



**HAL**  
open science

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

Jay R Yablon

► **To cite this version:**

Jay R Yablon. Theory of Fermion Masses, Mixing, Lagrangian Potentials and Weak Beta Decays, based on Higgs Bosons arising from the Scalar Fields of a Kaluza Klein Theory with Five-Dimensional General Covariance Provided by Dirac's Quantum Theory of the Electron. 2018. hal-01961238

**HAL Id: hal-01961238**

**<https://hal.science/hal-01961238>**

Preprint submitted on 19 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Jay R. Yablon, December 19, 2018

# **Theory of Fermion Masses, Mixing, Lagrangian Potentials and Weak Beta Decays, based on Higgs Bosons arising from the Scalar Fields of a Kaluza Klein Theory with Five-Dimensional General Covariance Provided by Dirac's Quantum Theory of the Electron**

Jay R. Yablon  
Einstein Centre for Local-Realistic Physics  
15 Thackley End  
Oxford OX2 6LB, United Kingdom  
[yablon@alum.mit.edu](mailto:yablon@alum.mit.edu)  
December 19, 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 to which their theoretical interconnections 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 scalar. After establishing a connection between this Kaluza-Klein scalar and the standard model Higgs scalar, we use the latter to theoretically connect the known masses of all the quarks and charged leptons to the CKM and PMNS mixing angles and matrix components and several other parameters which have not previously been 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 flavors of neutrino. 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. Uncovered as well, is a deep role for the cosmological neutrino background (CvB) and Higgs fields in triggering and facilitating weak interaction beta decay events.*

## Contents

Preface, and Guide for Efficient Reading and Study .....	1
1. Introduction – The Incompatibility of Kaluza-Klein and Dirac Theories .....	6
<b>PART I: THE MARRIAGE BETWEEN FIVE DIMENSIONAL KALUZA-KLEIN THEORY AND DIRAC’S QUANTUM THEORY OF THE ELECTRON .....</b>	<b>10</b>
2. The Kaluza-Klein Tetrad and Dirac Operators in Four Dimensional Spacetime, and the Covariant Fixing of Gauge Fields to the Photon .....	10
3. Derivation of the “Dirac-Kaluza-Klein” (DKK) Five-Dimensional Metric Tensor .....	16
4. Calculation of the Inverse Dirac-Kaluza-Klein Metric Tensor.....	19
5. The Dirac Equation with Five-Dimensional General Covariance .....	25
6. The Dirac-Kaluza-Klein Metric Tensor Determinant and Inverse Determinant .....	28
7. The Dirac-Kaluza-Klein Lorentz Force Motion .....	30
8. Luminosity and Internal Second-Rank Dirac Symmetry of the Dirac-Kaluza-Klein Scalar ...	40
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.....	46
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 .....	51
<b>PART II: THE DIRAC-KALUZA-KLEIN SCALAR, THE HIGGS FIELD, AND A THEORY OF FERMION MASSES, MIXING AND WEAK BETA DECAYS WHICH RUBUSTLY FITS THE EXPERIMENTAL DATA .....</b>	<b>59</b>
11. Spontaneous Symmetry Breaking of the Massless Luminous Dirac-Kaluza-Klein Scalar, and Integration to Deduce its Spacetime Behavior.....	59
12. The Fifth-Dimensional Component of the Dirac-Kaluza-Klein Energy Momentum Vector	69
13. Connection between the Dirac-Kaluza-Klein Scalar and the Higgs Field, and the Extraction of Energy from the Higgs Field by the Top Quark.....	74
<b>PART IIA: QUARKS .....</b>	<b>81</b>
14. Theory of Fermion Masses and Mixing: Up, Charm and Top Quarks .....	81
15. Theory of Fermion Masses and Mixing: Down, Strange and Bottom Quarks .....	90
16. Theoretical Relation amongst the Higgs Mass and the Isospin-Up and Isospin-Down Quark vevs; and the Two-Minimum, Two Maximum Lagrangian Potential for Quarks .....	101
17. The Role of the Higgs Boson and its Mass in Weak Beta-Decays Between Quarks .....	117
18. The CKM Quark Mixing Matrix Mass Parameterization, and the Fine-Tuning of Quark Masses, Mixing Angles and CKM Matrix Components.....	122
<b>PART IIB: LEPTONS .....</b>	<b>136</b>

19. Theory of Fermion Masses and Mixing: Electron, Mu and Tau Charged Leptons .....	136
20. Theory of Fermion Masses and Mixing: Prediction of the Neutrino Mass Sum and of the Individual Neutrino Masses .....	145
21. Prediction of a Second Leptonic Higgs Boson, and its Mass .....	152
22. The Two-Minimum, Two Maximum Lagrangian Potential for Leptons.....	155
PART IIC: COMPLETE THEORY OF WEAK BETA DECAY .....	162
23. How Weak Beta Decays are Triggered by Cosmological Neutrinos and Antineutrinos Interacting with Electrons, Neutrons and Protons via the Z Boson-Mediated Weak Neutral Current, with “Chiral Polarization” of Electrons .....	162
Conclusion .....	190
References.....	191

## Preface, and Guide for Efficient Reading and Study

### 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 led from one to the other, and why it is that this is all best-presented in one complete paper. But, given the very substantial length of this paper, it is also important to provide reader with a guide for efficient reading and study, which is contained a few pages hence in this preface.

As this manuscript will demonstrate, if we start with the Kaluza-Klein metric tensor denoted  $G_{MN}$  and then follow Dirac by finding set of gamma matrices  $\Gamma_M$ ,  $M=0,1,2,3,5$  defined such that the anticommutator  $\frac{1}{2}\{\Gamma_M, \Gamma_N\} \equiv G_{MN}$ , then not only are the most perplexing century-old problems of Kaluza-Klein theory repaired, but the  $\Gamma_M$  matrices so-defined can be used to formulate a modified Dirac equation  $(i\hbar c \Gamma^M \partial_M - mc^2)\Psi = 0$  in five dimensions, see (5.2), (5.4) and (5.6) infra. 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 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 the massive spin-1  $W$  and  $Z$  bosons of electroweak theory, and been gained additional support by the empirical observation of the Higgs scalar and its mass in the vicinity of 125 GeV. For spin- $\frac{1}{2}$  fermions, it is generally assumed that Higgs bosons are also the mainspring of gauge-invariant mass acquisition, but the specifics of how this occurs are not yet well-understood.

In this regard, where Kaluza-Klein theory is perhaps most prescient despite its genesis several years before gauge theory and Dirac theory and several decades before modern particle

physics and the standard model, is that its metric tensor denoted  $G_{MN}$  contain scalar field  $\phi$  for a spin-0 scalar boson, see (3.13) infra. This is in addition to  $G_{MN}$  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  $A_\mu$  for what in quantum theory are spin-1 photons. That is, Kaluza-Klein anticipated by decades, the modern view of interactions being mediated by bosons with even spins 0, 1 and 2. So, when we use the five-covariant  $\Gamma_M$  in a five-dimensional Dirac equation  $(i\hbar c \Gamma^M \partial_M - mc^2) \Psi = 0$ , a scalar field courtesy of Kaluza and Klein is implicitly contained in this Dirac equation. Consequently, it becomes 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 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 quarks with isospin-up, because of the uniquely-large mass of the top quark. These three quarks reveal the basic pattern for understanding fermion masses generally. This pattern is confirmed by being successfully extended to masses of the bottom, strange and down quarks with isospin-down. The lepton masses do not, however, follow as simply as do the quark masses. 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 at (19.11) supra. 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 the sum of the neutrino masses which turns out to be  $m_{\nu_e} + m_{\nu_\mu} + m_{\nu_\tau} = 0.13348 \text{ eV} / c^2$  at (20.2b). And from this it becomes possible to predict the individual neutrino masses at (20.4). Also predicted at (21.1) is a second Higgs boson for leptons, with a rest mass that turns out to be only a few MeV above the free proton and neutron rest masses, even though this second Higgs mass is derived independently.

When all of this is completed, at (21.5) 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 Newton and Fermi constants, the mass of the established Higgs boson, the mass of the new leptonic Higgs boson, and the value  $\alpha(M_w^{-2})$  of the electromagnetic running coupling at a probe energy equal to the  $W$  boson mass. 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  $W$  boson masses of about 80 GeV, but also by heretofore unknown, huge energy exchanges in the zone of 100 TeV. Additionally, the underlying Lagrangian potentials for the Higgs fields lead us to uncover a fundamental role for the cosmological neutrino background (CvB) as the trigger mechanism of weak beta decays.

The opening to be able to uncover these findings, however, as summarized, 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 to thereby crack one of the deepest puzzles that modern physics has to offer.

### Efficient Reading and Study Guide

In its original formulation to “repair” Kaluza-Klein theory, this manuscript was just under fifty 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, and also, to separating the finding about how the CvB triggers weak beta decays into a third paper. While the purely-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 not only as to how particle masses are acquired but also as to the underlying mechanism of weak beta decay, and how this all implicates the CvB in weak beta decays, would have been obscured with such a separation. So, it was decided to keep both parts along with the CvB findings in one manuscript, and to instead present below, a brief “guide” for efficient reading and study.

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 how much time, if any, to devote 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 masses 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 quark masses. 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 the coupling and mass sums (14.3) and (14.4) are indeed a true empirical relations within experimental error bars, and that (14.5) is therefore a warranted refinement for the precision for the top quark mass. The reader should next review the bi-unitary mass matrix transformations (14.8) to (14.10) and become convinced that the connections in

(14.13) between two of the mass mixing angles and two of the three real CKM angles are also true within experimental errors. And the reader should review how this all this data fits and is properly encapsulated in the couplings  $G_f$  for isospin-up quarks in (14.15).

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 (15.2) defining a second (local) vev minimum  $v_{\downarrow}$  for isospin-down quarks, now produces the first of the four relations in (15.6) whereby the third real CKM angle can also be related to a third mass mixing angle within experimental errors. Note that  $v_{\downarrow}$  in (15.2) is a different vacuum energy than the Fermi vev  $v \equiv 246.22$  GeV which we subsequently denote for distinction by  $v_{\uparrow} \equiv v$ . This becomes essential in developing the Lagrangian potential in section 16 with both a *global* vev  $v_{\uparrow}$  and a *local* vev  $v_{\downarrow}$  based on the mass sums of the isospin-up quarks and isospin-down quarks, respectively.

At this point, section 16 should be skimmed enough for the reader to become convinced that the theoretical relation  $m_h c^2 \equiv \frac{1}{2}(v_{\uparrow} + \frac{1}{\sqrt{2}}v_{\downarrow})$  precisely specifying the observed Higgs boson mass  $m_h$  in relation to these two quark mass-sum-based vev is also true within experimental errors. This means that  $v_{\downarrow}$  in (15.2) is not a new parameter, but rather, is a function  $v_{\downarrow}(v, m_h)$  of the Fermi vev and the Higgs mass. The reader should also be convinced that this now gives the parameter  $\lambda$  from the Lagrangian density the theoretical valuation of (16.8) in terms of the two vacuum minima. Like the Higgs mass, for decades  $\lambda$  been a theoretically-postulated parameter disconnected from other known data, and only discernable by experiment. The reader should also be convinced that as a result of all this, as reviewed prior to (16.6), the six quark masses become effectively reparameterized in terms of five heretofore-disconnected parameters, namely, the three CKM mixing angles  $\theta_{C21}$ ,  $\theta_{C23}$ ,  $\theta_{C31}$ , the Fermi vev, and the Higgs mass. Moreover, if the reader is willing to credit, at least on a preliminary basis, that the relation  $3(m_d - m_u)/(2\pi)^{1.5} = m_e$  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 true, period, then the electron rest mass itself becomes a sixth parameter. In this event, all six quark masses become fully reparameterized into other previously-disconnected parameters as summarized in (17.1) by  $m_u, m_c, m_t, m_d, m_s, m_b = F(v, \theta_{C31}, \theta_{C23}, \theta_{C21}, m_h, m_e)$ .

Next, the reader should review section 18, in which all of the foregoing results are globally fitted using the unitarity of the CKM quark mass mixing matrix. Specifically, using earlier-developed connections between the quark masses and the real CKM mixing angles restated at (18.1), it becomes possible to develop a “mass parameterization” of the CKM mixing matrix at (18.3). This is then used to obtain (18.10) which re-centers and tightens the data for five of the six quark masses and the CP-violating phase, (18.11) with refined values and error bars for the CKM

mixing angles in the standard parameterization, and (18.13) which refines the magnitudes and error bars of the elements of the CKM quark mass mixing matrix itself.

At this point we move to the lepton masses. For these the reader should next turn to section 19 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  $\theta_{P12}$  and  $\theta_{P13}$ . But, this can only be done only by postulating an extra energy  $\delta_{\downarrow}$  defined at (19.11), which is added to the rest energy sum  $m_{\tau}c^2 + m_{\mu}c^2 + m_e c^2$ . This means that we start with the masses  $m_{\tau}, m_{\mu}, m_e$ , supplemented with a new unknown parameter  $\delta_{\downarrow}$ . But we wash this out when we find at (19.16) that this sum with the extra energy can be related within experimental errors to the Fermi vev via  $m_{\tau}c^2 + m_{\mu}c^2 + m_e c^2 + \delta_{\downarrow} = \alpha(M_W^2)v_{\uparrow}$ , using the strength  $\alpha(M_W^2) \sim 1/128$  of the electromagnetic running coupling at a probe energy equivalent to the mass of the  $W$  boson. This boson, which interacts electromagnetically and weakly, must, of course, *always* be present at the vertex of any beta decay event. The Fermi vev is already a parameter used for quark masses. So effectively, as seen in (19.20),  $m_{\tau}, m_{\mu}, m_e, \delta_{\downarrow} = F(\theta_{P12}, \theta_{P13}, \alpha(M_W^2), \delta_{\downarrow})$  is the reparameterization, for charged leptons. 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 20 for the neutrino masses, where we seek to reparameterize the remaining set of mass/energy numbers  $\{m_{\nu_e}, m_{\nu_{\mu}}, m_{\nu_{\tau}}, \delta_{\downarrow}\}$ . The reader should first confirm that (20.1) is a correct numeric calculation, and then become convinced that the physical connection (20.2b) between the extra energy  $\delta_{\downarrow}$  and the sum of the neutrino rest masses is warranted within experimental errors and widely-accepted estimates for the upper limit of the neutrino mass sum. Because the ratio  $v/M_p c^2$  of the Fermi vev to the Planck energy is at the center of (20.2b), 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  $G$ . As a result, introduce *gravitation* into particle physics by way of the very tiny neutrino masses. As seen in directly in (20.2c), the extra energy  $\delta_{\downarrow}$  is equal to the sum  $(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2$  of the neutrino rest energies, times an amplification factor  $\sqrt{2M_p c^2 / v}$  which is at bottom based on the ratio of the Fermi constant  $G_F = 1.1663787(6) \times 10^{-5} \text{ GeV}^{-2}$  to the Newton constant  $G = 6.708 61(31) \times 10^{-39} \text{ GeV}^{-2}$ , in natural units. This sum of neutrino masses deduced via what is effectively the  $G/G_F$  ratio, is then used together with the known empirical square mass difference data (20.3) to predict the three neutrino masses at (20.4). The reader should also be convinced that these predictions are well within the ranges of what the neutrino masses are expected to be. Finally, the reader should become convinced that the connection between the neutrino masses and the final PMNS angle  $\theta_{P23}$  in (20.7), using the same type of bi-unitary transformations previously employed for the quark and charged lepton masses, correctly accords with the empirical data within experimental errors.

If the reader is convinced that the foregoing does represent a true reparameterization of the fermion rest masses, then it should be clear that the net result of all this, is that all twelve fermion rest masses will have been reparameterized according to (21.5), and that all told, twenty-two physics parameters which have 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 it was possible to obtain these results by first connecting the Higgs fields of particle physics to scalar fields of a Kaluza-Klein theory with five-dimensional general covariance provided by Dirac's quantum theory of the electron. As regards this connection, the upper equation (13.7) as well as Figure 1 for the example of the top quark, encapsulate the bridge between Kaluza-Klein theory and the theory of fermion masses, and explicitly show how energy is drawn out of the vacuum by the Higgs fields in accord with energy conservation principles to give rest mass to the fermions.

With the foregoing complete, the reader may next wish to review sections 16 and 22 in further depth to see how the Lagrangian potentials for quarks and leptons contain both a global and a local vev minimum and associated energy wells, and how weak beta decays for all but the third-generation fermions require the decaying fermion to make a "jump" between one well and the other over an energy barrier. For quarks, the  $x$ -axis position of this barrier is established by the usual Higgs boson mass, and for leptons it is established by the leptonic Higgs boson mass predicted at (21.1). Were this new leptonic Higgs boson with a mass just above that of the proton and neutron to be experimentally confirmed, this would provide direct support for separate quark and lepton Lagrangian potentials each having both these global and local vev minima.

Finally, readers for whom empirical data connections are paramount should review section 23, which could also be presented as a separate paper, but which is best presented together with all the other development here so that the context of its genesis is clear. This section reviews a wealth of empirical data which supports the proposition advanced as part of the overall study of beta decay, that  $\beta^-$  and  $\beta^+$  decays are respectively triggered by cosmological neutrinos and antineutrinos (CvB neutrinos) interacting with electrons in atomic shells and with protons and neutrons in atomic nuclei, via the  $Z$  boson-mediated weak neutral current. Not only does this lead to a possible solution of the so-called neutron lifetime puzzle and a very deep understanding of how neutrinos behave in the physical universe, but it shows how neutron beams, and protons and neutrons and atomic isotopes generally, can be used as detectors to empirically study the cosmological neutrino background.

## **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 geometrodynamics 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-geometrodynamics 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” geometrodynamics 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  $G^{MA}G_{AN} = \delta^M_N$  are specified in five dimensions with an index  $M=0,1,2,3,5$ , and may be represented in the 2x2 matrix format:

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

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  $[A_\mu, A_\nu] = 0$  for the vector potential gauge field  $A_\mu$ . 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, while the component  $G_{55} = \phi^2$  transforms as a scalar in spacetime. Because the metric tensor is dimensionless,  $G_{55} = \phi^2$  must be dimensionless, thus so too is  $\phi$ . And with  $\phi$  being a dimensionless scalar, the constant  $k$  must have dimensions of charge/energy because the metric tensor is dimensionless and because the gauge field  $A_\mu$  has dimensions of energy/charge.

It is very important to understand that when we set  $A_\mu = 0$  and  $\phi = 0$ ,  $G^{MN}$  in (1.1) becomes singular. This is indicated by the fact that  $\text{diag}(G_{MN}) = (g_{00}, g_{11}, g_{22}, g_{33}, 0)$  with a determinant  $|G_{MN}| = 0$  in this situation, and is seen directly from  $G^{55} = g_{\alpha\beta} A^\alpha A^\beta + 1/\phi^2 = 0 + \infty$ . Therefore, (1.1) relies upon  $\phi$  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 Einstein field equation, again with  $\phi$  taken to be dimensionless, the constant  $k$  is customarily found to be:

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

where  $k_e = 1/4\pi\epsilon_0 = \mu_0 c^2 / 4\pi$  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  $\text{diag}(\eta^{\mu\nu}) = (+1, -1, -1, -1)$  by defining ("≡") a set of four operator matrices  $\gamma^\mu$  via the anticommutator relation  $\frac{1}{2}\{\gamma^\mu, \gamma^\nu\} = \frac{1}{2}\{\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu\} \equiv \eta^{\mu\nu}$ . The lower-indexed gamma operators are likewise defined such that  $\frac{1}{2}\{\gamma_\mu, \gamma_\nu\} \equiv \eta_{\mu\nu}$ . 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  $g^{\mu\alpha} g_{\alpha\nu} \equiv \delta^\mu_\nu$ , we define a set of  $\Gamma^\mu$  with parallel definition  $\frac{1}{2}\{\Gamma^\mu, \Gamma^\nu\} \equiv g^{\mu\nu}$ . We simultaneously define a vierbein a.k.a. tetrad  $e_a^\mu$  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  $g^{\mu\nu} = \frac{1}{2}\{\Gamma^\mu, \Gamma^\nu\} = \frac{1}{2}\{\gamma^a \gamma^b + \gamma^b \gamma^a\} e_a^\mu e_b^\nu = \eta^{ab} e_a^\mu e_b^\nu$ . So just as the metric tensor  $g^{\mu\nu}$  transforms in four-dimensional spacetime as a contravariant (upper-indexed) tensor, these  $\Gamma^\mu$  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:

$$\frac{1}{2}\{\Gamma_M, \Gamma_N\} = \frac{1}{2}\{\Gamma_M \Gamma_N + \Gamma_N \Gamma_M\} \equiv G_{MN}; \quad \frac{1}{2}\{\Gamma^M, \Gamma^N\} = \frac{1}{2}\{\Gamma^M \Gamma^N + \Gamma^N \Gamma^M\} \equiv G^{MN}, \quad (1.3)$$

where  $\Gamma_M$  and  $\Gamma^M$  transform as *five-dimensional vectors* in five-dimensional spacetime. This would presumably include a five-dimensional definition  $\epsilon_A^M \gamma^A \equiv \Gamma^M$  for a tetrad  $\epsilon_A^M$ , where  $M=0,1,2,3,5$  is a world index and  $A=0,1,2,3,5$  is a local index, and where  $\Gamma^5$  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  $\Gamma^5$ . But this is a wrong presumption.

Specifically, as we shall now demonstrate, the Kaluza-Klein metric tensors in (1.1) *cannot* be deconstructed into  $\Gamma_M$  and  $\Gamma^M$  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  $A^\mu$ . This means that in fact, in view of Dirac theory which was developed afterwards, *the Kaluza-Klein metric tensors (1.1) are really not generally-covariant in five dimensions*. Rather, they only have a four-dimensional spacetime covariance represented in the components of  $G_{\mu\nu} = g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$ ,  $G^{\mu\nu} = g^{\mu\nu}$ ,  $G_{\mu 5} = \phi^2 k A_\mu$  and  $G^{\mu 5} = -A^\mu$ , and these are all patched together with fifth-dimensional components with which they are not generally-covariant. Moreover, even the spacetime components  $G_{\mu\nu} = g_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$  of (1.1) alone are not generally covariant even in the four spacetime dimensions alone, unless the gauge symmetry of the gauge field  $A_\mu$  is broken to remove two degrees of freedom and fixed to that of a photon. We represent this latter constraint by  $A^\mu = A_\gamma^\mu$ , with a subscripted  $\gamma$  which denotes a photon and which is not a spacetime index.

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  $\gamma^5$ , and the multitude of observed chiral and pseudoscalar and axial vector particle states that are centered about this  $\gamma^5$ . Finally, importantly, all of this happens without sacrificing the Kaluza "miracle" of placing electrodynamics onto a geometrodynamical footing. This is the focus of Part I of this paper.

Then, once we have a set of five  $\Gamma^M$  Dirac operators which possess five-dimensional covariance and via (1.3) reproduce a Kaluza-Klein metric tensor, we expect that these operators can be utilized in a new five-dimensional Dirac-type equation  $(i\hbar c \Gamma^M \partial_M - mc^2)\Psi = 0$  which will be found in (5.6) infra. Not only should this new Dirac equation govern fermion behavior, but the genesis of these  $\Gamma^M$  in (1.3) means that this Dirac equation will have the Kaluza-Klein scalar  $\phi$  built into it from the outset. Given that Dirac's equation governs fermion behaviors and that

fermions are expected to obtain their rest masses from a Higgs scalar, this naturally poses the question whether the Kaluza-Klein scalar can be connected to the Higgs scalar of the standard model, and whether the fermion masses known only from experimental observation can thereafter be given a theoretical foundation. As will also be seen, the fermions do acquire their masses from the Higgs field in the manner expected by the standard model, but with the Higgs field having additional properties rooted in its nexus with the Kaluza-Klein scalar. This will further reveal the role of the Higgs boson in weak interaction beta decays, and will in turn lead to an understanding of how the cosmological neutrino background (CvB) acts as a heretofore unknown triggering mechanism in these beta decays. This is the focus of Part II of this paper.

## 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:

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

The above uses 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  $\eta^{ab}$  and  $\eta_{ab}$ . To simplify calculation, we now set  $g_{\mu\nu} = \eta_{\mu\nu}$  thus  $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_\mu A_\nu$ . Later on, we will use the minimal-coupling principle to generalize back from  $\eta_{\mu\nu} \mapsto g_{\mu\nu}$ . With this, the spacetime becomes “flat” *except for* the curvature in  $G_{\mu\nu}$  brought about by the electrodynamic terms  $\phi^2 k^2 A_\mu A_\nu$ . 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:

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

Note, when we set either  $A_\mu = 0$  and / or  $\phi = 0$ , 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 solve:

$$\eta_{\mu a}\varepsilon'^a + \eta_{\nu b}\varepsilon'^b + \eta_{ab}\varepsilon'^a\varepsilon'^b = \phi^2 k^2 A_\mu A_\nu. \quad (2.3)$$

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  $\mu$  and  $\nu$  so in fact there are only ten (10) independent equations. Given that  $\text{diag}(\eta_{ab})=(1,-1,-1,-1)$ , the four  $\mu=\nu$  “diagonal” equations in (2.3) produce the relations:

$$\begin{aligned}\eta_{a_0}\varepsilon'_0{}^a + \eta_{0b}\varepsilon'^b{}_0 + \eta_{ab}\varepsilon'_0{}^a\varepsilon'^b{}_0 &= 2\varepsilon'^0{}_0 + \varepsilon'^0{}_0\varepsilon'^0{}_0 - \varepsilon'^1{}_0\varepsilon'^1{}_0 - \varepsilon'^2{}_0\varepsilon'^2{}_0 - \varepsilon'^3{}_0\varepsilon'^3{}_0 = \phi^2 k^2 A_0 A_0 \\ \eta_{a_1}\varepsilon'_1{}^a + \eta_{1b}\varepsilon'^b{}_1 + \eta_{ab}\varepsilon'_1{}^a\varepsilon'^b{}_1 &= -2\varepsilon'^1{}_1 + \varepsilon'^0{}_1\varepsilon'^0{}_1 - \varepsilon'^1{}_1\varepsilon'^1{}_1 - \varepsilon'^2{}_1\varepsilon'^2{}_1 - \varepsilon'^3{}_1\varepsilon'^3{}_1 = \phi^2 k^2 A_1 A_1 \\ \eta_{a_2}\varepsilon'_2{}^a + \eta_{2b}\varepsilon'^b{}_2 + \eta_{ab}\varepsilon'_2{}^a\varepsilon'^b{}_2 &= -2\varepsilon'^2{}_2 + \varepsilon'^0{}_2\varepsilon'^0{}_2 - \varepsilon'^1{}_2\varepsilon'^1{}_2 - \varepsilon'^2{}_2\varepsilon'^2{}_2 - \varepsilon'^3{}_2\varepsilon'^3{}_2 = \phi^2 k^2 A_2 A_2 \\ \eta_{a_3}\varepsilon'_3{}^a + \eta_{3b}\varepsilon'^b{}_3 + \eta_{ab}\varepsilon'_3{}^a\varepsilon'^b{}_3 &= -2\varepsilon'^3{}_3 + \varepsilon'^0{}_3\varepsilon'^0{}_3 - \varepsilon'^1{}_3\varepsilon'^1{}_3 - \varepsilon'^2{}_3\varepsilon'^2{}_3 - \varepsilon'^3{}_3\varepsilon'^3{}_3 = \phi^2 k^2 A_3 A_3\end{aligned}\tag{2.4a}$$

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

$$\begin{aligned}\eta_{a_1}\varepsilon'_0{}^a + \eta_{0b}\varepsilon'^b{}_1 + \eta_{ab}\varepsilon'_0{}^a\varepsilon'^b{}_1 &= -\varepsilon'^1{}_0 + \varepsilon'^0{}_1 + \varepsilon'^0{}_1\varepsilon'^0{}_0 - \varepsilon'^1{}_0\varepsilon'^1{}_1 - \varepsilon'^2{}_0\varepsilon'^2{}_1 - \varepsilon'^3{}_0\varepsilon'^3{}_1 = \phi^2 k^2 A_0 A_1 \\ \eta_{a_2}\varepsilon'_0{}^a + \eta_{0b}\varepsilon'^b{}_2 + \eta_{ab}\varepsilon'_0{}^a\varepsilon'^b{}_2 &= -\varepsilon'^2{}_0 + \varepsilon'^0{}_2 + \varepsilon'^0{}_2\varepsilon'^0{}_0 - \varepsilon'^1{}_0\varepsilon'^1{}_2 - \varepsilon'^2{}_0\varepsilon'^2{}_2 - \varepsilon'^3{}_0\varepsilon'^3{}_2 = \phi^2 k^2 A_0 A_2 \\ \eta_{a_3}\varepsilon'_0{}^a + \eta_{0b}\varepsilon'^b{}_3 + \eta_{ab}\varepsilon'_0{}^a\varepsilon'^b{}_3 &= -\varepsilon'^3{}_0 + \varepsilon'^0{}_3 + \varepsilon'^0{}_3\varepsilon'^0{}_0 - \varepsilon'^1{}_0\varepsilon'^1{}_3 - \varepsilon'^2{}_0\varepsilon'^2{}_3 - \varepsilon'^3{}_0\varepsilon'^3{}_3 = \phi^2 k^2 A_0 A_3\end{aligned}\tag{2.4b}$$

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

$$\begin{aligned}\eta_{a_2}\varepsilon'_1{}^a + \eta_{1b}\varepsilon'^b{}_2 + \eta_{ab}\varepsilon'_1{}^a\varepsilon'^b{}_2 &= -\varepsilon'^2{}_1 - \varepsilon'^1{}_2 + \varepsilon'^0{}_1\varepsilon'^0{}_0 - \varepsilon'^1{}_1\varepsilon'^1{}_2 - \varepsilon'^2{}_1\varepsilon'^2{}_2 - \varepsilon'^3{}_1\varepsilon'^3{}_2 = \phi^2 k^2 A_1 A_2 \\ \eta_{a_3}\varepsilon'_2{}^a + \eta_{2b}\varepsilon'^b{}_3 + \eta_{ab}\varepsilon'_2{}^a\varepsilon'^b{}_3 &= -\varepsilon'^3{}_2 - \varepsilon'^2{}_3 + \varepsilon'^0{}_2\varepsilon'^0{}_0 - \varepsilon'^1{}_2\varepsilon'^1{}_3 - \varepsilon'^2{}_2\varepsilon'^2{}_3 - \varepsilon'^3{}_2\varepsilon'^3{}_3 = \phi^2 k^2 A_2 A_3 \\ \eta_{a_1}\varepsilon'_3{}^a + \eta_{3b}\varepsilon'^b{}_1 + \eta_{ab}\varepsilon'_3{}^a\varepsilon'^b{}_1 &= -\varepsilon'^3{}_1 - \varepsilon'^1{}_3 + \varepsilon'^0{}_3\varepsilon'^0{}_0 - \varepsilon'^1{}_3\varepsilon'^1{}_1 - \varepsilon'^2{}_3\varepsilon'^2{}_1 - \varepsilon'^3{}_3\varepsilon'^3{}_1 = \phi^2 k^2 A_3 A_1\end{aligned}\tag{2.4c}$$

Now, we notice that the right-hand side of all ten of (2.4) have nonlinear second-order products  $\phi^2 k^2 A_\mu A_\nu$  of field terms, while on the left of each there is a mix of linear first-order and nonlinear second-order expressions containing the  $\varepsilon'^a{}_\mu$ . Our goal at the moment, therefore, is to eliminate all 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 (2.4a) the linear appearances are of  $\varepsilon'^0{}_0$ ,  $\varepsilon'^1{}_1$ ,  $\varepsilon'^2{}_2$  and  $\varepsilon'^3{}_3$  respectively. Noting again that the complete tetrad  $\varepsilon_\mu{}^a = \delta_\mu{}^a + \varepsilon'^a{}_\mu$  and that  $\varepsilon_\mu{}^a = \delta_\mu{}^a$  when  $A_\mu = 0$  or  $\phi = 0$ , 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  $\phi^2 k^2 A_\mu A_\nu$  will all appear in off-diagonal components of the tetrad. With this, (2.4a) reduce to:

$$\begin{aligned}-\varepsilon'^1{}_0\varepsilon'^1{}_0 - \varepsilon'^2{}_0\varepsilon'^2{}_0 - \varepsilon'^3{}_0\varepsilon'^3{}_0 &= \phi^2 k^2 A_0 A_0 \\ \varepsilon'^0{}_1\varepsilon'^0{}_1 - \varepsilon'^2{}_1\varepsilon'^2{}_1 - \varepsilon'^3{}_1\varepsilon'^3{}_1 &= \phi^2 k^2 A_1 A_1 \\ \varepsilon'^0{}_2\varepsilon'^0{}_2 - \varepsilon'^1{}_2\varepsilon'^1{}_2 - \varepsilon'^3{}_2\varepsilon'^3{}_2 &= \phi^2 k^2 A_2 A_2 \\ \varepsilon'^0{}_3\varepsilon'^0{}_3 - \varepsilon'^1{}_3\varepsilon'^1{}_3 - \varepsilon'^2{}_3\varepsilon'^2{}_3 &= \phi^2 k^2 A_3 A_3\end{aligned}\tag{2.5a}$$

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:

$$\begin{aligned} -\varepsilon_0^2 \varepsilon_1^2 - \varepsilon_0^3 \varepsilon_1^3 &= \phi^2 k^2 A_0 A_1 \\ -\varepsilon_0^1 \varepsilon_2^1 - \varepsilon_0^3 \varepsilon_2^3 &= \phi^2 k^2 A_0 A_2 . \\ -\varepsilon_0^1 \varepsilon_3^1 - \varepsilon_0^2 \varepsilon_3^2 &= \phi^2 k^2 A_0 A_3 \end{aligned} \tag{2.5b}$$

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:

$$\begin{aligned} \varepsilon_1^0 \varepsilon_2^0 - \varepsilon_1^3 \varepsilon_2^3 &= \phi^2 k^2 A_1 A_2 \\ \varepsilon_2^0 \varepsilon_3^0 - \varepsilon_2^1 \varepsilon_3^1 &= \phi^2 k^2 A_2 A_3 . \\ \varepsilon_3^0 \varepsilon_1^0 - \varepsilon_3^2 \varepsilon_1^2 &= \phi^2 k^2 A_3 A_1 \end{aligned} \tag{2.5c}$$

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:

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

Note, 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 next yields:

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

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

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

Now, the 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) while simultaneously satisfying (2.7c), is to set all of the pure-space components:

$$\varepsilon_1'^2 = \varepsilon_2'^1 = \varepsilon_2'^3 = \varepsilon_3'^2 = \varepsilon_3'^1 = \varepsilon_1'^3 = 0. \quad (2.8)$$

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

$$-A_1A_1 - A_2A_2 - A_3A_3 = A_0A_0, \quad (2.9a)$$

$$0 = \phi^2 k^2 A_0A_1 = \phi^2 k^2 A_0A_2 = \phi^2 k^2 A_0A_3. \quad (2.9b)$$

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:

$$A_0 = 0; \quad A_1A_1 + A_2A_2 + A_3A_3 = 0, \quad (2.10)$$

*These two constraints have removed two redundant degrees of freedom from the gauge field  $A_\mu$ , in a generally-covariant manner.* Moreover, for the latter constraint in  $A_1A_1 + A_2A_2 + A_3A_3 = 0$  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:

$$A_\mu = A\varepsilon_\mu \exp(-iq_\sigma x^\sigma / \hbar), \quad (2.11a)$$

with a polarization vector

$$\varepsilon_{R,L\mu}(\hat{z}) \equiv (0 \quad \pm 1 \quad +i \quad 0) / \sqrt{2}, \quad (2.11b)$$

and where  $A$  in (2.11a) 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  $\gamma$ ) with two helicity states propagating along the  $z$  axis, and having an energy-momentum vector:

$$cq^\mu(\hat{z}) = (E \quad 0 \quad 0 \quad cq_z) = (h\nu \quad 0 \quad 0 \quad h\nu). \quad (2.11c)$$

This also 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  $\Gamma_\mu$  to satisfy the Dirac anticommutator relation  $\frac{1}{2}\{\Gamma_\mu, \Gamma_\nu\} \equiv G_{\mu\nu}$  which is the spacetime subset of (2.1), *the gauge symmetry of  $A_\mu$  must be broken to correspond with that of the photon.* We designate required result simply by writing  $A_\mu = A_{\gamma\mu}$ . (Note: the subscript  $\gamma$  as used here denotes “photon,” and should not be confused as a spacetime index.) *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  $\Gamma_\mu$  transforming covariantly in spacetime with the anticommutator  $\frac{1}{2}\{\Gamma_\mu, \Gamma_\nu\} \equiv G_{\mu\nu}$ , 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 longstanding challenge of how to covariantly eliminate the redundancy inherent in using a four-component Lorentz vector  $A_\mu$  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  $\Gamma_\mu$  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.*

With this, we have now also deduced all the components of the tetrad  $\varepsilon_\mu^a = \delta_\mu^a + \varepsilon'_\mu{}^a$ . Pulling together all of  $\varepsilon_0^0 = \varepsilon_1^1 = \varepsilon_2^2 = \varepsilon_3^3 = 0$  and (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, this tetrad is deduced to have the components:

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

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

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

We in turn consolidate these into a set of  $\Gamma_\mu$  transforming as a four-vector in spacetime, namely:

$$\Gamma_\mu = \left( \gamma_0 + \phi k A_{\gamma_j} \gamma_j \quad \gamma_j + \phi k A_{\gamma_j} \gamma_0 \right). \tag{2.14}$$

It is a useful exercise to confirm that (2.14) above, inserted into (2.1), will reproduce  $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$ , 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:

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

The only change we have made in (1.1) is to replace  $A_\mu \mapsto A_{\gamma\mu}$ , which captures 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  $\Gamma_\mu$  defined in (2.1) without fixing the gauge field  $A_\mu$  to that of a photon  $A_{\gamma\mu}$ .

To avoid any possible confusion, it should be clearly stated that the photon gauge vector  $A_{\gamma\mu}$  obtained here does not differ in any way from the photon gauge vector of the standard model. Phenomenologically, the physical photons being represented are exactly the same. The difference is in how the phenomenological photon properties are formally derived from theoretical principles: It has long been known that although the vector gauge potential of classical electromagnetism has four degrees of freedom to accord with its four spacetime components, the observed photon field quanta have two degrees of freedom which accords with their being massless and having two transverse polarization states. In the standard model these redundant, unobservable two extra degrees of freedom in  $A_\mu$  are ordinarily removed by imposing the Lorentz and Coulomb gauge fixing conditions to arrive at  $A_{\gamma\mu}$ . Above, via (2.10) the exact same result is achieved merely by requiring the general covariance of Kaluza-Klein theory in five dimensions, implemented by deconstructing the Kaluza-Klein metric tensor into a set of five Dirac gamma matrix operators similarly to what is done via Dirac's Quantum Theory of the Electron. And as we have seen above, the gauge potential turns into a photon entirely via the subset requirement of Dirac-covariance in the four dimensions of spacetime. 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:

$$\frac{1}{2}\{\Gamma_M, \Gamma_N\} = \frac{1}{2}\{\Gamma_M\Gamma_N + \Gamma_N\Gamma_M\} \equiv G_{MN}. \quad (3.1)$$

As just shown, using (2.14) in the spacetime components of (3.1) with  $g_{\mu\nu} = \eta_{\mu\nu}$  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.

It is helpful to separate the time and space components of  $G_{MN}$  in (2.15) and write this as:

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

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:

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

Next, let us *define* a  $\Gamma_5$  to go along with the remaining  $\Gamma_\mu$  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  $G_{05} = G_{50}$ . This lays the foundation for the Kaluza miracles to remain intact. We impose this requirement though (3.1) by writing the  $\Gamma_5$  *definition* as:

$$\frac{1}{2}\{\Gamma_j, \Gamma_5\} = \frac{1}{2}\{\Gamma_j\Gamma_5 + \Gamma_5\Gamma_j\} \equiv G_{j5} = G_{5j} = \phi^2 k A_{\gamma j}. \quad (3.4)$$

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

$$0 + \phi^2 k A_{\gamma_j} \equiv \frac{1}{2} \{ \Gamma_j \Gamma_5 + \Gamma_5 \Gamma_j \} = \frac{1}{2} \{ \gamma_j, \Gamma_5 \} + \frac{1}{2} \phi k A_{\gamma_j} \{ \gamma_0, \Gamma_5 \}, \quad (3.5)$$

which reduces down to a pair of anticommutation constraints on  $\Gamma_5$ , namely:

$$\begin{aligned} 0 &= \frac{1}{2} \{ \gamma_j, \Gamma_5 \} \\ \phi &= \frac{1}{2} \{ \gamma_0, \Gamma_5 \} \end{aligned} \quad (3.6)$$

Now let's examine possible options for  $\Gamma_5$ .

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  $\Gamma_5$  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). So, using  $\Gamma_5 \equiv \gamma_X + Y$  in (3.6) we first deduce:

$$\begin{aligned} 0 &= \frac{1}{2} \{ \gamma_j \Gamma_5 + \Gamma_5 \gamma_j \} = \frac{1}{2} \{ \gamma_j \gamma_X + \gamma_j Y + \gamma_X \gamma_j + Y \gamma_j \} = \frac{1}{2} \{ \gamma_j, \gamma_X \} + \frac{1}{2} \{ \gamma_j, Y \} \\ 0 + \phi &= \frac{1}{2} \{ \gamma_0 \Gamma_5 + \Gamma_5 \gamma_0 \} = \frac{1}{2} \{ \gamma_0 \gamma_X + \gamma_0 Y + \gamma_X \gamma_0 + Y \gamma_0 \} = \frac{1}{2} \{ \gamma_0, \gamma_X \} + \frac{1}{2} \{ \gamma_0, Y \} \end{aligned} \quad (3.7)$$

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  $\gamma_0$  or  $\gamma_5$ . From the bottom line, however, we must also have the anticommutators  $\{ \gamma_0, \gamma_X \} = 0$  and  $\frac{1}{2} \{ \gamma_0, Y \} = \phi$ . 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  $\Gamma_5 = \gamma_5 + \phi \gamma_0$ . Including this in (2.14) now gives:

$$\Gamma_M = \begin{pmatrix} \gamma_0 + \phi k A_{\gamma_k} \gamma_k & \gamma_j + \phi k A_{\gamma_j} \gamma_0 & \gamma_5 + \phi \gamma_0 \end{pmatrix}. \quad (3.8)$$

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  $\Gamma_5 = \gamma_5 + \phi \gamma_0$  from (3.8) in (3.1) to ensure that these remaining components are also fully covariant over all five dimensions, then we determine that:

$$G_{05} = G_{50} = \frac{1}{2}\{\Gamma_0\Gamma_5 + \Gamma_5\Gamma_0\} = \frac{1}{2}\{(\gamma_0 + \phi k A_{\gamma_j} \gamma_j)(\gamma_5 + \phi \gamma_0) + (\gamma_5 + \phi \gamma_0)(\gamma_0 + \phi k A_{\gamma_j} \gamma_j)\} \quad (3.9)$$

$$= \phi \gamma_0 \gamma_0 + \frac{1}{2}\{\gamma_0, \gamma_5\} + \frac{1}{2}\phi k A_{\gamma_j} \{\gamma_j, \gamma_5\} + \frac{1}{2}\phi^2 k A_{\gamma_j} \{\gamma_j, \gamma_0\} = \phi$$

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

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:

$$G_{MN} = \begin{pmatrix} G_{00} & G_{0k} & G_{05} \\ G_{j0} & G_{jk} & G_{j5} \\ G_{50} & G_{5k} & G_{55} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \phi \\ 0 & \eta_{jk} + \phi^2 k^2 A_{\gamma_j} A_{\gamma_k} & \phi^2 k A_{\gamma_j} \\ \phi & \phi^2 k A_{\gamma_k} & 1 + \phi^2 \end{pmatrix}. \quad (3.11)$$

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 we set  $A_{\gamma_j} = 0$  and  $\phi = 0$  in (3.11), the metric signature becomes  $\text{diag}(G_{MN}) = (+1, -1, -1, -1, +1)$  with a determinant  $|G_{MN}| = -1$ , versus  $|G_{MN}| = 0$  in (1.1) as reviewed earlier. This means that the inverse obtained via  $G^{MA} G_{AN} = \delta^M_N$  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  $\phi$  which cause the fifth dimension in (1.1) to have a spacelike signature. And in fact, very importantly, 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 when  $\phi = 0$  from  $\gamma_5\gamma_5 = 1$ , *it now becomes possible to fully associate the Kaluza-Klein fifth dimension with the  $\gamma_5$  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  $G_{55} = \phi^2$  as in the usual (1.1) causes  $G_{55}$  to shrink or expand or even zero out entirely, based on the magnitude of  $\phi$ . In (3.11), there is no such problem. We review the physics consequences of all this more deeply in section 9 following other development.

To conclude this section, we wish to now consolidate (3.11) into a 2x2 matrix format akin to (1.1), by consolidating 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  $A_\mu = A_{\gamma\mu}$  as derived in section 2. 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) and taking advantage of  $A_{\gamma_0} = 0$ , we *define* a new covariant (lower-indexed) four-vector:

$$\Phi_\mu \equiv (\phi \quad \phi^2 k A_{\gamma_j}) = (\phi + \phi^2 k A_{\gamma_0} \quad \phi^2 k A_{\gamma_j}). \quad (3.12)$$

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  $\eta_{MN} \mapsto g_{MN}$  which also means accounting for the possibility of non-zero  $g_{\mu 5}$ ,  $g_{5\nu}$ , (3.11) may now be compacted via (3.12) to the 2x2 form:

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

This is the Dirac-Kaluza-Klein metric tensor which will form the basis for all continued development from here, and using (3.12) it should be closely contrasted with (1.1). The next step is to calculate the inverse  $G^{MN}$  of (3.13) above, in the next section.

#### 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  $G^{MA} G_{AN} = \delta^M_N$ . We already know this because when  $A_{\gamma_j} = 0$  and  $\phi = 0$  with  $g_{MN} = \eta_{MN}$ , we have a determinant  $|G_{MN}| = -1$ , and more generally  $|G_{MN}| = |g_{MN}|$ , 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  $\text{diag}(\eta_{MN}) = (+1, -1, -1, -1, +1)$  has two timelike and three spacelike dimensions when we set  $A_{\gamma j} = 0$  and  $\phi = 0$ , 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 (3.11) to show the entire 5x5 matrix, and we move the rows and columns so the ordering of the indexes is not  $M = 0, 1, 2, 3, 5$ , but rather is  $M = 0, 5, 1, 2, 3$ . With all this, (3.11) is:

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

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

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

with the matrix block assignments:

$$\mathbf{A} = \begin{pmatrix} 1 & \phi \\ \phi & 1+\phi^2 \end{pmatrix}; \quad \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ \phi^2 k A_{\gamma 1} & \phi^2 k A_{\gamma 2} & \phi^2 k A_{\gamma 3} \end{pmatrix};$$

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

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

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

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

$$\begin{aligned}
 -\mathbf{CA}^{-1}\mathbf{B} &= -\begin{pmatrix} 0 & \phi^2 k A_{\gamma 1} \\ 0 & \phi^2 k A_{\gamma 2} \\ 0 & \phi^2 k A_{\gamma 3} \end{pmatrix} \begin{pmatrix} 1+\phi^2 & -\phi \\ -\phi & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ \phi^2 k A_{\gamma 1} & \phi^2 k A_{\gamma 2} & \phi^2 k A_{\gamma 3} \end{pmatrix} \\
 &= -\begin{pmatrix} 0 & \phi^2 k A_{\gamma 1} \\ 0 & \phi^2 k A_{\gamma 2} \\ 0 & \phi^2 k A_{\gamma 3} \end{pmatrix} \begin{pmatrix} -\phi^3 k A_{\gamma 1} & -\phi^3 k A_{\gamma 2} & -\phi^3 k A_{\gamma 3} \\ \phi^2 k A_{\gamma 1} & \phi^2 k A_{\gamma 2} & \phi^2 k A_{\gamma 3} \end{pmatrix} = -\begin{pmatrix} \phi^4 k^2 A_{\gamma 1} A_{\gamma 1} & \phi^4 k^2 A_{\gamma 1} A_{\gamma 2} & \phi^4 k^2 A_{\gamma 1} A_{\gamma 3} \\ \phi^4 k^2 A_{\gamma 2} A_{\gamma 1} & \phi^4 k^2 A_{\gamma 2} A_{\gamma 2} & \phi^4 k^2 A_{\gamma 2} A_{\gamma 3} \\ \phi^4 k^2 A_{\gamma 3} A_{\gamma 1} & \phi^4 k^2 A_{\gamma 3} A_{\gamma 2} & \phi^4 k^2 A_{\gamma 3} A_{\gamma 3} \end{pmatrix}. \quad (4.5)
 \end{aligned}$$

Therefore:

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

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

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

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, see (2.10). 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:

$$\begin{aligned}
 (\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B})^{-1} &= \eta_{jk} - (\phi^2 - \phi^4) k^2 A_{\gamma j} A_{\gamma k} \\
 &= \begin{pmatrix} -1 - (\phi^2 - \phi^4) k^2 A_{\gamma 1} A_{\gamma 1} & -(\phi^2 - \phi^4) k^2 A_{\gamma 1} A_{\gamma 2} & -(\phi^2 - \phi^4) k^2 A_{\gamma 1} A_{\gamma 3} \\ -(\phi^2 - \phi^4) k^2 A_{\gamma 2} A_{\gamma 1} & -1 - (\phi^2 - \phi^4) k^2 A_{\gamma 2} A_{\gamma 2} & -(\phi^2 - \phi^4) k^2 A_{\gamma 2} A_{\gamma 3} \\ -(\phi^2 - \phi^4) k^2 A_{\gamma 3} A_{\gamma 1} & -(\phi^2 - \phi^4) k^2 A_{\gamma 3} A_{\gamma 2} & -1 - (\phi^2 - \phi^4) k^2 A_{\gamma 3} A_{\gamma 3} \end{pmatrix}. \quad (4.8)
 \end{aligned}$$

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  $\mathbf{C}$  from (4.3), with (4.4) and (4.8), to calculate:

$$\begin{aligned}
& -(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \\
& = \begin{pmatrix} 1+(\phi^2-\phi^4)k^2A_{\gamma_1}A_{\gamma_1} & (\phi^2-\phi^4)k^2A_{\gamma_1}A_{\gamma_2} & (\phi^2-\phi^4)k^2A_{\gamma_1}A_{\gamma_3} \\ (\phi^2-\phi^4)k^2A_{\gamma_2}A_{\gamma_1} & 1+(\phi^2-\phi^4)k^2A_{\gamma_2}A_{\gamma_2} & (\phi^2-\phi^4)k^2A_{\gamma_2}A_{\gamma_3} \\ (\phi^2-\phi^4)k^2A_{\gamma_3}A_{\gamma_1} & (\phi^2-\phi^4)k^2A_{\gamma_3}A_{\gamma_2} & 1+(\phi^2-\phi^4)k^2A_{\gamma_3}A_{\gamma_3} \end{pmatrix} \begin{pmatrix} 0 & \phi^2kA_{\gamma_1} \\ 0 & \phi^2kA_{\gamma_2} \\ 0 & \phi^2kA_{\gamma_3} \end{pmatrix} \begin{pmatrix} 1+\phi^2 & -\phi \\ -\phi & 1 \end{pmatrix}, \quad (4.9) \\
& = \begin{pmatrix} -\phi^3kA_{\gamma_1} - (\phi^2-\phi^4)\phi^3k^3A_{\gamma_1}A_{\gamma_k}A_{\gamma_k} & \phi^2kA_{\gamma_1} + (\phi^2-\phi^4)\phi^2k^3A_{\gamma_1}A_{\gamma_k}A_{\gamma_k} \\ -\phi^3kA_{\gamma_2} - (\phi^2-\phi^4)\phi^3k^3A_{\gamma_2}A_{\gamma_k}A_{\gamma_k} & \phi^2kA_{\gamma_2} + (\phi^2-\phi^4)\phi^2k^3A_{\gamma_2}A_{\gamma_k}A_{\gamma_k} \\ -\phi^3kA_{\gamma_3} - (\phi^2-\phi^4)\phi^3k^3A_{\gamma_3}A_{\gamma_k}A_{\gamma_k} & \phi^2kA_{\gamma_3} + (\phi^2-\phi^4)\phi^2k^3A_{\gamma_3}A_{\gamma_k}A_{\gamma_k} \end{pmatrix} = \begin{pmatrix} -\phi^3kA_{\gamma_1} & \phi^2kA_{\gamma_1} \\ -\phi^3kA_{\gamma_2} & \phi^2kA_{\gamma_2} \\ -\phi^3kA_{\gamma_3} & \phi^2kA_{\gamma_3} \end{pmatrix}
\end{aligned}$$

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

$$-\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \begin{pmatrix} -\phi^3kA_{\gamma_1} & -\phi^3kA_{\gamma_2} & -\phi^3kA_{\gamma_3} \\ \phi^2kA_{\gamma_1} & \phi^2kA_{\gamma_2} & \phi^2kA_{\gamma_3} \end{pmatrix}, \quad (4.10)$$

For the upper-left block in (4.2) we use  $\mathbf{B}$  from (4.3), with (4.4) and (4.9), to calculate:

$$\begin{aligned}
& \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \\
& = \begin{pmatrix} 1+\phi^2 & -\phi \\ -\phi & 1 \end{pmatrix} + \begin{pmatrix} 1+\phi^2 & -\phi \\ -\phi & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ \phi^2kA_{\gamma_1} & \phi^2kA_{\gamma_2} & \phi^2kA_{\gamma_3} \end{pmatrix} \begin{pmatrix} \phi^3kA_{\gamma_1} & -\phi^2kA_{\gamma_1} \\ \phi^3kA_{\gamma_2} & -\phi^2kA_{\gamma_2} \\ \phi^3kA_{\gamma_3} & -\phi^2kA_{\gamma_3} \end{pmatrix}, \quad (4.11) \\
& = \begin{pmatrix} 1+\phi^2 & -\phi \\ -\phi & 1 \end{pmatrix} + \begin{pmatrix} 1+\phi^2 & -\phi \\ -\phi & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \phi^5k^2A_{\gamma_k}A_{\gamma_k} & -\phi^4k^2A_{\gamma_k}A_{\gamma_k} \end{pmatrix} = \begin{pmatrix} 1+\phi^2 & -\phi \\ -\phi & 1 \end{pmatrix}
\end{aligned}$$

again using  $A_{\gamma_k}A_{\gamma_k} = 0$ . And, (4.8) already contains the complete lower-right block of (4.2).

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

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} 1+\phi^2 & -\phi & -\phi^3kA_{\gamma_1} & -\phi^3kA_{\gamma_2} & -\phi^3kA_{\gamma_3} \\ -\phi & 1 & \phi^2kA_{\gamma_1} & \phi^2kA_{\gamma_2} & \phi^2kA_{\gamma_3} \\ -\phi^3kA_{\gamma_1} & \phi^2kA_{\gamma_1} & -1-(\phi^2-\phi^4)k^2A_{\gamma_1}A_{\gamma_1} & -(\phi^2-\phi^4)k^2A_{\gamma_1}A_{\gamma_2} & -(\phi^2-\phi^4)k^2A_{\gamma_1}A_{\gamma_3} \\ -\phi^3kA_{\gamma_2} & \phi^2kA_{\gamma_2} & -(\phi^2-\phi^4)k^2A_{\gamma_2}A_{\gamma_1} & -1-(\phi^2-\phi^4)k^2A_{\gamma_2}A_{\gamma_2} & -(\phi^2-\phi^4)k^2A_{\gamma_2}A_{\gamma_3} \\ -\phi^3kA_{\gamma_3} & \phi^2kA_{\gamma_3} & -(\phi^2-\phi^4)k^2A_{\gamma_3}A_{\gamma_1} & -(\phi^2-\phi^4)k^2A_{\gamma_3}A_{\gamma_2} & -1-(\phi^2-\phi^4)k^2A_{\gamma_3}A_{\gamma_3} \end{pmatrix} \quad (4.12)$$

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

$$G^{MN} = \begin{pmatrix} 1+\phi^2 & -\phi^3 k A_{\gamma_1} & -\phi^3 k A_{\gamma_2} & -\phi^3 k A_{\gamma_3} & -\phi \\ -\phi^3 k A_{\gamma_1} & -1-(\phi^2-\phi^4)k^2 A_{\gamma_1} A_{\gamma_1} & -(\phi^2-\phi^4)k^2 A_{\gamma_1} A_{\gamma_2} & -(\phi^2-\phi^4)k^2 A_{\gamma_1} A_{\gamma_3} & \phi^2 k A_{\gamma_1} \\ -\phi^3 k A_{\gamma_2} & -(\phi^2-\phi^4)k^2 A_{\gamma_2} A_{\gamma_1} & -1-(\phi^2-\phi^4)k^2 A_{\gamma_2} A_{\gamma_2} & -(\phi^2-\phi^4)k^2 A_{\gamma_2} A_{\gamma_3} & \phi^2 k A_{\gamma_2} \\ -\phi^3 k A_{\gamma_3} & -(\phi^2-\phi^4)k^2 A_{\gamma_3} A_{\gamma_1} & -(\phi^2-\phi^4)k^2 A_{\gamma_3} A_{\gamma_2} & -1-(\phi^2-\phi^4)k^2 A_{\gamma_3} A_{\gamma_3} & \phi^2 k A_{\gamma_3} \\ -\phi & \phi^2 k A_{\gamma_1} & \phi^2 k A_{\gamma_2} & \phi^2 k A_{\gamma_3} & 1 \end{pmatrix}. \quad (4.13)$$

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

$$G^{MN} = \begin{pmatrix} G^{00} & G^{0k} & G^{05} \\ G^{j0} & G^{jk} & G^{j5} \\ G^{50} & G^{5k} & G^{55} \end{pmatrix} = \begin{pmatrix} 1+\phi^2 & -\phi^3 k A_{\gamma_k} & -\phi \\ -\phi^3 k A_{\gamma_j} & \eta^{jk} - (\phi^2 - \phi^4) k^2 A_{\gamma_j} A_{\gamma_k} & \phi^2 k A_{\gamma_j} \\ -\phi & \phi^2 k A_{\gamma_k} & 1 \end{pmatrix}. \quad (4.14)$$

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  $A_{\gamma}^M = G^{MN} A_{\gamma N}$ . 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:

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

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

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

which is the usual way of raising indexes in flat spacetime, then generalized to  $g^{\mu\nu}$  with minimal coupling. This means in view of (2.10), that  $A_{\gamma j} A_{\gamma}^j = 0$  and  $A_{\gamma\sigma} A_{\gamma}^{\sigma} = 0$  as well. As a result, with  $g^{\mu\nu} = \eta^{\mu\nu}$  we may raise the index in (3.12) to obtain:

$$\Phi^{\mu} = (\phi \quad \phi^2 k A_{\gamma}^j) = (\phi \quad -\phi^2 k A_{\gamma j}). \quad (4.17)$$

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

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

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:

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

So, we use (4.19) in (4.18), raise the indexes using (4.16) to obtain  $A_{\gamma j} A_{\gamma k} = A_{\gamma}^j A_{\gamma}^k$ , and write:

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

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:

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

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,  $G^{MA} G_{AN} = \delta^M_N$ .

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

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

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. All objects inside of (4.22) now have contravariant upper indexes, to match those of  $G^{MN}$ .

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  $\Gamma_\mu = (\gamma_0 - \phi k A_{\gamma_j} \gamma_j \quad \gamma_j - \phi k A_{\gamma_j} \gamma_0)$ , 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  $\Phi_\mu = (\phi \quad -\phi^2 k A_{\gamma_j})$  and  $\Phi^\mu = (\phi \quad -\phi^2 k A_{\gamma^j})$ . 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  $\Gamma_M$  in (3.8). In the next section, we will lay out the fundamental equations of 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  $\Gamma_M$  deduced in (3.8) via their anticommutator definitions (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  $\Gamma^M = G^{MN} \Gamma_N$  in two component form which consolidates the four spacetime operators  $\Gamma^\mu$  into a single four-covariant expression, then to do the same for the original  $\Gamma_M$  in (3.8).

As just noted, we may raise the indexes in the  $\Gamma_M$  of (3.8) by calculating  $\Gamma^M = G^{MN} \Gamma_N$ . 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  $\Gamma^0$ ,  $\Gamma^j$  and  $\Gamma^5$  as such:

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

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

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

To reduce the above, we have employed  $\Phi^\mu = (\phi \quad \phi^2 k A_{\gamma^j})$  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 (4.16), 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 its 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 with  $A_{\gamma^0} = 0$ . Making use of this, we consolidate all of (5.1) above into the two-part:

$$\boxed{\Gamma^M = \left( \gamma^\mu + \Phi^\mu k A_{\gamma^k} \gamma^k + k A_{\gamma^\mu} \Phi^0 \gamma^0 - \Phi^\mu \gamma^5 \quad \gamma^5 \right)}. \tag{5.2}$$

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

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

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

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:

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

Making use of  $\Phi_\mu \equiv (\phi \quad \phi^2 k A_{\gamma j})$  in (3.12), again mindful that  $A_{\gamma\mu} = 0$ , and noting that  $\Phi_\mu - \Phi_0 \Phi_0 k A_{\gamma\mu} = \Phi_0 = \phi$  for the  $\mu=0$  time component and  $\Phi_\mu - \Phi_0 \Phi_0 k A_{\gamma\mu} = \Phi_k - \phi^2 k A_{\gamma k} = 0$

for the  $\mu=k$  space components, it is a good exercise to confirm that (5.4) does reduce precisely to  $\Gamma_M = (\gamma_0 + \phi k A_{\gamma k} \gamma_k \quad \gamma_j + \phi k A_{\gamma j} \gamma_0 \quad \gamma_5 + \phi \gamma_0)$  obtained in (3.8). Using (5.2) and (5.4) and reducing with  $\Phi^\mu = (\phi \quad \phi^2 k A_{\gamma^j})$ ,  $\gamma^k \gamma_0 = -\gamma_0 \gamma^k$ ,  $A_{\gamma j} \gamma_j A_{\gamma^k} \gamma^k = 0$ ,  $\Phi_\mu A_{\gamma^\mu} = 0$  and  $A_{\gamma^\mu} A_{\gamma^\mu} = 0$ , it is also a good exercise to confirm that:

$$\Gamma_M \Gamma^M = \gamma_M \gamma^M = 5. \quad (5.5)$$

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:

$$\boxed{(i\hbar c \Gamma^M \partial_M - mc^2) \Psi = 0}. \quad (5.6)$$

If we then define a five-dimensional energy-momentum vector  $cp^M = (cp^\mu \quad cp^5)$  containing the usual four-dimensional  $cp^\mu = (E \quad \mathbf{cp})$ , and given that (3.13) and (4.22) now provide the means to lower and raise indexes at will, we may further define a wavefunction  $\Psi \equiv U_0(p^\Sigma) \exp(-ip_\Sigma x^\Sigma / \hbar)$  to include a Fourier kernel  $\exp(-ip_\Sigma x^\Sigma / \hbar)$  over all five dimensions  $x^\Sigma = (ct^0 \quad \mathbf{x} \quad ct^5)$ . These coordinates now include a *timelike*  $x^5 = ct^5$  which is distinguished from the ordinary time dimension  $x^0 = ct^0$  because as earlier reviewed, (3.13) has the tangent-space signature  $\text{diag}(G_{MN}) = (+1, -1, -1, -1, +1)$ . And  $U_0(p^\Sigma)$  is a Dirac spinor which is now a function of all five components of  $p^\Sigma$  but independent of the coordinates  $x^\Sigma$ . In other words,  $\partial_M U_0(p^\Sigma) = 0$ , 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, merely in five dimensions employing the  $\Gamma^M$  in (5.2), to obtain:

$$\boxed{(\Gamma^M cp_M - mc^2) U_0(p^\Sigma) = 0}. \quad (5.7)$$

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 the tensors in (1.1) are *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 massless photon with only two transverse degrees of freedom. Of course, it will be necessary as we progress from here to closely study the fifth component  $cp^5$  of the energy momentum and the second time dimension  $x^5 = ct^5$ . It is to be anticipated that the detailed development and study of the Dirac-Kaluza-Klein (DKK) equations (5.6) and (5.7) 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. And as we shall see in Part II of this paper, the detailed study of the momentum space (5.7) will enable us to develop of complete theory of fermion masses, mixing angles and weak beta decays, fitting the experimental data of the standard model, and in a number of cases refining this data to substantially-greater accuracy.

## 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  $\sqrt{-g} d^4x$  in four dimensions and which we anticipate will appear as  $\sqrt{-G} d^5x$  in five dimensions given the timelike signature of the fifth dimension. As we shall later elaborate in section 10, 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 these reasons and perhaps others, 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:

$$|G_{MN}| = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}| |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}|, \quad (6.1)$$

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 metric tensor  $g_{MN} = \eta_{MN}$ . As we found following (4.3),  $|\mathbf{A}| = 1 + \phi^2 - \phi^2 = 1$ , so (6.1) simplifies to  $|G_{MN}| = |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}|$ . 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  $|G_{MN}|$ .

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:

$$\begin{aligned} |m_{ij}| &= m_{11}m_{22}m_{33} + m_{12}m_{23}m_{31} + m_{13}m_{21}m_{32} - m_{13}m_{22}m_{31} - m_{12}m_{21}m_{33} - m_{11}m_{23}m_{32} \\ &= -1 + (\phi^2 - \phi^4)k^2 (A_{\gamma_1}A_{\gamma_1} + A_{\gamma_2}A_{\gamma_2} + A_{\gamma_3}A_{\gamma_3}) = -1 \end{aligned} \quad (6.2)$$

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  $|\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| = -1$ . Then, because  $|\mathbf{A}| = 1$ , that  $|G_{\text{MN}}| = -1 = |\eta_{\text{MN}}|$ . Moreover, because  $|M^{-1}| = |M|^{-1}$  for any square matrix, we likewise conclude that  $|G^{\text{MN}}| = -1 = |\eta^{\text{MN}}|$ . Then, because the blocks in (4.3) are based on having used  $g_{\text{MN}} = \eta_{\text{MN}}$ , starting with the flat Minkowski space results just reviewed, then employing minimal coupling to generalize from  $\eta_{\text{MN}} \mapsto g_{\text{MN}}$ , the complete five-dimensional determinant and its inverse are:

$$\boxed{\begin{aligned} G \equiv |G_{\text{MN}}| &= |\eta_{\text{MN}}| = -1 \mapsto G \equiv |G_{\text{MN}}| = |g_{\text{MN}}| \equiv g \\ G^{-1} \equiv |G^{\text{MN}}| &= -1 = |\eta^{\text{MN}}| \mapsto G^{-1} \equiv |G^{\text{MN}}| = |g^{\text{MN}}| \equiv g^{-1} \end{aligned}} \quad (6.3)$$

In the above, the massless, luminous  $A_\mu = A_{\gamma\mu}$  and the scalar field  $\phi$  wash entirely out of the determinant, leaving the determinants entirely-dependent upon  $g_{\text{MN}}$  which accounts for all curvatures *other than* those produced by  $A_{\gamma\mu}$  and  $\phi$ . In other words, even when  $A_{\gamma\mu} \neq 0$  and  $\phi \neq 0$ , these do not figure into the determinants which depend only on  $g_{\text{MN}}$ .

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  $\mathbf{A} = |\mathbf{A}| = 1$ , a 3x3 space block with the same  $\mathbf{D} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$ , and the 1x3 and 3x1 blocks  $\mathbf{B} = 0$  and  $\mathbf{C} = 0$ . So (6.1) becomes  $|G_{\mu\nu}| = |\mathbf{A}||\mathbf{D}| = |\mathbf{D}|$ . 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  $-\phi^4 k^2 A_{\gamma\mu} A_{\gamma\nu}$ , which tells us that the calculation of  $|\mathbf{D}|$  will produce the exact same result as (6.2) leading to  $|G_{\mu\nu}| = -1 = |\eta_{\mu\nu}|$ , with the inverse following suit. Consequently, after generalizing  $\eta_{\mu\nu} \mapsto g_{\mu\nu}$  via minimal coupling, we find that in the four dimensions of spacetime alone:

$$|G_{\mu\nu}| = |g_{\mu\nu}|; \quad |G^{\mu\nu}| = |g^{\mu\nu}|. \quad (6.4)$$

Here too, the massless, luminous  $A_\mu = A_{\gamma\mu}$  with two degrees of freedom and the scalar  $\phi$  are washed out entirely. Note, comparing (6.3) and (6.4), that we have reserved the notational definitions  $G \equiv |G_{\text{MN}}|$  and  $g = |g_{\text{MN}}|$  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 celebrates its centennial in 2019, has commanded attention for the past century for the very simple reason that despite its difficulties – all of which directly or indirectly stem from the degeneracy of the metric tensor (1.1) and its lack of five-dimensional covariance at the Dirac level, as will be reviewed in section 9 – it successfully explains Maxwell’s equations, the Lorentz Force motion and the Maxwell stress-energy tensor on an entirely geometrodynamical foundation. This successful geometrodynamical 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). This is 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  $\tilde{\Gamma}_{AB}^M$  (particularly in  $\tilde{\Gamma}_{\alpha 5}^{\mu}$  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 the Kaluza-Klein metric tensor (1.1), the 5-covariant metric tensor (3.13) does not sacrifice any aspects of the Kaluza miracle.

As just noted, in (1.1) it is the  $G_{\mu 5} = G_{5\mu} = \phi^2 k A_{\mu}$  metric tensor components which are responsible for the Kaluza miracle. In (3.13), these components are replaced by  $G_{\mu 5} = G_{5\mu} = g_{\mu 5} + \Phi_{\mu}$ . Using (3.12), this reduces to  $G_{\mu 5} = G_{5\mu} = \Phi_{\mu} = (\phi + \phi^2 k A_{\gamma 0} \quad \phi^2 k A_{\gamma j})$  in the flat Minkowski tangent spacetime with  $g_{MN} = \eta_{MN}$ . As such, the four-dimensional covariance of (1.1) in relation to the gauge fields is not changed. But, besides there being an  $\phi$  in the  $G_{05} = G_{50}$  terms, the other difference is that the gauge field  $A_{\mu} \mapsto A_{\gamma\mu}$  is now that of a photon for which  $\phi^2 k A_{\gamma 0} = 0$  as a result of (2.10). Because  $A_{\gamma\mu}$  still sits on the  $G_{\mu 5} = G_{5\mu}$  metric tensor components, now as a photon and merely with an extra  $\phi$  in  $G_{05} = G_{50}$ , we anticipate that the Kaluza “miracle” should remain intact. Our goal in this section is to show that this is so, as regards the Lorentz Force equation of classical motion for a charge in an electromagnetic field.

For a five-dimensional metric defined by:

$$c^2 dT^2 \equiv G_{MN} dx^M dx^N \quad (7.1)$$

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

$$\frac{d^2 x^M}{c^2 dT^2} = -\tilde{\Gamma}_{AB}^M \frac{dx^A}{cdT} \frac{dx^B}{cdT} = -\tilde{\Gamma}_{\alpha\beta}^M \frac{dx^{\alpha}}{cdT} \frac{dx^{\beta}}{cdT} - 2\tilde{\Gamma}_{\alpha 5}^M \frac{dx^{\alpha}}{cdT} \frac{dx^5}{cdT} - \tilde{\Gamma}_{55}^M \frac{dx^5}{cdT} \frac{dx^5}{cdT} \quad (7.2)$$

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

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

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

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

This is the equation of motion with regard to the ordinary invariant spacetime metric line element  $d\tau$ , with this four-dimensional proper time defined, using (3.13) and (3.12), 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  $\tau$  is then given by  $d^2 x^j/d\tau^2$  for the  $M=j=1,2,3$  components of (7.4). If we then multiply (7.4) through by  $d\tau^2/dt^0{}^2$  (mindful again that we now need to distinguish ordinary time  $t^0$  from time  $t^5$  in the second time dimension), we obtain the space acceleration  $d^2 x^j/dt^0{}^2$  with regard to the ordinary time coordinate.

Aside from minor notational changes intended to distinguish four- from five-dimensional objects, 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. 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 has a timelike fifth dimension. But the main reason 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  $\tilde{\Gamma}_{\alpha 5}^M$  are the particular ones responsible for the Kaluza-Klein representation of electrodynamics, whereby  $\tilde{\Gamma}_{\alpha 5}^\mu$  governs accelerations in the four spacetime dimensions and  $\tilde{\Gamma}_{\alpha 5}^5$  governs the fifth-dimensional acceleration. So, let's examine  $\tilde{\Gamma}_{\alpha 5}^\mu$  more closely. Using (3.13) and (4.22) in (7.3) along with the symmetric  $G_{MN} = G_{NM}$  we obtain:

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

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

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

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 start with (3.12) in the form of  $\Phi_{\mu} = (\phi + \phi^2 k A_{\gamma 0} \quad \phi^2 k A_{\gamma j})$ , 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:

$$\begin{aligned} \partial_0\Phi_k - \partial_k\Phi_0 &= \partial_0(\phi^2 k A_{\gamma k}) - \partial_k(\phi + \phi^2 k A_{\gamma 0}) \\ &= \phi^2 k (\partial_0 A_{\gamma k} - \partial_k A_{\gamma 0}) + 2\phi k (A_{\gamma k} \partial_0 - A_{\gamma 0} \partial_k) \phi - \partial_k \phi, \\ &= \phi^2 k F_{\gamma 0 k} - 2\phi k (A_{\gamma 0} \partial_k - A_{\gamma k} \partial_0) \phi - \partial_k \phi \end{aligned} \quad (7.8a)$$

$$\begin{aligned} \partial_j\Phi_k - \partial_k\Phi_j &= \partial_j(\phi^2 k A_{\gamma k}) - \partial_k(\phi^2 k A_{\gamma j}) \\ &= \phi^2 k (\partial_j A_{\gamma k} - \partial_k A_{\gamma j}) + 2\phi k (A_{\gamma k} \partial_j - A_{\gamma j} \partial_k) \phi. \\ &= \phi^2 k F_{\gamma j k} - 2\phi k (A_{\gamma j} \partial_k - A_{\gamma k} \partial_j) \phi \end{aligned} \quad (7.8b)$$

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  $\phi^2 k F_{\gamma\mu\nu}$ , modified with a  $\gamma$  subscript 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 (7.8) in a generally-covariant manner is the gradient  $-\partial_k\phi$  in (7.8a). For this, noting that with reversed indexes  $\partial_j\Phi_0 - \partial_0\Phi_j$  (7.8a) will produce a gradient  $+\partial_j\phi$  for the corresponding term, we define a four-component  $I_{\mu} \equiv (1 \quad \mathbf{0})$  and use this to form:

$$\begin{pmatrix} 0 & -\partial_k\phi \\ \partial_j\phi & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 \\ \partial_j\phi \end{pmatrix} (1 \quad \mathbf{0}) - \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} (0 \quad \partial_k\phi) = \partial_{\mu}\phi I_{\nu} - I_{\mu}\partial_{\nu}\phi = -(I_{\mu}\partial_{\nu} - I_{\nu}\partial_{\mu})\phi. \quad (7.9)$$

We then use this to covariantly combine both of (7.8) into:

$$\begin{aligned} \partial_{\mu}\Phi_{\nu} - \partial_{\nu}\Phi_{\mu} &= \phi^2 k F_{\gamma\mu\nu} - 2\phi k (A_{\gamma\mu}\partial_{\nu} - A_{\gamma\nu}\partial_{\mu})\phi - (I_{\mu}\partial_{\nu} - I_{\nu}\partial_{\mu})\phi \\ &= \phi^2 k F_{\gamma\mu\nu} - ((I_{\mu} + 2\phi k A_{\gamma\mu})\partial_{\nu} - (I_{\nu} + 2\phi k A_{\gamma\nu})\partial_{\mu})\phi \end{aligned} \quad (7.10)$$

The newly-appearing vector  $I_{\mu} + 2\phi k A_{\gamma\mu} = (1 \quad 2\phi k A_{\gamma j})$  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  $(1 \quad 2\phi k A_{\gamma j})$ , 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,  $\phi$  is taken to be dimensionless, and so  $2\phi k A_{\gamma j}$  is dimensionless as well. Compare also  $\Phi_\mu = (\phi \quad \phi^2 k A_{\gamma j})$ , then observe that  $\Phi_\mu + \phi^2 k A_{\gamma\mu} = \phi (I_\mu + 2\phi k A_{\gamma\mu})$ .

Most importantly, we now see in (7.10) 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  $\tilde{\Gamma}_{\alpha 5}^\mu$  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  $\phi_0$  being the proper potential (note: this is a different  $\phi$  from the Kaluza-Klein  $\phi$ ). 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  $qF_{\mu\nu} = q\partial_{[\mu} A_{\nu]} \rightarrow qF'_{\mu\nu} = q\partial_{[\mu} A_{\nu]} + \hbar c [\partial_{;\mu}, \partial_{;\nu}] \Lambda = qF_{\mu\nu}$ , the gauge transformation washes out because the commutator  $[\partial_{;\mu}, \partial_{;\nu}] \Lambda = 0$  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  $\partial_{;\mu} \partial_{;\nu} \Lambda = \partial_{;\mu} \partial_\nu \Lambda = \partial_\mu \partial_\nu \Lambda - \Gamma_{\mu\nu}^\sigma \partial_\sigma \Lambda$ , with a similar expression under  $\mu \leftrightarrow \nu$  interchange, and because  $\Gamma_{\mu\nu}^\sigma = \Gamma_{\nu\mu}^\sigma$  is symmetric under such interchange.

So even though we cannot Lorentz transform  $A_\gamma^\mu$  into  $A^\mu$ , and even though the gauge of  $A_\gamma^\mu$  is fixed so we cannot even gauge transform  $A_\gamma^\mu$  into  $A^\mu$ , 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  $\mathbf{E}$  and  $\mathbf{B}$ ) 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 (7.10) 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  $A^\mu = (\phi \ \mathbf{A})$ . Moreover, once we gauge transform  $F_\gamma^{\mu\nu} \rightarrow F^{\mu\nu}$ , the classical field strength  $F^{\mu\nu}$  will contain innumerable-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 (7.10) with a gauge-transformed  $F_{\gamma\mu\nu} \rightarrow F_{\mu\nu}$  into (7.7), to find that:

$$\begin{aligned} \tilde{\Gamma}_{\alpha 5}^\mu &= \frac{1}{2} \left( \eta^{\mu\sigma} - \Phi^0 \Phi^0 k^2 A_\gamma^\mu A_\gamma^\sigma + \Phi^\mu \Phi^\sigma \right) \phi^2 k F_{\alpha\sigma} \\ &+ \frac{1}{2} \left( \eta^{\mu\sigma} - \Phi^0 \Phi^0 k^2 A_\gamma^\mu A_\gamma^\sigma + \Phi^\mu \Phi^\sigma \right) \partial_5 \left( \Phi_0 \Phi_0 k^2 A_{\gamma\sigma} A_{\gamma\alpha} \right) \\ &- \frac{1}{2} \left( \eta^{\mu\sigma} - \Phi^0 \Phi^0 k^2 A_\gamma^\mu A_\gamma^\sigma + \Phi^\mu \Phi^\sigma \right) \left( \left( I_\alpha + 2\phi k A_{\gamma\alpha} \right) \partial_\sigma - \left( I_\sigma + 2\phi k A_{\gamma\sigma} \right) \partial_\alpha \right) \phi \\ &- \frac{1}{2} \Phi^\mu \partial_\alpha \left( \Phi_0 \Phi_0 \right) \end{aligned} \quad (7.11)$$

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  $i\hbar \partial_\Lambda A_{\gamma\mu} = q_\Lambda A_{\gamma\mu}$  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  $\sigma$  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  $i\hbar \partial_\Lambda A_{\gamma\mu} = q_\Lambda A_{\gamma\mu}$ , 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, (7.11) reduces to:

$$\begin{aligned}
\tilde{\Gamma}_{\alpha 5}^{\mu} &= \frac{1}{2} \left( \eta^{\mu\sigma} - \phi^2 k^2 A_{\gamma}^{\mu} A_{\gamma}^{\sigma} + \Phi^{\mu} \Phi^{\sigma} \right) \phi^2 k F_{\alpha\sigma} \\
&+ \frac{1}{2} \left( \eta^{\mu 0} + 2\eta^{\mu k} k A_{\gamma k} \phi - \Phi^{\mu} \phi \right) \partial_{\alpha} \phi \\
&- \frac{1}{2} \left( \eta^{\mu 0} + \Phi^{\mu} \phi \right) \left( I_{\alpha} + 2\phi k A_{\gamma\alpha} \right) \partial_0 \phi \\
&- \frac{1}{2} \left( \eta^{\mu k} - \phi^2 k^2 A_{\gamma}^{\mu} A_{\gamma}^k + \Phi^{\mu} \phi^2 k A_{\gamma}^k \right) \left( I_{\alpha} + 2\phi k A_{\gamma\alpha} \right) \partial_k \phi \\
&+ \phi \partial_5 \phi k^2 A_{\gamma}^{\mu} A_{\gamma\alpha} + \frac{1}{2} \phi^2 k^2 \partial_5 A_{\gamma}^{\mu} A_{\gamma\alpha} + \frac{1}{2} \phi^2 k^2 A_{\gamma}^{\mu} \partial_5 A_{\gamma\alpha}
\end{aligned} \tag{7.12}$$

Now, it is the upper  $\mu$  index in  $\tilde{\Gamma}_{\alpha 5}^{\mu}$  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 (7.12) into its time and space components, as such:

$$\begin{aligned}
\tilde{\Gamma}_{\alpha 5}^0 &= \frac{1}{2} \left( \eta^{0\sigma} + \phi \Phi^{\sigma} \right) \phi^2 k F_{\alpha\sigma} \\
&+ \frac{1}{2} \left( 1 - \phi^2 \right) \partial_{\alpha} \phi - \frac{1}{2} \left( 1 + \phi^2 \right) \left( I_{\alpha} + 2\phi k A_{\gamma\alpha} \right) \partial_0 \phi - \frac{1}{2} \phi^3 k A_{\gamma}^k \left( I_{\alpha} + 2\phi k A_{\gamma\alpha} \right) \partial_k \phi,
\end{aligned} \tag{7.13a}$$

$$\begin{aligned}
\tilde{\Gamma}_{\alpha 5}^j &= \frac{1}{2} \left( \eta^{j\sigma} - \phi^2 k^2 A_{\gamma}^j A_{\gamma}^{\sigma} + \phi^2 k A_{\gamma}^j \Phi^{\sigma} \right) \phi^2 k F_{\alpha\sigma} \\
&+ \left( 1 - \frac{1}{2} \phi^2 \right) k A_{\gamma}^j \phi \partial_{\alpha} \phi \\
&- \frac{1}{2} \phi^2 k A_{\gamma}^j \left( I_{\alpha} + 2\phi k A_{\gamma\alpha} \right) \phi \partial_0 \phi \\
&- \frac{1}{2} \left( \eta^{jk} - \left( \phi^2 - \phi^4 \right) k^2 A_{\gamma}^j A_{\gamma}^k \right) \left( I_{\alpha} + 2\phi k A_{\gamma\alpha} \right) \partial_k \phi \\
&+ k^2 A_{\gamma}^j A_{\gamma\alpha} \phi \partial_5 \phi + \frac{1}{2} \phi^2 k^2 A_{\gamma\alpha} \partial_5 A_{\gamma}^j + \frac{1}{2} \phi^2 k^2 A_{\gamma}^j \partial_5 A_{\gamma\alpha}
\end{aligned} \tag{7.13b}$$

It is noteworthy that all terms in (7.12) 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 (7.13a) because  $A_{\gamma}^0 = 0$ .

Now, as previewed prior to (7.11),  $A_{\gamma\alpha}$  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  $F_{\alpha\sigma}$ , 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  $k A_{\gamma\alpha}$  will be very small in relation to the numbers  $\pm 1$  contained in  $\eta^{\mu\nu}$ . With this in mind, we may set  $A_{\gamma\alpha} \cong 0$  as an extraordinarily-close approximation to zero in all terms which contain  $A_{\gamma\alpha}$  in (7.13). This includes for (7.13a), only retaining  $\Phi^0 = \phi$  in  $\phi \Phi^{\sigma} \phi^2 k F_{\alpha\sigma} = \phi \Phi^0 \phi^2 k F_{\alpha 0}$ . And in (7.13b) we further use  $\eta^{jk} \partial_k = -\partial_j$ . So now, both of (7.13) reduce to the much-simpler:

$$\tilde{\Gamma}_{\alpha 5}^0 = \frac{1}{2} \left( 1 + \phi^2 \right) \phi^2 k F_{\alpha 0} + \frac{1}{2} \left( 1 - \phi^2 \right) \partial_{\alpha} \phi - \frac{1}{2} \left( 1 + \phi^2 \right) I_{\alpha} \partial_0 \phi, \tag{7.14a}$$

$$\tilde{\Gamma}_{\alpha 5}^j = \frac{1}{2}\phi^2 k F_{\alpha}^j + \frac{1}{2}I_{\alpha} \partial_j \phi. \quad (7.14b)$$

Contrasting, we see that the former contains  $F_{\alpha 0}$  while the latter contains  $F_{\alpha}^j$  with a raised index. To properly compare these, we need to carefully raise the time index in (7.14a). To do this, we again 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  $\partial_{\alpha}$  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:

$$F_{\gamma \nu}^{\mu} = G^{\mu\sigma} F_{\gamma \sigma \nu} = G^{\mu\sigma} \partial_{\sigma} A_{\gamma \nu} - G^{\mu\sigma} \partial_{\nu} A_{\gamma \sigma} = (\eta^{\mu\sigma} + \Phi^{\mu} \Phi^{\sigma}) \partial_{\sigma} A_{\gamma \nu} - (\eta^{\mu\sigma} + \Phi^{\mu} \Phi^{\sigma}) \partial_{\nu} A_{\gamma \sigma}. \quad (7.15)$$

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

$$\begin{aligned} F_{\gamma \nu}^0 &= (\eta^{0\sigma} \partial_{\sigma} A_{\gamma \nu} + \Phi^0 \Phi^{\sigma} \partial_{\sigma} A_{\gamma \nu}) - (\eta^{0\sigma} \partial_{\nu} A_{\gamma \sigma} + \Phi^0 \Phi^{\sigma} \partial_{\nu} A_{\gamma \sigma}) \\ &= (\partial_0 A_{\gamma \nu} + \Phi^0 \Phi^0 \partial_0 A_{\gamma \nu}) - (\partial_{\nu} A_{\gamma 0} + \Phi^0 \Phi^0 \partial_{\nu} A_{\gamma 0}) \\ &= (1 + \phi^2) (\partial_0 A_{\gamma \nu} - \partial_{\nu} A_{\gamma 0}) = (1 + \phi^2) F_{\gamma 0 \nu} \end{aligned} \quad (7.16)$$

Using the gauge transformation  $F_{\gamma \mu \nu} \rightarrow F_{\mu \nu}$  discussed prior to (7.11) to write this as  $F_{\alpha}^0 = (1 + \phi^2) F_{\alpha 0}$ , then using this in (7.14a), now reduces the equation pair (7.14) to:

$$\tilde{\Gamma}_{\alpha 5}^0 = \frac{1}{2}\phi^2 k F_{\alpha}^0 + \frac{1}{2}(1 - \phi^2) \partial_{\alpha} \phi - \frac{1}{2}(1 + \phi^2) I_{\alpha} \partial_0 \phi, \quad (7.17a)$$

$$\tilde{\Gamma}_{\alpha 5}^j = \frac{1}{2}\phi^2 k F_{\alpha}^j + \frac{1}{2}I_{\alpha} \partial_j \phi. \quad (7.17b)$$

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  $\tilde{\Gamma}_{\alpha 5}^M$  term, we first write (7.4) as:

$$\frac{d^2 x^M}{c^2 d\tau^2} = -2\tilde{\Gamma}_{\alpha 5}^M \frac{dx^{\alpha}}{cd\tau} \frac{dx^5}{cd\tau} + \dots \quad (7.18)$$

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 (7.17) with  $F_{\alpha}^{\mu} = -F^{\mu}_{\alpha}$  and  $I_{\alpha} = (1 \quad \mathbf{0})$ . Importantly, we also use the differential chain rule on the  $\phi$  terms. We thus obtain:

$$\begin{aligned} \frac{d^2 x^0}{c^2 d\tau^2} &= -2\tilde{\Gamma}_{\alpha 5}^0 \frac{dx^\alpha}{cd\tau} \frac{dx^5}{cd\tau} + \dots = -\left(\phi^2 k F_\alpha^0 + (1-\phi^2)\partial_\alpha \phi - (1+\phi^2)I_\alpha \partial_0 \phi\right) \frac{dx^\alpha}{cd\tau} \frac{dx^5}{cd\tau} + \dots \\ &= \phi^2 k \frac{dx^5}{cd\tau} F_\alpha^0 \frac{dx^\alpha}{cd\tau} + 2 \frac{dx^5}{cd\tau} \phi^2 \frac{d\phi}{cd\tau} + \dots \end{aligned} \quad (7.19a)$$

$$\begin{aligned} \frac{d^2 x^j}{c^2 d\tau^2} &= -2\tilde{\Gamma}_{\alpha 5}^j \frac{dx^\alpha}{cd\tau} \frac{dx^5}{cd\tau} + \dots = -\left(\phi^2 k F_\alpha^j + I_\alpha \partial_j \phi\right) \frac{dx^\alpha}{cd\tau} \frac{dx^5}{cd\tau} \\ &= \phi^2 k \frac{dx^5}{cd\tau} F_\alpha^j \frac{dx^\alpha}{cd\tau} - \frac{dx^5}{cd\tau} \frac{dx^0}{dx^j} \frac{d\phi}{cd\tau} + \dots \end{aligned} \quad (7.19b)$$

In both of the above, for the scalar we find a *derivative along the curve*,  $d\phi/cd\tau$ . Note further that in (7.19b) 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 (7.19a) 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  $\phi$ .

This raises an important question and gives us our first piece of solid information about the *physical nature* of the Kaluza-Klein scalar  $\phi$ : Without their  $d\phi/cd\tau$  terms, (7.19) both easily consolidate into  $d^2 x^\mu / c^2 d\tau^2 = (\phi^2 k dx^5 / cd\tau) F^\mu_\alpha dx^\alpha / cd\tau$  following which we can make the usual ‘‘Kaluza miracle’’ association with the Lorentz Force law. However, with this term, if  $\phi$  is a *material* field or particle which can be Lorentz transformed to a rest frame with  $v^j = \mathbf{v} = 0$ , then we have a problem, because the  $d\phi/cd\tau$  term in (7.19b) will become infinite because  $dx^0/dx^j = 1/\mathbf{v} = \infty$ , causing the space acceleration to likewise become infinite. The only way to avoid this problem, is to understand the scalar  $\phi$  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  $d^2 x^\mu / c^2 d\tau^2 = (\phi^2 k dx^5 / cd\tau) F^\mu_\alpha dx^\alpha / cd\tau$ . Therefore, we use this observational evidence in view of (7.19b) to deduce that *the Kaluza-Klein scalar  $\phi$  must be luminous*. This also provides some additional fundamental perspective on Kaluza-Klein theory:

Referring back to the Kaluza-Klein metric tensors (1.1) and the DKK metric tensors derived here in (3.13) and (4.22), and to the discussion in the preface, we note again that the ingredients of the usual metric tensor (1.1) are the three fields  $g_{\mu\nu}$ ,  $A_\mu$  and  $\phi$ , which, respectively are a rank-2 symmetric tensor, a vector, and a scalar, all in ordinary four-dimensional spacetime. In quantum field theory,  $g_{\mu\nu}$  is associated with the spin-2 graviton, which is massless and luminous. But there is no *a priori* requirement that  $A_\mu$  or  $\phi$  be massless and luminous. This changed for  $A_\mu$  at (2.10), however, when the requirement for Dirac-level covariance merely in ordinary spacetime required us to set  $A_\mu \mapsto A_{\gamma\mu}$  for a massless, luminous, spin-1 photon. Now, to

bar (7.19b) from producing an infinite space acceleration when the scalar  $\phi$  is at rest, we are also required to have  $\phi$  become a massless, luminous spin-0 scalar boson which can never be at rest. The upshot, is that all three of the fields  $g_{\mu\nu}$ ,  $A_\mu$  and  $\phi$  in the DKK metric tensors (3.13) and (4.22) must now be massless and luminous, with respective boson spins 2, 1 and 0. Consequently, *the DKK metric tensors are constructed entirely from massless, luminous, bosonic fields.*

To implement this luminosity for  $\phi$ , we first write the four-dimensional spacetime metric for a luminous particle such as the photon, and now also the scalar  $\phi$ , 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:

$$\frac{dx^0}{dx^j} = -\frac{dx_j}{dx_0}. \quad (7.20)$$

This is the term of interest in (7.19b). Now, we want to raise indexes on the right side of (7.20) 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 that:

$$\begin{aligned} dx_0 &= G_{0\nu} dx^\nu = (\eta_{0\nu} + \phi^2 k^2 A_{\gamma 0} A_{\gamma\nu}) dx^\nu = \eta_{0\nu} dx^\nu = dx^0 \\ dx_j &= G_{j\nu} dx^\nu = (\eta_{j\nu} + \phi^2 k^2 A_{\gamma j} A_{\gamma\nu}) dx^\nu = -dx^j + \phi^2 k^2 A_\gamma^j A_\gamma^k dx^k \end{aligned} \quad (7.21)$$

Using the above in (7.20) then yields the *luminous* particle relation:

$$\frac{dx^0}{dx^j} = \frac{dx^j}{dx^0} - \phi^2 k^2 A_\gamma^j A_\gamma^k \frac{dx^k}{dx^0} = \hat{u}^j - \phi^2 k^2 A_\gamma^j A_\gamma^k \hat{u}^k, \quad (7.22)$$

were we 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  $\phi$ .

Inserting (7.22) for a luminous scalar into (7.19b) which first highlighted why the scalar must be luminous, then produces:

$$\frac{d^2 x^j}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F_\alpha^j \frac{dx^\alpha}{cd\tau} - \hat{u}^j \frac{dx^5}{cd\tau} \frac{d\phi}{cd\tau} + \frac{dx^5}{cd\tau} \phi^2 k^2 A_\gamma^j A_\gamma^k \hat{u}^k \frac{d\phi}{cd\tau} \quad (7.23)$$

As we did starting at (7.14) we then set  $A_{\gamma\alpha} \equiv 0$  because the gauge vector for a single photon will be swamped by the innumerable photons contained in the classical field strength  $F_\alpha^j$ . As a result, using (7.23) with  $A_{\gamma\alpha} \equiv 0$ , we find that (7.19) together now become:

$$\frac{d^2 x^0}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F_\alpha^0 \frac{dx^\alpha}{cd\tau} + 2 \frac{dx^5}{cd\tau} \phi^2 \frac{d\phi}{cd\tau} \quad (7.24a)$$

$$\frac{d^2 x^j}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F^j{}_{\alpha} \frac{dx^{\alpha}}{cd\tau} - \hat{u}^j \frac{dx^5}{cd\tau} \frac{d\phi}{cd\tau} \quad (7.24b)$$

In (7.24b),  $\phi$  is now made luminous, and there can no longer be an infinite space acceleration.

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

$$\frac{d^2 x^{\mu}}{c^2 d\tau^2} = \frac{q}{mc^2} F^{\mu}{}_{\alpha} \frac{dx^{\alpha}}{cdt}. \quad (7.25)$$

We start with the space components in (7.24b) combined with  $\mu = j$  in (7.25) and use these to *define* the association:

$$\frac{d^2 x^j}{c^2 d\tau^2} = \phi^2 k \frac{dx^5}{cd\tau} F^j{}_{\alpha} \frac{dx^{\alpha}}{cd\tau} - \hat{u}^j \frac{dx^5}{cd\tau} \frac{d\phi}{cd\tau} \equiv \frac{q}{mc^2} F^j{}_{\alpha} \frac{dx^{\alpha}}{cdt}. \quad (7.26)$$

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:

$$\boxed{\phi^2 k \frac{dx^5}{cd\tau} = \phi^2 k \frac{dt^5}{d\tau} = \frac{q}{mc^2}; \quad \frac{dx^5}{cd\tau} = \frac{dt^5}{d\tau} = \frac{q}{\phi^2 k mc^2}}. \quad (7.27)$$

This is why electric charge – and to be precise, the charge-to-rest energy ratio  $q/mc^2$  – 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, tiny curled-up fourth space dimension*. Rather, it is understood as being related to *rate of time flow  $dt^5/d\tau$  in a second time dimension*, with a factor  $\phi^2 k$  originating in the  $G_{\mu 5} = G_{5\mu}$  components of the usual Kaluza-Klein metric tensor (1.1) setting the proportionality.

Next, we substitute the above for  $dx^5/cd\tau$  in each of (7.24) and reduce to obtain:

$$\frac{d^2 x^0}{c^2 d\tau^2} = \frac{q}{mc^2} F^0{}_{\alpha} \frac{dx^{\alpha}}{cd\tau} + 2 \frac{q}{kmc^2} \frac{d\phi}{cd\tau} \quad (7.28a)$$

$$\frac{d^2 x^j}{c^2 d\tau^2} = \frac{q}{mc^2} F^j{}_{\alpha} \frac{dx^{\alpha}}{cd\tau} - \frac{\hat{u}^j}{\phi^2} \frac{q}{kmc^2} \frac{d\phi}{cd\tau} \quad (7.28b)$$

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, one might suppose that the luminous  $\phi$  is an extremely small field  $\phi \cong 0$  with  $d\phi/d\tau \cong 0$ , in order to minimize the physical impact of these final terms to the point that it is not observable. But this is problematic for two reasons: First, if  $k$  turns out to be the extremely small ratio  $k = (2/c^2)\sqrt{G/k_e}$  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 (7.28) 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  $\phi$  extremely small, the presence of  $1/\phi^2$  in (7.28b) still causes a problem, because an extremely small  $\phi \rightarrow 0$  implies an extremely large  $1/\phi^2 \rightarrow \infty$ . Ironically, the  $1/\phi^2$  which causes  $G^{MN} \rightarrow \infty$  in the usual Kaluza-Klein metric tensor (1.1) for very small  $\phi$  – which problem was solved by the non-singular (4.22) – nevertheless still persists, because of its appearance in (7.28b). 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 (7.28) both condense precisely into the Lorentz Force motion.

Now, on first appearance, the thought of requiring that  $d\phi/d\tau = 0$  seems to suggest that  $\phi$  must be a constant field with no gradient, which as pointed out in [11] imposes unwarranted constraints on the electromagnetic field. Indeed, this also defeats the purpose of a “field” if that field has to be constant. But in (7.28),  $d\phi/d\tau$  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  $d\phi/d\tau = (\partial\phi/\partial x^\mu)(dx^\mu/d\tau) = \partial_\mu\phi u^\mu$  with the usual four-velocity  $u^\mu \equiv dx^\mu/d\tau$ . Moreover, we have now learned at (7.19b) that  $\phi$  must be a massless, luminous scalar field, which requirement has been embedded in (7.28b). So, *this derivative along the curve will always be taken in frames of reference which travel with the luminous field. Moreover, these luminous reference frames can never 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  $d\phi/d\tau = 0$  in the luminous reference frame “along the curve” simultaneously with a non-zero gradient  $\partial_\mu\phi \neq 0$  defined from the coordinates of a material observer. As we now shall elaborate, this solves the “constant scalar 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  $\phi$ . Indeed, what we shall now learn about the Kaluza-Klein scalar  $\phi$  by requiring the derivative along its luminous propagation curve to be  $d\phi/d\tau = 0$ , will enable us to connect this luminous  $\phi$  to the massive Higgs boson  $h$  of the standard model, and in turn, will lay the foundation for the theory of fermion masses in Part II of this paper.

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

In view of the foregoing, we take the final step of connecting (7.28) to the observed Lorentz Force motion by formally setting the derivative along the curve for  $\phi$  to zero, thus:

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

With this, both of (7.28) immediately consolidate into and become synonymous with the Lorentz Force motion (7.25), and we achieve our goal of showing that the DKK metric tensors (3.13), (4.22) do indeed maintain the Kaluza “miracle” as regards the Lorentz Force motion of classical electromagnetism. From the standpoint of scientific method, we can take (7.28) together with (7.25) 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  $\phi$ .

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

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

Now, the four-dimensional spacetime metric (7.5) was formulated using the DKK metric tensor (3.13), and also uses (2.12). Next, let’s apply (7.5) with  $g_{\mu\nu} = \eta_{\mu\nu}$  to a luminous particle which by definition has  $d\tau^2 = 0$ , as such:

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

We already used a variant of this to obtain (7.22). Then, also appending a  $\phi^2$  and using an overall minus sign which will become useful momentarily, we restructure this to:

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

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:

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

Above,  $\tilde{\lambda} \equiv \lambda/2\pi$  is a reduced wavelength of the scalar  $\phi$ , needed and therefore introduced to balance the  $1/\text{length}^2$  dimension of  $\partial_\mu\phi\partial_\nu\phi$  with the dimensionless  $G_{\mu\nu} = \eta_{\mu\nu} + \phi^2 k^2 A_{\gamma\mu} A_{\gamma\nu}$ . To be clear: by (8.5) we are requiring the gradient  $\partial_\mu\phi$  (squared) to be non-zero from a *material* reference frame, then using this together with (8.4) to simultaneously allow (8.2) for  $d\phi/d\tau$  (squared) to be zero from the *luminous* reference frame. Now, all we need to do is determine a first-order (not squared)  $\partial_\mu\phi$  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 employed 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  $\Gamma_\mu = (\gamma_0 + kA_{\gamma_j} \gamma_j \quad \gamma_j + kA_{\gamma_j} \gamma_0)$  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  $\Gamma_\mu$  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:

$$\lambda \partial_\mu \phi = -i \Gamma_\mu \phi. \quad (8.6)$$

Now, just as the photon gauge field (2.11a) contains a Fourier kernel  $\exp(-iq_\sigma x^\sigma / \hbar)$  where  $q^\mu$  is the photon energy-momentum, and the fermion wavefunction used in (5.6) contains a Fourier kernel  $\exp(-ip_\Sigma x^\Sigma / \hbar)$  with a fermion five-momentum  $p^M$ , let us specify a Fourier kernel  $\exp(-is_\Sigma x^\Sigma / \hbar)$  with a five-dimensional  $s^M$  which we regard as the five-momentum of the luminous scalar  $\phi$ . Moreover, because  $\phi$  is dimensionless and so too is  $\exp(-is_\Sigma x^\Sigma / \hbar)$ , let us then assemble these ingredients to define:

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

Above,  $\exp(-is_\Sigma x^\Sigma / \hbar)$  is the above-noted Fourier kernel in five dimensions, while  $\phi_1 + i\phi_2$  is a dimensionless and complex-valued field. This complex field, 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  $\phi_h \equiv \frac{1}{\sqrt{2}} (\phi_{1h} + i\phi_{2h})$ . Specifically,  $\phi_1$  and  $\phi_2$  introduce two degrees of freedom which can be used to give mass to otherwise massless objects via a Higgs-like mechanism. Because  $(\phi_1 + i\phi_2)^* (\phi_1 + i\phi_2) = \phi_1^2 + \phi_2^2$ , the symmetry of the “circle” in the complex Euler plane of  $\phi_1$  and  $\phi_2$  can always be broken by choosing the  $\phi_2 = 0$  orientation, see, e.g., Figure 13.5 in [20], also analogously to the Higgs mechanism. In the standard model, once the symmetry is broken, the scalar field is expanded about the vacuum having an expectation value  $v$  via  $\phi_h(x^\mu) = \frac{1}{\sqrt{2}} (v + h(x^\mu))$ , with fluctuations provided by the Higgs field  $h(x^\mu)$ . In the standard model, this “vev” is taken to be  $v = 246.2196508$  GeV, namely, the Fermi vacuum expectation associated with the Fermi coupling via  $1/2v^2 = G_F / \sqrt{2} (\hbar c)^3$  based on the latest PDG data [21]. In (8.7), which we will connect directly to the standard model Higgs scalar in sections 11 through 13, the kernel  $\exp(-is_\Sigma x^\Sigma / \hbar) = \cos(s_\Sigma x^\Sigma / \hbar) - i \sin(s_\Sigma x^\Sigma / \hbar)$  provides two additional degrees of freedom in the complex Euler plane through which the angle  $\theta = s_\Sigma x^\Sigma / \hbar$  is oriented. In total, (8.7) thus provides four degrees of freedom.

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

$$\partial_M \phi = \left( \frac{\partial_M \phi_1 + i \partial_M \phi_2}{\phi_1 + i \phi_2} - i \frac{s_M}{\hbar} \right) \phi. \quad (8.8)$$

If we then covariantly extend (8.6) into the fifth dimension in the form of  $\tilde{\lambda} \partial_M \phi = -i \Gamma_M \phi$  and then apply (8.8) we find that:

$$\tilde{\lambda} \partial_M \phi = \tilde{\lambda} \left( \frac{\partial_M \phi_1 + i \partial_M \phi_2}{\phi_1 + i \phi_2} - i \frac{s_M}{\hbar} \right) \phi = -i \Gamma_M \phi. \quad (8.9)$$

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

$$c s_M = \hbar \omega \Gamma_M - i \hbar c \frac{\partial_M \phi_1 + i \partial_M \phi_2}{\phi_1 + i \phi_2}. \quad (8.10)$$

The time component of  $\hbar \omega \Gamma_0 = \hbar \omega (\gamma_0 + k A_{\gamma_j} \gamma_j)$  within the energy component  $c s_0$  above is positive for the upper (particle) components of  $\text{diag}(\gamma_0) = (+I, -I)$  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:

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

The product separation of exponentials in the lower line is possible in view of the Zassenhaus-Baker-Campbell-Hausdorff relation  $\exp(A + B) = \exp A \exp B \exp(-[A, B]/2) \dots$  because although  $\Gamma_\Sigma x^\Sigma$  is a 4x4 matrix operator, the second additive term inside the top line exponential is a 4x4 diagonal matrix which does commute with the first term, i.e.,  $[A, B] = 0$ . Because (8.10) contains an energy  $E = \hbar \omega = hf$ , we now must interpret  $\phi$  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  $\Gamma_\Sigma$  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  $\exp(-i\omega\Gamma_\Sigma x^\Sigma / c)$  into sine and cosine terms using the Maclaurin series  $\exp(-ix) = 1 - ix - \frac{1}{2!}x^2 + i\frac{1}{3!}x^3 + \frac{1}{4!}x^4 - \frac{1}{5!}ix^5 - \dots = \cos x - i\sin x$ . To do so, we first use the anticommutator (3.1) to calculate the square:

$$\left(\Gamma_\Sigma x^\Sigma\right)^2 = \frac{1}{2}\{\Gamma_M\Gamma_N + \Gamma_N\Gamma_M\}x^M x^N = G_{MN}x^M x^N \equiv c^2 T^2, \quad (8.12)$$

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

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

To get to the sin term in the bottom line, we multiplied through by  $1 = \omega T / \omega T$  in the top line. Inserting this into (8.11) yields the final expression for the luminous, dimensionless, massless Kaluza-Klein scalar, namely:

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

With Dirac indexes made explicit, the Dirac operator characteristics of  $\phi = \phi_{AB}$  are now seen to be isolated in and stem from the  $\Gamma_\Sigma x^\Sigma = \Gamma_{\Sigma AB} x^\Sigma$  matrix which multiplies the  $\sin(\omega T)$  term.

From (8.14) it is straightforward to then calculate the square modulus:

$$|\phi|^2 = \phi^* \phi = \frac{1}{2}(\phi_1^2 + \phi_2^2) \exp\left(-\frac{\partial_\Sigma (\phi_1^2 + \phi_2^2)}{\phi_1^2 + \phi_2^2} x^\Sigma\right) \quad (8.15)$$

Above, the first of the three multiplicative terms in (8.14) produce  $\frac{1}{2}(\phi_1^2 + \phi_2^2)$ , which although dimensionless, creates an analogue to the symmetry-breaking circle of the Higgs mechanism. The second of these three terms in (8.14) in view of (8.12) produces a 1. The third term first processed

using  $1/(\phi_1 + i\phi_2) = (\phi_1 - i\phi_2)/(\phi_1^2 + \phi_2^2)$  which is a standard trick for handling complex numbers in a denominator, and then after reduction using  $2(\phi_1\partial_z\phi_1 + \phi_2\partial_z\phi_2) = \partial_z(\phi_1^2 + \phi_2^2)$ . This reveals a  $\phi_1^2 + \phi_2^2$  “Mexican hat” “circle” analogous to what we obtain in Higgs theory to break a local U(1) gauge symmetry by choosing a  $\phi_2 = 0$  orientation in that circle. see, e.g., sections 14.7 and 14.8 in [20]. But here, the overall magnitude of  $\phi^*\phi$  is also modified by the exponential which contains the five-gradient  $\partial_z(\phi_1^2 + \phi_2^2)$  of the magnitude of this circle. We shall later see from (11.12) to (11.22) and thereafter at (13.7) and in Figure 1, how the gradient terms in (8.14) and (8.15) are at the heart of how the Higgs mechanism extracts energy from the Fermi vacuum to give rest mass to fermions. But as noted after (8.7), the kernel  $\exp(-is_zx^z/\hbar)$  which in (8.14) now contains the angle  $\theta = \omega T$ , provides additional third and fourth degrees of freedom. In sections 11 through 13 we shall see that (8.14) above further simplifies when we geometrize the fermion rest masses and break the symmetry such that the two degrees of freedom in  $\phi_1$  and  $\phi_2$  give rest masses to the gauge and the Higgs bosons in the usual way, while two degrees of freedom in the complex Euler plane with orientation angle  $\theta = \omega T$  give rest masses to the fermions.

For now, 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 of  $\phi$  can be non-zero,  $\partial_M\phi \neq 0$ , simultaneously with its derivative along the curve being 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,  $G_{MN} = G_{M N A B}$  owing to the structure (8.14) of the scalar field which sits 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. But as we shall see at (11.12), the symmetry of this Dirac-indexed scalar is broken to “hide” this internal Dirac structure, so that the DKK metric tensor (3.11) reverts to its present form without needing implied Dirac indexes. And of course, with (8.14) being derived to enable  $d\phi/d\tau = 0$ , (7.28) become synonymous with the electrodynamic Lorentz force motion, which is one of the key touchstones of Kaluza-Klein theory.

In conclusion, 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 see in Part II of this paper, once connected to the Higgs field, it can be used to spontaneously break symmetry, whereby the two degrees of freedom in the amplitude  $(\phi_1 + i\phi_2)/\sqrt{2}$  enable the gauge bosons and the DKK scalar to become massive with the latter

corresponding to the Higgs boson, and the additional degrees of freedom in the  $\theta = \omega T$  plane provide masses to fermions. This, in turn, enable us to develop a theory of fermion masses and mixing angles which fits all the known the experimental data, and which in a number of cases, refines the known data by two or more orders of magnitude. Moreover, a better understanding is obtained for the type of Higgs production paired with top quarks reported out of CERN in the middle of 2018 [22], [23], [24].

The next step in the development of the luminous Kaluza-Klein scalar (8.14) entails spontaneous symmetry breaking which will be pursued in section 11, followed by connecting this scalar to the standard model Higgs boson which will be pursued in sections 12 and 13, followed by the development of a theory of fermion masses and mixing angles and beta decays which will commence in section 14 and proceed throughout the balance of this paper. Because these next steps will develop and then cross the bridge from Dirac-Kaluza-Klein theory to elementary particle physics, we now pause the further development of the scalar (8.14) for the next two sections, so that we can conclude our exposition of Kaluza-Klein theory on its own terms before proceeding to how this connects to the elementary particle physics of the standard model. In short, in the next sections we shall conclude the first of the “two papers in one” referenced in the preface, so that we can then proceed on to the second paper.

## 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 regarding Kaluza-Klein theory. As has been previously pointed out, in the circumstance where  $A_{\gamma j} = 0$  and  $\phi = 0$  (which using the vector (3.12) is simply  $\Phi_\mu = 0$ ), then when  $g_{\mu\nu} = \eta_{\mu\nu}$ , (3.13) reduces to  $\text{diag}(G_{MN}) = (+1, -1, -1, -1, +1)$  with  $|G_{MN}| = -1$ . And we saw at (6.3) that this result does not change at all, even when all components of  $\Phi_\mu \neq 0$ . But in the similar circumstance where  $A_j = 0$  and  $\phi = 0$ , the usual Kaluza-Klein metric tensor (1.1) reduces to  $\text{diag}(G_{MN}) = (+1, -1, -1, -1, 0)$  with a determinant  $|G_{MN}| = 0$ . And even for  $A_{\gamma j} \neq 0$ , so long as  $\phi = 0$  we still have  $|G_{MN}| = 0$ . This of course means that whenever  $\phi = 0$  the Kaluza-Klein metric tensor (1.1) is not-invertible and therefore becomes singular. Again, this may be seen directly from the fact that when we set  $\phi = 0$  in (1.1) we obtain  $G^{55} = g_{\alpha\beta} A^\alpha A^\beta + 1/\phi^2 = g_{\alpha\beta} A^\alpha A^\beta + \infty$ . This degeneracy leads to a number of interrelated ills which have hobbled Kaluza-Klein as a viable theory of the natural world for an entire century:

First, with  $G_{55} = \phi^2$ , the scalar field  $\phi$  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  $\phi$  which would not exist if it was not doing “double duty” as both a scalar field with its own physics, and as a

structural element required to maintain the non-degeneracy of Minkowski spacetime extended to five dimensions.

Second, a metric that becomes singular when one of the physical fields becomes zero is thoroughly at odds with the most bedrock principles of theoretical physics first advanced by Einstein in the General Theory of Relativity [2]. In the General Theory, the space and time of real physical spacetime combine into a Minkowskian “union of these two” [25], with a flat spacetime signature  $\text{diag}(\eta_{\mu\nu}) = (+1, -1, -1, -1)$  that is structurally sound *even in the absence of any fields whatsoever*. But what does one make of a signature that is  $\text{diag}(\eta_{MN}) = (+1, -1, -1, -1, \phi^2)$  with a determinant  $|\eta_{MN}| = -\phi^2$  when  $g_{\mu\nu} = \eta_{\mu\nu}$  and  $A_{\gamma k} = 0$ ? How is one to explain the physicality of a Minkowski signature which, because  $G_{55} = \phi^2$ , is based entirely on a field, rather than being either a timelike +1 or a spacelike -1 Pythagorean metric component? Specifically, the Minkowski signature defines the *flat tangent spacetime* to curved spacetime at each event. So, how can a tangent space which *by definition* should not be curved, be dependent upon a field  $\phi$  which if it has even the slightest modicum of energy will cause curvature? This is an internal logical contradiction of the metric tensor (1.1) that has plagued Kaluza-Klein theory for an entire century. And it leads to such hard-to-justify oddities as a fifth dimensional metric tensor component  $G_{55} = \phi^2$  and determinant  $|\eta_{MN}| = -\phi^2$  which dilates or contracts (hence the name “dilaton”) in accordance with the behavior of  $\phi^2$ , and which vanishes entirely when  $\phi=0$  to give rise to a singular, degenerate theory. This all contradicts the fundamental theoretical proposition of the General Theory of Relativity that one starts with a non-singular flat spacetime metric tensor which has a Pythagorean signature with timelike and spacelike dimensions, then represents physical fields and their energies by Riemannian curvatures which amount to deviations against this flat spacetime background. If there is a singularity to be had, it should arise from the Einstein field equation for an energy tensor source giving rise to exceptionally-sharp curvatures (for example, as with black holes), not from a flat tangent-space metric tensor devoid of sources and fields.

Third, the DKK metric tensor (3.13) is obtained by *requiring* an “operator square root” deconstruction of the Kaluza-Klein metric tensor into a set of Dirac matrices obeying  $\frac{1}{2}\{\Gamma_M, \Gamma_N\} \equiv G_{MN}$  in (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 with the metric components  $G_{\mu 5} = G_{5\mu} = \phi^2 k A_\mu$  which lead to the Kaluza “miracle,” if at the same time  $G_{05} = G_{50} = 0$  and  $G_{55} = \phi^2$ , all as in (1.1). Rather, general 5-dimensional covariance with the “miracle” foundation  $G_{\mu 5} = G_{5\mu} = \phi^2 k A_\mu$  *requires* that  $G_{05} = G_{50} = \phi$  and  $G_{55} = 1 + \phi^2$  as in (3.13), see also (3.11). Further, to have general covariance *even in four spacetime dimensions alone*, we are *required* to gauge the electromagnetic potential to that of the photon, as first revealed at (2.10). 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.* No Dirac covariance  $\rightarrow$  no Dirac equation  $\rightarrow$  no fermions! Such an omission – even without any of its other known ills – most-assuredly renders the usual Kaluza-Klein metric (1.1) “unphysical.”

Fourth and finally, there is the century-old demand which remains unmet to this date: “Show me the fifth dimension!” There is no detected observational evidence at all to support the physical reality of the fifth dimension, either in direct experience or empirical experimentation, at least in the form specified by (1.1) which requires this fifth dimension to be spacelike. Nor do any later studies of Kaluza-Klein to date appear to have managed to remedy these problems. If anything, they only compound these problems by taking the fifth dimension *presumed to be spacelike*, developing this into a tiny curled up “string” orders of magnitude smaller than anything that one can ever hope to observe, and then adding even more spacelike dimensions.

In these regards, it should be obvious to anyone familiar with the Dirac  $\gamma_\mu$  and the axial operator  $\gamma_5 \equiv -i\gamma_0\gamma_1\gamma_2\gamma_3$  that one may easily use an anticommutator  $\eta_{MN} \equiv \frac{1}{2}\{\gamma_M, \gamma_N\}$  to form a five-dimensional metric tensor with  $\text{diag}(\eta_{MN}) = (+1, -1, -1, -1, +1)$  which has a Minkowski tangent signature with two timelike and three spacelike dimensions, such as in (3.13), see also (3.11). But it is not at all obvious how one might proceed to regard  $\gamma_5$  as the generator of a *truly-physical fifth dimension* on an absolute par with the generators  $\gamma_\mu$  of the four truly-physical dimensions which are time and space. This is because as pointed out in toward the end of [11], with  $G_{55} = \phi^2$  in the Kaluza-Klein metric tensor (1.1), if we require electromagnetic energy densities to be positive, then the fifth-dimension must have a *spacelike* signature. And this directly contradicts making  $\gamma_5$  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  $\gamma_5$  matrix, the Kaluza-Klein metric tensor (1.1) does *not* furnish a theoretical basis for associating  $\gamma_5$  with a fifth dimension, at the very least because of this timelike-versus-spacelike contradiction. This is yet another problem stemming from having  $\phi$  carry the burden of maintaining the fifth-dimensional signature and the fundamental Pythagorean character of the Minkowski tangent space, in addition to being a field within the spacetime-plus-fifth-dimension geometry.

This inability to connect  $\gamma_5$  with the fifth dimension persists, notwithstanding the clear observational evidence that  $\gamma_5$  has a multitude of observable physical impacts. The reality of  $\gamma_5$  is most notable in the elementary fermions that separate into  $\psi = \psi_R + \psi_L$  using the factor  $\frac{1}{2}(1 \pm \gamma_5)$  for right- and left-chirality; in the weak interaction of fermions always being 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 [26], all of which require the use of  $\gamma_5$  to underpin their theoretical origins. So  $\gamma_5$  is real and physical, as would therefore be any fifth dimension which can be *properly-connected* with  $\gamma_5$ . But again, there is a contradiction between the spacelike signature emerging from Kaluza-Klein’s (1.1) and the timelike signature of  $\gamma_5\gamma_5 = 1$ , so such a connection is simply not feasible based on (1.1). Instead, decades have been spent chasing tiny curled-up spacelike dimensions that have not a shred of observational support.

With all of these problems and perplexities, the reason Kaluza-Klein has nonetheless managed to maintain some measure of viability, is because of its “miracle” of providing an entirely-geometrodynamical foundation for Maxwell’s electrodynamics, the Lorentz Force Law, and the Maxwell stress-energy tensor. And also, as will be reviewed in the next section, because it offers the possibility of the fifth dimension being a “matter dimension” as has long been advocated by the 5D Space-Time-Matter Consortium [18].

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. Nor has it been possible to make credible sense of the dilation and contraction of that dimension based on the behavior of a scalar field, and especially of the Minkowski tensor singularity that occurs when  $\phi = 0$ . On the other hand, Dirac theory has an eminently-physical  $\gamma_5$  with pervasive observational manifestations on an equal footing with  $\gamma_\mu$ , but it has been impossible to connect this  $\gamma_5$  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  $\gamma_5$  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). Moreover, because there is no five-dimensional covariance, there is no way to theoretically represent Dirac fermions. All of these problems, weigh against what is otherwise the “miracle” of the geometrodynamical foundation which Kaluza-Klein does lay for electrodynamics.

*To maintain general 5-dimensional 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 that has been done, the metric tensors (3.13) and (4.22) lead to a whole other picture, and the problems reviewed above all evaporate. Following the same sequence as above, let us summarize how the problems and perplexities reviewed above are now “fixed” by (3.13) and (4.22):

First, with  $G_{55} = 1 + \phi^2$  the metric signature is now decoupled from the energy requirements for  $\phi$ . And with  $|G_{MN}| = |g_{MN}|$  from (6.3) the metric tensor determinant becomes entirely independent of both  $A_{\gamma\mu}$  and  $\phi$ . This means that  $\phi$  is no longer doing “double duty” as both a scalar field and as a structural element required to avoid a Minkowski metric singularity when  $\phi = 0$ . Rather, the metric is stabilized by the 1 in  $G_{55} = 1 + \phi^2$ , and as we see in (3.10), this  $1 = \gamma_5 \gamma_5$  is directly generated by the axial operator  $\gamma_5$ .

Second, as a direct result of this, there is no longer any singularity when  $\phi = 0$ . Implicitly, the litmus test for this is (6.3), whereby the five-dimensional determinant  $G = |G_{MN}| = |g_{MN}| = g$  no matter what the values of  $A_{\gamma\mu}$  and  $\phi$  may be. And when  $g_{MN} = \eta_{MN}$  this reduces to  $G \equiv |G_{MN}| = |\eta_{MN}| = -1$ . Explicitly, this is seen in (4.20) through (4.22) where  $G^{55} = g^{55}$  generally, and in flat spacetime where  $G^{55} = \eta^{55} = \gamma^5 \gamma^5 = 1$  is well-behaved, always. And as with the

determinant,  $G^{55}$  is entirely independent of  $A_{\gamma\mu}$  and  $\phi$ , always, in contrast to the usual  $G^{MN}$  in (1.1) which is singular and badly-behaved because  $G^{55} \rightarrow \infty$  when  $\phi \rightarrow 0$ . Then, because the tangent space when all fields are zeroed out now has  $G_{55} = G^{55} = +1$ , the Dirac-Kaluza-Klein theory based on (3.13) and (4.22) is wholly consistent with the General Theory of Relativity. This is because while physical fields do correlate with curvature and strong-enough sources can lead to extreme curvature based on the Einstein Equation, when all the fields are turned off, there is still a stable, well-behaved tangent space which has a Pythagorean signature with Minkowskian timelike and spacelike dimensions.

Third, the metric tensors (3.13) and (4.22) are formed by explicitly demanding five-dimensional Dirac covariance, from which we obtain a set of  $\Gamma_M$  defined by in (1.3) by  $\frac{1}{2}\{\Gamma_M, \Gamma_N\} \equiv G_{MN}$ . Therefore, *by definition*, a 5-covariant Dirac equation (5.6) can be formed, so there is no problem of incompatibility with Dirac theory. Working from (5.6), we may then anticipate that all aspects of fermion physics can be fully accounted for. And in Part II we will show that this is indeed so, wherein the Kaluza-Klein scalar which has thus far advanced to (8.14) becomes connected to the Higgs scalar, and it thereafter becomes possible to explain the masses, mixing angles and beta decays of the elementary fermions, entirely consistently with the standard model, and in a number of cases, with improved predictive accuracy.

Fourth, perhaps most importantly in terms of the scientific method requiring concurrence between theory and observation, when  $A_{\gamma_j} = 0$  and  $\phi = 0$  and  $g_{MN} = \eta_{MN}$ , and because of the foregoing decoupling of  $\phi$  from the metric signature, we now have a timelike  $\eta_{55} = \gamma_5 \gamma_5 = +1$  which is directly generated by  $\gamma_5$ . 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 which has  $\gamma_5$  as its generator, just as  $\gamma_0$  is the generator of a truly-physical time dimension and  $\gamma_j$  are the generators of truly-physical space dimensions. And again,  $\gamma_5$  has a wealth of empirical evidence to support its reality, such as weak interaction chirality, and the observed plethora of pseudo-scalar and pseudo-vector mesons.

Further in this regard, with a tangent space  $\text{diag}(\eta_{MN}) = (+1, -1, -1, -1, +1)$  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, never-to-be-seen “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  $\gamma_5$  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  $\gamma_5$  in particle physics, regardless of any other epistemological interpretations one may also arrive at for this fifth dimension. This is in contrast to the dearth of observational evidence for a spacelike fifth dimension.

Of course, it would not be beneficial if the foregoing problem solutions came at the cost of sacrificing the Kaluza “miracle.” Of course, 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. But 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. Rather, as already seen in sections 7 and 8 where the Lorentz Force motion “miracle” is faithfully reproduced, we simply expect some extra terms and some additional phenomenology to emerge in the equations of motion and the field equations because of these modifications. One example of this additional phenomenology is the requirement deduced at (7.19b) that to fully reproduce the Lorentz Force, the Kaluza-Klein spin-0 scalar must be luminous and massless alongside of the spin-2 graviton and the spin-1 photon, leading to this scalar having the explicit form of (8.14) prior to any of the mass-producing symmetry breaking still to be reviewed in Part II. And so, we expect that 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 we have already begun to see in sections 7 and 8.

Finally, given all of the foregoing, beyond the manifold observed impacts of  $\gamma_5$  in particle physics, there is every reason to believe that using the five-dimensional Einstein equation with the DKK metric tensors (3.13) and (4.22) 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 may thereby bring us ever-closer to uncovering the truly-geometric dynamical theoretical foundation at the heart of all of nature. These geometric dynamical interests are the subject of the next and final section devoted to Dirac-Kaluza-Klein theory on its own terms, before we turn in Part II to the connection between DKK theory and the particle physics of elementary fermions emerging from the 5-dimensional Dirac equation (5.6).

## 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  $\tilde{\Gamma}_{\alpha 5}^M$  in order to study the  $\tilde{\Gamma}_{\alpha 5}^M$  term in the equation of motion (7.4), because this is the term which provides the Lorentz Force motion which becomes (7.28) once  $\phi$  is understood to be a luminous field with  $d\phi/d\tau = 0$  as in (8.1) thereafter deduced explicitly in (8.14) to which we will soon return to break symmetry. 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, to demonstrate a similar favorable result in relation to the Einstein Equation for sources and fields.

First, it is necessary to calculate all of the other connections  $\tilde{\Gamma}_{AB}^M$  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 as regards the field equation, it is also necessary to obtain the five-dimensional Riemann tensor, the Ricci tensor in lower- and mixed-index form, and the Ricci scalar:

$$\begin{aligned}
 \hat{R}^A_{\text{BMN}} &= \partial_M \tilde{\Gamma}^A_{\text{NB}} - \partial_N \tilde{\Gamma}^A_{\text{MB}} + \tilde{\Gamma}^A_{\text{M}\Sigma} \tilde{\Gamma}^\Sigma_{\text{NB}} - \tilde{\Gamma}^A_{\text{N}\Sigma} \tilde{\Gamma}^\Sigma_{\text{MB}} \\
 \hat{R}_{\text{BM}} &= \hat{R}^T_{\text{BMT}} = \partial_M \tilde{\Gamma}^T_{\text{TB}} - \partial_T \tilde{\Gamma}^T_{\text{MB}} + \tilde{\Gamma}^T_{\text{M}\Sigma} \tilde{\Gamma}^\Sigma_{\text{TB}} - \tilde{\Gamma}^T_{\text{T}\Sigma} \tilde{\Gamma}^\Sigma_{\text{MB}} \\
 \hat{R}^A_{\text{M}} &= G^{\text{AB}} \hat{R}_{\text{BM}} = G^{\text{AB}} \partial_M \tilde{\Gamma}^T_{\text{TB}} - G^{\text{AB}} \partial_T \tilde{\Gamma}^T_{\text{MB}} + G^{\text{AB}} \tilde{\Gamma}^T_{\text{M}\Sigma} \tilde{\Gamma}^\Sigma_{\text{TB}} - G^{\text{AB}} \tilde{\Gamma}^T_{\text{T}\Sigma} \tilde{\Gamma}^\Sigma_{\text{MB}} \\
 \hat{R} &= \hat{R}^A_{\text{A}} = G^{\text{AB}} \partial_A \tilde{\Gamma}^T_{\text{TB}} - G^{\text{AB}} \partial_T \tilde{\Gamma}^T_{\text{AB}} + G^{\text{AB}} \tilde{\Gamma}^T_{\text{A}\Sigma} \tilde{\Gamma}^\Sigma_{\text{TB}} - G^{\text{AB}} \tilde{\Gamma}^T_{\text{T}\Sigma} \tilde{\Gamma}^\Sigma_{\text{AB}}
 \end{aligned} \tag{10.1}$$

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

$$-\text{K} \hat{T}_{\text{MN}} = \hat{R}_{\text{MN}} - \frac{1}{2} G_{\text{MN}} \hat{R}, \tag{10.2}$$

with a suitably-dimensioned constant  $\text{K}$  related to the usual  $\kappa$  as will 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 provide some confidence that the Kaluza miracle will remain 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 as first established at (7.12), notwithstanding any additional terms,  $\tilde{\Gamma}^{\mu}_{\alpha 5} = \frac{1}{2} \eta^{\mu\sigma} \phi^2 k F_{\alpha\sigma} + \dots$  contains the electromagnetic field strength exactly as before. So, we may be comfortable that the terms needed to reproduce the Maxwell tensor  $k_e T^{\mu\nu} = -\frac{1}{4\pi} c^2 \left( F^{\mu\sigma} F_{\sigma}^{\nu} - \frac{1}{4} g^{\alpha\beta} F_{\sigma\tau} F^{\sigma\tau} \right)$  via the field equation (10.2) will be embedded in the (10.1) terms housed originally in  $\Gamma^A_{\text{M}\Sigma} \Gamma^\Sigma_{\text{NB}} - \Gamma^A_{\text{N}\Sigma} \Gamma^\Sigma_{\text{MB}}$ , because these terms are of second order in the connections. 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  $\partial_M \Gamma^A_{\text{NB}} - \partial_N \Gamma^A_{\text{MB}}$ , because these terms contain gradients of the connections tied to the field strength. Moreover, 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, because  $\left( \hat{R}^{\text{MN}} - \frac{1}{2} G^{\text{MN}} \hat{R} \right)_{;\text{M}} = 0$  which via (10.2) ensures a locally-conserved energy  $\hat{T}^{\text{MN}}_{;\text{M}} = 0$ , is contracted from the second Bianchi identity  $\hat{R}^A_{\text{BMN};\text{P}} + \hat{R}^A_{\text{BNP};\text{M}} + \hat{R}^A_{\text{BPM};\text{N}} = 0$ , which has the same cycling of indexes and the correct differential order in relation to the connections. In short – although we leave this to a future paper for detailed development – we may be comfortable based on what has already been demonstrated 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 route to the view that the fifth dimension is a matter dimension [18]. Moreover, 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  $t^5$ . So, let us briefly preview these development paths.

The Einstein-Hilbert action reviewed for example in [27], in four dimensions, is given by:

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

where  $R = R^\sigma{}_\sigma$ . The derivation of the  $-\kappa T_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$  from (10.3) is well-known. So, in five dimensions, we expect that (10.2) will emerge from changing  $R \mapsto \hat{R}$  and  $\mathcal{L}_M \mapsto \hat{\mathcal{L}}_M$  and  $g \mapsto G$ , thus extending (10.3) to:

$$\hat{S} = \int \left( (1/2K)\hat{R} + \hat{\mathcal{L}}_M \right) \sqrt{-G} d^5x = \int \left( (1/2K)\hat{R}^\sigma{}_\sigma + (1/2K)\hat{R}^5{}_5 + \hat{\mathcal{L}}_M \right) \sqrt{-G} d^5x. \quad (10.4)$$

The above also uses  $\hat{R} = \hat{R}^\Sigma{}_\Sigma = \hat{R}^\sigma{}_\sigma + \hat{R}^5{}_5$  from (10.1) and the  $G$  already obtained in (6.3). And the new  $\kappa \mapsto K \equiv \lambda\kappa$  contains some suitable length  $\lambda$  to balance the extra space dimensionality in  $d^5x$  versus  $d^4x$ . In Kaluza-Klein theory based on (1.1)  $\lambda$  is normally the radius of the compactified fourth space dimension and is very small. Here, because there is a second time dimension, this will become associated with some suitable length =  $c$ \*time.

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:

$$T_{\mu\nu} \equiv -2 \frac{\delta \mathcal{L}_M}{\delta g^{\mu\nu}} + g_{\mu\nu} \mathcal{L}_M. \quad (10.5)$$

Therefore, in the Five-Dimensional Space-Time-Matter view of [18], and comparing (10.4) to (10.3), we see that the “wood” of  $\hat{\mathcal{L}}_M$  may be discarded entirely from (10.4) by setting  $\hat{\mathcal{L}}_M = 0$  leaving behind:

$$\hat{S} = \int \left( (1/2K)\hat{R}^\sigma{}_\sigma + (1/2K)\hat{R}^5{}_5 \right) \sqrt{-G} d^5x. \quad (10.6)$$

Then we associate

$$\hat{\mathcal{L}}_M \equiv (1/2K)\hat{R}^5{}_5 \quad (10.7)$$

with the matter Lagrangian density, giving this an entirely-geometric interpretation based on the curvature component  $\hat{R}^5{}_5$ . As a result, this is now also made of “marble.”

With this, (10.4) may be simplified to the 5-dimensional “vacuum” equation (see [28] at 428 and 429):

$$\hat{S} = \int (1/2K) \hat{R} \sqrt{-G} d^5x. \quad (10.8)$$

We also anticipate that the variation of the Ricci scalar will produce the usual relation:

$$\hat{R}_{MN} = \frac{\delta \hat{R}}{\delta G^{MN}} = \frac{\delta \hat{R}^\sigma_\sigma}{\delta G^{MN}} + \frac{\delta \hat{R}^5_5}{\delta G^{MN}}, \quad (10.9)$$

but now in five dimensions. Then, the field equation (10.2) derived from varying (10.8), combined with (10.9), becomes the vacuum equation:

$$0 = \hat{R}_{MN} - \frac{1}{2} G_{MN} \hat{R} = \frac{\delta \hat{R}}{\delta G^{MN}} - \frac{1}{2} G_{MN} \hat{R}. \quad (10.10)$$

And for the energy tensor, we generalize (10.5) to five dimensions, then apply (10.7) to obtain:

$$T_{MN} \equiv -2 \frac{\delta \hat{\mathcal{L}}_M}{\delta G^{MN}} + G_{MN} \hat{\mathcal{L}}_M = -\frac{1}{K} \frac{\delta \hat{R}^5_5}{\delta G^{MN}} + \frac{1}{2K} G_{MN} \hat{R}^5_5. \quad (10.11)$$

This makes clear how the energy tensor is now constructed out of the “marble” of  $\hat{R}^5_5$ .

It is useful to multiply the above through by  $-K$ , then use  $\hat{R} = \hat{R}^\sigma_\sigma + \hat{R}^5_5$ , to obtain:

$$-KT_{MN} = \frac{\delta \hat{R}^5_5}{\delta G^{MN}} - \frac{1}{2} G_{MN} \hat{R}^5_5 = -\frac{\delta \hat{R}^\sigma_\sigma}{\delta G^{MN}} + \frac{1}{2} G_{MN} \hat{R}^\sigma_\sigma. \quad (10.12)$$

Note that the equality of the latter two expressions can be used to immediately reproduce (10.10). The spacetime components of (10.12) of course are:

$$KT_{\mu\nu} = \frac{\delta \hat{R}^\sigma_\sigma}{\delta G^{\mu\nu}} - \frac{1}{2} G_{\mu\nu} \hat{R}^\sigma_\sigma = -\frac{\delta \hat{R}^5_5}{\delta G^{\mu\nu}} + \frac{1}{2} G_{\mu\nu} \hat{R}^5_5. \quad (10.13)$$

And if we take an  $\hat{S} = \int (1/2K) \hat{R} \sqrt{-G} d^4x$  action which integrates only over the four spacetime dimensions  $d^4x$  rather than  $d^5x$ , then we can remove the extra compensating space dimension from the overall constant and revert  $K \equiv \lambda\kappa \mapsto \kappa$  and so write the above as:

$$\kappa T_{\mu\nu} = \frac{\delta \hat{R}^\sigma_\sigma}{\delta G^{\mu\nu}} - \frac{1}{2} G_{\mu\nu} \hat{R}^\sigma_\sigma = -\frac{\delta \hat{R}^5_5}{\delta G^{\mu\nu}} + \frac{1}{2} G_{\mu\nu} \hat{R}^5_5. \quad (10.14)$$

So, with the above, the drill is to first obtain  $\hat{R}_{\text{BM}}$  from (10.1), then raise an index and obtain  $\hat{R}^\sigma_\sigma$  summed only over the four dimensions of spacetime, or alternatively obtain  $\hat{R}^5_5$ . Then we calculate either the variation  $\delta\hat{R}^\sigma_\sigma/\delta G^{\mu\nu}$  or  $\delta\hat{R}^5_5/\delta G^{\mu\nu}$ , then use all of this to obtain  $T_{\mu\nu}$ . To facilitate this calculation without having to calculate either of these variations directly, one may use the spacetime portion of (10.9) together with  $\hat{R}_{\mu\nu} = \hat{R}^\text{T}_{\mu\nu\text{T}} = \hat{R}^\sigma_{\mu\nu\sigma} + \hat{R}^5_{\mu\nu 5}$  to write:

$$\hat{R}_{\mu\nu} = \frac{\delta\hat{R}}{\delta G^{\mu\nu}} = \frac{\delta\hat{R}^\sigma_\sigma}{\delta G^{\mu\nu}} + \frac{\delta\hat{R}^5_5}{\delta G^{\mu\nu}} = \hat{R}^\sigma_{\mu\nu\sigma} + \hat{R}^5_{\mu\nu 5}. \quad (10.15)$$

We can then math the structural features of the summed indexes to separate out:

$$\frac{\delta\hat{R}^\sigma_\sigma}{\delta G^{\mu\nu}} = \hat{R}^\sigma_{\mu\nu\sigma}; \quad \frac{\delta\hat{R}^5_5}{\delta G^{\mu\nu}} = \hat{R}^5_{\mu\nu 5}. \quad (10.16)$$

We then use the above in (10.14) to obtain the alternative expressions for the energy tensor which are interrelated via (10.10):

$$\kappa T_{\mu\nu} = \hat{R}^\sigma_{\mu\nu\sigma} - \frac{1}{2} G_{\mu\nu} \hat{R}^\sigma_\sigma = -\hat{R}^5_{\mu\nu 5} + \frac{1}{2} G_{\mu\nu} \hat{R}^5_5. \quad (10.17)$$

Then we directly substitute the pertinent connection components from (10.1) into the above. For the former alternative:

$$\begin{aligned} \kappa T_{\mu\nu} = \hat{R}^\sigma_{\mu\nu\sigma} - \frac{1}{2} G_{\mu\nu} \hat{R}^\sigma_\sigma = & \partial_\nu \tilde{\Gamma}^\sigma_{\mu\sigma} - \partial_\sigma \tilde{\Gamma}^\sigma_{\mu\nu} + \tilde{\Gamma}^\sigma_{\Sigma\nu} \tilde{\Gamma}^\Sigma_{\mu\sigma} - \tilde{\Gamma}^\sigma_{\Sigma\sigma} \tilde{\Gamma}^\Sigma_{\mu\nu} \\ & - \frac{1}{2} G_{\mu\nu} \left( G^{\sigma\text{B}} \partial_\sigma \tilde{\Gamma}^\text{T}_{\text{TB}} - G^{\sigma\text{B}} \partial_\text{T} \tilde{\Gamma}^\text{T}_{\sigma\text{B}} + G^{\sigma\text{B}} \tilde{\Gamma}^\text{T}_{\sigma\text{B}} \tilde{\Gamma}^\Sigma_{\text{TB}} - G^{\sigma\text{B}} \tilde{\Gamma}^\text{T}_{\text{T}\Sigma} \tilde{\Gamma}^\Sigma_{\sigma\text{B}} \right), \end{aligned} \quad (10.18a)$$

and for the latter:

$$\begin{aligned} \kappa T_{\mu\nu} = -\hat{R}^5_{\mu\nu 5} + \frac{1}{2} G_{\mu\nu} \hat{R}^5_5 = & -\partial_\nu \tilde{\Gamma}^5_{\mu 5} + \partial_5 \tilde{\Gamma}^5_{\mu\nu} - \tilde{\Gamma}^5_{\Sigma\nu} \tilde{\Gamma}^\Sigma_{\mu 5} + \tilde{\Gamma}^5_{\Sigma 5} \tilde{\Gamma}^\Sigma_{\mu\nu} \\ & + \frac{1}{2} G_{\mu\nu} \left( G^{5\text{B}} \partial_5 \tilde{\Gamma}^\text{T}_{\text{TB}} - G^{5\text{B}} \partial_\text{T} \tilde{\Gamma}^\text{T}_{5\text{B}} + G^{5\text{B}} \tilde{\Gamma}^\text{T}_{5\text{B}} \tilde{\Gamma}^\Sigma_{\text{TB}} - G^{5\text{B}} \tilde{\Gamma}^\text{T}_{\text{T}\Sigma} \tilde{\Gamma}^\Sigma_{5\text{B}} \right). \end{aligned} \quad (10.18b)$$

Finally, we calculate all the necessary connections as we did in section 7 (see, for example, (7.17a) for  $\tilde{\Gamma}^\mu_{\alpha 5}$ ), then look amidst  $T_{\mu\nu}$  to find field configurations which, up to multiplicative coefficients, resemble the Maxwell tensor, which is part of the Kaluza miracle, and perhaps the tensors for dust, perfect fluids, and the like. Any such tensors will then have been created 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 via (10.7), we may anticipate that when the five-

dimensional Dirac equation (5.6) is fully developed, it will become possible to discard the hand-added “wood” mass  $m$  entirely just as we discarded  $\hat{\mathcal{L}}_M$  from (10.4), and in its place substitute a mass-dimensioned scalar number that is naturally contained in the “marble” of Dirac’s equation. Using the momentum space Dirac equation (5.7), this is precisely what will be developed from (11.1) to (11.3) supra, and thereafter.

Third, the action  $\hat{S} = \int (1/2\kappa) \hat{R} \sqrt{-G} d^5x$  of (10.8), like any action, is directly used in the quantum field path integral. Using (10.8), this is:

$$Z = \int DG_{MN} \exp(i\hat{S}/\hbar) = \int DG_{MN} \exp\left((i/\hbar) \int (1/2K) \hat{R} \sqrt{-G} d^5x\right). \quad (10.19)$$

Here, the only field over which the integration needs to take place is  $G_{MN}$ , because this intrinsically contains not only the usual  $g_{\mu\nu}$ , but also the photon  $A_{\gamma\mu}$  and the scalar  $\phi$ . But aside from the direct value of (10.19) 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 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 and the fermion masses 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  $A_\mu$  as well as a spin 0 scalar  $\phi$  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 massless, luminous  $\phi_{AB}$  of (8.14) for the DKK metric tensor (3.13) and inverse (4.22). So, it may be thought – and indeed has been 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  $\phi \rightarrow 0$  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 [29] 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  $G_\mu = T^i G_\mu^i$  with internal symmetry established by the group generators  $T^i$  which have a commutation relation  $[T^i, T^j] = if^{ijk}T^k$  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  $A_{\gamma\mu} \mapsto G_{\gamma\mu} = T^i G_{\gamma\mu}^i$  everywhere this field appears (with  $\gamma$  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  $G_{\gamma\mu} = T^i G_{\gamma\mu}^i$  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 an open question. As an example, in an earlier paper [30] 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  $(\nu, (u_R, d_G, d_B), e, (d_R, u_G, u_B))$  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 [31] and [32], rooted in [33], it was shown that the nuclear binding energies of fifteen distinct nuclides, namely  $^2\text{H}$ ,  $^3\text{H}$ ,  $^3\text{He}$ ,  $^4\text{He}$ ,  $^6\text{Li}$ ,  $^7\text{Li}$ ,  $^7\text{Be}$ ,  $^8\text{Be}$ ,  $^{10}\text{B}$ ,  $^9\text{Be}$ ,  $^{10}\text{Be}$ ,  $^{11}\text{B}$ ,  $^{11}\text{C}$ ,  $^{12}\text{C}$  and  $^{14}\text{N}$ , 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 really 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 [34], 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 [30]*. 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 (7.4) and the Einstein equation (10.10) so that the motion for all interactions is strictly geodesic motion and the field dynamics and energy tensors are at bottom strictly geometrodynamical "marble" rooted fully gravitational curvature.

In 2019, the scientific community celebrates 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 forebears who worked toward a unified geometrodynamical understanding of nature, as we come upon the Kaluza-Klein centennial.

## **PART II: THE DIRAC-KALUZA-KLEIN SCALAR, THE HIGGS FIELD, AND A THEORY OF FERMION MASSES, MIXING AND WEAK BETA DECAYS WHICH RUBUSTLY FITS THE EXPERIMENTAL DATA**

### **11. Spontaneous Symmetry Breaking of the Massless Luminous Dirac-Kaluza-Klein Scalar, and Integration to Deduce its Spacetime Behavior**

As we embark upon the theory of fermion masses which will be the focal point of Part II of this paper, it is natural and indeed required that we begin with Dirac's equation  $(i\hbar c \Gamma^M \partial_M - mc^2) \Psi = 0$  from (5.6), because this is the governing equation for fermions in the "spacetime plus one" dimensionality of DKK theory. However, because our interest is in the fermion masses  $m$  and their related rest energies  $mc^2$ , it is best to work from the momentum-space Dirac equation  $(\Gamma^M cp_M - mc^2) U_0(p^\Sigma) = 0$  of (5.7). To go from (5.6) to (5.7), recall that we first

defined a five-dimensional energy-momentum vector  $cp^M = (cp^\mu \quad cp^5)$  containing the usual four-dimensional  $cp^\mu = (E \quad \mathbf{c}\mathbf{p})$  plus a brand-new energy-dimensioned  $cp^5$ , then used this to define a wavefunction  $\Psi \equiv U_0(p^\Sigma) \exp(-ip_\Sigma x^\Sigma / \hbar)$ . We then converted (5.6) from configuration space to momentum space in the usual way, merely in five dimensions, to obtain (5.7).

In the context of Dirac's equation, this newly-defined  $cp^5$  play a very similar role to that of  $\hat{\mathcal{L}}_M \equiv (1/2K) \hat{R}^5_5$  in (10.7) from which the energy tensor was defined in (10.11). Specifically, just as we discarded the "wood" of  $\hat{\mathcal{L}}_M$  entirely from the action (10.4) and then replaced it with an  $\hat{\mathcal{L}}_M \equiv (1/2K) \hat{R}^5_5$  based entirely on the "marble"  $\hat{R}^5_5$  component that is part of the Ricci scalar  $\hat{R} = \hat{R}^\Sigma_\Sigma = \hat{R}^\sigma_\sigma + \hat{R}^5_5$ , here we will discard the "wood" of the  $mc^2$  in Dirac's equation and see if it can be replaced in some fashion by this new  $cp^5$  about which we presently know little beyond the fact that it arises out of the fifth dimension just as does  $\hat{R}^5_5$ , and so has a "marble" character.

First, leaving  $p^\Sigma$  in  $U_0(p^\Sigma)$  implicitly understood, we swap upper and lower indexes in (5.7) and expand using the three-part metric tensor (3.8) as such:

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

In the above,  $mc^2$  is the rest energy of the fermion, and it is placed into the Dirac equation "by hand." It would, however, 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 it would be even more desirable if the scalar  $\phi$  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 understand the observed pattern of fermion rest masses and mixing angles.

Toward these ends, we remove the hand-added mass entirely from (11.1) by setting  $mc^2 = 0$ , just as we did with  $\hat{\mathcal{L}}_M$  from (10.4). Then, in place of this now-removed "wood" mass, we instead use the  $cp^5$  terms in (11.1), *by postulate*, to define a "marble" fermion mass via the eigenvalue relation:

$$-mc^2 U_0 \equiv \Gamma_5 cp^5 U_0 = (\gamma_5 cp^5 + \phi \gamma_0 cp^5) U_0. \quad (11.2)$$

This is exactly what we did at (10.7) for the matter Lagrangian which defines the energy tensor at (10.11), but in the context of Dirac's equation rather than Einstein's equation. By this postulate from which we shall now develop 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  $\gamma_0$  and  $\gamma_5$  4x4 Dirac

operators as always, but from the result in (8.14), so too is the luminous scalar  $\phi = \phi_{AB}$ . These Dirac indexes highlight some very important points regarding spontaneous symmetry breaking.

At present, it is not known in detail how to arrive at *fermion* rest masses through spontaneous symmetry breaking without putting in mass numbers by hand to determine coupling to the Higgs field. But we do know how to arrive at *vector boson* rest masses from independent couplings, so let's first review this process as a template for proceeding: When a scalar field (also denoted  $\phi$ , but not the same scalar as in (8.14)) is used to break the symmetry, for example, of a triplet of three weak interaction gauge fields  $W^{a\mu}$  in the *adjoint* representation of a local SU(2) Yang-Mills [35] gauge group where  $a = 1, 2, 3$  is an internal symmetry index associated with the SU(2) generators  $\tau^a$  which have a commutator relation  $[\tau^a, \tau^b] = \varepsilon^{abc} \tau^c$  (see section 14.9 of [20]), the scalars are placed into the *fundamental* representation of SU(2) whereby  $\phi^T = (\phi_1 + i\phi_2 \quad \phi_3 + i\phi_4)$  is an SU(2) *doublet* of complex scalars providing four real scalar degrees of freedom. Note that  $\tau^a = \tau^{a\dagger}$  are Hermitian, and that the Yang-Mills gauge fields  $W_\mu \equiv \tau^a W_\mu^a$ . 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 in the Lagrangian density term  $g_W^2 \phi^\dagger \tau^a W_\mu^a \tau^b W^{b\mu} \phi = g_W^2 \phi^\dagger W_\mu W^\mu \phi$ , which coupling underlies the spontaneous symmetry breaking. If we make the spinor indexes of the fundamental representation explicit with  $A, B, C$ , all ranging from 1 to 2 for SU(2), this term is really  $g_W^2 \phi_A^\dagger W_{\mu AB} W_{BC}^\mu \phi_C$ .

With this as a guide, let us restore the Fourier kernel and thus  $\Psi \equiv U_0 \exp(-ip_\Sigma x^\Sigma / \hbar)$  as specified prior to (5.7), into (11.2). We know that the Lagrangian density term for a fermion rest mass has the form  $m \bar{\Psi} \Psi$ , and we know that the wavefunctions  $\Psi = \Psi_A$  contain normally-implicit Dirac spinor indexes. So, if we explicitly show the Dirac indexes  $A, B, C$  now all ranging from 1 to 4, from (11.2) we discern that a term  $\Psi \phi \gamma_0 c p^5 \Psi = \bar{\Psi}_A \phi_{AB} \gamma_{0BC} c p^5 \Psi_C$  will be part of the Lagrangian density. Importantly, this means that  $\phi = \phi_{AB}$  in (8.14) couples perfectly to Dirac fermion wavefunctions, so symmetry can be broken and the fermions can obtain rest masses. This is just as the scalar  $\phi = \phi_A$  used to break SU(2) weak interaction symmetry couples perfectly in the term  $g_W^2 \phi_A^\dagger W_{\mu AB} W_{BC}^\mu \phi_C$  to the gauge bosons for which this scalar is intended to generate a mass.

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 (7.28) can be covariantly combined to precisely reproduce the Lorentz force motion (7.25), actually makes perfect sense in view of (11.2): Gauge bosons have a Yang-Mills internal symmetry structure against which the internal symmetries of the scalars used to spontaneously break symmetry and give mass to these gauge bosons via the Higgs mechanism must be matched, so that the scalars properly couple to the bosons. Likewise, *fermions* have a Dirac spinor structure (in addition to their Yang-Mills internal symmetry 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, so that the scalars properly couple to the fermions. So, the luminous scalar (8.14) having a Dirac structure which

couples with the Dirac structure of fermions is in precisely the same league as the scalars used to break gauge boson symmetries having internal symmetry structure which couples with 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., [36]) for fermions to be able to covariantly couple to gravitation. So, notwithstanding the “oddity” of the scalar in (8.14) picking up a 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.* Moreover, if the scalar  $\phi = \phi_{AB}$  in (8.14) did *not* have Dirac indexes, then it would *not* be possible to use this scalar to break symmetry and provide mass to the fermions. So, in sum, the sequence of requiring  $d\phi/d\tau = 0$  at (8.1) to match the Kaluza-Klein equation of motion to the Lorentz Force law which ended up causing this scalar to take on a Dirac structure, serendipitously caused  $\phi = \phi_{AB}$  to be ready-made for breaking symmetry to reveal fermion masses.

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 boson. By way of contrast, with a subscript  $H$  used to denote the standard Higgs mechanism, a scalar which we write as  $\phi_h = (\phi_{1h} + i\phi_{2h})/\sqrt{2}$  used to break a local U(1) gauge symmetry starts out with two scalar degrees of freedom provided by  $\phi_{1H}$  and  $\phi_{2H}$ , with  $\phi_h^* \phi_h = \frac{1}{2}(\phi_{1h}^2 + \phi_{2h}^2)$  defining a “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 theoretically-anticipated massive Higgs scalar [37] is perhaps one of the most significant scientific events of the past few decades.

Here (8.14) contains the same form of expression  $(\phi_1 + i\phi_2)/\sqrt{2}$  used in the Higgs mechanism. Likewise,  $\frac{1}{2}(\phi_1^2 + \phi_2^2)$  defines the circle for symmetry breaking as seen explicitly in (8.15), where a gradient  $\partial_\Sigma(\phi_1^2 + \phi_2^2)$  also appears. So, these fields  $\phi_1$  and  $\phi_2$  carry two degrees of freedom available to be “swallowed” by other particles during symmetry breaking via the Goldstone mechanism [38]. But there are two important differences from the simple  $\phi_h = (\phi_{1h} + i\phi_{2h})/\sqrt{2}$  of the standard model, which we shall now study: First, (8.14) has an additional phase angle  $\theta = \omega T$ , which can be oriented in any direction as an additional aspect of symmetry breaking and used to provide a two additional degrees of freedom which can also be swallowed by other particles, and will be swallowed by fermions. Second,  $\phi_1$  and  $\phi_2$  in (8.14) are presently-dimensionless, whereas the usual Higgs scalars,  $\phi_h = (\phi_{1h} + i\phi_{2h})/\sqrt{2}$  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. This yields a Lagrangian term  $-\lambda v^2 h^2$  which, by comparison to the expected mass term form  $-\frac{1}{2}m_h^2 c^4 h^2$ , “reveals” a Higgs rest energy

$m_\mu c^2 = \sqrt{2\lambda v^2}$ , see sections 14.6 through 14.8 in [20]. So, we will need to find a way to introduce an energy dimensionality for the fields in to (8.14), which we shall do shortly.

These two extra degree of freedom is advantageous, because with four degrees of freedom rather than two, we have the means to add two degrees of freedom to a fermion to provide it with mass in a gauge-symmetric fashion, and we still have two degrees of freedom left to provide mass to the scalar itself and so produce the massive Higgs boson of the standard model, and well as to provide mass to a gauge boson. Specifically, 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  $\nu_L = \frac{1}{2}(1 - \gamma^5)v$  would have had only two degrees of freedom, with right-chirality nonexistent. This was the basis, for example, for the Georgi-Glashow GUT model [39] which entirely discarded  $\nu_R = \frac{1}{2}(1 + \gamma^5)v$ . So, for (8.14) to generate a fermion mass for an initially-massless chiral fermion, it is necessary that *both degrees of freedom* from represented by  $\theta = \omega T$  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.

With this overview, we now proceed with some further calculations. 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:

$$0 = \Gamma_\Sigma cp^\Sigma U_0 = (\Gamma_\sigma cp^\sigma + \Gamma_5 cp^5)U_0 = (\Gamma_\sigma cp^\sigma - mc^2)U_0 = (\Gamma_\sigma cp^\sigma + \gamma_5 cp^5 + \phi\gamma_0 cp^5)U_0. \quad (11.3)$$

This is precisely analogous to how we went from (10.4) to (10.8) for the Einstein-Hilbert action. The fermion mass term is no-longer hand-added, but rather, originates in the fifth-dimensional operator  $\Gamma_5 cp^5$ . It makes its usual appearance in the form of  $0 = (\Gamma_\sigma cp^\sigma - mc^2)U_0$  when the fifth-dimensional  $\Gamma_5 cp^5$  is replaced by  $-mc^2$  via (11.2). So, the momentum space Dirac equation (5.7) becomes  $\Gamma_\Sigma cp^\Sigma U_0 = 0$  and the configuration space equation (5.6) becomes  $i\hbar c \Gamma^M \partial_M \Psi = 0$ , now, importantly, 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 the anticommutator, strip off the operand, and apply (3.1) to obtain, respectively:

$$\Gamma_M cp^M \Gamma_N cp^N = \frac{1}{2} \{ \Gamma_M, \Gamma_N \} cp^M cp^N = G_{MN} cp^M cp^N = 0, \quad (11.4a)$$

$$\Gamma_\mu cp^\mu \Gamma_\nu cp^\nu = \frac{1}{2} \{ \Gamma_\mu, \Gamma_\nu \} cp^\mu cp^\nu = G_{\mu\nu} cp^\mu cp^\nu = m^2 c^4, \quad (11.4b)$$

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

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:

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

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

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

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), which is the anticommutator “square” of  $\Gamma_\Sigma cp^\Sigma U_0 = 0$  with a geometric “marble” fermion mass, leads to a very interesting and important consequence for the five-dimensional metric line element  $dS = cdT$  defined by:

$$c^2 dT^2 \equiv G_{MN} dx^M dx^N. \quad (11.7)$$

Specifically, if we further define the five-momentum in terms of mass and motion in the usual way by  $cp^M \equiv mc^2 dx^M / cd\tau$  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:

$$m^2c^4 \frac{dT^2}{d\tau^2} = G_{MN} mc^2 \frac{dx^M}{cd\tau} mc^2 \frac{dx^N}{cd\tau} = G_{MN} cp^M cp^N. \quad (11.8)$$

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:

$$\boxed{dS = cdT = 0}. \quad (11.9)$$

This is a very important and useful result, and it is one of the direct consequences of the postulated eigenvalue relation (11.2) to define the fermion rest mass out of the “marble” of geometry rather than hand-added “wood.” This means that with “marble” fermion masses, the five-dimensional metric line element (11.7) is null  $G_{MN} dx^M dx^N = 0$  and the five-dimensional relativistic energy momentum relation  $G_{MN} cp^M cp^N = 0$  is akin to that of a zero-mass body. This may be thought of as a form of five-dimensional masslessness and luminosity, which in four spacetime dimensions nevertheless reveals massive, subluminal material bodies.

Our first use of this result, will be to break the symmetry of the sine and cosine terms in (8.14) for the Kaluza-Klein scalar  $\phi = \phi_{AB}$ . In this regard, what we learn from (11.9) is that any *finite* five-dimensional proper metric interval  $S = cT = \int dS = \int cdT = S_0 = cT_0$  obtained from

integrating (11.9), whether of length or time dimensionality, must be equal to zero plus a constant of integration  $S_0 = cT_0$ . And this in turn means that  $T$  in the  $\omega T$  terms of (8.14) is zero up to a constant of integration, and specifically, that  $T = 0 + T_0$ . So, we now wish to use this finding to break symmetry in the most advantageous way possible.

Toward this end, starting with  $\cos(\omega T)$  in (8.14), let us break the symmetry in the plane of the angle  $\theta = \omega T$  by imposing the symmetry-breaking constraint  $\cos(\omega T) \equiv 1$ . This of course means that  $\omega T = \omega T_0 = 2\pi n$  is quantized, with  $n = 0, \pm(1, 2, 3, \dots)$  being any integer. Using  $c = \omega \lambda$  which we can do because  $\phi$  in (8.14) is massless and luminous, and also  $\lambda = 2\pi \tilde{\lambda}$ , this constraint  $\omega T = 2\pi n$  is alternatively formulated in terms of the five-dimensional space-dimensioned finite proper length element  $S = cT = n\lambda$  which is essentially quantized units of five-dimensional length set by  $\lambda$ . As well, this means that  $\sin(\omega T) = 0$ , but we need to be careful because there is also a  $cT$  in the denominator of the  $\sin(\omega T)$  term in (8.14).

So, for  $\sin(\omega T)$ , we insert the same  $\omega T = \omega T_0 = 2\pi n$ , then use  $c = \omega \lambda$  and  $\lambda = 2\pi \tilde{\lambda}$ , thus:

$$-i \frac{\Gamma_{\Sigma} x^{\Sigma}}{cT} \sin(\omega T) = -i \frac{\Gamma_{\Sigma} x^{\Sigma} \omega}{2\pi n c} \sin(2\pi n) = -i \frac{\Gamma_{\Sigma} x^{\Sigma}}{n\lambda} \sin(2\pi n). \quad (11.10)$$

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  $n = \pm(1, 2, 3, \dots)$ . With this final constraint (11.10) does become zero and (8.14) reduces to:

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

Having used  $S = cT = n\lambda$  to break symmetry with  $n = \pm(1, 2, 3, \dots)$  being a positive or negative non-zero integer, we see that *finite* five-dimensional proper lengths are quantized integer multiples of the wavelength  $\lambda$  first specified in (8.5) for the luminous Kaluza-Klein scalar field  $\phi$ . This follows a long tradition of quantization based on wavelength fitting which started with Bohr [40] and culminated with DeBroglie [41].

Importantly, with (11.11) we need no longer be concerned with the Dirac operator matrix  $\Gamma_{\Sigma}$  in  $\phi$ , because we have broken symmetry so as to effectively diagonalize the operation of this operator and remove it from explicitly appearing in (11.11). 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  $\lambda = 2\pi \tilde{\lambda}$  or energy  $hf = \hbar\omega$  of the scalar  $\phi$ , which were overt in (8.14). Particularly, we have hidden the dimensionless ratio  $x^{\Sigma} / \lambda$  in (11.10). This does not mean that the scalar no longer has a frequency or wavelength or energy or even an internal Dirac

structure. Rather, it means that the symmetry breaking has *hidden* these attributes. Also, importantly, the spontaneous non-appearance of this operator and its two degrees of freedom in (11.11), is the mechanism by which these degrees of freedom have now been swallowed by a fermion to provide it with a mass.

We complete the symmetry breaking in the usual way by again noting that  $\phi_1^2 + \phi_2^2$  seen in (8.15) defines a symmetry breaking circle, and by orienting the scalar in this circle by setting  $\phi_2 = 0$  in the customary manner. This further reduces (8.14) to its final symmetry-broken form:

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

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:

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

Of special interest in (11.13), is that whereas the Kaluza-Klein scalar  $\phi$ , thus  $\phi_1$ , has all along been physically dimensionless, in (11.13) this has now become multiplied by  $cp^5$  which has dimensions of energy. This means that  $cp^5 \phi_1$  now has precisely the same characteristics as  $\phi_h$  in the scalar field  $\phi_h = (\phi_{1h} + i\phi_{2h})/\sqrt{2} \mapsto \phi_{1h}/\sqrt{2} = (v + h(x^\mu))/\sqrt{2}$  employed in standard model Higgs field symmetry breaking. Specifically, being an energy-dimensioned scalar field,  $cp^5 \phi_1$  now presents the opportunity for a connection with the standard model Higgs field  $h$ .

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

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

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

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

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

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

The above now presents a differential equation for  $\phi_1$  as a function of the five-dimensional  $x^\Sigma$ .

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

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

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

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

Then we convert  $\partial \rightarrow d$  and finally restructure this into:

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

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

$$\begin{aligned}
 \int \frac{1}{x^\Sigma} dx^\Sigma &= \int \left( \frac{dx^0}{x^0} + \frac{dx^1}{x^1} + \frac{dx^2}{x^2} + \frac{dx^3}{x^3} + \frac{dx^5}{x^5} \right) = \ln x^0 + \ln x^1 + \ln x^2 + \ln x^3 + \ln x^5 + \ln \frac{1}{L^5} \\
 &= \ln \left( \frac{x^0 x^1 x^2 x^3 x^5}{L^5} \right) = \int -\frac{1}{A - \ln(By)} \frac{1}{y} dy = \ln(A - \ln(By)) \\
 &= \int -\frac{1}{\ln \left( \pm \sqrt{2} \sqrt{(mc^2)^2 - (cp^5)^2} \right) - \ln(cp^5 \phi_1)} \frac{1}{\phi_1} d\phi_1 \\
 &= \ln \left( \ln \left( \pm \sqrt{2} \sqrt{(mc^2)^2 - (cp^5)^2} \right) - \ln(cp^5 \phi_1) \right)
 \end{aligned} \tag{11.20}$$

The middle line includes using the generalized  $\int -\left(1/(A - \ln(By))\right) y dy = \ln(A - \ln(By)) + C'$ . But this integration constant  $C'$  is not needed because we have separately used  $C \equiv \ln(1/L^5)$  to integrate the left side of (11.19).

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

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

Exponentiating one final time, then isolating the energy-dimensioned field  $cp^5 \phi_1$  which we seek to connect to the standard model Higgs field as discussed following (11.13), the final result is:

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

The numerator inside the exponent, which we denote in consolidated form as  $V_{(5)} \equiv x^0 x^1 x^2 x^3 x^5$ , 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 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. The above (11.22) directly tells us how the Dirac-Kaluza-Klein scalar, now broken to  $\phi_1(x^M)$ , behaves as a function of the spacetime-plus-one coordinates. As we shall see in section 13, this result will play a central role in describing how fermions extract energy from the Fermi vacuum via the Higgs mechanism to obtain their rest masses.

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

As discussed following (11.13), the fact that the dimensionless  $\phi_1$  is now multiplied by the energy-dimensioned  $cp^5 = mc^2(dx^5/cd\tau)$  in (11.22) presents an opportunity to connect  $cp^5\phi_1$  with the standard model Higgs field  $h$ . So, the next step is to obtain a direct expression for  $cp^5$  which, of course, is the fifth component of the energy momentum  $cp^M = (cp^\mu \quad cp^5)$  first defined following (5.6). To directly study  $cp^5$ , recall that (7.27) connects electric charge to motion in the Kaluza-Klein fifth dimension. So, using  $dx^5/cd\tau = q/\phi^2 k mc^2$  from (7.27), and also “borrowing”  $k$  based on the standard Kaluza-Klein theory from (1.2), we obtain:

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

Formally speaking, we have not yet proved that (1.2) is the correct value of  $k$  for the DKK metric tensor (3.13). 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  $g_{MN} = \eta_{MN}$  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 the standard model photon. In particular, it adds a 1 to  $\phi^2$  in  $G_{55}$ , and it adds a  $\phi$  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 (7.11), and at (7.27) 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 to obtain the energy tensor via (10.18), and that  $k$  will likewise turn out to be exactly the same as 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, that we “borrow”  $k$  from (1.2). But we shall also continue to show  $k$  in our calculations, in order to also obtain results that apply without this borrowing.

Proceeding on this basis, we combine (12.1) with  $G_{55} cp^5 cp^5 = m^2 c^4$  from (11.4c) and  $G_{55} = 1 + \phi^2$  from (3.11) for  $g_{MN} = \eta_{MN}$ , and also borrow (1.2), to obtain:

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

This easily restructures into a quadratic for  $\phi^2$ , which we write as:

$$0 = \frac{m^2c^4k^2}{q^2}\phi^4 - \phi^2 - 1 = 4\frac{Gm^2}{k_eq^2}\phi^4 - \phi^2 - 1. \quad (12.3)$$

We see that this 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 momentum space 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  $\alpha = k_e e^2 / \hbar c$ , 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  $e, \mu, \tau$  leptons,  $Q = +2/3$  for the  $u, c, t$  quarks and  $Q = -1/3$  for the  $d, s, b$  quarks, and has reversed signs for the antiparticles. Also, we note that the Planck mass  $M_p = 1.220910 \times 10^{19} \text{ GeV} / c^2$  using the value reported in [42], 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:

$$\frac{Gm^2}{k_eq^2} = \frac{Gm^2}{Q^2k_e e^2} = \frac{Gm^2}{Q^2\alpha\hbar c} = \frac{Gm^2}{Q^2\alpha GM_p^2} = \frac{m^2}{Q^2\alpha M_p^2}. \quad (12.4a)$$

The square root of this will also be of interest. We write this as:

$$\sqrt{\frac{Gm^2}{k_eq^2}} = \sqrt{\frac{G}{k_e} \frac{m}{q}} = \sqrt{\frac{G}{k_e} \frac{m}{Qe}} = \frac{\sqrt{G}m}{Q\sqrt{\alpha}\sqrt{\hbar c}} = \frac{m}{Q\sqrt{\alpha}M_p} \quad (12.4b)$$

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 in several modes of expression are found to be at:

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

Placing these two root solutions into  $\phi^2$  in (12.1) we obtain two corresponding solutions for  $cp^5$ , also in several modes of expression:

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

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

$$\frac{dx_{\pm}^5}{cd\tau} = \frac{2mc^2k}{q} \frac{1}{1 \pm \sqrt{1 + 4 \frac{m^2c^4k^2}{q^2}}} = 4 \sqrt{\frac{Gm^2}{k_e q^2}} \frac{1}{1 \pm \sqrt{1 + 16 \frac{Gm^2}{k_e q^2}}} = 4 \frac{m}{Q\sqrt{\alpha M_p}} \frac{1}{1 \pm \sqrt{1 + 16 \frac{m^2}{Q^2 \alpha M_p^2}}} \quad (12.7)$$

for the “motion”  $dx_{\pm}^5 / cd\tau = dt_{\pm}^5 / d\tau$ , 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  $g_{MN} = \eta_{MN}$ , see (3.11), which we therefore now write as  $G_{55\pm} = 1 + \phi_{\pm}^2$ , it is also useful to employ (12.5) to write these two solutions as:

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

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, we can use the first three terms of the series expansion  $\sqrt{1+x^2} = 1 + \frac{1}{2}x^2 - \frac{1}{8}x^4 + \dots$  in each of (12.5) through (12.7) to a very close approximation, to obtain:

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

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

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

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:

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

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

$$\begin{aligned}cp_+^5 &\cong \frac{m^2c^4k}{q} = 2 \sqrt{\frac{Gm^2}{k_e q^2}} mc^2 = 2 \frac{m}{Q \sqrt{\alpha M_p}} mc^2 = 2 \frac{1}{Q \sqrt{\alpha M_p}} m^2 c^2 \ll mc^2 \\ cp_-^5 &\cong -\frac{q}{mc^2k} = -\frac{1}{2} \sqrt{\frac{k_e q^2}{Gm^2}} mc^2 = -\frac{1}{2} \frac{Q \sqrt{\alpha M_p}}{m} mc^2 = -\frac{1}{2} Q \sqrt{\alpha M_p} c^2 \gg mc^2\end{aligned}\quad (12.13)$$

And for (12.11) we similarly obtain:

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

Also, because the DKK metric tensor component  $G_{55} = 1 + \phi^2$  for  $g_{MN} = \eta_{MN}$ , 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:

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

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], in which the ‘‘movement’’ through the now-timelike fifth dimension is very rapid. Reciprocally,  $G_{55-}$  is very small, on the order of  $m^2 / M_p^2 \cong 10^{-40}$ , yet still retains a timelike rather than a spacelike signature. Therefore,  $dx_-^5 / cd\tau = dt_-^5 / d\tau$  is a very rapid rate of fifth dimensional time flow, and not a space velocity on the order of  $10^{20}c$ . Taken together, the upper result  $G_{55+} \cong 10^{40} \gg 1$  can be thought of as embodying in the DKK metric tensor, how in the Fermi vacuum the electromagnetic interaction is inordinately-stronger than the gravitational interaction, while the later result  $0 < G_{55-} \cong 10^{-40} \ll 1$  embodies how the Fermi vacuum is inordinately-less-energetic than the Planck vacuum.

Finally, we insert the next-to-last expression from the two quadratic solutions (12.13) into (11.22) which was the final result in the last section, and likewise split this into:

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

Recall again that these two solutions for  $cp_+^5\phi_1$  and  $cp_-^5\phi_1$  contain the two root solutions obtained starting at (12.5), to the quadratic (12.3). Above, we also split the integration constant into  $L_{\pm}$ , allowing for this to also be different for each of the two solutions. Note that the upper (+) solution above is independent of the fermion charge generator  $Q$  and  $\sqrt{\alpha}$ , while the lower (–) solution is not. However, the lower solution is independent of the fermion rest mass while the upper is not. As we shall see in the next section, the former solution applies in the *Fermi vacuum* with an energy

vev  $v = 246.2196508 \pm 0.0000633$  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  $M_p c^2 = 1.220910 \times 10^{19}$  GeV [42] is established from the Newton coupling via  $GM_p^2 \equiv \hbar c$ . Now, we have all that is needed to connect all of this to the standard model Higgs field.

### 13. Connection between the Dirac-Kaluza-Klein Scalar and the Higgs Field, and the Extraction of Energy from the Higgs Field by the Top Quark

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-dimensional scalar fields, as is the Higgs field  $h(\mathbf{X})$ . Inside the exponential  $V_{(5)} = x^0 x^1 x^2 x^3 x^5$  is a five-dimensional volume element. As to the ratio  $V_{(5)} / L^5 = x^0 x^1 x^2 x^3 x^5 / L_{\pm}^5$ , recall that  $L_{\pm}^5$  arose at (11.20) via the constant of integration  $C \equiv \ln(1/L^5)$ . So, like any integration constant, it must be determined by suitable “initial conditions.” Consequently, the length-to-the-fifth power dimensionality of  $L^5$  was required as an “initial condition” to provide a proper dimensionless ratio inside the exponential in view of  $V_{(5)}$  also having a fifth-order length dimension.

Another “initial condition” we now impose on this integration constant is that the overall ratio  $V_{(5)} / L^5$  in (12.16) must be invariant under rotations and boosts, which are linear, not general coordinate transformations. This is simply a requirement for symmetry under six of the ten parameters of the Poincare group. The reason for this is to ensure that  $\exp(-x^0 x^1 x^2 x^3 x^5 / L_{\pm}^5)$  and therefore the  $cp_{\pm}^5 \phi_1$  in (12.16) do not change simply because we rotate or boost the observational coordinate system. This in turn means that when the coordinate system is changed,  $L^5$  must rotate and Lorentz transform in exactly the same manner as  $V_{(5)}$ . To enforce this symmetry, we keep in mind that this exponential  $\exp(-x^0 x^1 x^2 x^3 x^5 / L_{\pm}^5)$  emerged from (11.12) following integration, and originated in the Fourier kernel  $\exp(-is_{\Sigma} x^{\Sigma} / \hbar)$  in (8.7). And, we note that  $s_{\Sigma} / \hbar$  inside the original kernel can be used to define a wavevector  $k_{\Sigma} \equiv s_{\Sigma} / \hbar$  with dimensions 1/length, in which event the kernel becomes  $\exp(-ik_{\Sigma} x^{\Sigma}) = \cos k_{\Sigma} x^{\Sigma} - i \sin k_{\Sigma} x^{\Sigma}$ . Rotation and boost symmetries are then naturally achieved by the construction of  $k_{\Sigma} x^{\Sigma}$ , because while  $x^M$  is not a vector under general coordinate transformations, it is a vector under the *linear* transformations of rotations and boosts, so that  $k_{\Sigma} x^{\Sigma}$  is likewise a scalar under the same linear transformations. If we also add a *phase* symmetry for these kernel waves – effectively translational invariance – we complete the Poincare group. Moreover, in the five DKK dimensions, we add one *time rotation* between  $t^0$  and  $t^5$ , three more Lorentz boosts between each of the three space coordinates and  $t^5$ , and one more translation through  $t^5$ , so that the Poincare group now has fifteen (15) parameters.

With all this in mind, we expect that to maintain Poincare symmetry in (12.16) each of the five  $L$  in  $L^5$  will be in the nature of one component of a five-dimensional wavevector  $k^M$  with 1/length dimensionality. Then,  $V_{(5)} / L^5$  will take the form of a *product of the five terms in  $k_\Sigma x^\Sigma$* , that is,  $V_{(5)} / L^5 = \Pi_{\Sigma=0,1,2,3,5} (k_\Sigma x^\Sigma) = k_0 k_1 k_2 k_3 k_5 x^0 x^1 x^2 x^3 x^5$ . This product of course, is merely what emerged at (11.20) during integration from what was originally the  $\exp(-ik_\Sigma x^\Sigma)$  kernel. So,  $V_{(5)} / L^5 = k_0 k_1 k_2 k_3 k_5 x^0 x^1 x^2 x^3 x^5 \rightarrow V'_{(5)} / L'^5 = k'_0 k'_1 k'_2 k'_3 k'_5 x'^0 x'^1 x'^2 x'^3 x'^5$  becomes the required symmetry under coordinate rotations and boosts. Again, all we have done here is simply impose “initial conditions” on the constant of integration in (12.16) to require a) proper physical dimensionality and b) rotational and Lorentz symmetry.

Now, for a wavevector  $k^\sigma = (\omega / c \quad \mathbf{k}) = (\omega / c \quad 1 / \tilde{\lambda})$  in four dimensions, the magnitudes of the frequency  $\omega$  and the wavelength components  $\lambda = 2\pi\tilde{\lambda}$  for a particle of mass  $m$  and reduced Compton wavelength  $\tilde{\lambda}_c = \lambda_c / 2\pi = \hbar / mc = h / 2\pi mc$  are determined by the Lorentz scalar  $g_{\mu\nu} k^\mu k^\nu = m^2 c^2 / \hbar = 1 / \tilde{\lambda}_c^2$ . With this in mind, the question arises as to the energy / length scale of the five-dimensional  $k^M$ , which has now inherited the role of the integration constant  $L^5$  which first entered at (11.20). For this, we return to the three relations (11.4), namely  $G_{MN} c p^M c p^N = 0$ ,  $G_{\mu\nu} c p^\mu c p^\nu = m^2 c^4$  and  $G_{55} c p^5 c p^5 = m^2 c^4$ , as well as  $G_{\mu 5} c p^\mu c p^5 = -m^2 c^4$  from (11.6). If we next define a wavevector  $k^\Sigma \equiv p^\Sigma / \hbar$  following the usual pattern generalized to five dimensions, then the foregoing relations produce the following five-dimensional DKK wavevector relations:

$$G_{MN} k^M k^N = 0; \quad G_{\mu\nu} k^\mu k^\nu = 1 / \tilde{\lambda}_c^2; \quad G_{55} k^5 k^5 = 1 / \tilde{\lambda}_c^2; \quad G_{\mu 5} k^\mu k^5 = -1 / \tilde{\lambda}_c^2. \quad (13.1)$$

Still, all we are doing with (13.1) is establishing the “initial conditions” on the constant of integration now represented in  $k^M$ , by pinning the 1/length magnitude of the wavevector components to the inverse of the Compton wavelength of the scalar bosons which are the field quanta of the scalar field  $c p_\pm^5 \phi_1$ . Although  $G_{MN} k^M k^N = 0$  is null in five dimensions, in the four-dimensions of spacetime, we recover the usual  $G_{\mu\nu} k^\mu k^\nu = 1 / \tilde{\lambda}_c^2$ . Given the two solutions leading to two  $L_\pm^5$  in (12.16), this means that we also need to denote two wavevectors  $k_\pm^M$  in correspondence with these two solutions.

Now we proceed to the question of what to use for the Compton wavelength  $\tilde{\lambda}_c = \hbar / mc$  and associated mass of the scalar bosons. For this, we recall that when we broke symmetry at (11.11) the energy / frequency  $E = hf = \hbar\omega$  and wavelength  $\lambda = 2\pi\tilde{\lambda}$  of the Kaluza-Klein scalar became hidden, along with this scalar’s Dirac operator properties which became diagonalized. But as discussed following (11.11), this does not mean that the scalar no longer has energy or a wavelength; these are just hidden at the moment. Moreover, although this scalar started as massless and luminous, the symmetry breaking at (11.11) released two degree of freedom that now enables the fermions to acquire mass. And at this point, the scalar fields are the two  $c p_\pm^5 \phi_1$  in (12.16) which have indeed had their symmetry broken. So, once we finish connecting these scalar

solutions  $cp_{\pm}^5\phi_1$  to the Higgs field (and actually, to two Higgs fields, one for the Fermi vacuum and one for the Planck vacuum), this will mean that  $\tilde{\lambda}_c$  and  $m$  in  $\tilde{\lambda}_c = \hbar / mc$  in  $G_{\mu\nu}k^{\mu}k^{\nu} = 1/\tilde{\lambda}_c^2$  will become the Compton wavelengths of the Higgs boson(s). This is the final “initial condition” imposed on the constant of integration, and it removes the physical energies of the scalar from hiding and makes them visible once again, in the  $k^{\mu}$  which have inherited the integration constant. Now, let’s proceed to finally make this connection to the Higgs field and boson.

Recall that in the standard model, we expect a fermion ( $f$ ) rest energy  $m_f c^2$  to be related to the Fermi vev  $v \cong 246.2196508 \text{ GeV}$  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 the actual physical lengths in  $L_{\pm}^5$  thus  $k_{\pm}^M$ , and irrespective of what  $cp_{\pm}^5\phi_1(\mathbf{X})$  physically represent, for a coordinate assignment  $x^M = \mathbf{0} = (0,0,0,0,0)$  to an origin, thus  $V_{(5)} = x^0 x^1 x^2 x^3 x^5 = 0$ , the term  $\exp(-V_{(5)} / L^5) = 1$ . Thus, with  $m = m_f$ , (12.16) reduce to:

$$\begin{aligned} cp_+^5\phi_1(\mathbf{0}) &= \pm\sqrt{2}m_f c^2 = \pm G_f v \\ cp_-^5\phi_1(\mathbf{0}) &= \pm\frac{1}{\sqrt{2}}iM_p c^2 Q\sqrt{\alpha} \end{aligned} \quad (13.2)$$

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

Conversely, again irrespective of  $L_{\pm}^5$  and  $k_{\pm}^M$ , for a coordinate assignment with  $V_{(5)} / L_{\pm}^5 \gg 1$ , the exponential  $\exp(-V_{(5)} / L_{\pm}^5) \rightarrow 0$  will approach zero, and (12.16) will reduce to:

$$\begin{aligned} cp_+^5\phi_1(V_{(5)} / L_+^5 \gg 1) &= 0 \\ cp_-^5\phi_1(V_{(5)} / L_-^5 \gg 1) &= 0 \end{aligned} \quad (13.3)$$

So, these energy-dimensioned fields  $cp_{\pm}^5\phi_1(x^M) / \sqrt{2}$  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 we finally turn to the standard model Higgs field.

The Higgs field which we represent in the five DKK dimensions by writing  $h(x^M)$ , is a scalar field with dimensions of energy. 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 any more than a statistical way. 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 these fermions will find “nests” at energetically-minimized locations in the Higgs field where the fermions are most likely to situate.

Now, for a U(1) gauge group, the standard model starts with a scalar field which we denote by  $\phi_h$  to distinguish from the Kaluza-Klein scalar  $\phi$ , and is given the following assignments at the various steps of symmetry breaking:

$$\phi_h(x^\mu) = \frac{1}{\sqrt{2}}(\phi_{1h}(x^\mu) + i\phi_{2h}(x^\mu)) \mapsto \frac{1}{\sqrt{2}}\phi_{1h}(x^\mu) = \frac{1}{\sqrt{2}}(v + h(x^\mu)). \quad (13.4)$$

That is, we first assign  $\phi_h = \frac{1}{\sqrt{2}}(\phi_{1h} + i\phi_{2h})$ . 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 Lagrangian potential  $V = \mu^2 \phi_h^* \phi_h + \lambda(\phi_h^* \phi_h)^2$  for the complex scalar field, we find that this potential has minima at  $\phi_{1h} = \pm\sqrt{-\mu^2/\lambda} \equiv \pm v$ . Finally, we perturbatively expand around the minimum at  $\phi_{1h} = v$  using the Higgs field  $h(x^\mu)$  which represents quantum fluctuations about the minima. Note that the potential  $V$  has physical dimensions of *energy to the fourth power*, because this is part of a Lagrangian density  $\mathcal{L} = T - V$  with  $T = \hbar^2 c^2 (D_\mu \phi_h)^* (D^\mu \phi_h)$  and  $D_\mu = \partial_\mu + iqA_\mu / \hbar c$ , and with  $V = \mu^2 \phi_h^* \phi_h + \lambda(\phi_h^* \phi_h)^2$  as above. For non-abelian Yang-Mills gauge theory [29], the kinetic portion of the Lagrangian density becomes  $T = \hbar^2 c^2 (D_\mu \phi_h)^\dagger (D^\mu \phi_h)$ . This is all nicely reviewed in sections 14.6 through 14.9 of [20].

Most importantly for the present discussion, because  $\phi_{1h} = \pm v$  are the minima of the potential  $V$  and because  $v = 246.2196508 \text{ GeV}$  is a constant energy, the expectation value  $\langle \phi_{1h} \rangle = \frac{1}{\sqrt{2}}(\pm v + \langle h \rangle) = \pm \frac{1}{\sqrt{2}} v$ , 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 vev 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 the  $V_{(5)}/L_\pm^5 = 0$  origin the exponential  $\exp(-V_{(5)}/L_\pm^5) = 1 > 0$  is above zero. Further, where  $V_{(5)}/L_\pm^5 \gg 1$  the exponential  $\exp(-V_{(5)}/L^5) \rightarrow 0$  drops to zero. So, if we want the origin at  $x^M = \mathbf{0}$  be the most energetically-favorable locale for a fermion to “nest” at, we must choose the  $-$  signs from the  $\pm$  in (12.16) for both solutions. Then, with this choice of sign, we assign  $cp_+^5 \phi_1(x^M) \equiv h(x^M)$  and  $cp_-^5 \phi_1(x^M) \equiv \frac{i}{2} Q \sqrt{\alpha} H(x^M)$ , with  $h$  and  $H$  representing Higgs fields associated with each respective solution. The former,  $\boxed{cp_+^5 \phi_1(x^M) \equiv h(x^M)}$ , represents the direct connection we have

been seeking since section 11 between that DKK scalar and the standard model Higgs field. So, showing  $m_f c^2 = G_f v / \sqrt{2}$ , and again with  $m = m_f$ , (12.16) now becomes:

$$\boxed{\begin{aligned} h(x^M) &\equiv cp_+^5 \phi_1 = -\sqrt{2} m_f c^2 \exp\left(-\frac{x^0 x^1 x^2 x^3 x^5}{L_+^5}\right) = -\sqrt{2} m_f c^2 \exp\left(-\frac{V_{(5)}}{L_+^5}\right) = -G_f v \exp\left(-\frac{V_{(5)}}{L_+^5}\right) \\ H(x^M) &\equiv -2icp_-^5 \phi_1 / Q\sqrt{\alpha} = -\sqrt{2} M_p c^2 \exp\left(-\frac{x^0 x^1 x^2 x^3 x^5}{L_-^5}\right) = -\sqrt{2} M_p c^2 \exp\left(-\frac{V_{(5)}}{L_-^5}\right) \end{aligned}} \quad (13.5)$$

Note from earlier in this section that  $V_{(5)} / L_{\pm}^5 = k_{\pm 0} k_{\pm 1} k_{\pm 2} k_{\pm 3} k_{\pm 5} x^0 x^1 x^2 x^3 x^5$ , which maintains Poincare symmetry. But in general, to keep things simple, we will continue using  $V_{(5)} / L_{\pm}^5$  unless we specifically need to use the  $k_{\pm}^M$  in a given circumstance. The lower assignment in (13.5) includes  $Q\sqrt{\alpha}$  to make the background field  $h_-(x^M)$  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  $1/2$  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. The capitalization of the latter Higgs field is to indicate the inordinately-higher energies rooted in the appearance of  $M_p c^2$  versus  $m_f c^2$  in the bottom versus the top relation (13.5).

We note in passing that rather than use the factor of  $i$  in the lower Planck-scale assignment, it may be possible given  $\phi_h = \frac{1}{\sqrt{2}}(\phi_{1h} + i\phi_{2h})$  to instead assign  $cp_-^5 i\phi_2(x^M) \equiv \frac{i}{2} Q\sqrt{\alpha} H(x^M)$ . In this case we borrow the  $i$  from  $i\phi_{2h}$  and thus have the Planck vacuum defined orthogonally to the Fermi vacuum in the complex Euler plane of  $\phi_{1h} + i\phi_{2h}$ . With this, after reaching (11.11) we break symmetry twice – once for the Planck vacuum by setting  $\phi_1 = 0$  and once for the Fermion vacuum by setting  $\phi_2 = 0$ . Either way, the net result is (13.5). Now let's examine the evidence in favor of the assignments (13.5).

With these assignments, and having chosen the  $-$  sign from the  $\pm$  in (12.16), if we denote  $x \equiv V_{(5)} / L_{\pm}^5 \geq 0$  to make (13.5) maximally-transparent, we see that both Higgs fields vary in accordance with  $h \propto -e^{-x}$ ,  $H \propto -e^{-x}$ . And, we learn explicitly from (13.5) that  $h(\mathbf{0}) = -\sqrt{2} m_f c^2$  and  $H(\mathbf{0}) = -\sqrt{2} M_p c^2$  at the origin, while  $h(x^M) = 0$  and  $H(x^M) = 0$  far from the origin. With the Higgs fields having energy minima at the origin, the origin provides the most energetically-favorable locale at which the fermions will “nest.” Again, this is why we chose the  $-$  sign. Given this, we next proceed albeit in five dimensions to make the standard model assignment  $\phi_h(x^M) = \frac{1}{\sqrt{2}}(v + h(x^M))$  from (13.4). We also assign  $\phi_H(x^M) = \frac{1}{\sqrt{2}}(v_p + H(x^M))$  for the Planck-scale Higgs field, where  $v_p \equiv M_p c^2$  is the Planck energy about which  $H(x^M)$  represents Planck-scale perturbations. Consequently, from (13.5) we obtain:

$$\phi_h(x^M) = \frac{1}{\sqrt{2}} \phi_{1h} = \frac{1}{\sqrt{2}} (v + h(x^M)) = \frac{1}{\sqrt{2}} v - m_f c^2 \exp\left(-\frac{V_{(5)}}{L_+^5}\right) = \frac{1}{\sqrt{2}} v - \frac{1}{\sqrt{2}} G_f v \exp\left(-\frac{V_{(5)}}{L_+^5}\right). \quad (13.6)$$

$$\phi_H(x^M) = \frac{1}{\sqrt{2}} (v_p + H(x^M)) = \frac{1}{\sqrt{2}} v_p - M_p c^2 \exp\left(-\frac{V_{(5)}}{L_-^5}\right)$$

Further, noting that  $h(x^M)/\sqrt{2}$  and  $H(x^M)/\sqrt{2}$  appear in the above cut by a  $\sqrt{2}$  factor, let's now take the integral of the Higgs field energies as they appear in (13.5), integrated from the origin at  $x^M = \mathbf{0}$  thus  $V_{(5)} = x^0 x^1 x^2 x^3 x^5 = 0$  where a fermion nests, out to infinity over the entire five-dimensional volume of the DKK spacetime-plus-one. This yields:

$$\boxed{\begin{aligned} \frac{1}{\sqrt{2}} \frac{1}{L_+^5} \int_0^\infty h(x^M) dV_{(5)} &= -\frac{1}{L_+^5} m_f c^2 \int_0^\infty \exp\left(-\frac{V_{(5)}}{L_+^5}\right) dV_{(5)} = m_f c^2 \exp\left(-\frac{V_{(5)}}{L_+^5}\right) \Big|_0^\infty = -m_f c^2 \\ \frac{1}{\sqrt{2}} \frac{1}{L_-^5} \int_0^\infty H(x^M) dV_{(5)} &= -\frac{1}{L_-^5} M_p c^2 \int_0^\infty \exp\left(-\frac{V_{(5)}}{L_-^5}\right) dV_{(5)} = M_p c^2 \exp\left(-\frac{V_{(5)}}{L_-^5}\right) \Big|_0^\infty = -M_p c^2 \end{aligned}}. \quad (13.7)$$

Here, in a very important result, we see directly how fermions draw energy out of the vacuum using the Higgs field to acquire their rest energies, consistently with energy conservation. Let us focus specifically on  $\phi_h(x^M)$  and  $h(x^M)$  in (13.6) and (13.7), which has the vev  $v$  of the Fermi vacuum  $v = 246.2196508$  GeV (not showing experimental errors) to see how this works:

At the  $V_{(5)} = x^0 x^1 x^2 x^3 x^5 = 0$  origin where the fermion is most-likely to nest because the energy of the Higgs field is at its lowest, the upper (13.6) reduces to  $\phi_h(\mathbf{0}) = \frac{1}{\sqrt{2}} v - m_f c^2$ . Far from the origin we have  $\phi_h(V_{(5)}/L_+^5 \gg 1) = \frac{1}{\sqrt{2}} v$ . So, where the fermion is nested there is a depression in the  $\phi_h(x^M)$  field which has dropped below  $\frac{1}{\sqrt{2}} v$  by an energy precisely equivalent to the fermion mass  $m_f c^2$ . Far from the origin  $\phi_h(x^M) = \frac{1}{\sqrt{2}} v$ , and there is no energy drop. Thus, the fermion has clearly extracted energy from the Fermi vacuum, and most of that energy is extracted close to the fermion. But what we learn from (13.7) is that the *total energy* extracted from the vacuum, integrated from the locale at the origin where the fermion is nested to the infinite reaches of the five-dimensional DKK volume element  $V_{(5)} = x^0 x^1 x^2 x^3 x^5$ , is precisely equal to the rest energy of the fermion! From throughout the spacetime-plus-one, the fermion has drawn an amount of energy from the vacuum that is precisely equal to the rest energy now retained by that fermion. This is energy conservation appearing in yet another guise, and it is the method by which fermions draw energy out of the vacuum using the Higgs field to acquire their rest masses.

At this point, we are ready to study the experimental masses of the observed elementary fermions, and we begin with the top quark, which is the heaviest of all. Using empirical mass data

from PDG's [43], this quark has a rest energy of  $m_t c^2 = 173.0 \pm 0.4$  GeV. We pointed out moments ago that at the origin for a given fermion,  $\phi_h(\mathbf{0}) = \frac{1}{\sqrt{2}} v - m_f c^2$ . With error bars now included, we may also calculate  $v / \sqrt{2} = 174.1035847 \pm 0.0000448$  GeV. The vev is known with four orders of magnitude greater precision than is the top mass, so the error spread is dominated by the top quark. Thus, from (13.6) we deduce that for the top quark:

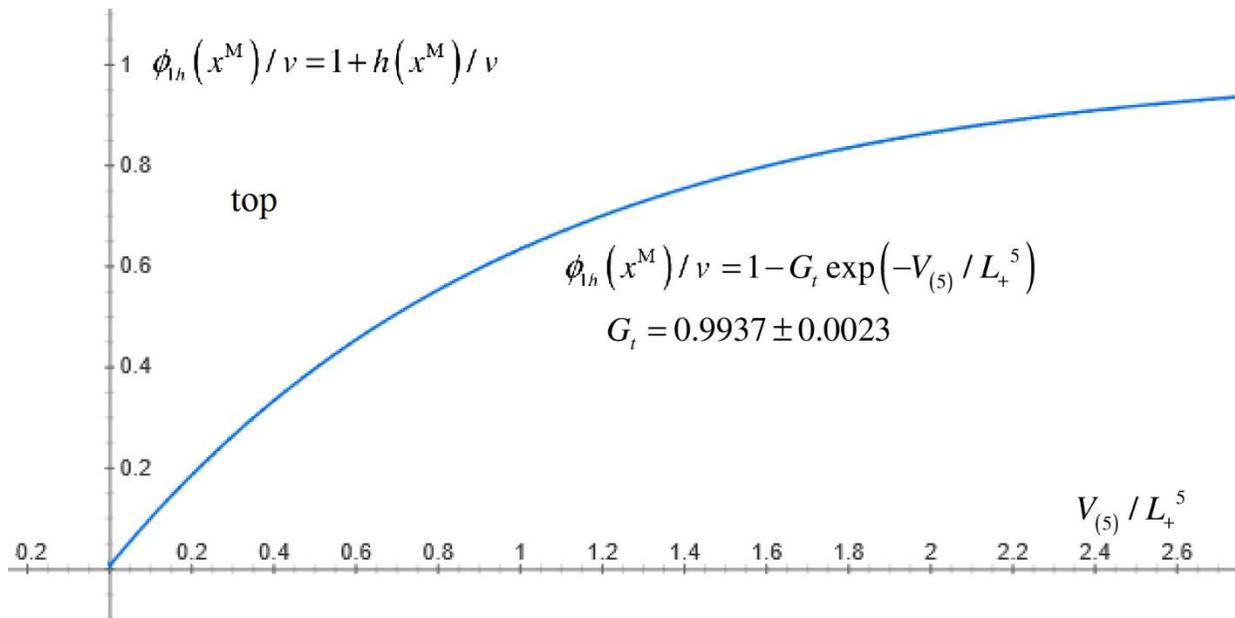
$$\phi_h(\mathbf{0}) = \frac{1}{\sqrt{2}} v - m_t c^2 = 1.1 \pm .04 \text{ GeV}. \quad (13.8)$$

In other words,  $\phi_h(x^M) = \frac{1}{\sqrt{2}} v \cong 174.1$  GeV far from the origin, but it drops all the way down to a mere  $\phi_h(\mathbf{0}) = 1.1 \pm .04$  GeV close to the origin. So, at the origin of the Higgs well in which the top quark nests, *almost all, but not all*, of the energy has been removed from the Fermi vacuum to give the top quark its rest mass.

It is very helpful to calculate  $G_t = \sqrt{2} m_f c^2 / v = 0.9937 \pm 0.0023$  for the top quark, then apply the upper (13.6) from the symmetry breaking in (13.4), to obtain:

$$\frac{\phi_{1h}(x^M)}{v} = \sqrt{2} \frac{\phi_h(x^M)}{v} = 1 + \frac{h(x^M)}{v} = 1 - \sqrt{2} \frac{m_f c^2}{v} \exp\left(-\frac{V_{(5)}}{L_+^5}\right) = 1 - G_t \exp\left(-\frac{V_{(5)}}{L_+^5}\right). \quad (13.9)$$

This, in turn, is illustrated graphically as shown below:



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

So, the top quark plot above together with (13.7) illustrates a first example of how the fermions generally, acquire their rest masses by quite literally sucking out of the vacuum via the Higgs field,

an energy exactly equal to their rest energies. But the top quark, as we shall see, is unique insofar as Figure 1 shows a very deep well which grows very close to zero where the top quark sits. Specifically, because the top quark coupling is very close to 1, the question first arises whether this coupling could in fact be equal to 1. But even when we consider the upper extremum of experimental error bars whereby  $0.9914 \leq G_t \leq 0.9960$ , this coupling is always less than 1. Thus, even at the locale where it nests, the top quark draws almost all, but not all, of the energy out of the Fermi vacuum. That this coupling is very close to 1, but slightly less than 1, in fact provides an extremely-important clue that will clear the way for the developing the theory of fermion masses, mixing angles, Lagrangian potentials and beta decays which is the central focus of the present Part II of this paper.

As we now start speak about quark masses and their “errors,” it must be noted that the error bars of the quark masses in from [43] 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.

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 – 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  $t\overline{H\overline{H}}$  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.

## PART IIA: QUARKS

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

As stated in the previous section, it is highly intriguing in its own right that the coupling  $G_t = 0.9937 \pm 0.0023$  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, or by some renormalization “scheme” by which it becomes 1. This leads us to raise the question whether the *sum* of the up-plus-charm-plus-top rest energies might yield a total energy for which the coupling is equal to 1 *within experimental and scheme-dependent errors*, and if so, whether this could be of theoretical

significance toward developing a viable theory which solves the presently-unsolved mystery puzzle of why the fermions have the rest masses that they do and how these masses relate to the CKM and PMNS mixing angles.

Using the empirical value  $v/\sqrt{2} = 174.1035847$  GeV sans error range, and using empirical mass data from PDG's [43], the dimensionless couplings  $G_f = m_f c^2 / \frac{1}{\sqrt{2}} v$  for the up, charm and top quarks are calculated to be:

$$G_u = 0.000013_{-0.000002}^{+0.000003}; \quad G_c = 0.00732_{-0.00020}^{+0.00014}; \quad G_t = 0.9937 \pm 0.0023, \quad (14.1)$$

In (14.1) we see again how  $G_t$  is just under 1, irrespective of the error bars. In other words, as already pointed out, 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.

It turns out, however, 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:

$$G_u + G_c + G_t = (m_u + m_c + m_t) c^2 / \frac{1}{\sqrt{2}} v = 1.00103_{-0.00250}^{+0.00244}, \quad (14.2)$$

i.e.,  $0.99853 < G_u + G_c + G_t < 1.00348$ . 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 relation of true physical significance. If this is so, then because  $v = 246.2196508 \pm 0.0000633$  GeV 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 theoretically 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, it becomes possible 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, presently unknown, masses of the neutrinos. Over the next several sections, we shall take up each of these four matters.

First, let us use the empirical data that  $0.99853 < G_{u+c+t} < 1.00348$  from (14.2) to *postulate* that in fact, this coupling sum

$$G_u + G_c + G_t = (m_u + m_c + m_t) c^2 / \frac{1}{\sqrt{2}} v \equiv 1, \quad (14.3)$$

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

$$m_u c^2 + m_c c^2 + m_t c^2 \equiv \frac{1}{\sqrt{2}} v = 174.1035847 \text{ GeV} . \quad (14.4)$$

We use the center of the Fermi vev error range in  $v = 246.2196508 \pm 0.0000633 \text{ GeV}$  because the vev precision is far greater than the precision in either  $m_u = .0022^{+.0005}_{-.0004} \text{ GeV}$  or  $m_c = 1.275^{+.025}_{-.035} \text{ GeV}$ , as well as in  $m_t = 173.0 \pm 0.4 \text{ GeV}$ , see [43]. So, we need not be concerned with the errors in  $v$ , but instead will account for the errors particularly in  $m_c$  by transferring these over to the top mass. Combining (14.4) with the known up and charm masses we then deduce that:

$$\boxed{m_t c^2 = 174.1035847 \text{ GeV} - m_c c^2 - m_u c^2 = 172.826^{+.035}_{-.026} \text{ GeV}} . \quad (14.5)$$

The .061 GeV error spread in the above is tighter than the currently-known 0.8 GeV spread from  $m_t = 173.0 \pm 0.4 \text{ GeV}$  by more than a full order of magnitude. *This result in (14.5) 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 (14.5) is true, then it is also convenient for the next step to collect all of the isospin-up quark masses together as such:

$$m_u c^2 = .0022^{+.0005}_{-.0004} \text{ GeV}; \quad m_c c^2 = 1.275^{+.025}_{-.035} \text{ GeV}; \quad m_t c^2 = 172.826^{+.035}_{-.026} \text{ GeV} . \quad (14.6)$$

We may also revise the isospin-up quark couplings  $G_f = m_f c^2 / \frac{1}{\sqrt{2}} v$  in (14.1) as such:

$$G_u = 0.000013^{+.0000003}_{-.0000002}; \quad G_c = 0.00732^{+.000014}_{-.000020}; \quad G_t = 0.99266^{+.000021}_{-.000014}, \quad (14.7)$$

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 isospin-up quarks 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 (13.7). But if there was to exist a single quark with the sum (14.4) 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 way 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 degrees of freedom), and for the symmetry

to then be broken in the manner reviewed leading to (11.12) whereby the top quark gains a mass of  $m_t c^2 \equiv \frac{1}{\sqrt{2}} v = 174.1035847 \text{ GeV}$  which depresses the vacuum down to a rock bottom 0 GeV at the origin of the Higgs well. Then, some of this mass may be 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 as reviewed four decades ago in [45].

To achieve this, we begin with a mass matrix defined by:

$$M_{uct} c^2 \equiv \begin{pmatrix} m_t & \sqrt{m_t m_c} & \sqrt{m_t m_u} \\ \sqrt{m_t m_c} & m_c & \sqrt{m_c m_u} \\ \sqrt{m_t m_u} & \sqrt{m_c m_u} & m_u \end{pmatrix} c^2 = \begin{pmatrix} \frac{1}{\sqrt{2}} v & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 174.1035847 \text{ GeV} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (14.8)$$

Then, we transform this into  $M_{uct} \rightarrow M'_{uct} = U^\dagger M_{uct} U$ , where  $U$  is a unitary matrix  $U^\dagger U = 1$ . An important point to note, is that under a bi-unitary transformation the trace  $\text{tr}(M_{uct}) = \text{tr}(M'_{uct})$  is preserved so that  $m_u c^2 + m_c c^2 + m_t c^2 \equiv \frac{1}{\sqrt{2}} v$  in (14.4) 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 mass mixing matrix  $U$ . As a 3x3 matrix this could have up the three real angles  $\theta_{21}$ ,  $\theta_{32}$ ,  $\theta_{31}$  and one imaginary phase  $\delta$  – thus four angles in total – in the same manner as the CKM and PMNS mixing matrices used to characterize generation-changing weak interaction beta decays for quarks, and leptonic neutrino oscillations. But the up, charm and top masses represent only three unknown mass parameters. Moreover, because the Fermi vev is known, (14.4) reduces this down to two unknown parameters. Thus, we ought not use more than two real angles without a phase 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  $\theta_{21}$ ,  $\theta_{32}$ ,  $\theta_{31}$  and structure the matrices accordingly.

Proceeding, we may choose a first parametrization using  $\theta_{32}$  and  $\theta_{21}$ , whereby some of the mass in  $m_t c^2 = \frac{1}{\sqrt{2}} v$  first is rotated using  $\theta_{32}$  into  $m_c c^2$ , then “downward cascades” using  $\theta_{21}$  into  $m_u c^2$ . Likewise, we may choose a second parameterization wherein the top quark mass is “distributed” to both the charm and up quarks by  $\theta_{32}$  and  $\theta_{31}$  respectively. A third parametrization uses first rotates the top mass rotates into the up quark using  $\theta_{31}$ , then “upward cascades” into the charm quark using  $\theta_{21}$ . For reasons that momentarily become apparent, it is fruitful to develop both the first and second “downward cascade” and the “distribution” parameterizations, while the third “upward cascade” parameterization turns out to be duplicative of the first but with a flip and 90-degree rotation of one of the angles.

Using the first, “downward cascade” parameterization, this bi-unitary transformation is:

$$\begin{aligned}
 M_{uct} c^2 &\rightarrow M'_{uct} c^2 = U^\dagger M_{uct} c^2 U \\
 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{21} & -s_{21} \\ 0 & s_{21} & c_{21} \end{pmatrix} \begin{pmatrix} c_{32} & -s_{32} & 0 \\ s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} v & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_{32} & s_{32} & 0 \\ -s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{21} & s_{21} \\ 0 & -s_{21} & c_{21} \end{pmatrix}. \quad (14.9a) \\
 &= \frac{1}{\sqrt{2}} v \begin{pmatrix} c_{32}^2 & c_{32} s_{32} c_{21} & c_{32} s_{32} s_{21} \\ c_{32} s_{32} c_{21} & s_{32}^2 c_{21}^2 & s_{32}^2 c_{21} s_{21} \\ c_{32} s_{32} s_{21} & s_{32}^2 c_{21} s_{21} & s_{32}^2 s_{21}^2 \end{pmatrix}_I = \begin{pmatrix} m'_t & \sqrt{m'_t m'_c} & \sqrt{m'_t m'_u} \\ \sqrt{m'_t m'_c} & m'_c & \sqrt{m'_c m'_u} \\ \sqrt{m'_t m'_u} & \sqrt{m'_c m'_u} & m'_u \end{pmatrix} c^2
 \end{aligned}$$

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 second, “distribution” parameterization:

$$\begin{aligned}
 M_{uct} c^2 &\rightarrow M'_{uct} c^2 = U^\dagger M_{uct} c^2 U \\
 &= \begin{pmatrix} c_{31} & 0 & -s_{31} \\ 0 & 1 & 0 \\ s_{31} & 0 & c_{31} \end{pmatrix} \begin{pmatrix} c_{32} & -s_{32} & 0 \\ s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} v & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_{32} & s_{32} & 0 \\ -s_{32} & c_{32} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_{31} & 0 & s_{31} \\ 0 & 1 & 0 \\ -s_{31} & 0 & c_{31} \end{pmatrix}. \quad (14.9b) \\
 &= \frac{1}{\sqrt{2}} v \begin{pmatrix} c_{32}^2 c_{31}^2 & c_{32} s_{32} c_{31} & c_{32}^2 c_{31} s_{31} \\ c_{32} s_{32} c_{31} & s_{32}^2 & c_{32} s_{32} s_{31} \\ c_{32}^2 c_{31} s_{31} & c_{32} s_{32} s_{31} & c_{32}^2 s_{31}^2 \end{pmatrix}_{II} = \begin{pmatrix} m'_t & \sqrt{m'_t m'_c} & \sqrt{m'_t m'_u} \\ \sqrt{m'_t m'_c} & m'_c & \sqrt{m'_c m'_u} \\ \sqrt{m'_t m'_u} & \sqrt{m'_c m'_u} & m'_u \end{pmatrix} c^2
 \end{aligned}$$

There are three mathematical points to now note in (14.9). First, as already mentioned, the trace is preserved under (14.9), 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 the mass sum (14.4) 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, although the masses and their associated couplings related by  $\frac{1}{\sqrt{2}} v G_f = m_f c^2$  are the same no matter which parameterization scheme we use, *the angles are defined differently depending on the scheme*. For this reason, outside the matrix containing the sines and cosines of these angles on the final lines of (14.9), we have denoted all of the angles by the *I* and *II* subscripts for the first and second schemes, respectively. Note also, if we use the coupling relation  $\frac{1}{\sqrt{2}} v G_f = m_f c^2$ , then dropping the primes of the transformations in (14.9) from here on, we can explicitly identify these couplings in the first and second parameterizations to be:

$$G_{uct} = \begin{pmatrix} G_t & \sqrt{G_t G_c} & \sqrt{G_t G_u} \\ \sqrt{G_t G_c} & G_c & \sqrt{G_c G_u} \\ \sqrt{G_t G_u} & \sqrt{G_c G_u} & G_u \end{pmatrix} = \begin{pmatrix} c_{32}^2 & c_{32} s_{32} c_{21} & c_{32} s_{32} s_{21} \\ c_{32} s_{32} c_{21} & s_{32}^2 c_{21}^2 & s_{32}^2 c_{21} s_{21} \\ c_{32} s_{32} s_{21} & s_{32}^2 c_{21} s_{21} & s_{32}^2 s_{21}^2 \end{pmatrix}_I, \quad (14.10a)$$

$$G_{uct} = \begin{pmatrix} G_t & \sqrt{G_t G_c} & \sqrt{G_t G_u} \\ \sqrt{G_t G_c} & G_c & \sqrt{G_c G_u} \\ \sqrt{G_t G_u} & \sqrt{G_c G_u} & G_u \end{pmatrix} = \begin{pmatrix} c_{32}^2 c_{31}^2 & c_{32} s_{32} c_{31} & c_{32}^2 c_{31} s_{31} \\ c_{32} s_{32} c_{31} & s_{32}^2 & c_{32} s_{32} s_{31} \\ c_{32}^2 c_{31} s_{31} & c_{32} s_{32} s_{31} & c_{32}^2 s_{31}^2 \end{pmatrix}_{II}. \quad (14.10b)$$

Note that for both of these, the trace  $\text{tr}G_{uct} = G_t + G_c + G_u = 1$ . This is another reflection of (14.4).

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 coupling data in (14.7) to calculate  $\theta_{I32}$  and  $\theta_{I21}$  in (14.10a), and  $\theta_{II32}$  and  $\theta_{II31}$  in (14.10b). From (14.10a) we first deduce  $c_{I32}^2 = G_t$ , then  $c_{I21}^2 = G_c / s_{I32}^2 = G_c / (1 - c_{I32}^2)$ , then ascertain the angles in both radians and degrees. From (14.10b) we likewise deduce  $s_{II32}^2 = G_c$  followed by  $c_{II31}^2 = G_t / c_{II32}^2 = G_t / (1 - s_{II32}^2)$  followed by the angles. In this way, we calculate that:

$$\begin{aligned} \theta_{I32} &= 0.08575_{-0.00120}^{+0.00085} \text{ rad} = 4.91338_{-0.06874}^{+0.04893} \circ \\ \theta_{II32} &= 0.08568_{-0.00119}^{+0.00084} \text{ rad} = 4.90914_{-0.06801}^{+0.04801} \circ \\ \theta_{I21} &= 0.04152_{-0.00343}^{+0.00403} \text{ rad} = 2.37864_{-0.19673}^{+0.23071} \circ \\ \theta_{II31} &= 0.00357_{-0.00034}^{+0.00038} \text{ rad} = 0.20442_{-0.01953}^{+0.02206} \circ \end{aligned} \quad (14.11)$$

The scheme-dependent  $\theta_{32}$  differ but slightly between these two parameterizations, and the angles of approximately  $4.91^\circ$  do not “ring any bells” regarding other known empirical data. *But for  $\theta_{I21}$  and  $\theta_{II31}$  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):

$$\begin{aligned} \theta_{C12} &= 0.2265 \pm 0.0005 \text{ rad} = 12.975 \pm 0.026 \circ \\ \theta_{C13} &= 0.0036_{-0.0002}^{+0.0003} \text{ rad} = 0.209_{-0.013}^{+0.015} \circ \\ \theta_{C23} &= 0.0422 \pm 0.0009 \text{ rad} = 2.415 \pm 0.053 \circ \\ \delta_C &= 1.2391_{-0.0335}^{+0.0348} \text{ rad} = 70.998_{-1.917}^{+1.995} \circ \end{aligned} \quad (14.12)$$

Doing the comparisons, we see that  $\theta_{I21} = 2.37864_{-0.19673}^{+0.23071} \circ$  versus  $\theta_{C23} = 2.415 \pm 0.053 \circ$  which overlap within the error bars, and that  $\theta_{II31} = 0.20442_{-0.01953}^{+0.02206} \circ$  versus  $\theta_{C13} = 0.209_{-0.013}^{+0.015} \circ$  which

likewise overlap within the error bars. In fact,  $\theta_{I_{21}}$  which has a wider error bar, has a central portion fitting entirely within the error range for  $\theta_{C_{23}}$ , and  $\theta_{II_{31}}$  with a wider spread also has a central region fitting entirely within the errors for  $\theta_{C_{13}}$ .

So, our goal was to see whether the mass mixing angles in the bi-unitary transformation  $M_{uct} \rightarrow M'_{uct} = U^\dagger M'_{uct} U$  bore any relation to any known data. And in the comparison between (14.11) and (14.12) we found that we have two “hits” directly in the middle of the empirical data for two of the three real CKM mixing angles. (In the next section we will likewise connect with the third real CKM angle using the isospin-down quark masses.) With two *simultaneous* such hits not just one, the statistical chances of this being a coincidence are extremely remote. Therefore, we now conclude that this concurrence between (14.11) and (14.12) in fact is the discovery of two fundamental physical relations:

$$\boxed{\begin{aligned} \theta_{I_{21}} &\equiv \theta_{C_{23}} = 2.415 \pm 0.053^\circ \\ \theta_{II_{31}} &\equiv \theta_{C_{13}} = 0.209^{+0.015}_{-0.013}^\circ \end{aligned}} \quad (14.13)$$

Above, we use the empirical data from the CKM angles because their error bars are smaller.

Once we have made the connections in (14.13), 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 (14.10) which were used to obtain (14.11), we may now use (14.13) to find that:

$$\begin{aligned} \cos^2 \theta_{C_{23}} = \cos^2 \theta_{I_{21}} &= \frac{G_c}{\sin^2 \theta_{I_{32}}} = \frac{G_c}{1 - \cos^2 \theta_{I_{32}}} = \frac{G_c}{1 - G_t} \\ \cos^2 \theta_{C_{13}} = \cos^2 \theta_{II_{31}} &= \frac{G_t}{\cos^2 \theta_{II_{32}}} = \frac{G_t}{1 - \sin^2 \theta_{II_{32}}} = \frac{G_t}{1 - G_c} \end{aligned} \quad (14.14)$$

Then, solving (14.14) as simultaneous equations in  $G_t$  and  $G_c$ , while also using  $G_u = s_{I_{32}}^2 s_{I_{21}}^2 = c_{II_{32}}^2 s_{II_{31}}^2$  from (14.10) along with (14.13), we are able to deduce that:

$$G_t = \frac{\sin^2 \theta_{C_{23}} \cos^2 \theta_{C_{31}}}{1 - \cos^2 \theta_{C_{23}} \cos^2 \theta_{C_{31}}}; \quad G_c = \frac{\cos^2 \theta_{C_{23}} \sin^2 \theta_{C_{31}}}{1 - \cos^2 \theta_{C_{23}} \cos^2 \theta_{C_{31}}}; \quad G_u = G_c \tan^2 \theta_{C_{23}} = G_t \tan^2 \theta_{C_{31}}. \quad (14.15)$$

The quark masses are then related to these by  $m_f c^2 = \frac{1}{\sqrt{2}} v G_f$ .

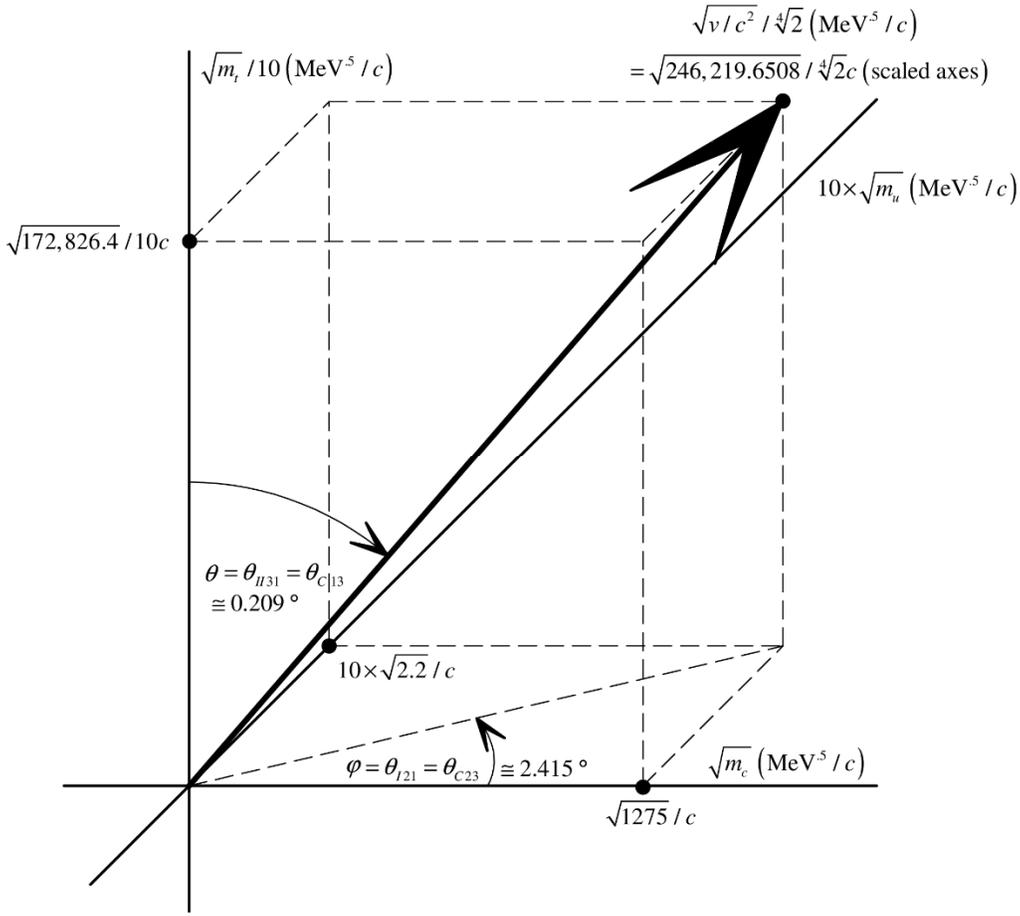
Consequently, (14.15) expresses the isospin-up quark masses entirely in terms of the CKM angles  $\theta_{C_{31}}$  and  $\theta_{C_{23}}$  which mix the third-generation quarks with the first and second generations, and the Fermi vev  $v$ . Only two of the three relations (14.15) are independent. But together with (14.5) 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  $m_u, m_c, m_t = F(v, \theta_{C31}, \theta_{C23})$  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 for 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.* This is the first step toward doing similarly, over the next several sections, for all twelve fermion masses.

Additionally, there is a very simple geometric interpretation of the results in (14.14). From (14.10) we may use  $G_t + G_c + G_u = 1$  then  $\frac{1}{\sqrt{2}}vG_f = m_f c^2$  to rewrite (14.14) as:

$$\begin{aligned} \cos^2 \theta_{C23} = \cos^2 \theta_{I21} &= \frac{G_c}{1-G_t} = \frac{G_c}{G_c + G_u} = \frac{\sqrt{G_c^2}}{\sqrt{G_c^2} + \sqrt{G_u^2}} = \frac{\sqrt{m_c^2}}{\sqrt{m_c^2} + \sqrt{m_u^2}} \\ \cos^2 \theta_{C13} = \cos^2 \theta_{II31} &= \frac{G_t}{1-G_c} = \frac{G_t}{G_t + G_u} = \frac{\sqrt{G_t^2}}{\sqrt{G_t^2} + \sqrt{G_u^2}} = \frac{\sqrt{m_t^2}}{\sqrt{m_t^2} + \sqrt{m_u^2}} \end{aligned} \quad (14.16)$$

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  $\theta$  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 (14.6):



**Figure 2: Isospin-Up Quark Mixing in Rest Mass Space**

Because the top quark mass is so much larger than the up quark mass, even after taking square roots the top-to-up ratio is still about 280-to-1. So, any visual representation drawn to scale would 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.*

Two final points are worth noting before we move on to examine the isospin-down quark masses. First, as between the first and second parameterizations, we also uncovered two other angles  $\theta_{132}$  and  $\theta_{112}$  in (14.11). And we did not develop an available third parameterization using what we shall denote as  $\theta_{1121}$  and  $\theta_{1131}$ . This is because as noted, the angles obtained from the bi-unitary transformations in (14.9) are parameter-dependent but the masses and their couplings are not nor can they be. So, once we attach a physical significance to  $\cos^2 \theta_{121}$  and  $\cos^2 \theta_{1131}$  in (14.13) we have squeezed all of the *independent* information we can out the bi-unitary transformations. The remaining angles  $\theta_{132}$  and  $\theta_{112}$  furnish no further information, as they are not independent of

the physical connections established in (14.13) but simply contain redundant information. Likewise, the third parameterization using what we shall denote as  $\theta_{III21}$  and  $\theta_{III31}$  produces a  $\theta_{III21} = 90^\circ - \theta_{I21}$  and  $\theta_{III31} = \theta_{I32}$  which effectively flips then rotates  $\theta_{I21}$  by 90 degrees, renames  $\theta_{I32}$  to  $\theta_{III31}$ , and in the process, also flips the signs of all the square roots which contain  $G_c$ . As such, this too is redundant and adds no new independent data.

Second, in Figure 1 we used  $G_t = \sqrt{2}m_f c^2 / v = 0.9937 \pm 0.0023$  to plot (13.9) for the top quark. At the origin  $V_{(5)} / L_+^5 = 0$  this plot bottomed out at  $\sqrt{2}\phi_h / v = 1 - G_t = 0.0063 \mp 0.0023$ , signifying the extraction of almost all the energy from the vacuum at the center of the well in which the top quark nests. With the tightened  $G_t = 0.99266_{-0.00014}^{+0.00021}$  in (14.7), Figure 1 is now slightly modified with the bottom at  $\sqrt{2}\phi_h / v = 1 - G_t = 0.00734_{+0.00021}^{-0.00014}$  to reflect this revised coupling. In view of (14.7), however, we may now also draw plots similar to Figure 1 for the up and charm quarks. These plots have the exact same form as Figure 1, with the only difference being how deeply the energy well descends from  $\sqrt{2}\phi_h / v = 1$  i.e.  $\phi_h = v / \sqrt{2}$ . For the charm quark with  $G_c = 0.00732_{-0.00020}^{+0.00014}$  the well at the origin bottoms out at  $\sqrt{2}\phi_h / v = 1 - G_c = 0.99268_{+0.00014}^{-0.00020}$ . Meanwhile for the up quark with  $G_u = 0.000013_{-0.000002}^{+0.000003}$  the well at the origin bottoms out at  $\sqrt{2}\phi_h / v = 1 - G_u = 0.999987_{+0.000003}^{-0.000002}$ . Because the top quark is far more massive than the charm and up quarks, it is no surprise to respectively see dimensionless origin bottoms of 0.00734, 0.99268 and 0.999987 at the center of the error bars: The top quark draws almost all of the local energy from the Fermi vacuum to supply its rest mass, while the charm and especially the up quark draw very little local energy to supply their rest masses.

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

If it is possible to express the three up quark masses as  $m_u, m_c, m_t = F(v, \theta_{C31}, \theta_{C23})$ , and given that the two CKM angles parameterize generation changes during weak beta decays between isospin-up and isospin-down quarks, and because  $\theta_{C12} = 12.975 \pm 0.026^\circ$  in (14.12) is still unaccounted for, it is natural to examine whether a carbon copy of the bi-unitary transformations in the last section can also be used to characterize the down, strange and bottom quark masses in a similar fashion, while also relating the unaccounted  $\theta_{C12} = 12.975 \pm 0.026^\circ$  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 clearly 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 further establish this distinction.

To cut right to the chase, let us replicate (14.10) identically, but with the substitutions  $u \mapsto d$ ,  $c \mapsto s$  and  $t \mapsto b$ , as well as a  $\Downarrow$  subscript for the angles, as such:

$$G_{dsb} = \begin{pmatrix} G_b & \sqrt{G_b G_s} & \sqrt{G_b G_d} \\ \sqrt{G_b G_s} & G_s & \sqrt{G_s G_d} \\ \sqrt{G_b G_d} & \sqrt{G_s G_d} & G_d \end{pmatrix} = \begin{pmatrix} c_{32}^2 & c_{32} s_{32} c_{21} & c_{32} s_{32} s_{21} \\ c_{32} s_{32} c_{21} & s_{32}^2 c_{21}^2 & s_{32}^2 c_{21} s_{21} \\ c_{32} s_{32} s_{21} & s_{32}^2 c_{21} s_{21} & s_{32}^2 s_{21}^2 \end{pmatrix}_{I\Downarrow}, \quad (15.1a)$$

$$G_{dsb} = \begin{pmatrix} G_b & \sqrt{G_b G_s} & \sqrt{G_b G_d} \\ \sqrt{G_b G_s} & G_s & \sqrt{G_s G_d} \\ \sqrt{G_b G_d} & \sqrt{G_s G_d} & G_d \end{pmatrix} = \begin{pmatrix} c_{32}^2 c_{31}^2 & c_{32} s_{32} c_{31} & c_{32}^2 c_{31} s_{31} \\ c_{32} s_{32} c_{31} & s_{32}^2 & c_{32} s_{32} s_{31} \\ c_{32}^2 c_{31} s_{31} & c_{32} s_{32} s_{31} & c_{32}^2 s_{31}^2 \end{pmatrix}_{II\Downarrow}. \quad (15.1b)$$

Here, it is clear that  $G_b + G_s + G_d = 1$ . As with (14.9) and (14.10), these coupling matrices utilizing first and second parameterizations with a  $\Downarrow$  subscript arise following a bi-unitary transformation  $M_{dsb} c^2 \rightarrow M'_{dsb} c^2 = U^\dagger_{\Downarrow} M'_{dsb} c^2 U_{\Downarrow}$  in which before the transformation,  $M_{dsb}$  places all of the rest mass into the bottom quark. Because the diagonals sum to 1,  $m_d c^2 + m_s c^2 + m_b c^2$  is invariant under these bi-unitary transformations. However, based on clear empirical data,  $m_d c^2 + m_s c^2 + m_b c^2 \neq m_u c^2 + m_c c^2 + m_t c^2 = \frac{1}{\sqrt{2}} v$ . So, to mirror the development of the last section we shall need to postulate a new, second, isospin-down vev *defined* by:

$$\frac{1}{\sqrt{2}} v_{\Downarrow} \equiv m_d c^2 + m_s c^2 + m_b c^2 = 4.2797^{+0.0495}_{-0.0333} \text{ GeV}. \quad (15.2)$$

Above, we have used the individual quark masses  $m_d c^2 = .0047^{+0.0005}_{-0.0003} \text{ GeV}$ ,  $m_s c^2 = .095^{+0.009}_{-0.003} \text{ GeV}$  and  $m_b c^2 = 4.18^{+0.04}_{-0.03} \text{ GeV}$  based on the empirical data in [43] to calculate the numerical value  $\frac{1}{\sqrt{2}} v_{\Downarrow} = 4.2797^{+0.0495}_{-0.0333} \text{ GeV}$ . The upside error of 49.5 MeV is based on the unlikely event of all three quarks having a mass at the upper end of their error bars and the downside error of 33.3 MeV conversely is based on all three quarks being at the lower end. Technically, because the bottom quark mass is only known to within 70 MeV, we should only show two digits past the decimal in (15.2), but to display the tighter strange and down mass errors, we show four digits after the decimal. Also, we now re-denote the fermi vev by  $v \mapsto v_{\Uparrow}$  to distinguish from the  $v_{\Downarrow}$  in (15.2). Thus, (14.4) re-denoted by  $\frac{1}{\sqrt{2}} v_{\Uparrow} = 174.1035847 \text{ GeV}$  now expresses how the Fermi vev, cut by the same  $\sqrt{2}$  factor, is equal to the sum of the isospin-up quarks. To 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  $\theta_{c12} = 12.975 \pm 0.026^\circ$  in the same way we have already accounted for the other two CKM angles. If we can, then the postulate is validated, and we need to then take further steps to properly accommodate two distinct vevs, one for isospin-up quark and one for isospin-down quarks, in our Lagrangian density potential for these quarks.

Specifically, by having both  $v_{\Uparrow}$  and  $v_{\Downarrow}$  vevs, with the latter being much smaller than the former, we are implicitly introducing the prospect that the Lagrangian potential  $V(\phi_h) = \frac{1}{2} \mu^2 \phi_h^2 + \frac{1}{4} \lambda \phi_h^4 + \dots$  not only has a first minimum at  $v_{\Uparrow} = \sqrt{2} \cdot 174.1035847 \text{ GeV}$ , but has

a *second minimum* at  $v_{\downarrow} = \sqrt{2} \cdot 4.2797 \text{ GeV}$  (sans error bars). 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 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  $v_{\downarrow} = \sqrt{2} \cdot 4.2797 \text{ GeV}$ , and indeed, which is ascertained *subject to the requirement* that  $V(\phi_h)$  have this second minimum at exactly at the domain point  $\phi_h = v_{\downarrow}$  as well as the usual first minimum at  $\phi_h = v_{\uparrow}$ .

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 isospin-down quark to  $G_{d,s,b} = m_{d,s,b} c^2 / \frac{1}{\sqrt{2}} v_{\downarrow}$ , using the second minimum at  $v_{\downarrow} = \sqrt{2} \cdot 4.2797_{-0.192}^{+0.284} \text{ GeV}$  rather than the first minimum at  $v_{\uparrow} = 246.2196508 \text{ GeV}$ . Then, similarly to the procedure followed prior to (14.11), we use (15.1a) to calculate  $c_{I\downarrow 32}^2 = G_b$  followed by  $c_{I\downarrow 21}^2 = G_s / s_{I\downarrow 32}^2$  followed by the two angles, and (15.1b) to calculate  $s_{II\downarrow 32}^2 = G_s$  followed by  $c_{II\downarrow 31}^2 = G_b / c_{II\downarrow 32}^2$  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  $\frac{1}{\sqrt{2}} v_{\downarrow} = 4.2797_{-0.0333}^{+0.0495} \text{ GeV}$  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 progressive calculations each based on different assumptions about the error bars in the isospin-down quark mass measurements and in the empirical CKM mixing angle  $\theta_{C12} = 12.975 \pm 0.026^\circ$  from (14.12).

Drawing again from PDG's [43], we start with the individual quark masses  $m_d c^2 = .0047_{-0.0003}^{+0.0005} \text{ GeV}$ ,  $m_s c^2 = .095_{-0.003}^{+0.009} \text{ GeV}$  and  $m_b c^2 = 4.18_{-0.03}^{+0.04} \text{ GeV}$  used in (15.2). In the first calculation we simply use the central value of the error bars in [43] for each of the three masses to calculate the four mass mixing angles as reviewed in the previous paragraph, obtaining:

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

At the lower end of the empirical  $\theta_{C12} = 12.975 \pm 0.026^\circ$  from (14.12) is  $\theta_{C12} = 12.949^\circ$ , which differs from  $\theta_{I\downarrow 21} = 12.540^\circ$  above by a mere  $0.409^\circ$ . Coupled with having already connected two mass mixing angles to the real CKM angles in (14.13), this leads us to suspect that  $\theta_{I\downarrow 21}$  is in fact physically equivalent to  $\theta_{C12}$ , 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 successful calculation summarized in (14.11).

It turns out that  $\theta_{I\downarrow 21}$  which was calculated to be  $12.540^\circ$  in (15.3) 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  $m_s c^2 = .095 \text{ GeV}$  and  $m_b c^2 = 4.18 \text{ GeV}$ , 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  $m_d c^2 = .0047_{-0.0003}^{+0.0005} \text{ GeV}$  and  $\theta_{C12} = 12.975 \pm 0.026^\circ$ . The combination of results turns out to be:

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

That is, now in MeV, with the bottom and strange quarks left at their centers, a down quark mass in the range  $m_d c^2 = 5.043 \pm .021 \text{ MeV}$  corresponds to the empirical range  $\theta_{C12} = 12.975 \pm 0.026^\circ$  for this CKM mixing angle. Because the error bars for the down quark mass allow this to be as large as  $m_d c^2 = 5.2 \text{ MeV}$ , we have now established that  $\theta_{I\downarrow 21} \equiv \theta_{C12}$  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  $m_d c^2 = 4.7 \text{ MeV}$ , it is necessary to reduce the strange quark mass somewhat to stay within the range of  $\theta_{C12} = 12.975 \pm 0.026^\circ$  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  $m_d c^2 = 4.7 \text{ MeV}$  to hit the CKM target of  $\theta_{C12} = 12.975 \pm 0.026^\circ$ , we start with a higher down quark mass assumed now to be  $m_d c^2 = 4.9 \text{ MeV}$ . Then, we examine the ranges of acceptable values for the strange quark mass which yield the empirical relation  $\theta_{C12} = 12.975 \pm 0.026^\circ$ . The result of this calculation are as follows:

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

Noting again that  $m_s c^2 = .095_{-0.003}^{+0.009} \text{ GeV}$ , 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  $m_d c^2 = 4.9 \text{ MeV}$  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  ${}^{+0.009}_{-0.003}$  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 the actual  $\theta_{C12}$  is likely on the low end of the  $\theta_{C12} = 12.975^\circ$  center value 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  $\theta_{C12} = 12.975 \pm 0.026^\circ$ .

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  $\theta_{C12}$ . Again, the bottom mass has negligible impact on any of this. So, we now take the step of defining  $\theta_{I\downarrow 21} \equiv \theta_{C12}$  as a true physical relationship based on this empirical concurrence, and we include this with (14.3) updated to differentiate isospin-up from isospin-down angles, to obtain:

$$\boxed{\begin{aligned} \theta_{I\downarrow 21} &\equiv \theta_{C12} = 12.975 \pm 0.026^\circ \\ \theta_{II\downarrow 31} &= 1.921^\circ \\ \theta_{I\uparrow 21} &\equiv \theta_{C23} = 2.415 \pm 0.053^\circ \\ \theta_{II\uparrow 31} &\equiv \theta_{C13} = 0.209^{+0.015}_{-0.013}^\circ \end{aligned}} \quad (15.6)$$

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 [33] and [10.1] of [31], in 2013 the author demonstrated the existence of a pair of simultaneous equations

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

through which the up and down quark masses may be deduced with extremely high precision based on the tightly-known, scheme-independent rest mass  $m_e$  of the electron and the tightly-known, scheme independent difference  $M_n - M_p$  between the free neutron and proton masses. 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 of various nuclides express *indirect* influences and manifestations of the objective quark masses, and are essentially “fingerprints” or parts of a “nuclear genome” to be “decoded,” from which the quark masses may be inferred. Using, (15.7) and the known data [47], [48] for the electron, neutron and proton masses, one may deduce very precise values for the up and down quark masses, which are:

$$\begin{aligned} m_u c^2 &= 0.002387339327 \text{ u} = 2.22379240 \text{ MeV} \\ m_d c^2 &= 0.005267312526 \text{ u} = 4.90647034 \text{ MeV} \end{aligned} \quad (15.8)$$

see [10.3] and [10.4] of [31] and prior to [3.12] in [34]. As will be reviewed in section 18, with updated data for the electron mass [47] and the neutron-minus-proton mass difference [49], these values in MeV are  $m_u c^2 = 2.22379229(55) \text{ MeV}$  and  $m_d c^2 = 4.90647034(55) \text{ MeV}$ , including error bars. So, in the fourth and final calculation we use this very precise value particularly of the down quark mass, which enables us to tighten up the error ranges for other quantities which are interconnected with this.

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

$$\begin{aligned} \text{if } m_b c^2 &= 4.18 \text{ GeV and } m_d c^2 = 4.90647034 \text{ MeV and } m_s c^2 = 92.039 \text{ MeV, then } \theta_{I\downarrow 21} = 13.001^\circ \\ \text{if } m_b c^2 &= 4.18 \text{ GeV and } m_d c^2 = 4.90647034 \text{ MeV and } m_s c^2 = 92.421 \text{ MeV, then } \theta_{I\downarrow 21} = 12.975^\circ \cdot (15.9) \\ \text{if } m_b c^2 &= 4.18 \text{ GeV and } m_d c^2 = 4.90647034 \text{ MeV and } m_s c^2 = 92.806 \text{ MeV, then } \theta_{I\downarrow 21} = 12.949^\circ \end{aligned}$$

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  $m_b c^2 = 4.18_{-0.03}^{+0.04} \text{ GeV}$ , 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 now related to this by  $\theta_{C12} = \theta_{I\downarrow 21}$  in (15.6) is virtually unaffected by variations in the bottom quark mass. Based on the match to the central empirical data  $\theta_{C12} = 12.975 \pm 0.026^\circ$  from (14.12), we shall henceforth use the middle line of (15.9) for the mid-range masses of the isospin-down quarks.

Because the results in (15.9) are impervious to bottom mass swings over the whole range of  $m_b c^2 = 4.18_{-0.03}^{+0.04} \text{ GeV}$ , 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  $m_s c^2 = 95_{-3}^{+9} \text{ MeV}$ , and is now tightened, using the ranges in (15.9), to:

$$\boxed{m_s c^2 = 92.42 \pm 0.38 \text{ MeV}} \quad (15.10)$$

*This is more than ten times as accurate as what is presently known for the strange quark mass, and constitutes an additional empirical prediction of this theory which can and should be tested.*

Because we have shown in (15.6) that  $\theta_{T\downarrow 21} = \theta_{C12} = 12.975 \pm 0.026^\circ$  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 empirical 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, recognizing that  $m_b c^2 = 4.18_{-0.03}^{+0.04} \text{ GeV}$  is the least-precise ingredient that goes into this vev, we use the very tight value for the down mass in (15.8) and the tightened strange mass in (15.10) to recalculate the numeric value in (15.2) to be  $\frac{1}{\sqrt{2}} v_{\downarrow} = 4.2773_{-0.0304}^{+0.0404} \text{ GeV}$ , with two digits still shown beyond those warranted by the bottom quark mass. Note, the high end of  $v_{\downarrow}$  corresponds to the low end of  $\theta_{C12}$  and vice-versa. Any further precision in this number will depend entirely upon 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:

$$\begin{aligned} v_{\downarrow} &= \sqrt{2} (m_d c^2 + m_s c^2 + m_b c^2) = 6.0491_{-0.0430}^{+0.0571} \text{ GeV} \\ v_{\uparrow} &= \sqrt{2} (m_u c^2 + m_c c^2 + m_t c^2) = 246.2196508 \pm 0.0000633 \text{ GeV} \end{aligned} \quad (15.11)$$

with a ratio  $v_{\uparrow} / v_{\downarrow} \equiv 40.7038$  at the center values. The Higgs field rest energy extraction plots for the isospin-down quarks look identical to that of Figure 1 with the depth dependent upon the particle mass, with the exception that while  $h / v = h / v_{\uparrow}$  in Figure 1, for the isospin-down quarks it becomes  $h / v = h / v_{\downarrow}$ , using the smaller vev. So, for the isospin-down quarks to acquire their masses, the energy extraction begins in a vacuum with a vev magnitude just over 40 times smaller than the Fermi vacuum vev, which directly reflects the ratio of the isospin-up mass sum to the isospin-down mass sum.

The connection in (15.6) whereby  $\theta_{T\downarrow 21} \equiv \theta_{C12} = 12.975 \pm 0.026^\circ$  also means that there are some additional theoretical relations between the CKM mixing angles and the  $\downarrow$  quark masses as represented by their couplings  $G_{d,s,b}(v_{\downarrow}) = m_{d,s,b} c^2 / \frac{1}{\sqrt{2}} v_{\downarrow}$ . These relations, assembled with the earlier (14.14) updated to reflect that these are for  $\uparrow$  quarks which use a different vev, and also using  $1 = G_u + G_c + G_t$  from (14.3) and  $1 = G_d + G_s + G_b$  from (15.1), are:

$$\begin{aligned}
 \cos^2 \theta_{C12} = \cos^2 \theta_{I\downarrow 21} &= \frac{G_s(v_\downarrow)}{\sin^2 \theta_{I\downarrow 32}} = \frac{G_s(v_\downarrow)}{1 - \cos^2 \theta_{I\downarrow 32}} = \frac{G_s(v_\downarrow)}{1 - G_b(v_\downarrow)} = \frac{G_s(v_\downarrow)}{G_d(v_\downarrow) + G_s(v_\downarrow)} \\
 \cos^2 \theta_{II\downarrow 31} &= \frac{G_b(v_\downarrow)}{\cos^2 \theta_{II\downarrow 32}} = \frac{G_b(v_\downarrow)}{1 - \sin^2 \theta_{II\downarrow 32}} = \frac{G_b(v_\downarrow)}{1 - G_s(v_\downarrow)} = \frac{G_b(v_\downarrow)}{G_d(v_\downarrow) + G_b(v_\downarrow)} \\
 \cos^2 \theta_{C23} = \cos^2 \theta_{I\uparrow 21} &= \frac{G_c(v_\uparrow)}{\sin^2 \theta_{I\uparrow 32}} = \frac{G_c(v_\uparrow)}{1 - \cos^2 \theta_{I\uparrow 32}} = \frac{G_c(v_\uparrow)}{1 - G_t(v_\uparrow)} = \frac{G_c(v_\uparrow)}{G_u(v_\uparrow) + G_c(v_\uparrow)} \\
 \cos^2 \theta_{C13} = \cos^2 \theta_{II\uparrow 31} &= \frac{G_t(v_\uparrow)}{\cos^2 \theta_{II\uparrow 32}} = \frac{G_t(v_\uparrow)}{1 - \sin^2 \theta_{II\uparrow 32}} = \frac{G_t(v_\uparrow)}{1 - G_c(v_\uparrow)} = \frac{G_t(v_\uparrow)}{G_u(v_\uparrow) + G_t(v_\uparrow)}
 \end{aligned} \tag{15.12}$$

We have also included the “leftover” angle  $\theta_{II\downarrow 31} = 0.034 \text{ rad} = 1.921^\circ$  which does not relate any other independently-known data, but which, like  $\cos^2 \theta_{C12} = \cos^2 \theta_{I\downarrow 21}$  above, is a function of the strange and bottom quark couplings.

Then, solving the top two (15.12) as simultaneous equations in  $G_b$  and  $G_s$ , while also using  $G_d = s_{I\downarrow 32}^2 s_{I\downarrow 21}^2 = c_{II\downarrow 31}^2 s_{II\downarrow 31}^2$  from (15.1) along with the results in (15.6), and assembling this with (14.15), we obtain:

$$\begin{aligned}
 G_b &= \frac{\sin^2 \theta_{C12} \cos^2 \theta_{II\downarrow 31}}{1 - \cos^2 \theta_{C12} \cos^2 \theta_{II\downarrow 31}}; \quad G_s = \frac{\cos^2 \theta_{C12} \sin^2 \theta_{II\downarrow 31}}{1 - \cos^2 \theta_{C12} \cos^2 \theta_{II\downarrow 31}}; \quad G_d = G_s \tan^2 \theta_{C12} = G_b \tan^2 \theta_{II\downarrow 31} \\
 G_t &= \frac{\sin^2 \theta_{C23} \cos^2 \theta_{C13}}{1 - \cos^2 \theta_{C23} \cos^2 \theta_{C13}}; \quad G_c = \frac{\cos^2 \theta_{C23} \sin^2 \theta_{C13}}{1 - \cos^2 \theta_{C23} \cos^2 \theta_{C13}}; \quad G_u = G_c \tan^2 \theta_{C23} = G_t \tan^2 \theta_{C13}
 \end{aligned} \tag{15.13}$$

In contrast to (14.15) where  $m_u, m_c, m_t = F(v_\uparrow, \theta_{C31}, \theta_{C23})$  so that all three quark masses may be expressed as a function of three independently-known 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  $\theta_{C12}$ . That is, for the isospin-down masses, we can write down a function  $m_d, m_s, m_b = F(v_\downarrow, \theta_{C12}, \theta_{II\downarrow 31})$  with the same pattern of relations, but both  $v_\downarrow$  and  $\theta_{II\downarrow 31}$  are only known by knowing the quark masses, so only the use of  $\theta_{C12}$  is truly a reparameterization. Another way to represent this is to write  $m_d, m_s, m_b = F(m_d, m_s, \theta_{C21})$  or alternatively  $m_d, m_s, m_b = F(m_d, m_b, \theta_{C21})$ , because  $\cos^2 \theta_{C21} = G_s / (1 - G_b)$  eliminates either  $G_s$  or  $G_b$  but not both as independent parameters. Thus, all told, we have now six previously-unexplained quark masses reduced to two unexplained quark masses plus the three CKM angles plus the Fermi vev. Part of the focus from here will be on reparameterizing these remaining two quark masses.

At (15.10) we tightened our knowledge of the strange quark mass using the very precise down quark mass in (15.8) and the tightened  $\theta_{I\downarrow 21} = \theta_{C12}$ . If we also use the up-quark mass

$m_u = 2.22379240$  MeV in (15.8) which is tighter than the presently-known  $m_u = .0022_{-0.0004}^{+0.0005}$  GeV by four orders of magnitude, then the bottom two relations in (15.12) which are the same as (14.14) enable us to also increase the precision of  $\theta_{C23}$  and  $\theta_{C13}$ . This is because in general, once we have relations between masses and mixing angles, we can use tighter angles to obtain tighter masses or tighter masses to obtain tighter angles, depending on whether we know the masses or the angles more tightly. So, with the up mass in (15.8) being very tight, we can tighten the angles. This calculation proceeds as follows: First, we calculate  $G_u = m_u c^2 / \frac{1}{\sqrt{2}} v_{\uparrow} = 0.00001277281$ , which is very precise because both  $m_u$  and  $v_{\uparrow}$  are very precise. Then, via (14.3), with equal precision, we may calculate that  $G_c + G_t = 1 - G_u = 0.99998722719$ .

Next, at (14.7) we recalculated  $G_t = 0.99266_{-0.00014}^{+0.00021}$  so as to inherit and reflect the error bars of the charm quark. So, using  $G_c + G_t = 0.99998722719$  to write the bottom two relations (15.12) entirely in terms of the top mass, then using this now-tightened mass, we obtain:

$$\begin{aligned} \cos^2 \theta_{C23} &= \frac{G_c}{1 - G_t} = \frac{0.99998722719 - G_t}{1 - G_t} = \frac{0.99998722719 - 0.99266_{-0.00014}^{+0.00021}}{1 - 0.99266_{-0.00014}^{+0.00021}} \\ \cos^2 \theta_{C13} &= \frac{G_t}{1 - G_c} = \frac{G_t}{0.00001277281 + G_t} = \frac{0.99266_{-0.00014}^{+0.00021}}{0.00001277281 + 0.99266_{-0.00014}^{+0.00021}} \end{aligned} \quad (15.14)$$

It is now the error bar in the charm quark mass, inherited by the up quark mass, which limits the precision of the numeric calculation in (15.14). What we then calculate, is that:

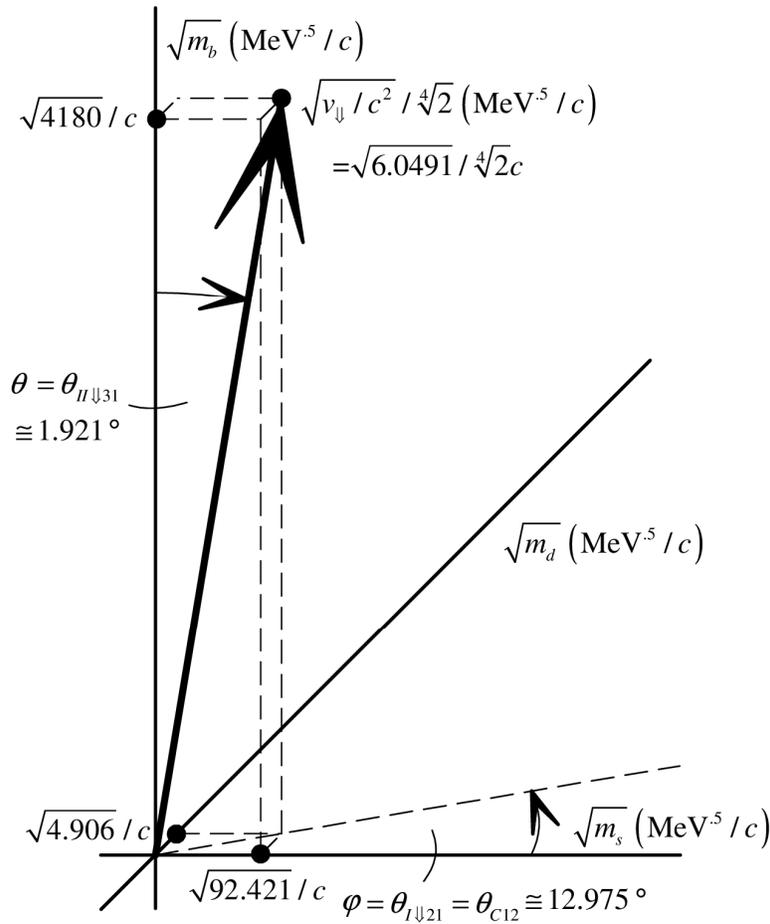
$$\begin{aligned} \theta_{C23} &= 0.0417_{-0.0004}^{+0.0006} \text{ rad} = 2.3911_{-0.0232}^{+0.0342} \circ \\ \theta_{C13} &= 0.0035871_{-0.0000004}^{+0.0000003} \text{ rad} = 0.20552_{-0.00002}^{+0.00001} \circ \end{aligned} \quad (15.15)$$

Comparing the upper relation (15.15) with presently-known empirical data  $\theta_{C23} = 2.415 \pm 0.053^\circ$  from (14.12), we see that the center value of this angle is reduced by  $0.0239^\circ = 1.4329'$  and that its error range is diminished from  $0.106^\circ$  to  $0.0574^\circ$ . This improves the accuracy by a factor of 1.8466 – almost 2. Comparing the lower (15.15) with the empirical  $\theta_{C13} = 0.209_{-0.013}^{+0.015} \circ$  also from (14.12), we see that the center value of this angle is reduced by  $0.00348^\circ = 0.2085'$ , while the error range is diminished from  $0.028^\circ$  to  $0.00004^\circ$  which improves the accuracy by a very large factor of 773.29 – almost three orders of magnitude. Thus, we see that  $\theta_{C13}$  which mixes the first and third quark generations is the recipient of this improved accuracy from both the up and top quark masses. *This is another prediction of the present theory which can and should be tested experimentally.*

Next, similarly to (14.16), using  $G_b + G_s + G_d = 1$  deduced from (15.1) we may rewrite the upper two relations (15.12) as:

$$\begin{aligned} \cos^2 \theta_{c21} = \cos^2 \theta_{l\downarrow 21} &= \frac{G_s}{1-G_b} = \frac{G_s}{G_s+G_d} = \frac{\sqrt{G_s^2}}{\sqrt{G_s^2} + \sqrt{G_d^2}} = \frac{\sqrt{m_s^2}}{\sqrt{m_s^2} + \sqrt{m_d^2}} \\ \cos^2 \theta_{ll\downarrow 31} &= \frac{G_b}{1-G_s} = \frac{G_b}{G_b+G_d} = \frac{\sqrt{G_b^2}}{\sqrt{G_b^2} + \sqrt{G_d^2}} = \frac{\sqrt{m_b^2}}{\sqrt{m_b^2} + \sqrt{m_d^2}} \end{aligned} \quad (15.16)$$

Then, as in Figure 2, 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, in contrast to Figure 2, 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 3: Isospin-Down Quark Mixing in Rest Mass Space**

Here too, what is remarkable, taken together with Figure 2, is that the azimuthal angle  $\varphi$  here corresponds also to the third of the three CKM mixing angles. Each of the four angles in Figures 2 and 3 is needed to specify projections of the vev vector associated into each of the individual masses, but only three of these angles are independent and used also for CKM mixing.

Taking stock of where we are at the moment, there are two reasons why, as reviewed, there are two free masses still remaining in the isospin-down portion of (15.13). First, there are only three real CKM mixing angles, not four. Two of those already went into the isospin-up quark masses  $m_u, m_c, m_t = F(v, \theta_{c31}, \theta_{c23})$ . All that was left for the isospin-down masses was  $\theta_{c21}$ . Second, for the isospin-up quarks, we had available  $v_{\uparrow} = v$  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 (15.11). 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  $v_{\downarrow}$ , separately from its value in (15.11) arrived at from the quark masses themselves. We will obtain this independent knowledge from the Higgs boson mass at (16.5) supra.

Related to this, prior to (15.3) we pointed out how the isospin-down coupling relation  $G_d + G_s + G_b = 1$  requires us to relate the isospin-down quark masses to the smaller vev  $v_{\downarrow}$  in (15.11) according to a relation we now denote by  $m_{\downarrow} c^2 = \frac{1}{\sqrt{2}} v_{\downarrow} G_{\downarrow}$ , while the isospin-up quark masses reviewed in the last section are related to the larger Fermi vev  $v_{\uparrow}$  by what we now denote as  $m_{\uparrow} c^2 = \frac{1}{\sqrt{2}} v_{\uparrow} G_{\uparrow}$ . Consequently, if we use the very tight down quark mass from (15.8) and the tightened strange quark mass from (15.10) along with the least-precise bottom quark mass  $m_b c^2 = 4.18_{-0.03}^{+0.04}$  GeV, then we are able to calculate:

$$G_d = 0.00115 \pm 0.00001; \quad G_s = 0.02161_{-0.00020}^{+0.00015}; \quad G_b = 0.97725_{+0.00021}^{-0.00016}. \quad (15.17)$$

In the above, recognizing that the error is predominantly driven by the bottom mass, we have used (15.8) and the center value of (15.10), and obtained the error variation entirely from the bottom quark mass. Effectively, this means we have now truncated the extra to digits shown in  $\frac{1}{\sqrt{2}} v_{\downarrow} = 4.2773_{-0.0304}^{+0.0404}$  GeV and used in (15.11), and simply used  $v_{\downarrow} / \sqrt{2} = 4.28_{-0.03}^{+0.04}$  GeV while only varying the value of the bottom quark mass, to obtain (15.17).

This also means that (13.9) which is (13.6) as graphed for the top quark in Figure 1, now becomes  $\phi_{h1}(x^M) / v_{\uparrow} = 1 - G_{\uparrow} \exp(-V_{(5)} / L_+^5)$  for the isospin-up quarks generally, while for the down quarks the counterpart to (13.9) is now  $\phi_{h1}(x^M) / v_{\downarrow} = 1 - G_{\downarrow} \exp(-V_{(5)} / L_+^5)$ . So, while the down quark curves will look the same as the curve in Figure 1, now we will have  $\phi_h(x^M) = v_{\downarrow} / \sqrt{2}$  far from the fermion, indicating a much smaller vev from which to draw energy for mass which corresponds directly to the smaller mass sum for the isospin-down quarks. Then, referring to (15.17), the bottom quark will draw out about 97.73% of the energy out of the energy made available from the smaller  $v_{\downarrow}$  (contrasted with the approximate 99.4% drawn by the top quark out of its larger  $v_{\uparrow}$  Fermi vev in Figure 1), the strange quark will draw about 2.16%, and the down quark will draw a mere 0.11%.

In conclusion, the standard model assumes that all fermion masses  $m_f$  are related to a single vev which is that of the Fermi vacuum,  $v=246.2196508\pm 0.0000633$  GeV, and that the associated couplings  $G_f$  are then calculated via the relation  $m_f c^2 = \frac{1}{\sqrt{2}} v G_f$  as against this one single vev. But again, these couplings are only known insofar as the fermion masses are known. The standard model has to date been incapable of relating these couplings to the CKM mixing angles as is done here via the relations (14.15), (14.16), (15.12), (15.13). However, in order to obtain these results, we have been required to utilize one relation  $m_{\uparrow} c^2 = \frac{1}{\sqrt{2}} v_{\uparrow} G_{\uparrow}$  for isospin-up quarks and a second relation  $m_{\downarrow} c^2 = \frac{1}{\sqrt{2}} v_{\downarrow} G_{\downarrow}$  for isospin-down quarks, and specifically, *we have been required to utilize one vev for the energy draw that produces the isospin-up quark masses and a second vev for the energy draw that produces the isospin-down quark masses.* And as we shall later see, the same pattern applies to leptons as well, so that we end up with a total of four distinct vevs, rather than just one.

Although this fits the empirical mass and mixing angle data in a way that has not heretofore been achieved by the standard model, one could fairly take the view that having two (and especially four) vevs rather than one vev is less “economical,” though one would also have to fairly acknowledge the inability of the standard model to relate the known masses to the known mixing angle as has been done here, for the quarks so far. So, if we are now to have two vevs rather than one vev to establish theoretical relationships between the quark masses and other independently-known parameters, it is necessary to properly and fully develop the relation between these two vevs. This is what brings us to the next stage in our development, which is to derive a Lagrangian potential which accounts for having two vev minima at both  $v_{\uparrow}$  and  $v_{\downarrow}$  in (15.11), rather than the usual single minimum at only the Fermi vev  $v_{\uparrow} = v$ . And as we shall see at (16.5) supra, it is the Higgs boson which sits at the center of the relation between these two vevs.

## 16. Theoretical Relation amongst the Higgs Mass and the Isospin-Up and Isospin-Down Quark vevs; and 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  $\theta_{T\downarrow 21} \equiv \theta_{C12} = 12.975 \pm 0.026^\circ$  with observed empirical data at (15.6) in addition to those already found for  $\theta_{C23}$  and  $\theta_{C13}$  at (14.13), this connection provided confirming evidence of this second vev. So, now that the empirical data apparently indicates that these two vevs do exist, it is important to understand how they are theoretically tied together. This brings us to the Lagrangian potential.

In the standard model for a U(1) gauge symmetry with a complex scalar field  $\phi_h$ , the Lagrangian potential is written as  $V = \mu^2 \phi_h^* \phi_h + \lambda (\phi_h^* \phi_h)^2 + \dots$  with higher-order terms above  $(\phi_h^* \phi_h)^2$  neglected. But if 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. Thus, it is desirable to start by briefly reviewing how the Fermi vacuum is used to

establish the vev at  $v_{\uparrow} = v$  in the standard model, as laid out, e.g., in sections 14.6 through 14.8 of [20] including Figure 13.3. Then we will take on the task of introducing a second vev at  $v_{\downarrow}$ .

In the standard model, with a gauge-covariant derivative  $D_{\mu} \equiv \partial_{\mu} + ieA_{\mu} / \hbar c$  and potential  $V = \mu^2 \phi_h^* \phi_h + \lambda (\phi_h^* \phi_h)^2 + \dots$  for complex scalar field  $\phi_h$ , one starts with a Lagrangian density:

$$\begin{aligned} \mathcal{L} = & \left( (\hbar c \partial_{\mu} + ieA_{\mu}) \phi_h \right)^* \left( (\hbar c \partial^{\mu} + ieA^{\mu}) \phi_h \right) - \mu^2 \phi_h^* \phi_h - \lambda (\phi_h^* \phi_h)^2 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \\ & + \frac{1}{2} m_A^2 c^4 A^{\mu} A_{\mu} - \frac{1}{2} m_{\phi_h}^2 c^4 \phi_h^2 \end{aligned} \quad (16.1)$$

In the above, for illustration only, we have also included hand-added terms  $+\frac{1}{2} m_A^2 c^4 A^{\mu} A_{\mu}$  for a massive gauge boson of mass  $m_A$  and  $-\frac{1}{2} m_{\phi_h}^2 c^4 \phi_h^2$  for a massive scalar boson of mass  $m_{\phi_h}$ . In four spacetime dimensions,  $\mathcal{L}$  has a mass dimensionality of +4, i.e., it has dimensions of energy to the fourth power. We then use  $\phi_h = \frac{1}{\sqrt{2}} (\phi_{1h} + i\phi_{2h})$  as reviewed at (13.4) to introduce the real and imaginary parts of  $\phi_h$ , and then break symmetry in favor of  $\phi_{1h}$  by setting  $\phi_{2h} = 0$ . Moreover, we also remove the hand-added vector and scalar boson mass terms entirely by setting  $m_A = 0$  and  $m_{\phi_h} = 0$  in the above, in favor of revealing these terms by other means. This is exactly how at (10.6) we set  $\hat{\mathcal{L}}_{\text{M}} = 0$  in (10.4) to require that the matter Lagrangian must emerge from the DKK geometry, as it did from the reassignment  $\hat{\mathcal{L}}_{\text{M}} \equiv (1/2\text{K}) \hat{R}^5_5$  in (10.5). This is also exactly how at (11.3) we set  $m = 0$  in (11.1) to likewise require that the fermion mass must emerge entirely from DKK geometry embodied in  $\Gamma_5$  and the fifth energy-momentum component  $cp^5$ , as it did via the reassignment  $-mc^2 U_0 \equiv \Gamma_5 cp^5 U_0 = (\gamma_5 cp^5 + \phi \gamma_0 cp^5) U_0$  in (11.2).

The upshot of all of these steps that that we end up setting  $\phi_h = \frac{1}{\sqrt{2}} \phi_{1h}$  in (16.1) while removing the illustrative mass terms, to next arrive at:

$$\mathcal{L} = \frac{1}{2} \hbar^2 c^2 \partial_{\mu} \phi_{1h} \partial^{\mu} \phi_{1h} + \frac{1}{2} e^2 \phi_{1h}^2 A_{\mu} A^{\mu} - \frac{1}{2} \mu^2 \phi_{1h}^2 - \frac{1}{4} \lambda \phi_{1h}^4 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu}. \quad (16.2)$$

The potential contained in the above has now become  $V = \frac{1}{2} \mu^2 \phi_{1h}^2 + \frac{1}{4} \lambda \phi_{1h}^4$ , expressed with  $\phi_{1h}$  rather than  $\phi_h$  as the domain points. It should also be cross-noted, and will become important shortly, that this very same  $\phi_{1h}$  is the range variable in (13.9) and Figure 1 for the top quark, and likewise for the other quarks in the manner reviewed near the end of the previous section.

The next step is to find the stationary points of  $V$ , i.e., those domain points at which  $V' = \partial V / \partial \phi_{1h} = \phi_{1h} (\mu^2 + \lambda \phi_{1h}^2) = 0$ . Clearly, these points occur where  $\phi_{1h} = 0$  and  $\mu^2 + \lambda \phi_{1h}^2$ , the latter of which means that  $\phi_{1h}^2 = -\mu^2 / \lambda$ . We assign the energy of this latter stationary point to the Fermi vev by definition, by setting  $\phi_{1h}^2 = -\mu^2 / \lambda \equiv v^2$ , thus  $\mu^2 = -\lambda v^2$ . Although the square

root of  $\phi_{1h}^2$  can be taken with either of the two possible sign choices  $\phi_{1h} = \pm v$ , because we are breaking symmetry we choose only one, customarily the positively signed  $\phi_{1h} = +v$ .

Next, we introduce the Higgs field as also reviewed at (13.4) by expanding  $\phi_{1h}$  in (16.2) about the Fermi vev using  $\phi_{1h} = v + h$ , with this  $\phi_{1h} = +v$  choice. We then substitute  $\phi_{1h} = v + h$  into (16.2) along with  $\mu^2 = -\lambda v^2$  just deduced. After consolidating terms, this produces:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \hbar^2 c^2 \partial^\mu h \partial_\mu h - \lambda v^2 h^2 - \lambda v h^3 - \frac{1}{4} \lambda h^4 + \frac{1}{4} \lambda v^4 \\ & + \frac{1}{2} e^2 v^2 A^\mu A_\mu + \frac{1}{2} e^2 h^2 A^\mu A_\mu + e^2 v h A^\mu A_\mu - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \end{aligned} \quad (16.3)$$

Comparing with (16.1) which makes us mindful that the Lagrangian term for a massive vector boson is expected to take the form  $+\frac{1}{2} m_A^2 c^4 A^\mu A_\mu$ , we identify  $m_A = ev$  (in natural units) as the mass of this boson, so as to now replace the hand-added mass term with one that emerged naturally from the Higgs field expansion. (Matching dimensionality explicitly, this association is really  $m_A^2 c^4 = k_e e^2 v^2 / \hbar c$  including Coulomb's constant  $k_e$ .) In non-Abelian gauge theory, for example weak interaction SU(2), an additional factor of  $1/2$  emerges so that the natural unit mass identification becomes  $m_W = \frac{1}{2} vg$  in relation to the coupling strength  $g$ . More importantly for present purposes, because the term for a massive scalar boson is expected to take the form  $-\frac{1}{2} m_{\phi_h}^2 c^4 \phi_h^2$  as also illustrated by this hand-added mass term in (16.1), we also identify  $\frac{1}{2} m_h^2 c^4 = \lambda v^2$  with the energy equivalent of the Higgs boson mass in the above. Thus, we have also replaced the hand-added scalar mass with a mass that also arises naturally from the Higgs expansion. Restructured,  $\lambda = m_h^2 c^4 / 2v^2$  informs us that the parameter  $\lambda$  is undetermined unless and until we know the mass of the Higgs boson.

As a result of the foregoing, focusing now on the potential  $V = \frac{1}{2} \mu^2 \phi_{1h}^2 + \frac{1}{4} \lambda \phi_{1h}^4$  and its derivative  $V' = \partial V / \partial \phi_{1h} = \phi_{1h} (\mu^2 + \lambda \phi_{1h}^2)$ , and using  $v = v_{\uparrow}$  to start distinguishing the Fermi vev from the  $v_{\downarrow}$  vev, we employ  $\mu^2 = -\lambda v_{\uparrow}^2$  and  $\lambda = m_h^2 c^4 / 2v_{\uparrow}^2$  to rewrite these as:

$$\begin{aligned} V = & \frac{1}{2} \mu^2 \phi_{1h}^2 + \frac{1}{4} \lambda \phi_{1h}^4 = \lambda \left( -\frac{1}{2} v_{\uparrow}^2 \phi_{1h}^2 + \frac{1}{4} \phi_{1h}^4 \right) = -\frac{1}{4} m_h^2 c^4 \phi_{1h}^2 + \frac{1}{8} \frac{m_h^2 c^4}{v_{\uparrow}^2} \phi_{1h}^4 \\ = & m_h^2 c^4 \left( -\frac{1}{4} \phi_{1h}^2 + \frac{1}{8} \frac{1}{v_{\uparrow}^2} \phi_{1h}^4 \right) \end{aligned} \quad (16.4)$$

$$V' = \frac{\partial V}{\partial \phi_{1h}} = \phi_{1h} (\mu^2 + \lambda \phi_{1h}^2) = \lambda \phi_{1h} (\phi_{1h}^2 - v_{\uparrow}^2) = \frac{m_h^2 c^4}{2v_{\uparrow}^2} \phi_{1h} (\phi_{1h}^2 - v_{\uparrow}^2)$$

If we calculate the second derivative then ascertain its value at each of the stationary points  $\phi_{1h} = 0$  and  $\phi_{1h} = v_{\uparrow}$ , we obtain  $V''(\phi_{1h} = 0) = -m_h^2 c^4 / 2$  and  $V''(\phi_{1h} = v_{\uparrow}) = +m_h^2 c^4$ , from which we

discern that the potential has a maximum at  $\phi_h = 0$  and a minimum at  $\phi_h = v_\uparrow$ . This  $V$  is, of course, the customary “Mexican hat” potential of the standard model Higgs sector. Now let’s move on from this review, and turn to how to incorporate the second vev at  $v_\downarrow$ .

Empirically,  $v_\uparrow = v = 246.2196508 \pm 0.0000633$  GeV is obtained from the Fermi coupling constant  $G_F$ . We calculated  $v_\downarrow = 6.0491_{-0.0430}^{+0.0571}$  GeV at (15.11) from the sum  $\frac{1}{\sqrt{2}} v_\downarrow = m_d c^2 + m_s c^2 + m_b c^2 = 4.2773_{-0.0304}^{+0.0404}$  GeV of the isospin-down quarks, with two extra decimal places shown to display the impact of the strange and down masses which known more precisely than the bottom mass, see (15.2), then (15.11) and the discussion following. 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  $m_h c^2 = 125.18 \pm 0.16$  GeV, see PDG’s [50]. It is noteworthy, and will momentarily become important, that  $m_h c^2$  is just a touch larger than half the Fermi vev, and to be precise, that  $m_h c^2 - v_\uparrow / 2 = 2.07 \pm 0.16$  GeV. Also, because we now know the Higgs mass empirically, we may deduce that the undetermined parameter  $\lambda = m_h^2 c^4 / 2v_\uparrow^2 = 0.1292 \pm 0.0003$ . 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  $\phi_h = v_\uparrow$ , using the center values of the data for  $m_h$  and  $v_\uparrow$ , the upper (16.4) yields the range value  $V(\phi_h^2 = v_\uparrow^2) = -\frac{1}{8} m_h^2 c^4 v_\uparrow^2 = -(104.39 \text{ GeV})^4$ .

Now, as noted just above, the Higgs mass in  $m_h c^2 = v_\uparrow / 2 + 2.07 \pm 0.16$  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, exceeding this vev by  $4.14 \pm 0.32$  GeV. Comparing  $\frac{1}{\sqrt{2}} v_\downarrow = m_d c^2 + m_s c^2 + m_b c^2 = \frac{1}{\sqrt{2}} v_\downarrow = 4.28_{-0.03}^{+0.04}$  GeV from (15.11) with the two extra decimal places removed and the error range now set by the bottom quark mass which is least-tightly-known, we see that these two numbers match up *within experimental errors*. This means that within experimental errors, *the Higgs mass is exactly halfway between*  $\frac{1}{\sqrt{2}} v_\downarrow = 4.28_{-0.03}^{+0.04}$  GeV *and*  $v_\uparrow = 246.2196508 \pm 0.0000633$  GeV, with the errors set by the former. Or, put differently, *if we now theoretically define the Higgs mass to be the average of*  $v_\uparrow = v = \sqrt{2}(m_u c^2 + m_c c^2 + m_t c^2)$  *and*  $\frac{1}{\sqrt{2}} v_\downarrow = m_d c^2 + m_s c^2 + m_b c^2$  *using the data from (15.11), we find that this relation*  $m_h c^2 = (v_\uparrow + \frac{1}{\sqrt{2}} v_\downarrow) / 2$  *is true within experimental errors, expressed as:*

$$\boxed{125.18 \pm 0.16 \text{ GeV} = m_h c^2 \equiv \frac{1}{2} \left( v_\uparrow + \frac{1}{\sqrt{2}} v_\downarrow \right) = 125.25 \pm 0.02 \text{ GeV} = \frac{1}{2} \left( \sqrt{2} (m_u c^2 + m_c c^2 + m_t c^2) + m_d c^2 + m_s c^2 + m_b c^2 \right)} \quad (16.5)$$

The question now becomes whether  $m_h c^2 = \frac{1}{2} \left( v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow} \right)$  above really is a relation of genuine physical significance, or is just a coincidence. There are a number of good reasons we shall now review, why this is likely a real relation:

First, if  $V(\phi_h)$  is to have a second (local, shallower) minimum at  $\phi_{1h} = v_{\downarrow}$  to provide a “nest” for isospin-down quarks along with its first (global, deeper) 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. This is not optional: elementary calculus demands that if a function has two minima, it inexorably must have a maximum somewhere between these two minima.

Second, given this required  $v_{\downarrow} < \phi_{1h} < v_{\uparrow}$  domain for the second maximum, it makes particular sense in the present context for the maximum to be reasonably close to the halfway point between  $v_{\downarrow}$  and  $v_{\uparrow}$ .

Third, given the requirement for a maximum in the domain  $v_{\downarrow} < \phi_{1h} < v_{\uparrow}$ , just as  $v_{\uparrow}$  and  $v_{\downarrow}$  are physically meaningful numbers, we expect that the energy of  $\phi_{1h}$  at this second maximum should have some physical meaning, for example, that it should be, or should at least be “based on,” 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 the second maximum is based on the Higgs mass itself.

Fourth, if this maximum is to be close to the halfway point between  $v_{\downarrow}$  and  $v_{\uparrow}$ , and is to be based on the Higgs mass, then  $m_h c^2 = \frac{1}{2} \left( v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow} \right)$  in (16.5) is indeed a good halfway point, because  $v_{\downarrow} \ll v_{\uparrow}$ . So, we infer that the Higgs itself mass may provide one suitable halfway point, whereby the maximum occurs at  $\phi_{1h} = m_h c^2$ , just above halfway. Another suitable halfway point would be at  $\phi_{1h} = v_{\uparrow} - m_h c^2$ , just below halfway. Both of these are clearly “based on” the Higgs mass. We will momentarily assess which of these options makes better physical sense.

Fifth, the Higgs mass itself and the related parameter  $\lambda = m_h^2 c^4 / 2v_{\uparrow}^2$  have long been entirely unexplained as a theoretical matter. Given that we now have good empirical data for the Higgs mass, and that  $m_h c^2 = \frac{1}{2} \left( v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow} \right)$  is confirmed by that data within experimental errors, regarding (16.5) as a new, correct theoretical relation of physical significance would allow us after more than four decades to finally place the value of  $\lambda$  on an entirely theoretical basis, as we shall detail further momentarily.

Sixth, the empirical data on the right i (16.5) has a tighter error bound than the data on the left:  $m_h c^2 = 125.25 \pm 0.02$  GeV is tighter than the presently-known  $m_h c^2 = 125.18 \pm 0.16$  GeV by a factor of almost 10, and raises the center value by .07 GeV. Thus, if we regard (16.5) as true, this contains a *prediction* that when the Higgs mass becomes measured more tightly than at present, it will be fit in the range  $m_h c^2 = 125.25 \pm 0.02$  GeV. So, this sixth reason to regard (16.5) as a true physical relation, is that it can likely be *experimentally tested* in the foreseeable future.

Seventh and finally, following (15.13) we noted using the parameterization  $m_d, m_s, m_b = F(m_d, m_s, \theta_{C21})$  or alternatively  $m_d, m_s, m_b = F(m_d, m_b, \theta_{C21})$ , that we had squeezed one degree of freedom from the isospin-down quark masses via the first relation (15.12) for the CKM mixing angle  $\theta_{C21}$  and the couplings for these masses. With the discovery of (16.5), we now have a basis for expressing the previously-undetermined number  $v_\downarrow$  as a function  $v_\downarrow = F(v_\uparrow, m_h)$ . In other words, given the Higgs mass and the Fermi vev, we may deduce  $v_\downarrow = \sqrt{2}(m_d c^2 + m_s c^2 + m_b c^2)$  from  $v_\uparrow$  and  $m_h$  via (16.5). This means that if we choose to regard the Higgs mass as a “given” number, related to the two vevs by (16.5), we can squeeze yet another unexplained energy number out of the parameters which drive the natural world. This would enable us to remove  $m_s$  or  $m_b$  from the above parameterizations and now write  $m_d, m_s, m_b = F(m_d, m_h, \theta_{C21})$  for the isospin-down quark masses. Together with  $m_u, m_c, m_t = F(v, \theta_{C31}, \theta_{C23})$ , this would mean that *we can now eliminate five (5) out of the six unexplained quark masses* and “explain” them so far as they relate to  $\theta_{C21}$ ,  $\theta_{C23}$ ,  $\theta_{C31}$ ,  $v$ , and  $m_h$ , leaving only  $m_d$  now unexplained. Of course, this would not “explain” why the five parameters  $\theta_{C21}$ ,  $\theta_{C23}$ ,  $\theta_{C31}$ ,  $v$ , and  $m_h$  have the empirical values that they have. But this would explain how these are related to the quark masses and so render five of these six quark mass numbers into the status of “redundant” data.

Accordingly, for all the reasons just reviewed, we shall now regard  $m_h c^2 = \frac{1}{2}(v_\uparrow + \frac{1}{\sqrt{2}}v_\downarrow)$  in (16.5) to be a true theoretical physical relation for the Higgs mass, and will use the tighter, raised-center value  $m_h c^2 = 125.25 \pm 0.02$  GeV for this mass from here on.

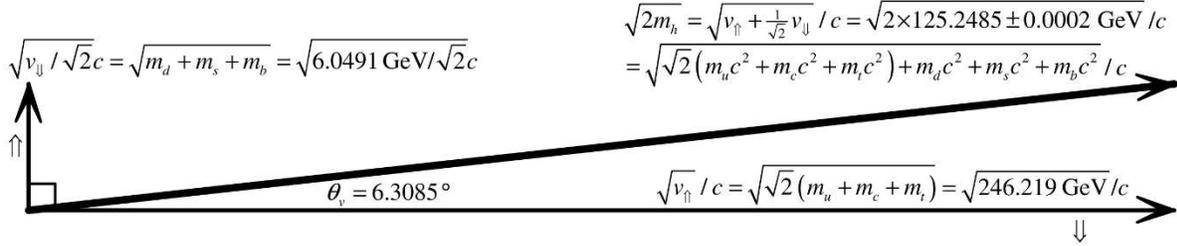
Next, turning to the square roots of masses, if we write (16.5) as:

$$\left(\sqrt{v_\uparrow} / c\right)^2 + \left(\sqrt{v_\downarrow} / \sqrt[4]{2}c\right)^2 = \left(\sqrt{2m_h}\right)^2, \quad (16.6)$$

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:

$$\sin \theta_v \equiv \frac{\sqrt{v_\downarrow} / \sqrt[4]{2}c}{\sqrt{2m_h}}; \quad \cos \theta_v = \frac{\sqrt{v_\uparrow} / c}{\sqrt{2m_h}}; \quad \tan \theta_v = \frac{\sqrt{v_\downarrow}}{\sqrt[4]{2}\sqrt{v_\uparrow}}, \quad (16.7)$$

such that  $\theta_v$  effectively measures the magnitude of each of the two vevs in relation to one another and the Higgs mass. Using the data from (15.11) and (16.5) we calculate that the central value for this angle is  $\theta_v = 6.3085^\circ$ . This can all be represented in the rather simple geometric Figure below:



**Figure 4: Vacuum and Higgs Mass Mixing in Quark Rest Mass Space**

Then, as noted in the fifth reason reviewed above, by advancing (16.5) to a meaningful relation, we also can deduce that the long-undetermined parameter  $\lambda$  in  $V = \frac{1}{2} \mu^2 \phi_{1h}^2 + \frac{1}{4} \lambda \phi_{1h}^4 + \dots$  is theoretically given, also using (16.7), by:

$$\lambda = \frac{m_h^2 c^4}{2v_{\uparrow}^2} = \frac{(v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow})^2}{8v_{\uparrow}^2} = \frac{1}{8} \left( 1 + \frac{1}{\sqrt{2}} \frac{v_{\downarrow}}{v_{\uparrow}} \right)^2 = \frac{1}{8} (1 + \tan^2 \theta_v)^2 = 0.1292 \pm 0.0003. \quad (16.8)$$

So physically, using Figure 4 and (16.8),  $\lambda$  and  $\theta_v = \tan^{-1} \left( \sqrt{v_{\downarrow}} / \sqrt{v_{\uparrow}} \sqrt[4]{2} \right)$  are now understood as measures of the ratio  $v_{\downarrow} / v_{\uparrow}$  of the two vevs. In the limiting case where  $v_{\downarrow} / v_{\uparrow} \rightarrow 0$  and  $\theta_v \rightarrow 0$ , we also have  $\lambda \rightarrow 1/8$ .

It is important to understand how Figures 2, 3 and 4 all tie together, wherein the Higgs rest energy is distributed into the two quark vevs in accordance with Figure 4, with these two vevs then parceling out their energies into the rest energies for each quark in their sector as illustrated in Figures 2 and 3. Specifically: In Figure 2,  $(\sqrt{v_{\uparrow}} / c) / \sqrt[4]{2}$  was the hypotenuse projected into each of three isospin-up mass roots and in Figure 3,  $(\sqrt{v_{\downarrow}} / c) / \sqrt[4]{2}$  was the hypotenuse projected into each of three isospin-up mass roots. This means that (16.6) and Figure 4 are the bridge between the two spaces in Figures 2 and 3, in a square root mass space that is overall *six dimensional*, as also seen from the bottom line of (16.5). So, with the coefficients and square roots as shown, geometrically, one starts with the Higgs mass  $m_h$  which is placed along the hypotenuse in Figure 4. This Higgs mass hypotenuse is then projected onto the two orthogonal axes, represented with  $\uparrow$  and  $\downarrow$  for the isospin-up and isospin-down, to arrive at the related 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 2, and in the other three of six dimensions  $v_{\downarrow}$  is projected into the bottom, strange and down masses as shown in Figure 3. The azimuthal and polar angles in the former, and the azimuthal angle in the latter, simultaneously are the three real CKM mixing angles.

Now we come to the point touched upon in the second through fourth reasons why (16.5) should be regarded as a true physical relation. From these, we discerned that  $\phi_{1h} = m_h c^2$  and alternatively  $\phi_{1h} = v_{\uparrow} - m_h c^2$  can suitably serve as the domain points for the Lagrangian potential maximum, because each is approximately halfway between  $v_{\downarrow}$  and  $v_{\uparrow}$ , and because each is clearly based upon the Higgs mass. In the former,  $\phi_{1h} = m_h c^2$  is at the energy equivalent of the Higgs mass. Because  $\phi_{1h} = v_{\uparrow} + h$ , the latter implies  $h = -m_h c^2$ , so that the energy of the Higgs field at the maximum domain point is equal to the negative energy equivalent of Higgs mass. Note from (13.5) and Figure 1, that the Higgs field energy is always negative close to a fermion. Now the question is: which of these two alternatives makes the best physical sense?

Using  $m_h c^2 = \frac{1}{2} \left( v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow} \right) = 125.25 \pm 0.02$  GeV from (16.5), and mindful of  $\phi_{1h} = v_{\uparrow} + h$ , for the former alternative there would be a Lagrangian potential maximum at the domain point:

$$\begin{aligned} \phi_{1h}(x^M) &\equiv m_h c^2 = \frac{1}{2} \left( v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow} \right) = 125.25 \pm 0.02 \text{ GeV}; \\ \text{i.e. } h(x^M) &= \phi_{1h}(x^M) - v_{\uparrow} = m_h c^2 - v_{\uparrow} = \frac{1}{2} \left( \frac{1}{\sqrt{2}} v_{\downarrow} - v_{\uparrow} \right) = -120.97 \pm .02 \text{ GeV} \end{aligned} \quad (16.9a)$$

But again, while the calculus 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. So, given the sensibility of this location being based on the Higgs boson mass, we also consider the alternative where the maximum is at:

$$\begin{aligned} h(x^M) &= -m_h c^2 = -\frac{1}{2} \left( v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow} \right) = -125.25 \pm 0.02 \text{ GeV}; \text{ i.e.} \\ \phi_{1h}(x^M) &= v_{\uparrow} + h(x^M) = v_{\uparrow} - m_h c^2 = \frac{1}{2} \left( v_{\uparrow} - \frac{1}{\sqrt{2}} v_{\downarrow} \right) = 120.97 \pm .02 \text{ GeV} \end{aligned} \quad (16.9b)$$

In effect, (16.9b) this shifts the maximum hypothesized in (16.9a) to the left, toward the isospin-down vev, by  $\frac{1}{2} \left( \frac{1}{\sqrt{2}} v_{\downarrow} - v_{\uparrow} \right) - \frac{1}{2} \left( v_{\uparrow} - \frac{1}{\sqrt{2}} v_{\downarrow} \right) = \frac{1}{\sqrt{2}} v_{\downarrow} = m_d c^2 + m_s c^2 + m_b c^2 = 4.28_{-0.03}^{+0.04}$  GeV, which is the sum of the charged lepton masses related to (15.11) with the error bar set by the bottom quark. Now, we are called upon to determine which of (16.9a) versus (16.9b) is the better hypothesis, and this is a physics question, not a mathematics question.

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  $h$  draws its rest energy out of the vacuum in a similar way. So, considered from the viewpoint of Figure 1, we see that the field  $h(x^M)$  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  $\phi_{1h}(x^M)$  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  $h(x^M) = -m_h c^2$ , 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

$\phi_{1h}(x^M) = m_h c^2$ , 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 domain point of the  $V$  maximum should be based upon energy removed from the vacuum, versus upon energy retained by the vacuum.*

Now we turn to (15.11) 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 “nests” for isospin-down quarks, that  $v_{\uparrow}$  will establish energetically-favored nests for isospin-up quarks. As we shall shortly examine in detail, we also anticipate 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 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 short-lived ( $\sim 10^{-22}$  s) Higgs boson is formed with a rest mass  $m_h c^2 \cong 125.25$  GeV, this is an amount of energy which is, briefly, no longer available for other purposes, such as for fermions to draw energy out of the vacuum to facilitate their beta decay over or through the  $V$  maximum. The top-line integral in (13.7) is illustrative of how this works for fermion masses. As such, this raises the barrier for beta decay, and can be analogized to having to climb up to the roof of a ten-story building by starting in the basement rather than on the fifth 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  $4.28$  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, recognizing that the Higgs field must give up an energy  $-m_h c^2 \cong -125.25$  GeV to provide a mass  $m_h c^2 \cong 125.25$  GeV to a Higgs boson in accordance with the energy conservation principles illustrated by (13.7), we now formally make the hypothesis as between (16.9a) and (16.9b), that *the maximum of the Lagrangian potential  $V$  between  $v_{\downarrow}$  and  $v_{\uparrow}$  is situated at the domain point where  $h(x^M) = -m_h c^2$* , that is, as the point where the Higgs field energy is equal to minus the Higgs mass, and has a negative sign because this represents energy which has been drawn out of the Fermi vacuum to briefly provide mass to a Higgs boson.

Now, we finally have all ingredients needed to revise the potential in (16.4) 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 (16.9b) 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:

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

This is constructed so that the leading terms  $(m_h^2 c^4 / 2v_{\uparrow}^2) \phi_{1h} (\phi_{1h}^2 - v_{\uparrow}^2)$  in the top line above precisely match the usual  $V'$  in (16.4). 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 (16.4) will continue to be  $-\frac{1}{4} m_h^2 c^4 \phi_{1h}^2$ , with all changes to  $V$  be introduced at higher order. This leading term we are matching stems from the “mass” term  $\frac{1}{2} \mu^2 \phi_{1h}^2$  in  $V = \frac{1}{2} \mu^2 \phi_{1h}^2 + \frac{1}{4} \lambda \phi_{1h}^4$ . It will be seen by inspection that the top line in the above will become zero at all four of  $\phi_{1h} = 0$ ,  $\phi_{1h} = m_h c^2$ ,  $\phi_{1h} = v_{\downarrow}$  and  $\phi_{1h} = v_{\uparrow} - m_h c^2$ . We should choose the overall sign in  $A$  so that the first two should provide maxima and the latter two should provide minima for  $V$  itself.

Next, we easily integrate the above. For the leading term to match  $-\frac{1}{4} m_h^2 c^4 \phi_{1h}^2$  in  $V$  from (16.4) we must set  $A = 1 / v_{\downarrow}^2 (v_{\uparrow} - m_h c^2)^2$ . Also based on the “initial condition” of matching (16.4), we discard any integration constant. We then consolidate and reduce to obtain:

$$V(\phi_{1h}) = m_h^2 c^4 \left( \begin{aligned} &-\frac{1}{4} \phi_{1h}^2 + \frac{1}{8} \frac{1}{v_{\uparrow}^2} \phi_{1h}^4 + \frac{1}{8} \left( \frac{1}{v_{\downarrow}^2} + \frac{1}{(v_{\uparrow} - m_h c^2)^2} \right) \phi_{1h}^4 \\ &-\frac{1}{12} \left( \frac{1}{v_{\uparrow}^2 v_{\downarrow}^2} + \frac{1}{(v_{\uparrow} - m_h c^2)^2} \frac{v_{\uparrow}^2 + v_{\downarrow}^2}{v_{\uparrow}^2 v_{\downarrow}^2} \right) \phi_{1h}^6 + \frac{1}{16} \frac{1}{(v_{\uparrow} - m_h c^2)^2} \frac{1}{v_{\uparrow}^2 v_{\downarrow}^2} \phi_{1h}^8 \end{aligned} \right). \quad (16.11)$$

Comparing with  $V$  in (16.4), we indeed see the original  $\phi_{1h}^2$  and  $\phi_{1h}^4$  terms. But there are some new additions to the  $\phi_{1h}^4$  term, and brand 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  $1/(v_{\uparrow} - m_h c^2)^2$  coefficient, and to then explicitly apply  $m_h c^2 = (v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow})/2$  from (16.5), thus  $v_{\uparrow} - m_h c^2 = \frac{1}{2}(v_{\uparrow} - \frac{1}{\sqrt{2}} v_{\downarrow})$ , as follows:

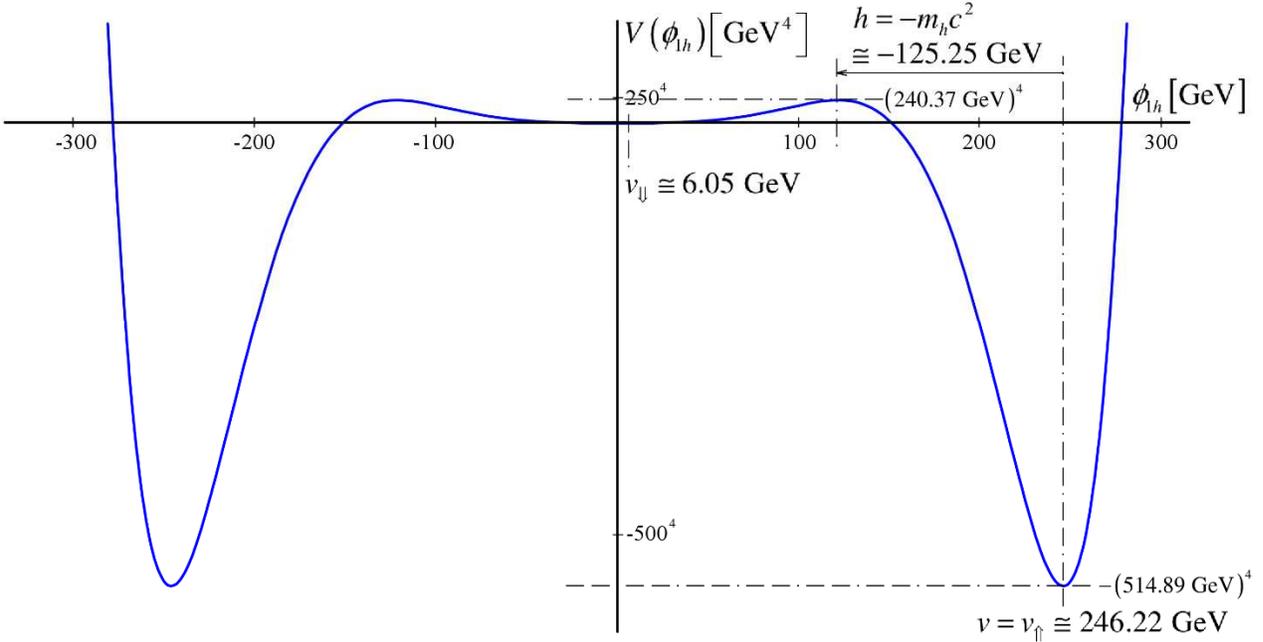
$$\begin{aligned}
 V(\phi_{1h}) &= m_h^2 c^4 \left( -\frac{1}{4} \phi_{1h}^2 + \frac{1}{8} \frac{v_{\uparrow}^2 + v_{\downarrow}^2}{v_{\uparrow}^2 v_{\downarrow}^2} \phi_{1h}^4 - \frac{1}{12} \frac{1}{v_{\uparrow}^2 v_{\downarrow}^2} \phi_{1h}^6 \right) \\
 &+ \frac{m_h^2 c^4}{(v_{\uparrow} - m_h c^2)^2}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

So, the behavior of  $V(\phi_{1h})$  is entirely driven by the two energy-dimensional numbers in (15.11). 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  $v_{\uparrow} = \sqrt{2}(m_u c^2 + m_c c^2 + m_t c^2)$ . 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  $v_{\downarrow} = \sqrt{2}(m_d c^2 + m_s c^2 + m_b c^2)$ . Additionally, the Higgs mass itself establishes a second maximum via (16.9b), but the new relation  $m_h c^2 = (v_{\uparrow} + \frac{1}{\sqrt{2}} v_{\downarrow})/2$  discovered in (16.5) means that only two of these energy numbers are truly independent of one another.

It is pedagogically-useful to graph the potential  $V(\phi_{1h})$  in (16.12) using the numerical values of  $v_{\uparrow}$  and  $v_{\downarrow}$  in (15.11), and / or the Higgs mass in (16.5). Substituting these into (16.12), reconsolidating terms at each order, and rounding the coefficient at each order to four significant digits, with  $\phi_{1h}$  expressed in GeV thus  $V(\phi_{1h})$  in  $\text{GeV}^4$ , we obtain:

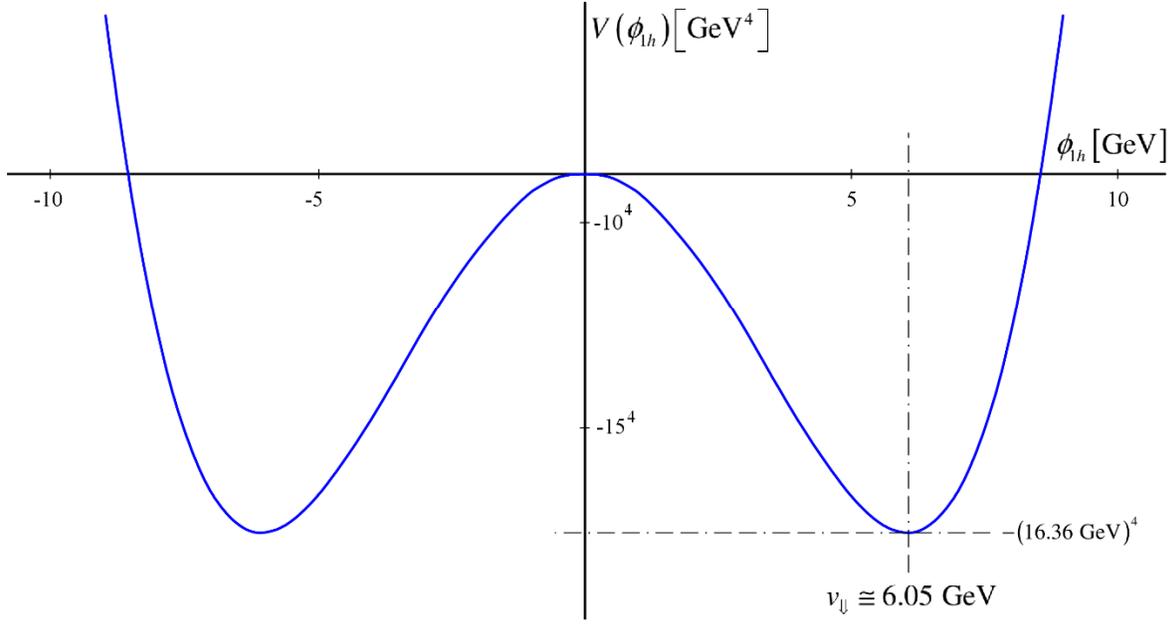
$$V(\phi_{1h}) [\text{GeV}^4] = -3922 \phi_{1h}^2 + 53.76 \phi_{1h}^4 - 0.003032 \phi_{1h}^6 + 3.020 \times 10^8 \phi_{1h}^8. \tag{16.13}$$

Keeping in mind from (16.1) through (16.4) that  $V(\phi_{1h})$  is part of the Lagrangian density and so has physical dimensions of quartic energy, and that  $\phi_{1h}$  is linear in energy, (16.13) can be easily graphed to produce the following plot:



**Figure 5: Lagrangian Potential for Quarks – Wide View**

Above we see the usual minimum at  $\phi_h = v_{\uparrow} \cong 246.22 \text{ GeV}$ , where along the y axis we have an energy “well” with a depth of  $V(\phi_h) \cong -(514.89 \text{ GeV})^4$ . But we now have a new maximum at  $\phi_h = v_{\uparrow} - m_h c^2 \cong 120.97 \text{ GeV}$  based on (16.9b), and at this maximum there is an energy “barrier” with a height of  $V(\phi_h) \cong (240.37 \text{ GeV})^4$ . Closer to the origin is the usual maximum at  $\phi_h = 0$  and the new minimum at  $\phi_h = v_{\downarrow} \cong 6.05 \text{ GeV}$ . But comparatively to the foregoing, these are extremely small, and are impossible to see in Figure 5. So, it is also useful to also magnify the domain from  $-10 \text{ GeV} < \phi_h < 10 \text{ GeV}$  in Figure 5, while also magnifying the range, to obtain the magnified view of the center of Figure 5, as shown below:

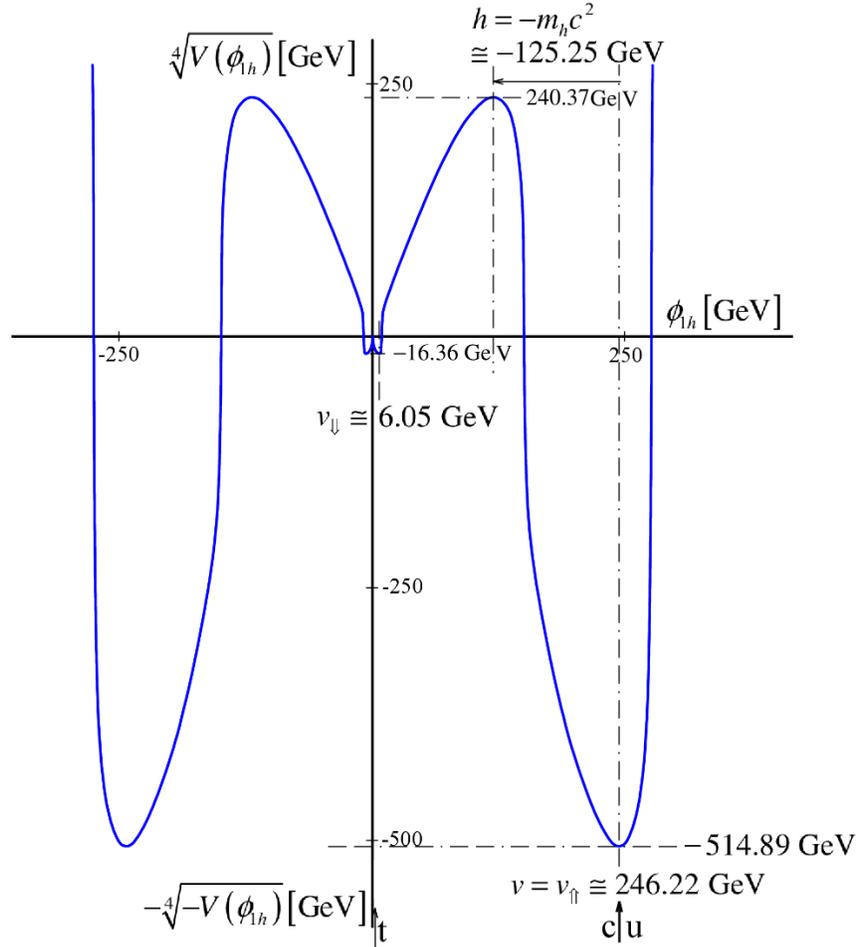


**Figure 6: Lagrangian Potential for Quarks – Magnified Center View**

Here, the usual maximum at  $V(\phi_h = 0) = 0$  is readily apparent, as is the new minimum at  $\phi_h = v_{\downarrow} \cong 6.05 \text{ GeV}$  where there is a second energy well of depth  $V(v_{\downarrow}) \cong -(16.36 \text{ GeV})^4$ . The above is simply an extremely magnified view of the region in Figure 5 close to the origin.

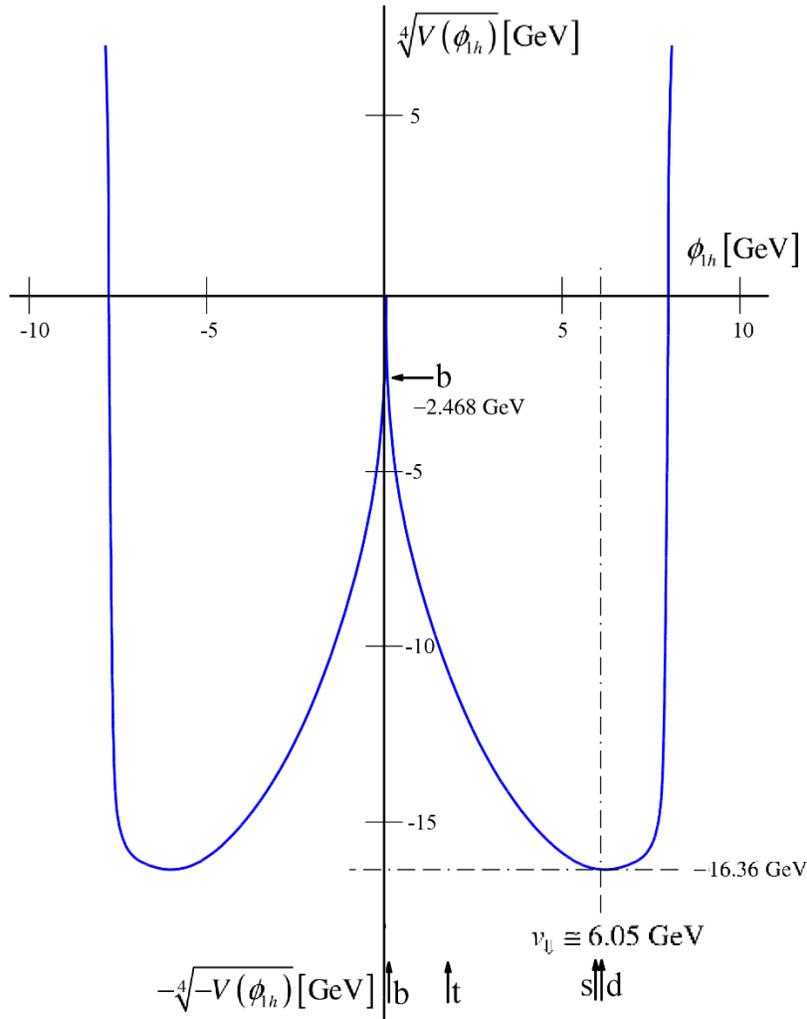
As reviewed shortly before (16.5), the well depth for the usual  $V = \frac{1}{2}\mu^2\phi_h^2 + \frac{1}{4}\lambda\phi_h^4$  in (16.4) at  $\phi_h = v_{\uparrow}$  was  $V(v_{\uparrow}^2) = -(104.39 \text{ GeV})^4$ . But of course, this was based only on square and quartic field terms. Now, in Figure 5, at the same  $\phi_h = v_{\uparrow}$  we have  $V(v_{\uparrow}^2) \cong -(514.89 \text{ GeV})^4$  which is deeper by a factor of almost 5 in linear energy dimensions. This substantially-increased depth is driven by the combination of setting  $A = 1/v_{\downarrow}^2(v_{\uparrow} - m_h c^2)^2$  going from (16.10) to (16.11) to preserve the leading  $-\frac{1}{4}m_h^2 c^4 \phi_h^2$  mass term in  $V = -\frac{1}{4}m_h^2 c^4 \phi_h^2 + \frac{1}{8}(m_h^2 c^4 / v_{\uparrow}^2)\phi_h^4$  from (16.4) without change, and from the new minimum at  $v_{\downarrow}$  and new maximum at  $\phi_h = v_{\uparrow} - m_h c^2$ . That is, this increased depth is driven entirely by the new higher-order  $\phi_h^4$ ,  $\phi_h^6$  and  $\phi_h^8$  terms, in combination with maintaining the existing  $\phi_h^2$  and  $\phi_h^4$  terms as is. Given the substantially-greater depth of  $V(v_{\uparrow}^2) \cong -(514.89 \text{ GeV})^4$  versus  $V(v_{\downarrow}) \cong -(16.36 \text{ GeV})^4$ , as was previously anticipated following (16.5), we see that the minimum at  $\phi_h = v_{\downarrow}$  is simply a *local* minimum, while that at  $\phi_h = v_{\uparrow}$  is a *global* minimum. Given that quarks situated in these wells will seek out the lowest available energy states, this means that will all else being equal, a quark will find it energetically-favorable to maintain an isospin-up state over and isospin-down state. As we shall later see, this is part of why free neutrons decay into free protons, rather than vice versa.

Even with Figures 5 and 6, however, the energetic behavior of quarks in these wells and the impact of the new maximum are not brought out as much as they could be, because  $\phi_{1h}$  is linear in energy while  $V$  is quartic in energy. So, it is also useful to reproduce Figures 5 and 6 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 5 and scaling what are now linear energy numbers along each axis to one another, we obtain the plot below:



**Figure 7: Lagrangian Potential for Quarks, Fourth Root – Wide View**

Note also the upward-pointing arrows designating the u, c and t quarks, which will momentarily be reviewed. 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 8 below, we also magnify Figure 7 over the domain  $-10 \text{ GeV} < \phi_{1h} < 10 \text{ GeV}$ , and again scale the energies on a 1:1 basis along the vertical and horizontal axes. This Figure 8 is equivalent to the fourth root of the magnified view of the potential in Figure 6.



**Figure 8: Lagrangian Potential for Quarks, Fourth Root – Magnified Center View**

Here, note also the upward-pointing arrows designating the b, s and d quarks as well as the t quark, which will also be reviewed momentarily.

These two plots in Figures 7 and 8 help provide a deeper understanding of how quarks behave in the Lagrangian potential. 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  $V(\phi_h = 0) = 0$  is not smooth as one might conclude looking at Figures 5 and 6. Rather, when comparing energies to energies at a 1:1 scale, this maximum comes to a sharp upward point with a slope that is infinite at the origin. Second, it is apparent, most clearly from Figure 7, that the  $v_\downarrow$  potential well establishes a *local* minimum while the  $v_\uparrow$  potential well presents a *global* minimum, as already noted. 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 most clearly from Figure 7 that there is high barrier between the two wells set by the new maximum

which has a height of +240.37 GeV. Using  $m_h c^2 = 125.25 \pm 0.02$  GeV as refined in (16.5), 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. First, however, let's review how Figures 1 through 8 all tie together, and explain how the upward-pointing arrows in Figure 7 and 8 lay the foundation for understanding the energetic behaviors of the individual quarks.

To fully understand Figures 5 through 8 and what they teach about the behavior of individual quarks, *it is essential to understand that  $\phi_{1h}$  plotted along the horizontal axis in all of Figure 5 through 8 is the very same  $\phi_{1h}$  which is plotted along the vertical axis of Figure 1.* It is most helpful to refer to Figure 7 to see this: Although Figure 1 applies to the up quark, we discussed following (15.17) how similar plots with the exact same character can be drawn for all of the other quarks as well. Far from the charm and up quarks, where  $V_{(5)}/L^5 \rightarrow \infty$ , this Figure 1-type plot will likewise level off at  $\phi_{1h} = v = v_{\uparrow} \cong 246.22$  GeV. But right at  $V_{(5)}/L^5 = 0$  which defines the energetic minimum at which a quark is most likely to nest, there will be far less energy taken out of the vacuum. This is because at  $V_{(5)}/L^5 = 0$ , based on (14.1), while the dimensionless coupling for the top quark  $G_t = 0.9937 \pm 0.0023$  removes about 99.4% of the energy from the Fermi vacuum with a global vev minimum  $v = v_{\uparrow}$  shown on the horizontal axis of Figure 7, in contrast, for the charm quark  $G_c = 0.00732^{+0.00014}_{-0.00020}$  removes only about 0.7% from the vacuum and for the up quark  $G_u = 0.000013^{+0.000003}_{-0.000002}$  removes a scant 0.001% from the vacuum.

For the isospin down quarks the vev itself is different, namely  $v_{\downarrow} \cong 6.05$  GeV. But this is also plotted on the horizontal axis of Figure 7, albeit far closer to the origin and much shallower than  $v_{\uparrow}$ . Now referring to (15.7), for  $V_{(5)}/L^5 \rightarrow \infty$  the Figure 1 analogs for each of the b, s, d quarks will level off at  $\phi_{1h} = v_{\downarrow} \cong 6.05$  GeV, because of the second vev local minimum. But right at  $V_{(5)}/L^5 = 0$  we have  $G_b = 0.97725^{+0.00016}_{-0.00021}$  which shows that the bottom quark takes an approximate 97.7% energy bite out of this diminished-magnitude vacuum, while  $G_s = 0.02161^{+0.00015}_{-0.00020}$  indicates a 2.1% bite from the strange quark and  $G_d = 0.00115 \pm 0.00001$  a 0.1% bite for the down quark. For both isospin-up and isospin-down quarks, we have the relations  $G_u + G_c + G_t = 1$  from (14.3) and  $G_b + G_s + G_d = 1$  from (15.1), which establish the hypotenuses in Figures 2 and 3, respectively. These two hypotenuses are then tied together by the Higgs mass in Figure 4 which also via (16.9b) sets the peak between the isospin-up and isospin-down vevs, which peak is in all of Figures 5 through 8, albeit most visually-clearly in Figure 7.

So with this understanding in place, we see that the upward-pointing arrows in Figure 7 and 8 for the individual quarks simply represent, for each quark, the value of  $\phi_{1h}$  at which their Figure 1-type plots bottom out at  $V_{(5)}/L^5 = 0$  where the quark is most likely to nest. But of course, in Figures 7 and 8  $\phi_{1h}$  is plotted on the horizontal axis, not the vertical axis as in Figure 1. So, in

Figures 7 and 8 (and 5 and 6) the removal of energy from the vacuum and its reappearance in the rest energy of an individual quark via the energy conservation relation (13.7), is represented by right-to-left displacement along the  $+\phi_{1h}$  horizontal axis, rather than by the downward displacement of a Figure 1-type plot the along  $+\phi_{1h}$  vertical axis. So in Figure 7 it is the small  $G_c = 0.00732^{+0.00014}_{-0.00020}$  for the charm quark and the even-smaller  $G_u = 0.000013^{+0.000003}_{-0.000002}$  for the up quark that places their nest-center arrows very-slightly to the left of the global minimum at  $\phi_{1h} = v = v_{\uparrow} \cong 246.22$  GeV . But the very large  $G_t = 0.9937 \pm 0.0023$  causes the top quark not only to nest far to the left of  $v_{\uparrow}$ , but even to the left of  $v_{\downarrow} \cong 6.05$  GeV , butting right up against the  $y$  axis. Then, in Figure 8 we have a magnified view of the  $v_{\downarrow}$  region. Here, we see the small  $G_s = 0.02161^{+0.00015}_{-0.00020}$  for the strange quark and the even smaller  $G_d = 0.00115 \pm 0.00001$  for the down quark causing these nest-center arrows to situate very-slightly to the left of the local minimum at  $\phi_{1h} = v_{\downarrow} \cong 6.05$  GeV . Bu the very large  $G_b = 0.97725^{+0.00016}_{-0.00021}$  moves the bottom quark nest-center arrow far to the left of  $v_{\downarrow}$ , and indeed, as we can see in this magnified view, also well to the left of the top quark.

Now, let's study what Figures 7 and 8 teach about the energetic behaviors of quarks in the Lagrangian potential, and about the role of the Higgs field not only in the acquisition of rest mass by quarks, but in the weak beta decays of quarks.

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

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 7 and 8 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. In addition for Figures 7 and 8, it is helpful for the ensuing discussion of this to also 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].

Keeping in mind principles of “least action” by which physical processes and particles always seek out the lowest energy states available, there are four general features in Figures 7 and 8 that should be noticed from the outset: First, as already noted, the Figure 1 nest-center of the top quark is perturbed so far to the left in Figure 7 that the top quark nest on the left side of the  $v_{\downarrow}$  energy well. *This crossover of the nest-center from one vev minimum to the other is unique to the top quark.* Second, the global minimum at  $\phi_{1h} = v_{\uparrow}$  presents a much deeper energy well than the local minimum at  $\phi_{1h} = v_{\downarrow}$ , which means that with all else being equal, up and charm quarks will be energetically favored over down, strange, bottom or (crossed-over) top quarks. This is one of the reasons why free neutrons decay into free protons with an extra up quark versus down quark, and not vice versa. Third, with the up quark slightly to the right of the charm quark and thus deeper

in the Lagrangian potential energy well at the  $v_{\uparrow}$  global minimum, and with the down quark likewise slightly to the right of the strange quark thus deeper in the potential energy well at the  $v_{\uparrow}$  local minimum, we also expect that with all else equal, the up flavor will be favored over the charm flavor, and the down flavor over the strange flavor. Moreover, the top quark and the bottom quark are both nest-centered well above the local  $v_{\downarrow}$  minimum, which makes these third-generation quarks the least-energetically-favored states. This is the general reason why first-generation baryons are energetically-favored over second-generation baryons, and second-generation over third-generation baryons. Moreover, the bottom quark, which is even further-left than the top quark, has the shallowest descent into a Lagrangian potential well, and so is the least-energetically favored of all the quarks. Fourth, via (16.9), the Higgs field and the Higgs boson mass clearly establish a steep energy peak, i.e., barrier between the  $v_{\uparrow}$  and the  $v_{\downarrow}$  Lagrangian potential wells at  $h(x^M) = -m_h c^2$ , and thus, between the up and charm quarks for the former, and the down, strange, bottom and (crossed-over) top quark for the latter. This means that any weak beta decay between a u or c quark and a d, s or b quark must either “jump over” or “tunnel through” this barrier. But also, because of the crossover by which the top quark is a “visitor” from the  $v_{\uparrow}$  well because of its exceptionally large mass, *this barrier does not need to be cleared* for weak beta decays between the top quark and any of the bottom, strange or down quarks. Now let’s turn to beta decay.

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” or “tunnel” barrier peak at  $h(x^M) = -m_h c^2$ . 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  $V_{ud} = 0.97446 \pm 0.00010$ . This is in fact clearly true, 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  $\sqrt[4]{V}(\phi_{1h} = v_{\uparrow} - m_h c^2) \cong 240.37 \text{ GeV}$ . Given that fermions acquire their masses from the Higgs field drawing energy out of the vacuum in accordance with the upper equation (13.7), it seems that the energy to jump this barrier at  $\phi_{1h} = v_{\uparrow} - m_h c^2$  a.k.a.  $h(x^M) = -m_h c^2$  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 7 and 8 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 7, 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  $m_t c^2 \cong 125.25$  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 7. 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 7. 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 8. Additionally, the bottom quark itself has a mass of  $m_b c^2 = 4.18_{-0.03}^{+0.04}$  GeV [43] 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  $u \rightarrow d, s, b$  or a  $c \rightarrow d, s, b$  decay from the  $v_{\uparrow} \rightarrow v_{\downarrow}$  well, it takes energy from three Higgs bosons to facilitate a  $d \rightarrow u, c$  or a  $s \rightarrow u, c$  decay from  $v_{\downarrow} \rightarrow v_{\uparrow}$ , and it takes energy from two Higgs bosons to facilitate a  $b \rightarrow u, c$  decay from  $v_{\downarrow} \rightarrow v_{\uparrow}$ . 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  $b \rightarrow t$  transition such as in [51] 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 helps explain why  $V_{tb} = 0.999105 \pm 0.000032$  is closest to 1 of all the CKM 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 quark 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 (if needed or a mass increase) 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 7. So, it is both easier to get from  $v_{\downarrow} \rightarrow v_{\uparrow}$  than the other way around, *and it is easier to stay in  $v_{\uparrow}$  after a  $v_{\downarrow} \rightarrow v_{\uparrow}$  decay has occurred.* This suggests that isospin-up quarks are *more energetically stable* than isospin-down quarks, again, with all else equal. Given that individual quarks (or, at least, quarks in the first and second generations) are always confined in hadrons, and that baryons contain three quarks, this is 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. This will be reviewed further in section 23, where we shall study

leptonic weak beta decay following the study of lepton masses, neutrino oscillation PMNS mixing angles, and the lepton Lagrangian potential.

Also, *ab initio*, the Higgs field  $h$  itself represents *quantum fluctuations* in the Fermi vacuum in which the scalar field  $\phi_h$  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  $\nu_{\downarrow} \rightarrow \nu_{\uparrow}$  decays, it may be possible to detect a smaller energy withdrawal followed by larger redeposit. For  $\nu_{\uparrow} \rightarrow \nu_{\downarrow}$  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-to-1  $V_{tb} = 0.999105 \pm 0.000032$ ,  $b \rightarrow t$  is simply a withdrawal and  $t \rightarrow b$  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  $\nu_{\downarrow}$  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, erect extremely high-energy barriers stemming from the large ratio of the charged-lepton-to-neutrino masses. The other aspect of beta decays which have been entirely ignored in this discussion because we have not yet developed the supporting evidence for this, is the central role that the low-energy, non-relativistic neutrinos in the cosmological neutrino background (CvB) play in triggering beta decays. All we shall preview for now, is that the experimental data regarding the CvB background shows that one neutrino flows through a one-barn cross section of space approximately every fifteen minutes, which is the same as the mean lifetime of a free neutron with an approximate .01 barn cross section; and that the weak  $Z$  interactions between neutrons and neutrinos enable neutrons to attract these CvB neutrinos to trigger beta decays. This will be reviewed in detail in section 23, after we have learned a great deal more about neutrinos in general.

Before concluding, let's take stock of all the reparameterizations we have found to this point. Prior to (16.6) 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 [33], [31]. 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  $m_d = F(m_e)$ , that is, to reparameterize the down mass as a function of the electron mass, given that we have also reparameterized the up quark mass. Therefore, we effectively used

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

to reparameterize *all six quark masses*. Of course, one of these parameters,  $m_e$ , is itself a fermion rest mass. But, it is the mass of a lepton not a quark. So, by (17.1), we effectively “kick down the road” to our study of the charged leptons, the completion of quark mass reparameterization. With this, following a review in the next section of the CKM quark mixing matrix based on all of the foregoing, we will be ready to begin our study of the lepton rest masses.

## 18. The CKM Quark Mixing Matrix Mass Parameterization, and the Fine-Tuning of Quark Masses, Mixing Angles and CKM Matrix Components

The unitary 3x3 CKM quark-mixing matrix  $V_{CKM}$  can be parametrized in several different ways, but has invariant magnitudes for each of its nine components irrespective of the parameterization choice. As reviewed at [52], these include the original parameterization of Kobayashi and Maskawa, as well as the “standard” and the Wolfenstein parameterizations. The latter two of these are reviewed in detail along with the latest empirical matrix data in PDG's 2018 survey [46]. However, with the advent of the relations (15.12) directly between the quark couplings  $G_q$  and the three real CKM mixing angles, it is possible to develop a fourth type of CKM parametrization which we shall refer to as the “mass parametrization.” As will be reviewed in this section, this can be used to express the CKM matrix directly in terms of the quark masses, and to further tighten the numeric data for the quark masses, CKM mixing angles, and CKM matrix components themselves.

We begin by recalling that  $m_{\uparrow}c^2 = \frac{1}{\sqrt{2}}v_{\uparrow}G_{\uparrow}$  and  $m_{\downarrow}c^2 = \frac{1}{\sqrt{2}}v_{\downarrow}G_{\downarrow}$  respectively relate the masses of the isospin-up and down quarks to their couplings and to their vev minima as reviewed in Figures 5 through 8. Using these, along with the final expressions in (15.12) and  $\sin^2\theta + \cos^2\theta = 1$ , we can write the cosines and sines of the CKM mixing angles directly in terms of the quark masses by:

$$\begin{aligned}
 \cos^2 \theta_{C12} &= \frac{m_s}{m_d + m_s}; & \sin^2 \theta_{C12} &= \frac{m_d}{m_d + m_s} \\
 \cos^2 \theta_{C23} &= \frac{m_c}{m_u + m_c}; & \sin^2 \theta_{C23} &= \frac{m_u}{m_u + m_c}. \\
 \cos^2 \theta_{C13} &= \frac{m_t}{m_u + m_t}; & \sin^2 \theta_{C13} &= \frac{m_u}{m_u + m_t}
 \end{aligned} \tag{18.1}$$

Now, let's turn to the CKM matrix itself. Given, e.g., in [12.3] of [46], this matrix is:

$$\begin{aligned}
 V_{CKM} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}.
 \end{aligned} \tag{18.2}$$

Clearly, this matrix satisfies the unitarity relation  $V_{CKM}^\dagger V_{CKM} = V_{CKM} V_{CKM}^\dagger = I_{3 \times 3}$ , where  $I_{3 \times 3}$  is a 3x3 unit (identity) matrix, which relation is true irrespective of the commutation order of  $V_{CKM}$  and its Hermitian conjugate  $V_{CKM}^\dagger$ . Given the sines and cosines in (18.2) and the connections of these to the quark masses in (18.1), this means that (18.2) can be rewritten exclusively in terms of the quark masses and the CP-violating phase  $\delta$  as:

$$V_{CKM} = \begin{pmatrix} \frac{\sqrt{m_s m_t}}{\sqrt{(m_d + m_s)(m_u + m_t)}} & \frac{\sqrt{m_d m_t}}{\sqrt{(m_d + m_s)(m_u + m_t)}} & \frac{\sqrt{m_u}}{\sqrt{m_u + m_t}} e^{-i\delta} \\ \frac{-\sqrt{m_d m_c} - \sqrt{m_u m_s} \sqrt{\frac{m_u}{m_u + m_t}} e^{i\delta}}{\sqrt{(m_u + m_c)(m_d + m_s)}} & \frac{\sqrt{m_c m_s} - \sqrt{m_u m_d} \sqrt{\frac{m_u}{m_u + m_t}} e^{i\delta}}{\sqrt{(m_u + m_c)(m_d + m_s)}} & \frac{\sqrt{m_u m_t}}{\sqrt{(m_u + m_c)(m_u + m_t)}} \\ \frac{\sqrt{m_u m_d} - \sqrt{m_c m_s} \sqrt{\frac{m_u}{m_u + m_t}} e^{i\delta}}{\sqrt{(m_u + m_c)(m_d + m_s)}} & \frac{-\sqrt{m_u m_s} - \sqrt{m_d m_c} \sqrt{\frac{m_u}{m_u + m_t}} e^{i\delta}}{\sqrt{(m_u + m_c)(m_d + m_s)}} & \frac{\sqrt{m_c m_t}}{\sqrt{(m_u + m_c)(m_u + m_t)}} \end{pmatrix}. \tag{18.3}$$

Note that this contains all of the quark masses *except* for the mass of the bottom quark. This is because of the “leftover” angle  $\theta_{ll\downarrow 31}$  in (15.12) which contains the bottom quark coupling but which is not directly associated with any of the CKM angles. This makes  $V_{CKM}$  independent of the bottom mass. The above (18.3) is what we shall now refer to as the “mass parameterization” of the CKM quark mass mixing matrix.

It is helpful on an element-by element basis to calculate each element times its own conjugate transpose. From this, in the four lower-left entries we obtain a common term with includes  $e^{i\delta} + e^{-i\delta} = 2 \cos \delta = 2c_\delta$ . All told, we find the nine square magnitudes:

$$\begin{aligned}
 |V_{CKM}|^2 &= \begin{pmatrix} |V_{ud}|^2 & |V_{us}|^2 & |V_{ub}|^2 \\ |V_{cd}|^2 & |V_{cs}|^2 & |V_{cb}|^2 \\ |V_{td}|^2 & |V_{ts}|^2 & |V_{tb}|^2 \end{pmatrix} = \begin{pmatrix} V_{ud}^* V_{ud} & V_{us}^* V_{us} & V_{ub}^* V_{ub} \\ V_{cd}^* V_{cd} & V_{cs}^* V_{cs} & V_{cb}^* V_{cb} \\ V_{td}^* V_{td} & V_{ts}^* V_{ts} & V_{tb}^* V_{tb} \end{pmatrix} \quad (18.4) \\
 &= \begin{pmatrix} \frac{m_s m_t}{(m_d + m_s)(m_u + m_t)} & \frac{m_d m_t}{(m_d + m_s)(m_u + m_t)} & \frac{m_u}{m_u + m_t} \\ \frac{m_d m_c + m_u m_s \frac{m_u}{m_u + m_t} + \sqrt{m_u m_d m_c m_s} \sqrt{\frac{m_u}{m_u + m_t}} 2c_\delta}{(m_u + m_c)(m_d + m_s)} & \frac{m_s m_c + m_u m_d \frac{m_u}{m_u + m_t} - \sqrt{m_u m_d m_c m_s} \sqrt{\frac{m_u}{m_u + m_t}} 2c_\delta}{(m_u + m_c)(m_d + m_s)} & \frac{m_u m_t}{(m_u + m_c)(m_u + m_t)} \\ \frac{m_d m_u + m_c m_s \frac{m_u}{m_u + m_t} - \sqrt{m_u m_d m_c m_s} \sqrt{\frac{m_u}{m_u + m_t}} 2c_\delta}{(m_u + m_c)(m_d + m_s)} & \frac{m_s m_u + m_c m_d \frac{m_u}{m_u + m_t} + \sqrt{m_u m_d m_c m_s} \sqrt{\frac{m_u}{m_u + m_t}} 2c_\delta}{(m_u + m_c)(m_d + m_s)} & \frac{m_c m_t}{(m_u + m_c)(m_u + m_t)} \end{pmatrix}
 \end{aligned}$$

It will be readily seen that the sum of elements in each of the three rows, and in each of the three columns, is identically equal to 1. This is another way of calculating the six relations in  $V_{CKM}^\dagger V_{CKM} = V_{CKM} V_{CKM}^\dagger = I_{3 \times 3}$  which produce the 1s on the diagonal of  $I_{3 \times 3}$ . For the three off-diagonal 0s in  $I_{3 \times 3}$ , there are three combinations 12, 23 and 31 for pairs of unlike rows, and likewise for pairs of unlike columns, that can be used to produce six unitary triangles with identical areas equal to  $\frac{1}{2}$  of the Jarlskog invariant. The magnitude of each of the nine CKM elements, denoted overall by  $|V_{CKM}|$ , is then obtained by taking the square root of each of the nine elements in (18.4), on an element-by element basis. Now we turn to the empirical data.

Equation [12.27] of PDG's [46] reports the magnitudes  $|V_{CKM}|$  of these nine elements based on a global fitting of the empirical data to the theoretically-presumed and experimentally-uncontradicted unitarity condition  $V_{CKM}^\dagger V_{CKM} = V_{CKM} V_{CKM}^\dagger = I_{3 \times 3}$ . For the present discussion, to facilitate comparison, we shall designate PDG's [12.27] of [46] as  $|V_{PDG}|$ . Against this, we will then compare data obtained from the mass parameterizations in (18.3) and (18.4). Moreover, because we will be studying both the center value of each element in  $|V_{CKM}|$  as well as the error bars in each, we separate  $|V_{PDG}|$  into  $|V_{PDG}| = |\bar{V}_{PDG}| + \mathcal{E}_{PDG}$ , with  $|\bar{V}_{PDG}|$  (overbar) denoting the data center and  $\mathcal{E}_{PDG}$  denoting the  $\pm$  error bars of each element. Although most of the PDG error bars have equal magnitudes above and below, those for  $V_{cs}$  and  $V_{td}$  do not. For those,  $\mathcal{E}_{cs} = \begin{smallmatrix} +0.00010 \\ -0.00011 \end{smallmatrix}$  and  $\mathcal{E}_{td} = \begin{smallmatrix} +0.00024 \\ -0.00023 \end{smallmatrix}$ . But because these are fairly close, to keep the calculation simple, we shall use the slightly tighter  $\mathcal{E}_{cs} = \pm 0.00010$  and  $\mathcal{E}_{td} = \pm 0.00023$ . In this notation, and with this change, we simply transcribe [12.27] of [46] below:

$$|V_{PDG}| = |\bar{V}_{PDG}| + \mathcal{E}_{PDG} = \begin{pmatrix} 0.97446 & 0.22452 & 0.00365 \\ 0.22438 & 0.97359 & 0.04214 \\ 0.00896 & 0.04133 & 0.999105 \end{pmatrix} + \begin{pmatrix} \pm 0.00010 & \pm 0.00044 & \pm 0.00012 \\ \pm 0.00044 & \pm 0.00010 & \pm 0.00076 \\ \pm 0.00023 & \pm 0.00074 & \pm 0.00032 \end{pmatrix}. \quad (18.5)$$

Now let use the following data for calculating with the mass parameterization (18.4) to independently obtain  $|V_{CKM}|$  and compare this to  $|V_{PDG}|$ : For the five quark masses, all in GeV, we use the very precise  $m_u c^2 = 0.00222379240$  GeV and  $m_d c^2 = 0.00490647034$  GeV from (15.8), which as reviewed, are obtained from the EPN scheme developed by the author in 2013. The only error bars in this data stem from those of the electron mass and the free neutron-minus-proton mass difference. We do not show errors for the moment because they are several orders of magnitude finer than those of any of the other data to be used, although shortly, we will look more closely at these. We will also utilize the center values and error bars  $m_u c^2 = .0022_{-0.0004}^{+0.0005}$  GeV and  $m_d c^2 = .0047_{-0.0003}^{+0.0005}$  GeV from PDG's [43]. For the strange mass, we know that PDG reports  $m_s c^2 = .095_{-0.003}^{+0.009}$  GeV. However, at (15.10) we were able to deduce a far-tighter value  $m_s c^2 = 0.09242 \pm 0.00038$  GeV which we shall use here. For the top and charm masses PDG reports  $m_c c^2 = 1.275_{-0.035}^{+0.025}$  GeV and  $m_t c^2 = 173.0 \pm 0.4$  GeV. However, at (14.5) we were able to re-center the top mass and also tighten its error bars to inherit those of the charm quark, such that  $m_t c^2 = 172.826_{-0.026}^{+0.035}$  GeV, which we shall use here for the top quark along with PDG's  $m_c c^2 = 1.275_{-0.035}^{+0.025}$  GeV for the charm quark. It is important to note – so that there are no errors in the calculations – that because we used  $m_t c^2 = 174.1035847$  GeV –  $m_c c^2 - m_u c^2$  in (14.5) to tighten the top mass, any upward variation of the top quark mass from its center must be accompanied by a downward variation of the same amount for the charm quark mass, and vice versa. Again, the bottom quark does not appear anywhere in (18.3) and (18.4). Finally, for the phase angle, we shall use  $\delta = \delta_C = 70.998_{-1.917}^{+1.995}$  ° from (14.12) which, as noted, is deduced from the Wolfenstein parameterization reviewed in PDG's [46].

For a first calculation (denoted “A”) we substitute the center values (denoted with an overbar) of all this data into (18.4). That is, we use  $\bar{m}_u c^2 = 0.00222379240$  GeV,  $\bar{m}_d c^2 = 0.00490647034$  GeV,  $\bar{m}_s c^2 = 0.09242$  GeV,  $\bar{m}_c c^2 = 1.275$  GeV,  $\bar{m}_t c^2 = 172.826$  GeV and  $\bar{\delta} = 70.998$  °. Then, we calculate the element-by element square roots. Finally, comparing to (18.5) to obtain the difference  $\Delta_A \equiv |V_{CKM}|_A - |\bar{V}_{PDG}|$ , what we obtain is:

$$\begin{aligned}
 |V_{CKM}|_A &= \begin{pmatrix} 0.97446 & 0.22453 & 0.00359 \\ 0.22438 & 0.97361 & 0.04173 \\ 0.00887 & 0.04093 & 0.999123 \end{pmatrix} \\
 \equiv |\bar{V}_{PDG}| + \Delta_A &= |\bar{V}_{PDG}| + \begin{pmatrix} +0.00000 & +0.00001 & -0.00006 \\ -0.00000 & +0.00002 & -0.00041 \\ -0.00009 & -0.00040 & +0.000018 \end{pmatrix}.
 \end{aligned} \tag{18.6}$$

It is also helpful using (18.5) and (18.6) to obtain the ratio  $\Delta_A / \varepsilon_{PDG}$  on a component-by-component basis using the positive (absolute) values from  $|\varepsilon_{PDG}|$ . This calculation shows that:

$$\frac{\Delta_A}{|\varepsilon_{PDG}|} \text{ (by-component)} = \begin{pmatrix} +0.01536 & +0.01325 & -0.52442 \\ -0.00169 & +0.18164 & -0.54422 \\ -0.39452 & -0.54004 & +0.551116 \end{pmatrix}. \tag{18.7}$$

A magnitude of 1 for any component in (18.7) would mean that this component is precisely at the edge of the error bars, with a + sign indicating above-center and a – sign indicating below-center. The + and – signs in front of the two zeros in (18.6) come from (18.7). That all of these magnitudes are less than 1 indicates that the five masses and the one phase used to obtain (18.6) produce a  $|V_{CKM}|_A$  which is entirely within the experimental errors for the magnitudes of all nine CKM component. The tightest fit to the PDG data center  $|\bar{V}_{PDG}|$  comes from the three upper-left and the center components, while the largest variation comes from the remaining five components in the third row and third column, with most of these elements about halfway to the error bars.

In statistical terms, the reason (18.7) is a helpful measure is that error bars in physics are often set at some number of standard deviations from center, and usually, to  $3\sigma$ , which is a 95% Gaussian confidence level. So if the error bars in each of the  $\varepsilon_{PDG}$  in (18.5) are set to some  $X\sigma$ , then  $X(\Delta_A / |\varepsilon_{PDG}|)\sigma$  provides a rough measure of the number of standard deviations from center represented by each of the  $\Delta_A$  in (18.6). For example, if  $\varepsilon_{PDG_{ub}} = \pm 0.00012$  in (8.5) is a  $3\sigma$  spread from center, then the corresponding  $-0.52442$  entry in (18.7) would represent an approximate  $-(3 \times 0.52442)\sigma$  deviation downward of center, or just larger than 1.5 sigma.

At this point, we now begin a series of calculations using the unitarity of the CKM matrix to further tune the center values and error bars of the five quark masses and the phase in (18.3), (18.4), and the standard parameterization mixing angles and CKM matrix components.

Each of the components of  $V_{CKM}$  is influenced in different ways by each of the five quark masses as well as the phase in (18.3). And as noted, variations in the top mass are tied to those in

the charm mass owing to (14.5). With this in mind, sample calculations reveal that the least-centered components in (18.6), (18.7) are most affected by, and can be brought much closer to center, by an upward revision of the top and thus a downward revision of the charm quark mass. Maintaining three decimal places for the top and charm masses in GeV, it turns out that by adding 0.025 GeV to  $m_t c^2 = 172.826$  GeV and subtracting the same from  $m_s c^2 = 0.09242$  GeV, while maintaining without change all other data used to obtain (18.6), a second (B) calculation reveals:

$$\begin{aligned} |V_{CKM}|_B &\equiv |\overline{V}_{PDG}| + \Delta_B \\ &= \begin{pmatrix} 0.97446 & 0.22453 & 0.00359 \\ 0.22438 & 0.97359 & 0.04214 \\ 0.00896 & 0.04133 & 0.999105 \end{pmatrix} = |\overline{V}_{PDG}| + \begin{pmatrix} +0.00000 & +0.00001 & -0.00006 \\ -0.00000 & +0.00000 & +0.00000 \\ -0.00000 & +0.00000 & +0.00000 \end{pmatrix}. \end{aligned} \quad (18.8)$$

With all but two of the magnitudes now perfectly-centered to five-digit accuracy, this is the “best fit” for the top and charm mass center values. The ratio calculation akin to (18.7) now reveals:

$$\frac{\Delta_B}{|\mathcal{E}_{PDG}|} \text{ (by-component)} = \begin{pmatrix} +0.01537 & +0.01325 & -0.52658 \\ -0.00950 & +0.01105 & +0.00113 \\ -0.01877 & +0.00562 & +0.007539 \end{pmatrix}. \quad (18.9)$$

This is now an extremely-tight fit to  $|\overline{V}_{PDG}|$ . In fact, with the exception of  $|V_{ub}|$  which is 52.658% of the way from the data center to the error bar edge, every single one of the components is less than 2% removed from the center, which means more than 98% from the extremes of the error range. We use this result in (18.8) and (18.9) to re-center the top and charm quark masses, for which we shall now use  $\boxed{\overline{m}_c c^2 = 1.250 \text{ GeV}}$  and  $\boxed{\overline{m}_t c^2 = 172.851 \text{ GeV}}$ . This 0.025 GeV downward adjustment of the charm mass center is within the lower  $-0.035$  GeV bound for  $m_c c^2 = 1.275_{-0.035}^{+0.025}$  GeV from PDG’s [43], which is a bound that was also inherited by the top quark at (14.5). We will also examine how to revise the error bars for the top and charm masses, but before we do so, some further calculation with the other quark masses and the phase is warranted.

As just noted, the only entry in (18.9) which is more than 2% removed from its center toward an extreme of its error bar is the upper-right  $|V_{ub}|$ , which is 52.658% removed. So, we start here to see what, if anything, can be done to bridge this gap. We see from taking the square root in (18.4) that the magnitude of this component is  $|V_{ub}| = m_u / \sqrt{m_u + m_t}$ . So, this is determined only by the top and up masses and not the phase, because the phase factor in (18.3) is removed via  $e^{-i\delta} e^{i\delta} = 1$  when its magnitude is taken. It can easily be calculated that if we try to bring this component closer to center by changing the top (thus charm) quark mass again, any favorable impact on  $|V_{ub}|$  would be very tiny, and greatly outweighed by the negative affect on many other

components, especially the lower-right entries. This is easily seen by observing that the above 0.025 GeV adjustment of the top and charm masses produced only a very slight change from  $\Delta_{Aub} / |\mathcal{E}_{PDGub}| = -0.52442$  in (18.7) to  $\Delta_{Bub} / |\mathcal{E}_{PDGub}| = -0.52658$  in (18.9), and that the beneficial effects of all other CKM matrix magnitudes was far greater.

So, the only other option for bridging the  $|V_{ub}|$  gap is to re-center the up-quark mass. Testing this possibility, it turns out that  $V_{ub}$  can be made to fit its center in (18.6) if the up-quark mass is increased to approximately 2.302837 MeV. But there are two problem with this: First, this is even outside the broad bar of  $m_u c^2 = 2.2_{-0.4}^{+0.5}$  MeV reported by PDG. Second, the cost of fitting  $V_{ub}$ , is that we ruin the fit in (18.8) for most of the other elements in CKM matrix, with the entire bottom row and the two right elements in the middle row approaching the outer extremes of their error bars. Our conclusion is that this  $|V_{ub}|$  gap cannot be remedied, without adverse impacts on several of the other very tight fits in (18.8) and (18.9). And from this, we conclude that the problem is not with the top or up quark masses which we leave as is, but rather, that  $|\overline{V}_{ubPDG}| = 0.00365$  in the PDG data is an outlier which itself will need to be re-centered downward, which will be confirmed in (18.13) below.

Now let's focus closely on the up and down quark masses. At (15.8) we introduced the very precise masses  $m_u c^2 = 0.00222379240$  GeV and  $m_d c^2 = 0.00490647034$  GeV for these quarks, based on the author's earlier-referenced 2013 work [33], [31]. Suppose, however, that these results were not known, and that we had to obtain these masses exclusively by a best fit using the mass parameterization matrix (18.3). To do so, we proceed as follows: Into (18.4), we insert  $\overline{m}_c c^2 = 1.250$  GeV and  $\overline{m}_t c^2 = 172.851$  GeV, as adjusted at (18.9). We continue to use  $\overline{m}_s c^2 = 0.09242$  GeV and  $\overline{\delta} = 70.998^\circ$ . Then, we turn to PDG's  $m_u c^2 = .0022_{-0.0004}^{+0.0005}$  GeV and  $m_d c^2 = .0047_{-0.0003}^{+0.0005}$  GeV for guidance. To get into the ballpark, we start by also inserting  $m_u c^2 = .0022$  GeV from the up-quark mass center value. Then, we substitute various "test" values for the down mass over the range  $m_d c^2 = .0047_{-0.0003}^{+0.0005}$  GeV to see where, exactly, a ratio calculation akin to (18.7) and (18.9) will cause any one of the CKM magnitudes to hit its outer error bar, as detected by an entry  $\Delta / |\mathcal{E}_{PDG}| = \pm 1$ . On the high side, it turns out that  $m_d c^2 = 0.004926476$  GeV will push  $\Delta_{us} / |\mathcal{E}_{PDGus}| \rightarrow +1$ , and will push the other three upper-left elements to also grow very close to their edges as well. On the low side, it turns out that  $m_d c^2 = 0.004888578$  GeV will push  $\Delta_{cs} / |\mathcal{E}_{PDGcs}| \rightarrow +1$  and likewise push other upper-left components to their error bar bounds as well. We then simply average this high and low result to obtain a center value  $m_d c^2 = 0.004907527$  GeV, and an overall spread whereby  $m_d c^2 = 4.907527 \pm 0.018949$  MeV when represented in MeV.

Then, to obtain a spread for the up mass, we leave the down mass at the foregoing  $m_d c^2 = 0.004907527$  GeV and sample various values for the up mass until one of the CKM components is pushed to its outer error bar. We find  $m_u c^2 = 0.002303843$  GeV pushes  $\Delta_{ts} / |\mathcal{E}_{PDGts}| \rightarrow +1$  with several other components also approaching their edges, and that  $m_u c^2 = 0.002153906$  GeV pushes  $\Delta_{ub} / |\mathcal{E}_{PDGub}| \rightarrow +1$  and moves several other components to their edges. Averaging high and low, the middle value is  $m_u c^2 = 0.002228875$  GeV. Accordingly, the overall result, represented in MeV, is  $m_u c^2 = 2.228875 \pm 0.074969$  MeV. If we repeat the exact same calculation but start with  $m_u c^2 = .0023$  GeV rather than  $m_u c^2 = .0022$  GeV, for the down quark we obtain  $m_d c^2 = 4.903723 \pm 0.016961$  MeV, and then using this center value in the same way as before, for the up quark we obtain  $m_u c^2 = 2.228836 \pm 0.074930$  MeV. To the first two digits in MeV, both of these test calculations coalesce to the common results  $m_u c^2 = 2.23 \pm 0.07$  MeV and  $m_d c^2 = 4.90 \pm 0.02$  MeV.

Now let's take stock of all this: Starting from the PDG data and working in MeV, we have the very broad ranges  $m_u c^2 = 2.2_{-4}^{+5}$  MeV and  $m_d c^2 = 4.7_{-0.3}^{+0.5}$  MeV which incorporate hadronic scattering experiments and so contain not only experimental errors but scheme-dependent errors. Then, in sampling calculations based on the mass parameterization (18.3) and (18.4) of the CKM matrix, these are already tightened to  $m_u c^2 = 2.23 \pm 0.07$  MeV and  $m_d c^2 = 4.90 \pm 0.02$  MeV. Compared to PDG's  $m_u c^2 = 2.2_{-4}^{+5}$  MeV the former has an error bar which is almost ten times as tight and so adds one digit of accuracy, and in this extra digit reveals a new center slightly-elevated by .03 MeV. Compared to PDG's  $m_d c^2 = 4.7_{-0.3}^{+0.5}$  MeV, the latter has a center substantially-elevated by a full .2 MeV, and an error bar more than ten times as tight. We may refer to the "scheme" for obtaining these tighter up and down masses as the "CKM scheme," because it is based on tying these masses to the observed CKM mixing angles rather than to hadronic scattering experiments which depend heavily on the impact scale  $\mu$ . In contrast to the scale-dependence of hadronic scattering, these CKM mixing angles may be regarded as constants below the weak scale at  $\mu = M_w$ , see [46] following [12.6], and so provide *scale-independent data* for establishing quark masses in this CKM scheme. Note, however, with the quark masses now tied to the CKM angles by (18.1), that the  $\mu$ -dependence of the quark masses and of the CKM elements can no longer be regarded independently.

Then, we compare this to the author's 2013 results  $m_u c^2 = 2.22379240$  MeV and  $m_d c^2 = 4.90647034$  MeV reproduced in (15.8) and obtained from the EPN (electron-proton-neutron) scheme represented by (15.7), and find that the foregoing fitting of the CKM mass parameterization are confirming of these earlier results. Specifically, these 2013 numeric results were originally derived in atomic mass units, then converted over using  $1 \text{ u} = 931.494061(21)$  MeV into less-precise MeV, see note [a] in [53]. Now, in MeV, let's

calculate the error bars from this EPN scheme: Calculating with (15.7) using PDG's 2018 data [47], the electron rest mass  $m_e c^2 = 0.5109989461(31)$  MeV is known to eight digits in MeV with error bars in the ninth and tenth positions, while the free proton and neutron masses are known to 5 digits in MeV with errors in the sixth position [48]. But it is the neutron minus proton *mass difference* which drives the latter equation (15.7), and this difference is known to six digits in MeV with errors in the seventh and eighth digits, namely  $M_n - M_p = 1.29333205(48)$  MeV, see the most recent NIST CODATA at [54]. So, with errors explicit, a recalculation of the up and down masses using (15.7) and the latest reported data yields  $m_u c^2 = 2.22379229(55)$  MeV and  $m_d c^2 = 4.90647034(55)$  MeV, inheriting the seventh-position errors from the n-p mass difference. This EPN scheme is similar in nature to the CKM scheme just reviewed, because each of these is independent of impact scale  $\mu$ . The CKM scheme instead uses the CKM elements which are effective constants for  $\mu \rightarrow 0$ , and the EPN scheme instead uses  $m_e c^2$  and  $M_n - M_p$  which likewise are effective constants for  $\mu \rightarrow 0$ .

So, the CKM mass parameterization calculation above produces  $m_u c^2 = 2.23(7)$  MeV versus  $m_u c^2 = 2.22379229(55)$  MeV from an updated EPN calculation based on (15.7). And, the CKM calculation yields  $m_d c^2 = 4.90(2)$  MeV versus  $m_d c^2 = 4.90647034(55)$  MeV from the EPN update. This means that calculation from the CKM mass parameterization provides *independent validation* that the EPN relations (15.7) obtained in 2013 are correct, because the EPN calculation falls within the substantially-tightened CKM error bars for both the up and down quark masses, and because all of these are well-within the PDG error bars. Because the EPN calculation is accurate to six digits in MeV while the CKM is accurate to only 1 digit in MeV, and because these data points are fully concurrent, from here we shall use the foregoing updated EPN values for the up and down masses, with error bars included. A six-digit accuracy in MeV is a nine-digit accuracy in GeV, so compared the all the other quark masses, this is an extreme degree of accuracy that enables us to effectively use these up and down quark masses without concern for their errors.

Now let's turn to the CP-violating phase. So, far, for this we have used  $\delta = \delta_C = 70.998_{-1.917}^{+1.995}$  ° calculated from [46]. As seen in (18.4), this phase only affects the magnitudes of the four lower-left CKM matrix components  $|V_{cd}|$ ,  $|V_{cs}|$ ,  $|V_{td}|$  and  $|V_{ts}|$ . Moreover, close study of (18.4) reveals that this phase term appears in each of these four components in identical form, stemming from their commonly-shared term  $s_{13} e^{i\delta}$  in (18.2). So, because the  $\Delta/|\mathcal{E}_{PDG}|$  ratios effectively measure statistical standard deviations as noted following (18.7), one way to measure a "best fit" for the phase is to simply sum the absolute values of  $\Delta/|\mathcal{E}_{PDG}|$  ratios for each of these components by calculating  $\sum_{cd,cs,td,ts} (|\Delta/|\mathcal{E}_{PDG}|)$ , then ascertain a phase magnitude which minimizes this sum. Doing so, we find a minimum sum at  $\delta = 71.066$  °. So, we now re-center the phase by  $0.068$  ° to this  $\bar{\delta} = 71.066$  °. Then, we proceed to determine outside

errors in the same way we did above, by finding which values of the phase push one of the CKM elements up to a  $\Delta/|\mathcal{E}_{PDG}| = \pm 1$  value corresponding to the edge of its error bars. When we do this, we find that  $\delta = 71.066^{+3.786}_{-3.775}^\circ$ , which slightly raises the center and just about doubles the  $^{+1.995}_{-1.917}^\circ$  spread from PDG.

Now, finding error bars by calculating mass and phase angles to produce  $\Delta/|\mathcal{E}_{PDG}| = \pm 1$  effectively uses CKM errors spreads  $\mathcal{E}_{PDG}$  and global unitarity of the CKM matrix to set error bars. The fact that the spread obtained using these  $\mathcal{E}_{PDG}$  is wider than the spread that PDG reported from CP experiments directly detecting  $\delta$  simply means that data used to directly determine  $\delta$  provides more-accurate information than does the global CKM fitting. As such, we shall keep the re-centered  $\bar{\delta} = 71.066^\circ$  from here forward, but will also keep the tighter  $^{+1.995}_{-1.917}^\circ$  spread from PDG. Specifically, because we have now raised the center angle by  $0.068^\circ$ , we shall keep the high-end error bar at its same absolute angle magnitude, thus subtracting  $(+1.995 - 0.068)^\circ = +1.927^\circ$  to establish the upper bound. For the lower bound we keep  $-1.985^\circ$ , however, we now pin this to the newly-raised center, effectively trimming off  $0.068^\circ$  from the bottom of the overall spread. Based on all this, from here forward we shall employ the revised phase angle  $\boxed{\delta = 71.066^{+1.927}_{-1.917}^\circ}$ .

Finally, at (15.10) we were able to re-center and tighten the strange quark mass to  $m_s c^2 = 92.42 \pm 0.38$  GeV. Using the fitting techniques that we just applied for the phase, we see if any further refinements are possible for the strange quark. From (18.3) and (18.4) it is seen that the strange quark mass affects the CKM matrix in the leftmost two columns. Accordingly, using all of the foregoing adjustments for the other masses and the phase, we calculate the sum  $\Sigma(|\Delta|/|\mathcal{E}_{PDG}|)$  over these two columns, then look for where this sum is minimized. Then, we see what values for the strange mass cause one of these elements to reach a  $\Delta/|\mathcal{E}_{PDG}| = \pm 1$  at its error bar extreme. The mildly-revised result of this calculation is that  $\boxed{m_s c^2 = 92.416^{+0.376}_{-0.371}$  MeV. This opens up the third decimal place in MeV, nominally reducing the mass center, and having a very tiny impact on the overall error spread.

Now, we return to the top and charm masses which we re-centered at (18.8) and (18.9) to  $\bar{m}_t c^2 = 172.851$  GeV and  $\bar{m}_c c^2 = 1.250$  GeV, to see if the subsequent adjustments to the other masses and the phase produce any discernable impact, and also, to establish error bars for these two masses. As to the center values, as we did for the phase and the strange mass, we take the sum  $\Sigma(|\Delta|/|\mathcal{E}_{PDG}|)$  over all of the CKM matrix elements which are affected by the top and charm masses (which happen to be all nine elements, see (18.3) and (18.4)), and look for where this sum is minimized. If we try to add an extra decimal point by adding  $\pm 0.0001$  GeV respectively to the top and charm masses, we find that even this small adjustment increases the  $\Sigma(|\Delta|/|\mathcal{E}_{PDG}|)$  sum.

So, we leave these centers exactly where they are. As to error bars, we again adjust these masses up and down to calculate where  $\Delta/|\varepsilon_{PDG}| = \pm 1$  for one of the CKM magnitudes. This reveals that based on global unitarity fitting,  $m_t c^2 = 172.851_{-0.046}^{+0.044}$  GeV and  $m_c c^2 = 1.250_{-0.044}^{+0.046}$  GeV, mindful again that a raised top mass correlates to a lowered charm mass and vice versa via (14.5). However, prior to the data adjustments we started to make at (18.7), we were using  $m_t c^2 = 172.826_{-0.026}^{+0.035}$  GeV and  $m_c c^2 = 1.275_{-0.035}^{+0.025}$  GeV, which has a tighter error bar than that provided by unitarity fitting. This indicates that direct determinations of the charm mass (which now drives the top mass error bars) are more accurate than unitarity fittings of these masses, so that we should use the  $_{-0.035}^{+0.025}$  GeV range from the direct PDG charm mass data rather than the  $_{-0.044}^{+0.046}$  charm range just ascertained by unitarity fitting. But, at (18.8), the unitarity fitting required us to raise the top quark and lower the charm quark masses by 0.025 GeV. So, as we just did for the phase, we should maintain the absolute upper mass value for the top and absolute lower mass value for the charm, while trimming this 0.025 GeV from the opposite ends of the range. Accordingly, our final center-values and ranges for the top and charm masses are  $m_t c^2 = 172.851_{-0.026}^{+0.010}$  GeV and  $m_c c^2 = 1.250_{-0.10}^{+0.025}$  GeV.

Accordingly, assembling all of the foregoing results, we have been able to use the mass parameterization (18.3), (18.4) of the CKM quark mass mixing matrix and the unitarity of this matrix, along with other PDG mass data and the EPN mass scheme of the author's [33], [31] to refine five of the six quark masses and the CP-violating phase as follows:

$$\boxed{m_u c^2 = 2.22379229(55) \text{ MeV}; \quad m_c c^2 = 1.250_{-0.10}^{+0.025} \text{ GeV}; \quad m_t c^2 = 172.851_{-0.026}^{+0.010} \text{ GeV} \\ m_d c^2 = 4.90647034(55) \text{ MeV}; \quad m_s c^2 = 92.416_{-0.371}^{+0.376} \text{ MeV}; \quad \delta = 71.066_{-1.917}^{+1.927} \text{ }^\circ} \quad (18.10)$$

Again, the bottom quark mass does not affect the CKM mass parameterization because it does not appear in (18.3), (18.4). Rather, the bottom quark mass is determined from the remaining five quark masses plus the Higgs boson mass using (16.5), review also the quark mass reparameterization summary at (17.1). As noted at (16.5), in reality, it is the PDG bottom quark mass  $m_b c^2 = 4.18_{-0.03}^{+0.04}$  GeV which sets the error bar for the Higgs mass, because at present, the bottom mass is more precisely-known than the Higgs mass.

Having this tightened data in (18.10), we may now also insert this into (18.1) to calculate revised CKM mixing angles for the standard parameterization, namely:

$$\boxed{\theta_{C12} = 12.975 \pm 0.025^\circ; \quad \theta_{C23} = 2.415_{-0.024}^{+0.010} \text{ }^\circ; \quad \theta_{C13} = 0.205510_{-0.000006}^{+0.000015} \text{ }^\circ; \quad \delta = 71.066_{-1.917}^{+1.927} \text{ }^\circ} \quad (18.11)$$

Compared with the values calculated in (14.12) from PDG's [46], we see that the previous  $\theta_{C12} = 12.975 \pm 0.026^\circ$  is virtually unchanged, while the previous  $\theta_{C23} = 2.415 \pm 0.053^\circ$  retains the same center but has its spread cut by about a factor of 3. However, there is a very substantial

change to the previous  $\theta_{C13} = 0.209_{-0.013}^{+0.015}$ °. First, its center is lowered by about 0.0035° which is substantial (just under one part in 60) in proportion to this already-small angle. Second, the error spread is tightened by a factor of about 1300, which is a precision improvement *exceeding three orders of magnitude*. Finally, as already reviewed, the foregoing does not alter the precision of the phase, other than raising its center by 0.068° which we also use to trim off the bottom of the error range by a like amount. Note also, importantly, that variations in  $\theta_{C23}$  and  $\theta_{C13}$  are *interdependent*, because of the interdependence of the top and charm masses found at (14.5), and because as seen in (18.1),  $\theta_{C23}$  and  $\theta_{C13}$  have the same form but for the interchange of the top and charm masses. Thus, the high end of  $\theta_{C23}$  corresponds to the low end of  $\theta_{C13}$  and vice versa, which we denote by  $13_-^+ \Leftrightarrow 23_+^-$ , just as for the top and charm masses themselves due to (14.5).

Finally, we are in a position to recalculate the magnitudes of all the components of the entire CKM matrix more tightly. For this, it is simplest to manage the high / low combinations of masses and angles by using the standard parameterization of (18.2) together with the angles and error bars found in (18.11). First, we write down the trigonometric content of the magnitude of each element in (18.2), in the usual way, as such:

$$|V_{CKM}| = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13} \\ \sqrt{s_{12}^2c_{23}^2 + c_{12}^2s_{23}^2s_{13}^2 + 2c_{12}s_{12}c_{23}s_{23}s_{13}c_\delta} & \sqrt{c_{12}^2c_{23}^2 + s_{12}^2s_{23}^2s_{13}^2 - 2c_{12}s_{12}c_{23}s_{23}s_{13}c_\delta} & s_{23}c_{13} \\ \sqrt{s_{12}^2s_{23}^2 + c_{12}^2c_{23}^2s_{13}^2 - 2c_{12}s_{12}c_{23}s_{23}s_{13}c_\delta} & \sqrt{c_{12}^2s_{23}^2 + s_{12}^2c_{23}^2s_{13}^2 + 2c_{12}s_{12}c_{23}s_{23}s_{13}c_\delta} & c_{23}c_{13} \end{pmatrix}. \quad (18.12)$$

Then, we simply insert the sines and cosines of the angles in (18.11) into (18.12) to obtain our results. However, given that  $\theta_{C12}$ ,  $\theta_{C13}$  and  $\delta$  can each range over their error bars independently of one another but at there is a  $13_-^+ \Leftrightarrow 23_+^-$  correlation between  $\theta_{C13}$  and  $\theta_{C23}$  because of (14.5) and (18.1), there are eight high (+) versus low (-) combinations that need to be calculated particularly for the four lower-left terms in (18.12), where we define “high” as the high value of an angle, and “low” as the low value of an angle. These correspond to a binary counting sequence from 7 downward to 0, namely,  $\theta_{C12}, \theta_{C13}, \delta = +++$ ,  $\theta_{C12}, \theta_{C13}, \delta = ++-$ ,  $\theta_{C12}, \theta_{C13}, \delta = +-+$ ,  $\theta_{C12}, \theta_{C13}, \delta = +--$ ,  $\theta_{C12}, \theta_{C13}, \delta = -++$ ,  $\theta_{C12}, \theta_{C13}, \delta = -+-$ ,  $\theta_{C12}, \theta_{C13}, \delta = --+$  and  $\theta_{C12}, \theta_{C13}, \delta = ---$ . The upper-right entry  $s_{13}$  has only a single high-low binary choice because it has a single angle. The right column also has a single binary choice because of the  $13_-^+ \Leftrightarrow 23_+^-$  correlation. Because the error bar for  $c_{13} = 0.9999935674_{-0.0000000010}^{+0.0000000004}$  does not change its value until the ninth digit, for all practical purposes at the five-to-six-digit accuracy of the CKM matrix elements the remaining two entries in the top row also effectively have only a single binary choice.

So, carrying out this calculation, we end up below with the final result for the refined CKM magnitude matrix, using a  $\pm$  convention for error bars with the high-side always on top and the

lower-side always on the bottom, along with a \* matrix coded to provide further compactly-coded details as will be explained below:

$$\boxed{
 \begin{aligned}
 |V_{CKM}| &= \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22453 \pm 0.00043 & 0.0035868^{+0.0000003}_{-0.0000001} \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00012}_{-0.00011} & 0.04214^{+0.00017}_{-0.00041} \\ 0.00896^{+0.00017}_{-0.00022} & 0.04133^{+0.00019}_{-0.00043} & 0.999105^{+0.000017}_{-0.000007} \end{pmatrix} * \\
 * & \begin{pmatrix} & 12_+^- & & & 12_+^+ & & & & 13_+^- \\ 12_+^-(0.00086), \delta_+^-(0.00001), 13_+^+(0.00001) & & & & 12_+^-(0.00020), 13_+^+(0.00002), \delta_+^+(0.00000) & & & & 13_+^- \\ \delta_+^-(0.00023), 13_+^-(0.00012), 12_+^+(0.00003) & & & & 13_+^-(0.00057), \delta_+^-(0.00005), 12_+^-(0.00001) & & & & 13_+^+ \end{pmatrix}
 \end{aligned}
 } \quad (18.13)$$

Comparing this with the usual PDG data reproduced in (18.5), in the top row  $|V_{ud}|$  is unchanged to five digits, and  $|V_{us}|$  is lowered by 1 part per  $10^5$  with its error bars substantially unchanged. But  $|V_{ub}|$  which we identified as an outlier after (18.9) is substantially changed in two ways: First, its precision is enhanced by more than two orders of magnitude, and more precisely, by a factor of about 640. Second, as anticipated, its center is indeed lowered by over 6 parts per  $10^5$ . In the remainder of the right column  $|V_{cb}|$  and  $|V_{tb}|$  have identical centers to those of the PDG data to five digits, but the error bar for each is tightened by a factor of just over 2.6.

The final point to be noted before we turn to the lower-left components which are functions of the phase, is the corresponding data in the \* matrix, which is coded to capture additional detail about how errors in  $\theta_{C12}$ ,  $\theta_{C13}$  and  $\delta$  impact these CKM matrix error bars. So for  $|V_{ud}|$  and  $|V_{us}|$ , the 12 designation in the \* matrix indicates that the high and low ranges of the error bar are determined entirely by  $\theta_{C12}$  (which  $c_{13} = 0.9999935674^{+0.0000000004}_{-0.0000000010}$  renders an allowable approximation, as noted above). But, the  $12_+^-$  designation for  $|V_{ud}| = 0.97446^{+0.00010}_{-0.00010}$  indicates that when  $\theta_{C12} = 12.975 - 0.025^\circ$  in (18.11) is at the low end of its error bar then  $|V_{ud}| = 0.97446 + 0.00010$  is at the high end of its error bar, and vice versa. In other words, the signs attendant to  $12_+^-$  indicated that this angle and the CKM component magnitude are reverse-correlated. Likewise, the  $12_+^+$  designation for  $|V_{us}| = 0.22453 \pm 0.00043$  means that the error-bar directions for  $\theta_{C12}$  and  $|V_{us}|$  are forward-correlated. So for the right column, given that the  $13_+^+ \Leftrightarrow 23_+^-$  correlation allows  $\theta_{C13}$  to serve as a reverse-correlated proxy for  $\theta_{C23}$  regarding error bars, the 13 designation indicates that all three of these error bars are driven solely by  $\theta_{C13}$ . The  $13_+^+$  for  $|V_{ub}|$  and  $|V_{tb}|$  indicates a forward correlation with  $\theta_{C13}$  (thus inverse with  $\theta_{C23}$ ), while the  $13_+^-$  for  $|V_{cb}|$  indicates an inverse correlation with  $\theta_{C13}$  (thus forward with  $\theta_{C23}$ ).

Turning to the four lower-left components in (18.13), in the main matrix we find that the center values for all these components are likewise identical to the PDG data to five digits. However, these error bars depend on the eight binary combinations noted earlier as to whether each of  $\theta_{C12}$ ,  $\theta_{C13}$  and  $\delta$  is at the top or bottom of its error bar. Moreover, each of these three angles has its own distinct effect on the error bars. Therefore, the forward and reverse correlations are coded for each of these angles in the same way as was reviewed in the previous paragraph, noting also that the  $13_-^+ \Leftrightarrow 23_+^-$  correlation always applies as well. But now, in addition, in parenthesis next to each of these angles is a five-decimal number which indicates the magnitude of the effect that each angle has on the CKM element magnitude, from one end to the other of that angle's error bars. Moreover, these are ordered for each CKM element from largest to smallest effect. So, for example, over the entire  $0.00088$  spread of  $|V_{cd}| = 0.22438 \pm 0.00044$ , the coded sequence  $12_+^+(0.00086), \delta_+^-(0.00001), 13_-^+(0.00001)$  indicates that the forward-correlated  $\theta_{C12}$  accounts for almost all ( $0.00086$ ) of this spread, and that  $\delta_+^-$  and  $\theta_{C13}$  can each swing this component spread by about  $0.00001$  apiece. The center component  $|V_{cs}|$  is most affected by the reverse-correlated  $\theta_{C12}$  with  $12_+^-(0.00020)$ , followed by a smaller contribution from swings in  $\theta_{C13}$  with  $13_-^+(0.00002)$ , and negligible impacts from phase swings with  $\delta_-^+(0.00000)$ . Where the CP-violating phase has its greatest impact is on  $|V_{td}|$  with  $\delta_-^+(0.00023)$ , followed by a less-impactful  $13_-^-(0.00012)$  and a least impactful  $12_+^-(0.00003)$ . And for  $|V_{ts}|$  the dominant angle is  $13_-^-(0.00057)$  followed by  $\delta_+^-(0.00005)$  followed by  $12_+^-(0.00001)$ . So, with this \* matrix it then becomes possible to rapidly ascertain how the error bars in each of the CKM standard parameterization angles correlates with the error bars in the CKM matrix component magnitudes themselves, and the magnitude impact each angle error bar has on each component error bar.

The scientific method holds that empirical confirmation is the repeated absence of empirical contradiction to the point where the possibility of contradiction is statistically eliminated. It is intended that the results in (18.10), (18.11) and (18.13), which clearly refine and tighten predictions about the center values and error bars of five of the six quark masses, the CKM standard parameterization mixing angles and the invariant CKM matrix component magnitudes, will provide a means for experimentally contradicting the results obtained here, with the expectation that no contradictions will be uncovered and the statistical possibility of contradiction will eventually be ruled out. With this, we have completed our study of quark masses, mixing and weak beta decay, and turn now to a similar study of leptons.

## PART IIB: LEPTONS

### 19. Theory of Fermion 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 [55]. 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 [56].

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 (14.9), we began to utilize three quark mass mixing angles denoted  $\theta_{21}$ ,  $\theta_{32}$ ,  $\theta_{31}$  which were later connected at (14.13) and (15.6) to the three real CKM quark mixing angles denoted  $\theta_{C12}$ ,  $\theta_{C23}$  and  $\theta_{C23}$ . Here, for leptons, we shall postulate three analogous mass mixing angles denoted  $\vartheta_{21}$ ,  $\vartheta_{32}$  and  $\vartheta_{31}$ , and will seek out a connection to the three real PMNS angles which we shall denote by  $\theta_{P12}$ ,  $\theta_{P23}$  and  $\theta_{P23}$ . Second, for the quarks, we found as crystallized at (15.11) that there are two minima for the vacuum which play a central role, namely, the well-known vev  $v_{\uparrow} = v \cong 246.22$  GeV established by the Fermi constant and a second  $v_{\downarrow} \cong 6.05$  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 (14.8) 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 (14.9) we perform a bi-unitary transformation  $M_{e\mu\tau}c^2 \rightarrow M'_{e\mu\tau}c^2 = U^\dagger M_{e\mu\tau}c^2 U$  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 (14.9):

$$\begin{aligned} m_\tau c^2 &= \frac{1}{\sqrt{2}} u_\downarrow c_{I\downarrow 32}^2 \\ m_\mu c^2 &= \frac{1}{\sqrt{2}} u_\downarrow c_{I\downarrow 21}^2 s_{I\downarrow 32}^2, \\ m_e c^2 &= \frac{1}{\sqrt{2}} u_\downarrow s_{I\downarrow 21}^2 s_{I\downarrow 32}^2 \end{aligned} \quad (19.1a)$$

$$\begin{aligned} m_\tau c^2 &= \frac{1}{\sqrt{2}} u_\downarrow c_{II\downarrow 31}^2 c_{II\downarrow 32}^2 \\ m_\mu c^2 &= \frac{1}{\sqrt{2}} u_\downarrow s_{II\downarrow 32}^2 \\ m_e c^2 &= \frac{1}{\sqrt{2}} u_\downarrow s_{II\downarrow 31}^2 c_{II\downarrow 32}^2 \end{aligned} \quad (19.1b)$$

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

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:

$$m_l c^2 \equiv \frac{1}{\sqrt{2}} G_l u_\downarrow. \quad (19.2)$$

Using these in (19.1) then yields:

$$\begin{aligned} G_\tau &= c_{I\downarrow 32}^2 \\ G_\mu &= s_{I\downarrow 32}^2 c_{I\downarrow 21}^2, \\ G_e &= s_{I\downarrow 32}^2 s_{I\downarrow 21}^2 \end{aligned} \quad (19.3a)$$

$$\begin{aligned} G_\tau &= c_{II\downarrow 32}^2 c_{II\downarrow 31}^2 \\ G_\mu &= s_{II\downarrow 32}^2 \\ G_e &= c_{II\downarrow 32}^2 s_{II\downarrow 31}^2 \end{aligned} \quad (19.3b)$$

From either (19.3a) or (19.3b), we use the trigonometric identity  $c^2 + s^2 = 1$  to find that:

$$G_\tau + G_\mu + G_e = 1. \quad (19.4)$$

Then, using (19.2) in (19.4) we find that:

$$\frac{1}{\sqrt{2}} u_\downarrow = m_\tau c^2 + m_\mu c^2 + m_e c^2 = 1883.029 \pm 0.120 \text{ MeV}. \quad (19.5)$$

These are identical in form with analogous relations (15.11) earlier found for the quarks. The numeric value of this vev is computed to three decimals using empirical data from [47], namely:

$$\begin{aligned} m_e c^2 &= 0.5109989461 \pm 0.0000000031 \text{ MeV}; \\ m_\mu c^2 &= 105.6583745 \pm 0.0000024 \text{ MeV}; \quad m_\tau c^2 = 1776.86 \pm 0.12 \text{ MeV} \end{aligned} \quad (19.6)$$

Next, we restructure (19.3) to isolate sines and cosines, then use (19.4) to obtain:

$$\begin{aligned} c_{I\downarrow 32}^2 &= G_\tau \\ c_{I\downarrow 21}^2 &= \frac{G_\mu}{s_{I\downarrow 32}^2} = \frac{G_\mu}{1 - c_{I\downarrow 32}^2} = \frac{G_\mu}{1 - G_\tau} = \frac{G_\mu}{G_\mu + G_e}, \end{aligned} \quad (19.7a)$$

$$\begin{aligned} s_{I\downarrow 21}^2 s_{I\downarrow 32}^2 &= G_e \\ c_{II\downarrow 31}^2 &= \frac{G_\tau}{c_{II\downarrow 32}^2} = \frac{G_\tau}{1 - s_{II\downarrow 32}^2} = \frac{G_\tau}{1 - G_\mu} = \frac{G_\tau}{G_\tau + G_e} \\ s_{II\downarrow 32}^2 &= G_\mu \\ s_{II\downarrow 31}^2 c_{II\downarrow 32}^2 &= G_e \end{aligned} \quad (19.7b)$$

Finally, we use (19.3) in (19.7) and combine with (19.4) and (19.5) to obtain:

$$\begin{aligned} c_{I\downarrow 32}^2 &= G_\tau = \frac{m_\tau c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \frac{m_\tau c^2}{\frac{1}{\sqrt{2}} u_\downarrow} \\ c_{I\downarrow 21}^2 &= \frac{G_\mu}{s_{I\downarrow 32}^2} = \frac{G_\mu}{G_\mu + G_e} = \frac{m_\mu c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \frac{m_\mu c^2}{(m_\tau c^2 + m_\mu c^2 + m_e c^2) - m_\tau c^2} = \frac{m_\mu c^2}{\frac{1}{\sqrt{2}} u_\downarrow - m_\tau c^2}, \\ s_{I\downarrow 21}^2 s_{I\downarrow 32}^2 &= G_e = \frac{m_e c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \frac{m_e c^2}{\frac{1}{\sqrt{2}} u_\downarrow} \end{aligned} \quad (19.8a)$$

$$\begin{aligned} c_{II\downarrow 31}^2 &= \frac{G_\tau}{c_{II\downarrow 32}^2} = \frac{G_\tau}{G_\tau + G_e} = \frac{m_\tau c^2}{m_\tau c^2 + m_e c^2} = \frac{m_\tau c^2}{(m_\mu c^2 + m_\tau c^2 + m_e c^2) - m_\mu c^2} = \frac{m_\tau c^2}{\frac{1}{\sqrt{2}} u_\downarrow - m_\mu c^2} \\ s_{II\downarrow 32}^2 &= G_\mu = \frac{m_\mu c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \frac{m_\mu c^2}{\frac{1}{\sqrt{2}} u_\downarrow} \\ s_{II\downarrow 31}^2 c_{II\downarrow 32}^2 &= G_e = \frac{m_e c^2}{m_\tau c^2 + m_\mu c^2 + m_e c^2} = \frac{m_e c^2}{\frac{1}{\sqrt{2}} u_\downarrow} \end{aligned} \quad (19.8b)$$

Proceeding from here, we use the mass data in (19.6) and the sum in (19.5) together with the relations for  $c_{I\downarrow 32}^2$ ,  $c_{I\downarrow 21}^2$ ,  $c_{II\downarrow 31}^2$  and  $s_{II\downarrow 32}^2$  to calculate that:

$$\begin{aligned}
 \vartheta_{I\downarrow 32} &= 0.23974 \pm 0.00001 \text{ rad} = 13.73605 \pm 0.00045^\circ \\
 \vartheta_{II\downarrow 32} &= 0.23915 \pm 0.00001 \text{ rad} = 13.70231 \pm 0.00045^\circ \\
 \vartheta_{I\downarrow 21} &= 0.06943 \pm 0 \text{ rad} = 3.97816 \pm 0^\circ \\
 \vartheta_{II\downarrow 31} &= 0.01696 \pm 0 \text{ rad} = 0.97155 \pm 0.00003^\circ
 \end{aligned} \tag{19.9}$$

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

The data in [56] lays out a best fit at both a  $1\sigma$  and  $3\sigma$  range. These spreads will become important momentarily. Therefore, without having more specific data we also *estimate* the  $2\sigma$  spread by taking the average of the  $1\sigma$  and  $3\sigma$  spreads. We then show the central observed value followed by successive ranges also shown for each of  $1\sigma$ , the estimated  $2\sigma$  as just mentioned, and  $3\sigma$ , respectively. Presented in this way, the four PMNS parameters, in degrees, are:

$$\begin{aligned}
 \theta_{P12} &= 33.62_{-0.76}^{+0.78} \text{ }_{-1.48}^{+1.605} \text{ }_{-2.2}^{+2.43} \text{ }^\circ \\
 \theta_{P13} &= 8.549_{-0.15}^{+0.15} \text{ }_{-0.3}^{+0.295} \text{ }_{-0.45}^{+0.44} \text{ }^\circ \\
 \theta_{P23} &= 47.2_{-3.9}^{+1.9} \text{ }_{-5.4}^{+3.1} \text{ }_{-6.9}^{+4.3} \text{ }^\circ \\
 \delta_P &= 234_{-31}^{+43} \text{ }_{-60.5}^{+91.5} \text{ }_{-90}^{+140} \text{ }^\circ
 \end{aligned} \tag{19.10}$$

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 (19.5) 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 (19.8) and referring to (19.5), we see that each numerator contains a specific lepton mass, while each denominator contains the sum  $\frac{1}{\sqrt{2}}u_\downarrow = m_\tau c^2 + m_\mu c^2 + m_e c^2 = 1883.029 \pm 0.120 \text{ MeV}$ . Because  $u_\downarrow$  is what we are postulating is a vev for the charged, isospin-down leptons, and because the angles deduced in (19.9) do not come anywhere near the empirical data in (19.10), 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 (19.5) 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:

$$\frac{1}{\sqrt{2}}u'_\downarrow \equiv \frac{1}{\sqrt{2}}u_\downarrow + \delta_\downarrow = m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_\downarrow = 1883.029 \pm 0.120 \text{ MeV} + \delta_\downarrow \text{ MeV}, \tag{19.11}$$

which *does* allow at least one of  $\vartheta_{I\downarrow 21}$  or  $\vartheta_{II\downarrow 31}$  to fit the empirical data in (19.10), 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 (19.8), but base them on  $u'_{\downarrow}$  defined in (19.11) rather than on  $u_{\downarrow}$ , by defining two new angles  $\vartheta'_{I\downarrow 21}$  and  $\vartheta'_{II\downarrow 31}$  according to:

$$\begin{aligned} \cos^2 \vartheta'_{I\downarrow 21} &\equiv \frac{m_{\mu}c^2}{\frac{1}{\sqrt{2}}u'_{\downarrow} - m_{\tau}c^2} = \frac{m_{\mu}c^2}{\frac{1}{\sqrt{2}}u_{\downarrow} + \delta_{\downarrow} - m_{\tau}c^2} = \frac{m_{\mu}c^2}{m_{\mu}c^2 + m_e c^2 + \delta_{\downarrow}} \\ \cos^2 \vartheta'_{II\downarrow 31} &\equiv \frac{m_{\tau}c^2}{\frac{1}{\sqrt{2}}u'_{\downarrow} - m_{\mu}c^2} = \frac{m_{\tau}c^2}{\frac{1}{\sqrt{2}}u_{\downarrow} + \delta_{\downarrow} - m_{\mu}c^2} = \frac{m_{\tau}c^2}{m_{\tau}c^2 + m_e c^2 + \delta_{\downarrow}} \end{aligned} \quad (19.12)$$

Then, we simply use the known mass data in (19.5) and (19.6), 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 (19.10).

Because error-bars are important in this calculation, let's us briefly comment on how we will approach these. The  $u_{\downarrow}$  in (19.12) 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  $3 \times 3 \times 3 = 27$  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 matters simple, we regard the electron and muon masses to be precisely at the center of their error spreads, and use the  $\pm 0.12$  MeV spread in the tau mass as the basis for calculating the spread in  $u_{\downarrow}$ . This is why there is a  $\pm 0.120$  MeV spread shown in (19.11), and also in (19.5), with one decimal place added.

Working from (19.12) 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  $\vartheta'_{II\downarrow 31} = 8.5490 \pm 0.0003^\circ$ , with a center conforming precisely with the center of the empirical  $\theta_{P13} = 8.549^{+0.15}_{-0.15} \text{ } ^{+0.295}_{-0.3} \text{ } ^{+0.44}_{-0.45} \text{ } ^\circ$  in (19.10). Simultaneously, with this same  $\delta_{\downarrow} = 39.642$  MeV we are able to obtain  $\vartheta'_{I\downarrow 21} = 31.65230 \pm 0^\circ$ . The empirical data in (19.10) tells us that  $\theta_{P12} = 33.62^{+0.78}_{-0.76} \text{ } ^{+1.605}_{-1.48} \text{ } ^{+2.43}_{-2.2} \text{ } ^\circ$ . Given that the  $3\sigma$  error permits an angle as low as  $\theta_{P12} = 31.42^\circ$ , we conclude that  $\delta_{\downarrow} = 39.642$  MeV matches  $\theta_{P13}$  right at the center, and comes in at about  $2.8\sigma$  on the low end of  $\theta_{P12}$ . 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  $3\sigma$  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  $\vartheta'_{I\downarrow 21} = 33.62 \pm 0^\circ$ , conforming precisely with the center of the empirical  $\theta_{P12} = 33.62^{+0.78}_{-0.76} \text{ } ^{+1.605}_{-1.48} \text{ } ^{+2.43}_{-2.2} \text{ }^\circ$  in (19.10). Simultaneously, with this same  $\delta_{\downarrow} = 46.199$  MeV we obtain  $\vartheta'_{II\downarrow 31} = 9.2096 \pm 0.0003^\circ$ . The  $3\sigma$  data puts the corresponding angle at  $\theta_{P13} = 8.989^\circ$  on the high side, so this value for  $\delta_{\downarrow}$  puts us above  $3\sigma$  data. But now we have a basis for interpolating between these two samples.

Because the first  $\delta_{\downarrow}$  sample gave us the center of  $\theta_{P13}$  but produced a low value for  $\theta_{P12}$ , while the second sample gave us the center of  $\theta_{P12}$  but produced a high value for  $\theta_{P13}$ , it appears as if the *actual*  $\theta_{P12}$  is below the center and the *actual*  $\theta_{P13}$  is above the center of what is shown in (19.10). So, for a third sample we take the following approach: Find a  $\delta_{\downarrow}$  which places the  $\theta_{P12}$  match below center and simultaneously places the  $\theta_{P13}$  match above center *by exactly the same statistical spread*. That is, find some  $\delta$  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  $\vartheta'_{II\downarrow 31} = 8.7945 \pm 0.0003^\circ$  versus the empirical  $\theta_{P13} = 8.549^{+0.15}_{-0.15} \text{ } ^{+0.295}_{-0.3} \text{ } ^{+0.44}_{-0.45} \text{ }^\circ$ , and  $\vartheta'_{I\downarrow 21} = 32.39 \pm 0^\circ$  versus the empirical  $\theta_{P12} = 33.62^{+0.78}_{-0.76} \text{ } ^{+1.605}_{-1.48} \text{ } ^{+2.43}_{-2.2} \text{ }^\circ$ . Estimating linearly between center values and  $3\sigma$  values, we find that  $\vartheta'_{II\downarrow 31}$  is about  $1.67\sigma$  above the  $\theta_{P13}$  center and  $\vartheta'_{I\downarrow 21}$  is about  $1.67\sigma$  below the  $\theta_{P12}$  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 (19.12) is able to *simultaneously connect both*  $\theta_{P13}$  and  $\theta_{P12}$  *within about*  $1.67\sigma$  *of their respective experimental centers*, as a physically meaningful relation. Consequently, based on this connection to the experimental date, we now establish:

$$\frac{1}{\sqrt{2}}u'_{\downarrow} = \frac{1}{\sqrt{2}}u_{\downarrow} + \delta_{\downarrow} = m_{\tau}c^2 + m_{\mu}c^2 + m_e c^2 + \delta_{\downarrow} \equiv 1925.047 \pm 0.120 \text{ MeV} . \quad (19.13)$$

Moreover, via (19.12) we also now use  $\vartheta'_{II\downarrow 31} = 8.795^\circ$  and  $\vartheta'_{I\downarrow 21} = 32.39^\circ$  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:

$$\boxed{\begin{array}{l} \theta_{P12} \equiv \vartheta'_{I\downarrow 21} = 32.39^{+2.01}_{+0.47} \text{ } ^{+2.83}_{-0.25} \text{ } ^{+3.66}_{-0.97} \text{ }^\circ \\ \theta_{P13} \equiv \vartheta'_{II\downarrow 31} = 8.795^{+0.050}_{-0.396} \text{ } ^{+0.195}_{-0.546} \text{ } ^{-0.096}_{-0.696} \text{ }^\circ \end{array}} \quad (19.14)$$

This is another way of showing that each of these is about  $1.67\sigma$  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 (19.11): Yes, the vev in (19.3) does allow *both* of  $\vartheta_{I\downarrow 21}$  and  $\vartheta_{II\downarrow 31}$  to fit the empirical data in (19.10), within about  $1.67\sigma$  for each, as precisely shown in (19.14). But now we have a seemingly-disconnected vev in (19.13), and this brings us to the second question whether this can be connected to other known data of independent origins. Because  $\frac{1}{\sqrt{2}}u'_{\downarrow} \equiv 1925.047 \pm 0.120$  MeV in (19.13) no longer is set by  $m_e c^2 + m_{\mu} c^2 + m_{\tau} c^2$  since it differs from this by  $\delta_{\downarrow} = 42.018$  MeV, the most obvious energy of comparison for (19.13) is the Fermi vev  $v_{\uparrow} = v = 246.2196508 \pm 0.0000633$  GeV given in (15.11). So, we simply calculate the ratio of these, and find that:

$$v_{\uparrow} / \frac{1}{\sqrt{2}}u'_{\downarrow} = 127.9032 \pm 0.0080. \quad (19.15)$$

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  $\alpha$ ] is  $\sim 1/128$ ,” see note <sup>†</sup> in PDG’s [21]. The closeness of (19.15) to this other empirical data raises the question whether  $\frac{1}{\sqrt{2}}u'_{\downarrow} = v_{\uparrow} \alpha(Q^2 = M_w^2 c^4)$  may be another relationship of genuine physical meaning. So, let us review the evidence in support:

First, the angles (19.14) originating in (19.10) 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  $e$ ,  $\mu$  and  $\tau$  leptons are the quintessential units of charge for the which interaction strength is set by  $\alpha(Q^2 = 0) = 1/137.035999139(31)$  in the low energy (fine structure constant) limit, and in general by the running  $\alpha(Q^2)$ . Fourth, because the  $e$ ,  $\mu$  and  $\tau$  leptons and the  $W^{\pm}$  boson both carry electric charge,  $\alpha(Q^2)$  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  $\alpha(M_w^2 c^4) \sim 1/128$ . Consequently, the unanticipated appearance of the number 127.9032 in (19.15) 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:

$$\frac{1}{\sqrt{2}}u'_{\downarrow} = m_{\tau}c^2 + m_{\mu}c^2 + m_e c^2 + \delta_{\downarrow} = \frac{1}{127.9032 \pm 0.0080} v_{\uparrow} = 1925.047 \pm 0.120 \text{ MeV} \equiv \alpha(M_w^2) v_{\uparrow}. \quad (19.16)$$

In the process, we tighten our knowledge of  $\alpha(M_w^2)$  to  $\alpha(M_w^2) = 1/(127.9032 \pm 0.0080)$ . 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  $\alpha(M_w^2)$ . Referring to (19.11), this means that:

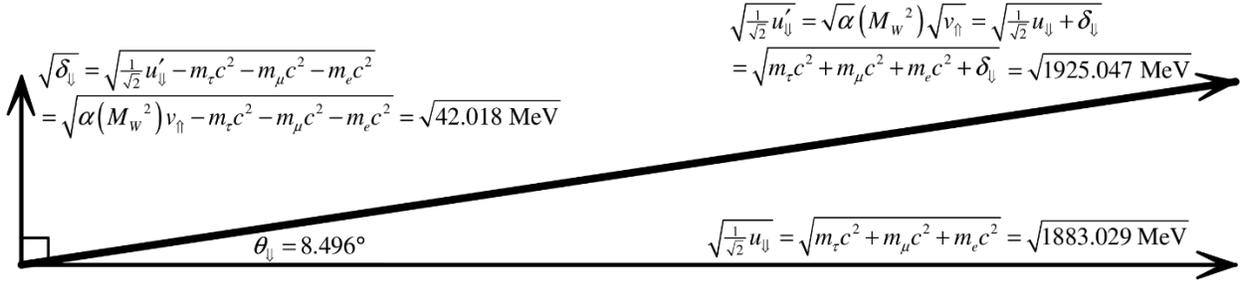
$$\delta_{\downarrow} = \frac{1}{\sqrt{2}}u'_{\downarrow} - m_{\tau}c^2 - m_{\mu}c^2 - m_e c^2 = \alpha(M_w^2) v_{\uparrow} - m_{\tau}c^2 - m_{\mu}c^2 - m_e c^2 = 42.018 \text{ MeV}. \quad (19.17)$$

Here,  $\delta_{\downarrow}$  no longer needs to be expressed as the energy difference which allows each of  $\vartheta_{l\downarrow 21}$  and  $\vartheta_{ll\downarrow 31}$  to fit the PMNS data in (19.10). 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 (19.17), namely, the charged lepton mass sum  $m_{\tau} + m_{\mu} + m_e$ , the Fermi vev, and the running  $\alpha(M_w^2)$  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:

$$m_{\tau}c^2 + m_{\mu}c^2 + m_e c^2 = \alpha(M_w^2) v_{\uparrow} - \delta_{\downarrow}, \quad (19.18)$$

wherein the mass sum  $m_{\tau} + m_{\mu} + m_e$  is seen to be a function of the independently-known parameters  $\alpha(M_w^2)$  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 2 through 4, we see that (19.17) lends itself to a geometric representation in the manner of Figure 4, with  $\sqrt{\frac{1}{\sqrt{2}}u'_{\downarrow}} = \sqrt{\alpha(M_w^2)} \sqrt{v_{\uparrow}} = \sqrt{m_{\tau}c^2 + m_{\mu}c^2 + m_e c^2 + \delta_{\downarrow}}$  on the hypotenuse, and with  $\sqrt{\delta_{\downarrow}}$  and  $\sqrt{\frac{1}{\sqrt{2}}u'_{\downarrow}} = \sqrt{m_{\tau}c^2 + m_{\mu}c^2 + m_e c^2}$  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,  $\sqrt{\alpha(M_w^2)} = \sqrt{k_e / \hbar c} e(M_w^2)$  is a direct measure of the electric charge strength at the beta decay vertex. Based on the numeric values from (19.13), (19.17) and (19.5), the small angle which we refer to as the charged lepton rotation angle and denote as  $\theta_l$ , has a value of  $\theta_l = 8.496^\circ$ . This may be illustrated as shown below:



**Figure 9: 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  $\sqrt{\frac{1}{\sqrt{2}}u_\downarrow} = \sqrt{m_\tau c^2 + m_\mu c^2 + m_e c^2}$  vector which is purely a function of the charged lepton masses, through an angle  $\theta_\downarrow = 8.496^\circ$ , into  $\sqrt{\frac{1}{\sqrt{2}}u'_\downarrow} = \sqrt{m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_\downarrow}$  which is a function of the charged lepton masses as well as  $\delta_\downarrow$ . While it also happens that  $\frac{1}{\sqrt{2}}u'_\downarrow = \alpha(M_w^2)v_\uparrow$  from (19.16), again, it will be important to acquire independent knowledge about  $\delta_\downarrow$ .

It is also very helpful to obtain mass relationships analogous to (14.15) and (15.13) which directly relate the charge lepton masses particularly to the two angles in (19.14). Solving the simultaneous equations which are (19.12), then using (19.14), for the tau and mu leptons we obtain:

$$m_\tau c^2 = \frac{1}{\sqrt{2}}u'_\downarrow \frac{\cos^2 \theta_{P13} \sin^2 \theta_{P12}}{1 - \cos^2 \theta_{P13} \cos^2 \theta_{P12}}; \quad m_\mu c^2 = \frac{1}{\sqrt{2}}u'_\downarrow \frac{\sin^2 \theta_{P13} \cos^2 \theta_{P12}}{1 - \cos^2 \theta_{P13} \cos^2 \theta_{P12}}. \quad (19.19a)$$

But because of the rotation (19.17) illustrated in Figure 9, the electron mass is *not* a direct function of these angles. For this mass, we need to use (19.3) and (19.2), then use (19.19a), to deduce:

$$\begin{aligned} m_e c^2 &= m_\tau c^2 \tan^2 \vartheta_{II\downarrow 31} = m_\mu c^2 \tan^2 \vartheta_{I\downarrow 21} \\ &= \frac{1}{\sqrt{2}}u'_\downarrow \frac{\cos^2 \theta_{P13} \sin^2 \theta_{P12}}{1 - \cos^2 \theta_{P13} \cos^2 \theta_{P12}} \tan^2 \vartheta_{II\downarrow 31} = \frac{1}{\sqrt{2}}u'_\downarrow \frac{\sin^2 \theta_{P13} \cos^2 \theta_{P12}}{1 - \cos^2 \theta_{P13} \cos^2 \theta_{P12}} \tan^2 \vartheta_{I\downarrow 21}. \end{aligned} \quad (19.19b)$$

We 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  $m_\tau, m_\mu, m_e$ . 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 (19.11) 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 (19.16) that this sum  $m_\tau c^2 + m_\mu c^2 + m_e c^2 + \delta_\downarrow = \alpha(M_w^2)v_\uparrow$  can be related within experimental errors to the

Fermi vev  $v_{\uparrow} \equiv v$  by the strength  $\alpha(M_W^2) = 1/(127.9032 \pm 0.0080)$  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:

$$m_\tau, m_\mu, m_e \mapsto m_\tau, m_\mu, m_e, \delta_{\downarrow} = F(\theta_{P12}, \theta_{P13}, \alpha(M_W^2), \delta_{\downarrow}). \quad (19.20)$$

In this way, we have now reparameterized all three charged lepton masses  $m_\tau, m_\mu, m_e$  over to  $\theta_{P12}, \theta_{P13}, \alpha(M_W^2)$ , but only by adding a new energy  $\delta_{\downarrow}$ . Taken together with (17.1) for the quark masses, and seeing in (19.20) how  $m_e$  is now included in the charged lepton mass reparameterization, all told we have now reparameterized:

$$\{m_u, m_c, m_t, m_d, m_s, m_b, m_e, m_\mu, m_\tau, \delta_{\downarrow}\} = F(v, m_h, \theta_{C31}, \theta_{C23}, \theta_{C21}, \theta_{P12}, \theta_{P13}, \alpha(M_W^2), \delta_{\downarrow}). \quad (19.21)$$

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  $v, m_h, \theta_{C31}, \theta_{C23}, \theta_{C21}, \theta_{P12}, \theta_{P13}, \alpha(M_W^2)$ , 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  $\{m_{\nu_e}, m_{\nu_\mu}, m_{\nu_\tau}, \delta_{\downarrow}\}$  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*.

## 20. Theory of Fermion 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 [57], 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 [58], “somewhere between 10 meV and 2eV is our playground.” And on page 12 of PDG’s 2018 review [57], 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 [57], “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  $v = v_\eta = 246.2196508 \pm 0.0000633$  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 (19.21) 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 2, 3, 4 and 9 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 (19.17) 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 (19.21) 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:

$$\delta_\downarrow \sqrt{\frac{v}{M_p c^2}} = 42.018 \text{ MeV} \times 4.492431 \times 10^{-09} = 0.18876 \text{ eV} . \quad (20.1)$$

*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  $\sim 2$  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 (20.1) to the neutrino masses.

In (14.4), (15.11) and (19.15), the vevs in relation the respective mass sums are  $v_{\uparrow} = \sqrt{2}(m_u c^2 + m_c c^2 + m_t c^2)$ ,  $v_{\downarrow} = \sqrt{2}(m_d c^2 + m_s c^2 + m_b c^2)$  and  $u_{\downarrow} = \sqrt{2}(m_{\tau} c^2 + m_{\mu} c^2 + m_e c^2)$ . So, for the neutrino sum we likewise define  $u_{\uparrow} \equiv \sqrt{2}(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2$ . The question now is whether the numeric result 0.18876 eV in (20.1) should be assigned to this new  $u_{\uparrow}$  or to the mass sum  $(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2$ . That is, where do we use the  $\sqrt{2}$  factor? Given that for the neutrinos,  $\Sigma_j m_j = (m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2 < 0.170$  eV with a 95% confidence level, this empirical data suggests that the appropriate assignment should be to the neutrino vev, namely:

$$u_{\uparrow} \equiv \sqrt{2}(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2 = \delta_{\downarrow} \sqrt{v / M_p c^2} = 0.18876 \text{ eV}, \quad (20.2a)$$

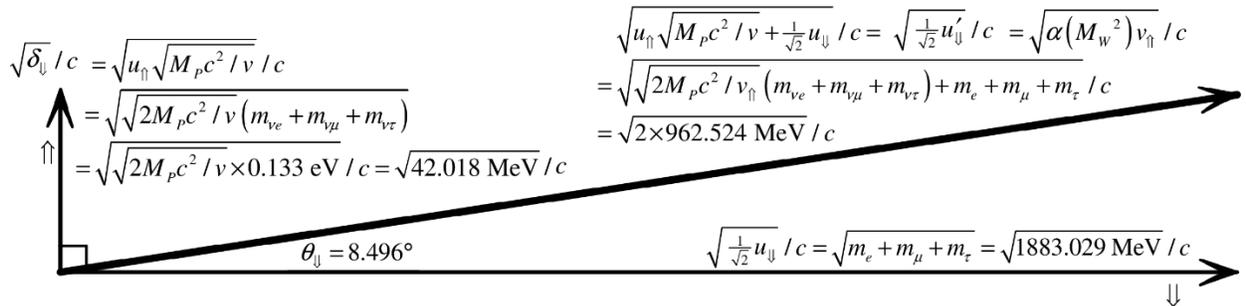
which means that for the neutrino mass sum we have:

$$\boxed{\frac{1}{\sqrt{2}}u_{\uparrow} \equiv (m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2 \equiv \frac{1}{\sqrt{2}}\delta_{\downarrow} \sqrt{v / M_p c^2} = 0.13348 \text{ eV} < 0.170 \text{ eV}}, \quad (20.2b)$$

clearly fitting the empirical data in [57]. Were we to assign  $(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2 = 0.189$  eV we would be somewhat-outside the 95% zone. It also helps to write the above in terms of  $\delta_{\downarrow}$  as:

$$\delta_{\downarrow} = u_{\uparrow} \sqrt{M_p c^2 / v} = \sqrt{2}(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2 \sqrt{M_p c^2 / v} = 42.018 \text{ MeV}. \quad (20.2c)$$

The above (20.2) 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 (14.4), (15.11) and (19.5) for quarks and the charged leptons. And, with (20.2c), we now have an independent understanding of  $\delta_{\downarrow} = \sqrt{2M_p c^2 / v}(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2$ , 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 (20.2c) we may update Figure 9 to display this new understanding, as seen below:



**Figure 10: 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  $\hat{\uparrow}$  and  $\hat{\downarrow}$ . Except for the “neutrino mass amplifier” factor  $\sqrt{M_p c^2 / \nu}$ , and the hypotenuse aligned toward isospin-down rather than up, this is identical in form to Figure 4 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 (20.2). Note also that the hypotenuse mirrors the Higgs mass relation  $2m_h = (v_{\uparrow} + \frac{1}{\sqrt{2}}v_{\downarrow})/c^2$  (16.5) 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 (19.11) – 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  $\sqrt{M_p c^2 / \nu}$  amplification.

At this point, having a predicted value  $(m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}})c^2 = 0.13348$  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 (19.1). 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 [56], 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 [56] contains both  $1\sigma$  and  $3\sigma$  data, as we did with the PMNS data, we interpolate that the  $2\sigma$  data is substantially equal to the  $1\sigma$  and  $3\sigma$  average. Accordingly, with  $l=1,2$ , the normal ordering data in [56] may be characterized by:

$$\begin{aligned} \Delta m_{21}^2 &= m_2^2 - m_1^2 = 7.40 \begin{matrix} +0.21 & +0.42 & +0.62 \\ -0.20 & -0.40 & -0.60 \end{matrix} \times 10^{-5} \text{ eV}^2 / c^4 \\ \Delta m_{3l}^2 &= m_3^2 - m_l^2 = 2.494 \begin{matrix} +0.033 & +0.066 & +0.099 \\ -0.031 & -0.063 & -0.095 \end{matrix} \times 10^{-3} \text{ eV}^2 / c^4 \end{aligned} \quad (20.3)$$

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 (20.3) 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 (20.3) we also use  $m_{\nu_e} + m_{\nu_{\mu}} + m_{\nu_{\tau}} = 0.13348$  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  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  spreads in (20.3). 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  $m_1 + \sqrt{\Delta m_{21}^2 + m_1^2} + \sqrt{\Delta m_{32}^2 + m_2^2 + \Delta m_{21}^2} = 0.13348 \text{ eV} / c^2$ , 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  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  spreads as follows:

$$\begin{array}{l} m_{\nu_e} c^2 = 0.03533 \begin{array}{l} +0.00012 \quad +0.00024 \quad +0.00035 \\ -0.00012 \quad -0.00025 \quad -0.00037 \end{array} \text{ eV} \\ m_{\nu_\mu} c^2 = 0.03637 \begin{array}{l} +0.00009 \quad +0.00017 \quad +0.00026 \\ -0.00009 \quad -0.00018 \quad -0.00027 \end{array} \text{ eV} \\ m_{\nu_\tau} c^2 = 0.06178 \begin{array}{l} -0.00020 \quad -0.00041 \quad -0.00062 \\ +0.00021 \quad +0.00043 \quad +0.00064 \end{array} \text{ eV} \end{array} \quad (20.4)$$

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  $3\sigma$  the tau generation rest energy  $m_{\nu_\tau} c^2 > 0.03610 \text{ eV}$  while the electron generation rest energy  $m_{\nu_e} c^2 < 0.03569 \text{ eV}$  is smaller by at least  $0.00041 \text{ eV}$ . This also highlights how (20.3) 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 [58] that “somewhere between 10 meV and 2eV is our playground,” we see that with the lightest neutrino mass predicted to be  $m_{\nu_e} c^2 \cong 35.33 \text{ meV}$  and a predicted mass sum  $\Sigma m \cong 0.13348 \text{ eV} / c^2$  versus the empirical constraint  $\Sigma_j m_j < 0.170 \text{ eV}$ , we have indeed landed right where we need to be in the “playground.” Also, we have obtained (20.4) 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 19. 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 (14.8) is then subjected to a bi-unitary transformation leading to relations which mirror (19.8) for the charged leptons. Specifically, borrowing the top two relations for the type *I* “downward cascade” parameterization in (19.8a) and for the type *II* “distribution” parameterization in (19.8b), and migrating  $\Downarrow$  to  $\Uparrow$  and the charged leptons to their neutrino partners, we write:

$$c_{I\uparrow 32}^2 = G_{\nu\tau} = \frac{m_{\nu\tau} c^2}{\frac{1}{\sqrt{2}} u_{\uparrow}} = \frac{m_{\nu\tau}}{m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}} \quad (20.5a)$$

$$c_{I\uparrow 21}^2 = \frac{G_{\nu\mu}}{s_{I\uparrow 32}^2} = \frac{m_{\nu\mu} c^2}{\frac{1}{\sqrt{2}} u_{\uparrow} - m_{\nu\tau} c^2} = \frac{m_{\nu\mu}}{m_{\nu e} + m_{\nu\mu}}$$

$$c_{II\uparrow 31}^2 = \frac{G_{\nu\tau}}{c_{II\uparrow 32}^2} = \frac{m_{\nu\tau} c^2}{\frac{1}{\sqrt{2}} u_{\uparrow} - m_{\nu\mu} c^2} = \frac{m_{\nu\tau}}{m_{\nu e} + m_{\nu\tau}} \quad (20.5b)$$

$$s_{II\uparrow 32}^2 = G_{\nu\mu} = \frac{m_{\nu\mu} c^2}{\frac{1}{\sqrt{2}} u_{\uparrow}} = \frac{m_{\nu\mu}}{m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}}$$

Then, as in (19.9), we simply use (20.2b) and the  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  in (20.4) to calculate each of these angles, as similarly to (19.14), to be:

$$\begin{aligned} \vartheta_{I\uparrow 32} &= 0.8226_{-0.0016}^{+0.0015} \quad +0.0031 \quad 0.0046_{-0.0048} \quad \text{rad} = 47.131_{-0.274}^{+0.087} \quad +0.177 \quad +0.266 \quad \circ \\ \vartheta_{I\uparrow 21} &= 0.7782_{-0.0002}^{+0.0003} \quad +0.0005 \quad +0.0008_{-0.0007} \quad \text{rad} = 44.585_{-0.039}^{+0.017} \quad +0.030 \quad +0.043 \quad \circ \\ \vartheta_{II\uparrow 31} &= 0.6475_{-0.0016}^{+0.0016} \quad +0.0032 \quad +0.0049_{-0.0049} \quad \text{rad} = 37.098_{-0.093}^{+0.092} \quad +0.185 \quad +0.278 \quad \circ \\ \vartheta_{II\uparrow 32} &= 0.5492_{-0.0008}^{+0.0007} \quad +0.0014 \quad +0.0022_{-0.0023} \quad \text{rad} = 31.466_{-0.045}^{+0.040} \quad +0.082 \quad +0.125 \quad \circ \end{aligned} \quad (20.6)$$

Now, at (19.14) 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 (19.10). The remaining real angle from (19.10) still to be fitted – presumably to the neutrino masses – is  $\theta_{p23} = 47.2_{-3.9}^{+1.9} \quad +3.1 \quad +4.3 \quad \circ$ . This is the least-tightly known of the three PMNS angles, varying even at  $1\sigma$  from  $43.3^\circ < \theta_{p23} < 49.1^\circ$ . So, in (20.6) there are actually two angles –  $\vartheta_{I\uparrow 32}$  and  $\vartheta_{I\uparrow 21}$  – which fit within  $1\sigma$  and so can be associated with the remaining angle  $\theta_{p23}$ . 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 (20.6). First for the quarks, among what was calculated leading to (14.11) were what we would now denote as  $c_{I\uparrow 21}^2 = G_c / s_{I\uparrow 32}^2$  and  $c_{II\uparrow 31}^2 = G_t / c_{II\uparrow 32}^2$  which led to the connections  $\theta_{I\uparrow 21} \equiv \theta_{C23} = 2.415 \pm 0.053^\circ$  and  $\theta_{II\uparrow 31} \equiv \theta_{C13} = 0.209_{-0.013}^{+0.015}^\circ$  in (14.13) to two of the three CKM angles within experimental errors. Then, leading to (15.3) among what was calculated were  $c_{I\downarrow 21}^2 = G_s / s_{I\downarrow 32}^2$  and  $c_{II\downarrow 31}^2 = G_b / c_{II\downarrow 32}^2$  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 (19.7) the calculations included  $c_{I\downarrow 21}^2 = G_\mu / s_{I\downarrow 32}^2$  and  $c_{II\downarrow 31}^2 = G_\tau / c_{II\downarrow 32}^2$ . After then having to introduce an energy difference  $\delta_{\downarrow}$  at (19.11) which as later shown in (20.1) and (20.2) is actually related to an amplified neutrino mass sum, we

calculated (19.12) which at (19.14) led to the connections  $\theta_{P12} \equiv \vartheta_{I\downarrow 21} = 32.39^{+2.01}_{+0.47} \text{ }^{+2.83}_{-0.25} \text{ }^{+3.66}_{-0.97} \circ$  and  $\theta_{P13} \equiv \vartheta_{II\downarrow 31} = 8.795^{+0.050}_{-0.396} \text{ }^{+0.195}_{-0.546} \text{ }^{+0.096}_{-0.696} \circ$  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 ( $G_c, G_s, G_\mu$ ) 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 ( $G_t, G_b, G_\tau$ ) 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  $c_{II\downarrow 31}^2 = G_b / c_{II\downarrow 32}^2$ .

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 (19.7) and (19.8) for guidance, it appears that  $c_{I\uparrow 21}^2 = G_{\nu\mu} / s_{I\uparrow 32}^2$  in (20.5a) (second generation, type-I, inverse sine-squared) is what should be connected to the final PMNS angle, and that  $c_{II\uparrow 31}^2 = G_{\nu\tau} / c_{II\uparrow 32}^2$  in (20.5b) (third generation, type II, inverse cosine-squared) should be regarded as the lepton “leftover.” Accordingly, we now formally connect  $\vartheta_{I\uparrow 21}$  in (20.6) to the remaining mixing angle  $\theta_{P23}$ , and regard  $\vartheta_{II\uparrow 31}$  in (20.6) as the leftover angle for leptons. Following a presentation form similar to what was used in (15.6) for quarks, we combine this with (19.14) whereby all three PMNS angles plus the lepton leftover are now related to the mass matrix mixing angles by:

$$\begin{array}{l} \theta_{P23} \equiv \vartheta_{I\uparrow 21} = 44.585^{+0.017}_{-0.039} \text{ }^{+0.030}_{-0.025} \text{ }^{+0.043}_{-0.039} \circ \\ \vartheta_{II\uparrow 31} = 37.098^{+0.092}_{-0.093} \text{ }^{+0.185}_{-0.188} \text{ }^{+0.278}_{-0.283} \circ \\ \theta_{P12} \equiv \vartheta_{I\downarrow 21} = 32.39^{+2.01}_{+0.47} \text{ }^{+2.83}_{-0.25} \text{ }^{+3.66}_{-0.97} \circ \\ \theta_{P13} \equiv \vartheta_{II\downarrow 31} = 8.795^{+0.050}_{-0.396} \text{ }^{+0.195}_{-0.546} \text{ }^{+0.096}_{-0.696} \circ \end{array} \quad (20.7)$$

From the second line of (20.5a), 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  $\theta_{P23} = 44.585^{+0.017}_{-0.039} \text{ }^{+0.030}_{-0.025} \text{ }^{+0.043}_{-0.039} \circ$  is tighter than the usual  $\theta_{P23} = 47.2^{+1.9}_{-3.9} \text{ }^{+3.1}_{-5.4} \text{ }^{+4.3}_{-6.9} \circ$  from [56], because it is rooted in the square-mass differences (20.3) which have been measured with tighter precision than  $\theta_{P23}$  directly. With the usual  $\theta_{P23}$  having a large error range especially on the low side, this new center at  $\theta_{P23} = 44.585^\circ$  is actually only at about  $.67\sigma$  below the usual  $\theta_{P23} = 47.2^\circ$  center. Moreover, the top-to-bottom  $3\sigma$  spread in this new valuation is a mere  $0.082^\circ$ , versus the usual  $11.2^\circ$  spread. Thus, this new valuation is about 135 times as precise at  $3\sigma$ , providing ample opportunity for experimental confirmation as it becomes possible to obtain more precise direct measurements of  $\theta_{P23}$ . This is why we are able to add two digits after the decimal in the new valuation of  $\theta_{P23}$  in (20.7). 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 (20.7) which have now been established, as well as the leftover angles, what we have now found is that within experimental errors we may associate:

$$\begin{aligned}
 \theta_{C12} &\equiv \theta_{I\downarrow 12}; & \theta_{P12} &\equiv \vartheta'_{I\downarrow 12} \\
 \theta_{C23} &\equiv \theta_{I\uparrow 12}; & \theta_{P23} &\equiv \vartheta'_{I\uparrow 12} \\
 \theta_{C13} &\equiv \theta_{II\uparrow 31}; & \theta_{P13} &\equiv \vartheta'_{II\downarrow 31} \\
 \text{leftover: } &\theta_{II\downarrow 31}; & \vartheta'_{II\uparrow 31}
 \end{aligned} \tag{20.8}$$

Finally, similarly to what we did at (14.15), (15.13) and (19.19), we may solve the simultaneous equations (20.5) and apply (20.7), and define a coupling  $G_\nu \equiv m_\nu c^2 / \frac{1}{\sqrt{2}} u_\uparrow$  for each neutrino type, to obtain:

$$G_{\nu\tau} = \frac{\cos^2 \vartheta'_{II\uparrow 31} \sin^2 \theta_{P23}}{1 - \cos^2 \vartheta'_{II\uparrow 31} \cos^2 \theta_{P23}}; \quad G_{\nu\mu} = \frac{\sin^2 \vartheta'_{II\uparrow 31} \cos^2 \theta_{P23}}{1 - \cos^2 \vartheta'_{II\uparrow 31} \cos^2 \theta_{P23}}; \quad G_{\nu e} = G_{\nu\mu} \tan^2 \theta_{P23} = G_{\nu\tau} \tan^2 \vartheta'_{II\uparrow 31}. \tag{20.9}$$

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.

## 21. Prediction of a Second Leptonic Higgs Boson, and its Mass

Back at (16.5) we showed how the mass of the Higgs boson can be described within experimental errors by  $m_h c^2 \equiv (v_\uparrow + \frac{1}{\sqrt{2}} v_\downarrow) / 2$  to the Fermi vev  $v = v_\uparrow = \sqrt{2} (m_u c^2 + m_c c^2 + m_t c^2)$  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 (15.11). And in Figure 4, 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 (19.5) and  $u_\uparrow \equiv \sqrt{2} (m_{\nu e} + m_{\nu\mu} + m_{\nu\tau}) c^2$  in (20.2a) for the leptons, we can likewise plot out a lepton analog to Figure 4 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 10, but it would lack the  $\sqrt{M_p c^2} / v$  amplifier, and so the angle corresponding to  $\theta_l = 8.496^\circ$  in Figures 9 and 10 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 (19.5) and (20.2a), 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  $\theta = 5.73653 \times 10^{-4} \text{ }^\circ$  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  $\sqrt{u_\uparrow + \frac{1}{\sqrt{2}} u_\downarrow} / c \equiv \sqrt{\frac{1}{\sqrt{2}} u_\downarrow}$  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  $m_{h_2}$  defined analogously to (16.5) by

$$\boxed{m_{h_2}c^2 \equiv \frac{u_\uparrow + \frac{1}{\sqrt{2}}u_\downarrow}{2} \cong \frac{1}{2\sqrt{2}}u_\downarrow = 941.515 \pm 0.060 \text{ MeV}}, \quad (21.1)$$

where we have used  $\frac{1}{\sqrt{2}}u_\downarrow = 1883.029 \pm 0.120 \text{ MeV}$  from (19.5) to supply the empirical data. Also using (20.2a), because  $u_\uparrow / \frac{1}{\sqrt{2}}u_\downarrow = 1.005 \times 10^{-09}$  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  $M_N = 939.565413 \pm 0.000006 \text{ MeV}$  [48] 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 (19.21), and will be further by (21.15) 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 (15.10) and (15.10) for tighter top and strange quark masses, (16.5) for a tighter Higgs mass, (19.14) for re-centered  $\theta_{p12}$  and  $\theta_{p13}$  values, (19.16) for a tighter  $\alpha(M_w^2)$ , and (20.7) for a far-tighter  $\theta_{p23}$ . 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, (20.2b), (20.4) and (21.1) contain *outright predictions* of four mass values which at present are not known. In (20.4) we are now told exactly the energies at which to look for the three neutrino masses, and in (20.2b) their mass sum. And in (21.1) 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  $m_{\nu\tau}c^2 + m_{\nu\mu}c^2 + m_{\nu e}c^2 = 0.13348 \text{ eV}$  and the separate masses in (20.4). And of course, finding a second Higgs boson at  $m_{h_2}c^2 = 941.515 \pm 0.060 \text{ MeV}$ , 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 (19.21) 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 (19.21), we summarize the complete reparameterization of *all twelve fermion masses*, including “leftover” angles, by:

$$\begin{aligned} & \{m_t, m_c, m_u, m_b, m_s, m_d, m_\tau, m_\mu, m_e, m_{\nu\tau}, m_{\nu\mu}, m_{\nu e}\} \\ & = F\left(G, G_F, m_h, \alpha(M_W^2), \theta_{C12}, \theta_{C23}, \theta_{C31}, \theta_{II\downarrow 31}, \theta_{P12}, \theta_{P23}, \theta_{P13}, \vartheta_{II\uparrow 31}\right) \end{aligned} \quad (21.2)$$

In the above, we have momentarily included the leftover angles  $\theta_{II\downarrow 31}$  of (15.6) and  $\vartheta_{II\uparrow 31}$  of (20.7) because these explicitly appear in (15.13) for isospin-down quarks and in (20.9) for neutrinos (isospin-up leptons). However, these leftover angles are redundant, which we can see specifically via (15.3) and the lower (20.5a) together with the upper (20.5b). 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, 19 and 20 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 (21.2) 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 (21.2). To see this, start with (19.21) which contains  $\delta_\downarrow$  as an added parameter. We of course found in (20.2c) 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 (20.2b) which we may combine with (19.18) to obtain:

$$(m_{\nu e} + m_{\nu\mu} + m_{\nu\tau})c^2 = \frac{1}{\sqrt{2}} \sqrt{v/M_P c^2} \left( \alpha(M_W^2) v_\uparrow - (m_\tau + m_\mu + m_e)c^2 \right). \quad (21.3)$$

Knowing the parameters  $G_F$  thereby  $v_\uparrow = v$ , and  $\alpha(M_W^2)$  in (21.2), we can of course use (19.16) and (20.2c) 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  $G$ ,  $G_F$ ,  $\alpha(M_W^2)$ . Therefore, with (21.3), 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 (21.2), 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 lepton Higgs mass  $m_{h_2}c^2 \cong \frac{1}{2\sqrt{2}}u_{\downarrow} = 941.515 \pm 0.060$  MeV as a proxy for charged lepton mass sum (19.5) because  $u_{\uparrow}$  in (21.1) is empirically indiscernible by comparison given that the ratio  $u_{\uparrow} / \frac{1}{\sqrt{2}}u_{\downarrow} = 1.005 \times 10^{-09}$ , then (21.3) now becomes:

$$(m_{\nu_e} + m_{\nu_\mu} + m_{\nu_\tau})c^2 \cong \sqrt{v / M_p c^2} \left( \frac{1}{2} \alpha (M_W^2) v_{\uparrow} - m_{h_2} c^2 \right). \quad (21.4)$$

Therefore, we choose the aesthetics of  $m_\tau + m_\mu + m_e$ , then use the new  $m_{h_2}$  as a proxy for this sum, while removing the redundant, phased-away leftover angles from (21.2), to finally write:

$$\boxed{\left\{ m_t, m_c, m_u, m_b, m_s, m_d, m_\tau, m_\mu, m_e, m_{\nu_\tau}, m_{\nu_\mu}, m_{\nu_e} \right\}} \\ = F \left( G, G_F, m_h, m_{h_2}, \alpha (M_W^2), \theta_{C12}, \theta_{C23}, \theta_{C31}, \theta_{P12}, \theta_{P23}, \theta_{P13} \right). \quad (21.5)$$

Consequently, we have finally reparameterized all *twelve* fermion rest masses into *eleven* previously-disconnected parameters. But  $m_{h_2}$  is a proxy for  $m_\tau + m_\mu + m_e$  given that  $u_{\uparrow} / \frac{1}{\sqrt{2}}u_{\downarrow} = 1.005 \times 10^{-09}$  and so  $u_{\uparrow}$  can be neglected in (21.1). 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  $m_{h_2}$  in (21.5) – down to eleven parameters, removing eleven independent unknowns from our understanding of the natural world.

## 22. 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  $\phi_h$  by  $\varphi_h$ . Consequently, given that (11.3) contains the symmetry-broken  $\phi_h = \frac{1}{\sqrt{2}}\phi_{1h} = \frac{1}{\sqrt{2}}(v+h)$ , we replace this for leptons with  $\varphi_h = \frac{1}{\sqrt{2}}\varphi_{1h} = \frac{1}{\sqrt{2}}(u+h_2)$ . 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 (16.7) for charged leptons, versus the enormously-smaller  $u_{\uparrow} = 0.189$  eV from (16.9a) for neutrinos. For the leptons, plots similar to Figure 1 may be drawn, but with the vevs established by one of the two foregoing vevs, not the Fermi vev, see the discussion following (13.20) 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:

$$U(\varphi_{1h}) = \lambda_7 \left( -\frac{1}{2} u_{\downarrow}^2 \varphi_{1h}^2 + \frac{1}{4} \varphi_{1h}^4 \right) = -\frac{1}{4} m_{h_2}^2 c^4 \varphi_{1h}^2 + \frac{1}{8} \frac{m_{h_2}^2 c^4}{u_{\downarrow}^2} \varphi_{1h}^4 = m_{h_2}^2 c^4 \left( -\frac{1}{4} \varphi_{1h}^2 + \frac{1}{8} \frac{1}{u_{\downarrow}^2} \varphi_{1h}^4 \right). \quad (22.1)$$

$$U'(\varphi_{1h}) = \lambda_7 \varphi_{1h} (\varphi_{1h}^2 - u_{\downarrow}^2) = \frac{m_{h_2}^2 c^4}{2u_{\downarrow}^2} \varphi_{1h} (\varphi_{1h}^2 - u_{\downarrow}^2)$$

Continuing with the notational distinctions which are entirely of form, we also use  $\lambda_7$  in the above to denote this parameter as it applies to leptons, while  $m_{h_2} = 941.515 \pm 0.060 \text{ MeV} / c^2$  is the second leptonic Higgs mass discovered in (19.1). The only substantive change made in (22.1) 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 (22.1) 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  $U(\varphi_{1h})$  with higher-order terms so as to require two minima. One of these is to be centered at  $\varphi_{1h} = u_{\downarrow} = 2663.005 \text{ MeV}$  the charged leptons, and the other at  $\varphi_{1h} = u_{\uparrow} = 0.18876 \text{ eV}$  for the neutrinos. We also require the usual maximum at  $\varphi_{1h} = 0$  and a new, second maximum that is established using  $m_{h_2}$ . In establishing the second maximum in this way, we apply the same rationale for why we used (14.9) and (14.10) to establish the second quark maximum as reviewed following those two equations. So, starting with  $\varphi_h = \frac{1}{\sqrt{2}} \varphi_{1h} = \frac{1}{\sqrt{2}} (u + h_2)$  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 (14.9) we establish the maximum at the domain point where:

$$h_2(x^M) = -m_{h_2} c^2 = -941.515 \pm 0.060 \text{ MeV}, \quad (22.2)$$

and therefore, as in (14.10), also using the new Higgs mass in (19.1), where:

$$\varphi_{1h}(x^M) = u_{\downarrow} + h_2(x^M) = u_{\downarrow} - m_{h_2} c^2 = \left(1 - \frac{1}{2\sqrt{2}}\right) u_{\downarrow} - \frac{1}{2} u_{\uparrow} \cong \left(1 - \frac{1}{2\sqrt{2}}\right) u_{\downarrow} = 1721.491 \pm 0.110 \text{ MeV}. \quad (22.3)$$

Next, we follow suit from (14.11) to build in these minima and maxima by defining:

$$\begin{aligned} U' &= B \frac{m_{h_2}^2 c^4}{2u_{\downarrow}^2} \varphi_{1h} (\varphi_{1h}^2 - u_{\downarrow}^2) \left( \varphi_{1h}^2 - (u_{\downarrow} - m_{h_2} c^2)^2 \right) (\varphi_{1h}^2 - u_{\uparrow}^2) \\ &= B \frac{m_{h_2}^2 c^4}{2u_{\downarrow}^2} \left( -u_{\downarrow}^2 u_{\uparrow}^2 (u_{\downarrow} - m_{h_2} c^2)^2 \varphi_{1h} + (u_{\downarrow}^2 u_{\uparrow}^2 + (u_{\downarrow}^2 + u_{\uparrow}^2) (u_{\downarrow} - m_{h_2} c^2)^2) \varphi_{1h}^3 \right. \\ &\quad \left. - (u_{\downarrow}^2 + u_{\uparrow}^2 + (u_{\downarrow} - m_{h_2} c^2)^2) \varphi_{1h}^5 + \varphi_{1h}^7 \right). \end{aligned} \quad (22.4)$$

with an overall coefficient  $B$  that will be used to match the leading term in the upper (22.1). As in (14.12) we then integrate, and find we must set  $B=1/u_{\uparrow}^2(u_{\downarrow}-m_{h2}c^2)^2$  for the leading term to match (22.1). And, also to match, we discard the integration constant. Thus, we obtain:

$$U(\varphi_{1h}) = m_{h2}^2 c^4 \left( \begin{aligned} & -\frac{1}{4}\varphi_{1h}^2 + \frac{1}{8}\frac{1}{u_{\downarrow}^2}\varphi_{1h}^4 + \frac{1}{8}\left(\frac{1}{u_{\uparrow}^2} + \frac{1}{(u_{\downarrow}-m_{h2}c^2)^2}\right)\varphi_{1h}^4 \\ & -\frac{1}{12}\left(\frac{1}{u_{\downarrow}^2 u_{\uparrow}^2} + \frac{1}{(u_{\downarrow}-m_{h2}c^2)^2}\frac{u_{\downarrow}^2+u_{\uparrow}^2}{u_{\downarrow}^2 u_{\uparrow}^2}\right)\varphi_{1h}^6 + \frac{1}{16}\frac{1}{(u_{\downarrow}-m_{h2}c^2)^2}\frac{1}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^8 \end{aligned} \right). \quad (22.5)$$

Then we separate terms as in (14.13) and use the approximations in (19.1) and (22.3), thus:

$$\begin{aligned} U(\varphi_{1h}) &= m_{h2}^2 c^4 \left( -\frac{1}{4}\varphi_{1h}^2 + \frac{1}{8}\frac{u_{\downarrow}^2+u_{\uparrow}^2}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^4 - \frac{1}{12}\frac{1}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^6 \right) \\ &+ \frac{m_{h2}^2 c^4}{(u_{\downarrow}-m_{h2}c^2)^2} \left( \frac{1}{8}\varphi_{1h}^4 - \frac{1}{12}\frac{u_{\downarrow}^2+u_{\uparrow}^2}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^6 + \frac{1}{16}\frac{1}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^8 \right) \\ &\cong \frac{1}{8}u_{\downarrow}^2 \left( -\frac{1}{4}\varphi_{1h}^2 + \frac{1}{8}\frac{u_{\downarrow}^2+u_{\uparrow}^2}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^4 - \frac{1}{12}\frac{1}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^6 \right) \\ &+ \frac{1}{(1-2\sqrt{2})^2} \left( \frac{1}{8}\varphi_{1h}^4 - \frac{1}{12}\frac{u_{\downarrow}^2+u_{\uparrow}^2}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^6 + \frac{1}{16}\frac{1}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^8 \right) \end{aligned} \quad (22.6)$$

Because  $u_{\uparrow}/\frac{1}{\sqrt{2}}u_{\downarrow}=1.005\times 10^{-09}$ , see following (19.1), it is possible within experimental errors for the charged lepton masses to drop some unobservable terms and so further reduce the above to:

$$U(\varphi_{1h}) = -\frac{1}{32}u_{\downarrow}^2\varphi_{1h}^2 + \frac{1}{64}\frac{u_{\downarrow}^2}{u_{\uparrow}^2}\varphi_{1h}^4 - \left( \frac{1}{96} + \frac{1}{12(1-2\sqrt{2})^2} \right) \frac{1}{u_{\uparrow}^2}\varphi_{1h}^6 + \frac{1}{16(1-2\sqrt{2})^2}\frac{1}{u_{\downarrow}^2 u_{\uparrow}^2}\varphi_{1h}^8. \quad (22.7)$$

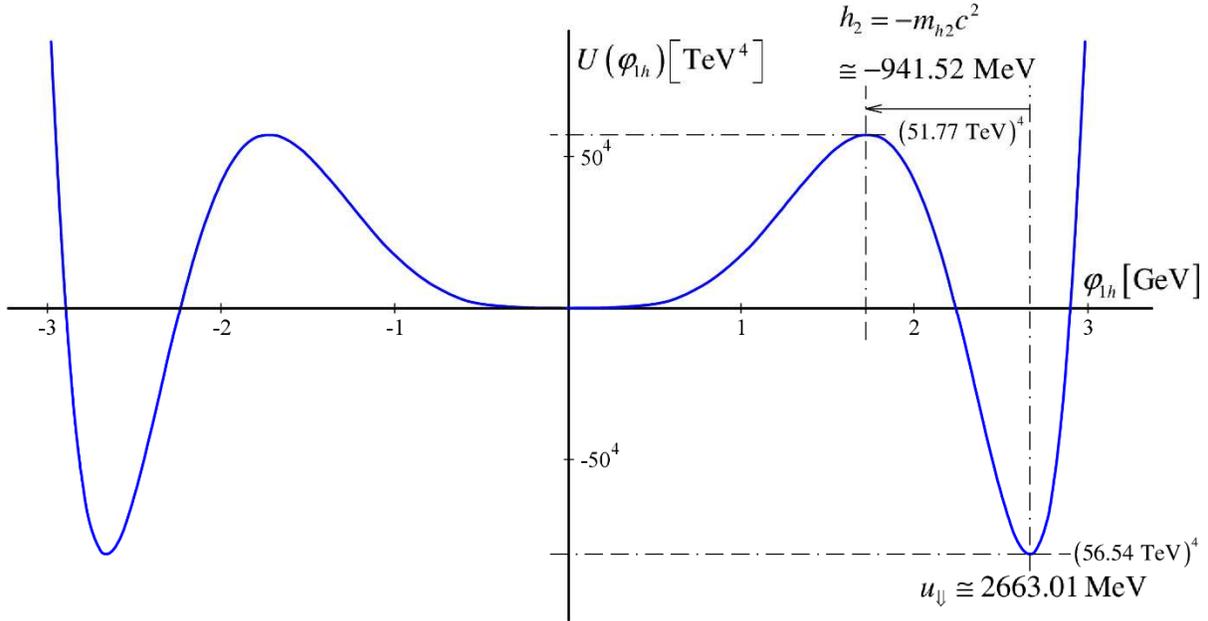
This sort of reduction has no analog for (14.13) because for quarks, the ratio  $v_{\uparrow}/\frac{1}{\sqrt{2}}v_{\downarrow}\cong 57.5635$ , see (13.20), whereby both  $v_{\uparrow}$  and  $v_{\downarrow}$  make all terms observable over at least some pertinent regions of the domain.

Then, as in (14.14), using the numerical values of  $u_{\downarrow}$  and  $u_{\uparrow}$  obtained from (16.7) and (16.9a), with  $\varphi_{1h}$  in both MeV and eV and thus  $U(\varphi_{1h})$  in MeV<sup>4</sup> and eV<sup>4</sup> respectively, we obtain:

$$\begin{aligned}
 U(\varphi_{1h})[\text{MeV}^4] &= -2.216 \times 10^5 \varphi_{1h}^2 + 3.102 \times 10^{18} \varphi_{1h}^4 - 9.894 \times 10^{11} \varphi_{1h}^6 + 7.380 \times 10^4 \varphi_{1h}^8 \\
 U(\varphi_{1h})[\text{eV}^4] &= -2.216 \times 10^{17} \varphi_{1h}^2 + 3.102 \times 10^{18} \varphi_{1h}^4 - 9.894 \times 10^{-1} \varphi_{1h}^6 + 7.380 \times 10^{-20} \varphi_{1h}^8.
 \end{aligned}
 \tag{22.8}$$

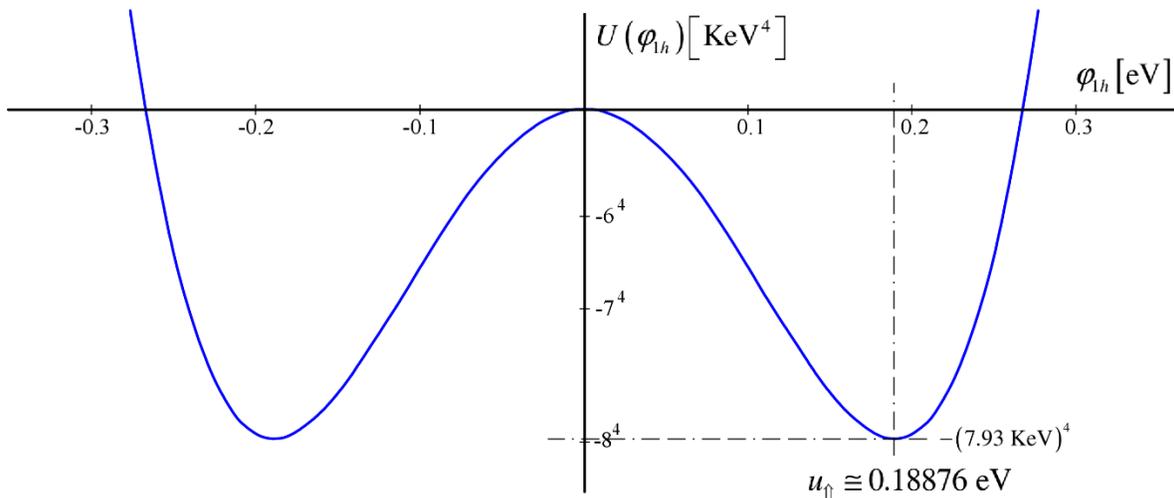
We show both MeV and eV because given the large chasm between the charged lepton and the neutrino vevs, the former is better for studying the charged lepton vev and the latter for studying the neutrino vev. As with (14.14) we may then draw plots of  $U(\varphi_{1h})$  similar to Figures 5 and 6, and may also draw fourth root plots similar Figures 7 and 8. The qualitative character of these plots is exactly the same as that of Figure 5 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 (22.2) and (22.3) much higher than their quark potential counterparts.

To see this in detail, we start with Figures 11, 12, 13 and 14 below which analogous and qualitatively-similar to Figures 5, 6, 7 and 8, but now for leptons not quarks. We see the minima and maxima of  $U(\varphi_{1h})$  at the domain points which were built in via (22.4). There are two primary quantitative contrasts with Figures 5 through 8: 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  $\varphi_{1h}$  – are far greater than what appears for the quarks in Figures 5 through 8. This is why the horizontal and vertical axes in the wide-view Figures 11 and 13 below are sized in GeV and TeV, while these same axes in the magnified center views of Figures 12 and 14 below are sized a billion times smaller in eV and KeV, and thus are magnified by a factor of a billion.



**Figure 11: Lagrangian Potential for Leptons – Wide View**

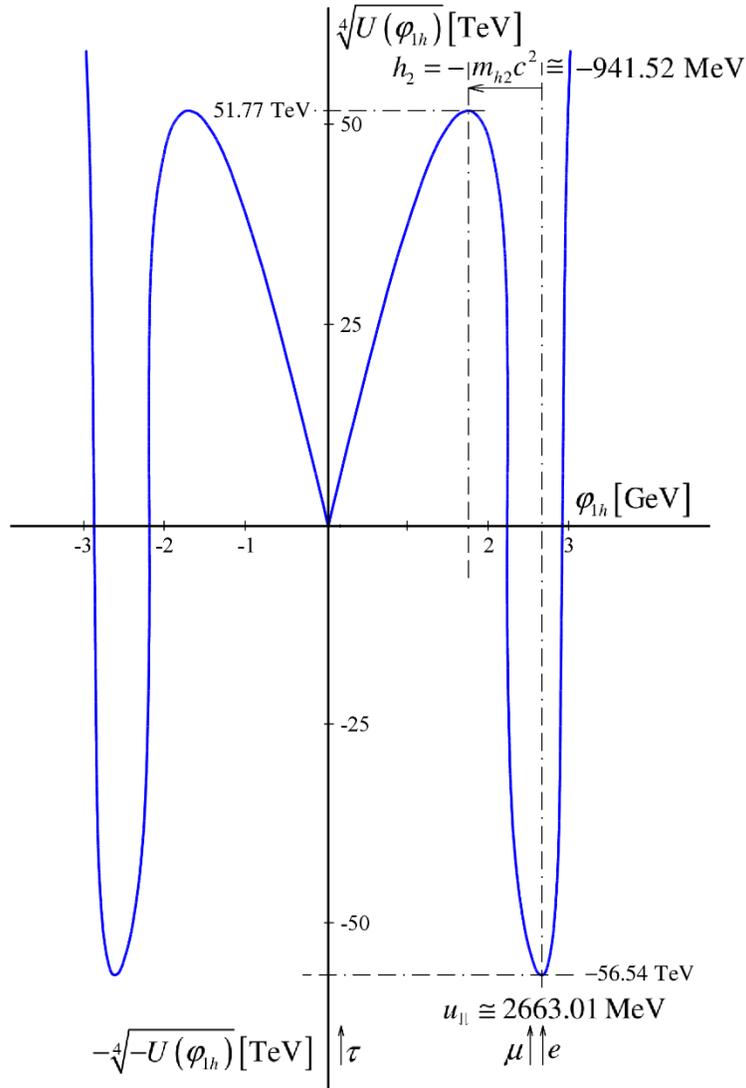
In Figure 11 above we see that the minimum at  $u_{\uparrow} \cong 2663.01 \text{ MeV}$  has an extremely large depth of  $U(u_{\uparrow}) \cong -(56.54 \text{ TeV})^4$ , and the maximum at  $h_2 = -m_{h_2}c^2 \cong -941.52 \text{ MeV}$  has an extremely large height of  $U(h_2 = -m_{h_2}c^2) \cong (51.77 \text{ TeV})^4$ . There is no possible way to visually represent the neutrino region of this plot, which is why we need the magnified figure below:



**Figure 12: Lagrangian Potential for Leptons – Magnified Center View**

In Figure 12 above, magnified by a factor of a billion over Figure 11, we see that the neutrino-well minimum at  $u_{\uparrow} \cong 0.18876 \text{ eV}$  also has – comparatively speaking – the extremely

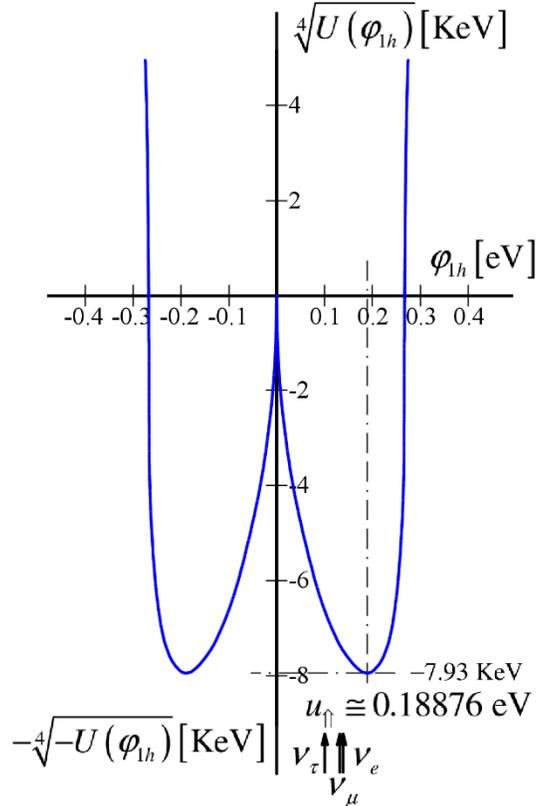
large depth  $U(u_{\downarrow}) \cong -(7.93 \text{ KeV})^4$ . But because  $U(\varphi_{1h})$  has dimensions of energy to the fourth power, we again take fourth roots as we did in Figure 7 and 9, so that we can compare energy-to-energy. Below, we take this fourth root on the vertical axis for Figure 11, as such:



**Figure 13: Lagrangian Potential for Leptons, Fourth Root – Wide View**

In this Figure 13, 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 Figure 1 (using  $u_{\downarrow}$

rather than  $\nu = \nu_{\uparrow}$  as the vev, see following (13.20)). 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 13 by a factor of a billion, we arrive at the Figure below:



**Figure 14: Lagrangian Potential for Leptons, Fourth Root – Magnified Center View**

In this final Figure 14 which is the fourth root of Figure 12, we see the neutrino portion of the potential. The vev minimum is at  $u_{\uparrow} \cong 0.18876 \text{ eV}$  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 13 and 14, similarly to the top quark behavior in Figure 8, that the tau lepton nests to the left of the peak set by (22.2) and (22.3), 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 (22.7), using  $u_{\downarrow} = 2663.005$  MeV from (16.7) and  $u_{\uparrow} = 0.18876$  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:

$$U(\varphi_{1h} = u_{\uparrow}) = -\frac{1}{64}u_{\downarrow}^2u_{\uparrow}^2 = -(7931.8 \text{ eV})^4, \quad (22.9a)$$

that the barrier between the two wells has a height of:

$$\begin{aligned} & U(\varphi_{1h} = (1 - \frac{1}{2\sqrt{2}})u_{\downarrow}) \\ &= (1 - \frac{1}{2\sqrt{2}})^4 \left[ \frac{1}{64} - \left( \frac{1}{96} + \frac{1}{12(1-2\sqrt{2})^2} \right) \left(1 - \frac{1}{2\sqrt{2}}\right)^2 + \frac{1}{16(1-2\sqrt{2})^2} \left(1 - \frac{1}{2\sqrt{2}}\right)^4 \right] \frac{u_{\downarrow}^2}{u_{\uparrow}^2} u_{\downarrow}^4 = (51.771 \text{ TeV})^4, \end{aligned} \quad (22.9b)$$

and that the charged lepton well depth is:

$$U(\varphi_{1h} = u_{\downarrow}) = \left[ \frac{1}{192} - \frac{1}{48(1-2\sqrt{2})^2} \right] \frac{u_{\downarrow}^2}{u_{\uparrow}^2} u_{\downarrow}^4 = -(56.537 \text{ TeV})^4. \quad (22.9c)$$

Again keeping in mind that  $U(\varphi_{1h} = u_{\uparrow})$  is quartic in energy, we see the mix of vev in  $u_{\downarrow}^2u_{\uparrow}^2$  in (22.9a) being responsible for the deep well in (22.9a) relative to the neutrino masses which are set exclusively by the much-smaller  $u_{\uparrow}$ . And in (22.9b) 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.

## PART IIC: COMPLETE THEORY OF WEAK BETA DECAY

### 23. How Weak Beta Decays are Triggered by Cosmological 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

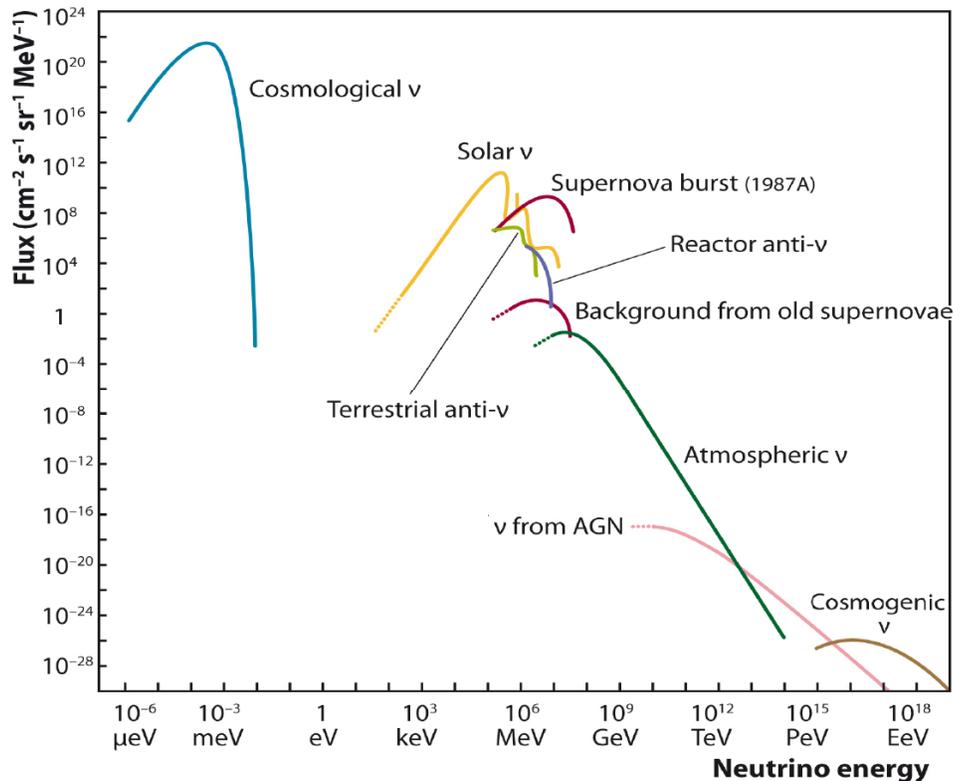
“visits” the isospin-down well), we showed using Figure 7 how it is necessary for any quark undergoing weak beta decay to cross the barrier at the domain point  $\phi_{1h} = v_{\uparrow} - m_h c^2 = 120.9712 \pm 0.0002$  GeV, and how this requires sufficient energy to clear the  $\sqrt[4]{V(\phi_{1h})} \cong 240.37$  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  $-\sqrt[4]{-V(\phi_{1h})} \cong -514.89$  GeV. At (14.4) 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 13 between the well depth in (22.9c) and the barrier height in (22.9b) 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-19<sup>th</sup> 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 why 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 isotope of an atom has beta-decayed after some elapsed time  $t$ , what was the cause 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 [59], which is reproduced below:



**Figure 15: Measured and expected fluxes of natural and reactor neutrinos, reproduced from Figure 1 of [59]**

The vertical axis above represents the *number flux* of neutrinos of various energies measured in neutrinos per  $\text{cm}^2$  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  $\text{MeV}^{-1}$  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  $\text{cm}^2$  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 15 that there is a peak flux of about  $10^{21}$  neutrinos per  $\text{cm}^2$  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.

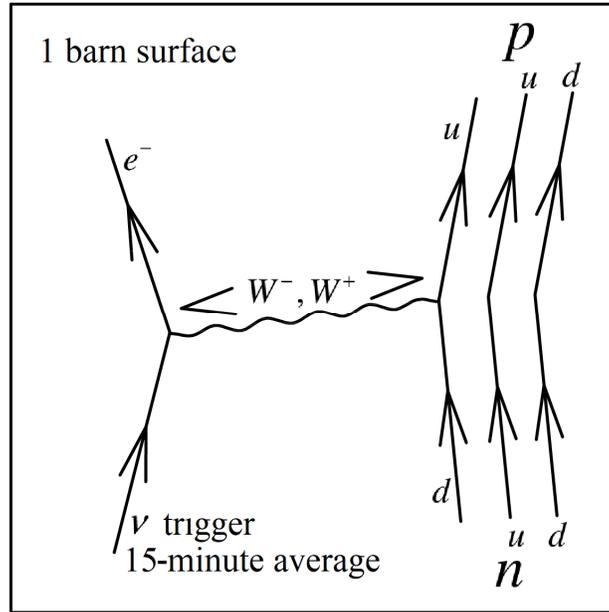
From this, let us do a rough “back of the envelope” calculation. To start, recognizing that the charge radii of the proton and neutron are roughly  $1 \text{ f} = 10^{-15} \text{ m}$ , let us regard 1 barn defined by  $1 \text{ b} = (10^{-14} \text{ m})^2 = (10 \text{ f})^2$  to be a very rough measure 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  $\text{cm}^2$ , the data just reviewed from Figure 15 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. [48], that the mean lifetime of a free neutron is  $880.2 \pm 1.0 \text{ s}$ , *which is also about 15 minutes*. So, the objective data tells us that every fifteen minutes, on average, one CvB neutrino busses 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 15, the answer is that it took an elapsed time  $t$  for one of the CvB 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  $\nu n \rightarrow e^- p$  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 16 below:



**Figure 16: Free neutron beta decay with neutrino trigger**

As is seen, this entails a CvB neutrino randomly entering the  $\sim 1$  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  $3 \times 10^{-25}$  s [60]. The net result following this very brief time period of  $3 \times 10^{-25}$  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 16 uses the  $\beta^-$  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 15 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  $\beta^-$  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  $\beta^+$  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  $5.29 \times 10^{-11}$  m, which we roughly approximate to an atomic *diameter* of  $10^{-10}$  m. So, whereas a barn with  $1 \text{ b} = (10^{-14} \text{ m})^2$  defines a measurement standard for *nuclear* cross sections, let us now define a “Bohr barn,” abbreviated bb, such that  $1 \text{ bb} = (10^{-10} \text{ m})^2$ . 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 pass 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  $\beta^-$  or  $\beta^+$  decay, and not by  $\alpha$  or  $\gamma$  decay or by merely jettisoning neutrons or protons. We start from the very-valuable Wikipedia Table of Nuclides [61] and use this to link over to the isotopes for various atoms through the  $Z$  atomic number listed above the symbolic name of each atom. In this way we find the following sampling of  $\beta^-$  decay data (without error bars), for the light nuclides from hydrogen (atomic number  $Z=1$ ) through sodium ( $Z=11$ ): For  $^3\text{H}$ , i.e. tritium, the half-life is about 12.32 y. For helium, for  $^6\text{He}$  the half-life is about 806.7 ms, while for  $^8\text{He}$  it is about 119.0 ms. For lithium, for  $^8\text{Li}$ ,  $^9\text{Li}$  and  $^{11}\text{Li}$ , the half-lives are 840.3 ms, 178.3 ms and 8.75 ms respectively. For beryllium, for  $^{10}\text{Be}$ ,  $^{11}\text{Be}$ ,  $^{12}\text{Be}$  and  $^{14}\text{Be}$ , the half-lives are  $1.39 \times 10^6$  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  $^{12}\text{B}$ , down to 2.92 ms for  $^{19}\text{B}$ , with consistent serial descent. For carbon, we of course have 5,730 years for  $^{14}\text{C}$  used in radioactive dating. Thereafter, the  $\beta^-$  half-lives range serially downward from a high of 2.449 s for  $^{15}\text{C}$  to 6.2 ms for  $^{16}\text{C}$ . For oxygen, there is a serial reduction from 26.464 s to 65 ms from  $^{19}\text{O}$  to  $^{24}\text{O}$ . For Fluorine, the range is serially-downward from 11.163 s to 2.6 ms from  $^{20}\text{F}$  to  $^{29}\text{F}$ , with very-mild exception at  $^{21}\text{F}$  and  $^{22}\text{F}$  which may be attributed simply to the growing complexity of the nuclide. For neon there is serial descent from 37.24 s for  $^{23}\text{Ne}$  to 3.5 ms for  $^{32}\text{Ne}$ , with a single exception at  $^{24}\text{Ne}$ . And going from the  $n=2$  shell to the  $n=3$  shell (principal quantum number, 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  $^{24}\text{Na}$ , down to 1.5 ms for  $^{35}\text{Na}$ , with mild exceptions attributable to nuclide complexity.

The same review for  $\beta^+$  decay evidences the following: for hydrogen there is no  $\beta^+$  decay channel. For helium there is a  $\beta^+$  decay channel for  $^2\text{He}$ , 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  $<0.01\%$   $\beta^+$  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  $\beta^+$  channels. So, we begin with boron, which is the first nucleus with a clear  $\beta^+$  channel, namely,  $^8\text{B}$  with a half-life of 770 ms. Turning to

carbon, for  $^{11}\text{C}$ ,  $^{10}\text{C}$  and  $^9\text{C}$ , the respective  $\beta^+$  decay half-lives are 20.334 min, 19.290 s and 126.5 ms, respectively. For nitrogen, we have  $^{13}\text{N}$  and  $^{12}\text{N}$  with the respective half-lives of 9.965 min and 11.000 ms. For oxygen we have  $^{15}\text{O}$ ,  $^{14}\text{O}$  and  $^{13}\text{O}$  with respective half-lives of 122.24 s, 70.598 s and 8.58 ms. For fluorine the two channels are for  $^{18}\text{F}$  and  $^{17}\text{F}$  with respective 109.771 min and 64.49 s half-lives. For neon the three isotopes with  $\beta^+$  channels are  $^{19}\text{Ne}$ ,  $^{18}\text{Ne}$  and  $^{17}\text{Ne}$  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  $\beta^+$  channels, namely,  $^{22}\text{Na}$ ,  $^{21}\text{Na}$  and  $^{20}\text{Na}$  with respective 2.6027 y, 22.49 s and 447.9 ms half-lives. Note that  $^{22}\text{Na}$  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  $\beta^-$  decay of  $^{35}\text{Na}$ . 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 [62]. 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  $\beta^-$  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  $\beta^-$  half-life is decreased. Third, for  $\beta^+$  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  $\beta^+$  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  $\beta^-$  correlation is broken by  $Z=12$  magnesium, wherein  $^{27}\text{Mg}$ ,  $^{28}\text{Mg}$  and  $^{29}\text{Mg}$  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  $\beta^+$  correlation is first broken by  $Z=17$  chlorine, wherein  $^{34}\text{Cl}$ ,  $^{33}\text{Cl}$  and  $^{32}\text{Cl}$  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  $\beta^-$  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  $\beta^-$  and  $\beta^+$ , in *any and all* atoms and atomic isotopes? In short, is it possible to understand all 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  $\beta^-$  or  $\beta^+$  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 photons 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  $3 \times 10^{-25}$  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  $I_3(f_R) = 0$  for the right-chiral projections  $f_R = \frac{1}{2}(1 + \gamma^5)f$  of all fermions  $f$ . Indeed, a very central finding in Part I of this paper as reviewed in section 9 is that the Dirac  $\gamma_5$  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  $\gamma_\mu$  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  $\eta_{MN} = \frac{1}{2}\{\gamma_M, \gamma_N\}$  for the flat spacetime Minkowski metric tensor  $\eta_{MN}$ .

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  $\mathfrak{M}_{em}(ee \rightarrow ee) \propto (Q(e)\bar{e}\gamma^\mu e)(Q(e)\bar{e}\gamma_\mu e) = +(\bar{e}\gamma^\mu e)(\bar{e}\gamma_\mu e)$ , while when an electron and a proton interact,  $\mathfrak{M}_{em}(ep \rightarrow ep) \propto Q(e)Q(p)(\bar{e}\gamma^\mu e)(\bar{p}\gamma_\mu p) = -(\bar{e}\gamma^\mu e)(\bar{p}\gamma_\mu p)$ . The overall positive sign for  $\mathfrak{M}_{em}(ee \rightarrow ee)$  indicates repulsion and the overall negative sign for  $\mathfrak{M}_{em}(ep \rightarrow ep)$  indicates attraction. Now let's turn to neutral currents mediated by Z bosons.

The weak neutral current, of course, has the form  $J_{NC}^\mu = J_3^\mu - J_{em}^\mu \sin^2 \theta_W$ , 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,  $f_R = \frac{1}{2}(1 + \gamma^5)f$  and  $f_L = \frac{1}{2}(1 - \gamma^5)f$ , 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:

$$\begin{aligned} \nu_L &= |I_3 = +\frac{1}{2}, \quad Q = 0, \quad Z = +\frac{1}{2}\rangle \\ \nu_R &= |I_3 = 0, \quad Q = 0, \quad Z = 0\rangle \\ e_L &= |I_3 = -\frac{1}{2}, \quad Q = -1, \quad Z = -\frac{1}{2} + \sin^2 \theta_W = -0.26845\rangle \\ e_R &= |I_3 = 0, \quad Q = -1, \quad Z = \sin^2 \theta_W = +0.23155\rangle \end{aligned} \tag{23.2a}$$

Likewise, for the quarks, also for all generations, the chiral charge quantum numbers are:

$$\begin{aligned} u_L &= |I_3 = +\frac{1}{2}, \quad Q = +\frac{2}{3}, \quad Z = +\frac{1}{2} - \frac{2}{3}\sin^2 \theta_W = +0.34563\rangle \\ u_R &= |I_3 = 0, \quad Q = +\frac{2}{3}, \quad Z = -\frac{2}{3}\sin^2 \theta_W = -0.15437\rangle \\ d_L &= |I_3 = -\frac{1}{2}, \quad Q = -\frac{1}{3}, \quad Z = -\frac{1}{2} + \frac{1}{3}\sin^2 \theta_W = -0.42282\rangle \\ d_R &= |I_3 = 0, \quad Q = -\frac{1}{3}, \quad Z = +\frac{1}{3}\sin^2 \theta_W = +0.07718\rangle \end{aligned} \tag{23.2b}$$

For the Z charge, it is then customary to define separate vertex and axial 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:

$$\begin{aligned} c_V(f) &\equiv Z_L + Z_R = I_{3L} - 2Q \sin^2 \theta_W \\ c_A(f) &\equiv Z_L - Z_R = I_{3L} \end{aligned} \quad (23.3)$$

as is well-known. Therefore, from (23.2) we may deduce that:

$$\begin{aligned} c_V(\nu) &= I_{3L}(\nu) - 2Q(\nu) \sin^2 \theta_W = +\frac{1}{2}; & c_A(\nu) &= I_{3L}(\nu) = +\frac{1}{2} \\ c_V(e) &= I_{3L}(e) - 2Q(e) \sin^2 \theta_W = -\frac{1}{2} + 2 \sin^2 \theta_W = -0.03690; & c_A(e) &= I_{3L}(e) = -\frac{1}{2} \\ c_V(u) &= I_{3L}(u) - 2Q(u) \sin^2 \theta_W = +\frac{1}{2} - \frac{4}{3} \sin^2 \theta_W = +0.19127; & c_A(u) &= I_{3L}(u) = +\frac{1}{2} \\ c_V(d) &= I_{3L}(d) - 2Q(d) \sin^2 \theta_W = -\frac{1}{2} + \frac{2}{3} \sin^2 \theta_W = -0.34563; & c_A(d) &= I_{3L}(d) = -\frac{1}{2} \end{aligned} \quad (23.4)$$

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 (23.4) to determine the couplings for the proton and neutron, then by simple addition we obtain:

$$\begin{aligned} c_V(p) = c_V(uud) &= +\frac{1}{2} - 2 \sin^2 \theta_W = +0.03690 & c_A(p) &= c_A(uud) = +\frac{1}{2} \\ c_V(n) = c_V(udd) &= -\frac{1}{2} & c_A(n) &= c_A(udd) = -\frac{1}{2} \end{aligned} \quad (23.5)$$

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:

$$\begin{aligned} c_R(\nu) &= 0; & c_L(\nu) &= +1 \\ c_R(e) &= 2 \sin^2 \theta_W = +0.46310; & c_L(e) &= -1 + 2 \sin^2 \theta_W = -0.53690 \\ c_R(u) &= -\frac{4}{3} \sin^2 \theta_W = -0.30873; & c_L(u) &= +1 - \frac{4}{3} \sin^2 \theta_W = +0.69127 \\ c_R(d) &= +\frac{2}{3} \sin^2 \theta_W = +0.15437; & c_L(d) &= -1 + \frac{2}{3} \sin^2 \theta_W = -0.84563 \end{aligned} \quad (23.6)$$

and for the proton and neutron:

$$\begin{aligned} c_R(p) &= -2 \sin^2 \theta_W = -0.46310; & c_L(p) &= +1 - 2 \sin^2 \theta_W = +0.53690 \\ c_R(n) &= 0; & c_L(n) &= -1 \end{aligned} \quad (23.7)$$

Now let's return to how the neutrino is attracted to or repelled by other fermions. In general,  $\mathcal{M}_Z(\nu f \rightarrow \nu f) \propto [\bar{\nu} \gamma^\mu (c_V(\nu) - c_A(\nu) \gamma^5) \nu] [\bar{f} \gamma_\mu (c_V(f) - c_A(f) \gamma^5) f]$  represents the invariant amplitude for a neutrino  $\nu$  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:

$$c_V(f) - c_A(f)\gamma^5 = (c_V(f) - c_A(f))\frac{1}{2}(1 + \gamma^5) + (c_V(f) + c_A(f))\frac{1}{2}(1 - \gamma^5), \quad (23.8)$$

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  $f_R = \frac{1}{2}(1 + \gamma^5)f$ ,  $\bar{f}_R = \bar{f}\frac{1}{2}(1 - \gamma^5)$  for right- and  $f_L = \frac{1}{2}(1 - \gamma^5)f$ ,  $\bar{f}_L = \bar{f}\frac{1}{2}(1 + \gamma^5)$  for left-chiral projections, together with  $\gamma^\mu\gamma^5 = -\gamma^5\gamma^\mu$  and the identities  $\frac{1}{2}(1 + \gamma^5) = \frac{1}{2}(1 + \gamma^5)\frac{1}{2}(1 + \gamma^5)$  and  $\frac{1}{2}(1 - \gamma^5) = \frac{1}{2}(1 - \gamma^5)\frac{1}{2}(1 - \gamma^5)$ , we may use (23.8) to rewrite the foregoing invariant amplitude as:

$$\begin{aligned} \mathfrak{M}_Z(\nu f \rightarrow \nu f) &\propto [\bar{\nu}\gamma^\mu(c_V(\nu) - c_A(\nu)\gamma^5)\nu] [\bar{f}\gamma_\mu(c_V(f) - c_A(f)\gamma^5)f] \\ &= [c_R(\nu)\bar{\nu}_R\gamma^\mu\nu_R + c_L(\nu)\bar{\nu}_L\gamma^\mu\nu_L] [c_R(f)\bar{f}_R\gamma_\mu f_R + c_L(f)\bar{f}_L\gamma_\mu f_L], \end{aligned} \quad (23.9)$$

Using  $c_R(\nu) = 0$  and  $c_L(\nu) = +1$  from (23.6) then distributing  $\bar{\nu}_L\gamma^\mu\nu_L$ , this further simplifies to:

$$\mathfrak{M}_Z(\nu f \rightarrow \nu f) \propto (\bar{\nu}_L\gamma^\mu\nu_L)c_R(f)(\bar{f}_R\gamma_\mu f_R) + (\bar{\nu}_L\gamma^\mu\nu_L)c_L(f)(\bar{f}_L\gamma_\mu f_L). \quad (23.10)$$

Note: the reason  $\bar{\nu}_R\gamma^\mu\nu_R$  drops out leaving only  $\bar{\nu}_L\gamma^\mu\nu_L$  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  $\bar{\nu}_R\gamma^\mu\nu_R$  from any amplitudes containing the neutrino.

For a neutrino interacting with an electron, and also with another neutrino, we use (23.6) in (23.10) to obtain:

$$\begin{aligned} \mathfrak{M}_Z(\nu e \rightarrow \nu e) &\propto +0.46310(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{e}_R\gamma_\mu e_R) - 0.53690(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{e}_L\gamma_\mu e_L) \\ \mathfrak{M}_Z(\nu\nu \rightarrow \nu\nu) &\propto +(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{\nu}_L\gamma_\mu\nu_L) \end{aligned} \quad (23.11)$$

For a neutrino interacting with up and down quarks we likewise obtain:

$$\begin{aligned} \mathfrak{M}_Z(\nu u \rightarrow \nu u) &\propto -0.30873(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{u}_R\gamma_\mu u_R) + 0.69127(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{u}_L\gamma_\mu u_L) \\ \mathfrak{M}_Z(\nu d \rightarrow \nu d) &\propto +0.15437(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{d}_R\gamma_\mu d_R) - 0.84563(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{d}_L\gamma_\mu d_L) \end{aligned} \quad (23.12)$$

Finally, for a neutrino interacting with a proton and neutron we use (23.7) in (23.10) to find:

$$\begin{aligned} \mathfrak{M}^{NC}(\nu p \rightarrow \nu p) &\propto -0.46310(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{p}_R\gamma_\mu p_R) + 0.53690(\bar{\nu}_L\gamma^\mu\nu_L)c_L(p)(\bar{p}_L\gamma_\mu p_L) \\ \mathfrak{M}^{NC}(\nu n \rightarrow \nu n) &\propto -(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{n}_L\gamma_\mu n_L) \end{aligned} \quad (23.13)$$

Now, in all of (23.11) through (23.13), 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  $\mathfrak{N}_{em}(ee \rightarrow ee) \propto +(\bar{e}\gamma^\mu e)(\bar{e}\gamma_\mu e)$  and  $\mathfrak{N}_{em}(ep \rightarrow ep) \propto -(\bar{e}\gamma^\mu e)(\bar{p}\gamma_\mu p)$  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 (23.11), owing to the plus sign in  $+0.46310(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{e}_R\gamma_\mu e_R)$  and the minus sign in  $-0.53690(\bar{\nu}_L\gamma^\mu\nu_L)(\bar{e}_L\gamma_\mu e_L)$ , we learn that the neutrino repels with the right-chiral projections 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 (23.6). We also see in (23.11) that neutrinos will repel other neutrinos.

From (23.12) 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 (23.13) 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 (23.13), the interaction between the neutrino and the neutron is exclusively, strongly-attractive.

Given  $c_R(\nu) = 0$  and  $c_L(\nu) = +1$  from (23.6), 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  $c_R(\bar{\nu}) = 0$  and  $c_L(\bar{\nu}) = -1$ . (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 (23.11) through (23.13) 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  $\beta^-$  decay is triggered by CvB neutrinos as shown in Figure 16, not only for free neutrons, but for all atomic isotopes in the periodic table which undergo  $\beta^-$  decay. And it also provides the basis for understanding how  $\beta^+$  decay is triggered by CvB antineutrinos (presumed to have a similar number flux to neutrinos) for all isotopes which undergo  $\beta^+$  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  $1 \text{ b} = (10^{-14} \text{ m})^2 = (10 \text{ f})^2$  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  $(2 \text{ f})^2 = .04 \text{ b}$ . 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  $\mathcal{N}^{NC}(\nu n \rightarrow \nu n) \propto -(\bar{\nu}_L \gamma^\mu \nu_L)(\bar{n}_L \gamma_\mu n_L)$  in (23.13): 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  $3 \times 10^{-25} \text{ s}$  mean life thus .1 f range of even an extreme-relativistic Z boson, more-likely through a  $Z \rightarrow \bar{f}f \rightarrow Z \rightarrow \bar{f}f \rightarrow Z \dots$  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  $\beta^-$  decay via  $\nu d \rightarrow W^+ e d \rightarrow e u$  thus an overall decay  $\nu n \rightarrow W^+ e n \rightarrow e p$ .

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  $10^6 \text{ b}$  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  $\mathcal{N}_Z(\nu e \rightarrow \nu e)$  from (23.11), 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  $Z \rightarrow \bar{f}f \rightarrow Z \rightarrow \bar{f}f \rightarrow Z \dots$  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 15 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  $Z \rightarrow \bar{f}f \rightarrow Z \rightarrow \bar{f}f \rightarrow Z\dots$  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, (23.13) 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  $\nu \rightarrow e^-W^+$  decay, followed about  $3 \times 10^{-25}$  s later by a  $W^+d \rightarrow u$  decay. The net result is that the  $\beta^-$  decay reaction  $\nu n \rightarrow e^-p$  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  $\beta^-$  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  $\beta^-$  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  $\beta^-$  decay event occurs.

Now we come to  $\beta^+$  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., [63]) that the density and flux of CvB antineutrinos is virtually the same as that of neutrinos, which means based on Figure 15 that 1 CvB antineutrino will also pass through a 1 b cross section approximately every fifteen minutes. If these antineutrinos are the trigger for  $\beta^+$  decay, then once an antineutrino gets in range of a proton, the reaction is either  $\nu_c \rightarrow e^+W^-$  followed momentarily thereafter by  $W^-p \rightarrow n$ , or the reverse-ordered  $p \rightarrow nW^+$  followed by  $\nu_c W^+ \rightarrow e^+$ . The positron can then annihilate one of the electrons. The next result in either case is  $\nu_c p \rightarrow e^+n$ , with the original electron included,  $\nu_c p e^- \rightarrow n$ . Just as with  $\beta^-$  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 (23.11) we obtained the neutral current cross section  $\mathfrak{N}_Z(\nu e \rightarrow \nu e)$  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  $\psi$  are related to their antifermion counterparts  $\psi_C$  by  $\psi_C = C\bar{\psi}^T$ . With  $C = i\gamma^2\gamma^0$  and  $\bar{\psi} = \psi^\dagger\gamma^0$  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 (23.6), we know that  $c_L(\nu_C) = 0$  and  $c_R(\nu_C) = -1$ . So, applying (23.9) to antineutrinos interacting with other fermions, we obtain a counterpart to (23.10), namely:

$$\begin{aligned} \mathfrak{N}_Z(\nu_C f \rightarrow \nu_C f) &\propto \left[ c_R(\nu_C)\bar{\nu}_{CR}\gamma^\mu\nu_{CR} + c_L(\nu_C)\bar{\nu}_{CL}\gamma^\mu\nu_{CL} \right] \left[ c_R(f)\bar{f}_R\gamma_\mu f_R + c_L(f)\bar{f}_L\gamma_\mu f_L \right] \\ &= -\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_R(f)\left(\bar{f}_R\gamma_\mu f_R\right) - \left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_L(f)\left(\bar{f}_L\gamma_\mu f_L\right) \end{aligned} \quad (23.14)$$

Using (23.6) and (23.7), the antineutrino counterparts to (23.11) through (23.13) are then:

$$\begin{aligned} \mathfrak{N}_Z(\nu_C e \rightarrow \nu_C e) &= -0.46310\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{e}_R\gamma_\mu e_R\right) + 0.53690\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{e}_L\gamma_\mu e_L\right), \\ \mathfrak{N}_Z(\nu_C \nu \rightarrow \nu_C \nu) &= -\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{\nu}_L\gamma_\mu\nu_L\right) \end{aligned} \quad (23.15)$$

$$\begin{aligned} \mathfrak{N}_Z(\nu_C u \rightarrow \nu_C u) &= +0.30873\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{u}_R\gamma_\mu u_R\right) - 0.69127\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{u}_L\gamma_\mu u_L\right) \\ \mathfrak{N}_Z(\nu_C d \rightarrow \nu_C d) &= -0.15437\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{d}_R\gamma_\mu d_R\right) + 0.84563\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_L(d)\left(\bar{d}_L\gamma_\mu d_L\right), \end{aligned} \quad (23.16)$$

$$\begin{aligned} \mathfrak{N}_Z(\nu_C p \rightarrow \nu_C p) &= +0.46310\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{p}_R\gamma_\mu p_R\right) - 0.53690\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)\left(\bar{p}_L\gamma_\mu p_L\right) \\ \mathfrak{N}_Z(\nu_C n \rightarrow \nu_C n) &= +\left(\bar{\nu}_{CR}\gamma^\mu\nu_{CR}\right)c_L(n)\left(\bar{n}_L\gamma_\mu n_L\right) \end{aligned} \quad (23.17)$$

So, as we expect, whatever is attractive to neutrinos is repulsive to antineutrinos, and vice versa. Most significantly, from  $\mathfrak{N}_Z(\nu_C e \rightarrow \nu_C e)$ , 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  $\beta^+$  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  $\gamma^5$  axial operator is the flat spacetime generator of

the Dirac-Kaluza-Klein timelike fifth dimension. And of course, the chiral operators  $R = \frac{1}{2}(1 + \gamma^5)$  and  $L = \frac{1}{2}(1 - \gamma^5)$  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,  $\psi_R$  and  $\psi_L$  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 (23.11) through (23.13) and (23.15) through (23.17), 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 physically polarized 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 (23.6), 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 the 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  $Z \rightarrow \bar{f}f \rightarrow Z \rightarrow \bar{f}f \rightarrow Z \dots$  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  $\nu_R$  spinor close to the  $\nu_L$  spinor. This is why all we ever observe is  $\nu_L$  seemingly-decoupled from its  $\nu_R$  counterpart. It is not that  $\nu_R$  and  $\nu_L$  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  $\nu_R$  to stray widely from its counterpart  $\nu_L$ . Indeed, the right-chiral  $\nu_R$  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  $\nu_R$  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 (23.11), 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 (23.15), the  $\mathfrak{M}_Z(\nu_C e \rightarrow \nu_C e)$  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 charge 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  $-0.53690(\bar{\nu}_L \gamma^\mu \nu_L)(\bar{e}_L \gamma_\mu e_L)$  term in (23.11) drawing the neutrino toward the nucleus and the repulsive  $+0.46310(\bar{\nu}_L \gamma^\mu \nu_L)(\bar{e}_R \gamma_\mu e_R)$  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  $-0.46310(\bar{\nu}_{CR} \gamma^\mu \nu_{CR})(\bar{e}_R \gamma_\mu e_R)$  from (23.15) having the dominant effect and  $+0.53690(\bar{\nu}_{CR} \gamma^\mu \nu_{CR})(\bar{e}_L \gamma_\mu e_L)$  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  $\beta^-$  and  $\beta^+$  decay for some of its isotopes, the  $\beta^-$  half-lives ought to be shorter than the  $\beta^+$  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  $\beta^-$  decays. But because of chiral polarization, both neutrino and antineutrinos can make it to the nucleus and we do observe both  $\beta^-$  and  $\beta^+$  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  $\beta^-$  than  $\beta^+$  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  $\beta^-$  and  $\beta^+$  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 7 and 8 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 13 and 14 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 [31], the author found in [10.6] and [10.7] that the rest energy of every free nucleon includes a latent binding energy given by  $B_p = 0.008200606481 u = 7.640679 \text{ MeV}$  and  $B_n = 0.010531999771 u = 9.812358 \text{ MeV}$  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,  $^{56}\text{Fe}$  with 26 protons and 30 neutrons – which has the distinction of having the highest average binding energy-per-nucleon than any other nuclide [64] – has available  $B(\text{Fe}^{56}) = 26 \times 7.640679 \text{ MeV} + 30 \times 9.812358 \text{ MeV} = 493.028394 \text{ MeV}$  of latent binding energy available be released for nucleon binding. This contrasts remarkably with the observed  $^{56}\text{Fe}$  binding energy of 492.253892 MeV, and shows that 99.8429093% of the latent binding energy goes into binding together the  $^{56}\text{Fe}$  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 [31] 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 [61] 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  $^1\text{H}$  protium which does not contain any neutrons. This of course, is the output of the  $\beta^-$  decay process  $\nu n \rightarrow e^- p$  decay process illustrated in Figure 16, because  $^1\text{H} = e^- p$ . So now, we wish to consider the  $\beta^+$  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  $^1\text{H}$  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 (23.16) and (23.17) 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  $\nu_c \rightarrow W^- e^+$  or  $p \rightarrow W^+ n$  emission, a  $\beta^+$  decay would proceed by the channel  $\nu_c e^- p \rightarrow e^+ e^- n \rightarrow n \gamma \gamma$  including electron / positron annihilation into photons, or into other bosons or mesons at higher energies. To explain this we turn to Figures 7, 8, 13 and 14 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  $\nu$  only nest at the local minimum of the Lagrangian potential. This explains why this  $\beta^+$  is not observed for an isolated proton and electron constituting  ${}^1\text{H}$ .

Next let's proceed to  ${}^2\text{H}$  deuterium and to  ${}^3\text{H}$  tritium. The former also is stable, but now contains a neutron which via (23.17) will strongly repel an incoming antineutrino. This makes it even harder for a  $\beta^+$  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  $\beta^-$  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,  ${}^3\text{H}$ , has one proton (1p) and two neutrons (2n). This does undergo  $\beta^-$  decay into  ${}^3\text{He}$  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  ${}^3\text{H}$  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  ${}^4\text{He}$  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  $\alpha$  particle is emitted or absorbed whole hog. This extreme stability, we ascribe to  ${}^4\text{He}$  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  $\beta^-$  or a  $\beta^+$  decay channel. The next-lower isotope,  ${}^3\text{He}$ , is also stable. Being poor by 1 neutron, were it to decay, the reaction would be  $\beta^+$  from  ${}^3\text{He} \rightarrow {}^3\text{H}$ . As between these two options, the nuclear binding energies render the neutron-poor  ${}^3\text{He}$  more stable than the neutron-rich  ${}^3\text{H}$  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  $\nu n$  state. The lightest isotope,  ${}^2\text{He}$  has two protons and no neutrons. Here, the alternative state is  ${}^2\text{H}$  which can be arrived at through a  $\beta^+$  decay. Here, with the two protons in  ${}^2\text{He}$  attract an incoming CvB antineutrino with no neutrons to repel, while the 1p and 1n in  ${}^2\text{H}$  provide a better attractive / repulsive balance. So, between the two options of  ${}^2\text{He}$  versus  ${}^2\text{H}$ , the latter is the stable option.

As to helium  $\beta^-$  decay, the isotopes  ${}^6\text{He}$  and  ${}^8\text{He}$  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  ${}^6\text{He}$  as an alpha plus two neutrons denoted  $\alpha + 2n$ , and  ${}^8\text{He}$  as  $\alpha + 4n$ , then with the  $\alpha$  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  $\mathcal{N}^{NC}(\nu n \rightarrow \nu n)$  in (23.13), 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  ${}^6\text{He}$  nucleus will beta decay with an 806.7 ms half-life, and  ${}^8\text{He}$  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  ${}^6\text{He}$  and  $10^4$  b for  ${}^8\text{He}$ . Note also that  ${}^7\text{He}$  with 5 neutrons and  ${}^9\text{He}$  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  ${}^6\text{Li}$  and  ${}^7\text{Li}$ , 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  ${}^7\text{Li}$  with 3p and 4n, which could in theory beta decay to  ${}^7\text{Be}$  with 4p and 3n. But the latter  ${}^7\text{Be}$  has a half-life of 53.22 d and the neutron-rich former  ${}^7\text{Li}$  is the one that is stable. Likewise,  ${}^6\text{Li}$  is stable. The alternatives would be  ${}^6\text{Be}$  with 4p and 2n which is neutron poor and so cannot sustain binding (and actually decays by releasing 2p), or  ${}^6\text{He}$  already reviewed which, with  $\alpha + 2n$ , will readily attract ensnared neutrinos and so is readily susceptible to  ${}^6\text{He} \rightarrow {}^6\text{Li}$  decay. The lighter isotopes  ${}^5\text{Li}$  and  ${}^4\text{Li}$  could in theory undergo  $\beta^+$  decay to  ${}^5\text{He}$  and  ${}^4\text{He}$  respectively, but apparently, nature follows a path in which simply discarding a proton is energetically-preferred. For  $\beta^-$  decay,  ${}^8\text{Li}$ ,  ${}^9\text{Li}$  and  ${}^{11}\text{Li}$  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  ${}^8\text{Li}$  down to a little under  $10^6$  b for  ${}^{11}\text{Li}$ .

### Beryllium

One might suppose that  ${}^8\text{Be}$  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 it as  $\alpha + 2p + 2n$ , the alpha both of its proton and neutrons in an  $n=1$  s-shell, whereas the extra 2p and 2n must be in an  $n=2$  p-shell. And it is more-favored to break this  ${}^8\text{Be}$  into  $2\alpha$  each with s-shells, than to maintain the p-shell. So,  ${}^9\text{Be}$  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,  ${}^{10}\text{Be}$  with 4p and 6n, in a pattern that will be repeated for heavier elements, is nearly stable, but does undergo  $\beta^-$  decay into  ${}^{10}\text{B}$  with a very long half-life of  $1.39 \times 10^6$  y. And of course,  ${}^{10}\text{B}$  is a stable isotope

of boron. For the lighter isotopes,  ${}^7\text{Be}$  does decay into the stable  ${}^7\text{Li}$  via electron capture, which is a form of  $\beta^+$  decay, with a comparatively-long 53.22 d. Were  ${}^6\text{Be}$  to beta decay it would become  ${}^6\text{Li}$  which is stable, but nature takes the route of ejecting 2p to arrive at  ${}^4\text{He}$ , apparently once again because of the extreme stability of alpha particles. For  $\beta^-$  decay, following the long-lived  ${}^{10}\text{Be}$  already discussed, we have  ${}^{11}\text{Be}$ ,  ${}^{12}\text{Be}$  and  ${}^{14}\text{Be}$  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  ${}^{13}\text{Be}$  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  ${}^{12}\text{Be}$  which is the isotope that undergoes beta decay. This is similar to what happens when  ${}^7\text{He}$  drops to  ${}^6\text{He}$  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  ${}^{10}\text{B}$  and  ${}^{11}\text{B}$  isotopes. The latter is rich by one neutron as is part of the pattern for elements heavier than helium. For light isotopes,  ${}^9\text{B}$  could in theory  $\beta^+$  decay into stable  ${}^9\text{Be}$  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  ${}^8\text{Be}$  which in turn immediately decays into two alpha particles. Next,  ${}^8\text{B}$  with 5p and 3n does indeed  $\beta^+$  decay into  ${}^8\text{Be}$  which in turn immediately decays into two alpha particles in  ${}^4\text{He}$ . All lighter isotopes also find a way to quickly decay toward the stability of  ${}^4\text{He}$ . As to heavier isotopes,  ${}^{12}\text{B}$ ,  ${}^{13}\text{B}$ ,  ${}^{14}\text{B}$ ,  ${}^{15}\text{B}$ ,  ${}^{17}\text{B}$  and  ${}^{19}\text{B}$  all exhibit  $\beta^-$  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  ${}^{16}\text{B}$  and  ${}^{18}\text{B}$  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  $\beta^+$  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  $\beta^-$  decays. The stable nuclides as  ${}^{12}\text{C}$  and  ${}^{13}\text{C}$  and at a half-life of 5730 y,  ${}^{14}\text{C}$  with 6p and 8n, well-known for its use in archeological dating, is almost but not quite stable. With its even number of neutrons,  ${}^{14}\text{C}$  is the second element to repeat the pattern that stated with  ${}^{10}\text{Be}$ , 4p and 6n, and its  $1.39 \times 10^6$  y half-life. The shorter-lived  $\beta^-$  decays begin in earnest with  ${}^{15}\text{C}$  through  ${}^{22}\text{C}$  with the exception of  ${}^{21}\text{C}$  with 6p and 15n which instead drops a neutron to  ${}^{20}\text{C}$  before it beta decays into  ${}^{20}\text{N}$ . Completely validating the neutrino trigger viewpoint, the respective half-lives of  ${}^{15}\text{C}$  through  ${}^{20}\text{C}$  are the successively-diminishing 2.449 s, 0.747 s, 193 ms, 92 ms, 46.2, ms and 16 ms, and for  ${}^{22}\text{C}$ , 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  $\beta^+$  decays. At the lightest isotope,  $^8\text{C}$  with 6p and 2n does not beta decay, but rather is alpha-driven. It sheds 2p down to  $^6\text{Be}$ , which in turn immediately sheds 2 more protons to an alpha particle. However,  $^{11}\text{C}$ ,  $^{10}\text{C}$  and  $^9\text{C}$  – all of which are proton-rich – do exhibit  $\beta^+$  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  $\beta^+$  decay is triggered by CvB antineutrinos being attracted to protons via  $\mathcal{O}_Z(\nu_c p \rightarrow \nu_c p)$  in (23.17), then as the number of neutrons which also by (23.17) would repel antineutrinos is reduced, the overall attraction balance over repulsion is increased. Thus, more antineutrinos will reach a proton to start a  $\beta^+$  decay. Second, because the  $\mathcal{O}_Z(\nu_c p \rightarrow \nu_c p)$  attraction of antineutrinos to protons in (23.17) is actually an attraction-weighted mix as between the two chiral states, while the  $\mathcal{O}^{NC}(\nu n \rightarrow \nu n)$  attraction of neutrinos to neutrons in (23.13) is purely attractive and thus stronger, this means that with all else equal, more neutrinos will be harvested per unit of time for  $\beta^-$  decay than will antineutrinos be harvested for  $\beta^+$  decay, with the result that the  $\beta^+$  half-lives ought to be longer, in general, than the  $\beta^-$  half-lives. Here, beyond the long  $^{14}\text{C}$  lifetime that occurs for reasons of nuclear structure and stability not neutrino availability, the  $\beta^-$  half-lives run from 2.449 s down to 6.2 ms, while the  $\beta^+$  half-lives run from 20.334 m to 19.290 s to 126.5 ms. These are consistently longer than the  $\beta^-$  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  $\beta^+$  decays. For nitrogen,  $^{14}\text{N}$  and  $^{15}\text{N}$  are stable. For  $^{16}\text{N}$  through  $^{22}\text{N}$  the  $\beta^-$  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  $^{13}\text{N}$  and  $^{12}\text{N}$  which are the only two isotopes with  $\beta^+$ , the half-lives of 9.965 min and 11.000 ms are also consistently diminishing. However, the latter for  $^{12}\text{N}$  decay does appear for the first time to buck the trend of  $\beta^+$  half-lives being longer in general than  $\beta^-$  half-lives. But on closer inspection we find that there are two channels for this decay. The dominant channel (96.5%) is for  $^{12}\text{N} \rightarrow ^{12}\text{C}$ . The less-frequent channel (3.5%) starts with  $^{12}\text{N}$  but includes both an alpha decay and a  $\beta^+$  decay. However, the  $\beta^+$  decay and the proton loss do not happen absolutely simultaneously. If the  $\beta^+$  occurs first, then we have  $^{12}\text{N} \rightarrow ^{12}\text{C}$ , and with  $^{12}\text{C}$  being stable, nothing more will happen. This just repeats the first channel. So, the  $\alpha$  drop must occur before the  $\beta^+$  to distinguish this channel, which means that the detailed sequence is  $^{12}\text{N} \rightarrow ^8\text{B} \rightarrow ^8\text{Be}$ . The  $^8\text{Be}$  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  ${}^8\text{Be} \rightarrow 2{}^4\text{He}$  has a half-life of  $6.7 \times 10^{-17}\text{s}$ , and many proton drops are even shorter. Therefore, it is to be expected that the less-frequent  ${}^{12}\text{N} \rightarrow {}^8\text{B} \rightarrow {}^8\text{Be}$  channel will actually be very-much much faster and the dominant  ${}^{12}\text{N} \rightarrow {}^{12}\text{C}$  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  ${}^{12}\text{N} \rightarrow {}^{12}\text{C}$  channel for a pure  $\beta^+$  decay, when segregated out, would indeed sustain the neutrino- and antineutrino-trigger trend of  $\beta^+$  decays having generally longer half-lives than  $\beta^-$  decays.

### Oxygen

Oxygen is stable at  ${}^{16}\text{O}$ ,  ${}^{17}\text{O}$  and  ${}^{18}\text{O}$ , continuing the neutron-rich stability for nucleons heavier than that of helium. The  $\beta^-$  decays from  ${}^{19}\text{O}$  through  ${}^{25}\text{O}$  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  ${}^{15}\text{O}$  to  ${}^{13}\text{O}$  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  $\beta^+$  half-lives are longer than all of the  $\beta^-$  times consistent with greater attraction between neutrinos and neutrons than between antineutrinos and protons. The  ${}^{13}\text{O}$  half-life again bucks the trend. But this mixes two channels, namely, the dominant channel (about 89.1%) with two serial  $\beta^+$  decays from  ${}^{13}\text{O} \rightarrow {}^{13}\text{N} \rightarrow {}^{13}\text{C}$ , and a less-frequent channel (10.9%) with both a  $\beta^+$  decay and a proton emission from  ${}^{13}\text{O} \rightarrow {}^{12}\text{C}$ . Similarly to what was reviewed for nitrogen, the  $\beta^+$  decay and the proton loss do not happen absolutely simultaneously. If the  $\beta^+$  occurs first then the detailed sequence is  ${}^{13}\text{O} \rightarrow {}^{13}\text{N} \rightarrow {}^{12}\text{C}$ . But if the proton drop occurs first then the sequence is  ${}^{13}\text{O} \rightarrow {}^{12}\text{N} \rightarrow {}^{12}\text{C}$ . But as just reviewed,  ${}^{12}\text{N}$  itself has two modes of proceeding, with the less-frequent (3.5%) mode being  ${}^{12}\text{N} \rightarrow {}^8\text{B} \rightarrow {}^8\text{Be}$ . So as with nitrogen, we expect that 8.58 ms for  ${}^{13}\text{O}$  actually averages the two decay channels, that the not-pure  $\beta^+$ , p decay  ${}^{13}\text{O} \rightarrow {}^{12}\text{C}$  actually occurs much faster, particularly when it makes brief passage through  ${}^{12}\text{N}$ . So as with nitrogen above, when the dominant and pure  ${}^{13}\text{O} \rightarrow {}^{13}\text{N} \rightarrow {}^{13}\text{C}$  decay is segregated out, we anticipate a longer lifetime consistent with  $\beta^+$  decays having taking longer than  $\beta^-$  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  ${}^{19}\text{F}$  with 9p and 10n. The  $\beta^-$  half-lives start at 11.163 s for  ${}^{20}\text{F}$  diminishing to 4.9 ms for  ${}^{27}\text{F}$ .  ${}^{28}\text{F}$  only emits a neutron, and  ${}^{29}\text{F}$  drops to a 2.6 ms half-life. The absolute trend of decreasing  $\beta^-$  life with increasing neutron number thus increased neutrino attraction is very-slightly broken for the first time by  ${}^{21}\text{F}$  with 9p and 12n and a 4.158 s half-life, followed by  ${}^{22}\text{F}$  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  $\beta^+$  half-lives are 109.771 min for  ${}^{18}\text{F}$  and 64.49 s for  ${}^{17}\text{F}$ , and

these do follow both the expected diminishing half-life pattern and also the pattern of  $\beta^+$  half-lives being generally longer than those for  $\beta^-$  decays.

### Neon

The three stable isotopes of neon are  $^{20}\text{Ne}$ ,  $^{21}\text{Ne}$  and  $^{22}\text{Ne}$ . The  $\beta^-$  half-lives begin with  $^{23}\text{Ne}$  and its 37.24 s half-life, then  $^{24}\text{Ne}$  with the *longer* 3.38 min, then  $^{25}\text{Ne}$  with 602(8) ms. So this too is a break in the half-life diminishment correlation. Thereafter the half-lives do serially descend from  $^{26}\text{Ne}$  with 197 ms to  $^{31}\text{Ne}$  with 3.4 ms, and then a mild break by  $^{32}\text{Ne}$  with 3.5 ms. The  $^{24}\text{Ne}$  isotope which has the longer half-life than  $^{23}\text{Ne}$  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  $\beta^+$  decays are  $^{19}\text{Ne}$  with 17.296 s,  $^{18}\text{Ne}$  with 1.672 s and  $^{17}\text{Ne}$  with 109.2 ms. This adheres to the correlation of diminishing antineutrino repulsion with reduced neutron number, and setting aside  $^{24}\text{Ne}$  with its 3.38 min that breaks the  $\beta^-$  sequence, it also adheres to the pattern of longer  $\beta^+$  over  $\beta^-$  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 15, fits the view of these neutrinos and antineutrinos as being harvested from a cross section within  $10^6$  b of the nucleus. With  $10^6$  b = .01 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  $\beta^-$  half-life on the former hand, and between decreased neutron number and decreased  $\beta^+$  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  $\beta^-$  half-lives being shorter than  $\beta^+$  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  $\beta^-$  decays taken together with that for  $\beta^+$  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., [65], [66], wherein on average, “bottle” neutrons decay after 14 minutes and 39 seconds while “beam” neutrons last 14 minutes and 48 seconds. If  $\beta^+$  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 somehow reduces the number flux of the CvB neutrinos which might otherwise be available to trigger their decay. 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.*

There are two mechanisms which may be considered to account for this number flux reduction, and both simply involve the motion of neutrons in the neutron beam: First, in electromagnetism, when we start with static charges in a pair of parallel wires and then apply a voltage to generate currents, there is an attraction between the wires when the currents flow in parallel and a repulsion when they flow antiparallel. This of course, results from the magnetic fields of moving charges, and it is well-known how to relate this to relativistic motion. So, the motion of the neutrino beam relative to the CvB background likewise will create neutral current “magnetic” fields which will alter the attractive and repulsive interactions between the CvB neutrinos and beam neutrons and thus impact number fluxes. Unlike parallel wires in electromagnetism, however, the interaction range for the weak neutral current is extremely short.

Second, perhaps decisively, although the CvB neutrinos are non-relativistic, they are still extraordinarily fast by human standards. Unlike light they do not orbit the earth 6.6 times every second, but they still do travel fast enough to cross an entire continent every second. So, by way of a loose analogy, one may think of a bottle neutrino as a car sitting by the side of a highway on which many other cars are driving by at 60 miles per hour, and may think of a beam neutrino as a car that starts to drive at 30 mph on the highway and so now sees a reduced “flux” of other cars going by because of its own motion. Similarly, if the CvB neutrinos are flowing past the “at rest” bottle neutrons from all directions in a statistically-isotopic manner, once these neutrons are imparted with motion in the form of a beam, one anticipates that the overall CvB number fluxes will diminished and / or the energy distribution will be altered.

In terms of Figure 15, and keeping in mind the Michaelson-Morley experiments which yielded a negative outcome for the detection of a rest frame for light, this is to say that: a) the CvB background *does have a rest frame*; b) the Figure 15 data curve for (at least) the cosmological neutrinos is taken *in this CvB rest frame*; c) when detected from other-than its rest frame the CvB background does *not* have the exact same data curve as that shown in Figure 15, because the motion either diminishes the height of the curve and / or alters the spread of the curve along the  $x$  axis and / or alters the position of the curve along the  $x$  axis. If the CvB background is in fact the trigger for beta decays, that any or all of the above will change the beta-decay lifetime for beam neutrinos in relation to bottle neutrinos. As such, the neutrino beam / lifetime experiments provide a very good vehicle to test the view that beta decays are triggered by the CvB background, especially if the effects of motion on this background can be calculated or detected and then correlated to neutron lifetime observations.

Once we understand beta decay half-lives as being environmentally-determined, for example, by the particular configuration of electron shells and nuclei inside of which the decays are taking place, or by the speed with which a beam of neutrons is moving through the CvB background, then any space or time anisotropies in the CvB background will also affect half-life. That is, the lifetime of a neutron or proton is determined not only by the atom of which it is a part or the speed with which it is beamed, but also, simply, by where it is sitting and when it is sitting there. So, if it becomes possible to detect irregularities in the CvB background, then simply moving atoms from one place to the next and finding that the lifetimes have changed becomes another way of validating the neutrino trigger hypothesis. Conversely, one can invert this logic and actually use beta decay-susceptible isotopes to detect anisotropies in the CvB background. Although in the foregoing discussion we shown the centers of the half-life data and ignored the error bars, there are actually two views one can take of these error bars: First, they may be regarded as ordinary experimental errors based on the limited precision of measuring equipment. Second, unbeknownst, *they may in fact already be measuring these CvB anisotropies*.

Specifically, while experimental precision is always a factor to be considered, it is not very difficult to measure elapsed times on the order of minutes or hours, nor does this require the patience of measuring years or centuries. So, some of the isotopes already reviewed may be providing us data regarding the CvB anisotropies even though we are unaware of this. For example, the half-life  $^{11}\text{C} \rightarrow ^{11}\text{B}$  is listed in the Wiki data as 20.334(24) min (error spread just under 3 s), that of  $^{13}\text{N} \rightarrow ^{13}\text{C}$  is 9.965(4) min (error just under .5 s), that of  $^{18}\text{F} \rightarrow ^{18}\text{O}$  is 109.771(20) min (spread of 1.2 s), and that of  $^{24}\text{Na} \rightarrow ^{24}\text{Mg}$  is 14.9590(12) h (spread of about 8.64 s). Certainly, using  $^{24}\text{Na}$  as an example for discussion, there is no problem at all finding a clock that can discern an 8 second time difference, which rules out equipment resolution as the source of the error bar. And while half-lives are determined using statistical data based on large numbers of individual decay events, it is difficult to believe that one cannot carry out enough experiments with enough samples to determine the  $^{24}\text{Na}$  half-life, in a way that statistically narrows this 8.64 second spread much more tightly. So, the only plausible explanation is that this 8.64

second spread is in fact, unbeknownst to us, a measurement of CvB anisotropy. Of course, translating an 8.64 s spread out of 14.9590 h overall into a percentage number for CvB anisotropy would require a very complex understanding of the structure of  $^{24}\text{Na}$  and its atomic and nuclear shell structures and binding energies and weak neutral current interactions with neutrinos. So, the simplest analysis is that which uses the simplest nuclear structure, and there is none simpler than a single free “bottle neutron” with a mean life of  $880.2 \pm 1.0$  s based on the PDG data [48]. This is a  $2/880.2 = 0.227\%$  error bar which, properly understood, tells us that on the surface of the earth where the experiments to arrive at this neutron mean life data were presumably conducted, there is a 0.227% anisotropy in the CvB neutrino background. Other experiments to directly detect CvB anisotropies – if consistent with this number – would confirm the neutrino trigger hypothesis.

Pedagogically, CvB neutrinos and antineutrinos being the triggers for weak beta decay also provides a more complete way to think about atoms and their nuclei: Historically, the neutrino was proposed by Pauli and Fermi in view of Chadwick’s discovery of the neutron as a means to ensure conservation not only of energy but of particle number and the other quantum numbers which distinguish one type of fermion flavor from another. So, the observed  $\beta^-$  decay written as  $n \rightarrow pe^-$  is incomplete, because on the left it has one baryon and on the right it has both a baryon and a lepton. Therefore, to provide proper balance, we add an antineutrino to the right and correctly write this as  $n \rightarrow pe^-\bar{\nu}$  (noting again that to be precise we should use the conjugate fermion  $\nu_c$  rather than the adjoint fermion  $\bar{\nu}$ ). Likewise, the  $\beta^+$  reaction, properly-balanced, is  $pe^- \rightarrow n\nu$  for electron capture or  $p \rightarrow ne^+\nu$  for positron emission. But in all instances, there is a tendency to show the neutrino or antineutrino on the right side amidst the *byproducts* of the reaction, as opposed to on the left side amidst the *ingredients* of the reaction. There appear to be two subliminal reasons for this: First, the foregoing historical context in which neutrinos and antineutrinos were added to the byproducts of a reaction to adhere to generalized conservation laws. Second, because we typically do think about electrons and protons and neutrons as “parts of an atom” but do not typically think about neutrinos in this way. Neutrinos are a sort of afterthought needed to balance out the conservation laws.

But once we take the CvB as a given feature of nature, then this background becomes the natural “environment” which all atoms exit. Nuclei naturally exist amidst a CvB background that permeates everywhere. To think about the process by which the electrons and protons and neutrons comprising an atom undergo beta decay without understanding this *metabolizing neutrino environment* is like trying to understand human biology without knowing that humans live in an atmosphere which provides oxygen required to metabolize their biological processes. So, once we think about atoms always being in a CvB atmosphere, it is natural to rewrite  $\beta^-$  decay as  $\nu n \rightarrow e^- p$  and  $\beta^+$  decay as either  $\bar{\nu} pe^- \rightarrow n$  or  $\bar{\nu} p \rightarrow ne^+$ , treating the triggering neutrinos and antineutrinos as the “oxygen” which metabolizes beta decay.

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 been 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.

## Conclusion

As a result of this theory of fermion masses, as summarized at (21.5), the known masses of the six quarks and the three charged leptons are all explained within experimental errors in terms of the real CKM and PMNS mixing angles and five other parameters. Four of these other parameters are known, namely the Fermi constant  $G_F$  and its related vev, the Newton gravitational constant  $G$  and its associated Planck energy, the Higgs boson mass, and the value  $\alpha(M_W^2)$  of the electromagnetic running coupling at an impact energy equal to the  $W$  boson mass. The final parameter is the rest mass of a predicted second Higgs boson from which leptons gain their masses. As a result, the known masses of the top and strange quarks are refined at (14.5) and (15.10) to over two orders of magnitude greater accuracy than what is known at present. The CKM mixing angles  $\theta_{c23}$  and  $\theta_{c13}$  have their values tightened at (15.15), the latter by almost three orders of magnitude. The Higgs boson mass is refined at (16.5) to almost three orders of magnitude greater accuracy, from the presently known  $m_h = 125.18 \pm 0.16 \text{ GeV}/c^2$  to a much-tighter  $m_h = 125.2485 \pm 0.0002 \text{ GeV}/c^2$ . The PMNS mixing angles  $\theta_{p12}$  and  $\theta_{p13}$  are re-centered at (19.14), and the remaining real angle  $\theta_{p23}$  is made tighter at (20.7) by more than two orders of magnitude. At (19.16) the value of  $\alpha(M_W^2)$ , presently known to be on the order of 1/128, is given the more-precise valuation  $\alpha(M_W^2) = 1/(127.9032 \pm 0.0080)$ . Further, the sum of the masses of the three flavors of neutrino is predicted at (20.2b) to be  $m_{\nu\tau} + m_{\nu\mu} + m_{\nu e} = 0.13348 \text{ eV}/c^2$ . This enables us at (20.4) to predict the separate masses of each neutrino flavor for a normal mass ordering, at  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  errors, to be  $m_{\nu e} c^2 = 0.03533^{+0.00012}_{-0.00012} {}^{+0.00024}_{-0.00025} {}^{+0.00035}_{-0.00037} \text{ eV}$ ,  $m_{\nu\mu} c^2 = 0.03637^{+0.00009}_{-0.00009} {}^{+0.00017}_{-0.00018} {}^{+0.00026}_{-0.00027} \text{ eV}$  and  $m_{\nu\tau} c^2 = 0.06178^{+0.00020}_{+0.00021} {}^{-0.00041}_{+0.00043} {}^{-0.00062}_{+0.00064} \text{ eV}$ . Finally, it becomes understood as a result of the foregoing that while the established Higgs boson is used to give mass to the six quark flavors, there is a second, separate Higgs scalar boson denoted  $h_2$  which gives mass to the six lepton flavors. The very existence of this new Higgs particle is a prediction of the theory, and the mass of this leptonic Higgs boson is predicted at (21.1) to be  $m_{h_2} c^2 = 941.515 \pm 0.060 \text{ MeV}$ , only a few MeV above the proton and neutron masses.

The theory presented is entirely consistent with the standard model, without contradiction. Indeed, the interrelations developed between fermion masses and other parameters of the standard model are what enable the tightening of multiple empirical numbers summarized above. This is because once these masses are interrelated with the other parameters, the masses and other parameters which are more-tightly-known can be used to adjust those which are less-accurately-known. Additionally, the underlying Lagrangian potentials for the Higgs fields lead us to uncover

a fundamental role for the cosmological neutrino background (CvB) as the trigger mechanism of weak beta decays.

The only place we go beyond the standard model, is to predict the neutrino mass sum, which at present has no theoretical basis in the standard model, and which is unknown other than via experiments which have established its upper limits. Specifically, the very-small ratio  $\nu / M_p c^2$  of the Fermi vev to the Planck energy is what establishes the neutrino mass sum in (20.2b) as against an energy magnitude set by the charged lepton masses and two of the three PMNS mixing angles. So, by including the Planck mass defined as usual by  $GM_p^2 \equiv \hbar c$ , this means that gravitation in the form of the Newton constant  $G$  becomes injected into particle physics, specifically to establish the exceptionally tiny masses and the beta-decay behaviors of neutrinos.

By reproducing multiple known experimental results of the standard model, by in several cases predicting existing experimental data with even-greater accuracy than what is known at present, by tightly predicting the neutrino masses which are presently known only within broad upper limits, and by predicting both the existence and the mass of a new leptonic Higgs boson, this theory provides multiple points for contact for experimental testing. And all of this is in addition to how this theory finally “repairs” all the long-standing perplexing problems of Kaluza-Klein theory in advance of the latter’s 2019 centenary.

## 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](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](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](https://en.wikipedia.org/wiki/Invertible_matrix#Blockwise_inversion)
- [15] [https://en.wikipedia.org/wiki/Invertible\\_matrix#Inversion\\_of\\_2\\_%C3%97\\_2\\_matrices](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\\_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_Renormalization) (2018)
- [17] [https://en.wikipedia.org/wiki/Einstein%E2%80%93Hilbert\\_action](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](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  $t\text{-}\bar{t}H$  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] H. Minkowski, *Raum und Zeit*, Jahresberichte der Deutschen Mathematiker-Vereinigung, 75-88 (1909)
- [26] <http://pdg.lbl.gov/2017/tables/rpp2017-qtab-mesons.pdf>
- [27] [https://en.wikipedia.org/wiki/Einstein%E2%80%93Hilbert\\_action#Derivation\\_of\\_Einstein's\\_field\\_equations](https://en.wikipedia.org/wiki/Einstein%E2%80%93Hilbert_action#Derivation_of_Einstein's_field_equations)
- [28] Zee, A. *Quantum Field Theory in a Nutshell*, Princeton (2003)
- [29] C.N. Yang and R. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Physical Review. 96 (1): 191–195. (1954)
- [30] 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](https://doi.org/10.4236/jmp.2013.44A011).
- [31] 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](https://doi.org/10.4236/jmp.2013.44A010).
- [32] 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](https://patentscope.wipo.int/search/docservicepdf_pct/id00000025517277/PAMPH/WO2014106223.pdf?psAuth=rD5Mq5vFTHPdQfjsYcI0wcNCzhFItduBNxW1oD9w9aw&download)
- [33] 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>
- [34] 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](https://doi.org/10.4236/jmp.2013.44A013).
- [35] C. N. Yang and R. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Physical Review. 96 (1): 191–195. (1954)
- [36] [https://en.wikipedia.org/wiki/Spin\\_connection](https://en.wikipedia.org/wiki/Spin_connection)
- [37] <https://home.cern/topics/higgs-boson>
- [38] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cimento. 19: 154–164 (1961)
- [39] H. Georgi, S. Glashow, *Unity of All Elementary-Particle Forces*, Physical Review Letters. 32 (8) (1974)
- [40] Bohr, N., *On the Constitution of Atoms and Molecules, Part I*, Philosophical Magazine. 26 (151): 1–24 (1913)
- [41] L. de Broglie, *A Tentative Theory of Light Quanta*, Phil. Mag. 47 446 (1924)
- [42] [https://physics.nist.gov/cgi-bin/cuu/Value?plkmc2gev|search\\_for=Planck+mass](https://physics.nist.gov/cgi-bin/cuu/Value?plkmc2gev|search_for=Planck+mass)
- [43] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), <http://pdg.lbl.gov/2018/tables/rpp2018-sum-quarks.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] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), <http://pdg.lbl.gov/2018/tables/rpp2018-sum-leptons.pdf>
- [48] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), <http://pdg.lbl.gov/2018/tables/rpp2018-sum-baryons.pdf>
- [49] <https://physics.nist.gov/cgi-bin/cuu/Value?mnmmpc2mev>
- [50] 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>
- [51] [https://en.wikipedia.org/wiki/Top\\_quark#Single\\_top\\_quarks](https://en.wikipedia.org/wiki/Top_quark#Single_top_quarks)

- [52] [https://en.wikipedia.org/wiki/Cabibbo%E2%80%93Kobayashi%E2%80%93Maskawa\\_matrix#Parameterizations](https://en.wikipedia.org/wiki/Cabibbo%E2%80%93Kobayashi%E2%80%93Maskawa_matrix#Parameterizations)
- [53] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), <http://pdg.lbl.gov/2018/tables/rpp2018-tab-baryons-N.pdf>
- [54] <https://physics.nist.gov/cgi-bin/cuu/Value?mnmmpc2mev>
- [55] M. J. Longo, *New Precision Tests of the Einstein Equivalence Principle from Sn1987a*, Phys. Rev. Lett. 60, 173 (1988)
- [56] NuFIT 3.2, <http://www.nu-fit.org/?q=node/166> (2018)
- [57] K. Nakamura, S.T. Petcov, *Neutrino Mass, Mixing, and Oscillations*, <http://pdg.lbl.gov/2018/reviews/rpp2018-rev-neutrino-mixing.pdf> (2018)
- [58] 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)
- [59] 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>
- [60] [https://en.wikipedia.org/wiki/W\\_and\\_Z\\_bosons](https://en.wikipedia.org/wiki/W_and_Z_bosons)
- [61] [https://en.wikipedia.org/wiki/Table\\_of\\_nuclides](https://en.wikipedia.org/wiki/Table_of_nuclides)
- [62] [https://en.wikipedia.org/wiki/List\\_of\\_radioactive\\_isotopes\\_by\\_half-life](https://en.wikipedia.org/wiki/List_of_radioactive_isotopes_by_half-life)
- [63] <http://www.astro.ucla.edu/~wright/neutrinos.html>
- [64] [https://en.wikipedia.org/wiki/Nuclear\\_binding\\_energy#Nuclear\\_binding\\_energy\\_curve](https://en.wikipedia.org/wiki/Nuclear_binding_energy#Nuclear_binding_energy_curve)
- [65] <https://www.quantamagazine.org/neutron-lifetime-puzzle-deepens-but-no-dark-matter-seen-20180213/>
- [66] 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)