}};\quad \tan\theta_v=\frac{\sqrt{v_{\Downarrow}}}{\sqrt[4]{2}\sqrt{v_{\Uparrow}}}, \tag{16.5}$$

such that θ_v effectively measures the magnitude of each vacuum in relation to one another and the Higgs mass. Using the data from (15.10) and (16.3) we calculate that the central value for this angle is $\theta_v=6.308519°$. This can be represented in the rather simple geometric Figure below:

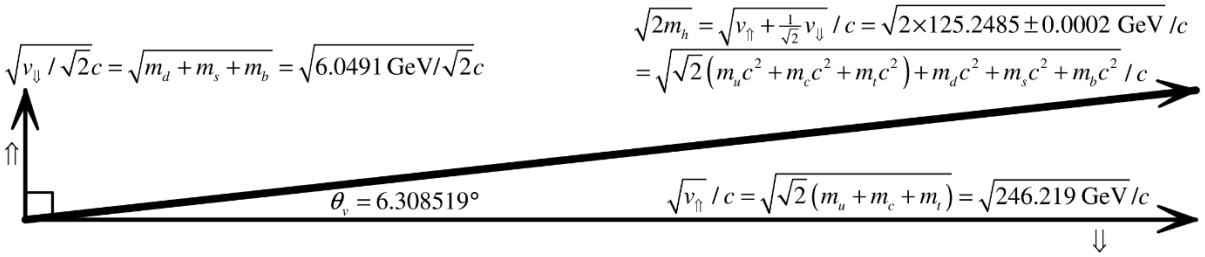

Figure 5: Vacuum and Higgs Mass Mixing in Quark Rest Mass Space

This is important, because in Figure 3 $\left(\sqrt{v_{\Uparrow}}/c\right)/\sqrt[4]{2}$ was the hypotenuse which was projected into each of three isospin-up mass roots and in Figure 4 $\sqrt{v_{\Downarrow}}/\sqrt[4]{2}c$ was the hypotenuse projected into each of three isospin-up mass roots. This means that (16.4) is the bridge between the two spaces in Figures 3 and 4, in a square root mass space that is overall six dimensional, as also seen from the bottom line of (16.3). So, with the coefficients and square roots as shown, one starts with the Higgs mass m_h axis which is the diagonal in Figure 5. That is projected into the two orthogonal axes, represented with $\Uparrow$ and $\Downarrow$ for the isospin-up and isospin-down vevs $v_{\Uparrow}$ and $v_{\Downarrow}$. Then, in three of the six dimensions $v_{\Uparrow}$ is further projected into the masses for the top, charm and up quarks as shown in the (not to scale) Figure 3, and in the other three of six dimensions $v_{\Downarrow}$ is projected into the bottom, strange and down masses. The azimuthal and polar angles in the former, and the azimuthal angle in the latter, then provide three the real angles for CKM mixing.

Once we advance $m_hc^2=\left(v_{\Uparrow}+\frac{1}{\sqrt{2}}v_{\Downarrow}\right)/2$ in (16.3) to a meaningful relation between the mass of the Higgs boson and the two vevs, we also may deduce that the long-undetermined parameter λ in $V=\frac{1}{2}\mu^2{\phi_{1h}}^2+\frac{1}{4}\lambda{\phi_{1h}}^4+...$, in view of (16.5), is theoretically given by:

$$\lambda = \frac{{m_h}^2 c^4}{2{v_{\Uparrow}}^2} = \frac{\left(v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}\right)^2}{8{v_{\Uparrow}}^2} = \frac{1}{8}\left(1 + \frac{1}{\sqrt{2}}\frac{v_{\Downarrow}}{v_{\Uparrow}}\right)^2 = \frac{1}{8}\left(1 + \tan^2\theta_v\right)^2 = 0.1292 \pm 0.0003 . \tag{16.6}$$

So physically, λ is now understood as another measure of how the energy equivalent of the Higgs rest mass is distributed into the two quark vevs in accordance with Figure 5, with these two vevs then parceling out their energies into the rest energies for each quark in their sector. In the limiting case where $\theta_v \to 0$ we will also have $\lambda \to 1/8$.

Before taking the next steps from here, let us now make a slight adjustment to the hypothesis of (16.2). Specifically, in light of what we now know from (16.3), and because $\phi_{1h}\left(x^{\mathrm{M}}\right) = v_{\Uparrow} + h\left(x^{\mathrm{M}}\right)$, having $\phi_{1h}\left(x^{\mathrm{M}}\right) = m_h c^2 = 125.2485 \pm 0.0002$ GeV as hypothesized in (16.2) with the more precise data from (16.3) be the maximum domain point, would mean that in terms of the Higgs field $h\left(x^{\mathrm{M}}\right)$, this maximum would be situated at:

$$h\left(x^{\mathrm{M}}\right) = m_h c^2 - v_{\Uparrow} = -m_h c^2 + \tfrac{1}{\sqrt{2}} v_{\Downarrow} = -\tfrac{1}{2}\left(v_{\Uparrow} - \tfrac{1}{\sqrt{2}} v_{\Downarrow}\right) = -120.9712 \pm 0.0002 \text{ GeV} . \tag{16.7}$$

While mathematics demands that there be a maximum *somewhere* in the domain $v_{\Downarrow} < \phi_{1h} < v_{\Uparrow}$, it does not tell us exactly where this maximum must be. The precise location is to be decided by physics. And another logical possibility is for the maximum to be where $h\left(x^{\mathrm{M}}\right) = m_h c^2$ rather than where $\phi_{1h}\left(x^{\mathrm{M}}\right) = m_h c^2$. So, we should also consider a modified hypothesis with a maximum at:

$$h\left(x^{\mathrm{M}}\right) = -m_h c^2 = -125.2485 \pm 0.0002 \text{ GeV} , \tag{16.8}$$

which, in terms of $\phi_{1h}\left(x^{\mathrm{M}}\right) = v_{\Uparrow} + h\left(x^{\mathrm{M}}\right)$, would mean, including (16.3), that the maximum is at:

$$\phi_{1h}\left(x^{\mathrm{M}}\right) = v_{\Uparrow} + h\left(x^{\mathrm{M}}\right) = v_{\Uparrow} - m_h c^2 = \tfrac{1}{2}\left(v_{\Uparrow} - \tfrac{1}{\sqrt{2}} v_{\Downarrow}\right) = 120.9712 \pm 0.0002 \text{ GeV} . \tag{16.9}$$

In effect, this shifts the maximum hypothesized in (16.2) to the left, toward the isospin-down vev, by the sum $\frac{1}{\sqrt{2}} v_{\Downarrow} = m_d c^2 + m_s c^2 + m_b c^2 = 4.2774 \pm 0.0003$ GeV of the charged lepton masses. Now, we are called upon to determine whether (16.9) is a better hypothesis than the preliminary (16.2), and this is a physics question, not a mathematics question.

The question requiring physics judgement is whether the maximum in the Lagrangian potential *V* ought to be at $h\left(x^{\mathrm{M}}\right) = -m_h c^2$ as in (16.8) versus at $\phi_{1h} = m_h c^2$ as in (16.2). Because standard model electroweak theory teaches that the *W* and *Z* bosons draw their rest energies from the Fermi vacuum, we anticipate the Higgs boson *H* draw its rest energy out of the vacuum in a similar way. Thus, we anticipate that all of these bosons will have a relation between $V_{(5)} / L^5$ on the horizontal axis and $\sqrt{2}\phi_{h+}$ on the vertical axis which is similar in character to that shown in

Figure 1 for the top quark which is a fermion. And in terms of the field we plan to use as the domain variable in the Lagrangian potential V as reviewed at (16.1b), this $\sqrt{2}\phi_{h+}$ is equivalent to ϕ_{1h}. So, considered from the viewpoint of Figure 1, we see that $h\left(x^{\mathrm{M}}\right)$ is a measure of how much energy has been *removed from the vacuum* in order to bestow a rest energy upon a particle, and that ϕ_{1h} is conversely a measure of how much energy is *retained by the vacuum* after the particle has acquired its rest energy. This means that if we choose $h\left(x^{\mathrm{M}}\right)=-m_h c^2$, then the V maximum will be based on the amount of energy *extracted* from the vacuum (with the minus sign indicating "extraction" or "removal"). Conversely, if we choose $\phi_{1h}=m_h c^2$, then the V maximum will be based on the amount of energy *retained* by the vacuum (with an implicit plus sign indicating "retention.") *So, the physics question is whether the V maximum should be based upon energy removed from the vacuum, versus upon energy retained by the vacuum.*

Now we turn to (15.10) which teaches that for quarks the Lagrangian potential V has *two minima*, one for isospin-up and one for isospin-down quarks. This is why we are needing to pinpoint a V maximum in the first place. We anticipate that $v_{\Downarrow}$ will establish energetically-favored states for isospin-down quarks, that $v_{\Uparrow}$ will establish energetically-favored states for isospin-up quarks, and as we shall momentarily examine in detail, that weak beta decays between isospin-up and isospin-down quarks will require the incoming quark at a beta-decay vertex to pass over or through the V maximum between $v_{\Downarrow}$ and $v_{\Uparrow}$ in order to decay into the outgoing quark. (We shall see that the top quark is an exception to what was just stated, because of its exceptionally-large rest mass.) We already know from the standard model that for a beta decay event to even take place, the incoming quark must draw sufficient energy out of the vacuum to cover the rest mass of the outgoing quark if the outgoing quark is more massive than the incoming quark, and that there must also be a sufficient energy draw on the order of 80 GeV to cover the rest mass of the W weak boson. If, however, there is also a V maximum sitting between $v_{\Downarrow}$ and $v_{\Uparrow}$ as is now being considered, then an energy draw from to the vacuum to enable beta decay will also be needed for a third purpose: *for the decaying fermion pass over or through this V maximum.*

Now, when a Higgs boson is formed with a rest mass $m_h c^2 \cong 125.2485$ GeV, this is an amount of energy which is no longer available for other purposes, such as for fermions to draw out of the vacuum to facilitate their beta decay over or through the V maximum. As such, this raises the barrier for beta decay, and can be analogized to having to climb up to the roof of a building starting in the basement rather than on the first floor. So, it is *the amount of energy drawn out of the vacuum, not the energy retained by the vacuum*, which is most-directly relevant to establishing the V maximum. Therefore, in view of all the foregoing, it makes the most physical sense for the V maximum to be defined at $h=-m_h c^2$ by the energy drawn out of the vacuum to bestow a mass upon the Higgs boson, and not at $\phi_{1h}=m_h c^2$ by the energy retained by the vacuum after the energy draw. The latter would place the V maximum at a domain point about $4.2774\ \text{GeV} \cong m_d c^2 + m_s c^2 + m_b c^2$ shy of the energy draw needed to give the Higgs boson its rest mass, and so would be close to the V maximum, but not right at the V maximum. Accordingly, we now modify our preliminary hypothesis of (16.2), and instead make the formal hypothesis based

on (16.8) and (16.9) that *the maximum of the Lagrangian potential V between* $v_{\Downarrow}$ *and* $v_{\Uparrow}$ *is situated at the domain point where* $h\left(x^{\mathrm{M}}\right)=-m_h c^2$, that is, as the point where the Higgs field is equal to the Higgs mass, and has a negative sign because this represents energy which has been drawn out of the Fermi vacuum.

Now we have all ingredients needed to revise the potential in (16.1b) with the higher-order terms necessary to provide the usual first minimum at $\phi_{1h}=v_{\Uparrow}=v$ and the usual first maximum at $\phi_{1h}=0$, as well as a second minimum at $\phi_{1h}=v_{\Downarrow}$ and, via (16.9) a second maximum at $\phi_{1h}=v_{\Uparrow}-m_h c^2$. We start with $V'=dV/d\phi_{1h}$ and build in these minima and maxima by *defining*:

$$\begin{aligned} V' &\equiv A\frac{{m_h}^2c^4}{2{v_{\Uparrow}}^2}\phi_{1h}\left({\phi_{1h}}^2-{v_{\Uparrow}}^2\right)\left({\phi_{1h}}^2-\left(v_{\Uparrow}-m_h c^2\right)^2\right)\left({\phi_{1h}}^2-{v_{\Downarrow}}^2\right) \\ &= A\frac{{m_h}^2c^4}{2{v_{\Uparrow}}^2}\begin{pmatrix} -{v_{\Downarrow}}^2{v_{\Uparrow}}^2\left(v_{\Uparrow}-m_h c^2\right)^2\phi_{1h}+\left({v_{\Downarrow}}^2{v_{\Uparrow}}^2+\left({v_{\Uparrow}}^2+{v_{\Downarrow}}^2\right)\left(v_{\Uparrow}-m_h c^2\right)^2\right){\phi_{1h}}^3 \\ -\left({v_{\Uparrow}}^2+{v_{\Downarrow}}^2+\left(v_{\Uparrow}-m_h c^2\right)^2\right){\phi_{1h}}^5+{\phi_{1h}}^7 \end{pmatrix}. \end{aligned} \tag{16.10}$$

This is constructed so that the leading terms $\left({m_h}^2c^4/2{v_{\Uparrow}}^2\right)\phi_{1h}\left({\phi_{1h}}^2-{v_{\Uparrow}}^2\right)$ in the top line above precisely match the usual V' in (16.1b). We also include an overall coefficient *A* which we will use to make certain that when we momentarily integrate (16.10), the leading ${\phi_{1h}}^2$ term of *V* in (16.1b) will continue to be $\frac{1}{4}{m_h}^2c^4{\phi_h}^2$, with all changes to *V* be introduced at higher order. This terms which we are matching stems from the "mass" term in $V=\frac{1}{2}\mu^2{\phi_h}^2+\frac{1}{4}\lambda{\phi_h}^4+...$ as reviewed at the start of this section. It will be seen by inspection that the top line in the above will be zero at all four of $\phi_{1h}=0$, $\phi_{1h}=m_h c^2$, $\phi_{1h}=v_{\Downarrow}$ and $\phi_{1h}=v_{\Uparrow}-m_h c^2$. The first two provide maxima and the latter two provide minima for *V* itself, or vice versa, depending on the overall sign in *A*.

Next, we easily integrate the above. For the leading term to match $\frac{1}{4}{m_h}^2c^4{\phi_{1h}}^2$ in *V* from (16.1b) we must set $A=1/{v_{\Downarrow}}^2\left(v_{\Uparrow}-m_h c^2\right)^2$. Also based on matching (16.1b) as an "initial condition," we discard any integration constant. We then consolidate and reduce to obtain:

$$V\left(\phi_{1h}\right)={m_h}^2c^4\begin{pmatrix} -\frac{1}{4}{\phi_{1h}}^2+\frac{1}{8}\frac{1}{{v_{\Uparrow}}^2}{\phi_{1h}}^4+\frac{1}{8}\left(\frac{1}{{v_{\Downarrow}}^2}+\frac{1}{\left(v_{\Uparrow}-m_h c^2\right)^2}\right){\phi_{1h}}^4 \\ -\frac{1}{12}\left(\frac{1}{{v_{\Uparrow}}^2{v_{\Downarrow}}^2}+\frac{1}{\left(v_{\Uparrow}-m_h c^2\right)^2}\frac{{v_{\Uparrow}}^2+{v_{\Downarrow}}^2}{{v_{\Uparrow}}^2{v_{\Downarrow}}^2}\right){\phi_{1h}}^6+\frac{1}{16}\frac{1}{\left(v_{\Uparrow}-m_h c^2\right)^2}\frac{1}{{v_{\Uparrow}}^2{v_{\Downarrow}}^2}{\phi_{1h}}^8 \end{pmatrix}. \tag{16.11}$$

Comparing with V in (16.1b), we indeed see the original $\phi_{1h}{}^2$ and $\phi_{1h}{}^4$ terms. But there are some new additions to the $\phi_{1h}{}^4$ tetm, and new $\phi_{1h}{}^6$ and $\phi_{1h}{}^8$ terms. These new terms, of course, are the ones we expect will deliver the second maximum and minimum as specified via (16.10).

To simplify calculation, it is very useful to restructure the above to separate terms which do not and which do have a $1/\left(v_{\Uparrow}-m_hc^2\right)^2$ coefficient, and to then explicitly show the result $m_hc^2=\left(v_{\Uparrow}+\frac{1}{\sqrt{2}}v_{\Downarrow}\right)/2$ from (16.3) and $v_{\Uparrow}-m_hc^2=\frac{1}{2}\left(v_{\Uparrow}-\frac{1}{\sqrt{2}}v_{\Downarrow}\right)$ from (16.7) as follows:

$$\begin{aligned}V\left(\phi_{1h}\right)&=m_h{}^2c^4\left(-\frac{1}{4}\phi_{1h}{}^2+\frac{1}{8}\frac{v_{\Uparrow}{}^2+v_{\Downarrow}{}^2}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^4-\frac{1}{12}\frac{1}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^6\right)\\&+\frac{m_h{}^2c^4}{\left(v_{\Uparrow}-m_hc^2\right)^2}\left(\frac{1}{8}\phi_{1h}{}^4-\frac{1}{12}\frac{v_{\Uparrow}{}^2+v_{\Downarrow}{}^2}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^6+\frac{1}{16}\frac{1}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^8\right)\\&=\frac{\left(v_{\Uparrow}+\frac{1}{\sqrt{2}}v_{\Downarrow}\right)^2}{4}\left(-\frac{1}{4}\phi_{1h}{}^2+\frac{1}{8}\frac{v_{\Uparrow}{}^2+v_{\Downarrow}{}^2}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^4-\frac{1}{12}\frac{1}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^6\right)\\&+\frac{\left(v_{\Uparrow}+\frac{1}{\sqrt{2}}v_{\Downarrow}\right)^2}{\left(v_{\Uparrow}-\frac{1}{\sqrt{2}}v_{\Downarrow}\right)^2}\left(\frac{1}{8}\phi_{1h}{}^4-\frac{1}{12}\frac{v_{\Uparrow}{}^2+v_{\Downarrow}{}^2}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^6+\frac{1}{16}\frac{1}{v_{\Uparrow}{}^2v_{\Downarrow}{}^2}\phi_{1h}{}^8\right)\end{aligned}\qquad(16.12)$$

So, the behavior of $V\left(\phi_h\right)$ is entirely driven by the two energy-dimensioned numbers in (15.10). The first is the Fermi vev $v_{\Uparrow}=v$ which establishes the usual minimum, and which we have learned here is related to the sum of the isospin-up quark masses via $v_{\Uparrow}=\sqrt{2}\left(m_uc^2+m_cc^2+m_tc^2\right)$. The second is the second vev $v_{\Downarrow}$ which establishes a second minimum and is related to the sum of the isospin-down quark masses via $v_{\Downarrow}=\sqrt{2}\left(m_dc^2+m_sc^2+m_bc^2\right)$. Additionally, the Higgs mass itself establishes a second maximum, but the new relation $m_hc^2=\left(v_{\Uparrow}+\frac{1}{\sqrt{2}}v_{\Downarrow}\right)/2$ uncovered in (16.3) means that only two of these energy numbers are truly independent of one another.

It is pedagogically-useful to graph the potential $V\left(\phi_{1h}\right)$ in (16.12) using the numerical values of $v_{\Uparrow}$ and $v_{\Downarrow}$ in (15.10), and / or the Higgs mass in (16.3). Substituting these into (16.12), reconsolidating terms at each order, and rounding the coefficient at each order to four significant digits, with ϕ_{1h} expressed in GeV thus $V\left(\phi_{1h}\right)$ in GeV^4, we obtain:

$$V\left(\phi_{1h}\right)\left[\text{GeV}^4\right]=-3922\phi_{1h}{}^2+53.76\phi_{1h}{}^4-0.003032\phi_{1h}{}^6+3.020\times10^8\phi_{1h}{}^8.\qquad(16.13)$$

Keeping in mind from following (13.3) that $V\left(\phi_{1h}\right)$ is a term in part of the Lagrangian density and so has physical dimensions of quartic energy, and that ϕ_{1h} is linear in energy, (16.13) produces:

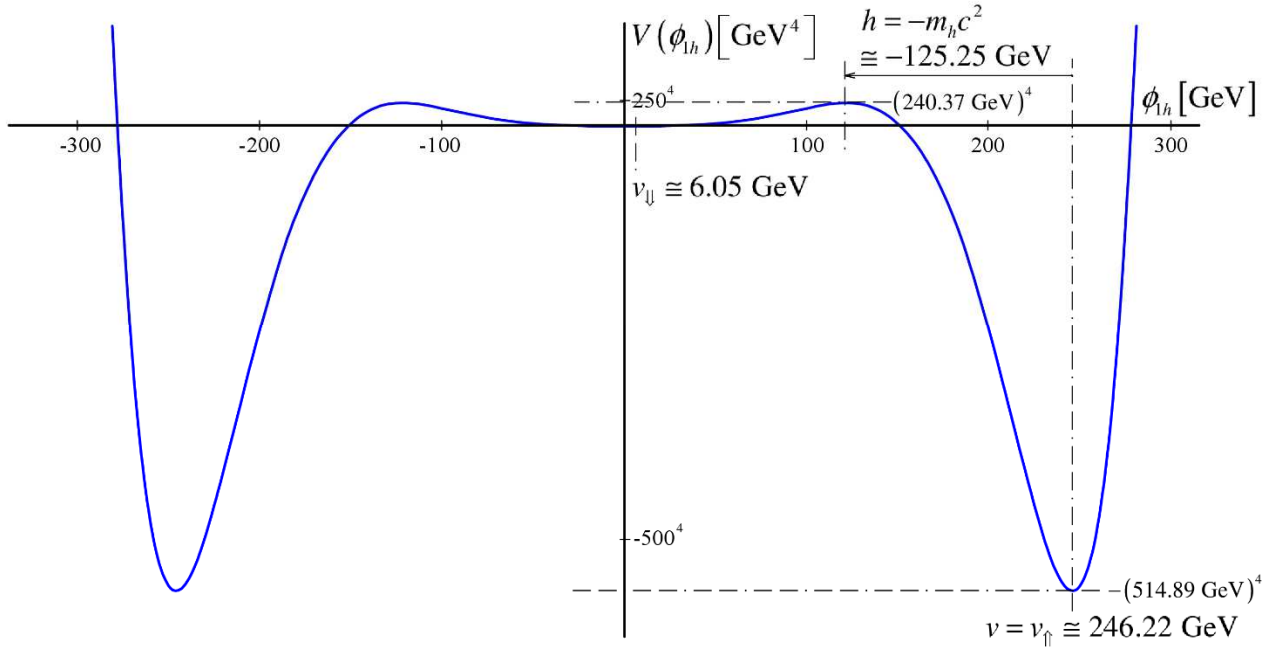

Figure 6: Lagrangian Potential for Quarks – Wide View

Above we see the usual minimum at $\phi_{1h}=v\cong 246.22$ GeV , where along the *y* axis we have $V\left(\phi_{1h}\right)\cong -\left(514.89\text{ GeV}\right)^4$. But we now have a new maximum at $\phi_{1h}=v_{\Uparrow}-m_hc^2=121.0$ GeV based on (16.9), and at this maximum, $V\left(\phi_{1h}\right)\cong\left(240.37\text{ GeV}\right)^4$. Closer to the origin is the usual maximum at $\phi_{1h}=0$ and the new minimum at $\phi_{1h}=v_{\Downarrow}\cong 6.05$ GeV , but by comparison, these are relatively extremely small, and impossible to see in Figure 6. So, it is also useful to also magnify the domain from $-10\text{ GeV}<\phi_{1h}<10\text{ GeV}$ in Figure 6, as is shown below:

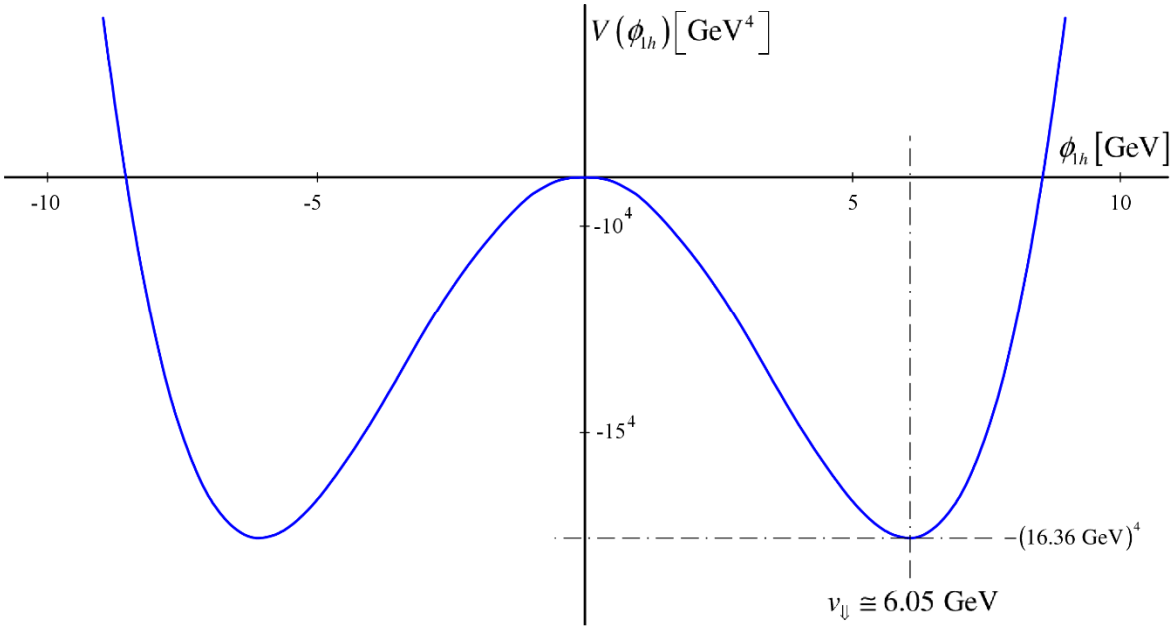

Figure 7: Lagrangian Potential for Quarks – Magnified Center View

Here, the usual maximum at $V(\phi_{1h}=0)=0$ is readily apparent, as is the new minimum at $\phi_{1h}=v_{\Downarrow}\cong 6.05$ GeV where $V(\phi_{1h})\cong -(16.36\text{ GeV})^4$. The above is simply an extremely magnified view of the region in Figure 6 close to the origin.

Note also that while the vertical depth of V at ${\phi_{1h}}^2={v_{\Uparrow}}^2$ was $-(104.39\text{ GeV})^4$ for $V=\lambda\left(-\frac{1}{2}{v_{\Uparrow}}^2{\phi_{1h}}^2+\frac{1}{4}{\phi_{1h}}^4\right)$ in (16.1b) based only on square and quartic field terms, now, in Figure 6, we have $V\cong -(514.89\text{ GeV})^4$ at the same Fermi vev global minimum ${\phi_{1h}}^2={v_{\Uparrow}}^2$. This substantially-increased depth is driven by the combination of setting $A=1/{v_{\Downarrow}}^2\left(v_{\Uparrow}-m_hc^2\right)^2$ going from (16.10) to (16.11) to preserve leading "mass term" $\frac{1}{2}\mu^2{\phi_h}^2$ in the original $V=\frac{1}{2}\mu^2{\phi_h}^2+\frac{1}{4}\lambda{\phi_h}^4+...$ with no change, and from the new local minimum at $v_{\Downarrow}$ and the new maximum at $\phi_{1h}=v_{\Uparrow}-m_hc^2$. That is, this increased depth is driven entirely by the new higher-order ${\phi_{1h}}^4$, ${\phi_{1h}}^6$ and ${\phi_{1h}}^8$ terms along with maintaining the ${\phi_{1h}}^2$ term as is.

But even with Figures 6 and 7, the energetic behavior of particles in these wells and the impact of the new maximum are not brought out as much as they could be, because ϕ_{1h} is linear in energy while V is quartic in energy. So, it is also useful to reproduce Figures 6 and 7 by taking the fourth root $\sqrt[4]{V(\phi_{1h})}$, and also by scaling the energies along the ordinate and the abscissa to match one another precisely. Of course, the fourth root of +1 has the quartic roots 1, –1, *i*, and –*i*, and below the *x* axis, to connect everything together, we wish to display what is really $-\sqrt[4]{-V(\phi_{1h})}$

using 1 for the quartic root. So, taking the fourth root along the vertical axis in Figure 6 and scaling what are now linear energy numbers in each axis to one another, we obtain the plot below:

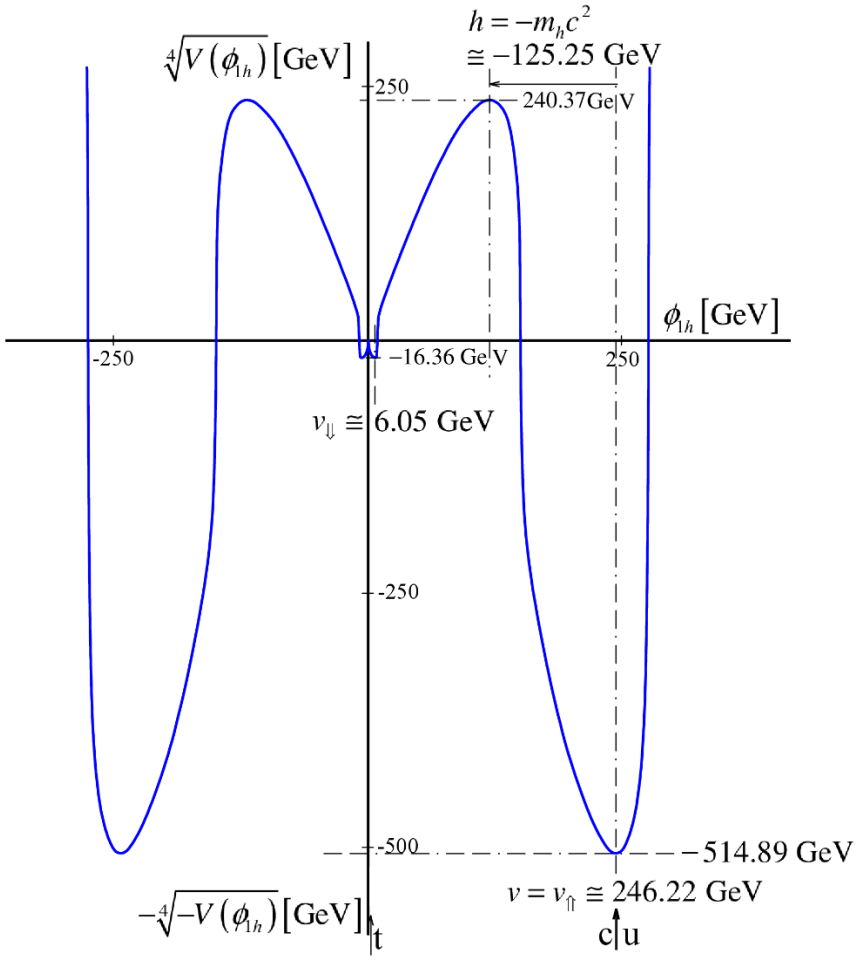

Figure 8: Lagrangian Potential for Quarks, Fourth Root – Wide View

Above, we are able to see both minima and both maxima in the same plot, although the central region is still rather small. Therefore, in Figure 9 below, we also magnify Figure 8 over the domain $-10 \text{ GeV} < \phi_{1h} < 10 \text{ GeV}$. This Figure 9 is equivalent to the fourth root of the magnified view of the potential in Figure 7.

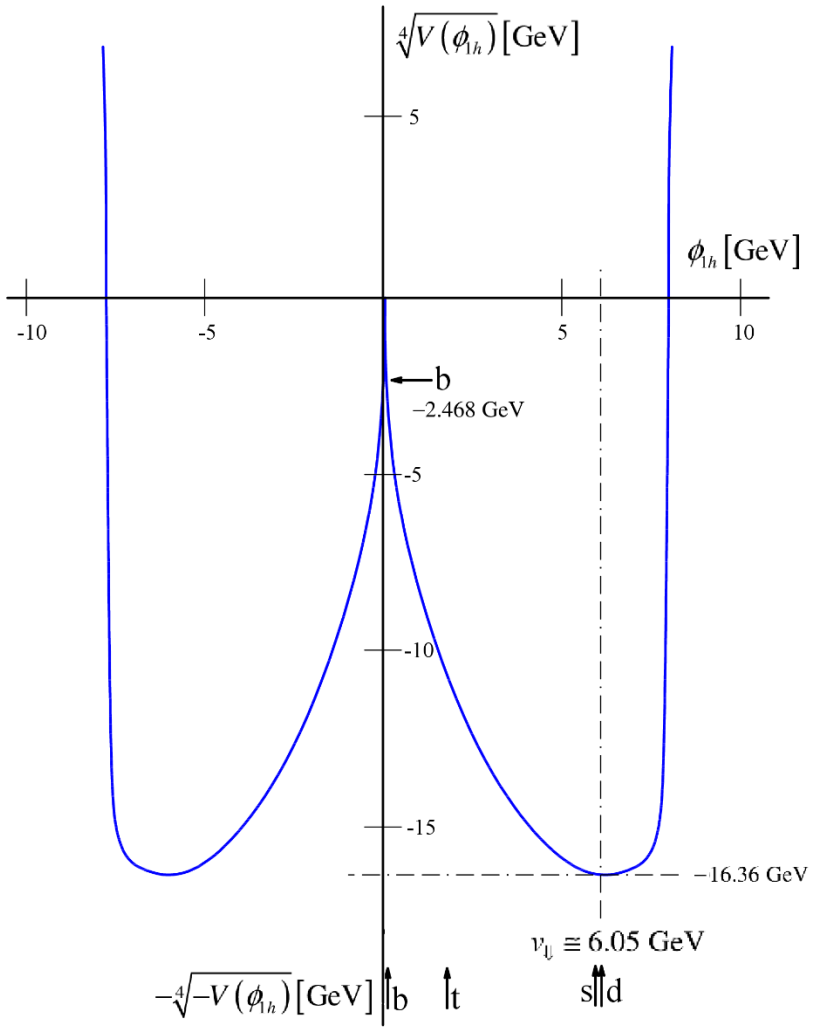

Figure 9: Lagrangian Potential for Quarks, Fourth Root – Magnified Center View

These two plots in Figures 8 and 9 help provide a deeper understanding of how quarks behave in the Higgs fields. First, it will be seen with energies linearized along both axes and scaled at 1:1, that the potential wells are very deep and steep. Moreover, it will be seen that the maximum at $V(\phi_{1h}=0)=0$ is not smooth as one might conclude looking at Figures 6 and 7. Rather, when comparing energies to energies to scale, this maximum is very steep, effectively coming to a sharp upward point with a slope that is infinite at the origin. Second, it is apparent, most clearly from Figure 8, that the $v_{\Downarrow}$ potential well establishes a *local* minimum while the $v_{\Uparrow}$ potential well presents a *global* minimum. The $v_{\Downarrow}$ local minimum has an energy depth of –16.36 GeV and the $v_{\Downarrow}$ global minimum has a depth of –514.89 GeV, about 31.47 times as large. Third, we see that there is high barrier between the two wells set by the new maximum which has a height of +240.37 GeV. Using $m_h c^2 = 125.2485 \pm 0.0002$ GeV as refined in (16.3), it should be noted that the ratio $240.37/125.25 = 1.919$, and so is slightly under twice the Higgs mass, and that the ratio

$514.89/125.25 = 4.111$ which is slightly over four times the Higgs mass. All this will be important shortly, to better understand the role of the Higgs field and boson in weak beta decay.

Now, let's cross- reference this with Figures 1 and 2, in which $\phi_{1h} = \sqrt{2}\phi_{h+}$. Recall how each quark uses the Higgs boson to extract energy from the vacuum and acquires its rest mass in accordance with (13.5) and (13.6). Figures similar to Figures 1 and 2 may be developed for all the other quarks and leptons, but all have the same basic character so we will not take the space to do so here. Also, recall as discussed after (15.10) that the rest energy extraction plots for the isospin-down quarks draw rest energy out of the vacuum in the much-less-energetic well at $\phi_{1h} = v_{\Downarrow}$, versus the isospin-up quarks drawing their rest energies in the more-energetic well at $\phi_{1h} = v_{\Uparrow}$ with $v_{\Uparrow} / v_{\Downarrow} \cong 40.70$. For each quark, ϕ_{1h} is plotted on the vertical axis of its own variant of Figures 1 and 2 and reaches minimum at $V_{(5)} / L^5 = 0$ along the horizontal axis. Recall that $V_{(5)} = x^0 x^1 x^2 x^3 x^5$ was defined at (11.22), and L^5 is a constant of integration with dimensions of length to the fifth power as an "initial condition," emerging from the integration (11.20), and that a further initial condition reviewed at the start of section 13 is that $V_{(5)} / L^5$ must have symmetry under the rotation and boost parameters of the Poincare group. Thus, the horizontal axes of Figures 1 and 2 are established by space and time (including t^5) coordinates, while ϕ_{1h} sits on the vertical axis of Figures 1 and 2. In contrast, in Figures 6 through 9 this exact same ϕ_{1h} is plotted on the horizontal axis. In short: ϕ_{1h} on the *vertical* axes of Figures 1 and 2 (and like-figures developed for other quarks) is synonymous with ϕ_{1h} on the *horizontal* axes of Figures 6 through 9, so these may be cross-referenced.

So, in Figures 8 and 9, we have also cross-referenced where the minima at the $V_{(5)} / L^5 = 0$ for each quark in their Figure 1 and 2 variants end up situating. First, in Figure 8, the up and charm quarks barely perturb away from the $v_{\Uparrow}$ vev minimum to extract the rest energy for their masses, as is indicated by the upward-pointing arrow designating their $V_{(5)} / L^5 = 0$ point of maximum energy extraction from the vacuum. However, the top quark, plotted in Figure 1, has the entirely unique characteristic of drawing almost all of the energy out of the vacuum, and this is from the larger well with $v_{\Uparrow} = v \cong 246.22$ GeV. Via (15.10), what remains in the vacuum and is not drawn out, is equivalent to the sum of the charm and up rest energies. Thus, for the top quark energy extraction in Figure 8, we see the top quark $V_{(5)} / L^5 = 0$ point perturbed so far to the left that it becomes nested on the left side of the $v_{\Downarrow}$ energy well. *This crossover characteristic is unique to the top quark.* Then in Figure 9 which has a closeup view of the $v_{\Downarrow}$ well, we see the down quark barely perturbed and the strange quark slightly more perturbed from the $v_{\Downarrow}$ minimum of this well, while the bottom quark is extremely perturbed almost to $\phi_{1h} = 0$, "hugging" the *x* axis. But in this $v_{\Downarrow}$ well, we also have the top quark which is a "visitor" from the $v_{\Uparrow}$ well because of its exceptionally large mass, and it too is perturbed well to the left of $v_{\Downarrow}$.

17. The Role of the Higgs Boson and its Mass in Weak Beta-Decays Between Quarks

Now, it has long been understood – at least in general if not specific terms – that the Higgs boson and associated fields are the responsible mechanism for giving rest masses to elementary particles, including fermions. What Figures 8 and 9 show is that the Higgs bosons and fields are also centrally involved in the mechanism for *weak interaction beta decays* between isospin-up and isospin-down quarks. Experimentally, this also means that close observations of beta decays may provide another good way to study the Higgs boson. It is helpful for the ensuing discussion of this to refer to the nine empirical components of the CKM mixing matrix V_{CKM}, such as may be found at [12.27] of PDG's [46].

Looking at Figure 8 and its centrally-magnified view of Figure 9, we see that there are two steep energy wells which, by design starting with (16.10), bottom out along the horizontal axis at $\phi_{1h} = v_{\Uparrow}$ and $\phi_{1h} = v_{\Downarrow}$. And by the same design there is also a high barrier between the wells which peaks at $\phi_{1h} = m_h c^2$. But only the down and strange quarks nest in the $v_{\Uparrow}$ well, because the top mass is so large that its Figure 1 minimum at $V_{(5)} / L^5 = 0$ has nested in the $v_{\Downarrow}$ well. So, the $v_{\Downarrow}$ well detailed in Figure 9 naturally nests the down, strange and bottom, *as well as the top quark*.

Beta decay, of course, only occurs between isospin-up and isospin-down quarks. For a decay event between an up or charm quark and a down, strange or bottom quark, the decaying quark must acquire enough energy to "jump" over the barrier peak at $\phi_{1h} = m_h c^2$. But uniquely, for a decay event between a top and any of the down, strange or bottom quarks, there is no need for the requisite energy to jump the barrier, because *both* the top and bottom quarks are nested in the same well, owing to the unique crossover properties of the top quark. This would suggest that for same-generation transitions the same-well diagonal CKM element $V_{tb} = 0.999105 \pm 0.000032$ ought to more energetically-favored thus substantially closer to 1 than either of the well-changing, barrier-jumping $V_{cs} = 0.97359^{+0.00010}_{-0.00011}$ or $V_{ud} = 0.97446 \pm 0.00010$, which in fact it clearly is by a ratio of just under 30-to-1 in both cases. (Note different use of *V* than for the Lagrangian potential.)

Now let's take a closer look at the well-changing transitions, in which a charm or up quark beta-decays into a down, strange or bottom quark, or vice-versa. All of these transitions – which are in the top two rows of V_{CKM} in [12.27] of [46] – cannot happen without the fermions drawing sufficient energy out of the vacuum via the Higgs fields and bosons to "jump" over the Lagrangian potential maximum at $\sqrt[4]{V}\left(\phi_{1h} = v_{\Uparrow} - m_h c^2\right) \cong 240.37$ GeV. Given that fermions acquire their masses from the Higgs field drawing energy out of the vacuum in accordance with the upper equation (13.5), it seems that the energy to jump this barrier at $\phi_{1h} = m_h c^2$ would come from the very same source: the Higgs field and bosons. This is where the vertical heights of both the wells and the new maximum in Figures 8 and 9 come into play.

First, start with an up or charm quark in the $v_{\Uparrow}$ well. As noted earlier the energy deficit at the bottom of the $v_{\Uparrow}$ well is –514.89 GeV. And as seen in Figure 8, the up and charm quarks for all practical purposes nest at the bottom of this well, which is an energetically-preferred state. Ignoring the error bars for the moment, with $m_h c^2 \cong 125.2485$ GeV ratio $514.89/125.25 = 4.111$. So, the energy equivalent of just over four Higgs boson masses is needed just to get from the bottom of $v_{\Uparrow}$ to $V = 0$. Then, with a height of +240.37 GeV and $240.37/125.25 = 1.919$, the energy of just under two additional Higgs bosons is needed to scale the wall and beta decay from an up or charm in the $v_{\Uparrow}$ well, to any of a down, strange or bottom in the $v_{\Downarrow}$ well, from right-to-left in Figure 8. So even if these quarks utilize all of their rest energy to clear the well barrier, calculating $6.03 = 4.111 + 1.919$, they still need an energy boost totaling just over the rest masses of six Higgs bosons. Additionally, because all of $m_u < m_d$ and $m_u < m_s$ and $m_u < m_b$, any beta decay that starts with the up quark will require the new quark at the end to retain for its new rest mass, some of the energy that was used to boost it over the well wall. For the charm quark with $m_c > m_d$ and $m_c > m_s$ but $m_c < m_b$, after the barrier transition some of its rest energy is released back into the vacuum for the former two transitions, but for charm-to-bottom, some of the barrier-jump energy will be retained for additional rest mass.

Now, let's start in the $v_{\Downarrow}$ well and go the opposite direction left-to-right. As noted, top to bottom and vice versa decays are *intra-well* and so occur most freely, which is why $V_{tb} = 0.999105 \pm 0.000032$. For *inter-well* transitions we start with one of down, strange or bottom and need to hop the barrier in Figure 8. Here, the energy deficit at the bottom of the well is only –16.36 GeV, which is much less than the energy deficit of the $v_{\Uparrow}$ well. For all practical purposes, the strange and the down quarks nest at the bottom of this well, which is an energetically-preferred state. To raise these two quarks to the $V = 0$ level, because $16.36/125.25 = 1/7.66$, one needs to extract a little more than 1/8 of the energy of a Higgs boson from the vacuum. But from there, one still needs the energy of $240.37/125.25 = 1.919$ Higgs bosons to scale the barrier and transition into an up or charm quark in the $v_{\Uparrow}$ well. Even if the strange or down quark was to apply all of its rest energy to getting over the barrier, it would still need an assist from a total of three Higgs bosons to get over the top of the well barrier, because they start at about –16.36 GeV. In all cases a bottom quark will release energy into the vacuum following the decay because it will end up with a lower mass, a strange quark will need to retain some energy if it is to become a more-massive charm but release energy if it becomes a less-massive up, and a down quark will release energy if converted to an up but retain energy if converted to a charm.

In the same way the top quark is unique insofar as it is a visitor in the $v_{\Downarrow}$ well, the bottom quark is also unique insofar that it hugs the vertical axis so closely that its $V_{(5)} / L^5 = 0$ energy in the Lagrangian potential is raised all the way up from –16.36 GeV to -2.468 GeV, as shown in Figure 9. Additionally, the bottom quark itself has a mass of $m_b c^2 = 4.18^{+0.04}_{-0.03}$ GeV [42] which can be contributed to scale the barrier. So, it only needs the energy equivalent of two, not three Higgs bosons to help it over the barrier to become a charm or up quark. Once the bottom quark does

decay into a charm or an up quark, it relinquishes most of its energy back into the vacuum because $m_b > m_c$ and $m_b \gg m_u$.

So to summarize, not yet counting the energy also needed to raise a *W* boson to mediate the beta decay, it takes the energy equivalent of just over six Higgs bosons to facilitate a $u \to d, s, b$ or a $c \to d, s, b$ decay from the $v_{\Uparrow} \to v_{\Downarrow}$ well, it takes energy from three Higgs bosons to facilitate a $d \to u, c$ or a $s \to u, c$ decay from $v_{\Downarrow} \to v_{\Uparrow}$, and it takes energy from two Higgs bosons to facilitate a $b \to u, c$ decay from $v_{\Downarrow} \to v_{\Uparrow}$. And in all these cases, after the decay, some of the energy used to jump the barrier is either released back into the vacuum or retained by the quark, depending respectively on whether the quark has lost or gained rest mass during the decay. Additionally, $t \leftrightarrow b$ decays require no additional energy to jump the barrier because they both nest in $v_{\Downarrow}$. However, because the top quark rest energy is about 169 GeV larger than that of the bottom quark, any $b \to t$ transition such as in [50] will need to be facilitated by two Higgs bosons – not for a barrier jump, but simply for the extra rest energy. However, this still takes less energy than the 240.37 GeV height of the well barrier, which again is why $V_{tb} = 0.999105 \pm 0.000032$ is the closest to 1 of all the CKM matrix components, by a substantial margin.

Consequently, keeping in mind that all of these quark decays are occurring inside a baryon which has very large internal energies due to gluon-mediated strong interactions, the picture we obtain for beta decay is that in the vicinity of a quark about to decay, some number of Higgs bosons spontaneously arise as fluctuations in the Fermi vacuum. The quark about to decay draws the energy out of the rest masses of these Higgs bosons in order to jump the barrier and / or acquire the additional rest energy needed to change its identity into a different type of quark, and the *W* boson also acquires its rest mass of about 80 GeV. Then, once the decay is complete, the excess energy beyond what is needed for the new rest mass is released back into the vacuum. Noting that Higgs bosons are their own antiparticles, if two Higgs bosons are needed to trigger a beta decay, these can each be supplied by a $q\bar{q}$ fluctuation inside a hadron. If three Higgs are needed, these can be supplied by each quark in a *qqq* baryon. And if six Higgs are needed, each of the three quarks in a baryon can precipitate a $q\bar{q}$ fluctuation to supply a pair of Higgs bosons. *The Higgs bosons therefore operate as the mechanism to transfer energy from the vacuum into the W boson and into both the rest energies of the fermions and the into barrier jump required for beta decays of the fermions in all but* $t \leftrightarrow b$ *decays.*

It is also important to keep in mind that the $v_{\Uparrow}$ well bottoms out at a *global* minimum with a depth of about –514.89 GeV while the $v_{\Downarrow}$ well has only a *local* minimum with a depth of about – 16.36 GeV, as seen in Figure 8. So, it is both easier to get from $v_{\Downarrow} \to v_{\Uparrow}$ than the other way around*, and it is easier to stay in* $v_{\Uparrow}$ *after a* $v_{\Downarrow} \to v_{\Uparrow}$ *decay has occurred*. This suggests that isospin-up quarks are *more energetically stable* than isospin-down quarks. Given that individual quarks (or, at least longer-lived quarks in the first and second generations) are always confined in hadrons, and that baryons contain three quarks, this may be part of the explanation for why free neutrons with a mean lifetime of about fifteen minutes, decay into completely-stable free protons.

Simply: becoming and staying an up quark is energetically-favored over becoming and staying a down quark.

Also, *ab initio*, the Higgs field *h* itself represents *quantum fluctuations* in the Fermi vacuum in which the scalar field ϕ_h is recast as $\phi_h = \frac{1}{\sqrt{2}}\phi_{1h} = \frac{1}{\sqrt{2}}(v+h)$. But everything we just described about beta decay entails Higgs bosons spontaneously arising in the Fermi vacuum while drawing energy out of the vacuum for their rest energies, transferring these energies to a fermion so it can jump the barrier and / or have the energy needed for its new masses in its new identity, further transferring energy into a *W* boson to mediate the transition, and the fermion and boson finally releasing and depositing any excess energy back into the vacuum. But *these ongoing draws and deposits of energy from and back into the vacuum energy bank are simply quantum fluctuations by another name*. Consequently, every time there is a beta decay event, it is accompanied by quantum fluctuations in which there is a quick withdrawal of energy from the vacuum, followed by a quick redeposit of energy into the vacuum, with the energy magnitudes of these withdrawals and deposits set by the depth of the two wells, the height of the well barrier, and the rest masses of the Higgs and *W* bosons and the involved fermions.

Experimentally, it would be highly desirable to closely observe various beta decay transitions associated with all nine components of the CKM matrix, in both directions, with a sharp focus on energy fluctuations in the vacuum. For $v_{\Downarrow} \to v_{\Uparrow}$ decays, it may be possible to detect a smaller energy withdrawal followed by larger redeposit. For $v_{\Uparrow} \to v_{\Downarrow}$ decays, it may be possible to detect a larger withdrawal followed by a smaller redeposit. And for the uniquely-situated $b \leftrightarrow t$ transitions that do not require jumping the well barrier and have the closest-to1 $V_{tb} = 0.999105 \pm 0.000032$, $b \to t$ is simply a withdrawal and $t \to b$ is simply a deposit. So, ironically, $b \leftrightarrow t$ decays between the most-massive quarks involve smaller energies than all other decays because the requisite energies are determined solely by the mass difference between these two quarks and their heights in the $v_{\Downarrow}$ well and not by the larger magnitude of the well barrier height. So, it may be possible to detect that there are *smaller* energy fluctuations in $b \leftrightarrow t$ than in any other type of beta decay event between quarks.

Finally, to be clear, although all forms of beta-decay are mediated by weak *W* bosons, the foregoing discussion applies only to beta decays of quarks, not to leptonic beta decays involving charged lepton and neutrinos. As we shall see in the upcoming development, leptonic beta decays have further unique characteristics stemming from the close-to-but-not-quite-zero masses of the neutrinos, which, also somewhat ironically, involve extremely high-energy barriers stemming from the large ratio of the charged-lepton-to-neutrino masses.

Before concluding, let's take stock of all the reparameterizations we have found to this point. Following (16.3) we noted that we had reparameterized the six quark masses as $m_u, m_c, m_t = F(v, \theta_{C31}, \theta_{C23})$ and $m_d, m_s, m_b = F(m_d, m_h, \theta_{C21})$ with $v_{\Uparrow} = v$, leaving only m_d unconnected to some other known observed empirical energy or mixing angle. But at (15.7) we also made use of the relation $3(m_d - m_u)/(2\pi)^{1.5} = m_e$ separately discovered by the author in 2013 [47], [48]. So, having reparameterized the up mass in $m_u, m_c, m_t = F(v, \theta_{C31}, \theta_{C23})$, and knowing

the electron mass, this 2013 relation allows us to reparameterize $m_d = F\left(m_e\right)$, that is, to reparameterize the down mass as a function of the electron mass. Therefore, we effectively used

$$m_u, m_c, m_t, m_d, m_s, m_b = F\left(v, \theta_{C31}, \theta_{C23}, \theta_{C21}, m_h, m_e\right) \tag{17.1}$$

to reparameterize *all six quark masses*. But one of these parameters, m_e, is itself a rest mass, but of a lepton not a quark. So, we have effectively "kicked down the road" to our study of the charged leptons, the completion of quark mass reparameterization. So, we now turn to lepton rest masses.

18. Theory of Fermion Rest Masses and Mixing: Electron, Mu and Tau Charged Leptons

Having studied the quark masses and their mixing and beta decay mechanisms in relation to Higgs fields and bosons, we now turn to the leptons. Just like quarks, it is well known that leptons also mix generations during beta decays, utilizing the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix which has an identical mathematical structure to the CKM quark mixing matrix. The existence of a PMNS matrix with non-zero off-diagonal elements provides the central empirical indication that neutrinos are not massless as was considered possible a generation ago, but have an extremely small rest energy on the order of a fraction of a single electron volt (eV). This is also borne out by cosmological observations of a slight but definite time delay between the arrivals of photons and neutrinos from supernova events following a transit times of more than 100,000 years, such as described in [51]. However, direct observations as to what the masses of these neutrinos actually are, or at least as to the mass *ratios* of the various neutrino types (electron, mu or tau partner), are still wanting as of the present day. What has been established directly, are upper limits on these neutrino masses, on the order of less than a single electron volt. By way of comparison, the electron, which is the lightest charged lepton, has a rest energy of just over half a million electron volts (MeV). In the discussion following, we shall utilize the PMNS matrix and related leptonic mixing angles laid out in the most recent NuFIT data at [52].

Because the leptons are known to parallel the quarks insofar as they are both elementary fermions and have identical weak isospin structures, we shall begin by seeing whether the results for sections 14 through 16 for the quarks can be carried over in identical form to the leptons, with the only difference being the numeric values of the various mixing angles and fermion masses. However, now that everything that was developed for quarks will be replicated for leptons, let us make some notational choices which will help avoid confusion as between quark parameters and similar lepton parameters. First, starting at (14.8), we began to utilize three quark mass mixing angles denoted θ_{21}, θ_{32}, θ_{31} which were later connected at (14.12) and (15.6) to the three real CKM quark mixing angles denoted θ_{C12}, θ_{C23} and θ_{C23}. Here, for leptons, we shall postulate three analogous mass mixing angles denoted ϑ_{21}, ϑ_{32} and ϑ_{31}, and will seek out a connection to the three real PMNS angles which we shall denote by θ_{P12}, θ_{P23} and θ_{P23}. Second, for the quarks, we found as crystallized at (15.10) that there are two minima for the vacuum which play a central role, namely, the well-known vev $v_{\Uparrow} = v \cong 246.22 \text{ GeV}$ established by the Fermi constant and a second $v_{\Downarrow} \cong 6.05 \text{ GeV}$. Importantly, each was shown to relate by $\frac{1}{\sqrt{2}} v_{\Uparrow} = m_u c^2 + m_c c^2 + m_t c^2$ and

$\frac{1}{\sqrt{2}}v_{\Downarrow} = m_d c^2 + m_s c^2 + m_b c^2$ to a sum of quark masses. For leptons, for notational distinctness, we shall use *u* rather than *v* to denote any similar vacuums. Now we begin the calculations.

As at (14.7) we postulate a 3x3 charged lepton mass matrix $M_{e\mu\tau}c^2$ with all energy concentrated in $\frac{1}{\sqrt{2}}u_{\Downarrow}$ for the upper-left component, with $u_{\Downarrow}$ denoting a vev for isospin-down leptons, that is, the electron and the mu and tau leptons. At the moment, the magnitude of $u_{\Downarrow}$ is yet to be determined. Then, as at (14.8) we perform a bi-unitary transformation $M_{e\mu\tau}c^2 \rightarrow M'_{e\mu\tau}c^2 = U^{\dagger}M_{e\mu\tau}c^2U$ on $M_{e\mu\tau}c^2$ using both the type *I* "downward cascade" parameterization and the type *II* "distribution" parameterization. As a result, we arrive at relations analogous to those contained in (14.8):

$$\begin{aligned} m_\tau c^2 &= \tfrac{1}{\sqrt{2}}u_{\Downarrow}\,\mathrm{c}_{I\Downarrow 32}{}^2 \\ m_\mu c^2 &= \tfrac{1}{\sqrt{2}}u_{\Downarrow}\,\mathrm{c}_{I\Downarrow 21}{}^2\,\mathrm{s}_{I\Downarrow 32}{}^2 \\ m_e c^2 &= \tfrac{1}{\sqrt{2}}u_{\Downarrow}\,\mathrm{s}_{I\Downarrow 21}{}^2\,\mathrm{s}_{I\Downarrow 32}{}^2 \end{aligned} \quad , \tag{18.1a}$$

$$\begin{aligned} m_\tau c^2 &= \tfrac{1}{\sqrt{2}}u_{\Downarrow}\,\mathrm{c}_{II\Downarrow 31}{}^2\,\mathrm{c}_{II\Downarrow 32}{}^2 \\ m_\mu c^2 &= \tfrac{1}{\sqrt{2}}u_{\Downarrow}\,\mathrm{s}_{II\Downarrow 32}{}^2 \\ m_e c^2 &= \tfrac{1}{\sqrt{2}}u_{\Downarrow}\,\mathrm{s}_{II\Downarrow 31}{}^2\,\mathrm{c}_{II\Downarrow 32}{}^2 \end{aligned} \quad . \tag{18.1b}$$

These sines and cosines are associated with the leptonic mass mixing angles ϑ_{21}, ϑ_{32} and ϑ_{31}

Next, we define a relation amongst each of the lepton masses m_l, associated dimensionless couplings G_l and the vev $u_{\Downarrow}$ in the customary form as follows:

$$m_l c^2 \equiv \tfrac{1}{\sqrt{2}}G_l u_{\Downarrow} \, . \tag{18.2}$$

Using these in (18.1) then yields:

$$\begin{aligned} G_\tau &= \mathrm{c}_{I\Downarrow 32}{}^2 \\ G_\mu &= \mathrm{s}_{I\Downarrow 32}{}^2\,\mathrm{c}_{I\Downarrow 21}{}^2 \\ G_e &= \mathrm{s}_{I\Downarrow 32}{}^2\,\mathrm{s}_{I\Downarrow 21}{}^2 \end{aligned} \quad , \tag{18.3a}$$

$$\begin{aligned} G_\tau &= \mathrm{c}_{II\Downarrow 32}{}^2\,\mathrm{c}_{II\Downarrow 31}{}^2 \\ G_\mu &= \mathrm{s}_{II\Downarrow 32}{}^2 \\ G_e &= \mathrm{c}_{II\Downarrow 32}{}^2\,\mathrm{s}_{II\Downarrow 31}{}^2 \end{aligned} \quad . \tag{18.3b}$$

From either (18.3a) or (18.3b), we use the trigonometric identity $c^2 + s^2 = 1$ to find that:

$$G_\tau + G_\mu + G_e = 1 . \tag{18.4}$$

Then, using (18.2) in (18.4) we find that:

$$\tfrac{1}{\sqrt{2}} u_\Downarrow = m_\tau c^2 + m_\mu c^2 + m_e c^2 = 1883.029 \pm 0.120 \text{ MeV} . \tag{18.5}$$

These are identical in form with analogous relations (15.10) earlier found for the quarks. The numeric value of this vev is computed to three decimals using empirical data from [43], namely:

$$\begin{array}{l} m_e c^2 = 0.5109989461 \pm 0.0000000031 \text{ MeV}; \\ m_\mu c^2 = 105.6583745 \pm 0.0000024 \text{ MeV}; \quad m_\tau c^2 = 1776.86 \pm 0.12 \text{ MeV} \end{array} . \tag{18.6}$$

Next, we restructure (18.3) to isolate sines and cosines, then use (18.4) to obtain:

$$\begin{array}{l} \mathrm{c}_{I\Downarrow 32}{}^2 = G_\tau \\ \mathrm{c}_{I\Downarrow 21}{}^2 = \dfrac{G_\mu}{\mathrm{s}_{I\Downarrow 32}{}^2} = \dfrac{G_\mu}{1-\mathrm{c}_{I\Downarrow 32}{}^2} = \dfrac{G_\mu}{1-G_\tau} = \dfrac{G_\mu}{G_\mu + G_e} , \\ \mathrm{s}_{I\Downarrow 21}{}^2 \mathrm{s}_{I\Downarrow 32}{}^2 = G_e \end{array} \tag{18.7a}$$

$$\begin{array}{l} \mathrm{c}_{II\Downarrow 31}{}^2 = \dfrac{G_\tau}{\mathrm{c}_{II\Downarrow 32}{}^2} = \dfrac{G_\tau}{1-\mathrm{s}_{II\Downarrow 32}{}^2} = \dfrac{G_\tau}{1-G_\mu} = \dfrac{G_\tau}{G_\tau + G_e} \\ \mathrm{s}_{II\Downarrow 32}{}^2 = G_\mu \\ \mathrm{s}_{II\Downarrow 31}{}^2 \mathrm{c}_{II\Downarrow 32}{}^2 = G_e \end{array} . \tag{18.7b}$$

Finally, we use (18.3) in (18.7) and combine with (18.4) and (18.5) to obtain:

$$\begin{array}{l} \mathrm{c}_{I\Downarrow 32}{}^2 = G_\tau = \dfrac{m_\tau c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \dfrac{m_\tau c^2}{\frac{1}{\sqrt{2}} u_\Downarrow} \\ \mathrm{c}_{I\Downarrow 21}{}^2 = \dfrac{G_\mu}{\mathrm{s}_{I\Downarrow 32}{}^2} = \dfrac{G_\mu}{G_\mu + G_e} = \dfrac{m_\mu c^2}{m_\mu c^2 + m_e c^2} = \dfrac{m_\mu c^2}{\left(m_\tau c^2 + m_\mu c^2 + m_e c^2\right) - m_\tau c^2} = \dfrac{m_\mu c^2}{\frac{1}{\sqrt{2}} u_\Downarrow - m_\tau c^2} , \\ \mathrm{s}_{I\Downarrow 21}{}^2 \mathrm{s}_{I\Downarrow 32}{}^2 = G_e = \dfrac{m_e c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \dfrac{m_e c^2}{\frac{1}{\sqrt{2}} u_\Downarrow} \end{array} \tag{18.8a}$$

$$\begin{aligned}
\mathrm{c}_{II\Downarrow 31}{}^2 &= \frac{G_\tau}{\mathrm{c}_{II\Downarrow 32}{}^2} = \frac{G_\tau}{G_\tau + G_e} = \frac{m_\tau c^2}{m_\tau c^2 + m_e c^2} = \frac{m_\tau c^2}{\left(m_\mu c^2 + m_\tau c^2 + m_e c^2\right) - m_\mu c^2} = \frac{m_\tau c^2}{\frac{1}{\sqrt{2}}u_\Downarrow - m_\mu c^2} \\
\mathrm{s}_{II\Downarrow 32}{}^2 &= G_\mu = \frac{m_\mu c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \frac{m_\mu c^2}{\frac{1}{\sqrt{2}}u_\Downarrow} \\
\mathrm{s}_{II\Downarrow 31}{}^2\,\mathrm{c}_{II\Downarrow 32}{}^2 &= G_e = \frac{m_e c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \frac{m_e c^2}{\frac{1}{\sqrt{2}}u_\Downarrow}
\end{aligned} \quad . \tag{18.8b}$$

Proceeding from here, we use the mass data in (18.6) and the sum in (18.5) together with the relations for $\mathrm{c}_{I\Downarrow 32}{}^2$, $\mathrm{c}_{I\Downarrow 21}{}^2$, $\mathrm{c}_{II\Downarrow 31}{}^2$ and $\mathrm{s}_{II\Downarrow 32}{}^2$ to calculate that:

$$\begin{aligned}
\vartheta_{I\Downarrow 32} &= 0.23974 \pm 0.00001 \text{ rad} = 13.73605 \pm 0.00045\,^\circ \\
\vartheta_{II\Downarrow 32} &= 0.23915 \pm 0.00001 \text{ rad} = 13.70231 \pm 0.00045\,^\circ \\
\vartheta_{I\Downarrow 21} &= 0.06943 \pm 0 \text{ rad} = 3.97816 \pm 0\,^\circ \\
\vartheta_{II\Downarrow 31} &= 0.01696 \pm 0 \text{ rad} = 0.97155 \pm 0.00003\,^\circ
\end{aligned} \quad . \tag{18.9}$$

Then we are ready to compare this to the empirical data for the PMNS mixing angles.

The data in [52] lays out a best fit at both a 1σ and 3σ range. These spreads will become important momentarily. Therefore, without having more specific data we also *estimate* the 2σ spread by taking the average of the 1σ and 3σ spreads. We then show the central observed value followed by successive ranges also shown for each of 1σ, the estimated 2σ as just mentioned, and 3σ, respectively. Presented in this way, the four PMNS parameters, in degrees, are:

$$\begin{aligned}
\theta_{P12} &= 33.62^{+0.78\ +1.605\ +2.43}_{-0.76\ -1.48\ -2.2}\,^\circ \\
\theta_{P13} &= 8.549^{+0.15\ +0.295\ +0.44}_{-0.15\ -0.3\ -0.45}\,^\circ \\
\theta_{P23} &= 47.2^{+1.9\ +3.1\ +4.3}_{-3.9\ -5.4\ -6.9}\,^\circ \\
\delta_P &= 234^{+43\ +91.5\ +140}_{-31\ -60.5\ -90}\,^\circ
\end{aligned} \quad . \tag{18.10}$$

Based on what we saw for the quarks, it is $\vartheta_{I\Downarrow 21} = 3.97816^\circ$ and $\vartheta_{II\Downarrow 31} = 0.97155^\circ$ for which we would anticipate a match. But comparing with (18.5) there is nothing close. So at least one of the suppositions we used to obtain a correct data match for the quarks, does not apply to the leptons.

Taking a close look at final term in each of the six relations (18.8) and referring to (18.5), we see that each numerator contains a specific lepton mass, while each denominator contains the sum $\frac{1}{\sqrt{2}}u_\Downarrow = m_\tau c^2 + m_\mu c^2 + m_e c^2 = 1883.029 \pm 0.120$ MeV. Because $u_\Downarrow$ is what we are postulating is a vev for the charged, isospin-down leptons, and because the angles deduced in (18.9) do not come anywhere near the empirical data in (18.10), we conclude that this postulate – although its analogue worked for the quarks – is incorrect for leptons. In other words, we conclude based on the failure to obtain an empirical match that $u_\Downarrow$ as specified in (18.5) is in fact *not* the correct vev

to be using when it comes to the charged leptons. So if $u_{\Downarrow}$ is not the correct vev, the question now becomes, what is the correct vev? More precisely there are two questions: First, denoting an energy difference by $\delta_{\Downarrow}$, is there some other vev denoted $u'_{\Downarrow}$ and *defined* such that:

$$\tfrac{1}{\sqrt{2}}u'_{\Downarrow} \equiv \tfrac{1}{\sqrt{2}}u_{\Downarrow} + \delta_{\Downarrow} = m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_{\Downarrow} = 1883.029 \pm 0.120\ \text{MeV} + \delta_{\Downarrow}\ \text{MeV}\,, \tag{18.11}$$

which *does* allow at least one of $\vartheta_{I\Downarrow 21}$ or $\vartheta_{II\Downarrow 31}$ to fit the empirical data in (18.10), and even better, which allows *both* of these to fit the data? Second, if there does exist some $\frac{1}{\sqrt{2}}u'_{\Downarrow}$ which fits the data, this would initially be an independent, unexplained energy number not based solely on the separately-known data $m_\tau c^2 + m_\mu c^2 + m_e c^2$, but rather on $m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_{\Downarrow}$. Therefore, can this new $\frac{1}{\sqrt{2}}u'_{\Downarrow}$ be connected to *other known data of independent origins*, for example, the Fermi vev once again, so that we will not have added any new unexplained data?

Because the angles of interest are $\vartheta_{I\Downarrow 21}$ and $\vartheta_{II\Downarrow 31}$, let us use these angles as shown in (18.8), but base them on $u'_{\Downarrow}$ defined in (18.11) rather than on $u_{\Downarrow}$, by defining two new angles $\vartheta'_{I\Downarrow 21}$ and $\vartheta'_{II\Downarrow 31}$ according to:

$$\begin{aligned}
\cos^2\vartheta'_{I\Downarrow 21} &\equiv \frac{m_\mu c^2}{\frac{1}{\sqrt{2}}u'_{\Downarrow} - m_\tau c^2} = \frac{m_\mu c^2}{\frac{1}{\sqrt{2}}u_{\Downarrow} + \delta_{\Downarrow} - m_\tau c^2} = \frac{m_\mu c^2}{m_\mu c^2 + m_e c^2 + \delta_{\Downarrow}} \\
\cos^2\vartheta'_{II\Downarrow 31} &\equiv \frac{m_\tau c^2}{\frac{1}{\sqrt{2}}u'_{\Downarrow} - m_\mu c^2} = \frac{m_\tau c^2}{\frac{1}{\sqrt{2}}u_{\Downarrow} + \delta_{\Downarrow} - m_\mu c^2} = \frac{m_\tau c^2}{m_\tau c^2 + m_e c^2 + \delta_{\Downarrow}}
\end{aligned}\,. \tag{18.12}$$

Then, we simply use the known mass data in (18.5) and (18.6), and sample various values for $\delta_{\Downarrow}$ using a spreadsheet or the like, until the values deduced for $\vartheta_{I\Downarrow 21}$ or $\vartheta_{II\Downarrow 31}$ appear to bear a statistically-meaningful relation to the empirical data in (18.10).

Because error-bars are important in this calculation, let's us briefly comment on how we will approach these. The $u_{\Downarrow}$ in (18.12) is related to the sum of the three charged lepton masses. Because the error spread for each of the masses is independent of the other two, there are 3x3x3=27 different ways of calculating $u_{\Downarrow}$ for each individual lepton being high, medium or low on its error spread. But the muon mass is known about 50,000 times as precisely as the tau mass, and the electron mass is known just shy of 40 million times as tightly as the tau mass. Therefore, to keep maters simple, we regard the electron and muon masses to be precisely at the center of their error spreads, and use the $\pm 0.12\ \text{MeV}$ spread in the tau mass as the basis for calculating the spread in $u_{\Downarrow}$. This is why there is a $\pm 0.120\ \text{MeV}$ spread shown in (18.11), and also in (18.5), with one decimal place added.

Working from (18.12) and sampling various $\delta_{\Downarrow}$, we find that when we set $\delta_{\Downarrow} = 39.642$ MeV thus $\frac{1}{\sqrt{2}}u'_{\Downarrow} = 1922.671 \pm 0.120$ MeV, we are able to obtain $\vartheta'_{II\Downarrow 31} = 8.5490 \pm 0.0003°$, with a center conforming precisely with the center of the empirical $\theta_{P13} = 8.549^{+0.15\ +0.295\ +0.44}_{-0.15\ -0.3\ -0.45}\,°$ in (18.10). Simultaneously, with this same $\delta_{\Downarrow} = 39.642$ MeV we are able to obtain $\vartheta'_{I\Downarrow 21} = 31.65230 \pm 0°$. The empirical data in (18.10) tells us that $\theta_{P12} = 33.62^{+0.78\ +1.605\ +2.43}_{-0.76\ -1.48\ -2.2}\,°$. Given that the 3σ error permits an angle as low as $\theta_{P12} = 31.42°$, we conclude that $\delta_{\Downarrow} = 39.642$ MeV matches θ_{P13} right at the center, and comes in at about 2.8σ on the low end of θ_{P12}. This is very important, because this means that in fact we are able to *simultaneously* match $\vartheta'_{II\Downarrow 31} \leftrightarrow \theta_{P13}$ and $\vartheta'_{I\Downarrow 21} \leftrightarrow \theta_{P12}$ within 3σ error bars for *both* items of data, and more closely if we move $\vartheta'_{II\Downarrow 31}$ upward somewhat from its center value.

For a second sample, we find that when we set $\delta_{\Downarrow} = 46.199$ MeV thus $\frac{1}{\sqrt{2}}u'_{\Downarrow} = 1929.229 \pm 0.120$ MeV, we are able to obtain $\vartheta'_{I\Downarrow 21} = 33.62 \pm 0°$, conforming precisely with the center of the empirical $\theta_{P12} = 33.62^{+0.78\ +1.605\ +2.43}_{-0.76\ -1.48\ -2.2}\,°$ in (18.10). Simultaneously, with this same $\delta_{\Downarrow} = 46.199$ MeV we obtain $\vartheta'_{II\Downarrow 31} = 9.2096 \pm 0.0003°$. The 3σ data puts the corresponding angle at $\theta_{P13} = 8.989°$ on the high side, so this value for $\delta_{\Downarrow}$ puts us above 3σ data. But now we have a basis for interpolating between these two samples.

Because the first $\delta_{\Downarrow}$ sample gave us the center of θ_{P13} but produced a low value for θ_{P12}, while the second sample gave us the center of θ_{P12} but produced a high value for θ_{P13}, it appears as if the *actual* θ_{P12} is below the center and the *actual* θ_{P13} is above the center of what is shown in (18.10). So, for a third sample we take the following approach: Find a $\delta_{\Downarrow}$ which places the θ_{P12} match below center and simultaneously places the θ_{P13} match above center *by exactly the same statistical spread*. That is, find some δ for which $x\sigma(\theta_{P13}) = x\sigma(\theta_{P12})$ above and below respectively, with $x\sigma < 3\sigma$ and preferably with $x\sigma < 2\sigma$.

In accordance with this prescription, it turns out that when we set $\delta_{\Downarrow} = 42.018$ MeV thus $\frac{1}{\sqrt{2}}u'_{\Downarrow} = 1925.047 \pm 0.120$ MeV, we simultaneously obtain $\vartheta'_{II\Downarrow 31} = 8.7945 \pm 0.0003°$ versus the empirical $\theta_{P13} = 8.549^{+0.15\ +0.295\ +0.44}_{-0.15\ -0.3\ -0.45}\,°$, and $\vartheta'_{I\Downarrow 21} = 32.39 \pm 0°$ versus the empirical $\theta_{P12} = 33.62^{+0.78\ +1.605\ +2.43}_{-0.76\ -1.48\ -2.2}\,°$. Estimating linearly between center values and 3σ values, we find that $\vartheta'_{II\Downarrow 31}$ is about 1.67σ above the θ_{P13} center and $\vartheta'_{I\Downarrow 21}$ is about 1.67σ below the θ_{P12} center. Accordingly, regard this threading of the needle whereby for a lepton vev of $\frac{1}{\sqrt{2}}u'_{\Downarrow} = 1925.047 \pm 0.120$ MeV *(18.12)* is able to *simultaneously connect both* θ_{P13} *and* θ_{P12} *within*

about 1.67σ *of their respective experimental centers*, as a physically meaningful relation. Consequently, based on this connection to the experimental date, we now establish:

$$\tfrac{1}{\sqrt{2}}u'_{\Downarrow} = \tfrac{1}{\sqrt{2}}u_{\Downarrow} + \delta_{\Downarrow} = m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_{\Downarrow} \equiv 1925.047 \pm 0.120 \text{ MeV} . \tag{18.13}$$

Moreover, via (18.12) we also now use $\vartheta'_{II\Downarrow 31} = 8.795°$ and $\vartheta'_{I\Downarrow 21} = 32.39°$ to establish both a connection with, and new center value for, the lepton mixing angles. We represent this, with the empirical sigma spreads left untouched except as they are adjusted from these new centers, as:

$$\boxed{\begin{array}{l} \theta_{P12} \equiv \vartheta'_{I\Downarrow 21} = 32.39^{\,+2.01\ \ +2.83\ \ +3.66}_{\,+0.47\ \ -0.25\ \ -0.97}\ ^\circ \\ \theta_{P13} \equiv \vartheta'_{II\Downarrow 31} = 8.795^{\,-0.096\ \ +0.050\ \ +0.195}_{\,-0.396\ \ -0.546\ \ -0.696}\ ^\circ \end{array}} . \tag{18.14}$$

This is another way of showing that each of these is about 1.67σ away from their previous centers, with the former moved up thus leaving a larger downside range, and the latter moved down thus leaving a larger upside range.

With these results, we answer the first of the two questions posed at (18.11): Yes, the vev in (18.3) does allow *both* of $\vartheta_{I\Downarrow 21}$ *and* $\vartheta_{II\Downarrow 31}$ to fit the empirical data in (18.10), within about 1.67σ for each, as precisely shown in (18.14). But now we have a seemingly-disconnected vev in (18.13), and this brings us to the second question whether this can be connected to other known data of independent origins. Because $\tfrac{1}{\sqrt{2}}u'_{\Downarrow} \equiv 1925.047 \pm 0.120$ MeV in (18.13) no longer is set by $m_\tau c^2 + m_\mu c^2 + m_e c^2$ since it differs from this by $\delta_{\Downarrow} = 42.018$ MeV, the most obvious energy of comparison for (18.13) is the Fermi vev $v_{\Uparrow} = v = 246.2196508 \pm 0.0000633$ GeV given in (15.10). So, we simply calculate the ratio of these, and find that:

$$v_{\Uparrow} \,/\, \tfrac{1}{\sqrt{2}}u'_{\Downarrow} = 127.9032 \pm 0.0080 . \tag{18.15}$$

This numerical result is extremely pregnant, because it is well known that "at $Q^2 \approx M_W^{\,2}$ the value [of the electromagnetic running coupling α] is ~1/128," see note † in PDG's [21]. The closeness of (18.15) to this other empirical data raises the question whether $\tfrac{1}{\sqrt{2}}u'_{\Downarrow} = v_{\Uparrow}\alpha\left(Q^2 = M_W{}^2 c^4\right)$ may be another relationship of genuine physical meaning. So, let us review the evidence in support:

First, the angles (18.14) originating in (18.10) are distinctively related to weak interaction beta decays between the electron and the mu and tau leptons, and their respective neutrino partners, and the mixing (so-called neutrino oscillations) which crosses from one generation into another. Second, while electroweak interactions are mediated by both neutral-current *Z* bosons and charged $W^\pm$ bosons, it is the latter, with a rest energy of $M_W c^2 = 80.379(12)$ GeV (again see [21]), which is the sole mediator of these weak interaction beta decays. Third, the *e*, μ and τ leptons are the quintessential units of charge for the which interaction strength is set by $\alpha\left(Q^2 = 0\right) = 1/137.035999139(31)$ in the low energy (fine structure constant) limit, and in general

by the running $\alpha\left(Q^2\right)$. Fourth, because the *e*, μ and τ leptons and the $W^{\pm}$ boson both carry electric charge, $\alpha\left(Q^2\right)$ is in fact distinctly relevant to the strength of the electromagnetic interaction which occurs during the beta decay. Fifth, given that these beta decays are all mediated by a $W^{\pm}$ which has a rest energy $M_W c^2$, the pertinent energy scale at the decay event is not $Q^2 = 0$ but rather $Q^2 = {M_W}^2 c^4$, and so the pertinent electromagnetic coupling is $\alpha\left({M_W}^2 c^4\right) \sim 1/128$. Consequently, the unanticipated appearance of the number 127.9032 in (18.15) does not look to be a simple coincidental appearance of some other number that happens to be close to 128. Rather, this supports the conclusion that this is in fact, yet another physically-meaningful connection.

Therefore, we now connect these two numbers, and conclude that the vev which is pertinent to leptons is in fact given by:

$$\boxed{\tfrac{1}{\sqrt{2}} u'_{\Downarrow} = m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_{\Downarrow} = \frac{1}{127.9032 \pm 0.0080} v_{\Uparrow} = 1925.047 \pm 0.120 \text{ MeV} \equiv \alpha\left({M_W}^2\right) v_{\Uparrow}} \quad (18.16)$$

In the process, we tighten our knowledge of $\alpha\left({M_W}^2\right)$ to $\alpha\left({M_W}^2\right) = 1/\left(127.9032 \pm 0.0080\right)$. This result has the extremely beneficial consequence of being able to express $\delta_{\Downarrow}$ directly from the sum $m_\tau c^2 + m_\mu c^2 + m_e c^2$ and Fermi vev and $\alpha\left({M_W}^2\right)$. Referring to (18.11), this means that:

$$\delta_{\Downarrow} = \tfrac{1}{\sqrt{2}} u'_{\Downarrow} - m_\tau c^2 - m_\mu c^2 - m_e c^2 = \alpha\left({M_W}^2\right) v_{\Uparrow} - m_\tau c^2 - m_\mu c^2 - m_e c^2 = 42.018 \text{ MeV}. \quad (18.17)$$

Here, $\delta_{\Downarrow}$ no longer needs to be expressed as the energy difference which allows each of $\vartheta_{I \Downarrow 21}$ and $\vartheta_{II \Downarrow 31}$ to fit the PMNS data in (18.10). Rather, to answer the second question posed at (5.11): No, this vev difference $\delta_{\Downarrow}$ does *not* add any new unexplained data, because it is entirely specified by the other known data in (18.17), namely, the charged lepton mass sum $m_\tau + m_\mu + m_e$, the Fermi vev, and the running $\alpha\left({M_W}^2\right)$ which is the strength of the electromagnetic interaction at the lepton-to-$W^{\pm}$ beta decay event (Feynman diagram vertex). It is also helpful to write this as:

$$m_\tau c^2 + m_\mu c^2 + m_e c^2 = \alpha\left({M_W}^2\right) v_{\Uparrow} - \delta_{\Downarrow}, \quad (18.18)$$

wherein the mass sum $m_\tau + m_\mu + m_e$ is seen to be a function of the independently-known parameters $\alpha\left({M_W}^2\right)$ and $v_{\Uparrow}$, but also of $\delta_{\Downarrow}$ about which we do not yet have independent knowledge. As we shall see in the next section, $\delta_{\Downarrow}$ is in fact directly driven by the neutrino masses and – of all things – the Newton gravitational constant.

Recalling the importance of the square roots of the various rest energies and vev energies reviewed in Figures 3 through 5, we see that (18.17) lends itself to a geometric representation in

the manner of Figure 5, with $\sqrt{\tfrac{1}{\sqrt{2}}u'_{\Downarrow}} = \sqrt{\alpha}\left(M_W{}^2\right)\sqrt{v_{\Uparrow}} = \sqrt{m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_{\Downarrow}}$ on the hypotenuse, and with $\sqrt{\delta_{\Downarrow}}$ and $\sqrt{\tfrac{1}{\sqrt{2}}u_{\Downarrow}} = \sqrt{m_\tau c^2 + m_\mu c^2 + m_e c^2}$ on each of the legs. Because $\alpha = k_e e^2 / \hbar c$ where k_e is Coulomb's constant and e is the charge strength of a single charge quantum (such as a charged lepton and such as the $W^{\pm}$ bosons which mediate the beta decay, $\sqrt{\alpha}\left(M_W{}^2\right) = \sqrt{k_e / \hbar c}\, e\left(M_W{}^2\right)$ is a direct measure of the electric charge strength at the beta decay vertex. Based on the numeric values from (18.13), (18.17) and (18.5), the small angle which we refer to as the charged lepton rotation angle and denote as θ_l, has a value of $\theta_l = 8.496°$. This may be illustrated as shown below:

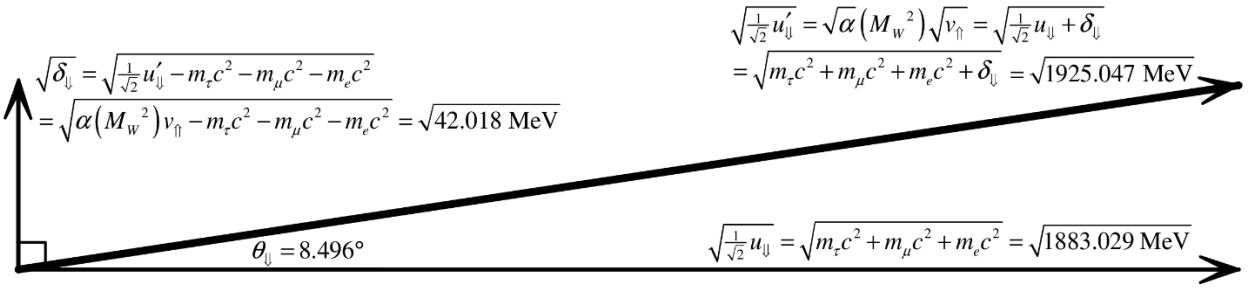

Figure 10: Projection of the Lepton vev onto the Lepton Mass Sum

Viewed in this light, the energy difference taken in its square root form $\sqrt{\delta_{\Downarrow}}$ rotates the $\sqrt{\tfrac{1}{\sqrt{2}}u_{\Downarrow}} = \sqrt{m_\tau c^2 + m_\mu c^2 + m_e c^2}$ vector which is purely a function of the charged lepton masses, through an angle $\theta_{\Downarrow} = 8.496°$, into $\sqrt{\tfrac{1}{\sqrt{2}}u'_{\Downarrow}} = \sqrt{m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_{\Downarrow}}$ which is a function of the charged lepton masses as well as $\delta_{\Downarrow}$. While it also happens that $\tfrac{1}{\sqrt{2}}u'_{\Downarrow} = \alpha\left(M_W{}^2\right)v_{\Uparrow}$ from (18.16), again, it will be important to acquire independent knowledge about $\delta_{\Downarrow}$.

It is also very helpful to obtain mass relationships analogous to (14.14) and (15.14) which directly relate the charge lepton masses particularly to the two angles in (18.14). Solving the simultaneous equations which are (18.12), then using (18.14), for the tau and mu leptons we obtain:

$$m_\tau c^2 = \frac{1}{\sqrt{2}}u'_{\Downarrow}\frac{\cos^2\theta_{P13}\sin^2\theta_{P12}}{1-\cos^2\theta_{P13}\cos^2\theta_{P12}};\quad m_\mu c^2 = \frac{1}{\sqrt{2}}u'_{\Downarrow}\frac{\sin^2\theta_{P13}\cos^2\theta_{P12}}{1-\cos^2\theta_{P13}\cos^2\theta_{P12}}. \tag{18.19a}$$

But because of the rotation (18.17) illustrated in Figure 10, the electron mass is *not* a direct function of thee angles. For this mass, we need to use (18.3) and (18.2), then use (18.19a), to deduce:

$$\begin{aligned} m_e c^2 &= m_\tau c^2 \tan^2\vartheta_{II\Downarrow 31} = m_\mu c^2 \tan^2\vartheta_{I\Downarrow 21} \\ &= \frac{1}{\sqrt{2}}u'_{\Downarrow}\frac{\cos^2\theta_{P13}\sin^2\theta_{P12}}{1-\cos^2\theta_{P13}\cos^2\theta_{P12}}\tan^2\vartheta_{II\Downarrow 31} = \frac{1}{\sqrt{2}}u'_{\Downarrow}\frac{\sin^2\theta_{P13}\cos^2\theta_{P12}}{1-\cos^2\theta_{P13}\cos^2\theta_{P12}}\tan^2\vartheta_{I\Downarrow 21}. \end{aligned} \tag{18.19b}$$

W noted at the end of section 16 that one of the parameters used to reparameterize the quark masses, m_e, is effectively "kicked down the road" to our study of the charged leptons, which is now just completed. So, let us review what we now know about the electron mass.

In this section, we started with the three lepton masses m_τ, m_μ, m_e. The latter, m_e, had been "kicked down the road" from the quark mass study. To connect these with the PMNS angles we were required at (18.11) to postulate a fourth, entirely-new energy number $\delta_\Downarrow$ to be added to the sum of the charged lepton rest energies. But we "recover" this when we find at (18.16) that this sum $m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_\Downarrow = \alpha\left(M_W{}^2\right) v_\Uparrow$ can be related within experimental errors to the Fermi vev $v_\Uparrow \equiv v$ by the strength $\alpha\left(M_W{}^2\right) = 1/\left(127.9032 \pm 0.0080\right)$ of the electromagnetic running coupling at a probe energy $Q^2 = M_W^2 c^4$. And, we note the clear relevance of this coupling strength to beta decays between charged leptons and neutrinos, because these must always be mediated by charged $W^\pm$ bosons and so will always have an $M_W c^2$ present at the interaction vertex of the decay to provide an elevated Q^2. Thus, the reparameterization of this section is:

$$m_\tau, m_\mu, m_e \mapsto m_\tau, m_\mu, m_e, \delta_\Downarrow = F\left(\theta_{P12}, \theta_{P13}, \alpha\left(M_W{}^2\right), \delta_\Downarrow\right). \tag{18.20}$$

In this way, we have now reparameterized all three charged lepton masses m_τ, m_μ, m_e over to $\theta_{P12}, \theta_{P13}, \alpha\left(M_W{}^2\right)$, *but only by adding a new energy* $\delta_\Downarrow$. Taken together with (17.1) for the quark masses, and seeing in (18.20) how m_e is now included in the charged lepton mass reparameterization, all told we have now reparameterized:

$$\left\{m_u, m_c, m_t, m_d, m_s, m_b, m_e, m_\mu, m_\tau, \delta_\Downarrow\right\} = F\left(v, m_h, \theta_{C31}, \theta_{C23}, \theta_{C21}, \theta_{P12}, \theta_{P13}, \alpha\left(M_W{}^2\right), \delta_\Downarrow\right). \tag{18.21}$$

So at this point, the set of *nine* elementary fermion masses exclusive of neutrinos as well as the new parameter $\delta_\Downarrow$, becomes a function of the *eight* independently-known energies, angles, and couplings $v, m_h, \theta_{C31}, \theta_{C23}, \theta_{C21}, \theta_{P12}, \theta_{P13}, \alpha\left(M_W{}^2\right)$, while we "kick" our direct understanding of $\delta_\Downarrow$ "down the road" to the study of neutrinos. Specifically, what we have left to do, is to now reparameterize the set $\left\{m_{\nu e}, m_{\nu\mu}, m_{\nu\tau}, \delta_\Downarrow\right\}$ of the three neutrino masses plus the extra energy $\delta_\Downarrow$. But because the neutrino masses – unlike all the other elementary fermion masses – are not known, we will also show how, in the process of reparameterizing the neutrino masses and seeking a direct physical understanding of $\delta_\Downarrow$, *it is also possible to predict the neutrino masses with specificity*.

19. Theory of Fermion Rest Masses and Mixing: Prediction of the Neutrino Mass Sum and of the Individual Neutrino Masses

The neutrinos are unique among the elementary fermions. Not only was it believed for a long time that these were massless fermions – which was disproved by neutrino oscillations which

we are in the midst of studying here – but there remains debate to this day as to their fundamental character, that is, whether they are Dirac fermions in the same way as all other fermions, or are Majorana fermions with the distinctive property of being their own antiparticles. From a practical standpoint, there is one very striking difference which affects how we approach the question of neutrino masses: while upper limits have been established for the neutrino masses, *we have limited empirical data available to tell us what the precise neutrino masses actually are.*

Let us start on page 11 of PDG's [53], where it is stated that "determining, or obtaining significant constraints on, the absolute scale of neutrino masses remains a very significant research problem at the present time." But as noted in [54], "somewhere between 10 meV and 2eV is our playground." And on page 12 of PDG's 2018 review [53], it is reported that the sum of the neutrino masses is $\Sigma_j m_j < 0.170$ eV at a 95% confidence level. So perhaps the most striking feature of what we do know about neutrino masses, is that these masses are so immensely-small in comparison with other fermion masses. With the lightest non-neutrino fermion – the electron – having a mass of just over *half a million* eV, the largest possible mass for a neutrino is over a million times smaller than the electron mass. And the magnitude of this ratio is even greater for other fermions; for the GeV scale fermions, it is 10^9 or larger. As stated also on page 12 of [53], "it is natural to suppose that the remarkable smallness of neutrino masses is related to the existence of a new fundamental mass scale in particle physics, and thus to new physics beyond that predicted by the Standard Model." Indeed, the only natural energy ratios which come to mind as able to produce a mass scale this small, involve the Fermi vev $v = v_{\Uparrow} = 246.2196508 \pm 0.0000633$ GeV relative to $M_P c^2 = 1.220910 \times 10^{19}$ GeV, which is the Planck energy. The former of course is a proxy for the Fermi constant G_F, and the latter for the Newton gravitational constant *G*.

In this regard, when we look at (18.21) and take inventory of parameters, we see of course that the Fermi *v* is one of the parameters already used, which means that G_F has already been used. But the Newton constant *G* and its associated Planck energy $M_P c^2$ with the Planck mass defined by $GM_P{}^2 \equiv \hbar c$ is not yet used. Given the need for a very small energy ratio to bridge the chasm from other fermion masses to neutrino masses, we proceed from the viewpoint that the dimensionless ratio $v / M_P c^2 = 2.018194 \times 10^{-17}$ may provide the basis for supplying the requisite very small energy ratio. And in view of the important role that square roots of energy numbers appear to play in connecting masses to mixing angles and other parameters – for example, see the Pythagorean axes in Figures 3, 4, 5 and 10 and all the prior equations which contain energy square roots – we also consider using the ratio $\sqrt{v / M_P c^2} = 4.492431 \times 10^{-09}$. Then, we need a baseline energy against which to apply this ratio.

Now, the energy parameter $\delta_{\Downarrow} = 42.018$ MeV deduced in (18.17) to fit the charged lepton masses to two of the PMNS mixing angles is brand new. Aside from its origin as a necessity to fit this empirical data, we still have no *independent knowledge* about its direct physical meaning. In contrast, all the other parameters in (18.21) do have separate status as physical quantities with well-understood, independent meaning. So, supposing that $\delta_{\Downarrow}$ is, perhaps, the baseline energy against which to use $\sqrt{v / M_P c^2} = 4.492431 \times 10^{-09}$, we simply do the exploratory calculation:

$$\delta_{\Downarrow}\sqrt{\frac{v}{M_P c^2}} = 42.018 \text{ MeV} \times 4.492431 \times 10^{-09} = 0.18876 \text{ eV} . \tag{19.1}$$

This is a bullseye! Not only is this number at the right order of magnitude to describe the neutrino mass sum based on the knowledge we have to date of these masses, but within the correct order of magnitude, it is at the correct ~2 eV upper limit which empirical data has placed on this sum. It seems highly unlikely that arriving at 0.189 eV from across nine orders of magnitude when our target energy is near .2 eV is merely a coincidence. As a result, we conclude that this is no coincidence, and regard this as a relation of true physical meaning. So now, we need to make a formal assignment of the result in (19.1) to the neutrino masses.

In (14.3), (15.10) and (18.15), the vevs in relation the respective mass sums are $v_{\Uparrow} = \sqrt{2}\left(m_u c^2 + m_c c^2 + m_t c^2\right)$, $v_{\Downarrow} = \sqrt{2}\left(m_d c^2 + m_s c^2 + m_b c^2\right)$ and $u_{\Downarrow} = \sqrt{2}\left(m_\tau c^2 + m_\mu c^2 + m_e c^2\right)$. So, for the neutrino sum we likewise define $u_{\Uparrow} \equiv \sqrt{2}\left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2$. The question now is whether the numeric result 0.18876 eV in (19.1) should be assigned to this new $u_{\Uparrow}$ or to the mass sum $\left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2$. That is, where do we use the $\sqrt{2}$ factor? Given that for the neutrinos, $\Sigma_j m_j = \left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2 < 0.170$ eV with a 95% confidence level, this empirical data suggests that the appropriate assignment should be to the neutrino vev, namely:

$$u_{\Uparrow} \equiv \sqrt{2}\left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2 = \delta_{\Downarrow}\sqrt{v / M_P c^2} = 0.18876 \text{ eV} , \tag{19.2a}$$

which means that for the neutrino mass sum we have:

$$\boxed{\tfrac{1}{\sqrt{2}} u_{\Uparrow} \equiv \left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2 \equiv \tfrac{1}{\sqrt{2}}\delta_{\Downarrow}\sqrt{v / M_P c^2} = 0.13348 \text{ eV} < 0.170 \text{ eV}}, \tag{19.2b}$$

clearly fitting the empirical data in [53]. Were we to assign $\left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2 = 0.189$ eV we would be somewhat-outside the 95% zone. It also helps to write the above in terms of $\delta_{\Downarrow}$ as:

$$\delta_{\Downarrow} = u_{\Uparrow}\sqrt{M_P c^2 / v} = \sqrt{2}\left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2\sqrt{M_P c^2 / v} = 42.018 \text{ MeV} . \tag{19.2c}$$

The above (19.2) provide a theoretical prediction about the true sum of the physical neutrino rest masses, and a definition of a new vev $u_{\Uparrow}$ for the neutrinos which parallels the previous (14.3), (15.10) and (18.5) for quarks and the charged leptons. And, with (19.2c), we now have an *independent* understanding of $\delta_{\Downarrow} = \sqrt{2M_P c^2 / v}\left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2$, and see that this is a not an independent new parameter, but rather is simply a function of the neutrino masses and the Newton gravitational constant in $GM_P^{\,2} \equiv \hbar c$. Indeed, using (19.2c) we may update Figure 10 to display this new understanding, as seen below:

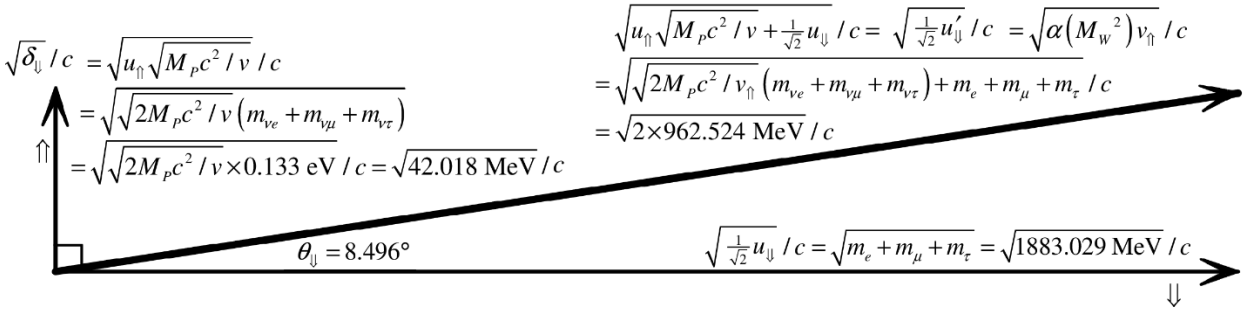

Figure 11: Charged Lepton and Amplified Neutrino Masses, and Rotation of the Charge Lepton Mass Space Vector

Above, we see the isospin-down and isospin-up leptons on orthogonal axes, labelled as such with $\Uparrow$ and $\Downarrow$. Except for the "neutrino mass amplifier" factor $\sqrt{M_P c^2/v}$, and the hypotenuse aligned toward isospin-down rather than up, this is identical in form to Figure 5 for quarks. This includes the $\sqrt{2}$ appearing as a multiplying factor for the isospin-up mass sum and not the isospin down mass sum, which shows theoretical consistency in addition to empirical confidence in the use of this factor in (19.2). Note also that the hypotenuse mirrors the Higgs mass relation $2m_h = \left(v_{\Uparrow}+\frac{1}{\sqrt{2}}v_{\Downarrow}\right)/c^2$ (16.3) as well, but for the neutrino mass amplifier. We see that the relation $\frac{1}{\sqrt{2}}u'_{\Downarrow} \equiv \frac{1}{\sqrt{2}}u_{\Downarrow}+\delta_{\Downarrow}$ in (18.11) – which does *not* have an analogue for quarks – effectively causes a rotation of the horizontal vector $\sqrt{\frac{1}{\sqrt{2}}u_{\Downarrow}}$ for isospin-down charged leptons, toward the vertical vector $\sqrt{u_{\Uparrow}}$ for isospin-up neutrinos with $\sqrt{M_P c^2/v}$ amplification.

At this point, having a predicted value $\left(m_{\nu e}+m_{\nu\mu}+m_{\nu\tau}\right)c^2 = 0.13348\text{ eV}$ for the sum of the neutrino rest masses, we follow the approach previously used for quark and charged lepton masses. Specifically, just as at before we postulate that all of the rest mass for the neutrinos starts off in a single neutrino, and then is subjected to a bi-unitary transformation leading to relations which mirror (18.1). However, unlike for the quarks and the charged leptons, we do not know the neutrino masses at the outset. Therefore, we need to first use the empirical data for the square mass differences defined by $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$ with $m_1 \equiv m_{\nu e}$, $m_2 \equiv m_{\nu\mu}$, $m_3 \equiv m_{\nu\tau}$ to get a better handle on ranges of the individual neutrino masses. Again turning to the data in [52], we work from the reasonable hypothesis that the neutrino masses have a "normal ordering" in which $m_{\nu e} < m_{\nu\mu} < m_{\nu\tau}$. And because [52] contains both 1σ and 3σ data, as we did with the PMNS data, we interpolate that the 2σ data is substantially equal to the 1σ and 3σ average. Accordingly, with $l = 1, 2$, the normal ordering data in [52] may be characterized by:

$$\begin{aligned}\Delta m_{21}^2 &= m_2^2 - m_1^2 = 7.40\ {}^{+0.21\ +0.42\ +0.62}_{-0.20\ -0.40\ -0.60}\times 10^{-5}\ \text{eV}^2/c^4 \\ \Delta m_{3l}^2 &= m_3^2 - m_l^2 = 2.494\ {}^{+0.033\ +0.066\ +0.099}_{-0.031\ -0.063\ -0.095}\times 10^{-3}\ \text{eV}^2/c^4\end{aligned}. \tag{19.3}$$

Now, we conduct the following calculation: We start with the top line above in the form $m_2 = \sqrt{m_1^2 + \Delta m_{21}^2}$. Given that (19.3) will cause m_1 and m_2 to be much closer to another than either of them is to m_3, we set $l = 2$ in the bottom line above which we now write as $m_3 = \sqrt{m_2^2 + \Delta m_{32}^2}$, also using $\Delta m_{32}^2 = \Delta m_{3l}^2$. In all cases, irrespective of the error spreads in (19.3) we also use $m_{\nu e} + m_{\nu\mu} + m_{\nu\tau} = 0.13348 \text{ eV}/c^2$ as a constraint to be applied in all cases to the sum of the three neutrino masses. Then, using a spreadsheet or the like, we sample various values of m_1 using the center values and each of the 1σ, 2σ, and 3σ spreads in (19.3). Specifically, we use our m_1 samples in $m_2 = \sqrt{m_1^2 + \Delta m_{21}^2}$ to determine m_2, simultaneously use m_2 in $m_3 = \sqrt{m_2^2 + \Delta m_{32}^2}$ to determine m_3, and keep sampling until the sum of all three masses always turns out to be $m_{\nu e} + m_{\nu\mu} + m_{\nu\tau} = 0.13348 \text{ eV}/c^2$ for the center values and the error spreads. The analytical calculation is $m_1 + \sqrt{\Delta m_{21}^2 + m_1^2} + \sqrt{\Delta m_{32}^2 + m_1^2 + \Delta m_{21}^2} = 0.13348 \text{ eV}/c^2$, but there is no straightforward way to analytically isolate m_1 which is why we use computational sampling. In this way we are able to predict the neutrino masses and obtain corresponding 1σ, 2σ, and 3σ spreads as follows:

$$\boxed{\begin{aligned} m_{\nu e}c^2 &= 0.03533 \,{}^{+0.00012\ +0.00024\ +0.00035}_{-0.00012\ -0.00025\ -0.00037} \text{ eV} \\ m_{\nu\mu}c^2 &= 0.03637 \,{}^{+0.00009\ +0.00017\ +0.00026}_{-0.00009\ -0.00018\ -0.00027} \text{ eV} \\ m_{\nu\tau}c^2 &= 0.06178 \,{}^{-0.00020\ -0.00041\ -0.00062}_{+0.00021\ +0.00043\ +0.00064} \text{ eV} \end{aligned}} \qquad (19.4)$$

It will be seen that this is a normal ordering, because the tau mass is clearly greater than the other two masses, and because even at 3σ the tau generation rest energy $m_{\nu\mu}c^2 > 0.03610$ eV while the electron generation rest energy $m_{\nu e}c^2 < 0.03569$ eV is smaller by at least 0.00041 eV. This also highlights how (19.3) causes the first- and second-generation neutrinos to have very close masses, and the third-generation neutrino to have a definitively-larger mass. Note also that the superscripted spreads for $m_{\nu e}$ and $m_{\nu\mu}$ are positive and those for $m_{\nu\tau}$ are negative. This is because the overall constraint $m_{\nu e} + m_{\nu\mu} + m_{\nu\tau} = 0.13348 \text{ eV}/c^2$ means that as the masses for the first two generations are increased, the third-generation is mass lowered, and vice versa. A similar calculation can be done for inverted and other possible ordering, but we shall leave such an exercise to the reader. Noting again from [54] that "somewhere between 10 meV and 2eV is our playground," we see that with the lightest neutrino mass predicted to be $m_{\nu e}c^2 \cong 35.33$ meV and a predicted mass sum $\Sigma m \cong 0.13348 \text{ eV}/c^2$ versus the empirical constraint $\Sigma_j m_j < 0.170$ eV, we have indeed landed right where we need to be in the "playground." Also, we have obtained (19.4) by regarding the neutrinos to be Dirac fermions insofar as we have approached these masses in exactly the same way as the quark and charged lepton masses. So empirical observation of these masses would serve to validate that the neutrinos are in fact Dirac fermions.

From here we follow the precise development that we used to previously reparameterize the quark masses in sections 14 and 15 and the charged lepton masses in section 18. Given the postulated normal ordering, we further postulate that all of the rest mass for the neutrinos starts off in the tau neutrino, and that a neutrino mass matrix analogous to that in (14.7) is then is subjected to a bi-unitary transformation leading to relations which mirror (18.8) for the charged leptons. Specifically, borrowing the top two relations for the type *I* "downward cascade" parameterization in (18.8a) and for the type *II* "distribution" parameterization in (18.8b), and migrating $\Downarrow$ to $\Uparrow$ and the charged leptons to their neutrino partners, we write:

$$\begin{aligned} \mathrm{c}_{I\Uparrow 32}{}^2 &= G_{\nu\tau} = \frac{m_{\nu\tau}c^2}{\frac{1}{\sqrt{2}}u_{\Uparrow}} = \frac{m_{\nu\tau}}{m_{\nu e}+m_{\nu\mu}+m_{\nu\tau}} \\ \mathrm{c}_{I\Uparrow 21}{}^2 &= \frac{G_{\nu\mu}}{\mathrm{s}_{I\Uparrow 32}{}^2} = \frac{m_{\nu\mu}c^2}{\frac{1}{\sqrt{2}}u_{\Uparrow}-m_{\nu\tau}c^2} = \frac{m_{\nu\mu}}{m_{\nu e}+m_{\nu\mu}} \end{aligned}, \tag{19.5a}$$

$$\begin{aligned} \mathrm{c}_{II\Uparrow 31}{}^2 &= \frac{G_{\nu\tau}}{\mathrm{c}_{II\Uparrow 32}{}^2} = \frac{m_{\nu\tau}c^2}{\frac{1}{\sqrt{2}}u_{\Uparrow}-m_{\nu\mu}c^2} = \frac{m_{\nu\tau}}{m_{\nu e}+m_{\nu\tau}} \\ \mathrm{s}_{II\Uparrow 32}{}^2 &= G_{\nu\mu} = \frac{m_{\nu\mu}c^2}{\frac{1}{\sqrt{2}}u_{\Uparrow}} = \frac{m_{\nu\mu}}{m_{\nu e}+m_{\nu\mu}+m_{\nu\tau}} \end{aligned}. \tag{19.5b}$$

Then, as in (18.9), we simply use (19.2b) and the 1σ, 2σ, and 3σ in (19.4) to calculate each of these angles, as similarly to (18.14), to be:

$$\begin{aligned} \vartheta_{I\Uparrow 32} &= 0.8226^{+0.0015\ +0.0031\ 0.0046}_{-0.0016\ -0.0032\ -0.0048}\ \mathrm{rad} = 47.131^{+0.087\ +0.177\ +0.266}_{-0.274\ -0.274\ -0.274}{}^\circ \\ \vartheta_{I\Uparrow 21} &= 0.7782^{+0.0003\ +0.0005\ +0.0008}_{-0.0002\ -0.0004\ -0.0007}\ \mathrm{rad} = 44.585^{+0.017\ +0.030\ +0.043}_{-0.039\ -0.025\ -0.039}{}^\circ \\ \vartheta_{II\Uparrow 31} &= 0.6475^{+0.0016\ +0.0032\ +0.0049}_{-0.0016\ -0.0033\ -0.0049}\ \mathrm{rad} = 37.098^{+0.092\ +0.185\ +0.278}_{-0.093\ -0.188\ -0.283}{}^\circ \\ \vartheta_{II\Uparrow 32} &= 0.5492^{+0.0007\ +0.0014\ +0.0022}_{-0.0008\ -0.0016\ -0.0023}\ \mathrm{rad} = 31.466^{+0.040\ +0.082\ +0.125}_{-0.045\ -0.089\ -0.132}{}^\circ \end{aligned}. \tag{19.6}$$

Now, at (18.14) we were able to connect two of the three PMNS angles to the charged lepton mass mixing angles, namely, $\theta_{P13} = \vartheta'_{II\Downarrow 31}$ and $\theta_{P12} = \vartheta'_{I\Downarrow 21}$, while in the process obtaining tighter fits than those known at (18.10). The remaining real angle from (18.10) still to be fitted – presumably to the neutrino masses – is $\theta_{P23} = 47.2^{+1.9\ +3.1\ +4.3}_{-3.9\ -5.4\ -6.9}{}^\circ$. This is the least-tightly known of the three PMNS angles, varying even at 1σ from $43.3^\circ < \theta_{P23} < 49.1^\circ$. So, in (19.6) there are actually two angles – $\vartheta_{I\Uparrow 32}$ and $\vartheta_{I\Uparrow 21}$ – which fit within 1σ and so can be associated with the remaining angle θ_{P23}. So, we need now to discern which is the more suitable association.

For this, we review the connections earlier made for the quarks and charged leptons to see which association would be most consistent in relation to the angles in (19.6). First for the quarks, among what was calculated leading to (14.10) were what we would now denote as $\mathrm{c}_{I\Uparrow 21}{}^2 = G_c / \mathrm{s}_{I\Uparrow 32}{}^2$ and $\mathrm{c}_{II\Uparrow 31}{}^2 = G_t / \mathrm{c}_{II\Uparrow 32}{}^2$ which led to the connections

$\theta_{I\Uparrow 21} \equiv \theta_{C23} = 2.415 \pm 0.053\,^\circ$ and $\theta_{II\,31} \equiv \theta_{C13} = 0.209^{+0.015}_{-0.013}\,^\circ$ in (14.12) to two of the three CKM angles within experimental errors. Then, leading to (15.3) among what was calculated were $\mathrm{c}_{I\Downarrow 21}{}^2 = G_s / \mathrm{s}_{I\Downarrow 32}{}^2$ and $\mathrm{c}_{II\Downarrow 31}{}^2 = G_b / \mathrm{c}_{II\Downarrow 32}{}^2$ which after further analysis led in (15.6) to the connection $\theta_{I\Downarrow 21} \equiv \theta_{C12} = 12.975 \pm 0.026\,^\circ$ within errors, and one "leftover" angle $\theta_{II\Downarrow 31} = 1.921\,^\circ$. For the charged leptons, at (18.7) the calculations included $\mathrm{c}_{I\Downarrow 21}{}^2 = G_\mu / \mathrm{s}_{I\Downarrow 32}{}^2$ and $\mathrm{c}_{II\Downarrow 31}{}^2 = G_\tau / \mathrm{c}_{II\Downarrow 32}{}^2$. After then having to introduce an energy difference $\delta_\Downarrow$ at (18.11) which as later shown in (19.1) and (19.2) is actually related to an amplified neutrino mass sum, we calculated (18.12) which at (18.14) led to the connections $\theta_{P12} \equiv \vartheta'_{I\Downarrow 21} = 32.39^{+2.01\ +2.83\ +3.66}_{+0.47\ -0.25\ -0.97}\,^\circ$ and $\theta_{P13} \equiv \vartheta'_{II\Downarrow 31} = 8.795^{-0.096\ +0.050\ +0.195}_{-0.396\ -0.546\ -0.696}\,^\circ$ for two of the PMNS angles, within errors. *In all cases*, the mass mixing angles which connected to a CKM or PMNS angle took the form of a second-generation coupling (G_c, G_s, G_μ) divided by the sine-squared of a type-*I* mass mixing angle ($\sin^2\theta_{I\Uparrow 32}$, $\sin^2\theta_{I\Downarrow 32}$, $\sin^2\vartheta_{I\Downarrow 32}$), or of a third-generation coupling (G_t, G_b, G_τ) divided by the cosine-squared of a type-*II* mass mixing angle ($\cos^2\theta_{II\Uparrow 32}$, $\cos^2\theta_{II\Downarrow 32}$, $\cos^2\vartheta_{II\Downarrow 32}$), with the leftover angle coming from $\mathrm{c}_{II\Downarrow 31}{}^2 = G_b / \mathrm{c}_{II\Downarrow 32}{}^2$.

If the pattern which held for isospin-up and isospin-down quarks and for charged leptons is to also carry through for neutrinos, then using (18.7) and (18.8) for guidance, it appears that $\mathrm{c}_{I\Uparrow 21}{}^2 = G_{\nu\mu} / \mathrm{s}_{I\Uparrow 32}{}^2$ in (19.5a) (second generation, type-*I*, inverse sine-squared) is what should be connected to the final PMNS angle, and that $\mathrm{c}_{II\Uparrow 31}{}^2 = G_{\nu\tau} / \mathrm{c}_{II\Uparrow 32}{}^2$ in (19.5b) (third generation, type *II*, inverse cosine-squared) should be regarded as the lepton "leftover." Accordingly, we now formally connect $\vartheta_{I\Uparrow 21}$ in (19.6) to the remaining mixing angle θ_{P23}, and regard $\vartheta_{II\Uparrow 31}$ in (19.6) as the leftover angle for leptons. Following a presentation form similar to what was used in (15.6) for quarks, we combine this with (18.14) whereby all three PMNS angles plus the lepton leftover are now related to the mass matrix mixing angles by:

$$\boxed{\begin{aligned} \theta_{P23} \equiv \vartheta_{I\Uparrow 21} &= 44.585^{+0.017\ +0.030\ +0.043}_{-0.039\ -0.025\ -0.039}\,^\circ \\ \vartheta_{II\Uparrow 31} &= 37.098^{+0.092\ +0.185\ +0.278}_{-0.093\ -0.188\ -0.283}\,^\circ \\ \theta_{P12} \equiv \vartheta'_{I\Downarrow 21} &= 32.39^{+2.01\ +2.83\ +3.66}_{+0.47\ -0.25\ -0.97}\,^\circ \\ \theta_{P13} \equiv \vartheta'_{II\Downarrow 31} &= 8.795^{-0.096\ +0.050\ +0.195}_{-0.396\ -0.546\ -0.696}\,^\circ \end{aligned}} \qquad (19.7)$$

From the second line of (19.5a), we see that this angle is slightly *less than 45 degrees* because the rest mass of the mu neutrino is *slightly greater than* the rest mass of the electron neutrino, thus preserving normal ordering. This new valuation $\theta_{P23} = 44.585^{+0.017\ +0.030\ +0.043}_{-0.039\ -0.025\ -0.039}\,^\circ$ is tighter than the usual $\theta_{P23} = 47.2^{+1.9\ +3.1\ +4.3}_{-3.9\ -5.4\ -6.9}\,^\circ$ from [52], because it is rooted in the square-mass differences (19.3) which have been measured with tighter precision than θ_{P23} directly. With the usual θ_{P23} having a large error range especially on the low side, this new center at $\theta_{P23} = 44.585°$ is actually only at about $.67\sigma$ below the usual $\theta_{P23} = 47.2°$ center. Moreover, the top-to-bottom

3σ spread in this new valuation is a mere $0.082°$, versus the usual $11.2°$ spread. Thus, this new valuation is about 135 times as precise at 3σ, providing ample opportunity for experimental confirmation as it becomes possible to obtain more precise direct measurements of θ_{P23}. This is why we are able to add two digits after the decimal in the new valuation of θ_{P23} in (19.7). To highlight the parallels between quarks and leptons, pulling together all six of the CKM and PMNS mass-mixing to flavor-mixing connections from (15.6) and (19.7) which have now been established, as well as the leftover angles, what we have now found is that within experimental errors we may associate:

$$\begin{array}{l} \theta_{C12} \equiv \theta_{I\Downarrow 12};\quad \theta_{P12} \equiv \vartheta'_{\ I\Downarrow 12} \\ \theta_{C23} \equiv \theta_{I\Uparrow 12};\quad \theta_{P23} \equiv \vartheta_{I\Uparrow 12} \\ \theta_{C13} \equiv \theta_{II\Uparrow 31};\quad \theta_{P13} \equiv \vartheta'_{\ II\Downarrow 31} \\ \text{leftover: } \theta_{II\Downarrow 31};\quad \vartheta_{II\Uparrow 31} \end{array} . \tag{19.8}$$

Finally, similarly to what we did at (14.14), (15.14) and (18.19), we may solve the simultaneous equations (19.5) and apply (19.7), and define a coupling $G_\nu \equiv m_\nu c^2 / \frac{1}{\sqrt{2}} u_{\Uparrow}$ for each neutrino type, to obtain:

$$G_{\nu\tau} = \frac{\cos^2 \vartheta_{II\Uparrow 31} \sin^2 \theta_{P23}}{1-\cos^2 \vartheta_{II\Uparrow 31} \cos^2 \theta_{P23}}; G_{\nu\mu} = \frac{\sin^2 \vartheta_{II\Uparrow 31} \cos^2 \theta_{P23}}{1-\cos^2 \vartheta_{II\Uparrow 31} \cos^2 \theta_{P23}}; G_{\nu e} = G_{\nu\mu} \tan^2 \theta_{P23} = G_{\nu\tau} \tan^2 \vartheta_{II\Uparrow 31}. \tag{19.9}$$

Now let's review in totality how we have been able to reparameterize all twelve of the fermion masses. In the process of doing so, we are led to predict a second Higgs boson associated with lepton masses and beta decays.

20. Prediction of a Second Leptonic Higgs Boson, and its Mass

Back at (16.3) we showed how the mass of the Higgs boson can be described within experimental errors by $m_h c^2 \equiv \left(v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}\right)/2$ to the Fermi vev $v = v_{\Uparrow} = \sqrt{2}\left(m_u c^2 + m_c c^2 + m_t c^2\right)$ and the sum of isospin-down quark masses $\frac{1}{\sqrt{2}} v_{\Downarrow} = m_d c^2 + m_s c^2 + m_b c^2$, see (15.10). And in Figure 5, it was shown how $\sqrt{v_{\Uparrow} + \frac{1}{\sqrt{2}} v_{\Downarrow}} / c = \sqrt{2m_h}$ actually specifies the hypotenuse of the orthogonal mass spaces for $v_{\Uparrow}$ and $v_{\Downarrow}$. Now the we have similar expressions $\frac{1}{\sqrt{2}} u_{\Downarrow} = m_\tau c^2 + m_\mu c^2 + m_e c^2$ in (18.5) and $u_{\Uparrow} \equiv \sqrt{2}\left(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}\right)c^2$ in (19.2a) for the leptons, we can likewise plot out a lepton analog to Figure 5 in which the larger number $\sqrt{u_{\Downarrow} / \sqrt{2}}\, c$ is drawn along the horizontal axis and the smaller number and the smaller number $\sqrt{u_{\Uparrow}} / c$ is drawn vertically. Such a figure would be similar to Figure 11, but it would lack the $\sqrt{M_P c^2 / v}$ amplifier, and so the angle corresponding to $\theta_l = 8.496°$ in Figures 10 and 11 would be exceedingly small, amounting in effect to merely

drawing a horizontal line of length $\sqrt{u_{\Downarrow}/\sqrt{2}}c$. To be precise, given the values we have computed in (18.5) and (19.2a), the ratio would be $\sqrt{1883.029\text{ MeV}/0.18876\text{ eV}} = 99878.85$ between the two axis lengths, with an easily computed angle of $\theta = 5.73653\times10^{-4}\,^\circ$ or $\theta = 2.0652"$. So, for example, if the vertical leg was drawn at about a half an inch in height, the horizontal leg if drawn to scale would have to run for about a mile. And the hypotenuse would have a length of $\sqrt{u_{\Uparrow} + \frac{1}{\sqrt{2}}u_{\Downarrow}}/c \cong \sqrt{\frac{1}{\sqrt{2}}u_{\Downarrow}}$ due to the scant 2" angle just noted

Taken together with the parallels formulated throughout between the quark and lepton masses spaces, and the need – to be explored in the next section – to develop a Lagrangian potential for leptons with a second maximum parallel to that for the section 16 quark potential, this is highly suggestive that there exists a second leptonic Higgs field denoted h_2 with a second Higgs boson having a mass m_{h2} defined analogously to (16.3) by

$$\boxed{m_{h2}c^2 \equiv \frac{u_{\Uparrow} + \frac{1}{\sqrt{2}}u_{\Downarrow}}{2} \cong \frac{1}{2\sqrt{2}}u_{\Downarrow} = 941.515 \pm 0.060\text{ MeV}}, \tag{20.1}$$

where we have used $\frac{1}{\sqrt{2}}u_{\Downarrow} = 1883.029 \pm 0.120\text{ MeV}$ from (18.5) to supply the empirical data. Also using (19.2a), because $u_{\Uparrow}/\frac{1}{\sqrt{2}}u_{\Downarrow} = 1.005\times10^{-09}$ the above approximation sets $u_{\Uparrow} \cong 0$, since any effects this may have are six digits outside of the experimental error range for $\frac{1}{\sqrt{2}}u_{\Downarrow}$. This new Higgs mass differs from the proton and neutron masses $M_P = 938.272081 \pm 0.000006\text{ MeV}$ and $M_N = 939.565413 \pm 0.000006\text{ MeV}$ [55] by only a few MeV – in the former case by 3.243 MeV and in the latter by 1.950 MeV.

Now, in general, there are three types of predictions that can be made for empirical data. First, there is *retrodiction*, in which empirical data which is already known is explained in relation to other known data. This reduces the number of independent data numbers in our physical theories, and is often accompanied by better theoretical understanding of the observed physics. This is exemplified here, so far, by (17.1) and (18.21), and will be further by (20.15) below. Second, there is *tuning prediction*, in which a prediction is made about how the experimental error bars for already-known data will be affected as it becomes possible to obtain tighter measurements of this data, owing to better experiments and / or better theory. This is exemplified here by (15.11) and (15.12) for tighter top and strange quark masses, (16.3) for a tighter Higgs mass, (18.14) for re-centered θ_{P12} and θ_{P13} values, (18.16) for a tighter $\alpha\left(M_W{}^2\right)$, and (19.7) for a far-tighter θ_{P23}. Third, there is *outright prediction*, in which data which is known to exist but has not yet been successfully measured is predicted, or in which some data which is not even known to exist is predicted to exist, along with a prediction as to how it will be measured. This is most important, because absent theoretical information telling us where to target our detection efforts, experiments to detect such data are often carried out "scattershot" over a broad range of possible values.

Here, (19.2b), (19.4) and (20.1) contain *outright predictions* of four mass values which at present are not known. In (19.4) we are now told exactly the energies at which to look for the three neutrino masses, and in (19.2b) their mass sum. And in (20.1) we are told not only that a new Higgs boson exists, but we are told that to find it, one should be looking in the zone of energies just a few MeV higher than the proton and neutron rest energies. Now, *knowing precisely where to look*, experimental efforts to pinpoint neutrino masses can be focused on confirming the mass sum $m_{\nu\tau}c^2 + m_{\nu\mu}c^2 + m_{\nu e}c^2 = 0.133 \text{ eV}$ and the separate masses in (19.4). And of course, finding a second Higgs boson at $m_{h2}c^2 = 941.515 \pm 0.060 \text{ MeV}$, just above the proton and neutron rest energies, would be entirely new, because the very existence of such a new particle – much less its mass value – is entirely unanticipated based on present knowledge.

As to retrodiction, we now supplement (18.21) with the neutrino and the leptonic Higgs developments, using $G_F = 1.1663787(6)\times 10^{-5} \text{ GeV}^{-2}$ and $G = 6.708\ 61(31)\times 10^{-39} \text{ GeV}^{-2}$ [21] in natural units as proxies for the Fermi vev and Planck mass. Starting from (18.21), we summarize the complete reparameterization of *all twelve fermion masses*, including "leftover" angles, by:

$$\begin{aligned}&\left\{m_t, m_c, m_u, m_b, m_s, m_d, m_\tau, m_\mu, m_e, m_{\nu\tau}, m_{\nu\mu}, m_{\nu e}\right\} \\ &= F\left(G, G_F, m_h, \alpha\left(M_W{}^2\right), \theta_{C12}, \theta_{C23}, \theta_{C31}, \theta_{II\Downarrow 31}, \theta_{P12}, \theta_{P23}, \theta_{P13}, \vartheta_{II\Uparrow 31}\right),\end{aligned} \tag{20.2}$$

In the above, we have momentarily included the leftover angles $\theta_{II\Downarrow 31}$ of (15.6) and $\vartheta_{II\Uparrow 31}$ of (19.7) because these explicitly appear in (15.14) for isospin-down quarks and in (19.9) for neutrinos (isospin-up leptons). However, these leftover angles are redundant, which we can see specifically via (15.3) and the lower (19.5a) together with the upper (19.5b). The mathematical origin of this redundancy is based on what is discussed from [12.114] to [12.116] of [20]: For an $N\times N$ unitary matrix mixing N generations of quarks or of leptons there are of course N^2 real elements. But because we can change the phase of each of $2N$ quark or lepton states independently without altering the observable physics, such a matrix will only contain $N^2 - (2N-1)$ real parameters. So, for N=3 there are 4 real parameters, which in the case of the mass mixing matrices used in the bi-unitary transformations of sections 14, 15, 18 and 19 can be parameterized into $\theta_{I\Uparrow 21}$, $\theta_{II\Uparrow 31}$, $\theta_{I\Downarrow 21}$, $\theta_{II\Downarrow 31}$ which we have used for quarks and $\vartheta'_{I\Downarrow 21}$, $\vartheta'_{II\Downarrow 31}$, $\vartheta_{I\Uparrow 21}$ and $\vartheta_{II\Uparrow 31}$ which we have used for leptons. However, in each case an overall phase can be omitted while the unitary matrix remains invariant. Thus, we drop from 4 to 3 real parameters for each of the quarks and leptons, and this accounts for the leftover angles. Accordingly, these redundant angles may be removed from (20.2) by an overall phase omission, in which case we will have actually reparameterized twelve fermion masses with only ten parameters.

However, we still need an overall energy scale which *cannot* be independently deduced from the parameters in (20.2). To see this, start with (18.21) which contains $\delta_\Downarrow$ as an added parameter. We of course found in (19.2c) that we can relate this to the neutrino mass sum $m_e + m_\mu + m_\tau = 0.133 \text{ eV}/c^2$ using G, G_F. So, it is not that we do not know the value of this parameter, because now we do. It is that this parameter is only known because of our knowledge,

among other things, of the charged lepton rest mass sum. That is, it is only known because of the mass sum in (19.2b) which we may combine with (18.18) to obtain:

$$\left(m_{\nu e}+m_{\nu\mu}+m_{\nu\tau}\right)c^2=\tfrac{1}{\sqrt{2}}\sqrt{v/M_P c^2}\left(\alpha\left(M_W{}^2\right)v_{\Uparrow}-\left(m_\tau+m_\mu+m_e\right)c^2\right). \tag{20.3}$$

Knowing the parameters G_F thereby $v_{\Uparrow}=v$, and $\alpha\left(M_W{}^2\right)$ in (20.2), we can of course use (18.16) and (19.2c) to deduce $m_\tau c^2+m_\mu c^2+m_e c^2+\delta_{\Downarrow}$. But this gives us neither $m_\tau+m_\mu+m_e$ nor $m_{\nu e}+m_{\nu\mu}+m_{\nu\tau}$ separately, but only a combination of the two together with G, G_F, $\alpha\left(M_W{}^2\right)$. Therefore, with (20.3), we could regard either $m_{\nu\tau}+m_{\nu\mu}+m_{\nu e}$ or $m_\tau+m_\mu+m_e$ as the mass sum still not reparameterized in (20.2), and then deduce the other. But one of these sums must be given at the start to be able to infer all of the fermion masses.

Which of these two mass sums we choose to "seed" an overall energy scale is really an aesthetic matter. But if we use the leptin Higgs mass $m_{h2}c^2\cong\frac{1}{2\sqrt{2}}u_{\Downarrow}=941.515\pm0.060$ MeV as a proxy for charged lepton mass sum (18.5) because $u_{\Uparrow}$ in (20.1) is empirically indiscernible by comparison given that the ratio $u_{\Uparrow}/\frac{1}{\sqrt{2}}u_{\Downarrow}=1.005\times10^{-09}$, then (20.3) now becomes:

$$\left(m_{\nu e}+m_{\nu\mu}+m_{\nu\tau}\right)c^2\cong\sqrt{v/M_P c^2}\left(\tfrac{1}{2}\alpha\left(M_W{}^2\right)v_{\Uparrow}-m_{h2}c^2\right). \tag{20.4}$$

Therefore, we choose the aesthetics of $m_\tau+m_\mu+m_e$, then use the new m_{h2} as a proxy for this sum, while removing the redundant, phased-away leftover angles from (20.2), to finally write:

$$\boxed{\begin{aligned}&\left\{m_t,m_c,m_u,m_b,m_s,m_d,m_\tau,m_\mu,m_e,m_{\nu\tau},m_{\nu\mu},m_{\nu e}\right\}\\&=F\left(G,G_F,m_h,m_{h2},\alpha\left(M_W{}^2\right),\theta_{C12},\theta_{C23},\theta_{C31},\theta_{P12},\theta_{P23},\theta_{P13}\right)\end{aligned}}. \tag{20.5}$$

Consequently, we have finally reparameterized all *twelve* fermion rest masses into *eleven* previously-disconnected parameters. But m_{h2} is a proxy for $m_\tau+m_\mu+m_e$ given that $u_{\Uparrow}/\frac{1}{\sqrt{2}}u_{\Downarrow}=1.005\times10^{-09}$ and so $u_{\Uparrow}$ can be neglected in (20.1). And $m_\tau+m_\mu+m_e$ is known as soon as we start with all the fermion masses. Consequently, we have really reduced twenty-two physics parameters – twelve masses and the ten parameters other than m_{h2} in (20.5) – down to eleven parameters, removing eleven independent unknowns from our understanding of the natural world.

21. The Two-Minimum, Two Maximum Lagrangian Potential for Leptons

Now let us turn to the Lagrangian potential for leptons which, in contrast with the quark potential reviewed in section 16, we shall denote by *U* rather than *V*, and for which we shall replace ϕ_h by φ_h. Consequently, given that (11.3) contains the symmetry-broken $\phi_h=\frac{1}{\sqrt{2}}\phi_{1h}=\frac{1}{\sqrt{2}}\left(v+h\right)$

, we replace this for leptons with $\varphi_h = \frac{1}{\sqrt{2}}\varphi_{1h} = \frac{1}{\sqrt{2}}\left(u + h_2\right)$. This is entirely a notational replacement intended to clearly distinguish quarks from leptons, and nothing more. In the above, as with the quarks, *u* will be the *larger* of the two lepton vevs, namely, $u_{\Downarrow} = 2663.005 \pm 0.170$ MeV obtained from (16.5) for charged leptons, versus the enormously-smaller $u_{\Uparrow} = 0.189$ eV from (16.2a) for neutrinos. For the leptons, plots similar to Figures 1 and 2 may be drawn, but with the vevs established by one of the two foregoing vevs, not the Fermi vev, see the discussion following (13.10) which applies here also. Then, borrowing from the recalibrated (14.1b), and with $U' = dU / d\varphi_{1h}$, the Lagrangian potential to be studied for leptons is specified in leading order by:

$$U\left(\varphi_{1h}\right) = \lambda_l\left(-\tfrac{1}{2}{u_{\Downarrow}}^2{\varphi_{1h}}^2 + \tfrac{1}{4}{\varphi_{1h}}^4\right) = -\frac{1}{4}{m_{h2}}^2c^4{\varphi_{1h}}^2 + \frac{1}{8}\frac{{m_{h2}}^2c^4}{{u_{\Downarrow}}^2}{\varphi_{1h}}^4 = {m_{h2}}^2c^4\left(-\frac{1}{4}{\varphi_{1h}}^2 + \frac{1}{8}\frac{1}{{u_{\Downarrow}}^2}{\varphi_{1h}}^4\right) \quad (21.1)$$
$$U'\left(\varphi_{1h}\right) = \lambda_l\varphi_{1h}\left({\varphi_{1h}}^2 - {u_{\Downarrow}}^2\right) = \frac{{m_{h2}}^2c^4}{2{u_{\Downarrow}}^2}\varphi_{1h}\left({\varphi_{1h}}^2 - {u_{\Downarrow}}^2\right)$$

Continuing with the notational distinctions which are entirely of form, we also use λ_l in the above to denote this parameter as it applies to leptons, while $m_{h2} = 941.515 \pm 0.060$ MeV / c^2 is the second leptonic Higgs mass discovered in (18.1). The only substantive change made in (21.1) versus (14.1b) which is not merely notational to distinguish quarks from leptons, is the use of $u_{\Downarrow}$ rather than $u_{\Uparrow}$. This is because for quarks $v_{\Uparrow}$ is the larger vacuum versus $v_{\Downarrow}$, while for leptons $u_{\Downarrow} \gg u_{\Uparrow}$. So, in (21.1) we have utilized the larger vev, and will develop *U* to ensure that this vev supplies the *global* minimum with the neutrino vev supplying a second, *local* minimum.

From here we follow the same path that was taken in section 16 to develop the Lagrangian potential for quarks. We construct $U\left(\varphi_{1h}\right)$ with higher-order terms so as to require two minima. One of these is to be centered at $\varphi_{1h} = u_{\Downarrow} = 2663.005$ MeV the charged leptons, and the other at $\varphi_{1h} = u_{\Uparrow} = 0.18876$ eV for the neutrinos. We also require the usual maximum at $\varphi_{1h} = 0$ and a new, second maximum that is established using m_{h2}. In establishing the second maximum in this way, we apply the same rationale for why we used (14.8) and (14.9) to establish the second quark maximum as reviewed following those two equations. So, starting with $\varphi_h = \frac{1}{\sqrt{2}}\varphi_{1h} = \frac{1}{\sqrt{2}}\left(u + h_2\right)$ in the preceding paragraph, we set $u = u_{\Downarrow}$ to the larger of the two vevs, so that $\varphi_{1h} = u_{\Downarrow} + h_2$. Then, as in (14.8) we establish the maximum at the domain point where:

$$h_2\left(x^{\mathrm{M}}\right) = -m_{h2}c^2 = -941.515 \pm 0.060 \text{ MeV}, \quad (21.2)$$

and therefore, as in (14.9), also using the new Higgs mass in (18.1), where:

$$\varphi_{1h}\left(x^{\mathrm{M}}\right) = u_{\Downarrow} + h_2\left(x^{\mathrm{M}}\right) = u_{\Downarrow} - m_{h2}c^2 = \left(1 - \tfrac{1}{2\sqrt{2}}\right)u_{\Downarrow} - \tfrac{1}{2}u_{\Uparrow} \cong \left(1 - \tfrac{1}{2\sqrt{2}}\right)u_{\Downarrow} = 1721.491 \pm 0.110 \text{ MeV}. \quad (21.3)$$

Next, we follow suit from (14.10) to build in these minima and maxima by defining:

$$\begin{aligned} U' &= B\frac{{m_{h2}}^2c^4}{2{u_{\Downarrow}}^2}\varphi_{1h}\left({\varphi_{1h}}^2-{u_{\Downarrow}}^2\right)\left({\varphi_{1h}}^2-\left(u_{\Downarrow}-m_{h2}c^2\right)^2\right)\left({\varphi_{1h}}^2-{u_{\Uparrow}}^2\right) \\ &= B\frac{{m_{h2}}^2c^4}{2{u_{\Downarrow}}^2}\begin{pmatrix} -{u_{\Downarrow}}^2{u_{\Uparrow}}^2\left(u_{\Downarrow}-m_{h2}c^2\right)^2\varphi_{1h}+\left({u_{\Downarrow}}^2{u_{\Uparrow}}^2+\left({u_{\Downarrow}}^2+{u_{\Uparrow}}^2\right)\left(u_{\Downarrow}-m_{h2}c^2\right)^2\right){\varphi_{1h}}^3 \\ -\left({u_{\Downarrow}}^2+{u_{\Uparrow}}^2+\left(u_{\Downarrow}-m_{h2}c^2\right)^2\right){\varphi_{1h}}^5+{\varphi_{1h}}^7 \end{pmatrix}. \end{aligned} \tag{21.4}$$

with an overall coefficient B that will be used to match the leading term in the upper (21.1). As in (14.11) we then integrate, and find we must set $B=1/{u_{\Uparrow}}^2\left(u_{\Downarrow}-m_{h2}c^2\right)^2$ for the leading term to match (21.1). And, also to match, we discard the integration constant. Thus, we obtain:

$$U\left(\varphi_{1h}\right)={m_{h2}}^2c^4\begin{pmatrix} -\frac{1}{4}{\varphi_{1h}}^2+\frac{1}{8}\frac{1}{{u_{\Downarrow}}^2}{\varphi_{1h}}^4+\frac{1}{8}\left(\frac{1}{{u_{\Uparrow}}^2}+\frac{1}{\left(u_{\Downarrow}-m_{h2}c^2\right)^2}\right){\varphi_{1h}}^4 \\ -\frac{1}{12}\left(\frac{1}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}+\frac{1}{\left(u_{\Downarrow}-m_{h2}c^2\right)^2}\frac{{u_{\Downarrow}}^2+{u_{\Uparrow}}^2}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}\right){\varphi_{1h}}^6+\frac{1}{16}\frac{1}{\left(u_{\Downarrow}-m_{h2}c^2\right)^2}\frac{1}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^8 \end{pmatrix}. \tag{21.5}$$

Then we separate terms as in (14.12) and use the approximations in (18.1) and (21.3), thus:

$$\begin{aligned} U\left(\varphi_{1h}\right) &= {m_{h2}}^2c^4\left(-\frac{1}{4}{\varphi_{1h}}^2+\frac{1}{8}\frac{{u_{\Downarrow}}^2+{u_{\Uparrow}}^2}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^4-\frac{1}{12}\frac{1}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^6\right) \\ &\quad+\frac{{m_{h2}}^2c^4}{\left(u_{\Downarrow}-m_{h2}c^2\right)^2}\left(\frac{1}{8}{\varphi_{1h}}^4-\frac{1}{12}\frac{{u_{\Downarrow}}^2+{u_{\Uparrow}}^2}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^6+\frac{1}{16}\frac{1}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^8\right) \\ &\cong \frac{1}{8}{u_{\Downarrow}}^2\left(-\frac{1}{4}{\varphi_{1h}}^2+\frac{1}{8}\frac{{u_{\Downarrow}}^2+{u_{\Uparrow}}^2}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^4-\frac{1}{12}\frac{1}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^6\right) \\ &\quad+\frac{1}{\left(1-2\sqrt{2}\right)^2}\left(\frac{1}{8}{\varphi_{1h}}^4-\frac{1}{12}\frac{{u_{\Downarrow}}^2+{u_{\Uparrow}}^2}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^6+\frac{1}{16}\frac{1}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^8\right) \end{aligned}. \tag{21.6}$$

Because $u_{\Uparrow}/\frac{1}{\sqrt{2}}u_{\Downarrow}=1.005\times10^{-09}$, see following (18.1), it is possible within experimental errors for the charged lepton masses to drop some unobservable terms and so further reduce the above to:

$$U\left(\varphi_{1h}\right)=-\frac{1}{32}{u_{\Downarrow}}^2{\varphi_{1h}}^2+\frac{1}{64}\frac{{u_{\Downarrow}}^2}{{u_{\Uparrow}}^2}{\varphi_{1h}}^4-\left(\frac{1}{96}+\frac{1}{12\left(1-2\sqrt{2}\right)^2}\right)\frac{1}{{u_{\Uparrow}}^2}{\varphi_{1h}}^6+\frac{1}{16\left(1-2\sqrt{2}\right)^2}\frac{1}{{u_{\Downarrow}}^2{u_{\Uparrow}}^2}{\varphi_{1h}}^8. \tag{21.7}$$

This sort of reduction has no analog for (14.12) because for quarks, the ratio $v_{\Uparrow} / \frac{1}{\sqrt{2}} v_{\Downarrow} \cong 57.5635$, see (13.10), whereby both $v_{\Uparrow}$ and $v_{\Downarrow}$ make all terms observable over at least some pertinent regions of the domain.

Then, as in (14.13), using the numerical values of $u_{\Downarrow}$ and $u_{\Uparrow}$ obtained from (16.5) and (16.2a), with φ_{1h} in both MeV and eV and thus $U\left(\varphi_{1h}\right)$ in MeV^4 and eV^4 respectively, we obtain:

$$\begin{aligned} U\left(\varphi_{1h}\right)\left[\text{MeV}^4\right] &= -2.216\times 10^{5}\varphi_{1h}{}^{2} + 3.102\times 10^{18}\varphi_{1h}{}^{4} - 9.894\times 10^{11}\varphi_{1h}{}^{6} + 7.380\times 10^{4}\varphi_{1h}{}^{8} \\ U\left(\varphi_{1h}\right)\left[\text{eV}^4\right] &= -2.216\times 10^{17}\varphi_{1h}{}^{2} + 3.102\times 10^{18}\varphi_{1h}{}^{4} - 9.894\times 10^{-1}\varphi_{1h}{}^{6} + 7.380\times 10^{-20}\varphi_{1h}{}^{8} \end{aligned} . \qquad (21.8)$$

We show both MeV and eV because given the large chasm between the charged lepton and the neutrino vevs, the former is better for studying the charged lepton vev and the latter for studying the neutrino vev. As with (14.13) we may than draw plots of $U\left(\varphi_{1h}\right)$ similar to Figures 6 and 7, and may also draw fourth root plots similar Figures 8 and 9. The qualitative character of these plots is exactly the same as that of Figure 6 through 9. Quantitatively, however there are two significant differences: First, the two vev minima for leptons are widely-separated by the ratio $u_{\Uparrow} / \frac{1}{\sqrt{2}} u_{\Downarrow} = 1.005\times 10^{-09}$ versus the much-closer $v_{\Uparrow} / \frac{1}{\sqrt{2}} v_{\Downarrow} \cong 57.5635$ for quarks. Second, as a direct result of this, the wells in the lepton Lagrangian potential are much deeper and the barrier set by (21.2) and (21.3) much higher than their quark potential counterparts.

To see this in detail, we start with Figures 12, 13, 14 and 15 below which analogous and qualitatively-similar to Figures 6, 7, 8 and 9, but now for leptons not quarks. We see the minima and maxima of $U\left(\varphi_{1h}\right)$ at the domain points which were built in via (21.4). There are two primary quantitative contrasts with Figures 6 through 9: First, whereas the vev minima for the quarks have a ratio $v_{\Uparrow} / \frac{1}{\sqrt{2}} v_{\Downarrow} = 57.5635$ and so are somewhat close to one another, for the leptons the analogous ratio $u_{\Uparrow} / \frac{1}{\sqrt{2}} u_{\Downarrow} = 1.005\times 10^{-09}$ produces an extraordinarily wide gulf between the two minima along the horizontal axis. This of course, is directly reflective of the very tiny masses of the neutrinos. Second, as a direct consequence of this wide vev separation, the depths of the two vev minima and the height of the intermediate maximum have magnitudes which – in relation to φ_{1h} – are far greater than what appears for the quarks in Figures 6 through 9. This is why the horizontal and vertical axes in the wide-view Figures 12 and 14 below are sized in GeV and TeV, while these same axes in the magnified center views of Figures 13 and 15 below are sized a billion times smaller in eV and KeV, and thus are magnified by a factor of a billion.

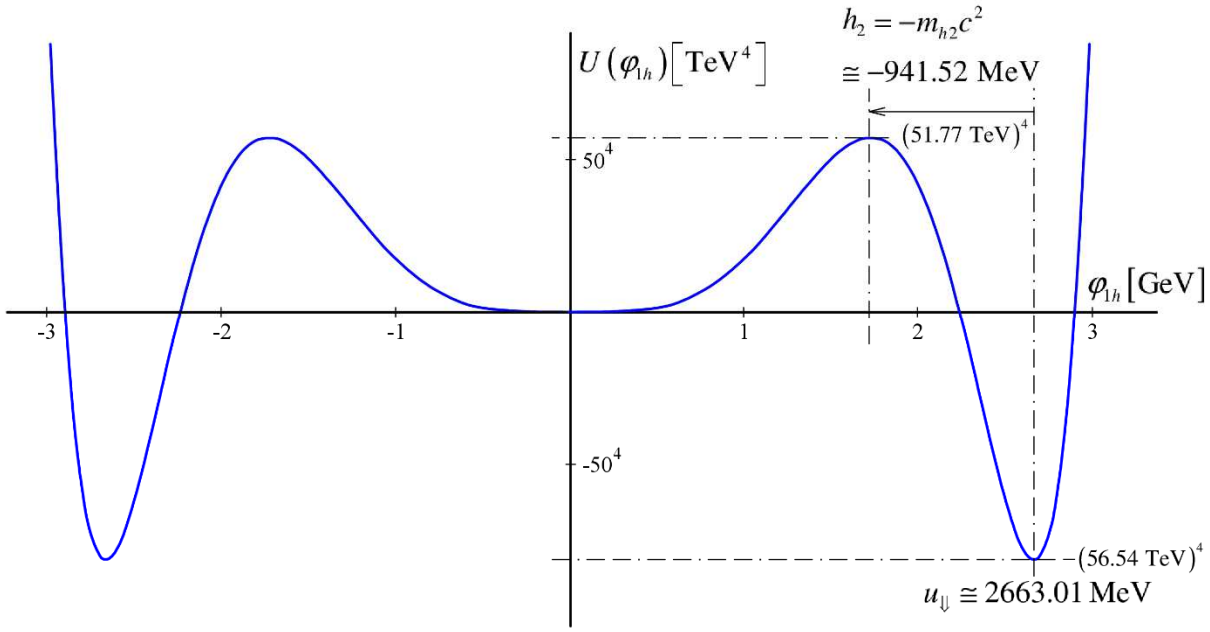

Figure 12: Lagrangian Potential for Leptons – Wide View

In Figure 12 above we see that the minimum at $u_{\Downarrow} \cong 2663.01\,\text{MeV}$ has an extremely large depth of $U(u_{\Uparrow}) \cong -(56.54\ \text{TeV})^4$, and the maximum at $h_2 = -m_{h2}c^2 \cong -941.52\ \text{MeV}$ has an extremely large height of $U(h_2 = -m_{h2}c^2) \cong (51.77\ \text{TeV})^4$. There is no possible way to visually represent the neutrino region of this plot, which is why we need the magnified figure below:

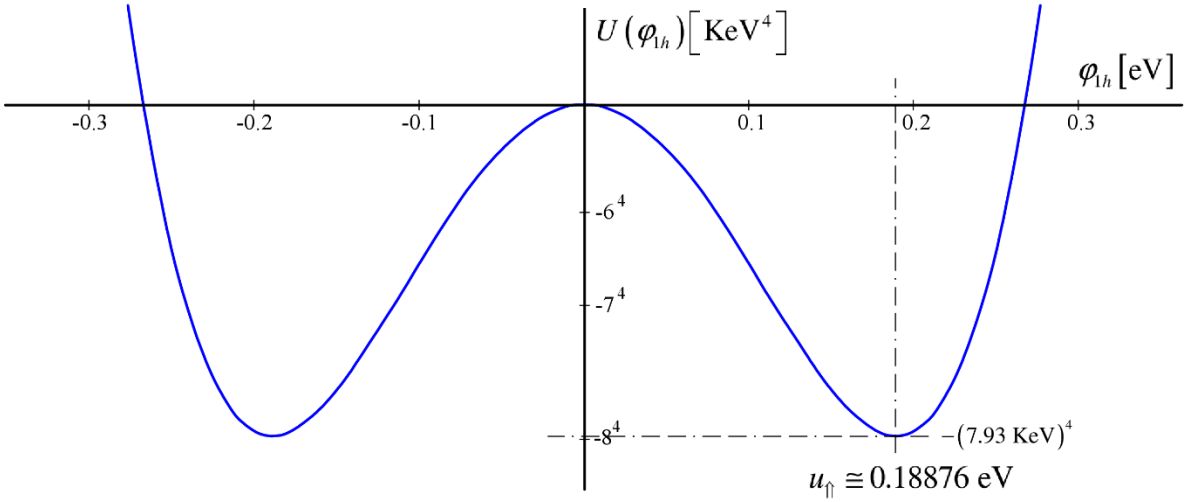

Figure 13: Lagrangian Potential for Leptons – Magnified Center View

In Figure 13 above, magnified by a factor of a billion over Figure 12, we see that the neutrino-well minimum at $u_{\Uparrow} \cong 0.18876\ \text{eV}$ also has – comparatively speaking – the extremely

large depth $U\left(u_{\Downarrow}\right) \cong -\left(7.93 \text{ KeV}\right)^4$. But because $U\left(\varphi_{1h}\right)$ has dimensions of energy to the fourth power, we again take fourth roots as we did in Figure 8 and 9, so that we can compare energy-to-energy. Below, we take this fourth root on the vertical axis for Figure 12, as such:

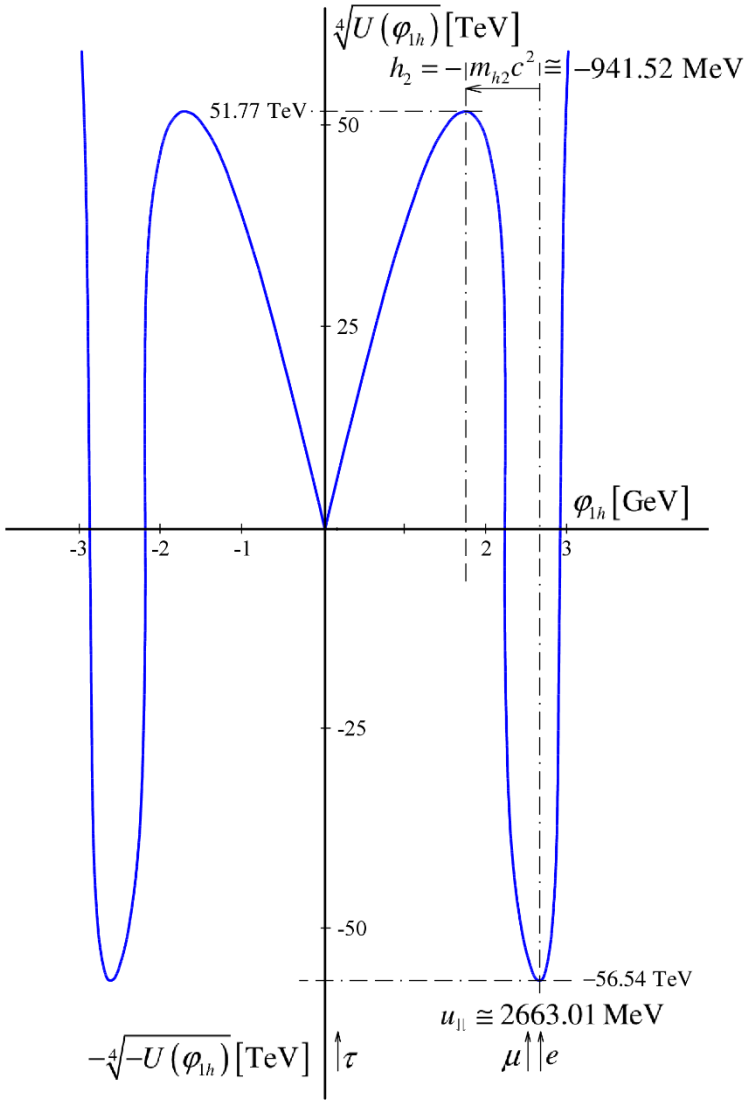

Figure 14: Lagrangian Potential for Leptons, Fourth Root – Wide View

In this Figure 14, even taking the vertical fourth root, the well depths and barrier height are so comparatively large, that we cannot draw the two energy axes to scale. Rather, the vertical axis is drawn to the scale of the horizontal axis with a compression factor of 10^4. That is, 1 GeV on the horizontal axis has the same linear scale as 10 TeV on the vertical axis. This makes clear that if these drawings were to scale both axes together as we were able to do for quarks, aside from the height of the drawing being close to a mile, the wells and the barrier would be extremely steep, with first derivatives far more vertical than even what is depicted. We also show the energetic placements of the three charged leptons in this well based on their versions of Figures 1 and 2

(with Figure 2 modified to use $u_{\Downarrow}$ rather than $v = v_{\Uparrow}$ as the vev, see following (13.10)). Similarly to the up and charm quarks, the electron and the muon nest very close to the vev minima, though the muon is somewhat more removed from its vev than is the charm quark from its vev. Now zooming into the center of Figure 14 by a factor of a billion, we arrive at the Figure below:

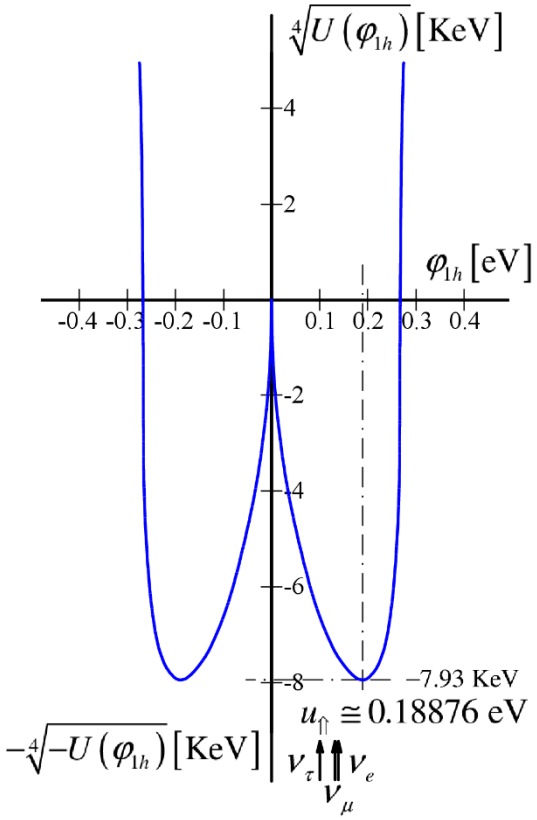

Figure 15: Lagrangian Potential for Leptons, Fourth Root – Magnified Center View

In this final Figure 15 which is the fourth root of Figure 13, we see the neutrino portion of the potential. The vev minimum is at $u_{\Uparrow} \cong 0.18876$ eV which is the energy first found at (17.1), and all three neutrinos situate discernably-displaced to the left of this minimum. This is in comparison to the quarks and charged leptons for which the first-generation (and more or less the second-generation) fermion does sit substantially right at the bottom of its potential well. Note also, considering both Figures 14 and 15, similarly to the top quark behavior in Figure 9, that the tau lepton nests to the left of the peak set by (21.2) and (21.3), inside the neutrino well, albeit well to the right of the neutrino vev minimum by what is still a factor on the order of a billion. Here too, although the two axes compare energy-to-energy, we cannot draw the axes to scale without the drawing approaching a mile in height. So, we again compress the vertical axis by a factor of 10^4. Now, for example, 0.1 eV on the horizontal axis scales to 1 KeV on the vertical axis.

In a sharp contrast to what we saw for quarks, it warrants attention that the neutrinos – which via have (17.4) masses from about 35 meV to 62 meV (milli-electron volts) – sit in a well that is close to 8 KeV deep, and that the charged leptons – with masses from about .5 MeV to 2 GeV – sit in a well that is over 50 TeV deep. Moreover, the barrier between the charged lepton

and the neutrino wells, set by the leptonic Higgs mass, itself peaks at over 50 TeV, which means that there is an energy difference of over 100 TeV between this peak and the bottom of the charged lepton well. This has important implications for how we must understand lepton beta decays between charged leptons and neutrinos, as will be explored in the next section.

We stated earlier that the very large magnitudes of the lepton well depth and barrier height are a direct consequence of the very wide chasm by which $u_{\Uparrow} / \frac{1}{\sqrt{2}} u_{\Downarrow} = 1.005 \times 10^{-09}$. It is good to explicitly see how this comes about, by analytically calculating this height and these depths. Working from (21.7), using $u_{\Downarrow} = 2663.005 \text{ MeV}$ from (16.5) and $u_{\Uparrow} = 0.18876 \text{ eV}$ from (17.2a), and applying the very small ratio $u_{\Uparrow} / \frac{1}{\sqrt{2}} u_{\Downarrow} = 1.005 \times 10^{-09}$ to set comparatively extremely small terms to zero, we may analytically calculate that the neutrino vev well depth:

$$U\left(\varphi_{1h} = u_{\Uparrow}\right) = -\tfrac{1}{64} u_{\Downarrow}{}^{2} u_{\Uparrow}{}^{2} = -\left(7931.8 \text{ eV}\right)^{4}, \tag{21.9a}$$

that the barrier between the two wells has a height of:

$$\begin{aligned} &U\left(\varphi_{1h} = \left(1 - \tfrac{1}{2\sqrt{2}}\right) u_{\Downarrow}\right) \\ &= \left(1 - \tfrac{1}{2\sqrt{2}}\right)^{4} \left[\tfrac{1}{64} - \left(\tfrac{1}{96} + \frac{1}{12\left(1-2\sqrt{2}\right)^{2}} \right) \left(1 - \tfrac{1}{2\sqrt{2}}\right)^{2} + \frac{1}{16\left(1-2\sqrt{2}\right)^{2}} \left(1 - \tfrac{1}{2\sqrt{2}}\right)^{4} \right] \frac{u_{\Downarrow}{}^{2}}{u_{\Uparrow}{}^{2}} u_{\Downarrow}{}^{4} = \left(51.771 \text{ TeV}\right)^{4}, \end{aligned} \tag{21.9b}$$

and that the charged lepton well depth is:

$$U\left(\varphi_{1h} = u_{\Downarrow}\right) = \left[\tfrac{1}{192} - \frac{1}{48\left(1-2\sqrt{2}\right)^{2}} \right] \frac{u_{\Downarrow}{}^{2}}{u_{\Uparrow}{}^{2}} u_{\Downarrow}{}^{4} = -\left(56.537 \text{ TeV}\right)^{4}. \tag{21.9c}$$

Again keeping in mind that $U\left(\varphi_{1h} = u_{\Uparrow}\right)$ is quartic in energy, we see the mix of vev in $u_{\Downarrow}{}^{2} u_{\Uparrow}{}^{2}$ in (21.9a) being responsible for the deep well in (21.9a) relative to the neutrino masses which are set exclusively by the much-smaller $u_{\Uparrow}$. And in (21.9b) and (9.9c) we see the gigantic ratio $u_{\Downarrow}{}^{2} / u_{\Uparrow}{}^{2} = 1.9903 \times 10^{20}$ being responsible for barrier height and charged lepton well depth having >50 TeV-scale energies that are huge in relation to the charged lepton masses.

22. How Weak Beta Decays are Triggered by Neutrinos and Antineutrinos Interacting with Electrons, Neutrons and Protons via the *Z* Boson-Mediated Weak Neutral Current, with "Chiral Polarization" of Electrons

In section 17 we studied the mechanics of weak beta decays between quarks. Specifically (with the exception of beta decays between top and bottom quarks because of how the top quark "visits" the isospin-down well), we showed using Figure 8 how it is necessary for any quark undergoing weak beta decay to cross the barrier at the domain point $\phi_{1h} = v_{\Uparrow} - m_h c^2 = 120.9712 \pm 0.0002 \text{ GeV}$, and how this requires sufficient energy to clear the

$\sqrt[4]{V\left(\phi_{1h}\right)} \cong 240.37$ GeV peak at this domain point. Moreover, if an up or charm quark is to decay into any of the isospin-down quarks, it also needs additional energy to emerge out of the well at $\phi_{1h} = v_{\Uparrow} \cong 246.22$ GeV which bottoms out at a depth of $-\sqrt[4]{-V\left(\phi_{1h}\right)} \cong -514.89$ GeV. At (14.3) we tightened the Higgs boson mass to $m_h c^2 = 125.2485 \pm 0.0002$ GeV. We thereafter came understand how a small number of Higgs bosons may be involved in providing the energies needed to: a) facilitate excitation out of a well and clearance of the barrier between the wells, b) provide the mass also needed to excite a *W* boson with a mass of about 80 GeV out of the vacuum, and c) also to supply the mass, if needed, for any beta decay where the fermion needs to gain rest mass after the decay. And, we came to understand this activity as a form of vacuum fluctuation wherein energy is briefly withdrawn from the vacuum to facilitate beta decay, then returned to the vacuum after the decay event has completed, with all of this occurring inside a baryon containing very large internal energies arising from strong interactions between quarks.

Now, as if these high barriers and deep wells for quark beta decays are not large enough, the >100 TeV difference shown in Figure 14 between the well depth in (21.9c) and the barrier height in (21.9b) is in a whole other league, because this energy difference is on the order of 865 Higgs boson masses. This means that any time there is to be a beta decay between a neutrino and a charged lepton (with the exception of the tau lepton which "visits" the neutrino well analogously to the behavior of the top quark), it is necessary to raise over 50 TeV of energy to decay from a neutrino to a charged lepton. And it is necessary to raise over 100 TeV for the reverse-decay from a charged lepton to a neutrino. Moreover, importantly, we know that charged leptons can and do beta decay into neutrinos and vice versa all the time, and especially, that they apparently do so *spontaneously*. But, if it is necessary to amass over 50 TeV of energy for a neutrino to decay into a charged lepton, and over 100 TeV for the reverse reaction, then this clearly raises the question: From where is all this >50 TeV of energy acquired? And especially, where does this energy come from for *spontaneous* beta decays where we are not using particle accelerators, or nuclear reactors or weapons, or other human technology, to facilitate these decays?

Related to this, we know very well – dating all the way back to the late-19th century work of Henri Becquerel and Marie and Pierre Curie – that weak beta decays occur all the time in the natural world, without human technologies having to precipitate these decays. Most notably, as just stated, these decays appear to occur *spontaneously*, without *apparent* cause. So, for example, if we have a free neutron, we know that on average, this neutron will last for about 15 minutes before it decays into a proton. But we also know that this 15-minute period is a *mean time period*, and that there is a probabilistic spread about this 15-minute mean. Any given neutron might decay after 8 minutes, or 20 minutes, or any other period of time *t* in accordance with a temporal probability distribution for such decay. But the *causal* question as to <u>*why*</u> *any particular decay* takes 8 or 15 or 20 minutes or any other time to occur, has never been satisfactorily answered in the 120+ years since Becquerel and Curie's discovery. This leads us to pose two related questions: First, when a particular neutron or proton or atomic isotype of an atom has beta-decayed after some elapsed time *t*, what was the <u>*cause*</u> of *why* that decay happened exactly when it did? Second, what it is, exactly, that determines the 15-minute half-life of a free neutron, and the half-lives of various atomic isotopes which undergo beta decay? We begin here with the latter questions about lifetimes. Then, we later return to the questions about the >50 TeV energies.

For reasons that will momentarily become apparent, we start by considering the natural background flux of neutrinos observed in the physical world. For this, we refer to Figure 1 from [56], which is reproduced below:

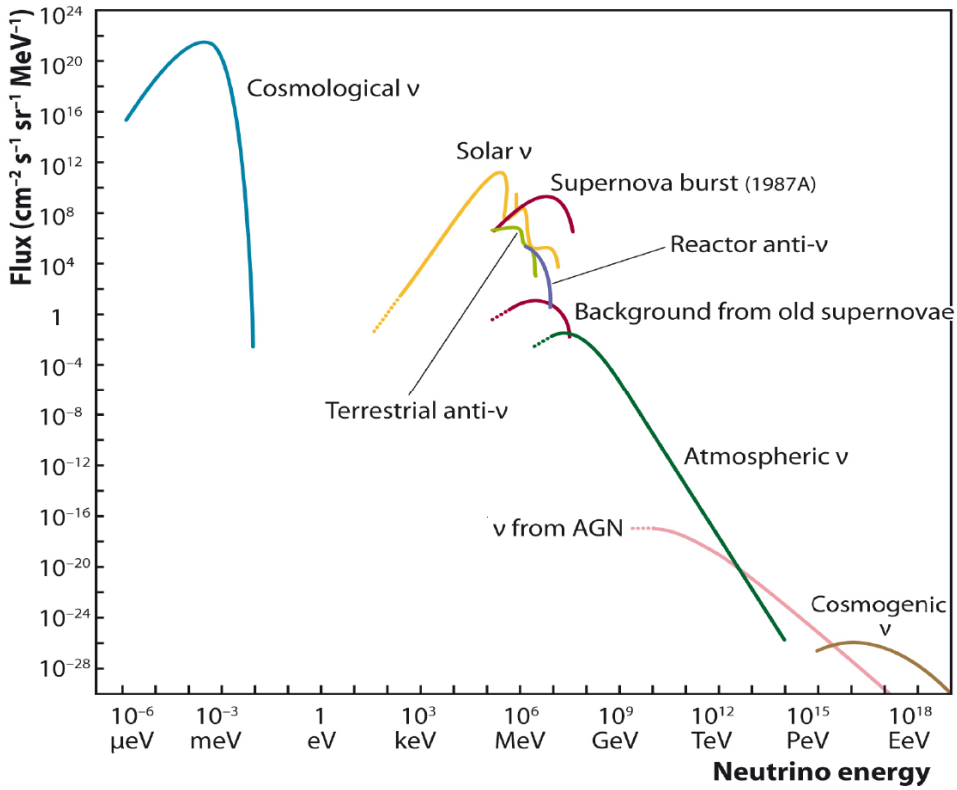

Figure 16: Measured and expected fluxes of natural and reactor neutrinos, reproduced from Figure 1 of [56]

The vertical axis above represents the *number flux* of neutrinos of various energies measured in neutrinos per cm^2 per second. This is plotted against the horizontal axis for neutrino kinetic energy. That the vertical axis represents a number flux is discerned from the MeV^{-1} in the vertical axis dimensionality which divides neutrino energy flux by energy to obtain number flux. In short, this is a plot for neutrino number flux as a function of neutrino kinetic energy. So, for example, for solar neutrinos, the vertical axis informs us that there is a peak of about 10^{11} solar neutrinos per cm^2 per second, while the horizontal axis informs us that these solar neutrinos have kinetic energies on the order of 10^5 to 10^6 eV. Of course, implicit in the above is that these measurements are taken at a particular locale in the universe, in this instance, at the surface of the earth. There is no reason to suppose that the exact same plot would be observed if measurements were taken say, on the surface of the planet mercury where the solar neutrino flux would certainly be greatly increased.

Of particular interest for the present discussion, however, are the much-more abundant cosmological neutrinos, often referred to as the cosmic neutrino background (CvB). For these, we are informed from Figure 16 that there is a peak flux of about 10^{21} neutrinos per cm^2 per second, and that these neutrinos have kinetic energies on the order of 10^{-3} eV = 1 meV or less. Given the individual neutrino rest masses ranging deduced in (17.4), namely 35.33 meV, 36.37 meV and 61.78 meV for the electron, mu and tau neutrinos respectively, we see that these cosmological

neutrinos have kinetic energies which are on the order of a few percent or less, of their rest masses. Thus, these are low energy, comparatively-nonrelativistic, neutrinos, travelling also at only a few percent of the speed of light, but which is still fast enough to cross the United States from east to west in under a second. Clearly, these CvB neutrinos comprise the vast abundance of neutrinos flowing through our everyday environment, by a factor of 10^{10} or more versus any of the other types of much-higher-energy neutrino shown.

From this, let us do a rough "back of the envelope" calculation. To start, recognizing that the charge radii of the proton and neutron are roughly $1\text{ f} = 10^{-15}\text{ m}$, let us regard 1 barn defined by $1\text{ b} = \left(10^{-14}\text{m}\right)^2 = \left(10\text{ f}\right)^2$ to be a very rough measure of the cross-sectional area for any stray particle to interact with a nucleus, being non-specific at the outset as to the particular particle or the particular nucleus. So, if we use barns rather than cm^2, the data just reviewed from Figure 16 tells us that there is a peak flux of about 10^{-3} CvB neutrinos per barn per second, or about 1 neutrino per barn per thousand seconds. And we may approximate 1000 seconds to fifteen minutes. So, as a rough calculation, we can say that in our day-to-day existence, one CvB neutrino flows through any one-barn cross sectional area approximately every 15 minutes.

Against this we also consider from, e.g. [55], that the mean lifetime of a free neutron is 880.2 ± 1.0 s, *which is also about 15 minutes.* So, the objective data tells us that every fifteen minutes, on average, one CvB neutrino basses through a 1 barn cross sectional area, and also, on average, a free neutron beta decays into a free proton. So, the question now presents itself: are these two seemingly-independent fifteen-minute natural episodes concurrent by sheer coincidence? Or, given the indispensable role of neutrinos in weak beta decay, is this no coincidence at all, but rather, a deep, heretofore unrecognized physical connection? In view of the fact that beta-decay appears to be spontaneous, and that the question of why a particular neutron happens to decay at any particular moment has never been explained since the days of Becquerel and Curie, we should at least consider the possibility that these two fifteen-minute natural episodes of free neutron decay and the passing of a neutrino through the "side of a barn" are in fact no coincidence at all. Doing so, we then we have the basis to introduce the following fundamental *hypothesis* as to *why individual beta decay events occur when they do*:

> *Neutrino Trigger Hypothesis: Semi-leptonic beta decays, such as that of a free neutron into a free proton with the concurrent decay of a neutrino into an electron, are in fact triggered when a stray neutrino – every fifteen minutes or so according to observed empirical data – randomly flows through an approximately 1 barn surface which contains the neutron, and thereby precipitates the latter's beta decay into a proton, with the triggering neutrino concurrently decaying into an electron.*

With this, the question of *why* a specific neutron decayed after a particular elapsed time *t* has a very intuitive and causal answer: With neutrinos randomly flying through space all the time and having the fluxes shown in Figure 16, the answer is that it took an elapsed time *t* for one of the CnB neutrinos permeating our natural environment to actually arrive and pass through the 1 barn cross section in which that neutron was centered, and accordingly, this is *why* it took the same time *t* for that neutron to decay. Recognizing that the $\nu n \to e^- p$ decay really takes place via the

quark decay $\nu d \rightarrow e^{-} u$, the Feynman diagram for this hypothesized neutrino-triggered beta decay is then shown in Figure 17 below:

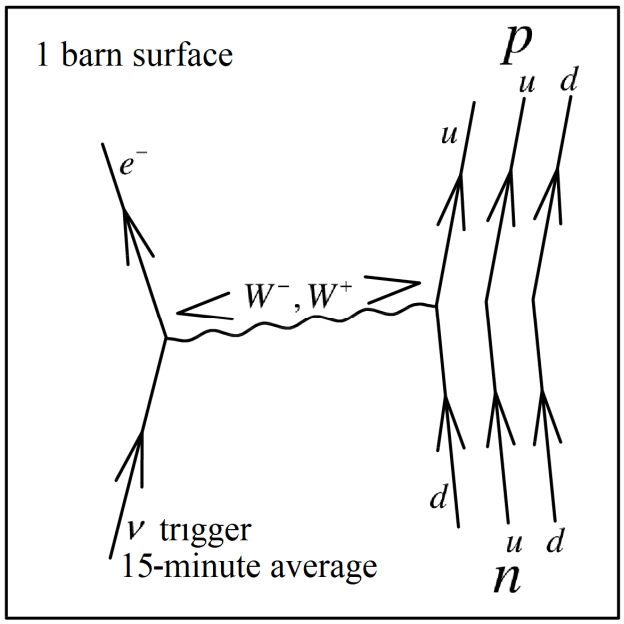

Figure 17: Free neutron beta decay with neutrino trigger

As is seen, this entails a CvB neutrino randomly entering the ~ 1 barn cross-sectional zone of the neutron following an average elapsed time of about 15 minutes, and getting close enough to the neutron to induce a *W* boson decay. It is possible for this decay to proceed in either direction, as illustrated. That is, left-to-right the decay can start with $\nu \rightarrow e^{-} W^{+}$ then finish with $W^{+} d \rightarrow u$. Or, right-to-left, it can start with $d \rightarrow W^{-} u$ then finish with $W^{-} \nu \rightarrow e^{-}$. The *W* boson which is the mediator of this interaction, has a very brief mean lifetime of about 3 x 10^{-25} s [57]. The net result following this very brief time period of 3 x 10^{-25} s, in either direction, is the beta decay $\nu n \rightarrow e^{-} p$ of the neutron and neutrino into an electron and a proton.

Now, Figure 17 uses the β^{-} decay reaction $\nu n \rightarrow e^{-} p$ as one very important example of our hypothesized neutrino-triggered beta decay. And it is based on observing from Figure 16 that one neutrino flows through 1 barn every 15 minutes or so, closely corresponding to the mean life of a free neutron. But if neutrinos are the trigger for free-neutron beta decays, then they should likewise be the trigger for other beta decays occurring in complex nuclides and atoms which contain multiple protons and neutrons and have numerous isotopes. And, of course, these other beta decays have mean lifetimes which are not 15 minutes apart, but which are variable depending on the specific isotopes being considered. For this neutrino trigger hypothesis to stand up, therefore, we must tackle the further question whether these other β^{-} decays can be explained in this way, and what additional factors may come into play. Moreover, we need to tackle the question whether and how β^{+} decays are triggered, suspecting based on similar principles that we would have to utilize antineutrinos as the trigger.

Because all atoms contain at least one proton, any beta decays of any atom from hydrogen on up will occur inside the nucleus of an atom containing at least one electron. Of course, electrons in an atom do not "orbit" in the sense of the early atomic models of a planet traversing the sun, but rather, they form a "probability density cloud" about the nucleus. The Bohr radius of a hydrogen atom is on the order of 5.29 x 10^{-11} m, which we roughly approximate to an atomic *diameter* of 10^{-10} m. So, whereas a barn with $1\text{ b} = \left(10^{-14}\text{m}\right)^2$ defines a measurement standard for *nuclear* cross sections, let us now define a "Bohr barn," abbreviated bb, such that $1\text{ bb} = \left(10^{-10}\text{m}\right)^2$. This may be thought of as a measurement standard for *atomic* cross sections, and it is larger than an ordinary barn by a factor of 10^8. So, because one neutrino passes through a nuclear barn every 10^3 seconds (approximately 15 minutes) as found earlier, one neutrino will bass through an atomic Bohr barn every 10^{-5} seconds, which is .01 milliseconds (ms).

As against this bb measure, let us review the half-lives of various isotopes which decay through pure β^- or β^+ decay, and not by α or γ decay or by merely jettisoning neutrons or protons. We start from the very valuable Wikipedia Table of Nuclides [58] and use this to link over to the isotopes for various atoms. For any individual atom, at the bottom of the summary box on the right side of the screen there is a review of the "main isotopes" of the given atom as well as link to *all* the "isotopes of" that atom. In this way we find the following sampling of β^- decay data (without error bars), for the light nuclides from hydrogen (atomic number Z=1) through sodium (Z=11): For ^{3}H, i.e. tritium, the half-life is about 12.32 y. For helium, for ^{6}He the half-life is about 806.7 ms, while for ^{8}He it is about 119.0 ms. For lithium, for ^{8}Li, ^{9}Li and ^{11}Li, the half-lives are 840.3 ms. 178.3 ms and 8.75 ms respectively. For beryllium, for ^{10}Be, ^{11}Be, ^{12}Be and ^{14}Be, the half-lives are 1.39×106 y, 13.81 s, 21.49 ms and 4.84 ms respectively. For boron, the range is from a high of 20.20 ms for ^{12}B, down to 2.92 ms for ^{19}B, with consistent serial descent. For carbon, we of course have 5,730 years for ^{14}C used in radioactive dating. Thereafter, the β^- half-lives range serially downward from a high of 2.449 s for ^{15}C to 6.2 ms for ^{16}C. For oxygen, there is a serial reduction from 26.464 s to 65 ms from ^{19}O to ^{24}O. For Fluorine, the range is serially-downward from 11.163 s to 2.6 ms from ^{20}F to ^{29}F, with very-mild exception at ^{21}F and ^{22}F which may be attributed simply to the growing complexity of the nuclide. For neon there is serial descent from 37.24 s for ^{23}Ne to 3.5 ms for ^{32}Ne, with a single exception at ^{24}Ne. And going from the n=2 shell to the n=3 shell (principal quantum umber, second to third row in the periodic table) to ensure the pattern holds, for sodium we again see a range serially diminishing from 14.9590 h for ^{24}Na, down to 1.5 ms for ^{35}Na, with mild exceptions attributable to nuclide complexity.

The same review for β^+ decay evidences the following: for hydrogen there is no β^+ decay channel. Fer helium there is a β^+ decay channel for ^{2}He, but the greatly-favored channel by >99.99% is to jettison a proton, with a half-life under 10^{-9} s. The half-life for the <0.01% β^+ decay channel is not clearly shown in this data, or any other data that the author could uncover. Likewise, neither lithium nor beryllium have any β^+ channels. So, we begin with boron, which is the first nucleus with a clear β^+ channel, namely, ^{8}B with a half-life of 770 ms. Turning to carbon, for ^{11}C, ^{10}C and ^{9}C, the respective β^+ decay half-lives are 20.334 min, 19.290 s and 126.5

ms, respectively. For nitrogen, we have ^{13}N and ^{12}N with the respective half-lives of 9.965 min and 11.000 ms. For oxygen we have ^{15}O, ^{14}O and ^{13}O with respective half-lives of 122.24 s, 70.598 s and 8.58 ms. For fluorine the two channels are for ^{18}F and ^{17}F with respective 109.771 min and 64.49 s half-lives. For neon the three isotopes with β^+ channels are ^{19}Ne, ^{18}Ne and ^{17}Ne with 17.296 s, 1.672 s and 109.2 ms in series. Finally, moving to the next shell, for sodium there are three isotopes with β^+ channels, namely, ^{22}Na, ^{21}Na and ^{20}Na with respective 2.6027 y, 22.49 s and 447.9 ms half-lives. Note that ^{22}Na is the first isotope with a half-life measured in times as long as years.

There are three very striking and consistent patterns revealed by the above light nuclide data. First, while the beta-decay half lives in a few cases run as high as years, in most cases they run in minutes or seconds and at bottom, milliseconds. The very shortest half-life in the data above was 1.5 ms for the β^- decay of ^{35}Na. And, studying higher up the periodic table, there does not appear to be *any* beta decay with a half-life less than 1 ms, for *any* isotope of *any* atom. Of course, there are many decays with half-lives shorter than 1 ms, see [59]. But none of these are *beta* decays, which informs us that beta decay lifetimes are comparatively long relative to other types of decays such as alpha decays and neutron or proton emission. Second, for β^- decay, for the light nuclides, there is a consistent and unbroken correlation whereby whenever the number of neutrons is *increased* for an atom of a given atomic number, the β^- half-life is decreased. Third, for β^+ decay, there is a likewise consistent and unbroken correlation whereby whenever the number of neutrons is *decreased* for an atom of a given atomic number, the β^+ half-life is decreased. That is, working from stable atoms in the middle of neutron-rich or neutron-poor isotopes, the more an isotope is either neutron-rich or neutron-poor, the shorter will be its half-life for beta decay.

Now, to be sure, as was already seen starting with fluorine, these correlations do get partially-broken for heavier nuclides. For example, the β^- correlation is broken by Z=12 magnesium, wherein ^{27}Mg, ^{28}Mg and ^{29}Mg have respective half-lives of 9.458 min, then a longer 20.915 h, then 1.30 s which is shorter and returns to pattern. And, the β^+ correlation is first broken by Z=17 chlorine, wherein ^{34}Cl, ^{33}Cl and ^{32}Cl have respective half-lives of 1.5264 s, then a longer 2.511 s, then 298 ms, which is shorter and returns to pattern. Given that this the correlation between isotopes becoming either more neutron-rich or neutron-poor and a diminishment of half-life is an unbroken pattern for light nuclides up to magnesium, it is fair to regard breaks in this pattern for the heavier nuclides as being less a break in pattern, and more as a masking of pattern by the more-complex atomic and nuclear shell structures. The factors involved in this will become clearer momentarily.

Now we come to the key question: In view of this data, how do we apply the hypothesis that neutrinos are the trigger for the β^- decay of a free neutron into a free proton with the trigger neutrino also decaying into an electron, to *any and all* beta decays, both β^- and β^+, in *any and all* atoms and atomic isotopes? In short, is it possible to understand <u>*all*</u> *beta decay events* – which randomly occur with known half-lives – as occurring at precise particular times t when a CvB

neutrino (or an antineutrino) passes close enough to a neutron (or a proton) in an atomic nucleus to become the *triggering cause* of that decay event?

We have reviewed that the observed half-lives for beta decay are greater than 1 ms for *all nuclides in the periodic table*. As reviewed, 1 ms corresponds to a cross section of .01 bb or less using our new Bohr barn yardstick, i.e., a cross section that is less than 1% of the cross section for a hydrogen atom. But even so, a 1 ms half-life corresponding to 10^{-2} bb thus a 10^{6} b cross section for nuclear events is much closer to atomic shell rather than to nuclear cross sections. And yet, it is a neutron or a proton inside the nucleus which decays in all of the nuclear isotope data just reviewed. Therefore, for short-lived beta-decays closer to 1 ms than to seconds or years to be triggered by a neutrino or antineutrino, there must be some mechanism which ensnares a neutrino or antineutrino entering the atomic shells at 10^{-2} bb a.k.a. 10^{6} b and so still at some distance from the nucleus, and nevertheless guides that neutrino or antineutrino through the atomic shells to find a neutron or proton within the nucleus and trigger that neutron or proton to β^- or β^+ decay.

This brings us to the *Z* boson for the weak neutral current, which, aside from gravitation, is the *only* means by which a neutrino can interact with an electron or a quark *while each maintains its identity*. Specifically, if a low-energy CvB neutrino is going to enter the electron shells of an atom inside a 10^{6} b cross section about the nucleus and end up beta-decaying with that nucleus sitting within a 1 b-or-less cross section, then the neutrino will need to be attracted to the nucleus through the weak neutral current *Z* boson, analogously to how electrons are attracted to the up and down quarks inside of protons via electromagnetic interactions mediated by photons. But there are two important differences: First, electromagnetism is an inverse-square interaction because the mediating photos are massless and so have unlimited range, which the electroweak neutral current integration has a very short range because of the very short lifetime on the order of 3 x 10^{-25} s for the *Z* boson. Even if travelling close to the speed of light $c = 299792458$ m/s, exactly [21], the mean range of this boson is in the order of 10^{-16} m = .1 f, which, squared, corresponds to a .01 b cross section. Second, electromagnetism is a chiral-symmetric interaction for which the left- and right-chiral components of fermions each have the same charge strength *Q*, whereas V-A weak interactions are distinctly non-chiral. Specifically, the third component of the weak isospin $I_3\left(f_R\right) = 0$ for the right-chiral projections $f_R = \frac{1}{2}\left(1+\gamma^5\right) f$ of all fermions *f*. Indeed, a very central finding in Part I of this paper as reviewed in section 9 is that the Dirac γ_5 used to project left-and right-chiral components out of a fermion is the generator of the fifth Kaluza-Klein dimension in exactly the same way that the first four γ_μ generate the one time and the three space dimensions of ordinary spacetime, all with a fifth dimension that is timelike not spacelike, via the five-dimensional relation $\eta_{\mathrm{MN}} = \frac{1}{2}\{\gamma_{\mathrm{M}}, \gamma_{\mathrm{N}}\}$ for the flat spacetime Minkowski metric tensor η_{MN}.

Because our interest is in the attraction of a neutrino into a nucleus (and of course, any offsetting repulsive forces), let us start by considering electromagnetic attraction and repulsion which does not introduce the complexities of either limited range or chiral non-symmetry. The charges of an electron and proton, respectively, are $Q(e) = -1$ and $Q(p) = +1$, so that when an electron interacts with another electron we have $Q(e)Q(e) = +1$ which is repulsive while when an electron interacts with a proton we have $Q(e)Q(p) = -1$, with repulsion versus attraction

determined by the sign. A slightly-more complicated way of saying the same thing – which provides a baseline for considering attraction and repulsion under the weak neutral current interaction – is to say the following: When two electrons interact, the invariant amplitude is proportional ($\propto$) to $\mathcal{M}_{em}(ee\to ee)\propto\left(Q(e)\bar{e}\gamma^{\mu}e\right)\left(Q(e)\bar{e}\gamma_{\mu}e\right)=+\left(\bar{e}\gamma^{\mu}e\right)\left(\bar{e}\gamma_{\mu}e\right)$, while when an electron and a proton interact, $\mathcal{M}_{em}(ep\to ep)\propto Q(e)Q(p)\left(\bar{e}\gamma^{\mu}e\right)\left(\bar{p}\gamma_{\mu}p\right)=-\left(\bar{e}\gamma^{\mu}e\right)\left(\bar{p}\gamma_{\mu}p\right)$. The overall positive sign for $\mathcal{M}_{em}(ee\to ee)$ indicates repulsion and the overall negative sign for $\mathcal{M}_{em}(ep\to ep)$ indicates attraction. Now let's turn to neutral currents mediated by *Z* bosons.

The weak neutral current, of course, has the form $J^{\mu}_{NC}=J^{\mu}_{3}-J^{\mu}_{em}\sin^2\theta_W$, where $\sin^2\theta_W=0.23155(4)$ is the effective weak mixing angle [21]. As with any other "charge," this means that the "*Z* charge" of a particle is $Z=I_3-Q\sin^2\theta_W$, where I_3 is the third component of the weak isospin and *Q* is the electrical charge of that particle. Weak interactions, however, are not chiral symmetric. For a given fermion state, $f_R=\frac{1}{2}\left(1+\gamma^5\right)f$ and $f_L=\frac{1}{2}\left(1-\gamma^5\right)f$, thus $f=f_L+f_R$. Also, $I_3(f_R)=0$ for all fermions. So, separating left- and right-chiral projections, and using the center-value $\sin^2\theta_W=0.23155$, the pertinent electroweak charge quantum numbers for each lepton chiral projection, which apply to all three generations, are as follows:

$$\begin{aligned}
\nu_L&=\left|I_3=+\tfrac{1}{2},\quad Q=0,\quad Z=+\tfrac{1}{2}\right\rangle\\
\nu_R&=\left|I_3=0,\quad Q=0,\quad Z=0\right\rangle\\
e_L&=\left|I_3=-\tfrac{1}{2},\quad Q=-1,\quad Z=-\tfrac{1}{2}+\sin^2\theta_W=-0.26845\right\rangle\\
e_R&=\left|I_3=0,\quad Q=-1,\quad Z=\sin^2\theta_W=+0.23155\right\rangle
\end{aligned}\qquad(22.2a)$$

Likewise, for the quarks, also for all generations, the chiral charge quantum numbers are:

$$\begin{aligned}
u_L&=\left|I_3=+\tfrac{1}{2},\quad Q=+\tfrac{2}{3},\quad Z=+\tfrac{1}{2}-\tfrac{2}{3}\sin^2\theta_W=+0.34563\right\rangle\\
u_R&=\left|I_3=0,\quad Q=+\tfrac{2}{3},\quad Z=-\tfrac{2}{3}\sin^2\theta_W=-0.15437\right\rangle\\
d_L&=\left|I_3=-\tfrac{1}{2},\quad Q=-\tfrac{1}{3},\quad Z=-\tfrac{1}{2}+\tfrac{1}{3}\sin^2\theta_W=-0.42282\right\rangle\\
d_R&=\left|I_3=0,\quad Q=-\tfrac{1}{3},\quad Z=+\tfrac{1}{3}\sin^2\theta_W=+0.07718\right\rangle
\end{aligned}\qquad(22.2b)$$

For the *Z* charge, it is then customary to define separate vertex and axal couplings according to $c_V(f)\equiv Z_L+Z_R$ and $c_A(f)\equiv Z_L-Z_R$, that is, the sum and difference of the charge contributions from each of the chiral parts of the fermion. Given that $I_3(f_R)=0$ for all fermions, and that $Z=I_3-Q\sin^2\theta_W$ generally so that $Z_L=I_{3L}-Q\sin^2\theta_W$ and $Z_R=I_{3R}-Q\sin^2\theta_W=-Q\sin^2\theta_W$, this means:

$$\begin{aligned} c_V(f) &\equiv Z_L + Z_R = I_{3L} - 2Q\sin^2\theta_W \\ c_A(f) &\equiv Z_L - Z_R = I_{3L} \end{aligned}, \tag{22.3}$$

as is well-known. Therefore, from (22.2) we may deduce that:

$$\begin{aligned} c_V(\nu) &= I_{3L}(\nu) - 2Q(\nu)\sin^2\theta_W = +\tfrac{1}{2}; & c_A(\nu) &= I_{3L}(\nu) = +\tfrac{1}{2} \\ c_V(e) &= I_{3L}(e) - 2Q(e)\sin^2\theta_W = -\tfrac{1}{2} + 2\sin^2\theta_W = -0.03690; & c_A(e) &= I_{3L}(e) = -\tfrac{1}{2} \\ c_V(u) &= I_{3L}(u) - 2Q(u)\sin^2\theta_W = +\tfrac{1}{2} - \tfrac{4}{3}\sin^2\theta_W = +0.19127; & c_A(u) &= I_{3L}(u) = +\tfrac{1}{2} \\ c_V(d) &= I_{3L}(d) - 2Q(d)\sin^2\theta_W = -\tfrac{1}{2} + \tfrac{2}{3}\sin^2\theta_W = -0.34563; & c_A(d) &= I_{3L}(d) = -\tfrac{1}{2} \end{aligned}, \tag{22.4}$$

Finally, a neutron contains two down and one up quarks, while a proton contains two up quarks and one down quark. If we use the couplings in (22.4) to determine the couplings for the proton and neutron, then by simple addition we obtain:

$$\begin{aligned} c_V(p) &= c_V(uud) = +\tfrac{1}{2} - 2\sin^2\theta_W = +0.03690 & c_A(p) &= c_A(uud) = +\tfrac{1}{2} \\ c_V(n) &= c_V(udd) = -\tfrac{1}{2} & c_A(n) &= c_A(udd) = -\tfrac{1}{2} \end{aligned}. \tag{22.5}$$

It will be noted that the protons couplings are equal in magnitude and opposite in sign to those of the electron, and the neutron couplings are likely opposite the neutrino couplings.

Now, it is customary to also define $c_R(f) \equiv c_V(f) - c_A(f)$ and $c_L(f) \equiv c_V(f) + c_A(f)$ for couplings of the right- and left-chiral projections of a fermion. Thus, for elementary fermions:

$$\begin{aligned} c_R(\nu) &= 0; & c_L(\nu) &= +1 \\ c_R(e) &= 2\sin^2\theta_W = +0.46310; & c_L(e) &= -1 + 2\sin^2\theta_W = -0.53690 \\ c_R(u) &= -\tfrac{4}{3}\sin^2\theta_W = -0.30873; & c_L(u) &= +1 - \tfrac{4}{3}\sin^2\theta_W = +0.69127 \\ c_R(d) &= +\tfrac{2}{3}\sin^2\theta_W = +0.15437; & c_L(d) &= -1 + \tfrac{2}{3}\sin^2\theta_W = -0.84563 \end{aligned}, \tag{22.6}$$

and for the proton and neutron:

$$\begin{aligned} c_R(p) &= -2\sin^2\theta_W = -0.46310; & c_L(p) &= +1 - 2\sin^2\theta_W = +0.53690 \\ c_R(n) &= 0; & c_L(n) &= -1 \end{aligned}. \tag{22.7}$$

Now let's return to how the neutrino is attracted to or repelled by other fermions. In general, $\mathcal{M}_Z(\nu f \to \nu f) \propto \left[\overline{\nu}\gamma^\mu\left(c_V(\nu) - c_A(\nu)\gamma^5\right)\nu\right]\left[\overline{f}\gamma_\mu\left(c_V(f) - c_A(f)\gamma^5\right)f\right]$ represents the invariant amplitude for a neutrino ν interacting with a second fermion f via the weak neutral current Z boson, where we neglect the fermion masses in the propagators and use a proportionality because all we are interested in is the overall sign. First, we may construct the identity:

$$c_V(f)-c_A(f)\gamma^5=\left(c_V(f)-c_A(f)\right)\tfrac{1}{2}\left(1+\gamma^5\right)+\left(c_V(f)+c_A(f)\right)\tfrac{1}{2}\left(1-\gamma^5\right), \tag{22.8}$$

see [13.59] in [20]. Then, Using $c_R\equiv c_V-c_A$ and $c_L\equiv c_V+c_A$ as laid out above, as well as the well-known relations $f_R=\tfrac{1}{2}\left(1+\gamma^5\right)f$, $\overline{f}_R=\overline{f}\,\tfrac{1}{2}\left(1-\gamma^5\right)$ for right- and $f_L=\tfrac{1}{2}\left(1-\gamma^5\right)f$, $\overline{f}_L=\overline{f}\,\tfrac{1}{2}\left(1+\gamma^5\right)$ for left-chiral projections, together with $\gamma^\mu\gamma^5=-\gamma^5\gamma^\mu$ and the identities $\tfrac{1}{2}\left(1+\gamma^5\right)=\tfrac{1}{2}\left(1+\gamma^5\right)\tfrac{1}{2}\left(1+\gamma^5\right)$ and $\tfrac{1}{2}\left(1-\gamma^5\right)=\tfrac{1}{2}\left(1-\gamma^5\right)\tfrac{1}{2}\left(1-\gamma^5\right)$, we may use (22.8) to rewrite the foregoing invariant amplitude as:

$$\begin{aligned}\mathcal{M}_Z(\nu f\to\nu f)&\propto\left[\overline{\nu}\gamma^\mu\left(c_V(\nu)-c_A(\nu)\gamma^5\right)\nu\right]\left[\overline{f}\gamma_\mu\left(c_V(f)-c_A(f)\gamma^5\right)f\right]\\&=\left[c_R(\nu)\overline{\nu}_R\gamma^\mu\nu_R+c_L(\nu)\overline{\nu}_L\gamma^\mu\nu_L\right]\left[c_R(f)\overline{f}_R\gamma_\mu f_R+c_L(f)\overline{f}_L\gamma_\mu f_L\right]\end{aligned}, \tag{22.9}$$

Using $c_R(\nu)=0$ and $c_L(\nu)=+1$ from (22.6) then distributing $\overline{\nu}_L\gamma^\mu\nu_L$, this further simplifies to:

$$\mathcal{M}_Z(\nu f\to\nu f)\propto\left(\overline{\nu}_L\gamma^\mu\nu_L\right)c_R(f)\left(\overline{f}_R\gamma_\mu f_R\right)+\left(\overline{\nu}_L\gamma^\mu\nu_L\right)c_L(f)\left(\overline{f}_L\gamma_\mu f_L\right). \tag{22.10}$$

Note: the reason $\overline{\nu}_R\gamma^\mu\nu_R$ drops out leaving only $\overline{\nu}_L\gamma^\mu\nu_L$ is not because the neutrino is massless (which is isn't) and not because the neutrino is only left-chiral (which it also is not because it has a mass), but merely because $c_R(\nu)=0$ as a consequence of the V-A character of weak interactions which entirely eliminates the $\overline{\nu}_R\gamma^\mu\nu_R$ from any amplitudes containing the neutrino.

For a neutrino interacting with an electron, and also with another neutrino, we use (22.6) in (22.10) to obtain:

$$\begin{aligned}\mathcal{M}_Z(\nu e\to\nu e)&\propto+0.46310\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{e}_R\gamma_\mu e_R\right)-0.53690\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{e}_L\gamma_\mu e_L\right)\\\mathcal{M}_Z(\nu\nu\to\nu\nu)&\propto+\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{\nu}_L\gamma_\mu\nu_L\right)\end{aligned}. \tag{22.11}$$

For a neutrino interacting with up and down quarks we likewise obtain:

$$\begin{aligned}\mathcal{M}_Z(\nu u\to\nu u)&\propto-0.30873\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{u}_R\gamma_\mu u_R\right)+0.69127\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{u}_L\gamma_\mu u_L\right)\\\mathcal{M}_Z(\nu d\to\nu d)&\propto+0.15437\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{d}_R\gamma_\mu d_R\right)-0.84563\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{d}_L\gamma_\mu d_L\right)\end{aligned}. \tag{22.12}$$

Finally, for a neutrino interacting with a proton and neutron we use (22.7) in (22.10) to find:

$$\begin{aligned}\mathcal{M}^{NC}(\nu p\to\nu p)&\propto-0.46310\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{p}_R\gamma_\mu p_R\right)+0.53690\left(\overline{\nu}_L\gamma^\mu\nu_L\right)c_L(p)\left(\overline{p}_L\gamma_\mu p_L\right)\\\mathcal{M}^{NC}(\nu n\to\nu n)&\propto-\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{n}_L\gamma_\mu n_L\right)\end{aligned}. \tag{22.13}$$

Now, in all of (22.11) through (22.13), a plus sign is indicative of neutral current repulsion and a minus sign is indicative of neutral current attraction. Again, this is just as how $\mathcal{M}_{em}\left(ee\to ee\right)\propto+\left(\bar{e}\gamma^{\mu}e\right)\left(\bar{e}\gamma_{\mu}e\right)$ and $\mathcal{M}_{em}\left(ep\to ep\right)\propto-\left(\bar{e}\gamma^{\mu}e\right)\left(\bar{p}\gamma_{\mu}p\right)$ tell us that for electromagnetic interactions, electrons repel other electrons but attract protons. But what we see above for *Z*-mediated interactions is that the neutrino will have different interactions with the superposed chiral spinors in $f=f_L+f_R$. In (22.11), owing to the plus sign in $+0.46310\left(\bar{\nu}_L\gamma^{\mu}\nu_L\right)\left(\bar{e}_R\gamma_{\mu}e_R\right)$ and the minus sign in $-0.53690\left(\bar{\nu}_L\gamma^{\mu}\nu_L\right)\left(\bar{e}_L\gamma_{\mu}e_L\right)$, we learn that the neutrino repels with the right-chiral protections in $\bar{e}_R\gamma_{\mu}e_R$ and attracts with the left-chiral projections in $\bar{e}_L\gamma_{\mu}e_L$. In other words, one consequence of the chiral asymmetry of weak interactions is that when a neutrino and electron get close-enough to interact via a limited-range *Z* boson, *the neutrino will attract the left-chiral components and repel the right-chiral components of the electron*. Overall, there is a small weighting favoring attraction over repulsion by 0.53690 versus 0.46310, so the *net interaction* is attractive. Note the origin of these numbers in (22.6). We also see in (22.11) that neutrinos will repel other neutrinos.

From (22.12) we have a similar chiral interaction asymmetry: For neutrino / up quark interactions, there is a -0.30873 factor for right-chiral attraction weighted against a $+0.69127$ factor for left-chiral repulsion, so that in net, neutrinos and up quarks repel. For the down quark interaction, there is $+0.15437$ for right-component repulsion versus -0.84563 for left-component attraction, with the net result being that neutrinos and down quarks attract. And from (22.13) we see that the neutrino mildly repels with the proton by an amount similarly weighted by 0.53690 versus 0.46310, with behavior opposite that of neutrino / electron interaction. Finally, from (22.13), the interaction between the neutrino and the neutron is exclusively, strongly-attractive.

Given $c_R\left(\nu\right)=0$ and $c_L\left(\nu\right)=+1$ from (22.6), if neutrinos are Dirac not Majorana fermions, then we expect the charge quantum numbers of the antineutrinos to be opposite those of neutrinos. Thus, specifically, we expect that $c_R\left(\bar{\nu}\right)=0$ and $c_L\left(\bar{\nu}\right)=-1$. (To be formally-precise with notation, rather than as the oft-employed $\bar{\nu}$, we will hereafter designate the antineutrino as $\nu_C=C\bar{\nu}^T$ using the Dirac conjugation operator $C=i\gamma^2\gamma^0$.) This means that antineutrinos will attract and repel other fermions via the weak neutral current in a manner opposite what is shown in (22.11) through (22.13) for neutrinos. This also all means that neutrinos strongly repel other neutrinos, and strongly attract antineutrinos. The latter can pull in-range neutrinos and antineutrinos together for annihilation. The same applies for other fermions as well.

All of this provides the basis for understanding how β^- decay is triggered by CvB neutrinos as shown in Figure 17, not only for free neutrons, but for all atomic isotopes in the periodic table which undergo β^- decay. And it also provides the basis for understanding how β^+ decay is triggered by CvB antineutrinos (presumed to have a similar number flux to neutrinos) for all isotopes which undergo β^+ decay. First, let's return to free-neutron decay.

In our earlier "back of the envelope" calculation, we found that one CvB neutrino flows through a $1\text{ b} = \left(10^{-14}\text{m}\right)^2 = \left(10\text{ f}\right)^2$ cross section every fifteen minutes or so, which happens to correspond to the mean lifetime of a free neutron, and which caused us to suspect a non-coincidental tie between these two seemingly-independent pieces of data. However, the radii of the neutron and the proton – irrespective of the precise details of how these are determined – are approximately 1 f, and thus their diameters are about 2 f. And so, their cross sections based on their diameters are about $\left(2\text{ f}\right)^2 = .04\text{ b}$. So, the 1 b cross section that we used to arrive at a fifteen-minute mean life is about 25 times as large as the actual cross of a physical neutrino. Thus, as we fine tune this rough calculation, the question arises how to account for this factor of 25 discrepancy.

Now, having reviewed *Z* boson-mediated attraction and repulsion, this is accounted for by the minus sign in $\mathcal{M}^{NC}\left(\nu n \to \nu n\right) \propto -\left(\bar{\nu}_L \gamma^\mu \nu_L\right)\left(\bar{n}_L \gamma_\mu n_L\right)$ in (22.13): If a CvB neutrino flows into the 1 b cross section centered about the neutrino, it is now close enough to be strongly-attracted to the neutrino via a sufficiently long-lived *Z* boson, or, given the 3 x 10^{-25} s mean life thus .1 f range of even an extreme-relativistic *Z* boson, more-likely through a $Z \to \bar{f} f \to Z \to \bar{f} f \to Z...$ chain of *Z* bosons with intermediate $\bar{f} f$ pairs extending the range of the *Z* boson. In short, the *Z* boson exchanges, likely with intermediate virtual $\bar{f} f$ pairs to extend range, operate as a "trap" to "ensnare" this non-relativistic neutrino and attract it toward the nucleus, until it is finally close enough to β^- decay via $\nu d \to W^+ ed \to eu$ thus an overall decay $\nu n \to W^+ en \to ep$.

Next, let's proceed to beta decays of atomic isotopes. Here, any stray CvB neutrino which passes into the atomic shell structure of an atom – even through a cross section of 10^6 b which corresponds to the shortest 1 ms half-lives in the periodic table – will be attracted *by the electrons in that atom*, as a whole, albeit mildly, toward the center of that atom, by the electroweak neutral current force, via the overall negative sign in $\mathcal{M}_Z\left(\nu e \to \nu e\right)$ from (22.11), with left-chiral attraction outweighing right-chiral repulsion as regards the spinors in $e = e_L + e_R$. So, for example, when a neutrino enters the atomic shells to the "left" of the nucleus, the preponderance of the electron cloud will be to the "right" of the neutrino, the preponderance of probabilities for the electron locations will thereby also be to the right of the neutrino, and by the weak neutral current attraction of the neutrino to the electrons as a whole, the neutrino will be attracted to the right, toward the nucleus. This is further made possible by the very fact that these CvB neutrinos have such low kinetic energies. As with the free neutron interactions just reviewed, the range of the *Z* boson can be extended through a $Z \to \bar{f} f \to Z \to \bar{f} f \to Z...$ chain with virtual fermion pairs.

In this way, the electron cloud interacting with the neutrino via the *Z* boson interactions acts as a sort of a "spider's web," snaring the neutrino into the atom and attracting it toward the nucleus at the center of the atom. This works, in part, because the cosmological neutrinos in Figure 16 have kinetic energies on the order of 1 meV and rest mass energy equivalents from about 35 meV to 62 meV, which means that these neutrinos are travelling slowly enough to have their trajectories changed by the neutral current attractions of the electrons. Moreover, this initial "snaring" of the neutrino is *not done by the neutron*, but rather, is *done by the electrons*. So as

long as the neutrino and one of the electrons are in a close-enough range to one another, they can interact without changing flavor via a neutral *Z* boson or a $Z \to \overline{f} f \to Z \to \overline{f} f \to Z...$ chain.

Then, after the neutrino finally draws close to the nucleus through this web of *Z* interactions with the electrons, the nucleons themselves finally come into play. For a nucleus with a rough balance of protons and neutrons, (22.13) makes clear that neutrinos will be attracted toward neutrons much-more-strongly than they will be repelled by protons. That is, overall, neutrinos are net attracted toward nuclei via the weak neutral current interaction. The end result, following very large numbers of *Z* boson for boson chain exchanges, will find the neutrino drawn closely-enough within range of one of the neutrons (and specifically within range of a down quark inside one of the neutrons), so that the neutrino may undergo a $\nu \to e^- W^+$ decay, followed about 3 x 10^{-25} s later by a $W^+ d \to u$ decay. The net result is that the β^- decay reaction $\nu n \to e^- p$ is now complete, and has been triggered by the low-energy neutrino that was initially ensnared by the electron shells from a much-larger atomic-scale cross section of about 10^{-2} bb = 10^6 b.

Very importantly, the foregoing fully explains the data reviewed earlier, whereby there is a consistent and unbroken correlation for light nuclides such that whenever the number of neutrons is increased for an atom of a given atomic number, the β^- half-life is decreased: When we add neutrons to a nucleus, we increase the attraction of electron-ensnared neutrinos toward the nucleus. This enable neutrinos from a larger cross section to reach the nucleus, which means that there are more neutrinos per unit of time available to decay a neutron into a proton, which means that the lifetime between beta decays is reduced. So, the very short β^- half-lives of the neutron-rich isotopes of any atom are directly reflective of the fact that these atoms have more neutrons available to attract snared neutrinos without them exiting the atomic shells, casting a wider net in the nearby space, and thereby reducing the elapsed time until a β^- decay event occurs.

Now we come to β^+ decays of neutron-poor isotopes, where a proton gets decayed into a neutron. Here too, with the half-lives being no less than milliseconds but ranging up to seconds and hours and years, we also anticipate a triggering mechanism. But given the quantum numbers that need to be conserved here, we postulate that *antineutrinos* comprise the trigger mechanism, and adopt the prevailing view (e.g., [60]) that the density and flux of CvB antineutrinos is virtually the same as that of neutrinos, which means based on Figure 16 that 1 CvB antineutrino will also pass through a 1 b cross section approximately every fifteen minutes. If these antineutrinos are the trigger for β^+ decay, then once an antineutrino gets in range of a proton, the reaction is either $\nu_C \to e^+ W^-$ followed momentarily thereafter by $W^- p \to n$, or the reverse-ordered $p \to n W^+$ followed by $\nu_C W^+ \to e^+$. The positron can then annihilate one of the electrons. The next result in either case is $\nu_C p \to e^+ n$, with the original electron included, $\nu_C p e^- \to n$. Just as with β^- decay, we expect that the electrons in the atomic shells will first use *Z* boson interactions to ensnare the antineutrino to draw it toward the nucleus, until it gets close enough for beta decay. But here, we encounter a bit of difficulty that we must sort out:

In (22.11) we obtained the neutral current cross section $\mathcal{M}_Z(\nu e \to \nu e)$ for neutrino electron interactions, and found that on balance this interaction is slightly-weighted toward attraction, with a coefficient -0.53690 for attraction of the neutrino to left-chiral electron components weighted against the mildly-smaller coefficient $+0.46310$ for repulsion of the neutrino from the right-chiral electron components. In general, Dirac fermions ψ are related to their antifermion counterparts ψ_C by $\psi_C = C\overline{\psi}^T$. With $C = i\gamma^2\gamma^0$ and $\overline{\psi} = \psi^\dagger\gamma^0$ this means $\psi_C = i\gamma^2\psi^*$. When we separate this into the chiral parts of $\psi = \psi_L + \psi_R$ and apply $\gamma^5\gamma^2 = -\gamma^2\gamma^5$ we obtain $\psi_{CL} = i\gamma^2\psi_R{}^*$ and $\psi_{CR} = i\gamma^2\psi_L{}^*$. So, comparing these to $\psi_C = i\gamma^2\psi^*$ for the entire fermion is how we know that right-chiral antifermions are the antiparticles of left-chiral fermions and left-chiral antifermions are the antiparticles of right-chiral fermions. And because for any Dirac fermion the interaction charges of antifermions are opposite those of fermions, starting with $c_R(\nu) = 0$ and $c_L(\nu) = +1$ in (22.6), we know that $c_L(\nu_C) = 0$ and $c_R(\nu_C) = -1$. So, applying (22.9) to antineutrinos interacting with other fermions, we obtain the counterpart to (22.10), namely:

$$\begin{aligned}\mathcal{M}_Z(\nu_C f \to \nu_C f) &\propto \left[c_R(\nu_C)\overline{\nu}_{CR}\gamma^\mu\nu_{CR} + c_L(\nu_C)\overline{\nu}_{CL}\gamma^\mu\nu_{CL}\right]\left[c_R(f)\overline{f}_R\gamma_\mu f_R + c_L(f)\overline{f}_L\gamma_\mu f_L\right] \\ &= -\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_R(f)\left(\overline{f}_R\gamma_\mu f_R\right) - \left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_L(f)\left(\overline{f}_L\gamma_\mu f_L\right)\end{aligned}. \quad (22.14)$$

Using (22.6) and (22.7), the antineutrino counterparts to (22.11) through (22.13) are then:

$$\begin{aligned}\mathcal{M}_Z(\nu_C e \to \nu_C e) &= -0.46310\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{e}_R\gamma_\mu e_R\right) + 0.53690\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{e}_L\gamma_\mu e_L\right) \\ \mathcal{M}_Z(\nu_C\nu \to \nu_C\nu) &= -\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{\nu}_L\gamma_\mu\nu_L\right)\end{aligned}, \quad (22.15)$$

$$\begin{aligned}\mathcal{M}_Z(\nu_C u \to \nu_C u) &= +0.30873\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{u}_R\gamma_\mu u_R\right) - 0.69127\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{u}_L\gamma_\mu u_L\right) \\ \mathcal{M}_Z(\nu_C d \to \nu_C d) &= -0.15437\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{d}_R\gamma_\mu d_R\right) + 0.84563\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_L(d)\left(\overline{d}_L\gamma_\mu d_L\right)\end{aligned}, \quad (22.16)$$

$$\begin{aligned}\mathcal{M}_Z(\nu_C p \to \nu_C p) &= +0.46310\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{p}_R\gamma_\mu p_R\right) - 0.53690\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{p}_L\gamma_\mu p_L\right) \\ \mathcal{M}_Z(\nu_C n \to \nu_C n) &= +\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_L(n)\left(\overline{n}_L\gamma_\mu n_L\right)\end{aligned}. \quad (22.17)$$

So, as we expect, whatever is attractive to neutrinos is repulsive to antineutrinos, and vice versa. Most significantly, from $\mathcal{M}_Z(\nu_C e \to \nu_C e)$, we see that on the balance of the factor $+0.53690$ versus -0.46310, antineutrinos and electrons will repel. So, if we wish to use *Z* bosons to ensnare an antineutrino and draw it toward the nucleus as the trigger for β^+ decay we have a problem, because at least on a superficial first impression, the electrons in atomic shells will *repel* the antineutrino. To solve this problem, we must dig into the physical relation between the L and R chiral states of fermions, and introduce a *physical* process of "chiral polarization."

Fermion chirality has long been a somewhat murky subject. As developed in detail in Part I and reviewed in section 9 of this paper, the γ^5 axial operator is the flat spacetime generator of the Dirac-Kaluza-Klein timelike fifth dimension. And of course, the chiral operators $R=\frac{1}{2}\left(1+\gamma^5\right)$ and $L=\frac{1}{2}\left(1-\gamma^5\right)$ are built using this fifth-dimension generator. Now, if a massless fermion were to exist in nature which was thought possible for neutrinos until their oscillations were discovered, then the massless fermion would be entirely-chiral, either left- or right. For fermions with mass which is all that we appear to have in nature, when these fermions are highly-relativistic so that their propagation direction is not easily overtaken, the chirality operator is synonymous with the helicity operator, which is why chirality is often likened to spin. But for low-velocity fermions this is not so, and chirality has to be approached independently of any other physics concept, and thought of merely as one very important consequence of the fifth dimension. In fact, the existence of chiral fermions and axial vectors, pseudo-vectors, etc., as reviewed in section 9, provides clear physical evidence of this timelike fifth dimension – and certainly is infinitely superior to the complete lack of evidence of a spacelike fifth dimension curled-up into compactified strings.

Now, we know that because all fermions have mass, they are all four-component spinors, with the only question being whether neutrinos are Dirac or Majorana fermions. These four-component spinors contain a superposition $\psi=\psi_R+\psi_L$ of a right- and a left-chiral spinor, and in the Weyl representation of the Dirac matrices each of these two chiral spinors can be written as a two-component spinor. Because of the fermions having mass, ψ_R and ψ_L are not separate and distinct fermions. Rather they are simply superposed into the single massive fermion via $\psi=\psi_R+\psi_L$. But, because right-chiral and left-chiral spinors do interact differently under *Z* boson exchange as clearly manifest in (22.11) through (22.13) and (22.15) through (22.17), with R repelling and L attracting or vice versa, we now pose the question: is it possible for the R and L components of a single $\psi=\psi_R+\psi_L$ fermion to become *physically-separated* as a consequence of their having different weak neutral *Z*-mediated current charges? This is *not* to suggest that R and L chiral spinors would physically decouple into separate fermions because this cannot happen for any fermion with rest mass. Rather – similarly to what happens when a charged body moves into the middle of an electrically-neutral assemblage of positive and negative electrical charges – the question is this: *Can the R and L projections of a single fermion with mass, move into physically-separate spatial positions when they are each having different interactions with another fermion? And, in particular, is it possible for the R and L parts of a single fermion to become <u>physically polarized</u> whereby the chiral component which is attracting the other fermion moves closer to that fermion and the chiral component which is repelling the other fermion moves farther from that fermion?* This is what we are introducing as the *physical process* of "chiral polarization."

To set a baseline, let us consider each elementary fermion with $\psi=\psi_R+\psi_L$ when that fermion it is not interacting with any other fermion. Referring to (22.6), for the electron, up quark and down quark, respectively, we can calculate $c_R(e)c_L(e)=-0.24864$, $c_R(u)c_L(u)=-0.21342$ and $c_R(d)c_L(d)=-0.13054$. Because the sign of all three of these interactions measures between he left-chiral components is negative, this means the R and L chiral components of each of these fermions are attracted to one another by the weak neutral current interaction, so long as they are

within about .1 f of one another given that this is the maximum range for *Z* bosons. As noted, with a $Z \to \overline{f} f \to Z \to \overline{f} f \to Z...$ chain, this range can be extended. Therefore, the L and R chiral parts of fermions can be and likely are held together by the neutral current *Z* bosons of the very same interaction which is responsible for the absence of weak interaction symmetry between these left- and right-chiral components. For a neutrino, $c_R(\nu)c_L(\nu)=0$, which means there is nothing to hold the ν_R spinor close to the ν_L spinor. This is why all we ever observe is ν_L seemingly-decoupled from its ν_R counterpart. It is not that ν_R and ν_L are separate fermions, because having a mass all we have is a single $\nu = \nu_R + \nu_L$. Rather, it is that there is nothing other than gravitation to bind these together, which enables ν_R to stray widely from its counterpart ν_L. Indeed, the right-chiral ν_R is a true "ghost," having no interactions except for gravitational interaction with any other particle, and given tiny masses of the neutrinos, even this interaction is extraordinarily weak.

Now, let us presuppose that the L and R chiral parts of a fermion are held together by the weak neutral current except for neutrinos because ν_R has checked out from all interactions except gravitation, and that neutral current interactions with another fermion can cause a separation and polarization of the L and R components. Then, let's move a CvB trigger neutrino within *Z*-range of the electron. From (22.11), the amplitude contains a -0.53690 coefficient for attraction with e_L and a $+0.46310$ coefficient for repulsion with e_R, each of which has a larger magnitude than $c_R(e)c_L(e)=-0.24864$ which attracts the two components of $e = e_R + e_L$ to one another. As a result, e_R and e_L physically separate, with the former moving further from and the latter moving closer to the neutrino. By way of contrast, suppose we instead move a CvB *antineutrino* arriving within *Z*-range of an electron. From (22.15), the $\mathcal{M}_Z(\nu_C e \to \nu_C e)$ contains a -0.46310 coefficient for attraction with e_R and a $+0.53690$ coefficient for repulsion from e_L, each of which are still larger numbers than $c_R(e)c_L(e)=-0.24864$ which attracts the two components of $e = e_R + e_L$ to one another. The overall result in either case is that the superposed chiral projections of the electron with rest mass will separate and become polarized, with e_R closer and e_L further away. Then, the interaction of the antineutrino with the further e_L will be substantially weaker than the interaction with the closer e_R, not because of the change magnitude, but because of the very-limited range of the *Z* boson. At some later time once the neutrino or antineutrino is removed, the $c_R(e)c_L(e)=-0.24864$ attraction will pull the two chiral projections back together through *Z* or *Z*-chain interaction, into its non-interacting default. This is the mechanism for chiral polarization.

So, for neutrinos entering the electron shell cloud of an atom, the electrons within *Z* range or *Z*-chain range will polarize, with the $-0.53690\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{e}_L\gamma_\mu e_L\right)$ term in (22.11) drawing the neutrino toward the nucleus and the repulsive $+0.46310\left(\overline{\nu}_L\gamma^\mu\nu_L\right)\left(\overline{e}_R\gamma_\mu e_R\right)$ term polarized farther away and so having a greatly-diminished effect. For antineutrinos entering the same electron cloud, the electrons in range will again polarize, but now with $-0.46310\left(\overline{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\overline{e}_R\gamma_\mu e_R\right)$ from

(22.15) having the dominant effect and $+0.53690\left(\bar{\nu}_{CR}\gamma^{\mu}\nu_{CR}\right)\left(\bar{e}_{L}\gamma_{\mu}e_{L}\right)$ polarized further away with consequent diminished effect. So, because of this chiral polarization, the electron cloud can still ensnare antineutrinos toward the nucleus, but with somewhat less strength than it ensnares neutrinos. Specifically, in a polarized setting where the repulsive chiral spinors have been shunted far-enough away to have minimal effect, the neutrino draw is stronger than the antineutrino draw by a factor of 0.53690 to 0.46310. This means that during any period of time, even if there are similar number flux rates for CvB neutrino and antineutrinos, more neutrinos than antineutrinos will be attracted into the nucleus by the electrons surrounding the nucleus. And this in turn means that as a general trend subject to the vagaries of the shell structures of more complex nuclei, for a given atom which exhibits both β^{-} and β^{+} decay for some of its isotopes, the β^{-} half-lives ought to be shorter than the β^{+} half-lives for comparable nuclides. In other words, if there was no chiral polarization, then antineutrino triggers would never make it to the nucleus and we would only observe β^{-} decays. But because of chiral polarization, both neutrino and antineutrinos can make it to the nucleus and we do observe both β^{-} and β^{+} decay. But because there is a modestly-stronger polarized attraction of neutrinos over antineutrinos, the general empirical trend – with all other things being equal – should be toward shorter β^{-} than β^{+} decay half-lives.

Finally, every element in the periodic table from hydrogen (Z=1) through lead (Z=82) has at least one stable isotope. Thereafter, all isotopes of all elements are unstable. Of course, if neutrinos and antineutrinos are the triggers for both β^{-} and β^{+} decay, then even these stable elements will have CvB neutrinos and antineutrinos passing nearby. So the question here is how and why these stable nuclides are absolutely shielded from beta decay. And this reduces to the question of how CvB neutrinos and antineutrinos – even after they are drawn toward the nucleus through atomic electron shell structures that are substantially identical for any given atomic number Z – are entirely blocked from penetrating the nuclei of these stable nuclei to precipitate a beta decay, even as they are able to penetrate the nuclei of other isotopes.

For elements are stable or near stable (beta decay half-lives in years or in many years) or are unstable but decay through channels through other-than weak beta decay, we cannot consider CvB neutrinos and antineutrinos as triggers in isolation. We must also consider the nuclides themselves, and specifically, the energetic characteristics of each nuclide as relates to nuclear binding energies, and shells structures characterized by the principal, azimuthal, magnetic and spin quantum numbers n, l, m and s and the fermion Exclusion Principle that they reflect. And we must also consider Figures 8 and 9 which show that the up quark nests in a global minimum of the Lagrangian potential well while the down quark only nests in a local minimum, and Figures 14 and 15 which show that an electron nests in a global but the neutrinos only in a local minimum. Thus, we must expect that the proton and neutron balance in a given nuclide will cause that nuclide to stay stable, as is, *even if a neutrino or antineutrino closely approaches*, simply because a beta decay into a different state would lead to a nuclide that is very energetically-disfavored based on these binding energies and shell structures and Lagrangian potentials. Put simply: the arrival of a trigger neutrino or antineutrino is a *necessary* condition to trigger beta decay, but it is not *sufficient*. For sufficiency, the nuclear shell and energy conditions must also be favorable.

As to binding energies, we note that in an earlier publication [48], the author found in [10.6] and [10.7] that the rest energy of every free nucleon includes a latent binding energy given by $B_{\mathrm{P}} = 0.008200606481\,u = 7.640679\ \mathrm{MeV}$ and $B_N = 0.010531999771\,u = 9.812358\ \mathrm{MeV}$ for the proton and neutron respectively. This latent energy "see-saws" whereby some of this energy is always retained to confine quarks, while some is released in the form of fusion energy to bind nucleons together into nuclides. For example, ^{56}Fe with 26 protons and 30 neutrons – which has the distinction of having the highest average binding energy-per-nucleon than any other nuclide [61] – has available $\mathrm{B}\left(\mathrm{Fe}^{56}\right) = 26\times 7.640679\ \mathrm{MeV} + 30\times 9.812358\ \mathrm{MeV} = 493.028394\ \mathrm{MeV}$ of latent binding energy available be released for nucleon binding. This contrasts remarkably with the observed 56Fe binding energy of 492.253892 MeV, and shows that 99.8429093% of the latent binding energy goes into binding together the ^{56}Fe nucleus, with a small 0.1570907% balance reserved for confining quarks within each nucleon. As such, this constitutes an energy-based explanation of why quarks always remain confined even in this most-tightly-bound of nuclides. But the key point is that with each neutron containing about 9.81 MeV and each proton only about 7.64 MeV of energy that can be used for inter-nucleon binding, as a nuclide grows larger in its nucleon number, neutrons will be better-able to bind than protons, which explains the manifest excess of neutrons over protons as atomic number grows. So even if this particular research result in [48] is not considered, this does not obviate the fact that larger stable nuclides are neutron-rich, as are most isotopes of larger nuclides in general.

With all this in mind, let's again return to the periodic table and see how the foregoing might be use to explain the observed data trends, working from the nuclide table [58] and isotope listings it links to. It is important to keep in mind at the outset that for any given element with the atomic number Z and thus Z protons, there will also be Z electrons, and that the atomic shell structure of these electrons will be substantially identical *regardless of the particular isotope under consideration*. This means that the initial step of ensnaring a CvB neutrino toward the nucleus will proceed in essentially the same way for all isotopes of a given element with Z, with the same flux of CvB fermions drawn close to the nucleus *irrespective of isotope*. Therefore, the half-life of any particular isotope and the question of which are stable and which are not, will depend virtually exclusively upon the particular nuclide under review.

Hydrogen

We begin with Hydrogen, with one proton and one electron, and specifically with ^{1}H protium which does not contain any neutrons. This of course, is the output of the β^- decay process $\nu n \to e^- p$ decay process illustrated in Figure 17, because $^1\mathrm{H} = e^- p$. So now, we wish to consider the β^+ inverse of this process, which has never been observed and for which the data rules out a half-life below 10^{34} years. So, suppose that we now have a ^{1}H atom and a CvB antineutrino approaches, which by random statistical good fortune happens to be aimed dead-on toward the proton so that chiral polarization of the electron is not even needed to ensnare the antineutrino toward the nucleus. From (22.16) and (22.17) this antineutrino will be attracted, on balance, toward the proton and toward an up quark in the proton. Once they grow close enough for a $\nu_C \to W^- e^+$ or $p \to W^+ n$ emission, a β^+ decay would proceed by the channel $\nu_C e^- p \to e^+ e^- n \to n\gamma\gamma$ including electron / positron annihilation into photons, or into other bosons

or mesons at higher energies. To explain this we turn to Figures 8, 9, 14 and 15 which make clear that with all other things being equal – and *here they are equal* because all we have is a proton and an electron in isolation of any other nucleons – the state with pe^- thus ue^- for its distinguishing quark content, is energetically favored over the state with $n\nu$ thus $d\nu$ for its distinguishing quark content, because each of u and e^- nest at the global minimum while each of d and ν only nest at the local minimum of the Lagrangian potential. This explains why this β^+ is not observed for an isolated proton and electron constituting ^{1}H.

Next let's proceed to ^{2}H deuterium and to ^{3}H tritium. The former also is stable, but now contains a neutron which via (22.17) will strongly repel an incoming antineutrino. This makes it even harder for a β^+ decay to occur because in addition to the energy considerations of the preceding paragraph, there is also repulsion to ward off the incoming antineutrino. In the opposite direction, a β^- decay of $^2\text{H} \rightarrow {}^2\text{He}$ would produce a helium atom with no neutrons, which is energetically barred based on considerations we will momentarily consider regarding helium. The latter, ^{3}H, has one proton (1p) and two neutrons (2n). This does undergo β^- decay into ^{3}He with one neutron becoming a proton, but with a comparatively-large 12.32 y lifetime that is about half a million times longer than the 15-minute lifetime of a free neutron into a proton. The 2 neutrons in this instance fill a complete 1s shell ($n=1$, $l=0$) while the 1 proton occupies a 1s shell with an open proton position. As we shall see when we get to some heavier nuclides, the *relative* stability of ^{3}H with 2n appears to be part of a general trend wherein all of the monoisotopic elements except for beryllium contain an even number of neutrons, i.e., complete neutron shells as regards the spin quantum number *s*.

Helium

Now we turn to helium. It is best to start at ^{4}He for which the nucleus is an alpha particle. This is something of a paradigm for atomic stability, because there are numerous decay channels in which an entire α particle is emitted or absorbed whole hog. This extreme stability, we ascribe to ^{4}He having a complete 1s shell for both the protons and the neutrons, which is energetically stable enough to ward off any incoming CvB neutrinos or antineutrinos from precipitating either a β^- or a β^+ decay channel. The next-lower isotope, ^{3}He, is also stable. Being poor by 1 neutron, were it to decay, the reaction would be β^+ from $^3\text{He} \rightarrow {}^3\text{H}$. As between these two options, the nuclear bending energies render the neutron-poor ^{3}He more stable that the neutron-rich ^{3}H because these elements are still too light to require extra neutrons for effective binding, and relatedly, for the same reasons just reviewed as to why with all other things being equal, a e^-p state is energetically favored over a νn state. The lightest isotope, ^{2}He has two protons and no neutrons. Here, the alternative state is ^{2}H which can be arrived at through a β^+ decay. Here, with the two protons in ^{2}He attract an incoming CvB antineutrino with no neutrons to repel, while the 1p and 1n in ^{2}H provide a better attractive / repulsive balance. So, between the two options of ^{2}He versus ^{2}H, the latter is the stable option.

As to helium β^- decay, the isotopes ^{6}He and ^{8}He have two and four excess neutrons respectively. These excess neutrons – especially because they form complete spin pairs – are

highly attractive to incoming neutrinos. If we regard ^{6}He as an alpha plus two neutrons denoted $\alpha + 2\text{n}$, and ^{8}He as $\alpha + 4n$, then with the α having its neutrons and protons in n=1 shells, the additional neutrons in an n=2 shell with complete spins provide a great deal of additional attraction to CvB neutrinos via $\mathcal{M}^{NC}\left(\nu n \to \nu n\right)$ in (22.13), because there are no n=2 protons at all to offset this attraction. Thus, after the electron shells have ensnared passing CvB neutrinos and drawn them toward the nucleus, the ^{6}He nucleus will beta decay with an 806.7 ms half-life, and ^{8}He which will attract the neutrinos even more strongly, will decay with an even-shorter 119.0 ms half-life. In these two data points, we see a clear empirical correlation whereby as we increase the number of neutrons, we attract more neutrinos from a wider cross section, reduce the time required for one of these neutrinos to reach a neutron to decay, and thus reduce the half-life. Moreover, with 10^6 b corresponding to 1 ms, a half-life on the order of 1 s corresponds to 10^3 b and 100 ms corresponds to 10^4 b. So, the nuclides in these helium isotopes have enough attractive juice to decay with neutrinos ensnared by the two helium electrons from within about 10^3 b for ^{6}He and 10^4 b for ^{8}He. Note also that ^{7}He with 5 neutrons and ^{9}He with 7 neutrons do not beta decay. Rather, they favor shedding the odd neutron which is not spin-paired, then beta decaying from the lighter isotope.

Lithium

When we now turn to lithium, which has the stable isotopes ^{6}Li and ^{7}Li, for the first time we open an n=2 proton shell. And at the same time, we cross a natural nuclear physics threshold where "all other things" are "no longer equal," and neutron-rich nuclides begin to provide more stability than those which have more protons. In terms of the latent binding energies of about 9.81 MeV per neutron and about 7.64 MeV per proton, this is the threshold at which more energy is required for stable binding, and thus, more neutrons are needed. This is seen in ^{7}Li with 3p and 4n, which could in theory beta decay to ^{7}Be with 4p and 3n. But the latter ^{7}Be has a half-life of 53.22 d and the neutron-rich former ^{7}Li is the one that is stable. Likewise, ^{6}Li is stable. The alternatives would be ^{6}Be with 4p and 2n which is neutron poor and so cannot sustain binding (and actually decays by releasing 2p), or ^{6}He already reviewed which, with $\alpha + 2\text{n}$, will readily attract ensnared neutrinos and so is readily susceptible to $^6\text{He} \to {}^6\text{Li}$ decay. The lighter isotopes ^{5}Li and ^{4}Li could in theory undergo β^+ decay to ^{5}He and ^{4}He respectively, but apparently, nature follows a path in which simply discarding a proton is energetically-preferred. For β^- decay, ^{8}Li, ^{9}Li and ^{11}Li with 5, 6 and 8 neutrons respectively are respectively more-attractive to ensnared neutrinos. And, correlating fully with the neutrino trigger viewpoint, these have respective half-lives of 840.3 ms, 178.3 ms and 8.75 ms. with CvB neutrinos harvested the electron shells over cross sections ranging from about 10^4 b for ^{8}Li down to a little under 10^6 b for ^{11}Li.

Beryllium

One might suppose that ^{8}Be ought to be stable, but it is not. This is because this is already in the domain where extra neutrons are required for stability, and also because if we write is as $\alpha + 2\text{p} + 2\text{n}$, the alpha both of its proton and neutrons in an n=1 s-shell, where as the extra 2p and 2n must be in an n=2 p-shell. And it is more-favored to break this ^{8}Be into 2α each with s-shells, than to maintain the p-shell. So, ^{9}Be with 4p and 5n is the only stable isotope, making beryllium the first monoisotopic element, and the only one with an even number of protons and an odd number of neutrons. All others have odd proton and even neutron numbers. Moreover, ^{10}Be with

4p and 6n, in a pattern that will be repeated for heavier elements, is nearly stable, but does undergo β^- decay into ^{10}B with a very long half-life of 1.39x10^6 y. And of course, ^{10}B is a stable isotope of boron. For the lighter isotopes, ^{7}Be does decay into the stable ^{7}Li via electron capture, which is a form of β^+ decay, with a comparatively-long 53.22 d. Were ^{6}Be to beta decay it would become ^{6}Li which is stable, but nature takes the route of ejecting 2p to arrive at ^{4}He, apparently once again because of the extreme stability of alpha particles. For β^- decay, following the long-lived ^{10}Be already discussed, we have ^{11}Be, ^{12}Be and ^{14}be with respective half-lives of 13.81 s, 21.49 ms and 4.84 ms, once again correlating fully a neutrino trigger where excess neutrons greatly enhance the neutral current attraction of triggering neutrinos. We note that ^{13}Be with 4p and 9n does not beta decay. This would seem to be because the ninth neutron is exposed by itself in an incomplete-spin 2p shell, instead is rapidly shed to drop down to ^{12}Be which is the isotope that undergoes beta decay. This is similar to what happens when ^{7}He drops to ^{6}He before the latter undergoes beta decay, and it begins to establish a pattern wherein heavy isotope with an even number of neutrons (complete spins) will beta decay while those with an odd number of neutrons (incomplete spins) prefer to first very quickly drop a neutron and then beta decay.

Boron

Boron is stable in its ^{10}B and ^{11}B isotopes. The latter is rich by one neutron as is part of the pattern for elements heavier than helium. For light isotopes, ^{9}B could in theory β^+ decay into stable ^{9}Be with 4p and 5n. But instead, again because of the extreme stability of the alpha, the much-more rapid decay in lieu of waiting around for a CvB neutrino, is to drop a proton down to ^{8}Be which in turn immediately decays into two alpha particles. Next, ^{8}B with 5p and 3n does indeed β^+ decay into ^{8}Be which in turn immediately decays into two alpha particles in ^{4}He. All lighter isotopes also find a way to quickly decay toward the stability of ^{4}He. As to heavier isotopes, ^{12}B, ^{13}B, ^{14}B, ^{15}B, ^{17}B and ^{19}B all exhibit β^- decay, with respective half-lives of 20.20 ms, 17.33 ms, 12.5 ms, 9.87 ms 5.08 ms and 2.92 ms. This is a particularly striking validation of the correlation in which beta decay is precipitated by a neutrino trigger, with larger numbers of neutrons causing greater neutrino attraction and thus a decreased half-life. Reinforcing the pattern from Helium and Beryllium, we see that ^{16}B and ^{18}B with 11n and 13n respectively do not beta decay but instead much-more quickly shed a neutron, then beta decay from the lighter isotope.

Carbon

Carbon is the first nuclide which has multiple isotopes that undergo β^+ decay, and which as a result provides data that can be used to confirm the view that CvB antineutrinos trigger these types of beta decays just as CvB neutrinos trigger β^- decays. The stable nuclides as ^{12}C and ^{13}C and at a half-life of 5730 y, ^{14}C with 6p and 8n, well-known for its use in archeological dating, is almost but not quite stable. With its even number of neutrons, ^{14}C is the second element to repeat the pattern that stated with ^{10}Be, 4p and 6n, and its 1.39x10^6 y half-life. The shorter-lived β^- decays begin in earnest with ^{15}C through ^{22}C with the exception of ^{21}C with 6p and 15n which instead drops a neutron to ^{20}C before it beta decays into ^{20}N. Completely validating the neutrino trigger viewpoint, the respective half-lives of ^{15}C through ^{20}C are the successively-diminishing

2.449 s, 0.747 s, 193 ms, 92 ms, 46.2, ms and 16 ms, and for ^{22}C, 6.2 ms, owing to the successively-increasing neutral current attraction of neutrino to the added neutrons.

But as just noted, carbon exhibits a half-life pattern that also validates the antineutrino viewpoint for β^+ decays. At the lightest isotope, ^{8}C with 6p and 2n does not beta decay, but rather is alpha-driven. It sheds 2p down to ^{6}Be, which in turn immediate sheds 2 more protons to an alpha particle. However, ^{11}C, ^{10}C and ^{9}C – all of which are proton-rich – do exhibit β^+ decay, and their respective half-lives are the successively-diminishing 20.334 m, 19.290 s and 126.5 ms. There are two features of this data which are striking. First, if β^+ decay is triggered by CvB antineutrinos being attracted to protons via $\mathcal{M}_Z\left(\nu_C p \to \nu_C p\right)$ in (22.17), then as the number of neutrons which also by (22.17) would repel antineutrinos is reduced, the overall attraction balance over repulsion is increased. Thus, more antineutrinos will reach a proton to start a β^+ decay. Second, because the $\mathcal{M}_Z\left(\nu_C p \to \nu_C p\right)$ attraction of antineutrinos to protons in (22.17) is actually an attraction-weighted mix as between the two chiral sates, while the $\mathcal{M}^{NC}\left(\nu n \to \nu n\right)$ attraction of neutrinos to neutrons in (22.13) is purely attractive and thus stronger, this means that with all else equal, more neutrinos will be harvested per unit of time for β^- decay than will antineutrinos be harvested for β^+ decay, with the result that the β^+ half-lives ought to be longer, in general, than the β^- half-lives. Here, beyond the long ^{14}C lifetime that occur for reasons of nuclear structure and stability not neutrino availability, the β^- half-lives run from 2.449 s down to 6.2 ms, while the β^+ half-lives run from 20.334 m to 19.290 s to 126.5 ms. These are consistently longer than the β^- half-lives, and directly exhibit how the *Z*-mediated neutral current neutron-neutrino attraction is definitively stronger than the antineutrino-proton attraction, and how this directly impacts the observed beta-decay half-lives.

Nitrogen

The same pattern just reviewed for carbon remains intact for all of N, O, Li and Ne, but with some interesting details for the lightest-isotope β^+ decays. For nitrogen, ^{14}N and ^{15}N are stable. For ^{16}N through ^{22}N the β^- half-lives correlated to increased neutron number are the consistently-diminishing 7.13 s, 4.173 s, 622 ms, 271 ms, 130 ms, 87 ms, and 13.9 ms., supporting neutrino triggering with weak neutral current attraction to neutrons. For ^{13}N and ^{12}N which are the only two isotopes with β^+, the half-lives of 9.965 min and 11.000 ms are also consistently diminishing. However, the latter for ^{12}N decay does appear for the first time to buck the trend of β^+ half-lives being longer in general than β^- half-lives. But on closer inspection we find that there are two channels for this decay. The dominant channel (96.5%) is for $^{12}\text{N} \to {}^{12}\text{C}$. The less-frequent channel (3.5%) starts with ^{12}N but includes both an alpha decay and a β^+ decay. However, the β^+ decay and the proton loss do not happen absolutely simultaneously. If the β^+ occurs first, then we have $^{12}\text{N} \to {}^{12}\text{C}$, and with ^{12}C being stable, nothing more will happen. This just repeats the first channel. So, the α drop must occur before the β^+ to distinguish this channel,

which means that the detailed sequence is $^{12}\mathrm{N} \rightarrow {}^{8}\mathrm{B} \rightarrow {}^{8}\mathrm{Be}$. The ^{8}Be then further alpha decays toward stable helium. Now, alpha-driven decays, or drops of individual protons or neutrons, are typically shorter than beta-decays by many orders of magnitude. For example, many neutron drops take nanoseconds, the paradigmatic alpha decay $^{8}\mathrm{Be} \rightarrow 2\,{}^{4}\mathrm{He}$ has a half-life of $6.7\text{x}10^{-17}$s, and many proton drops are even shorter. Therefore, it is to be expected that the less-frequent $^{12}\mathrm{N} \rightarrow {}^{8}\mathrm{B} \rightarrow {}^{8}\mathrm{Be}$ channel will actually be very-much much faster and the dominant $^{12}\mathrm{N} \rightarrow {}^{12}\mathrm{C}$ somewhat slower, with 11.000 ms half-life being a statistical averaging of these two channels. If that is the case, it may well be – and likely is the case – that the $^{12}\mathrm{N} \rightarrow {}^{12}\mathrm{C}$ channel for a pure β^{+} decay, when segregated out, would indeed sustain the neutrino- and antineutrino-trigger trend of β^{+} decays having generally longer half-lives than β^{-} decays.

Oxygen

Oxygen is stable at ^{16}O, ^{17}O and ^{18}O, continuing the neutron-rich stability for nucleons heaver than that of helium. The β^{-} decays from ^{19}O through ^{25}O maintain the steadily-decreasing half-life sequence of 26.464 s, 13.51 s, 3.42 s, 2.25 s, 82 ms and 65 ms for increasing attraction between neutrinos and neutrons. Beta decays from ^{15}O to ^{13}O are 122.24 s, 70.598 s and 8.58 ms likewise maintaining the pattern for decreasing repulsion between neutrinos and antineutrinos. The first two β^{+} half-lives are longer than all of the β^{-} times consistent with greater attraction between neutrinos and neutrons than between antineutrinos and protons. The ^{13}O half-life again bucks the trend. But this mixes two channels, namely, the dominant channel (about 89.1%) with two serial β^{+} decays from $^{13}\mathrm{O} \rightarrow {}^{13}\mathrm{N} \rightarrow {}^{13}\mathrm{C}$, and a less-frequent channel (10.9%) with both a β^{+} decay and a proton emission from $^{13}\mathrm{O} \rightarrow {}^{12}\mathrm{C}$. Similarly to what was reviewed for nitrogen, the β^{+} decay and the proton loss do not happen absolutely simultaneously. If the β^{+} occurs first then the detailed sequence is $^{13}\mathrm{O} \rightarrow {}^{13}\mathrm{N} \rightarrow {}^{12}\mathrm{C}$. But if the proton drop occurs first then the sequence is $^{13}\mathrm{O} \rightarrow {}^{12}\mathrm{N} \rightarrow {}^{12}\mathrm{C}$. But as just reviewed, ^{12}N itself has two modes of proceeding, with the less-frequent (3.5%) mode being $^{12}\mathrm{N} \rightarrow {}^{8}\mathrm{B} \rightarrow {}^{8}\mathrm{Be}$. So as with nitrogen, we expect that 8.58 ms for ^{13}O actually averages the two decay channels, that the not-pure β^{+}, p decay $^{13}\mathrm{O} \rightarrow {}^{12}\mathrm{C}$ actually occurs much faster, particularly when it makes brief passage through ^{12}N. So as with nitrogen above, when the dominant and pure $^{13}\mathrm{O} \rightarrow {}^{13}\mathrm{N} \rightarrow {}^{13}\mathrm{C}$ decay is segregated out, we anticipate a longer lifetime consistent with β^{+} decays having taking longer than β^{-} decays.

Fluorine

Fluorine is the second monoisotopic element, and is the first with even neutron and odd proton numbers which trend is followed by all succeeding monoisotopes. The stable isotope is ^{19}F with 9p and 10n. The β^{-} half-lives start at 11.163 s for ^{20}F diminishing to 4.9 ms for ^{27}F. ^{28}F only emits a neutron, and ^{29}F drops to a 2.6 ms half-life. The absolute trend of decreasing β^{-} life with increasing neutron number thus increased neutrino attraction is very-slightly broken for the first time by ^{21}F with 9p and 12n and a 4.158 s half-life, followed by ^{22}F with 9p and 13n and a 4.23 s half-life. But now we are at a place in the periodic table where the nuclides are large and complex enough that mild pattern breaks in serial half-life decreases can be attributed to the

vagaries of atomic structure, just as occurs when general nuclear binding energy trends are broken for more complex nuclides. The β^+ half-lives are 109.771 min for ^{18}F and 64.49 s for ^{17}F, and these do follow both the expected diminishing half-life pattern and also the pattern of β^+ half-lives being generally longer that those for β^- decays.

Neon

The three stable isotopes of neon are ^{20}Ne, ^{21}Ne and ^{22}Ne. The β^- half-lives begin with ^{23}Ne and its 37.24 s half-life, then ^{24}Ne with the *longer* 3.38 min, then ^{25}Ne with 602(8) ms. So this too is a break in the half-life diminishment correlation. Thereafter the half-lives do serially descend from ^{26}Ne with 197 ms to ^{31}Ne with 3.4 ms, and then a mild break by ^{32}Ne with 3.5 ms. The ^{24}Ne isotope which has the longer half-life than ^{23}N has 10p and 14n, which may be attributable to the complete neutron shell, which is a $3p_0$ shell in atomic parlance, with the 0 subscript denoting that m_l=0. The isotopes with β^+ decays are ^{19}Ne with 17.296 s, ^{18}Ne with 1.672 s and ^{17}Ne with 109.2 ms. This adheres to the correlation of diminishing antineutrino repulsion with reduced neutron number, and setting aside ^{24}Ne with its 3.38 min that breaks the β^- sequence, it also adheres to the pattern of longer β^+ over β^- half-lives.

To summarize, all of the foregoing data for H through Ne does appear to confirm the viewpoint that weak beta decays are triggered by CvB neutrinos or antineutrinos entering the electron shells of atoms, being lured by weak neutral current *Z* boson or boson-chain interactions with chiral-polarized electrons toward the nucleus, being additionally attracted to the nucleus via neutral current interactions with individual nucleons once they are in range, and finally precipitating decay via *W* boson exchange. First, *all* beta decay lifetimes in the periodic table are no shorter than single digits of milliseconds, which, given the empirical number fluxes and low kinetic energies of CvB neutrinos shown in Figure 16, fits the view of these neutrinos and antineutrinos as being harvested from a cross section within 10^6 b of the nucleus. With $10^6 \text{ b} = .01 \text{ bb}$, this has linear dimensions on the order of 1/10 the Bohr diameter. That this number is .01 bb and not 1 bb also fits with the view that for CvB fermion to be drawn to a nucleus there needs to be at least some original penetration of a CvB fermion into the atomic shell based on its random travels through space, and that CvB fermions which merely "glance" the electron shells will not interact sufficiently to become ensnared. In short, the CvB number fluxes and low kinetic energies, diameters of atoms, and beta-decay nuclide half-lives all match up.

Second, given that the weak neutral current interaction between neutrinos and neutrons are strongly-attractive and between neutrinos and protons are repulsive albeit less strongly, and that these same interactions between antineutrinos and neutrons are strongly-repulsive and between antineutrinos and protons are attractive albeit less strongly, through the first eight elements from H to O there is an unbroken correlation between increased neutron number and decreased β^- half-life on the former hand, and between decreased neutron number and decreased β^+ half-life on the latter hand. Moreover, when we carefully consider the dual channels for the lightest isotopes of N and O, there is also an unbroken pattern of β^- half-lives being shorter than β^+ half-lives based on neutrino interactions with neutron-rich isotope nuclides being more strongly attractive than

antineutrino interactions with proton-rich isotope nuclides. As one considers larger nuclides there are occasional mild breaks in these basic patterns, but these may be fairly attributed to the complex nuclear shell structures masking these patterns, not really breaking them.

It is also important to observe that all of the foregoing appears to rule out neutrinos being Majorana fermions which are their own antiparticles, and to favor them being Dirac fermions just like all other elementary fermions. Simply stated: it would not be possible neutrinos to attract and antineutrinos to repel neutrons, and also for neutrinos to repel and antineutrinos to attract protons, if neutrinos and antineutrinos were one and the same. Thus, one can then make the broad statement that *if beta decays are in fact triggered by CvB neutrinos and antineutrinos and their weak-neutral current interactions with electrons, neutrons and protons in the manner proposed, then the empirically-observed half-life data for* β^- *decays taken together with that for* β^+ *decays definitively and empirically rules out Majorana neutrinos.*

Understanding weak beta-decay as the consequence of triggering by the flux of CvB neutrinos may also help to solve the neutron lifetime puzzle as reviewed in, e.g., [62], [63], wherein on average, "bottle" neutrons decay after 14 minutes and 39 seconds while "beam" neutrons last 14 minutes and 48 seconds. If β^+ decays occur when a CvB neutrino passes through the 100 b cross section about a free neutron and then is attracted toward the neutron via the weak neutral current, this would indicate that the motion of the beam neutrons enables them to evade some of the CvB neutrinos which might otherwise be available to trigger their decay, i.e., that the relative motion between the neutrinos and the CvB background decreases the number flux. This in turn suggests a refinement of beam experiments to test the CvB trigger explanation: If the motion of neutrons in the beam is reducing the number flux of CvB neutrinos and this in turn correlates to lifetime, then *performing neutron beam experiments which vary the speed of the neutrons should reveal a correlation wherein the faster the neutron beam, the longer the lifetime*.

In closing, it is often said that neutrinos pass through matter and in most cases pass through the entire earth without interacting, which is why they are so hard to detect. But if the foregoing is confirmed, this means that weak beta-decays are in fact the best and most-prolific detectors there are, of neutrinos flowing through our daily environment. Indeed, whenever a nuclear beta decay is observed – although previously unbeknownst – this means that a neutrino or antineutrino has struck and been absorbed by a nucleus, with one of the nucleons in that nucleus having acted as a neutrino detector. Knowing that there is a neutrino or antineutrino passing through and being snared by the electron shells toward a nucleus whenever a beta-decay event is about to occur, may also open new paths for better technological "management" of neutrinos.

To be continued

References

[1] Planck, M., *Über das Gesetz der Energieverteilung im Normalspektrum*, Annalen der Physik. **309** (3): 553–563 (1901) Bibcode:1901AnP...309..553P. doi:10.1002/andp.19013090310.
[2] A. Einstein, *The Foundation of the Generalised Theory of Relativity*, Annalen der Physik 354 (7), 769-822 (1916)

[3] J. A. Wheeler, Geometrodynamics, New York: Academic Press (1963)
[4] See, e.g., https://www.aps.org/publications/apsnews/200512/history.cfm
[5] H. Weyl, *Gravitation and Electricity*, Sitzungsber. Preuss. Akad.Wiss., 465-480. (1918).
[6] H. Weyl, *Space-Time-Matter* (1918)
[7] H. Weyl, *Electron und Gravitation*, Zeit. f. Physik, 56, 330 (1929)
[8] Kaluza, T, *Zum Unitätsproblem in der Physik, Sitzungsber*, Preuss. Akad. Wiss. Berlin. (Math. Phys.): 966–972 (1921)
[9] Klein, O. Quantentheorie und fünfdimensionale Relativitätstheorie, Zeitschrift für Physik A. 37 (12): 895–906 (1926). Bibcode:1926ZPhy...37..895K. doi:10.1007/BF01397481
[10] Klein, O, The Atomicity of Electricity as a Quantum Theory Law, *Nature*. **118**: 516 (1926) Bibcode:1926Natur.118..516K. doi:10.1038/118516a0
[11] https://en.wikipedia.org/wiki/Kaluza%E2%80%93Klein_theory#Field_equations_from_the_Kaluza_hypothesis
[12]https://en.wikipedia.org/wiki/Kaluza%E2%80%93Klein_theory#Equations_of_motion_from_the_Kaluza_hypothesis
[13] 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)
[14] https://en.wikipedia.org/wiki/Invertible_matrix#Blockwise_inversion
[15] https://en.wikipedia.org/wiki/Invertible_matrix#Inversion_of_2_%C3%97_2_matrices
[16] Yablon, J. R., *Quantum Theory of Individual Electron and Photon Interactions: Geodesic Lorentz Motion, Electromagnetic Time Dilation, the Hyper-Canonical Dirac Equation, and Magnetic Moment Anomalies without Renormalization*, https://www.researchgate.net/publication/325128699_Quantum_Theory_of_Individual_Electron_and_Photon_Interactions_Geodesic_Lorentz_Motion_Electromagnetic_Time_Dilation_the_Hyper-Canonical_Dirac_Equation_and_Magnetic_Moment_Anomalies_without_Renormaliza (2018)
[17] https://en.wikipedia.org/wiki/Einstein%E2%80%93Hilbert_action
[18] https://tigerweb.towson.edu/joverdui/5dstm/pubs.html
[19] https://en.wikipedia.org/wiki/Determinant#Block_matrices
[20] Halzen, F. and Martin A. D., *Quarks & Leptons: An Introductory Course in Modern Particle Physics*, Wiley (1984)
[21] http://pdg.lbl.gov/2018/reviews/rpp2018-rev-phys-constants.pdf
[22] *The Higgs boson reveals its affinity for the top quark*, https://press.cern/press-releases/2018/06/higgs-boson-reveals-its-affinity-top-quark (4 June 2018)
[23] A. M. Sirunyan et al. (CMS Collaboration), *Observation of t-bar t H Production*, Phys. Rev. Lett. 120, 231801 https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.231801 (4 June 2018)
[24] *Physicists cheer rendezvous of Higgs boson and top quark*, https://www.nature.com/articles/d41586-018-05348-x (8 June 2018)
[25] http://pdg.lbl.gov/2017/tables/rpp2017-qtab-mesons.pdf
[26] https://en.wikipedia.org/wiki/Einstein%E2%80%93Hilbert_action#Derivation_of_Einstein's_field_equations
[27] Zee, A. *Quantum Field Theory in a Nutshell*, Princeton (2003)
[28] C.N. Yang and R. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Physical Review. 96 (1): 191–195. (1954)
[29] Yablon, J. R., *Grand Unified SU(8) Gauge Theory Based on Baryons which Are Yang-Mills Magnetic Monopoles Journal of Modern Physics*, Vol. 4 No. 4A, pp. 94-120. (2013) doi: 10.4236/jmp.2013.44A011.
[30] Yablon, J. R, *A Nuclear Fusion System for Extracting Energy from Controlled Fusion Reaction*, World Intellectual Property Organization, WO/2014/106223 (2014) https://patentscope.wipo.int/search/docservicepdf_pct/id00000025517277/PAMPH/WO2014106223.pdf?psAuth=rD5Mq5vFTHPdqfjsYcI0wcNCzhFItduBNxW1oD9w9aw&download
[31] Yablon, J. R., *Predicting the Neutron and Proton Masses Based on Baryons which Are Yang-Mills Magnetic Monopoles and Koide Mass Triplets*, Journal of Modern Physics, Vol. 4 No. 4A, pp. 127-150 (2013) doi: 10.4236/jmp.2013.44A013.
[32] C. N. Yang and R. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Physical Review. **96** (1): 191–195. (1954)
[33] https://en.wikipedia.org/wiki/Spin_connection
[34] https://home.cern/topics/higgs-boson
[35] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cimento. 19: 154–164 (1961)
[36] Bohr, N., *On the Constitution of Atoms and Molecules, Part I*, Philosophical Magazine. 26 (151): 1–24 (1913)

[37] L. de Broglie, *A Tentative Theory of Light Quanta*, Phil. Mag. 47 446 (1924)
[38] https://physics.nist.gov/cgi-bin/cuu/Value?plkmc2gev|search_for=Planck+mass
[39] https://physics.nist.gov/cgi-bin/cuu/Value?plkmc2gev|search_for=Planck+mass
[40] J. A. Wheeler, *On the nature of quantum geometrodynamics*, Ann. Phys. 2 (6): 604–614 (1957)
[41] S. W. Hawking, *Black hole explosions?*, Nature. 248 (5443): 30–31 (1974)
[42] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), http://pdg.lbl.gov/2018/tables/rpp2018-sum-quarks.pdf
[43] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), http://pdg.lbl.gov/2018/tables/rpp2018-sum-leptons.pdf
[44] A.V. Manohar, C.T. Sachrajda and R.M. Barnett, *Quark Masses* (2018) http://pdg.lbl.gov/2018/reviews/rpp2018-rev-quark-masses.pdf
[45] R. Gatto, *The Mass Matrix*, in To Fulfill a Vision: Jerusalem Einstein Centennial Symposium on Gauge Theories and the Unification of Physical Forces, Y. Ne'eman ed., Addison-Wesley, section 14, 185-196 (1981)
[46] A. Ceccucci (CERN), Z. Ligeti (LBNL), and Y. Sakai (KEK), *CKM Quark-Mixing Matrix*, http://pdg.lbl.gov/2018/reviews/rpp2018-rev-ckm-matrix.pdf (2018)
[47] Yablon, J. R., *Why Baryons are Yang-Mills Magnetic Monopoles*, Hadronic Journal, Volume 35, Number 4, 399-467 (2012), https://jayryablon.files.wordpress.com/2013/03/hadronic-journal-volume-35-number-4-399-467-20121.pdf
[48] Yablon, J. R, *Predicting the Binding Energies of the 1s Nuclides with High Precision, Based on Baryons which Are Yang-Mills Magnetic Monopoles*, Journal of Modern Physics, Vol. 4 No. 4A, pp. 70-93 (2013) doi: 10.4236/jmp.2013.44A010.
[49] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), http://pdg.lbl.gov/2018/tables/rpp2018-sum-gauge-higgs-bosons.pdf
[50] https://en.wikipedia.org/wiki/Top_quark#Single_top_quarks
[51] M. J. Longo, *New Precision Tests of the Einstein Equivalence Principle from Sn1987a*, Phys. Rev. Lett. 60, 173 (1988)
[52] NuFIT 3.2, http://www.nu-fit.org/?q=node/166 (2018)
[53] K. Nakamura, S.T. Petcov, *Neutrino Mass, Mixing, and Oscillations*, http://pdg.lbl.gov/2018/reviews/rpp2018-rev-neutrino-mixing.pdf (2018)
[54] J. Chu, *3 Questions: Pinning down a neutrino's mass: KATRIN experiment investigates the ghostly particle*, https://news.mit.edu/2018/3-questions-pinning-down-neutrino-mass-0611 (2018)
[55] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), http://pdg.lbl.gov/2018/tables/rpp2018-sum-baryons.pdf
[56] C. Spiering, *Towards high-energy neutrino astronomy. A Historical Review*, EPJ H (2012) 37: 515. https://doi.org/10.1140/epjh/e2012-30014-2; https://arxiv.org/pdf/1207.4952.pdf
[57] https://en.wikipedia.org/wiki/W_and_Z_bosons
[58] https://en.wikipedia.org/wiki/Table_of_nuclides
[59] https://en.wikipedia.org/wiki/List_of_radioactive_isotopes_by_half-life
]60] http://www.astro.ucla.edu/~wright/neutrinos.html
[61] https://en.wikipedia.org/wiki/Nuclear_binding_energy#Nuclear_binding_energy_curve
[62] https://www.quantamagazine.org/neutron-lifetime-puzzle-deepens-but-no-dark-matter-seen-20180213/
[63] R. W. Pattie Jr. et al., *Measurement of the neutron lifetime using an asymmetric magnetogravitational trap and in situ detection*, https://arxiv.org/ftp/arxiv/papers/1707/1707.01817.pdf (2018)