

Jay R. Yablon

# **Why the Composite Magnetic Monopoles of Yang-Mills Gauge Theory have all the Required Chromodynamic and Confinement Symmetries of Baryons, how these may be developed into Topologically-Stable Protons and Neutrons, and how to Analytically Path Integrate the Yang-Mills Action**

Jay R. Yablon  
Schenectady, New York  
[jyablon@nycap.rr.com](mailto: jyablon@nycap.rr.com)

February 16, 2014, Revised February 23, March 17, April 1 and 18, June 5, 2014

*Abstract: We develop in detail, the classical magnetic monopoles of non-abelian Yang-Mills gauge theory and show how these monopoles, when analyzed using Gauss' / Stokes' theorem, appear to confine their gauge fields, and also, appear to be composite objects. Of course, baryons, which include the protons and neutrons at the heart of nuclear physics, also confine their gauge fields and are similarly-composite objects. This raises the question whether the magnetic monopoles of Yang-Mills theory are in some fashion related to the observed physical baryons. After developing inverse solutions for the non-abelian electric charge densities while carefully examining uniqueness and gauge fixing, we use these solutions together with Dirac theory to "populate" these classical monopoles with fermions. Applying the Fermi-Dirac-Pauli Exclusion Principle to these fermions forces the selection of a dimension-3 gauge group initially chosen to be  $SU(3)$ . We then find that these non-abelian magnetic monopoles have the exact chromodynamic symmetries of baryons and interact via colored magnetic fields with the exact chromodynamic symmetries of mesons. We show that a required  $U(1)$  factor ensures that these monopoles are topologically stable, and also "flavors" these monopole as protons and neutrons. Because this exposition is classical, we also discuss the extent to which classical field theory can be used to effectively analyze baryons and confinement. We point out how a recursive aspect of the non-abelian electric charge solution may be used to perform an analytically-exact quantum path integration for Yang-Mills theory, proving the existence of a non-trivial quantum Yang-Mills theory on  $R^4$  for any simple gauge group  $G$ . Finally, we use the results of this path integration to develop four examples of the application of analytical non-linear quantum field theory, which includes a quantum field explanation of confinement, a fitting of the running QCD curve to the known empirical data within experimental error bars, and a careful review of single and double slit experiments.*

PACS: 12.38.Aw; 12.40.-y; 14.20.-c; 14.40.-n

**Contents and Figures**

PART I: CLASSICAL YANG-MILLS THEORY ..... 5

1. Introduction: The Field Strength Curvature Tensor in Gauge Theory, and a Review of Gauge-Covariant Derivatives ..... 5
2. Classical Field Equations for the Yang-Mills Magnetic Monopole ..... 9
3. Confinement of Gauge fields within, and the Composite Nature of, Yang-Mills Magnetic Monopoles ..... 12
4. Can a Classical Field Equation Really Teach us Anything Useful about Baryons and Confinement? ..... 16
5. Classical Field Equations for the Yang-Mills Electric Charge..... 18
6. Abelian and non-Abelian Massive Gauge Boson Inverses for the Electric Charge Density, Using the “Coleman-Zee” Method..... 24
7. Abelian and non-Abelian Massless Gauge Boson Inverses for the Electric Charge Density, Using the Faddeev-Popov Method ..... 37
8. The Recursive Nature of Non-Abelian Gauge Theory, and what it may Teach about Quantizing Yang-Mills Gauge Theory ..... 44
9. Populating the Composite Yang-Mills Magnetic Monopoles with Chromodynamically-Colored Fermions ..... 53
10. Why the Composite Faux Magnetic Monopoles of Yang-Mills Gauge Theory have all of the Required Chromodynamic Symmetries of Baryons, and how these are Flavored into being Topologically-Stable Protons and Neutrons ..... 66

PART II: QUANTUM YANG-MILLS THEORY ..... 77

11. Quantum Yang-Mills Theory: Exact Analytical Path Integration ..... 77
12. The Baryon Candidate Lagrangian Density and Action ..... 82
13. The Baryon Candidate Quantum Path Integral ..... 90
14. Direct Quantum Field Theory Demonstration of Confinement – Abelian Calculation ..... 97
  - Figure 1: The Ordinary  $R^{-1}$  Potential ..... 109
  - Figure 2: The Yang-Mills Potential  $|E_1|$  of (14.39) at First Recursive Order ..... 110
  - Figure 3: The Yang-Mills Potential  $\text{Re } E_1$  of (14.40) at First Recursive Order ..... 111
15. Direct Quantum Field Theory Demonstration of Confinement – Non-Abelian Generalization ..... 115

PART III: ANALYTICAL NON-LINEAR QUANTUM FIELD THEORY: SPECIFIC EXAMPLES ..... 134

16. Constant Probability, Zero Probability Density Fields: Introduction to Analytical Non-Linear Quantum Field Theory ..... 134

17. Constant, Isotropic Probability Densities and Confining Stable Quantum Potentials in Non-Linear Quantum Field Theory .....	137
Figure 4: The Yang-Mills Potential $ E_1 $ of (17.13) at First Recursive Order .....	142
Figure 5: Equation (17.14) for $\text{Re } E_1$ , showing a First-Order Confinement Peak at $R_{\text{peak}} \cong 8.245$ .....	143
Figure 6: The Yang-Mills Potential $ E_1 $ of (17.33) at First Recursive Order (Figure 4 with Physical Energies and Length Scales) .....	148
Figure 7: Constant Radial Probability Density .....	148
Figure 8: Dimensionless Probability Field for Constant Probability Density $g_s^{\frac{1}{3}}\mu_r$ over $0 \leq r \leq r_\Lambda$ .....	149
18. Asymptotic Freedom and Asymptotic Confinement: Fitting and Extending the QCD Running Coupling Curve.....	151
Figure 9: The Running Strong Coupling $\alpha_s(Q = \hbar c / r)$ (reproduced from PDG's [24], Figure 9.4).....	152
Figure 10: The Running Strong Coupling $\alpha_s$ Modelled from (18.8) based on the Unit Circle.....	155
Figure 11: Superimposition of Figure 10 Theoretical Curve over Figure 9 Empirical Curve for $\alpha_s$ .....	156
Figure 12: The Bare Probability Density Fitted to the Empirical $\alpha_s(Q = \hbar c / r)$ Data of Figure 9 .....	158
Figure 13: Fitting of the PDG QCD Curve of Figure 9 to Extended Domain and Range for $\alpha_s(Q = \hbar c / r)$ , based on the Theoretical Function (18.20).....	161
Figure 14: Plot of (18.22) for $\alpha_s(Q)$ fitted to $\Lambda = \Lambda^{(6)} = .0906\text{GeV}$ and $\alpha_s(M_Z) = .1185$ at $M_Z = 91.876\text{GeV}$ , Superimposed on Figure 9 from PDG.....	164
Figure 15: The First-Order Quantum Potential Well, as a function of $Q$ .....	169
Figure 16: Variation of Observed $r = \hbar c / Q$ in relation to $r_\Lambda = \hbar c / \Lambda = 2.178F$ for a "Super-Radial" Observer Situated at $r > r_\Lambda = 2.178F$ .....	173
Figure 17: Variation of Observed $r = \hbar c / Q$ in relation to $r_\Lambda = \hbar c / \Lambda = 2.178F$ for a "Sub-Radial" Observer Situated at $r < r_\Lambda = 2.178F$ .....	174
19. Gaussian Probability Densities in Non-Linear Quantum Field Theory.....	176
Figure 18: The Discrete Cumulative Probability Field for Rolling a Pair of Dice – A Simple Example .....	177

Figure 19: Graph of Equation (19.10) Revealing Apparent $-1/R$ Potential for Gaussian Probability Density .....	182
Figure 20: Graph of $R \exp(-R^2)$ which Drives (19.6) and (19.10) and Figure 19....	182
Figure 21: Fractional Deviation of (19.10) and Figure 19 from a $-1/R$ Potential.....	183
20. Single and Double Slit Probability Densities and Guiding Potentials in Non-Linear Quantum Field Theory .....	188
Figure 22: Schematic Illustration of Single and / or Double Slit Diffraction.....	189
Figure 23: Probability Density / Intensity for Single-Slit Diffraction .....	191
Figure 24: Probability Density / Intensity for Double-Slit Diffraction, $B=10$ Example .....	193
Figure 25: Single-Slit Potential for $\Theta = 30^\circ = \pi / 6$ , Equation (20.29) .....	200
Figure 26: Single-Slit Potential for $\Theta = 30^\circ = \pi / 6$ , Equation (20.29), Wide View..	201
Figure 27: Double-Slit Potential for $\Theta = 30^\circ = \pi / 6$ , $B = 10$ , Equation (20.30) .....	202
Figure 28: Double-Slit Potential for $\Theta = 30^\circ = \pi / 6$ , $B = 10$ , Equation (20.30), Wide View .....	202
Figure 29: Buildup over Time of Double-Slit “Interference” Pattern (Reproduced from [32]).....	205
Figure 30: The Guiding Potential $ E_1 $ and the Consequent Probability Density $\partial_x (hP_0)$ for the Double-Slit Experiment .....	213
21. Summary and Conclusion .....	217
References .....	224

## PART I: CLASSICAL YANG-MILLS THEORY

### 1. Introduction: The Field Strength Curvature Tensor in Gauge Theory, and a Review of Gauge-Covariant Derivatives

In 1918, [1], [2] Hermann Weyl first conceived the idea that electrodynamics might be unified with Einstein's recently-developed geometric theory of gravitation [3], by analyzing a "twisting" of vectors under parallel transport to measure the geometric curvature of a gauge space. While Weyl first conceived of this as a local "gauge" symmetry, in 1929 [4] he corrected his original misconception into the modern view of a local "phase" symmetry. Notwithstanding, the original misnomer "gauge" is still used to name Weyl's theory, perhaps as a reminder to posterity that even the most foundational physical theories are sometimes properly-conceived in the abstract but misconceived in some details that need to be worked out over time.

In gravitational theory the Riemann curvature tensor  $R^\sigma_{\alpha\mu\nu}$  may of course be *defined* as a measure of the degree to which the gravitationally-covariant derivative  $\partial_{;\mu}$  is non-commuting when it operates on an arbitrary vector  $A_\sigma$ , that is, as  $R^\sigma_{\alpha\mu\nu}A_\sigma \equiv [\partial_{;\mu}, \partial_{;\nu}]A_\alpha$ . What Weyl essentially found is that the antisymmetric, second rank, field strength tensor / bivector  $F_{\mu\nu}$  which appears in electromagnetic theory may be defined as a measure of the extent to which the gauge-covariant derivative  $D_\mu$  is not self-commuting when it operates on an arbitrary scalar field  $\phi$ . That is,  $F_{\mu\nu}$  may be *defined* analogously to  $R^\sigma_{\alpha\mu\nu}$ , as a type of curvature in "gauge space," by:

$$F_{\mu\nu}\phi \equiv i[D_\mu, D_\nu]\phi = iD_\mu(D_\nu\phi) - iD_\nu(D_\mu\phi). \quad (1.1)$$

It is instructive to review how the explicit relationship between the field strength  $F_{\mu\nu}$  and a gauge / vector potential  $G_\mu$  then arises from this definition (1.1).

Gauge-covariant derivatives, like covariant derivatives in Riemannian geometry, take a form that depends on the representation of the object they act upon. Taking the gauge field as the defining (fundamental) representation, the form of the gauge-covariant derivatives in (1.1) is  $D_\mu = \partial_\mu - iG_\mu$ . But in other situations to be reviewed, it is a bit more complicated than this. (In general, for compactness, we scale the interaction charge strength  $g$  into the gauge field via  $gG_\mu \rightarrow G_\mu$ . This  $g$  can always be extracted back out when explicitly needed.) So, applying  $D_\mu = \partial_\mu - iG_\mu$  in (1.1), we may write:

$$\begin{aligned} iD_\mu(D_\nu\phi) &= i(\partial_\mu - iG_\mu)((\partial_\nu - iG_\nu)\phi) = i\partial_\mu(\partial_\nu\phi - iG_\nu\phi) + G_\mu(\partial_\nu\phi - iG_\nu\phi), \\ &= i\partial_\mu\partial_\nu\phi + \partial_\mu G_\nu\phi + G_\nu\partial_\mu\phi + G_\mu\partial_\nu\phi - iG_\mu G_\nu\phi \end{aligned} \quad (1.2)$$

as well as the reverse-signed, transposed-indexed:

$$-iD_\nu(D_\mu\varphi) = -i\partial_\nu\partial_\mu\varphi - \partial_\nu G_\mu\varphi - G_\mu\partial_\nu\varphi - G_\nu\partial_\mu\varphi + iG_\nu G_\mu\varphi. \quad (1.3)$$

Using (1.2) and (1.3) in (1.1) then yields:

$$F_{\mu\nu}\varphi \equiv i[D_\mu, D_\nu]\varphi = iD_\mu(D_\nu\varphi) - iD_\nu(D_\mu\varphi) = i[\partial_\mu, \partial_\nu]\varphi + \partial_{[\mu}G_{\nu]}\varphi - i[G_\mu, G_\nu]\varphi. \quad (1.4)$$

In flat spacetime where  $R^\sigma{}_{\alpha\mu\nu}A_\sigma \equiv [\partial_{;\mu}, \partial_{;\nu}]A_\alpha = [\partial_\mu, \partial_\nu]A_\alpha = 0$  and removing the arbitrary operand field  $\varphi$ , the above becomes the more familiar:

$$F_{\mu\nu} = \partial_{[\mu}G_{\nu]} - i[G_\mu, G_\nu] = (\partial_{[\mu} - iG_{[\mu})G_{\nu]} = D_{[\mu}G_{\nu]}. \quad (1.5)$$

Again,  $D_\mu = \partial_\mu - iG_\mu$  above is the gauge-covariant derivative when it acts upon gauge field objects  $G_\nu$  in the fundamental representation, but in general, when operating on other representations, it is a bit more complicated as we shall now see.

If the gauge fields commute, i.e., if  $[G_\mu, G_\nu] = 0$ , then (1.5) reduces to  $F_{\mu\nu} = \partial_{[\mu}G_{\nu]} = \partial_\mu G_\nu - \partial_\nu G_\mu$  and the gauge theory is known as an *abelian* gauge theory. If the gauge fields do *not* commute,  $[G_\mu, G_\nu] \neq 0$ , then (1.5) becomes the field strength for a *non-abelian* gauge theory, often also referred to as Yang-Mills [5] gauge theory.

Using differential forms, we may write the abelian field strength as:

$$F = \frac{1}{2!}F_{\mu\nu}dx^\mu \wedge dx^\nu = \frac{1}{2!}\partial_{[\mu}G_{\nu]}dx^\mu \wedge dx^\nu = \partial_\mu G_\nu dx^\mu \wedge dx^\nu = dG. \quad (1.6)$$

In general, the wedge product  $dx^\mu \wedge dx^\nu = dx^\mu dx^\nu - dx^\nu dx^\mu = [dx^\mu, dx^\nu]$  is antisymmetric under adjacent index interchange, and the differential elements are anticommuting,  $dx^\mu dx^\nu = -dx^\nu dx^\mu$ . So, by inspection from (1.5) in view of (1.6), the non-abelian field strength is:

$$F = \frac{1}{2!}F_{\mu\nu}dx^\mu \wedge dx^\nu = \frac{1}{2!}(\partial_{[\mu}G_{\nu]} - i[G_\mu, G_\nu])dx^\mu \wedge dx^\nu = dG - i[G, G] \equiv DG. \quad (1.7)$$

Here, compacted into differential forms, the gauge-covariant derivative is not separable from its operand as was  $D_\mu = \partial_\mu - iG_\mu$  when operating on  $G_\nu$  in (1.1) to (1.5), but rather involves the commutator of  $G$  with the operand which, in this case, just so happens to also be  $G$ . That is, it involves  $[G, G]$ . This in fact reveals the more-general form of the gauge-covariant derivative as we shall review next.

Now, focusing on non-abelian gauge theories, we introduce a set of traceless Hermitian generators  $t^i = t^{\dagger i}$  which form a closed group under multiplication via  $[t^i, t^j] = if^{ijk}t^k$ , where  $f^{ijk}$  are the group structure constants and are antisymmetric under the transposition of any two adjacent indexes. For any simple group SU(N), the internal symmetry indexes of the adjoint representation  $i, j, k = 1 \dots N^2 - 1$ . We may then define  $F_{\mu\nu} \equiv t^k F^k_{\mu\nu}$  and  $G_\mu \equiv t^i G^i_\mu$  and use these in (1.5) to expand:

$$F_{\mu\nu} = t^k F^k_{\mu\nu} = \partial_{[\mu} G_{\nu]} - i[G_\mu, G_\nu] = t^k \partial_{[\mu} G^k_{\nu]} - i[t^i, t^j] G^i_\mu G^j_\nu = t^k \partial_{[\mu} G^k_{\nu]} + f^{ijk} t^k G^i_\mu G^j_\nu. \quad (1.8)$$

Factoring out  $t^k$  this simplifies to the recognizable:

$$F^k_{\mu\nu} = \partial_{[\mu} G^k_{\nu]} + f^{ijk} G^i_\mu G^j_\nu. \quad (1.9)$$

Now, for illustration, let us momentarily consider the situation where the  $t^i$  are one half ( $1/2$ ) times the three (3) Pauli spin matrix generators of SU(2),  $t^i = \frac{1}{2}\sigma^i$ , so that  $f^{ijk}$  simply becomes the rank-3 Levi-Civita tensor,  $f^{ijk} \rightarrow \epsilon^{ijk}$ , which again, is antisymmetric in all indexes. In spacetime, if we were to write  $\epsilon^{ijk} A^i B^j$  for any two vectors  $A^i$  and  $B^j$  and were to regard  $i, j, k$  as indexes for the *space* dimensions  $x, y, z$ , then, for example,  $\epsilon^{ij3} A^i B^j = A^1 B^2 - A^2 B^1 = (\mathbf{A} \times \mathbf{B})^3$  is the z-component of the cross product  $\mathbf{A} \times \mathbf{B}$ , and more generally,  $\epsilon^{ijk} A^i B^j = (\mathbf{A} \times \mathbf{B})^k$ . But of course, the  $i, j, k$  indexes in (1.9) are not space indexes, but are *internal symmetry* indexes. So rather than using the cross-product symbol “ $\times$ ” which is used for vectors in physical space, and because we still wish to be able compactly represent the fundamentally-antisymmetric character of  $f^{ijk}$  in the form of a “cross-like product” in internal symmetry space, we instead employ the wedge symbol “ $\wedge$ .” Although  $G^i_\mu$  and  $G^j_\nu$  in (1.9) both are gauge fields  $G$ , they have different spacetime indexes  $\mu$  and  $\nu$ , so we may still think of them as two different vectors just like  $A^i$  and  $B^j$  above. So analogously to  $\epsilon^{ijk} A^i B^j = (\mathbf{A} \times \mathbf{B})^k$  in the three space dimensions of spacetime, we write  $f^{ijk} G^i_\mu G^j_\nu = (G_\mu \wedge G_\nu)^k$  in internal symmetry space. Then, we use this in (1.9) to write  $F^k_{\mu\nu} = \partial_{[\mu} G^k_{\nu]} + (G_\mu \wedge G_\nu)^k$ . Because the general form of this equation holds in SU(N) for each of the indexes  $k = 1 \dots N^2 - 1$ , we may suppress the  $k$  index throughout to write:

$$F_{\mu\nu} = \partial_{[\mu} G_{\nu]} + G_\mu \wedge G_\nu. \quad (1.10)$$

Then, compacting (1.10) to differential forms as in (1.6), we have:

$$F = \frac{1}{2!} F_{\mu\nu} dx^\mu \wedge dx^\nu = \frac{1}{2!} (\partial_{[\mu} G_{\nu]} + G_\mu \wedge G_\nu) dx^\mu \wedge dx^\nu = dG + G \wedge G = (d + G \wedge) G \equiv DG. \quad (1.11)$$

Now, Jaffe and Witten point out at pages 1 and 2 of [6], that:

“If  $A$  denotes the  $U(1)$  gauge connection, locally a one-form on space-time, then the curvature or electromagnetic field tensor is the two-form  $F = dA$  [see (1.6) above], and Maxwell’s equations in the absence of charges and currents read  $0 = dF = d * F$  .”

They then proceed to explain that in “non-abelian gauge theory”:

“at the classical level one replaces the gauge group  $U(1)$  of electromagnetism by a compact gauge group  $G$ . The definition of the curvature arising from the connection must be modified to  $F = dA + A \wedge A$  and Maxwell’s equations are replaced by the Yang–Mills equations,  $0 = d_A F = d_A * F$ , where  $d_A$  is the gauge-covariant extension of the exterior derivative.”

Equation (1.11) is precisely  $F = dA + A \wedge A$  with the gauge field simply renamed from  $A$  to  $G$ , and what Jaffe and Witten write above is a condensed explanation for what we have laid out above in equations (1.1) through (1.11). When we use the generalized one-form  $G$  and two-form  $F$  without any particular generator set  $t^i$ , then the differential forms equation is written as  $F = dG - i[G, G]$  in (1.7). But when one does introduce a set of group generators  $t^i$  and the antisymmetric structure constants  $f^{ijk} \rightarrow \wedge$ , the differential forms equation is  $F = dG + G \wedge G$  in (1.11). To display the particular  $i = 1 \dots N^2 - 1$  field components for a compact simple gauge group  $SU(N)$ , this equation is  $F^i = dG^i + (G \wedge G)^i$ . So  $F = dG - i[G, G]$  (commutator form) and  $F = dG + G \wedge G$  (wedge form) are just alternative ways of saying the same thing. But a benefit of the wedge form is that we may write  $F = (d + G \wedge)G \equiv DG$  so as to define a gauge-covariant derivative  $D \equiv d + G \wedge$  ( $= d_A$ ) in a form which is fully-separable from its operand, and which is generally applicable to *any and all operands*. We will find it useful in general to develop both these forms.

Indeed, the reason we have gone through the exercise of (1.8) through (1.11), is to explore the question of how one generally performs  $d_A = D$ , independently of its operand, “where  $d_A$  is the gauge-covariant extension of the exterior derivative.” That is, we want to be able to generalize the taking of these derivatives, and especially, to ascertain the correct way to derive the equations  $*J = d_A * F = D * F$  and  $P = d_A F = DF$  in the presence of the electric and magnetic three-form charge densities  $*J$  and  $P$ .

Specifically, as already stated, if we write equation (1.11) as  $F = (d + G \wedge)G \equiv DG$  with  $D \equiv d + G \wedge$ , we find that  $D \equiv d + G \wedge$  is in fact the generalized definition of the gauge-covariant derivative which tells us how to take higher-rank gauge derivatives, independent of the representation of the operand. Thus, the Maxwell equations for Yang-Mills theory, with electric and magnetic sources, in differential forms, where  $t^i$  and  $f^{ijk}$  are specified, with index  $i$

suppressed, for SU(N), where we use the duality operator  $*$ , and with  $F = dG + G \wedge G$ , are merely the  $i = 1 \dots N^2 - 1$  equations:

$$\begin{aligned}
 *J &= D*F = D*DG = (d + G \wedge)*F = d*F + G \wedge *F = d*(dG + G \wedge G) + G \wedge *(dG + G \wedge G) \\
 &= d*dG + d*(G \wedge G) + G \wedge *dG + G \wedge *(G \wedge G) \\
 P &= DF = DDG = (d + G \wedge)F = dF + G \wedge F = d(dG + G \wedge G) + G \wedge (dG + G \wedge G) \\
 &= ddG + d(G \wedge G) + G \wedge dG + G \wedge G \wedge G
 \end{aligned} \tag{1.12}$$

The duality operator  $*$  was first developed by Reinich [7] and later elaborated by Wheeler [8], and it makes integral use of the Levi-Civita tensor as laid out in [9] at pages 87-89.

In this paper, we shall develop the classical Yang-Mills magnetic monopole density  $P$  and a related “faux” magnetic charge density  $P'$  in detail, and shall show how this related density  $P'$ , when analyzed using Gauss' / Stokes' theorem, appears to confine its gauge fields. Of course, baryons, which include the protons and neutrons at the heart of nuclear physics, also confine their gauge fields. So this will raise the question we thereafter explore in detail, whether these magnetic monopoles of Yang-Mills theory are in some fashion related to baryons.

## 2. Classical Field Equations for the Yang-Mills Magnetic Monopole

To further develop the monopole density  $P$ , first, akin to the derivation (1.1) through (1.5), we calculate the commutator:

$$\begin{aligned}
 [D_\sigma, F_{\mu\nu}] \varphi &= D_\sigma (F_{\mu\nu} \varphi) - F_{\mu\nu} D_\sigma \varphi = (\partial_\sigma - iG_\sigma)(F_{\mu\nu} \varphi) - F_{\mu\nu} (\partial_\sigma - iG_\sigma) \varphi \\
 &= \partial_\sigma F_{\mu\nu} \varphi + F_{\mu\nu} \partial_\sigma \varphi - iG_\sigma F_{\mu\nu} \varphi - F_{\mu\nu} \partial_\sigma \varphi + iF_{\mu\nu} G_\sigma \varphi = \partial_\sigma F_{\mu\nu} \varphi - i[G_\sigma, F_{\mu\nu}] \varphi
 \end{aligned} \tag{2.1}$$

We *can* use  $D_\sigma = \partial_\sigma - iG_\sigma$  in the above, precisely because this is a commutator, and so the gauge field will be commuted with the operand  $F_{\mu\nu}$  as in  $F = dG - i[G, G]$  a.k.a.  $F = dG + G \wedge G$ . Removing  $\varphi$  we see that (2.1) contains the useful identity:

$$[D_\sigma, F_{\mu\nu}] = \partial_\sigma F_{\mu\nu} - i[G_\sigma, F_{\mu\nu}] = D_\sigma F_{\mu\nu}, \tag{2.2}$$

with the commutator included in the gauge-covariant derivative. In differential wedge form this is  $DF = (d + G \wedge)F$ , which is part of the monopole density in (1.12). Then, combining (2.2) with (1.1) in the form  $F_{\mu\nu} = i[D_\mu, D_\nu]$  first yields:

$$D_\sigma F_{\mu\nu} = [D_\sigma, F_{\mu\nu}] = i[D_\sigma, [D_\mu, D_\nu]] \tag{2.3}$$

containing an anticommuting succession of gauge-covariant derivatives. This in turn means that the index-cyclical combination:

$$P_{\sigma\mu\nu} = D_\sigma F_{\mu\nu} + D_\mu F_{\nu\sigma} + D_\nu F_{\sigma\mu} = i\left(\left[D_\sigma, \left[D_\mu, D_\nu\right]\right] + \left[D_\mu, \left[D_\nu, D_\sigma\right]\right] + \left[D_\nu, \left[D_\sigma, D_\mu\right]\right]\right) = 0, \quad (2.4)$$

by the Jacobian identity  $\left[a, \left[b, c\right]\right] + \left[b, \left[c, a\right]\right] + \left[c, \left[a, b\right]\right] = 0$ . So we see that the *Yang-Mills magnetic monopole densities vanish, just like those of abelian gauge theory*. Consequently, we can append  $P = 0$  from (2.4) onto (1.12), and so write  $P = DF = DDG = 0$ . This is the non-abelian analog to the abelian  $ddG = 0$ .

But there is another zero in the monopole  $P$  of (1.12), and that is the zero which comes from this very same abelian  $ddG = 0$ . This is rooted in the geometric relationship  $dd = \mathbf{0}$  of exterior calculus in spacetime: “the exterior derivative of an exterior derivative is zero.” In general in this paper, we shall highlight the zero of  $dd = \mathbf{0}$  to distinguish it from the (not highlighted) zero of the Jacobian identity  $DDG = 0$  which is established by the combination of (1.12) and (2.4). The highlighted zero in  $dd = \mathbf{0}$  is a “subset” identity contained within (1.12), which we may now rewrite as:

$$\begin{aligned} 0 = P = DF = DDG = ddG + d(G \wedge G) + G \wedge dG + G \wedge G \wedge G \\ = \mathbf{0} + d(G \wedge G) + G \wedge dG + G \wedge G \wedge G \end{aligned} \quad (2.5)$$

Of course, in an abelian gauge theory such as Maxwell’s electrodynamics where  $\left[G_\mu, G_\nu\right] = 0$  so that  $F_{\mu\nu} = \partial_{[\mu} G_{\nu]}$  in (1.5) thus  $F = dG$ , the Magnetic monopole densities are themselves specified by  $P_{\text{abelian}} = dF = ddG = \mathbf{0}$ . This means that the Yang-Mills monopole density in (2.5), although it too is equal to zero, contains a number of apparently non-zero terms embedded within, as well as the term  $ddG = \mathbf{0}$  which we associate with the vanishing monopoles of electrodynamics. This will be very important to keep in mind as we develop this monopole, because this “abelian subset” embedding of  $ddG = \mathbf{0}$  within (2.5) will be directly responsible for *confining* the gauge fields within the Yang-Mills monopole, and will lead us to consider whether there is some connection between Yang-Mills monopoles and baryons.

Next let us ascertain the commutator form for the monopole (2.5). Via the exact same type of calculation we used to turn (1.5) a.k.a. (1.7) into (1.11), one may demonstrate that  $P = DF = dF - i[G, F]$  is equivalent to  $P = DF = (d + G \wedge)F$ . So, combining the former,  $P = DF = dF - i[G, F]$ , with  $F = DG = dG - i[G, G]$  from (1.7) a.k.a.  $F = DG = (d + G \wedge)G$  from (1.11), we may translate (2.5) into the commutator expression:

$$\begin{aligned} P = DF = DDG = dF - i[G, F] &= d(dG - i[G, G]) - i[G, dG - i[G, G]] \\ &= ddG - id[G, G] - i[G, dG] - [G, [G, G]] \\ &= \mathbf{0} - id[G, G] - i[G, dG] - [G, [G, G]] = 0 \end{aligned} \quad (2.6)$$

Let us now expand (2.6) above into tensor components term-by-term, and then do some additional reductions. For  $P$  and  $-id[G, G]$  we have:

$$P = \frac{1}{3!} P_{\sigma\mu\nu} dx^\sigma \wedge dx^\mu \wedge dx^\nu = P_{\sigma\mu\nu} dx^\sigma dx^\mu dx^\nu, \quad (2.7)$$

$$\begin{aligned} -id[G, G] &= -\frac{1}{3!} i \left( \partial_\sigma [G_\mu, G_\nu] + \partial_\mu [G_\nu, G_\sigma] + \partial_\nu [G_\sigma, G_\mu] \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= -\frac{1}{2!} i \partial_\sigma [G_\mu, G_\nu] dx^\sigma \wedge dx^\mu \wedge dx^\nu = -i \partial_\sigma (G_\mu G_\nu) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= -i (\partial_\sigma G_\mu G_\nu + G_\mu \partial_\sigma G_\nu) dx^\sigma \wedge dx^\mu \wedge dx^\nu = (-i \partial_\sigma G_\mu G_\nu + i G_\sigma \partial_\mu G_\nu) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= -idGG + iGdG \end{aligned} \quad (2.8)$$

The sign reversal in the third line of (2.8) reveals the identity  $d[G, G] = dGG - GdG$ , in contrast to scalar product rule  $d(a \cdot b) = da \cdot b + a \cdot db$ . For  $-i[G, dG]$  in (2.6) we further have:

$$\begin{aligned} -i[G, dG] &= -\frac{1}{3!} i \left( [G_\sigma, \partial_{[\mu} G_{\nu]}] + [G_\mu, \partial_{[\nu} G_{\sigma]}] + [G_\nu, \partial_{[\sigma} G_{\mu]}] \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= -\frac{1}{2!} i [G_\sigma, \partial_{[\mu} G_{\nu]}] dx^\sigma \wedge dx^\mu \wedge dx^\nu = -i [G_\sigma, \partial_\mu G_\nu] dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= -i (G_\sigma \partial_\mu G_\nu - \partial_\mu (G_\nu G_\sigma)) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= -i (G_\sigma \partial_\mu G_\nu - G_\nu \partial_\mu G_\sigma - \partial_\mu G_\nu G_\sigma) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= (-2i G_\sigma \partial_\mu G_\nu + i \partial_\sigma G_\mu G_\nu) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= -2iGdG + idGG \end{aligned} \quad (2.9)$$

in which the  $GdG$  doubles by a similar sign reversal in the fifth line. Finally, by the Jacobian identity  $[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0$ , for  $[G, [G, G]]$  in (2.6), we find (cf. (2.4)) that:

$$-[G, [G, G]] = -\frac{1}{3!} \left( [G_\sigma, [G_\mu, G_\nu]] + [G_\mu, [G_\nu, G_\sigma]] + [G_\nu, [G_\sigma, G_\mu]] \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu = 0, \quad (2.10)$$

In (2.6), we then use  $-id[G, G] = -idGG + iGdG$  and  $-i[G, dG] = -2iGdG + idGG$  and  $-[G, [G, G]] = 0$  from (2.8) to (2.10) to restructure and consolidate the monopole density as much as possible while retaining n Gauss / Stokes integrable  $d[G, G]$  term, into:

$$\begin{aligned} P &= \mathbf{0} - id[G, G] - i[G, dG] \\ &= \mathbf{0} - idGG + iGdG - 2iGdG + idGG \\ &= \mathbf{0} - iGdG \\ &= -\mathbf{0} + id[G, G] - idGG = 0 \end{aligned} \quad (2.11)$$

This in turn reveals the additional identities  $d[G, G] = dGG$  and  $GdG = 0$ . The former identity  $d[G, G] = dGG$  will be very important in the development to follow, and will be shown to be the density for a baryon including the proton and neutron flavors of baryon.

Now, of central interest in the discussion to follow, the monopole density in the final line above contains a Gauss/Stokes-integrable term  $d[G, G]$  (and the  $\mathbf{0} = ddG$ ) together with the non-integrable term  $dGG$ . Applying Gauss' / Stokes Theorem  $\iint dX = \oint X$  for any differential form  $X$  to the final line above, we may ascertain the classical surface flux associated with this non-abelian magnetic monopole, namely:

$$\begin{aligned} \iiint P &= \iiint (-ddG + id[G, G] - idGG) = \iiint (-\mathbf{0} + id[G, G] - idGG) \\ &= -\oint dG + i\oint [G, G] - i\iiint dGG = -\mathbf{0} + i\oint [G, G] - i\iiint dGG = 0 \end{aligned} \quad (2.12)$$

By then writing (2.12) using the not-highlighted 0 of  $\iiint P = 0$  rooted in the Jacobian identity (2.4) as:

$$\begin{aligned} -\oint dG + i\oint [G, G] &= i\iiint dGG \\ -\mathbf{0} + i\oint [G, G] &= i\iiint dGG \end{aligned} \quad (2.13)$$

we clearly see the relationship between what is contained within the three-dimensional volume  $\iiint$  and what net flows through the closed two-dimensional surface  $\oint$  enclosing that volume. Now, we wish to interpret what is being taught by (2.13).

### 3. Confinement of Gauge fields within, and the Composite Nature of, Yang-Mills Magnetic Monopoles

We start with the term  $\oint dG = \mathbf{0}$  which is embedded in (2.13). In electrodynamics, Gauss' law for magnetism and Faraday's law are both contained within:

$$\iiint P = \iiint dF = \iiint ddG = \oint F = \oint F^{\mu\nu} dx_\mu dx_\nu = \oint dG = \mathbf{0}. \quad (3.1)$$

At rest, this tells us that while magnetic fields may flow across some surfaces, there is never a *net* flux of a magnetic field through any *closed* two dimensional surface. In the form  $P = dF = ddG = \mathbf{0}$ , this simply says there are no observed magnetic charges. So how might we interpret the presence of  $\oint dG = \mathbf{0}$  as *one of the terms* among a number of *non-vanishing* terms in equations (2.12) and (2.13) for the Yang-Mills magnetic monopoles?

To find out, let us return to the *non-abelian, Yang-Mills* field strength (1.5), namely  $F_{\mu\nu} = \partial_{[\mu} G_{\nu]} - i[G_{\mu}, G_{\nu}]$ , and rewrite this using the differential forms equation:

$$\begin{aligned} \oint\!\!\!\oint F &= \frac{1}{2!} \oint\!\!\!\oint F_{\mu\nu} dx^{\mu} \wedge dx^{\nu} = \oint\!\!\!\oint \partial_{\mu} G_{\nu} dx^{\mu} \wedge dx^{\nu} - \frac{1}{2!} i \oint\!\!\!\oint [G_{\mu}, G_{\nu}] dx^{\mu} \wedge dx^{\nu} \\ &= \oint\!\!\!\oint dG - i \oint\!\!\!\oint [G, G] = \mathbf{0} - i \oint\!\!\!\oint [G, G] \end{aligned} \quad (3.2)$$

We may then use (3.2) to rewrite (2.13) with a sign reversal as:

$$\begin{aligned} \oint\!\!\!\oint F &= -i \oint\!\!\!\oint [G, G] = -i \iiint dGG \\ &= -i \iiint \frac{1}{3!} (\partial_{\sigma} [G_{\mu}, G_{\nu}] + \partial_{\mu} [G_{\nu}, G_{\sigma}] + \partial_{\nu} [G_{\sigma}, G_{\mu}]) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} . \\ &= -i \iiint \frac{1}{3!} (\partial_{[\sigma} G_{\mu]} G_{\nu} + \partial_{[\mu} G_{\nu]} G_{\sigma} + \partial_{[\nu} G_{\sigma]} G_{\mu}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \neq 0 \end{aligned} \quad (3.3)$$

So, while (3.1) tells us that there is no net magnetic field flux over of any closed surface in *abelian* electrodynamics, (3.3) tells us that in non-Abelian, Yang-Mills gauge theory, there is indeed a *non-vanishing* net flux across closed surfaces,  $\oint\!\!\!\oint F \neq 0$ , of *whatever the Yang-Mills analog is to an ordinary abelian magnetic field*.

Now, we have a puzzle: any time we see a term  $\oint\!\!\!\oint F$ , we know that we are talking about a magnetic monopole, and that whatever is contained within the associated volume integral is a magnetic charge. Indeed, (3.3) may be thought of as *the very definition of a magnetic charge*, which in (3.3) is *not* zero. At the same time, we found in (2.4) a.k.a. (2.6) that  $P = DF = DDG = 0$ , which is to say, that the magnetic charge density is zero, just as it is in electrodynamics. So if  $P = DF = DDG = 0$  but  $\oint\!\!\!\oint F \neq 0$ , how do we reconcile the former equation which says the magnetic charge density is zero with the latter equation which says there is a non-zero magnetic charge?

One way to think this through, is take the Yang-Mills electric charge field equation (1.12),  $*J = D*F$ , revert this (merely for pedagogic simplicity) to its abelian form  $*J = d*F$  which contains Gauss' law for electricity, and then apply Gauss' / Stokes' Theorem to obtain  $\oint\!\!\!\oint *F = \iiint *J$  ( $= \iiint d*F$ ). Just as  $\oint\!\!\!\oint F$  in the rest frame represents a net flux of magnetic field through a closed surface,  $\oint\!\!\!\oint *F$  in the rest frame represents a net flux of electric field through a closed surface. And this  $\oint\!\!\!\oint *F$  then becomes the very definition of the *electric* charge. But here, electric charge density is defined by  $*J$  inside  $\iiint *J$ , while in (3.3) magnetic charge density is defined by  $-idGG$  inside  $-i \iiint dGG$ . That is, we have a magnetic charge density  $-idGG$  which we need to think about in comparison to an electric charge density  $*J$ .

The answer to this puzzle is that the magnetic charge density in (3.3) is *not* the  $P$  of  $P = DF = DDG = 0$ , it is the  $P' \equiv -idGG$  which, via (2.11) can be extended to  $P' = -id[G, G] = -idGG$ . The magnetic charge as defined by the enclosure surface  $\oint\!\!\!\oint F$  is a three-form just like  $*J$  and  $P$ , but it is not an *elementary* three-form source. Rather, it is a three-form constructed from  $-idGG$  which includes some dynamical behavior of the gauge fields inside the volume integral. That is, the magnetic charge  $P' = -id[G, G] = -idGG$  is a *composite three-form* built out of gauge fields, rather than an elementary three form like the abelian electric charge source  $*J$ . Indeed, we may take this a step further:

In electrodynamics, the three-form  $*J$  which in tensor language is related to the electric source current density vector  $J^\alpha$  by  $*J_{\sigma\mu\nu} = (-g)^5 \epsilon_{\alpha\sigma\mu\nu} J^\alpha$ , is a *true electric source* which then gives rise to gauge fields in abelian gauge theory via  $*J = d*F = d* dG$ , and per (1.12), via  $*J = D*F = D* DG$  in Yang-Mills gauge theory. On the other hand, the  $P' = -idGG$  in (3.3), written in tensor form as  $P'_{\sigma\mu\nu} = -i(\partial_{[\sigma} G_{\mu]} G_\nu + \partial_{[\mu} G_{\nu]} G_\sigma + \partial_{[\nu} G_{\sigma]} G_\mu)$  and converted over to a one form via the related general identities  $*P'^\alpha = \frac{1}{3!}(-g)^{-5} \epsilon^{\sigma\mu\nu\alpha} P'_{\sigma\mu\nu}$  and  $*\partial^{[\nu} G^{\alpha]} = \frac{1}{2}(-g)^{-5} \epsilon^{\sigma\mu\nu\alpha} \partial_{[\sigma} G_{\mu]}$ , will result in a *faux magnetic source*:

$$\begin{aligned}
 *P'^\alpha &= \frac{1}{3!}(-g)^{-5} \epsilon^{\sigma\mu\nu\alpha} P'_{\sigma\mu\nu} = -(-g)^{-5} \epsilon^{\sigma\mu\nu\alpha} \frac{1}{3!} i (\partial_{[\sigma} G_{\mu]} G_\nu + \partial_{[\mu} G_{\nu]} G_\sigma + \partial_{[\nu} G_{\sigma]} G_\mu) \\
 &= -(-g)^{-5} \frac{1}{3} i \left( \frac{1}{2} \epsilon^{\sigma\mu\nu\alpha} \partial_{[\sigma} G_{\mu]} G_\nu + \frac{1}{2} \epsilon^{\sigma\mu\nu\alpha} \partial_{[\mu} G_{\nu]} G_\sigma + \frac{1}{2} \epsilon^{\sigma\mu\nu\alpha} \partial_{[\nu} G_{\sigma]} G_\mu \right) \\
 &= -\frac{1}{3} i \left( *\partial^{[\nu} G^{\alpha]} G_\nu + *\partial^{[\sigma} G^{\alpha]} G_\sigma + *\partial^{[\mu} G^{\alpha]} G_\mu \right) \\
 &= -i *\partial^{[\sigma} G^{\alpha]} G_\sigma
 \end{aligned} \tag{3.4}$$

which is constructed solely out of gauge fields  $G$  which themselves are sourced by  $*J = D*F = D* DG$ . So, there is only one *elementary* source  $J$ , not two sources  $J$  and  $P$ . From this one source  $J$ , gauge fields  $G$  are emitted from interaction vertices. From these gauge fields  $G$ , a faux magnetic source  $P' = -idGG$  is assembled. And finally, from this faux magnetic source,  $\oint\!\!\!\oint F \neq 0$  flows across closed surfaces as in (3.3). The electric source  $J^\alpha$ , whether in abelian or non-abelian gauge theory, has its own independent existence, and it is the source of any and all gauge fields. But the faux magnetic source charge in (3.3) has *no independent existence* apart from the gauge fields  $G$ . Rather, it is built out of the gauge fields. *So the Yang-Mills monopoles are composite, not elementary, objects.* And, by the way, so too are baryons.

Having resolved the puzzle of how to reconcile  $P = DF = DDG = 0$  with  $\oint\!\!\!\oint F \neq 0$ , we next pose the following question: what happens to the total flux  $\oint\!\!\!\oint F$  in (3.2) under the local gauge-like transformation  $F^{\mu\nu} \rightarrow F'^{\mu\nu} = F^{\mu\nu} - \partial^{[\nu} G^{\mu]}$ ? In differential forms, this transformation is  $F \rightarrow F' = F - dG$ , which means, precisely because  $\oint\!\!\!\oint dG = \mathbf{0}$ , that:

$$\oint\!\!\!\oint F \rightarrow \oint\!\!\!\oint F' = \oint\!\!\!\oint (F - dG) = \oint\!\!\!\oint F, \quad (3.5)$$

So, the net surface flux in the monopole equation (3.3) is *invariant* under the transformation  $F^{\mu\nu} \rightarrow F'^{\mu\nu} = F^{\mu\nu} - \partial^{[\nu} G^{\mu]}$ , which means that the gauge field is *not observable* with respect to net flux across closed surfaces of the monopole. The abelian expression  $\oint\!\!\!\oint dG = \mathbf{0}$ , expanded to show the Riemann tensor, may be written as  $\oint\!\!\!\oint F = \oint\!\!\!\oint dG = \iiint R^{\tau}_{\nu\sigma\mu} G_{\tau} dx^{\sigma} dx^{\mu} dx^{\nu} = \mathbf{0}$ , and explicitly shows how individual gauge fields  $G_{\tau}$  couple with spacetime geometry as represented by  $R^{\tau}_{\nu\sigma\mu}$ . This represents an *absence* of monopoles in electrodynamics, and yields the *symmetry principle* (3.5) for the behavior of magnetic monopoles in Yang-Mills theory generally.

But if the non-zero flux in the Yang-Mills monopole equation (3.3) is invariant under the gauge-like transformation  $F^{\mu\nu} \rightarrow F'^{\mu\nu} = F^{\mu\nu} - \partial^{[\nu} G^{\mu]}$  which means that the gauge fields  $G^{\mu}$  are not net observables over a closed monopole surface, *this would seem to suggest that the Yang-Mills monopole inherently confine their gauge fields*. This is another hint that the monopole equation (3.3) could be the classical field equation for a baryon, in integral form.

The final point is that because the *faux* magnetic source  $P' = -idGG$  is constructed out of gauge fields, and because the gauge fields are in turn sourced by  $*J = D*F = D*DG$ , and because electric sources may be represented in vector form in terms of Dirac fermion wavefunctions  $\psi$  via  $J^{\mu} = \bar{\psi}\gamma^{\mu}\psi$ , it should be possible in principle, and would certainly be desirable in practice, to rewrite the *faux* magnetic source  $-idGG$  in terms of the *true* source currents  $J^{\mu}$  from which they arise, and then to rewrite the  $J^{\mu} = \bar{\psi}\gamma^{\mu}\psi$  in terms of their fermion wavefunctions  $\psi$ . The upshot of all this, is that while  $\oint\!\!\!\oint F$  in (3.3) is presently expressed in terms of gauge fields as  $\oint\!\!\!\oint F(G)$ , once we obtain the gauge fields  $G(J)$  in terms of sources and the sources  $J(\psi)$  in terms of fermions, we will end up with  $\oint\!\!\!\oint F(G(J(\psi)))$ . Then, if we happen to find more than one fermion (maybe even three fermions) within the enclosed  $\oint\!\!\!\oint F$  “system” in its “ground” state, we would need to apply the Exclusion Principle of Fermi-Dirac-Pauli statistics to maintain the  $\psi$  in distinct quantum eigenstates, which would give us the opportunity, for example, to introduce a color degree of freedom to do so and thus make a connection to  $SU(3)_C$  Chromodynamics, with  $\oint\!\!\!\oint F(G(J(\psi_R, \psi_G, \psi_B)))$ . So this means that the Yang-Mills monopoles are not only composite objects, but are composite objects which contain fermions and gauge fields, and that these fermions will need to obey some form of quantum exclusion which may include  $SU(3)_C$ . And, by the way, all of the same the same is true of baryons, and as to fermion exclusion, quarks.

It is for these reasons, that it may be fruitful to entertain the prospect that (3.3) is not only the classical field equation for a Yang-Mills magnetic monopole, but may be synonymous with the classical field equation for a baryon. All of the development in sections 5 through 10 serves the singular purpose of proving that this is true. But first, we need to discuss whether a classical

analysis along the lines of (3.3) can really teach us anything useful about baryons and confinement.

#### 4. Can a Classical Field Equation Really Teach us Anything Useful about Baryons and Confinement?

Given that (3.3) is a classical field equation, we must pose the question whether such a classical equation can really have anything of interest to say about baryons and confinement, which have many features that arise only out of quantum field theory. For example, it might be observed that a classical analysis which seeks to understand baryons and confinement in no way takes account of quantum field theory with operator-valued fields. This, it might be argued, is despite the fact that there are many reasons to believe confinement and the existence of a mass gap are related to the running of the coupling constant, which is an inherently quantum effect.

Certainly, (3.3) above is a completely classical field equation, not yet taking into account any aspects (or the need to prove existence) of a non-trivial relativistic quantum Yang–Mills theory on  $\mathcal{R}^4$  [6]. And, of course, there are many reasons to believe that confinement is related to the running of the strong coupling constant, which is an inherently quantum effect, and which manifests in asymptotic freedom at “ultraviolet” energy and infrared slavery at low energy [10]. However, just like electrodynamics, Yang-Mills gauge theory has a classical formulation and (is expected once quantum Yang-Mills existence is proven, to have) a quantum field formulation. This means that (3.3) may reveal inherently-confining attributes for the magnetic monopoles of Yang-Mills gauge theory which appear at the classical level and which are rooted in the relationship  $dd = \mathbf{0}$  of Riemannian spacetime exterior geometry, as well as inherently-composite attributes expressed by  $\oint\!\!\!\oint F(G(J(\psi)))$ . That opens up the question how these same attributes translate through to quantum Yang-Mills theory.

Specifically, *if* in fact (3.3) for  $\oint\!\!\!\oint F$  is an equation for baryon-like gauge field confinement properties of Yang-Mills magnetic monopoles based upon their abelian-subset behaviors rooted in the classical equation  $ddG = \mathbf{0}$  and its integral form  $\oint\!\!\!\oint dG = \mathbf{0}$  and the consequent symmetry (3.5), and if the composite faux magnetic charge  $P' = -idGG$  in (3.3) in some way represents a baryon charge, then the classical baryons that would be represented by (3.3) would not suddenly become “not baryons” in quantum field theory. Rather, there would *two sets of behaviors* that need to be studied: a) how these monopoles behave in a classical formulation, which includes (3.3) and (3.5) above, and b) how these monopoles additionally behave in quantum field theory. So if we can demonstrate that the classical behaviors appear to be confining and appear to involve a non-elementary, composite charge that includes some amalgam of fermions and gauge fields, one should expect that this will “bleed” through to yield quantum amplitudes and running couplings and color symmetries that buttress, not defy, these classical behaviors, just as abelian magnetic monopoles do not suddenly appear and ordinary magnetic fields do not suddenly net flow through closed surfaces, once one goes from classical to quantum electrodynamics.

Further, one might take the perspective that the *cause* for confinement and baryon compositeness is the classical field equation (3.3) for a Yang-Mills monopole which has the symmetry (3.5), and that one of the *effects* of this is that in a quantum field treatment of these baryon monopoles, the strong coupling will weaken for ultraviolet and strengthen for infrared probes. And, it can be argued that this is a more natural approach than simply trying to figure out how to “glue” together disparate quarks into baryons *without knowing to begin with what sorts of covariant objects baryons actually are in spacetime*. Indeed, if the hints of baryons and confinement that arise in (3.3) and (3.5) are correct, then we would need to start thinking of baryons as third-rank antisymmetric tensors and related three-forms in spacetime governed by the classical equation (3.3) with the symmetry (3.5), and then see how that connects to everything else we know about baryons. The “let’s glue together the quarks” approach, notwithstanding many opportunities to do so, has thus far failed to explain why QCD “must have ‘quark confinement, that is, even though the theory is described in terms of elementary fields, such as the quark fields, that transform non-trivially under SU(3), the physical particle states—such as the proton, neutron, and pion—are SU(3)-invariant,” see [6] at page 3. This SU(3)-invariance of *physical particle states* is a *symmetry principle*, and while not every classical symmetry carries through to quantum field theory, for example, the chiral anomaly (e.g., [11], section IV.7), there is no apparent *a priori* reason to believe that whatever classical symmetries are found for these monopoles (such as (3.5)) will *only* manifest in the classical but not the quantum field theory. At the very least, the question for study becomes: do these symmetries carry over from classical to quantum field theory, and if not, why not, and in what manner are they altered? Further, if the baryon charge really is  $P' = -idGG$ , then as we turn  $\oint\!\!\!\oint F(G) \rightarrow \oint\!\!\!\oint F(G(J(\psi)))$ , so too would we turn  $P'(G) \rightarrow P'(G(J(\psi)))$ . This may reveal that the inherently-composite nature of this  $P' = -idGG$  charge is in fact the long-sought “glue” to aggregate quarks and gluons together into a single *charge system, ab initio*.

Additionally, approaching confinement starting from a classical treatment of baryons has validating precedent in the MIT Bag Model reviewed in, e.g., [12], section 18. Irrespective of the specifics of any particular bag-type model of confinement, the MIT Bag Model very correctly makes one very important point: *focus carefully on what flows and does not flow across any closed two-dimensional surface*. And it does so using the *classical* formulation of Gauss’ / Stokes’ theorem. This is why the integral form of Maxwell’s equations in classical field theory may well be a very sensible starting point studying confinement, because from the Bag Model viewpoint, confinement is all about what passes and does not pass through closed surfaces containing the extended field configuration within the baryon volume.

Further, by talking about the “classical level” of “non-abelian gauge theory” right on page 1 of [6], Jaffe and Witten themselves recognize that Yang-Mills theory *has a classical level*, and that a reasonable starting point for developing quantum Yang-Mills theory, is to first fully and properly develop and understand Yang-Mills gauge theory at this classical level.

Finally, it is certainly unrealistic to expect that a classical-only treatment of baryons based on Yang-Mills magnetic monopoles will explain *all* of the observed phenomenology of baryons. It cannot and will not. Only a proper quantum field treatment may be expected to do so. Yet, at the same time, there are some important physics insights to be gained even from a classical treatment of the Yang-Mills monopole equation (3.3). And we know, if we can fully

develop a classical theory on its own terms, and then obtain its Lagrangian density  $\mathcal{L}(\phi)$  and action  $S(\phi)$  in terms of its fields  $\phi$ , that we can then convert over to a quantum field theory via the path integration  $Z = \int D\phi \exp i \int \mathcal{L} d^4x = \int D\phi \exp iS$ . While carrying out the path integration of a non-linear theory such as Yang-Mills gauge theory (and especially gravitational theory) is still an exceptionally challenging problem, that does not mean one ought not make the effort to find the correct road for doing so, which road is revealed in section 8 and used to carry out an analytically-exact path integration in section 11. But this all this begins by finding and fleshing out, the right classical theory to quantize.

So what is most important is for researchers in particle, baryon and nuclear theory to be aware of the possibility of modelling baryons as Yang-Mills magnetic monopoles to gain possible insight into confinement and related QCD symmetries, so that this possible connection can be further developed, vetted, and empirically-tested by anyone who finds it interesting or promising. We now explore the next several steps in this development.

## 5. Classical Field Equations for the Yang-Mills Electric Charge

Now let us develop the electric charge density  $*J$  in (1.12). Once again, via the same type of calculation used to go from (1.5) a.k.a. (1.7) to (1.11), which was also used to go from (2.5) to (2.6), together with  $F = DG = dG - i[G, G]$ , we write (1.12) for  $*J$  in commutator form:

$$\begin{aligned} *J &= D*F = D*DG = d*F - i[G, *F] = d*(dG - i[G, G]) - i[G, *(dG - i[G, G])] \\ &= d*dG - id*[G, G] - i[G, *dG] - [G, *[G, G]] \end{aligned} \quad (5.1)$$

This should be contrasted with the analog for  $P$  in the middle line of (2.6). Above, however, we do not have all the zeroes that were in (2.6), namely,  $ddG = 0$ ,  $P = 0$ , and  $[G, [G, G]] = 0$ .

As in (2.7) to (2.10), we expand the differential forms of each term. We first have:

$$*J = \frac{1}{3!} *J_{\sigma\mu\nu} dx^\sigma \wedge dx^\mu \wedge dx^\nu = *J_{\sigma\mu\nu} dx^\sigma dx^\mu dx^\nu, \quad (5.2)$$

$$\begin{aligned} d*dG &= \frac{1}{3!} (\partial_{;\sigma} * \partial_{;[\mu} G_{\nu]} + \partial_{;\mu} * \partial_{;[\nu} G_{\sigma]} + \partial_{;\nu} * \partial_{;[\sigma} G_{\mu]}) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= \frac{1}{2!} \partial_{;\sigma} * \partial_{;[\mu} G_{\nu]} dx^\sigma \wedge dx^\mu \wedge dx^\nu = \frac{1}{2!} \frac{1}{2!} \partial_{;\sigma} \left( (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial^{;\alpha} G^{\beta 1} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\ &= \frac{1}{2!} \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial_{;\sigma} \partial^{;\alpha} G^{\beta 1} dx^\sigma \wedge dx^\mu \wedge dx^\nu = \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial_{;\sigma} \partial^{;\alpha} G^{\beta} dx^\sigma \wedge dx^\mu \wedge dx^\nu \end{aligned} \quad (5.3)$$

Above, we have used the duality relationship  $*\partial_{;[\mu} G_{\nu]} = \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial^{;\alpha} G^{\beta 1}$ . We have also allowed for a curved spacetime by using the covariant derivatives, as well as the product rule

which simplifies to  $\partial_{;\sigma} \left( (-g)^5 \partial^{[\alpha} G^{\beta]} \right) = (-g)^5 \partial_{;\sigma} \partial^{[\alpha} G^{\beta]}$  because of the metricity  $g_{\mu\nu;\sigma} = 0$ . In flat spacetime,  $\partial_{;\sigma} \rightarrow \partial_{\sigma}$  and  $(-g)^5 = 1$ .

Next, in contrast to (2.8), using  $*[G_{\mu}, G_{\nu}] = \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} [G^{\alpha}, G^{\beta}]$  and of course  $g_{\mu\nu;\sigma} = 0$ , with the analogous sign reversal at the sixth line as in (2.8), we have:

$$\begin{aligned}
& -id * [G, G] = -\frac{1}{2!} i \partial_{;\sigma} * [G_{\mu}, G_{\nu}] dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -\frac{1}{2!} \frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial_{;\sigma} [G^{\alpha}, G^{\beta}] dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} = -\frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial_{;\sigma} (G^{\alpha} G^{\beta}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -\frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} (\partial_{;\sigma} G^{\alpha} G^{\beta}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} - \frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} (G^{\alpha} \partial_{;\sigma} G^{\beta}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -\frac{1}{2!} \frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} (\partial_{;\sigma} G^{[\alpha} G^{\beta]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} - \frac{1}{2!} \frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} (G^{[\alpha} \partial_{;\sigma} G^{\beta]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -\frac{1}{2!} i (\partial_{;\sigma} * G_{[\mu} G_{\nu]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} - \frac{1}{2!} i (*G_{[\mu} \partial_{;\sigma} G_{\nu]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -\frac{1}{2!} i (\partial_{;\sigma} * G_{[\mu} G_{\nu]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} + \frac{1}{2!} i (*G_{[\sigma} \partial_{;\mu} G_{\nu]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -i (\partial_{;\sigma} * G_{\mu} G_{\nu}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} + i (*G_{\sigma} \partial_{;\mu} G_{\nu}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -\frac{1}{3!} i (\partial_{;\sigma} * G_{[\mu} G_{\nu]} + \partial_{;\mu} * G_{[\nu} G_{\sigma]} + \partial_{;\nu} * G_{[\sigma} G_{\mu]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& \quad + \frac{1}{3!} i (*G_{[\sigma} \partial_{;\mu} G_{\nu]} + *G_{[\mu} \partial_{;\nu} G_{\sigma]} + *G_{[\nu} \partial_{;\sigma} G_{\mu]}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -\frac{1}{3!} i (*\partial_{;\sigma} G_{\mu} G_{\nu} + *\partial_{;\mu} G_{\nu} G_{\sigma} + *\partial_{;\nu} G_{\sigma} G_{\mu}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& \quad + \frac{1}{3!} i (G_{\sigma} * \partial_{;\mu} G_{\nu} + G_{\mu} * \partial_{;\nu} G_{\sigma} + G_{\nu} * \partial_{;\sigma} G_{\mu}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = (-i * \partial_{;\sigma} G_{\mu} G_{\nu} + i G_{\sigma} * \partial_{;\mu} G_{\nu}) dx^{\sigma} \wedge dx^{\mu} \wedge dx^{\nu} \\
& = -i * dGG + iG * dG \tag{5.4}
\end{aligned}$$

Note that within the differential forms, and given  $*F_{\mu\nu} = \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} F^{\alpha\beta}$  and  $g_{\mu\nu;\sigma} = 0$ , we are able to “transfer” the duality operation, i.e., that we are able to set  $\partial_{;\sigma} * G_{[\mu} G_{\nu]} \rightarrow * \partial_{;\sigma} G_{[\mu} G_{\nu]}$ , etc. and  $*G_{[\sigma} \partial_{;\mu} G_{\nu]} \rightarrow G_{\sigma} * \partial_{;\mu} G_{\nu}$ , etc. This reveals  $d * [G, G] = *dGG - G * dG$  as a duality product-rule identity, contrast  $d [G, G] = dGG - GdG$  from (2.8).

Similarly, in contrast to (2.9), using  $*\partial_{;\mu} G_{\nu} = \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial^{[\alpha} G^{\beta]}$ , with a sign reversal as previously in the sixth line, and transferring  $*G_{[\sigma} \partial_{;\mu} G_{\nu]} \rightarrow G_{\sigma} * \partial_{;\mu} G_{\nu}$  in the eighth line as was done in (5.4) above without repeating the expansion to third rank tensor form, we obtain:

$$\begin{aligned}
& -i[G, *dG] = -\frac{1}{3!}i \left( [G_\sigma, * \partial_{;\mu} G_\nu] + [G_\mu, * \partial_{;\nu} G_\sigma] + [G_\nu, * \partial_{;\sigma} G_\mu] \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -\frac{1}{2!}i [G_\sigma, * \partial_{;\mu} G_\nu] dx^\sigma \wedge dx^\mu \wedge dx^\nu = -\frac{1}{2!} \frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} [G_\sigma, \partial^{;\alpha} G^{\beta}] dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -\frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} [G_\sigma, \partial^{;\alpha} G^\beta] dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -\frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \left( G_\sigma \partial^{;\alpha} G^\beta - \partial^{;\alpha} (G^\beta G_\sigma) \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -\frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \left( G_\sigma \partial^{;\alpha} G^\beta - G^\beta \partial^{;\alpha} G_\sigma - \partial^{;\alpha} G^\beta G_\sigma \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \quad (5.5) \\
& = -\frac{1}{2!} \frac{1}{2!} i (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \left( G_\sigma \partial^{;\alpha} G^{\beta}] + G^{[\alpha} \partial^{;\beta]} G_\sigma - \partial^{;\alpha} G^{\beta]} G_\sigma \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -\frac{1}{2!} i \left( G_\sigma * \partial_{;\mu} G_\nu + * G_{[\sigma} \partial_{;\mu]} G_\nu - * \partial_{;\sigma} G_\mu G_\nu \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -\frac{1}{2!} i \left( 2G_\sigma * \partial_{;\mu} G_\nu - * \partial_{;\sigma} G_\mu G_\nu \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = \left( -2iG_\sigma * \partial_{;\mu} G_\nu + i * \partial_{;\sigma} G_\mu G_\nu \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -2iG * dG + i * dGG
\end{aligned}$$

Finally, in contrast to (2.10), using  $*[G_\mu, G_\nu] = \frac{1}{2!}(-g)^5 \varepsilon_{\alpha\beta\mu\nu} [G^\alpha, G^\beta]$ ,

$$\begin{aligned}
& -[G, *[G, G]] = -\frac{1}{3!} \left( [G_\sigma, *[G_\mu, G_\nu]] + [G_\mu, *[G_\nu, G_\sigma]] + [G_\nu, *[G_\sigma, G_\mu]] \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
& = -\frac{1}{2!} [G_\sigma, *[G_\mu, G_\nu]] dx^\sigma \wedge dx^\mu \wedge dx^\nu = -\frac{1}{2!} \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} [G_\sigma, [G^\alpha, G^\beta]] dx^\sigma \wedge dx^\mu \wedge dx^\nu \quad (5.6) \\
& = -\frac{1}{2!} \frac{1}{3!} (-g)^5 \left( \varepsilon_{\alpha\beta\mu\nu} [G_\sigma, [G^\alpha, G^\beta]] + \varepsilon_{\alpha\beta\nu\sigma} [G_\mu, [G^\alpha, G^\beta]] + \varepsilon_{\alpha\beta\sigma\mu} [G_\nu, [G^\alpha, G^\beta]] \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \neq 0
\end{aligned}$$

Unlike (2.10), this does not map into the Jacobian identity  $[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0$ , and so is not zero.

So now we use  $-id*[G, G] = -i*dGG + iG*dG$  and  $-i[G, *dG] = -2iG*dG + i*dGG$  found in (5.4) and (5.5), in (5.1). Analogously to (2.11) we obtain:

$$\begin{aligned}
& *J = d*dG - id*[G, G] - i[G, *dG] - [G, *[G, G]] \\
& = d*dG - i*dGG + iG*dG - 2iG*dG + i*dGG - [G, *[G, G]] \\
& = d*dG - iG*dG - [G, *[G, G]] \\
& = d*dG + id*[G, G] - i*dGG - [G, *[G, G]] \quad (5.7) \\
& = d*F - i[G, *dG] - [G, *[G, G]] \\
& = d*F - 2iG*dG + i*dGG - [G, *[G, G]]
\end{aligned}$$

This corresponds to (2.11), however, here: a)  $*J \neq 0$  in contrast to  $P = 0$ ; b)  $d*dG \neq 0$  in contrast to  $ddG = 0$ ; c)  $[G, *[G, G]] \neq 0$  in contrast to  $[G, [G, G]] = 0$ , and d) the terms  $id[G, G] \rightarrow id*[G, G]$  and  $-idGG \rightarrow -i*dGG$ . Starting from on the top line, we also employ  $*F = *(dG - i[G, G])$  which is the differential form for  $*F_{\mu\nu} = *(\partial_{[\mu}G_{\nu]} - i[G_{\mu}, G_{\nu}])$  in the final two lines.

Now we wish to apply Gauss' / Stokes' theorem to (5.7), as we earlier did to (2.11). Using the last two lines of (5.7) with the integrable term  $d*F$  separated on the left, we have:

$$\begin{aligned} \oint\!\!\!\oint *F &= \iiint d*F \\ &= \iiint (*J + i[G, *dG] + [G, *[G, G]]) \\ &= \iiint (*J - i*dGG + 2iG*dG + [G, *[G, G]]) \end{aligned} \quad (5.8)$$

The Abelian portion of this equation,  $\oint\!\!\!\oint *F = \iiint *J$  which we used for pedagogic simplicity in the analysis following (3.3), is clearly included when the gauge fields are set to zero. Putting the Yang-Mills electric charge equation (5.8) together with the magnetic charge equation (3.3), we find that Maxwell's Yang-Mills equations in integral form are:

$$\begin{aligned} \oint\!\!\!\oint *F &= \iiint *J - i\iiint *dGG + \iiint (2iG*dG + [G, *[G, G]]) \\ \oint\!\!\!\oint F &= \iiint P' = -i\iiint dGG = -i\oint\!\!\!\oint [G, G] \end{aligned} \quad (5.9)$$

In this form, the parallels and differences are manifestly clear.  $\oint\!\!\!\oint *F$  is the net electric field flux and  $\oint\!\!\!\oint F$  the net magnetic field flux over a closed surface.  $*J$  is the electric source charge density and it is non-vanishing, while the magnetic source density  $P = 0$  vanishes by the Jacobian (2.4). Similarly, while  $G*dG \neq 0$  and  $[G, *[G, G]] \neq 0$  in the electric field equation, their duality counterparts  $GdG = 0$  and  $[G, [G, G]] = 0$  are also part of the magnetic charge equation, but vanish by the respective identities found in (2.11) and (2.10). We see how the only true, elementary source is  $*J$  and that there are then a number of faux sources which include  $P' = -idGG = -id[G, G]$  for the net magnetic field flux  $\oint\!\!\!\oint F$ , and  $*J' \equiv -i*dGG + 2iG*dG + [G, *[G, G]]$  which is a faux electric source which contributes to the net electric field flux beyond that contributed by "true" electric source  $J$  in the abelian portion  $\oint\!\!\!\oint *F = \iiint *J$  of (5.9).

Because the only elementary, real, not-faux source in the Yang-Mills equations (5.9) is the electric source  $*J$ , it will be desirable to solve the electric charge density equation (5.7) for the gauge field  $G$  in terms of  $*J$ . Particularly, as laid out at the end of section 3, our eventual

goal is to find  $\oint F(G(J(\psi)))$ . So a key step along the way is to obtain the gauge fields  $G(J)$  in terms of sources. Equation (5.7) has a number of alternative ways to express  $*J(G)$ , but the most compact way is on the third line. So we expand those differential forms to obtain:

$$\begin{aligned}
 *J &= \frac{1}{3!} *J_{\sigma\mu\nu} dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= d *dG - iG *dG - [G, *[G, G]] \\
 &= \frac{1}{3!} (\partial_\sigma * \partial_{[\mu} G_{\nu]} + \partial_\mu * \partial_{[\nu} G_{\sigma]} + \partial_\nu * \partial_{[\sigma} G_{\mu]}) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &\quad - \frac{1}{3!} i (G_\sigma * \partial_{[\mu} G_{\nu]} + G_\mu * \partial_{[\nu} G_{\sigma]} + G_\nu * \partial_{[\sigma} G_{\mu]}) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &\quad - \frac{1}{3!} ([G_\sigma, *[G_\mu, G_\nu]] + [G_\mu, *[G_\nu, G_\sigma]] + [G_\nu, *[G_\sigma, G_\mu]]) dx^\sigma \wedge dx^\mu \wedge dx^\nu
 \end{aligned} \tag{5.10}$$

Stripping off the forms, we obtain the tensor equation:

$$\begin{aligned}
 *J_{\sigma\mu\nu} &= (\partial_\sigma * \partial_{[\mu} G_{\nu]} + \partial_\mu * \partial_{[\nu} G_{\sigma]} + \partial_\nu * \partial_{[\sigma} G_{\mu]}) \\
 &\quad - i (G_\sigma * \partial_{[\mu} G_{\nu]} + G_\mu * \partial_{[\nu} G_{\sigma]} + G_\nu * \partial_{[\sigma} G_{\mu]}) \\
 &\quad - ([G_\sigma, *[G_\mu, G_\nu]] + [G_\mu, *[G_\nu, G_\sigma]] + [G_\nu, *[G_\sigma, G_\mu]])
 \end{aligned} \tag{5.11}$$

Then, we apply the duality operations  $*J_{\sigma\mu\nu} = (-g)^5 \varepsilon_{\alpha\sigma\mu\nu} J^\alpha$ ,  $*\partial_{[\mu} G_{\nu]} = \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} \partial^{[\alpha} G^{\beta]}$  and  $*[G_\mu, G_\nu] = \frac{1}{2!} (-g)^5 \varepsilon_{\alpha\beta\mu\nu} [G^\alpha, G^\beta]$ , and the metricity  $g_{\mu\nu;\sigma} = 0$  as discussed after (5.3), to obtain (a good summary of the use of duality is contained in [9], pages 87-89):

$$\begin{aligned}
 &(-g)^5 \varepsilon_{\alpha\sigma\mu\nu} J^\alpha \\
 &= \frac{1}{2!} (-g)^5 (\varepsilon_{\alpha\beta\mu\nu} \partial_\sigma \partial^{[\alpha} G^{\beta]} + \varepsilon_{\alpha\beta\nu\sigma} \partial_\mu \partial^{[\alpha} G^{\beta]} + \varepsilon_{\alpha\beta\sigma\mu} \partial_\nu \partial^{[\alpha} G^{\beta]}) \\
 &\quad - \frac{1}{2!} i (-g)^5 (\varepsilon_{\alpha\beta\mu\nu} G_\sigma \partial^{[\alpha} G^{\beta]} + \varepsilon_{\alpha\beta\nu\sigma} G_\mu \partial^{[\alpha} G^{\beta]} + \varepsilon_{\alpha\beta\sigma\mu} G_\nu \partial^{[\alpha} G^{\beta]}) \\
 &\quad - \frac{1}{2!} (-g)^5 (\varepsilon_{\alpha\beta\mu\nu} [G_\sigma, [G^\alpha, G^\beta]] + \varepsilon_{\alpha\beta\nu\sigma} [G_\mu, [G^\alpha, G^\beta]] + \varepsilon_{\alpha\beta\sigma\mu} [G_\nu, [G^\alpha, G^\beta]])
 \end{aligned} \tag{5.12}$$

Factoring out  $(-g)^5$  and multiplying through by  $\varepsilon^{\kappa\sigma\mu\nu}$  next yields:

$$\begin{aligned}
 \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\sigma\mu\nu} J^\alpha &= -3! \delta^\kappa_\alpha J^\alpha = -6J^\kappa \\
 &= \frac{1}{2!} \left( \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\mu\nu} \partial_\sigma \partial^{[\alpha} G^{\beta]} + \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\nu\sigma} \partial_\mu \partial^{[\alpha} G^{\beta]} + \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\sigma\mu} \partial_\nu \partial^{[\alpha} G^{\beta]} \right) \\
 &\quad - \frac{1}{2!} i \left( \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\mu\nu} G_\sigma \partial^{[\alpha} G^{\beta]} + \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\nu\sigma} G_\mu \partial^{[\alpha} G^{\beta]} + \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\sigma\mu} G_\nu \partial^{[\alpha} G^{\beta]} \right) \\
 &\quad - \frac{1}{2!} \left( \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\mu\nu} \left[ G_\sigma, \left[ G^\alpha, G^\beta \right] \right] + \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\nu\sigma} \left[ G_\mu, \left[ G^\alpha, G^\beta \right] \right] + \varepsilon^{\kappa\sigma\mu\nu} \varepsilon_{\alpha\beta\sigma\mu} \left[ G_\nu, \left[ G^\alpha, G^\beta \right] \right] \right). \quad (5.13) \\
 &= - \left( \delta^{\kappa\sigma}_{\alpha\beta} \partial_\sigma \partial^{[\alpha} G^{\beta]} + \delta^{\kappa\mu}_{\alpha\beta} \partial_\mu \partial^{[\alpha} G^{\beta]} + \delta^{\kappa\nu}_{\alpha\beta} \partial_\nu \partial^{[\alpha} G^{\beta]} \right) \\
 &\quad + i \left( \delta^{\kappa\sigma}_{\alpha\beta} G_\sigma \partial^{[\alpha} G^{\beta]} + \delta^{\kappa\mu}_{\alpha\beta} G_\mu \partial^{[\alpha} G^{\beta]} + \delta^{\kappa\nu}_{\alpha\beta} G_\nu \partial^{[\alpha} G^{\beta]} \right) \\
 &\quad + \left( \delta^{\kappa\sigma}_{\alpha\beta} \left[ G_\sigma, \left[ G^\alpha, G^\beta \right] \right] + \delta^{\kappa\mu}_{\alpha\beta} \left[ G_\mu, \left[ G^\alpha, G^\beta \right] \right] + \delta^{\kappa\nu}_{\alpha\beta} \left[ G_\nu, \left[ G^\alpha, G^\beta \right] \right] \right)
 \end{aligned}$$

Using  $\delta^{\kappa\sigma}_{\alpha\beta} \equiv \delta^\kappa_\alpha \delta^\sigma_\beta - \delta^\kappa_\beta \delta^\sigma_\alpha$  and the like, with  $\kappa \rightarrow \nu$  index renaming, this reduces to:

$$-J^\nu = \partial_\sigma \partial^{[\sigma} G^{\nu]} - i G_\sigma \partial^{[\sigma} G^{\nu]} - \left[ G_\sigma, \left[ G^\sigma, G^\nu \right] \right]. \quad (5.14)$$

Contrasting to the original  $*J = d* dG - iG* dG - \left[ G, *[G, G] \right]$ , we see that aside from the sign reversal, the  $*$  between two objects essentially results in an index contraction between those two objects when they are written as tensors. If we then expand all the commutators and reorganize terms in a familiar way, we obtain:

$$\begin{aligned}
 -J^\nu &= \partial_\sigma \partial^{[\sigma} G^{\nu]} - i G_\sigma \partial^{[\sigma} G^{\nu]} - \left[ G_\sigma, \left[ G^\sigma, G^\nu \right] \right] \\
 &= \left( \partial_\sigma \partial^\sigma - i G_\sigma \partial^\sigma - G_\sigma G^\sigma \right) G^\nu - \left( \partial^\sigma \partial^\nu - i G^\sigma \partial^\nu - 2G^\sigma G^\nu + G^\nu G^\sigma \right) G_\sigma \\
 &= g^{\nu\sigma} \left( \partial_\tau \partial^\tau - i G_\tau \partial^\tau - G_\tau G^\tau \right) G_\sigma - \left( \partial^\sigma \partial^\nu - i G^\sigma \partial^\nu - 2G^\sigma G^\nu + G^\nu G^\sigma \right) G_\sigma \\
 &\equiv \left( g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu \right) G_\sigma, \quad (5.15)
 \end{aligned}$$

with a configuration space operator  $g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu$  where in the final line we have defined the second rank tensor operator:

$$D^\sigma D^\nu \equiv \partial^\sigma \partial^\nu - i G^\sigma \partial^\nu - 2G^\sigma G^\nu + G^\nu G^\sigma \quad (5.16)$$

which, upon contraction, does yield the scalar also appearing in(5.15), namely:

$$D_\tau D^\tau = \partial_\tau \partial^\tau - i G_\tau \partial^\tau - G_\tau G^\tau. \quad (5.17)$$

By way of contrast, in Abelian gauge theory  $-J^\nu = \left( g^{\nu\sigma} \partial_\tau \partial^\tau - \partial^\sigma \partial^\nu \right) G_\sigma$ . So (5.15) for  $J^\nu(G_\sigma)$ , is now in a familiar form which we can use to approach taking the inverse  $G_\sigma(J^\nu)$ . This is the first step toward being able to obtain  $\oint F(G(J(\psi)))$ .

Finally, let us find the continuity equation for conservation of the electric source density and current, based on (5.15). Equation (5.15) will clearly be recognized as another way to express  $-J^\nu = D_\sigma F^{\sigma\nu}$  which may be similarly derived from  $*J = D*F$  in (1.12). Particularly, we wish to show that  $-D_\nu J^\nu = D_\nu D_\sigma F^{\sigma\nu} = 0$ , by identity. Similarly to (2.1), we may take the gauge-covariant derivative of  $J^\nu$  via the commutation:

$$\begin{aligned} [D_\nu, J^\nu] \varphi &= D_\nu (J^\nu \varphi) - J^\nu D_\nu \varphi = (\partial_\nu - iG_\nu)(J^\nu \varphi) - J^\nu (\partial_\nu - iG_\nu) \varphi \\ &= \partial_\nu J^\nu \varphi + J^\nu \partial_\nu \varphi - iG_\nu J^\nu \varphi - J^\nu \partial_\nu \varphi + iJ^\nu G_\nu \varphi = \partial_\nu J^\nu \varphi - i[G_\nu, J^\nu] \varphi = D_\nu J^\nu \varphi \end{aligned} \quad (5.18)$$

Stripping off the  $\varphi$ , we see the correct derivative:

$$[D_\nu, J^\nu] = \partial_\nu J^\nu - i[G_\nu, J^\nu] = D_\nu J^\nu \quad (5.19)$$

which includes the commutator  $[G_\nu, J^\nu]$ . So, we start with  $-D_\nu J^\nu = D_\nu D_\sigma F^{\sigma\nu}$  and apply  $[D_\nu, J^\nu] = D_\nu J^\nu$  from (5.19),  $-J^\nu = D_\sigma F^{\sigma\nu}$ ,  $D_\sigma F_{\mu\nu} = [D_\sigma, F_{\mu\nu}]$  from (2.2), and  $iF_{\sigma\nu} = [D_\nu, D_\sigma]$  from (1.1) to show via simple index commutativity that the continuity equation, due finally to the scalar contraction  $F_{\sigma\nu} F^{\sigma\nu}$  of like-objects in a commutator  $[F_{\sigma\nu}, F^{\sigma\nu}] = 0$ , is:

$$\begin{aligned} -D_\nu J^\nu &= -[D_\nu, J^\nu] = [D_\nu, D_\sigma F^{\sigma\nu}] = [D_\nu, [D_\sigma, F^{\sigma\nu}]] \\ &= D_\nu D_\sigma F^{\sigma\nu} - D_\nu F^{\sigma\nu} D_\sigma - D_\sigma F^{\sigma\nu} D_\nu + F^{\sigma\nu} D_\sigma D_\nu \\ &= D_\nu D_\sigma F^{\sigma\nu} + F^{\sigma\nu} D_\sigma D_\nu = [D_\nu D_\sigma, F^{\sigma\nu}] = \frac{1}{2} [[D_\nu, D_\sigma], F^{\sigma\nu}] \\ &= \frac{1}{2} i [F_{\sigma\nu}, F^{\sigma\nu}] = 0 \end{aligned} \quad (5.20)$$

The continuity equation in differential forms, therefore, is  $D*J = DD*F = 0$ . This equation for the conservation of the non-abelian charge density will play a very central role the development to follow.

## 6. Abelian and non-Abelian Massive Gauge Boson Inverses for the Electric Charge Density, Using the ‘‘Coleman-Zee’’ Method

The next stage in our development to demonstrate that  $\oint\!\!\!\oint F = -i \int\!\!\!\int dGG$  in (5.9) is the integral-form classical equation for a baryon, is to invert the configuration space operator  $g^{\nu\sigma} D_\nu D^\sigma - D^\sigma D^\nu$  of (5.15) to obtain  $G_\sigma(J^\nu)$ , so we can obtain  $\oint\!\!\!\oint F(G(J))$ . This inverse, which we denote by  $I_{\mu\nu}$ , may be defined by  $G_\mu \equiv I_{\mu\nu} J^\nu$ . In general,  $I_{\mu\nu} \neq I_{\nu\mu}$  is not necessarily symmetric, so  $G_\mu \equiv I_{\mu\nu} J^\nu$  is an inner product definition not necessarily the same as an outer

product definition  $G_\mu \equiv I'_{\nu\mu} J^\nu$ . Making use of  $G_\mu \equiv I_{\mu\nu} J^\nu$  to left-multiply (5.15) by  $-I_{\mu\nu}$  allows us to write:

$$I_{\mu\nu} J^\nu = -I_{\mu\nu} (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) G_\sigma = G_\mu = \delta^\sigma_\mu G_\sigma, \quad (6.1)$$

from which we may extract a more-directly defined inverse:

$$-I_{\mu\nu} (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) = \delta^\sigma_\mu. \quad (6.2)$$

Now the task is to show that this inverse exists, to understand the degree to which any particular inverse which does exist is non-unique, to review the options for fixing the gauge of these inverses, and to select the inverse or inverses with suitable gauge choices or better yet, *unique gauge requirements* which best illustrate why  $\oint\!\!\!\oint F = -i\iiint dGG$  based on a faux magnetic charge  $P' = -idGG$  of (3.4) has all of the key symmetries of a baryon.

Taking inverses in gauge theory is a tricky business, because one is often free to choose the gauge resulting in non-unique inverses, and because particularly for massless gauge bosons – which include the gluons of QCD – the inverse *may not even exist* without a careful selection and fixing of the gauge, see, e.g., [11] chapter III.4. Additionally, because the gauge field is the field of integration used to turn a classical action  $S$  into a quantum field amplitude  $W$ , a symmetry that exists classically may not be a symmetry of the related quantum field theory, see, e.g., [11] chapter IV.7 (Chiral Anomaly). Specifically, a classical symmetry exists if some transformation leaves the action  $S(\varphi)$  invariant. A quantum symmetry exists (and inherits the classical symmetry) if the same transformation leaves the path integral  $Z = \int D\varphi \exp iS(\varphi)$  invariant. But this may not always be the case. Therefore, let us start by carefully parsing out the various issues that come into play when taking inverses of the form (6.2).

First, as to classical versus quantum fields, we consider the local non-abelian gauge transformation which is  $G_\mu \rightarrow G'_\mu = G_\mu + \partial_\mu \theta - i[G_\mu, \theta]$  in tensors,  $G \rightarrow G' = G + d\theta - i[G, \theta]$  in differential commutator forms, and  $G \rightarrow G' = G + d\theta + G \wedge \theta = G + (d + G \wedge) \theta$  in differential wedge forms. These are all alternative but equivalent ways of saying the same thing. All of the classical field equations developed thus far including (1.12), (2.11), (3.3), (5.1), (5.7) and (5.9) are symmetric under such a gauge transformation. So too, the electric charge field equation (5.15) with the specific  $D^\sigma D^\nu$  and  $D_\tau D^\tau$  identified in (5.16) and (5.17) is symmetric under this non-abelian gauge transformation. This should be no surprise: all of these equations were developed with the express purpose of preserving this gauge symmetry. This means that the action  $S(G) = \int \mathcal{L}(G) d^4x$  is similarly invariant. But when we take a path integral  $Z = \int DG \exp iS(G) \equiv \mathcal{C} \exp iW(J)$  to obtain the associated quantum field theory for the amplitude  $W(J)$ , we are not necessarily assured that the measure  $DG$  will inherit this same symmetry. And this in turn means that the quantum field theory may not share all of the

symmetries of the classical field theory. Typically, ensuring that the path integral also carries forward the gauge symmetry under  $DG \rightarrow D(G + d\theta - i[G, \theta])$  is what gives rise to gauge-fixing measures such as Faddeev-Popov [13] including anticommuting scalar “ghost” fields, see some concise development of this in [11], chapters III.4, and VII.1. However, so long as we restrict ourselves to classical field theory, which we are doing at the moment, we can develop inverses without this particular worry. We just need to be prepared to address this issue once we are ready to calculate the path integral, which is to be done only after the classical theory has been fully elaborated. Again, as to why there is both validity and benefit to doing taking this approach of fully elaborating the classical theory in advance of the quantum theory, see the discussion of section 4.

Second, as to why we need to take inverses when going from classical to quantum field theory, this is because the mathematical exercise of calculating a path integral revolves around clever extrapolations of the Gaussian integral  $\int dx \exp(-\frac{1}{2}Ax^2 - Jx) = (-2\pi/A)^{-5} \exp(J^2/2A)$  into  $Z = \int DG \exp(iS(G)) \equiv C \exp(iW(J))$ , with the correspondence  $W(J) \sim J^2/2A$ . Because the abstracted coefficient  $A$  of  $Ax^2$  gets inverted in  $J^2/2A$ , and because  $A$  ends up corresponding with the configuration space operator  $g^{\nu\sigma}D_\tau D^\tau - D^\sigma D^\nu$  in (6.2) which then gets inverted via  $J^2/2A$  into  $I_{\nu\mu}$  which then becomes proportionately related to the quantum propagator assuming we can find a way as we will in sections 8 and 11 to deal with  $g^{\nu\sigma}D_\tau D^\tau - D^\sigma D^\nu$  *not* being quadratic in  $G_\mu$ , one must expect to have to obtain  $(g^{\nu\sigma}D_\tau D^\tau - D^\sigma D^\nu)^{-1}$  to arrive at quantum field theory, in addition to having to deal with the invariance of the measure under  $DG \rightarrow D(G + d\theta - i[G, \theta])$ . Thus, it is desirable to have a number of inverses already developed “on the shelf” when it comes time to use them to calculate a path integral. But, as we see in (6.2), even before we start approaching path integration, we still need this inverse *even to develop the classical theory*, and specifically, in order to obtain  $\oint\!\!\!\oint F(G(J))$ .

Third, even in classical theory, as already mentioned, configuration space operators of the form  $g^{\nu\sigma}\partial_\tau\partial^\tau - \partial^\sigma\partial^\nu$  simply have no inverse! Although often couched in mystery, this problem arises from the simple fact that for a massless gauge boson, a Lorentz vector  $G_\mu$  with *four* spacetime components is used to describe physical fields – for example the photon in electrodynamics and the gluons in chromodynamics – which only have *two* physical degrees of freedom. That is, a *mathematical object*  $G_\mu$  with four degrees of freedom is used to represent a *physical object* which only has half as many degrees of freedom. This is an inherent redundancy in how we describe gauge fields that causes inverses to be non-unique and brings about the need for gauge fixing. Gauge fixing and related methods are then used to create a menu of gauge-fixed solutions out of the non-uniqueness stemming from this redundancy. *This gauge non-uniqueness is a separate and distinct issue from gauge symmetry*. For example, the field equation  $-J^\nu = (g^{\nu\sigma}\partial_\tau\partial^\tau - \partial^\sigma\partial^\nu)A_\sigma$  for a photon field  $A_\sigma$  sourced by a current density  $J^\nu$  is

fully symmetric under an abelian gauge transformation  $A_\sigma \rightarrow A'_\sigma = A_\sigma + \partial_\sigma \theta$ . But  $A_\sigma$  is still redundant insofar as it has four spacetime degrees of freedom while a photon only has two transverse degrees of freedom. Additionally, as mentioned, the operator  $g^{\nu\sigma} \partial_\tau \partial^\tau - \partial^\sigma \partial^\nu$  has no inverse, or, to be more precise, has an inverse which is of infinite magnitude and so is completely indeterminate.

Now, following Zee on page 30 of [11]:

“In order to avoid complications at this stage associated with gauge invariance [we] will consider instead the field theory of a massive spin 1 meson, or vector meson. . . . We can adopt a pragmatic attitude: Calculate a photon mass  $m$  and set  $m = 0$  at the end, and if the result does not blow up in our faces, we will presume that it is OK.”

Zee states in a footnote to this passage that when he “took a field theory course as a student with Sidney Coleman this was how he treated QED to avoid discussing gauge invariance.” So to simplify the development here, we shall take this same pragmatic approach as Coleman and Zee: We shall introduce a non-zero “Proca mass” for the gauge fields  $G$ , develop the classical monopole  $\oint\!\!\!\oint F = -i \iiint dGG$  of (5.9) to show how it has all of the classical symmetries that one would expect of a baryon, and then set  $m = 0$  at the appropriate point in the development (which will come at (9.15) infra) and explore the massive / massless correspondences.

In this section, we shall develop the inverse of the *massive* boson configuration space operators  $g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu$  for non-abelian gauge theory and  $g^{\nu\sigma} (\partial_\tau \partial^\tau + m^2) - \partial^\sigma \partial^\nu$  for abelian gauge theory, and then follow Coleman and Zee by setting the mass to zero to see what results. In the next section we will take the more formal approach of developing the inverses  $g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu$  and  $g^{\nu\sigma} \partial_\tau \partial^\tau - \partial^\sigma \partial^\nu$  for a *massless* particle directly, using the Faddeev-Popov method. We will then contrast the both approaches and see where they meet, to give us some guidance about how to then use these inverses in the non-abelian magnetic monopole field equation  $\oint\!\!\!\oint F = -i \iiint dGG$ .

So, following the Coleman-Zee approach, let us add a Proca mass  $m$  to (5.15), thus:

$$-J^\nu = \left( g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu \right) G_\sigma. \quad (6.3)$$

Let us then consider (6.3) in flat spacetime where gradient operators  $[\partial_\mu, \partial_\nu] = 0$  commute. Let us also momentarily revert  $D \rightarrow \partial$  to ordinary derivatives to make a pedagogical point, and so write (6.3) as its abelian subset  $-J^\nu = \left( g^{\nu\sigma} (\partial_\tau \partial^\tau + m^2) - \partial^\sigma \partial^\nu \right) G_\sigma$ . The current density is conserved by the continuity equation  $\partial_\nu J^\nu = 0$ , so if we take the gradient of each side and reduce, we find that  $m^2 \partial_\nu G^\nu = 0$ . Because we take the mass to be non-zero, this means that

$\partial_\nu G^\nu = 0$ , which is a fully-covariant equation known as the Lorenz gauge. Here,  $\partial_\nu G^\nu = 0$  is not a gauge condition at all; it is a *requirement* needed to ensure continuity for a massive vector boson. The number of degrees of freedom in the *mathematical* object  $G^\nu$  is covariantly reduced from four to three by  $\partial_\nu G^\nu = 0$ , and this matches precisely with the three polarization degrees of freedom – one longitudinal and two transverse – possessed by the *physical* gauge boson. So now, most of the gauge redundancy is squeezed out from  $G^\nu$ . Even here, however, there is still a residual redundancy that requires gauge fixing. For, if we transform  $G^\nu \rightarrow G^\nu + \partial^\nu \theta$ , then the Lorenz condition becomes  $\partial_\nu (G^\nu + \partial^\nu \theta) = 0$ , or  $\partial_\nu G^\nu = -\partial_\nu \partial^\nu \theta$ . So to maintain  $\partial_\nu G^\nu = 0$  under any such gauge transformation, we may this fix the gauge completely by the gauge condition  $\partial_\nu \partial^\nu \theta = 0$ . Thus, with everything taken together,  $-J^\nu = (g^{\nu\sigma} (\partial_\tau \partial^\tau + m^2) - \partial^\sigma \partial^\nu) G_\sigma$  is invariant under a gauge transformation  $G^\nu \rightarrow G^\nu + \partial^\nu \theta$ , the four degrees of freedom in  $G^\nu$  are covariantly-reduced down to three degrees of freedom by  $\partial_\nu G^\nu = 0$  which is required to match the three polarization degrees of freedom of the physical field, and the residual gauge freedom is fixed and thereby removed by  $\partial_\nu \partial^\nu \theta = 0$ . The field equation  $-J^\nu = (g^{\nu\sigma} (\partial_\tau \partial^\tau + m^2) - \partial^\sigma \partial^\nu) G_\sigma$  remains invariant under the gauge transformation  $G^\nu \rightarrow G^\nu + \partial^\nu \theta$  and this invariance does not depend in any way on  $\partial_\nu \partial^\nu \theta = 0$  because nowhere does the non-observable gauge (really, phase) angle  $\theta$  appear in the field equation.

In the non-abelian (6.3) it is a bit more complicated, because we have  $D$  from (5.15) to (5.17), not  $\partial$ , and because the proper way to take the gauge-derivative of the current density is by  $[D_\nu, J^\nu] = \partial_\nu J^\nu - i[G_\nu, J^\nu] = D_\nu J^\nu$  derived in (5.19). But we already saw that the continuity equation  $D_\nu J^\nu = 0$  of (5.20) which we now combine with (5.15), by identity, is:

$$-D_\nu J^\nu = D_\nu (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) G_\sigma = \mathbf{0}. \quad (6.4)$$

So if we simply add a Proca mass to (6.4) and maintain continuity, we must have:

$$\begin{aligned} -D_\nu J^\nu &= D_\nu (g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu) G_\sigma = D_\nu (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) G_\sigma + m^2 D_\nu g^{\nu\sigma} G_\sigma = 0 \\ &= \mathbf{0} + m^2 D_\nu G^\nu = 0 \end{aligned} \quad (6.5)$$

This includes  $D_\nu (g^{\nu\sigma} m^2 G_\sigma) = D_\nu (m^2 G^\nu) = m^2 D_\nu G^\nu = 0$ , where the highlighted zero in (6.4) and (6.5) is the zero-by-identity of the continuity equation (5.20). But the symmetries of the term  $D_\nu G^\nu$  in the above are driven by those of (5.19) which is  $D_\nu J^\nu = \partial_\nu J^\nu - i[G_\nu, J^\nu]$ . Consequently,  $D_\nu G^\nu = \partial_\nu G^\nu - i[G_\nu, G^\nu]$  because of (5.19). Additionally, because of (6.5) and the assumed non-zero mass,  $D_\nu G^\nu = \partial_\nu G^\nu - i[G_\nu, G^\nu] = 0$ . As in the abelian case just discussed, for a massive gauge boson, and this is not a mere gauge condition. It is *required* to ensure continuity. As in abelian theory this reduces the gauge freedom of a four-component spacetime

object  $G_\nu$  down to three to match the three massive boson polarizations. Additionally, here the commutator  $[G_\nu, G^\nu] = 0$  because of the scalar contraction  $G_\nu G^\nu$  of like objects. This means in turn that  $D_\nu G^\nu = \partial_\nu G^\nu = 0$ . And this means that  $\partial_\nu G^\nu = 0$  *still applies even to the non-abelian theory* and is not a gauge condition but is a requirement for a massive gauge boson.

As to the residual gauge freedom, because  $G^\nu \rightarrow G'^\nu = G^\nu + \partial^\nu \theta - i[G^\nu, \theta] = G^\nu + D^\nu \theta$  is the non-abelian gauge transformation,  $D_\nu G'^\nu = D_\nu G^\nu + D_\nu D^\nu \theta = \partial_\nu G^\nu + \partial_\nu D^\nu \theta = 0$  is the required covariant gauge condition for  $G'^\nu$ . Taken with  $D_\nu G^\nu = 0$  this means that for a non-abelian theory,  $D_\nu D^\nu \theta = 0$  replaces  $\partial_\nu \partial^\nu \theta = 0$  as the residual gauge condition. Taken with  $\partial_\nu G^\nu = 0$ , this means that  $\partial_\nu D^\nu \theta = \partial_\nu \partial^\nu \theta - i\partial_\nu [G^\nu, \theta] = 0$ , which means that  $D_\nu D^\nu \theta = 0$  may be written out with ordinary derivatives as  $\partial_\nu \partial^\nu \theta - i\partial_\nu [G^\nu, \theta] = 0$ . So while (6.3) is invariant under a non-abelian gauge transformation  $G^\nu \rightarrow G'^\nu = G^\nu + \partial^\nu \theta - i[G^\nu, \theta]$ , we are *required* to have  $D_\nu G^\nu = \partial_\nu G^\nu = 0$  because the boson in (6.3) is presumed to be massive and subject to continuity, and the remaining gauge freedom is fixed by imposing  $D_\nu D^\nu \theta = 0$  which as just seen is equivalent to the expression  $\partial_\nu \partial^\nu \theta - i\partial_\nu [G^\nu, \theta] = 0$ . Nonetheless, as in the abelian theory, this invariance does not depend in any way on  $D_\nu D^\nu \theta = 0$  a.k.a.  $\partial_\nu \partial^\nu \theta - i\partial_\nu [G^\nu, \theta] = 0$  because nowhere does the non-observable gauge / phase angle  $\theta$  appear in the field equation (6.3).

Now, let us stop for a moment to take a close look at the gauge-covariant, second-rank, second-derivative operator  $D^\sigma D^\nu$  in (5.16) and its gauge-covariant d'Alembertian  $\square = D_\tau D^\tau$  of (5.17). Close study of  $D^\sigma D^\nu$  will reveal that there is no apparent way to separate each of  $D^\sigma$  and  $D^\nu$  to make  $D^\sigma D^\nu$  a product of two separate expressions for  $D^\sigma$ ,  $D^\nu$ . Even the commutator of (5.16), which we can calculate to be  $i[D^\sigma, D^\nu] = G^{[\sigma} \partial^{\nu]} - 3i[G^\sigma, G^\nu]$  in flat spacetime, is different from  $F_{\mu\nu} \phi = i[D_\mu, D_\nu] \phi = (\partial_{[\mu} G_{\nu]} - i[G_\mu, G_\nu]) \phi$  which is the field strength defined in (1.1), (1.5). This is because in (5.15)  $D^\sigma D^\nu$  is operating on  $G_\sigma$  not  $\phi$  and because, as noted at the outset following (1.1), gauge-covariant derivatives, like covariant derivatives in Riemannian geometry, take a form that depends on the representation of the object they act upon.

However, for  $\square = D_\tau D^\tau$  we may make use of the very recent finding after (6.5) that  $\partial_\nu G^\nu = 0$  for a massive gauge boson *even in non-abelian gauge theory*, and specifically, may add this “zero” to (5.17) and thus write:

$$\begin{aligned} D_\tau D^\tau &= \partial_\tau \partial^\tau - i\partial_\tau G^\tau - iG_\tau \partial^\tau - G_\tau G^\tau = \partial_\tau (\partial^\tau - iG^\tau) - iG_\tau (\partial^\tau - iG^\tau) = (\partial_\tau - iG_\tau) (\partial^\tau - iG^\tau), \\ &= \partial_\tau \partial^\tau + V \end{aligned} \quad (6.6)$$

where in the final line we have defined the gauge field *perturbation* (see, e.g., [14] eq. [4.4]):

$$V \equiv -i\left(\partial_\tau G^\tau + G_\tau \partial^\tau\right) - G_\tau G^\tau \xrightarrow{i\partial \rightarrow k} -k_\tau G^\tau - G_\tau k^\tau - G_\tau G^\tau. \quad (6.7)$$

This use of  $\partial_\nu G^\nu = 0$  *does* allow a clean separation  $D_\tau D^\tau = (\partial_\tau - iG_\tau)(\partial^\tau - iG^\tau)$ , and it enables us to explicitly introduce and identify gauge field perturbations. This will be very useful throughout the subsequent development. And again, because we are considering a massive gauge boson,  $\partial_\nu G^\nu = 0$  is not just an optional gauge condition; it is required for continuity. At the end of (6.7) we convert into momentum space by the useful substitution  $i\partial \rightarrow k$ .

With these preliminaries behind us, it is time to calculate the inverse of (5.15) for a massive gauge boson. We start with the inverse  $I_{\mu\nu}$  of (6.2), for which we follow Coleman and Zee and add the Proca mass as follows:

$$I_{\mu\nu} \left( g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu \right) = -\delta^\sigma_\mu. \quad (6.8)$$

It is well-known how to calculate inverses of the form (6.8), but we do need to be cognizant of two important points because the  $D$  are not the same as ordinary  $\partial$  especially in flat spacetime. First, while  $[\partial^\sigma, \partial^\nu] = 0$  in flat spacetime, we cannot treat  $D^\sigma D^\nu$  as commuting here, that is,  $[D^\sigma, D^\nu] \neq 0$ . In fact, as noted prior to (6.6),  $i[D^\sigma, D^\nu] = G^{[\sigma} \partial^{\nu]}$   $- 3i[G^\sigma, G^\nu] \neq 0$  when the operand is  $G_\sigma$ . So we need to be very careful throughout to maintain strict commutation ordering. Second, we cannot just put expressions involving  $D^\sigma D^\nu$  or  $D_\tau D^\tau$  into a denominator. Rather, we have to treat carefully, as inverses and not mere denominators, inverse expressions which contain  $D^\sigma D^\nu$  as well as the gauge-covariant d'Alembertian  $\square = D_\tau D^\tau$ .

With that in mind, let us calculate  $I_{\mu\nu}$ . First, we specify  $I_{\mu\nu}$  using the general form:

$$I_{\mu\nu} \equiv Ag_{\mu\nu} + BD_\mu D_\nu, \quad (6.9)$$

with  $A$  and  $B$  unknown and to-be-deduced. Given that  $I_{\mu\nu} \neq I_{\nu\mu}$  (to see this, simply note that  $D_\mu D_\nu \neq D_\nu D_\mu$ ), the above definition together with  $G_\mu \equiv I_{\mu\nu} J^\nu$  leads to  $G_\mu \equiv (Ag_{\mu\nu} + BD_\mu D_\nu) J^\nu = Ag_{\mu\nu} J^\nu + BD_\mu D_\nu J^\nu = AJ_\mu$  once the continuity relation  $D_\nu J^\nu = 0$  of (5.20) is applied. So the inner-product definition  $G_\mu \equiv I_{\mu\nu} J^\nu$  combined with the inverse definition (6.9) will eventually allow the important simplification of setting  $BD_\mu D_\nu \rightarrow 0$  by continuity, which is analogous to what happens in abelian gauge theory when the continuity equation  $\partial_\nu J^\nu = 0$  is applied.

So, the task now is to find the unknowns  $A$  and  $B$ . If we place (6.9) into (6.8) we obtain:

$$\begin{aligned}
 -\delta^\sigma{}_\mu &= (Ag_{\mu\nu} + BD_\mu D_\nu) (g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu) \\
 &= Ag_{\mu\nu} g^{\nu\sigma} (D_\tau D^\tau + m^2) - Ag_{\mu\nu} D^\sigma D^\nu + BD_\mu D_\nu g^{\nu\sigma} (D_\tau D^\tau + m^2) - BD_\mu D_\nu D^\sigma D^\nu. \quad (6.10) \\
 &= A\delta^\sigma{}_\mu (D_\tau D^\tau + m^2) - AD^\sigma D_\mu + BD_\mu D^\sigma (D_\tau D^\tau + m^2) - BD_\mu D_\nu D^\sigma D^\nu
 \end{aligned}$$

Matching up the terms with  $\delta^\sigma{}_\mu$  we first obtain  $-1 = A(D_\tau D^\tau + m^2)$ , or inverting:

$$A = -(D_\tau D^\tau + m^2)^{-1}. \quad (6.11)$$

We then use (6.11) in (6.10) and reduce, to next obtain:

$$0 = (D_\tau D^\tau + m^2)^{-1} D^\sigma D_\mu + B(D_\mu D^\sigma (D_\tau D^\tau + m^2) - D_\mu D_\nu D^\sigma D^\nu), \quad (6.12)$$

or, rearranged:

$$B = -(D_\tau D^\tau + m^2)^{-1} D^\sigma D^\alpha (D^\alpha D^\sigma (D_\tau D^\tau + m^2) - D^\alpha D_\tau D^\sigma D^\tau)^{-1}. \quad (6.13)$$

Finally, we use (6.11) and (6.13) in (6.9) to find that:

$$I_{\mu\nu} = -(D_\tau D^\tau + m^2)^{-1} \left[ g_{\mu\nu} + D^\sigma D^\alpha (D^\alpha D^\sigma (D_\tau D^\tau + m^2) - D^\alpha D_\tau D^\sigma D^\tau)^{-1} D_\mu D_\nu \right]. \quad (6.14)$$

Above, each derivative pair is defined by  $D^\sigma D^\nu \equiv \partial^\sigma \partial^\nu - iG^\sigma \partial^\nu - 2G^\sigma G^\nu + G^\nu G^\sigma$  in (5.16) and  $\square = D_\tau D^\tau = \partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau$  in (5.17) (remember too, that  $\partial_\tau G^\tau = 0$  which produces (6.6) and (6.7)). We may then substitute (6.14) into the original definition  $G_\mu \equiv I_{\mu\nu} J^\nu$  to conclude that:

$$\begin{aligned}
 G_\mu &= I_{\mu\nu} J^\nu = -(D_\tau D^\tau + m^2)^{-1} \left[ g_{\mu\nu} + D^\sigma D^\alpha (D^\alpha D^\sigma (D_\tau D^\tau + m^2) - D^\alpha D_\tau D^\sigma D^\tau)^{-1} D_\mu D_\nu \right] J^\nu \\
 &= -(D_\tau D^\tau + m^2)^{-1} g_{\mu\nu} J^\nu - (D_\tau D^\tau + m^2)^{-1} D^\sigma D^\alpha (D^\alpha D^\sigma (D_\tau D^\tau + m^2) - D^\alpha D_\tau D^\sigma D^\tau)^{-1} D_\mu D_\nu J^\nu. \quad (6.15) \\
 &= -(D_\tau D^\tau + m^2)^{-1} J_\mu = -(\partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau + m^2)^{-1} J_\mu
 \end{aligned}$$

In an essential step, we get to the final line by enforcing continuity  $D_\nu J^\nu = 0$  from (5.20), and then making use of the d'Alembertian  $\square = D_\tau D^\tau$  of (5.17). We shall shortly add a term  $-i\partial_\tau G^\tau = 0$  to the expression for which the inverse is being taken, so that we can take advantage of (6.6) and explicitly identify the perturbations  $V$ .

To make all of this appear a bit more familiar to the way such inverses are usually written, let us set  $D \rightarrow \partial$  in (6.14), and let us assume flat spacetime so all derivatives commute,  $[\partial_\mu, \partial_\nu] = 0$ . With these assumptions, the inverses can be treated as regular denominators. With all this, we find embedded in (6.15), the very familiar abelian ( $A$  subscript) inverse  $I_{\mu\nu} \rightarrow I_{A\mu\nu}$ :

$$I_{A\mu\nu} = -\frac{g_{\mu\nu} + \frac{\partial^\sigma \partial^\alpha \partial_\mu \partial_\nu}{\partial^\alpha \partial^\sigma (\partial_\tau \partial^\tau + m^2) - \partial^\alpha \partial_\tau \partial^\sigma \partial^\tau}}{\partial_\tau \partial^\tau + m^2} = -\frac{g_{\mu\nu} + \frac{\partial_\mu \partial_\nu}{m^2}}{\partial_\tau \partial^\tau + m^2} \xrightarrow{i\partial \rightarrow k} \frac{g_{\mu\nu} - \frac{k_\mu k_\nu}{m^2}}{k_\tau k^\tau - m^2} \xrightarrow{+i\epsilon} \frac{g_{\mu\nu} - \frac{k_\mu k_\nu}{m^2}}{k_\tau k^\tau - m^2 + i\epsilon}. \quad (6.16)$$

With the first arrow, we convert to momentum space via  $i\partial_\mu \rightarrow k_\mu$ . With the second arrow, we then add the  $+i\epsilon$  prescription. Using the final term above with  $G_{A\mu} = I_{A\mu\nu} J^\nu$ , we may write:

$$G_{A\mu} = I_{A\mu\nu} J^\nu = \frac{g_{\mu\nu} - \frac{k_\mu k_\nu}{m^2}}{k_\tau k^\tau - m^2 + i\epsilon} J^\nu \stackrel{k_\nu J^\nu = 0}{=} \frac{g_{\mu\nu}}{k_\tau k^\tau - m^2 + i\epsilon} J^\nu = \frac{1}{k_\tau k^\tau - m^2 + i\epsilon} J_\mu, \quad (6.17)$$

where  $k_\nu J^\nu = i\partial_\nu J^\nu = 0$ , which is just another version of the continuity equation, is used for the reduction after the third equal sign. If we set  $m=0$  in (6.16) we then obtain the clearly indeterminate result:

$$I_{A\mu\nu} = \frac{g_{\mu\nu} - \frac{k_\mu k_\nu}{0}}{k_\tau k^\tau + i\epsilon} = \frac{g_{\mu\nu} - \infty}{k_\tau k^\tau + i\epsilon} = -\infty. \quad (6.18)$$

But in contrast, doing the same in (6.17) simply yields the finite:

$$G_{A\mu} = \frac{1}{k_\tau k^\tau + i\epsilon} J_\mu. \quad (6.19)$$

The infinite result in (6.18) is tamed in (6.19) *because of the continuity imposed in (6.17)*. If we then put the boson on mass shell,  $k_\tau k^\tau = 0$ , we finally have:

$$G_{A\mu} = \frac{1}{i\epsilon} J_\mu. \quad (6.20)$$

This only stays finite because of the  $+i\epsilon$  prescription. Equation (6.18) explicitly illustrates why  $g^{\nu\sigma} \partial_\tau \partial^\tau - \partial^\sigma \partial^\nu$  has no inverse, or more precisely, why the abelian inverse for a massless gauge boson in flat spacetime is indeterminately-infinite. Equation (6.20) explicitly illustrates why this inverse is also indeterminately-infinite for on-shell bosons, unless one uses the  $+i\epsilon$  prescription.

Now let us do the same in the *non-abelian* inverse (6.14) to see whether the same infinities are encountered. Setting  $m = 0$  in (6.14) we simply obtain:

$$I_{\mu\nu} = -\left(D_\tau D^\tau\right)^{-1} \left[ g_{\mu\nu} + D^\sigma D^\alpha \left( D^\alpha D^\sigma D^\tau D_\tau - D^\alpha D^\tau D^\sigma D_\tau \right)^{-1} D_\mu D_\nu \right]. \quad (6.21)$$

The term  $D^\alpha D^\sigma D^\tau D_\tau - D^\alpha D^\tau D^\sigma D_\tau = D^\alpha \left[ D^\sigma, D^\tau \right] D_\tau$  must be evaluated using the  $D^\sigma D^\nu$  and  $D_\tau D^\tau$  of (5.16) and (5.17), that is, as a second order quadratic rather than a fourth order linear term. That is because these derivatives were obtained prior to inversion by operating on  $G_\sigma$  and because the explicit form of a gauge-covariant derivative depends upon its operand. Thus, from (5.16) and (5.17):

$$\begin{aligned} D^\alpha D^\sigma D^\tau D_\tau - D^\alpha D^\tau D^\sigma D_\tau &= (D^\alpha D^\sigma) (D^\tau D_\tau) - (D^\alpha D^\tau) (D^\sigma D_\tau) \\ &= (\partial^\alpha \partial^\sigma - iG^\alpha \partial^\sigma - 2G^\alpha G^\sigma + G^\sigma G^\alpha) (\partial^\tau \partial_\tau - iG^\tau \partial_\tau - G^\tau G_\tau) \\ &\quad - (\partial^\alpha \partial^\tau - iG^\alpha \partial^\tau - 2G^\alpha G^\tau + G^\tau G^\alpha) (\partial^\sigma \partial_\tau - iG^\sigma \partial_\tau - 2G^\sigma G_\tau + G_\tau G^\sigma) \end{aligned} \quad (6.22)$$

If it was possible to commute  $\left[ D^\sigma, D^\tau \right] = 0$ , then this term would become zero and (6.21) would contain  $\left( D^\alpha \left[ D^\sigma, D^\tau \right] D_\tau \right)^{-1} = 0^{-1} = \infty$  and become indeterminate when the mass is zero for the same reason as (6.18). But the defining feature of non-abelian gauge theory is that the gauge fields do not commute, i.e., that  $\left[ G^\sigma, G^\tau \right] \neq 0$ . So the term (6.22) is *not* zero and thus (6.21) does not become infinite *even when the mass is set to zero*. It is the non-commuting nature of non-Abelian gauge theory that bears direct responsibility for maintaining a finite inverse (6.21) for the configuration space operator  $g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu$  in (6.1) *even when the gauge boson has no mass*. As we see in (6.15), however, none of this matters at all once we apply  $D_\nu J^\nu = 0$  continuity, because that zeroes out the term in (6.22) entirely. Indeed, setting  $m = 0$  in the final line of the non-abelian relation (6.15) for  $G(J)$  simply yields

$$G_\mu = -\left(D_\tau D^\tau\right)^{-1} J_\mu = -\left(\partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau\right)^{-1} J_\mu. \quad (6.23)$$

Now let us examine what happens for on-shell bosons in non-abelian gauge theory. The relativistic energy relationship is  $p_\sigma p^\sigma - m^2 = 0$ . Via  $\eta^{\sigma\tau} = \frac{1}{2}(\gamma^\sigma \gamma^\tau + \gamma^\tau \gamma^\sigma) = \frac{1}{2}\{\gamma^\sigma, \gamma^\tau\}$  this becomes  $(\not{p} - m)u = 0 \Leftrightarrow (i\partial - m)\psi = 0$  when operating on a free, non-interacting Dirac spinor / wavefunction. But for interaction via a gauge field  $G^\tau$ ,  $p_\sigma p^\sigma - m^2 = 0$  becomes  $\pi_\sigma \pi^\sigma - m^2 = 0$  with  $\pi^\tau \equiv p^\tau + G^\tau$  defining the *kinetic momentum*  $\pi^\tau$  in relation to the canonical momentum  $p^\tau$  and the gauge field  $G^\tau$ . This means that  $(\not{\pi} - m)u = (\not{p} + \not{G} - m)u = 0$ , or, with  $p \rightarrow i\partial$  and  $u \rightarrow \psi$ , that  $(i\partial + \not{G} - m)\psi = 0$ . This is just Dirac's equation for an *interacting* fermion. The

key point of all this – with  $p^\tau$  and  $k^\sigma$  respectively used to denote fermion and boson momentum vectors – is that a free on-shell fermion is described by  $p_\sigma p^\sigma - m^2 = 0$  and a free on-shell gauge boson by  $k_\sigma k^\sigma - m^2 = 0$ . But for an interacting on-shell particle with  $\pi^\tau \equiv p^\tau + G^\tau$  for fermions and  $\pi^\tau \equiv k^\tau + G^\tau$  for bosons, the exact form of the on-shell equation depends on whether  $G^\tau$  is an abelian or a non-abelian gauge field. Let us see why:

Suppose that  $G^\tau$  is a U(1) photon / electromagnetic potential  $A^\tau$ . Here the on-shell relationship, referring also to the perturbation (6.7) and noting that  $\pi_\sigma \pi^\sigma = k_\sigma k^\sigma - V$  because  $k_\tau G^\tau = 0$ , is:

$$\begin{aligned} 0 &= \pi_\sigma \pi^\sigma - m^2 = (k_\sigma + A_\sigma)(k^\sigma + A^\sigma) - m^2 = k_\sigma k^\sigma + k_\sigma A^\sigma + A_\sigma k^\sigma + A_\sigma A^\sigma - m^2 \\ &= -V + k_\sigma k^\sigma - m^2 \end{aligned} \quad (6.24)$$

This perturbation  $-V = k_\tau A^\tau + A_\tau k^\tau + A_\tau A^\tau$  is a 1x1 scalar number which can be added to the number  $k_\sigma k^\sigma - m^2$ , so that (6.24) is a sensible equation. But suppose now that  $G^\tau = \lambda^i G^{i\tau}$  is an NxN object formed using the generators  $\lambda^i$  of the simple gauge group SU(N). To be explicit, showing Yang-Mills indexes  $A, B = 1 \dots N$  for the fundamental SU(N) representation, suppose now that  $G^\tau_{AB} = \lambda^i_{AB} G^{i\tau}$ . Then, if carelessly generalized, (6.24) would become:

$$\begin{aligned} 0 &= \pi_\sigma \pi^\sigma - m^2 = (k_\sigma + G_\sigma)(k^\sigma + G^\sigma) - m^2 = k_\sigma k^\sigma + k_\sigma G^\sigma_{AB} + G_{\sigma AB} k^\sigma + (G_\sigma G^\sigma)_{AB} - m^2 \\ &= -V_{AB} - \delta_{AB} (m^2 - k_\sigma k^\sigma) \quad (= -V + k_\sigma k^\sigma - m^2) \end{aligned} \quad (6.25)$$

But this expression is not quite right. The  $k_\sigma k^\sigma - m^2$  is still a scalar number, and because  $V_{AB}$  is now taken to be an NxN object for SU(N), the  $k_\sigma k^\sigma - m^2$  will occupy the *diagonal* positions in the overall expression (6.25), hence the explicit showing of  $\delta_{AB} (m^2 - k_\sigma k^\sigma)$ . At the same time,  $-V_{AB} = k_\sigma G^\sigma_{AB} + G_{\sigma AB} k^\sigma + (G_\sigma G^\sigma)_{AB}$  will now be an NxN Hermitian matrix with off-diagonal elements. The perturbation  $V_{AB}$  is a matrix, while  $k_\sigma k^\sigma - m^2$  is a scalar number that we also know is part of an inverse abelian propagator. So the only way to make sense out of (6.25) is to use this as an *eigenvalue equation* in which  $m^2 - k_\sigma k^\sigma$  represents the scalar eigenvalues of the perturbation  $-V_{AB}$ .

Now, one way to write (6.25) as an eigenvalue equation, is to have it operate on an N-component column vector  $\varphi$ , and to rewrite the non-abelian on-shell condition as  $[V_{AB} - \delta_{AB} (m^2 - k_\sigma k^\sigma)]\varphi = 0$ . But because expressions such as (6.25) will show up in the context of equations such as (6.15), we want to be able to express the on-shell condition

independently of any  $\varphi$ . We can do so by taking the determinant  $|A| = \det A$  of (6.25), in the form of the eigenvalue equation:

$$0 = \left| \pi_\sigma \pi^\sigma - m^2 \right| = - \left| V_{AB} - \delta_{AB} (k_\sigma k^\sigma - m^2) \right| \quad \left( = \left| -V + k_\sigma k^\sigma - m^2 \right| \right). \quad (6.26)$$

This is what specifies an on-shell gauge boson in non-abelian gauge theory: on shell, the scalar number  $k_\sigma k^\sigma - m^2$  gives the eigenvalue solutions of the perturbation  $V_{AB}$ .

In view of this, if we therefore write (6.15) with  $+i\mathcal{E}$  and  $\pi^\tau \equiv k^\tau + G^\tau$  as

$$\begin{aligned} G_\mu &= - \left( D_\tau D^\tau + m^2 - i\mathcal{E} \right)^{-1} J_\mu = - \left( \partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau + m^2 - i\mathcal{E} \right)^{-1} J_\mu \\ &\stackrel{i\partial \rightarrow k}{\Rightarrow} \left( k_\tau k^\tau + G_\tau k^\tau + G_\tau G^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\mu \\ &\stackrel{k_\tau G^\tau = 0}{\Rightarrow} \left( k_\tau k^\tau + k_\tau G^\tau + G_\tau k^\tau + G_\tau G^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\mu = \left( \pi_\tau \pi^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\mu \\ &= \left( -V + k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\mu \end{aligned}, \quad (6.27)$$

we see by writing (6.17) in the form of an inverse:

$$G_{A\mu} = \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\mu, \quad (6.28)$$

that the *sole* difference between the abelian and non-abelian solutions for  $G_\mu(J_\mu)$  is that the canonical scalar  $k_\tau k^\tau$  of abelian gauge theory is replaced by the kinetic scalar  $\pi_\tau \pi^\tau$  in non-abelian gauge theory, or, alternatively and equivalently, that a perturbation  $-V = -V_{AB}$  is added to the abelian (6.28) to arrive at the non-abelian (6.27), which then turns the usual inverse propagator  $k_\tau k^\tau - m^2 + i\mathcal{E}$  into  $-V_{AB} + k_\tau k^\tau - m^2 + i\mathcal{E}$  for which on-shell particles are described by  $\left| V_{AB} - \delta_{AB} (k_\sigma k^\sigma - m^2) \right| = 0$  in (6.26).

If the “careless”  $\pi_\sigma \pi^\sigma - m^2 = 0$  in (6.25) were to describe the on-shell condition for an interacting particle in non-abelian gauge theory – which it does *not* – then for an on-shell particle, (6.27) in the form  $G_\mu = \left( \pi_\tau \pi^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\mu$  would reduce to  $G_\mu = (+i\mathcal{E})^{-1} J_\mu$  which is exactly the same as the abelian (6.20). So in either abelian or non-abelian gauge theory, we would require the  $+i\mathcal{E}$  prescription to avoid the poles for an on-shell particle. However,  $\pi_\sigma \pi^\sigma - m^2 = 0$  is *not* the on-shell condition for non-abelian gauge theory. Rather, on-shell bosons are specified by the eigenvalue equation  $\left| \pi_\sigma \pi^\sigma - m^2 \right| = 0$  of (6.26). So even with  $\left| \pi_\sigma \pi^\sigma - m^2 \right| = 0$ , the expression  $G_\mu = \left( \pi_\tau \pi^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\mu$  will generally remain finite in non-

abelian gauge theory even if we use  $G_\mu = (\pi_\tau \pi^\tau - m^2)^{-1} J_\mu$  absent  $+i\epsilon$ . Because on shell particles are described by  $|\pi_\sigma \pi^\sigma - m^2| = 0$  and not  $\pi_\sigma \pi^\sigma - m^2 = 0$  in non-abelian gauge theory, the non-abelian theory can remain finite on shell even absent  $+i\epsilon$ .

Before studying massless gauge bosons using the more formal approach of Faddeev-Popov, it is important to see that the continuity relation  $\partial_\nu J^\nu = 0$  which tames  $G_\mu(J_\mu)$  in the classical massless boson inverse (6.19) notwithstanding the infinite inverse (6.18), plays a similar role in taming the quantum field amplitude obtained from the QED path integral. Specifically, the action corresponding to the field equation  $-J^\nu = (g^{\nu\sigma}(\partial_\tau \partial^\tau + m^2) - \partial^\sigma \partial^\nu) G_\sigma$  which is the abelian version of (6.3), for which the inverse was found in (6.16), is:

$$S(G) = \int d^4x \mathcal{L} = \int d^4x \left[ \frac{1}{2} G_\nu (g^{\nu\sigma}(\partial_\tau \partial^\tau + m^2) - \partial^\nu \partial^\sigma) G_\sigma - G_\sigma J^\sigma \right]. \quad (6.29)$$

When the Gaussian integral  $\int dx \exp(-\frac{1}{2} Ax^2 - Jx) = (-2\pi/A)^{1/2} \exp(J^2/2A)$  is employed as the template to use (6.29) in  $Z = \int DG \exp(iS(G)) \equiv C \exp(iW(J))$ , the inverse in  $J^2/2A$  is based on the abelian inverse  $I_{A\mu\nu}$  in (6.16), and we obtain (see, e.g., [11], pages 30-31):

$$W(J) = -\frac{1}{2} \int \frac{d^4k}{(2\pi)^4} J^\mu(k) * \frac{-g_{\mu\nu} + \frac{k_\mu k_\nu}{m}}{k_\tau k^\tau - m + i\epsilon} J^\nu(k) = \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} J^\mu(k) * I_{A\mu\nu} J^\nu(k). \quad (6.30)$$

This too looks like it will become singular for  $m = 0$ , just like (6.18). But there too, as in (6.17), the continuity relationship  $k_\nu J^\nu = i\partial_\nu J^\nu = 0$  rescues the path integral from an indeterminate fate, and facilitates the reduction:

$$W(J) = +\frac{1}{2} \int \frac{d^4k}{(2\pi)^4} J^\mu(k) * \frac{1}{k_\tau k^\tau - m + i\epsilon} J_\mu(k) \stackrel{m=0}{\Rightarrow} +\frac{1}{2} \int \frac{d^4k}{(2\pi)^4} J^\mu(k) * \frac{1}{k_\tau k^\tau + i\epsilon} J_\mu(k). \quad (6.31)$$

This also tells us that the electromagnetic force between like-charges is *repulsive*.

But the key feature of interest in both (6.17) which is for a classical field and (6.31) which is for a quantum field, is that even though the *mathematical* abelian inverse (6.16) becomes infinite if  $m = 0$ , when this inverse is placed into the context of a *physical* equation such as  $G_{A\mu} = I_{A\mu\nu} J^\nu$  in (6.17) or  $\dots J^\mu * I_{A\mu\nu} J^\nu$  in (6.30), the seemingly-infinite result becomes finite and well-behaved. This is because the physical context – in this case the continuity relation  $k_\nu J^\nu = i\partial_\nu J^\nu = 0$  – causes the otherwise singular term  $k_\mu k_\nu / m \rightarrow k_\mu k_\nu / 0 = \infty$  to be zeroed out *before it ever gets to wreak any havoc*. This contextual finiteness is very important, because even though the mathematical object – the inverse – becomes singular, the physical

result remains finite. In the discussion to now be developed, where we use the more formal approach of Faddeev-Popov to develop the massive gauge bosons, this will lead to what we shall call “contextual gauge fixing.” In Faddeev-Popov, where a gauge number  $\xi$  enables an unlimited array of non-unique *mathematical* inverses, the continuity relation forces the *physical* results into a very definite and unique choice of gauge. When we use these same inverses in  $\oint\!\!\!\int F(G(J))$  to show why  $\oint\!\!\!\int F = -i\int\!\!\!\int dGG$  looks very much like a baryon, this type of “contextual gauge fixing” coupled with Fermi-Dirac-Pauli Exclusion will not only result in unique solutions for  $G(J)$ , but will give mass to the fermions of  $J^\mu = \bar{\psi}\gamma^\mu\psi$  and turn them into quarks, while rendering the massive gauge bosons massless just like gluons.

## 7. Abelian and non-Abelian Massless Gauge Boson Inverses for the Electric Charge Density, Using the Faddeev-Popov Method

In the last section we took the “pragmatic” Coleman-Zee approach of obtaining the classical field equation inverse for a massive gauge boson and then setting the mass to zero to see what happens under a variety of circumstances. Now, we take the more formal, direct approach of using the Faddeev-Popov method to calculate the inverse for a massless gauge boson *ab initio*, without the intermediate stop for a massive boson.

If we take the “non-pragmatic” route and start out with a *massless* gauge boson for which we apply Faddeev-Popov, and to open simplified discussion revert (5.15) to its abelian limit  $D \rightarrow \partial$ , then along the way the *effective* field equation becomes (see [11], after (III.4(8))):

$$-J^\nu = \left( g^{\nu\sigma}\partial_\tau\partial^\tau - (1-1/\xi)\partial^\sigma\partial^\nu \right) G_\sigma, \quad (7.1)$$

where  $\xi$  is a gauge number. While for the moment we treat the introduction of  $\xi$  simply as a mathematical manipulation of the classical field equation  $-J^\nu = \left( g^{\nu\sigma}\partial_\tau\partial^\tau - \partial^\sigma\partial^\nu \right) G_\sigma$  of (5.15) to which (7.1) reduces for  $\xi = \infty$ , we keep in mind that  $\xi$  actually arises when we start with a path integral  $Z = \int DG \exp(iS(G))$  and turn this into  $Z = \int DG \exp\left(i\left[S(G) - (i/2\xi)\int d^4x(\partial G)^2\right]\right)$  through a change of the integration variable which maintains the invariance of the  $Z$  under the abelian gauge transformation  $G \rightarrow G' = G + d\theta$ . So by introducing  $\xi$  in this way, and knowing that this carries over to non-abelian gauge theory but for the further introduction of ghost fields  $c^\dagger, c$  with a path integral  $Z = \int DGDcDc^\dagger \exp\left(i\left[S(G) - (1/2\xi)\int d^4x(\partial G)^2\right] + S(c^\dagger, c)\right)$  containing a ghost action  $S(c^\dagger, c)$ , we have a “hook” by which this can eventually be used to set up a quantum path integration for non-abelian theory. But for now, as discussed at length in section 4, we continue to develop the classical theory.

Once again using an inner-product definition  $G_{A\mu} \equiv I_{A\mu\nu}J^\nu$  for the abelian inverse, in flat spacetime we may multiply through by  $-I_{A\mu\nu}$  and write (7.1) as (contrast (6.1)):

$$I_{A\mu\nu}J^\nu = -I_{A\mu\nu} \left( g^{\nu\sigma} \partial_\tau \partial^\tau - (1-1/\xi) \partial^\sigma \partial^\nu \right) G_\sigma = G_\mu = \delta^\sigma{}_\mu G_\sigma \quad (7.2)$$

from which we extract (contrast (6.2)):

$$I_{A\mu\nu} \left( g^{\nu\sigma} \partial_\tau \partial^\tau - (1-1/\xi) \partial^\sigma \partial^\nu \right) = -\delta^\sigma{}_\mu. \quad (7.3)$$

Then using  $I_{A\mu\nu} \equiv Ag_{\mu\nu} + B\partial_\mu \partial_\nu$  based on (6.9), this becomes (contrast (6.10)):

$$\begin{aligned} -\delta^\sigma{}_\mu &= \left( Ag_{\mu\nu} + B\partial_\mu \partial_\nu \right) \left( g^{\nu\sigma} \partial_\tau \partial^\tau - (1-1/\xi) \partial^\sigma \partial^\nu \right) \\ &= Ag_{\mu\nu} g^{\nu\sigma} \partial_\tau \partial^\tau - Ag_{\mu\nu} (1-1/\xi) \partial^\sigma \partial^\nu + B\partial_\mu \partial_\nu g^{\nu\sigma} \partial_\tau \partial^\tau - B\partial_\mu \partial_\nu (1-1/\xi) \partial^\sigma \partial^\nu. \\ &= A\delta^\sigma{}_\mu \partial_\tau \partial^\tau - A(1-1/\xi) \partial^\sigma \partial_\mu + B\partial_\mu \partial^\sigma \partial_\tau \partial^\tau - B(1-1/\xi) \partial_\mu \partial_\nu \partial^\sigma \partial^\nu \end{aligned} \quad (7.4)$$

From this we match up the  $\delta^\sigma{}_\mu$  terms to find (contrast (6.11)):

$$A = -1/\partial_\tau \partial^\tau. \quad (7.5)$$

so that (cf. (6.12)):

$$0 = \frac{(1-1/\xi) \partial^\sigma \partial_\mu}{\partial_\tau \partial^\tau} + B \left( \partial_\mu \partial^\sigma \partial_\tau \partial^\tau - (1-1/\xi) \partial_\mu \partial_\tau \partial^\sigma \partial^\tau \right), \quad (7.6)$$

or, commuting and cancelling derivatives freely (cf. (6.13)):

$$B = -\frac{(1-1/\xi) \frac{\partial^\sigma \partial_\mu}{\partial_\mu \partial^\sigma \partial_\tau \partial^\tau - (1-1/\xi) \partial_\mu \partial_\tau \partial^\sigma \partial^\tau}}{\partial_\tau \partial^\tau} = -\frac{\left( \frac{1-1/\xi}{1/\xi} \right) \frac{1}{\partial_\alpha \partial^\alpha}}{\partial_\tau \partial^\tau} = \frac{(1-\xi) \frac{1}{\partial_\alpha \partial^\alpha}}{\partial_\tau \partial^\tau}. \quad (7.7)$$

Thus, using (7.5) and (7.7) in  $I_{A\mu\nu} \equiv Ag_{\mu\nu} + B\partial_\mu \partial_\nu$  we obtain (cf. (6.14) and (6.16)):

$$I_{A\mu\nu} = \frac{-g_{\mu\nu} + (1-\xi) \frac{\partial_\mu \partial_\nu}{\partial_\alpha \partial^\alpha}}{\partial_\tau \partial^\tau} \xrightarrow{i\partial \rightarrow k} -\frac{-g_{\mu\nu} + (1-\xi) \frac{k_\mu k_\nu}{k_\alpha k^\alpha}}{k_\tau k^\tau} \xrightarrow{+i\epsilon} \frac{g_{\mu\nu} - (1-\xi) \frac{k_\mu k_\nu}{k_\alpha k^\alpha}}{k_\tau k^\tau + i\epsilon}. \quad (7.8)$$

We then use this in  $G_{A\mu} \equiv I_{A\mu\nu} J^\nu$  to write:

$$G_{A\mu} \equiv I_{A\mu\nu} J^\nu = \frac{g_{\mu\nu} - (1-\xi) \frac{k_\mu k_\nu}{k_\alpha k^\alpha}}{k_\tau k^\tau + i\epsilon} J^\nu. \quad (7.9)$$

Now let us follow two different routes to reduce (7.9). First, let us apply the continuity relation  $k_\nu J^\nu = i\partial_\nu J^\nu = 0$  as we did in (6.17). This causes (7.9) to become:

$$G_{A\mu} = \frac{g_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k_\alpha k^\alpha}}{k_\tau k^\tau + i\epsilon} J^\nu \xrightarrow{k_\nu J^\nu = 0} \frac{g_{\mu\nu} - (1-\xi)0}{k_\tau k^\tau + i\epsilon} J^\nu = \frac{g_{\mu\nu}}{k_\tau k^\tau + i\epsilon} J^\nu = \frac{1}{k_\tau k^\tau + i\epsilon} J_\mu. \quad (7.10)$$

Alternatively, let us embark upon the different path of selecting the Feynman gauge  $\xi = 1$  in (7.9). Now we have:

$$G_{A\mu} = \frac{g_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k_\alpha k^\alpha}}{k_\tau k^\tau + i\epsilon} J^\nu \xrightarrow{\xi=1} \frac{g_{\mu\nu} - 0}{k_\tau k^\tau + i\epsilon} J^\nu = \frac{g_{\mu\nu}}{k_\tau k^\tau + i\epsilon} J^\nu = \frac{1}{k_\tau k^\tau + i\epsilon} J_\mu, \quad (7.11)$$

which is the exact same result as in (7.10). And both of these are exactly the same as the result in (6.19). These are three routes to the exact same result. In (7.10), the expression  $(1-\xi)0$  which emerges from requiring continuity via  $k_\nu J^\nu = i\partial_\nu J^\nu = 0$  has *forced* this term to be zeroed out. Just as in (6.17) (and analogously in the non-abelian (6.15)), there is no choice other than to zero out the term containing the gauge number  $\xi$ . But if we were unaware of continuity, we could get to the same *effective* inverse  $I_{A\mu\nu} = g_{\mu\nu} / (k_\tau k^\tau + i\epsilon)$  *in general*, by the different route of selecting the Feynman gauge  $\xi = 1$ . Importantly, this means that after we find the inverse and then use it in  $G_{A\mu} \equiv I_{A\mu\nu} J^\nu$ , we are forced into an equation for  $G_{A\mu}$  which could be independently arrived at by selecting the Feynman gauge  $\xi = 1$  for the standalone inverse.

The point here is that for a *massless* gauge boson, there is a complete freedom to select *any* gauge number  $-\infty \leq \xi \leq \infty$  for the inverse  $I_{A\mu\nu}$ , which means that this inverse is *infinitely non-unique* when regarded as a *mathematical* entity. This is because of the redundancy whereby  $G_\mu$  contains four degrees of freedom despite the associated massless physical field having only two degrees of freedom. And here, unlike for a massless boson, we do not even have  $D_\sigma G^\sigma = 0$  mandated as a covariant condition which at least takes out one degree of freedom. So the mathematical inverse is highly nonunique. Nevertheless, once we use this inverse in a *physical* equation such as  $G_{A\mu} \equiv I_{A\mu\nu} J^\nu$  in (7.9) to (7.11), the continuity equation forces us to fix the gauge of the inverse into  $\xi = 1$ , or more precisely, forces a result that can equivalently be achieved by selecting  $\xi = 1$  for the standalone inverse before it is ever inserted into  $G_{A\mu} \equiv I_{A\mu\nu} J^\nu$ . This is a specific example of the “*contextual gauge fixing*” mentioned at the end of section 6, wherein a gauge which is completely non-unique and thus an associated inverse which is also non-unique as a mathematical matter, is forced to be unique *when placed into a physical context*, in this case, the context of a conserved current density enforced by continuity.

In this way, we may think of the Feynman gauge as the “continuity gauge,” because it *uniquely* fixes the inverse in the exact same manner as does the continuity equation  $k_\nu J^\nu = i\partial_\nu J^\nu = 0$ .

With (7.1) to (7.11) as a backdrop, we return to the field equation (5.15) with  $D^\sigma D^\nu$  and  $D_\tau D^\tau$  defined as in (5.16) and (5.17) when the operand is  $G_\sigma$ , and introduce the gauge number  $\xi$  exactly as we did in (7.1). Thus, we write:

$$-J^\nu = \left( g^{\nu\sigma} D_\tau D^\tau - (1-1/\xi) D^\sigma D^\nu \right) G_\sigma. \quad (7.12)$$

As with (7.1), we treat the introduction of  $\xi$  simply as a mathematical manipulation of (5.15) to which (7.12) will revert for  $\xi = \infty$ , which allows us to solve this classical equation (7.12) for  $G_\sigma$  as a function of  $J^\nu$ . Since  $Z = \int DGDcDc^\dagger \exp\left(i\left[S(G) - (1/2\xi) \int d^4x (\partial G)^2\right] + S(c^\dagger, c)\right)$  is the path integral for non-abelian gauge theory, it should be clear that the inverse obtained from (7.12) will be a useful item to have “on the shelf” when it comes time to try to calculate the non-ghost portion of this path integral. But for now, we are still working classically, so our imminent goal is to solve the classical equation (7.12) for  $G_\sigma$  as a function of  $J^\nu$ .

As we have done previously, we use  $G_\mu \equiv I_{\mu\nu} J^\nu$  to define  $I_{\mu\nu}$ , and then multiply each side of (7.12) by  $-I_{\mu\nu}$  to write:

$$I_{\mu\nu} J^\nu = -I_{\mu\nu} \left( g^{\nu\sigma} D_\tau D^\tau - (1-1/\xi) D^\sigma D^\nu \right) G_\sigma = G_\mu = \delta^\sigma_\mu G_\sigma. \quad (7.13)$$

From this we extract:

$$I_{\mu\nu} \left( g^{\nu\sigma} D_\tau D^\tau - (1-1/\xi) D^\sigma D^\nu \right) = -\delta^\sigma_\mu. \quad (7.14)$$

Then we combine the above with (6.9) to write (cf. (6.10) and (7.4)):

$$\begin{aligned} -\delta^\sigma_\mu &= \left( Ag_{\mu\nu} + BD_\mu D_\nu \right) \left( g^{\nu\sigma} D_\tau D^\tau - (1-1/\xi) D^\sigma D^\nu \right) \\ &= Ag_{\mu\nu} g^{\nu\sigma} D_\tau D^\tau - Ag_{\mu\nu} (1-1/\xi) D^\sigma D^\nu + BD_\mu D_\nu g^{\nu\sigma} D_\tau D^\tau - BD_\mu D_\nu (1-1/\xi) D^\sigma D^\nu. \quad (7.15) \\ &= A\delta^\sigma_\mu D_\tau D^\tau - A(1-1/\xi) D^\sigma D_\mu + BD_\mu D^\sigma D_\tau D^\tau - BD_\mu D_\nu (1-1/\xi) D^\sigma D^\nu \end{aligned}$$

Here, the reductions used twice earlier (cf. (6.11) to (6.13) and (7.5) to (7.7)) yield:

$$A = -\left( D_\tau D^\tau \right)^{-1}, \quad (7.16)$$

$$0 = (1-1/\xi) \left( D_\tau D^\tau \right)^{-1} D^\sigma D_\mu + BD_\mu D^\sigma D_\tau D^\tau - B(1-1/\xi) D_\mu D_\tau D^\sigma D^\tau, \quad (7.17)$$

$$B = -(1-1/\xi)(D_\tau D^\tau)^{-1} D^\sigma D^\alpha \left( D^\alpha D^\sigma D_\beta D^\beta - (1-1/\xi) D^\alpha D^\beta D^\sigma D_\beta \right)^{-1}, \quad (7.18)$$

thus leading via  $I_{\mu\nu} \equiv Ag_{\mu\nu} + BD_\mu D_\nu$  from (6.9), to:

$$I_{\mu\nu} = -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + (1-1/\xi) D^\alpha D^\beta \left( D^\beta D^\alpha D^\sigma D_\sigma - (1-1/\xi) D^\beta D^\sigma D^\alpha D_\sigma \right)^{-1} D_\mu D_\nu \right]. \quad (7.19)$$

Reducing (7.19) is a bit tricky because of the inverse. But if we momentarily put the latter inverse into a “denominator” and use a  $\surd$  marker to hold the commutation position of the inverse, all just to aid in visualization, we may reduce this to:

$$\begin{aligned} I_{\mu\nu} &= -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + \frac{(1-1/\xi) D^\alpha D^\beta \surd D_\mu D_\nu}{D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma + (1/\xi) D^\beta D^\sigma D^\alpha D_\sigma} \right] \\ &= -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + \frac{(\xi-1) D^\alpha D^\beta \surd D_\mu D_\nu}{\xi (D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma) + D^\beta D^\sigma D^\alpha D_\sigma} \right], \quad (7.20) \\ &= -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + (\xi-1) D^\alpha D^\beta \left( \xi (D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma) + D^\beta D^\sigma D^\alpha D_\sigma \right)^{-1} D_\mu D_\nu \right] \end{aligned}$$

where in the middle line we multiply each of the “numerator” and “denominator” by  $\xi$ , then in the final line revert to the inverse formulation.

In this form, we see that the redundancy of  $G_\mu$  with four degrees of freedom to describe a massless field that has two degrees of freedom permits an infinite non-uniqueness  $-\infty \leq \xi \leq \infty$  in the choice of the gauge number, just as it does in abelian gauge theory, see after (7.11). But now, as before, let us insert this inverse (7.20) into  $G_\mu = I_{\mu\nu} J^\nu$  to obtain:

$$G_\mu = -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + (\xi-1) D^\alpha D^\beta \left( \xi (D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma) + D^\beta D^\sigma D^\alpha D_\sigma \right)^{-1} D_\mu D_\nu \right] J^\nu. \quad (7.21)$$

As in (7.10) and (7.11) we now take two routes to reduce (7.21). For the first route, we apply the non-abelian continuity relationship  $D_\nu J^\nu = 0$  deduced in (5.20) to obtain:

$$\begin{aligned} G_\mu &= -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + (\xi-1) D^\alpha D^\beta \left( \xi (D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma) + D^\beta D^\sigma D^\alpha D_\sigma \right)^{-1} D_\mu D_\nu \right] J^\nu \\ &\stackrel{D_\nu J^\nu=0}{\Rightarrow} -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} J^\nu + (\xi-1) D^\alpha D^\beta \left( \xi (D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma) + D^\beta D^\sigma D^\alpha D_\sigma \right)^{-1} D_\mu (0) \right]. \quad (7.22) \\ &= -(D_\tau D^\tau)^{-1} J_\mu \end{aligned}$$

For the second route, we simply select the Feynman gauge  $\xi = 1$  in (7.21). Now we obtain:

$$\begin{aligned}
 G_\mu &= -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + (\xi - 1) D^\alpha D^\beta \left( \xi (D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma) + D^\beta D^\sigma D^\alpha D_\sigma \right)^{-1} D_\mu D_\nu \right] J^\nu \\
 &\stackrel{\xi=1}{\Rightarrow} -(D_\tau D^\tau)^{-1} \left[ g_{\mu\nu} + (0) D^\alpha D^\beta \left( \xi (D^\beta D^\alpha D^\sigma D_\sigma - D^\beta D^\sigma D^\alpha D_\sigma) + D^\beta D^\sigma D^\alpha D_\sigma \right)^{-1} D_\mu D_\nu \right] J^\nu \quad .(7.23) \\
 &= -(D_\tau D^\tau)^{-1} J_\mu
 \end{aligned}$$

These two results (7.22) and (7.23) are exactly the same. So just as in the abelian (7.10) and (7.11), the Feynman gauge acts as a continuity gauge, because when used in the standalone inverse of (7.20), it leads us to the exact same result as the non-abelian continuity relationship  $D_\nu J^\nu = 0$ . Additionally, if we now return to (6.15) in which we have also employed continuity, and follow the Coleman-Zee approach of setting the gauge field mass  $m = 0$ , we also find just as in (7.22) and (7.23) that:

$$G_\mu = -(D_\tau D^\tau)^{-1} J_\mu = -(\partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau)^{-1} J_\mu \quad (7.24)$$

which we have already seen in (6.23), with  $D_\tau D^\tau = \partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau$  as found in (5.17), see also (6.6) and (6.7) which make use of  $\partial_\nu G^\nu = 0$  for a massive gauge boson and so are able to also provide a connection to the perturbation  $V$ .

So we see that in contextual setting of the continuity relationship  $D_\nu J^\nu = 0$ , the *unique* solution to the massless non-abelian field equation  $-J^\nu = (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) G_\sigma$  of (5.15) is *always* going to be  $G_\mu = -(D_\tau D^\tau)^{-1} J_\mu$ . Whether we arrive at (6.23) / (7.24) by starting with a *massive* gauge field, obtaining the inverse, applying continuity, and then setting  $m = 0$  via Coleman-Zee; whether we start with a *massless* gauge field, use Faddeev-Popov to find the inverse, and then apply continuity; or whether we start with a *massless* gauge field, use Faddeev-Popov to find the inverse, and then *choose* the Feynman/continuity gauge  $\xi = 1$ ; we will always end up with the same *unique* solution (7.22) to (7.24).

The point is that even for non-abelian gauge theory, while the *mathematical* inverse for a massless gauge field gives us the freedom to select *any* gauge number  $-\infty \leq \xi \leq \infty$ , the *physical* continuity condition  $D_\nu J^\nu = 0$  forces us to put the inverse into the Feynman gauge. This *contextual gauge fixing* removes the arbitrariness of the mathematical inverse, and forces us into the specific gauge  $\xi = 1$  the moment we use the inverse in  $G_\mu = I_{\mu\nu} J^\nu$  and then apply  $D_\nu J^\nu = 0$ .

Before concluding this section, let us compare the non-abelian results (7.22) to (7.24) all of which are equivalent to one another, with the abelian results (7.10) and (7.11) both of which are also equivalent to one another. The chief difference at this point is that we have not yet introduced the  $+i\epsilon$  prescription into the non-abelian inverses. Comparing (7.22) to (7.24) with (7.10) and (7.11), we see that the way to introduce  $+i\epsilon$  is to amend (7.24) as such:

$$G_\mu = -(D_\tau D^\tau - i\mathcal{E})^{-1} J_\mu = -(\partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau - i\mathcal{E})^{-1} J_\mu \xrightarrow{i\partial \rightarrow k} (k_\tau k^\tau + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu. \quad (7.25)$$

Above, we have also gone over into momentum space via  $i\partial \rightarrow k$ . This is just the second line of (6.27) with  $m=0$ . In the  $k_\tau G^\tau = 0$  gauge, which for a massless boson is a choice and not a requirement, this becomes:

$$G_\mu = (k_\tau k^\tau + k_\tau G^\tau + G_\tau k^\tau + G_\tau G^\tau + i\mathcal{E})^{-1} J_\mu = (\pi_\tau \pi^\tau + i\mathcal{E})^{-1} J_\mu = (-V + k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu. \quad (7.26)$$

In contrast, if we write (7.10) / (7.11) in the form of an inverse relation, these become:

$$G_{A\mu} = (k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu, \quad (7.27)$$

which is just (6.28) with  $m=0$ . Of course, the abelian  $(k_\tau k^\tau + i\mathcal{E})^{-1}$  can be written as an ordinary denominator, while the non-abelian  $(k_\tau k^\tau + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1}$  cannot because the  $G_\tau \pi^\tau = G_\tau k^\tau + G_\tau G^\tau$  term in general will have a matrix form which must be inverted rather than placed in a denominator.

Insofar as on-shell bosons are concerned, as noted in (6.28) and the discussion following, an on-shell boson in non-abelian gauge theory will be described by the eigenvalue equation (6.26), which for  $m=0$  and using (6.7) and  $\pi_\sigma \pi^\sigma = k_\sigma k^\sigma - V$  in the  $k_\tau G^\tau = 0$  gauge becomes:

$$0 = |\pi_\sigma \pi^\sigma| = -|V_{AB} - \delta_{AB} k_\sigma k^\sigma| \quad ( = |-V + k_\sigma k^\sigma| = |k_\sigma k^\sigma + k_\tau G^\tau + G_\tau k^\tau + G_\tau G^\tau| ). \quad (7.28)$$

Note again that while  $i\partial_\tau G^\tau = k_\tau G^\tau = 0$  is a *required* relation for a *massive* gauge boson as found in (6.5) and the ensuing discussion, it is an *optional gauge condition* for a *massless* gauge boson. So the relation  $G_\mu = (k_\tau k^\tau + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu = (k_\tau k^\tau - V)^{-1} J_\mu$  without mass, whenever it is used, *assumes* the gauge condition  $k_\tau G^\tau = 0$ . With this gauge condition this can also be written in terms of the kinetic momentum as  $G_\mu = (\pi_\tau \pi^\tau)^{-1} J_\mu = (k_\tau k^\tau - V)^{-1} J_\mu$  and it will not become singular even on-shell because  $|\pi_\sigma \pi^\sigma| = 0$  above, and not  $\pi_\sigma \pi^\sigma = 0$ , is the on-shell condition for a massless gauge boson in non-abelian theory in the chosen, not required,  $k_\tau G^\tau = 0$  gauge. This does introduce a degree of non-uniqueness into the inverse relationship for a massless gauge boson even with continuity. This is because the freedom to vary  $k_\tau G^\tau$  to non-zero states, unlike the residual gauge condition  $D_\nu D^\nu \theta = 0$  a.k.a.  $\partial_\nu \partial^\nu \theta - i\partial_\nu [G^\nu, \theta] = 0$  discussed after (6.5), *does* affect the form of the equations whenever one wishes to write them with the perturbation  $V$  or the kinetic scalar  $\pi_\sigma \pi^\sigma$ . As such, we will wish to find ways to avoid situations in which

$i\partial_\tau G^\tau = k_\tau G^\tau = 0$  is an optional gauge condition, in favor of always having it be a required relationship, which will lead us down the Coleman-Zee path of choosing massive solutions wherever they can physically justified.

## 8. The Recursive Nature of Non-Abelian Gauge Theory, and what it may Teach about Quantizing Yang-Mills Gauge Theory

Now we look for the first time at a very important *recursive* feature of non-abelian gauge theory. If we write the massive boson solution as  $G_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu$  from the second line of (6.27) and recognize that the perturbation  $V = -k_\tau G^\tau - G_\tau k^\tau - G_\tau G^\tau$  in (6.7) may also be written as  $V = -G_\tau k^\tau - G_\tau G^\tau$  because  $k_\tau G^\tau = 0$  is a *required* condition for a massive gauge boson, see (6.5) et seq., then a preferred way to write and use (6.27) will be the following:

$$G_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E} - V)^{-1} J_\mu. \quad (8.1)$$

Again, it bears emphasis, this uses the fact that  $k_\tau G^\tau = 0$  is required, but *only* for a massive, not massless, gauge boson. Now, although (8.1) appears on the surface to solve for  $G_\mu(J_\mu)$ , this is not a *closed* solution. Rather, it is really a *recursive* solution for  $G_\tau(G_\tau, J_\tau)$  which can be recursed into itself *ad infinitum*. Let us see exactly how this is done.

To do recursion, one generally needs two inputs: first, a recursive kernel; second, a terminal condition. A quintessential example is the recursive definition of the factorial function: The recursive kernel says that  $n! = n \times (n-1)!$ . The terminal condition says that  $0! = 1$ . We shall pursue a similar approach to understand  $G_\mu$  in (8.1).

To keep track of things, let us develop some notations. We shall generally use the double-nested symbol  $(( ))$  to denote a recursion. If we recurse  $G_\mu$  into itself  $n$  times, we shall denote this as  $G_\mu((( )))_n$ . If after  $n$  recursions we leave the perturbation  $V$  in the equation, then we shall write this as  $G_\mu(((V)))_n$ . If, however, after  $n$  recursive iterations we set  $V = 0$ , then we shall write this as  $G_\mu(((0)))_n \equiv G_\mu(((V=0)))_n$ . In this notation, this means that we write (8.1) as  $G_\mu(((V)))_0 = (k_\tau k^\tau - m^2 + i\mathcal{E} - V)^{-1} J_\mu$ . So, at the zeroth order of recursion, we simply set  $-V = G_\tau k^\tau + G_\tau G^\tau = 0$  in (8.1) which removes all of the terms containing  $G_\tau$  and reduces (8.1) to

$$G_\mu(((0)))_0 = (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu. \quad (8.2)$$

This is simply the abelian solution (6.28).

But now, let us perform the first order of recursion. Here, we substitute (8.1) back into itself one time and then set  $V = -G_\tau k^\tau - G_\tau G^\tau = 0$ . This exercise yields:

$$\begin{aligned}
 G_\mu((V))_1 &= (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu \\
 &= \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau k^\tau \\ + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J^\tau \end{array} \right)^{-1} J_\mu, \quad (8.3) \\
 \stackrel{V=0}{\Rightarrow} G_\mu((0))_1 &= \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} \\ + (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\tau k^\tau \\ + (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\tau (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J^\tau \end{array} \right)^{-1} J_\mu
 \end{aligned}$$

In leading order, this solution of course still contains (8.2) which is  $(k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu$ . But inside the overall inverse we now also have a new  $J_\tau k^\tau$  ( $J^1$ ) and a new  $J_\tau J^\tau$  ( $J^2$ ) term. This is now an expression strictly for  $G_\mu(J_\mu)$  not  $G_\tau(G_\tau, J_\tau)$ , because we have cut off the recursion at the first iteration by setting the perturbation  $V = -G_\tau k^\tau - G_\tau G^\tau = 0$  in the final line.

Now, let us go to the second order of recursion. Here, we start with the middle line of (8.3), do a second substitution of (8.1) to arrive at the second order recursion, and then cut things off by setting the perturbation  $V = 0$ . Now we obtain:

$$\begin{aligned}
 G_\mu((V))_2 &= \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau k^\tau \\ + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J^\tau \end{array} \right)^{-1} J_\mu \\
 &= \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} \\ + \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau k^\tau \\ + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J^\tau \end{array} \right)^{-1} J_\tau k^\tau \\ + \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau k^\tau \\ + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J^\tau \end{array} \right)^{-1} J_\tau \\ \times \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau k^\tau \\ + (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\tau (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J^\tau \end{array} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\mu, \quad (8.4)
 \end{aligned}$$

which, upon setting  $V = -G_\tau k^\tau - G_\tau G^\tau = 0$  reduces to:

$$G_\mu((0))_2 = \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\epsilon \\ + \left( k_\tau k^\tau - m^2 + i\epsilon + \frac{J_\tau k^\tau}{k_\tau k^\tau - m^2 + i\epsilon} + \frac{J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\epsilon)^2} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\epsilon + \frac{J_\tau k^\tau}{k_\tau k^\tau - m^2 + i\epsilon} + \frac{J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\epsilon)^2} \right)^{-1} J_\tau \\ \times \left( k_\tau k^\tau - m^2 + i\epsilon + \frac{J_\tau k^\tau}{k_\tau k^\tau - m^2 + i\epsilon} + \frac{J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\epsilon)^2} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\mu, \quad (8.5)$$

It will be appreciated this second recursive iteration contain terms in  $J$ ,  $J^2$ ,  $J^3$  and  $J^4$ . A third iteration would be expected to produce terms up to  $J^6$ , and in general,  $n$  iterations should produce terms over the entire gamut of  $J^1 \dots J^{2n}$ . As with (8.3),  $G_\mu((0))_2$  is an expression strictly for  $G_\mu(J_\mu)$  (really,  $G_\mu(J_\mu, k_\mu, m, \epsilon)$ ), not  $G_\tau(G_\tau, J_\tau)$  because we have cut off the recursion at the second iteration by setting the perturbation  $V = 0$ . But, having done two iterations rather than one, we have some new terms that we did not have at the first iteration. So in general the technique is to iterate as many times as one wishes, and then set  $V = 0$  to end the recursion. Each iteration will add new terms of yet higher order in  $J$ , and the result will be an expression for  $G_\mu(J_\mu)$  with terms of order  $J^1 \dots J^{2n}$ . And, of course, mathematically, theoretically, to obtain an *exact, closed* expression for  $G_\mu(J_\mu)$  not  $G_\tau(G_\tau, J_\tau)$ , one would iterate an *infinite* number of times and then set  $V = 0$ . But, of course, the real method we now need to pursue is not to iterate to infinity, but to figure out the pattern.

To discern the overall pattern, we do one more recursion to the  $n=3$  level by substituting (8.1) into the each and every  $G_\mu$  in (8.4). The expression for  $G_\mu((V))_3$  takes up over a page, and is not shown here. But upon setting  $V = 0$  to arrive at  $G_\mu((0))_3$ , this reduces to the still very large expression:

$$G_\mu((0))_3 = \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} \\ + \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau k^\tau \\ + \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau \\ \times \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau k^\tau \\ + \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau k^\tau \\ + \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau \\ \times \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J^\tau \\ \times \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau k^\tau \\ + \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau \\ \times \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J^\tau \\ \times \left( \begin{array}{l} k_\tau k^\tau - m^2 + i\mathcal{E} + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J_\tau \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} J^\tau \end{array} \right)^{-1} J^\tau \end{array} \right)^{-1} J^\tau \quad J_\mu \quad . \quad (8.6)$$

Even this is rather formidable, but now we have enough information to establish a definite pattern that can be generalized to any order of recursion.

Recognizing that the abelian boson propagator  $\pi$  may be denoted  $\pi^{-1} \equiv k_\tau k^\tau - m^2 + i\mathcal{E}$  up to a factor of  $i$ , we rewrite the abelian (8.2) simply as:

$$G_\mu((0))_0 = \pi J_\mu . \quad (8.7)$$

We also use this to write (8.3) as:

$$G_\mu((0))_1 = \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\mu, \quad (8.8)$$

and to write (8.5) as:

$$G_\mu((0))_2 = \left( \begin{array}{c} \pi^{-1} + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau k^\tau \\ + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J^\tau \end{array} \right)^{-1} J_\mu, \quad (8.9)$$

Now we see that  $\left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1}$  from (8.8) appears three times in (8.9). Given this, let us next define  $\Pi^{-1} \equiv \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau$ . This allows us to rewrite (8.8) as:

$$G_\mu((0))_1 = \Pi J_\mu, \quad (8.10)$$

and (8.9) as:

$$G_\mu((0))_2 = \left( \pi^{-1} + \Pi J_\tau k^\tau + \Pi J_\tau \Pi J^\tau \right)^{-1} J_\mu, \quad (8.11)$$

Now we see that (8.11) looks just like (8.8), except that each  $\pi$  which is in a term with  $J$  has advanced to a  $\Pi$ . So now let's go to that rather large (8.6) to nail down the pattern. Using  $\pi^{-1} \equiv k_\tau k^\tau - m^2 + i\mathcal{E}$  we first reduce (8.6) to:

$$G_\mu((0))_3 = \left( \begin{array}{c} \pi^{-1} + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau k^\tau \\ + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau k^\tau \left( \begin{array}{c} \pi^{-1} + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau k^\tau \\ + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J^\tau \end{array} \right)^{-1} J_\tau \left( \begin{array}{c} \pi^{-1} + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau k^\tau \\ + \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J_\tau \left( k\pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)^{-1} J^\tau \end{array} \right)^{-1} J^\tau \right)^{-1} J_\mu. \quad (8.12)$$

Now, using  $\Pi^{-1} \equiv \left( \pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau \right)$ , we may further reduce (8.12) to:

$$G_\mu((0))_3 = \left( \begin{array}{l} \pi^{-1} + (\pi^{-1} + \Pi J_\tau k^\tau + \Pi J_\tau \Pi J^\tau)^{-1} J_\tau k^\tau \\ + (\pi^{-1} + \Pi J_\tau k^\tau + \Pi J_\tau \Pi J^\tau)^{-1} J_\tau (\pi^{-1} + \Pi J_\tau k^\tau + \Pi J_\tau \Pi J^\tau)^{-1} J^\tau \end{array} \right)^{-1} J_\mu. \quad (8.13)$$

But now, we see that  $(\pi^{-1} + \Pi J_\tau k^\tau + \Pi J_\tau \Pi J^\tau)^{-1}$  from (8.11) appears three times in (8.13). So now, we define yet another  $\bar{\Pi}^{-1} \equiv \pi^{-1} + \Pi J_\tau k^\tau + \Pi J_\tau \Pi J^\tau$  and use this to rewrite (8.11) as:

$$G_\mu((0))_2 = \bar{\Pi} J_\mu \quad (8.14)$$

and (8.13) as:

$$G_\mu((0))_3 = (\pi^{-1} + \bar{\Pi} J_\tau k^\tau + \bar{\Pi} J_\tau \bar{\Pi} J^\tau)^{-1} J_\mu \equiv \bar{\bar{\Pi}} J_\mu. \quad (8.15)$$

This now has the form of (8.11) but with  $\Pi \rightarrow \bar{\Pi}$ . Seeing the pattern, we further define  $\bar{\bar{\Pi}}^{-1} = \pi^{-1} + \bar{\Pi} J_\tau k^\tau + \bar{\Pi} J_\tau \bar{\Pi} J^\tau$ . It is now inductively-clear that this is the pattern which will continue for higher recursive order. Now, let us systematize this pattern.

Pulling together the various results from (8.7), (8.10), (8.14), (8.15) and the various notational definitions made along the way, we have:

$$\begin{aligned} G_\mu((0))_0 &= \pi J_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu \\ G_\mu((0))_1 &= \Pi J_\mu = (\pi^{-1} + \pi J_\tau k^\tau + \pi J_\tau \pi J^\tau)^{-1} J_\mu \\ G_\mu((0))_2 &= \bar{\Pi} J_\mu = (\pi^{-1} + \Pi J_\tau k^\tau + \Pi J_\tau \Pi J^\tau)^{-1} J_\mu \\ G_\mu((0))_3 &= \bar{\bar{\Pi}} J_\mu = (\pi^{-1} + \bar{\Pi} J_\tau k^\tau + \bar{\Pi} J_\tau \bar{\Pi} J^\tau)^{-1} J_\mu \end{aligned} \quad (8.16)$$

Of course, for notational economy we do not want to have to keep adding bars or primes or any other qualifier to each of the “propagators.” So let us denote each “propagator” with a subscript that simply declares its recursive order, thus,  $\pi \equiv \pi_0$ ,  $\Pi \equiv \pi_1$ ,  $\bar{\Pi} \equiv \pi_2$ ,  $\bar{\bar{\Pi}} \equiv \pi_3$ , etcetera. Then, we can inductively compact (8.16) into a fully recursive solution just like the recursive kernel  $n! = n \times (n-1)!$  and the terminal condition  $0! = 1$  for factorial. Specifically, starting with  $G_\mu((0))_3$  and working down, the recursive kernel and the terminal condition are induced to be:

$$\begin{cases} G_\mu((0))_n = \pi_n J_\mu = (\pi_0^{-1} + \pi_{n-1} J_\tau k^\tau + \pi_{n-1} J_\tau \pi_{n-1} J^\tau)^{-1} J_\mu \\ G_\mu((0))_0 = \pi_0 J_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu \end{cases} \quad (8.17)$$

If we wish to separate the propagators from the gauge fields in (8.17), the recursive kernel and the abelian terminal condition may be written also as:

$$\begin{cases} \pi_n = (\pi_0^{-1} + \pi_{n-1} J_\tau k^\tau + \pi_{n-1} J_\tau \pi_{n-1} J^\tau)^{-1} \\ \pi_0 = (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} \end{cases} \quad (8.18)$$

So with all of this in mind, let us now return to (8.1) which is an expression for  $G(G, J)$ . But at any recursive order, we now know how to turn this into  $G(J)$  without any gauge field residual: Just zero out the perturbation. Of course, nature will not stop at some arbitrary order and then zero out perturbations. She will recurse ad infinitum and the physics we observe should be for an infinite-order recursion. So in the natural world, we expect that the observed non-linear solution for  $G(J)$  will be the one which recurses to infinity, thus contains terms up to infinite order in  $J$  and in  $k$  (really,  $2 \times \infty$  in  $J$ ), and then sets the perturbation  $V$  to zero. That is, we expect that nature's *physical* solution (8.1) will be  $G_\mu \equiv G_\mu((0))_\infty$ , or in detail:

$$\begin{aligned} G_\mu &= (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E} - V)^{-1} J_\mu \\ &= -(D_\tau D^\tau + m^2 - i\mathcal{E})^{-1} J_\mu = -(\partial_\tau \partial^\tau + m^2 - i\mathcal{E} - iG_\tau \partial^\tau - G_\tau G^\tau)^{-1} J_\mu \quad , \\ &\equiv G_\mu((0))_\infty = \pi_\infty J_\mu = (\pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau)^{-1} J_\mu \end{aligned} \quad (8.19)$$

Above, for future use in doing an analytical path integral in section 11, we have also included the earlier solution (6.27) to the field equation  $-J^\nu = (g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu) G_\sigma$  of (5.15) with a Proca massive boson and  $D^\sigma D^\nu \equiv \partial^\sigma \partial^\nu - iG^\sigma \partial^\nu - 2G^\sigma G^\nu + G^\nu G^\sigma$  from (5.16) and  $D_\tau D^\tau = \partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau$  from (5.17). We especially wish to take note of the correspondence  $\pi_\infty \leftrightarrow -(D_\tau D^\tau + m^2 - i\mathcal{E})^{-1}$ . And we also note the embedded correspondences  $G_\tau k^\tau \leftrightarrow \pi_{\infty-1} J_\tau k^\tau$  and  $G_\tau G^\tau \leftrightarrow \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau$ , which both contain the elemental correspondence  $G_\tau \leftrightarrow \pi_{\infty-1} J_\tau \cong \pi_\infty J_\tau$ .

Very importantly, written as:

$$\begin{cases} G_\mu = G_\mu((0))_\infty = \pi_\infty J_\mu = (\pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau)^{-1} J_\mu \\ \pi_0 = (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} \end{cases} \quad (8.20)$$

we have an expression for  $G(J, k, m, \mathcal{E})$  rather than  $G(G, J, k, m, \mathcal{E})$ , with all gauge fields removed. What is left of the gauge field is its momentum vector  $k$ , interacting with the current

density in the terms  $J_\tau k^\tau$  and contracted with itself in the linear terms  $k_\tau k^\tau$ , as well as its mass  $m$  and its  $\mathcal{E}$  which as an imaginary mass-type term, is related to lifetime, see, e.g., [14] page 150.

Why is this all so very important? First, it points out that although (8.1) appears on the surface to solve for  $G_\mu(J_\mu)$ , this is not a *closed* solution. Rather, it is really a *recursive* solution for  $G_\tau(G_\tau, J_\tau) = G_\tau(G_\tau(G_\tau, J_\tau), J_\tau) = G_\tau(G_\tau(G_\tau(G_\tau, J_\tau), J_\tau), J_\tau) \dots$  which can be iteratively recursed *ad infinitum*, but at any order can be cut off and turned into  $G_\mu(J_\mu)$  not  $G_\tau(G_\tau, J_\tau)$  by setting  $V = 0$ , i.e., by ceasing any further perturbations. This makes the non-linear nature of Yang-Mills theory very apparent from a different view than  $F^k_{\mu\nu} = \partial_{[\mu} G^k_{\nu]} + f^{ijk} G^i_\mu G^j_\nu$  of (1.9) or  $F = dG + G \wedge G$  of (1.11) which are the usual expressions used to highlight the non-linear nature of Yang-Mills theory.

Secondly, and of very deep importance, this recursion may well point the way toward being able to *analytically and exactly* quantize Yang-Mills theory. Specifically, we now return to Jaffe and Witten who on page 7 of [6], state:

“Since the inception of quantum field theory, two central methods have emerged to show the existence of quantum fields on non-compact configuration space (such as Minkowski space). These known methods are (i) Find an exact solution in closed form; (ii) Solve a sequence of approximate problems, and establish convergence of these solutions to the desired limit.”

The foregoing suggests a third method which is really a hybrid of (i) and (ii): find an exact *recursive kernel* in closed form (which is  $G_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu$ ) and then expand that kernel in successive iterations to see how the recursion behaves in the limit of infinite recursive nesting. That is exactly what we have done in (8.17), (8.18) and (8.20).

Specifically, regarding  $G_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu$  as the zero<sup>th</sup> order solution for  $G_\tau(G_\tau, J_\tau)$ , with each iteration of  $G_\tau(G_\tau, J_\tau)$  from the  $n^{\text{th}}$  to the  $(n+1)^{\text{th}}$  recursive order we are effectively replacing all gauge fields  $G_\tau$  at the  $n^{\text{th}}$  order with current densities  $J_\tau$  up to the  $2(n+1)^{\text{th}}$  order, and at the same time injecting a new set of gauge fields  $G_\tau$  at the  $(n+1)^{\text{th}}$  order. But at any time we can stop introducing new gauge fields by simply setting the perturbation to zero. So at each order, whenever we decide to do so, we may effectively strip out the gauge fields and replace them with current densities. *This means that in the limit  $n \rightarrow \infty$  we may effectively replace all gauge fields with current densities by stopping perturbation at  $n = \infty$ .*

Very similarly, when we take a path integral  $Z = \int DG \exp iS(G) = \mathcal{C} \exp iW(J)$ , because  $G$  is the integration variable, we effectively strip off the  $G$  and obtain a quantum amplitude  $W(J)$  expressed in terms of the current density  $J$ . So the infinite recursion has the same effect as a path integral in terms of trading  $G$  for  $J$ . But as pointed out at the start of

section 6, the mathematical exercise of analytically calculating a path integral revolves around clever extrapolations of the Gaussian integral  $\int dx \exp(-\frac{1}{2}Ax^2 - Jx) = (-2\pi/A)^{-1/2} \exp(J^2/2A)$  into  $Z = \int DG \exp iS(G) = \mathcal{C} \exp iW(J)$ . The calculation impediment we run into is that  $\int dx \exp(-\frac{1}{2}Ax^2 - Jx)$  is integrable because it is quadratic, but becomes quite intractable once this integral involves a polynomial of  $x^3$  and higher order, which is exactly what happens in Yang-Mills theory and indeed, any non-linear interaction theory. Why is this intractable? Because nobody knows how to calculate such integrals exactly and analytically!

The usual and best workaround is to employ what Zee [11] in Appendix A refers to as the “central identity of quantum field theory”:

$$\int D\phi \exp(-\frac{1}{2}\phi \cdot K \cdot \phi - V(\phi) + J \cdot \phi) = \mathcal{C} \exp(-V(\delta/\delta J)) \exp(\frac{1}{2}J \cdot K^{-1} \cdot J). \quad (8.25)$$

This method uses the functional variation  $G_\mu \rightarrow \delta/\delta J^\mu$  to remove all terms which are polynomial (greater than second order) in the gauge field  $G_\mu$ , and replace them with terms  $\delta/\delta J^\mu$  that contain only the current density. This allows  $\exp(V(\delta/\delta J))$  to be removed from inside the integral, so that the only terms left inside the integral are quadratic in  $G_\mu$ . Then, the integral is performed to obtain  $\exp(\frac{1}{2}J \cdot K^{-1} \cdot J)$ , and the operation of  $\exp(-V(\delta/\delta J))$  on  $\exp(\frac{1}{2}J \cdot K^{-1} \cdot J)$  is thereafter used to extract order-by-order terms in the quantum amplitude to reveal various Green’s and Wick’s coefficients in this amplitude.

The very important point is that an infinitely-iterative application of the recursive kernel  $G_\mu = (k_\tau k^\tau - m^2 + i\varepsilon + G_\tau k^\tau + G_\tau G^\tau)^{-1} J_\mu$  of (8.1) serves a purpose totally analogous to  $G_\mu \rightarrow \delta/\delta J^\mu$ . But  $G_\mu \rightarrow (\pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau)^{-1} J_\mu$  from (8.20) is now the replacement we use in lieu of  $G_\mu \rightarrow \delta/\delta J^\mu$ . In the limit of infinite recursion, this will allow us in section 11 to do an analytically-exact calculation of the path integral by turning  $G_\mu$  into  $J_\mu$  on an order-by-order basis such that in the limit of infinite nesting, all of the gauge fields have been replaced by current densities which then pose no problem to carrying out a Gaussian integration which is simply of quadratic form  $\int dx \exp(-\frac{1}{2}Ax^2 - Jx)$  in the gauge fields.

Now, let us return to the Yang-Mills monopoles  $\oint\!\!\!\oint F = -i \iiint dGG \neq 0$  of (3.3) and (5.9), and particularly the identity  $P' = d[G, G] = dGG$  of (2.11) upon which this is based. It will be our goal to use one or more of the inverses  $G(J)$  that we have developed in sections 6 through 8 to replace each  $G$  in this monopole with its source current  $J$ , then to replace each  $J$  with fermions via  $J^\mu = \bar{\psi} \gamma^\mu \psi$ , then to apply exclusion to the fermions, and then to show that this faux

magnetic charge  $P' = d[G, G] = dGG$  – at least in the classical theory – has the exact same chromodynamic symmetries as a baryon.

## 9. Populating the Composite Yang-Mills Magnetic Monopoles with Chromodynamically-Colored Fermions

Let us start the present discussion with the identity  $d[G, G] = dGG$  uncovered in (2.11), which we combine with (3.3) and then expand into tensor component expressions (see also (2.8) and (2.9)) while also including the faux magnetic charge  $P' = -idGG = -id[G, G]$ , as such:

$$\begin{aligned}
 \oint\!\!\!\oint F &= \iiint P' = -i \iiint dGG = -i \iiint d[G, G] = -i \oint\!\!\!\oint [G, G] \\
 &= \oint\!\!\!\oint \frac{1}{2!} F_{\mu\nu} dx^\mu \wedge dx^\nu = \iiint \frac{1}{3!} P'_{\sigma\mu\nu} dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \iiint \frac{1}{3!} (\partial_{[\sigma} G_{\mu]} G_\nu + \partial_{[\mu} G_{\nu]} G_\sigma + \partial_{[\nu} G_{\sigma]} G_\mu) dx^\sigma \wedge dx^\mu \wedge dx^\nu \quad . \quad (9.1) \\
 &= -i \iiint \frac{1}{3!} (\partial_\sigma [G_\mu, G_\nu] + \partial_\mu [G_\nu, G_\sigma] + \partial_\nu [G_\sigma, G_\mu]) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \oint\!\!\!\oint \frac{1}{2!} [G_\mu, G_\nu] dx^\mu \wedge dx^\nu \neq 0
 \end{aligned}$$

Let us now further develop (9.1) using the inverses reviewed in sections 6 and 7.

For a *massless* gauge boson in non-abelian gauge theory, we found that the relationship  $G_\mu = -(D_\tau D^\tau)^{-1} J_\mu$  is the *unique* solution to the field equation  $-J^\nu = (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) G_\sigma$  of (5.15) with  $D^\sigma D^\nu$  and  $D_\tau D^\tau$  given by (5.16) and (5.17), in the circumstance where the current density is conserved according to  $D_\nu J^\nu = 0$  as found in (5.20), because this continuity contextually fixes the gauge to the Feynman / continuity gauge  $\xi = 1$ , see (7.22) and (7.23). We further found in (7.24) that by setting the mass  $m = 0$  in (6.15) for a *massive* gauge boson, we arrive at exactly the same solution  $G_\mu = -(D_\tau D^\tau)^{-1} J_\mu$ . And, we found that in (7.25), in order to include the  $+i\mathcal{E}$  prescription in the non-Abelian theory, we need simply migrate  $D_\tau D^\tau \Rightarrow D_\tau D^\tau - i\mathcal{E}$ . So as shown in (6.27), the non-abelian solution for a massive gauge boson is  $G_\mu = (\pi_\tau \pi^\tau - m + i\mathcal{E})^{-1} J_\mu$ , while as shown in (6.28), the corresponding abelian solution for a massive gauge boson is  $G_{A\mu} = (k_\tau k^\tau - m + i\mathcal{E})^{-1} J_\mu$ . So again, we are reminded that the non-abelian solution is identical in form to the abelian relation for a massive gauge boson, but for the replacement of the canonical  $k_\tau k^\tau$  with the kinetic  $\pi_\tau \pi^\tau$  momentum scalar, which replacement can be made in the massive theory because  $\partial_\tau G^\tau = 0$  is a requirement, and which replacement *may* be made in the massless theory if one chooses  $\partial_\tau G^\tau = 0$  although one does not have to. So the massive solution is more unique in this way than the massless solution.

Now we wish to replace each  $G_\mu$  in (9.1) with its unique continuity solution, i.e., with the gauge contextually fixed to  $\xi=1$  because of requiring continuity, either  $\partial_\sigma J^\sigma = 0$  for abelian theory, or  $D_\sigma J^\sigma = 0$  for non-abelian theory, and to have the result be as uniquely-determined as possible. Based on the development in sections 6 and 7, we have four choices of solution: a) the massive non-abelian solution  $G_\mu = (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu$  of (6.27); b) the massive abelian solution  $G_\mu = (k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu$  of (6.28) which is simply solution (a) with  $V = 0$ ; c) the massless non-abelian solution  $G_\mu = (-V + k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu$  of (7.26) in the  $k_\tau G^\tau = 0$  gauge which is simply solution (a) with  $m=0$ ; and d) the massless abelian solution  $G_\mu = (k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu$  of (7.27) which is simply solution (b) with  $m=0$  or solution (c) with  $V = 0$ . Because one can follow Coleman-Zee as shown in sections 6 and 7 to include a massive boson solution  $m \neq 0$  and then arrive at the massless solution simply by setting  $m = 0$ , and because the massless solution is uniquely forced to the  $\xi=1$  gauge to preserve continuity and thus we arrive at the exact same point whether we start with a massive or a massless solution, it makes more sense to first include the mass  $m \neq 0$ . This is a more general approach, and as we have seen, this mass can always be zeroed out later at the appropriate time, whereby the requirement for continuity will contextually fix the gauge into the Feynman / continuity gauge  $\xi=1$ .

But there is also another more specific reason for starting with  $m \neq 0$  beyond its generality, and that has specifically to do with the *uniqueness* of the massive solutions. Even though the continuity relationships  $D_\sigma J^\sigma = 0$  and  $\partial_\sigma J^\sigma = 0$  do zero out the terms containing the gauge number  $\xi$  from the massless bosons and contextually fix the gauge to  $\xi=1$ , see (7.22) and (7.23), the condition  $k_\tau G^\tau = 0$  is *required* for a massive boson but is simply a covariant choice of gauge condition for a massless gauge boson. So if we start with massive solution (a) which is  $G_\mu = (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu$ , we know that the gauge condition  $k_\tau G^\tau = 0$  *must* be in place because that is a requirement to ensure continuity for the massive solution, and that the perturbation  $V$  appears in simple form in this solution precisely because  $k_\tau G^\tau = 0$ , see (6.6) and (6.7), and (6.24). On the other hand, if we start with massless solution (c) which is  $G_\mu = (-V + k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu$ , we know even though the gauge number is contextually fixed to  $\xi=1$  by continuity, again, (7.22) and (7.23), that  $k_\tau G^\tau = 0$  is merely a *choice* of gauge, and that the manner in which the perturbation  $V$  appears in  $G_\mu = (-V + k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu$  is itself dependent upon this choice of  $k_\tau G^\tau = 0$  gauge. If we choose  $k_\tau G^\tau \neq 0$ , then  $G_\mu = (-V + k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu$  will have to include this  $k_\tau G^\tau \neq 0$ , and so its very form will change. So solution (a) is *uniquely* determined in all respects up to the covariant gauge condition  $D_\nu D^\nu \theta = 0$  a.k.a.  $\partial_\nu \partial^\nu \theta - i\partial_\nu [G^\nu, \theta] = 0$  developed after (6.5), while solution (c) is contextually fixed to the  $\xi=1$

gauge by continuity but  $D_\nu G^\nu = \partial_\nu G^\nu$  remains a free scalar object which is *not required to be zero* and so renders the massless solutions weaker, i.e., less-unique than the massive solutions. Again, this solution will only be  $G_\mu = (-V + k_\tau k^\tau + i\mathcal{E})^{-1} J_\mu$  if we choose  $k_\tau G^\tau = 0$  and *will change in form* in the event we choose a  $k_\tau G^\tau \neq 0$  whereby we will explicitly have to include a  $k_\tau G^\tau$  term.

So to preserve generality and maximize uniqueness, we shall now use solution (a), namely  $G_\mu = (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu$  of (6.27) to replace each occurrence of  $G_\mu$  with  $(-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu$  in (9.1). This has a *required* gauge relation  $k_\tau G^\tau = 0$ , and a selected gauge condition  $D_\nu D^\nu \theta = 0$  which does not change the form of the solution in the event one chooses  $D_\nu D^\nu \theta \neq 0$ , see (6.5) and thereafter. As noted, this becomes solution (b) if we set  $V=0$ , this becomes solution (c) if we set  $m=0$  and choose  $k_\nu G^\nu = 0$  as a gauge condition, and it becomes solution (d) if we set  $V=0$  and  $m=0$  and again choose  $k_\nu G^\nu = 0$ . Thus, inserting  $G_\mu = (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu$  into (9.1) we obtain:

$$\begin{aligned}
 \oint\!\!\!\oint F &= \iiint P' = -i \iiint dGG = -i \iiint d[G, G] = -i \oint\!\!\!\oint [G, G] \\
 &= \oint\!\!\!\oint \frac{1}{2!} F_{\mu\nu} dx^\mu \wedge dx^\nu = \iiint \frac{1}{3!} P'_{\sigma\mu\nu} dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \iiint \frac{1}{3!} \left( \begin{array}{l} \partial_{[\sigma} \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_{\mu]} \right) \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\nu \right) \\ + \partial_{[\mu} \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_{\nu]} \right) \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\sigma \right) \\ + \partial_{[\nu} \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_{\sigma]} \right) \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu \right) \end{array} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \iiint \frac{1}{3!} \left( \begin{array}{l} \partial_\sigma \left[ \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu, \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\nu \right) \right] \\ + \partial_\mu \left[ \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\nu, \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\sigma \right) \right] \\ + \partial_\nu \left[ \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\sigma, \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu \right) \right] \end{array} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \oint\!\!\!\oint \frac{1}{2!} \left[ \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\mu, \left( (-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1} J_\nu \right) \right) dx^\mu \wedge dx^\nu \neq 0
 \end{aligned} \tag{9.2}$$

This is the complete expression for the *net-flux*  $\oint\!\!\!\oint F$  of the non-abelian magnetic field over a *closed* two-dimensional surface, and as we just learned in section 8, it is highly nonlinear, and indeed, contains an infinite recursion of  $G_\tau(G_\tau, J_\tau)$  which is ultimately made into  $G_\tau(J_\tau)$  by

recursing to infinity then setting  $V=0$  as shown in (8.20). Indeed, we could also have employed  $G_\mu = \pi_\infty J_\mu$  in from (8.20) in (9.1) to alternatively and equivalently obtain:

$$\begin{aligned}
 \oint\!\!\!\oint F((0))_\infty &= \iiint P'((0))_\infty = -i \iiint dGG((0))_\infty = -i \iiint d[G, G]((0))_\infty = -i \oint\!\!\!\oint [G, G]((0))_\infty \\
 &= \oint\!\!\!\oint \frac{1}{2!} F_{\mu\nu}((0))_\infty dx^\mu \wedge dx^\nu = \iiint \frac{1}{3!} P'_{\sigma\mu\nu}((0))_\infty dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \iiint \frac{1}{3!} (\partial_{[\sigma} \pi_\infty J_{\mu]} \pi_\infty J_\nu + \partial_{[\mu} \pi_\infty J_{\nu]} \pi_\infty J_\sigma + \partial_{[\nu} \pi_\infty J_{\sigma]} \pi_\infty J_\mu) dx^\sigma \wedge dx^\mu \wedge dx^\nu \quad .(9.3) \\
 &= -i \iiint \frac{1}{3!} (\partial_\sigma [\pi_\infty J_\mu, \pi_\infty J_\nu] + \partial_\mu [\pi_\infty J_\nu, \pi_\infty J_\sigma] + \partial_\nu [\pi_\infty J_\sigma, \pi_\infty J_\mu]) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \oint\!\!\!\oint \frac{1}{2!} [\pi_\infty J_\mu, \pi_\infty J_\nu] dx^\mu \wedge dx^\nu \neq 0
 \end{aligned}$$

We will eventually return at the end of section 10 to discuss (9.3) above in more detail. But at the moment, (9.2) is in a form that better facilitates understanding the connection between  $P'$  and a baryon density, because we can set  $V=0$  at any order  $n$  of recursion we choose and thereby obtain  $\oint\!\!\!\oint F((0))_n$ .

Before trying to tackle the highly-nonlinear (9.2), see the section 8 discussion of recursion that is inherent in the above because (9.2) contains the perturbation  $-V = k_\tau G^\tau + G_\tau k^\tau + G_\tau G^\tau$  of (6.7) throughout, let us now do what is commonly done in many other situations in particle physics: consider the zero-perturbation limit by setting  $V=0$  throughout (9.2) right away. That is, we obtain and explore  $\oint\!\!\!\oint F((0))_0$ . This will of course remove the non-linear physics occurring in (9.2), but it will readily reveal why these faux magnetic monopoles have the symmetries that one expects to see in a baryon. Moreover, surprisingly enough, when we use  $\oint\!\!\!\oint F((0))_0$  to calculate the energies associated with the flux equation  $\iiint P' = -i \oint\!\!\!\oint [G, G]$  after some development of the baryon into protons and neutrons, we find a surprising, very tight concurrence with the binding energies that are experimentally-observed in nuclear physics, which suggests that the nuclear binding energies are in fact expressive of the behaviors of (9.2) in this zero-perturbation limit, i.e., in the linear / abelian approximation (see [15] sections 6 through 12 and all of [16]).

Once we set  $V=0$  in each of the  $(-V + k_\tau k^\tau - m^2 + i\mathcal{E})^{-1}$  in (9.2), these each become the ordinary denominator  $1/(k_\tau k^\tau - m^2 + i\mathcal{E})$ , because as developed in (6.26), it is  $-V_{AB} = k_\sigma G^\sigma_{AB} + G_{\sigma AB} k^\sigma + (G_\sigma G^\sigma)_{AB}$  which is responsible for our having to write (9.2) with inverses rather than denominators. Thus, setting  $V=0$  and rearranging somewhat, (9.2) for  $\oint\!\!\!\oint F((0))_0$  and  $P'((0))_0$  becomes:

$$\begin{aligned}
 \oint\!\!\!\oint F((0))_0 &= \iiint P'((0))_0 = -i \iiint dGG((0))_0 = -i \iiint d[G, G]((0))_0 = -i \oint\!\!\!\oint [G, G]((0))_0 \\
 &= \oint\!\!\!\oint \frac{1}{2!} F_{\mu\nu}((0))_0 dx^\mu \wedge dx^\nu = \iiint \frac{1}{3!} P'_{\sigma\mu\nu}((0))_0 dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \iiint \frac{1}{3!} \left( \frac{\partial_{[\sigma} J_{\mu]} J_\nu}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial_{[\mu} J_\nu] J_\sigma}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial_{[\nu} J_\sigma] J_\mu}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu. \quad (9.4) \\
 &= -i \iiint \frac{1}{3!} \left( \frac{\partial_\sigma [J_\mu, J_\nu]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial_\mu [J_\nu, J_\sigma]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial_\nu [J_\sigma, J_\mu]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \oint\!\!\!\oint \frac{1}{2!} \frac{[J_\mu, J_\nu]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} dx^\mu \wedge dx^\nu \neq 0
 \end{aligned}$$

Although the complete non-linear physics of  $\oint\!\!\!\oint F \neq 0$  is described by (9.2) and alternatively (9.3), the simplified (9.4) enables us to reveal certain key symmetries for  $\oint\!\!\!\oint F \neq 0$  which will support the view that the faux magnetic monopole density  $P'$  is in fact a baryon density, which symmetries carry over fully to the more-complete, highly-perturbed (9.2), (9.3). We shall refer to (9.4) as the ‘‘ground state’’ monopole equation, because the perturbations are zeroed out immediately before any levels of recursion are carried out.

Of particular interest, let us now focus on the  $= -i \iiint d[G, G]((0))_0$  term in (9.4), which we restructure into:

$$\begin{aligned}
 \oint\!\!\!\oint F((0))_0 &= \oint\!\!\!\oint \frac{1}{2!} F_{\mu\nu}(0) dx^\mu \wedge dx^\nu \\
 &= \iiint P'((0))_0 = \iiint \frac{1}{3!} P'_{\sigma\mu\nu}((0))_0 dx^\sigma \wedge dx^\mu \wedge dx^\nu = -i \iiint d[G, G]((0))_0. \quad (9.5) \\
 &= -i \iiint \frac{1}{3!} \left( \frac{\partial_\sigma [J_\mu, J_\nu]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial_\mu [J_\nu, J_\sigma]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial_\nu [J_\sigma, J_\mu]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu
 \end{aligned}$$

From this we extract the faux magnetic monopole density raised to contravariant indexes:

$$P'^{\sigma\mu\nu}((0))_0 = -i \left( \frac{\partial^\sigma [J^\mu, J^\nu]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial^\mu [J^\nu, J^\sigma]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} + \frac{\partial^\nu [J^\sigma, J^\mu]}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right). \quad (9.6)$$

Now we take the crucial step of developing the current sources densities  $J^\mu$  in terms of the underlying fermion wavefunctions  $\psi$  which arise in Dirac theory. Specifically, in abelian gauge theory, Dirac’s equation says that  $(i\gamma^\mu \partial_\mu - m)\psi = 0$ . For the adjoint spinor  $\bar{\psi} = \psi^\dagger \gamma^0$  the

field equation is  $i\partial_\mu \bar{\psi}\gamma^\mu + m\bar{\psi} = 0$ . Adding yields  $\partial_\mu (\bar{\psi}\gamma^\mu\psi) = 0$  as is well known. And because the conserved current is expressed by  $\partial_\mu J^\mu = 0$ , we identify the current density with  $J^\mu = \bar{\psi}\gamma^\mu\psi$ , where each Dirac wavefunction  $\psi$  in a U(1) theory is of course a four-component column vector.

In non-abelian gauge theory, for the compact simple gauge group SU(N) (or for the product group SU(N)xU(1) with a U(1) factor that is required for magnetic monopoles to be topological stability as will be reviewed in section 10), the generalized wavefunction  $\Psi = \Psi_A$ ,  $A = 1 \dots N$  is an Nx4 column vector of 4-component Dirac wavefunctions  $\psi$ . This non-abelian wavefunction  $\Psi$  may then subsist in any one of  $N$  distinct eigenstates. For example, for the SU(3)<sub>C</sub> group of chromodynamic strong interactions, the three (3) eigenstates are generally denoted (R)ed, (G)reen, (B)lue, and these distinct eigenstates are used to enable a baryon containing three quarks to satisfy the Fermi-Dirac-Pauli Exclusion Principle. Explicitly defined, using the SU(N) group generators  $\lambda^i = \lambda_{AB}^i$ ,  $i = 1 \dots N^2 - 1$ , the current density generalizes to  $J^\mu = \lambda_{AB}^i J^{i\mu} = \lambda_{AB}^i \bar{\Psi}_C \lambda_{CD}^i \gamma^\mu \Psi_D \equiv \bar{\Psi} \gamma^\mu \Psi$ , with Yang-Mills adjoint  $i$  and fundamental  $A, B, C, D$  indexes explicitly shown for illustration, and where as already stated  $\Psi = \Psi_A$  is an N-component column vector of N fermion eigenstates. As has been reviewed at length earlier starting at (5.20), this current density satisfies the continuity relationship  $D_\nu J^\nu = 0$ . For SU(N)xU(1), we may for simplicity use  $\lambda_{AB}^i$  with  $i = 0 \dots N^2 - 1$ , where we denote the U(1) generator as  $\lambda_{AB}^0$  with the “0” index. If we suppress the  $A, B, C, D$  indexes, then  $J^\mu = \lambda^i J^{i\mu} = \lambda^i \bar{\Psi} \lambda^i \gamma^\mu \Psi \equiv \bar{\Psi} \gamma^\mu \Psi$ .

So now, into (9.6), we first substitute  $J^\mu = \lambda^i J^{i\mu}$ , then  $J^{i\mu} = \bar{\Psi} \lambda^i \gamma^\mu \Psi$ , and then use  $[\lambda^i, \lambda^j] (\bar{\Psi} \lambda^i \gamma^\mu \Psi) (\bar{\Psi} \lambda^j \gamma^\nu \Psi) = [\bar{\Psi} \gamma^\mu \Psi, \bar{\Psi} \gamma^\nu \Psi]$  (just a variant of  $[\lambda^i, \lambda^j] A^{i\mu} B^{j\nu} = [A^\mu, B^\nu]$ ) in (9.6) to “populate” the faux Yang-Mills magnetic monopole with fermions. The result is:

$$\begin{aligned}
 P'^{\sigma\mu\nu}((0))_0 &= -i \left( \frac{\partial^\sigma([\lambda^i, \lambda^j] J^{i\mu} J^{j\nu})}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} + \frac{\partial^\mu([\lambda^i, \lambda^j] J^{i\nu} J^{j\sigma})}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} + \frac{\partial^\nu([\lambda^i, \lambda^j] J^{i\sigma} J^{j\mu})}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} \right) \\
 &= -i \left( \frac{\partial^\sigma([\lambda^i, \lambda^j] (\bar{\Psi} \lambda^i \gamma^\mu \Psi) (\bar{\Psi} \lambda^j \gamma^\nu \Psi))}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} \right. \\
 &\quad + \frac{\partial^\mu([\lambda^i, \lambda^j] (\bar{\Psi} \lambda^i \gamma^\nu \Psi) (\bar{\Psi} \lambda^j \gamma^\sigma \Psi))}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} \\
 &\quad \left. + \frac{\partial^\nu([\lambda^i, \lambda^j] (\bar{\Psi} \lambda^i \gamma^\sigma \Psi) (\bar{\Psi} \lambda^j \gamma^\mu \Psi))}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} \right) \quad . \quad (9.7) \\
 &= -i \left( \frac{\partial^\sigma [\bar{\Psi} \gamma^\mu \Psi, \bar{\Psi} \gamma^\nu \Psi]}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} + \frac{\partial^\mu [\bar{\Psi} \gamma^\nu \Psi, \bar{\Psi} \gamma^\sigma \Psi]}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} + \frac{\partial^\nu [\bar{\Psi} \gamma^\sigma \Psi, \bar{\Psi} \gamma^\mu \Psi]}{(k_\tau k^\tau - m^2 + i\varepsilon)^2} \right)
 \end{aligned}$$

We could just as readily have just inserted  $J^\mu = \bar{\Psi} \gamma^\mu \Psi$  into (9.6) to arrive directly at the bottom line of (9.7), but it is helpful to see the intermediate calculations which explicitly contain the group generators. Given that  $\oint\!\!\!\oint F = \iiint P'$ , and referring back to the discussion at the end of section 3, we now see for the first time the manner in which  $\oint\!\!\!\oint F(G(J(\psi)))$ , that is, the manner in which the *composite* faux magnetic monopole  $\oint\!\!\!\oint F$  arising from the faux magnetic source  $P' = -idGG = -id[G, G]$  does indeed contain fermion wavefunctions  $\Psi$ . Now, we shall show how these fermion wavefunction in fact possess all of the key symmetries required to qualify them as colored quarks, and how  $P'^{\sigma\mu\nu}$  possesses all of the key symmetries of a baryon.

The first thing we observe is that  $P'^{\sigma\mu\nu}((0))_0$  contains three additive terms. And, as discussed moments ago, for SU(N) or for SU(N)xU(1), each  $\Psi = \Psi_A$  is an N-component column vector of 4-component Dirac wavefunctions  $\psi$  which may subsist in any one of  $N$  distinct eigenstates. So if we regard  $P'^{\sigma\mu\nu}((0))_0$  as a composite system of more than one fermion, then each fermion in this system must be placed into a distinct eigenstate in order to satisfy the Fermion Exclusion Principle. The three additive terms in (9.7) advise us that there are a total of three such fermion eigenstates which constitute  $P'^{\sigma\mu\nu}((0))_0$ , and so we label these eigenstates among the three additive terms as  $\Psi_1, \Psi_2, \Psi_3$ . With this we now rewrite (9.7), including a restructuring  $[\bar{\Psi} \gamma^\mu \Psi, \bar{\Psi} \gamma^\nu \Psi] = \bar{\Psi} \gamma^\mu \Psi \bar{\Psi} \gamma^\nu \Psi$  of the commutators in the bottom line below, as:

$$\begin{aligned}
 & P'^{\sigma\mu\nu} \left( (0) \right)_0 \\
 &= -i \left( \frac{\partial^\sigma \left[ \overline{\Psi}_1 \gamma^\mu \Psi_1, \overline{\Psi}_1 \gamma^\nu \Psi_1 \right]}{\left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^2} + \frac{\partial^\mu \left[ \overline{\Psi}_2 \gamma^\nu \Psi_2, \overline{\Psi}_2 \gamma^\sigma \Psi_2 \right]}{\left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^2} + \frac{\partial^\nu \left[ \overline{\Psi}_3 \gamma^\sigma \Psi_3, \overline{\Psi}_3 \gamma^\mu \Psi_3 \right]}{\left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^2} \right). \quad (9.8) \\
 &= -i \left( \frac{\partial^\sigma \left( \overline{\Psi}_1 \gamma^{\mu\nu} \Psi_1 \overline{\Psi}_1 \gamma^{\nu\mu} \Psi_1 \right)}{\left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^2} + \frac{\partial^\mu \left( \overline{\Psi}_2 \gamma^{\nu\sigma} \Psi_2 \overline{\Psi}_2 \gamma^{\sigma\nu} \Psi_2 \right)}{\left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^2} + \frac{\partial^\nu \left( \overline{\Psi}_3 \gamma^{\sigma\mu} \Psi_3 \overline{\Psi}_3 \gamma^{\mu\sigma} \Psi_3 \right)}{\left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^2} \right)
 \end{aligned}$$

Because we must be able to place the fermions into one of three distinct eigenstates in order to satisfy Exclusion for the composite ground state faux monopole  $P'^{\sigma\mu\nu} \left( (0) \right)_0$ , we must now chose a dimension-3 gauge group in order to enforce this exclusion. There are two apparent choices. First is the simple group SU(3). Second is the product group SU(3)×U(1). But as we shall see in the next section, there really is not a choice and we actually must choose SU(3)×U(1). But to start simply, let us *assume* the simpler choice of SU(3) until contradicted, and then see why we are later compelled by contradiction to amend this choice to SU(3)×U(1). Choosing SU(3), we first label eigenstates. Because the labels are arbitrary, we use the names of some colors, say, (R)ed, (G)reen, (B)lue. Thus, using the SU(3) generators  $\lambda^i$  normalized to  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$  we define:

$$\Psi_1 \equiv \left| \lambda^8 = \frac{1}{\sqrt{3}}; \lambda^3 = 0 \right\rangle = \begin{pmatrix} \psi_R \\ 0 \\ 0 \end{pmatrix}; \Psi_2 \equiv \left| \lambda^8 = -\frac{1}{2\sqrt{3}}; \lambda^3 = \frac{1}{2} \right\rangle = \begin{pmatrix} 0 \\ \psi_G \\ 0 \end{pmatrix}; \Psi_3 \equiv \left| \lambda^8 = -\frac{1}{2\sqrt{3}}; \lambda^3 = -\frac{1}{2} \right\rangle = \begin{pmatrix} 0 \\ 0 \\ \psi_B \end{pmatrix}. \quad (9.9)$$

Now, all of a sudden, in a very consequential step, we see how these  $P'^{\sigma\mu\nu} \left( (0) \right)_0$  ground state magnetic monopole densities contain three fermions in one of three eigenstates R, G, B, and how SU(3) (or really, SU(3)×U(1) as we shall see in the next section) emerges as a *required* gauge group in order to force exclusion upon the fermions that comprise  $P'^{\sigma\mu\nu} \left( (0) \right)_0$ . In other words, we have never had to *postulate* SU(3) *per se* in order to force exclusion on the quarks within *experimentally*-observed baryons. Rather, *we have been forced to introduce* SU(3) (or at least a dimension-3 gauge group) in order to ensure proper Exclusion for the fermions of the *theoretically*-motivated  $P'^{\sigma\mu\nu}$  which first emerged back in (3.3) when we found that  $\oint\!\!\!\oint F \neq 0$  in a non-abelian gauge theory, and when we found that the underlying magnetic charge density was the composite  $P' = -idGG = -id[G, G]$  which is faux-assembled from the gauge fields  $G$ . At the same time, because we are *required* to select a dimension-3 gauge group which for now is SU(3), and because we have labelled the eigenstates with the names of colors, there are now eight gauge bosons  $G_\mu^i$  in  $G_\mu = \lambda^i G_\mu^i$  associated with (9.8), and each of these will be bi-colored, just as are the gluons of chromodynamic theory. This means that we may be able to obviate the need for a separate postulation of classical or quantum chromodynamics, such that *chromodynamics no longer a fundamental theory, but rather is a corollary, secondary theory* that emerges in the

process of enforcing fermion Exclusion upon the fermions contained in the non-abelian faux magnetic monopole density (9.8).

Now we focus on the terms of the form  $\Psi\bar{\Psi}$  which appear in the bottom line of (9.8). These terms have a column vector to the left of a row vector, and using (9.9), these may be explicitly written in 3x3 matrix form as:

$$\Psi_1\bar{\Psi}_1 = \begin{pmatrix} \psi_R\bar{\psi}_R & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad \Psi_2\bar{\Psi}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \psi_G\bar{\psi}_G & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad \Psi_3\bar{\Psi}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \psi_B\bar{\psi}_B \end{pmatrix}. \quad (9.10)$$

We may then use this to rewrite (9.8) in explicit 3x3 matrix form:

$$P'^{\sigma\mu\nu}((0))_0 = -i \begin{pmatrix} \frac{\partial^\sigma (\bar{\Psi}_1 \gamma^{\lambda\mu} \psi_R \bar{\psi}_R \gamma^{\nu\lambda} \Psi_1)}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} & 0 & 0 \\ 0 & \frac{\partial^\mu (\bar{\Psi}_2 \gamma^{\lambda\nu} \psi_G \bar{\psi}_G \gamma^{\sigma\lambda} \Psi_2)}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} & 0 \\ 0 & 0 & \frac{\partial^\nu (\bar{\Psi}_3 \gamma^{\lambda\sigma} \psi_B \bar{\psi}_B \gamma^{\mu\lambda} \Psi_3)}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \end{pmatrix}. \quad (9.11)$$

Next, we focus in on  $\psi_R\bar{\psi}_R = u_R\bar{u}_R$ ,  $\psi_G\bar{\psi}_G = u_G\bar{u}_G$  and  $\psi_B\bar{\psi}_B = u_B\bar{u}_B$  which involve ordinary, four-component Dirac wavefunctions  $\psi$  and spinors  $u$ , and we focus especially on the  $u\bar{u}$  which contain a column spinor to the left of a row spinor. Often, the Dirac spin sum relationship is normalized to  $N^2 = E + m$  and so is written as  $\sum_{\text{spins}} u\bar{u} = (p + m)$ . But if we wish to be more general and defer a decision on normalization, we may employ in (9.11) the spin sum *prior to normalization*, which is (see, e.g., [14] exercise 5.9):

$$\sum_{\text{spins}} u\bar{u} = \frac{N^2}{E + m} (p + m). \quad (9.12)$$

So, if we now take the sum over all spins  $\sum_{\text{spins}} P'^{\sigma\mu\nu}((0))_0$  of the faux monopole (9.11), and if we apply (9.12) in the form  $\sum_{\text{spins}} u_C\bar{u}_C = N^2 (p_C + m_C) / (E_C + m_C)$  to each color  $C = R, G, B$  of fermion, we may use (9.12) to rewrite (9.11), for the moment without  $+i\mathcal{E}$ , as:

$$\sum_{\text{spins}} P^{\sigma\mu\nu} ((0))_0 = -i \begin{pmatrix} \frac{N^2}{E_R + m_R} \frac{\partial^\sigma (\overline{\Psi}_1 \gamma^{\lambda\mu} (p_R + m_R) \gamma^{\nu 1} \Psi_1)}{(k_\tau k^\tau - m^2)^2} & 0 & 0 \\ 0 & \frac{N^2}{E_G + m_G} \frac{\partial^\mu (\overline{\Psi}_2 \gamma^{\lambda\nu} (p_G + m_G) \gamma^{\sigma 1} \Psi_2)}{(k_\tau k^\tau - m^2)^2} & 0 \\ 0 & 0 & \frac{N^2}{E_B + m_B} \frac{\partial^\nu (\overline{\Psi}_3 \gamma^{\lambda\sigma} (p_B + m_B) \gamma^{\mu 1} \Psi_3)}{(k_\tau k^\tau - m^2)^2} \end{pmatrix}. \quad (9.13)$$

Next we next turn our attention to the expressions  $(p_C + m_C)/(k_\tau k^\tau - m^2)$  which appear in each diagonal entry above. We simultaneously take note of the fact that the fermion propagator  $i(p - m)^{-1}$  sans  $+i\epsilon$  is related by a constant factor  $i$  to:

$$\frac{p + m}{p^\tau p_\tau - m^2} = \frac{p + m}{(p + m)(p - m)} = (p - m)^{-1}. \quad (9.14)$$

So we are motivated to see if there is a basis upon which we may set the  $(p_C + m_C)/(k_\tau k^\tau - m^2)$  terms in (9.13) to  $(p - m)^{-1}$  and thereby introduce the propagator for each of these fermions directly into (9.13). For this, we return to the discussion of sections 6 and 7 during which we developed inverse solutions to the electric charge equation  $-J^\nu = (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) G_\sigma$  in both massive and massless form, and where we also reviewed the degrees of freedom of various solutions and related questions of uniqueness.

Each term in equation (9.13) contains  $1/(k_\tau k^\tau - m^2)^2$ , that is  $1/(k_\tau k^\tau - m^2)$  times itself. As noted in the mass shell discussion prior to (6.24), we are using  $p^\tau$  and  $k^\sigma$  respectively to denote fermion and boson momentum vectors. And, of course, each  $1/(k_\tau k^\tau - m^2)$  entered (9.13) back at (9.2) when we inserted the massive boson inverse solution  $G_\mu = (-V + k_\tau k^\tau - m^2 + i\epsilon)^{-1} J_\mu$  of (6.27) into (9.1). As reviewed in sections 6 and 7, this solution, in view of the continuity requirement  $D_\nu J^\nu = 0$  of (5.20) and the consequently-mandated covariant gauge  $D_\nu G^\nu = 0$  of (6.5) is *unique* up to the gauge condition  $D_\nu D^\nu \theta = 0$  a.k.a.  $\partial_\nu \partial^\nu \theta - i \partial_\nu [G^\nu, \theta] = 0$ . And this solution is unchanged in form under a non-abelian gauge transformation because nowhere does the unphysical parameter  $\theta$  appear in any of the covariant physics equations. So in trying to match up  $(p_C + m_C)/(k_\tau k^\tau - m^2)$  which appears in (9.13) with  $(p + m)/(p^\tau p_\tau - m^2)$  in the fermion propagator-related (9.14), we see that the numerators match up perfectly but there is a mismatch in the denominators. Particularly, each

$k_\tau k^\tau - m^2$  in (9.13) is the propagator denominator for a massive gauge boson which has *three* degrees of freedom, while  $p^\tau p_\tau - m^2$  in (9.14) is the propagator denominator for a massive fermion which has *four* degrees of freedom. So, how do we match these up, and what impact, if any, might this have on the uniqueness of the solution  $G_\mu = (-V + k_\tau k^\tau - m^2 + i\varepsilon)^{-1} J_\mu$  upon which (9.13) is based?

Because each of the boson propagator denominators  $1/(k_\tau k^\tau - m^2)$  in (9.13) represents a massive boson with three degrees of freedom, the term  $1/(k_\tau k^\tau - m^2)^2$  which is a product of two boson propagator denominators thus represents six degrees of freedom. So we now take each  $1/(k_\tau k^\tau - m^2)(k_\tau k^\tau - m^2)$  and shift one degree of freedom from the first  $1/(k_\tau k^\tau - m^2)$  into the second  $1/(k_\tau k^\tau - m^2)$ . That is, keeping in mind that  $p^\tau$  and  $k^\sigma$  respectively denote fermion and boson momentum vectors and that the former has four degrees of freedom (particle / antiparticle in each of spin up and spin down states) and the latter when massive has three degrees of freedom (two transverse polarizations, one longitudinal), we rewrite  $1/(k_\tau k^\tau - m^2)^2$  as:

$$\frac{1}{(k_\tau k^\tau - m^2)^2} = \frac{1}{(k_\tau k^\tau - m^2)(k_\tau k^\tau - m^2)} = \frac{1}{k_\tau k^\tau (p_\tau p^\tau - m^2)}. \quad (9.15)$$

What we have effectively done is to take the 6=3+3 degrees of freedom represented in the first term, and redistribute them into 6=2+4 degrees of freedom represented in the final term. In the final term, therefore, we have turned one originally-massive gauge boson propagator denominator  $1/(k_\tau k^\tau - m^2)$  into a massless gauge boson propagator denominator  $1/k_\tau k^\tau$ . But at the same time, we have turned the other originally-massive gauge boson propagator denominator  $1/(k_\tau k^\tau - m^2)$  into a massive fermion propagator denominator  $1/(p_\tau p^\tau - m^2)$ . This is very analogous to the Goldstone mechanism used to give mass to massless gauge bosons by shifting a degree of freedom from a scalar field into a boson field. *Here, we are simply shifting a degree of freedom from a boson field into a fermion field.*

Now we saw of course in sections 6 and 7 that the solution for a massless gauge boson was less-unique than that for a massive boson, precisely because the massless gauge boson has one less degree of freedom. But we also saw how context matters, and how the context of a conserved current  $D_\nu J^\nu = 0$  contextually fixed the massless boson into the Feynman / continuity gauge  $\xi = 1$ . The only *contextual* loss of uniqueness in the massless solution, therefore, was that  $D_\nu G^\nu = 0$  was no longer a mandatory constraint but instead was relegated to a mere choice of gauge, which meant that  $\partial_\nu G^\nu = 0$  was also demoted from a requirement of continuity to an optional gauge condition. And all of the non-uniqueness of the massless solution, even before the application of continuity  $D_\nu J^\nu = 0$  fixed the gauge number to  $\xi = 1$ , emanated from *removing* a degree of freedom when going from a massive to a massless gauge boson. But in

(9.15) we are not *removing* any degrees of freedom as we did in going from section 6 to section 7. We are merely shifting them around in the overall context of (9.13) according to the recipe of (9.15). So after we apply (9.15) to (9.13), we will not in any way alter the uniqueness of (9.13). It will remain just as uniquely-specified after (9.15), as before (9.15). Effectively, we *contextually* dodge the additional non-uniqueness that emerges in going from the massive solutions of section 6 to the massless solutions of section 7, by *moving* rather than *removing* a degree of freedom, in the *context* of (9.13).

So let us now do exactly what we just said. We now use (9.15) in (9.13) to shift around the six degrees of freedom in each diagonal element from a 3+3 to a 2+4 configuration, and at the same time we label the  $p_\tau$  and the  $m$  in relation to the color of the fermion in each term. Thus, without any loss of uniqueness, simply by shifting a degree of freedom, (9.13) becomes:

$$\sum_{\text{spins}} P^{\sigma\mu\nu} ((0))_0 = -i \begin{pmatrix} \frac{N^2}{E_R + m_R} \frac{\partial^\sigma (\overline{\Psi}_1 \gamma^{\lambda\mu} (p_R + m_R) \gamma^{\nu\lambda} \Psi_1)}{k_\tau k^\tau (p_{R\tau} p_{R^\tau} - m_R^2)} & 0 & 0 \\ 0 & \frac{N^2}{E_G + m_G} \frac{\partial^\mu (\overline{\Psi}_2 \gamma^{\lambda\nu} (p_G + m_G) \gamma^{\sigma\lambda} \Psi_2)}{k_\tau k^\tau (p_{G\tau} p_{G^\tau} - m_G^2)} & 0 \\ 0 & 0 & \frac{N^2}{E_B + m_B} \frac{\partial^\nu (\overline{\Psi}_3 \gamma^{\lambda\sigma} (p_B + m_B) \gamma^{\mu\lambda} \Psi_3)}{k_\tau k^\tau (p_{B\tau} p_{B^\tau} - m_B^2)} \end{pmatrix}. \quad (9.16)$$

Importantly, in the process of shifting degrees of freedom, the remaining boson propagator denominator in each term has become  $1/k_\tau k^\tau$  which is the propagator for a *massless* gauge boson. So now, the eight bi-colored gauge bosons of the required  $SU(3)_C$  group have become massless, at the same time the fermions have acquired mass since they have four degrees of freedom following application of (9.15). Because the eight bi-colored gluons of QCD are also massless, this means that the gauge bosons associated with (9.16) have now have three very important symmetries that match up with the gluons of QCD: 1) there are eight of them, 2) they are bi-colored, and 3) they are massless. Yet, because of using a Goldstone-like method for what is a variant of the contextual gauge shifting discussed in section 7, *no uniqueness has been lost*.

Now we return to the normalization which we deferred back at (9.12). Often, as noted, the chosen normalization is  $N^2 = E + m$ . Let us instead, however, for each term in (9.16), choose to *include the  $k_\tau k^\tau$  massless boson term in the normalization*. That is, for each term in (9.16) let us now normalize to:

$$N^2 \equiv (E_C + m_C) k_\tau k^\tau. \quad (9.17)$$

So, applying the normalization (9.17), and propagator expression (9.14) for each fermion color  $C = R, G, B$ , we reduce (9.16) to:

$$\sum_{\text{spins}} P'^{\sigma\mu\nu} ((0))_0 = -i \begin{pmatrix} \partial^\sigma \left( \overline{\Psi}_1 \gamma^{\lambda\mu} (\mathbf{p}_R - m_R)^{-1} \gamma^{\nu\lambda} \Psi_1 \right) & 0 & 0 \\ 0 & \partial^\mu \left( \overline{\Psi}_2 \gamma^{\lambda\nu} (\mathbf{p}_G - m_G)^{-1} \gamma^{\sigma\lambda} \Psi_2 \right) & 0 \\ 0 & 0 & \partial^\nu \left( \overline{\Psi}_3 \gamma^{\lambda\sigma} (\mathbf{p}_B - m_B)^{-1} \gamma^{\mu\lambda} \Psi_3 \right) \end{pmatrix}. \quad (9.18)$$

Next we look closely at one of the terms above, say, the term  $\overline{\Psi}_1 \gamma^{\lambda\mu} (\mathbf{p}_R - m_R)^{-1} \gamma^{\nu\lambda} \Psi_1$  on the upper left. Making explicit use of (9.9), this term, is:

$$\overline{\Psi}_1 \gamma^{\lambda\mu} (\mathbf{p}_R - m_R)^{-1} \gamma^{\nu\lambda} \Psi_1 = \begin{pmatrix} \overline{\psi}_R & 0 & 0 \end{pmatrix} \gamma^{\lambda\mu} (\mathbf{p}_R - m_R)^{-1} \gamma^{\nu\lambda} \begin{pmatrix} \psi_R \\ 0 \\ 0 \end{pmatrix} = \overline{\psi}_R \gamma^{\lambda\mu} (\mathbf{p}_R - m_R)^{-1} \gamma^{\nu\lambda} \psi_R. \quad (9.19)$$

A similar result obtains for the other two terms, which now allows us to rewrite (9.18) as:

$$\sum_{\text{spins}} P'^{\sigma\mu\nu} ((0))_0 = -i \begin{pmatrix} \partial^\sigma \left( \overline{\psi}_R \gamma^{\lambda\mu} (\mathbf{p}_R - m_R)^{-1} \gamma^{\nu\lambda} \psi_R \right) & 0 & 0 \\ 0 & \partial^\mu \left( \overline{\psi}_G \gamma^{\lambda\nu} (\mathbf{p}_G - m_G)^{-1} \gamma^{\sigma\lambda} \psi_G \right) & 0 \\ 0 & 0 & \partial^\nu \left( \overline{\psi}_B \gamma^{\lambda\sigma} (\mathbf{p}_B - m_B)^{-1} \gamma^{\mu\lambda} \psi_B \right) \end{pmatrix}. \quad (9.20)$$

Any time we wish to calculate with the propagator terms  $i(\mathbf{p} - m)^{-1}$  and also include  $+i\epsilon$ , we set these to  $i(\mathbf{p} - m)^{-1} = i(\mathbf{p} + m) / (p_\sigma p^\sigma - m^2 + i\epsilon)$ .

Finally, in another important step that will lead us to topological stability, we take the trace of the above. This yields the fully-developed, spin-summed trace of the faux monopole density  $P' = -idGG = -id[G, G]$  in the zero-recursion, zero-perturbation limit  $((0))_0$ , namely:

$$\text{Tr} \sum_{\text{spins}} P'^{\sigma\mu\nu} ((0))_0 = -i \left( \partial^\sigma \left( \overline{\psi}_R \gamma^{\lambda\mu} (\mathbf{p}_R - m_R)^{-1} \gamma^{\nu\lambda} \psi_R \right) + \partial^\mu \left( \overline{\psi}_G \gamma^{\lambda\nu} (\mathbf{p}_G - m_G)^{-1} \gamma^{\sigma\lambda} \psi_G \right) + \partial^\nu \left( \overline{\psi}_B \gamma^{\lambda\sigma} (\mathbf{p}_B - m_B)^{-1} \gamma^{\mu\lambda} \psi_B \right) \right). \quad (9.21)$$

We shall now show how this has the identical symmetries as a baryon, how this leads directly to meson mediators of interactions between monopoles, how this requires us to choose  $\text{SU}(3) \times \text{U}(1)$  rather than  $\text{SU}(3)$  as our dimension-3 gauge group, how this leads to topological stability, and how the above becomes flavored into protons and neutrons.

## 10. Why the Composite Faux Magnetic Monopoles of Yang-Mills Gauge Theory have all of the Required Chromodynamic Symmetries of Baryons, and how these are Flavored into being Topologically-Stable Protons and Neutrons

In the trace form of (9.21), we see clearly that  $\text{Tr}\Sigma_{\text{spins}}P'^{\sigma\mu\nu}((0))_0$  is a third rank antisymmetric tensor in spacetime which will reverse sign under the interchange of any two adjacent indexes. From here, we simplify by just writing  $\Sigma_{\text{spins}} \rightarrow \Sigma$ . Let us denote this fundamental antisymmetry, which is an inherent feature of any magnetic monopole in spacetime, using the wedge-product notation  $\sigma \wedge \mu \wedge \nu$ . If we now associate each color wavefunction with the spacetime index in the related  $\partial^\sigma$  operator in (9.21), i.e.,  $\sigma \sim R$ ,  $\mu \sim G$  and  $\nu \sim B$ , and keeping in mind that  $\text{Tr}\Sigma P'^{\sigma\mu\nu}((0))_0$  is antisymmetric in all spacetime indexes, we may use  $\sigma \wedge \mu \wedge \nu \sim R \wedge G \wedge B = R[G, B] + G[B, R] + B[R, G]$  to express this antisymmetry. *But this is the exact colorless wavefunction that is expected of a baryon.* Indeed, *the antisymmetric character of the spacetime indexes in a magnetic monopole should have been a good tipoff that magnetic monopoles would naturally make good baryons.* So, we now may assert that the non-abelian composite faux monopole density  $\text{Tr}\Sigma P'^{\sigma\mu\nu}((0))_0$  in the ground state (9.21) has the exact same antisymmetric colorless chromodynamic symmetry as does a baryon!

Now, let us lower the indexes in (9.21) and write this as the differential form relation:

$$\begin{aligned}
 \text{Tr}\Sigma P'((0))_0 &= \text{Tr} \frac{1}{3!} \Sigma P'_{\sigma\mu\nu}((0))_0 dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -\frac{1}{3!} i \left( \begin{array}{l} \partial_\sigma \left( \overline{\psi}_R \gamma_{[\mu} (\mathbf{p}_R - m_R)^{-1} \gamma_{\nu]} \psi_R \right) \\ + \partial_\mu \left( \overline{\psi}_G \gamma_{[\nu} (\mathbf{p}_G - m_G)^{-1} \gamma_{\sigma]} \psi_G \right) \\ + \partial_\nu \left( \overline{\psi}_B \gamma_{[\sigma} (\mathbf{p}_B - m_B)^{-1} \gamma_{\mu]} \psi_B \right) \end{array} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu. \tag{10.1} \\
 &= -\frac{1}{3!} i \partial_\sigma \left( \begin{array}{l} \overline{\psi}_R \gamma_{[\mu} (\mathbf{p}_R - m_R)^{-1} \gamma_{\nu]} \psi_R \\ + \overline{\psi}_G \gamma_{[\mu} (\mathbf{p}_G - m_G)^{-1} \gamma_{\nu]} \psi_G \\ + \overline{\psi}_B \gamma_{[\mu} (\mathbf{p}_B - m_B)^{-1} \gamma_{\nu]} \psi_B \end{array} \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu
 \end{aligned}$$

In the bottom expression, a  $\partial_\sigma$  with the same  $\sigma$  index has been factored out of the entire expression. So now we can apply Gauss' / Stokes theorem to (10.1), and can use the forms in the top line of (9.1) to help us out.

Specifically, by expanding some of the forms in the top line of (9.1), we may write:

$$\begin{aligned}
 \oint\!\!\!\oint F &= \iiint P' = \oint\!\!\!\oint \frac{1}{2!} F_{\mu\nu} dx^\mu \wedge dx^\nu = \iiint \frac{1}{3!} P'_{\sigma\mu\nu} dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \iiint d[G, G] = -i \iiint \frac{1}{3!} \left( \partial_\sigma [G_\mu, G_\nu] + \partial_\mu [G_\nu, G_\sigma] + \partial_\nu [G_\sigma, G_\mu] \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu. \quad (10.2) \\
 &= -i \oint\!\!\!\oint [G, G] = -i \oint\!\!\!\oint \frac{1}{2!} [G_\mu, G_\nu] dx^\mu \wedge dx^\nu
 \end{aligned}$$

Therefore, taking the zero perturbation limit  $V = 0$ , summing all spins, taking the trace, and then injecting in the final expression from (10.1), we may write this as:

$$\begin{aligned}
 \oint\!\!\!\oint \text{Tr} \Sigma F((0))_0 &= \oint\!\!\!\oint \frac{1}{2!} \text{Tr} \Sigma F_{\mu\nu}((0))_0 dx^\mu \wedge dx^\nu = \iiint \text{Tr} \Sigma P'((0))_0 = \iiint \frac{1}{3!} \text{Tr} \Sigma P'_{\sigma\mu\nu}((0))_0 dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \iiint \text{Tr} \Sigma d[G, G]((0))_0 = -i \iiint \frac{1}{3!} \text{Tr} \Sigma \left( \partial_\sigma [G_\mu, G_\nu] + \partial_\mu [G_\nu, G_\sigma] + \partial_\nu [G_\sigma, G_\mu] \right) ((0))_0 dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \oint\!\!\!\oint \text{Tr} \Sigma [G, G]((0))_0 = -i \oint\!\!\!\oint \frac{1}{2!} \text{Tr} \Sigma [G_\mu, G_\nu]((0))_0 dx^\mu \wedge dx^\nu \quad (10.3) \\
 &= -i \iiint \frac{1}{3!} \partial_\sigma \left( \overline{\psi}_R \gamma_{[\mu} (\mathbf{p}_R - m_R)^{-1} \gamma_{\nu]} \psi_R + \overline{\psi}_G \gamma_{[\mu} (\mathbf{p}_G - m_G)^{-1} \gamma_{\nu]} \psi_G + \overline{\psi}_B \gamma_{[\mu} (\mathbf{p}_B - m_B)^{-1} \gamma_{\nu]} \psi_B \right) dx^\sigma \wedge dx^\mu \wedge dx^\nu \\
 &= -i \oint\!\!\!\oint \frac{1}{2!} \left( \overline{\psi}_R \gamma_{[\mu} (\mathbf{p}_R - m_R)^{-1} \gamma_{\nu]} \psi_R + \overline{\psi}_G \gamma_{[\mu} (\mathbf{p}_G - m_G)^{-1} \gamma_{\nu]} \psi_G + \overline{\psi}_B \gamma_{[\mu} (\mathbf{p}_B - m_B)^{-1} \gamma_{\nu]} \psi_B \right) dx^\mu \wedge dx^\nu
 \end{aligned}$$

From this we extract several integrands with an overall multiplication by  $i$ :

$$\begin{aligned}
 \text{Tr} \Sigma i F_{\text{eff} \mu\nu}((0))_0 &\equiv \text{Tr} \Sigma [G_\mu, G_\nu]((0))_0 \\
 &= \overline{\psi}_R \gamma_{[\mu} (\mathbf{p}_R - m_R)^{-1} \gamma_{\nu]} \psi_R + \overline{\psi}_G \gamma_{[\mu} (\mathbf{p}_G - m_G)^{-1} \gamma_{\nu]} \psi_G + \overline{\psi}_B \gamma_{[\mu} (\mathbf{p}_B - m_B)^{-1} \gamma_{\nu]} \psi_B. \quad (10.4)
 \end{aligned}$$

This includes defining an “effective”  $\text{Tr} \Sigma i F_{\text{eff} \mu\nu}((0))_0$ . This is because while (1.5) tells us that  $F_{\mu\nu} = \partial_{[\mu} G_{\nu]} - i [G_\mu, G_\nu]$  so that  $\text{Tr} \Sigma [G_\mu, G_\nu] = \text{Tr} \Sigma i F_{\mu\nu} - \text{Tr} \Sigma i \partial_{[\mu} G_{\nu]}$ , as found in (3.5) the total net flux  $\oint\!\!\!\oint F$  is invariant under the transformation  $F^{\mu\nu} \rightarrow F'^{\mu\nu} = F^{\mu\nu} - \partial^{[\nu} G^{\mu]}$ . This means that the gauge field is *not observable* with respect to net flux across closed surfaces of the monopole precisely because of the abelian subset expression  $\oint\!\!\!\oint dG = \mathbf{0}$  which is responsible for there being no net flux of magnetic fields *at all* across a closed surface in abelian gauge theory. So while  $\oint\!\!\!\oint \text{Tr} \Sigma F((0))_0 = -i \iiint \text{Tr} \Sigma d[G, G]((0))_0$  in the integral formation of (10.3) by virtue of the symmetry principle (3.5), when the integrands are separately extracted as in (10.4), the actual relationship is  $F_{\mu\nu} = \partial_{[\mu} G_{\nu]} - i [G_\mu, G_\nu]$ . But the *effective* relationship in terms of what actually becomes *net observable flux across closed surfaces*, is  $F_{\text{eff} \mu\nu} = -i [G_\mu, G_\nu]$ . That is the basis for the definition of  $F_{\text{eff} \mu\nu}$  in (10.4).

By inspection,  $\text{Tr} \Sigma [G_\mu, G_\nu]((0))_0$  in (10.4) has the color wavefunction  $\overline{R}R + \overline{G}G + \overline{B}B$  of a meson. But look at the context in which this meson wavefunction has appeared in (10.3): Using selected terms from (10.3), especially  $\oint\!\!\!\oint \text{Tr} \Sigma F((0))_0$ , we see that:

$$\begin{aligned} \oint\!\!\!\oint \text{Tr}\Sigma F((0))_0 &= -i\oint\!\!\!\oint \text{Tr}\Sigma[G, G]((0))_0 = -i\oint\!\!\!\oint \frac{1}{2!} \text{Tr}\Sigma[G_\mu, G_\nu]((0))_0 dx^\mu \wedge dx^\nu \\ &= -i\oint\!\!\!\oint \frac{1}{2!} \left( \overline{\Psi}_R \gamma_{1\mu} (\not{p}_R - m_R)^{-1} \gamma_\nu \Psi_R + \overline{\Psi}_G \gamma_{1\mu} (\not{p}_G - m_G)^{-1} \gamma_\nu \Psi_G + \overline{\Psi}_B \gamma_{1\mu} (\not{p}_B - m_B)^{-1} \gamma_\nu \Psi_B \right) dx^\mu \wedge dx^\nu \end{aligned} \quad (10.5)$$

So we see that the Yang-Mills magnetic fields which net-flow across closed surfaces of the composite, faux magnetic monopole density  $P' = -idGG = -id[G, G]$  of non-abelian gauge theory in the form of  $\oint\!\!\!\oint \text{Tr}\Sigma F((0))_0$ , have the  $\overline{RR} + \overline{GG} + \overline{BB}$  color symmetry of mesons!

*This is a very important finding.* Back at (3.3) we identified a puzzle: We found that in non-abelian Yang-Mills gauge theory there is a non-zero net flow of magnetic fields across closed surfaces,  $\oint\!\!\!\oint F \neq 0$ , yet at the same time the magnetic charge density completely vanished  $P = DF = DDG = 0$  just like in abelian gauge theory. To reconcile this, we determined that the magnetic charge density in non-abelian gauge theory is not the elementary  $P = DF = DDG = 0$ , but rather is a composite *faux* magnetic charge density  $P' = -id[G, G] = -idGG$  constructed from gauge fields, and particularly, that the net flux of magnetic field is given by  $\oint\!\!\!\oint F = -i\oint\!\!\!\oint [G, G] \neq 0$  in (3.3).

Ever since then, we have known that non-abelian gauge theory gives rise to a non-zero  $\oint\!\!\!\oint F \neq 0$ , but beyond a few vague hints pointing in the possible direction of baryons and confinement, it has not been known what the physics of this  $\oint\!\!\!\oint F \neq 0$  might be. Now, we see in (10.5) that  $\oint\!\!\!\oint \text{Tr}\Sigma F(0) = -i\oint\!\!\!\oint \text{Tr}\Sigma[G, G](0) \sim \overline{RR} + \overline{GG} + \overline{BB}$ . In other words, the composite faux magnetic fields which net flow across closed surfaces in non-abelian gauge theory are simply colorless mesons with the symmetric  $\overline{RR} + \overline{GG} + \overline{BB}$  wavefunction. Colorless  $\overline{RR} + \overline{GG} + \overline{BB}$  mesons – which, once flavored, include such things as the pions that mediate nuclear interactions – are simply the  $\oint\!\!\!\oint F \neq 0$  faux magnetic monopole fields of Yang-Mills gauge theory. That means that these  $\text{Tr}\Sigma iF_{\text{eff}\mu\nu} = \text{Tr}\Sigma[G_\mu, G_\nu]$  objects in (10.4) – which are the only objects which flow in and out of the monopoles – *must be the mediators of interactions between the monopoles*. So if those monopoles are baryons as suggested by their  $R[G, B] + G[B, R] + B[R, G]$  wavefunctions, and if these baryons can be turned into protons and neutrons as well shall show how to do momentarily, then these  $\text{Tr}\Sigma iF_{\text{eff}\mu\nu} = \text{Tr}\Sigma[G_\mu, G_\nu]$  fields are also the mediators of the nuclear interaction. And this also means that we should look to  $\text{Tr}\Sigma iF_{\text{eff}\mu\nu} = \text{Tr}\Sigma[G_\mu, G_\nu]$  when studying anything that might pass in and out of a proton or neutron through a closed  $\oint\!\!\!\oint$  surface including energies released during nuclear fusion and fission which of course are intimately related to nuclear binding energies.

Related to this, to ensure Exclusion for the fermions in (9.8), we were forced to introduce a dimension-3 gauge group which we assumed to be  $\text{SU}(3)_C$ . As pointed out after (9.16), after

shifting the degrees of freedom using a Goldstone-like mechanism, this yielded eight associated gauge fields, which are bi-colored and massless, just like the strong interaction gluons. As had been earlier shown at (3.5), the abelian properties of the differential geometry via  $dd = 0$  which is responsible in electrodynamics for the absence of magnetic monopoles entirely, prevents individual gauge fields – now these eight bi-colored massless gauge fields – from net flowing across any closed surface of the faux magnetic monopole  $P'$  because of  $\oint\oint dG = \iiint R^\tau{}_{\nu\sigma\mu} G_\tau dx^\sigma dx^\mu dx^\nu = \mathbf{0}$ . So in this way, these eight bi-colored massless gauge fields appeared to be *confined*. What we now see more explicitly and deeply in (10.5) is that the only thing which *does* net flow across these closed surfaces, are mesons which possess a color wavefunction  $\overline{RR} + \overline{GG} + \overline{BB}$ . And finally we saw at the start of this section that the faux magnetic monopoles themselves possess the totally-antisymmetric color wavefunction of a baryon, namely,  $R[G, B] + G[B, R] + B[R, G]$ . While one may think of this as color “confinement,” what it really says is that is that the non-abelian faux magnetic monopoles  $P'$  and the mesons  $[G, G]$  which net flow across closed surfaces of these monopoles, respectively, are antisymmetrically and symmetrically *color neutral*, and that nothing is permitted to net-flow across a closed monopole surface *unless* it has a  $\overline{RR} + \overline{GG} + \overline{BB}$  neutral color configuration. So individual gauge fields, because they are bi-colored and not color neutral, are confined.

With all of this, we see multiple symmetries which are highly reminiscent of hadron physics: We are forced to introduce three fermion eigenstates which can be arbitrarily named as three “colors” just like the quark fields which transform non-trivially under SU(3) in the chromodynamic theory of strong interactions. What is arbitrary are the names; *what is not arbitrary is that we require three such names*. This simultaneously produces eight bi-colored gauge fields, also transforming non-trivially under SU(3), just as is the case for the strong interaction gluons, and so *derives* the chromodynamic requirement for a theory with three colors of fermion and eight bi-colors of gluon, and shows why baryons contain three quarks. These gluons after using the Goldstone-like mechanism in (9.16) must become massless just like the strong interaction gluons. The faux magnetic monopoles (9.21) have the antisymmetric, color-neutral symmetry of a baryon, and so are SU(3)-invariant. No gauge fields are allowed to net flow across any closed surface of this monopole, which means that the gauge fields are “confined” within the closed monopole surface, just like individual gluons. Yet there *is* a net flux of a non-abelian magnetic field across the closed monopole surfaces, as we found all the way back in section 3. Now, we see that these net-flowing magnetic fields have the symmetric, color-neutral symmetry of a meson, which means that they too are SU(3)-invariant, and that interactions between the faux monopoles will take place via colorless meson exchange, exactly as occurs in strong hadronic interactions between baryons.

Or, as Jaffe and Witten make clear at page 3 of [6], “quark confinement” is evidenced when:

“even though the theory is described in terms of elementary fields, such as the quark fields, that transform non-trivially under SU(3), the physical particle states—such as the proton, neutron, and pion—are SU(3)-invariant.”

This is exactly what transpires if one regards the composite faux magnetic monopole of (9.21) as a zero-perturbation, ground state baryon density! Given all of these symmetries, from here we shall regard the monopole  $\text{Tr}\Sigma P'((0))_0$  as a ground state baryon. And this means that (9.3), and specifically  $\text{Tr}\Sigma P'((0))_\infty$ , which contains  $\pi_\infty = (\pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau)^{-1}$  which can be expanded using (8.18) to reveal an exceptionally-non-linear system with perturbations up to infinite order in current density  $J$  and gauge field momentum  $k$ , is the *physical baryon* with all of its non-linear quark and gluon field behaviors.

Proceeding forward, we now expand the differential forms relationship for the faux magnetic charge density  $P' = -id[G, G]$  ( $= -idGG$ ) uncovered after (3.3) into tensor form, expand  $G_\mu = \lambda^i G^i_\mu$ , and then, having extracted the group generators, finally apply the SU(3) group relation  $[\lambda^i, \lambda^j] = if^{ijk} \lambda^k$ . This yields:

$$\begin{aligned} P'_{\sigma\mu\nu} &= -i(\partial_\sigma [G_\mu, G_\nu] + \partial_\mu [G_\nu, G_\sigma] + \partial_\nu [G_\sigma, G_\mu]) \\ &= -i[\lambda^i, \lambda^j](\partial_\sigma (G^i_\mu G^j_\nu) + \partial_\mu (G^i_\nu, G^j_\sigma) + \partial_\nu (G^i_\sigma, G^j_\mu)) \\ &= f^{ijk} \lambda^k (\partial_\sigma (G^i_\mu G^j_\nu) + \partial_\mu (G^i_\nu, G^j_\sigma) + \partial_\nu (G^i_\sigma, G^j_\mu)) \end{aligned} \quad (10.6)$$

Let us now *assume* as we have since after (9.9) that our gauge group is the simple *subgroup* SU(3) with the eight traceless generators  $\lambda^k$ ,  $k=1\dots 8$  often referred to as the Gell-Mann matrices. If we now take the trace of the above, given that the eight  $\lambda^k$  of the *subgroup* SU(3) are all traceless,  $\text{Tr}\lambda^k = 0$ , (10.6) tells us that  $\text{Tr}P'_{\sigma\mu\nu} = 0$ .

But (9.21) has a non-zero trace, and so it is worthwhile understanding how it is that even when we assume an SU(3) subgroup with  $\text{Tr}\lambda^k = 0$ , we can still end up with a non-zero trace equation (9.21). The key is to closely examine (9.7), which is why we chose to display the intermediate terms even though we could have gone directly from (9.6) to the bottom line (9.7) using  $J^\mu = \bar{\Psi}\gamma^\mu\Psi$  without showing generators or internal symmetry indexes. The key is that (9.6) contains commutators  $[J^\mu, J^\nu]$ , and so contains a very specific type of second-order expression for the currents  $J^\nu$ . Although the generators are traceless, when any generator is squared and then traced, the result in the customary normalization is the non-zero  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$ .

In the intermediate terms (9.7), we see multiple sums  $\lambda^i \lambda^i$  of a generator with itself. When all of the anti-symmetries in these intermediate terms are accounted for, the result is the bottom line of (9.7) which, by the time it is worked into (9.21), reflects in a deeper way of the general result that  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$  is not zero.

Nonetheless, (10.6) appears to contradict this non-zero trace result obtained in (9.21) wherein  $\text{Tr}\Sigma P'^{\sigma\mu\nu}(0) \neq 0$ . This is another puzzle. But think about this more closely: In (9.9)

we were compelled to introduce a dimension-3 gauge group to enforce exclusion for each of the fermion wavefunctions in (9.8). But all we really knew is that we needed three mutually-exclusive eigenstates and therefore required a dimension-3 gauge group. Although we could have just as readily chosen  $SU(3)\times U(1)$ , we *assumed* that the gauge group could be  $SU(3)$  unless and until contradicted. But now this assumption *is* contradicted. Specifically, based on the development up to (9.8), the choice of a gauge group appeared to be non-unique. Any dimension-3 group would do. But by the time we reached (9.21), it became clear that we had a  $\text{Tr}P'_{\sigma\mu\nu} \neq 0$ , i.e., that  $P'_{\sigma\mu\nu}$  must have a non-vanishing trace. If one tries to write (9.21) in the same way as (10.6) to extract out an overall  $f^{ijk} \lambda^k$ , it cannot be done, other than by backtracking to (9.7). The development from (9.7) (where this still could be done) to (9.21) removed the ability to do so, and in particular, that started to happen once we used (9.12) in (9.13) and summed spins to remove two wavefunctions using the fermion spin sum.

Now, (10.6) informs us that if the gauge group is  $SU(3)$  then the trace will vanish. So now, what appeared at (9.9) to be a non-unique choice of  $SU(3)$  or  $SU(3)\times U(1)$  is forced by (9.21) in view of (10.6) to be a *unique* choice of  $SU(3)\times U(1)$ , with  $\lambda^0$  used to denote the new  $U(1)$  generator, which now also adds one more degree of freedom to the (9.21) system. Of course, we will now need to determine what this additional  $U(1)$  generator represents, and as we shall see, it represents the baryon number  $B=1/3$  for each of the three colored fermions appearing in (9.21) and may be used to more formally turn the faux magnetic monopole density (10.6) into a baryon density. As we shall also see, while the gauge group  $SU(3)$  by itself is simply the usual color group  $SU(3)_C$  of strong interaction chromodynamic theory, once this group gets crossed with  $U(1)$  it becomes a “modified” color group which mixes color and *flavor* because the introduction of baryon number also facilitates the introduction of the flavor-distinguishing electric charge generator  $Q$ . But before we discuss this, there is a more general point that must be made, and this has to do with topological stability.

Cheng and Li point out at 472-473 of [17] that “topological considerations lead to the general result that stable monopole solutions occur for any gauge theories in which a *simple* gauge group  $G$  is broken down to a smaller group  $H = h \times U(1)$  containing an explicit  $U(1)$  factor.” Further, “the stable grand unified monopole . . . is expected to have both the ‘ordinary’ and the colour magnetic charges.” So, while  $SU(3)$  alone is incapable of supporting a topologically-stable colored magnetic monopole, the group  $SU(3)\times U(1)$  – when understood to be the residual group following symmetry breaking of a larger simple grand unified gauge group  $G \supset SU(3)\times U(1)$  – *will support topologically stable configurations*. This is an essential requirement if the faux monopole (10.6) can ever be regarded as a physically-stable entity like a baryon, and especially a distinctively-stable proton, and a neutron which is comparatively stable when free, and very stable when part of many lighter atomic nuclei.

Weinberg makes a similar point to Cheng and Li in his definitive treatise [18] at 442:

“The Georgi-Glashow model was ruled out as a theory of weak and electromagnetic interactions by the discovery of neutral currents, but magnetic monopoles are expected to occur in other theories, where a simply connected gauge group  $G$  is spontaneously broken not to  $U(1)$ , but to some subgroup

$H' \times U(1)$ , where  $H'$  is simply connected. . . . There are no monopoles produced in the spontaneous breaking of the gauge group  $SU(2) \times U(1)$  of the standard electroweak theory, which is not simply connected. . . . But we do find monopoles when the simply connected gauge group  $G$  of theories of unified strong and electroweak interactions, such as  $SU(4) \times SU(4)$  or  $SU(5)$  or  $Spin(10)$ , is spontaneously broken to the gauge group  $SU(3) \times SU(2) \times U(1)$  of the standard model. . . .”

Consequently, not only does (9.8) force us to uniquely select a dimension-3 gauge group to enforce Exclusion on the faux magnetic monopole density of (9.8), but the non-vanishing trace of (9.21) forces us into the *specific, unique selection* of  $SU(3) \times U(1)$  over  $SU(3)$ . This then ensures that these faux monopoles will be topologically stable so long as we arrive at this product group following the spontaneous symmetry breaking of a larger simple gauge group  $G = SU(N \geq 4) \supset SU(3) \times U(1)$ , as yet undetermined. Topologically speaking, referring again to Weinberg’s [18] at 442, the homotopy groups associated with this symmetry breaking would be:

$$\pi_2(G / SU(3) \times U(1)) = \pi_1(SU(3) \times U(1)) = \pi_1(SU(3)) \times \pi_1(U(1)) = \pi_1(U(1)) = Z. \quad (10.7)$$

So there are really two questions raised by the non-vanishing trace in (9.21). First, as already stated, what is the physical meaning of the new  $U(1)$  generator? Second, what is the larger group  $G = SU(N \geq 4) \supset SU(3) \times U(1)$  from which we arrive at  $SU(3) \times U(1)$  following symmetry breaking so as to achieve topological stability? There is also a third question, not yet apparent, but linked to the first question, which is this: what is the meaning of the  $SU(3)$  group which is multiplied by the new  $U(1)$  gauge group as part of  $SU(3) \times U(1)$ , and how does this relate to the usual color group  $SU(3)_C$ ?

For the new  $U(1)$  group which provides topological stability, the generator  $\lambda^0$  must be a constant multiple of the  $3 \times 3$  identity (unit) matrix  $I_{3 \times 3}$ . If we normalize this to  $\text{Tr}(\lambda^0)^2 = \frac{1}{2}$  just like all the other generators, then we must have  $\lambda^0 = \frac{1}{\sqrt{6}} I_{3 \times 3}$ . Taken together with the two remaining diagonalized generators of  $SU(3)$  normalized to  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$ , we have:

$$\lambda^0 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \lambda^8 = \frac{1}{2\sqrt{3}} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}; \quad \lambda^3 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (10.8)$$

But that is only the mathematics: now we need a *physical* interpretation for  $\lambda^0$ . Because each of the three fermion eigenstates in (9.9) will have identical  $\lambda^0$  eigenvalues, because the monopole in (9.21) exhibits many of symmetries of a baryon and the fermions exhibit many of the symmetries of quarks, it would appear fruitful to assign the  $U(1)$  generator to baryon number  $B$  according to:

$$B \equiv 2 \frac{1}{\sqrt{6}} \lambda^0 = \frac{1}{3} I_{3 \times 3} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}. \quad (10.9)$$

This is our first explicit introduction of *flavor* into the color eigenstates that were introduced at (9.9). Following (10.9), the monopole (9.21) will now have  $B=1$  and each of the R, G, B fermions will now have  $B = \frac{1}{3}$ , which brings these even a step closer to being identifiable with baryons and quarks.

Next, if these monopoles (9.21) are to be baryons and the fermions are to be quarks, let us see if there is some way to identify the *electric charge*  $Q$  of these baryons, and specifically to produce a proton with  $Q = +1$  which has a duu configuration of quark *flavors*, and a neutron with  $Q = 0$  which has a udd configuration of quark flavors, wherein the up (u) quark has  $Q = +\frac{2}{3}$  and the down (d) quark has  $Q = -\frac{1}{3}$ .

For the proton, we may form the combination:

$$Q_p \equiv B - \frac{2}{\sqrt{3}} \lambda^8 = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} - \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}, \quad (10.10)$$

Following (10.9), each of the R, G, B colored fermions in (9.9) has a flavored baryon number  $B = \frac{1}{3}$ . Now, with (10.10), the red *color* of fermion is assigned  $Q = -\frac{1}{3}$  and so is a down *flavor* of fermion in addition to its red color assignment, the green and blue *colors* of quark are assigned  $Q = +\frac{2}{3}$  and so are up *flavors* of fermion in addition to their green and blue color assignments. So the  $SU(3) \times U(1)$  quark triplet is now  $(d_R, u_G, \mu_B)$ . Further, the entire faux monopole  $\text{Tr} \Sigma P^{\sigma\mu\nu} ((0))_0$  of (9.21) which comprises all of these fermions has a baryon number  $B=1$  and an electric charge  $Q = +1$  and so is a proton-flavored baryon with the color-neutral wavefunction  $R[G, B] + G[B, R] + B[R, G]$ . To use a parlance familiar from electroweak theory, we see in (10.10) that the electric charge generator for the proton and for the quarks within the proton *sit across* baryon number  $B$  and the  $\lambda^8$  color generator, that is, they sit across  $SU(3) \times U(1)$  in a *non-compact* manner. In similar fashion, in electroweak theory a  $U(1)_Y$  generator is crossed with the three  $SU(2)_W$  isospin generators  $I^i$ ,  $i=1,2,3$  to form  $SU(2)_W \times U(1)_Y$  with the (left-chiral) quark doublets having the  $U(1)_Y$  2x2 weak hypercharge matrix generator  $Y = \frac{1}{3} I_{2 \times 2}$ , the (left-chiral) lepton doublets having the 2x2 weak hypercharge matrix generator  $Y = -1 I_{2 \times 2}$ , and a *non-compact* embedding of the electromagnetic group with charge generator  $Q = Y/2 + I^3$  *sitting across*  $SU(2)_W \times U(1)_Y$ .

For the neutron it is even simpler. We simply make the *compact* assignment:

$$Q_N \equiv \frac{2}{\sqrt{3}} \lambda^8 = \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & -\frac{1}{3} \end{pmatrix}. \quad (10.11)$$

Here, all of the fermions still have baryon number  $B = \frac{1}{3}$ . But now the red fermion is assigned  $Q = +\frac{2}{3}$  thus is an up flavored-fermion, the green and blue fermions are assigned  $Q = -\frac{1}{3}$  and so are down flavored. So the quark triplet is now  $(u_R, d_G, d_B)$ . The overall faux monopole of (9.21) now has baryon number  $B=1$  and electric charge  $Q=0$  and so is a neutron-flavored baryon. So the electric charge generator for the neutron and its quarks is compactly-embedded in  $\lambda^8$  which now serves the dual role of one of two  $SU(3)_C$  generators *and* the electric charge generator.

Of course, the fact that we must employ a different charge assignment (10.10) for the proton than (10.11) for the neutron is symptomatic that there is a larger yet-to-be-found gauge group which encompasses the  $SU(3) \times U(1)$  group developed in (10.8) through (10.11). That is  $Q_P = B - \frac{2}{\sqrt{3}} \lambda^8$  and  $Q_N = \frac{2}{\sqrt{3}} \lambda^8$  is not invariant whereby one relationship, not two, defines the relationship between the electric charge and the group generators. This disconnection between the proton and neutron electric charges is analogous to how in electroweak theory, the  $Y_q = \frac{1}{3}$  for the quark (q) doublets is disconnected from the  $Y_l = -1$  lepton (l) doubles which there too, signifies the need for a larger unifying group. So the question is now raised: what is the nature of the gauge group that provides a unified basis for the proton and neutron electric charges  $Q$ , and can this same group *also* provide the basis for unifying the separate  $Y$  charges as between quarks and leptons while also dealing with chiral symmetry (breaking) issues?

While we shall not explore this here, the author has studied these exact questions in [19] and shown how a simple  $SU(8)$  group with the fundamental fermion multiplet  $(\nu, u_R, d_G, d_B, e, d_R, u_G, u_B)$  provides a complete unification which breaks down at low energies to the phenomenological  $SU(3)_C \times SU(2)_W \times U(1)_Y$  with protons and neutrons, and at the same time – because two of the diagonalized  $SU(8)$  generators themselves become “fractured” apart from the other five diagonalized generators during symmetry breaking – leads to an explanation of why the known fermions appear to exist in exactly three generations, which answers Isador Rabi’s famous quip about the muon “who ordered this?” That is because these two “fractured” generators provide the precise freedom needed to accommodate three horizontal generational eigenstates.

But what we now know from the development within this paper and specifically (10.10) and (10.11) is that the  $SU(3)_C$  group which we introduced at (9.9) to enforce Exclusion actually becomes modified into a *hybrid color and flavor group* in view of the requirement to use  $SU(3) \times U(1)$  because of the non-vanishing trace in (9.21). We shall thus refer to this as a “flavor-enhanced color group” which we denote generally by  $SU(3)_C$ . When we use this group to represent a proton (P) quark triplet  $(d_R, u_G, \mu_B)$  with the charge assignments (10.10) we shall

further denote this by  $SU(3)_{PC}$ , while when we use this to represent a neutron (N) quark triplet  $(u_R, d_G, d_B)$  with the charge assignments (10.11) we shall denote this by  $SU(3)_{NC}$ . Finally, in all cases, the  $U(1)$  factor is associated with baryon number  $B$ , so we shall denote this as  $U(1)_B$ . So to summarize, once the  $U(1)$  factor is in place, the group developed thus far is  $SU(3)_C \times U(1)_B$ . For protons it is specialized via (10.10) to  $SU(3)_{PC} \times U(1)_B$ . For neutrons it is specialized via (10.11) to  $SU(3)_{NC} \times U(1)_B$ .

Next, keeping in mind (10.7), it also becomes important to find a larger simple gauge group  $G = SU(N \geq 4) \supset SU(3)_C \times U(1)_B$  which breaks down spontaneously to  $SU(3)_C \times U(1)_B$ . As the author details in section 7 of [15], there are two disconnected  $G = SU(4)$  groups, but we are able to use  $B - L \equiv -\sqrt{\frac{8}{3}}\lambda^{15}$  as the generator of baryon minus lepton number for both. This follows Volovok from [20] Section 12.2.2 who also uses the  $\lambda^{15}$  of  $SU(4)$  for a  $B - L$  generator, but in the context of a preon model. The first group, denoted  $SU(4)_P$ , places the proton's quarks and the electron into a  $(e, d_R, u_G, u_B)$  quadruplet in the fundamental representation. The second group, denoted  $SU(4)_N$ , places the neutron's quarks and the neutrino into a  $(\nu, u_R, d_G, d_B)$  quadruplet in the fundamental representation. Then, each of these disconnected proton and neutron groups gets broken at GUT energies via  $G = SU(4)_{B-L} \rightarrow SU(3)_C \times U(1)_B$  to produce the stable magnetic monopole baryons via:

$$\pi_2(SU(4)_{B-L} / SU(3)_C \times U(1)_B) = \pi_1(SU(3)_C \times U(1)_B) = \pi_1(SU(3)_C) \times \pi_1(U(1)_B) = \pi_1(U(1)_B) = Z. \quad (10.7)$$

Then, as the author details throughout [19], the disconnected  $SU(4)_N$  and  $SU(4)_P$  groups become unified together in the  $(\nu, u_R, d_G, d_B, e, d_R, u_G, u_B)$  of  $SU(8)$  mentioned moments ago, such that two of the seven generators ( $\lambda^{48}$  and  $\lambda^{35}$ ) become fractured from the remaining generators between the Planck and the GUT energy scales to provide the "horizontal" degrees of freedom needed to accommodate replication of the fermions into three generations, and there is also just enough freedom provided to also support chiral symmetry breaking. Additionally, all of the observed features of left-chiral Cabibbo / CKM mixing naturally emerge. The overall sequence of symmetry breaking is:

$$SU(8) \rightarrow SU(6)_B \times SU(2)_L \rightarrow SU(3)_C \times SU(2)_W \times U(1)_{Y_L=B-L} \rightarrow SU(3)_C \times U(1)_{em}. \quad (10.12)$$

Simultaneously with and as part of the  $SU(8) \rightarrow SU(6)_B \times SU(2)_L$  symmetry breaking, the two isospin-differing  $SU(4)_{B-L} \rightarrow SU(3)_C \times U(1)_B$  symmetry breaks also take place to form the topologically-stable proton and neutron. There is also an earlier breaking of  $SU(8) \rightarrow SU(7) \times U(1)$  at or near Planck energies which separates the neutrino from all the other fermions right at the very start and causes the neutrino to behave very differently from all the other fermions as it clearly does at observable energies. The symmetry breaking sequences found in [19] are then utilized in [21] to explain the observed proton and neutron masses themselves in relation to the current up and down quark masses and the CKM mixing matrices based on [16], *within all experimental errors*.

Next, let us return to (9.4) where we set the perturbation to  $V = 0$  in (9.2). Because everything that has been developed since (9.2), most notably the  $\text{Tr}\Sigma P'^{\sigma\mu\nu} \left( (0) \right)_0$  monopole / baryon of (9.21) was developed for  $V = 0$ , the question may be asked whether all of these results carry through when we no longer set  $V = 0$  but allow all of the perturbations to occur. Section 8 answers this question. What we learn in section 8 is that including perturbations really means recursing  $G_\mu = \left( k_\tau k^\tau - m^2 + i\mathcal{E} + G_\tau k^\tau + G_\tau G^\tau \right)^{-1} J_\mu$  as many times as one chooses, then cutting off the recursion by setting  $V = -G_\tau k^\tau - G_\tau G^\tau = 0$  at some chosen recursive order. Of course, recursing to some order  $n$  and then setting  $V=0$  as in (8.17) and (8.18) to arrive at a  $\left( (0) \right)_n$  expression is a calculation technique. But it is to be expected that nature does not cut off the recursion at all, but rather, recurses to infinity before setting  $V = 0$ , so that  $G_\mu = \pi_\infty J_\mu$  as in (8.20). So if the monopole  $\text{Tr}\Sigma P'^{\sigma\mu\nu} \left( (0) \right)_0$  of (9.21) is the ground state of the baryon, it will be the infinite recursion of (8.20), not some arbitrarily truncated recursion, which will drive what nature herself does in physical reality. This means that (9.3) in the form  $\text{Tr}\Sigma P'^{\sigma\mu\nu} \left( (0) \right)_\infty$ , is really the equation for the *physical* baryon, with a teeming non-linear mix of quarks and gauge fields in a “sea” perturbing through all finite orders up to infinite order, which is exactly what one observes in the complex composite system that is a proton or a neutron or any other baryon.

Finally, although (9.21), if it represents a baryon, only does so in the zero-perturbation, no-recursion limit, it is important to ask whether there is anything about this limit that is observed in nature. Put differently, while cutting off the perturbations at the zeroth recursive order may see arbitrary, it is the only order beside infinite order that would seem to have some distinctive claim to not being arbitrary. And so we raise the question whether there are any phenomena observed in nuclear or particle physics which manifest the linear, non-perturbative behavior of the  $\text{Tr}\Sigma P'^{\sigma\mu\nu} \left( (0) \right)_0$  baryon (9.21)? To use an analogy, although gravitation is a highly non-linear theory, we do observe certain aspects of the linear behavior of gravitation theory in the real world, namely, whenever we observe what was first discovered by Kepler and Newton. So while we would most certainly need to describe the complete proton and neutron and other baryons without removing the perturbations from (9.2) a.k.a. (9.3), we should also look to see if certain aspects of nuclear behavior that might be very-definitively described by the “linear approximation” (9.21).

In this regard,  $F_{\text{eff}\mu\nu} \left( (0) \right)_0$  in (10.4) is very important for pursuing experimental validation, because it does describe what “effectively” net flows in and out of the closed monopole surfaces in the ground state linear theory. Specifically, it is well-known that one can calculate electrodynamic energies from the pure gauge field  $\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F_{\sigma\tau} F^{\sigma\tau}$  by using this in  $E = -\iiint \mathcal{L}_{\text{gauge}} d^3x$ . So one should do a similar exercise using what in non-abelian theory becomes the Lagrangian density  $\mathcal{L}_{\text{gauge}} = -\frac{1}{2} \text{Tr} \left( F_{\sigma\tau} F^{\sigma\tau} \right)$ , using  $F_{\text{eff}\mu\nu} \left( (0) \right)_0$ . If we compare (10.4) which is a trace equation to (9.21) which is another trace equation from which it was

derived, then by backtracking to (9.20), we see that (we have now removed the  $\Sigma$  spin sum designation, which now is taken to be implied):

$$F_{\text{eff}\mu\nu}((0))_0 = -i \begin{pmatrix} \overline{\Psi}_R \gamma_{[\mu} (\not{p}_R - m_R)^{-1} \gamma_{\nu]} \Psi_R & 0 & 0 \\ 0 & \overline{\Psi}_G \gamma_{[\mu} (\not{p}_G - m_G)^{-1} \gamma_{\nu]} \Psi_G & 0 \\ 0 & 0 & \overline{\Psi}_B \gamma_{[\mu} (\not{p}_B - m_B)^{-1} \gamma_{\nu]} \Psi_B \end{pmatrix}. \quad (10.13)$$

This is now a 3x3 matrix expression with all diagonal elements. From this, there are two trace expressions that can be formed. One is  $\text{Tr}(F_{\sigma\tau} F^{\sigma\tau})$  which is what is usually found in the Yang-Mills Lagrangian density. The other is  $\text{Tr}F_{\sigma\tau} \text{Tr}F^{\sigma\tau}$ .

It turns out as the author has detailed in sections 11 and 12 of [15], and greatly expanded upon throughout [16], that the expression (10.13) when used in  $E = -\iiint \mathcal{L}_{\text{gauge}} d^3x$  with a combination of  $\text{Tr}(F_{\sigma\tau} F^{\sigma\tau})$  and  $\text{Tr}F_{\sigma\tau} \text{Tr}F^{\sigma\tau}$  inner and outer products, can be used to retrodict *nuclear binding energies*, including the heretofore unexplained binding energies of the lightest nuclides  ${}^2\text{H}$ ,  ${}^3\text{H}$ ,  ${}^3\text{He}$  and  ${}^4\text{He}$ , as well as the  ${}^{56}\text{Fe}$  binding energy, with parts per  $10^5$  or even  $10^6$  AMU precision, and the neutron minus proton mass difference to *under one part per million AMU*. Note that in general, the trace of a product of two square matrices is *not* the product of traces. The only circumstance in which “trace of a product” equals “product of traces” is when one forms a tensor outer product using  $\text{Tr}(A \otimes B) = \text{Tr}(A)\text{Tr}(B)$ , and as shown in [16] the observed binding energies contain both inner and outer products. This line of development in sections 11 and 12 of [15] and throughout [16] also explains why the per-nucleon binding energy seems to be limited for any nucleus to a maximum of about 8.75 MeV for  ${}^{56}\text{Fe}$ , and yields a dynamical, energy-based understanding of confinement.

While all of the formal understandings of the color symmetries of baryons and mesons and quarks are important, direct experimental validation is even more important. It is the experimental concurrences that can be confirmed starting with (10.13) to perform various energy calculations  $E = -\iiint \mathcal{L}_{\text{gauge}} d^3x$  with  $\text{Tr}(F_{\sigma\tau} F^{\sigma\tau})$  and  $\text{Tr}F_{\sigma\tau} \text{Tr}F^{\sigma\tau}$ , that leads to the direct phenomenological confirmation that the faux magnetic monopoles of non-abelian gauge theory really are baryons including protons and neutrons.

## PART II: QUANTUM YANG-MILLS THEORY

### 11. Quantum Yang-Mills Theory: Exact Analytical Path Integration

Finally, let us make use of the recursion developed in section 8, and particularly the substitution  $G_\mu \rightarrow (\pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau)^{-1} J_\mu$  from (8.20) in lieu of the usual  $G_\mu \rightarrow \delta / \delta J^\mu$ , to perform an *exact analytical* deduction of the quantum path integral associated

with the classical field equation  $-J^\nu = (g^{\nu\sigma} D_\tau D^\tau - D^\sigma D^\nu) G_\sigma$  of (5.15) in order to “prove that for any compact simple gauge group  $G$ , a non-trivial quantum Yang–Mills theory exists on  $\mathbb{R}^4$ ,” see page 6 of Jaffe and Witten’s [6].

In abelian gauge theory, the classical electric charge field equation is of course  $*J = d*dG$  which is an abelian subset equation embedded in (1.12). When fully expanded for a massive boson this becomes the abelian  $-J^\nu = (g^{\nu\sigma} (\partial_\tau \partial^\tau + m^2) - \partial^\sigma \partial^\nu) G_\sigma$  of (5.15). The related action after integration-by-parts is thus  $S(G) = \frac{1}{2} G_\mu (g^{\mu\nu} (\partial_\sigma \partial^\sigma + m^2) - \partial^\mu \partial^\nu) G_\nu + J^\mu G_\mu$ , and this is what is used in the path integral  $Z = \int DG \exp i \int d^4 x S(G) \equiv \exp iW(J)$  to deduce the quantum amplitude  $W(J) = \frac{1}{2} \int (d^4 k / (2\pi)^4) J_\sigma (k_\sigma k^\sigma - m^2 + i\epsilon)^{-1} J^\sigma$  with  $+i\epsilon$  using the contextual reduction that also occurs from the continuity relation  $k_\nu J^\nu = 0$  as reviewed at length in section 6 and 7. If we use the terminal condition  $\pi_0 = (k_\tau k^\tau - m^2 + i\epsilon)^{-1}$  of the (8.20) recursion, then this simplifies to  $W(J) = \frac{1}{2} \int (d^4 k / (2\pi)^4) J_\sigma \pi_0 J^\sigma$ .

In non-abelian gauge theory the classical electric charge field equation is the entirety of (1.12), that is  $*J = D*D G$  which as shown expands to  $-J^\nu = (g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu) G_\sigma$  derived in (5.15). Without going through a detailed exposition of how to derive the associated Lagrangian because this is well-known, it will be appreciated that as the result of this exercise the non-abelian action will found to be:

$$\begin{aligned} S(G, J) &= \int d^4 x \mathcal{L}(G, J) = \int d^4 x \text{Tr} \left[ G_\mu (g^{\mu\nu} (D_\tau D^\tau + m^2) - D^\nu D^\mu) G_\nu + 2J^\tau G_\tau \right] \\ &= \int d^4 x \text{Tr} \left[ G_\mu (g^{\mu\nu} ((\partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau) + m^2) - (\partial^\nu \partial^\mu - iG^\nu \partial^\mu - 2G^\nu G^\mu + G^\nu G^\mu)) G_\nu + 2J^\tau G_\tau \right], \end{aligned} \quad (11.1)$$

where we have also included (5.16) and (5.17).

When we now take the next step of using this action in  $Z = \int DG \exp iS(G, J)$ , there are two new issues that come into play which are not present in the abelian gauge theory. The first is that the non-abelian gauge transformation  $G^\nu \rightarrow G'^\nu = G^\nu + \partial^\nu \theta - i[G^\nu, \theta]$  gives rise to ghost fields due to the introduction of the additional term  $-i[G^\nu, \theta]$  into the integration measure  $DG$  in order to ensure that  $Z \rightarrow Z' = Z$  remains invariant under this gauge transformation, and so we need to employ  $DGDcDc^\dagger$  not just  $DG$  as the integration measure. But the second issue is that even before we get to worrying about ghost fields, it is simply not known, as a mathematical matter, how to use an expression like (11.1) in a path integral to calculate:

$$\begin{aligned}
 Z &= \int DG \exp i \int d^4 x \text{Tr} \left[ G_\mu \left( g^{\mu\nu} \left( D_\tau D^\tau + m^2 \right) - D^\nu D^\mu \right) G_\nu + 2J^\tau G_\tau \right] \\
 &= \int DG \exp i \int d^4 x \text{Tr} \left[ G_\mu \begin{pmatrix} g^{\mu\nu} \left( \left( \partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau \right) + m^2 \right) \\ - \left( \partial^\nu \partial^\mu - iG^\nu \partial^\mu - 2G^\nu G^\mu + G^\mu G^\nu \right) \end{pmatrix} G_\nu + 2J^\tau G_\tau \right]. \quad (11.2)
 \end{aligned}$$

This is because, as will be apparent from studying the lower expression in (11.2), this is a fourth-order polynomial in  $G$ , but known mathematical techniques for calculating integrals of this form use the second order  $\int dx \exp\left(-\frac{1}{2}Ax^2 - Jx\right) = (-2\pi/A)^5 \exp(J^2/2A)$ . Why? Put plainly and simply, it is known how to calculate  $\int dx \exp\left(-\frac{1}{2}Ax^2 - Jx\right)$ , but *not* how to calculate the higher order  $\int dx \exp\left(Bx^4 + Cx^3 - \frac{1}{2}Ax^2 - Jx\right)$ . To date, this is an intractable *mathematics* problem. Normally, of course, the approach is to turn every gauge field inside the configuration space operator  $g^{\mu\nu} \left( D_\tau D^\tau + m^2 \right) - D^\mu D^\nu$  into a current term  $G_\mu \rightarrow \delta / \delta J^\mu$  via  $G_\mu = \delta(G_\nu J^\nu) / \delta J^\mu$  and then use (8.25) to apply  $\exp(-V(\delta / \delta J))$  to  $\exp\left(\frac{1}{2}J \cdot K^{-1} \cdot J\right)$  the latter of which is obtained in the usual way from  $\int dx \exp\left(-\frac{1}{2}Ax^2 - Jx\right) = (-2\pi/A)^5 \exp(J^2/2A)$ .

But now the recursion developed in section 8 gives us a new *mathematical* approach. Now, we are able to use (8.20) to turn every occurrence of  $G$  inside  $g^{\mu\nu} \left( D_\tau D^\tau + m^2 \right) - D^\mu D^\nu$  into a function solely of  $G(J, k)$  via  $G_\mu = \pi_\infty J_\mu = \left( \pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau \right)^{-1} J_\mu$  with the abelian terminal condition  $\pi_0 = \left( k_\tau k^\tau - m^2 + i\varepsilon \right)^{-1}$ . *None of  $G_\mu \rightarrow \pi_\infty J_\mu$  these contains  $G_\mu$ !* So, making this replacement in (11.2), we now have:

$$\begin{aligned}
 Z &= \int DG \exp i \int d^4 x \text{Tr} \left[ G_\mu \left( g^{\mu\nu} \left( D_\tau D^\tau + m^2 \right) - D^\nu D^\mu \right) G_\nu + 2J^\tau G_\tau \right] \\
 &= \int DG \exp i \int d^4 x \text{Tr} \left[ G_\mu \begin{pmatrix} g^{\mu\nu} \left( \left( \partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau \right) + m^2 \right) \\ - \left( \partial^\nu \partial^\mu - iG^\nu \partial^\mu - 2G^\nu G^\mu + G^\mu G^\nu \right) \end{pmatrix} G_\nu + 2J^\tau G_\tau \right] \\
 &= \int DG \exp i \int d^4 x \text{Tr} \left[ G_\mu \begin{pmatrix} g^{\mu\nu} \left( \left( \partial_\tau \partial^\tau - i\pi_\infty J_\tau \partial^\tau - \pi_\infty J_\tau \pi_\infty J^\tau \right) + m^2 \right) \\ - \left( \partial^\nu \partial^\mu - i\pi_\infty J^\nu \partial^\mu - 2\pi_\infty J^\nu \pi_\infty J^\mu + \pi_\infty J^\nu \pi_\infty J^\mu \right) \end{pmatrix} G_\nu + 2J^\tau G_\tau \right]. \quad (11.3) \\
 &\stackrel{i\partial \rightarrow k}{\Rightarrow} \int DG \exp i \int d^4 x \text{Tr} \left[ G_\mu \begin{pmatrix} -g^{\mu\nu} \left( \left( k_\tau k^\tau + \pi_\infty J_\tau k^\tau + \pi_\infty J_\tau \pi_\infty J^\tau \right) - m^2 \right) \\ + k^\nu k^\mu + \pi_\infty J^\nu k^\mu + 2\pi_\infty J^\nu \pi_\infty J^\mu - \pi_\infty J^\nu \pi_\infty J^\mu \end{pmatrix} G_\nu + 2J^\tau G_\tau \right]
 \end{aligned}$$

Lo and behold, we have removed all the gauge fields from the path integral except for  $G_\mu(\dots^{\mu\nu})G_\nu$  and  $J^\tau G_\tau$ . This leaves us with the tractable quadratic form

$\int dx \exp\left(-\frac{1}{2} Ax^2 - Jx\right) = (-2\pi/A)^{5/2} \exp(J^2/2A)$ . So we can integrate (11.3) analytically and exactly, so long as we know the inverse for  $(\dots)^{\mu\nu} = g^{\mu\nu} (D_\tau D^\tau + m^2) - D^\mu D^\nu$  or any of its other variants in (11.3). But this, of course, was a central focus of what we studied in section 6 and 7, and this was one of the reasons we studied this so closely. Particularly, for the field equation  $-J^\nu = \left(g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu\right) G_\sigma$ , as seen in (8.19), with the context afforded by the continuity relation  $D_\sigma J^\sigma = 0$ , the inverse solution is simply  $G_\mu = \pi_\infty J_\mu$ . So we recognize immediately that the exact analytical solution to (11.3) is:

$$\begin{aligned} Z &= \int DG \exp i \int d^4x \text{Tr} \left[ G_\mu \left( g^{\mu\nu} (D_\tau D^\tau + m^2) - D^\nu D^\mu \right) G_\nu + 2J^\tau G_\tau \right] \\ &= \int DG \exp i \int d^4x \text{Tr} \left[ G_\mu \left( \begin{array}{l} -g^{\mu\nu} \left( (k_\tau k^\tau + \pi_\infty J_\tau k^\tau + \pi_\infty J_\tau \pi_\infty J^\tau) - m^2 \right) \\ + k^\nu k^\mu + \pi_\infty J^\nu k^\mu + 2\pi_\infty J^\nu \pi_\infty J^\mu - \pi_\infty J^\mu \pi_\infty J^\nu \end{array} \right) G_\nu + 2J^\tau G_\tau \right]. \quad (11.4) \\ &= \exp i \int \left( d^4k / (2\pi)^4 \right)^{1/2} \text{Tr} (J_\sigma \pi_\infty J^\sigma) \equiv \exp i W(J) \end{aligned}$$

This, again, is an *exact analytical solution*. Expressed directly in terms of the amplitude and using (8.18), this means that:

$$\left\{ \begin{array}{l} (2\pi)^4 W(J) = \int d^4k \text{Tr} (J_\sigma \pi_\infty J^\sigma) = \int d^4k \text{Tr} \left( J_\sigma \left( \pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau \right)^{-1} J^\sigma \right) \\ \pi_0 = \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} \end{array} \right. . \quad (11.5)$$

If it is desired to see explicitly how this gives us the non-linear propagator and current and momentum terms that we expect to find in a Yang-Mills path integral, it suffices, just for illustration, to examine the amplitude  $W(J)_2$  for a second-order recursion, using the terminal condition  $\pi_0 = \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1}$ . This is (cf. (8.5)):

$$\begin{aligned}
 (2\pi)^4 W(J)_2 &= \int d^4 k \text{Tr} [J_\sigma \pi_2 J^\sigma] = \int d^4 k \text{Tr} [J_\sigma (\pi_0^{-1} + \pi_1 J_\tau k^\tau + \pi_1 J_\tau \pi_1 J^\tau)^{-1} J^\sigma] \\
 &= \int d^4 k \text{Tr} \left[ J_\sigma \left( \begin{array}{c} \pi_0^{-1} + (\pi_0^{-1} + \pi_0 J_\tau k^\tau + \pi_0 J_\tau \pi_0 J^\tau)^{-1} J_\tau k^\tau \\ + (\pi_0^{-1} + \pi_0 J_\tau k^\tau + \pi_0 J_\tau \pi_0 J^\tau)^{-1} J_\tau (\pi_0^{-1} + \pi_0 J_\tau k^\tau + \pi_0 J_\tau \pi_0 J^\tau)^{-1} J^\tau \end{array} \right)^{-1} J^\sigma \right] \\
 &= \int d^4 k \text{Tr} \left[ J_\sigma \left( \begin{array}{c} k_\tau k^\tau - m^2 + i\mathcal{E} \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} + \frac{J_\tau k^\tau}{k_\tau k^\tau - m^2 + i\mathcal{E}} + \frac{J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right)^{-1} J_\tau k^\tau \\ + \left( k_\tau k^\tau - m^2 + i\mathcal{E} + \frac{J_\tau k^\tau}{k_\tau k^\tau - m^2 + i\mathcal{E}} + \frac{J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right)^{-1} J_\tau \\ \times \left( k_\tau k^\tau - m^2 + i\mathcal{E} + \frac{J_\tau k^\tau}{k_\tau k^\tau - m^2 + i\mathcal{E}} + \frac{J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right)^{-1} J^\tau \end{array} \right)^{-1} J^\sigma \right] . \quad (11.6)
 \end{aligned}$$

With this being only the second-order recursion, it will be appreciated how this will expand rapidly in a highly-nonlinear way to include all orders of  $J$ ,  $k$ ,  $m$  and  $+i\mathcal{E}$ , right through infinity for  $W(J) \equiv W(J)_\infty$ . For doing practical calculations, including those with computers, one can use expressions with a few more orders of recursion to obtain results fairly close to those that would be obtained upon an infinite recursion, assuming convergence. So let us now look at that.

We can ascertain the general trend toward convergence or divergence simply using the  $n=1$  recursive order, because as we have seen here and in section 8, the basic pattern for higher orders is already established at first order. For  $W(J)_1$  we have:

$$\begin{aligned}
 (2\pi)^4 W(J)_1 &= \int d^4 k \text{Tr} [J_\sigma \pi_1 J^\sigma] = \int d^4 k \text{Tr} [J_\sigma (\pi_0^{-1} + \pi_0 J_\tau k^\tau + \pi_0 J_\tau \pi_0 J^\tau)^{-1} J^\sigma] \\
 &= \int d^4 k \text{Tr} \left[ J_\sigma \left( k_\tau k^\tau - m^2 + i\mathcal{E} + \frac{J_\tau k^\tau}{k_\tau k^\tau - m^2 + i\mathcal{E}} + \frac{J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right)^{-1} J^\sigma \right] . \quad (11.7) \\
 &= \int d^4 k \text{Tr} \left[ J_\sigma \left( \frac{(k_\tau k^\tau - m^2 + i\mathcal{E})^3 + (k_\tau k^\tau - m^2 + i\mathcal{E}) J_\tau k^\tau + J_\tau J^\tau}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right)^{-1} J^\sigma \right]
 \end{aligned}$$

Given  $J^\mu = \lambda_{AB}^i J^{i\mu} = J_{AB}^\mu$  for  $\text{SU}(3) \times \text{U}(1)$  has dimension 3 at the same time that  $(k_\tau k^\tau - m^2 + i\mathcal{E})^3 = \delta_{AB} (k_\tau k^\tau - m^2 + i\mathcal{E})^3$  sits on the 3x3 diagonal, a naive look at (11.7) tells us that the dominant term in the numerator will be  $(k_\tau k^\tau - m^2 + i\mathcal{E})^3$  for  $J_\tau k^\tau < (k_\tau k^\tau - m^2 + i\mathcal{E})^2$

and  $J_\tau J^\tau < (k_\tau k^\tau - m^2 + i\mathcal{E})^3$ . But when considering the matrix equations, a more precise statement would say that  $-(k_\tau k^\tau - m^2 + i\mathcal{E})^3$  represents eigenvalues of  $\delta \equiv (k_\tau k^\tau - m^2 + i\mathcal{E})J_\tau k^\tau + J_\tau J^\tau$ , and will dominate when these eigenvalues are larger rather than smaller. In the case where  $J_\tau k^\tau$  and  $J_\tau J^\tau$  are small and substantially negligible in relation to  $(k_\tau k^\tau - m^2 + i\mathcal{E})^2$  and  $(k_\tau k^\tau - m^2 + i\mathcal{E})^3$ , the overall expression (11.7) will be:

$$(2\pi)^4 W(J)_1 = \int d^4 k \text{Tr} \left[ J_\sigma \left( \frac{(k_\tau k^\tau - m^2 + i\mathcal{E})^3 + \delta}{(k_\tau k^\tau - m^2 + i\mathcal{E})^2} \right)^{-1} J^\sigma \right] \cong \int d^4 k \text{Tr} \left[ J_\sigma \frac{1}{k_\tau k^\tau - m^2 + i\mathcal{E}} J^\sigma \right], \quad (11.8)$$

which is of the same form as the abelian propagator. So the solution (11.6) would appear to be fully convergent (or, at least no more divergent than the abelian path integral) for  $J_\tau k^\tau$  and  $J_\tau J^\tau$  which are small in comparison to eigenvalues which are specific powers of  $k_\tau k^\tau - m^2 + i\mathcal{E}$ . It is also worth noting that the positive sign in (11.8) is the quantum field explanation for why the electromagnetic force between like charges is *repulsive*. In electrodynamics this means that like charges repel. In chromodynamics this means that charges of the same color repel, e.g., red repels red, etc., which is another way of viewing Exclusion.

Finally, because (11.5) is an exact analytical calculation using a closed recursive kernel, per [6] page 6, this “prove(s) that for any compact simple gauge group G, a non-trivial quantum Yang–Mills theory exists on  $\mathbb{R}^4$ .”

## 12. The Baryon Candidate Lagrangian Density and Action

The quantum path integral for the non-abelian  $W(J)$  derived in (11.5) is for the current density  $J^\mu = \lambda_{AB}^i J^{i\mu} = \lambda_{AB}^i \bar{\Psi}_C \lambda_{CD}^i \gamma^\mu \Psi_D \equiv \bar{\Psi} \gamma^\mu \Psi$  of a dimension-3 column vector of fermion wavefunctions which we were required by Exclusion to introduce at (9.9) and which we named with Red, Green and Blue eigenstates. As stated between (10.5) and (10.6), what is arbitrary are the names. What is not arbitrary is that we require three such names. At (9.21) and in the opening discussion of section 10 we showed that the faux monopole density  $P'^{\sigma\mu\nu}$  has the colorless SU(3)-invariant antisymmetric wavefunction  $R \wedge G \wedge B$  of a baryon, and at (10.5) we found that the only fields which net flow across closed surfaces of these monopoles have the SU(3)-invariant symmetric wavefunction  $\overline{RR} + \overline{GG} + \overline{BB}$  of a meson. It was on this basis, as well as the topological discussion in section 10 which led to additional flavored associations of these monopoles with protons and neutrons, that we have identified these faux monopoles  $P'^{\sigma\mu\nu}$ , at least at the classical level, with a baryon density, and thus the colored fermions with quarks. We shall sometimes refer to these as “candidate” baryons and quarks.

As stated at the end of section 4, it is certainly unrealistic to expect that a classical-only treatment of baryons based on Yang-Mills magnetic monopoles will explain *all* of the observed

phenomenology of baryons. Nonetheless, finding the complete and correct quantum description of baryons begins by finding and fleshing out, the right classical theory to quantize. Sections 10 and 11 advance the thesis that the right classical theory is one in which these non-abelian faux monopoles  $P'^{\sigma\mu\nu}$  – which arise as the underlying density behind a non-vanishing Yang-Mills magnetic field flux  $\oint\!\!\!\oint F \neq 0$  which we showed in section 11 has a  $\overline{RR} + \overline{GG} + \overline{BB}$  meson wavefunction – are taken to be classical baryons. So having found and suitably developed a candidate for the right classical theory to quantize, it is now time to quantize that theory.

As made clear in (5.9), the classical theory can be summarized in two Gauss' / Stokes' integral equations which are the non-abelian generalization of Maxwell's equations. One is for a net flow  $\oint\!\!\!\oint *F \neq 0$  of a non-abelian electric field across closed surfaces, sourced by the elementary, non-composite electric charge and current density three-form  $*J = d*dG - iG*dG - [G, *[G, G]]$ , see the third line of (5.7). The other is for a net flow  $\oint\!\!\!\oint F \neq 0$  of a non-abelian magnetic field across closed surfaces sourced by the faux magnetic density three-form  $P' = -idGG = -id[G, G]$  which arises as a non-elementary, composite function of the gauge fields  $G$  in turn sourced by  $*J$ . We have already obtained the highly-non-linear quantum amplitude  $W(J)$  in (11.5) which characterizes the quantum interactions of the non-abelian electric currents  $J^\mu = \overline{\Psi}\gamma^\mu\Psi$  and thus, of the quark eigenstates  $\psi_R, \psi_G, \psi_B$ . That is,  $W(J)$  is the quantum amplitude / potential energy, in momentum space, for quark interactions. We now seek  $W(P')$  which, if  $P' = -idGG = -id[G, G]$  is indeed a baryon density, would be the quantum amplitude for *baryon interactions*. Because we have shown later in section 10 that these classical baryons can also be made topologically stable and given the flavored properties of protons and neutrons, finding  $W(P')$  would lay the foundation for developing a quantum field theory description of nuclear interactions. So let us proceed.

The classical field equation for a Lagrangian density  $\mathcal{L}$  with a gauge field  $G_\nu$  is of course given by the Euler-Lagrange equation:

$$0 = \partial_\sigma \left( \frac{\partial}{\partial(\partial_\sigma G_\nu)} \mathcal{L} \right) - \frac{\partial}{\partial G_\nu} \mathcal{L}. \quad (12.1)$$

The first step is to obtain the Lagrangian density  $\mathcal{L}(P')$  for the monopole density  $P' = -idGG = -id[G, G]$ , that is, to find an  $\mathcal{L}(P')$  which causes (12.1) to be one and the same as the classical field equation  $P' = -idGG = -id[G, G]$ . We implicitly did the same thing prior to (11.1) for  $\mathcal{L}(J)$  and the classical field equation  $-J^\nu = \left( g^{\nu\sigma} (D_\tau D^\tau + m^2) - D^\sigma D^\nu \right) G_\sigma$  without going through the detailed exercise of doing so via (12.1) because it is already well-known how to do this. But for  $P' = -idGG = -id[G, G]$  we shall trace through the entire exercise because

this does not appear to be known and certainly it does not appear that  $P' = -idGG = -id[G, G]$  has ever before been regarded as the classical field equation for a baryon. The second step is to then write down the monopole / baryon action  $S(G, P') = \int d^4x \mathcal{L}(G, P')$  parallel to what we did in (11.1) for the electric charge / quark field action. And the final step is to do the path integral  $Z = \int DG \exp iS(G, P')$  and thus deduce the quantum amplitude  $W(P')$  for the baryon field as we did in the balance of section 11 to find  $W(J)$ . The advantage we now have in view of section 11, is knowing that a clever use of recursive kernel  $G_\mu = \pi_\infty J_\mu$  from (8.20) will enable us to replace some of the gauge fields  $G_\mu$  with current densities  $J_\mu$  to maintain a quadratic form expression which can then be integrated exactly and analytically. And, given this, and studying (11.5), we anticipate that the amplitude we obtain will really be  $W(P'(J(\psi_R, \psi_G, \psi_B), k, m, \varepsilon))$ , which will describe the quantum interactions of a baryon in a fashion which includes their quark fields  $\psi_R, \psi_G, \psi_B$  as well as their gauge fields as represented  $k^\tau$ .

For  $-J^\mu = (g^{\mu\nu}(D_\tau D^\tau + m^2) - D^\nu D^\mu)G_\nu$  which is the electric charge / quark current density field equation first obtained in (5.15), the Lagrangian density is  $\mathcal{L}(G, J) = \text{Tr} \left[ G_\mu (g^{\mu\nu}(D_\tau D^\tau + m^2) - D^\nu D^\mu) G_\nu + 2J^\tau G_\tau \right]$  as seen in (11.1). These both contain the identical configuration space operator  $g^{\mu\nu}(D_\tau D^\tau + m^2) - D^\mu D^\nu$ . The only difference is that in the action this operator is sandwiched between two gauge fields  $G_\mu(\dots)G_\nu$ , there is the additional term  $J^\tau G_\tau$  which contains the source current, and because we are using non-abelian gauge theory, a trace (Tr) is required to maintain  $\mathcal{L}$  as a scalar. Earlier, way back in (3.4), we derived the four-vector current density  $*P'^\alpha$  for first-rank dual of the faux monopole  $P'_{\sigma\mu\nu}$  which we now see from (9.21) has the classical color symmetries of a baryon. In (3.4) the calculation was illustrative. But now, a calculation akin to (3.4) is an essential step to deriving  $\mathcal{L}(P')$ .

Specifically, we keep in mind that field equation  $-J^\mu = (g^{\mu\nu}(D_\tau D^\tau + m^2) - D^\nu D^\mu)G_\nu$  which we abbreviate as  $-J^\mu = (\dots)G_\nu$  with  $\dots = g^{\mu\nu}(D_\tau D^\tau + m^2) - D^\nu D^\mu$  maps to Lagrangian density  $\mathcal{L}(G, J) = \text{Tr} \left[ G_\mu (g^{\mu\nu}(D_\tau D^\tau + m^2) - D^\mu D^\nu) G_\nu + 2J^\tau G_\tau \right]$  for the electric current density  $J$ , which  $\mathcal{L}$  we abbreviate as  $\mathcal{L}(G, J) = \text{Tr} \left[ G_\mu(\dots)G_\nu + 2J^\tau G_\tau \right]$ . We also are mindful that it is easier to start with a Lagrangian and find its field equation rather than the other way around. We note that (2.11) contains several way to write  $P$ , but that the only terms without commutators are  $dGG$  and  $GdG$ . Noting the identities  $-i[G, dG] = -2iGdG + idGG$  and  $d[G, G] = dGG - GdG$  which were found in (2.8) and (2.9), we see that any expression containing a commutator can always be separated into an expression with some linear combination of terms  $AdGG + BGdG$  where  $A$  and  $B$  are constant numbers. We also know that  $GdG = 0$  from (12.11), but for the moment we will include terms with  $GdG$  to establish formal concurrence between the

Lagrangian density and the classical monopole field equation and only apply this at the very end to reduce our results.

Thus, with all of this in mind, we now fashion a “test” Lagrangian density for the pure field term absent any sources in which  $G_\tau$  left-multiplies  $dGG \Leftrightarrow * \partial^{[\mu} G^{\tau]} G_\mu$  and  $GdG \Leftrightarrow G_\mu * \partial^{[\mu} G^{\tau]}$ , while also including a term  $G_\tau * P^\tau$  with the dual  $*P^\alpha = \frac{1}{3!} \epsilon^{\sigma\mu\nu\alpha} P_{\sigma\mu\nu}$  of the original (not faux) monopole, all with the required trace, of the form:

$$\mathcal{L}_{\text{test}} \equiv -iA\text{Tr}\left(G_\tau * \partial^{[\mu} G^{\tau]} G_\mu\right) - iB\text{Tr}\left(G_\tau G_\mu * \partial^{[\mu} G^{\tau]}\right) + C\text{Tr}\left(G_\tau * P^\tau\right). \quad (12.2)$$

Of course, because  $*P^\alpha = \frac{1}{3!} \epsilon^{\sigma\mu\nu\alpha} P_{\sigma\mu\nu} = 0$  via the Jacobian identity (2.4), in the last term  $G_\tau * P^\tau = 0$ . But this is still needed because the faux monopole which appears to be a classical baryon density is  $P' = -id[G, G] = -idGG$ . Thus, we can eventually use this last term to introduce  $P'$  via  $0 = P = id[G, G] - idGG = -P' - idGG = id[G, G] + P'$ . We also note that  $B\text{Tr}\left(G_\tau G_\mu * \partial^{[\mu} G^{\tau]}\right) = 0$  because  $G_\mu * \partial^{[\mu} G^{\tau]}$  is an alternate way to write  $GdG = 0$ , but for now, as stated, we also carry this as a term without yet setting it to zero. In (12.2) above,  $A$ ,  $B$ , and  $C$  are unknown constant numbers that we shall now determine by using (12.2) in (12.1) to reproduce the classical field equation  $P = id[G, G] - idGG = 0$  of (2.11). In other words,  $P = id[G, G] - idGG = 0$  is the target equation we seek to derive by placing (12.2) into (12.1) and then choosing  $A$ ,  $B$ , and  $C$  to math up (12.1) to  $P = id[G, G] - idGG = 0$ .

For the latter term in the Euler-Lagrange (12.1) we may use (12.2) to first calculate:

$$\begin{aligned} -\frac{\partial}{\partial G_\nu} \mathcal{L}_{\text{test}} &= -\frac{\partial}{\partial G_\nu} \left[ -iA\text{Tr}\left(G_\tau * \partial^{[\mu} G^{\tau]} G_\mu\right) - iB\text{Tr}\left(G_\tau G_\mu * \partial^{[\mu} G^{\tau]}\right) + C\text{Tr}\left(G_\tau * P^\tau\right) \right] \\ &= iA\text{Tr}\left(\frac{\partial G_\tau}{\partial G_\nu} * \partial^{[\mu} G^{\tau]} G_\mu + G_\tau * \partial^{[\mu} G^{\tau]} \frac{\partial G_\mu}{\partial G_\nu}\right) \\ &\quad + iB\text{Tr}\left(\frac{\partial G_\tau}{\partial G_\nu} G_\mu * \partial^{[\mu} G^{\tau]} + G_\tau \frac{\partial G_\mu}{\partial G_\nu} * \partial^{[\mu} G^{\tau]}\right) - C\text{Tr}\left(\frac{\partial G_\tau}{\partial G_\nu} * P^\tau\right) \quad (12.3) \\ &= iA\text{Tr}\left(\delta^\nu_\tau * \partial^{[\mu} G^{\tau]} G_\mu + G_\tau * \partial^{[\mu} G^{\tau]} \delta^\nu_\mu\right) + iB\text{Tr}\left(\delta^\nu_\tau G_\mu * \partial^{[\mu} G^{\tau]} + G_\tau \delta^\nu_\mu * \partial^{[\mu} G^{\tau]}\right) - C\text{Tr}\left(\delta^\nu_\tau * P^\tau\right) \\ &= iA\text{Tr}\left(*\partial^{[\mu} G^{\nu]} G_\mu + G_\tau * \partial^{[\nu} G^{\tau]}\right) + iB\text{Tr}\left(G_\mu * \partial^{[\mu} G^{\nu]} + G_\tau * \partial^{[\nu} G^{\tau]}\right) - C\text{Tr} * P^\nu \\ &= iA\text{Tr}\left(*\partial^{[\sigma} G^{\nu]} G_\sigma\right) + iA\text{Tr}\left(G_\sigma * \partial^{[\nu} G^{\sigma]}\right) - C\text{Tr} * P^\nu \end{aligned}$$

For the operand of the former term in (12.1) we use (12.2) together with second rank duality in the form  $*\partial^{[\mu} G^{\tau]} = \frac{1}{2!} \epsilon^{\alpha\beta\mu\tau} \partial_{[\alpha} G_{\beta]}$ , being very careful with signs, to calculate:

$$\begin{aligned}
 \frac{\partial}{\partial(\partial_\sigma G_\nu)} \mathcal{L}_{\text{test}} &= \frac{\partial}{\partial(\partial_\sigma G_\nu)} \left( -iA\text{Tr}(G_\tau * \partial^{[\mu} G^{\tau]} G_\mu) - iB\text{Tr}(G_\tau G_\mu * \partial^{[\mu} G^{\tau]}) + C\text{Tr}(G_\tau * P^\tau) \right) \\
 &= -\frac{1}{2!} i \varepsilon^{\alpha\beta\mu\tau} \frac{\partial}{\partial(\partial_\sigma G_\nu)} \left( A\text{Tr}(G_\tau \partial_{[\alpha} G_{\beta]} G_\mu) + B\text{Tr}(G_\tau G_\mu \partial_{[\alpha} G_{\beta]}) \right) \\
 &= -\frac{1}{2!} i \varepsilon^{\alpha\beta\mu\tau} \frac{\partial}{\partial(\partial_\sigma G_\nu)} \left( A\text{Tr}(G_\tau (\partial_\alpha G_\beta - \partial_\beta G_\alpha) G_\mu) + B\text{Tr}(G_\tau G_\mu (\partial_\alpha G_\beta - \partial_\beta G_\alpha)) \right) \\
 &= -\frac{1}{2!} i \varepsilon^{\alpha\beta\mu\tau} \left( A\text{Tr} \left( G_\tau \left( \frac{\partial(\partial_\alpha G_\beta)}{\partial(\partial_\sigma G_\nu)} - \frac{\partial(\partial_\beta G_\alpha)}{\partial(\partial_\sigma G_\nu)} \right) G_\mu \right) + B\text{Tr} \left( G_\tau G_\mu \left( \frac{\partial(\partial_\alpha G_\beta)}{\partial(\partial_\sigma G_\nu)} - \frac{\partial(\partial_\beta G_\alpha)}{\partial(\partial_\sigma G_\nu)} \right) \right) \right) \\
 &= -\frac{1}{2!} i \varepsilon^{\alpha\beta\mu\tau} \left( A\text{Tr}(G_\tau (\delta^\sigma_\alpha \delta^\nu_\beta - \delta^\sigma_\beta \delta^\nu_\alpha) G_\mu) + B\text{Tr}(G_\tau G_\mu (\delta^\sigma_\alpha \delta^\nu_\beta - \delta^\sigma_\beta \delta^\nu_\alpha)) \right) \\
 &= -\frac{1}{2!} i \left( A\text{Tr}(\varepsilon^{\sigma\nu\mu\tau} G_\tau G_\mu - \varepsilon^{\nu\sigma\mu\tau} G_\tau G_\mu) + B\text{Tr}(\varepsilon^{\sigma\nu\mu\tau} G_\tau G_\mu - \varepsilon^{\nu\sigma\mu\tau} G_\tau G_\mu) \right) \\
 &= i \left( A\text{Tr} \left( \frac{1}{2!} \varepsilon^{\sigma\nu\mu\tau} [G_\tau, G_\mu] \right) + B\text{Tr} \left( \frac{1}{2!} \varepsilon^{\sigma\nu\mu\tau} [G_\tau, G_\mu] \right) \right) \\
 &= iA\text{Tr} * [G^\sigma, G^\nu] + iB\text{Tr} * [G^\sigma, G^\nu]
 \end{aligned} \tag{12.4}$$

This of course means that:

$$\partial_\sigma \left( \frac{\partial}{\partial(\partial_\sigma G_\nu)} \mathcal{L}_{\text{test}} \right) = iA\text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + iB\text{Tr} \partial_\sigma * [G^\sigma, G^\nu]. \tag{12.5}$$

Finally, combining (12.5) and (12.3) into (12.1) yields the classical field equation:

$$\begin{aligned}
 0 &= \partial_\sigma \left( \frac{\partial}{\partial(\partial_\sigma G_\nu)} \mathcal{L}_{\text{test}} \right) - \frac{\partial}{\partial G_\nu} \mathcal{L}_{\text{test}} \\
 &= iA\text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + iB\text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + iA\text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma) + iA\text{Tr}(G_\sigma * \partial^{[\nu} G^{\sigma]}) - C\text{Tr} * P^\nu
 \end{aligned} \tag{12.6}$$

which we then rewrite with  $*P^\nu = 0$  on the left using the Jacobian zero of (2.4) to obtain:

$$C\text{Tr} * P^\nu = Ai\text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + Bi\text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + Ai\text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma) + Ai\text{Tr}(G_\sigma * \partial^{[\nu} G^{\sigma]}) = 0 \tag{12.7}$$

But in (2.11) we uncovered the identity  $0 = GdG$  which in tensor form reads:

$$G_\sigma \partial_{[\mu} G_{\nu]} + G_\mu \partial_{[\nu} G_{\sigma]} + G_\nu \partial_{[\sigma} G_{\mu]} = 0 \tag{12.8}$$

Multiplying through by  $\frac{1}{3!} \varepsilon^{\alpha\sigma\mu\nu}$  to take the first rank dual then yields:

$$\begin{aligned}
0 &= *(G_\sigma \partial_{[\mu} G_{\nu]} + G_\mu \partial_{[\nu} G_{\sigma]} + G_\nu \partial_{[\sigma} G_{\mu]}) = \frac{1}{3!} \varepsilon^{\sigma\mu\nu\alpha} (G_\sigma \partial_{[\mu} G_{\nu]} + G_\mu \partial_{[\nu} G_{\sigma]} + G_\nu \partial_{[\sigma} G_{\mu]}) \\
&= \frac{1}{3} (G_\sigma \frac{1}{2} \varepsilon^{\sigma\mu\nu\alpha} \partial_{[\mu} G_{\nu]} + G_\mu \frac{1}{2} \varepsilon^{\sigma\mu\nu\alpha} \partial_{[\nu} G_{\sigma]} + G_\nu \frac{1}{2} \varepsilon^{\sigma\mu\nu\alpha} \partial_{[\sigma} G_{\mu]}) \\
&= \frac{1}{3} (G_\sigma * \partial^{[\sigma} G^{\alpha]} + G_\mu * \partial^{[\mu} G^{\alpha]} + G_\nu * \partial^{[\nu} G^{\alpha]}) = G_\sigma * \partial^{[\sigma} G^{\alpha]} = -G_\sigma * \partial^{[\alpha} G^{\sigma]}
\end{aligned} \tag{12.9}$$

So the final term in (12.7) zeroes out and the overall field equation now becomes:

$$C\text{Tr} * P^\nu = A i \text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + B i \text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + A i \text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma) = 0 \tag{12.10}$$

Now we need to match this up against the target field equation  $P = id[G, G] - idGG = 0$  of (2.11) which we write in tensor form as:

$$P_{\sigma\mu\nu} = i(\partial_\sigma [G_\mu, G_\nu] + \partial_\mu [G_\nu, G_\sigma] + \partial_\nu [G_\sigma, G_\mu]) - i(\partial_{[\sigma} G_{\mu]} G_\nu + \partial_{[\mu} G_{\nu]} G_\sigma + \partial_{[\nu} G_{\sigma]} G_\mu) = 0 \tag{12.11}$$

We then use  $*P^\alpha = \frac{1}{3!} \varepsilon^{\sigma\mu\nu\alpha} P_{\sigma\mu\nu}$  to write the first rank dual as:

$$\begin{aligned}
*P^\alpha &= \frac{1}{3!} \varepsilon^{\sigma\mu\nu\alpha} P_{\sigma\mu\nu} \\
&= i \frac{1}{3} (\partial_\sigma \frac{1}{2!} \varepsilon^{\sigma\mu\nu\alpha} [G_\mu, G_\nu] + \partial_\mu \frac{1}{2!} \varepsilon^{\sigma\mu\nu\alpha} [G_\nu, G_\sigma] + \partial_\nu \frac{1}{2!} \varepsilon^{\sigma\mu\nu\alpha} [G_\sigma, G_\mu]) \\
&\quad - i \frac{1}{3} (\frac{1}{2!} \varepsilon^{\sigma\mu\nu\alpha} \partial_{[\sigma} G_{\mu]} G_\nu + \frac{1}{2!} \varepsilon^{\sigma\mu\nu\alpha} \partial_{[\mu} G_{\nu]} G_\sigma + \frac{1}{2!} \varepsilon^{\sigma\mu\nu\alpha} \partial_{[\nu} G_{\sigma]} G_\mu) \\
&= i \frac{1}{3} (\partial_\sigma * [G^\sigma, G^\alpha] + \partial_\mu * [G^\mu, G^\alpha] + \partial_\nu * [G^\nu, G^\alpha]) - i \frac{1}{3} (*\partial^{[\nu} G^{\alpha]} G_\nu + *\partial^{[\sigma} G^{\alpha]} G_\sigma + *\partial^{[\mu} G^{\alpha]} G_\mu) \\
&= i \partial_\sigma * [G^\sigma, G^\alpha] - i * \partial^{[\sigma} G^{\alpha]} G_\sigma = 0
\end{aligned} \tag{12.12}$$

Taking the trace and renaming the free index, this is:

$$\text{Tr} * P^\nu = i \text{Tr} (\partial_\sigma * [G^\sigma, G^\nu]) - i \text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma) = 0 \tag{12.13}$$

Now we compare this with (12.7) which was derived from the Euler-Lagrange equation (12.1) and the Lagrangian density (12.2). First we set  $C = 1$  to match the left side. Thereafter, on the right, we must match  $A i \text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma)$  to  $-i \text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma)$  which tells us that  $A = -1$ . So (12.10) becomes:

$$\text{Tr} * P^\nu = -i \text{Tr} \partial_\sigma * [G^\sigma, G^\nu] + B i \text{Tr} \partial_\sigma * [G^\sigma, G^\nu] - i \text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma) = 0 \tag{12.14}$$

To complete the match, we must set  $B = +2$ . Then (12.14) becomes:

$$\text{Tr} * P^\nu = i \text{Tr} \partial_\sigma * [G^\sigma, G^\nu] - i \text{Tr} (*\partial^{[\sigma} G^{\nu]} G_\sigma) = 0 \tag{12.15}$$

which perfectly matches (12.13). So we use these findings in (12.2) to write the Lagrangian density as:

$$\mathcal{L} = i\text{Tr}\left(G_\tau * \partial^{[\mu} G^{\tau]} G_\mu\right) - i2\text{Tr}\left(G_\tau G_\mu * \partial^{[\mu} G^{\tau]}\right) + \text{Tr}\left(G_\tau * P^\tau\right). \quad (12.16)$$

Finally, having established formal equivalence of the Lagrangian density with the monopole field equation via Euler-Lagrange, we now take advantage of  $0 = GdG$  which is written as  $G_\mu * \partial^{[\nu} G^{\tau]} = 0$  in (12.9), and we also use the Jacobian-based  $*P^\tau = 0$ , to further reduce this Lagrangian density to:

$$\mathcal{L} = i\text{Tr}\left(G_\tau * \partial^{[\mu} G^{\tau]} G_\mu\right) + \text{Tr}\left(G_\tau * P^\tau\right) = i\text{Tr}\left(G_\tau * \partial^{[\mu} G^{\tau]} G_\mu\right). \quad (12.17)$$

The above, when used in the Euler-Lagrange equation (12.1), will indeed reproduce the classical field equation  $P = id[G, G] - idGG = 0$  of (2.11), or more precisely, will reproduce its trace equation  $\text{Tr}P = i\text{Tr}\left(d[G, G]\right) - i\text{Tr}\left(dGG\right) = 0$ .

Yet, (12.17) is not our final result, because it still contains the original monopole  $*P^\tau$  which is zero, and we wish to now inject the faux monopole density  $*P'^\tau$  which is the dual of the candidate baryon density  $P'_{\sigma\mu\nu}$ . Now, as noted following (12.2), we will want to apply the relation  $0 = P = id[G, G] - idGG = -P' - idGG = id[G, G] + P'$  to get  $P'$  into the Lagrangian density (12.17). Were we to use  $P = -P' - idGG$  in (12.17), this would offset the  $G_\tau * \partial^{[\mu} G^{\tau]} G_\mu$  term and we would lose important information about the gauge fields, just as if we had stopped at  $P = 0$  in (2.11) rather than proceeding to make use of  $P = id[G, G] - idGG = 0$  to develop the identity  $P' = -id[G, G] = -idGG$  from which we learned a lot more including in (9.21) that  $P'$  has the color symmetries of a baryon and that  $\oint\!\!\!\oint F = \iiint P' = -i\iiint dGG = -i\oint\!\!\!\oint [G, G] \neq 0$  has the color symmetries of a meson. So we expand  $P = id[G, G] + P' = 0$  into tensor form and use this in (12.17) to create a mix of both  $id[G, G]$  and  $idGG$  terms which has already been fruitful elsewhere. Thus, we expand  $P = id[G, G] + P' = 0$  into:

$$P_{\sigma\mu\nu} = i\left(\partial_\sigma [G_\mu, G_\nu] + \partial_\mu [G_\nu, G_\sigma] + \partial_\nu [G_\sigma, G_\mu]\right) + P'_{\sigma\mu\nu} = 0. \quad (12.18)$$

The dual relation is then:

$$\begin{aligned}
*P^\alpha &= \frac{1}{3!} \mathcal{E}^{\sigma\mu\nu\alpha} P_{\sigma\mu\nu} \\
&= i \frac{1}{3} \left( \partial_\sigma \frac{1}{2} \mathcal{E}^{\sigma\mu\nu\alpha} [G_\mu, G_\nu] + \partial_\mu \frac{1}{2} \mathcal{E}^{\sigma\mu\nu\alpha} [G_\nu, G_\sigma] + \partial_\nu \frac{1}{2} \mathcal{E}^{\sigma\mu\nu\alpha} [G_\sigma, G_\mu] \right) + \frac{1}{3!} \mathcal{E}^{\sigma\mu\nu\alpha} P'_{\sigma\mu\nu} \\
&= i \frac{1}{3} \left( \partial_\sigma * [G^\sigma, G^\alpha] + \partial_\mu * [G^\mu, G^\alpha] + \partial_\nu * [G^\nu, G^\alpha] \right) + *P'^\alpha \\
&= i \partial_\sigma * [G^\sigma, G^\alpha] + *P'^\alpha = 0
\end{aligned} \tag{12.19}$$

We then use (12.19) in (12.17) to finally obtain:

$$\mathcal{L}(G, *P') = \text{Tr} \left( i G_\sigma * \partial^{[\tau} G^{\sigma]} G_\tau + i G_\sigma \partial_\tau * [G^\tau, G^\sigma] + G_\sigma * P'^\sigma \right) = i \text{Tr} \left( G_\sigma * \partial^{[\tau} G^{\sigma]} G_\tau \right). \tag{12.21}$$

This is the Lagrangian density for the Yang-Mills monopole, written in terms of the dual of the candidate baryon density  $P'^\sigma$ . By replicating  $i \text{Tr} \left( G_\sigma * \partial^{[\tau} G^{\sigma]} G_\tau \right)$  all by itself in the final term, we are asserting that  $i G_\sigma \partial_\tau * [G^\tau, G^\sigma] + G_\sigma * P'^\sigma = 0$  i.e., that  $*P'^\sigma = -i \partial_\tau * [G^\tau, G^\sigma]$ , which is  $P' = -id[G, G]$ . Then, because  $-id[G, G] = -idGG$  is also an identity, we end up via this final term implicitly stating that  $0 = P = id[G, G] - idGG$ . And, given that  $*\partial^{[\tau} G^{\sigma]} G_\tau \Leftrightarrow dGG$  and  $\partial_\tau * [G^\tau, G^\sigma] \Leftrightarrow d[G, G]$  are both embedded in (12.21), we see how (12.21) is just a Lagrangian statement of  $P' = -id[G, G] = -idGG$ .

Now let us rework (12.21) a little bit so we can identify its configuration space operator as a first step to calculating the path integral. Here, we use the dual relationships  $*\partial^{[\tau} G^{\sigma]} = \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} \partial_{[\alpha} G_{\beta]}$  and  $*[G^\tau, G^\sigma] = \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} [G_\alpha, G_\beta]$  to write:

$$\begin{aligned}
\mathcal{L}(G, *P') &= \text{Tr} \left( i G_\sigma * \partial^{[\tau} G^{\sigma]} G_\tau + i G_\sigma \partial_\tau * [G^\tau, G^\sigma] + G_\sigma * P'^\sigma \right) \\
&= \text{Tr} \left( \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} \left( i G_\sigma \partial_{[\alpha} G_{\beta]} G_\tau + i G_\sigma \partial_\tau [G_\alpha, G_\beta] \right) + G_\sigma * P'^\sigma \right) \\
&= \text{Tr} \left( \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} \left( i G_\sigma \partial_\alpha G_\beta G_\tau - i G_\sigma \partial_\beta G_\alpha G_\tau + i G_\sigma \partial_\tau (G_\alpha G_\beta) - i G_\sigma \partial_\tau (G_\beta G_\alpha) \right) + G_\sigma * P'^\sigma \right) \\
&= \text{Tr} \left( \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} \left( i G_\sigma \partial_\alpha G_\beta G_\tau - i G_\sigma \partial_\beta G_\alpha G_\tau \right. \right. \\
&\quad \left. \left. + i G_\sigma \partial_\tau G_\alpha G_\beta + i G_\sigma G_\alpha \partial_\tau G_\beta - i G_\sigma \partial_\tau G_\beta G_\alpha - i G_\sigma G_\beta \partial_\tau G_\alpha \right) + G_\sigma * P'^\sigma \right) \tag{12.22} \\
&= \text{Tr} \left( \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} \left( 4i G_\sigma \partial_\alpha G_\beta G_\tau - 2i G_\sigma G_\alpha \partial_\beta G_\tau \right) + G_\sigma * P'^\sigma \right) \\
&= \text{Tr} \left( \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} \left( i G_\sigma \left( 4\partial_\alpha G_\beta - 2G_\alpha \partial_\beta \right) G_\tau \right) + G_\sigma * P'^\sigma \right) \\
&= \text{Tr} \left( \frac{1}{2!} \mathcal{E}^{\alpha\beta\tau\sigma} \left( i G_\sigma \left( 2\partial_{[\alpha} G_{\beta]} - G_{[\alpha} \partial_{\beta]} \right) G_\tau \right) + G_\sigma * P'^\sigma \right) \\
&= \text{Tr} \left( i G_\sigma \left( 2 * \partial^{[\tau} G^{\sigma]} - * G^{[\tau} \partial^{\sigma]} \right) G_\tau + G_\sigma * P'^\sigma \right)
\end{aligned}$$

So now we see that for the faux monopole / baryon dual  $*P'^\sigma$ , the configuration space operator is  $2*\partial^{[\tau}G^{\sigma]} - *G^{[\tau}\partial^{\sigma]}$ , contrast  $g^{\nu\sigma}D_\tau D^\tau - D^\sigma D^\nu$  of (5.15) for the current density  $J^\nu$ , with  $D^\sigma D^\nu \equiv \partial^\sigma \partial^\nu - iG^\sigma \partial^\nu - 2G^\sigma G^\nu + G^\nu G^\sigma$  in (5.16) and  $D_\tau D^\tau = \partial_\tau \partial^\tau - iG_\tau \partial^\tau - G_\tau G^\tau$  in (5.17).

Finally, based on (12.22), the faux monopole / baryon action is written in the configuration space format:

$$S(G, *P') = \int d^4x \mathcal{L}(G, *P') = \int d^4x \text{Tr} \left( iG_\sigma \left( 2*\partial^{[\tau}G^{\sigma]} - *G^{[\tau}\partial^{\sigma]} \right) G_\tau + G_\sigma *P'^\sigma \right). \quad (12.23)$$

This is what we will now seek to plug into the path integral  $Z(G, *P') = \int DG \exp iS(G, *P') \equiv \mathcal{C} \exp iW(*P')$  to develop the quantum field properties of our faux monopole / candidate baryon  $P'$ .

Before turning to the path integral, given that we now have a faux monopole action  $S(G, *P')$  in contrast to the electric charge action  $S(G, J)$  in (11.1), it also helps by way of contrast to similarly take the first rank dual of (9.21) to obtain:

$$\begin{aligned} \text{Tr} *P'_\Sigma{}^\alpha((0))_0 &= \frac{1}{3!} \mathcal{E}^{\sigma\mu\nu\alpha} \text{Tr} P'_\Sigma{}^{\sigma\mu\nu}((0))_0 \\ &= -i \begin{pmatrix} \partial_\sigma \frac{1}{3!} \mathcal{E}^{\sigma\mu\nu\alpha} \left( \overline{\psi}_R \gamma_{[\mu} (\mathbf{p}_R - m_R)^{-1} \gamma_{\nu]} \psi_R \right) \\ + \partial_\mu \frac{1}{3!} \mathcal{E}^{\sigma\mu\nu\alpha} \left( \overline{\psi}_G \gamma_{[\nu} (\mathbf{p}_G - m_G)^{-1} \gamma_{\sigma]} \psi_G \right) \\ + \partial_\nu \frac{1}{3!} \mathcal{E}^{\sigma\mu\nu\alpha} \left( \overline{\psi}_B \gamma_{[\sigma} (\mathbf{p}_B - m_B)^{-1} \gamma_{\mu]} \psi_B \right) \end{pmatrix} = -i \begin{pmatrix} \partial_\sigma \frac{1}{3} * \left( \overline{\psi}_R \gamma^{\lambda\sigma} (\mathbf{p}_R - m_R)^{-1} \gamma^{\alpha\lambda} \psi_R \right) \\ + \partial_\mu \frac{1}{3} * \left( \overline{\psi}_G \gamma^{\lambda\mu} (\mathbf{p}_G - m_G)^{-1} \gamma^{\alpha\lambda} \psi_G \right) \\ + \partial_\nu \frac{1}{3} * \left( \overline{\psi}_B \gamma^{\lambda\nu} (\mathbf{p}_B - m_B)^{-1} \gamma^{\alpha\lambda} \psi_B \right) \end{pmatrix}. \quad (12.24) \\ &= -\frac{1}{3} i \partial_\sigma * \left( \overline{\psi}_R \gamma^{\lambda\sigma} (\mathbf{p}_R - m_R)^{-1} \gamma^{\alpha\lambda} \psi_R + \overline{\psi}_G \gamma^{\lambda\sigma} (\mathbf{p}_G - m_G)^{-1} \gamma^{\alpha\lambda} \psi_G + \overline{\psi}_B \gamma^{\lambda\sigma} (\mathbf{p}_B - m_B)^{-1} \gamma^{\alpha\lambda} \psi_B \right) \end{aligned}$$

So in the ground state, this first rank monopole / baryon dual has the symmetric  $\overline{RR} + \overline{GG} + \overline{BB}$  color symmetry of a duality-transformed meson, in the form

$$\text{Tr} *P'_\Sigma{}^\alpha((0))_0 \propto \partial_\sigma * \left( \overline{RR} + \overline{GG} + \overline{BB} \right)^{\sigma\alpha}.$$

### 13. The Baryon Candidate Quantum Path Integral

In order to path integrate the classical action (12.23), we will again make use of the recursion developed in section 8 to recast the path integral into a form that is quadratic, i.e., no higher than second order, in the gauge field  $G_\mu$ . We first substitute the recursive kernel

$G_\mu = \pi_\infty J_\mu = \left( \pi_0^{-1} + \pi_{\infty-1} J_\tau k^\tau + \pi_{\infty-1} J_\tau \pi_{\infty-1} J^\tau \right)^{-1} J_\mu$  with  $\pi_0 = \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1}$  as a terminal condition from (8.20) into the configuration space operator in (12.23) and then convert over to momentum space via  $i\partial \rightarrow k$  to obtain:

$$\begin{aligned}
S(G, *P') &= \int d^4 x \text{Tr} \left( iG_\sigma \left( 2 * \partial^{[\tau} G^{\sigma]} - * G^{[\tau} \partial^{\sigma]} \right) G_\tau + G_\sigma * P'^\sigma \right) \\
&= \int d^4 x \text{Tr} \left( iG_\sigma \left( 2\pi_\infty * \partial^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} \partial^{\sigma]} \right) G_\tau + G_\sigma * P'^\sigma \right) \quad . \\
&= \int d^4 x \text{Tr} \left( G_\sigma \left( 2\pi_\infty * k^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} k^{\sigma]} \right) G_\tau + G_\sigma * P'^\sigma \right)
\end{aligned} \tag{13.1}$$

Note that  $k^\tau$  and  $\pi_\infty$  readily transpose because  $k^\tau$  is an ordinary abelian vector and everything else on the final line is also in momentum space (thus averting any terms arising from the canonical commutation operation  $[x, k_x] = i\hbar$ ), and the duality can be moved over by segregating out the Levi-Civita tensor and then integrating it back in. The (non-Ghost portion of the) path integral over  $DG$  (omitting  $DcDc^\dagger$ ) is then specified by:

$$\begin{aligned}
Z(G, *P') &= \int DG \exp iS(G, *P') \\
&= \int DG \exp i \int d^4 x \text{Tr} \left( iG_\sigma \left( 2 * \partial^{[\tau} G^{\sigma]} - * G^{[\tau} \partial^{\sigma]} \right) G_\tau + G_\sigma * P'^\sigma \right) \\
&= \int DG \exp i \int d^4 x \text{Tr} \left( iG_\sigma \left( 2\pi_\infty * \partial^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} \partial^{\sigma]} \right) G_\tau + G_\sigma * P'^\sigma \right). \\
&= \int DG \exp i \int d^4 x \text{Tr} \left( G_\sigma \left( 2\pi_\infty * k^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} k^{\sigma]} \right) G_\tau + G_\sigma * P'^\sigma \right) \\
&\equiv \mathcal{C} \exp iW(*P')
\end{aligned} \tag{13.2}$$

As in section 8, the infinite recursion has enabled us to remove all gauge fields above quadratic order, and render (13.2) into a form that can be integrated exactly, analytically. The mathematical Gaussian integral that we will want to use as template for this is:

$$\int dx \exp(iAx^2 + iPx) = (\pi i / A)^{5/2} \exp(-iP^2 / 2A) \quad . \tag{13.3}$$

with  $x \sim G$ ,  $A \sim 2\pi_\infty * k^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} k^{\sigma]}$ , and  $P \sim *P'$ . Clearly, we need to now obtain the inverse of  $A \sim 2\pi_\infty * k^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} k^{\sigma]}$  because it shows up on the right-hand side in  $P^2 / A$ .

To find the inverse of  $2\pi_\infty * k^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} k^{\sigma]}$ , we note the procedure employed throughout sections 6 and 7. But here,  $A \equiv 2\pi_\infty * k^{[\tau} J^{\sigma]} - \pi_\infty * J^{[\tau} k^{\sigma]}$  is an *antisymmetric* matrix. Therefore, as a special case, its inverse is equal to *inverse* of the *negative* of its *transpose*, that is:

$$\frac{1}{2} A^{-1} = \frac{1}{2} \left( -2\pi_\infty * k^{[\sigma} J^{\tau]} + \pi_\infty * J^{[\sigma} k^{\tau]} \right)^{-1} \quad . \tag{13.4}$$

Therefore, using (13.3) as a template, we find from (13.2) and (13.4) that:

$$\begin{aligned}
Z(G, *P') &= \int DG \exp i \int d^4x \text{Tr} \left( G_\sigma \left( 2\pi_\infty *k^{[\tau} J^{\sigma]} - \pi_\infty *J^{[\tau} k^{\sigma]} \right) G_\tau + G_\sigma *P'^\sigma \right) \\
&= \mathcal{C} \exp i \frac{1}{2} \int \left( d^4k / (2\pi)^4 \right) \text{Tr} \left( *P'^\sigma \left( 2\pi_\infty *k^{[\sigma} J^{\tau]} - \pi_\infty *J^{[\sigma} k^{\tau]} \right)^{-1} *P'^\tau \right). \\
&\equiv \mathcal{C} \exp i W(*P')
\end{aligned} \tag{13.5}$$

And from this we extract the amplitude:

$$(2\pi)^4 W(*P') = \frac{1}{2} \int d^4k \text{Tr} \left( *P'^\sigma \left( 2\pi_\infty *k^{[\sigma} J^{\tau]} - \pi_\infty *J^{[\sigma} k^{\tau]} \right)^{-1} *P'^\tau \right). \tag{13.6}$$

We further define  $\mathcal{W}$  which is the momentum space amplitude density by  $\int d^4k \mathcal{W} \equiv (2\pi)^4 W$  to extract both  $\int d^4k \mathcal{W}(*P') \equiv (2\pi)^4 W(*P')$  from (13.6) and  $\int d^4k \mathcal{W}(J) \equiv (2\pi)^4 W(J)$  from (11.5) with comparable coefficients. In (11.5) we also use  $\mathcal{W}(J)_\infty \equiv \mathcal{W}(J)_{\infty-1}$  to set  $\infty-1 \rightarrow \infty$ . Thus, in an apples-to-apples comparison, shown together with the recursive kernel and the terminal condition from (8.18), we may write:

$$\left\{ \begin{aligned}
\mathcal{W}(J) &= \frac{1}{2} \text{Tr} \left( J_\sigma \left( 2\pi_0^{-1} + 2\pi_\infty J_\tau k^\tau + 2\pi_\infty J_\tau \pi_\infty J^\tau \right)^{-1} J^\sigma \right) \\
\mathcal{W}(*P') &= \frac{1}{2} \text{Tr} \left( *P'^\sigma \left( 2\pi_\infty *k^{[\sigma} J^{\tau]} - \pi_\infty *J^{[\sigma} k^{\tau]} \right)^{-1} *P'^\tau \right) \\
\pi_n &= \left( \pi_0^{-1} + \pi_{n-1} J_\tau k^\tau + \pi_{n-1} J_\tau \pi_{n-1} J^\tau \right)^{-1} \\
\pi_0 &= \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1}
\end{aligned} \right. \tag{13.7}$$

Earlier, we wrote out (5.9) which displayed Maxwell's classical equations in non-abelian gauge theory in integral form. The above (13.7) are the quantum field theory counterpart of these non-abelian Maxwell equations. Together, these two equations – which are in momentum space – should furnish the basis for properly deciphering the quantum field behaviors of the electric current densities which are quark current candidates, and the quantum field behaviors of the faux magnetic current densities which are baryon candidates. In short, (13.7) is a candidate quantum field theory for quark and baryons which parallels Maxwell's equations in non-abelian form.

However,  $*P'^\sigma$  appearing in the monopole amplitude (13.7) is the first rank dual of the third-rank faux monopole  $P'_{\sigma\mu\nu}$  which, formally speaking, is itself the candidate baryon density with the  $R \wedge G \wedge B$  color symmetries found in (9.21). So we now seek to have  $P'_{\sigma\mu\nu}$  explicitly appear in  $\mathcal{W}(*P')$ . Toward this end, we make use of second-rank duality in the form of  $*k^{[\sigma} J^{\tau]} = \frac{1}{2!} \mathcal{E}^{\alpha\beta\sigma\tau} k_{[\alpha} J_{\beta]}$  and  $*J^{[\sigma} k^{\tau]} = \frac{1}{2!} \mathcal{E}^{\alpha\beta\sigma\tau} J_{[\alpha} k_{\beta]}$  to write:

$$\mathcal{W}(*P') = \frac{1}{2} \left( \frac{1}{2!} \mathcal{E}^{\alpha\beta\sigma\tau} \right)^{-1} \text{Tr} \left( *P'^\sigma \left( 2\pi_\infty k_{[\alpha} J_{\beta]} - \pi_\infty J_{[\alpha} k_{\beta]} \right)^{-1} *P'^\tau \right). \tag{13.8}$$

But it is also an identity of the Levi-Civita tensor in flat spacetime that  $\varepsilon^{\alpha\beta\sigma\tau}\varepsilon_{\alpha\beta\sigma\tau} = -\frac{1}{4!}$ , and of course, by definition,  $(\varepsilon^{\alpha\beta\sigma\tau})^{-1}\varepsilon^{\alpha\beta\sigma\tau} \equiv 1$ . Combining these two expressions gives  $-\frac{1}{4!}\varepsilon^{\alpha\beta\sigma\tau}(\varepsilon^{\alpha\beta\sigma\tau})^{-1} = \varepsilon^{\alpha\beta\sigma\tau}\varepsilon_{\alpha\beta\sigma\tau} = -\frac{1}{4!}$  or  $-\frac{1}{4!}(\varepsilon^{\alpha\beta\sigma\tau})^{-1} = \varepsilon_{\alpha\beta\sigma\tau}$ . We then use the later expression in the form of  $\frac{1}{2}\left(\frac{1}{2!}\varepsilon^{\alpha\beta\sigma\tau}\right)^{-1} = -4!\varepsilon_{\alpha\beta\sigma\tau}$  to rewrite (13.8) as:

$$\mathfrak{N}(*P') = -4!\varepsilon_{\alpha\beta\sigma\tau}\text{Tr}\left(*P'^{\sigma}\left(2\pi_{\infty}k_{[\alpha}J_{\beta]} - \pi_{\infty}J_{[\alpha}k_{\beta]}\right)^{-1}*P'^{\tau}\right). \quad (13.9)$$

Because the minus sign in  $\varepsilon^{\alpha\beta\sigma\tau}\varepsilon_{\alpha\beta\sigma\tau} = -\frac{1}{4!}$  arises from the fact that the Minkowski metric tensor  $\text{diag}(\eta_{\mu\nu}) = (1, -1, -1, -1)$  has a determinant  $\det\eta = -1$ , the sign reversal in (13.9) emanates from the underlying structure of Minkowski space.

Next we use  $*P'^{\sigma} = \frac{1}{3!}\varepsilon^{\mu\nu\rho\sigma}P'_{\mu\nu\rho}$  to further rewrite (13.9) as:

$$\begin{aligned} \mathfrak{N}(*P') &= -4!\varepsilon_{\alpha\beta\sigma\tau}\text{Tr}\left(*P'^{\sigma}\left(2\pi_{\infty}k_{[\alpha}J_{\beta]} - \pi_{\infty}J_{[\alpha}k_{\beta]}\right)^{-1}*P'^{\tau}\right) \\ &= -4!\frac{1}{3!}\varepsilon^{\mu\nu\rho\sigma}\varepsilon_{\alpha\beta\sigma\tau}\text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\alpha}J_{\beta]} - \pi_{\infty}J_{[\alpha}k_{\beta]}\right)^{-1}*P'^{\tau}\right) \\ &= -4!\frac{1}{3!}\delta^{\mu\nu\rho}_{\alpha\beta\tau}\text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\alpha}J_{\beta]} - \pi_{\infty}J_{[\alpha}k_{\beta]}\right)^{-1}*P'^{\tau}\right) \\ &= -4!\frac{1}{3!}\left(\delta^{\mu}_{\alpha}\delta^{\nu}_{\beta}\delta^{\rho}_{\tau} + \delta^{\mu}_{\beta}\delta^{\nu}_{\tau}\delta^{\rho}_{\alpha} + \delta^{\mu}_{\tau}\delta^{\nu}_{\alpha}\delta^{\rho}_{\beta} - \delta^{\mu}_{\beta}\delta^{\nu}_{\alpha}\delta^{\rho}_{\tau} - \delta^{\mu}_{\alpha}\delta^{\nu}_{\tau}\delta^{\rho}_{\beta} - \delta^{\mu}_{\tau}\delta^{\nu}_{\beta}\delta^{\rho}_{\alpha}\right)\text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\alpha}J_{\beta]} - \pi_{\infty}J_{[\alpha}k_{\beta]}\right)^{-1}*P'^{\tau}\right) \\ &= -2 \times 4! \frac{1}{3!} \left( \text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\mu}J_{\nu]} - \pi_{\infty}J_{[\mu}k_{\nu]}\right)^{-1}*P'^{\rho}\right) \right. \\ &\quad \left. + \text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\rho}J_{\mu]} - \pi_{\infty}J_{[\rho}k_{\mu]}\right)^{-1}*P'^{\nu}\right) \right. \\ &\quad \left. + \text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\nu}J_{\rho]} - \pi_{\infty}J_{[\nu}k_{\rho]}\right)^{-1}*P'^{\mu}\right) \right). \end{aligned} \quad (13.10)$$

Next we use  $*P'^{\rho} = \frac{1}{3!}\varepsilon^{\alpha\beta\gamma\rho}P'_{\alpha\beta\gamma}$  and the like for the remaining first rank duals.

$$\mathfrak{N}(P') = -2 \times 4! \frac{1}{3!} \frac{1}{3!} \left( \varepsilon^{\alpha\beta\gamma\rho}\text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\mu}J_{\nu]} - \pi_{\infty}J_{[\mu}k_{\nu]}\right)^{-1}P'_{\alpha\beta\gamma}\right) \right. \\ \left. + \varepsilon^{\alpha\beta\gamma\nu}\text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\rho}J_{\mu]} - \pi_{\infty}J_{[\rho}k_{\mu]}\right)^{-1}P'_{\alpha\beta\gamma}\right) \right. \\ \left. + \varepsilon^{\alpha\beta\gamma\mu}\text{Tr}\left(P'_{\mu\nu\rho}\left(2\pi_{\infty}k_{[\nu}J_{\rho]} - \pi_{\infty}J_{[\nu}k_{\rho]}\right)^{-1}P'_{\alpha\beta\gamma}\right) \right). \quad (13.11)$$

But because  $P_{\mu\nu\rho}$  is antisymmetric in all indexes and given the way it is summed, this reduces to:

$$\begin{aligned} \mathfrak{N}(P') &= -2 \times 4! \frac{1}{3!} \times 3 \mathcal{E}^{\alpha\beta\gamma\rho} \text{Tr} \left( P'_{\mu\nu\rho} \left( 2\pi_\infty k_{[\mu} J_{\nu]} - \pi_\infty J_{[\mu} k_{\nu]} \right)^{-1} P'_{\alpha\beta\gamma} \right) \\ &= -4 \mathcal{E}^{\alpha\beta\gamma\rho} \text{Tr} \left( P'_{\mu\nu\rho} \pi_\infty^{-1} \left( 2k_{[\mu} J_{\nu]} - J_{[\mu} k_{\nu]} \right)^{-1} P'_{\alpha\beta\gamma} \right) \end{aligned} \quad (13.12)$$

Note that in the bottom line we have factored out  $\pi_\infty^{-1}$  from  $\left( 2\pi_\infty k_{[\mu} J_{\nu]} - \pi_\infty J_{[\mu} k_{\nu]} \right)^{-1}$ . So we now use (13.12) to rewrite  $\mathfrak{N}(*P')$  in (13.7) with  $\infty \rightarrow \infty - 1$  directly in terms of the third rank baryon density candidate  $P'_{\alpha\beta\gamma}$ , as:

$$\left\{ \begin{aligned} \mathfrak{N}(J) &= \frac{1}{2} \text{Tr} \left( 2J_\sigma \pi_\infty J^\sigma \right) \\ \mathfrak{N}(P') &= -2 \mathcal{E}^{\alpha\beta\gamma\rho} \text{Tr} \left( 2P'_{\mu\nu\rho} \pi_\infty^{-1} \left( 2k_{[\mu} J_{\nu]} - J_{[\mu} k_{\nu]} \right)^{-1} P'_{\alpha\beta\gamma} \right) \\ \pi_n &= \left( \pi_0^{-1} + \pi_{n-1} J_\tau k^\tau + \pi_{n-1} J_\tau \pi_{n-1} J^\tau \right)^{-1} \\ \pi_0 &= \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} \end{aligned} \right. \quad (13.13)$$

This describes the quantum field interactions of candidate quark and baryon charge / current densities, up to ghost fields  $c$  that we omitted at (13.2) and also in section 11. That is, the complete, gauge-invariant path integral involves the integral and measure  $\int DG Dc Dc^\dagger$ . So by integrating over  $\int DG$  we have effectively integrated in one of two “dimensions,” in the nature of taking  $\int dx F(x)$  as a partial integration in an overall integral of the form  $\int dx dy F(x, y)$ . The much more challenging, seemingly-intractable problem is how to do the mathematics of a Gaussian integral which contains terms higher than second order quadratic in the integration variable, as does the Yang-Mills path integrals with regard to the gauge field. With (13.13) containing the very explicit solution to this key mathematical problem, any number of individuals with ordinary knowledge should be able to fill in the ghost fields. When truncated at a *finite* recursive order  $n$ , cf. (8.17), from (13.13) we may write:

$$\left\{ \begin{aligned} \mathfrak{N}(J)_n &= \frac{1}{2} \text{Tr} \left( 2J_\sigma \pi_n J^\sigma \right) \\ \mathfrak{N}(P')_n &= -2 \mathcal{E}^{\alpha\beta\gamma\rho} \text{Tr} \left( 2P'_{\mu\nu\rho} \pi_n^{-1} \left( 2k_{[\mu} J_{\nu]} - J_{[\mu} k_{\nu]} \right)^{-1} P'_{\alpha\beta\gamma} \right) \\ \pi_n &= \left( \pi_0^{-1} + \pi_{n-1} J_\tau k^\tau + \pi_{n-1} J_\tau \pi_{n-1} J^\tau \right)^{-1} \\ \pi_0 &= \left( k_\tau k^\tau - m^2 + i\mathcal{E} \right)^{-1} \end{aligned} \right. \quad (13.14)$$

There are some additional reductions that can be made in (13.14). First because everything is in momentum space we may commute  $J_\tau k^\tau = k^\tau J_\tau$  and then set this to zero in  $\pi_n$  because  $i\partial_\nu G^\nu = k_\nu G^\nu = 0$  as first found in (6.5). Thus, we simplify to:

$$\pi_n = (\pi_0^{-1} + \pi_{n-1} J_\tau \pi_{n-1} J^\tau)^{-1} \quad (13.15)$$

which now contains only second order terms  $\pi_{n-1} J_\tau \pi_{n-1} J^\tau$  in  $J$ , but no first-order  $Jk$  terms.

Next, if we write out the term  $2k_{[\mu} J_{\nu]} - J_{[\mu} k_{\nu]}$  from  $\mathfrak{N}(P')$  component-by-component and again keep in mind that we are in momentum space so that the commutator  $[x, k_x] = i\hbar$  does not come into play, we find that  $2k_{[\mu} J_{\nu]} - J_{[\mu} k_{\nu]} = 2k_{[\mu} J_{\nu]} + J_{[\nu} k_{\mu]} = 3k_{[\mu} J_{\nu]}$ . As a result,

$$\mathfrak{N}(P')_n = -2\varepsilon^{\alpha\beta\gamma\rho} \text{Tr} \left( 2P'_{\mu\nu\rho} \pi_n^{-1} (3k_{[\mu} J_{\nu]})^{-1} P'_{\alpha\beta\gamma} \right) = -\frac{4}{3} \varepsilon^{\alpha\beta\gamma\rho} \text{Tr} \left( P'_{\mu\nu\rho} \pi_n^{-1} (k_{[\mu} J_{\nu]})^{-1} P'_{\alpha\beta\gamma} \right). \quad (13.16)$$

It is interesting to note the natural emergence of the QCD color factor  $C_F = \frac{4}{3}$ , see [14] eq. [2.98]. And it is also very interesting to note that the relationship between an abelian gauge field  $A_\nu$  and the associated abelian (A) field strength is  $iF_{A\mu\nu} = i\partial_{[\mu} A_{\nu]} = k_{[\mu} A_{\nu]}$  in view of  $i\partial \rightarrow k$ . But  $k_{[\mu} J_{\nu]}$  in (13.15) has *exactly the same form as*  $iF_{A\mu\nu}$ . It simply contains  $J_\nu$  rather than  $A_\nu$  and so is two differential orders lower than  $iF_{A\mu\nu}$ . We shall thus *define* the antisymmetric tensor  $i\Phi_{\mu\nu} \equiv k_{[\mu} J_{\nu]}$  as the ‘‘echo field strength’’ tensor because in a sense it is merely an ‘‘echo’’ of  $iF_{A\mu\nu}$  at two orders lower. And to be able to think about this tensor in a familiar way, we may define its contravariant components by analogy to  $E$  and  $B$  via bivectors  $\mathbf{E} \rightarrow \boldsymbol{\varepsilon}$  and  $\mathbf{B} \rightarrow \boldsymbol{\beta}$  as:

$$\Phi^{\mu\nu} \equiv \begin{pmatrix} 0 & -\varepsilon_x & -\varepsilon_y & -\varepsilon_z \\ \varepsilon_x & 0 & -\beta_z & \beta_y \\ \varepsilon_y & \beta_z & 0 & -\beta_x \\ \varepsilon_z & -\beta_y & \beta_x & 0 \end{pmatrix}. \quad (13.17)$$

Because (13.16) contains  $(k_{[\mu} J_{\nu]})^{-1}$ , we see that to further develop (13.16), we will need to take the inverse of this tensor. Generally, to invert a 4x4 matrix  $\mathbf{A}$ , the formula is [22]:

$$\mathbf{A}^{-1} = |\mathbf{A}|^{-1} \left[ \frac{1}{6} ((\text{tr}\mathbf{A})^3 - 3\text{tr}\mathbf{A}\text{tr}\mathbf{A}^2 + 2\text{tr}\mathbf{A}^3) - \frac{1}{2} \mathbf{A} ((\text{tr}\mathbf{A})^2 - \text{tr}\mathbf{A}^2) + \mathbf{A}^2 \text{tr}\mathbf{A} - \mathbf{A}^3 \right]. \quad (13.18)$$

But because (13.17) is an antisymmetric tensor its trace is zero and so the above simplifies to:

$$\mathbf{A}^{-1} = -|\mathbf{A}|^{-1} \mathbf{A}^3. \quad (13.19)$$

The third and final simplification is to recognize that if (13.14) does indeed describe the amplitude densities of QCD, then we should also set  $m=0$  in  $\pi_0$  because this is associated with the ‘‘mass’’ of the gluons, and we know that the QCD gluons are massless. Ordinarily, when a gauge boson mass is set to zero, some uniqueness is lost because a degree of freedom is removed

from the system, even when we enforce the contextual gauge fixing of a conserved continuous current density. In particular,  $\partial_\sigma G^\sigma = 0$  is no longer a requirement but is relegated to an optional gauge condition. This was a central aspect of the discussion in section 6 and 7. But we also saw how contextual gauge fixing restores lost uniqueness, and in effect forces the massless inverses into the Feynman / continuity gauge. As regards (13.13), we then showed at (9.15) how when  $k_\tau k^\tau - m^2$  are associated with the gluons within a baryon, the gluon mass can be set to zero at the same time the quarks are given a mass, by simply shifting one degree of freedom from a gluon into a fermion in a type of Goldstone mechanism, so that no uniqueness is lost in the context of the overall baryon system. So as long as we associate (13.13) with the gluons within a baryon that contains massive quarks, we can set these boson masses in (13.13) to zero with the implicit transfer of a degree of freedom that makes the quarks massive, and without any loss of uniqueness. So in this context, we now set  $m=0$  in (13.13) and (13.14). Together (13.15) and (13.16), we rewrite (13.13) and (13.14) respectively as:

$$\left\{ \begin{array}{l} \mathfrak{N}(J) = \text{Tr}(J_\sigma \pi_\infty J^\sigma) \\ \mathfrak{N}(P') = -\frac{4}{3} \varepsilon^{\alpha\beta\gamma\rho} \text{Tr}\left(P'_{\mu\nu\rho} \pi_\infty^{-1} (k_{[\mu} J_{\nu]})^{-1} P'_{\alpha\beta\gamma}\right) \\ \pi_n = (\pi_0^{-1} + \pi_{n-1} J_\tau \pi_{n-1} J^\tau)^{-1} \\ \pi_0 = (k_\tau k^\tau + i\varepsilon)^{-1} \end{array} \right. , \quad (13.20)$$

$$\left\{ \begin{array}{l} \mathfrak{N}(J)_n = \text{Tr}(J_\sigma \pi_n J^\sigma) \\ \mathfrak{N}(P')_n = -\frac{4}{3} \varepsilon^{\alpha\beta\gamma\rho} \text{Tr}\left(P'_{\mu\nu\rho} \pi_n^{-1} (k_{[\mu} J_{\nu]})^{-1} P'_{\alpha\beta\gamma}\right) \\ \pi_n = (\pi_0^{-1} + \pi_{n-1} J_\tau \pi_{n-1} J^\tau)^{-1} \\ \pi_0 = (k_\tau k^\tau + i\varepsilon)^{-1} \end{array} \right. . \quad (13.21)$$

With this, we have the quark density and baryon density amplitudes which should explain the phenomenology of quantum chromodynamics for which the gluons are indeed massless.

It is worth noting that since  $\mathfrak{N}(J)_0 = \text{Tr}(J^\mu \pi_0 g_{\mu\nu} J^\nu)$  for the zeroth recursive order is simply the abelian amplitude density, we normally associate  $\pi_0 g_{\mu\nu} = g_{\mu\nu} / (k_\tau k^\tau + i\varepsilon)$  with the abelian propagator (up to a factor of  $i$ ). This means that  $\pi_n g_{\mu\nu}$  generally represents the non-abelian gluon propagator for a given recursive order  $n$ , and that  $\pi_\infty g_{\mu\nu}$  therefore represents the *physical* gluon propagator with all non-linear effects accounted for. But if  $\pi_\infty g_{\mu\nu}$  represents the propagator for a gluon field between two  $J$  each of which represents a quark current density, then what can we say about  $\mathfrak{N}(P')$  which is the amplitude density for two monopoles  $P$  which we have shown have the color symmetries of baryons? We know that interactions between baryons

are mediated, not by gluons, but by mesons. So by analogy to  $\mathfrak{N}(J)$ , we expect that  $\pi_n^{-1}(k_{[\mu}J_{\nu]})^{-1}$  will be the *meson propagator* at recursive order  $n$ , and that  $\pi_\infty^{-1}(k_{[\mu}J_{\nu]})^{-1} = \pi_\infty^{-1}(i\Phi_{\mu\nu})^{-1}$  will define the propagator for the *physical* mesons with all non-linear effects accounted for. So we see that the “echo” tensor is not just an interesting analog to  $iF_{A\mu\nu} = k_{[\mu}A_{\nu]}$ , but rather, is an important physical tensor which in inverse form, see (13.19), plays a definitive and central role in the propagation of the mesons which mediate interactions between baryons and most importantly, nucleons. That is,  $\pi_\infty^{-1}(i\Phi_{\mu\nu})^{-1}$  fundamentally represents the strong nuclear force between nucleons and other baryons.

So if  $\pi_\infty g_{\mu\nu}$  defines the propagator for gluons mediating interactions between quark currents  $J$ , and if  $(\pi_\infty \Phi_{\mu\nu})^{-1}$  likewise defines the propagator for mesons mediating interactions between monopole / baryon currents  $P$ , then a full development of  $\mathfrak{N}(J)$  should establish the confinement of quarks within baryons, while the full development of  $\mathfrak{N}(P')$  should establish the short range of the nuclear interaction. We now develop  $\mathfrak{N}(J)$  in configuration space to demonstrate confinement.

#### 14. Direct Quantum Field Theory Demonstration of Confinement – Abelian Calculation

While the physical amplitude density for interactions between quark currents  $J$  is given by  $\mathfrak{N}(J) = \text{Tr}(J_\sigma \pi_\infty J^\sigma)$ , the basic character of confinement is already demonstrated at the first recursive order, that is, by  $\mathfrak{N}(J)_1 = \text{Tr}(J_\sigma \pi_1 J^\sigma)$ . Because this calculation can be completed on a wholly analytical basis, this will be our starting point for dynamically demonstrating confinement, and in particular, for developing a potential energy  $E$  between any two  $J$  separated in configuration space by a distance  $r$  which tightly confines the  $J$  together as  $r$  is increased beyond a certain length on the order of the nuclear scale close to 1 Fm. The calculation we shall now develop directly mirrors the calculation shown in chapter I.4 of [11], with the exception that it uses  $\pi_1$  rather than  $\pi_0$ , and with the further difference that in (13.20) and (13.21) we have set  $m=0$  because the non-linear current densities will themselves take on a role analogous to the mass  $m$  as used in the calculation of chapter I.4 of [11].

To start with, we use (13.21) to write  $\pi_1$  from the gluon propagator  $\pi_1 g_{\mu\nu}$  as:

$$\pi_1 = \left( \pi_0^{-1} + \pi_0 J_\tau \pi_0 J^\tau \right)^{-1} = \left( k_\tau k^\tau + i\mathcal{E} + \frac{J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^2} \right)^{-1} = \left( \frac{(k_\tau k^\tau + i\mathcal{E})^3 + J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^2} \right)^{-1}. \quad (14.1)$$

Also, back before (1.2) we scaled  $gG_\mu \rightarrow G_\mu$ , noting that this  $g$  can always be extracted back out when explicitly needed. Now, it is explicitly needed. Because the currents  $J$  have been shown to have the features of quark currents, we associate  $g$  with the running strong charge  $g_s$ , and because all of the gauge fields  $G_\mu$  were replaced with like-indexed currents  $J_\mu$  via  $G_\mu \rightarrow \pi_\infty J_\mu$  back at (11.3), we may rescale  $J_\tau J^\tau \rightarrow g_s^2 J_\tau J^\tau$  and then use the dimensionless strong running coupling  $\alpha_s = g_s^2 / 4\pi\hbar c$  in natural units  $\hbar = c = 1$  to write (14.1) as:

$$\pi_1 = \left( \frac{(k_\tau k^\tau + i\varepsilon)^3 + 4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\varepsilon)^2} \right)^{-1} = \left( k_\tau k^\tau + i\varepsilon + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\varepsilon)^2} \right)^{-1}. \quad (14.2)$$

So now the running strong coupling is now part of this first-recursive-order propagator term.

Next we may use (13.21) and (14.2) to write the amplitude density:

$$\mathcal{N}(J)_1 = \text{Tr}(J_\sigma \pi_1 J^\sigma) = \text{Tr} \left( 4\pi\alpha_s J_\sigma \left( k_\tau k^\tau + i\varepsilon + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\varepsilon)^2} \right)^{-1} J^\sigma \right). \quad (14.3)$$

If we use the SU(3) generators (the Gell-Mann matrices  $\lambda^i$ ) to expand  $J^\tau = \lambda^i J^{i\tau}$ , we see that:

$$\mathcal{N}(J)_1 = \text{Tr}(J_\sigma \pi_1 J^\sigma) = \text{Tr} \left( 4\pi\alpha_s \lambda^k J^k_\sigma \left( k_\tau k^\tau + i\varepsilon + \frac{4\pi\alpha_s \lambda^i \lambda^j J^i_\tau J^{j\tau}}{(k_\tau k^\tau + i\varepsilon)^2} \right)^{-1} \lambda^l J^{l\sigma} \right). \quad (14.4)$$

So inverting involves taking  $\left( 4\pi\alpha_s \lambda^i \lambda^j J^i_\tau J^{j\tau} + (k_\tau k^\tau + i\varepsilon)^3 \right)^{-1}$ , which is the inverse of the 3x3 matrix  $4\pi\alpha_s \lambda^i \lambda^j J^i_\tau J^{j\tau} - \left[ -\delta^{ij} (k_\tau k^\tau + i\varepsilon)^3 \right]$  for which  $-(k_\tau k^\tau + i\varepsilon)^3$  represent the eigenvalues of

$$4\pi\alpha_s \lambda^i \lambda^j J^i_\tau J^{j\tau} = 4\pi\alpha_s J_\tau J^\tau \quad \text{via the determinant} \quad \left| 4\pi\alpha_s \lambda^i \lambda^j J^i_\tau J^{j\tau} - \left[ -\delta^{ij} (k_\tau k^\tau + i\varepsilon)^3 \right] \right| = 0.$$

Further, commuting  $J^\sigma$  from the right to the left of  $\left( k_\tau k^\tau + i\varepsilon + 4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\varepsilon)^2 \right)^{-1}$ , which we will need to do at one point in the calculation following (at (4.10 supra)) is made much more difficult because  $J^\sigma = \lambda^i J^{i\sigma}$  and  $J^\tau = \lambda^j J^{j\tau}$  and do not commute with one another, but rather commute via  $[J^\sigma, J^\tau] = J^{i\sigma} J^{j\tau} [\lambda^i, \lambda^j] = if^{ijk} \lambda^k J^{i\sigma} J^{j\tau} \neq 0$  based on the group relation  $[\lambda^i, \lambda^j] = if^{ijk} \lambda^k$ .

As a result, we shall organize the mathematical calculation proceeding from (14.3) into two main steps. First, in this section, we shall treat all of the  $J$  in (14.3) as 1x1 matrices, rather

than as the 3x3 matrices  $J^\tau = \lambda^i J^{i\tau}$  which they are in  $SU(3)_C$ . This means two things: first, we can now set  $\left(k_\tau k^\tau + i\mathcal{E} + 4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\mathcal{E})^2\right)^{-1} \rightarrow 1 / \left(k_\tau k^\tau + i\mathcal{E} + 4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\mathcal{E})^2\right)$  as if this is an ordinary denominator. Second, we can now treat the current densities  $J^\sigma$  as ordinary abelian currents and so commute them past one another using  $[J^\sigma, J^\tau] = 0$ , at will. Second, in the next section, we shall review how this abelian simplification affected the overall calculation, and use that review to generalize the overall abelian result of this section, to non-abelian theory.

Accordingly, as an *abelian simplification*, we may now use the inverse in (14.3) as if it was an ordinary denominator, and so write (14.3) as:

$$\mathcal{D}\mathcal{L}(J)_1 = (J_\sigma \pi_1 J^\sigma) = 4\pi\alpha_s J_\sigma \frac{1}{k_\tau k^\tau + i\mathcal{E} + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^2}} J^\sigma. \quad (14.5)$$

Because we now take these  $J$  to be 1x1 objects, we remove the trace. At (14.8) below, we will also restore the overall coefficient of  $\frac{1}{2}$  which is eliminated once one introduces generator matrices normalized to  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$ . The above, (14.5), will now be our starting point for demonstrating a confining potential between the two  $J^\sigma$  sources.

Using the above, the amplitude density integrated over momentum space will then be:

$$W(J)_1 = \int \frac{d^4 k}{(2\pi)^4} \mathcal{D}\mathcal{L}(J)_1 = \int \frac{d^4 k}{(2\pi)^4} (J_\sigma \pi_1 J^\sigma) = \int \frac{d^4 k}{(2\pi)^4} 4\pi\alpha_s J_\sigma \frac{1}{k_\tau k^\tau + i\mathcal{E} + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^2}} J^\sigma. \quad (14.6)$$

But this is all in momentum space, and we now need to do the Fourier transforms over into configuration space. We know that in configuration space, the propagator  $D_1(x-y)$  to first recursive order, based on (14.6) is found via the Fourier transform:

$$D_1(x-y) = \int \frac{d^4 k}{(2\pi)^4} \frac{1}{k_\tau k^\tau + i\mathcal{E} + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^2}} e^{ik_\sigma(x-y)^\sigma}. \quad (14.7)$$

Additionally,  $W(J)_1$  in (14.6) is also given in configuration space with  $4\pi\alpha_s$  by:

$$W(J)_1 = -\frac{1}{2} \iint d^4 x d^4 y 4\pi\alpha_s J_\sigma(x) D_1(x-y) J^\sigma(y). \quad (14.8)$$

This now includes the factor of  $\frac{1}{2}$  which doubles to compensate for generator matrices normalized to  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$  as noted after (14.5). It is included because we are approximating past the generator matrices by using the inverse as an ordinary denominator. Using (14.7) in (14.8) then gives us:

$$W(J)_1 = -\frac{1}{2} \iint d^4x d^4y 4\pi\alpha_s J_\sigma(x) \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik_\sigma(x-y)^\sigma}}{k_\tau k^\tau + i\epsilon + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^2}} J^\sigma(y). \quad (14.9)$$

Now let's get to work evaluating (14.9) as a definite integral. The first thing is to separate the time from the space components and the spacetime-dependent objects from the momentum space-dependent objects, and so write (14.9) as:

$$W(J)_1 = -\frac{1}{2} \iint dx^0 dy^0 \iint d^3x d^3y 4\pi\alpha_s J_\sigma(\mathbf{x}) J^\sigma(\mathbf{y}) \int \frac{dk^0}{2\pi} e^{ik_0(x-y)^0} \int \frac{d^3k}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{k_\tau k^\tau + i\epsilon + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^2}}. \quad (14.10)$$

In going from (14.9) to (14.10), we moved  $J(y)$  from the right to the left of the overall denominator  $1/\left(k_\tau k^\tau + i\epsilon + 4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\epsilon)^2\right)$ , which was only possible because we are using an abelian simplification in which  $[J^\sigma, J^\tau] = 0$ . So in the next section when we seek the non-abelian generalization of the results to be derived here, we shall return to dissect the step of going from (14.9) to (14.10) in much more detail to "reconstruct" whatever we are foregoing because of the abelian simplification.

The expression  $\iint d^3x d^3y 4\pi\alpha_s J_\sigma(\mathbf{x}) J^\sigma(\mathbf{y}) = \int d^3x g_s J_\sigma(x) \int d^3y g_s J^\sigma(y)$  can effectively be set to 1 given that  $g_s J_0 = g_s \psi^\dagger \psi \equiv \rho$  defines the probability density. So in the rest frame,  $\int d^3x g_s \rho(x) \int d^3y g_s \rho(y) = \rho_0(x) \int d^3y g_s \rho_0(y) = 1 \cdot 1 = 1$ , where  $\rho_0$  is the proper, scalar probability density. Thus the integral  $\iint d^3x d^3y 4\pi\alpha_s J(\mathbf{x}) J(\mathbf{y}) = 1$ . Additionally, as Zee does at the top of page 26 in [11], we add a factor of 2 to account for both of the interactions  $J_x J_y$  and  $J_y J_x$ . Thus, (14.10) becomes:

$$W(J)_1 = -\iint dx^0 dy^0 \int \frac{dk^0}{2\pi} e^{ik_0(x-y)^0} \int \frac{d^3k}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{k_\tau k^\tau + i\epsilon + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^2}}. \quad (14.11)$$

Next we use the Dirac delta. This infinite Gaussian spike of area 1 is defined as the Fourier transform of the number 1, that is, as  $\delta(x) \equiv \int (dk / 2\pi) 1 e^{ikx}$  with  $\int \delta(x) dx = 1$ . So taking the entire term  $\iint dx^0 dy^0 \int (dk^0 / 2\pi) e^{ik_0(x-y)^0}$  in (14.11), and in the final step setting  $k_0 = 0$  as at the top of page 26 in [11], we may rework this term into:

$$\begin{aligned} & \iint dx^0 dy^0 \int \frac{dk^0}{2\pi} e^{ik_0(x-y)^0} = \iint dx^0 dy^0 e^{ik_0 x^0} \int \frac{dk^0}{2\pi} e^{-ik_0 y^0} \\ & = \iint dx^0 dy^0 e^{ik_0 x^0} \delta(-y^0) = \int dx^0 e^{ik_0 x^0} \int \delta(-y^0) dy^0 = \int dx^0 e^{ik_0 x^0} = \int dx^0 \end{aligned} \quad (14.12)$$

Because  $k_0 = 0$ ,  $k_\tau k^\tau = -\mathbf{k}^2$ . Using this together with (14.12), and also removing  $+i\epsilon$  because with  $k_0 = 0$  we are not on-shell, there is an overall sign reversal, and (14.11) simplifies to:

$$W(J)_1 = \int dx^0 \int \frac{d^3 k}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}. \quad (14.13)$$

Now for  $\int dx^0$ , we note that the path integral  $Z = \mathcal{C} \exp iW(J)$  represents the quantum operator  $\langle 0 | \exp(-iHT) | 0 \rangle = \exp(-iET)$  so we may in this context use  $\int dx^0 = T$ . Thus, setting  $iW = -iET = -iE \int dx^0$  we reduce (14.13) to:

$$E_1 = - \int \frac{d^3 k}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}. \quad (14.14)$$

The subscript in  $E_1$  designates that this is taken at the first recursive order, and this should be contrasted with (I.4.6) in [11] with which it is precisely analogous. The only difference is that at the first recursive order the term  $m'^2 \equiv -4\pi\alpha_s J_\tau J^\tau / \mathbf{k}^4$  (which means that  $m'$  is imaginary in the nature of  $i\epsilon$ ) plays a role analogous to the mass in (I.4.6) of [11], which we see very clearly if we use  $m'^2 \equiv -4\pi\alpha_s J_\tau J^\tau / \mathbf{k}^4$  to write (14.14) as:

$$E_1 = - \int \frac{d^3 k}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2 + m'^2}. \quad (14.15)$$

Now, our goal is to get from (14.14) which is an analog to (I.4.6) of [11] as we see just above, to an analog of (I.4.7) of [11], namely, the abelian potential  $E(r) = -(1/4\pi r) e^{-mr}$ . In particular, just as  $m$  in this expression alters the inverse square nature of the potential, we expect that  $m'^2 \equiv -4\pi\alpha_s J_\tau J^\tau / \mathbf{k}^4$  in (14.14) will also modify the potential away from an inverse square

potential. And in particular, one would hope – as we shall now show – that the modification stemming from  $m'^2 \equiv -4\pi\alpha_s J_\tau J^\tau / \mathbf{k}^4$  in (14.14) might lead to confinement. So we now proceed.

The challenge presented by (14.13) is that although it can be put into the form of (14.14), this term  $m'^2 \equiv -4\pi\alpha_s J_\tau(k) J^\tau(k) / \mathbf{k}^4$  is still a function of  $k$  whereas an ordinary mass  $m$  is not. So this must be accounted for in the integral over  $d^3k$ , which makes the integration much more difficult than if  $m'^2 \equiv -4\pi\alpha_s J_\tau(k) J^\tau(k) / \mathbf{k}^4$  was not a function of  $k$ , as regards both the explicitly-appearing  $\mathbf{k}^4$ , and the implicit  $J^\tau(k)$ .

As a first step to analytically calculate (14.15), let us transform over from Cartesian into spherical coordinates,  $x^i = (x, y, z) \rightarrow x'^i = (r, \theta, \varphi)$ . With this the volume element transforms over to  $d^3k = dk_x dk_y dk_z \rightarrow d^3k' = \mathbf{k}^2 dk_r \sin\theta d\theta d\varphi = -\mathbf{k}^2 dk_r d\varphi \sin\theta d\theta$ . The sign reversal in the last term arises from the differential geometry relation  $dx^2 dx^3 = -dx^3 dx^2$  because area elements not only have area magnitude, but also vector direction. So the transformed (14.14) is:

$$E_1 = -\int \frac{d^3k}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}} = \int \frac{\mathbf{k}^2 dk_r d\varphi \sin\theta d\theta}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}. \quad (14.16)$$

Now,  $\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})$  in the exponent is a scalar (dot) product in three space dimensions. So if the angle between  $\mathbf{k}$  and  $\mathbf{x}-\mathbf{y}$  is defined to be  $\theta$ , we may write  $\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) = |\mathbf{k}||\mathbf{x}-\mathbf{y}|\cos\theta$ . Then we may define  $r$  as the radial length  $r = |\mathbf{x}-\mathbf{y}|$ , so  $\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) = |\mathbf{k}|r\cos\theta$ . Further, let us do a further coordinate transformation from  $x'^2 = \theta \rightarrow x''^2 = u = \cos\theta$ . Thus,  $du = -\sin\theta d\theta$ , and also,  $\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) = |\mathbf{k}|ru$ . With all of this, and carefully attending to the ranges for the definite integrals, (14.16) becomes:

$$E_1 = \int \frac{\mathbf{k}^2 dk_r d\varphi \sin\theta d\theta}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}} = -\frac{1}{(2\pi)^2} \int_0^\infty \mathbf{k}^2 dk_r \int_0^{2\pi} \frac{d\varphi}{2\pi} \int_{-1}^1 du \frac{e^{i|\mathbf{k}|ru}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}. \quad (14.17)$$

Of course  $\int_0^{2\pi} (d\varphi/2\pi) = 1$  so that term comes out. And we can also readily perform the integral over  $du$ . Doing all of this and using  $e^{ix} - e^{-ix} = 2i \sin x$  turns (14.17) into:

$$\begin{aligned}
 E_1 &= -\frac{1}{(2\pi)^2} \int_0^\infty \mathbf{k}^2 dk_r \frac{1}{i|\mathbf{k}|r} \frac{e^{i|\mathbf{k}|ru}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}} \Bigg|_{u=-1}^{u=1} = -\frac{1}{(2\pi)^2} \frac{1}{i r} \int_0^\infty dk_r |\mathbf{k}| \frac{e^{i|\mathbf{k}|r} - e^{-i|\mathbf{k}|r}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}. \quad (14.18) \\
 &= -\frac{2}{(2\pi)^2} \frac{1}{r} \int_0^\infty dk_r |\mathbf{k}| \frac{\sin|\mathbf{k}|r}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}
 \end{aligned}$$

We now have our  $1/r$  dependence for the potential, and the final integral we need to do is the one over  $dk$ . First, using  $\int_0^{+\infty} \sin \theta d\theta = \frac{1}{2} \int_{-\infty}^{+\infty} \sin \theta d\theta$  we extend the range of the definite integral and divide by 2. Then, we may use  $i \int_{-\infty}^{\infty} \sin x = \int_{-\infty}^{\infty} e^{ix}$  because  $\cos x$  is an even function which cancels out when the magnitudes of the top and bottom of the integration range are equal, as they now are with the extension. Thus, (14.18) now becomes:

$$\begin{aligned}
 E_1 &= -\frac{2}{(2\pi)^2} \frac{1}{r} \int_0^\infty dk_r |\mathbf{k}| \frac{\sin|\mathbf{k}|r}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}} = -\frac{1}{(2\pi)^2} \frac{1}{r} \int_{-\infty}^{\infty} dk_r |\mathbf{k}| \frac{\sin|\mathbf{k}|r}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}. \quad (14.19) \\
 &= -\frac{1}{(2\pi)^2} \frac{1}{i r} \int_{-\infty}^{\infty} dk_r \frac{|\mathbf{k}| \exp i|\mathbf{k}|r}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}}.
 \end{aligned}$$

Now there is one final matter we must first resolve before we can integrate (14.19), which is that  $J^\tau(k^\sigma)$  in momentum space is itself also a function of momentum, and we are integrating over  $dk_r$ , which based on how we arrived at (14.19) is the momentum in the radial coordinate direction. Here we keep in mind that  $dk_r$  is not the same as the original  $d^4k$ , and that we have already integrated out over three of the four spacetime dimensions including time  $t$  and the two angles  $\theta = \cos^{-1} u$  and  $\phi$ . So the way to resolve the  $J^\tau(k^\sigma)$  momentum-dependency problem is to choose a frame in which  $J^\tau(k^\sigma)$  is independent of the measure  $dk_r$  in this integral. One way to do this is to transform  $J^\tau$  to the rest frame,  $J^\tau = (\rho_0, 0, 0, 0)$ , where  $\rho_0$  is the proper current density. Then, the spatial momentum vector for  $J$  will be zero,  $\mathbf{k} = 0$ , and  $J_\tau J^\tau$  in this rest frame will be independent of the  $dk_r$  measure. Then, after we have done the integral, we can use general covariance to move back out of the rest frame. So, setting  $J^\tau = (\rho_0, 0, 0, 0)$  to rest, we may write  $J_\tau J^\tau = \rho_0^2$  with  $\rho_0$  independent of the radial space integration measure  $dk_r$ , and (14.19) finally becomes:

$$E_1 = -\frac{1}{(2\pi)^2} \frac{1}{i r} \int_{-\infty}^{\infty} dk_r \frac{|\mathbf{k}| \exp i |\mathbf{k}| r}{\mathbf{k}^2 - \frac{4\pi\alpha_s \rho_0^2}{\mathbf{k}^4}} = -\frac{1}{(2\pi)^2} \frac{1}{i r} \int_{-\infty}^{\infty} dk_r \frac{|\mathbf{k}| \mathbf{k}^4 \exp i |\mathbf{k}| r}{\mathbf{k}^6 - 4\pi\alpha_s \rho_0^2}. \quad (14.20)$$

The only  $\mathbf{k}$  dependence left in (14.20) is that which explicitly appears in the  $\mathbf{k}$ .

Now we must embark upon the remaining integral, which will utilize the method of contour integration. As now constituted,  $|\mathbf{k}| = \sqrt{\mathbf{k}^2}$  and  $\mathbf{k}^2$  hence  $\mathbf{k}^4$  are all ordinary variables. Thus, we may turn every real  $\mathbf{k}$  in the above into a complex variable  $z$  and then do the integral over a suitable contour. Specifically, the integral we now seek to evaluate based on extending into the complex plane via  $|\mathbf{k}| \rightarrow z$  is:

$$\oint_C f(z) dz = \oint_C dz \frac{z \exp izr}{z^2 - \frac{4\pi\alpha_s \rho_0^2}{z^4}} = \oint_C dz \frac{z^5 \exp izr}{z^6 - 4\pi\alpha_s \rho_0^2}. \quad (14.21)$$

We now need to separate this out using the method of partial fractions. But as a predicate for doing this, we first need to work with the denominator  $z^6 - 4\pi\alpha_s \rho_0^2$  which, with  $z = x^2$  and  $d = -4\pi\alpha_s \rho_0^2$ , takes the general form of a cubic equation  $f(x) = ax^3 + bx^2 + cx + d$  with  $a = 1$  and  $b = c = 0$ . A good online reference to help evaluate the roots of this function is [23], from which it may be shown that:

$$z^6 - 4\pi\alpha_s \rho_0^2 = \left( z^2 - (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right) \left( z^2 - \frac{-1+i\sqrt{3}}{2} (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right) \left( z^2 - \frac{-1-i\sqrt{3}}{2} (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right). \quad (14.22)$$

Therefore, the contour integral is:

$$\begin{aligned} \oint_C f(z) dz &= \oint_C dz \frac{z^5 \exp izr}{z^6 - 4\pi\alpha_s \rho_0^2} \\ &= \oint_C dz \frac{z^5 \exp izr}{\left( z^2 - (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right) \left( z^2 - \frac{-1+i\sqrt{3}}{2} (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right) \left( z^2 - \frac{-1-i\sqrt{3}}{2} (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right)}. \end{aligned} \quad (14.23)$$

Now we can separate this into a sum of three distinct contour integrals via partial fractions. The result of this exercise is:

$$\begin{aligned} \oint_C f(z) dz &= \oint_C dz \frac{z^5 \exp izr}{z^6 - 4\pi\alpha_s \rho_0^2} \\ &= \oint_{C_1} dz \frac{z \exp izr}{3 \left( z^2 - (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right)} + \oint_{C_2} dz \frac{z \exp izr}{3 \left( z^2 - \frac{-1+i\sqrt{3}}{2} (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right)} + \oint_{C_3} dz \frac{z \exp izr}{3 \left( z^2 - \frac{-1-i\sqrt{3}}{2} (4\pi\alpha_s)^{\frac{1}{3}} \rho_0^{\frac{2}{3}} \right)}. \end{aligned} \quad (14.24)$$

Now we need to directly reveal the poles of the first order  $z$ , so we can obtain the residues and complete the Cauchy integration. It is readily appreciated that with an explicit pole separation, (14.24) may be further written as:

$$\begin{aligned} \oint_C f(z) dz &= \oint_C dz \frac{z^5 \exp izr}{z^6 - 4\pi\alpha_s \rho_0^2} \\ &= \oint_{C_1} dz \frac{z \exp izr}{3 \left( z + (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right) \left( z - (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)} \\ &+ \oint_{C_2} dz \frac{z \exp izr}{3 \left( z + \sqrt{\frac{-1+i\sqrt{3}}{2}} (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right) \left( z - \sqrt{\frac{-1+i\sqrt{3}}{2}} (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)} \\ &+ \oint_{C_3} dz \frac{z \exp izr}{3 \left( z + \sqrt{\frac{-1-i\sqrt{3}}{2}} (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right) \left( z - \sqrt{\frac{-1-i\sqrt{3}}{2}} (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)}. \end{aligned} \quad (14.25)$$

This contains six poles and it also contains square roots of the cubed roots of unity, which are thus sixth roots of unity. Additionally, in  $(4\pi\alpha_s)^{\frac{1}{6}}$  and  $\rho_0^{\frac{1}{3}}$  we see that other sixth roots have been taken to arrive at (14.25). For two of these sixth roots of unity, it is readily seen that:

$$\sqrt{\frac{-1+i\sqrt{3}}{2}} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i; \quad \sqrt{\frac{-1-i\sqrt{3}}{2}} = -\frac{1}{2} - \frac{\sqrt{3}}{2}i \quad (14.26)$$

which enables us to rewrite (14.25) as:

$$\begin{aligned}
 \oint_C f(z) dz &= \oint_C dz \frac{z^5 \exp izr}{z^6 - 4\pi\alpha_s \rho_0^2} \\
 &= \oint_{C_1} dz \frac{z \exp izr}{3 \left( z + (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right) \left( z - (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)} \\
 &+ \oint_{C_2} dz \frac{z \exp izr}{3 \left( z + \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right) \left( z - \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)} \\
 &+ \oint_{C_3} dz \frac{z \exp izr}{3 \left( z + \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right) \left( z - \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)} .
 \end{aligned} \tag{14.27}$$

Now, we use the three roots  $z_1 = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$ ,  $z_2 = \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$  and  $z_3 = \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$  to extract the residue and evaluate the integral which includes an overall multiplication by  $2\pi i$  which is standard in such integrals. What we obtain is:

$$\begin{aligned}
 \oint_C f(z) dz &= \oint_C dz \frac{z^5 \exp izr}{z^6 - 4\pi\alpha_s \rho_0^2} \\
 &= 2\pi i \frac{1}{6} \left[ \exp \left( i (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \right) + \exp \left( i \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \right) + \exp \left( i \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \right) \right] .
 \end{aligned} \tag{14.28}$$

The overall integral evaluated above is  $\oint_C f(z) dz = \int_{-\infty}^{\infty} f(z) dz + \int_{\text{Arc}} f(z) dz$ , which includes both the entire range over the real arguments  $\int_{-\infty}^{\infty} f(z) dz$  as well as  $\int_{\text{Arc}} f(z) dz$  which represents the contour arc through the complex plane. But it is readily shown and is well-known that for an integral of the form (14.27),  $\int_{\text{Arc}} f(z) dz = 0$ . So (14.28) is a complete result, and it may therefore be equated back to the original integral in (14.20). So we now have:

$$\begin{aligned}
 \oint_C f(z) dz &= \oint_C dz \frac{z^5 \exp izr}{z^6 - 4\pi\alpha_s \rho_0^2} = \int_{-\infty}^{\infty} dk_r \frac{|\mathbf{k}| \exp i|\mathbf{k}|r}{\mathbf{k}^2 - \frac{4\pi\alpha_s \rho_0^2}{\mathbf{k}^4}} \\
 &= 2\pi i \frac{1}{6} \left[ \exp \left( i (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \right) + \exp \left( i \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \right) + \exp \left( i \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \right) \right]
 \end{aligned} \tag{14.29}$$

Then, we use the above in (14.20) to compute the potential, which is:

$$E_1 = -\frac{1}{(2\pi)^2} \frac{1}{i r} \int_{-\infty}^{\infty} dk_r \frac{|\mathbf{k}| \mathbf{k}^4 \exp i|\mathbf{k}|r}{\mathbf{k}^6 - 4\pi\alpha_s \rho_0^2} \quad (14.30)$$

$$= -\frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \left[ \exp\left(i(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r\right) + \exp\left(i\left(\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r\right) + \exp\left(i\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r\right) \right]$$

We see that for  $\alpha_s \rightarrow 0$  which is the regime in which quarks are asymptotically free, this will reduce to:

$$E_1 \xrightarrow{\alpha_s=0} -\frac{1}{4\pi} \frac{1}{r} \frac{1}{3} (3) = -\frac{1}{4\pi} \frac{1}{r} = -\frac{1}{4\pi} r^{-1}. \quad (14.31)$$

which is the inverse-square law potential that is in [I.4.7] of [11]. *This is an important check that our calculation properly reduces to the expected result when the strong coupling is small.* But (14.30) also contains a radial dependence inside the square brackets, which should give us confinement if all is well.

Proceeding, we further simplify (14.30) by separating the roots of unity in the latter two of the three terms in (14.30) into real and imaginary parts and then using the hyperbolic function  $2 \cosh x = e^x + e^{-x}$ . We also note that  $\rho_0^2 = J_\sigma J^\sigma$ . Although we earlier set  $J_\sigma$  to be at rest in order to get through the integration over  $dk_r$ , after the integration everything is safely in configuration space and so we can use general covariance to transform back out of the rest frame and insert  $\rho_0^{\frac{1}{3}} = (J_\sigma J^\sigma)^{\frac{1}{6}}$  into (14.30). But the proper current density cubed root  $\rho_0^{\frac{1}{3}}$  is itself a scalar number in spacetime, so it is simpler to leave this as is. This yields our final result:

$$E_1 = -\frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \left[ \exp\left(i(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r\right) + 2 \exp\left(i\frac{1}{2}(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r\right) \right] \quad (14.32)$$

It now helps to graph the behavior of this function. If we set the “frequency” coefficient which is common to all three terms to  $f \equiv (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$  and regard this at this time to be a constant (later we shall examine other behaviors for this), and also scale out the lead amplitude coefficient  $A \equiv 1/12\pi$  except for the negative sign of the potential, then we can somewhat unclutter the above by writing:

$$E_1 = -Ar^{-1} \left[ \exp(ifr) + 2 \exp\left(i\frac{1}{2}fr\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}fr\right) \right]. \quad (14.33)$$

Defining a dimensionless  $R \equiv fr \equiv (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r$  we further rewrite this as:

$$E_1 / Af = -R^{-1} \left[ \exp(iR) + 2 \exp\left(i \frac{1}{2} R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} R\right) \right]. \quad (14.34)$$

Now, this is a complex number. Because observables energies in physics are real numbers, we will wish to ascertain the square magnitude (modulus)  $|E_1|^2 = E_1 * E_1$  and then obtain  $|E_1| = \pm \sqrt{|E_1|^2}$ . (We use  $\pm$  because as with any square root of a positive number, this can be either positive or negative. The  $| |$  symbol here is not for the absolute value but for the real magnitude.) Therefore, the real magnitude of (14.34) may be written as:

$$\begin{aligned} |E_1| / Af &= -R^{-1} \left[ \cos(R) + 2 \cos\left(\frac{1}{2} R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} R\right) + i \left[ \sin(R) + 2 \sin\left(\frac{1}{2} R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} R\right) \right] \right] \\ &= -R^{-1} [a + bi] = -R^{-1} \sqrt{|a + bi|^2} = -R^{-1} \sqrt{(a + bi)(a - bi)} = -R^{-1} \sqrt{a^2 + b^2} \end{aligned} \quad (14.35)$$

where we have defined the real and imaginary portions of the complex  $E_1$  via:

$$\begin{aligned} a &\equiv \cos(R) + 2 \cos\left(\frac{1}{2} R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} R\right) \\ b &\equiv \sin(R) + 2 \sin\left(\frac{1}{2} R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} R\right) \end{aligned} \quad (14.36)$$

It is readily seen that the square magnitude:

$$\begin{aligned} |a + bi|^2 &= a^2 + b^2 \\ &= \cos^2(R) + 4 \cos^2\left(\frac{1}{2} R\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{2} R\right) + 4 \cos(R) \cos\left(\frac{1}{2} R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} R\right) \\ &\quad + \sin^2(R) + 4 \sin^2\left(\frac{1}{2} R\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{2} R\right) + 4 \sin(R) \sin\left(\frac{1}{2} R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} R\right) \\ &= 1 + 4 \cosh^2\left(\frac{\sqrt{3}}{2} R\right) + 4 \left[ \cos(R) \cos\left(\frac{1}{2} R\right) + \sin(R) \sin\left(\frac{1}{2} R\right) \right] \cosh\left(\frac{\sqrt{3}}{2} R\right) \end{aligned} \quad (14.37)$$

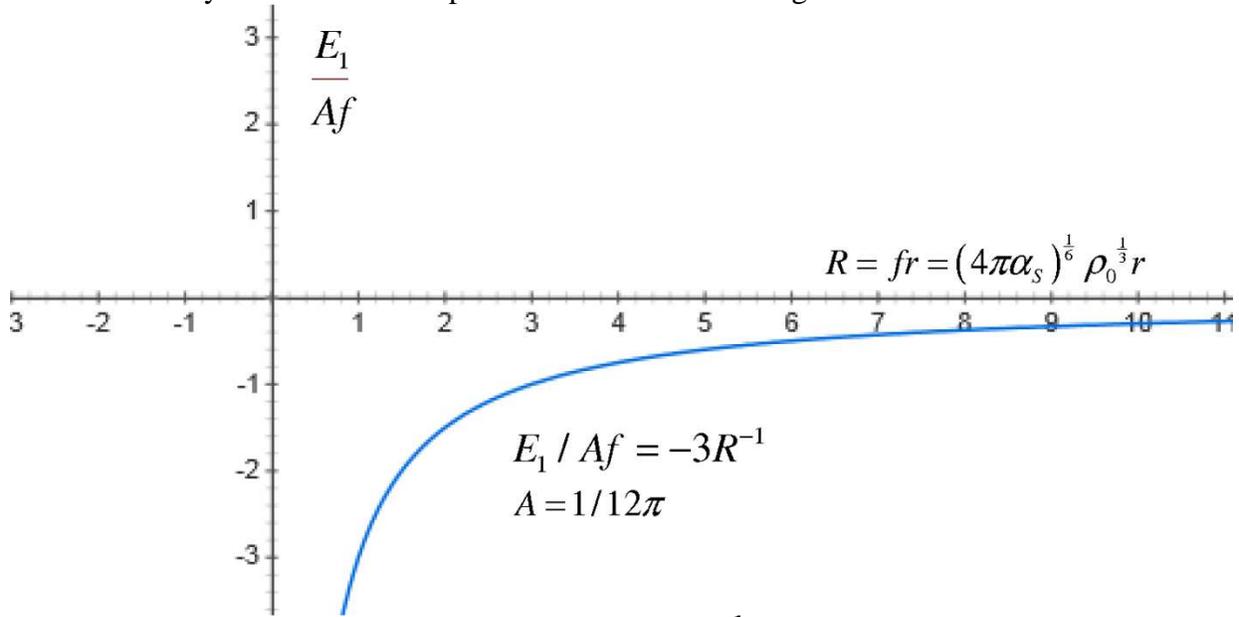
This is not separable via the quadratic equation because it does not cross the  $R$  axis and so has no real roots. So  $|a + bi| = \pm \sqrt{|a + bi|^2}$  is just the positive and negative square roots of the above:

$$|a + bi| = \pm \sqrt{1 + 4 \cosh^2\left(\frac{\sqrt{3}}{2} R\right) + 4 \left[ \cos(R) \cos\left(\frac{1}{2} R\right) + \sin(R) \sin\left(\frac{1}{2} R\right) \right] \cosh\left(\frac{\sqrt{3}}{2} R\right)}. \quad (14.38)$$

Consequently, we use this in (14.35) to write the observable magnitude  $|E_1|$  as:

$$|E_1| / Af = \mp R^{-1} \sqrt{1 + 4 \cosh^2\left(\frac{\sqrt{3}}{2} R\right) + 4 \left[ \cos(R) \cos\left(\frac{1}{2} R\right) + \sin(R) \sin\left(\frac{1}{2} R\right) \right] \cosh\left(\frac{\sqrt{3}}{2} R\right)}. \quad (14.39)$$

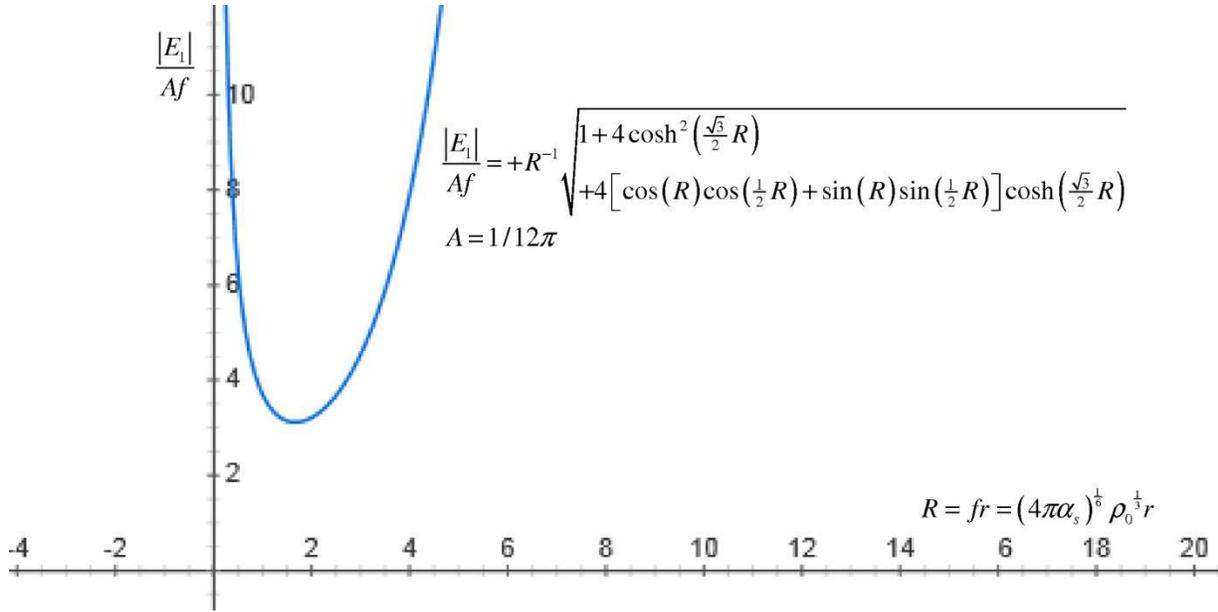
At this point we would like to graph  $|E_1|$  as a function of  $R$ . But before we do so, as a baseline for discussion we plot the ordinary  $E_1 = -(1/4\pi)r^{-1}$  potential of (14.31). To plot this on scales and with variables that can be compared directly to those in (14.39) we use  $A \equiv 1/12\pi$  and  $R \equiv fr \equiv (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r$  to rewrite this potential as  $E_1 / Af = -3R^{-1}$ . The radial distance in spherical coordinates is always taken to be a positive number, so we only show the curve for  $r > 0$ . This very familiar baseline potential is illustrated in Figure 1 below.



**Figure 1: The Ordinary  $R^{-1}$  Potential**

For a charge situated at  $r$  which is unlike-charged in relation to the charge sourcing this potential, the natural “geodesic” tendency will always be to seek the lowest possible potential, so that a charge at  $R$  will trend toward the left of the above graph and move closer to  $R=0$ . The like-charge potential is simply the mirror image of Figure 1 flipped about the  $R$  axis, i.e.,  $E_1 / Af = +3R^{-1}$ . So, two like-charges will naturally tend to push further apart. Additionally, a charge situated at large  $R$  does not require a whole lot of energy to separate even further, because of the manner in which  $E_1 / Af = -3R^{-1}$  asymptotically approaches the  $r$  axis from below. Because of this asymptotic behavior of the potential for large  $R$ , there is nothing in the Figure 1 potential to “confine” this charge. With the provision of sufficient finite, small energy, this charge is free to move all the way out to  $R \rightarrow \infty$ .

Now let’s graph (14.39). We see from (14.38) that it is possible to use either the negative or positive sign. One choice will yield a like-charge potential, the other an unlike-charge potential, and we will need to ascertain which is which. For reasons that will become momentarily apparent, we graph this using the negative sign from (14.38) which produces an overall positive sign. The graph is shown below in Figure 1a:



**Figure 2: The Yang-Mills Potential  $|E_1|$  of (14.39) at First Recursive Order**

The  $+1/R$  potential dominates the behavior of this curve at small  $R$ . But at large  $R$  the terms containing  $\cosh\left(\frac{\sqrt{3}}{2}R\right)$  take over and force the curve to reverse and become infinitely large in the same way as does the hyperbolic cosine function. Between these two domains there is a minimum in the potential at approximately  $(R, |E_1|/Af) \approx (1.668, 3.118)$ . So by least action / least potential principles, a charge situated in this potential will tend to seek this minimum point at  $R \approx 1.668$ . Starting from this minimum, energy is required to pull the charge further away from  $R=0$  and also to push the charge closer to  $R=0$ . So the force  $F_1 = \partial|E_1|/\partial R$  associated with this potential is attractive for  $R > 1.668$ , repulsive for  $R < 1.668$  and zero at  $R \approx 1.668$ .

Although (14.39) is only an abelian simplification as noted prior to (14.5), we nonetheless see in Figure 1a the first indication that the Yang-Mills potential, when  $f \equiv (4\pi\alpha_s)^{1/6} \rho_0^{1/3}$  is defined to be constant, is a *confining* potential. Additionally, it is a *stable* potential, because at the same time it operates against a charge being removed to a separation much greater than  $R \approx 1.668$ , it also prevents a charge from getting too close to the source of the potential, because the potential  $\propto +1/r$  for  $r \rightarrow 0$ . We shall develop and explore these two aspects of the Yang-Mills potential in great depth in the next few sections, together with exploring its asymptotic freedom.

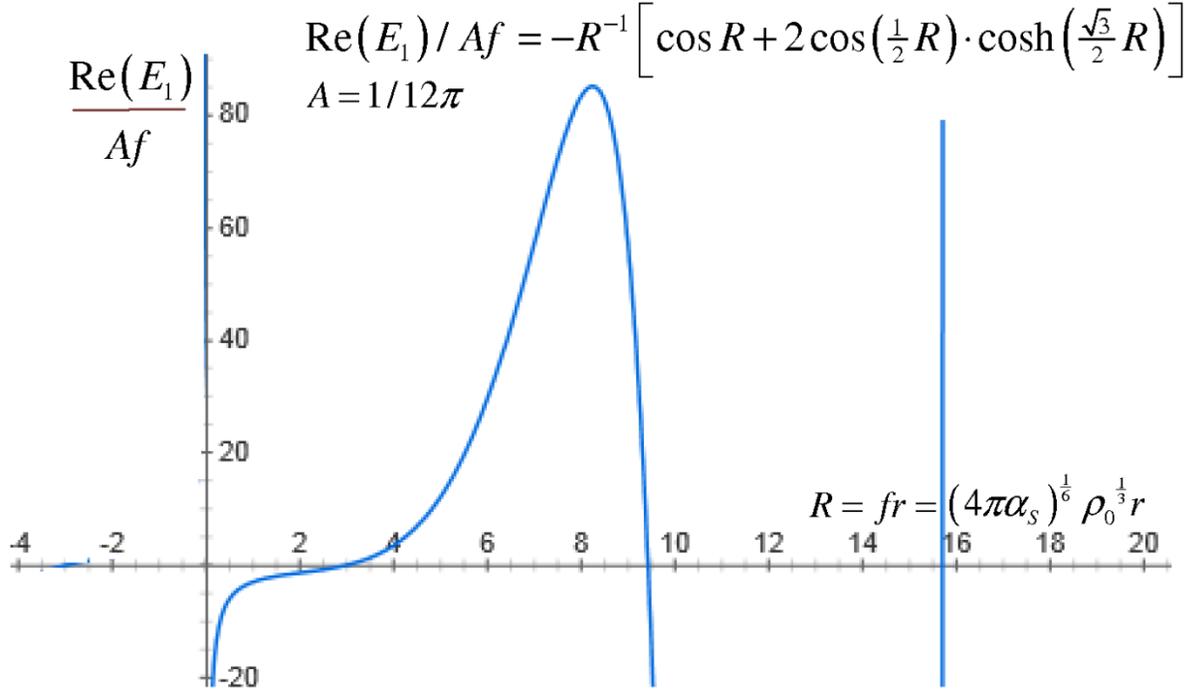
We proceed to gain additional insight into this potential if we examine the real portion of (14.34), (14.35) by itself, namely:

$$\text{Re}(E_1)/Af = -R^{-1} \left[ \cos R + 2 \cos\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \right]. \quad (14.40)$$

Because  $R = fr = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r$ , in the  $\alpha_s \rightarrow 0$  limit the square bracket term is still equal to 3 as in (14.31) and the dimensionless version of (14.31) becomes.

$$E_1 / Af = -3R^{-1}. \tag{14.41}$$

As is to be expected, this is the  $1/r$  potential of (14.31) and Figure 1. If we now plot (14.40) using  $R = fr = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r$  while still treating  $f \equiv (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$  as a *constant* frequency as we did for Figures 1 and 2, the result is Figure 3 below:



**Figure 3: The Yang-Mills Potential  $\text{Re} E_1$  of (14.40) at First Recursive Order**

This is the exact same function based on (14.34) as that shown in Figure 2, except here we are looking at  $\text{Re} E_1$  rather than  $|E_1|$ . We expect that  $|E_1|$  will be an abelian-simplified, first-recursive order approximation to an observable potential, and that  $\text{Re} E_1$  in Figure 3, although not observable because it truncates the imaginary contributions to  $|E_1|$ , can still give us some valuable insights into the observable Figure 2 potential. Specifically, in Figure 3 we see the usual  $-1/R$  potential melded with a confining potential that begins its uptick in the vicinity of  $R \approx 2$ . Both Figure 2 and Figure 3 show this confining potential, but in Figure 2 the potential continues to rise to infinity without ever retreating, while in Figure 3 this potential peaks at around  $(R, \text{Re}(E_1)/Af) \approx (8.245, 85.184)$  and then retreats due to the sinusoidal behaviors that disappear in Figure 2, see the term reductions in (14.37). What Figure 3 clues us into, which Figure 2 does not, is a *length scale* for these confining behaviors. Specifically, because the first peak at  $R \approx 8.245$  is a natural length scale embedded in (14.34), and because both Figures 2 and

3 show a confining potential, we shall now wish to associate  $R \approx 8.245$  with some physically-based radius  $r$  which is indicative of confinement.

This brings us to the question whether Figure 2 is the potential for like or unlike charges. On the one hand, for  $R \rightarrow 0$ , this approximates to a  $+1/R$  potential, which is the potential for like-charges repelling. The  $-1/R$  potential for unlike-charges attracting is what appears for  $R \rightarrow 0$  in Figures 1 and 3. So, one might conclude that Figure 2 is a potential for like charges repelling. But at the same time, for  $R > 1.668$  Figure 2 becomes confining, and so looks like a potential for unlike charges attracting. And so we come to the crux: in Figure 2 there is a region of repulsion for  $R < 1.668$ , a region of attraction for  $R > 1.668$ , and a region of stability at  $R \approx 1.668$ , as already pointed out. But the charges do not switch from being like to unlike; it is the potential that changes its character as a function of  $r$ . Put differently, we normally take the view that unlike charges attract and like charges repel because the potentials we usually have available, such as  $1/r$ , do not at any point switch from repulsive to attractive. But the potential in Figure 3 does switch from repulsive to attractive, and that is a very desirable feature of this potential because this renders it both confining and stable. So how do we interpret this desirable feature of Figure 2?

Even since the time of Newton, and later Coulomb,  $-1/r$  potentials have been a central fixture of theoretical physics. This is because  $1/r$  potentials are central to both gravitation and electromagnetism. But at the same time, it has been widely understood – at least qualitatively – that if one truly were able to experimentally study a  $-1/r$  potential for  $r \rightarrow 0$ , *where zero really meant zero*, e.g., where  $r$  was a length even smaller than the Planck length, or maybe even a nuclear or atomic length, at some point the  $-1/r$  potential would no longer apply. That is, it has long been understood that while  $-1/r$  potential has very wide applicability to macroscopic length scales, its range of validity in the smaller scale is not expected to be unlimited. This was one of the problems confronted after the turn of the 20<sup>th</sup> century, when the Bohr model of the atom began to provide a mechanism for stopping the electron from otherwise losing energy and spiraling into the nucleus as its potential  $1/4\pi r \rightarrow -\infty$ . So in general, it has been shown that it is wise to treat with caution, the applicability of a  $-1/r$  for extremely small  $r$  in the atomic domain and below. On the other hand, a  $+1/r$  potential, often associated with the repulsion of like-charges, does prevent a physically viable picture for small  $r$ . This potential, for example, would prevent two protons from ever collapsing into one another. But if there was nothing further, the protons would repel and fly apart, and there would be no atomic nuclei. This is really the inverse of the problem of the electron spiral, and of course, today we know that the strong interaction is what prevents this from occurring.

But what is still missing from present-day understanding, is a *single* potential curve *derived from quantum field theory* (as opposed to a potential which is simply postulated such as the  $V(\phi) \equiv \mu^2 \phi^* \phi + \lambda (\phi^* \phi)^2$  potential commonly employed in scalar field theory) which permits systems of unlike charges to simultaneously a) not collapse at extremely close distance, and b) not disintegrate at larger distances. This is not dissimilar to problem that Max Planck confronted at the turn of the 20<sup>th</sup> century in trying to meld together the Wein curve for short wavelengths with the Rayleigh–Jeans curve for larger wavelengths. Figure 2, while an abelian simplification limited to the first recursive order of non-linear quantum field theory, is the type of stable potential that is required to seamlessly meld attraction and repulsion into a single

potential curve that provides stability and avoids either system collapse or system disintegration. And because it has features of both attraction and repulsion, Figure 2 appears to be a stable, confining-yet-collapse-averting potential for *unlike* charges, and more specifically, the type of potential which is qualitatively suitable to govern the behaviors of the R, G, B quarks of unlike color, inside a baryon. Again, while Figures 2 and 3 are approximations based on an abelian simplification and consideration of only the first recursive order of non-linear quantum field theory, these very desirable features will cause us to develop study this form of potential much more closely in the next several sections.

Before concluding, let us do some order of magnitude calculations using dimensional analysis, based on all of the foregoing. We see in Figure 3 for the real portion of the potential that a sharp rise in the potential starts to occur in the vicinity of  $R \approx 2$  to  $R \approx 4$  and peaks in the vicinity of the dimensionless  $R = fr = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \approx 8.245$ . All experimental evidence suggests that there are exactly six quark flavors existing in nature, and we know from empirical data that for six quarks,  $\Lambda^{(6)}_{QCD} = (90.6 \pm 3.4) MeV$  is the strong interaction cutoff arrived at through dimensional transmutation, see [9.24a] from PDG's [24]. The deBroglie relation  $E = \hbar c / \lambda$  enables any mass / energy value  $E = mc^2$  to be represented by an equivalent reduced length scale  $\lambda$ . In natural  $\hbar = c = 1$  units,  $E = 1 / \lambda$ . The PDG data at [25] states that the "wavelength of a 1 eV/c particle"  $hc / (1eV) = 1.239\ 841\ 875(31) \times 10^{-6} m$ . Via a reduced  $\hbar = h / 2\pi$  this is alternatively expressed as the GeV-to-Fermi conversion  $1GeV^{-1} = .197\ 326\ 963\ 1F$  or  $1F = 5.067\ 731\ 163\ GeV^{-1}$ , see also [11], Table 1.2b. These conversions effectively provide a shortcut to transpose between  $E$  and  $\lambda$  in  $E = \hbar c^3 / \lambda$  without having to explicitly use  $\hbar$  and  $c$ , which is the practical upshot of using  $\hbar = c = 1$  units. So via  $1GeV = 5.067\ 731\ 163F^{-1}$ , the mean empirical value  $\Lambda^{(6)}_{QCD} = .0906GeV$  has a radial length equivalent via  $E = 1 / \lambda$ , of  $\Lambda^{(6)}_{QCD} = .0906GeV = .4591F^{-1} = 1 / 2.178F \equiv 1 / r_\Lambda$ . So we are able to define  $r_\Lambda \equiv 2.178F$  as the approximate radial distance associated with the six quark QCD cutoff  $\Lambda^{(6)}_{QCD}$  which we henceforth denote simply as  $\Lambda$ .

Because this cutoff radius is the approximate empirical radial length at which the QCD coupling  $\alpha_s \rightarrow \infty$  and thereby confines quarks in six quark models, this  $r_\Lambda \equiv 2.178F$  is the length scale at which we expect the confining potential inside a nucleon to peak. Although Figure 3 only accounts for the real portion of the potential and is abelian and is based on the first recursive order only, it does provide a dimensionless  $R \approx 8.245$  at which the potential peaks in Figure 3. So in order to introduce an observable physical length / energy scale into Figures 2 and 3, we now associate  $r_\Lambda \equiv 2.178F$  at which  $\alpha_s \rightarrow \infty$  in six-quark theory with  $R \approx 8.245$  at which the real portion of the potential peaks in Figure 1. That is, we now regard  $r_\Lambda = 2.178F \Leftrightarrow R \approx 8.245$  to be two equivalent ways of expressing the same *physical* radius, the former dimensional based on empirical data, and the latter dimensionless and based on the theory that led to Figures 2 and 3. So using the definition  $R \equiv fr \equiv (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \equiv \text{constant}$  leading to (14.34), this means that we can implement the association  $r_\Lambda = 2.178F \Leftrightarrow R \approx 8.245$  by way of:

$$8.245 \equiv R = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} 2.178F . \quad (14.42)$$

We may then invert this and use  $1F^{-1} = .197\ 326\ 963\ 1\text{GeV}$  to obtain:

$$(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} = 8.245 \cdot .459F^{-1} = 3.784F^{-1} = .747\text{GeV} = 747\text{MeV} . \quad (14.43)$$

So we see a (cubed root) density number that, on an order-of-magnitude basis, is what we expect to see when we are talking about baryons such as protons and neutrons with  $m_p = 938.272046(21)\ \text{MeV}$  and  $m_n = 939.565379(21)\ \text{MeV}$  respectively, see, e.g., [26], which establish the lower range of baryon masses which generally run from about 1GeV to about 6GeV, see, e.g., [27].

There is also another way to understand (14.43), which in view of  $4\pi\alpha_s = g_s^2 / \hbar c$ , is to cube everything and then write (14.43) as:

$$\sqrt{4\pi\alpha_s} \rho_0 = g_s \rho_0 = (.747\text{GeV})^3 . \quad (14.44)$$

This contains the running of the coupling  $\alpha_s$  and the bare (uncoupled) proper density  $\rho_0$  of the quark currents in natural units. Because  $.747\text{GeV}$  is a constant energy, as the strong charge  $g_s = \sqrt{4\pi\alpha_s}$  grows the bare proper density diminishes in inverse proportion. As  $g_s = \sqrt{4\pi\alpha_s}$  weakens the bare proper density grows larger in inverse proportion. So the conclusion is clear: *bare* quark current densities  $\rho_0$  will be more greatly-concentrated where the running coupling  $\alpha_s$  is smaller, and less-concentrated where the running coupling is larger. For an infinite running coupling  $\alpha_s \rightarrow \infty$ , the bare proper density  $\rho_0 \rightarrow 0$ . There are of course many reasons to believe that confinement and the existence of a mass gap are related to the running of the coupling constant, which is an inherently quantum effect. The above is yet another way of using the dimensional analysis in (14.42) to (14.44) to better understand the nature of confinement in relation to this running of the strong coupling.

So, with all of these results, we have fully converted from the classical analysis of sections 1 through 10 which gave us evidence that the magnetic monopoles of Yang-Mills theory have many of the symmetry features of baryons and that the electric current densities of Yang-Mills theory similarly mirror the quark currents, to a complete analysis relying upon quantum field theory. We confirm via (14.32) and its visualization in Figures 2 and 3 that even at first recursive order, and even with an abelian simplification, there is a very definitive appearance of confinement in the form of a potential that grows increasingly rapidly at distances larger than  $R = fr = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r \cong 3$ , which based on  $r_\Lambda = 2.178F \Leftrightarrow R \cong 8.245$ , has the correspondence  $r = .792F \Leftrightarrow R = 3$ . In other words, all of this suggests that confinement starts to kick in, in earnest, once we try to separate two  $J$  by more than  $R \cong 3$ , which is the dimensional length

$r_\Lambda = .792F$  based on associating  $r_\Lambda = 2.178F \Leftrightarrow R \cong 8.245$ . So as predicted by the theoretical results here in combination with the empirical  $\Lambda_{QCD}^{(6)} = (90.6 \pm 3.4) MeV$ , it is at about  $.792F$  that one crosses over from asymptotic freedom and starts to encounter an uptick in the quantum potential  $E_1$  which we associate with confinement.

## 15. Direct Quantum Field Theory Demonstration of Confinement – Non-Abelian Generalization

At (14.5) in the prior section, we introduced an abelian simplification of the amplitude  $\mathcal{N}(J)_1 = (J_\sigma \pi_1 J^\sigma)$  which enabled us to treat  $4\pi\alpha_s J_\sigma (k_\tau k^\tau + i\varepsilon + 4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\varepsilon)^2)^{-1}$  as an ordinary denominator and treat the currents as commuting  $[J^\sigma, J^\tau] = 0$ . Based on this simplification we arrived in (14.32). Now we wish to extend (14.32) to the generalized non-abelian relationship  $[J^\sigma, J^\tau] = J^{i\sigma} J^{j\tau} [\lambda^i, \lambda^j] = if^{ijk} \lambda^k J^{i\sigma} J^{j\tau} \neq 0$ . To do so, we will identify precisely those points in the calculation of the previous section where the term  $if^{ijk} \lambda^k J^{i\sigma} J^{j\tau} \neq 0$  was neglected, and thereby pinpoint what would be the generalized form of (14.32) had we not neglected this term. There are two aspects to what we did which now need to be updated: first treating the inverse as an ordinary denominator, and second, commuting with  $[J^\sigma, J^\tau] = 0$ . We take these in turn.

As to inverses, for a square matrix  $M$  the inverse  $M^{-1}$  is *defined* generally (and often deduced) by  $M \cdot M^{-1} \equiv \delta$ , where  $\delta$  is a diagonal unit matrix. It is customary to use the mathematical notation  $M^{-1}$  rather than  $1/M$  for two reasons. First, this serves as a mnemonic reminder that the object  $M$  is a matrix and not an ordinary number. But this is just a symbolic convenience, and one could still write  $1/M$  rather than  $M^{-1}$  so long as one was very careful to keep in mind that an object is a matrix and make sure that whenever  $1/M$  was in fact calculated, this calculation was performed using  $M \cdot (1/M) \equiv \delta$ . The more serious issue is that matrices in general are not commuting, and so the use of  $M^{-1}$  rather than  $1/M$  generally serves as a placeholder to hold the commutation position of the matrix inverse in what may otherwise be a string of matrix multiplications for which left-right ordering matters. For example, suppose we have three square matrices  $A$ ,  $B$  and  $C$  which have the relationship  $A = BC^{-1}$ . If we wanted to rearrange, we could multiply *from the right* by  $C$ , and thus obtain  $AC = BC^{-1}C = B$ . The fact that we use  $A = BC^{-1}$  rather than  $A = B/C$  tells us that a right multiplication is in order. Had we instead started merely with  $A = B/C$  we could end up with either  $AC = B$  or  $CA = B$ , but the former would be right and the latter would be wrong. Here too, one could still write  $A = B/C$  and try to *remember* that  $A = BC^{-1}$  and not  $A = C^{-1}B$  is the original relationship, but that is not best practice for two reasons. First, this illustrates that  $A = B/C$  is notationally *ambiguous*, and mathematical notation should be unambiguous. Second, in a complicated calculation (such as the one in the last section) where original expressions undergo substantial

metamorphosis, it may be difficult to properly account for or reconstruct left and right ordering even with a good memory.

And yet, there are benefits to the  $1/M$  notation, especially when it comes to visualizing cancellations of terms between a numerator and denominator, as well as carrying out other mathematical operations which do not involve an explicit calculation of  $M^{-1}$  via  $M \cdot M^{-1} \equiv \delta$ . We saw an example of this in (7.20) where we used the subscripted down-arrow “ $\sphericalangle$ ” symbol as a marker to denote and hold commutativity position in combination with the usual “divide by” symbol “/” the eliminate this ambiguity in lieu of using the “ $^{-1}$ ” notation. So with the present example, we could write  $A = BC^{-1}$  as  $A = B_{\sphericalangle} / C$  and in this way continue to use an ordinary “/” symbol without ambiguity. If the posited relationship was  $A = C^{-1}B$  we could then write  $A = \sphericalangle B / C$  and similarly have an unambiguous expression. As an additional benefit, it may turn out that after a complicated calculation is complete, a matrix  $M$  which started out in the form  $M^{-1}$  ends up being reinverted back to  $M$  without it ever becoming necessary to do the explicit calculation of  $M^{-1}$  via  $M \cdot M^{-1} \equiv \delta$ . A good example of this is the result (14.32) presently under review: We already pointed out prior to (14.32) that one could revert to  $\rho_0^2 = J_\sigma J^\sigma$ . Now we do exactly that to rewrite this as:

$$E_1 = -\frac{1}{12\pi} r^{-1} \left[ \exp\left(i(4\pi\alpha_s J_\sigma J^\sigma)^{\frac{1}{6}} r\right) + 2 \exp\left(i\frac{1}{2}(4\pi\alpha_s J_\sigma J^\sigma)^{\frac{1}{6}} r\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}(4\pi\alpha_s J_\sigma J^\sigma)^{\frac{1}{6}} r\right) \right]. \quad (15.1)$$

Although  $4\pi\alpha_s J_\tau J^\tau$  originated in the inverse term  $\left(k_\tau k^\tau + i\varepsilon + 4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\varepsilon)^2\right)^{-1}$  in (14.3), by the time we completed the complex set of calculations that led to (14.32), (15.1), this inverted appearance of  $4\pi\alpha_s J_\tau J^\tau$  ended up reinverted in the form of (15.1) and there was no need to explicitly calculate  $(J_\tau J^\tau)^{-1}$  via  $J_\tau J^\tau \cdot (J_\tau J^\tau)^{-1} \equiv \delta$  or make use of this inverse in a string of other matrix multiplications. Instead, we found that the cubic and sixth root mathematics that started with the denominator  $z^6 - 4\pi\alpha_s \rho_0^2$  in (14.21), ended up with the original  $4\pi\alpha_s J_\tau J^\tau$  appearing in an uninverted sixth root  $(4\pi\alpha_s J_\sigma J^\sigma)^{\frac{1}{6}}$ .

As to the second aspect, commutation, we now ask: how would (15.1) change if we had done the calculation of the last section using  $[J^\sigma, J^\tau] = J^{i\sigma} J^{j\tau} [\lambda^i, \lambda^j] = i f^{ijk} \lambda^k J^{i\sigma} J^{j\tau} \neq 0$  rather than  $[J^\sigma, J^\tau] = 0$ ? Or, stated differently, how does (15.1) generalize if the  $J$  which it contains commute via  $[J^\sigma, J^\tau] \neq 0$  rather than  $[J^\sigma, J^\tau] = 0$ ? If the calculation of the last section had used no commutations between two or more  $J$  with different spacetime indexes, then (15.1) would remain as is. But if there was a commutation – as there was in in going from (14.9) to (14.10) – then (15.1) needs to be reviewed and possibly amended for  $[J^\sigma, J^\tau] \neq 0$ . Let us now trace this through.

The abelian simplification began at (14.5) where we treated the inverse  $I$  which we reorder as  $I = \left( 4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\mathcal{E})^2 + k_\tau k^\tau + i\mathcal{E} \right)^{-1}$ , as a denominator. Now, let's treat this like a true inverse. The mathematical skeleton of this inverse comes from substituting  $4\pi\alpha_s J_\tau J^\tau / (k_\tau k^\tau + i\mathcal{E})^2 \rightarrow x$  and  $k_\tau k^\tau + i\mathcal{E} \rightarrow -k$  and representing this inverse as  $I = (x - k)^{-1}$ . Then, taking the series expansion, we have:

$$I = (x - k)^{-1} = -\frac{1}{k} \left( 1 + \frac{x}{k} + \left(\frac{x}{k}\right)^2 + \left(\frac{x}{k}\right)^3 + \left(\frac{x}{k}\right)^4 + \dots \right) = -\frac{1}{k} \sum_{n=0}^{\infty} \left(\frac{x}{k}\right)^n. \quad (15.2)$$

So reversing this substitution and again reordering in the inverse now tells us that the inverse written as a series expansion is:

$$\begin{aligned} I &= \left( k_\tau k^\tau + i\mathcal{E} + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^2} \right)^{-1} \\ &= \frac{1}{k_\tau k^\tau + i\mathcal{E}} \left( 1 - \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^3} + \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^3} \right)^2 + \dots \right) = \frac{1}{k_\tau k^\tau + i\mathcal{E}} \sum_{n=0}^{\infty} (-1)^n \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^3} \right)^n. \end{aligned} \quad (15.3)$$

So let us now return to (14.7) for the Fourier transform  $D_1(x - y)$  and use the above inverse instead, thus replacing (14.7) with:

$$D_1(x - y) = \int \frac{d^4 k}{(2\pi)^4} \frac{e^{ik_\sigma(x-y)^\sigma}}{k_\tau k^\tau + i\mathcal{E}} \sum_{n=0}^{\infty} (-1)^n \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^3} \right)^n. \quad (15.4)$$

We then remove the leading coefficient  $\frac{1}{2}$  from (14.8) because we are now reintroducing  $J^\alpha = \lambda^i J^{i\alpha}$  with generator matrices normalized to  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$  and restore the trace that was removed at (14.5), and so write:

$$W(J)_1 = -\text{Tr} \iint d^4 x d^4 y 4\pi\alpha_s J_\sigma(x) D_1(x - y) J^\sigma(y). \quad (15.5)$$

Then using (15.4) in (15.5) we obtain:

$$W(J)_1 = -\text{Tr} \iint d^4 x d^4 y 4\pi\alpha_s J_\sigma(x) \int \frac{d^4 k}{(2\pi)^4} \frac{e^{ik_\sigma(x-y)^\sigma}}{k_\tau k^\tau + i\mathcal{E}} \sum_{n=0}^{\infty} (-1)^n \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\mathcal{E})^3} \right)^n J^\sigma(y). \quad (15.6)$$

This now replaces what was earlier (14.9).

Now we can pinpoint exactly how the abelian approximation of the last section lost certain information due to the fact that we treated the current densities as commuting  $[J^\sigma, J^\tau] = 0$ . This is because in going from (14.9) to (14.10) we moved  $J^\sigma(y)$  from the far right, over to the left past the inverse without concern for commutation. But now, we are treating  $[J^\sigma, J^\tau] = J^{i\sigma} J^{j\tau} [\lambda^i, \lambda^j] = if^{ijk} \lambda^k J^{i\sigma} J^{j\tau} \neq 0$  as non-commuting, and in (15.6) we see that to move  $J^\sigma(y)$  over to the left past the series sum  $\Sigma$ , we must effectively commute  $J^\sigma(y)$  past  $(J_\tau J^\tau)^n$  for any  $n$  right up to infinity, that is, we must move  $(J_\tau J^\tau)^n J^\sigma(y) \rightarrow J^\sigma(y) (J_\tau J^\tau)^n$ . Accordingly, we see that what was neglected in the abelian calculation of the previous section was the non-zero commutator  $[(J_\tau J^\tau), J^\sigma] \neq 0$ .

One might think to attack the required commutation in (15.6) by actually trying to calculate  $[(J_\tau J^\tau), J^\sigma] \neq 0$ , and then generalize to  $[(J_\tau J^\tau)^n, J^\sigma]$  for larger  $n$ . But that leads to some very unwieldy expressions, and there is a much better way. Instead, we make use of the fact that for SU(N) generally, each  $J^\alpha = \lambda^i J^{i\alpha}$ , where  $\lambda^i$  are the group generators. Of course, for SU(3)<sub>C</sub> we use the eight Gell-Mann matrices, but there is no reason for the present discussion to limit ourselves to one particular gauge group, and we can be perfectly general. So in general, the number of group generators for SU(N) is  $N^2 - 1$  and the  $i = 1 \dots N^2 - 1$  generators sit in an adjoint representation of the group.

Now, although  $J^\alpha = \lambda^i J^{i\alpha}$  is a perfectly good way to expand  $J^\alpha$ , let us be even more pedantic about this, and use the bra-ket notation to make one of  $\lambda^i$  and  $J^{i\alpha}$  a row/bra object, and the other a ket/column object. It does not matter which is which because the result is identical in either case, that is:

$$J^\alpha = \lambda^i J^{i\alpha} = \langle \lambda^i | J^{i\alpha} \rangle = \langle J^{i\alpha} | \lambda^i \rangle. \quad (15.7)$$

Just to illustrate explicitly, suppose the group is SU(2). Then (15.7) would be expanded to:

$$J^\alpha = \sigma^i J^{i\alpha} = \langle \sigma^i | J^{i\alpha} \rangle = \begin{pmatrix} \sigma^1 & \sigma^2 & \sigma^3 \end{pmatrix} \begin{pmatrix} J^{1\alpha} \\ J^{2\alpha} \\ J^{3\alpha} \end{pmatrix} = \langle J^{i\alpha} | \sigma^i \rangle = \begin{pmatrix} J^{1\alpha} & J^{2\alpha} & J^{3\alpha} \end{pmatrix} \begin{pmatrix} \sigma^1 \\ \sigma^2 \\ \sigma^3 \end{pmatrix}, \quad (15.8)$$

and we see why it does not matter which is the row/bra and which is the column/ket.

So, going back to (15.6), we set and  $J_\sigma(x) = \langle \lambda^i | J^i_\sigma(x) \rangle$  and  $J^\sigma(y) = \langle J^{j\sigma}(y) | \lambda^j \rangle$  with opposite alternatives (15.7) and so rewrite this as:

$$W(J)_1 = -\text{Tr} \iint d^4x d^4y 4\pi\alpha_s \langle \lambda^i | J^i_\sigma(x) \rangle \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik_\sigma(x-y)^\sigma}}{k_\tau k^\tau + i\epsilon} \sum_{n=0}^{\infty} (-1)^n \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^3} \right)^n \langle J^{j\sigma}(y) | \lambda^j \rangle. \quad (15.9)$$

In the above, we are now free to move the  $\langle J^{j\sigma}(y) |$  bra over to the left past the  $(J_\tau J^\tau)^n$ , so long as we leave the  $|\lambda^j\rangle$  ket right where it is way over on the right. This is because it is the  $\lambda^j$  which hold the commutation position, not the  $J^{j\sigma}$ . The only restriction on moving  $\langle J^{j\sigma}(y) |$  to the left is that we cannot move it to the left of  $|J^i_\sigma(x)\rangle$  because now the ket and the bra will “butt heads.” But most importantly, because  $\langle J^{j\sigma}(y) |$  is a function of configuration space while  $J_\tau J^\tau$  inside the series is a function of momentum  $k$ , we can move  $\langle J^{j\sigma}(y) |$  far enough left to get it outside the integral over  $d^4k$ . Doing this move, and also moving the  $\langle \lambda^i |$  (constant) bra all the way over to left outside of the  $d^4x d^4y$  integral, (15.9) now becomes:

$$W(J)_1 = -\text{Tr} \langle \lambda^i | \iint d^4x d^4y 4\pi\alpha_s |J^i_\sigma(x)\rangle \langle J^{j\sigma}(y) | \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik_\sigma(x-y)^\sigma}}{k_\tau k^\tau + i\epsilon} \sum_{n=0}^{\infty} (-1)^n \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^3} \right)^n | \lambda^j \rangle. \quad (15.10)$$

Contrasting with (14.9) and (14.10), the difference wrought by non-abelian gauge theory now rests in the bras and the kets appearing above.

Now, let us focus on reducing:

$$\begin{aligned} & \iint d^4x d^4y 4\pi\alpha_s |J^i_\sigma(x)\rangle \langle J^{j\sigma}(y) | = \iint dx^0 dy^0 \iint d^3x d^3y 4\pi\alpha_s |J^i_\sigma(\mathbf{x})\rangle \langle J^{j\sigma}(\mathbf{y}) | \\ & = \iint dx^0 dy^0 \int d^3x g_s |J^i_\sigma(\mathbf{x})\rangle \int d^3y g_s \langle J^{j\sigma}(\mathbf{y}) | \end{aligned} \quad (15.11)$$

As we did between (14.10) and (14.10), we can move into the rest frame where  $g_s J_0 = g_s \psi^\dagger \psi \equiv \rho$  is the probability density. But now, we have  $g_s J^i_0 \equiv \rho^i$  defining a total of  $N^2 - 1$  such probability densities, with the result that  $\int d^3x g_s |J^i_\sigma(\mathbf{x})\rangle \int d^3y g_s \langle J^{j\sigma}(\mathbf{y}) | = |1^i\rangle \langle 1^j|$ . For the SU(2) example, to be explicit:

$$|1^i\rangle \langle 1^j| = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1 \quad 1 \quad 1) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \equiv 1^{ij}. \quad (15.12)$$

So  $\int d^3x g_s |J^i_\sigma(\mathbf{x})\rangle \int d^3y g_s \langle J^{j\sigma}(\mathbf{y}) |$  is an  $(N^2 - 1) \times (N^2 - 1)$  matrix of ones  $1^{ij}$ , as opposed to a diagonal unit matrix  $\delta^{ij}$ . Thus, (15.11) reduces to:

$$\iint d^4x d^4y 4\pi\alpha_s |J^i{}_\sigma(x)\rangle \langle J^{j\sigma}(y)| = \iint dx^0 dy^0 1^{ij}, \quad (15.13)$$

and in view also of  $\langle \lambda^i | 1^{ij} = (N^2 - 1) \langle \lambda^j |$  (use (15.12) to see this explicitly for SU(2)), (15.10) reduces further to:

$$\begin{aligned} W(J)_1 &= -2\text{Tr} \iint dx^0 dy^0 \langle \lambda^i | 1^{ij} \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik_\sigma(x-y)^\sigma}}{k_\tau k^\tau + i\epsilon} \sum_{n=0}^{\infty} (-1)^n \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^3} \right)^n | \lambda^j \rangle \\ &= -2(N^2 - 1) \text{Tr} \iint dx^0 dy^0 \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik_\sigma(x-y)^\sigma}}{k_\tau k^\tau + i\epsilon} \langle \lambda^j | \sum_{n=0}^{\infty} (-1)^n \left( \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^3} \right)^n | \lambda^j \rangle \end{aligned} \quad (15.14)$$

Above, we have also added a factor of 2 as we did at (4.11) to account for both of the interactions  $J_x J_y$  and  $J_y J_x$ .

Now that the series  $\Sigma$  in (15.14) has served its function by showing us exactly what commutations need to be carefully considered, let us revert to remove the series via (15.3) and rewrite the above it renaming the summed indexes  $i \rightarrow j$  as:

$$W(J)_1 = -2(N^2 - 1) \text{Tr} \iint dx^0 dy^0 \int \frac{d^4k}{(2\pi)^4} e^{ik_\sigma(x-y)^\sigma} \langle \lambda^i | \left( k_\tau k^\tau + i\epsilon + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^2} \right)^{-1} | \lambda^i \rangle. \quad (15.15)$$

Now, as in (14.10), we separate  $dk^0$  from  $d^3k$  in the Fourier terms to write:

$$W(J)_1 = -2(N^2 - 1) \text{Tr} \iint dx^0 dy^0 \int \frac{dk^0}{2\pi} e^{ik_0(x-y)^0} \int \frac{d^3k}{(2\pi)^3} \langle \lambda^i | e^{ik \cdot (x-y)} \left( k_\tau k^\tau + i\epsilon + \frac{4\pi\alpha_s J_\tau J^\tau}{(k_\tau k^\tau + i\epsilon)^2} \right)^{-1} | \lambda^i \rangle. \quad (15.16)$$

We then apply (14.12) with  $k_0 = 0$  so  $k_\tau k^\tau = -\mathbf{k}^2$  and we can remove  $+i\epsilon$  exactly as before, so (15.16) becomes:

$$W(J)_1 = 2(N^2 - 1) \text{Tr} \int dx^0 \int \frac{d^3k}{(2\pi)^3} \langle \lambda^i | e^{ik \cdot (x-y)} \left( \mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4} \right)^{-1} | \lambda^i \rangle. \quad (15.17)$$

Finally, as in (14.14), we set  $\int dx^0 = T$  and  $W = -ET$  to obtain:

$$E_1 = -2(N^2 - 1) \text{Tr} \int \frac{d^3k}{(2\pi)^3} \langle \lambda^i | e^{ik \cdot (x-y)} \left( \mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4} \right)^{-1} | \lambda^i \rangle. \quad (15.18)$$

This is the non-abelian counterpart to (14.14). It is identical in form in all respects. The *substantive* differences are as follows: A) There is an overall trace (Tr) as expected for a non-abelian gauge theory. B) There is an overall factor of 2 to account for the generator normalization  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$ . C) There is a further overall factor of  $N^2 - 1$  which accounts for the dimension of the adjoint representation of SU(N). D) The integrand inside the Fourier  $d^3k$  is exactly the same as before, with the exception that it is bracketed inside of a  $\langle \lambda^i |$  on the left and a  $| \lambda^i \rangle$  on the right. We also note that (15.18) employs the customary inverse notation  $M^{-1}$  rather than the ordinary  $1/M$  employed in (14.14), but this is only a difference of form. If we wish, we can as a matter of mathematical notation represent (15.18) exactly as we did (14.14), by writing this as:

$$E_1 = 2(N^2 - 1) \text{Tr} \langle \lambda^i | \left( - \int \frac{d^3k}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2 - \frac{4\pi\alpha_s J_\tau J^\tau}{\mathbf{k}^4}} \right) | \lambda^i \rangle = 2(N^2 - 1) \text{Tr} \langle \lambda^i | E_{1 \text{ abelian}} | \lambda^i \rangle. \quad (15.19)$$

It is now to be seen that the term inside the rounded brackets above is *exactly the same* as the integral for  $E_1$  in the abelian simplification (14.14). We simply need to remember at suitable points in the development that it is really an inverse not a denominator. It is the sandwiching of this inverse between  $\langle \lambda^i |$  and  $| \lambda^i \rangle$  which now contains the residue of the commutativity issues that we first started to tackle in (15.6) and (15.7). This is to say, in the non-abelian gauge theory, the inverse must be and must remain sandwiched between  $\langle \lambda^i |$  and  $| \lambda^i \rangle$  in order to take proper account of  $[J^\sigma, J^\tau] = J^{i\sigma} J^{j\tau} [\lambda^i, \lambda^j] = if^{ijk} \lambda^k J^{i\sigma} J^{j\tau} \neq 0$  in non-abelian gauge theory. These  $\langle \lambda^i |$  and  $| \lambda^i \rangle$  carry and preserve the fact that in (15.6),  $J_\sigma(x)$  was to the left and  $J^\sigma(y)$  was to the right of the inverse. Everything else has now been distilled out from (15.19).

Now let us fast forward a few steps from (14.14) to (14.20) during which we transformed the coordinates to eliminate all but the integral over  $dk_r$  and in a final step set  $J_\tau J^\tau = \rho_0^2$  where  $\rho_0$  is the proper probability density. Nothing occurred during that stretch of equations as regards the inverse. So we substitute the result (14.20) for  $E_{1 \text{ abelian}}$  in the above and now explicitly show the inverse  $(\mathbf{k}^6 - 4\pi\alpha_s \rho_0^2)^{-1}$ , to write:

$$E_1 = - \frac{2(N^2 - 1)}{(2\pi)^2} \frac{1}{i} \frac{1}{r} \text{Tr} \langle \lambda^i | \int_{-\infty}^{\infty} dk_r |\mathbf{k}| \mathbf{k}^4 \exp i|\mathbf{k}|r (\mathbf{k}^6 - 4\pi\alpha_s \rho_0^2)^{-1} | \lambda^i \rangle. \quad (15.20)$$

This is the non-abelian counterpart to (14.20).

Now we need to again integrate over  $dk_r$ , and will look to contour integration via  $|\mathbf{k}| \rightarrow z$  to help us do so. But there is an important difference between this integral and (14.20), because in an SU(N) non-abelian gauge theory  $\rho_0^2$  has an NxN matrix representation  $J_\tau J^\tau = \lambda^i \lambda^j J_\tau^i J^\tau^j = \lambda^i \lambda^j \rho_0^i \rho_0^j = \rho_0^2$ , which we emphasize using inverse notation. So as we did in (14.21), we extend this integral to complex numbers and write the contour integral which is identical to (14.20) in form, but for the denominator now being an inverse:

$$\oint_C f(z) dz = \int_C dz z^5 \exp izr (z^6 - 4\pi\alpha_s \rho_0^2)^{-1}. \quad (15.21)$$

Using the same partial fraction separation employed from (14.21) to (14.27), we then arrive at:

$$\begin{aligned} \oint_C f(z) dz &= \oint_C dz z^5 \exp izr (z^6 - 4\pi\alpha_s \rho_0^2)^{-1} \\ &= \frac{1}{3} \oint_{C_1} dz z \exp izr \left( z + (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)^{-1} \left( z - (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)^{-1} \\ &+ \frac{1}{3} \oint_{C_2} dz z \exp izr \left( z + \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)^{-1} \left( z - \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)^{-1} \\ &+ \frac{1}{3} \oint_{C_3} dz z \exp izr \left( z + \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)^{-1} \left( z - \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right)^{-1}. \end{aligned} \quad (15.22)$$

This is the counterpart to (14.27), but now, because  $\rho_0 = \lambda^i \rho_0^i$  with  $i=1\dots N^2-1$  for SU(N) which stems from  $J^\sigma = \lambda^i J^{i\sigma}$ , the steps we must take to further develop (15.22) into an analog of (14.28) will bring us into qualitatively new, and very deep territory in a number of ways.

First of all, because  $\rho_0 = \lambda^i \rho_0^i$  is an NxN Hermitian matrix and the next step is to use the three roots  $z_1 = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$ ,  $z_2 = \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$  and  $z_3 = \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$  to arrive at an analog of (14.28), we must inquire about both the nature of the roots  $z_1, z_2, z_3$  as well as about the nature of  $\rho_0^{\frac{1}{3}}$ . In the contour integration of the last section we took  $z$  to be an ordinary complex number  $z = A + iB$ . But if we need to take, e.g.,  $z_1 = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$ , and if  $\rho_0 = \left( \rho_0^{\frac{1}{3}} \right)^3$  is an NxN Hermitian matrix, then we must regard  $z$  as an NxN Hermitian matrix for SU(N), because this is the only way to make these root expressions zero which is required to perform to contour integral. Otherwise, these would simply be eigenvalue equations for  $z$  in the form of, e.g.,  $\left| z_1 - (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} \right| = 0$  using  $|A| \equiv \det A$ . Second, if  $\rho_0 = \lambda^i \rho_0^i = \left( \rho_0^{\frac{1}{3}} \right)^3$ , then we must find out more about the cubed root object  $\rho_0^{\frac{1}{3}}$ , because this is not an ordinary number but is rather is the cubed root of an NxN Hermitian matrix  $\rho_0 = \lambda^i \rho_0^i$ . So what we really must do here is

harmonize the structure of  $z$  with that of  $\rho_0^{\frac{1}{3}}$  and with that of  $\rho_0$  in a way that all makes sense in a quantum field theory. Two pieces of historical context will help set the stage for this.

First, in 1843 imaginary  $i = \sqrt{-1}$  and complex  $A + Bi$  numbers were still fairly new when William Rowan Hamilton sought to generalize  $i^2 = -1$  to three dimensions by creating two more numbers  $j, k$  different from  $i$  which also are specified by  $j^2 = k^2 = -1$ . Among other things, this helped to describe rotations in three space dimensions. In a seminal flash, he conceived the answer to his quest, and used his penknife to carve in the side of the Brougham Bridge the quaternions  $i^2 = j^2 = k^2 = ijk = -1$ . In so doing, he extended complex numbers into three-dimensional spaces. These quaternions are still very much in use throughout physics, but in modern parlance they take the form of the Pauli spin matrices  $[\sigma_i, \sigma_j] = i\epsilon_{ijk}\sigma_k$  normalized to  $Tr(\sigma_i)^2 = \frac{1}{2}$ , which have the quaternion relationship  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = -i\sigma_1\sigma_2\sigma_3 = \delta_{2 \times 2}$ . Then, in 1954 Chen Ning Yang and Robert Mills took the next step and generalized all of this to even higher dimensionality  $N^2 - 1$  via the generators  $[\lambda_i, \lambda_j] = if_{ijk}\lambda_k$  of whatever compact simple traceless gauge group  $SU(N)$  one may wish to consider. So, for example, the color group  $SU(3)$  is an eight-dimensional quaternion-like extension of Hamilton's original complex analysis. So it is natural, when trying to solve (15.20) by extending into complex numbers via  $|\mathbf{k}| \rightarrow z$  and  $dk_r \rightarrow dz$ , to take  $z$  to be not the simple complex number  $z = A + iB$ , but a Hermitian matrix of  $N \times N$  dimensionality. And, after all, the matrices of  $SU(N)$  are simply matrices containing several real and complex numbers, rather than just a single  $z = A + iB$ . So if  $dk_r \rightarrow dz$  is now an  $N \times N$  Hermitian matrix, we are simply doing  $N \times N$  contour integrals all at once and packaging them up in a single matrix. The off-diagonal  $N \times N - N$  of these integrals are over complex measures  $z$ , and the on-diagonal  $N$  of these integrals are over real measures. This is important to keep in mind because the  $\rho_0 = \lambda^i \rho_0^i$  which appear in (15.22) and which we need to set to roots such as  $z_1 = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}}$ , make use of these very same generators which are the progeny of Yang-Mill's extension of Pauli's and Hamilton's extension of complex numbers into higher dimensionality.

Second, in 1928 Paul Dirac was attempting to obtain the non-trivial square root of the relativistic energy relationship  $p_\sigma p^\sigma = m^2$  but wanted to find a relationship which – unlike the Klein-Gordon equation a.k.a. relativistic Schrödinger equation – was linear in the spacetime gradient  $\partial^\sigma$ . In the process, he found that although the equations of special and general relativity were based on a Minkowski metric tensor  $\eta^{\mu\nu}$  which generalized to the spacetime metric  $g_{\mu\nu}$ , there is an underlying fermion structure to spacetime that lays hidden in a set of  $\gamma^\mu$  defined via a Clifford Algebra such that  $\eta^{\mu\nu} = \frac{1}{2}\{\gamma^\mu, \gamma^\nu\}$ . In this way, Dirac's equation  $\gamma^\sigma p_\sigma |u\rangle = m|u\rangle$  is just the square root of  $p_\sigma p^\sigma = m^2$ , but with a much richer substructure than is revealed by the trivial root equation  $\sqrt{p_\sigma p^\sigma} = \pm m$ . This is also important to keep in mind here, because channeling Dirac, the presence of  $\rho_0^{\frac{1}{3}}$  in the root equations residing in (15.22) is telling us that there is a cubed root substructure resting “beneath the hood” of Yang-Mills theory.

This is a substructure that we shall now find and develop. And as we shall see, this substructure will bring us to the very heart of quantum field theory.

Both of these historical threads converge here, because now that we are taking  $z$  to be a Hermitian matrix so that we can set the roots to  $z_1 = (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} = (g_s \rho_0)^{\frac{1}{3}}$ ,  $z_2 = \left(\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} = \left(\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)(g_s \rho_0)^{\frac{1}{3}}$  and  $z_3 = \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} = \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)(g_s \rho_0)^{\frac{1}{3}}$  and complete the contour integration (15.22). Because of this we must find out about these  $\rho_0^{\frac{1}{3}}$  objects to which these three  $z_i$  are proportional up to numeric factors. So now let us closely study  $\rho_0$  and  $\rho_0^{\frac{1}{3}}$ .

Insofar as  $\rho_0^{\frac{1}{3}}$  is concerned, this is not just any ordinary object. The  $\rho_0 = \lambda^i \rho_0^i$  of which this is the cubed root is itself an inherently three-space-dimensional object. Specifically, in the rest frame,  $\rho_0 = 4\pi\alpha_s J^0 = 4\pi\alpha_s \psi^\dagger \psi$  is a *proper probability density of the source current  $J^\sigma$  in the three space dimensions of the natural world*. In natural units  $\hbar = c = 1$ ,  $\rho_0$  has a mass dimensionality of +3 and a length dimensionality of -3, or 1/volume. So this means that  $\rho_0^{\frac{1}{3}}$  has a mass dimensionality of +1 and a length dimensionality of -1, i.e., that its dimension is of 1/length. If we use  $\mu \equiv \rho_0^{\frac{1}{3}}$  to denote an object with mass dimension of +1, which when scaled with the running charge to  $g_s^{\frac{1}{3}} \mu \equiv (g_s \rho_0)^{\frac{1}{3}}$  may be interpreted as a density along a single length dimension, then we may write  $g_s \rho_0 = g_s \mu^3$  and thus  $\rho_0 = \mu^3$ . But these three dimensions for which  $\rho_0$  is a 1/length<sup>3</sup> measure are not just some abstract space: they are the physical space of physical experience which we often refer to the Cartesian coordinates x, y and z when we seek to talk about measurements in that space. So rather than just write  $\rho_0 = \mu^3$ , let us define three distinct  $\mu_{0x}, \mu_{0y}, \mu_{0z}$  which define proper linear densities along each of these three measurement axes, and thus write  $\rho_0 \equiv \mu_{0x} \mu_{0y} \mu_{0z}$  as the *definition* of  $\mu_{0x}, \mu_{0y}, \mu_{0z}$ . However, so as to not introduce any bias toward a particular x, y, z ordering in view of the structure of antisymmetric field theory, we should really define these  $\mu_0$  using wedge products as  $\rho_0 \equiv \frac{1}{3!} \mu_{0x} \wedge \mu_{0y} \wedge \mu_{0z}$ . Finally, because  $\rho_0 = \lambda^i \rho_0^i$  is itself an NxN Hermitian matrix, let us similarly define each of the one-dimensional  $\mu_{0x,y,z}$  in like fashion, and then ascertain the detailed  $\rho_0(\mu_{0x}, \mu_{0y}, \mu_{0z})$  relationship. That is we now define  $\mu_{0x} \equiv \lambda^i \mu_{0x}^i$ ,  $\mu_{0y} \equiv \lambda^j \mu_{0y}^j$  and  $\mu_{0z} \equiv \lambda^k \mu_{0z}^k$ .

Tending for a moment to notation, we now define the notations  $\vec{\lambda} \equiv \lambda^k$  and  $\vec{\mu}_0 \equiv \mu_0^k = (\mu_{0x}^k, \mu_{0y}^k, \mu_{0z}^k)$  to represent each of  $\lambda^k$  and  $\mu_0^k$  as a vector in the adjoint representation of the internal symmetry (not space) indexes  $k = 1 \dots N^2 - 1$ . This is to say, vectors  $V$  in *experiential three space* will be represented in boldface type  $V$  and vectors  $I$  in the *internal symmetry space* will be represented by  $\vec{I}$  with an arrow above the object. We also use “ $\circ$ ” to represent a dot product in internal symmetry space, versus the “ $\cdot$ ” reserved for experiential

space. Further, in the same way that we use the Feynman slash  $V = \gamma^\mu \bar{V}_\mu$  to represent the “scalar product” of a vector  $V_\mu$  with the Dirac gamma matrices  $\gamma^\mu$ , we will also define  $\not{\mu}_0 \equiv \bar{\lambda} \circ \bar{\mu}_0$  with a horizontal slash to represent the internal symmetry space scalar product of  $\bar{\mu}_0$  with the group generators  $\bar{\lambda}$ . Note that  $\bar{\mu}_0$  is itself a vector in *both* internal symmetry and experiential space. Thus, we now denote the x, y, z components of  $\not{\mu}_0$  by  $\not{\mu}_{0x} \equiv \lambda^i \mu_{0x}^i$ ,  $\not{\mu}_{0y} \equiv \lambda^j \mu_{0y}^j$  and  $\not{\mu}_{0z} \equiv \lambda^k \mu_{0z}^k$ . In these notations, this means that we are really defining  $\not{\mu}_0$  by  $\not{\rho}_0 \equiv \frac{1}{3!} \not{\mu}_{0x} \wedge \not{\mu}_{0y} \wedge \not{\mu}_{0z}$ .

All of this now means, using the group structure relationship  $[\lambda^i, \lambda^j] = if^{ijm} \lambda^m$ , that:

$$\begin{aligned}
 \not{\rho}_0 &= \frac{1}{3!} \not{\mu}_{0x} \wedge \not{\mu}_{0y} \wedge \not{\mu}_{0z} = \frac{1}{2!} \left( \not{\mu}_{0x} [\not{\mu}_{0y}, \not{\mu}_{0z}] + \not{\mu}_{0y} [\not{\mu}_{0z}, \not{\mu}_{0x}] + \not{\mu}_{0z} [\not{\mu}_{0x}, \not{\mu}_{0y}] \right) \\
 &= \rho_0^l \lambda^l = \frac{1}{2!} \left( \lambda^i \mu_{0x}^i [\lambda^j \mu_{0y}^j, \lambda^k \mu_{0z}^k] + \lambda^j \mu_{0y}^j [\lambda^k \mu_{0z}^k, \lambda^i \mu_{0x}^i] + \lambda^k \mu_{0z}^k [\lambda^i \mu_{0x}^i, \lambda^j \mu_{0y}^j] \right) \quad (15.23) \\
 &= \frac{1}{2!} \left( \lambda^i [\lambda^j, \lambda^k] + \lambda^j [\lambda^k, \lambda^i] + \lambda^k [\lambda^i, \lambda^j] \right) \mu_{0x}^i \mu_{0y}^j \mu_{0z}^k \\
 &= \frac{1}{2!} \mu_{0x}^i \mu_{0y}^j \mu_{0z}^k \left( \lambda^i if^{jkl} + \lambda^j if^{kil} + \lambda^k if^{ijl} \right) \lambda^l
 \end{aligned}$$

From this we can factor out the  $\lambda^l$  generator from very right, and simplify to:

$$\begin{aligned}
 \rho_0^m &= \frac{1}{2!} \mu_{0x}^i \mu_{0y}^j \mu_{0z}^k \left( \lambda^i if^{jkm} + \lambda^j if^{kim} + \lambda^k if^{ijm} \right) \\
 &= \frac{1}{2!} if^{mij} \left( \mu_{0x}^i \mu_{0y}^j \mu_{0z}^k + \mu_{0y}^i \mu_{0z}^j \mu_{0x}^k + \mu_{0z}^i \mu_{0x}^j \mu_{0y}^k \right) \lambda^k \quad (15.24) \\
 &= \frac{1}{2!} if^{mij} \left( \mu_{0x}^i \mu_{0y}^j \mu_{0z}^k + \mu_{0y}^i \mu_{0z}^j \mu_{0x}^k + \mu_{0z}^i \mu_{0x}^j \mu_{0y}^k \right)
 \end{aligned}$$

For a particular gauge group SU(N) with free index  $m=1\dots N^2-1$ , this contains  $N^2-1$  simultaneous equations.

If we wish to gain a better geometric understanding of this relationship in the three-space of spacetime, we may choose the simplest internal symmetry group SU(2) just for illustration. Here,  $\lambda^k \rightarrow \sigma^k$  become the Pauli spin matrices (normalized with a  $\frac{1}{2}$  factor) and  $f^{mij} \rightarrow \epsilon^{mij}$  becomes the Levi-Civita tensor. With a free index  $m$ , (15.24) for SU(2) now contains three simultaneous equations. To garner the pattern, it suffices to explore one of these three equations, say, for  $\rho_0^1$ . From the middle line of (15.25), using boldface type to represent vectors  $V$  in the three-space of ordinary experience, and then using ordinary cross and dot products in these three space dimensions, we obtain:

$$\begin{aligned}
\rho_0^1 &= \frac{1}{2!} i \varepsilon^{1ij} \left( \mu_{0x}^i \mu_{0y}^j \mu_{0z}^k + \mu_{0y}^i \mu_{0z}^j \mu_{0x}^k + \mu_{0z}^i \mu_{0x}^j \mu_{0y}^k \right) \sigma^k \\
&= \frac{1}{2!} i \left( \left( \mu_{0y}^2 \mu_{0z}^3 - \mu_{0z}^2 \mu_{0y}^3 \right) \sigma^k \mu_{0x}^k + \left( \mu_{0z}^2 \mu_{0x}^3 - \mu_{0x}^2 \mu_{0z}^3 \right) \sigma^k \mu_{0y}^k + \left( \mu_{0x}^2 \mu_{0y}^3 - \mu_{0y}^2 \mu_{0x}^3 \right) \sigma^k \mu_{0z}^k \right) \\
&= \frac{1}{2!} i \left( \left( \mu_0^2 \times \mu_0^3 \right)^x \sigma^k \mu_{0x}^k + \left( \mu_0^2 \times \mu_0^3 \right)^y \sigma^k \mu_{0y}^k + \left( \mu_0^2 \times \mu_0^3 \right)^z \sigma^k \mu_{0z}^k \right) \\
&= \frac{1}{2!} i \left( \mu_0^2 \times \mu_0^3 \right) \cdot \left( \sigma^k \mu_0^k \right) = \frac{1}{4} i \varepsilon^{1ij} \left( \mu_0^i \times \mu_0^j \right) \cdot \left( \sigma^k \mu_0^k \right) = \frac{1}{4} i \varepsilon^{1ij} \left( \mu_0^i \times \mu_0^j \right) \cdot \boldsymbol{\mu}_0
\end{aligned} \tag{15.25}$$

With this notation we can write out  $\boldsymbol{\mu}_0 = \sigma^k \mu_0^k$  in the final expression in (15.25) as:

$$\boldsymbol{\mu}_0 = \vec{\sigma} \circ \vec{\mu}_0 = \sigma^k \mu_0^k = \begin{pmatrix} \mu_0^3 & \mu_0^1 - i \mu_0^2 \\ \mu_0^1 + i \mu_0^2 & \mu_0^3 \end{pmatrix}. \tag{15.26}$$

So in sum, (15.25) may now be written as  $\rho_0^1 = \frac{1}{4} i \varepsilon^{1ij} \left( \mu_0^i \times \mu_0^j \right) \cdot \boldsymbol{\mu}_0$ . Given that this is just for the  $i=1$  component, we can generalize (15.25) to the  $i=2,3$  components by writing  $\rho_0^m = \frac{1}{4} i \varepsilon^{mij} \left( \mu_0^i \times \mu_0^j \right) \cdot \boldsymbol{\mu}_0$ . Then, generalizing this back to any SU(N) by  $\sigma^k \rightarrow \lambda^k$  and  $\varepsilon^{mij} \rightarrow f^{mij}$ , we find that for any SU(N), with  $\boldsymbol{\mu}_0 = \vec{\lambda} \circ \vec{\mu}_0$ , the relationship (15.24) is:

$$\rho_0^m = \frac{1}{4} i f^{mij} \left( \mu_0^i \times \mu_0^j \right) \cdot \boldsymbol{\mu}_0 = \frac{1}{4} i f^{mij} \left( \mu_0^i \times \mu_0^j \right) \cdot \vec{\lambda} \circ \vec{\mu}_0. \tag{15.27}$$

As a final notational consolidation, recall the discussion back between (1.9) and (1.10) regarding how the wedge symbol  $\wedge$  is used to represent a cross product in internal symmetry space. Specifically, we observed that while  $\varepsilon^{ijk} A^i B^j = (\mathbf{A} \times \mathbf{B})^k$  is used in experiential space, the analogous  $f^{ijk} A^i B^j = (\vec{A} \wedge \vec{B})^k$  is used in internal symmetry space. So we wish to use this to compact  $f^{mij} \left( \mu_0^i \times \mu_0^j \right)$  to include the internal symmetry space cross product symbol  $\wedge$ . But there is already a spatial cross product  $\times$  in this expression. So the final notation we introduce will be  $\otimes \equiv \times \cup \wedge$  to denote a ‘‘super-cross product’’ for a situation such as we have at present, where we unite ( $\cup$ ) a cross product in *both* the internal symmetry and the experiential spaces. With this notation, we employ  $(\vec{\mu}_0 \otimes \vec{\mu}_0)^m \equiv f^{mij} \left( \mu_0^i \times \mu_0^j \right)$  to write (15.27) as:

$$\rho_0^m = \frac{1}{4} i \left( \vec{\mu}_0 \otimes \vec{\mu}_0 \right)^m \cdot \boldsymbol{\mu}_0. \tag{15.28}$$

Then we may suppress internal symmetry vector index  $m$  via  $\rho_0^m \rightarrow \vec{\rho}_0$  and  $\otimes^m \rightarrow \vec{\otimes}$  as we did at (1.10) to implement one final consolidation of the internal symmetry space vectors to:

$$\vec{\rho}_0 = \frac{1}{4} i \left( \vec{\mu}_0 \vec{\otimes} \vec{\mu}_0 \right) \cdot \boldsymbol{\mu}_0. \tag{15.29}$$

The above is (15.24), reduced down to the geometrically-understandable Maxwellian language of vectors and vector products in internal symmetry and experiential spaces.

Now, returning to the original probability density  $\rho_0 = \lambda^m \rho_0^m$  with which all of this started, we may take the scalar product of (15.28) with  $\lambda^m$  to finally write:

$$\rho_0 = \lambda^m \rho_0^m = \frac{1}{4} i \lambda^m (\bar{\mu}_0 \otimes \bar{\mu}_0)^m \cdot \boldsymbol{\mu}_0 = \frac{1}{4} i \bar{\lambda} \circ (\bar{\mu}_0 \bar{\otimes} \bar{\mu}_0) \cdot \boldsymbol{\mu}_0 = \frac{1}{4} i (\bar{\mu}_0 \bar{\otimes} \bar{\mu}_0) \cdot \boldsymbol{\mu}_0. \quad (15.30)$$

In the final expression, we have set  $\lambda^i \wedge^i = \bar{\lambda} \circ \bar{\lambda} = \bar{\wedge}$ , which because we are symbolically using  $\bar{\otimes} \equiv \times \cup \wedge$  to represent a unified cross product, results in  $\lambda^m \bar{\otimes}^m = \bar{\lambda} \circ \bar{\otimes} = \bar{\otimes}$ . With the running charge strength scaled back in, this becomes  $g_s \rho_0 = \frac{1}{4} i (g_s^{\frac{1}{3}} \bar{\mu}_0 \bar{\otimes} g_s^{\frac{1}{3}} \bar{\mu}_0) \cdot g_s^{\frac{1}{3}} \boldsymbol{\mu}_0$ .

So if we want to see what this three-space probability density which in turn is situated inside the current density term  $\mathcal{J}_\sigma \mathcal{J}^\sigma = \rho_0^2$  looks like in terms of the proper probability density along each of the x, y, z space dimensions, SU(2) provides a good illustration. From (15.28) and (15.30) and  $\boldsymbol{\mu}_0 = \bar{\lambda} \circ \bar{\mu}_0 \rightarrow \bar{\sigma} \circ \bar{\mu}_0$ , we have:

$$\begin{aligned} \rho_0 &= \sigma^m \rho_0^m = \begin{pmatrix} \rho_0^3 & \rho_0^1 - i \rho_0^2 \\ \rho_0^1 + i \rho_0^2 & -\rho_0^3 \end{pmatrix} = \frac{1}{4} i \sigma^m (\bar{\mu}_0 \otimes \bar{\mu}_0)^m \cdot \boldsymbol{\mu}_0 = \frac{1}{4} i \bar{\sigma} \circ (\bar{\mu}_0 \bar{\otimes} \bar{\mu}_0) \cdot \boldsymbol{\mu}_0 \\ &= \frac{1}{4} i \begin{pmatrix} (\bar{\mu}_0 \otimes \bar{\mu}_0)^3 & (\bar{\mu}_0 \otimes \bar{\mu}_0)^1 - i (\bar{\mu}_0 \otimes \bar{\mu}_0)^2 \\ ((\bar{\mu}_0 \otimes \bar{\mu}_0)^1 + i (\bar{\mu}_0 \otimes \bar{\mu}_0)^2) & -(\bar{\mu}_0 \otimes \bar{\mu}_0)^3 \end{pmatrix} \cdot \begin{pmatrix} \mu_0^3 & \mu_0^1 - i \mu_0^2 \\ \mu_0^1 + i \mu_0^2 & -\mu_0^3 \end{pmatrix}. \end{aligned} \quad (15.31)$$

The construction of the probability density for higher dimensional groups SU(N) will follow this same basic pattern.

When we used the analogy following (15.22) of Paul Dirac having discovered a Clifford Algebra substructure to spacetime in the form of  $\eta^{\mu\nu} = \frac{1}{2} \{ \gamma^\mu, \gamma^\nu \}$ , it was (15.31) above that we had in mind. Normally the consideration of the probability density ends with  $\rho_0 = \lambda^m \rho_0^m$  above and goes no further. In (15.31) which uses SU(2) for illustration, we deconstruct the probability density which is a density in a three-dimensional volume, into its component probability density vector  $\bar{\mu}$  over each of the x, y and z dimensions (and over the internal symmetry dimensions). Given that  $J_\sigma J^\sigma = \rho_0^2$  in turn, this is a level of substructure within a current density four-vector  $J^\mu \equiv \rho_0 u^\mu$  which does not appear to have previously been uncovered.

Now, we come to the  $\bar{\mu} = (\bar{\mu}_x, \bar{\mu}_y, \bar{\mu}_z) = (\mu_x^i, \mu_y^j, \mu_z^k)$  themselves. This three-space vector of  $N^2 - 1$  internal symmetry vectors has a mass dimension of +1, i.e., of 1/length. Because these represent probability one-densities, the question now arises as to the origin of these one-densities. Here, we simply introduce a proper “*probability field*”  $P_0^i(x, y, z) = \bar{P}_0(\mathbf{x})$

which for SU(N) represents  $N^2 - 1$  dimensionless probability distributions spread over all three space dimensions. Therefore  $\vec{\lambda} \circ \vec{P}_0 = \vec{P}_0$ . We very deliberately refer to this as a probability *field*, because as we shall explore in the next several sections, the abelian relationship (14.32) which we are presently seeking to extend to non-abelian gauge theory is a *quantum field* equation. So the question naturally arises: what are the “fields” in quantum *field* theory which are analogs of the gauge fields  $G^\mu$  of classical theory which are the variable of integration in the path integral  $Z = \int DG \exp i \int d^4x S(G) \equiv \exp i W(J)$ ? That is, because the classical fields  $G^\mu$  disappear during path integration, *by definition*, what then are the “fields” of *quantum* field theory as represented in  $W(J)$  following path integration?

The answer to this, is that *the probability fields  $P_0^i(x, y, z)$  are the fundamental fields underlying quantum field theory*. The coupled linear probability densities will be normalized such that  $\iiint g_s^{\frac{1}{3}}(x) \vec{\mu}_x(x) dx = \vec{1}$ ,  $\iiint g_s^{\frac{1}{3}}(y) \vec{\mu}_y(y) dy = \vec{1}$  and  $\iiint g_s^{\frac{1}{3}}(z) \vec{\mu}_z(z) dz = \vec{1}$ , which is to say that coupled the field associated with  $g_s^{\frac{1}{3}} \vec{P}_0(x, y, z)$  will definitely be located *somewhere* along the x axis, *somewhere* along the y axis and *somewhere* along the z axis. We earlier did the same thing between (15.11) and (15.12) with the per-three-space-volume probability density. If we then wish to ascertain the one-density of this probability along each of x, y, z, we simply take the partial derivatives of each. This means that the  $\vec{\mu}_0$  from which we constructed  $\vec{\rho}_0$  in (15.31) above are themselves rooted in the space gradient  $\nabla = (\partial_x, \partial_y, \partial_z)$  according to:

$$\begin{aligned} g_s^{\frac{1}{3}} \vec{\mu}_0 &= g_s^{\frac{1}{3}} (\vec{\mu}_{0x}, \vec{\mu}_{0y}, \vec{\mu}_{0z}) = g_s^{\frac{1}{3}} (\mu_x^i, \mu_y^i, \mu_z^i) = g_s^{\frac{1}{3}} \boldsymbol{\mu}^i \\ &\equiv (\partial_x, \partial_y, \partial_z) (h(x, y, z) P_0^i(x, y, z)) = \nabla (h(\mathbf{x}) \vec{P}_0(\mathbf{x})) = \nabla (h \vec{P}_0) = \nabla h \vec{P}_0 + h \nabla \vec{P}_0 \end{aligned} \quad (15.32)$$

where a dimensionless running probability field coupling  $h(\mathbf{x})$  is *defined* in terms of  $g_s$  and  $\vec{\mu}_0$  by the above first order differential relationship, and where the  $\vec{\mu}_0$  are defined by  $\vec{\rho}_0 \equiv \frac{1}{3!} \boldsymbol{\mu}_{0x} \wedge \boldsymbol{\mu}_{0y} \wedge \boldsymbol{\mu}_{0z}$  prior to (15.23) which became by  $\vec{\rho}_0 = \frac{1}{4} i (\vec{\mu}_0 \otimes \vec{\mu}_0) \cdot \boldsymbol{\mu}_0$  in (15.30). This  $h(\mathbf{x})$  not to be confused with the Higgs field  $h(\mathbf{x})$ . We simply use  $h$  here because it is the next letter after  $g$  in the Roman alphabet. Taking (15.32) in the compacted form  $g_s^{\frac{1}{3}} \vec{\mu}_0 = \nabla (h \vec{P}_0)$ , we also have  $\vec{\lambda} \circ g_s^{\frac{1}{3}} \vec{\mu}_0 = \vec{\lambda} \circ \nabla (h \vec{P}_0) = g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 = \nabla (h \vec{P}_0)$ .

In terms of this bare probability  $\vec{P}_0(\mathbf{x})$  and its coupling  $h(\mathbf{x})$ , (15.28) and (15.29) may now be written together as:

$$g_s \rho_0^m = \frac{1}{4} i (\nabla (h \vec{P}_0) \otimes \nabla (h \vec{P}_0))^m \cdot \nabla (h \vec{P}_0) = g_s \vec{\rho}_0 = \frac{1}{4} i (\nabla (h \vec{P}_0) \otimes \nabla (h \vec{P}_0)) \cdot \nabla (h \vec{P}_0). \quad (15.33)$$

Further, we may take the internal symmetry space  $\circ$  product of the above with  $\vec{\lambda} = \lambda^m$ , thus:

$$\begin{aligned}
 g_s \lambda^m \rho_0^m &= g_s \vec{\lambda} \circ \vec{\rho}_0 = g_s \mathcal{P}_0 \\
 &= \frac{1}{4} i \lambda^m \left( \nabla \left( h \vec{P}_0 \right) \otimes \nabla \left( h \vec{P}_0 \right) \right)^m \cdot \nabla \left( h \mathcal{P}_0 \right) = \frac{1}{4} i \vec{\lambda} \circ \left( \nabla \left( h \vec{P}_0 \right) \otimes \nabla \left( h \vec{P}_0 \right) \right) \cdot \nabla \left( h \mathcal{P}_0 \right). \\
 &= \frac{1}{4} i \left( \nabla \left( h \vec{P}_0 \right) \otimes \nabla \left( h \vec{P}_0 \right) \right) \cdot \nabla \left( h \mathcal{P}_0 \right)
 \end{aligned} \tag{15.34}$$

Again,  $\vec{P}_0(\mathbf{x})$  are dimensionless probability distributions. Fundamentally, *these  $\vec{P}_0$  are the fields of quantum field theory.* Each of the three  $g_s^{\frac{1}{3}} \vec{\mu}_0 = \nabla \left( h \vec{P}_0 \right)$  in (15.32) has a mass dimension of +1 by virtue of the gradient  $\nabla = (\partial_x, \partial_y, \partial_z)$ , so  $g_s \mathcal{P}_0 = \frac{1}{4} i \left( \nabla \left( h \vec{P}_0 \right) \otimes \nabla \left( h \vec{P}_0 \right) \right) \cdot \nabla \left( h \mathcal{P}_0 \right)$  which is the overall three-volume density has the required mass dimension of +3.

We may also calculate out  $\nabla \left( h \vec{P}_0 \right) = (\nabla h + h \nabla) \vec{P}_0$  as shown in the last two expressions of (15.32), to write the above as:

$$\begin{aligned}
 g_s \mathcal{P}_0 &= \frac{1}{4} i \left( \nabla \left( h \vec{P}_0 \right) \otimes \nabla \left( h \vec{P}_0 \right) \right) \cdot \nabla \left( h \mathcal{P}_0 \right) = \frac{1}{4} i \left( (\nabla h + h \nabla) \vec{P}_0 \otimes (\nabla h + h \nabla) \vec{P}_0 \right) \cdot (\nabla h + h \nabla) \mathcal{P}_0 \\
 &= \frac{1}{4} i \left[ \begin{aligned} &h^3 \left( \nabla \vec{P}_0 \otimes \nabla \vec{P}_0 \right) \cdot \nabla \mathcal{P}_0 \\ &+ h^2 \left( (\nabla h \vec{P}_0 \otimes \nabla \vec{P}_0) \cdot \nabla \mathcal{P}_0 + (\nabla \vec{P}_0 \otimes \nabla h \vec{P}_0) \cdot \nabla \mathcal{P}_0 + (\nabla \vec{P}_0 \otimes \nabla \vec{P}_0) \cdot \nabla h \mathcal{P}_0 \right) \\ &+ h \left( (\nabla h \vec{P}_0 \otimes \nabla h \vec{P}_0) \cdot \nabla \mathcal{P}_0 + (\nabla h \vec{P}_0 \otimes \nabla \vec{P}_0) \cdot \nabla h \mathcal{P}_0 + (\nabla \vec{P}_0 \otimes \nabla h \vec{P}_0) \cdot \nabla h \mathcal{P}_0 \right) \\ &+ (\nabla h \vec{P}_0 \otimes \nabla h \vec{P}_0) \cdot \nabla h \mathcal{P}_0 \end{aligned} \right] \tag{15.35}
 \end{aligned}$$

$$\nabla \left( h \vec{P}_0 \right) = (\nabla h + h \nabla) \vec{P}_0$$

which now includes in the form of a first order differential equation, the running of the usual interaction charge  $g_s(\mathbf{x})$  in relation to that of the newly-established probability coupling  $h(\mathbf{x})$ .

In the next section we shall discuss the physical interpretation of these results, but at the moment, our goal is to complete the mathematics of doing the contour integral (15.22). So we return to the  $z_1 = (g_s \rho_0)^{\frac{1}{3}}$ ,  $z_2 = \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (g_s \rho_0)^{\frac{1}{3}}$  and  $z_3 = \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (g_s \rho_0)^{\frac{1}{3}}$  poles, and now approach these in view of what we have uncovered since (15.22). We now see that each of these  $\rho_0^{\frac{1}{3}} \rightarrow \mu_0$  represents a probability density along one of three linear space dimensions x, y, z, and that  $\mathcal{P}_0 \equiv \frac{1}{3!} \mathcal{H}_{0x} \wedge \mathcal{H}_{0y} \wedge \mathcal{H}_{0z}$ . We see that each of these linear probability densities has a Yang-Mills structure  $\mathcal{H}_{0x} \equiv \lambda^i \mu_{0x}^i$ ,  $\mathcal{H}_{0y} \equiv \lambda^j \mu_{0y}^j$  and  $\mathcal{H}_{0z} \equiv \lambda^k \mu_{0z}^k$ . We see that the probability densities  $\vec{\rho}_0 = \rho_0^i$ ,  $i = 1 \dots N^2 - 1$  for the gauge group SU(N) are related to the linear probability densities by (15.29), so that the original probability density  $\mathcal{P}_0 = \lambda^m \rho_0^m$  which appears as  $\rho_0^{\frac{1}{3}}$  in

(15.22) is related to the  $\vec{\mu}_0$  according to (15.30). And we finally see in (15.32) that the  $\vec{\mu}_0$  themselves are part of the gradient relationship  $g_s^{\frac{1}{3}}\vec{\mu}_0 = \nabla(h\vec{P}_0)$  of a dimensionless probability distribution field  $\vec{P}_0(\mathbf{x}) = P_0^i(\mathbf{x})$ , so that the usual representations  $\vec{\rho}_0 = \rho_0^m$ ,  $\rho_0 = \lambda^i \rho_0^i$  for the three-dimensional (per-volume) probability density are related to the probability field via (15.33) and (15.34), which is intertwined with the relationship between the running charge strength  $g_s(\mathbf{x}) = \sqrt{4\pi\alpha(\mathbf{x})}$  and a coupling  $h(\mathbf{x})$  associated with the probability field  $\vec{P}_0(\mathbf{x})$  in the overall form of  $h(\mathbf{x})\vec{P}_0(\mathbf{x})$ .

Given all of this, let us commence the contour integration of (15.22) by associating each of these three poles  $z_1$ ,  $z_2$  and  $z_3$  in (15.22), which we take to be Hermitian matrix extensions of the  $|\mathbf{k}|$  of (15.20) into one or more complex planes, with one of the three  $\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$  which have been defined, in essence, as the cubed roots of the density  $\rho_0 = \rho_0$  according to what was originally  $\rho_0 \equiv \frac{1}{3!}\mu_{0x} \wedge \mu_{0y} \wedge \mu_{0z}$  and what we now recognize following the development of (15.23) to (15.30) should be denoted  $\rho_0 \equiv \frac{1}{3!}\boldsymbol{\mu}_0 \wedge \boldsymbol{\mu}_0 \wedge \boldsymbol{\mu}_0$ , and which, in (15.30), calculates out to be  $\rho_0 = \frac{1}{4}i(\vec{\mu}_0 \otimes \vec{\mu}_0) \cdot \boldsymbol{\mu}_0$ . First, we update (15.22) to include all that we have learned through (15.35), including the fact that the contour integration variable  $z$  is to be regarded as an NxN Hermitian matrix. So we specifically establish  $N^2 - 1$  real  $z^i$  for SU(N), and then extend into one or more imaginary planes via setting  $z \rightarrow \lambda^i z^i = \vec{\lambda} \circ \vec{z} = \bar{z}$ , and so write (15.22) with  $4\pi\alpha_s = g_s^2$  as:

$$\begin{aligned}
 \oint_C f(\bar{z}) d\bar{z} &= \oint_C d\bar{z} \bar{z}^5 \exp izr \left( \bar{z}^6 - (g_s \rho_0)^2 \right)^{-1} \\
 &= \frac{1}{3} \oint_{C_1} d\bar{z} \bar{z} \exp izr \left( \bar{z} + (g_s \rho_0)^{\frac{1}{3}} \right)^{-1} \left( \bar{z} - (g_s \rho_0)^{\frac{1}{3}} \right)^{-1} \\
 &+ \frac{1}{3} \oint_{C_2} d\bar{z} \bar{z} \exp izr \left( \bar{z} + \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (g_s \rho_0)^{\frac{1}{3}} \right)^{-1} \left( \bar{z} - \left( \frac{1}{2} - \frac{\sqrt{3}}{2} i \right) (g_s \rho_0)^{\frac{1}{3}} \right)^{-1} \\
 &+ \frac{1}{3} \oint_{C_3} d\bar{z} \bar{z} \exp izr \left( \bar{z} + \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (g_s \rho_0)^{\frac{1}{3}} \right)^{-1} \left( \bar{z} - \left( \frac{1}{2} + \frac{\sqrt{3}}{2} i \right) (g_s \rho_0)^{\frac{1}{3}} \right)^{-1}.
 \end{aligned} \tag{15.36}$$

Now we come to  $\rho_0^{\frac{1}{3}}$  which first motivated the development from (15.23) through (15.35). We now know that each of these  $\rho_0^{\frac{1}{3}}$  should be associated with one of the  $g_s^{\frac{1}{3}}\boldsymbol{\mu}_0 = \nabla(h\vec{P}_0)$  in  $\rho_0 \equiv \frac{1}{3!}\boldsymbol{\mu}_0 \wedge \boldsymbol{\mu}_0 \wedge \boldsymbol{\mu}_0$ , because these  $\boldsymbol{\mu}_0$  were *defined* to be the cubed roots of  $\rho_0$  in recognition that this density  $\rho_0$  subsists in three space dimensions so that its cube root naturally has an x, y, and z aspect, in other words, because  $dV = \frac{1}{3!}dx \wedge dy \wedge dz$  is the differential volume element within which  $\rho_0$  is a density. So now we have migrated from the originally-

appearing  $\rho_0^{\frac{1}{3}} \rightarrow \mathcal{P}_0^{\frac{1}{3}} = (\rho_{0x}^{\frac{1}{3}}, \rho_{0y}^{\frac{1}{3}}, \rho_{0z}^{\frac{1}{3}}) \rightarrow \mathcal{H}_0 = (\mathcal{H}_{0x}, \mathcal{H}_{0y}, \mathcal{H}_{0z})$ , and the question now becomes how to do the assignments in (15.36) as among the three contour integrals over  $C_1$ ,  $C_2$  and  $C_3$ . As with  $dV = \frac{1}{3!} dx \wedge dy \wedge dz$  we will want a balanced x, y and z assignment, which we will do after the contour integration when the right way to do this will become clear. First, merely to get started, we assign the  $\rho_0^{\frac{1}{3}}$  in  $C_1$ ,  $C_2$  and  $C_3$  successively to  $(\mathcal{H}_{0x}, \mathcal{H}_{0y}, \mathcal{H}_{0z})$  and then set pole values  $z_z = g_s^{\frac{1}{3}} \mathcal{H}_{0z}$ ,  $z_x = \left(\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) g_s^{\frac{1}{3}} \mathcal{H}_{0x}$  and  $z_y = \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) g_s^{\frac{1}{3}} \mathcal{H}_{0y}$ . With this, we advance (15.36) to:

$$\begin{aligned}
 \oint_C f(z) dz &= \oint_C dz z^5 \exp izr \left( z^6 - (g_s \rho_0)^2 \right)^{-1} \\
 &= \frac{1}{3} \oint_{C_1} dz z \exp izr \left( z + g_s^{\frac{1}{3}} \mathcal{H}_{0x} \right)^{-1} \left( z_x - g_s^{\frac{1}{3}} \mathcal{H}_{0x} \right)^{-1} \\
 &+ \frac{1}{3} \oint_{C_2} dz z \exp izr \left( z + \left( \frac{1}{2} - \frac{\sqrt{3}}{2}i \right) g_s^{\frac{1}{3}} \mathcal{H}_{0y} \right)^{-1} \left( z_y - \left( \frac{1}{2} - \frac{\sqrt{3}}{2}i \right) g_s^{\frac{1}{3}} \mathcal{H}_{0y} \right)^{-1} \\
 &+ \frac{1}{3} \oint_{C_3} dz z \exp izr \left( z + \left( \frac{1}{2} + \frac{\sqrt{3}}{2}i \right) g_s^{\frac{1}{3}} \mathcal{H}_{0z} \right)^{-1} \left( z_z - \left( \frac{1}{2} + \frac{\sqrt{3}}{2}i \right) g_s^{\frac{1}{3}} \mathcal{H}_{0z} \right)^{-1}. \tag{15.37}
 \end{aligned}$$

Now we are finally ready to do the integral. Having set the three poles we remove those terms and set  $z = z_x$ ,  $z = z_x$  and  $z = z_x$  in the respective residue terms in  $C_1$ ,  $C_2$  and  $C_3$ . We then tack on the usual  $2\pi i$  factor, and relate this back to (15.20) with the all-important non-abelian indication  $\rho_0 \rightarrow \mathcal{P}_0$ . Finally discard the contour arc  $\int_{\text{Arc}} f(z) dz = 0$  through the complex planes. We thus obtain:

$$\begin{aligned}
 \oint_C f(z) dz &= \oint_C dz z^5 \exp izr \left( z^6 - 4\pi\alpha_s \rho_0^2 \right)^{-1} = \int_{-\infty}^{\infty} dk_r |\mathbf{k}| \mathbf{k}^4 \exp i|\mathbf{k}|r \left( \mathbf{k}^6 - 4\pi\alpha_s \rho_0^2 \right)^{-1} \\
 &= 2\pi i \frac{1}{6} \left[ \exp i g_s^{\frac{1}{3}} \mathcal{H}_{0x} r + \exp i \left( \frac{1}{2} - \frac{\sqrt{3}}{2}i \right) g_s^{\frac{1}{3}} \mathcal{H}_{0y} r + \exp i \left( \frac{1}{2} + \frac{\sqrt{3}}{2}i \right) g_s^{\frac{1}{3}} \mathcal{H}_{0z} r \right] \tag{15.38} \\
 &= 2\pi i \frac{1}{6} \left[ \exp i g_s^{\frac{1}{3}} \mathcal{H}_{0x} r + \exp \left( i \frac{1}{2} g_s^{\frac{1}{3}} \mathcal{H}_{0y} r \right) \exp \left( \frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \mathcal{H}_{0y} r \right) + \exp \left( i \frac{1}{2} g_s^{\frac{1}{3}} \mathcal{H}_{0z} r \right) \exp \left( -\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \mathcal{H}_{0z} r \right) \right]
 \end{aligned}$$

This should be contrasted to (14.29). Now let's turn to the x, y, z balancing. The structure of this integral clarifies that an  $\frac{1}{3}(xyz + yzx + zxy)$  and *not* a  $\frac{1}{3!}x \wedge y \wedge z$  combination is the suitable way to spatially balance this equation, so we advance the above to:

$$\begin{aligned}
 & \int_{-\infty}^{\infty} dk_r |\mathbf{k}| \mathbf{k}^4 \exp i |\mathbf{k}| r \left( \mathbf{k}^6 - (g_s \boldsymbol{\rho}_0)^2 \right)^{-1} \\
 &= 2\pi i \frac{1}{6} \frac{1}{3} \left[ \begin{aligned}
 & \exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) + \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right) \exp\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right) + \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right) \exp\left(-\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right) \\
 & + \exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right) + \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right) \exp\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right) + \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) \exp\left(-\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) \\
 & + \exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right) + \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) \exp\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) + \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right) \exp\left(-\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right)
 \end{aligned} \right]. \quad (15.39)
 \end{aligned}$$

The reason why  $\frac{1}{3!} x \wedge y \wedge z$  is not appropriate is clear: it would zero out the entire integral by identity, and this is because there is never any place in (15.38) where *three* terms with spatial aspects are multiplied. All we have are terms like  $\exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_y r\right) \exp\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_y r\right)$  where *two* terms are multiplied. So we discard  $-\frac{1}{3}(xzy + zyx + yxz)$  as inappropriate to the way (15.38) is structured and use only  $\frac{1}{3}(xyz + yzx + zxy)$  which is positive signed based on the right-hand-rule convention. The above may now be consolidated using  $2 \cosh x = e^x + e^{-x}$  to:

$$\begin{aligned}
 & \int_{-\infty}^{\infty} dk_r |\mathbf{k}| \mathbf{k}^4 \exp i |\mathbf{k}| r \left( \mathbf{k}^6 - (g_s \boldsymbol{\rho}_0)^2 \right)^{-1} \\
 &= 2\pi i \frac{1}{6} \frac{1}{3} \left[ \begin{aligned}
 & \exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) + \exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right) + \exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right) \\
 & + 2 \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) \cosh\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0x} r\right) \\
 & + 2 \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right) \cosh\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0y} r\right) \\
 & + 2 \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right) \cosh\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_{0z} r\right)
 \end{aligned} \right]. \quad (15.39)
 \end{aligned}$$

Then if we define a unit vector  $\mathbf{U} = U^i \equiv (1,1,1)$  as a notational convenience to consolidate the line with  $\exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_0 r\right)$ , we can use the ordinary dot product in experiential three-dimensional space to further consolidate this to:

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} dk_r |\mathbf{k}| \mathbf{k}^4 \exp i |\mathbf{k}| r \left( \mathbf{k}^6 - (g_s \boldsymbol{\rho}_0)^2 \right)^{-1} \\
 &= 2\pi i \frac{1}{6} \frac{1}{3} \left[ \mathbf{U} \cdot \exp\left(ig_s^{\frac{1}{3}} \boldsymbol{\mu}_0 r\right) + 2 \exp\left(i\frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 r\right) \cdot \cosh\left(\frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 r\right) \right]. \quad (15.40)
 \end{aligned}$$

Plugging this result into (15.20) with  $g_s^{\frac{1}{3}} = (4\pi\alpha_s)^{\frac{1}{6}}$ , then yields our final non-abelian result, which should be carefully contrasted with the abelian (14.32):

$$\begin{aligned}
 E_1 &= -\frac{2(N^2-1)}{(2\pi)^2} \frac{1}{i} \frac{1}{r} \text{Tr} \langle \lambda^i \left| \int_{-\infty}^{\infty} dk_r |\mathbf{k}| \mathbf{k}^4 \exp i|\mathbf{k}|r (\mathbf{k}^6 - g_s^2 \rho_0^2)^{-1} \right| \lambda^i \rangle \\
 &= -\frac{2(N^2-1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \langle \lambda^i \left| \left[ \mathbf{U} \cdot \exp \left( i g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 r \right) + 2 \exp \left( i \frac{1}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 r \right) \cdot \cosh \left( \frac{\sqrt{3}}{2} g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 r \right) \right] \right| \lambda^i \rangle
 \end{aligned} \tag{15.41}$$

To express this in terms of the coupled probability we finally substitute  $g_s^{\frac{1}{3}} \vec{\mu}_0 = \nabla(h\vec{P}_0)$  from (15.32) into the above in the form  $g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 = \nabla(h\mathcal{P}_0)$  to write our final result:

$$\boxed{E_1 = -\frac{2(N^2-1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \langle \lambda^i \left| \left[ \mathbf{U} \cdot \exp \left( i \nabla(h\mathcal{P}_0) r \right) + 2 \exp \left( i \frac{1}{2} \nabla(h\mathcal{P}_0) r \right) \cdot \cosh \left( \frac{\sqrt{3}}{2} \nabla(h\mathcal{P}_0) r \right) \right] \right| \lambda^i \rangle} \tag{15.42}$$

This expresses the potential  $E_1$  at the first order of recursion, for the non-abelian, non-linear gauge theory, as a direct function of the coupled probability field  $h\mathcal{P}_0$ .

Given that  $g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 = \nabla(h\mathcal{P}_0)$  is a definition of both the bare probability  $\mathcal{P}_0$  field and its coupling  $h$  and that  $g_s^{\frac{1}{3}}$  has a definitive interpretation as the cubed root of a running interaction coupling and that each of  $h$  are running scalar coupling numbers, we may also define  $h$  directly in terms of the known  $g_s$  rather simply, by:

$$h \equiv g_s^{\frac{1}{3}}. \tag{15.43}$$

Then, in view of the definition  $g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 = \nabla(h\mathcal{P}_0)$  a.k.a.  $g_s^{\frac{1}{3}} \vec{\mu}_0 = \nabla(h\vec{P}_0)$  in (15.32), (15.43) leads to the differential equation:

$$g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 = \nabla \left( g_s^{\frac{1}{3}} \mathcal{P}_0 \right) = g_s^{\frac{1}{3}} \nabla \mathcal{P}_0 + \frac{1}{3} g_s^{-\frac{2}{3}} \nabla g_s \mathcal{P}_0. \tag{15.44}$$

relating *bare* probability  $\mathcal{P}_0$  field to *bare* probability density  $\boldsymbol{\mu}_0$ . This is easily rewritten as:

$$\boldsymbol{\mu}_0 = \nabla \mathcal{P}_0 + \frac{1}{3} \frac{\nabla g_s}{g_s} \mathcal{P}_0. \tag{15.45}$$

In general, we shall not find the need to explicitly use (15.45) for the bare probability and bare probability density, but shall work with the *coupled* probability  $h\mathcal{P}_0$  and the *coupled* probability density  $g_s^{\frac{1}{3}} \boldsymbol{\mu}_0$  as interrelated by  $g_s^{\frac{1}{3}} \boldsymbol{\mu}_0 = \nabla(h\mathcal{P}_0)$ . Although  $h \equiv g_s^{\frac{1}{3}}$  and we could write the coupled probability as  $g_s^{\frac{1}{3}} \mathcal{P}_0$ , we shall generally opt to use the form  $h\mathcal{P}_0$  so that the equations are not filled with a plethora of  $g_s^{\frac{1}{3}}$  cubed roots. More generally, this will help in resetting us to

think about the probability densities not in terms of the three-dimensional coupled probability densities  $g_s \rho_0$  with which we are of course familiar, but in terms of three independent one-dimensional coupled probability densities  $g_s^{\frac{1}{3}} \mu_0 = h \mu_0$  into which a careful analysis of the content of (15.22) has required us to deconstruct  $g_s \rho_0$ .

Now we shall study (15.42) as well as its abelian counterpart (14.30) from a variety of different viewpoints, to see what these teach us about quantum fields in a non-linear quantum field theory (NLQFT) such as Yang-Mills gauge theory.

### PART III: ANALYTICAL NON-LINEAR QUANTUM FIELD THEORY: SPECIFIC EXAMPLES

#### 16. Constant Probability, Zero Probability Density Fields: Introduction to Analytical Non-Linear Quantum Field Theory

Beginning in this section, we shall use equation (15.42) above to analytically explore the workings of non-linear quantum field theory (NLQFT), using several different illustrative examples of coupled probability fields  $hP_0$ . While  $E_1$  in (15.42) is based on the amplitude density  $\mathfrak{N}(J)_1 = \text{Tr}(J_\sigma \pi_1 J^\sigma)$  at first order of recursion, where from (13.21),  $\pi_1 = (\pi_0^{-1} + \pi_0 J_\tau \pi_0 J^\tau)^{-1}$  and  $\pi_0 = (k_\tau k^\tau + i\varepsilon)^{-1}$ , and while *physical* amplitude densities  $\mathfrak{N}(J) = \text{Tr}(J_\sigma \pi_\infty J^\sigma)$  in (13.20) are expected to have non-linear contributions through infinite recursive order, the use of  $\mathfrak{N}(J)_1$  does introduce a first order non-linearity which is very helpful to flesh out a deep understanding of NLQFT as an *analytical field theory*, versus simply doing numerical calculations with NLQFT without being able to obtain analytical functions involving fields and source potentials.

In this section, starting with (15.42), let us first consider a region of experiential space in which the coupled probability field  $hP_0 = \text{constant}$ . This means that  $g_s^{\frac{1}{3}} \mu_0 = \nabla(hP_0) = \boldsymbol{\theta}$  even for  $g_s \neq 0$  which is the approximation we considered to obtain the  $E_1 = -(1/4\pi)r^{-1}$  potential in (14.31) graphed in Figure 1. So for a region of constant coupled probability  $hP_0 = \text{constant}$ , (15.42) reduces to:

$$\begin{aligned}
 E_1 &= -\frac{2(N^2-1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \langle \lambda^i | \left[ \mathbf{U} \cdot \exp\left(ig_s^{\frac{1}{3}}(\boldsymbol{\theta})r\right) + 2 \exp\left(i\frac{1}{2}g_s^{\frac{1}{3}}(\boldsymbol{\theta})r\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}g_s^{\frac{1}{3}}(\boldsymbol{\theta})r\right) \right] | \lambda^i \rangle \\
 &= -\frac{2(N^2-1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \langle \lambda^i | [\mathbf{U} \cdot \mathbf{U} + 2\mathbf{U} \cdot \mathbf{U}] | \lambda^i \rangle = -\frac{2(N^2-1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \langle \lambda^i | 9 | \lambda^i \rangle \quad .(16.1) \\
 &= -2(N^2-1) \frac{1}{4\pi} \frac{1}{r} \text{Tr} \langle \lambda^i | \lambda^i \rangle = -(N^2-1)^2 \frac{1}{4\pi} \frac{1}{r}
 \end{aligned}$$

This makes use of  $\text{Tr}\langle\lambda^i|\lambda^i\rangle = \frac{1}{2}(N^2 - 1)$  which is just another way of stating the normalization  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$  for SU(N) with  $N^2 - 1$  generators. Contrasting to (14.31), we see that in this limit of the 1/r potential, an overall factor of 1 in the abelian U(1) theory becomes an overall factor of  $(N^2 - 1)^2$  for SU(N), and we explicitly see how the factor of 2 in  $2(N^2 - 1)$  compensates for the generator normalization  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$ .

But what is far more intriguing than comparing the coefficients as between abelian and non-abelian gauge theory is the very deep statement that (16.1) makes about the non-linear quantum field underlying the 1/r potential of linear, abelian gauge theory: In non-linear quantum field theory, a 1/r potential goes hand in hand with a constant coupled probability density  $\hbar P_0 = \text{constant}$ . Restated: *in non-linear quantum field theory, a 1/r potential is the source of a probability field which is constant, while any and all spatially-varying probability fields are sourced by other than a 1/r potential*. Of course, we can certainly set the charge  $g_s^{\frac{1}{3}} = 0$  a.k.a.  $\alpha_s = 0$  in (15.41) a.k.a.  $h = 0$  in (15.42) to arrive at the very same result we see in (16.1), or as we did in (14.30) to arrive at  $E_1 = -(1/4\pi)r^{-1}$  in (14.31). But setting the coupling  $\alpha_s = 0$  – or for that matter setting *any* coupling precisely to zero – is just a mathematical idealization which may be approached but never precisely attained in the real physical world as evidenced by the fact that  $\alpha_{em} \rightarrow 1/137.036\dots$  and no smaller. But one could very readily have a constant probability density over a given region of space in the physical world. And, *it is an absolute certainty that we do have and do observe 1/r potentials in the real physical world*: this is the precise, very well-studied potential of electrodynamics! So (16.1) is telling us something very deep and physically real about non-linear quantum field theory, and we need to explore this. We shall now begin to do so from a number of different viewpoints.

Let us start with Quantum Electrodynamics, QED, which is the paradigm of an abelian gauge theory. Going back to (13.21), the QED amplitude is simply  $\mathfrak{N}(J)_n = \text{Tr}(J_\sigma \pi_n J^\sigma)$  with  $\pi_n = \pi_0 = (k_\tau k^\tau + i\mathcal{E})^{-1}$  and with the trace removed and a factor of  $\frac{1}{2}$  restored because there are no  $\text{Tr}(\lambda^i)^2 = \frac{1}{2}$ -normalized generators. That is, in QED,  $\mathfrak{N}(J)_0 = \frac{1}{2} J_\sigma (k_\tau k^\tau + i\mathcal{E})^{-1} J^\sigma$ , with  $\mathfrak{N}(J)_0$  designating that we are taking (13.21) through zero recursions. Starting with  $\mathfrak{N}(J)_0$ , one can repeat the calculations of sections 14 and 15, and in either case, one will arrive at a potential energy  $E_0 = -(4\pi r)^{-1}$  for a Coulomb charge, as we did in (14.31), which then leads to the Coulomb force law  $\partial E_0 / \partial r = (1/4\pi)r^{-2}$ . See again [11] chapter I.4. If we scale in the electric charge  $e$  which is related to the running coupling by  $4\pi\alpha = e^2 / \hbar c$  with the familiar  $\alpha \rightarrow 1/137.036\dots$  at low probe energy / larger  $r$ , then these become  $E_0 = -e^2 (4\pi r)^{-1} = -\alpha r^{-1}$  along with the inverse-square force law  $\partial E_0 / \partial r = \alpha r^{-2}$  of Coulomb.

Comparing  $E_0 = -e^2(4\pi r)^{-1}$  with (15.42), we see that both of these equations are for a potential energy  $E$ . And, we see that both of these equations are functions of a running charge  $e \Leftrightarrow g_s$  and a radial length  $r$ . But what (15.42) contains which  $E_0 = -e^2(4\pi r)^{-1}$  does not contain, is  $g_s^{\frac{1}{3}} \mu_0 = \nabla(h\mathcal{P}_0)$  which is the gradient of the coupled probability density. So mathematically, in non-Abelian, non-linear gauge theory, the coupled probability field  $h\mathcal{P}_0$  and its gradient  $\nabla(h\mathcal{P}_0)$  are the “new” elements which do not appear at all in a linear theory such as QED. But the reason for this, as we learn from (16.1), is that in QED,  $\nabla(h\mathcal{P}_0) = \mathbf{0}$ . That is, when viewed in the context of a non-linear quantum field theory (15.42), QED has a  $E_0 = -e^2(4\pi r)^{-1}$  potential because its coupled probability density  $h(\mathbf{x})\mathcal{P}_0(\mathbf{x}) = \text{constant}$ . This leads us to conclude: *A seemingly-linear quantum field theory such as QED is actually a special type of a non-linear quantum field theory in which the coupled probability field  $h\mathcal{P}_0$  is constant.*

This is a very fundamental observation. It tells us that all quantum field theory in the observed physical world – even QED – is a non-linear, but that there are special cases such as QED which *appear* linear because the coupled probability field  $h\mathcal{P}_0$  is constant. So while we might ordinarily state that QED is a linear quantum field theory, we may with absolute equivalence assert that QED is a *non-linear* quantum field theory for which the coupled probability – the dimensionless  $h\mathcal{P}_0$  field – is constant. Either viewpoint ends up with the observed potential  $E = -e^2(4\pi r)^{-1}$ . But the latter view allows us to consider QED in the broader context of non-linear quantum field theories, which may be essential, for example, when we consider how electrodynamics results from the breaking of  $SU(2)_w \times U(1)_y \rightarrow U(1)_{em}$ , where  $SU(2)_w$  indubitably is a non-linear quantum field theory, and  $U(1)_y$  is similarly expected to be a non-linear quantum field theory once it is made part of a larger group  $G$  breaking down to include  $U(1)_y$ , see, e.g., [28]. In other words, the latter view enables us to view QED as a non-linear quantum field theory just like all of the other  $SU(N)$  gauge theories that one encounters in particle physics. *QED is then distinguished from all other phenomenological gauge theories not by its being a linear quantum field theory, but by its coupled probability field being constant.* From this view, *all* quantum field theories are nonlinear, but there happen to be particular quantum field theories for which the probability field is constant, and these are the ones with  $1/r$  potentials. This means that electrodynamics is one such non-linear quantum field theory. And, so too, is gravitation in its Newtonian limit!

This also means that potentials other than  $1/r$ , such as the confinement potentials shown in Figures 2 and 3, and presumably the short-range potentials of nuclear interactions, all arise from the common feature that their coupled probability fields  $h(\mathbf{x})\mathcal{P}_0(\mathbf{x})$  are *variable* over the spatial regions being considered. So it is the spatial behavior of  $h(\mathbf{x})\mathcal{P}_0(\mathbf{x})$  which drives the spatial behavior of the potential  $E(\mathbf{x})$ , and vice versa. Consequently, this means that when one compares one type of interaction to another – electromagnetic, weak, strong, nuclear, hadronic – one is in all cases dealing with a common non-linear structure. The dynamical feature which

distinguishes the specifics of one interaction from those of another is the nature of the behavior of the probability density  $h(\mathbf{x})P_0(\mathbf{x})$ , and via (15.42) which applies to any SU(N) interaction, this ties directly to the potential  $E(\mathbf{x})$ . Sometimes we may posit a probability or probability density and deduce the potential. Other times we may posit a potential and deduce a probability density and via integration a probability. But in all cases, non-linear quantum field theory via (15.42) fundamentally links a potential  $E$  to a probability density  $P$  just as assuredly as the classical field equation  $*J = D*F = D*DG$  of (1.12) fundamentally links a source  $J$  to a gauge field  $G$ .

The foregoing is not just an analogy. It is a deep and fundamental feature of quantum reality. Classical theory contains coupled gauge fields  $g\mathcal{G}$ . But these are the variables of integration in the path integral, so by definition they are stripped away and there is no gauge field left in quantum field theory. Something else takes the place of the gauge field, and (15.42) tells us that this is the coupled probability field  $hP$ . And, while the source  $J$  does survive the path integration into  $\mathfrak{N}(J)_n = \text{Tr}(J_\sigma \pi_n J^\sigma)$ , the dynamical source object in the quantum field theory turns out not be  $J$ , but is rather the momentum-space integrated *quantum action*  $W(J) = \int (d^4k / (2\pi)^4) \mathfrak{N}(J)$  which we see, for example, in (14.6). And, by the time one carries out not only the  $d^4k$  but also the spacetime integrals  $\iint d^4x d^4y g_s^2 J_\sigma(x) D_1(x-y) J^\sigma(y)$  as in (14.8) and (15.5), what survives is  $W = -ET$ , see (14.14) and (15.18). If we factor out the time  $T$  as we did in the section 14 and 15 calculations, then the *quantum action source*  $W$  is replaced by a *quantum potential source*  $E$  as in (15.42). If we were to do some alternative calculation that did not factor out time-dependency, then we would remain with  $W = -ET$  which is dimensioned as angular momentum a.k.a. action, which is also the dimension of Planck's constant  $\hbar$ . So just as the field  $g\mathcal{G} \Rightarrow hP$  when going from classical to quantum field theory, so too, the source  $J \Rightarrow W$ . To sum up: in quantum field theory, the quantum action  $W$  is the source of the quantum probability field  $P$ , just as in classical field theory the current density  $J$  is the source of classical gauge field  $G$ .

The next several sections will explore and deepen all of these analytical statements, using the respective examples of constant isotropic probability densities, isotropic Gaussian probability densities, and the observed probability densities of single and double slit experiments.

## 17. Constant, Isotropic Probability Densities and Confining Stable Quantum Potentials in Non-Linear Quantum Field Theory

In the last section we considered the special case of a spatially-constant coupled probability field  $hP_0$  for which the probability density  $\nabla(hP_0) = 0$  and showed how this corresponds with an inverse-square potential  $E = -(1/4\pi r)$ . As we now start to consider a variety of non-zero densities  $\nabla(hP_0) \neq 0$  and thus spatially-varying coupled probability fields  $hP_0$ , it will help to first transform (15.42) from Cartesian into spherical coordinates

$x^i = (x, y, z) \rightarrow x'^i = (r, \theta, \phi)$ . The first step is to expand (15.42) into each of its additive terms using the gradient vector  $\nabla = \partial_i = (\partial_x, \partial_y, \partial_z)$  in Cartesian coordinates, thus:

$$E_1 = -\frac{2(N^2 - 1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \left\langle \lambda^i \left| \begin{array}{l} \exp(i\partial_x (hP_0) r) + \exp(i\partial_y (hP_0) r) + \exp(i\partial_z (hP_0) r) \\ + 2 \exp(i\frac{1}{2}\partial_x (hP_0) r) \cosh\left(\frac{\sqrt{3}}{2}\partial_x (hP_0) r\right) \\ + 2 \exp(i\frac{1}{2}\partial_y (hP_0) r) \cosh\left(\frac{\sqrt{3}}{2}\partial_y (hP_0) r\right) \\ + 2 \exp(i\frac{1}{2}\partial_z (hP_0) r) \cosh\left(\frac{\sqrt{3}}{2}\partial_z (hP_0) r\right) \end{array} \right. \right\rangle. \quad (17.1)$$

Second, since  $\nabla = \partial_i = (\partial_x, \partial_y, \partial_z) = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$  is a three-vector in physical space, it will have the same transformation properties as reciprocals of  $d\mathbf{x} = dx^i = (dx, dy, dz)$ . Specifically, leaving the time as is,  $dt' = dt$ , the invariant differential length element in the physical three-space is  $dl^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2$ . So we orient  $r \geq 0$  to align with the positive x axis, so that  $dx = dr$ . With the positive z-axis at  $\theta = 0$ , we define  $0 \leq \theta \leq \pi$  to represent the descent angle from this +z axis, and  $0 \leq \phi \leq 2\pi$  to represent right-handed rotation about the z axis with the positive x axis at  $\phi = 0$ . We can then pick off the components  $dx'^i = (dr, r d\theta, r \sin \theta d\phi)$  from the square roots of terms in  $dl^2$ . Then, because  $\nabla = \partial_i = \partial/\partial x^i$ , the transformed  $dx'^i$  will take on the form  $\partial'_i = (\partial_r, \partial_\theta / r, \partial_\phi / r \sin \theta)$ . Thus, throughout (16.1) we may substitute  $\partial_i = (\partial_x, \partial_y, \partial_z) \rightarrow \partial'_i = (\partial_r, \partial_\theta / r, \partial_\phi / r \sin \theta)$  to obtain:

$$E_1 = -\frac{2(N^2 - 1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \left\langle \lambda^i \left| \begin{array}{l} \exp(i\partial_r (hP_0) r) + \exp(i\partial_\theta (hP_0)) + \exp(i\partial_\phi (hP_0) / \sin \theta) \\ + 2 \exp(i\frac{1}{2}\partial_r (hP_0) r) \cosh\left(\frac{\sqrt{3}}{2}\partial_r (hP_0) r\right) \\ + 2 \exp(i\frac{1}{2}\partial_\theta (hP_0)) \cosh\left(\frac{\sqrt{3}}{2}\partial_\theta (hP_0)\right) \\ + 2 \exp(i\frac{1}{2}\partial_\phi (hP_0) / \sin \theta) \cosh\left(\frac{\sqrt{3}}{2}\partial_\phi (hP_0) / \sin \theta\right) \end{array} \right. \right\rangle. \quad (17.2)$$

Third, keeping in mind that in spherical coordinates  $r \geq 0$ , let us consider only coupled probabilities which are isotropic under rotations about  $r=0$ . Thus, we consider the special set of coupled probabilities for which  $\partial_\theta (hP_0) = \partial_\phi (hP_0) = 0$ . Given that  $\exp(0) = \cosh(0) = 1$ , such an isotropic probability further simplifies (17.2) to:

$$\begin{aligned}
 E_1 &= -\frac{2(N^2-1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \left\langle \lambda^i \left[ \exp(i\partial_r(hP_0)r) + 2\exp(i\frac{1}{2}\partial_r(hP_0)r) \cdot \cosh\left(\frac{\sqrt{3}}{2}\partial_r(hP_0)r\right) + 6 \right] \lambda^i \right\rangle \\
 &= -\frac{2(N^2-1)}{3} \frac{1}{4\pi} \frac{1}{r} \frac{1}{3} \text{Tr} \left\langle \lambda^i \left[ \exp(i\partial_r(hP_0)r) + 2\exp(i\frac{1}{2}\partial_r(hP_0)r) \cdot \cosh\left(\frac{\sqrt{3}}{2}\partial_r(hP_0)r\right) \right] \lambda^i \right\rangle \quad (17.3) \\
 &\quad -\frac{2}{3}(N^2-1)^2 \frac{1}{4\pi} \frac{1}{r}
 \end{aligned}$$

where we have used  $\text{Tr} \langle \lambda^i | 6 | \lambda^i \rangle = 6 \cdot \frac{1}{2} (N^2 - 1)$  to separate out a pure  $1/r$  potential which is added to the rest of the expression, using  $N^2 - 1$  independent coupled probability fields  $hP_0^i$ .

Fourth, and finally, the expressions  $\partial_r(hP_0) = \lambda^i \partial_r(hP_0^i)$  will have an  $N \times N$  dimensionality for  $SU(N)$ . A main purpose of section 15 was to obtain an equation for the potential which contains the non-abelian probability  $P_0 = \lambda^i P_0^i$ , but in the process, we also achieved a decomposition of the three-dimensional probability density into each of its one-dimensional components. This makes (15.42) distinctly different from (14.32) even without  $P_0 = \lambda^i P_0^i$ . So to gain an appreciation of the general behavior of (15.42) in a variety of forms including (17.2) and (17.3), let us simplify (17.2) to its abelian form akin to (14.32). The easiest way to make sure we match up the overall coefficients is to work from (17.2) and keep in mind the correspondences laid out just after (16.1). Specifically, we reverse-migrate  $2\text{Tr} \langle \lambda^i | \dots | \lambda^i \rangle \rightarrow N^2 - 1$ ,  $(N^2 - 1)^2 \rightarrow 1$ , and  $P_0 \rightarrow P_0$  to turn (17.2) into the abelian form:

$$E_1 = -\frac{1}{4\pi} \frac{1}{r} \frac{1}{9} \left[ \begin{array}{l} \exp(i\partial_r(hP_0)r) + \exp(i\partial_\theta(hP_0)) + \exp(i\partial_\phi(hP_0)/\sin\theta) \\ +2\exp(i\frac{1}{2}\partial_r(hP_0)r) \cdot \cosh\left(\frac{\sqrt{3}}{2}\partial_r(hP_0)r\right) \\ +2\exp(i\frac{1}{2}\partial_\theta(hP_0)) \cdot \cosh\left(\frac{\sqrt{3}}{2}\partial_\theta(hP_0)\right) \\ +2\exp(i\frac{1}{2}\partial_\phi(hP_0)/\sin\theta) \cdot \cosh\left(\frac{\sqrt{3}}{2}\partial_\phi(hP_0)/\sin\theta\right) \end{array} \right]. \quad (17.4)$$

This now corresponds directly to (14.32) transformed to polar coordinates, and in recognition of the finding developed throughout section 15 that the  $(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} = (g_s \rho_0)^{\frac{1}{3}}$  in (14.32) all represent probability densities deconstructed to the three space coordinates. We see that for a constant probability density  $\partial'_i = (\partial_r, \partial_\theta, \partial_\phi) P_0 = 0_i$ , this will reduce to the electrodynamic Coulomb potential  $E = -(1/4\pi)r^{-1}$ , which is identical to the result in (16.1) except that we have used a spherical coordinate system and we have accounted for the correspondences laid out just after (16.1). For an isotropic  $\partial_\theta(hP_0) = \partial_\phi(hP_0) = 0$ , (17.4) reduces to an abelian counterpart of (17.3), namely:

$$E_1 = -\frac{1}{4\pi} \frac{1}{r} \left( \frac{2}{3} + \frac{1}{9} \left[ \exp(i\partial_r (hP_0) r) + 2 \exp(i\frac{1}{2}\partial_r (hP_0) r) \cdot \cosh\left(\frac{\sqrt{3}}{2}\partial_r (hP_0) r\right) \right] \right). \quad (17.5)$$

Now, in Figures 2 and 3 we showed how the potential (14.32) behaves if we regard  $f \equiv (4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} = (g_s \rho_0)^{\frac{1}{3}}$  as a *constant* frequency. Regarding  $f = (g_s \rho_0)^{\frac{1}{3}}$  as a *constant* was an *assumption* we made at the time to gain a sense for the behavior of this potential, and we saw that even at first recursive order, this potential exhibited attributes of confinement. In the interim, we have seen that  $(g_s \rho_0)^{\frac{1}{3}}$  really needs to be thought of as a probability density along a single one of the space coordinates x, y, z. So prior to (15.23) and thereafter we established that for non-abelian gauge theory  $\rho_0 \equiv \frac{1}{3!} \mu_{0x} \wedge \mu_{0y} \wedge \mu_{0z}$  defines a spatial vector of probability densities  $\mu_0 = (\mu_{0x}, \mu_{0y}, \mu_{0z})$  along each of the three space dimensions, which relationship for abelian theory simplifies to  $\rho_0 \equiv \mu_{0x} \mu_{0y} \mu_{0z}$ . Because (17.5) is an abelian equation, we may use the simpler  $\rho_0 \equiv \mu_{0x} \mu_{0y} \mu_{0z}$ . Therefore,  $g_s \rho_0 \equiv (g_s^{\frac{1}{3}} \mu_{0x}) (g_s^{\frac{1}{3}} \mu_{0y}) (g_s^{\frac{1}{3}} \mu_{0z})$ , so the earlier treatment assuming  $f = (g_s \rho_0)^{\frac{1}{3}} = \text{constant}$  corresponds in light of what we have learned since, to regarding  $g_s^{\frac{1}{3}} \mu_0 = \text{constant}$ . Further, at (15.32) we linked this to the gradient of the coupled probability field via  $g_s^{\frac{1}{3}} \bar{\mu}_0 = \nabla(h\bar{P}_0)$  which for abelian theory is simply  $g_s^{\frac{1}{3}} \mu_0 = \nabla(hP_0)$ . So the assumption back at (14.32) that  $f = (g_s \rho_0)^{\frac{1}{3}} = \text{constant}$  from which we then proceeded to develop Figures 2 and 3 corresponds to regarding  $g_s^{\frac{1}{3}} \mu_0 = \nabla(hP_0) = \text{constant}$  in (17.4) and (17.5). Further, with the isotropic probability  $\partial_\theta(hP_0) = \partial_\phi(hP_0) = 0$  assumed in (17.5), the former  $f = (g_s \rho_0)^{\frac{1}{3}} = \text{constant}$  assumption translates into assuming that  $\partial_r(hP_0) = \text{constant}$  in (17.5). So, we shall now examine (17.5) for the condition that  $\partial_r(hP_0) = \text{constant}$ , which is the same condition used in our earlier consideration of (14.32) for  $f = (g_s \rho_0)^{\frac{1}{3}} = \text{constant}$ , which was then drawn out in Figures 2 and 3. This allows an ‘‘apples-to-apples’’ comparison. Finally, because we are considering an isotropic probability, this means that  $hP_0$  is a function *exclusively* of  $r$ , and *not* of  $\theta$  and  $\phi$ . So while  $\rho_0 = \mu_{0x} \mu_{0y} \mu_{0z}$  is expressed in Cartesian coordinates, for this isotropic, radial-only-dependent probability field, we have  $g_s^{\frac{1}{3}} \mu_{0x} = g_s^{\frac{1}{3}} \mu_{0y} = g_s^{\frac{1}{3}} \mu_{0z} = g_s^{\frac{1}{3}} \mu_{0r}$ , with the consequence that  $\rho_0 = \mu_{0r}^3$ .

Given the above, if we now posit that  $\partial_r(hP_0) = \text{constant} \equiv A$  where  $A$  is a positive, non-zero, real constant, that means that  $y = hP_0$  will have the general form of a linear equation  $y = Ax + B$  with  $x = r$  and with  $A$  and  $B$  being constants. To simplify as much as possible, let us discard the constant of integration  $B$ , and simply posit an isotropic coupled probability field of the general linear form with  $B = 0$ , thus:

$$hP_0 = Ar . \quad (17.6)$$

Clearly,  $\partial_r (hP_0) = A$ , consistent with our supposition of a constant radial probability density. Because  $hP_0$  has a mass dimension of zero and  $r$  has mass dimension -1, the constant  $A$  must have a mass dimensionality of +1. Now our task will be to link this constant  $A$  to a confinement potential and to phenomenological strong interaction data, and to see whether discarding the integration constant runs into any contradictions. We do this recognizing that (17.5) is still an abelian approximation, but a better one and with more information developed than was (14.32).

The first thing we may now do is substitute  $\partial_r (hP_0) = A$  into (17.5) to write:

$$E_1 = -\frac{1}{4\pi} \frac{1}{r} \left( \frac{2}{3} + \frac{1}{9} \left[ \exp(iAr) + 2 \exp\left(i\frac{1}{2}Ar\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}Ar\right) \right] \right). \quad (17.7)$$

Next, so we can study (17.7) in a dimensionless fashion, let us define a dimensionless radial coordinate via the to-be-determined constant  $A$  with a mass dimension of +1 as such:

$$R \equiv Ar \quad (17.8)$$

which also means that  $r^{-1} = AR^{-1}$ . Therefore, we use (17.8) in (17.7) to write:

$$4\pi \frac{E_1}{A} = -\frac{1}{R} \left( \frac{2}{3} + \frac{1}{9} \left[ \exp(iR) + 2 \exp\left(i\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \right] \right). \quad (17.9)$$

Now, keeping in mind the way in which the potential  $E_1$  in Figures 2 and 3 appeared to be confining, we will want to similarly examine the behavior of (17.9) above.

As with (14.34), this is a complex number, so we again wish to ascertain the  $|E_1|^2 = E_1 * E_1$  and then obtain  $|E_1| = \pm \sqrt{|E_1|^2}$ . This equation has a similar form to (14.34), but here, it may be written as:

$$\begin{aligned} 4\pi \frac{E_1}{A} &= -\frac{1}{R} \left( \frac{2}{3} + \frac{1}{9} \cos(R) + \frac{2}{9} \cos\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) + i \left[ \frac{1}{9} \sin(R) + \frac{2}{9} \sin\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \right] \right) \\ &= -R^{-1} (a + bi) \end{aligned} \quad (17.10)$$

where:

$$\begin{aligned} a &\equiv \frac{2}{3} + \frac{1}{9} \cos(R) + \frac{2}{9} \cos\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \\ b &\equiv \frac{1}{9} \sin(R) + \frac{2}{9} \sin\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \end{aligned} \quad (17.11)$$

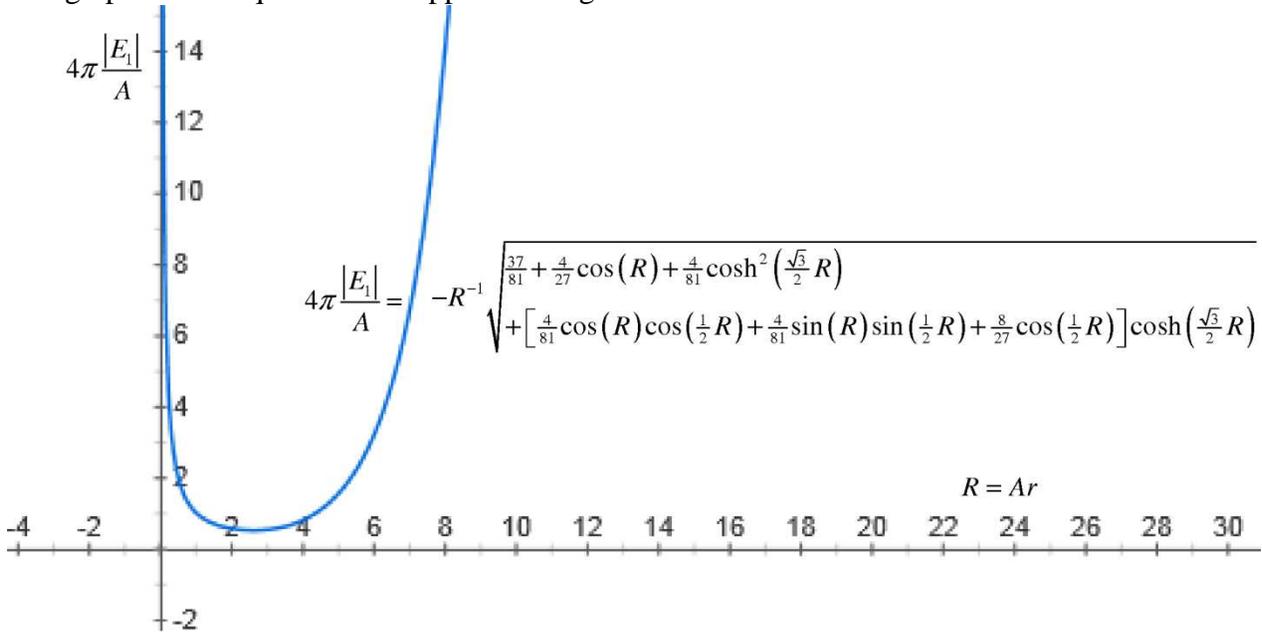
The decomposition of the probability density into three separate spatial components introduces a new term of  $\frac{2}{3}$  into the real portion of (17.9), which carries contributions from the two coordinates  $\theta, \phi$  over which the present example is isotropic. So now, contrast (14.37):

$$\begin{aligned}
 |a + bi|^2 &= a^2 + b^2 = \frac{4}{9} + \frac{4}{3} \left[ \frac{1}{9} \cos(R) + \frac{2}{9} \cos\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \right] \\
 &+ \frac{1}{81} \cos^2(R) + \frac{4}{81} \cos^2\left(\frac{1}{2}R\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{2}R\right) + \frac{4}{81} \cos(R) \cos\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \\
 &+ \frac{1}{81} \sin^2(R) + \frac{4}{81} \sin^2\left(\frac{1}{2}R\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{2}R\right) + \frac{4}{81} \sin(R) \sin\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) . \tag{17.12} \\
 &= \frac{37}{81} + \frac{4}{27} \cos(R) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{2}R\right) \\
 &+ \left[ \frac{4}{81} \cos(R) \cos\left(\frac{1}{2}R\right) + \frac{4}{81} \sin(R) \sin\left(\frac{1}{2}R\right) + \frac{8}{27} \cos\left(\frac{1}{2}R\right) \right] \cosh\left(\frac{\sqrt{3}}{2}R\right)
 \end{aligned}$$

Making use of the above in (17.10) while taking  $|E_1|$  then yields:

$$4\pi \frac{|E_1|}{A} = \mp R^{-1} \sqrt{\frac{37}{81} + \frac{4}{27} \cos(R) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{2}R\right) + \left[ \frac{4}{81} \cos(R) \cos\left(\frac{1}{2}R\right) + \frac{4}{81} \sin(R) \sin\left(\frac{1}{2}R\right) + \frac{8}{27} \cos\left(\frac{1}{2}R\right) \right] \cosh\left(\frac{\sqrt{3}}{2}R\right)} . \tag{17.13}$$

The graph of this equation now appears in Figure 4 below:



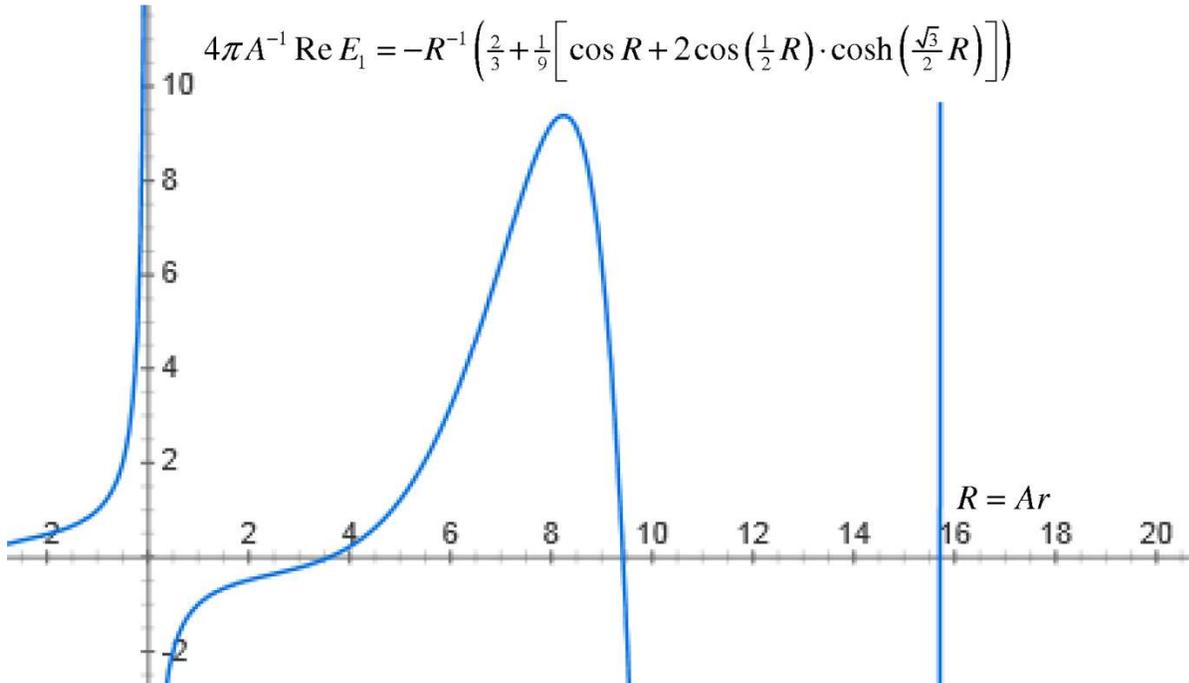
**Figure 4: The Yang-Mills Potential  $|E_1|$  of (17.13) at First Recursive Order**

We see that this curve exhibits characteristics of confinement and collapse-averting stability, just like Figure 2. But the decomposition of the three-dimensional probability density into three separate one-dimensional components and the isotropy along the  $\theta, \phi$  space coordinates has reduced the amplitude of the curve somewhat, moving the minimum of the potential to the right

and down from  $(R, |E_1| / Af) \approx (1.668, 3.118)$  in Figure 2 to  $(R, |E_1| / Af) \approx (2.623, 0.532)$  above, and somewhat flattened the minimum region from about  $R = 2$  to  $R = 4$ . As with Figure 2, however, there is nothing in Figure 4 which gives us a basis upon which to introduce a mass scale. As we did with Figure 3, let us now graph the real portion only, of (17.9). We may obtain this directly from (17.10) with the imaginary terms set to zero, namely:

$$4\pi A^{-1} \text{Re } E_1 = -R^{-1} \left( \frac{2}{3} + \frac{1}{9} \cos(R) + \frac{2}{9} \cos\left(\frac{1}{2}R\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \right). \quad (17.14)$$

This function is shown in Figure 5 below:



**Figure 5: Equation (17.14) for  $\text{Re } E_1$ , showing a First-Order Confinement Peak at  $R_{\text{peak}} \cong 8.245$**

Although the amplitude of Figure 5 is significantly reduced from that of Figure 3, as was the Figure 4 amplitude reduced from Figure 2, the  $R$ -coordinate of the peak appears to have stayed essentially the same. The potential peaked at  $(R, \text{Re}(E_1) / Af) \approx (8.245, 85.184)$  in Figure 3. In Figure 5 it peaks at  $(R, \text{Re}(E_1) / Af) \approx (8.245, 9.384)$ . And the next upward sinusoidal crossing of the  $R$  axis appears near  $R \approx 15.71$ , just as in Figure 3. So the decomposition of the probability three-density into three one-densities appears to substantially diminish the amplitude of these curves and thus the magnitude of the potential energy, but has very minimal effect, if any, on the  $R$ -dependent “frequency” aspects of these curves. It is also interesting to note in passing that with the peak in Figure 5 situated at  $R_{\text{peak}} = 8.245$  and the minimum in Figure 4 situated at  $R_{\text{min}} = 2.623$  (which Figures are both from the same underlying equation (17.9)), the ratio  $R_{\text{peak}} / R_{\text{min}} = 8.245 / 2.623 = 3.143 \approx \pi$  to the third decimal place. Of course, with the

various sinusoidal activities going on in (17.9), it is not surprising to find the number  $\pi$  is some places, but this ratio is worth keeping this in mind.

Based on Figure 5 and (17.8), let us now define the approximate number:

$$R_{\text{peak}} \equiv Ar_{\text{peak}} \equiv 8.245, \quad (17.15)$$

to be the radial distance at which the first peak occurs in Figure 5. This means we may now use (17.15) to write (17.6) as:

$$hP_0 = Ar = R = R_{\text{peak}} \frac{r}{r_{\text{peak}}} = 8.245 \frac{r}{r_{\text{peak}}}. \quad (17.16)$$

Now, we want to associate the peak at  $R = Ar \cong 8.245$  with confinement, at least at the first recursive order. There are several steps we take to do this. First, we require that the only domain over which the radial probability density  $\partial_r(hP_0)$  is *not zero*, is the domain from  $0 \leq R \leq R_{\text{peak}} \cong 8.245$ , i.e., from  $0 \leq r \leq r_{\text{peak}}$ . So we effectively regard all of the portions of Figures 4 and 5 *outside* the domain  $0 \leq R \leq R_{\text{peak}}$  as having a zero probability density,  $\partial_r(hP_0) = 0$ , while all regions inside this domain are regarded to have a constant density  $\partial_r(hP_0) = A \neq 0$ .

Next, we recall the discussion just prior to (14.42) where we made the association  $r_\Lambda = 2.178F \Leftrightarrow R \cong 8.245$ , that is where we associated the peak potential at  $R_{\text{peak}}$  with the length  $r_\Lambda$  that is in turn associated with  $\Lambda_{QCD}$ . Earlier, we simply made this association to arrive at some order of magnitude estimates. Now, we seek to embed  $\Lambda_{QCD}$  directly into the equations for the one-recursion potential  $E_1$ . So let's do this again: we take the six-quark cutoff  $\Lambda_{QCD}^{(6)} = .0906\text{GeV}$  to be the energy at which the strong coupling grows infinite and confinement takes place. From here simplifying notation to  $\Lambda \equiv \Lambda_{QCD}^{(6)}$ , the associated deBroglie length  $r_\Lambda = \hbar c / \Lambda^{(6)}_{QCD}$  is explicitly calculated to be  $r_\Lambda = 2.178F$  in natural  $\hbar = c = 1$  units. So we now define:

$$r_{\text{peak}} \equiv r_\Lambda = \frac{1}{\Lambda} = 2.178F, \quad (17.17)$$

$$R_{\text{peak}} \equiv R_\Lambda \cong 8.245 \quad (17.18)$$

By so-identifying the peak in Figure 5 with the QCD cutoff we may rewrite (17.16) as:

$$hP_0 = Ar = R_\Lambda \frac{r}{r_\Lambda} \cong 8.245 \frac{r}{r_\Lambda}. \quad (17.19)$$

This means, including  $\Lambda r_\Lambda = 1$ , that:

$$g_s^{\frac{1}{3}} \mu_{0r} = \partial_r (hP_0) = A = \frac{R_\Lambda}{r_\Lambda} = R_\Lambda \Lambda \cong \frac{8.245}{r_\Lambda}. \quad (17.20)$$

Also, from (17.8) and  $A = R_\Lambda / r_\Lambda$  embedded in (17.19) and  $\Lambda r_\Lambda = 1$ , we have:

$$R = Ar = \frac{R_\Lambda}{r_\Lambda} r = R_\Lambda \Lambda r \quad (17.21)$$

So with all of this, we return to (17.5) and use  $\partial_r (hP_0) = R_\Lambda \Lambda$  from (17.20), and divide both sides by  $R_\Lambda \Lambda = R / r$  which is a variant of (17.21), to obtain:

$$E_1 = -\frac{1}{4\pi} \frac{1}{r} \left( \frac{2}{3} + \frac{1}{9} \left[ \exp(iR_\Lambda \Lambda r) + 2 \exp(i \frac{1}{2} R_\Lambda \Lambda r) \cdot \cosh\left(\frac{\sqrt{3}}{2} R_\Lambda \Lambda r\right) \right] \right). \quad (17.22)$$

Now the QCD cutoff is embedded into (17.22) to provide physical mass and length scales.

The final step is to ensure that the density of this probability integrates to 1 over the now-relevant domain of  $0 \leq r \leq r_{\text{peak}}$ , that is, to ensure that  $\int_0^{r_{\text{peak}}} \partial_r (hP_0) dr = 1$ . To do this, we now introduce a normalization constant  $N$ , and normalize (17.16), also using (17.6), (17.17) and (17.18) and  $\Lambda r_\Lambda = 1$ , to:

$$hP_0 \equiv NAr = NR = NR_\Lambda \frac{r}{r_\Lambda} = 8.245N \frac{r}{r_\Lambda} = NR_\Lambda \Lambda r = 8.245N \Lambda r. \quad (17.23)$$

This effectively is a *definition* of  $N$ . Then to ascertain the *value* of  $N$ , we first obtain:

$$g_s^{\frac{1}{3}} \mu_{0r} = \partial_r (hP_0) = NA = N \frac{R_{\text{peak}}}{r_{\text{peak}}} = N \frac{8.245}{r_{\text{peak}}}. \quad (17.24)$$

And then we evaluate the definite integral:

$$\int_0^{r_{\text{peak}}} g_s^{\frac{1}{3}} \mu_{0r} dr = \int_0^{r_{\text{peak}}} \partial_r (hP_0) dr = \int_0^{r_{\text{peak}}} N \frac{R_{\text{peak}}}{r_{\text{peak}}} dr = N \frac{R_{\text{peak}}}{r_{\text{peak}}} r \Big|_0^{r_{\text{peak}}} = NR_{\text{peak}} = 1. \quad (17.25)$$

So from this, we fix the normalization constant also via (17.18) to be:

$$N = \frac{1}{R_{\text{peak}}} = \frac{1}{R_{\Lambda}} \cong \frac{1}{8.245}. \quad (17.26)$$

So having determined  $N$ , and writing (17.23) as  $hP_0 \equiv NAr = NR = R/R_{\Lambda}$  with (17.26) applied in the final step, we can achieve this normalization by rescaling:

$$R \Rightarrow R/R_{\Lambda} \cong R/8.245. \quad (17.27)$$

So as a result of the normalization (17.27), the relation in (17.16) is renormalized using (17.17) and (17.18) and  $\Lambda r_{\Lambda} = 1$  via:

$$hP_0 = R = R_{\Lambda} \frac{r}{r_{\Lambda}} = R_{\Lambda} \Lambda r \Rightarrow hP_0 = R/R_{\Lambda} = \frac{r}{r_{\Lambda}} = \Lambda r. \quad (17.28)$$

Also, based directly on  $R \Rightarrow R/R_{\Lambda}$  in (17.27), the relation in (17.21) is renormalized via:

$$R = \frac{R_{\Lambda} r}{r_{\Lambda}} = R_{\Lambda} \Lambda r \Rightarrow R = \frac{r}{r_{\Lambda}} = \Lambda r \quad (17.29)$$

From the above (17.29), we also deduce:

$$\frac{1}{r} = \frac{1}{R} R_{\Lambda} \Lambda \Rightarrow \frac{1}{r} = \Lambda \frac{1}{R} \quad (17.30)$$

So finally we return to (17.22). From (17.29) we renormalize with  $R_{\Lambda} \Lambda r \Rightarrow \Lambda r$  and from (17.30) we renormalize with  $r^{-1} \Rightarrow R_{\Lambda}^{-1} r^{-1}$ . Thus we have:

$$E_1 = -\frac{1}{4\pi} \frac{1}{R_{\Lambda} r} \left( \frac{2}{3} + \frac{1}{9} \left[ \exp(i\Lambda r) + 2 \exp(i\frac{1}{2}\Lambda r) \cdot \cosh\left(\frac{\sqrt{3}}{2}\Lambda r\right) \right] \right). \quad (17.31)$$

We then divide both sides through by  $\Lambda$  while moving over the  $4\pi$  and the  $R_{\Lambda}$ . Then with the renormalization complete, in the second line below we substitute  $R = \Lambda r$  from (17.19) to obtain:

$$\begin{aligned} 4\pi R_{\Lambda} \frac{E_1}{\Lambda} &= -\frac{1}{\Lambda r} \left( \frac{2}{3} + \frac{1}{9} \left[ \exp(i\Lambda r) + 2 \exp(i\frac{1}{2}\Lambda r) \cdot \cosh\left(\frac{\sqrt{3}}{2}\Lambda r\right) \right] \right) \\ &= -\frac{1}{R} \left( \frac{2}{3} + \frac{1}{9} \left[ \exp(iR) + 2 \exp(i\frac{1}{2}R) \cdot \cosh\left(\frac{\sqrt{3}}{2}R\right) \right] \right). \end{aligned} \quad (17.32)$$

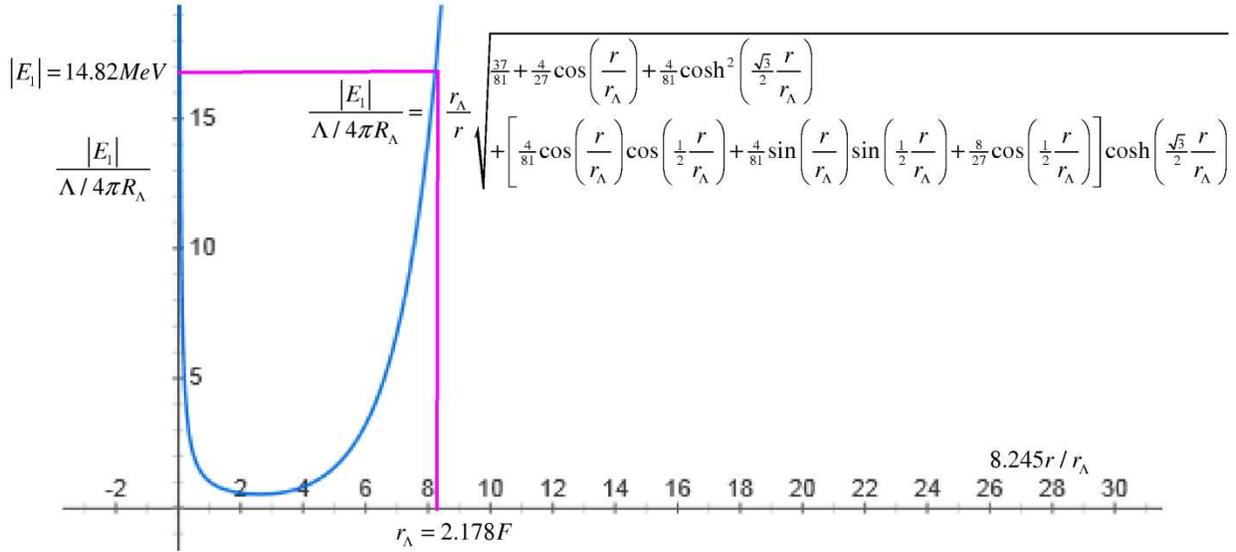
It will be appreciated that the right hand side of the second line of (17.32) is absolutely identical to the right hand side of (17.9), the magnitude and real portion of which was graphed in Figures

4 and 5. But what was previously  $R = Ar$  in (17.8) is now  $R = \Lambda r$ , with  $A$  replaced by  $\Lambda = \Lambda_{\text{QCD}}$ . On the left hand side of the second line above, as a result of the connections (17.17) and (17.18) to  $\Lambda$  and the renormalization (17.26) to ensure that the probability density integrates to 1, we have had the constant divisor  $A$  in (17.8) replaced by  $A \Rightarrow \Lambda / R_\Lambda \cong \Lambda / 8.245$ , with the result that the normalization ultimately bleeds through to a rescaling of the energy by  $E_1 \Rightarrow R_\Lambda E_1$ . As to the probability field, the original relationship  $hP_0 = Ar$  of (17.6) has been replaced by  $hP_0 = \Lambda r$  in (17.30). And finally, following (17.16), and in view of (17.18), we originally set the domain for the non-zero probability density to run from  $0 \leq (R = Ar) \leq R_\Lambda \cong 8.245$ . With the renormalization (17.29), we now set the this domain to commensurately run over  $0 \leq (R = \Lambda r) \leq R_\Lambda \cong 8.245$ .

Now, although (17.32) is the same function as (17.9) graphed in Figures 4 and 5, because it now contains the physical content of  $\Lambda_{\text{QCD}}$  following renormalization, let us again graph this to include this new information as to physical scale. First, we write the magnitude of the second line of (17.32), which we can obtain directly from (17.13) with no more than an  $A \rightarrow \Lambda$  and  $R = r / r_\Lambda$  substitution, thus:

$$\frac{|E_1|}{\Lambda / 4\pi R_\Lambda} = \mp \frac{r_\Lambda}{r} \sqrt{\frac{37}{81} + \frac{4}{27} \cos\left(\frac{r}{r_\Lambda}\right) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{2} \frac{r}{r_\Lambda}\right) + \left[ \frac{4}{81} \cos\left(\frac{r}{r_\Lambda}\right) \cos\left(\frac{1}{2} \frac{r}{r_\Lambda}\right) + \frac{4}{81} \sin\left(\frac{r}{r_\Lambda}\right) \sin\left(\frac{1}{2} \frac{r}{r_\Lambda}\right) + \frac{8}{27} \cos\left(\frac{1}{2} \frac{r}{r_\Lambda}\right) \right] \cosh\left(\frac{\sqrt{3}}{2} \frac{r}{r_\Lambda}\right)}. \quad (17.33)$$

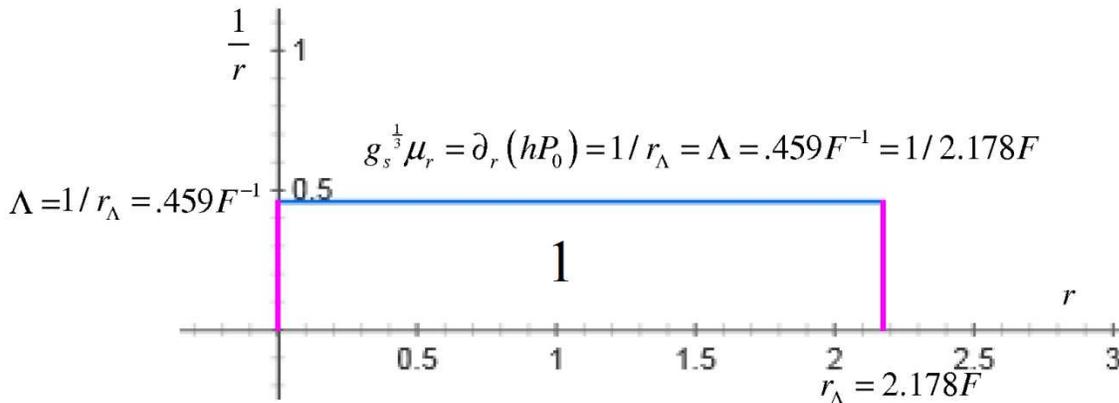
As noted prior to (17.17), the mean empirical value of the QCD cutoff is  $\Lambda^{(6)}_{\text{QCD}} = 90.6 \text{ MeV}$  with a corresponding  $r_\Lambda = 2.178F$ . So  $\Lambda / 4\pi R_\Lambda = 0.874 \text{ MeV}$ . Thus, for example, when  $4\pi R_\Lambda |E_1| / \Lambda = 10$ , this means  $|E_1| = 10 \cdot \Lambda / 4\pi R_\Lambda = 8.74 \text{ MeV}$ . So while we still use (17.33) to graph the dimensionless range  $4\pi R_\Lambda |E_1| / \Lambda$  against the dimensionless domain  $r / r_\Lambda$ , we can also show the actual physical energies and lengths along the axes which correspond to this plot. Since the cutoff is designed to be at  $r_{\text{peak}} \equiv r_\Lambda$  via (17.17), and since  $r \geq 0$ , we also limit the domain to  $0 \leq r \leq r_\Lambda$ . Thus, as in Figure 4, we may graph the unlike-charge potential:



**Figure 6: The Yang-Mills Potential  $|E_1|$  of (17.33) at First Recursive Order (Figure 4 with Physical Energies and Length Scales)**

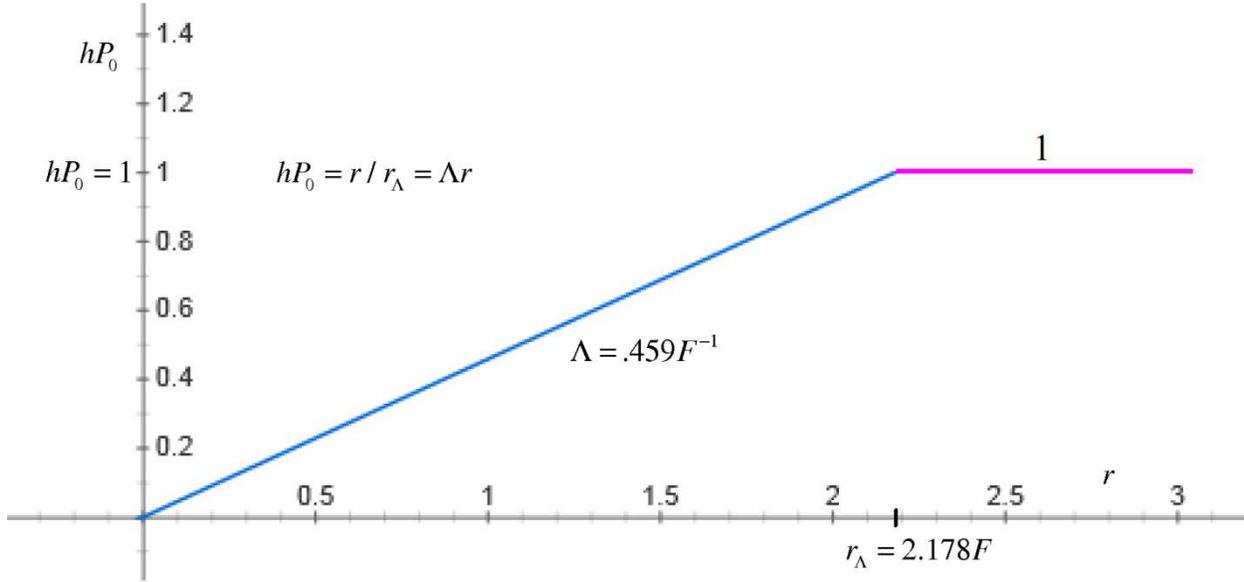
The peak found in Figure 5 at  $R_{\text{peak}} = 8.245$  corresponds to the cutoff length  $r_\Lambda = 2.178F$ , and is introduced into Figure 6. In Figure 6,  $R_{\text{peak}} = 8.245$  has a dimensionless amplitude  $4\pi R_\Lambda |E_1|/\Lambda = 16.953$ , which via  $\Lambda/4\pi R_\Lambda = 0.874\text{MeV}$  translates to a potential with the magnitude  $|E_1| = 14.82\text{MeV}$ , at the  $R_{\text{peak}} = 8.245$  of Figure 5 which has been set to correspond to  $r_\Lambda = 2.178F$ . All of these correspondences are clearly shown in Figure 6.

Now, we went out of our way in the last two sections to also develop the probability field and the probability density which go along with Figure 6. So we should now show those. Following normalization the probability field is given by (17.28), namely,  $hP_0 = r/r_\Lambda = \Lambda r$ . Therefore  $g_s^{1/3} \mu_r = \partial_r (hP_0) = 1/r_\Lambda = \Lambda = .459F^{-1} = 1/2.178F$  is the probability density. So over the domain  $0 \leq r \leq r_\Lambda$ , this density graphs out to:



**Figure 7: Constant Radial Probability Density**

The total probability density integrated over the relevant domain  $0 \leq r \leq r_\Lambda$  is equal to 1, as was implemented in (17.25). Figure 7 simply contains an area  $.459F^{-1} \times 2.178F = 1$ . Now we integrate over  $r$  to arrive at the coupled probability field  $hP_0 = r/r_\Lambda + B = \Lambda r + B$ , where  $B$  is a constant of integration. When we began the present exercise in (6.7), we posited the linear form  $hP_0 = \Lambda r$  with  $B = 0$ . If we maintain this, then Figure 7 integrates to  $hP_0 = r/r_\Lambda = \Lambda r$ , which is the normalized (17.28). The integral of the probability density of Figure 7 into a dimensionless probability field, with what is now a slope of  $\Lambda$  is shown below in Figure 8.



**Figure 8: Dimensionless Probability Field for Constant Probability Density  $g_s^{\frac{1}{3}}\mu_r$  over  $0 \leq r \leq r_\Lambda$**

Because the probability density of Figure 7 is required to integrate out to 1 as we clearly see by the dimensionless area of 1 inside this density, once the integral is taken as in Figure 8, then beyond the upper extremity  $r_\Lambda$  of the non-zero probability density domain, the probability field is required to also be equal to 1. This is what fixes the constant of integration to  $B=0$ , which was the assumption that we sought to test starting at (6.7). Specifically, while the mathematics permits the plot in Figure 8 to be raised or lowered by a constant of integration  $B$ , the *physical interpretation of Figure 7 as a probability density* which must integrate to 1 requires us to discard the constant of integration so that for the domain beyond  $r_\Lambda$  the probability field carries forward this same 1 that is shown in Figure 7. Finally, it should be made very clear that Figures 6, 7 and 8 all describe exactly the same physics from three different views. Figure 6 shows the real magnitude  $|E_1|$  of the potential of (17.33), while Figure 7 shows its associated probability density and Figure 8 the associated probability field, all as a function of radius  $r$ . There is a one-to-one isomorphic mapping among Figures 6, 7 and 8. If any of these are changed, then the other two are changed as well.

Now, let us review what we may learn about analytical, non-linear quantum field theory from the results in this section. We began at (17.6) by positing a constant probability density  $\partial_r(hP_0) = \text{constant} \equiv A$  with  $A$  being real and non-zero and positive. But because  $\partial_r(hP_0)$  is a probability density, it operates under the very important constraint that its integral over  $r$  *must* be equal to unity,  $\int_0^\infty \partial_r(hP_0) dr = 1$ . So by positing that the probability density is a positive real constant, and by positing that *the probability density is a probability density*, we are *inherently and necessarily* positing that the probability density is constant *over a finite, bounded domain of the radial coordinate  $r$* . Why? Because the domain of  $r$  runs over  $0 \leq r \leq \infty$ . If the probability density  $\partial_r(hP_0)$  were to be constant over the entire domain from  $0 \leq r \leq \infty$ , then in order for it to integrate to  $\int_0^\infty \partial_r(hP_0) dr = 1$ , we would have to have  $A = \text{constant} = 0$ , and in that instance, we would have a zero probability density everywhere except at  $r = 0$ , which is the problem we reviewed in the last section for a  $1/r$  potential. To visualize this, just look at Figure 7 and suppose that we were to have  $\partial_r(hP_0) = \Lambda$  not over  $0 \leq r \leq r_\Lambda$ , but over the entirety of  $0 \leq r \leq \infty$ . What would happen? As we stretched  $r_\Lambda$  further to the right and had it approach  $r_\Lambda \rightarrow \infty$ , the value of  $\Lambda$  would diminish in order to maintain a total area of 1 within the Figure 7 rectangle. And at  $r_\Lambda = \infty$ , the constant probability density would necessarily become the constant *zero* probability density we examined in the last section, and no longer be a constant *positive* probability density. So, a positive, *non-zero* constant probability density *necessarily* implies a radially-bounded probability density.

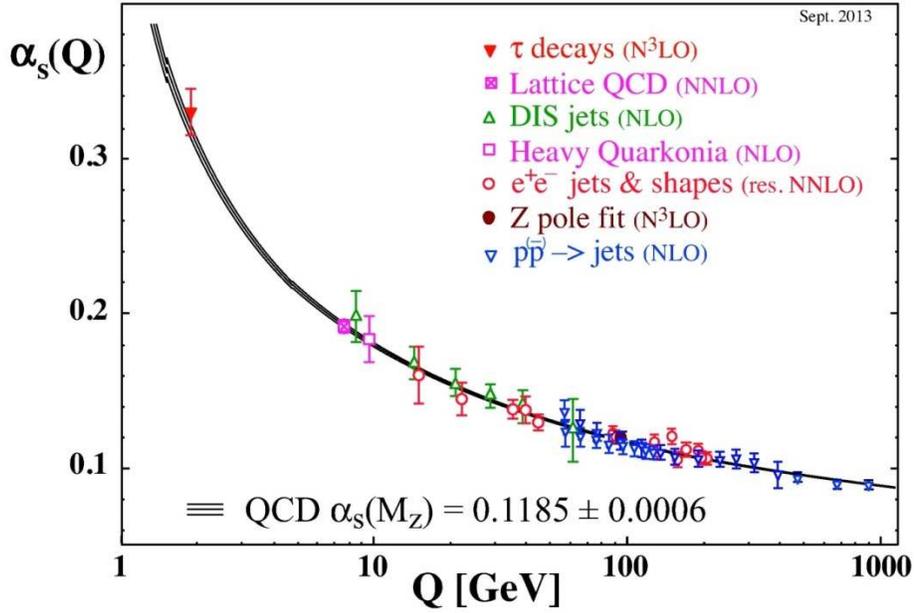
Now, it is one thing to posit a positive, constant radially-bounded coupled probability density, and quite another to assemble a *physical system* what has such a probability density. After all, at bottom, we are still doing physics, not just mathematics. What Figure 6 illustrates is that if one is going to have a physical system with a positive, constant radially-bounded coupled probability density, it will be necessary to assemble a suitable potential energy distribution to hold that probability distribution in place. Figure 6, and equation (17.33) which is derived from positing a constant, non-zero  $\partial_r(hP_0) = \Lambda$ , tells us what that potential must be. Specifically, in order to have  $\partial_r(hP_0) = \Lambda = \text{constant} > 0$  over the bounded domain  $0 \leq r \leq r_\Lambda$ , we are required, as seen in Figure 6, to have a potential well which “pulls” all of the physical fields constituting the probability density together, and creates a “least action” or “least potential” environment in the domain from about  $R = 1.75$  to  $R = 4$ . If the fields were to try to wander to a larger distance, say  $R = 6$  or  $R = 8$ , they would need to acquire extra energy to do so. If they were to try to compress themselves to a smaller distance, say  $R = .5$ , they would need to acquire extra energy to do so. By least action principles, the elements of a system will move toward and congregate near positions that can be maintained with a minimum of energy.

So the potential in Figure 6 provides the geodesic, least action environment which physically enables the probability density to remain constant and bounded in the manner of Figure 7. Because it is a stable, confining potential, the potential of Figure 6 “holds together both ends at the middle,” and tells us about the energetics of the *physical environment* which is required to support the posited  $\partial_r(hP_0) = \Lambda = \text{constant} > 0$ . So, while we can posit any

probability distribution we want, we cannot actually realize that probability distribution *in the physical universe* without creating the potential energy environment to assemble and maintain that probability distribution. Figure 6 which is equation (17.33) tells us the energies we need to maintain Figure 7. But more generally, (17.1) in Cartesian coordinates a.k.a. (17.2) in spherical coordinates tells us the energies required to maintain *whatever probability distribution we may wish to posit*, with whatever radial and angular distributions we may wish to posit, because for any posited probability distribution, there is an associated quantum potential which can be deduced from these equations. And these two equations, (17.1) and (17.2) contain non-abelian probability fields  $\mathcal{P}_0$  and so provide a complete set of tools to do the exact same development we did here, for, e.g., the  $SU(3)_c$  group of QCD, and thus to understand with precision, to first recursive order, the dynamics within a hadron that cause quarks and gluons to be confined and also subsist in a stable system, i.e., to live in a space outside of which their probability densities are zero and to not collapse together in the nature of the ultraviolet catastrophe and atomic spiraling that so-plagued physicists at the opening of the 20<sup>th</sup> century. As stated at the end of the last section, in non-linear quantum field theory, probability densities and potential energies go hand in hand, with a one-to-one isomorphic mapping between them.

## 18. Asymptotic Freedom and Asymptotic Confinement: Fitting and Extending the QCD Running Coupling Curve

While the probability density of Figure 7 and its isomorphically-related potential energy curve of Figure 6 clearly show features of confinement and stability, we have not yet discussed the third critical aspect of QCD, which is asymptotic freedom. As is well known, the running strong coupling  $\alpha_s = g_s^2 / 4\pi$  becomes very large and indeed asymptotically approaches infinity for small probe energies below about 1GeV, and becomes relatively flat (asymptotically free) for large probe energies in the deep TeV area and beyond. Via the deBroglie relation  $E = \hbar c / \lambda$ , this is inverted when talking about length rather than energy as we are doing in Figures 6 through 8: the running strong coupling  $\alpha_s = g_s^2 / 4\pi$  becomes very large and tends toward infinity while approaching the larger radial lengths  $r \rightarrow r_\Lambda$  from smaller lengths, and flattens out approaching very short length scales  $r \rightarrow 0$ . To provide a common point of reference, we reproduce below, Figure 9.4 from PDG's [24] which illustrates all of this based on a range of the most-current empirical data, and make note that we have based Figures 6 through 8 on  $\Lambda_{QCD}^{(6)} = (90.6 \pm 3.4) MeV$  for six quarks and thus used the length equivalent  $r_\Lambda = 2.178F$  to bound the probability distribution. Which is to say, we have worked from the view that  $\alpha_s = g_s^2 / 4\pi$  asymptotically approaches infinity at around  $90.6 MeV$ , which is about one order of magnitude to the left where  $Q = 1 GeV$  is shown on Figure 9 below. The prevailing view is also that the curve in Figure 9 will tend to asymptotically flatten moving to the right beyond the  $1000 GeV$  shown below to higher and higher energies.



**Figure 9: The Running Strong Coupling  $\alpha_s(Q = \hbar c / r)$  (reproduced from PDG's [24], Figure 9.4)**

So, the question now is, precisely how does the potential of Figure 6 and the bounded constant probability density of Figure 7 connect with the  $\alpha_s(Q = \hbar c / r)$  curve of Figure 9? Here, it is important to keep in mind that Figures 6 through 8 as well as much of the development here is expressed in terms of a *coupled* proper probability density  $g_s^{\frac{1}{3}}\mu_{0r} = \partial_r(hP_0)$ , and more generally  $g_s^{\frac{1}{3}}\mu_0 = \nabla(hP_0)$ . That is, we have a *bare* probability field  $P_0$  and a *bare* probability density  $\mu_0$  which then couples through a dimensionless running charge  $g_s^{\frac{1}{3}} = h$ , see (15.43), which is related to the strong coupling by  $\alpha_s = g_s^2 / 4\pi\hbar c$  (or generally, to any given interaction coupling  $g$  via  $\alpha = g^2 / 4\pi\hbar c$ ). So, for the constant probability density we have developed here, the interrelationships, rather simply, are:

$$g_s^{\frac{1}{3}}\mu_{0r} = \partial_r(hP_0) = \Lambda, \quad (18.1)$$

or alternatively:

$$\mu_{0r} = \Lambda g_s^{-\frac{1}{3}} = \Lambda (4\pi\hbar c \alpha_s)^{-\frac{1}{6}} = \partial_r P_0 + \frac{1}{3} \frac{\partial_r g_s}{g_s} P_0, \quad (18.2)$$

where in the final expression, we have also applied the differential equation (15.45) to (18.1).

In other words, deepening the earlier discussion from (14.44), the *bare* radial probability density  $\mu_{0r}$  will run in inverse proportion to the sixth root of the  $\alpha_s$  illustrated in Figure 9 above, because  $g_s^{\frac{1}{3}} = (4\pi\alpha_s)^{\frac{1}{6}}$ , and by (18.1),  $\mu_{0r} = \Lambda g_s^{-\frac{1}{3}} = \Lambda (4\pi\alpha_s)^{\frac{1}{6}}$ . Where  $\alpha_s \rightarrow \infty$  asymptotically in the neighborhood of  $r_\Lambda = 2.178F$ , the bare probability density will asymptotically tend toward zero,  $\mu_{0r} \rightarrow 0$ . This means that the potential of Figure 6 effectively causes any field quantum to have a near-zero probability of situating near  $r_\Lambda = 2.178F$ . But, if any field quantum should happen to situate near  $r_\Lambda = 2.178F$ , it will be very-highly-coupled so as to maintain the *coupled* probability density at the constant  $\Lambda$  per (18.1). This also means that the *bare* probability of where one might find a field quantum is maximized at  $r=0$ , but that at  $r=0$  the coupled density still remains asymptotically constant because  $\alpha_s$  approaches its asymptotically-flat minimum value. (We shall for now ignore any GUT effects that might come into play near  $10^{15} GeV$  probe energies and especially any effects that may arise at the scale  $Gm_p^2 = \hbar c$  of the Planck mass  $m_p$ , and will return to consider possible GUT effects at the very end of this discussion.) In order to obtain the differential equation which precisely governs the behavior of the bare probability field  $P_0$  as a function of  $r$  and  $g_s$ , we rewrite (18.2) as:

$$0 = \frac{\partial P_0}{\partial r} + \frac{1}{3} \frac{1}{g_s} \frac{\partial g_s}{\partial r} P_0 - \Lambda g_s^{-\frac{1}{3}}. \quad (18.3)$$

The bare  $P_0$  obtained through this equation is the bare counterpart to the coupled  $hP_0$  illustrated in Figure 8.

Of extremely high importance, all of the forgoing provides us with the tools we need to actually fit the running QCD curve of Figure 9 very precisely to some very simple mathematical functions by focusing on the bare probability density  $\mu_{0r}$  given in (18.1) by  $g_s^{\frac{1}{3}}\mu_{0r} = \Lambda$ . We make use of two well-established theoretical premises to do this, while at the outset neglecting any GUT effects: First, based on asymptotic freedom, we assume that approaching  $r \rightarrow 0$  from  $r > 0$  the curve in Figure 9 flattens completely, i.e., that  $\partial\alpha_s / \partial Q \rightarrow 0$  as  $r \rightarrow 0$ . Second, we assume when approaching  $r \rightarrow r_\Lambda = \hbar c / \Lambda$  from  $r < r_\Lambda$  that the slope of  $\alpha_s$  asymptotically becomes infinite,  $\partial\alpha_s / \partial r \rightarrow \infty$ . Simply put: we accept and utilize the commonly-held twin premises a) that asymptotic freedom is really asymptotic along the  $r$  or  $Q$  axis for  $r \rightarrow 0$  and  $Q \rightarrow \infty$ , and b) that confinement really is asymptotic along the vertical  $\alpha_s$  axis as one approaches  $r \rightarrow r_\Lambda = \hbar c / \Lambda$  from smaller  $r$ , i.e., as  $Q \rightarrow \Lambda$  from higher  $Q$  when moving to the left of the domain in Figure 9. Via  $g_s^{\frac{1}{3}}\mu_{0r} = \Lambda$ , these twin premises tell us that  $\mu_{0r}$  must be flat for  $r \rightarrow 0$  and that  $\mu_{0r}$  must have a slope of negative infinity for  $r \rightarrow r_\Lambda$ . Mathematically, this narrows the scope of plausible  $\mu_{0r}(r)$  because  $\mu_{0r}(r)$  can only be a function for which  $\partial\mu_{0r} / \partial r = 0$  as  $r \rightarrow 0$  and  $\partial\mu_{0r} / \partial r \rightarrow -\infty$  as  $r \rightarrow r_\Lambda$  from smaller  $r$ . So now the question is well framed: what set of mathematical functions have these required properties to simultaneously

facilitate asymptotic freedom and asymptotic confinement? As it happens, the most basic mathematical function which fits these two requirements is a simple ellipse. And the simplest ellipse is a circle. So let us now see if we can find a way to fit the running coupling in Figure 9 to what is effectively a mathematical ellipse or even a circle.

First, we go to (18.1), but because we want to find  $\alpha_s$  and (18.1) only contains the sixth root of  $\alpha_s$ , let us raise everything to the sixth power. Thus we may write  $g_s^2 \mu_{0r}^6 = \Lambda^6$ , or, with  $4\pi\alpha_s = g_s^2 / \hbar c$ , in natural units:

$$\alpha_s \mu_{0r}^6 = \Lambda^6 / 4\pi. \quad (18.4)$$

Now we are dealing with  $\mu_{0r}^6$  rather than just  $\mu_{0r}$ , but this too should have the same requirements as  $\mu_{0r}$ : a flat slope as  $r \rightarrow 0$  and a slope of negative infinity as  $r \rightarrow r_\Lambda$  from smaller  $r$ . So let us now set up the ellipse for  $\mu_{0r}^6$ . And to keep things very simple, let us use the simplest ellipse of all, a circle. So, for a circle of constant fixed radius  $R_0$  we have the familiar  $x^2 + y^2 = R_0^2$ , or  $y = \sqrt{R_0^2 - x^2}$ . Let us now make the proportionality associations  $\Lambda^6 / 4\pi \propto R_0$ ,  $\mu_{0r}^6 \propto y$  and  $r \propto x$ . And to simplify further, let us set  $R_0 = 1$ . So we first write the very simple, Pythagorean relationship for a right triangle with a hypotenuse of 1 which defines the unit circle:

$$y = \sqrt{1 - x^2}. \quad (18.5)$$

We shall now fit this to the QCD curve in Figure 9.

By (18.4)  $\alpha_s \propto 1 / \mu_{0r}^6 \propto 1 / y$ . So the relationship that now becomes of interest based on (18.5), including  $x \rightarrow r \rightarrow r / r_\Lambda = R$ , is:

$$\alpha_s = \frac{1}{\sqrt{1 - (r/r_\Lambda)^2}} = \frac{1}{\sqrt{1 - R^2}}. \quad (18.6)$$

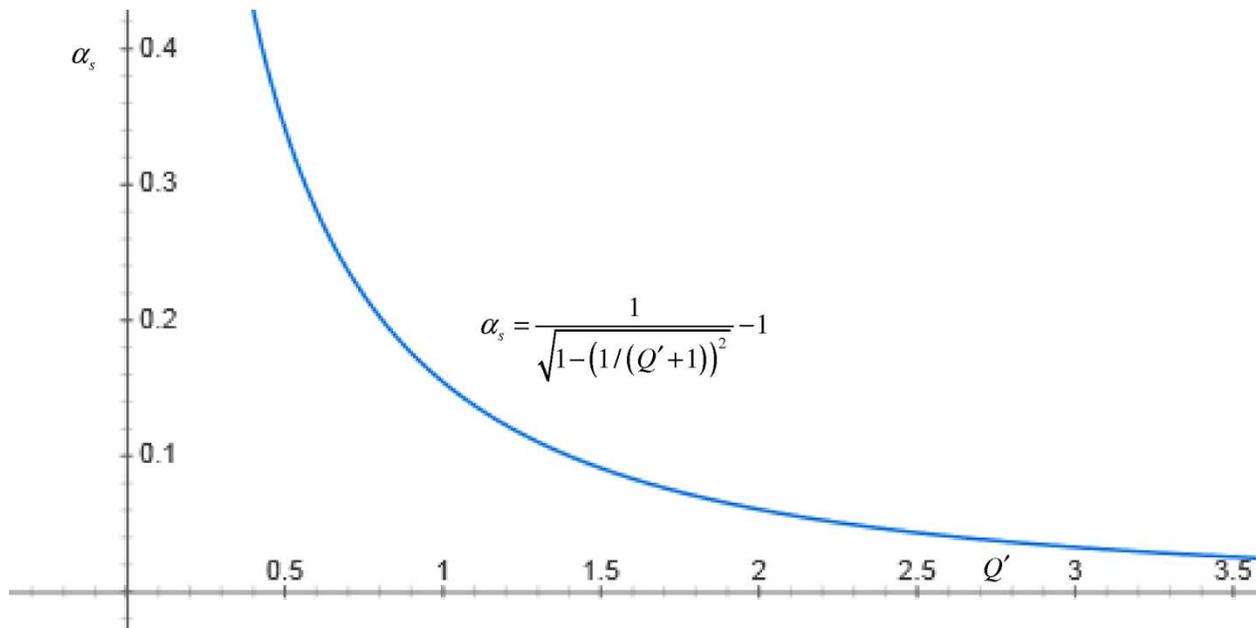
We note that in the form  $\gamma = 1 / \sqrt{1 - (v/c)^2}$ , this exact same mathematical function – the inverted circle – sets another asymptotic limit in the natural material world, namely that of the speed of light. Indeed,  $r_\Lambda$  above takes on the *exact same mathematical character* as  $r \rightarrow r_\Lambda$  as does the speed of light  $c$  as  $v \rightarrow c$ . This is how we set up the asymptotic limit for confinement. By deBroglie, we can also substitute  $r = \hbar c / Q'$  where  $Q'$  has a mass dimension of +1. The reason for  $Q'$  rather than  $Q$  will become momentarily apparent. So now (18.6) becomes:

$$\alpha_s = \frac{1}{\sqrt{1-(1/Q')^2}}. \quad (18.7)$$

This function (18.7) will have two asymptotes: a vertical asymptote at  $Q'=1$  as  $Q' \rightarrow 1$  from right to left, and a horizontal asymptote approaching  $\alpha_s = 1$  from above as  $Q' \rightarrow \infty$ . So as a final step to move the asymptotes onto the vertical and horizontal axes at  $Q'=0$  and  $\alpha_s = 0$ , we simply displace this curve one unit to the left and one unit down, by rewriting (18.7) as:

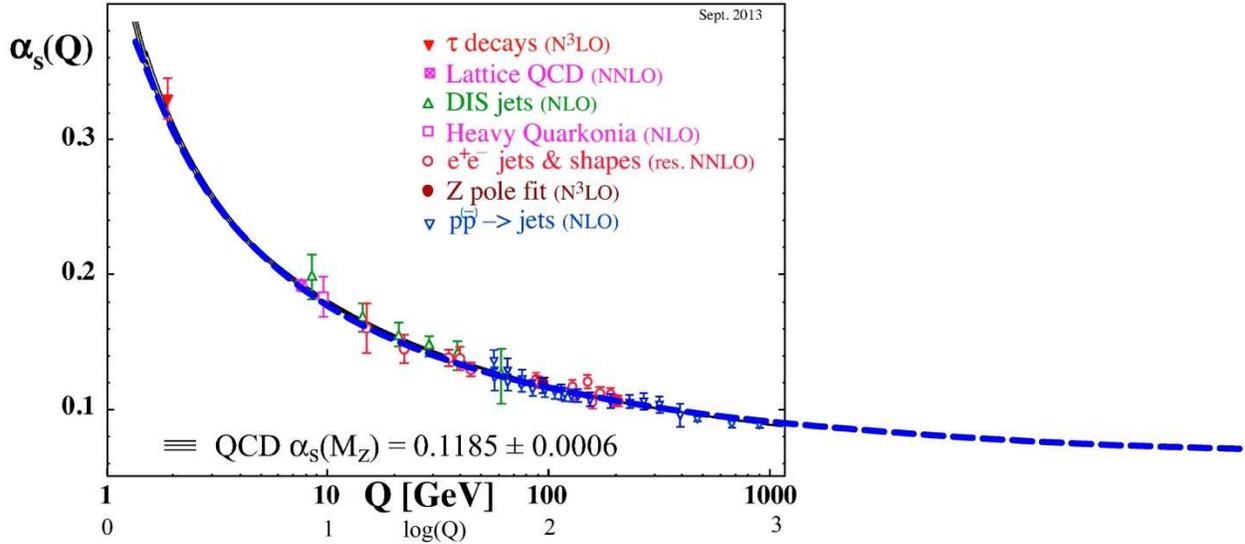
$$\alpha_s = \frac{1}{\sqrt{1-(1/(Q'+1))^2}} - 1. \quad (18.8)$$

Now, let graph (18.8) and compare it to Figure 9:



**Figure 10: The Running Strong Coupling  $\alpha_s$  Modelled from (18.8) based on the Unit Circle**

This looks very much like it has the same form as the empirical PDG data curve in Figure 9. And in fact it would be the exact same curve should it be possible to match up the height of Figure 10 with Figure 9 and then rescale Figure 10 along the horizontal axis to get the shapes of the curves to match. So, let us do just that. We match up the heights, and then stretch the curve of Figure 10 along the  $Q'$  axis to the degree required to match Figure 9. Then we move the stretched Figure 10 curve left/right and up/down as needed to superimpose it data point-to-data point over Figure 9. What we obtain is Figure 11 below, with the curve of Figure 10 in the blue dashed line superimposed over the empirical PDG curve of Figure 9:



**Figure 11: Superimposition of Figure 10 Theoretical Curve over Figure 9 Empirical Curve for  $\alpha_s$**

The possible match is clear, and we also now see clearly that  $Q'$  bears a logarithmic relation to  $Q$ , that is,  $Q' = 1/r \propto \log Q = \ln Q / \ln 10$ . If we write this simply as  $Q' = f \ln Q$  where  $f$  is a data “fitting” constant to stretch the horizontal axis on move it onto a logarithmic rather than linear scale, if we then replace  $(Q' + 1) \Rightarrow (f \ln Q + A)$  in (18.8) to permit horizontal movement of the whole curve via an unknown, to-be-fitted constant  $A$ , and if we finally replace  $1 \rightarrow B$  at the end of (18.8) to permit vertical movement of the whole curve by an also unknown and to-be-fitted constant  $B$ , then (18.8) now becomes:

$$\alpha_s(Q) = \frac{1}{\sqrt{1 - \left(\frac{1}{f \ln Q + A}\right)^2}} - B. \quad (18.9)$$

As we shall now show, with suitable choice of the three fitting parameters  $f$ ,  $A$  and  $B$ , (18.9) can be fitted to an exact match with the empirical running PDG data of Figure 9 for  $\alpha_s(Q)$  *within all the indicated experimental error-bars*. What is especially significant about the fit in Figure 11 beyond the fact that it does fit, is that this curve provides definitive predictions as to how the strong coupling will continue to run in the region above 1TeV, and based just on what is seen in Figure 11 above, the curve is now extended out to somewhat over 100TeV. In fact, we shall return to this point shortly, because now that (18.9) can indeed be fitted to Figure 9 over the entirety of the domain and range of Figure 9, it should be possible to use (18.9) to extend Figure 9 way beyond the domain and range of the data that it contains.

Shortly, we shall indeed use this to extrapolate Figure 9 to both higher and lower  $Q$ . But first let us backtrack to our starting point and fully develop the logarithmic fitting of the radial

probability density. By (18.4)  $\mu_{0r}^6 = \Lambda^6 / 4\pi\alpha_s \propto 1/\alpha_s$ . So without the horizontal or vertical shifting, we combine (18.4) and (18.6) together and use  $r = 1/Q' = 1/f \ln Q$  to write:

$$\mu_{0r}^6 = \frac{\Lambda^6}{4\pi\alpha_s} = \frac{\Lambda^6}{4\pi} \sqrt{1-R^2} = \frac{\Lambda^6}{4\pi} \sqrt{1 - \left(\frac{r}{r_\Lambda}\right)^2} = \frac{\Lambda^6}{4\pi} \sqrt{1 - \left(\frac{1}{f \ln Q}\right)^2}, \quad (18.10)$$

Now, taking the sixth root and also including  $r = 1/Q' = 1/f \ln Q$ , we arrive at the following for the bare radial probability density  $\mu_{0r} = \partial_r P_0$ :

$$\frac{(4\pi)^{\frac{1}{6}}}{\Lambda} \partial_r P_0 = (4\pi)^{\frac{1}{6}} \frac{\mu_{0r}}{\Lambda} = \sqrt{1-R^2}^{\frac{1}{6}} = \sqrt{1 - \left(\frac{r}{r_\Lambda}\right)^2}^{\frac{1}{6}} = \sqrt{1 - \left(\frac{1}{fr_\Lambda \ln Q}\right)^2}^{\frac{1}{6}}, \quad (18.11)$$

Next, since  $Q$  has dimensions of energy, we set  $\ln Q \rightarrow \ln(Q/Q_0)$  where  $Q_0$  is some suitably chosen scale against which to measure  $Q$ , so that the quantity inside the natural log is dimensionless. But, of course,  $\ln(Q/Q_0) = \ln Q - \ln Q_0$ . So, also defining  $f' \equiv fr_\Lambda$  and  $A' \equiv -f' \ln Q_0$ , we can extend (18.11) to:

$$\begin{aligned} \frac{(4\pi)^{\frac{1}{6}}}{\Lambda} \partial_r P_0 &= (4\pi)^{\frac{1}{6}} \frac{\mu_{0r}}{\Lambda} = \sqrt{1-R^2}^{\frac{1}{6}} = \sqrt{1 - \left(\frac{r}{r_\Lambda}\right)^2}^{\frac{1}{6}} = \sqrt{1 - \left(\frac{1}{fr_\Lambda \ln Q - fr_\Lambda \ln Q_0}\right)^2}^{\frac{1}{6}}, \\ &= \sqrt{1 - \left(\frac{1}{f'r_\Lambda \ln Q - f'r_\Lambda \ln Q_0}\right)^2}^{\frac{1}{6}} = \sqrt{1 - \left(\frac{1}{f' \ln Q + A'}\right)^2}^{\frac{1}{6}}, \end{aligned} \quad (18.12)$$

This square root now is of the exact same form as the square root in (18.9) for  $\alpha_s(Q)$ , except that it has  $f' \equiv fr_\Lambda$  and  $A' \equiv -f' \ln Q_0$ . But the  $f$  and  $A$  in (18.9) were simply unknown data-fixing constants. So we can replace  $f \rightarrow f'$  and  $A \rightarrow A'$  in (18.9) to write:

$$\alpha_s(Q) = \frac{1}{\sqrt{1 - \left(\frac{1}{f' \ln Q + A'}\right)^2}} - B = \frac{1}{\sqrt{1 - \left(\frac{1}{f'r_\Lambda \ln Q - f'r_\Lambda \ln Q_0}\right)^2}} - B. \quad (18.13)$$

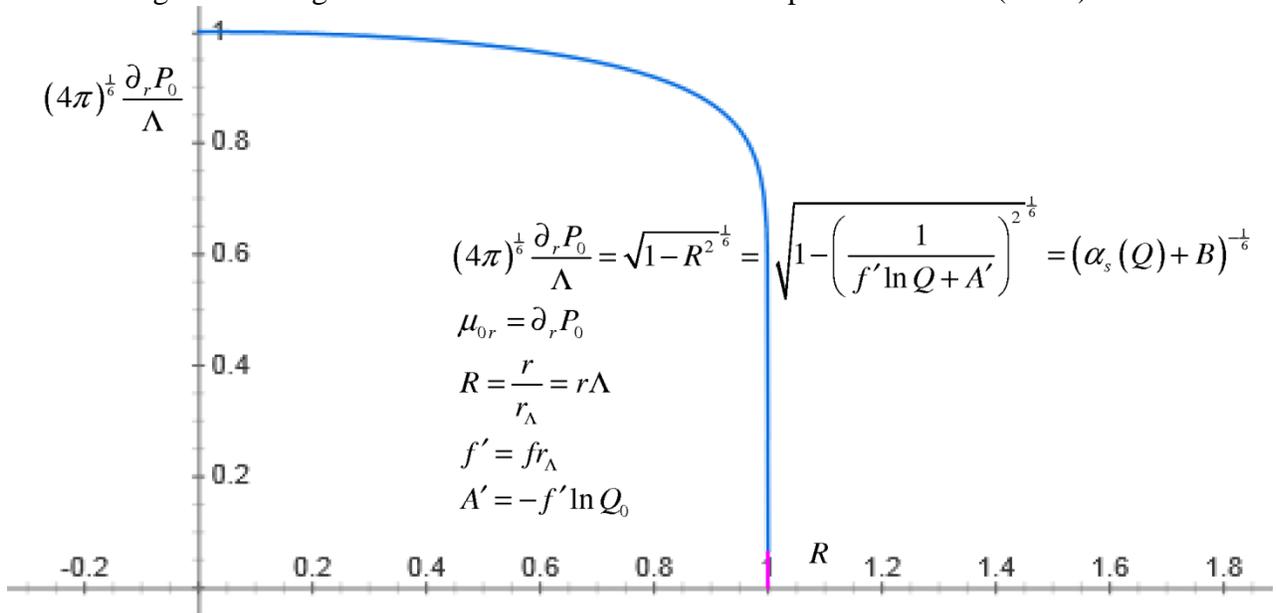
Then, we restructure (18.13) into:

$$\frac{1}{\alpha_s(Q) + B} = \sqrt{1 - \left(\frac{1}{f'r_\Lambda (\ln Q - \ln Q_0)}\right)^2}, \quad (18.14)$$

which means finally, that the multiple alternative expressions of (18.12) can now be supplemented to directly include  $\alpha_s(Q)$ , as follows:

$$\begin{aligned} \frac{(4\pi)^{\frac{1}{6}}}{\Lambda} \partial_r P_0 &= (4\pi)^{\frac{1}{6}} \frac{\mu_{0r}}{\Lambda} = \sqrt{1-R^2}^{\frac{1}{6}} = \sqrt{1-\left(\frac{r}{r_\Lambda}\right)^2}^{\frac{1}{6}} = \sqrt{1-\left(\frac{1}{fr_\Lambda \ln Q - fr_\Lambda \ln Q_0}\right)^2}^{\frac{1}{6}}, \\ &= \sqrt{1-\left(\frac{1}{fr_\Lambda \ln Q - fr_\Lambda \ln Q_0}\right)^2}^{\frac{1}{6}} = \sqrt{1-\left(\frac{1}{f' \ln Q + A'}\right)^2}^{\frac{1}{6}} = (\alpha_s(Q) + B)^{-\frac{1}{6}} \end{aligned} \quad (18.15)$$

The simplest way to graph this is with  $(4\pi)^{\frac{1}{6}} \partial_r P_0 / \Lambda = \sqrt{1-R^2}^{\frac{1}{6}}$ . This is illustrated below in Figure 12 along with all of the other interrelationships embedded in (18.15):



**Figure 12: The Bare Probability Density Fitted to the Empirical  $\alpha_s(Q = \hbar c / r)$  Data of Figure 9**

Here, we see the running coupling  $\alpha_s(Q = \hbar c / r)$  represented in terms of the bare radial probability density  $\mu_{0r} = \partial_r P_0$ , such that  $\mu_{0r} = \Lambda(4\pi\alpha_s + 4\pi B)^{-\frac{1}{6}}$ , or via  $4\pi\alpha_s = g_s^2$  and  $\partial_r(hP_0) = \Lambda$  from (17.28) and some rearrangement:

$$\partial_r(hP_0) = \mu_{0r} (g_s^2 + 4\pi B)^{\frac{1}{6}} = \Lambda. \quad (18.16)$$

The constant coupled probability density function of Figure 7 for  $\partial_r(hP_0) = \Lambda$  looks the same as before, but when expressed in terms of  $\mu_r$  it also acquires the fitting term  $4\pi B$  used simply to set the height of  $\alpha_s$  in Figures 10 and 11.

The final matter to point out, of course, is that because  $\partial_r P_0$  is a probability density, it must normalize to 1 when integrated over its entire domain  $0 \leq r \leq r_\Lambda$  a.k.a.  $0 \leq R \leq 1$ , that is, we must have  $\int_0^{r_\Lambda} \partial_r P_0 dr = 1$  a.k.a.  $\int_0^1 \partial_R P_0 dR = 1$  in Figure 12. It can be found via numerical calculation by computer that the definite integral:

$$\int_0^1 \sqrt{1-R^2}^{\frac{1}{6}} dR = 0.952354, \quad (18.17)$$

which makes visual sense when noting that Figure 12 basically contains a unit square with a shaved corner on the upper right. So if we go back to the  $R$ -based expressions in (18.15) and use  $\partial_r = \partial_R / r_\Lambda$  and  $\Lambda r_\Lambda = 1$  and integrate over the domain of  $0 \leq R \leq 1$  we obtain:

$$\int_0^1 \partial_R P_0 dR = \int_0^1 \mu_{0R} dR = \frac{1}{(4\pi)^{\frac{1}{6}}} \int_0^1 \sqrt{1-R^2}^{\frac{1}{6}} dR = \frac{0.952354}{(4\pi)^{\frac{1}{6}}} = .624592 = \frac{1}{1.60104} \equiv \frac{1}{N}, \quad (18.18)$$

which defines a normalization constant  $N = 1.60104 = (4\pi)^{\frac{1}{6}} / 0.952354$ . Then, we normalize by multiplying everything in (18.15) beyond the first two terms by  $N$  to ensure that  $\int_0^1 \partial_R P_0 dR = 1$ . We then factor out the  $(4\pi)^{\frac{1}{6}}$  which then appears in front of every term, and again use  $\partial_r = \partial_R / r_\Lambda$  and  $\Lambda r_\Lambda = 1$ , with the net result:

$$\begin{aligned} \partial_R P_0 = \mu_{0r} &= \frac{\sqrt{1-R^2}^{\frac{1}{6}}}{0.952354} = \frac{1}{0.952354} \sqrt{1 - \left(\frac{r}{r_\Lambda}\right)^2}^{\frac{1}{6}} = \frac{1}{0.952354} \sqrt{1 - \left(\frac{1}{fr_\Lambda \ln Q - fr_\Lambda \ln Q_0}\right)^2}^{\frac{1}{6}} \\ &= \frac{1}{0.952354} \sqrt{1 - \left(\frac{1}{fr_\Lambda \ln Q - fr_\Lambda \ln Q_0}\right)^2}^{\frac{1}{6}} = \frac{1}{0.952354} \sqrt{1 - \left(\frac{1}{f' \ln Q + A'}\right)^2}^{\frac{1}{6}} = \frac{(\alpha_s(Q) + B)^{\frac{1}{6}}}{0.952354} \end{aligned} \quad (18.19)$$

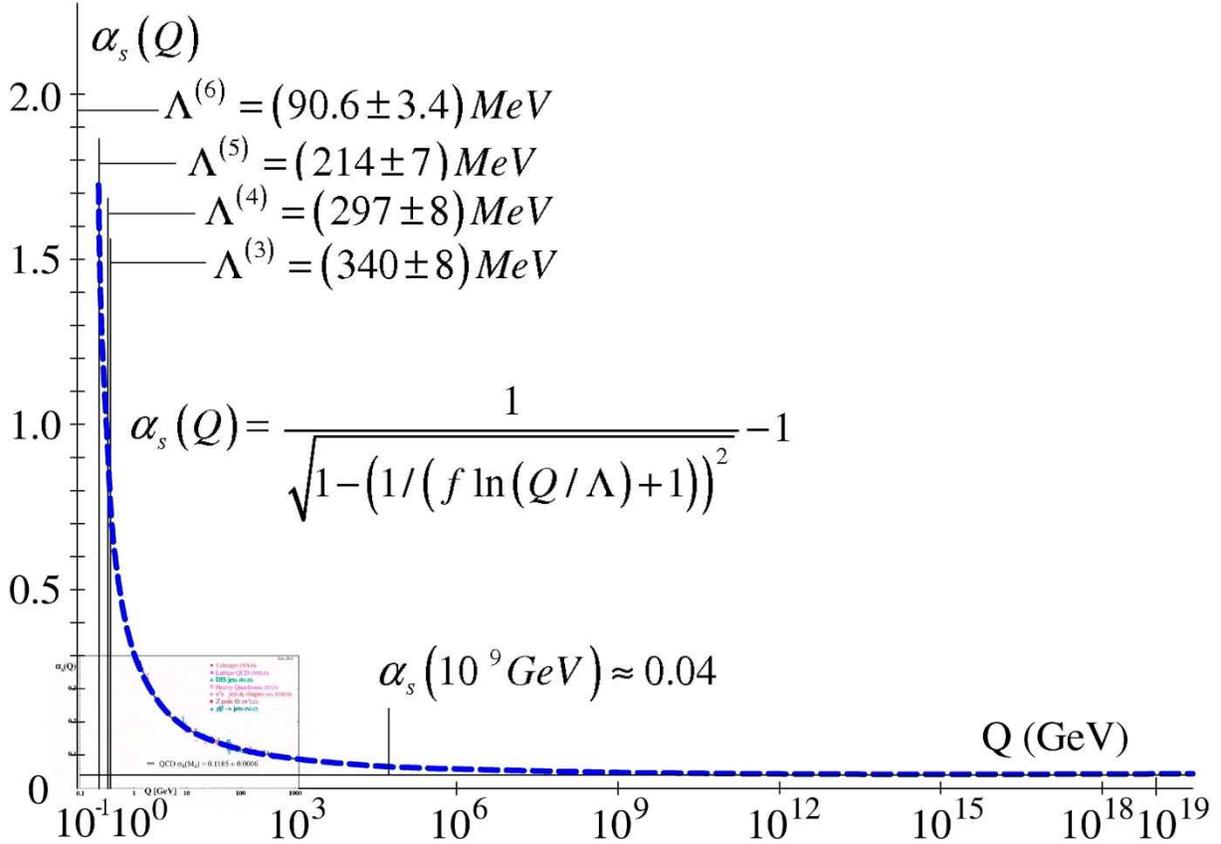
When established in this way, Figure 12, subject to precise empirical determination of the various fitting parameters, becomes an alternative way to express the phenomenological running coupling curve  $\alpha_s(Q)$  of Figure 9 in terms of (the sixth root of) a circle, which is one of the simplest objects in the mathematical world. Physically, this centers around the expression  $\sqrt{1-(r/r_\Lambda)^2}$ . Contrasting with  $\gamma = 1/\sqrt{1-(v/c)^2}$  which set sets the speed of light  $c$  as a natural limit in the material world, we see that (18.19) places the QCD cutoff  $r_\Lambda = 1/\Lambda$  into a precisely analogous role as the speed of light, as a natural, material limitation. And, of course,

if we write the mathematical crux of (18.19) as  $\left(\left(0.952354\mu_{0,r}\right)^6\right)^2 + R^2 = 1$ , we see once again – as is often the case – that the Pythagorean theorem sits in yet another guise at the root of the inner workings of the material universe.

Finally, as we discussed following (18.9), because of the clear fit in Figure 11, it should be possible to graph the curve of Figure 10 which is equation (18.8) over a greatly-extended domain and range, and then fit a limited piece of this curve from about 1 GeV to 1000 GeV and from  $a_s \approx 0.4$  to  $a_s \approx 0.1$  to the PDG curve of Figure 9. The theoretical curve we shall graph is (18.8) with  $Q' = f \ln(Q/\Lambda)$  and with  $f' \rightarrow f$ ,  $Q \rightarrow Q/\Lambda$ ,  $A' \rightarrow 1$  and  $B \rightarrow 1$  in (18.13), namely:

$$\alpha_s(Q) = \frac{1}{\sqrt{1 - (1/(Q'+1))^2}} - 1 = \frac{1}{\sqrt{1 - (1/(f \ln(Q/\Lambda) + 1))^2}} - 1. \quad (18.20)$$

Specifically, in Figure 13 below, we extend the PDG curve of Figure 9 one order of magnitude to the left, i.e. over to  $10^{-1} GeV = 100 MeV$ , and we also extend it downward to include the  $\alpha_s = 0$  axis. We then use the fitting parameter  $f$  to fit this to the PDG curve of Figure 9 by stretching the curve horizontally as needed to create a tight fit over about 1 GeV to 1000 GeV and from  $a_s \approx 0.4$  to  $a_s \approx 0.1$ . We also bring to bear additional data from [9.24a] through [9.24d] of PDG's [24] for the QCD cutoffs  $\Lambda^{(n_f)}$  for  $n_f = 6, 5, 4, 3$  quark flavors, by showing what would be the confinement asymptotes for each of these  $\Lambda^{(n_f)}$ . This is all illustrated in Figure 13 below:



**Figure 13: Fitting of the PDG QCD Curve of Figure 9 to Extended Domain and Range for  $\alpha_s(Q = \hbar c / r)$ , based on the Theoretical Function (18.20)**

Now let's discuss Figure 13 in some depth. We see the PDG data from Figure 9.4 of [24] which is reproduced in Figure 9 here, occupying a very small corner at the lower-left of Figure 13. When magnified, this corner has the same fit between the empirical PDG data and the mathematical curve of (18.20) as is illustrated in Figure 11. But of course, this curve extends that data well beyond the domain or range of Figures 9 or 11, so that the PDG curve fits to a small yet important corner of the theoretical curve of (18.20). Of special interest is the extrapolated curve toward the vertical axes on the left. It is very desirable to know the magnitude of the  $Q$  at which the vertical confinement asymptote reaches infinity. If the PDG curve of Figure 9 is accurate as we must assume it is, and if the  $\Lambda^{(n_f)}$  for  $n_f = 6, 5, 4, 3$  in [9.24a] through [9.24d] of PDG's [24] are accurate as we also must assume they are, then these are not two disconnected pieces of empirical data. They are interrelated such that one of the  $\Lambda^{(n_f)}$  in [9.24a] through [9.24d] of [24] *must* be the asymptote of Figure 9.4 of [24]. And this in turn should confirm from yet another view, the number of quark flavors that exist in nature. What we see from the extrapolation in Figure 13 is that the PDG data, when extended with (18.20), moves definitively to the left of the  $n_f = 3, 4$  asymptote candidates, so that we must have either 5 or 6 quark flavors. But, this curve even moves to the left past the  $n_f = 5$  asymptote candidate right around  $\alpha_s \approx 1.5$ . Since there is still a long way to go from  $\alpha_s \approx 1.5$  to  $\alpha_s = \infty$ , it look highly

plausible that in fact it is  $\Lambda^{(6)} = (90.6 \pm 3.4) MeV$  which is the asymptote for this curve. And this, of course, fits perfectly with our in fact observing six quark flavors in nature. So when we in fact engage in the data fitting represented in Figure 6, the first step is to set  $\Lambda = \Lambda^{(6)}$  in (18.20). Then, at  $Q = \Lambda = \Lambda^{(6)}$ , we will have  $\alpha_s(\Lambda^{(6)}) = \infty$ . From there, we *choose* the fitting parameter  $f$  such that the remainder of the curve passes properly through the empirical PDG data of Figure 9.

Having set the vertical asymptote to  $\Lambda = \Lambda^{(6)}$  and having set  $f$  to yield the proper fit so as to pass through the PDF data in Figure 13 by graphically compressing or expanding the horizontal aspect of the (18.20) curve as needed (our next task will be to numerically determine  $f$ ), all that remains is to explore the asymptotic freedom region. The use of  $B=1$  in (18.20) for the more general height-fitting parameter  $B$  of (18.9) ensures that  $\alpha_s \rightarrow 0$  as  $Q \rightarrow \infty$ . So now the question is whether this also accords with empirical data. Left to its own devices, the curve which uses  $B=1$  will tend to  $\alpha_s \rightarrow 0$  as  $Q \rightarrow \infty$ . But this is just mathematical. We know that physics intervenes, because at around  $Q \approx 10^{15} GeV$  we expect a GUT to bring together the running coupling of the strong, weak and electromagnetic interactions, and we expect that near the Planck energy, these will all meet up as well with the gravitational coupling and will reverse course and start to increase in magnitude. So, physically, we do not expect to ever reach  $\alpha_s = 0$ , because other things will intervene by then, including the exceedingly high-energy fluctuations of quantum gravitation in the geometrodynamical vacuum. This means that must expect Figure 15 to lose its predictive ability as to  $\alpha_s$ , possibly near  $Q \approx 10^{15} GeV$ , and definitely near  $Q \approx 1.22 \times 10^{19} GeV$ .

So the way to determine if we are correct to use  $B=1$  as the vertical fitting parameter in (18.20), is not to expect  $\alpha_s$  to ever become zero, but to study the empirical behavior of  $\alpha_s$  in whatever energy domains become experimentally-accessible at TeV energies and higher. When studied more closely, Figure 13 reveals that at  $Q=1TeV$ ,  $\alpha_s$  is slightly less than  $\alpha_s \approx 0.08$ . Close study also reveals that at  $Q=1PeV = 10^6 GeV$ , we have  $\alpha_s \approx 0.035$ , and that at  $Q=1EeV = 10^9 GeV$ , we have  $\alpha_s \approx 0.020$ . So by mapping out the running of  $\alpha_s$  at deeper and deeper probe energies, one can confirm whether the fit using  $B=1$  is the correct fit or whether  $B < 1$  by some very tiny amount. In the event that  $B=1$  is confirmed to be correct, and given that setting  $\Lambda = \Lambda^{(6)}$  for the vertical confinement asymptote obviates the need for any horizontal fitting via the parameter  $A'$  in (18.9) which is set to  $A'=1$  in (18.20), we have by the analysis and fitting of Figure 3 already determined that  $A'=B=1$  insofar as horizontal and vertical fitting is concerned and as seen in (18.20) and how it graphs out in Figure 13. Any fitting with  $A'$  other than 1 is absorbed into how we define  $f' \rightarrow f$ . So the fitting parameter  $f$  in (18.20) is the only parameter which remains to be empirically fitted via empirical data. This  $f$  truly is an empirical parameter, and this is the parameter which causes the curve of Figure 13 to compress or expand (scale) as a logarithmic function of  $Q$ . Only the empirical data can determine the proper numerical value of  $f$ , and this now becomes a key number to determine with experimental data. So let us now do just that.

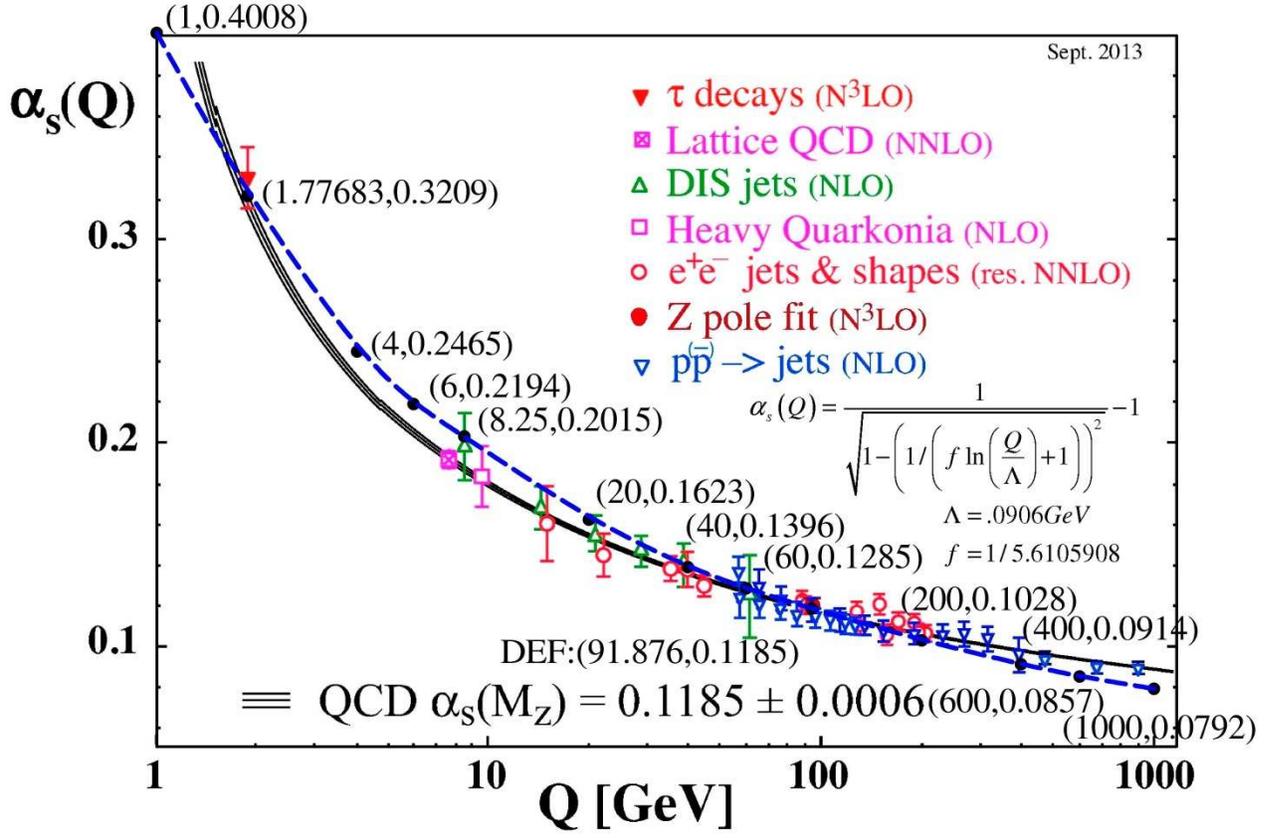
Starting with (18.20), our immediate objective is to use empirical data to determine the value of  $f$ . To do this, we fit (18.20) to two data points which then uniquely determine  $f$  and as a result, the entirety of the mathematical  $\alpha_s(Q)$  curve. The first data point to which we fit the curve is the vertical confinement asymptote, which we set to  $\Lambda = \Lambda^{(6)} = 90.6MeV$ , right at the mean of the empirical data reported in [9.24a] of [24]. The second data point we use is the very same one that is used in Figure 9.4 of [24] which is Figure 9 here, namely, that  $\alpha_s(M_Z) = .1185$ , wherein we take the  $Z$ -mass to be  $M_Z = 91.876GeV$  based on the mean empirical data reported to be  $M_Z = 91.1876(21)GeV$  at [25]. This is the common choice of convention for defining the running strong coupling. So, we place  $\Lambda = \Lambda^{(6)} = .0906GeV$  in (18.20), and then we sample various numeric values for  $f$  in (18.20) until finding a value for  $f$  which fits  $\alpha_s(M_Z) = .1185$  at  $M_Z = 91.1876(21)GeV$  to five decimal places beyond  $\alpha_s(M_Z) = .1185$ , that is, such that  $\alpha_s(M_Z) = .118500000$ . The value we determine for  $f$  via this empirical fitting is:

$$f = \frac{1}{5.6105908}. \quad (18.21)$$

We then use this and  $\Lambda = \Lambda^{(6)} = .0906GeV$  in (18.20) to write:

$$\alpha_s(Q) = \frac{1}{\sqrt{1 - (1/(f \ln(Q/\Lambda) + 1))^2}} - 1 = \frac{1}{\sqrt{1 - \left(1 / \left(\frac{\ln(Q/.0906GeV)}{5.6105908} + 1\right)\right)^2}} - 1. \quad (18.22)$$

Now, we simply graph (18.22) above, and to permit close comparison to the empirical data, we superimpose this over Figure 9. The result is below in Figure 14:



**Figure 14: Plot of (18.22) for  $\alpha_s(Q)$  fitted to  $\Lambda = \Lambda^{(6)} = .0906 GeV$  and  $\alpha_s(M_Z) = .1185$  at  $M_Z = 91.876 GeV$ , Superimposed on Figure 9 from PDG**

In the above, we have used (18.22) as fitted to  $\Lambda = \Lambda^{(6)} = .0906 GeV$  and  $\alpha_s(M_Z) = .1185$  at  $M_Z = 91.876 GeV$ , to determine  $\alpha_s(Q)$  at several  $Q$  values, specifically, 1 GeV,  $m_\tau = 1.77683 GeV$ , 4 GeV, 6 GeV, 8.25 GeV, 20 GeV, 40 GeV, 60 GeV, the defined  $\alpha_s(M_Z)$  at  $M_Z = 91.876 GeV$ , 200 GeV, 400 GeV, 600 GeV and 1 TeV. And as stated at (18.21), this fitting yields the empirical  $f = 1/5.6105908$ . We then plot each of the  $\alpha_s(Q)$  determined at these several  $Q$  by placing a  $\bullet$  dot over the PDG Figure and labelling the associated  $(Q, \alpha_s)$  ordered pair next to each dot. We have generally chosen at each order of magnitude  $n = 0, 1, 2$  to plot the  $1 \times 10^n GeV$ ,  $2 \times 10^n GeV$ ,  $4 \times 10^n GeV$  and  $6 \times 10^n GeV$  data points, with certain exceptions. The exceptions are: we plot right at  $m_\tau = 1.77683 GeV$  rather than the nearby  $2 GeV$ , because there have been direct and explicit studies conducted of  $\alpha_s(m_\tau)$ , see, e.g., Figure 9.2a in [24]; we plot at  $\approx 8.25 MeV$  rather than 10 MeV because the PDG drawing shows an explicit empirical data point at this locale, and this is the lowest- $Q$  DIS jet point plotted by PDG; and we plot at  $M_Z = 91.876 GeV$  rather than the nearby 100 GeV because the former is part of the curve definition (the other part being  $\Lambda = \Lambda^{(6)} = .0906 GeV$ ). We then

connect these dots with a bezier curve plotter to yield the blue dashed curve overlaid on the PDG curve, to provide a sense for the smooth curve of (18.22) which connects these dots in comparison with the PDG interpolated / extrapolated curve.

Generally, we find that (18.22) fits the empirical data within many of the error-bars as shown in Figure 14. Of course the curves match exactly at  $M_Z = 91.876\text{GeV}$  because they are both defined such that  $\alpha_s(M_Z) = .1185$ . From  $m_\tau = 1.77683\text{GeV}$  to  $M_Z = 91.876\text{GeV}$  the predicted Bezier curve is slightly higher than PDG's interpolated curve, yet it passes directly through some key empirical data points. At  $m_\tau = 1.77683\text{GeV}$ , (18.22) predicts that  $\alpha_s(m_\tau) = 0.3209$ . This is slightly below the mean of  $\alpha_s(m_\tau) = 0.33$  shown in Figure 9.2a of [24], yet well within the error bars (and more-so than the PDG curve). It is also noted that three of the ten studies plotted in Figure 9.2a of [24] (Baikov, Davier and Boito) are substantially above the mean, and that the other seven (Beneke, Caprini, Maltman, Narison, Boito and SM review) are below the mean and clustered more consistently. If one discards the three above-mean studies as statistical outliers which they seem to be when relative clustering of the seven below-mean studies are considered, then the predicted  $\alpha_s(m_\tau) = 0.3209$  appears very close to the new mean, and is in fact right in the center of the Caprini study. So all told, the predicted fit at  $\alpha_s(m_\tau) = 0.3209$  is supportable by the empirical data.

The next point of interest is the predicted  $\alpha_s(\approx 8.25\text{GeV}) = 0.2015$ . Although this is somewhat higher than the lattice QCD prediction and at the high end of the nearby heavy quarkonia study, this data point fits very closely to the high center of DIS jets study. In fact, in general, the predicted blue-line curve similarly fits well within the high-center of all of the next three DIS jets data, and just below the center of the DIS jets data near  $\alpha_s(\approx 40\text{GeV}) = 0.1396$ . Further, the  $\alpha_s(\approx 40\text{GeV}) = 0.1396$  prediction sits at the very center of the ee jets and shapes data as well as within the high error bars for most of the other ee data.

As to the pp jets, the predicted curve fits fairly well within this data from about 60GeV to 400GeV. Above 400 GeV, the predicted  $\alpha_s$  drops more sharply than is indicated by the final three data points from the pp jets, but not to a degree that rules out the accuracy of the prediction in this domain. For example, at 400 GeV, the prediction is  $\alpha_s(\approx 400\text{GeV}) = 0.0914$ . At this very same data point, the pp data ranges from a low of  $\alpha_s = 0.088$  to a high of  $\alpha_s = 0.104$  with a mean of about  $\alpha_s = 0.096$ , which puts the prediction below the mean but well within the error bars. Certainly, under all circumstances  $\alpha_s$  will diminish further from here. So if we extrapolate the  $\alpha_s = 0.088$  low at 400GeV over to 600GeV, the predicted  $\alpha_s(\approx 600\text{GeV}) = 0.0857$  also appears to be highly feasible. And indeed, if one were to pass the interpolated PDG curve through the mean of the final four pp points before the very last point at about  $\alpha_s(\approx 900\text{GeV}) = 0.089$ , and exclude this final point as an outlier, then the predicted curve appears to be supported by the low end of the pp data as well.

Finally, returning to the low- $Q$  domain at the left end of the curve, it is of interest to observe that the predicted plot point  $\alpha_s(1\text{GeV})=0.4008$  veers the predicted curve sharply to the left of the extrapolated PDG curve. This is required to get over to the asymptote at  $\Lambda = \Lambda^{(6)} = .0906\text{GeV}$ , and it can already been seen by looking very closely at this same region of Figure 11, and seen also in Figure 13. As we start to study data such as the predicted  $\alpha_s(1\text{GeV})=0.4008$ , we are entering a region where perturbation theory simply no longer applies, and one might surmise that the sharper uptick in the PDG extrapolated curve versus the predicted curve is something of an artistic flourish to show that this curve will become vertically asymptotic. However, if one hews tightly to the two empirical data points at  $\alpha_s(m_\tau)=0.3209$  and  $\alpha_s(\approx 8.25\text{GeV})=0.2015$  in the PDG Figure, rather than to the interpolation between those points which *is not tied to specific empirical data points*, then the leftward movement of the predicted curve to make its way over to  $\Lambda = \Lambda^{(6)} = .0906\text{GeV}$  actually does gain support from these two empirical data points.

So in total, the predicted curve of (18.22) and Figure 14 does appear to be supported and is certainly not ruled out, by the weight of empirical data. If this predicted curve is taken to be a correct representation of how nature behaves, then in general the PDG extrapolated / interpolated curve is slightly on the lower between  $m_\tau=1.77683\text{GeV}$  and  $M_Z=91.876\text{GeV}$  and slightly higher above  $M_Z$ . These deviations are systematically interrelated because  $M_Z$  is used as a defining data point and so becomes something of a “fulcrum” for the rest of the curve. And, by virtue of selecting  $\Lambda = \Lambda^{(6)} = .0906\text{GeV}$  as the other data item to establish the curve, which  $\Lambda^{(6)}$  drags the vertical confinement asymptote well to the left of the  $\Lambda^{(n_f)}$  asymptotes for  $n_f = 5, 4, 3$ , see Figure 13, we see that the uptick drawn in the PDG curve is too extreme, and needs to veer more to the left. As to the entire curve, this tells us that the predicted curve has a curvature which is slightly gentler than the curve shown in the PDG extrapolation / interpolation.

All of this gives us the foundation to now reflect on physics at GUT and Planck scale energies. As already discussed at length, Figure 6 based on (17.33) contain a deep potential well which stabilizes the syem to which it relates and confines the probabilities within the system so that that are all between  $0 \leq r \leq r_\Lambda$ . The coupled probability density is that of Figure 7, which is the constant probability density postulated for study prior to (17.6) and eventually illustrated in Figure 7 after normalization and fitting to  $\Lambda = 1/r_\Lambda$ . That this should be a constant over the whole domain  $0 \leq r \leq r_\Lambda$  does not cause any consternation at least insofar as confinement is concerned, because as later shown in Figure 12, the *bare* probability density  $\mu_{0,r} = \partial_r P_0 \rightarrow 0$  goes sharply to zero at the same time that  $\alpha_s \rightarrow \infty$  because of the inverse relationship they bear in (18.4) based on the constant  $\Lambda$ . In other words, very near  $r \rightarrow r_\Lambda$ , the sharp rise in the potential well of Figures 4 and 6 drives the *bare* probability density down to near zero, which is exactly what is to be expected from least action principles.

But what about  $\mu_{0,r} = \partial_r P_0$  as  $r \rightarrow 0$ ? Here too Figures 4 and 6 show a sharp rise in the potential, which one expects will also drive the bare probability density  $\mu_{0,r} = \partial_r P_0 \rightarrow 0$  as  $r \rightarrow 0$ . But this is not so: in Figure 12,  $\mu_{0,r} = \partial_r P_0 \rightarrow 1$  as  $r \rightarrow 0$ , notwithstanding that the potential  $E_1 \rightarrow \infty$  as  $r \rightarrow 0$ , as is clearly seen in Figures 4 to 6. Of course, the fact that  $\mu_{0,r} = \partial_r P_0 \rightarrow 1$  as  $r \rightarrow 0$  is rooted in one of the two premises prior to (18.4) namely, that of asymptotic freedom; that as  $r \rightarrow 0$ ,  $\alpha_s \rightarrow 0$  asymptotically. In other words, we built in a constant bare probability density  $\mu_{0,r} = \partial_r P_0 \rightarrow 1$  and  $\partial_r \mu_{0,r} = \partial_r^2 P_0 \rightarrow 0$  as  $r \rightarrow 0$  from the start, in order to build in asymptotic freedom. But then we used the non-linear quantum field equations (17.13) and (17.33) to deduce the potentials associated with the implied asymptotic freedom, and discovered that deep into the small-scale zone where  $R = r/r_\Lambda \approx .125$  the potential rises sharply which would undoubtedly cause us to have  $\mu_{0,r} = \partial_r P_0 \rightarrow 0$ , and *not*  $\mu_{0,r} = \partial_r P_0 \rightarrow 1$  as  $r \rightarrow 0$ . So the non-linear equations of quantum field theory – via the steep rise in the potential near  $R = r/r_\Lambda \approx .125$  – are themselves contradicting the premise of asymptotic freedom!

Let's get right to the point: Figure 6 is telling us that there is asymptotic freedom *up to a point*, and that that point is in the area of about  $R = r/r_\Lambda \approx .125$ . It is telling us that at around  $R = r/r_\Lambda \approx .125$ , the asymptotic freedom ceases, because the probabilities via the steeply rising potential will be barred by least action from taking on any substantial density when  $R = r/r_\Lambda < .125$ , and that the probability densities will have approach zero as  $R = r/r_\Lambda \ll .125$ . As directly as possible: *Figures 4 and 6 are telling us to expect new physics in which  $\alpha_s$  is not asymptotically free, as  $R = r/r_\Lambda < .125$* . The question we now ask is this: at what  $Q$  is Figure 6 predicting that should we expect this new physics? Previously, we were not equipped to answer this question. But now that we have fitted the QCD curve and found in (18.21) that the fitting parameter  $f = 1/5.6105908$ , we do have the ability to ask about the  $Q$  associated with  $R = r/r_\Lambda \approx .125$ . As we shall now see,  $R = .13186$  corresponds to a GUT energy of  $Q = 10^{15} GeV$ , and  $R = .10798$  corresponds to the Planck energy  $Q = 1.22 \times 10^{19} GeV$ . So Figure 6 in view of the fitting of Figure 14 via  $f = 1/5.6105908$  is predicting new physics precisely in the GUT-to-Planck energy domain. This is exactly what we expect to see *a priori*, so let us now examine specifically how this comes about.

If we compare (18.6) with which we started the current development to (18.22) which is fitted to the PDG curve in Figure 14, we see inside the square root term that what originally started as the radial coordinate  $R = r/r_\Lambda$  eventually became  $1/(f \ln(Q/\Lambda) + 1)$ , i.e., that the latter occupies the same position in the square root that was initially occupied by the former, which is also the same position in the same square root occupied by  $v/c$  in special relativity wherein the speed of light becomes a material limitation. The 1 at the end of (18.22) is beside the point here; it is just used to shift the curve to the down so that the vertical and horizontal confinement and freedom asymptotes approach the vertical and horizontal axes rather than approach one unit away from the axes. So let us represent this migration of the original  $R$  by:

$$R = \frac{r}{r_\Lambda} = \frac{1}{f \ln(Q/\Lambda) + 1} \sim \frac{v}{c} \quad (18.23)$$

Clearly, when  $Q/\Lambda=1$ , we have  $\ln(Q/\Lambda)=0$ , and so  $R=1$  and  $r=r_\Lambda$ , which is the desired correspondence. And when we have  $f=1/5.6105908$  as use this as in (18.22) we can fit the PDG curve fairly within the empirical error bars over the 1 GeV to 1 TeV domain as seen in Figure 14. We include the similarity to  $v/c$  as a reminder that  $r \rightarrow r_\Lambda$  from below, “sub-radially,” has the same effect in this square root as  $v \rightarrow c$  from below, “sub-luminally,” in special relativity. If we use (18.23) to go backwards from (18.22), then with the vertical shift we have:

$$\alpha_s(Q) = \frac{1}{\sqrt{1-R^2}} - 1. \quad (18.24)$$

It is simple to invert (18.23) and rewrite this as:

$$Q = \Lambda \exp \frac{1}{f} \left( \frac{1}{R} - 1 \right) = \Lambda \exp \frac{1}{f} \left( \frac{r_\Lambda}{r} - 1 \right) \equiv \frac{\hbar c}{r'} \quad (18.25)$$

where in the final term we have *defined*  $r' = 1/Q$  in natural  $\hbar = c = 1$  units. Because  $Q$  is clearly an *observable* energy as we see in Figure 14, this means by deBroglie that  $r'$  is the associated observable length scale. We may also restructure (18.25) via  $\Lambda r_\Lambda = 1$  and defining  $R' \equiv r'/r_\Lambda$ , into the form:

$$R' \equiv \frac{r'}{r_\Lambda} = \exp \frac{1}{f} \left( 1 - \frac{r_\Lambda}{r} \right) = \exp \frac{1}{f} \left( 1 - \frac{1}{R} \right) = \frac{\Lambda}{Q} \quad (18.26)$$

So now we have two different radial coordinates,  $R = r/R_\Lambda$  and  $R' = r'/r_\Lambda$ . With one more inversion, (18.26) becomes:

$$R = \frac{r}{r_\Lambda} = \frac{1}{1 - f \ln R'} = \frac{1}{1 - f \ln(r'/r_\Lambda)} = \frac{1}{1 - f \ln(\Lambda/Q)} \sim \frac{v}{c}, \quad (18.27)$$

which is a variant of (18.23) via  $\ln(x) = -\ln(1/x)$ , with some further alternative terms including the  $v/c$  similarity.

Now, it will be appreciated that  $R$  and  $R' = \Lambda/Q$  in (18.26) and (18.27) are simply different coordinates against which to plot  $\alpha_s$  in (18.24). Simply put, they are related to one another by a general coordinate transformation which we can make explicit by deducing from (18.26) that:

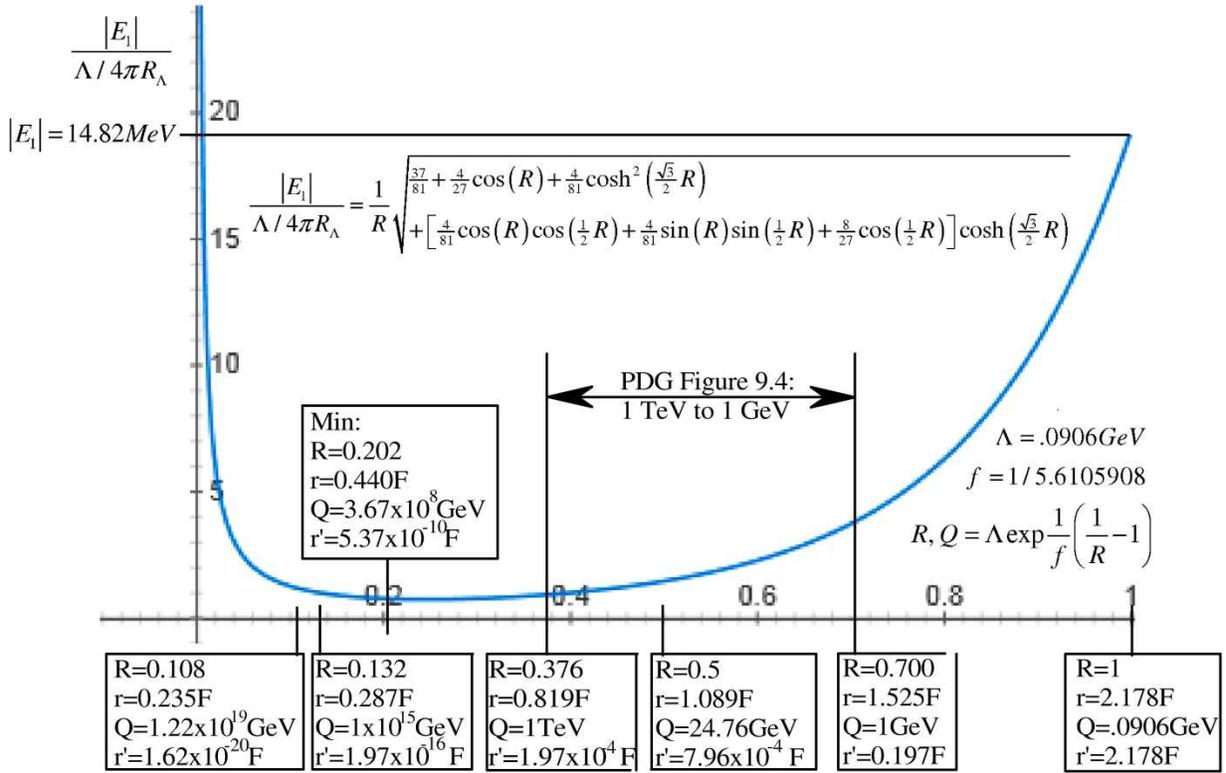
$$dR' = \frac{\partial R'}{\partial R} dR = \frac{1}{f} \frac{1}{R^2} \exp \frac{1}{f} \left( 1 - \frac{1}{R} \right) dR \quad (18.28)$$

Similarly, so too is  $Q$  just another coordinate which helps us to plot  $\alpha_s$ . From (18.25),  $Q$  is arrived at from  $R$  by the general coordinate transformation:

$$dQ = \frac{\partial Q}{\partial R} dR = -\Lambda \frac{1}{f} \frac{1}{R^2} \exp \frac{1}{f} \left( \frac{1}{R} - 1 \right) dR. \quad (18.29)$$

We will not directly use either (18.28) or (18.29); we simply write those here to illustrate the point about how these are simply general coordinate transformations.

Now, let us return to Figure 6 for the confining, stable potential, rescale that Figure so it is expressed in term of the coordinate  $R$  over the domain  $0 \leq R \leq 1$ , and show this domain simultaneously in all four coordinates  $R, r, Q$  and  $r' = R' r_\Lambda$ .



**Figure 15: The First-Order Quantum Potential Well, as a function of  $Q$**

Figure 15 raises at least two very intriguing points, one about GUTs and the Planck scale, the other about measurements in spacetime on the sub-nuclear scale.

As to the first point about GUT and Planck scales, because  $Q$  bears an exponential relationship to  $1/R$ , as  $R = r/r_\Lambda = 2.178F$  diminishes from 1 down toward 0, at a certain point,  $Q$  will rise very rapidly. From right to left, over the domain of approximately  $1 \geq R \geq 0.703$

things start slowly, and all we do is get from  $Q = \Lambda = .0906 GeV$  to  $Q = 1 GeV$  which starts the left side of Figure 9. The entire domain for the PDG Figure 9 from 1 GeV to 1 TeV is covered over  $0.703 \geq R \geq 0.376$ . At the coordinate  $R = r/r_\Lambda = .5$ , which is  $r = 1.098F$ , the corresponding  $Q = 24.76 GeV$ . But moving further to the left, the exponential kicks in and things become very interesting. The minimum of the potential is found to be at  $R = 0.202$ , which is the coordinate  $r = 0.440F$ . But the probe energy for this is now a whopping  $Q = 3.67 \times 10^8 GeV$ , which is equivalent to  $r' = 5.37 \times 10^{-10} F$ . This is over  $10^5 TeV$  and so is way beyond any foreseeable direct experimental observation. But, because this point is the minimum, by least action, one would expect that the *bare* probability density reaches its maximum at this point, and given that the overall context for this analysis is a constant *coupled* probability density  $\partial_r(hP_0) = \Lambda$ , see Figure 8, this would mean that  $\alpha_s$  reaches its asymptotic minimum. But now things really get interesting, because the non-linear quantum field theory which gives us Figure 15 has somehow managed to tell us that above  $Q = 3.67 \times 10^8 GeV$  the asymptotic freedom ends, and the running coupling slowly starts to rise again as the bare probability diminishes due to a now-rising potential which by least action lessens the probability for being in a higher-potential state. A GUT unifying electroweak and strong interactions is expected to emerge at about  $Q = 10^{15} GeV$  which corresponds to  $R = 0.132$ , which is  $r = 0.28F$  but  $r' = 1.97 \times 10^{-16} F$ . And the potential begins a substantial rise in the region of  $R = 0.108$  which happens to map to the Planck energy  $Q = 1.22 \times 10^{19} GeV$ , which has the coordinates  $r = .236F$  and  $r' = 1.62 \times 10^{-20} F$ , namely, the Planck length. So it is a point of deep fascination that Figure 15 together with the fitting  $f = 1/5.6105908$  motivated by empirical strong interaction data is actually telling us via the rising potential to expect substantially new physics to occur in the region of the Planck scale.

This also tells us that our supposition that asymptotic freedom continues right down to  $R = 0$  is an incorrect supposition that needs to be revisited, and that that freedom bottoms out in the region of  $Q = 3.67 \times 10^8 GeV$  which is still way beyond foreseeable detection. As to how one corrects the asymptotic freedom supposition to accord with the clear rise in potential at the Planck scale, Figure 12 provides perhaps the best visual example for how to approach this. Figure 12 is based on the +x, +y quadrant of the sixth root of a circle  $y = \sqrt{1-x^2}$ . We deliberately built this such that  $\partial y / \partial x = 0$  at  $x = 0$  to create asymptotic freedom at  $x = 0$  and such that  $\partial y / \partial x = -\infty$  at  $x = 1$  to create confinement at  $x = 1$ , see (18.5) et seq. Now, if we want the asymptotic freedom to bottom out around the same place that the potential reaches its minimum at  $R = 0.202$  i.e.,  $Q = 3.67 \times 10^8 GeV$  and for  $\alpha_s$  to then start to rise and soon thereafter meet up with the other three interaction couplings and become asymptotically infinite at  $R = 0$  to ensure that the bare probability becomes zero at  $R = 0$  because the  $R = 0$  potential in Figure 15 is infinite, we will need a function for which  $\partial y / \partial x = +\infty$  at  $R = 0$ . So in sum, whatever function we choose as our mathematical “seed” for the *bare* probability density, will really need to satisfy three constraints:  $\partial y / \partial x = +\infty$  at  $x = 0$ ;  $\partial y / \partial x = 0$  at  $x \approx 0.202$ , and  $\partial y / \partial x = -\infty$  at  $x = 1$ . We leave this to future study, but note that a simple shifting of a circle to the right, e.g., using a function such as  $y = \sqrt{1-(x-1)^2}$  may be too simplistic, because this hits

$\partial y / \partial x = 0$  at  $x = .5$ , and the potential in Figure 15 appears to provide clear marching orders to maximize the probability density near  $x \approx 0.202$ , not  $x = .5$ . Given that a circle is a special case of an ellipse, it would be possible to rotate an ellipse such that  $\partial y / \partial x = +\infty$  at  $x = 0$  and  $\partial y / \partial x = 0$  at  $x \approx 0.202$  and  $\partial y / \partial x = -\infty$  at  $x = 1$ , but it is not immediately apparent how one would define the probability densities, because the  $\partial y / \partial x = +\infty$  and  $\partial y / \partial x = -\infty$  points then have different heights along the  $y$ -axis. Again, as noted, we simply point this out, but leave this melding the  $\alpha_s$  curve at observable energies with the running coupling curves near the Planck scale for the future.

The one final observation about the GUT and Planck scales we will make is this: it was noted earlier that because  $M_Z$  is used as a defining data point in Figure 14, it becomes something of a “fulcrum” for the rest of the curve. Although we have pointed out how the predicted curve of Figure 14 does fit within experimental error bars for much of the data over the 1 GeV to 1 TeV domain, it still must be noted that the predicted curve in Figure 14 “pivots” about the  $M_Z$  point such that it is slightly higher than the statistical mean of all the data as shown in the PDG curve for  $m_\tau < Q < M_Z$ , and slightly lower for  $Q > M_Z$ . If the asymptotic freedom does indeed bottom out near  $Q = 3.67 \times 10^8 \text{ GeV}$  – and indeed *we know that it will and must bottom out at some energy before we get to the Planck scale* – then these GUT type effects will already make their presence slightly known at lower energies by slightly raising  $\alpha_s(Q)$  in the  $Q > M_Z$  domain. So if one were to take Figure 14 and posit that there is some nominal increase in the predicted  $\alpha_s(Q)$  in the  $Q > M_Z$  domain *once these GUT effects are accounted for*, then along with this slight upward shift for  $Q > M_Z$ , the fulcrum at  $Q = M_Z$  coupled with the fixed asymptote at  $\Lambda = \Lambda^{(6)} = .0906 \text{ GeV}$  will cause a slight *downward* shift in the predicted  $\alpha_s(Q)$  curve in the  $m_\tau < Q < M_Z$ , and really in the entire  $\Lambda < Q < M_Z$ , domain. So, to the extent that the predicted curve (18.22) graphically illustrated in Figure 14 is found to slightly deviate from empirical data as represented in the PDG curve, it appears highly likely that this slight deviation may be fully accounted for by the fact that in the predicted curve we are treating asymptotic freedom as if it goes on forever right to  $r = 0$  and  $Q = \infty$ , when in fact the more realistic physical supposition is that asymptotic freedom goes on only up to a certain, definite  $Q$  which is less than the Planck energy and – based on Figure 15 – possibly less than the GUT energy. So, in short, any true deviation between the two curves in Figure 14 is likely the result of not taking into account Planck-scale and / or GUT effects upon asymptotic freedom when computing the predicted curve of Figure 14.

The second point of intrigue raised by Figure 15 has to do with subnuclear measurement. This point is raised simply by noting, for example, that at  $Q = 10^{15} \text{ GeV}$ , which corresponds to  $R = 0.132$ , the first radial coordinate is  $r = 0.28F$  but the transformed radial coordinate via (18.26) is  $r' = 1.97 \times 10^{-16} F$ . Or, for example, by noting that at the Planck energy  $Q = 1.22 \times 10^{19} \text{ GeV}$ , the first radial coordinate  $r = .236F$  while the transformed radial coordinate  $r' = 1.62 \times 10^{-20} F$ , namely, the Planck length. So the theory has given us two coordinates which measure a length. One is the original radial coordinate  $r$  which came out of

the path intergration and its related  $R = r / r_\Lambda$ . The second is the transformed radial coordinate  $r'$  defined in (18.26) in relation to  $r$  via the parameters  $\Lambda = 90.6 \text{ MeV}$  and  $f = 1/5.6105908$  which were used to fit the circle-based QCD curve to the observed QCD data, and which relates to the first radial coordinate system of  $r$  via the general coordinate transformation (18.28).

Now, we have known for almost a century, since the advent of General Relativity, that a coordinate system can be chosen completely arbitrarily, and that the laws of nature must be invariant with respect to any and every choice of coordinates that might be made. Sometimes a coordinate system is chosen because it facilitates a mathematical calculation such as taking an integral, for example, the coordinate system  $x'^2 = \theta \rightarrow x''^2 = u = \cos \theta$  which enabled us to do the integral in (14.17). But eventually, we need to find and choose a coordinate system which matches up with the clocks and measuring rods and scales that we use to measure what we are observing. And when we observe sub-nuclear interactions, the measurable observable is the energy scale  $Q = \hbar c / r'$  and its associated length  $r'$ . So it is the transformations (18.28) and (18.29) which get us from a length coordinate  $r$  which we do not measure directly, to a length coordinate  $r'$  which we do measure directly. How do we know that we measure  $r'$  and not  $r$ ? Because the empirical data in Figure 14 tells us so!

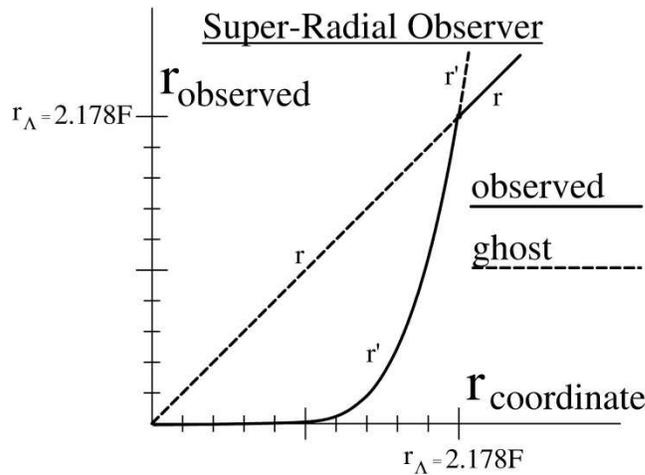
*A priori*, there is no reason why  $r$  should not provide the proper measuring rod to measure lengths in the sub nuclear scale. And in fact,  $r$  would be a perfectly good representation of a directly measurable length if the strong coupling  $\alpha_s(Q)$  ran linearly with  $Q$  rather than linearly with  $\ln Q$ . But we know, for example, that  $r' = 1.97 \times 10^{-16} F$  which corresponds to  $Q = 10^{15} \text{ GeV}$  is a much better radial measure of the physics at the coordinate  $R = 0.132$  than is  $r = 0.28 F$ . And in the directly observable domain, we know that that  $r' = .197 F$  is a better radial measure of the physics we observe at  $1 \text{ GeV}$  than is  $r = 1.52 F$ ; just as  $r' = 1.97 \times 10^{-4} F$  is a better measure of the observed physics at  $1 \text{ TeV}$  than is  $r = .819 F$ .

But look at what we have just said: the coordinate  $r$ , which is certainly a radial coordinate that corresponds – or so we thought – to a measurable physical length, *does not work as a direct observable measuring rod on the sub-nuclear scale*. Rather, if we want a coordinate that maps to what we measure on the sub-nuclear scale, the empirical data tells us that *we must choose  $r'$* . Again: both  $r$  and  $r'$  are perfectly acceptable and valid coordinate systems. But only one of them,  $r'$ , is directly equal to a radial length which we observe when we do empirical experiments on the sub-nuclear scale. So we must inquire: what has happened to turn  $r$  from a perfectly good coordinate which corresponds directly to observable lengths, into a perfectly good coordinate which *no longer corresponds to observable lengths and must be transformed into  $r'$  to yield an observable length measure?*

*What has happened is that in measuring sub nuclear physics, we have crossed an asymptotic barrier at  $\Lambda = 90.6 \text{ MeV}$  and  $r_\Lambda = 2.178 F$ . When we take “super-radial” measurements at  $r > r_\Lambda = 2.178 F$ , the original radial coordinate  $r$  is a perfectly good measure of observable length. But when we take “sub-radial” measurements at  $r < r_\Lambda = 2.178 F$ , the physics itself – of crossing through the  $\Lambda = 90.6 \text{ MeV}$  barrier with our measuring instrumentation –*

requires us to now use  $r'$  to properly represent the measurements we are taking. And what happens precisely at  $r = r_\Lambda = 2.178F$ ? Studying (18.26), we see that at  $r = r_\Lambda$ , we also have  $r' = r_\Lambda$ . So right at  $r_\Lambda$  these two coordinates are identical,  $r = r' = r_\Lambda$ . But as soon as we move away from  $r_\Lambda$  in either direction, these two coordinate systems diverge exponentially / logarithmically from one another, and the only other place where they meet up again is at  $r = r' = 0$ .

Let us try to understand this seemingly-required coordinate transformation from the standpoint of a “super-radial” observer – an observer such as ourselves situated in the  $r > r_\Lambda = 2.178F$  world – taking measurements of “sub-radial” physics in the  $r < r_\Lambda = 2.178F$  behind the physical asymptotic confinement barrier at  $\Lambda = 90.6MeV$  a.k.a.  $r_\Lambda = 2.178F$ . For that observer, the radial coordinate which correspondd to what is measured is  $r$  for  $r > r_\Lambda = 2.178F$ ,  $r'$  for  $r < r_\Lambda = 2.178F$ , and either coordinate at  $r = r' = r_\Lambda$ . This is laid out in Figure 16 below:



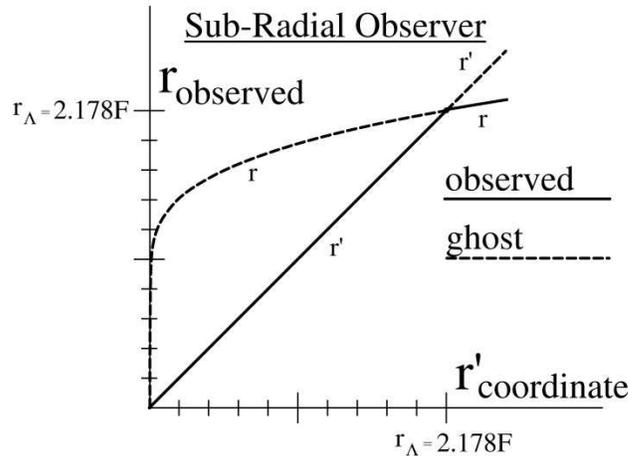
**Figure 16: Variation of Observed  $r = \hbar c / Q$  in relation to  $r_\Lambda = \hbar c / \Lambda = 2.178F$  for a “Super-Radial” Observer Situated at  $r > r_\Lambda = 2.178F$**

So what has transpired at  $r_\Lambda = \hbar c / \Lambda = 2.178F$  is that nature herself seems to have forced a change of the coordinate system which corresponds to the observed radial length  $r_{\text{observed}}$ , from the original  $r$  to the  $r'$  of (18.26). Of course,  $r$  is still a perfectly good coordinate; it is simply not a coordinate which any longer corresponds to the radial lengths which are observed via the relation  $r' = \hbar c / Q$  based on Figure 14. So the question which now occurs is this: is any other physical precedent for this sort of situation in which nature herself, as the results of crossing some physically-meaningful, observable barrier, seems to force a change in the coordinate system needed to describe what is observed?

Actually, the answer appears to be yes: refraction of light at a surface between two unlike media is very good analogy to what is seen in Figure 16. If we view the confinement asymptote at  $r_\Lambda = \hbar c / \Lambda = 2.178F$  as analogous to the top surface of a body of water over which the water surface interfaces with the air, then a light ray at a first non-normal angle  $\theta_1$  will, upon striking

the surface, be altered to so as to refract to a second angle  $\theta_2 \neq \theta_1$ . The incoming  $r$  for the light rays will become *physically altered* to a different  $r'$  just as in Figure 16 above. Although one usually describes this by using one set of coordinates and simply changing the angle in those coordinates at the interface, one could alternatively describe refraction by leaving the angle unchanged *and instead simply rescaling the coordinates* at the interface. The only difference is that for ordinary refraction, assuming a constant medium, the angle does not keep varying as one goes deeper into the medium but rather stays constant, whereas in Figure 16, via (18.26), the  $r'$  rapidly becomes exponentially smaller. As with refraction, it is possible in Figure 16 to talk about what “would have been” the line for  $r_{\text{observed}}$  if the  $r_\Lambda$  asymptote had not interceded, just as one can talk about the angle at which light would have travelled had it not struck the water and been refracted. That “would have been” line in Figure 16 is the “ghost”  $r$  line. And the ghost  $r'$  curve shows what the observed radius would have been had the  $r_\Lambda$  asymptote interceded even further out than it did. It should also be noted that although  $r'$  in Figure 16 has a sharp upward slope at  $r = r_\Lambda$ , this trend only continues for a limited range and domain. If one were to look at this curve for  $r \gg r_\Lambda$ , it would be seen that this curve veers sharply to the right right around  $r'/r_\Lambda \approx 200$ , the then approaches a *horizontal asymptote* which is fixed by the fitting constant  $f = 1/5.6105908$  at  $\exp(1/f) = 273.3054407$ .

If the viewpoint of a “super-radial” observer is akin to that of a person situated in the air and observing light refract once it strikes the water, then one can equally adopt the viewpoint of an observer who is underwater and viewing refraction from that perspective. The equation which describes this is the inverted relationship (18.27), and this is graphed in Figure 17 below.



**Figure 17: Variation of Observed  $r = \hbar c / Q$  in relation to  $r_\Lambda = \hbar c / \Lambda = 2.178F$  for a “Sub-Radial” Observer Situated at  $r < r_\Lambda = 2.178F$**

Here, we might imagine that the “observer” is situated with a quark confined inside the asymptote at  $r < r_\Lambda = 2.178F$ , trying to design up a coordinate system which not only describes what is seen inside the nuclide, but also, maps over to what is seen to be beyond the confinement horizon. Again, while any coordinate system one might choose to use to describe this physics is equally valid, *this does not mean that every coordinate system has a direct linear relationship to*

*something observable*. If the radial length which someone measures with instrumentation is  $r_{\text{observed}}$ , then the only radial coordinate  $r_{\text{coordinate}}$  which matches up with what is observed is one for which  $\partial r_{\text{coordinate}} / \partial r_{\text{observed}} = 1$ , or, without an integration constant,  $r_{\text{coordinate}} = r_{\text{observed}}$ . In Figure 16 for the super-radial observer,  $r_{\text{coordinate}} = r_{\text{observed}}$  is the 45 degree line toward the upper right where  $r > r_{\Lambda}$ . In Figure 17  $r_{\text{coordinate}} = r_{\text{observed}}$  is the 45 degree line in the middle of the Figure where  $r < r_{\Lambda}$ . Note that Figure 17 is just Figure 16 flipped around the line at 45 degrees. It should also be noted that the  $r$  curve, although headed to the right at  $r = r_{\Lambda}$  will start to veer sharply upward, and will then approach a vertical asymptote fixed by the fitting constant  $f = 1/5.6105908$  at  $\exp(1/f) = 273.3054407$ .

All of this suggests that it may be fruitful to view the confinement asymptote of a nuclide at  $r = r_{\Lambda} = 2.178F$  as a physical boundary at which a *medium change* occurs, just as it does when light strikes water or glass or some other diffractive medium. Although one is permitted to choose any coordinate system one wishes, if one wishes to describe the effects of this medium change using coordinates *for which*  $\partial r_{\text{coordinate}} / \partial r_{\text{observed}} = 1$ , one is required at the interface to switch between the  $r$  and  $r'$  coordinates via the relationships (18.26) and (18.27), which may be couched as the general coordinate transformations (18.28) and (18.29). So the observable physics of a nucleon at the physical barrier  $r = r_{\Lambda} = 2.178F$  forces a general coordinate transformation upon any observer who wishes to employ coordinates for which  $\partial r_{\text{coordinate}} / \partial r_{\text{observed}} = 1$  in that observer's frame of observation. Again: any choice of coordinates is just as valid as any other choice. But, if we ourselves decide that we prefer a coordinate for which  $\partial r_{\text{coordinate}} / \partial r_{\text{observed}} = 1$ , for *all*  $r_{\text{observed}}$  whether large or small, then nature herself – right at the physical barrier at  $r = r_{\Lambda} = 2.178F$  – will force us to make the transformations (18.26) through (18.29) illustrated by Figures 16 and 17.

Referring back to (18.22) and Figure 14, one final point should also be made before concluding this section: we have observed, for example in (18.25), and after (18.6) and (18.19), that  $R = r/r_{\Lambda}$  within the radical  $1/\sqrt{1-R^2}$  is exactly analogous to  $v/c$  within the radical  $\gamma = 1/\sqrt{1-(v/c)^2}$  which is central to special relativity, and that in each circumstance, there is an asymptotic limit being set which is associated with an observed physical limitation. In the case of  $1/\sqrt{1-R^2}$ , one has a confining potential and quarks are not allowed to cross a radial boundary at  $r_{\Lambda}$ . In the case of  $\gamma = 1/\sqrt{1-(v/c)^2}$  one has a confining limitation which bars material bodies from ever reaching or exceeding the speed of light. In this analogy, the super-radial observer of Figure 16 may be analogized to a *super-luminal* observer and the sub-radial-observer of Figure 17 may be analogized to a *sub-luminal* observer, while (18.26) through (18.29) are analogized to coordinate transformations which relate between what is observed by a superluminal observer and what is observed by a sub-luminal observer.

Now, to be very clear, this is only an analogy. And a radial coordinate  $r$  has a different physical meaning than a velocity  $v = \partial r / \partial t$  which measures a change in the radial coordinate  $r$

in relation to a change in the time coordinate  $t$ . And even if this analogy has some deeper physics behind it, there is no *a priori* reason to conclude that (18.26) through (18.29) would in fact be the form of the coordinate transformations which relate superluminal to subluminal observations, that  $f=1/5.6105908$  would be the empirical fitting constant for super/sub luminal transformations, and that Figures 16 and 17 would be the correct visual pictures for this. Nevertheless, with all of these caveats, it is a matter of intrigue that the speed of light  $c$  and the QCD cutoff  $\Lambda=1/r_\Lambda$  are each understood to be material, physical limitations which exist in nature, and that they each enter into critical physics equations in the same way:  $1/\sqrt{1-R^2}$  for the mathematical root of the running QCD curve fitted in Figure 14, and  $\gamma=1/\sqrt{1-(v/c)^2}$  through special relativity. This at least raises the prospect of studying the material barrier posed by the speed of light and long-thought to bar superluminous material transport, as a sort of medium change analogous to the nuclear confinement surface or to a refraction surface between varying media. Then, we may embark to enquire about the transformation laws which nature has so far hidden from human comprehension, as between the subluminal and the superluminal universe.

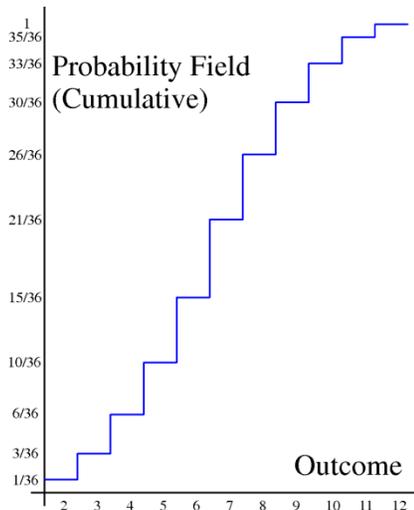
## 19. Gaussian Probability Densities in Non-Linear Quantum Field Theory

The next example of a quantum probability density we shall consider is a Gaussian probability distribution. But in preparation for that, this is a good time to take stock of what the examples reviewed thus far teach about the nature of probability and probability densities in non-linear quantum field theory.

In classical field theory, the “field” is a coupled gauge field  $gG^\sigma$  which has a mass dimension of +1. A field strength tensor defined according to  $F^{\mu\nu} = \partial^{[\mu}G^{\nu]} - ig^2[G^\mu, G^\nu]$ , see (1.5), may then be thought of as the spacetime “field density” of this coupled classical field. This has a mass dimension of +2. And finally, the “source” of this field is a current density  $-J^\mu = \partial_\sigma F^{\sigma\mu} - i[G_\sigma, F^{\sigma\mu}] = D_\sigma F^{\sigma\mu}$ , see (5.15) et seq. This has a mass dimension of +3. Quantum field theory has an identical structure, but with different mass dimensionalities. In quantum field theory, the “field” is a dimensionless coupled “probability field”  $hP_0$ . When this field is time-independent and isotropic and varies linearly with radial distance and stops its ascent at a given  $r_\Lambda$ , its character is illustrated in Figure 8. The quantum “field density” is the spacetime gradient of the probability field  $\partial_\mu(hP_0)$  and it has a mass dimension +1. In the situation where the quantum probability field is that of Figure 8, the radial field *density* is that of Figure 7. And, in quantum field theory, the “source” of the quantum field is  $W(J)$  obtained from path integration. This source has dimensions of action, which is energy×time. In  $\hbar=c=1$  units, this action source is also dimensionless. But if we remove the time dependency as we did in sections 14 and 15, then the quantum field source is a potential energy which has a mass dimension of +1. For the first recursive order of a non-linear quantum field theory, this potential is  $E_1$  in (15.42). And for the linear quantum probability field of Figure 8 and the constant field density of Figure 7, the magnitude  $|E_1|$  of that first-recursive-order potential is illustrated in

Figure 6 and later, Figure 15. Further, it is because the *source* of Figure 5 has a definitive peak in the potential at  $R_{\text{peak}} \cong 8.245$ , we identify this with the six quark QCD cutoff length at  $r_\Lambda = 2.178F$  and therefore concentrate all of the probability density of Figure 7 inside the  $0 \leq r \leq r_\Lambda$  range, as was all developed in the last section. So what does all of this teach us about non-linear quantum field theory in general, and even more so, about the nature of “probability” in physics?

Let us first talk about probability, and dimensionality in probabilistic systems. While in colloquial language one often hears talk that the “probability” of some “event” or “outcome” is  $P$ , where  $0 \leq P \leq 1$ , it is important when dealing with probabilistic / stochastic physics to keep firmly in mind that such talk is really referring to a probability *density*. The probability itself is then the cumulative integral of all of the probability densities, and this integral sums over the domain of the density to the dimensionless number 1. For a *discrete* probability system – let us use the specific example of rolling a pair of dice with a probabilistic result from 2 through 12 – the “dimensionality” of the probability density is measured in an “event” or “outcome” space. One might say by analogy to physics that the probability density has an “outcome dimensionality” of +1. For the pair of dice, the outcomes are the discrete, closed event group wherein one rolls a number from the closed set of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. As is well known, the probability *density* for each of these successive outcomes in the outcome space of outcome dimensionality +1 is  $1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36$  and  $1/36$ . And of course, when summed over the entire event space, the *cumulative distribution* all sums to a probability of 1. This rather simple, but very illustrative example of the discrete “probability field” for rolling a pair of dice, is shown in Figure 18 below.



**Figure 18: The Discrete Cumulative Probability Field for Rolling a Pair of Dice – A Simple Example**

In non-linear quantum field theory, the probability field, such as the field shown in Figure 8, is the exact analog of Figure 8, because it too is a *cumulative distribution*. The only difference is that the physical field in Figure 8 is continuous and linear, while the physical field of Figure 18 is discrete and so contains step functions. In Figure 18 the x-axis has the dimensionality of an

“outcome”, while in Figure 8 the x-axis has dimensionality of a radial position in physical space. Yet, this radial position space is really just another type outcome space, with the outcome simply measured by a continuously-variable spatial position rather than the discrete sum of the two die. Specifically, if we take the gradient of Figure 18 in the outcome space, we obtain the familiar  $1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36$  and  $1/36$  outcome density curve with the peak at  $1/6=6/36$  for the outcome “7,” while if we take the radial gradient  $\partial_r (hP_0)$  of Figure 8 we arrive at Figure 7 which says that the “outcome” of a field quantum being detected at  $r = 1F$  has the same probability density as the “outcome” wherein the field quantum is detected at  $r = 2F$ , or indeed, as the “outcome” wherein the quantum field is detected at any other radial position over the outcome continuum measured by  $0 \leq r \leq r_\Lambda$ .

What is most important to understand based on Figure 18 in comparison to Figure 8 is that the probability field in quantum field theory is not an outcome density field, *but is the cumulative integral over outcome densities*. Again, the quantum probability field is a dimensionless cumulative distribution. As the integral over outcomes, its value at the highest part of the domain along the outcome axis must *always* be equal to unity. At the outcome of a 12 roll of the dice, the probability field is equal to 1. At the outcome of  $r = r_\Lambda$  in Figure 8 for a constant density physical probability field, the probability field is 1.

It is important to be cognizant of this, because nature of *all* cumulative probability fields is such that they must *always* starts at zero at the minimum extremum of the domain and become equal to 1 at the highest extremum of the domain, and because this vastly narrows the sorts of *mathematical* functions which are suitable to be used as physical probability fields. For a *continuously-varying* probability density, the probability field must asymptotically approach 1 at the upper domain extremum. This means that  $hP_0$  as a function of a dimensionless  $R$  must be some type of *sigmoid* function such as the logisitic function  $hP_0(R) = (1 + \exp(-R))^{-1}$ , the Gompertz function  $hP_0(r) = \exp(-\exp(-R))$ , the arc tangent  $hP_0(R) = 2(\arctan(R))/\pi$ , the hyperbolic tangent  $hP_0(R) = \tanh(R)$ , and the error function  $hP_0(R) = \text{erf}(R)$  which is of particular interest as the integral of a Gaussian normal distribution. See, e.g., [29]. Indeed, it will be recognized that the probability field in Figure 18 is itself also a sigmoid function, albeit a discrete sigmoid. Alternatively, for a probability density which has a *discontinuous variation*, i.e., a cutoff in the density such as the one shown in Figure 7,  $hP_0$  will then have a similar character to  $hP_0(r) = r/r_\Lambda = \Lambda r$  in Figure 8. Here, there is a broader set of possible mathematics functions which one may choose from, with the caveat that the function  $hP_0$  must be a probability field, and so must start at zero on the leftmost-extremum of its domain and end up at one on the rightmost portion of its domain. That exhausts the mathematical options: there are only two types of mathematical functions which are suitable as quantum probability fields: sigmoid functions for a continuous probability density, and other functions with cutoffs designed to range from zero to 1 between the low and high extrema of their domains for a density with a discontinuous first derivative such as the one in Figure 7.

We have seen at first recursive order how an isotropic probability density which is constant over the radius  $0 \leq r \leq r_\Lambda$  as shown in Figure 7, using the abelian equation (17.5), will yield a potential  $E_1$  in (17.32) the magnitude of which is given in (17.33) and graphed in Figure 6. By way of contrast, let us again start with (17.5), but now, let us consider the illustrative probability field  $hP_0(R) = \text{erf}(R)$ , which means that the probability density is a Gaussian normal distribution. Here, because we are starting with a sigmoid probability field which asymptotically approaches at 1 at the extremal maximum of infinity, we may assure right at the outset that the probability will be normalized to 1.

So, we start with  $hP_0(R) = \text{erf}(R)$  and its associated Gaussian. Mathematically, we will base the probability density on the function  $y(x) = (1/\sigma\sqrt{\pi})\exp(-x^2/\sigma^2)$ , for which  $\int_{-\infty}^{\infty} y dx = \int_{-\infty}^{\infty} (1/\sigma\sqrt{\pi})\exp(-x^2/\sigma^2) dx = 1$ , and for which  $\sigma$  is the standard deviation. But we will be operating in spherical coordinates and will wish to associate  $x$  with the radial coordinate  $r \geq 0$ . Because these Gaussians are symmetric about  $x=0$ , we can double the function for  $y$  and integrate only over the positive domain. Doing so, we instead use  $y(r) = (2/\sigma\sqrt{\pi})\exp(-r^2/\sigma^2)$ , for which  $\int_0^{\infty} y dr = \int_0^{\infty} (2/\sigma\sqrt{\pi})\exp(-r^2/\sigma^2) dx = 1$ . The standard deviation is still  $\sigma$  and the probability density still integrates to 1, but we have discarded the  $r < 0$  portion of the domain because of the coordinate choice. In Figure 7 we knew to set  $r = r_\Lambda$  as the cutoff for the probability density, because in Figure 5 we had already found that there was a natural peak in potential at  $R_{\text{peak}} \cong 8.245$  which we then identified with  $r_\Lambda$ . So knowing what we now know, let us use  $r_\Lambda$ , not as a cutoff, but as the *standard deviation* for this Gaussian. So with the educated choice  $\sigma = r_\Lambda$ , we now postulate an isotropic probability density which has its peak at  $r=0$  and a radial behavior over the domain  $r \geq 0$  given by:

$$\partial_r (hP_0) \equiv \frac{2}{r_\Lambda \sqrt{\pi}} \exp\left(-\frac{r^2}{r_\Lambda^2}\right). \quad (19.1)$$

Then, we simply calculate the other physics that is associated with this, especially, the quantum potential  $E_1(r)$  from (17.5).

As to the integral of (19.1), we know as just pointed out that:

$$\int_0^{\infty} \partial_r (hP_0) dr = \frac{2}{r_\Lambda \sqrt{\pi}} \int_0^{\infty} \exp\left(-\frac{r^2}{r_\Lambda^2}\right) dr = hP_0|_0^\infty = 1. \quad (19.2)$$

So the cumulative dimensionless coupled probability field as a function of radial coordinate  $r$  is:

$$hP_0 = \text{erf}(r) = \frac{2}{r_\Lambda \sqrt{\pi}} \int_0^r \exp\left(-\frac{r'^2}{r_\Lambda^2}\right) dr'. \quad (19.3)$$

It is also worth keeping in mind that the Gaussian  $y(x) = (1/\sigma\sqrt{\pi})\exp(-x^2/\sigma^2)$  will become the Dirac delta function  $\delta_\sigma(x) = \lim_{\sigma \rightarrow 0} (1/\sigma\sqrt{\pi})\exp(-x^2/\sigma^2)$ , that is, in the limit where the standard deviation approaches zero. Once again, because this Gaussian is centered about  $x=0$  and we will wish to use this in spherical coordinates for which  $r \geq 0$  we can likewise define a “half-delta” by doubling everything whereby  $\delta_\sigma(r) = \lim_{\sigma \rightarrow 0} (2/\sigma\sqrt{\pi})\exp(-r^2/\sigma^2)$ , with the recognition that  $r \geq 0$  cuts off half the area under the curve. Thus,  $\int_0^\infty \delta_\sigma(r) dr = 1$  as is required. Consequently, we will want to have available based on (19.1), the half-delta function:

$$\delta_{r_\Lambda}(r) = \lim_{r_\Lambda \rightarrow 0} \partial_r (hP_0) = \lim_{r_\Lambda \rightarrow 0} \frac{2}{r_\Lambda \sqrt{\pi}} \exp\left(-\frac{r^2}{r_\Lambda^2}\right). \quad (19.4)$$

With these preliminaries, we simply use (19.1) in (17.5) to specify the one-recursion potential:

$$E_1 = -\frac{1}{4\pi r} \left[ \frac{2}{3} + \frac{1}{9} \left[ \exp\left(i \frac{2}{\sqrt{\pi}} \frac{r}{r_\Lambda} \exp\left(-\frac{r^2}{r_\Lambda^2}\right)\right) + 2 \exp\left(i \frac{1}{\sqrt{\pi}} \frac{r}{r_\Lambda} \exp\left(-\frac{r^2}{r_\Lambda^2}\right)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} \frac{r}{r_\Lambda} \exp\left(-\frac{r^2}{r_\Lambda^2}\right)\right) \right] \right]. \quad (19.5)$$

It is then simpler to use the dimensionless  $R \equiv r/r_\Lambda$  and  $\Lambda r_\Lambda = 1$  to rewrite this as:

$$4\pi \frac{E_1}{\Lambda} = -\frac{1}{R} \left[ \frac{2}{3} + \frac{1}{9} \left[ \exp\left(i \frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + 2 \exp\left(i \frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \right] \right]. \quad (19.6)$$

Similarly now we graph the magnitude of this energy. First, we write the above as:

$$4\pi \frac{E_1}{\Lambda} = -\frac{1}{R} \left[ \frac{2}{3} + \frac{1}{9} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{2}{9} \cos\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \right] + i \left[ \frac{1}{9} \sin\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{2}{9} \sin\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \right], \quad (19.7)$$

$$= -R^{-1}(a + bi)$$

where:

$$\begin{aligned}
 a &\equiv \frac{2}{3} + \frac{1}{9} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{2}{9} \cos\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \\
 b &\equiv \frac{1}{9} \sin\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{2}{9} \sin\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right)
 \end{aligned} \tag{19.8}$$

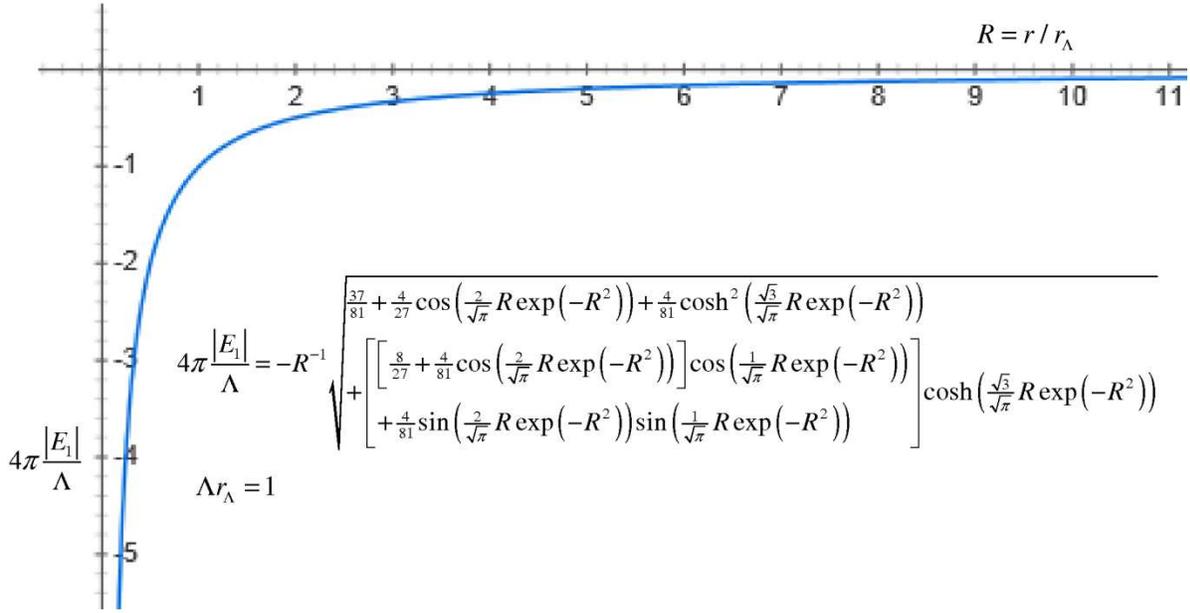
The square magnitude, contrast (17.12), is then:

$$\begin{aligned}
 |a + bi|^2 &= a^2 + b^2 = \frac{4}{9} + \frac{4}{3} \left[ \frac{1}{9} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{2}{9} \cos\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \right] \\
 &+ \frac{1}{81} \cos^2\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{4}{81} \cos^2\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \\
 &+ \frac{1}{81} \sin^2\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{4}{81} \sin^2\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \\
 &+ \frac{4}{81} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) \cos\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \\
 &+ \frac{4}{81} \sin\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) \sin\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \\
 &= \frac{37}{81} + \frac{4}{27} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \\
 &+ \left[ \begin{aligned} &\left[ \frac{8}{27} + \frac{4}{81} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) \right] \cos\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \\ &+ \frac{4}{81} \sin\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) \sin\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \end{aligned} \right] \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right)
 \end{aligned} \tag{19.9}$$

Using this in (19.7) then yields:

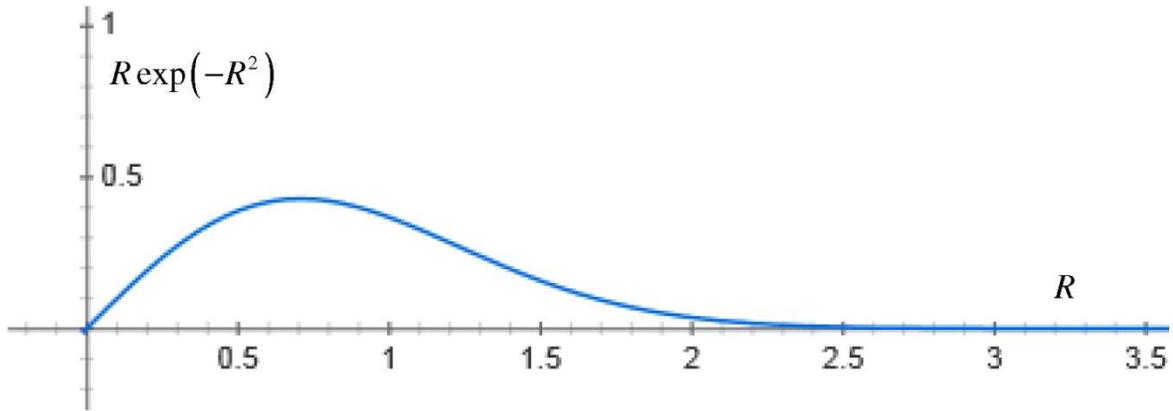
$$4\pi \frac{|E_1|}{\Lambda} = \mp R^{-1} \sqrt{\begin{aligned} &\frac{37}{81} + \frac{4}{27} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \\ &+ \left[ \begin{aligned} &\left[ \frac{8}{27} + \frac{4}{81} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) \right] \cos\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \\ &+ \frac{4}{81} \sin\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) \sin\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) \end{aligned} \right] \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) \end{aligned}}, \tag{19.10}$$

We then graph (19.10) using the negative root, which is shown below in Figure 19:



**Figure 19: Graph of Equation (19.10) Revealing Apparent -1/R Potential for Gaussian Probability Density**

We see immediately that this looks just like an ordinary -1/R potential for an abelian gauge theory such as Quantum Electrodynamics? Might it be just that? The key term driving (19.10)  $R \exp(-R^2)$ . It is clear that in all regions where  $R \exp(-R^2) \rightarrow 0$ , equations (19.6) and (19.10) will both reduce to  $E_1 / \Lambda = -1/4\pi R$  a.k.a.  $E_1 = -1/4\pi r$  which is the Coulomb potential. And  $R \exp(-R^2) \rightarrow 0$  for both  $R \rightarrow 0$  and for  $R = r/r_\Lambda \gg 3$ . So because  $\sigma = r_\Lambda$  is the standard deviation, this means  $R \rightarrow 0$  and  $r \gg 3\sigma$ . Indeed, it is helpful to look at a graph for  $R \exp(-R^2)$  which is shown below and illustrates all of this:



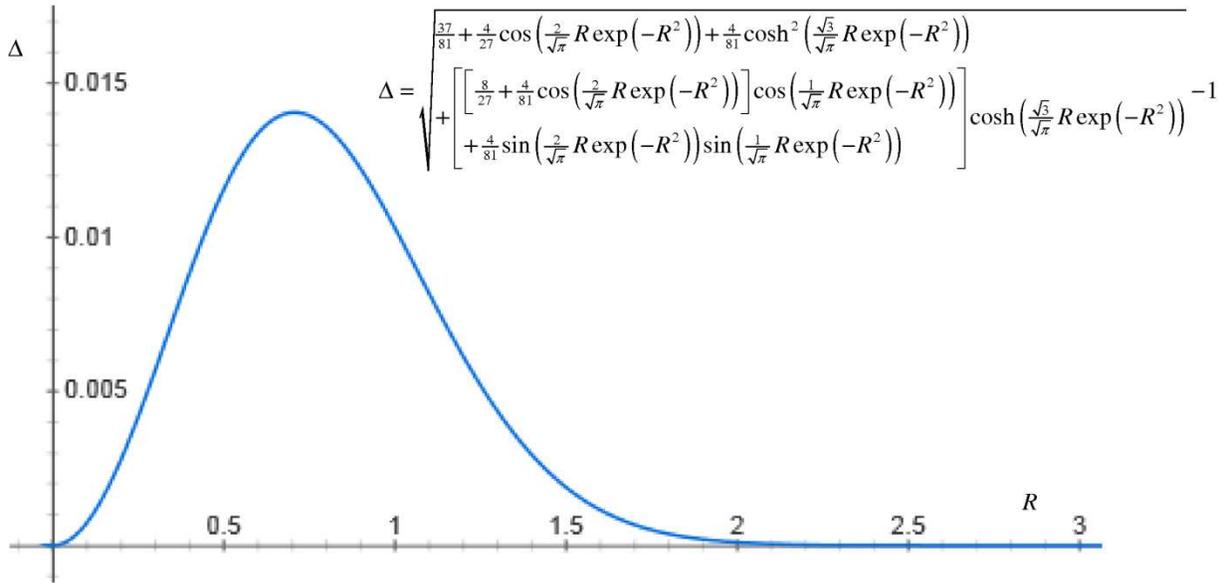
**Figure 20: Graph of  $R \exp(-R^2)$  which Drives (19.6) and (19.10) and Figure 19**

The peak in this function occurs at  $R = 1/\sqrt{2}$  a.k.a.  $r = r_\Lambda / \sqrt{2}$ .

So if Figure 19 and thus equation (19.10) yields the potential  $E_1 = -1/4\pi r$  except for in the regions where  $R \exp(-R^2)$  in Figure 20 is further from zero, the next question is this: by what order of magnitude does  $R \exp(-R^2)$  when it is in its non-zero range cause the Figure 19, equation (19.7) potential to *deviate* from  $E_1 = -1/4\pi r$ ? To answer this, we may use (19.10) to define a deviation parameter  $\Delta \equiv -4\pi R |E_1| / \Lambda - 1 = -4\pi r |E_1| - 1$ , so that:

$$\Delta = \sqrt{\frac{\frac{37}{81} + \frac{4}{27} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right)}{\left[\frac{\frac{8}{27} + \frac{4}{81} \cos\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right)\right] \cos\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right) + \frac{4}{81} \sin\left(\frac{2}{\sqrt{\pi}} R \exp(-R^2)\right) \sin\left(\frac{1}{\sqrt{\pi}} R \exp(-R^2)\right)}} \cosh\left(\frac{\sqrt{3}}{\sqrt{\pi}} R \exp(-R^2)\right) - 1. \quad (19.11)$$

We will have  $\Delta + 1 \rightarrow 1$  when  $R \exp(-R^2) \rightarrow 0$ , so  $\Delta$  measures the difference between the entire parenthetical expression above and 1, i.e., the fractional deviation of (19.10) from a  $1/r$  potential. A graph of  $\Delta$  shows that:



**Figure 21: Fractional Deviation of (19.10) and Figure 19 from a  $-1/R$  Potential**

Just like  $R \exp(-R^2)$  in Figure 20, this deviation peaks at  $R = 1/\sqrt{2}$  a.k.a.  $r = r_\Lambda / \sqrt{2}$ . So within about 1 standard deviation  $\sigma = r_\Lambda$  of the peak at  $R = r/r_\Lambda = 1/\sqrt{2}$ , Figure 19 and thus (19.10) deviates from the Coulomb potential  $E_1 = -1/4\pi r$  by no more than .015 out of 1, or by less than two percent. Over the limited domain where this deviation from a  $-1/r$  potential occurs, the potential slightly decreases, that is, the potential in Figure 19 is slightly lower than it would be if it was a strict  $-1/r$  potential.

All of this brings us full circle back to, and is very informative about, our examination in section 16 of constant probability, zero probability density fields in non-linear quantum field theory. We saw in (16.1) that for  $hP_0 = \text{constant}$ ,  $E_1 = -(N^2 - 1)^2 / 4\pi r$ , which becomes  $E_1 = -1/4\pi r$  when we go over to abelian gauge theory as in (14.31). At the time of finding (16.1), we had not yet established that  $hP_0 = \text{constant}$  is a *cumulative* probability field, as we subsequently did see in Figure 8 and in the discrete analogy of Figure 18. So this means not only that  $hP_0 = \text{constant}$ , but that  $hP_0 = 1$  at  $r \rightarrow \infty$ . And because we are using radial coordinates for which  $r \geq 0$ , we must also have  $hP_0 = 0$  at  $r = 0$  (at which  $E_1 = -1/4\pi(0) = -\infty$  becomes singular also, as has long been known), because a cumulative probability starts at zero at the minimum domain extremum and ends up at unity at the maximum domain extremum. So, how do we start with  $hP_0 = 0$  at  $r = 0$  and end up with  $hP_0 = 1$  at  $r \rightarrow \infty$  all while maintaining  $hP_0 = \text{constant}$  over the entire domain  $r < 0 \leq \infty$ ? There is only one way to do this: the coupled probability density  $\partial_r(hP)$  for a pure  $E_1 = -1/4\pi r$  potential must be a Dirac “half-delta” precisely at  $r = 0$ , so that  $hP_0$  can instantaneously step up from 0 to 1 right at  $r = 0$ .

This is why we pointed out at (19.4) that the Gaussian probability (19.1) becomes a half-delta  $\delta_{r_\Lambda}(r) = \lim_{r_\Lambda \rightarrow 0} \partial_r(hP_0)$  in the limit where we take  $\sigma = r_\Lambda \rightarrow 0$ , or  $\Lambda = 1/r_\Lambda \rightarrow \infty$ . The only way in the context of non-linear quantum field theory to have a potential which is *precisely*  $E_1 = -1/4\pi r$  over the *entire* non-zero domain  $r < 0 \leq \infty$  *without any deviation*, is to have a probability density:

$$\partial_r(hP_0) = \delta_{r_\Lambda}(r). \quad (19.12)$$

So  $\partial_r(hP_0) = 0$  for  $r > 0$ , *but not at*  $r = 0$ . Then, we will have a coupled, dimensionless probability field:

$$hP_0 = 0 \text{ at } r = 0; \quad hP_0 = 1 = \text{constant for } r > 0. \quad (19.13)$$

And then, via (16.1) and (14.31), we will have a potential:

$$E_1 = -(N^2 - 1)^2 \frac{1}{4\pi r} \stackrel{\text{abelian}}{\Rightarrow} -\frac{1}{4\pi r} \text{ for } r > 0 \quad (19.14)$$

This means the potential is *undefined* at  $r = 0$ , not because  $E_1 = -1/4\pi(0) = -\infty$  at  $r = 0$ , but because the probability density  $\partial_r(hP_0) = \delta_{r_\Lambda}(r)$  has an infinite spike of total area 1 at  $r = 0$  and so the relationship (19.14) does not apply at  $r = 0$  because  $\partial_r(hP_0) \neq 0$  at  $r = 0$ . In this sense,

we counteract the singularity  $E_1 = -1/4\pi(0) = -\infty$  at  $r=0$  by having a controlled singularity  $\partial_r(hP_0) = \delta_{r_\Lambda}(r)$  for the probability density at  $r=0$ .

Now, let's talk about the observable physics associated with all of this. The Coulomb potential  $E = -1/4\pi r$  has been well-known ever since work in the late 18<sup>th</sup> century by Henry Cavendish and Charles-Augustin de Coulomb. And, this potential has always been taken, without any apparent empirical contradiction, to be precisely  $E = -1/4\pi r$  over all measurable radial distances outside the charge density, i.e., where the charge density is zero. It has been learned in the intervening centuries that this is based on the photon being a massless, luminous entity, and that spacetime curvature can impact the radius  $r$  due not to anything relating to  $E = -1/4\pi r$  per-se, but to the effects of spacetime curvature upon measurements of length and time. But in flat spacetime, insofar as is known,  $E = -1/4\pi r$  for all finite, non-zero  $r$  outside the charge density.

We see in (19.14) that a precise  $E = -1/4\pi r$  potential over all measurable radial distances, in the context of non-linear quantum field theory, presupposes the  $\partial_r(hP_0) = \delta_{r_\Lambda}(r)$  Dirac delta probability density of (19.12). Via (19.4), this means that the Coulomb potential presupposes the Gaussian probability density (19.1) with  $r_\Lambda = 0$ , where  $r_\Lambda$  is a previously-“invisible” parameter which is precisely set to zero. But while Dirac deltas are sometime-helpful mathematical entities, it seems unrealistic to expect that the real physical world will present us with probability densities which are infinitely tall and infinitesimally narrow, enclosing a total area of 1. This is a mathematical idealization. In the real physical world, we expect that  $r_\Lambda$  will be very small, but not precisely equal to zero. For example, if  $r_\Lambda$  in, say, Figure 19 were to be based on  $r_\Lambda = 2.178F$  which is in turn based on  $\Lambda^{(6)}_{QCD} = .0906GeV$ , then Figure 19 would represent a potential which is extremely close to  $E = -1/4\pi r$  over all  $r$ , but with a very slight reduction in the potential from  $E = -1/4\pi r$  on the order of up to 1.5 percent within 1 standard deviation of  $r_\Lambda/\sqrt{2} = 1.54F$ . So, if we are willing as a matter of physics to entertain the possibility of horizontally stretching the Dirac delta such that  $r_\Lambda = 0$  instead becomes  $r_\Lambda =$  “very small but finite,” then *the physical potential of quantum electrodynamics would become that of Figure 19*. As seen in Figures 11 and 12, within about 1 standard deviation of  $\sigma/\sqrt{2} = r_\Lambda/\sqrt{2} = 1.54F$ , this deviates ever-so-slightly from Coulomb's  $E = -1/4\pi r$ .

In fact, if one steps back, the deviation from a strict  $-1/r$  potential which occurs in Figure 19 and which is highlighted in Figure 21 ought not to be at all surprising. It has long been understood that the strict  $-1/r$  potential of Cavendish and Coulomb only applies *outside* the region in which the source charge is distributed. Inside the charge distribution, the potential is not expected to be  $-1/r$ , but rather something else. So if we use a probability density  $\partial_r(hP_0) = \delta_{r_\Lambda}(r)$  from (9.4) which is a Dirac delta right at  $r=0$ , then every spatial location for which  $r > 0$  will be outside the  $\delta_{r_\Lambda}(r)$  spike, and so the potential will be Coulomb's  $-1/r$  for all  $r > 0$ . On the other hand, if we spread the Gaussian out from being a spike by using a probability density (9.1) in which  $\sigma = r_\Lambda$  is very small but finite, then in the domain close to

$\sigma = r_\Lambda$ , one is no longer “outside” the source charge, but rather, is right in the middle of the source charge. Therefore, it is to be expected that the potential will measurably deviate from that of Coulomb within a few  $\sigma$  of  $r_\Lambda$ . Once we go more than a few standard deviations beyond  $\sigma = r_\Lambda$ ,  $\partial_r(hP_0) \rightarrow 0$  in (19.1). This takes us effectively outside the charge density, and contemporaneously, brings us into a region where the  $-1/r$  potential now applies just as much as does the charge density  $\partial_r(hP_0)=0$ . So the quantum field equation (17.5) which becomes (19.5) when applied to (19.1) inherently takes account of all of this, by causing the potential to deviate from  $-1/r$  where the probability density for the source charge is measurably non-zero, and by causing the potential to return to  $-1/r$  where the probability density for the source charge becomes zero – or to be precise – where our measuring instrumentation is no longer able to detect the extremely tiny extent to which  $\partial_r(hP_0)$  is not zero outside of a few standard deviations from  $r = 0$ .

Given all of this, let us make several postulates, leading to a possible experimental test to confirm the ever-so-slight modification to the Coulomb potential as very close quarters, as predicted by Figure 19 and highlighted in Figure 21. First, just as there is a  $\Lambda_{QCD}$  associated with the QCD strong interaction, we postulate some  $\Lambda_{QED}$  associated with the QED electromagnetic interaction. There will then be a related radial length  $r_{\Lambda QED} = hc / \Lambda_{QED}$  via the deBroglie relation. Either  $r_{\Lambda QED} = 0$ , or  $r_{\Lambda QED}$  = “very small but finite.” Because the former requires a *physical* Dirac delta for the probability density, and we take the view that nature entertains no such singularities as a matter of physics, we shall now adopt the *first postulate* that:

$$r_{\Lambda QED} = \text{very small but finite} . \quad (19.15)$$

Next, if we adopt this first postulate, then the question becomes, what is the precise magnitude of  $r_{\Lambda QED}$ ? Either  $r_{\Lambda QED} = r_{\Lambda QCD}$ , or  $r_{\Lambda QED} \neq r_{\Lambda QCD}$ . That is, either the cutoff / standard deviation length for QED is the same as the one for QCD, or it is different. If it is different, then there is some heretofore unknown  $\Lambda_{QED}$  which represents yet another fundamental constant of nature. But nature is economical, and one should consider the prospect that  $r_{\Lambda QED} = r_{\Lambda QCD}$ . So, we now ask, would  $r_{\Lambda QED} = r_{\Lambda QCD}$  make sense, or at least, would this not run into any apparent contradiction with known theory and data?

What we learn from Figure 6 is that a confining potential appears when the coupled probability density  $\partial_r(hP_0) = \Lambda_{QCD} = 1/r_{\Lambda QCD} = \text{constant}$ . What we learn from Figures 19 through 21 is that a very-close-to  $-1/r$  potential emerges when the coupled probability density is the Gaussian  $\partial_r(hP_0) \equiv 2/r_\Lambda \sqrt{\pi} \exp(-r^2/r_\Lambda^2)$  of (19.1), where  $r_\Lambda$  may or may not be synonymous with  $r_{\Lambda QCD}$ , which is question under consideration at this moment. And what we learn from (16.1) and (14.31) and (19.4) is that this very-close-to  $-1/r$  potential will become the precise  $E = -1/4\pi r$  potential of Coulomb when the Gaussian coupled probability density becomes a Dirac delta. So when one boils this all down, we see that in the context of non-linear

quantum field theory, the difference between a confining potential and a  $-1/r$  potential is the difference between a constant coupled probability density in which  $r_{\Lambda QCD}$  is a sharp cutoff, and a Gaussian coupled probability density in which  $r_{\Lambda QED}$  is the standard deviation of the Gaussian rather than a sharp cutoff. But there is no apparent contradiction in having the same  $r_{\Lambda QED} = r_{\Lambda QCD}$  be a cutoff for  $\partial_r(hP_0)$  in QCD, and the standard deviation for a Gaussian  $\partial_r(hP_0)$  in QED. Furthermore, while QED of course applies to the electrodynamic interactions of electrons, it also applies to the electrodynamic interactions of quarks which have  $Q = +\frac{2}{3}, -\frac{1}{3}$  for the up and down quark respectively. And for quarks which are confined within  $r_{\Lambda QCD}$  or some radial length close to this, it would certainly make some sense if  $r_{\Lambda QED} = r_{\Lambda QCD}$ . Then, once this is done, there would be no reason to have an exception for the electron, because electrodynamic interactions are electrodynamic interactions, whether the fermions involved are quarks or leptons. So, we shall now adopt the *second postulate* that:

$$r_{\Lambda QED} = r_{\Lambda QCD}, \text{ i.e., } \Lambda_{QED} = \Lambda_{QCD}. \quad (19.16)$$

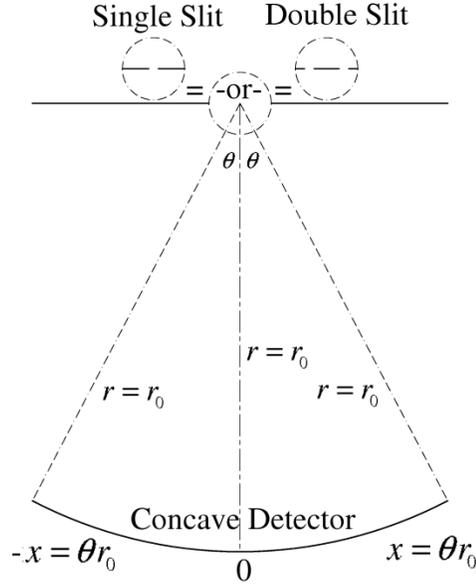
If this second postulate – which we now propose to empirically test – is true, then although  $\Lambda_{QCD}$  comes onto the physics radar and is derived when studying strong interactions, it also plays a role in QED by establishing the standard deviation  $\sigma = r_{\Lambda QCD}$  for the Gaussian coupled probability (19.1), which we take in lieu of  $\delta_{r_\Lambda}(r)$  with  $r_\Lambda = 0$  in (19.4) to be the coupled probability density for QED by the first postulate (19.15). Then, (19.10) which is graphed out in Figure 19 predicts a very slight modification to Coulomb's law centered within one standard deviation  $\sigma = r_{\Lambda QCD}$  of  $r_{\Lambda QCD} / \sqrt{2} = 1.54F$ .

This can be empirically tested if one can devise an experiment, or find within existing experimental data, the up to 1.5% modification to the Coulomb potential shown in Figure 21 close to a 1.54 Fermi separation. Keeping in mind that the Bohr radius  $a_0 = 1/m_e = 5.29 \times 10^4 F$ , we see that this is not a difference that would ever be detectable on the atomic scale. This can only be detected when two charged fermions are brought together with separations on the nuclear scale. It would seem to be quite a challenge for multiple reasons to ever get two electrons compressed to a 1.54 Fermi separation. Consequently, it appears that the best way to detect this predicted parts-per-million modification to the Coulomb potential – which is rooted non-linear quantum field theory – is by studying the electromagnetic interactions of quarks. The challenge for this sort of experiment is that although quarks will naturally have the requisite separation on the order of 1 Fermi which is required to test this, they are confined. So it would be necessary to in some way to be able to study the electromagnetic interactions of quarks within the nucleus and see if their Coulomb potential varies from, and is lower than, what is expected from a strict  $E = -1/4\pi r$  potential by about 1.5% at an approximate 1.54 Fermi separation.

## 20. Single and Double Slit Probability Densities and Guiding Potentials in Non-Linear Quantum Field Theory

Richard Feynman referred to the double slit experiment as “a phenomenon which is impossible . . . to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the *only* mystery [of quantum mechanics].” [30] And indeed, the double slit experiment was recently ranked at #1 among “Science's 10 Most Beautiful Physics Experiments” at [31]. Having reviewed how to apply the quantum field equation (15.42) which in spherical coordinates is (17.2) to the several examples of a constant probability field (section 16), a constant probability density (sections 17 and 18) and a Gaussian probability density including a Dirac delta (section 19), it is time to turn our attention to the single and double slit experiments. When viewed in light of, e.g., equation (17.2) which relates the quantum probability density of a source or sink to a quantum potential  $E_1$  in the first recursive order of non-linear quantum field theory, the single and double slit experiments teach us some very deep and previously unseen insights into the nature of quantum reality.

We start with equation (17.4) in spherical coordinates for the probability density  $\partial_{r,\theta,\phi}(hP_0)$  in Abelian gauge theory to study both single and double slit diffraction in the context of non-linear quantum field theory. Figure 22 below is a schematic illustration of the envisioned simple experimental configuration upon which we shall base the discussion to follow. We use this Figure 22 for either of a single or a double slit experiment as illustrated. We regard the slits as elongated rectangular slits, with the elongation normal to the page. We assume that the radial distance  $r$  from the slit(s) to the detector is *much greater than* the size and spacing of the slits, and we show a circularly-concave detector so that each locale on the detector is at substantially the same constant distance  $r = r_0 = \text{constant}$  from the detector. We regard  $r$  to be an observable, macroscopic separation which could be conveniently measured, for example, in meters or centimeters. We denote the center of the detector by  $x = 0$ , so that the circumferential length from there to any other locale on the detector will be  $x = \theta r_0$  in radian measure. So for single slit diffraction the intensity will be vary in approximate proportion to  $\text{sinc}^2 x = \sin^2 x / x^2$  in accordance with the formulation of Fraunhofer diffraction (note:  $\text{sinc}(x) \equiv \sin(x) / x$ ), while for double slit diffraction the intensity in the central region may be approximated by  $\cos^2 x$ , and overall the single slit envelop contains this so the complete double-slit diffraction intensity varies with the product  $\text{sinc}^2 x \cos^2 x = (\sin x \cos x / x)^2$ . It is also noted that the  $\text{sinc}^2$  function is the Fourier transform of the triangular function, in other words, that if  $f(x) = \text{sinc}^2(ax)$ , then  $\hat{f}(p) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x p} dx = (1/|a|) \text{tri}(p/a)$ .



**Figure 22: Schematic Illustration of Single and / or Double Slit Diffraction**

These two diffraction configurations are of great interest for many reasons. But they are of particular interest here because the intensities  $\propto \text{sinc}^2(x)$  for single slit and  $\propto \text{sinc}^2 x \cos^2 x$  for double slit are *synonymous* with the probability densities  $\partial_{r,\theta,\phi}(hP_0)$ . That is, when we observe the intensity spread for photons or electrons or any other field quanta striking the detector, we are *directly observing* the probability densities  $\partial_{r,\theta,\phi}(hP_0)$  which appear in (17.4). Thus, it becomes possible to via (17.4) derive a commensurate *quantum potential*  $E_1$  in the first recursive order of non-linear quantum field theory. So our goal in this section is to do exactly that for both single and double slit diffraction, and then to use these results together with those of the last few sections to arrive at a better understanding of quantum field theory and the role of probability in quantum field theory, in general.

Now, to keep things simple, let us set  $r = \text{constant}$  so as to represent the fact that all points on the detector are at approximately the same distance from the slit(s). Even for a flat detector, so long as the size and spacing of the slits is much less than  $r$ , we can approximate the locales near the center of the detector as being approximately equidistant from the slit(s). So, if we situate the slit(s) in Figure 22 at  $r = 0$ , and the detector at  $r = r_0$  where  $r_0$  is *much larger* than any either the slit width or the slit spacing, we may regard  $r = r_0 = \text{constant}$  in the mathematical equations. So we start with (17.4) and do the following: First, we set  $r \equiv r_0$  and then move  $r_0$  from the  $1/r$  portion of the equation over to the left side. Now, both sides of this equation are dimensionless. Here, we are no longer looking at the probability density  $\partial_{r,\theta,\phi}(hP_0)$  and the potential  $E_1$  as a function of  $r$ , because the probability density is projected right onto the detector itself, and all points of the detector are assumed to be approximately equidistant  $r = r_0$  from the slit(s). Since there is no  $r$ -variation in the probabilities striking the detector in Figure 22, the radial  $\partial_r(hP_0) = 0$  along the detector. Similarly, because the elongated slit(s) are aligned

vertically perpendicularly to the page, we know that for the detector regions of interest which subtend a very small  $\phi$ , we will also have  $\partial_\phi(hP_0)=0$ . Finally, we make use of the deBroglie relation  $mc^2 = \hbar c / r$  to define an energy parameter  $\Lambda_0 \equiv \hbar c / r_0$  which is associated with the length  $r_0$ . Specifically, we define  $\Lambda_0 \equiv \hbar c / r_0$ , or in natural units and inverted, simply  $r_0 = 1 / \Lambda_0$ . All of this allows us to specialize (17.4) for the configuration of Figure 22 to (compare (17.5) for the radially-dependent  $E_1$ ):

$$\frac{E_1}{\Lambda_0} = -\frac{1}{4\pi} \frac{1}{9} \left[ 6 + \exp(i\partial_\theta(hP_0)) + 2 \exp\left(i\frac{1}{2}\partial_\theta(hP_0)\right) \cdot \cosh\left(\frac{\sqrt{3}}{2}\partial_\theta(hP_0)\right) \right]. \quad (20.1)$$

So now let's turn to the single and double slit probability densities  $\partial_\theta(hP_0)$ .

As pointed out earlier, for single slit Fraunhofer diffraction in which the detector is at a distance from the slit which greatly exceeds the slit size, the intensity, which is proportional to the probability density  $\partial_x(hP_0)$ , will vary with  $\text{sinc}^2(x) = \sin^2 x / x^2$ . Because  $x = \theta r_0$  in Figure 22,  $x \propto \theta$ , and therefore  $\partial_\theta(hP_0)$  in (20.2) will also vary with  $\text{sinc}^2(\theta)$ . Now, we need to pin this down more exactly. Because  $\partial_\theta(hP_0)$  is a probability density, it is necessary that  $\int_{-\infty}^{\infty} \partial_\theta(hP_0) d\theta = 1$ . Therefore, let us assign:

$$\partial_\theta(hP_0) = \frac{A}{2\pi} \text{sinc}^2\left(\frac{A}{2}\theta\right), \quad (20.2)$$

where  $A$  is some constant number that determines the width of the diffraction pattern. We choose this definition because it correctly normalizes to:

$$\int_{-\infty}^{\infty} \partial_\theta(hP_0) d\theta = \int_{-\infty}^{\infty} \frac{A}{2\pi} \text{sinc}^2\left(A\frac{\theta}{2}\right) d\theta = 1, \quad (20.3)$$

and also because  $A$  enables us to vary the width of the diffraction pattern. Physically, the wavelength  $\lambda$  of the signal being diffracted in relation to the width  $a$  of the slit will determine the value of  $A$ . For this discussion, it is not necessary to formalize this relation, though for a rough qualitative correspondence, we note that the diffraction width will narrow as  $A$  increases, and will also narrow with decreasing  $\lambda$  and increasing  $a$ , so that  $A \approx a / \lambda$ .

To simplify the mathematical and graphical development, let us now transform  $\theta$  to a different dimensionless coordinate  $X$ , defined according to:

$$A\frac{\theta}{2} \equiv \pi X. \quad (20.4)$$

Because  $x = \theta r_0$  in Figure 22, this new coordinate  $X$  is related to the  $x$  with length dimension in Figure 22 according to:

$$x = \theta r_0 = \frac{2\pi r_0}{A} X ; \text{ alternatively, } X = \frac{A\theta}{2\pi} = \frac{Ax}{2\pi r_0}. \quad (20.5)$$

Making use of (20.4) as well as  $\partial_\theta = (A/2\pi)\partial_x$  deduced from (20.4) as well as (20.5), we may rewrite (20.2) as:

$$\partial_\theta (hP_0) = \frac{A}{2\pi} \partial_x (hP_0) = \frac{A}{2\pi} \text{sinc}^2\left(A\frac{\theta}{2}\right) = \frac{A}{2\pi} \text{sinc}^2\left(A\frac{x}{2r_0}\right) = \frac{A}{2\pi} \text{sinc}^2(\pi X), \quad (20.6)$$

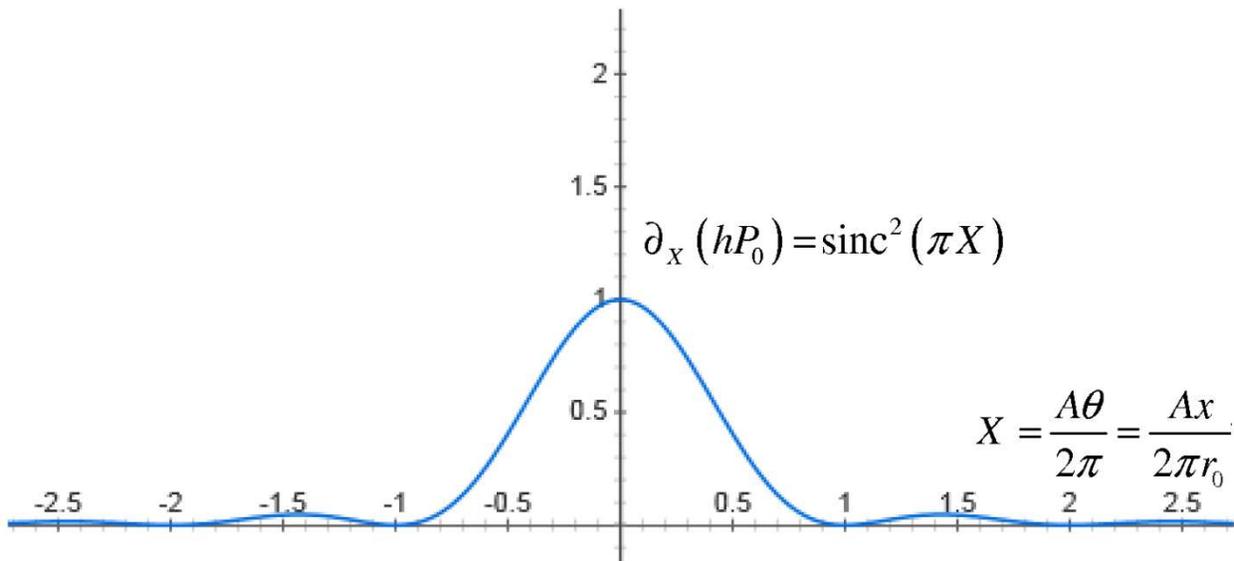
which contains the somewhat less-cluttered equation:

$$\partial_x (hP_0) = \text{sinc}^2(\pi X) \quad (20.7)$$

with the very simple normalization:

$$\int_{-\infty}^{\infty} \partial_x (hP_0) dX = \int_{-\infty}^{\infty} \text{sinc}^2(\pi X) dX = 1 \quad (20.8)$$

It is for these reasons that  $X$  presents an attractive coordinate choice for carrying out the mathematical and graphical development. Indeed, including (20.5), Figure 23 below shows the well-known configuration of the single slit diffraction probability density given by (20.7). We see that another attractive feature of the  $X$  coordinate is that minima occur at integer values of  $X$ .



**Figure 23: Probability Density / Intensity for Single-Slit Diffraction**

For a double slit, as we noted at the outset, what is now  $\partial_x(hP_0) = \text{sinc}^2(\pi X)$  for a single slit is the envelope for what we now take to be a  $\cos^2(B\pi X)$  oscillation within the envelope, where  $B$  is an independent parameter from  $A$  in the definition  $X = A\theta / 2\pi$  in (20.4). Mathematically, the definite integral  $\int_{-\infty}^{\infty} 2\text{sinc}^2(\pi X)\cos^2(B\pi X)dX = 1$  *independently* of the value of  $B$ . Therefore, we may assign

$$\partial_x(hP_0) = 2\text{sinc}^2(\pi X)\cos^2(B\pi X) \quad (20.9)$$

to the probability density  $\partial_x(hP_0)$ , and thereby be assured that the integral:

$$\int_{-\infty}^{\infty} \partial_x(hP_0)dX = \int_{-\infty}^{\infty} 2\text{sinc}^2(\pi X)\cos^2(B\pi X)dX = 1 \quad (20.10)$$

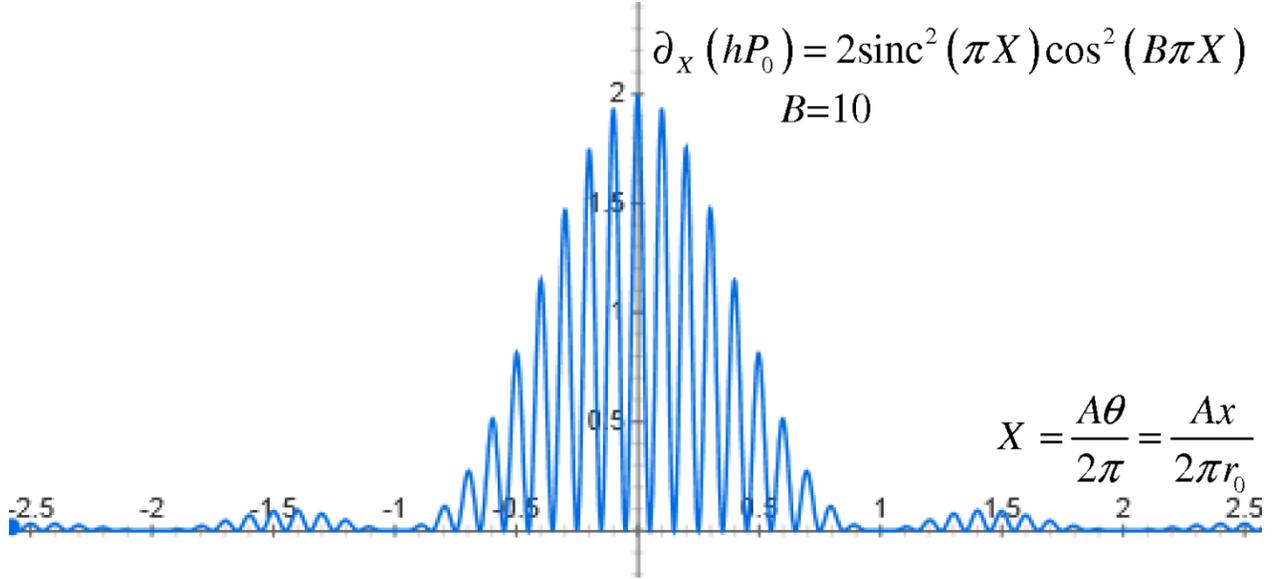
as is required for a normalized probability density. Of course, (20.1) in which we will want to make use of this contains  $\partial_\theta(hP_0)$ . So, using  $\partial_x = (2\pi/A)\partial_\theta$  deduced from (20.4) and well as (20.4) itself and  $\theta = x/r_0$  from Figure 22, we may write (20.9) as:

$$\begin{aligned} \partial_x(hP_0) &= \frac{2\pi}{A}\partial_\theta(hP_0) = 2\text{sinc}^2(\pi X)\cos^2(B\pi X) = 2\text{sinc}^2\left(A\frac{\theta}{2}\right)\cos^2\left(BA\frac{\theta}{2}\right) \\ &= 2\text{sinc}^2\left(A\frac{x}{2r_0}\right)\cos^2\left(BA\frac{x}{2r_0}\right) \end{aligned} \quad (20.11)$$

This contains the equation:

$$\begin{aligned} \partial_\theta(hP_0) &= \frac{A}{2\pi}\partial_x(hP_0) = \frac{A}{\pi}\text{sinc}^2\left(A\frac{\theta}{2}\right)\cos^2\left(BA\frac{\theta}{2}\right) = \frac{A}{\pi}\text{sinc}^2\left(A\frac{x}{2r_0}\right)\cos^2\left(BA\frac{x}{2r_0}\right), \\ &= \frac{A}{\pi}\text{sinc}^2(\pi X)\cos^2(B\pi X) \end{aligned} \quad (20.12)$$

which should be contrasted with its counterpart (20.6) for a single slit. The new parameter  $B$  determines the number of peaks within the overall domain  $0 \leq X \leq 1$ , and indeed, the number of overall peak within any unit spread of  $X$ , e.g., there are  $B$  peaks from  $1 \leq X \leq 2$ , from  $2 \leq X \leq 3$  etc., with the peak at the very center shared between the  $-1 \leq X \leq 0$  and the  $0 \leq X \leq 1$  domains. So, as a concrete example, if we graph (20.9) for  $B=10$ , the curve is as follows:



**Figure 24: Probability Density / Intensity for Double-Slit Diffraction,  $B=10$  Example**

We see that there are in fact  $B = 10$  peaks within each unit of domain, and that the center peak is shared and thus “double-counted” for both  $X < 0$  and  $X > 0$ . Contrasting this with the single slit Figure 23, we see a doubling of the amplitude owing to the coefficient of 2 that is needed in the normalization (20.10) versus the implied coefficient of 1 in (20.8). Physically, this results in a doubling of the overall height of the envelope owing to there now being twice as many slits – two rather than one – through which the diffraction signal may pass. And we directly see how parameter  $B$  affects peak packing within the envelope. The overall spread of the envelope, as noted earlier, is determined by the parameter  $A$  which roughly varies as  $A \approx a / \lambda$  with the slit width  $a$  and the signal wavelength  $\lambda$ . Denoting the spacing between the two slits as  $d$ , because increased  $B$  leads to a proportionate increase in the density with which the peaks are packed within the envelope, we know that both increased  $d$  and decreased  $\lambda$  also increase this packing density. So while it is not essential to the development here, we may also write down the rough correspondence  $B \approx d / \lambda$ .

Now let’s use (20.1) to calculate the potential  $E_1$  in the first recursive order of non-linear quantum field theory. For *single-slit diffraction* we use (20.6) in terms of the mathematically-simplifying  $X$  coordinate in (20.5) to obtain:

$$\frac{E_1}{\Lambda_0} = -\frac{1}{4\pi} \frac{1}{9} \left[ 6 + \exp\left(i \frac{A}{2\pi} \text{sinc}^2(\pi X)\right) + 2 \exp\left(i \frac{A}{4\pi} \text{sinc}^2(\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}A}{4\pi} \text{sinc}^2(\pi X)\right) \right]. \quad (20.13)$$

For *double-slit diffraction* we use (20.12) in (20.1) also in  $X$ , to obtain:

$$\frac{E_1}{\Lambda_0} = -\frac{1}{4\pi} \frac{1}{9} \left[ \begin{aligned} &6 + \exp\left(i \frac{A}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\ &+ 2 \exp\left(i \frac{A}{2\pi} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}A}{2\pi} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \end{aligned} \right]. \quad (20.14)$$

Now as we have done three times previously, contrast (14.35), (17.10) and (19.7), we wish to obtain and then graph the real magnitude  $|E_1| = \pm\sqrt{E_1 * E_1}$  for each of (20.13) and (20.14), using some realistic exemplary values for  $A$  and  $B$ . In Figure 24 we used  $B=10$  for illustration, so to maintain consistency, we continue with  $B=10$ . As to  $A$ , we go back to (20.4) and write this as  $A = (2\pi/\theta)X$ . As we see in both Figures 23 and 24, the coordinate  $X$  is designed among other things to place the first minimum of the diffraction envelope at  $X=1$ , so that the width of the central peak is equal to 2 along the  $X$  axis in the domain from  $-1 \leq X \leq +1$ . So via (20.5) in the form of  $X = A\theta/2\pi$ , over the domain of the central peak we have  $-1 \leq A\theta/2\pi \leq +1$ , or:

$$-\frac{2\pi}{A} \leq \theta \leq +\frac{2\pi}{A}. \quad (20.15)$$

Now,  $\theta$  is the angle illustrated in Figure 22, and  $A$  in the above lends itself to being discussed in degrees rather than radians. First, even if the central peak was to be spread over the entire 180 degrees ( $=\pi$  radians) emanating from the Figure 22 slits, we would have  $-\pi/2 \leq \theta \leq +\pi/2$ . So for anything to even make sense, we *must* have  $A \geq 4$ . For  $A=4$ , (20.15) becomes  $-\pi/2 \leq \theta \leq +\pi/2$ , which represents a full 180 degree spread for the central peak, which we denote by  $\Theta = 180^\circ$ . For  $A=8$  we would have a  $\Theta = 90^\circ$  degree spread for the central peak, 45 degrees on each side of the vertical centerline of Figure 22.  $A=12$  is the parameter for a  $\Theta = 60^\circ$  total spread,  $A=16$  for  $\Theta = 45^\circ$ ,  $A=24$  for a  $\Theta = 30^\circ$  spread. So with  $A=360$  there is a one-degree spread on each side of the centerline totaling  $\Theta = 2^\circ$  of peak spread, and with  $A=720$  there is a total spread of  $\Theta = 1^\circ$  with .5 degrees on either side of the centerline. Thus, with reference to Figure 22, if we use  $\Theta = 2\theta$  to denote the *total* angular spread of the *entire* central peak which sits in the domain  $-1 \leq X \leq +1$  in Figures 14 and 15, we deduce from (20.15) that, in general:

$$A = \frac{720^\circ}{\Theta_{\text{deg}}} = \frac{4\pi}{\Theta_{\text{rad}}}. \quad (20.16)$$

Consequently, to make things more physically meaningful, we now use (20.16) above to replace  $A = 4\pi/\Theta$  with a direct measure of the peak spread angle  $\Theta$  in each of (20.13) and (20.14). So for *single slit diffraction*, also moving  $-4\pi$  to the left side, we now have:

$$-4\pi \frac{E_1}{\Lambda_0} = \frac{2}{3} + \frac{1}{9} \exp\left(i \frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{2}{9} \exp\left(i \frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right), \quad (20.17)$$

while for *double-slit diffraction* we obtain:

$$\begin{aligned}
 -4\pi \frac{E_1}{\Lambda_0} = & \frac{2}{3} + \frac{1}{9} \exp\left(i \frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
 & + \frac{2}{9} \exp\left(i \frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right).
 \end{aligned} \tag{20.18}$$

To obtain the real magnitudes for each of the above, as previously done at (14.35), (17.10) and (19.7), we *define* each of  $a$  and  $b$  according to  $a + bi \equiv -4\pi E_1 / \Lambda_0$ . For the single slit (20.17) this means that:

$$\begin{aligned}
 a = & \frac{2}{3} + \frac{1}{9} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{2}{9} \cos\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right) \\
 b = & \frac{1}{9} \sin\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{2}{9} \sin\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right)
 \end{aligned} \tag{20.19}$$

and therefore that:

$$\begin{aligned}
 |a + bi|^2 = & a^2 + b^2 \\
 = & \frac{4}{9} + \frac{4}{27} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{8}{27} \cos\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right) \\
 & + \frac{1}{81} \cos^2\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{4}{81} \cos^2\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right) \\
 & + \frac{1}{81} \sin^2\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{4}{81} \sin^2\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh^2\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right) \\
 & + \frac{4}{81} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cos\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right) \\
 & + \frac{4}{81} \sin\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) \sin\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \cdot \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right)
 \end{aligned} \tag{20.20}$$

This reduces down to:

$$\begin{aligned}
 & |a + bi|^2 = a^2 + b^2 \\
 & = \frac{37}{81} + \frac{4}{27} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right) \\
 & + \left[ \frac{4}{81} \sin\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) \sin\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \right. \\
 & \left. + \left[ \frac{8}{27} + \frac{4}{81} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) \right] \cos\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \right] \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right)
 \end{aligned} \tag{20.21}$$

So when we use (20.21) in (20.17) such that  $|a + bi| \equiv -4\pi|E_1|/\Lambda_0$ , we now use the square root of the above, which has  $\pm$  roots, obtain the real magnitude for the single-slit potential, namely:

$$-4\pi \frac{|E_1|}{\Lambda_0} = \pm \sqrt{\frac{37}{81} + \frac{4}{27} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \frac{4}{81} \cosh^2\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \left[ \frac{4}{81} \sin\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) \sin\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) + \left[ \frac{8}{27} + \frac{4}{81} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X)\right) \right] \cos\left(\frac{1}{\Theta} \operatorname{sinc}^2(\pi X)\right) \right] \cosh\left(\frac{\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X)\right)} \tag{20.22}$$

As to the double-slit (20.18), defining  $a + bi \equiv -4\pi E_1/\Lambda_0$ , we start with:

$$\begin{aligned}
 a & = \frac{2}{3} + \frac{1}{9} \cos\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
 & + \frac{2}{9} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
 b & = \frac{1}{9} \sin\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
 & + \frac{2}{9} \sin\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right)
 \end{aligned} \tag{20.23}$$

This means that:

$$\begin{aligned}
& |a + bi|^2 = a^2 + b^2 \\
& = \frac{4}{9} + \frac{4}{27} \cos\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& + \frac{8}{27} \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& + \frac{1}{81} \cos^2\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) + \frac{1}{81} \sin^2\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& + \frac{4}{81} \cos^2\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh^2\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& + \frac{4}{81} \sin^2\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh^2\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& + \frac{4}{81} \cos\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& + \frac{4}{81} \sin\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \sin\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \cdot \cosh\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right), \quad (20.24)
\end{aligned}$$

which reduces to:

$$\begin{aligned}
& |a + bi|^2 = a^2 + b^2 \\
& = \frac{37}{81} + \frac{4}{27} \cos\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) + \frac{4}{81} \cosh^2\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& + \left[ \frac{4}{81} \sin\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \sin\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \right] \\
& + \left[ \frac{8}{27} + \frac{4}{81} \cos\left(\frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \right] \cos\left(\frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \\
& \times \cosh\left(\frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X)\right) \quad (20.25)
\end{aligned}$$

So when we use (20.25) in (20.18) such that  $|a + bi| \equiv -4\pi|E_1|/\Lambda_0$ , we use the square roots of the above obtain real magnitude for the double-slit potential, namely:

$$\begin{aligned}
 & -4\pi \frac{|E_1|}{\Lambda_0} \\
 & = \pm \sqrt{ \left[ \frac{37}{81} + \frac{4}{27} \cos \left( \frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X) \right) + \frac{4}{81} \cosh^2 \left( \frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X) \right) \right. } \\
 & \quad \left. + \left[ \frac{4}{81} \sin \left( \frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X) \right) \sin \left( \frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X) \right) \right. \right. \\
 & \quad \left. \left. + \left[ \frac{8}{27} + \frac{4}{81} \cos \left( \frac{4}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X) \right) \right] \cos \left( \frac{2}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X) \right) \right] \right. } \\
 & \quad \left. \times \cosh \left( \frac{2\sqrt{3}}{\Theta} \operatorname{sinc}^2(\pi X) \cos^2(B\pi X) \right) \right]. \quad (20.26)
 \end{aligned}$$

So let us now graph the respective single and double slit potentials (20.22) and (20.26). In both cases, we have some flexibility to choose the angle  $\Theta$  which was defined so as to measure the overall spread of the central peaks in Figures 23 and 24 about the centerline of Figure 22. Simply for illustration, let us continue to choose  $\Theta = 30^\circ = \pi/6$  and use this for both the single and double slit potentials, so that the potentials we obtain will match up with the probability densities graphed in Figures 23 and 24. So for the *single slit* potential of (20.22), using  $\Theta = 30^\circ = \pi/6$ , which is  $A = 4\pi/\Theta = 24$  when expressed in terms of the original envelope spread parameter  $A$ , we now obtain:

$$\begin{aligned}
 & -4\pi \frac{|E_1|}{\Lambda_0} = \pm \sqrt{ \left[ \frac{37}{81} + \frac{4}{27} \cos \left( \frac{12}{\pi} \operatorname{sinc}^2(\pi X) \right) + \frac{4}{81} \cosh^2 \left( \frac{6\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X) \right) \right. } \\
 & \quad \left. + \left[ \frac{4}{81} \sin \left( \frac{12}{\pi} \operatorname{sinc}^2(\pi X) \right) \sin \left( \frac{6}{\pi} \operatorname{sinc}^2(\pi X) \right) \right. \right. \\
 & \quad \left. \left. + \left[ \frac{8}{27} + \frac{4}{81} \cos \left( \frac{12}{\pi} \operatorname{sinc}^2(\pi X) \right) \right] \cos \left( \frac{6}{\pi} \operatorname{sinc}^2(\pi X) \right) \right] \cosh \left( \frac{6\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X) \right) \right]. \quad (20.27)
 \end{aligned}$$

As to the double slit potential of (20.26) we continue to use  $\Theta = 30^\circ = \pi/6$  for consistent illustration. At the same time, we choose  $B = 10$  also for illustration, which means that we are choosing to have 10 peaks within each unit of domain including the double-counting of the very center peak at  $X = 0$ . Using  $\Theta = 30^\circ = \pi/6$  and  $B = 10$  in (20.26), for the *double-slit* we obtain:

$$-4\pi \frac{|E_1|}{\Lambda_0} = \pm \sqrt{ \left[ \frac{37}{81} + \frac{4}{27} \cos \left( \frac{24}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X) \right) + \frac{4}{81} \cosh^2 \left( \frac{12\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X) \right) \right. } \quad (20.28)$$

$$+ \left[ \frac{4}{81} \sin \left( \frac{24}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X) \right) \sin \left( \frac{12}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X) \right) \right.$$

$$\left. + \left[ \frac{8}{27} + \frac{4}{81} \cos \left( \frac{24}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X) \right) \right] \cos \left( \frac{12}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X) \right) \right]$$

$$\times \cosh \left( \frac{12\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X) \right)$$

Now, before we graph the illustrative  $\Theta = 30^\circ = \pi/6$  single slit potential (20.27) and the illustrative  $\Theta = 30^\circ = \pi/6$ ,  $B = 10$  double slit potential (20.28) there are two points which we first need to discuss. First, consulting with Figures 23 and 24, the probability peaks are the places to which the maximum number of photons, electrons, etc. congregate during single or double slit experiments. How do we know this? We observe it in our experiments, directly. By least action principles, all of the *maximum probability* locations in Figures 23 and 24 must therefore correspond to the *minimum energy* locations in the potential. How do we know this? Because particles moving in a potential will be drawn toward and thus congregate near the minima in the potential. Second, in general, the magnitude of a spatially-varying potential such as  $E_1$  is not itself a physical observable. A voltage, for example, has no independent meaning. This is also one of the consequences of gauge symmetry. What does have observable meaning is a *difference* of potential, that is, a *voltage drop* between point A and point B.

Keeping both of these points in mind, first, it turns out that when one graphs (20.27) and (20.28), the choice of the positive root  $\pm$  when taken with the  $-$  sign on the left side  $-4\pi|E_1|/\Lambda_0$  correctly matches up the maximum probability locations to the minimum potential locations. In other words, we get the correct least-action mapping of probability to potential by choosing the positive roots in (20.27) and (20.28), which maintains the overall negative sign. Second, it turns out that  $4\pi|E_1|/\Lambda_0 \approx -1$  throughout most of the curve, and only drops definitively below -1 at the probability density peaks closest to the center of the  $X$  axis. Because it is the relative difference between two potentials, i.e., the voltage drop which is observable, we will also add 1 to each of the inverted potentials, to make zero the “ground” potential, and to thereby have the probability density peaks correspond to a “voltage drop” below zero. So, choosing the positive roots to maintain the overall negative sign, and then adding 1 to set zero to ground, we rewrite the *single-slit* (20.27) and the *double-slit* (20.28), respectively, as:

$$1 + 4\pi \frac{|E_1|}{\Lambda_0} = 1 - \sqrt{\left[ \frac{37}{81} + \frac{4}{27} \cos\left(\frac{12}{\pi} \operatorname{sinc}^2(\pi X)\right) + \frac{4}{81} \cosh^2\left(\frac{6\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X)\right) \right.} \quad (20.29)$$

$$\left. + \left[ \frac{4}{81} \sin\left(\frac{12}{\pi} \operatorname{sinc}^2(\pi X)\right) \sin\left(\frac{6}{\pi} \operatorname{sinc}^2(\pi X)\right) \right] \cosh\left(\frac{6\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X)\right) \right.}$$

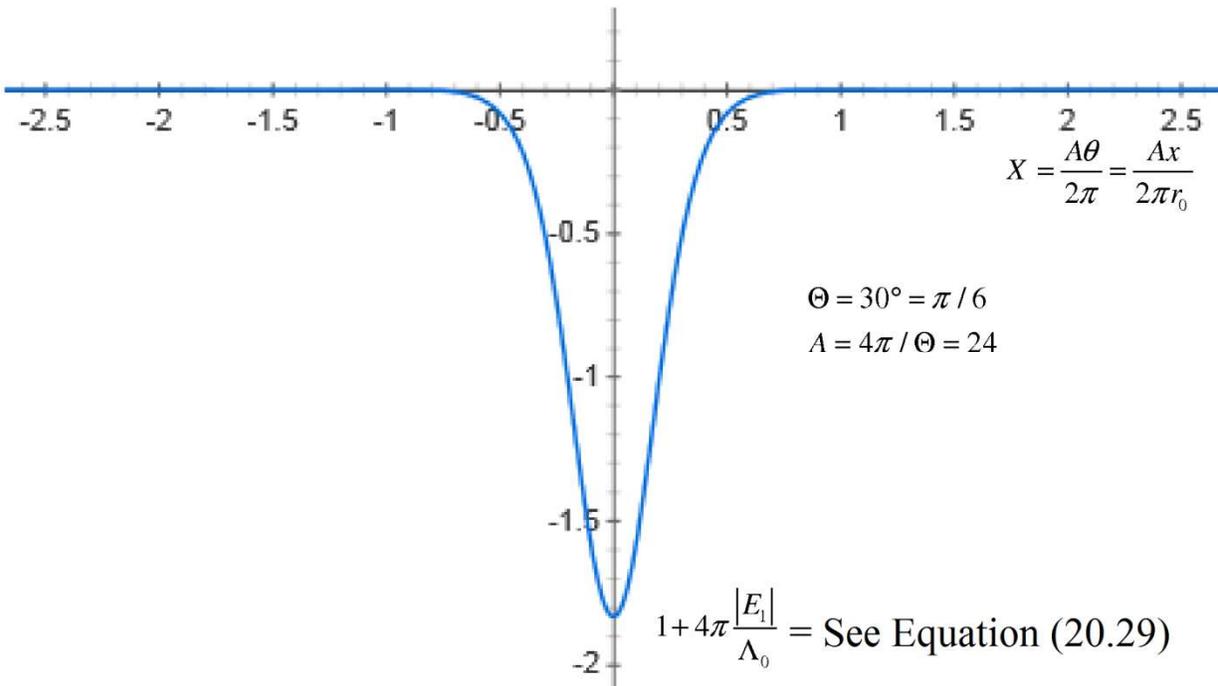
$$\left. + \left[ \frac{8}{27} + \frac{4}{81} \cos\left(\frac{12}{\pi} \operatorname{sinc}^2(\pi X)\right) \right] \cos\left(\frac{6}{\pi} \operatorname{sinc}^2(\pi X)\right) \right] \cosh\left(\frac{6\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X)\right)$$

$$1 + 4\pi \frac{|E_1|}{\Lambda_0} = - \sqrt{\left[ \frac{37}{81} + \frac{4}{27} \cos\left(\frac{24}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X)\right) + \frac{4}{81} \cosh^2\left(\frac{12\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X)\right) \right.} \quad (20.30)$$

$$\left. + \left[ \frac{4}{81} \sin\left(\frac{24}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X)\right) \sin\left(\frac{12}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X)\right) \right] \right.}$$

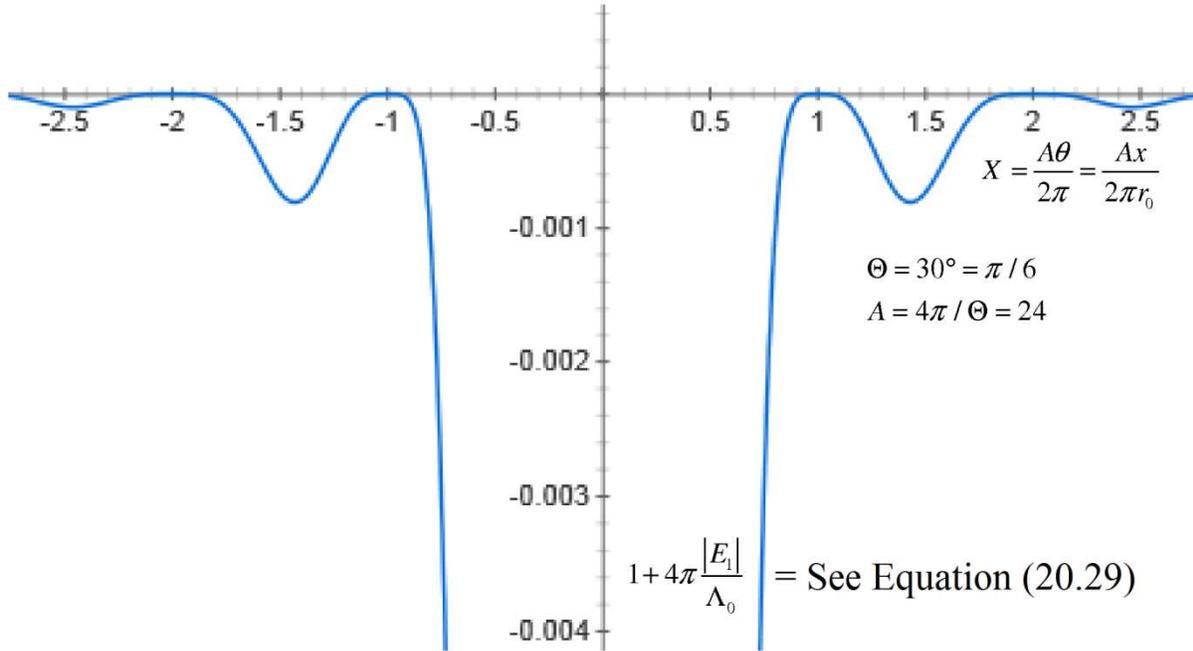
$$\left. + \left[ \frac{8}{27} + \frac{4}{81} \cos\left(\frac{24}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X)\right) \right] \cos\left(\frac{12}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X)\right) \right] \cosh\left(\frac{12\sqrt{3}}{\pi} \operatorname{sinc}^2(\pi X) \cos^2(10\pi X)\right)$$

Graphing the single slit potential of (20.29) for which  $\Theta = 30^\circ = \pi / 6$ , which corresponds directly to the single slit probability density graphed in Figure 23, we have:



**Figure 25: Single-Slit Potential for  $\Theta = 30^\circ = \pi / 6$ , Equation (20.29)**

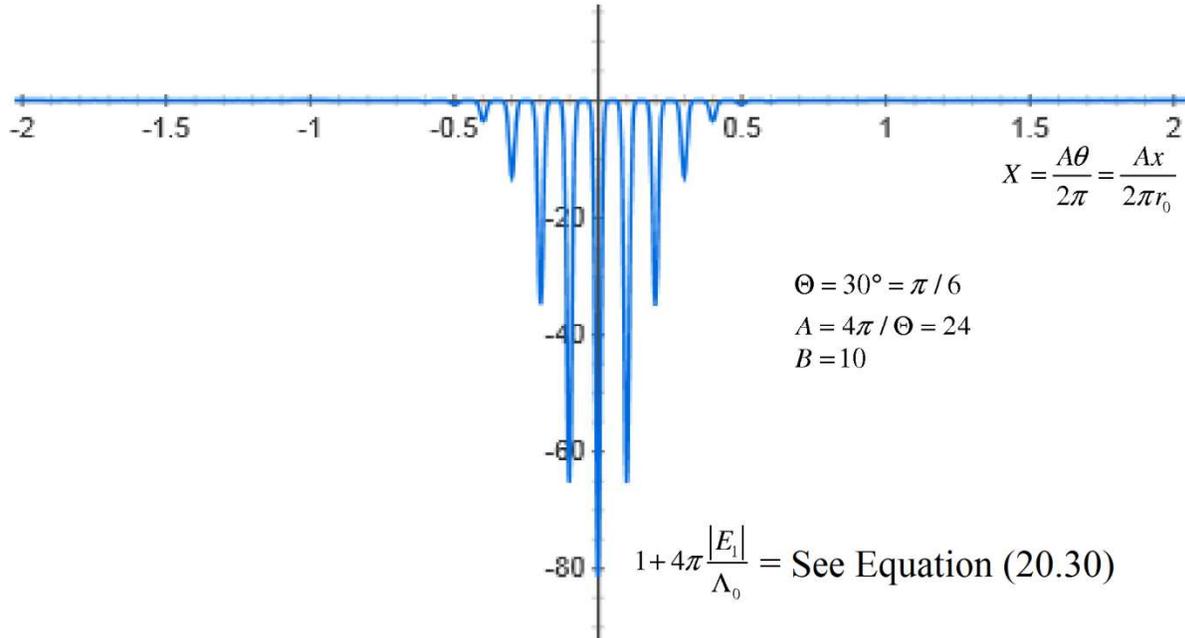
If we wish to see the behaviors for the envelope in the domain  $|X| > 1$  we need to zoom out from, i.e., widen the aspect ratio for the vertical axis. Doing this, we graph the exact same function shown in Figure 25, but from a wider vertical view:



**Figure 26: Single-Slit Potential for  $\Theta = 30^\circ = \pi / 6$ , Equation (20.29), Wide View**

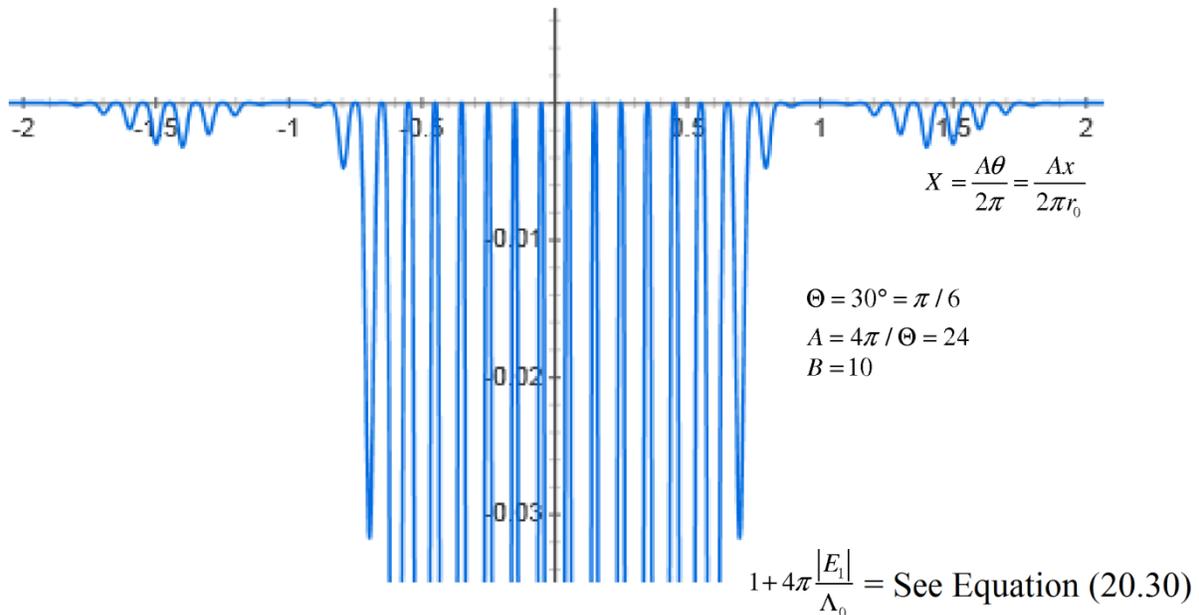
So we see how the potential draws the field quanta of Figure 23 toward their maximum probability positions via least action, which we shall discuss in further depth momentarily. Figure 25 shows how the definitive minimum of the potential is at  $X = 0$ , and Figure 26 shows the secondary minima at just shy of  $|X|=1.5$  and  $|X|=2.5$ , matching up with all of the probability density maxima in Figure 23.

Now let's graph the double-slit potential of (20.30) for which  $\Theta = 30^\circ = \pi / 6$  and  $B = 10$ . This corresponds directly to the probability density graphed in Figure 24. Here we have the potential corresponding directly to the double-slit probability density graphed in Figure 24:



**Figure 27: Double-Slit Potential for  $\Theta = 30^\circ = \pi / 6$ ,  $B = 10$ , Equation (20.30)**

As with Figure 25, the potential minima away from  $X = 0$  are much smaller in magnitude than the minima near the center. So we again draw a wider view of the above for the vertical axis.



**Figure 28: Double-Slit Potential for  $\Theta = 30^\circ = \pi / 6$ ,  $B = 10$ , Equation (20.30), Wide View**

Here too, the potential draws the field quanta of Figure 24 toward their maximum probability positions via least action, which we shall discuss in further depth momentarily. Figure 27 shows a series of dominant minima near  $X = 0$  which reduce as  $|X|$  increases, and Figure 28 shows the

smaller, further minima under magnification. Here too, these potential minima match up with all of the probability density maxima in Figure 24.

Now, let us use some actual physical assumptions to calculate the magnitude of the voltage drops associated with the single and double slit experiments. The single and double slit voltage drops in Figures 25 through 28 are expressed in terms of  $1+4\pi|E_1|/\Lambda_0$ , where  $\Lambda_0$  has dimensions of energy and is defined prior to (20.1) by  $\Lambda_0 \equiv \hbar c / r_0$ , where  $r_0$  is defined in Figure 22 as the approximate distance from the slit(s) to the detector, which includes macroscopic distances. So let's now choose a macroscopic distance  $r_0$ , and to keep things simple, let's choose  $r = 1m$ . Once we make this final illustrative physical assumption (the other illustrative assumptions were  $\Theta = 30^\circ = \pi/6$  and  $B = 10$ ), we need to calculate the deBroglie energy  $\Lambda_{r=1} = \hbar c / r_{=1}$  associated with one meter. As we laid out prior to (14.42), one can use the shortcut  $1F = 5.067\,731\,163\text{GeV}^{-1}$  in  $\hbar = c = 1$  natural units without ever having to use  $\hbar$  and  $c$  explicitly in the calculation. Because  $1F = 10^{-15}m$  and  $1\text{GeV} = 10^9 eV$ , the natural units shortcut can be written as  $10^{-15}m = 5.067\,731\,163 \times 10^{-7} eV^{-1}$ . Or  $1m = 5.067\,731\,163 \times 10^6 eV^{-1}$ . Or finally,  $1m = 1/1.973\,269\,631 \times 10^{-7} eV$ . This means that the energy equivalent of  $r_0 = 1$  meter, is:

$$\Lambda_0 = \hbar c / r_0 = 1.973\,269\,631 \times 10^{-7} eV. \quad (20.31)$$

So, now referring to Figure 25, close inspection reveals that the minimum at  $X = 0$  has the four-digit value  $1+4\pi|E_1|_{\min} / \Lambda_0 = -1.8309$ , or  $|E_1|_{\min} = -2.8309\Lambda_0 / 4\pi$ . At the same time, wherever  $1+4\pi|E_1|_{\max} / \Lambda_0 = 0$  at the potential maxima which are the probability density minima, we have  $|E_1|_{\max} = -\Lambda_0 / 4\pi$ . So because it is the *drop in potential* which is observable, and not the potential itself, we can calculate that the single-slit *voltage drop* at  $X = 0$  is:

$$|E_1|_{\min} - |E_1|_{\max} = -1.8309\Lambda_0 / 4\pi = -2.8750 \times 10^{-8} eV = -4.6063 \times 10^{-27} J, \quad (20.32)$$

where we also include the eV to Joules unit conversion  $1eV = 1.6022 \times 10^{-19} J$ . At the same time, referring to Figure 27, close inspection reveals that the minimum at  $X = 0$  has the four-digit value  $1+4\pi|E_1|_{\min} / \Lambda_0 = -81.3854$  or  $|E_1|_{\min} = -82.3854\Lambda_0 / 4\pi$ . Once again the potential maxima are at  $|E_1|_{\max} = -\Lambda_0 / 4\pi$ . So here, the double slit voltage drop at  $X = 0$  is:

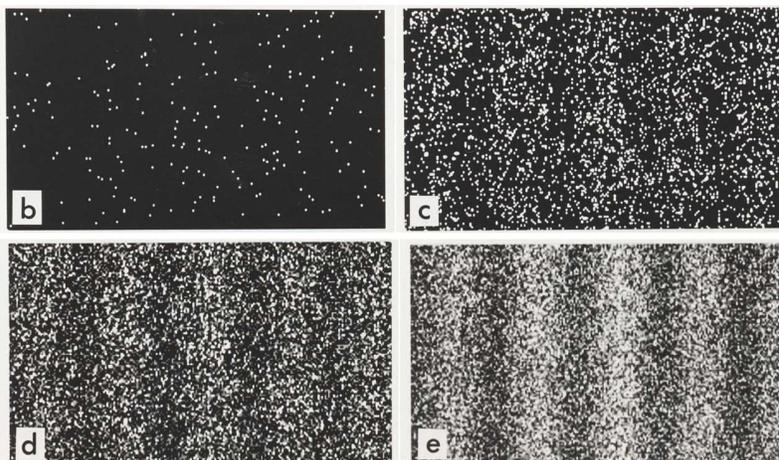
$$|E_1|_{\min} - |E_1|_{\max} = -81.3854\Lambda_0 / 4\pi = -1.2780 \times 10^{-6} eV = -2.0476 \times 10^{-25} J. \quad (20.33)$$

These are the *largest* voltage drops which occur for a slit experiment at 1 meter. For the smaller peaks away from  $X = 0$ , the voltage drops are smaller, as seen in Figures 25 through 28. Because the absorption of a field quantum on a detector inherently creates a photovoltaic reaction, it may well be possible to detect these predicted voltage drops by measuring the photovoltaic drop between peaks and troughs in the probability density while field quanta are

accumulating at the detector, keeping in mind that  $|E_1|$  is the potential at the approximation of the first recursive order in non-linear quantum field theory. Due to the isomorphic  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$  mapping between the probability density for a given  $\Theta$  and  $B$  over a range of  $q$ ,  $a$ ,  $d$  and  $E$  input parameters, another confirmation is to show that the photovoltaic behaviors at the detector are dependent solely on the values of  $\Theta$  and  $B$  for the detected probability distribution, and are otherwise invariant with respect to  $q$ ,  $a$ ,  $d$  and  $E$ .

It is worth noting that the maximum voltage drop (20.33) for the double slit experiment is  $44.4510 = 81.3854 / 1.8309$  times as large as the maximum drop for (20.32) for the single slit experiment. Part of this – like the doubling of the probability density peak from Figure 23 to Figure 24 – is accounted for by having two slits rather than one. But most of this is accounted for because the total voltage drop when integrated over the entire domain, i.e., the total area between the horizontal axis and the potential curve, must also be twice as large for a double slit potential as for a like-slit, like-signal-energy single-slit potential. Because the double slit potential drops comprise a series of very narrow spikes seen in Figure 27 while the single slit potential drop seen in Figure 25 does not lose total area due to the sharp spikes, the lost area under the horizontal axis of the double slit potential needs to be recouped, and this occurs by the distinctively larger voltage drop which occurs for the double slit potential in the narrow spikes where there is any drop at all.

Now we return to Richard Feynman who we quoted at the start of this section to see what the foregoing teaches us about one of the great mysteries of quantum theory, namely, the interference pattern of Figure 24 which we obtain in the double slit experiment, notwithstanding that photons reach and hit the detector one at a time. This goes to the heart of wave-particle duality, because the interference pattern of Figure 24 can be and was fully accounted for for many years by a wave model of light with constructive and destructive interference yielding the peaks and troughs in Figure 24. The mystery arose when it was discovered that light exists in quanta which each have an energy  $hf$  for light of a given frequency  $f$ , and that they reach a detector one at a time, and that the pattern of Figure 24 thus builds itself up one photon at a time. This can be visually seen, for example, in widely-available pictures such as Figure 29 below which is reproduced from [32], and in visual simulations such as [33] which highlight that this is not only a phenomenon for photons, but for electrons and even for composite particles such as hadrons.



**Figure 29: Buildup over Time of Double-Slit “Interference” Pattern (Reproduced from [32])**

Because Figure 24 is the constructive / destructive interference pattern for light waves at a distance from a double slit, but because this pattern builds up one field quantum at a time as shown in Figure 29, this cannot be the result of constructive and destructive interference even though the patterns are the same. This must be the result of “something else,” and to this day, the nature of the “something else” which is responsible for the buildup in Figure 29 which leads to the interference-like probability density for Figure 24 is not really understood.

Here, we obtained in equation (20.30) and graphed in Figures 27 and 28, a double slit quantum potential  $|E_1|$  in the first recursive order of non-linear quantum field theory. The minima of the potential in Figures 27 and 28 match up with the maxima in the double slit probability density of Figure 24. The same holds true with the single slit potential (20.29) graphed in Figures 25 and 26, for which the minima match up with the maxima in the single slit probability density of Figure 23. If the “something else” which explains why field quanta build up into the interference pattern of Figure 29 is to be some quantum field theory version of the geodesic principle of *least action*, then the quantum potentials of Figures 27 and 28 for a double slit configuration, and of Figures 25 and 26 for a single slit configuration, may well be that “something else” which explains the type of buildup shown in Figure 29 which looks like a wave interference pattern but cannot be so because it is built up one quantum at a time.

So, let us now pose the central question: take a single photon with an energy  $E = hf = hc / \lambda$  or a single electron or proton or neutron or meson with energy  $E = hf = hc / \lambda$  and fire that field quantum through a double slit. Figure 24 shows the probability density for where that field quantum will strike the detector. That probability density happens to be the probability density for constructive and destructive wave interference, even though there does not appear to be any way for a single field quantum to interfere with itself. So the question: why will the field quantum strike the detector with a probability density given by the interference density of Figure 24, even though there seems to be no basis for explaining the one-at-a-time buildup of Figure 29 as an interference phenomenon? Restated: what is the physical cause and / or explanation for the observed fact that singly-fired field quanta will build themselves up over

time as shown in Figure 29 into the interference pattern of Figure 24, when the buildup of Figure 29 appears to remove interference itself as the cause and / or explanation?

This question brings us to the center of philosophical challenges which have confronted physicists over the past century. There exist schools of thought which maintain that one cannot even ask the question “what is the physical cause” for the interference-like buildup of Figure 29, because as an epistemological matter, one cannot know anything definite about the location or momentum of a field quantum unless and until that field quantum has struck a detector and been detected. In other words, Figure 29 gives us epistemological certainty, because each dot tells us that a field quantum *has struck the detector* at the position indicated by the dot. But to ask how it is that the dots came to build up in the overall interference pattern of Figure 29 which when made smooth and continuous is that of Figure 24, these schools of thought will maintain, is a question which has no answer, because it requires one to talk about what happens to the field quantum *while it is in flight to the detector* which is epistemologically unknowable.

But the quantum field potential  $|E_1|$  of equation (20.30) graphed in Figures 27 and 28 raises the prospect that this school of thought can be reversed, and that one can ask “what is the physical cause” for the interference-like buildup of Figure 29 and can ask what happens to the field quanta while they are in flight to the detector and can obtain an answer / explanation rooted in the highly conservative physics principles of least action and geodesic motion. Least action principles, in their simplest form, state that a particle moves through space, over time, the way it does, because it is supposed to. Because it follows the simplest, most direct path of least resistance. Because it seeks a path which requires it to expend or have to acquire, as little energy as possible. When there is a potential involved, least action principles state that a particle seeks the minimum of the potential. And because the potential of Figures 27 and 28 has its minima right at the locations where the probability density of Figure 24 has its maxima, the answer to the question “what is the physical cause of the interference-like double slit probability buildup” of Figure 24 appears to be rooted in the potential of Figures 27 and 28: Field quanta build up, one at a time, into the interference-like probability density of Figure 24, because they follow least action principles and seek to find to minima of the “guiding potential” of Figures 27 and 28.

Although it is attractive to think that the double-slit probability density can perhaps be explained on the basis of least action principles based on a guiding potential like those of Figures 27 and 28, putting this explanation rigorously into place is not as simple as it may seem at first. This is because the single and double slit quantum potentials (20.29) and (20.30), which are rooted in (20.13) and (20.14) which are in turn rooted in (20.1) which in turn derives from the generalized spherical coordinate potential (17.4), are not quite the same as the classical potentials such as the Coulomb potential  $E_1 = -(1/4\pi)r^{-1}$  which we studied in section 16 and further in section 19 in relation to sources  $\delta_{r_\lambda}(r)$  in (19.4) which are Dirac deltas. These  $E_1$  are *quantum* potentials of an analytical, non-linear quantum field theory, not *classical* potentials, and it now becomes crucial to understand the difference between these two types of potential.

To explore this question, we continue to consider the double slit probability density of Figure 24, and its associated quantum potential of Figures 27 and 28. These figures were developed to illustrate the particular case where the overall spread of the major envelope running

from  $-1 \leq X \leq 1$  is given by the angle  $\Theta = 30^\circ = \pi / 6$ , and where there are  $B = 10$  probability peaks over each unit of the  $X$  domain including double counting of the very center peak at  $X = 0$ . This is actually the reverse of how one usually discusses double slit experiments: Usually, one starts by stating four parameters: type of field quantum  $q$  (e.g., photon, electron, hadron, etc.), slit width  $a$ , slit separation  $d$ , and the energies / wavelengths  $E = hf = hc / \lambda$  of the field quanta  $q$  being fired through the slits. Based on these four parameters, and taking  $r_0$  in Figure 22 to be very much larger than either of  $a$  or  $d$ , we will end up seeing some pattern on the detector with some  $\Theta$  and some  $B$ . Here, we have postulated that one has chosen *some specific but unspecified set* of  $q$ ,  $a$ ,  $d$  and  $E$  parameters *such that*  $\Theta = 30^\circ = \pi / 6$  and  $B = 10$ , which then produces the pattern of Figure 24 on the detector. That is why, as noted after (20.3) and (20.12), it is not essential to the development here to formalize the specific  $q$ ,  $a$ ,  $d$  and  $E$  which lead to  $\Theta = 30^\circ = \pi / 6$  and  $B = 10$ , or which lead to *whatever*  $\Theta$  and  $B$  one may observe on the detector.

In fact it is very important to have organized the approach to the slit experiments in this non-specific way as to  $q$ ,  $a$ ,  $d$  and  $E$ , because of the fact that there is a *non-unique* relation between the “input parameters”  $q$ ,  $a$ ,  $d$  and  $E$ , and the output parameters  $\Theta$  and  $B$ . Specifically, for any posited  $\Theta$  and  $B$ , there are a variety of  $q$ ,  $a$ ,  $d$  and  $E$  which will lead to the posited  $\Theta$  and  $B$  being observed on the detector. One can make a change, for example, in  $a$  and / or  $d$ , and then make a compensatory change in  $E$ , so as to produce precisely the same  $\Theta$  and  $B$ . And, *one can even change the type of field quantum* which is being fired at the detector, and by a compensating change in  $a$ ,  $d$  and  $E$ , one can maintain the exact same  $\Theta$  and  $B$ . Put into invariance language, we may say that any given  $\Theta$  and  $B$  output parameter pair will be *invariant* with regards to certain changes in the input parameters  $q$ ,  $a$ ,  $d$  and  $E$ , which changes in the input parameters can be specified with mathematical precision. And, as we see in comparing the double slit probability density of Figure 24 with the double slit quantum potential of Figures 27 and 28, there is also a one-to-one isomorphic mapping between probability density. We see the same thing comparing the single slit Figure 23 with Figures 25 and 26, and we saw this also in sections 16 through 19 for other probability densities and potentials including the  $-1/r$  potential, confining potentials, and Gaussian and Dirac delta probability densities. So this means that not only is the double slit probability density invariant with regards to certain changes in the input parameters  $q$ ,  $a$ ,  $d$  and  $E$ , but so too is the quantum potential! That is, we can change  $q$ ,  $a$ ,  $d$  and  $E$ , and so long as we do so carefully and cleverly, we can leave the quantum guiding potential  $E_1$  invariant. Put differently, any given  $\Theta$  and  $B$  – which means any given double slit probability distribution  $\partial_x(hP_0)$  and its isomorphically-associated quantum potential  $E_1$  – will map on a one-to-many basis over to a range of  $q$ ,  $a$ ,  $d$  and  $E$  which will produce  $\partial_x(hP_0)$  and  $E_1$  which are parameterized by  $\Theta$  and  $B$ , without any change. As pointed out after (20.33), because of this isomorphic  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$  mapping between the probability density for a given  $\Theta$  and  $B$  over a range of  $q$ ,  $a$ ,  $d$  and  $E$  input parameters, the quantum field equations which lead to results such as (20.32) and (20.33) could be confirmed by showing that the photovoltaic activities at the detector are dependent solely on the values of  $\Theta$  and  $B$  for the detected probability distribution, *and are otherwise invariant with respect to  $q$ ,  $a$ ,  $d$  and  $E$ .*

Where the challenge arises in understanding how even without any overt interference occurring, the quantum potential  $E_1$  serves to “guide” the field quanta toward the observed interference-like probability distribution, is not in the invariance reviewed in the last paragraph, but in the fact that while some changes in change  $q$ ,  $a$ ,  $d$  and  $E$  will leave  $\Theta$  and  $B$  and therefore  $\partial_x(hP_0)$  and  $E_1$  invariant, *there are other changes in  $q$ ,  $a$ ,  $d$  and  $E$  which will not*. For example, it is well-known that if one makes a slight change in the energy of the field quanta being fired at the double slit, and / or changes the slit configuration itself without changing any other input parameter in a compensatory manner, or changes the type of field quanta being fired without a compensating change elsewhere, that the probability distribution  $\partial_x(hP_0)$  will also change. This means due to the isomorphism between  $\partial_x(hP_0) \Leftrightarrow E_1$ , that the potential  $E_1$  will also change. Now, as to the input parameters  $q$ ,  $a$ ,  $d$  and  $E$  which affect the isomorphic  $\partial_x(hP_0) \Leftrightarrow E_1$ , two of these parameters,  $q$  and  $E$  are *intrinsic* to the field quanta themselves. The  $q$  parameter tells us what type of field quanta is being fired, e.g., photon, electron, proton, neutron, meson, helium nucleus, hydrogen atom, carbon molecule, etc. And the  $E = hf = hc / \lambda$  parameter tells us about the energy of those field quanta. The other two parameters  $a$  and  $d$  are the slit width and the slit separation, and these are completely independent of, i.e., *extrinsic* to, the type of field quanta being fired through the double slit.

But all four of the input parameters  $q$ ,  $a$ ,  $d$  and  $E$  would appear to be independent of, i.e., extrinsic to, the potential energy  $E_1$ , at least if we think of  $E_1$  as a potential in the classical sense as a preexisting background field which is not affected – or is at most minimally affected – by the objects which travel through the potential. Here, in non-linear quantum field theory, something very different is going on than what we see for a classical potential: We can change the type of field quantum  $q$  or its energy  $E$ , and by so doing, we will change the isomorphic  $\partial_x(hP_0) \Leftrightarrow E_1$  which results from the overall  $q$ ,  $a$ ,  $d$  and  $E$  parameter set. And so, the change in  $q$  or  $E$  will change the guiding quantum potential  $E_1$ . Even more unusual if we think in terms of a classical potential, is that we can even change  $a$  and  $d$ , which are the slit width and the slit separation, and that this too will cause a change in the guiding potential  $E_1$ , again because it will change the probability density  $\partial_x(hP_0)$  and because of the isomorphism  $\partial_x(hP_0) \Leftrightarrow E_1$ . In classical theory, there is no way in which one would expect a potential to be changed because of the changing of a slit configuration. Yet in quantum field theory, changing the slit configuration definitively changes the guiding quantum potential  $E_1$ . This is what we must now understand if we are to place the prospect that  $E_1$  is a guiding potential for the field quanta, onto a rigorous basis.

As just discussed, *the non-linear quantum potential  $E_1$  appears to very different character from a classical potential*. In classical field theory, potentials are regarded as being background fields which are not affected by the particles which travel through them, except perhaps at very nominal order due to the fields of the particles themselves. But in quantum field theory, a quantum potential such as  $E_1$  is not preconfigured. It only becomes configured when a quantum particle (a single field quantum) travels through that the quantum “vacuum.” When this

happens, both the particle itself *and the configuration of spacetime* in nearby locales, and even in distant locales, coact in their entirety to “light up” the potential which then dynamically gives rise to the probability distribution observed at a sink (e.g., detector), because the travelling quantum is itself disturbing and changing the quantum vacuum and giving rise to the potential. Specifically, once the potential is lit up, in rather Bohm-ian fashion (referring to David Bohm), the field quantum simply propagates through the quantum vacuum so as to find the lowest potential in a geodesic / least action / least potential sort of manner, and strikes (sinks at) the detector after so doing. Then, after that quantum particle has travelled through the potential which is partially induced by that particle type and particle energy and partially induced by everything else in the universe, and is detected, the vacuum potential goes back into a latent, non-configured state, awaiting the arrival of the next quantum particle to travel through the vacuum and do the same thing.

In sum, this is a case of the travelling particle itself acting as an “observer” to the vacuum in which it is travelling by disturbing the vacuum and causing the vacuum to respond by configuring its vacuum potential. Then, the field quantum follows a least action path *through the potential to which it itself has given rise*, to its final detected location on the detector. After enough field quanta have each traversed the vacuum in this way, and have followed paths of least action based on the self-created potentials  $E_1$  of Figures 27 and 28, the accumulated result accrued one quantum at a time as in Figure 29, will be the probability distribution of Figure 24. As to any individual field quantum, we cannot predict precisely where this will land on the detector, except probabilistically via Figure 24. But this is not because of any quantum-epistemological bar to knowing what happens while the field quantum is in flight toward the detector for why the probability distribution of Figure 24 is an interference pattern even though there is no discernable interference when we consider one quantum at a time. It is because of the statistical distribution which it is most convenient to ascribe to the field quanta as they are being fired from the source and as they make their way through the slits over a range of trajectories and positions which are described statistically, rather than particle-by-particle, in the same fashion as the 19<sup>th</sup> century probabilities which were used to describe the kinetic motions of the particles in a gas.

Now, it might be intuited how this sort of self-creation of a guiding potential via the interaction of a field quantum with the vacuum could explain how a travelling field quantum might affect the quantum probability and the potential based its own energy and type, i.e., based on the input parameters  $E$  and  $q$ . But the real quantum theory-based mystery is how it is that changing the *slit configuration*, which is part of the configuration of spacetime in nearby locales, can also affect a change the potential, since the slit configuration is *independent* of the particle type or its energy. This turns out to be rooted in the fundamental and qualitative difference between a quantum action  $W$  or a quantum potential  $E$  in  $W = ET$ , and a classical action or a classical potential. Let talk about the quantum potential  $E$  rather than the quantum action  $W$  because that is what we have in Equation (20.30) which is used to graph Figures 27 and 28.

A *classical* potential is a one-body potential sourced by a classical object which is posited to be situated in a particular location at a particular time (which it will be appreciated already raises quantum issues of “uncertainty”). But even more importantly, the classical potential is used to calculate in a “controlled” or “sterile” environment where anything else which might

influence a particle travelling through the potential *other than* the one body generating the potential, is segregated out and ignored. As soon as the controls are relaxed – for example, when one starts to consider a three-body problem, or worse, a many-body problem – it becomes much more difficult if not impossible to actually do a useful calculation. In sum, classical field theory idealizes away the effects of all but the single localized source creating the potential and the single localized particle posited to be travelling through the potential, and when there is a deviation from these idealizations, classical field theory rapidly loses its ability to make useful quantitative predictions without inordinate heavy lifting. To be sure, there are some classical simplifications which do enable multiple sources to be treated accurately. For example, a uniform spherical distribution of multiple charges can be idealized to a single charge at the center of the sphere which has a charge that is the sum total of all the charges, and the Coulomb potential may then be applied to anything travelling outside the sphere. But these require very clear symmetry assumptions and lose their predictive capacity as soon as the postulated symmetries are removed.

In contrast, the *quantum* potential in the path-integrated  $W = ET$  is the precise opposite of a classical one-body potential. It is an all-body potential. And  $W$  is an *all-body action*. To get to  $W = ET$  one has already path integrated over *all* of the possible configurations of the classical field via the measure  $DG$ , see (11.4), so all possible field configurations are included. To then get to  $E_1$  in (15.42) or all of its progeny such as (20.30) which is used to graph Figures 27 and 28, one has further Fourier integrated over all possible momenta via the measure  $d^4k$ , and one has also integrated over all of spacetime for both the source and the sink via the measure  $d^4x d^4y$ , see (15.9). So the  $E_1$  in (15.42) already contains implicit “quantum knowledge” about the spacetime of which it is a function, because it was arrived at in the first place by integrating over all possible spacetime, energy-momentum, and field configurations. The quantum potential  $E_1$  is therefore not the product any one particular source as is a potential in classical field theory. Rather, it is the product of any and all sources and sinks and configurations of matter and energy throughout all of space and time. This, of course, includes any nearby slit configurations.

While  $E_1$  is an all-body potential, this does not mean that we feed everything in the universe at all times and places into the equation such as (15.42) or (17.2) or their progeny which determine  $E_1$ , or that  $E_1$  is somehow omniscient enough to possess quantum knowledge of everything that has ever happened or will happen anywhere. Rather, it means that whatever probability densities  $\partial_x(hP_0)$  we do feed in to determine  $E_1$  via the isomorphic mapping  $\partial_x(hP_0) \Leftrightarrow E_1$ , and whether this probability density is the known probability density for a source or the known probability for a sink (at a detector), we will come away with an associated quantum guiding potential  $E_1$  that fully explains and indeed is the *physical result* of the observed probability density for the source, or the *physical cause* of the observed probability density for the sink, via the isomorphic mapping  $\partial_x(hP_0) \Leftrightarrow E_1$ . Using the examples already developed here, let us see exactly how this works.

First, as we saw in section 19, suppose we posit the source probability density  $\partial_r(hP_0) = \delta_{r_\Lambda}(r)$  of (19.12), which is a Dirac half-delta defined in (19.4) right at  $r = 0$ , and which is zero everywhere else. Then, as seen in (19.14) which we started to see at (16.1), the quantum guiding potential  $E_1 = -(1/4\pi)r^{-1}$  turns out to coincide precisely with the Coulomb potential. That is, the isomorphic  $\partial_x(hP_0) \Leftrightarrow E_1$  which we generalize to  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$  becomes a probability density  $\partial_r(hP_0) = \delta_{r_\Lambda}(r)$  mapping to a potential  $E_1 = -(1/4\pi)r^{-1}$ , that is, it becomes the mapping  $\delta_{r_\Lambda}(r) \Leftrightarrow -(1/4\pi)r^{-1}$ . This is the correspondence of non-linear quantum field theory, to classical field theory. A source which is a Dirac half-delta right at  $r = 0$ , yields a potential  $-(1/4\pi)r^{-1}$  which is finite everywhere but at  $r = 0$ , which is the precise same place where we have posited the controlled singularity  $\delta_{r_\Lambda}(r)$ . The source at  $\delta_{r_\Lambda}(r)$  is the *physical cause* of the  $E_1 = -(1/4\pi)r^{-1}$  Coulomb potential. Further, in classical theory, if we had multiple charge sources, one would model this by placing multiple deltas  $\delta_{r_\Lambda}(\mathbf{x}_1)$ ,  $\delta_{r_\Lambda}(\mathbf{x}_2)$ ,  $\delta_{r_\Lambda}(\mathbf{x}_3)$  etc. at various positions  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$  etc., and each of these sources would then generate Coulomb potentials  $-(1/4\pi)r^{-1}$  centered at  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ . Then the overall potential would be a linear superposition of the individual potentials.

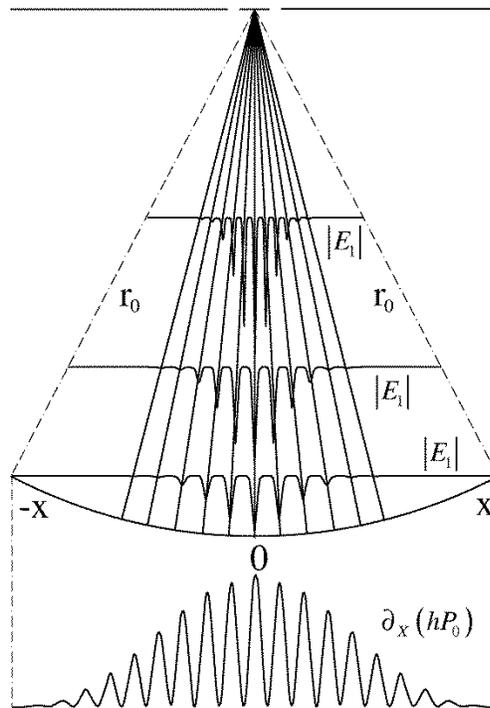
But in non-linear quantum field theory this works differently. As our second example, we also saw in section 19 that if we posit a source probability density which is the half-Gaussian  $\partial_r(hP_0) = (2/r_\Lambda\sqrt{\pi})\exp(-r^2/r_\Lambda^2)$  of (19.1), this will map over via the  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$  isomorphism to the quantum potential (19.5) / (19.6) which has the real magnitude (19.10) and which is graphed in Figure 19 and deviates from a strict  $-(1/4\pi)r^{-1}$  Coulomb potential in the  $r_\Lambda/\sqrt{2}$  region to the degree illustrated in Figure 21. So now, the source probability density is a Gaussian, which unlike the Dirac delta, has a *non-zero spatial expanse*. But to deduce its potential  $|E_1|$  graphed in Figure 19, we simply plug the Gaussian probability density (19.1) into the generalized radial relationship (17.5) and out pops the isomorphically-corresponding potential of (19.5) for which we graph the “almost -1/r” magnitude in Figure 19. When we refer to the Figure 19 potential as an “all-body” potential, what we are saying is that the Gaussian (19.1) is *not* an idealized, localized half-delta  $\delta_{r_\Lambda}(r)$ , but has a spatial expanse most of which is concentrated within a few standard deviations of  $r_\Lambda$ , and all of which spreads out in its “tail” throughout the domain from  $0 \leq r \leq \infty$ . Once we specify the “entire body” of the Gaussian source probability density (19.1) as one holistic entity with finite, unrestricted spatial expanse, we immediately pop out the associated “all body” potential which accounts for the entire Gaussian from body to tail, and find a small deviation from -1/r in the body, and a virtually unchanged -1/r potential throughout the tail. So, as long as we are able to specify the entire body of the probability density for which we wish to ascertain a quantum potential, we can indeed deduce the entire potential and do not have to go about this task piecemeal as we would in classical theory.

Our third example was that the constant isotropic coupled probability density explored in section 17, which led to fitting the running strong coupling curve  $\alpha_s(Q)$  in section 18. Here, we posited a constant coupled probability field  $hP_0 = \Lambda r$  in (17.28) following normalization, for which the associated probability density  $\partial_r(hP_0) = \Lambda$  is constant over the domain  $0 \leq r \leq r_\Lambda$  as illustrated in Figure 7. Here, via the isomorphism  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$ , we obtained the quantum guiding potential of (17.33) graphed in Figure 6 and further studied in Figure 15. Here, we begin to really see the nature of  $E_1$  as a guiding potential. How so? As noted earlier, it is one thing as a mathematical exercise to posit a constant density  $\partial_r(hP_0) = \Lambda$  over the domain  $0 \leq r \leq r_\Lambda$  which drops to  $\partial_r(hP_0) = 0$  for  $r > r_\Lambda$ ; it is quite another to specify the physics which will hold together the density  $\partial_r(hP_0) = \Lambda$  within  $0 \leq r \leq r_\Lambda$  and not allow that any of the probability density to “leak out” beyond  $r > r_\Lambda$ . Here is where the physics of the Figure 6 and 15 potential goes hand in hand with the probability density. The potential in Figures 6 and 15 makes it impossible for any field quantum to move to  $r > r_\Lambda$ , because the potential  $E_1$  at the first-recursive order rises to a maximum at  $r_\Lambda$ , and because of the expectation that the potential actual physical potential  $E_\infty$  to infinite recursive order will itself grow asymptotically to infinity. The postulation of  $\partial_r(hP_0) = \Lambda$  within  $0 \leq r \leq r_\Lambda$  carries with it, isomorphically, the confining potential needed to physically-enforce itself through the natural tendency of physics systems to tend toward least action. Here, the “entire body” is the density  $\partial_r(hP_0) = \Lambda$  within  $0 \leq r \leq r_\Lambda$ , and the all-body potential which enforces the entire density configuration  $\partial_r(hP_0) = \Lambda$  is the one in Figures 6 and 15.

So, what we really do in section 17 is postulate a probability density which is confined to  $0 \leq r \leq r_\Lambda$ , and the quantum field equations themselves then pop out the potential required to enforce that confinement! This provides the physical, least action, guiding potential necessary to explain and cause confinement. And beyond confinement and the resulting fitting of the QCD running coupling in Figure 14, and beyond the fact that the potential itself as studied in Figure 15 points toward new physics precisely at the GUT and Planck scales where new physics is to be expected, the key lesson taught by the potential in Figures 6 and 15 is that field quanta such as quarks do not become confined just because we posit their confinement. They become confined because the positing of confinement carries with it an associated potential which enforces confinement via principles of least action. And this means more globally that notwithstanding all of the conceptual and epistemological challenges presented by the quantum reality that has been discovered and developed over the past century and longer, *physics is still physics; it is not magic. One who practices physics as a discipline still must find the causal reasons which explain why we observe what we observe, and least action principles remain a very central and important cause for the natural behaviors that we observe, even in the quantum domain.*

So now we return to the double slit experiment of the present section, armed with the understanding that in non-linear quantum field theory, a coupled probability density goes hand-in-hand with a quantum potential with which it is isomorphic, which completely causes and / or

is caused by every aspect of the fully specified coupled probability density. And it is in the context of the double slit experiment that this all comes into full bloom. To illustrate this, let us return to the detector apparatus schematically laid out in Figure 22, apply this specifically for the double slit experiment, and superimpose both the guiding potential of Figure 27 in the region where the field quanta propagate from the slits to the detector, and the probability density of Figure 24 which is ultimately viewed at the detector following a buildup of field quanta at the detector in sufficient numbers to fill in the one-at-a-time strikes of Figure 29 and bring about a curve with a the continuous appearance of double slit wave interference. We illustrate all of this in Figure 30 below:



**Figure 30: The Guiding Potential  $|E_1|$  and the Consequent Probability Density  $\partial_x(hP_0)$  for the Double-Slit Experiment**

Now, as we did for confinement just discussed, we posit the probability density  $\partial_x(hP_0)$  seen at the detector. But even here, for a quantum treatment of the double-slit experiment, *physics is still physics; and it is still not magic*. We still require a causal explanation of why it is that the individual field quanta aggregate together to produce the observed interference-like probability density, just as we required a causal explanation for what held together the confined system with a posited density  $\partial_r(hP_0) = \Lambda$  over  $0 \leq r \leq r_\Lambda$  and found the guiding potential of Figures 6 and 15 to do so. We see that sufficiently far from the slits themselves, the guiding potential  $|E_1|$  creates what may be thought of as least action grooves in the vacuum which cause the individual field quanta to strike the detector with the observed probability density. To use a physical analogy, these are in the nature of “bobsled tracks.” Not every bobsled will register the same time in a race, and not every bobsled will end up in the same place at the end of a race, but

the bobsled tracks do definitively establish the probability density curve for where any particular bobsled is more likely and less likely to end up.

However, lest one think that Figure 30 yields a fully classical way to explain the double slit experiment and removes all of the quantum challenges from the required explanation, rest assured, it does not. Unlike a classical potential, the potential “grooves” do not preexist in the vacuum. If one changes the field quanta  $q$  or if one changes the energy  $E$  of the field quanta without a compensatory change to another of the  $q$ ,  $a$ ,  $d$  and  $E$  parameters, *then the guiding potential itself will change*. This does not happen with a classical potential. Even more quantum-indicative, if one changes the slit size  $a$  or the slit separation  $d$  without a compensating change to another parameter, *then the guiding potential itself will again change*. This most certainly does not happen with a classical potential.

But the non-classical aspects of quantum theory *do not* come into play via an abandonment of cause and effect or via relinquishing the requirement to explain why the observed probability densities are what they are on some causal foundation. Nor do they come into play via an abandonment of least action / geodesic principles as the mechanism to explain the causation of why things happen the way they happen and why the field quanta statistically end up where they end up. Nor do the non-classical aspects of quantum theory come into play by taking the epistemological view that what a field quantum does while it is propagating has no meaning and that the only meaningful thing we can say about an individual field quantum is where and when it observably struck a detector. Figure 30 enables us to talk very definitively, and indeed classically albeit statistically, about the behaviors of individual field quanta as guided by a potential during propagation which directly causes the observed probability distribution at the detector.

The non-classical aspects of quantum theory *do* come into play when it comes to understanding the non-linear guiding potential  $E_1$  itself, and how this potential itself arises in response to, i.e., is caused by, the  $q$ ,  $a$ ,  $d$  and  $E$  parameters in double slit experiments. (Physically, one should always keep in mind that  $E_\infty$  deduced to infinite recursive order is the physical guiding potential in the real universe and  $E_1$  deduced here is just the potential containing the first non-linear order of recursion.) One might very well fire a number of field quanta, say, electrons, through a given double slit configuration, and end up with what is seen in Figure 30. Then, one might do nothing other than change the energy  $E$  of a new set of electrons. Thereafter, the observed probability density  $\partial_x(hP_0)$  will change, and the isomorphically mapped guiding potential  $E_1$  will change, and all of this change will be a causal result of the change in the electron energy. This means that a first electron  $A$  with energy  $E_A$  will actually give rise to and propagate through a *different guiding potential* than a second electron  $B$  with different energy  $E_B$ . In a very quantum, non-classical sort of maneuver, the propagating field quantum itself is affecting and changing the vacuum in which it is propagating, almost as if a bobsled managed to inform the bobsled course as to which particular bobsled was riding through the course, with the bobsled course then reconfiguring its bobsled tracks accordingly. Similarly, one might leave all else the same, but change the electron to photons, and again, the vacuum

itself will be influenced by the fact that it now has a photon rather than an electron riding through, and will lay out a different guiding potential than the one that guides the electron.

And, the non-classical aspects of quantum theory *do* come into play in superlative terms, when one does nothing more than change the slit width  $a$  or the slit spacing  $d$  without any other change. Here too, if one changes the slit configuration in the middle of an experiment, then the observed probability density  $\partial_x(hP_0)$  will change and its guiding potential  $E_1$  will change in consequence, *all because of the change to the slit configuration*. This is a truly non-classical sort of phenomenon, and it suggests that the quantum vacuum actually has “quantum knowledge” as to the nearby slit configuration and updates its knowledge when the slit configuration is changed. But this is understood by keeping in mind that the nonlinear quantum field equations which generally relate  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$  on an isomorphic basis are arrived at following a path integration over all classical fields  $\int DG$ , following an integration over all momenta  $\int d^4k$ , and following a integration over all of spacetime for both the source  $\int d^4x$  and the sink  $\int d^4y$ . By Huygens, all points along the front of a signal propagation can be viewed as the source of a new signal propagation. For a freely-propagating wave, this is often a triviality, because constructive and destructive interference serve to cancel each other out. But when a source emits a signal and then that signal travels through a slit, by Huygens we can re-source the source to the slit, and thereby regard the slit as the source which was integrated over  $\int d^4x$ .

So from this viewpoint, by Huygens, the source of a double slit experiment is the double slit itself. The probability density at this source has .5 of its cumulative probability situated at the first slit, and the other .5 of its cumulative probability situated at the second slit. This then maps via  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$  to a related quantum potential. Meanwhile, the sink of a double slit experiment which was integrated over  $\int d^4y$  is the detector. This displays the probability densities illustrated in Figure 30 which are caused by the potentials also illustrated in Figure 30. What Figure 30 does not display explicitly, is the  $1 = .5 + .5$  probability at the source slits, and the related potential. Mathematically, to *marry the source to the sink via the propagation and bridge the wave view to the particle view of quantum field theory*, one may start with classical wave theory, specify the probability densities over the entire continuum for both the coordinates  $r$  and  $\theta \rightarrow X$ , and then use the non-linear quantum field equations for  $\nabla(hP_0) \Leftrightarrow E_1(\mathbf{x})$  to deduce the  $E_1(\mathbf{x})$  which guides a field quantum all the way from source to sink. At that point, although we have taken advantage of classical wave theory to specify  $\nabla(hP_0)$  and then deduce  $E_1(\mathbf{x})$ , the  $E_1(\mathbf{x})$  so deduced now provides a least-action foundation for guiding individual field quanta over to the detector such that statistically, they will display the observed interference-like pattern of the double slit experiment.

That is what we have in effect done in this section, by starting with the wave-densities (20.7) and (20.9) for the single slit and double slit sink graphed in Figures 23 and 24, respectively; and then deriving the associated  $|E_1|$  guiding potentials (20.29) and (20.30) graphed

in Figures 25 through 28. At this point, we are equipped to pass over to the particle view, because now, as shown in Figure 30, these guiding potentials  $|E_1|$  provide a least action foundation to guide individual field quanta through their propagation so as to statistically aggregate at the detector sink in the interference-like probability pattern shown at the bottom of Figure 30 for the double slit experiment. However, the non-classical features of quantum theory remain, but not in an abandonment of cause and effect or relinquishing of the need to explain why we observe what we observe by least-action principles. They remain insofar as the vacuum itself gives rise to the guiding potential used to propagate any particular field quantum from the source to the sink, by interacting with and thereby having quantum knowledge of and being affected by: a) the type of field quantum that is propagating through the vacuum; b) the energy of that field quantum; and c) the configuration of the nearby slits which by Huygens become the sources. So each field quantum while propagating, coupled with the nearby slit configuration: a) induces (causes and creates) the guiding potential in the vacuum which will then influence its own propagation from source to sink; b) propagates from source to sink under the least action influence of the self-induced and slit-induced guiding potential based on interaction with and quantum knowledge within the vacuum, and c) strikes the sink detector in the observed interference-like probability distribution as a causal consequence of having travelled through the self-induced and slit-induced guiding potential.

In this way, nonlinear quantum field theory differs from classical field theory because in classical field theory the potentials are sourced by a localized source and are taken to be virtually or completely unaffected by what travels through those potentials; while in quantum field theory a guiding potential arises in a vacuum in definitive response to both the individual field quanta propagating through the vacuum and to the quantum knowledge held by the vacuum about the surrounding spacetime including such matters as slit configurations, such that the propagation of an individual field quantum follows a path of least action in relation to its self-induced and slit-induced guiding potential. *In both quantum and classical field theory*, individual field quanta propagate as they do because of least action principles, including a natural tendency toward the minima of a potential, and when one is dealing with large numbers of field quanta the overall pattern of field quantum arrival at a sink / detector is described as the statistical accumulation of large numbers of statistically-distributed field quanta each pursuing least action paths. *The difference between classical and quantum field theory rests solely in the way in which the guiding potential comes about.* In classical field theory the potential is preexisting and is taken to be totally or virtually unaffected by anything but the source. This limits the analytical predictive reach of classical field theory to sources which are either highly localized as in the Dirac deltas which implicitly underlie application of the Coulomb potential, or have high degrees of symmetry such as a uniform, spherically symmetric charge distribution, and creates practical challenges analytically solving problems involving three or more bodies absent some exploitable symmetry. In quantum field theory the potential is induced in the quantum vacuum in real time as field quanta propagate and takes full account of the types of quanta which are propagating, what their energies are, and the configuration of the nearby spacetime in the form of slits through which the quanta are propagating. There is no epistemological problem with discussing the propagation of individual field quanta from source to sink, and so long as one can mathematically specify as source and / or sink in a complete fashion, one will be able to specify the isomorphically-associated least-action guiding potential in commensurately complete fashion.

## 21. Summary and Conclusion

This concludes the formal development of this paper, so let us summarize what we have learned: In non-abelian gauge theory with gauge fields  $G$ , although the magnetic charge density  $P = DF = DDG = 0$  by a Jacobian identity (2.4) just as the abelian magnetic charge density  $P = dF = ddG = 0$  because of the differential forms geometry, there is still a non-vanishing magnetic field flux  $\oint\!\!\!\oint F = -i\oint\!\!\!\oint [G, G] = -i\iiint dGG \neq 0$  (3.3) across closed surfaces which contrasts to the zero net flux  $\oint\!\!\!\oint F = 0$  that one has in abelian gauge theory. These apparently-conflicting features of non-abelian theory – namely a non-zero magnetic flux over closed surfaces but no magnetic sources – are reconciled by realizing that the magnetic field flux is not sourced  $\oint\!\!\!\oint F(P)$  by any elementary magnetic charge density which is  $P = DF = DDG = 0$ , but rather is sourced  $\oint\!\!\!\oint F(G)$  by a “faux” magnetic source  $P' = -id[G, G] = -idGG$  which arises totally from the gauge fields,  $P'(G)$ . But real gauge fields do not arise spontaneously. They must be sourced by an electric charge density  $J$ , and in non-abelian gauge theory, the differential equation which governs this is  $*J = D*F = D*DG$ . Further, we also know that in Dirac theory, electric charge densities are in turn sourced by fermion wavefunctions  $\psi$  via Dirac’s  $J^\mu = \bar{\psi}\gamma^\mu\psi$ . Thus, we now need to set upon obtaining the inverse solution to  $*J = D*F = D*DG$  for  $G(J)$  to enable us to find  $\oint\!\!\!\oint F(G(J))$  and  $\oint\!\!\!\oint F(G(J(\psi)))$ .

So in section 5 we develop the electric source field equation  $*J = D*F = D*DG$ , and in sections 6 and 7 respectively, we carefully develop the inverse solutions  $G(J)$  for massive and massless gauge bosons respectively, paying very close attention to issues involving uniqueness and gauge-invariance and gauge fixing and “contextual gauge fixing” wherein a *mathematical* inverse which is non-unique becomes unique when placed into the *physical* context of a conserved current density. And in section 8 we see how  $G(J)$  is not really a solution involving  $J$  alone, but is a highly-non-linear, recursive function  $G(G, J)$  which can be recursed as often as desired, and then turned from  $G(G, J)$  into  $G(J)$  by setting the perturbation  $V = 0$  at any desired order. We also noted how the physical inverse ought not to depend on an arbitrary cutoff of the recursion, but rather, ought rather to be based on the series (8.20) that results from recursing an infinite number of times before zeroing the perturbation.

So starting in section 9 we made use of the non-abelian solution for a massive gauge boson, namely  $G_\mu = (-V + k_\tau k^\tau - m^2 + i\epsilon)^{-1} J_\mu$  of (6.27) to write out  $\oint\!\!\!\oint F(G(G, J))$  in (9.2). Then to keep the initial development simple and develop the “ground state” symmetries, we immediately set  $V = 0$  in (9.4) to write  $\oint\!\!\!\oint F(G(J))$  in the zeroth recursive order  $((0))_0$ , which is the same thing as having used the abelian massive solution  $G_\mu = (k_\tau k^\tau - m^2 + i\epsilon)^{-1} J_\mu$  (6.17)

a.k.a. (6.28). After then using  $J^\mu = \bar{\Psi}\gamma^\mu\Psi$  to replace currents with fermions and thus arrive at  $\oint\!\!\!\oint F(G(J(\psi)))$  in (9.7), we turned to the fermion Exclusion Principle of Fermi-Dirac-Pauli.

It is the Exclusion Principle that drives the introduction of a dimension-3 gauge group to ensure that all of the fermions within the  $\oint\!\!\!\oint F(G(J(\psi)))$  system are in three distinct eigenstates, turning this now into  $\oint\!\!\!\oint F(G(J(\psi_R, \psi_G, \psi_B)))$  which reaches the goal established at the end of section 3. The reason for there being three colored quarks in the ground state of a baryon is then seen to be very simple: because there are three additive terms in the covariant tensor expression (9.7) for a magnetic monopole. This also brings with it, eight bi-colored gauge fields. After applying a Goldstone-like mechanism (9.15) to reallocate degrees of freedom and force the gauge fields to be massless and give mass to the fermions while contextually-preserving the uniqueness of the underlying solution for  $(G(J))$ , we arrive at the ground state monopole density of (9.21). This monopole has the antisymmetric  $R[G, B] + G[B, R] + B[R, G]$  color-neutral wavefunction of a baryon although it does also contain fermions in three colored eigenstates, and as we had already found in (3.5), it permits no net flux of individual gauge fields across its closed surfaces. But then we find in (10.4) and (10.5) that this monopole does permit a net flux only of color-neutral  $\bar{R}R + \bar{G}G + \bar{B}B$  mesons, which further cements the confinement of gauge fields first suspected in section 3 because nothing *other than* colorless  $\bar{R}R + \bar{G}G + \bar{B}B$  fields are permitted to net flow in across closed surfaces. And we further find from (10.6) that the dimension-3 gauge group must be  $SU(3)\times U(1)$ , not just  $SU(3)$ , and that this provides the magnetic monopoles with topological stability so long as this  $SU(3)\times U(1)$  group emerges following the spontaneous symmetry breaking of a larger simple group  $G \supset SU(3)\times U(1)$ . We learn at (10.9) that the  $U(1)$  generator provides a natural platform for equipping each fermion with a baryon number  $B = \frac{1}{3}$  and the overall monopole with  $B = 1$ , which now introduces *flavor* to these color-neutral monopoles and mesons and their colored fermions and gauge bosons. And we see in (10.10) and (10.11) that one can thereafter arrive at suitable generator assignments which give rise to the correct electric charges  $Q = +1$  for the proton and by a disconnected assignment (which then requires a larger unifying group)  $Q = 0$  for the neutron, as well as the  $Q = +\frac{2}{3}$  for the up and  $Q = -\frac{1}{3}$  for the down flavors of quark.

Although nuclear and particle physics are often discussed as if they are one and the same discipline, in fact, they are very distinct based on present understandings of each. This fault line which separates nuclear and hadron physics from particle physics is concisely captured by Jaffe and Witten when they state at page 3 of the “Yang-Mills and Mass Gap” problem [6] that:

“... for QCD to describe the strong force successfully . . . It must have ‘quark confinement,’ that is, even though the theory is described in terms of elementary fields, such as the quark fields, that transform non-trivially under  $SU(3)$ , the physical particle states—such as the proton, neutron, and pion—are  $SU(3)$ -invariant.”

It is this difference between “elementary fields, such as the quark [and the gluon] fields, that transform non-trivially under  $SU(3)$ ” and “the physical particle states—such as the proton, neutron, and pion—[which] are  $SU(3)$ -invariant,” as well as the need to give flavor to color-neutral baryons and understand the origins of the specific baryon flavors which are protons and neutrons, which separates the elementary particle physics of colored quarks and gluons, from the hadron physics of the colorless baryons and mesons, and the nuclear physics of proton- and neutron-flavored baryons.

As detailed in the discussion following (10.5), if one advances the thesis that the non-abelian faux magnetic monopole of (9.21) is in fact synonymous with a baryon, then the results reviewed in detail in section 10 would appear to solve this *confinement leg of the mass gap problem*, at least in the classical context. Moreover, the results presented here take a critical step forward toward unifying elementary particle physics with hadron physics and nuclear physics. It is equation (9.21) which operates as a “bridge” between the elementary particle physics of colored quarks and gluons and the hadron physics of the colorless baryons and mesons. This is because (9.21), together with its related consequence (10.5), demonstrates how quark and gluon fields that transform non-trivially under  $SU(3)$  assemble together into the *colorless,  $SU(3)$ -invariant particle states* which are baryons and mesons, that is, hadrons. Then, the non-vanishing trace of (9.21) forces us to employ  $SU(3)\times U(1)$ . This ensures topological stability which is required if (9.21) is to be associated with stable physical particles such as the neutron and especially the proton. Further, via the new  $U(1)$  generator, this introduces flavor which then allows these baryons to be flavored into the protons and neutrons at the heart of nuclear physics.

Of course, as discussed in section 4 there are many reasons to believe confinement is related to the running of the coupling constant which is an inherently quantum effect. But as also argued in section 4, one might take the perspective that the *cause* for confinement and baryon compositeness is the classical field equation (3.3) for a Yang-Mills monopole which has the symmetry (3.5), and that one of the *effects* of this is that in a quantum field treatment of these baryon monopoles, the strong coupling will weaken for ultraviolet and strengthen for infrared probes. Without more, however, one could fairly conclude that the connections suggested between some identities of the classical Yang-Mills equation and confinement in the quantum theory are simply still too speculative or weakly supported to constitute a viable theory of hadronic physics, especially since quarks are alluded to but not shown to be required.

But sections 9 and 10 overcome any such conclusion. These sections deepen support for the argument made in sections 3 and 4 by demonstrating that a further *cause* for confinement is the color-neutral  $SU(3)$ -invariance of both the monopole (9.21) and the meson (10.5), which might then be expected in a quantum field treatment to reveal the *effect* of a running coupling constant which is consistent with these root causes that are *already seen in the classical theory*. It is certainly true that an important view of confinement is the quantum view of a running coupling. But so too is Jaffe and Witten’s complementary symmetry view of confinement as utilized here, in which “even though [a] theory is described in terms of elementary fields, such as the quark fields, that transform non-trivially under  $SU(3)$ , the physical particle states—such as the proton, neutron, and pion—are  $SU(3)$ -invariant.” Sections 9 and 10 here make clear that Yang-Mills monopoles manifest these required confinement symmetries. And, this underscores the value as argued in section 4, of finding and fleshing out, the right classical theory to quantize,

before trying to leap unarmed into quantization. Colloquially speaking, classical theory is the “horse” which must one must precede the “cart” of quantization.

As to the “cart” of quantization, two further points may now be made in light of the development after section 4, to supplement those already made in section 4. First, as noted in section 4, the chiral anomaly provides an object lesson that not every symmetry which appears in a classical theory carries through to the associated quantum theory. As pointed out in section 7, any divergence there may be between classical and quantum symmetries emanates from the measure  $D\varphi$  which is the integration variable in the path integral. A classical symmetry exists if some transformation leaves the action  $S(\varphi)$  invariant. A quantum symmetry exists (and inherits the classical symmetry) if the same transformation leaves the path integral  $Z = \int D\varphi \exp iS(\varphi)$  invariant. So, for example, although the classical monopole (9.21) has the color-neutral baryon wavefunction  $R[G, B] + G[B, R] + B[R, G]$  and the classical net-flowing magnetic field (10.5) has the color-neutral meson wavefunction  $\bar{R}R + \bar{G}G + \bar{B}B$ , i.e., are classically invariant under an  $SU(3)$  gauge transformation, it is valid to ask whether these symmetries will carry through to the related quantum objects. This cannot be answered with absolute certainty until one has the complete quantum theory corresponding to the foregoing classical development, but it is encouraging to note that the observed baryons and the mesons of *quantum* physics are also known to be color-neutral with the same respective  $R[G, B] + G[B, R] + B[R, G]$  and  $\bar{R}R + \bar{G}G + \bar{B}B$  wavefunctions. Thus for example, when Jaffe and Witten state on page 3 of [6] that “the physical particle states—such as the proton, neutron, and pion—are  $SU(3)$ -invariant,” they are not qualifying or restricting this statement to classical particles. *QCD is a quantum theory, and the invariance of baryons and mesons, i.e., hadrons, under  $SU(3)$  is a well-known feature not only of classical, but of quantum, chromodynamics. That these symmetries appear to emerge very naturally and inexorably from classical Yang-Mills theory without having to make any separate postulates about  $SU(3)$  being a theory of strong interactions, is highly compelling.*

Second, the most important result pertaining to quantization in this paper, is the finding in section 8 and its application in sections 11 and 13 that the inverse solution  $G(J)$  is actually a recursive solution for  $G(G, J)$ , but that this can be turned into a  $G(J)$  solution by recursing to any desired order and then setting the perturbation  $V = 0$ . This is important because, referring to page 6 of [6], the difficulty of being able to:

“Prove that for any compact simple gauge group  $G$ , a non-trivial quantum Yang–Mills theory exists on  $\mathbb{R}^4 \dots$ ”

is not a physics problem, it is a *mathematics* problem, and more particularly, it is a *calculation* problem of not knowing how to perform an exact analytical calculation of the quantum path integral for Yang-Mills theory in particular, and for non-linear physics theories in general.

Specifically, as discussed in section 8, the technique of analytically calculating a path integral  $Z = \int DG \exp iS(G) = \mathcal{C} \exp iW(J)$  revolves around clever extrapolations of the

Gaussian integral  $\int dx \exp\left(-\frac{1}{2}Ax^2 - Jx\right) = (-2\pi/A)^{-\frac{1}{2}} \exp\left(J^2/2A\right)$  which only contains  $x$  and  $x^2$  and no higher order in the integration variable  $x$ . Put an  $x^3$  or an  $x^4$  into this integral, or even worse, put any higher-order polynomial into this integral, and it is simply not known mathematically how to calculate this integral at all. So the *physics* recipe for quantizing Yang-Mills is very clear: find the action, and use it in a path integral. But the mathematical technique for how to calculate this is not known. The best anybody had been able to do thus far is to make use of (8.25) to replace gauge fields with  $G_\mu \rightarrow \delta/\delta J^\mu$  and then remove  $\exp(V(\delta/\delta J))$  from the integral so all that remains behind to integrate is the simple  $\int dx \exp\left(-\frac{1}{2}Ax^2 - Jx\right)$ . Generally speaking, we need to replace the gauge fields  $G$  with current densities  $J$ , and leave behind the simple quadratic form  $\int dx \exp\left(-\frac{1}{2}Ax^2 - Jx\right)$ . What we find in section 8 is a new and different way to make a  $G \rightarrow J$  substitution in lieu of the usual  $G_\mu \rightarrow \delta/\delta J^\mu$ : recurse  $G_\mu = \left(k_\tau k^\tau - m^2 + i\epsilon + G_\tau k^\tau + G_\tau G^\tau\right)^{-1} J_\mu$  to any desired order, then set  $V = -G_\tau k^\tau - G_\tau G^\tau$  (because  $k_\tau G^\tau = 0$ ) to zero so that all gauge fields are removed. By recursing to infinite order and removing these gauge fields, we can arrive at an expression for  $G(J)$  with all the gauge fields removed, and be left with only having to integrate  $\int dx \exp\left(-\frac{1}{2}Ax^2 - Jx\right)$ . In short, the recursion preliminarily developed in section 8 provides the needed *mathematical tools* to carry out exact analytical calculations of what are now seemingly-intractable path integrations for non-linear *physical* field theories.

In sections 11 through 13, we then show how to apply these recursive results to calculate the non-linear Yang-Mills path integrals for both quark current and faux magnetic monopole densities over the gauge field portion  $DG$  of the path integral measure analytically and exactly, thereby proving the existence of a non-trivial relativistic quantum Yang-Mills theory exists on  $\mathbb{R}^4$  for any compact simple gauge group  $G$  by solving a mathematical challenge for which the solution has not previously been known. The results of this are in (13.13) / (13.16) for infinite recursion and (13.14) for finite recursion. Having used recursive technique to prove a quantum field theory for Yang-Mills, the question now arises whether recursive technique may be similarly applied to other non-linear field theories, most notably, gravitation.

In section 14, we starts at (14.3) with the amplitude density  $\mathfrak{N}(J)_1 = \text{Tr}\left(J_\sigma \pi_1 J^\sigma\right)$  at first recursive order, and uses this to derive (14.32) for the potential energy  $E_1$  between two  $J^\sigma$  as a function of radial distance. In Figures 1 through 4 we see how this leaves intact the normal  $r^{-1}$  character of this potential at very short distances, yet how at around  $1/6$  F based on  $\Lambda_{QCD}^{(6)} = (90.6 \pm 3.4) \text{ MeV}$  the potential changes its qualitative character and is dominantly driven by a  $\cosh\left(\frac{\sqrt{3}}{2}(4\pi\alpha_s)^{\frac{1}{6}} \rho_0^{\frac{1}{3}} r\right)$  contribution into being a confining potential. Given that we have now been able to use  $\mathfrak{N}(J)_1 = \text{Tr}\left(J_\sigma \pi_1 J^\sigma\right)$  from (13.21) to analytically show the qualitative features of confinement as between two  $J$ , and given that

$\mathcal{N}(P')_1 = -\frac{4}{3}\epsilon^{\alpha\beta\gamma\rho}\text{Tr}\left(P'_{\mu\nu\rho}\pi_1^{-1}(k_{1\mu}J_{\nu 1})^{-1}P'_{\alpha\beta\gamma}\right)$  is the amplitude density in this same (13.21) for two monopole baryons  $P'$ , a good next step would be to develop this amplitude in similar fashion (that is, Fourier transform it over from momentum space to configuration space) to see if it reveals a short range interaction between two  $P'$  densities. If this can be shown similarly to what was done in section 14, one would be able to more or less rest the case that baryons are indeed the magnetic monopoles of Yang-Mills gauge theory.

Another important step is to see if this can be connected to numerically-precise empirical observations relating to protons and neutrons. Among the important unexplained data that we already know about for protons and neutrons are their masses, as well as their binding energies in a wide variety of nuclei. Thus, it becomes important to calculate energies and as pointed out at (10.13), the way to do so is to use (10.13) in the general energy formulation  $E = -\iiint \mathcal{L}_{\text{gauge}} d^3x$ , using a combination of  $\text{Tr}(F_{\sigma\tau}F^{\sigma\tau})$  inner and  $\text{Tr}F_{\sigma\tau}\text{Tr}F^{\sigma\tau}$  outer product terms. While we do not do so in this paper, the author has done so before, and published these results in [15], [16] and [21]. Beyond the clear symmetry concurrences developed in section 10, these empirical concurrences provide compelling experimental support for the concluding that the non-zero faux magnetic source densities  $P' = -id[G, G] = -idGG$  are baryon densities, that  $\iiint P'$  is a baryon, that  $F_{\text{eff}\mu\nu} = -i[G_\mu, G_\nu]$  in (10.4) is a meson field, and that the  $\oint\oint F \neq 0$  which originally actuated this whole line of development represents the interaction of these baryons via mesons, and indeed the nuclear interaction protons and neutrons at classical level. As discussed, although these symmetries were all developed using the classical theory, there is no apparent reason why these symmetries would be lost in the  $DGDcDc^\dagger$  measure of the complete path integral  $Z = \int DGDcDc^\dagger \exp\left(i\left[S(G) - (1/2\xi)\int d^4x(\partial G)^2\right] + S(c^\dagger, c)\right)$  and would not carry over to the quantum field theory.

In fact, it is well known that the same color symmetries which have been classically developed in the present treatment solely emergent from classical Yang-Mills theory, do carry over to Quantum Chromodynamics.

Because the calculation of section 14 using an Abelian simplification which in which certain matrix inverses are treated as ordinary denominators, section 15 proceeds directly into a full non-abelian calculation with no simplification of the inverses. We discover in section 15 that the probability density for the source current  $J$  which is of course a density in a three dimensional space, i.e., which has dimensions of  $1/\text{volume}=1/\text{length}^3$ , cannot be properly treated in the non-abelian case without deconstructing the three-density into its one-dimensional component densities, i.e., into three separate  $1/x$ ,  $1/y$  and  $1/z$  densities in  $x$ ,  $y$ ,  $z$  Cartesian coordinates. The need to engage in such a deconstruction to properly develop the non-abelian theory reveals a substructure in the quantum field equations not dissimilar to the type of substructure that Dirac found when he pursued a linear expression for the energy-momentum relationship  $p_\sigma p^\sigma - m^2 = 0$  and came upon the fermion substructure represented in  $(\gamma^\sigma p_\sigma - m)u = 0$ .

Indeed, the upshot of the overall result obtained in (15.42) that there is a direct isomorphic relationship between a coupled quantum probability field designated  $h\mathcal{P}_0$  and a quantum potential  $E_1$  at the first order of recursion in the non-linear quantum field theory. This teaches that analytical non-linear quantum field theory effectively involves studying relationships between coupled probability fields and their associated underlying potentials which operate on propagating field quanta via least action principles. Thus, it is desirable to study what is found in section 15 using some specific, paradigmatic quantum probability densities to help flesh out something of an “operator’s manual” for applying post-path-integration analytical non-linear quantum field theory.

Sections 16 through 20 develop what are effectively four specific examples of the application of non-linear quantum field theory, each of which is informative as to how non-linear quantum field theory actually works when analytically applied. Section 16 shows that the classical  $-1/r$  potential of Coulomb emerges in the special case where the probability density is taken to be zero for all  $r$  except  $r=0$ , and a portion of section 19 later shows that the probability density for the Coulomb potential must be taken to be a Dirac (half-)delta right at  $r=0$ .

Section 17, which deepens the development of section 14, shows how a constant isotropic probability density bounded within a limited spatial region is isomorphically associated with a confining potential. This may appear at first impression to be a tautology, because if we posit a spatially-bounded probability density then we are of course confining the locales at which a field quantum contributing to that bounded probability may situate. But the point is that one cannot simply posit a constant bounded probability density and then expect that density to hold together by itself. Physics is still physics; it is not magic. Even in quantum theory there must be some explanatory cause and effect. So if one is going to posit a bounded probability density, then one must at the same time demonstrate that there is an associated potential which is capable of dynamically causing and enforcing that confinement. It is the isomorphism between the posited bounded constant probability density and the confining potential which provides us with the potential that is needed as a causal matter of dynamic physics, to maintain the constant bounded probability density.

Section 18 specifically examines the running of the strong QCD interaction coupling based on the confining potential and the constant bounded probability density developed in Section 17. The results obtained in this Figure 14 based on the theoretical equation (18.22) are able to match the theoretical results developed here, to within the observed error bars of the empirical running coupling data. And, to the extent that the predicted curve in Figure 14 may be nominally higher than the fitted PDG curve below  $M_Z$  and nominally lower above  $M_Z$ , it is shown via Figure 15 how this is both indicative of, and accounted for, by new physics which we know is to be expected in the GUT and Planck energy domains of  $10^{15}$  to  $10^{20}$  GeV, and results from positing that asymptotic freedom remains asymptotic right up to  $Q \rightarrow \infty$  rather than accounting for the likelihood that the asymptotic freedom behavior is likely to change once we enter the GUT to Planck-scale domain.

Section 19 develops the specific example of a Gaussian probability distribution, and shows that this distribution is isomorphically linked to a potential which is very close to the

Coulomb potential, differing only within small standard deviations about the Gaussian peak. This is how we show that in the limit where the Gaussian becomes a Dirac delta, the potential maps precisely over to the Coulomb potential as noted above in relation to section 16.

Section 20 tackles perhaps the most intriguing challenges of quantum field theory, namely, those presented by single and double slit experiments. Here we start with the sinc-based probability densities which are observed by detectors at a distance from the slits which greatly exceeds slit width and separation, and as was done in the previous three examples, we obtain the isomorphically-related potential. What we find is that the accumulation of field quanta on a detector in an interference-like probability density is in fact guided by this potential via long-standing principles of least action propagation. The quantum paradoxes of slit experiments come into play not by discarding least action as a causal principle to explain the dynamical evolution and propagation of a system of large numbers of field quanta, but by the fact that this guiding potential is itself induced in the quantum vacuum by interaction with the field quanta themselves as well as by the slit configuration, in contrast to a classical potential which is taken to be a preexisting background potential influenced minimally if at all by anything other than the source of that potential.

Taken together, the ability to analytically complete the path integration of the classical action and then study some specific examples of physical systems in non-linear quantum field theory, gives us some deep new insights into some of the most perplexing problems which have confronted theoretical physicists since the dawning of the quantum era at the outset of the 20<sup>th</sup> century.

## References

- 
- [1] H. Weyl, *Gravitation and Electricity*, Sitzungsber. Preuss. Akad.Wiss., 465-480. (1918).
  - [2] H. Weyl, *Space-Time-Matter* (1918)
  - [3] A. Einstein, *The Foundation of the General Theory of Relativity*, Annalen der Physik (ser. 4), 49, 769–822 (1916)
  - [4] H. Weyl, Electron und Gravitation, Zeit. f. Physik, 56 (1929), 330
  - [5] Yang, C. N.; Mills, R. (1954). *Conservation of Isotopic Spin and Isotopic Gauge Invariance*. *Physical Review* 96 (1): 191–195. Bibcode:1954PhRv...96..191Y. doi:10.1103/PhysRev.96.191
  - [6] A. Jaffe and E. Witten, *Quantum Yang–Mills Theory*, Clay Mathematics Institute (2000)
  - [7] Reinich, G.Y., *Electrodynamics in the General Relativity Theory*, Trans. Am. Math. Soc., Vol. 27, pp. 106-136 (1925)
  - [8] Wheeler, J. A., *Geometrodynamics*, Academic Press, pp. 225-253 (1962)
  - [9] Misner, C. W., Thorne, K. S., and Wheeler, J. A., *Gravitation*, W. H. Freeman & Co. (1973)
  - [10] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. 30, 1343–1346 (1973)
  - [11] Zee, A., *Quantum Field Theory in a Nutshell*, Princeton (2003)
  - [12] Close, F. E., *An Introduction to Quarks and Partons*, Academic Press (1979)
  - [13] L.D. Faddeev and V.N. Popov, Feynman Diagrams for the Yang-Mills Field, Phys. Lett. B25 (1967) 29
  - [14] Halzen, F., and Martin A. D., *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, J. Wiley & Sons (1984)
  - [15] Yablon, J. R., *Why Baryons Are Yang-Mills Magnetic Monopoles*, Hadronic Journal, Volume 35, Number 4, 399-467 (2012) <http://www.hadronicpress.com/issues/HJ/VOL35/HJ-35-4.pdf>

- 
- [16] Yablon. J. R., *Predicting the Binding Energies of the Is Nuclides with High Precision, Based on Baryons which are Yang-Mills Magnetic Monopoles*, Journal of Modern Physics (JMP) Vol.4 No.4A (April 2013) <http://www.scirp.org/journal/PaperDownload.aspx?paperID=30817>
- [17] Cheng, T-P and Li, L-F, *Gauge theory of elementary particle physics*, Oxford (1984) (2011 reprint)
- [18] Weinberg, S., *The Quantum Theory of Fields, Volume II, Modern Applications*, Cambridge (1996)
- [19] J. Yablon, *Grand Unified SU(8) Gauge Theory Based on Baryons which Are Yang-Mills Magnetic Monopoles*, Journal of Modern Physics, Vol. 4 No. 4A, 2013, pp. 94-120. doi: 10.4236/jmp.2013.44A011. <http://www.scirp.org/journal/PaperDownload.aspx?paperID=30822>
- [20] Volovok, G. E., *The Universe in a Helium Droplet*, Clarendon Press – Oxford (2003)
- [21] J. Yablon, *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, 2013, pp. 127-150. doi: 10.4236/jmp.2013.44A013. <http://www.scirp.org/journal/PaperDownload.aspx?paperID=30830>
- [22] [http://en.wikipedia.org/wiki/Invertible\\_matrix#Inversion\\_of\\_4.C3.974\\_matrices](http://en.wikipedia.org/wiki/Invertible_matrix#Inversion_of_4.C3.974_matrices)
- [23] [http://en.wikipedia.org/wiki/Cubic\\_function](http://en.wikipedia.org/wiki/Cubic_function)
- [24] <http://pdg.lbl.gov/2013/reviews/rpp2013-rev-qcd.pdf>
- [25] <http://pdg.lbl.gov/2010/reviews/rpp2010-rev-phys-constants.pdf>
- [26] <http://physics.nist.gov/cuu/Constants/index.html>
- [27] <http://pdg.lbl.gov/2013/tables/rpp2013-qtab-baryons.pdf>
- [28] J. Yablon, *Grand Unified SU(8) Gauge Theory Based on Baryons which Are Yang-Mills Magnetic Monopoles*, Journal of Modern Physics, Vol. 4 No. 4A, 2013, pp. 94-120. doi: 10.4236/jmp.2013.44A011. <http://www.scirp.org/journal/PaperDownload.aspx?paperID=30822>
- [29] [http://en.wikipedia.org/wiki/Sigmoid\\_function](http://en.wikipedia.org/wiki/Sigmoid_function)
- [30] Feynman, Richard P.; Robert B. Leighton; Matthew Sands. *The Feynman Lectures on Physics, Vol. 3*. US: Addison-Wesley. pp. 1.1–1.8 (1965).
- [31] <http://physics-animations.com/Physics/English/top10.htm>
- [32] [http://en.wikipedia.org/wiki/File:Double-slit\\_experiment\\_results\\_Tanamura\\_2.jpg#mediaviewer/File:Double-slit\\_experiment\\_results\\_Tanamura\\_four.jpg](http://en.wikipedia.org/wiki/File:Double-slit_experiment_results_Tanamura_2.jpg#mediaviewer/File:Double-slit_experiment_results_Tanamura_four.jpg)
- [33] <http://phys.educ.ksu.edu/vqm/html/doubleslit/index.html>