Reasons the photon mass should be exactly zero

Warren D. Smith
WDSmith@fastmail.fm

3 December 2003

Abstract —
We argue from general relativity that if photons had mass, then charged black holes could not exist. But there is evidence they do exist. Also, big violations of Maxwell's laws would occur, even in regions far from black holes. This is the first evidence that photons really have *zero*, as opposed to just a very small, mass. We also point out that the presence of extragalactic magnetic fields suggests $|m_\gamma| \lesssim 10^{-64}$kg. Two independent lines of argument, one based on extragalactic magnetic fields and very conservative bounds on current densities, the other based on Voyager spaceprobe magnetometer measurements, both find $|m_\gamma| \lesssim 1.7 \times 10^{-56}$kg. These numbers are respectively 10^{10} and 60 times stronger than the best previous experimental bound. Finally, we consider the fact that massless quantum fields have additional symmetries. While it is tempting to try to use this as a further justification for masslessness, it ultimately seems wrong.

Keywords — Photon, black holes, Proca equations, conformal invariance.

1 Introduction

The photon is usually assumed to be massless. But it might have some very small mass m_γ. There have been many experiments to detect or constrain a photon mass. The latest and tightest such experimental bound is $|m_\gamma| < 1.2 \times 10^{-54}kg= 7 \times 10^{-19}eV/c^2$ by Luo et al [10]. If photons had mass, then the Maxwell-Proca equations would be obeyed instead of the Maxwell equations (section 12.9 of [9]). The intercharge potential would be of Yukawa (i.e. $qe^{-r/\ell}/r$, ultimately exponentially declining with distance r) rather than Coulomb (q/r) form.

2 Improvements to the best known experimental bound

Before mentioning arguments that m_γ is exactly 0, let me mention several astronomy-based approaches to get big improvements in the Luo bound on $|m_\gamma|$. That bound may be rewritten as this bound on the Yukawa e-folding length $\ell = \hbar/(m_\gamma c) > 2$AU, where an AU ("astronomical unit") is 1.5×10^{11}meters, i.e. the rough distance from the Earth to the sun.

According to NASA press releases, the sun has a magnetic field of order 5milliTesla (i.e. about 100 times stronger than the Earth's field of 20-70μTesla) which reverses about once every 11 years. It is confined within a roughly spherical region called the "heliosphere" of radius 110-160 AU whose boundaries are now believed to be being detected by magnetometers on the Voyager 1 and 2 spacecraft. The absence of any Yukawa exponential decline in this field tells us that the Luo bound may be improved by a factor of at least 60. (The only possible reason for any doubt is the fact that the situation is complicated by solar wind and plasma effects. Since those effects have been heavily modeled and measured, a careful analysis of them should be possible.)

Supernovae are believed to generate much larger magnetic "bubbles" containing much larger fields. For example, the Crab Nebula is the remnant of a supernova observed in 1054 AD. It is now about 10 lightyears across and contains an average magnetic field (17 ± 3)nanoTesla, as judged from observations of gamma rays from synchrotron radiation [1]. This should yield an improvement over the Luo bound by a factor of at least 10^5.

The fact that our *galaxy* has a magnetic field of about 600 picoTesla at our location [13][1] has been shown by observations of Faraday rotation from polarized astronomical sources and by synchrotron radiation from charged cosmic ray particles. Furthermore, there are measurements of the same nature of *extragalactic* magnetic fields, even including inter*cluster* magnetic fields, of 200-1000 picoTesla [17][5][4]. However, considering the diameter of the Milky way is 6×10^9 AU, the distance to the Andromeda (our closest neighbor) galaxy is 1.3×10^{11} AU, and the distance to the Virgo cluster (comparable to its size) is 4×10^{12} AU, this suggests that the Luo bound may be

[1] And there are similar values at other locations and in other galaxies; their magnetic fields tend to be along the spiral arms

improved by factors of at least 10^9, 10^{10}, and 10^{11} respectively. Again, the only reason for doubt is the fact that the structures and sources of these fields are successively less and less well understood.

It seems hard to believe, though, that whatever process is generating intercluster magnetic fields is operating on much smaller length scales than the size of those clusters themselves. So, multiplying our final upper bound by a further safety factor of 10 to be conservative about these uncertainties, we find $|m_\gamma| \lesssim 10^{-64}$kg.

It is possible to use the lower limit $B > 200$ picoTesla on the size of the cited magnetic fields in extragalactic space to get a *rock solid* photon mass bound almost independent of any beliefs about the sources of those fields. Use the fact that the baryonic (i.e. charged) matter density of the universe (i.e., of intergalactic space) is < 0.2amu/meter^3. Since the temperature out there is 2.7°K, that means all that matter (assuming it consists of hydrogen atoms) is moving at an RMS thermal speed of 260meter/second. Even under the outrageous assumption that all the electrons in that matter are drifting at 260 meter/second in one direction relative to the plasma center of mass frame[2],for a current density of $j = 8 \times 10^{-18}$ ampere/meter^2, it still would be necessary, in order for the resulting currents to generate a 200 picoTesla field (the lowest of the several claimed observations of extragalactic fields cited) to have this happening in a cylinder of radius at least 256 AU (with drift direction axial), assuming Maxwell's equations hold. If the photon has mass, then even in a cylinder of unboundedly large radius, this current density would be unable[3] to generate a magnetic field exceeding the Maxwellian magnetic field from a cylinder of radius 2ℓ. Therefore the Yukawa length ℓ for photons *certainly* exceeds 128 AU. The same argument in one of the extragalactic regions with 1μTesla field would yield $\ell > 638$ AU. The 128 AU result improves the Luo bound by a factor of 64. This confirms the improvement in the Luo bound that comes from the Voyager magnetometer heliopause measurements. We conclude that it should be entirely safe to say $m_\gamma < 1.7 \times 10^{-56}$kg.

It also is possible to bound the current densities in extragalactic space by forbidding the ohmic heating they would cause from exceeding observed power fluxes. Since the results of that approach seem considerably weaker, we shall not discuss it.

3 Masslessless via Black holes

We now point out that even an arbitrarily small nonzero photon mass would have very large and observable astrophysical consequences. Specifically, by the same standard reasoning [2][7][16] as was used in the 1970s to show that black holes have no weak-force "hair," it would then follow from classical general relativity that it would be *impossible for black holes to have electric charge.* Intuitively, that is because the virtual photons that transmit the Coulomb intercharge force, would suffer infinite red shift trying to climb out of the black hole, and therefore effectively would have to travel for a longer distance than the Yukawa *e*-folding length, reducing the electric force felt by an external test charge essentially to zero. Hartle [7] and Teitelboim [16] argued that any nonelectromagnetic force between baryons or leptons transmitted by various kinds of possible carrier particles, would drop to zero as one of the particles approached a black hole event horizon. Bekenstein [2] further argued that, in the exterior of a static black hole, neither massive scalar nor vector (e.g. Maxwell-Proca) fields of that hole could even *exist.*

Charge would therefore not be conserved. Hawking [8], in now-standard reasoning, had argued that the process of black hole formation and evaporation would violate conservation of baryon and lepton number. For the same reason, after a black hole containing charge inside its event horizon later eventually evaporated, the charge of the universe would be different from its original value before the formation of the hole.

This would also violate Gauss's theorem that the total electric flux through a spherical surface always is equal to the charge inside, and always is constant while no current passes through that surface. But any sufficiently large-radius charged mass of matter will inevitably collapse into a black hole [11]. This is even if that mass initially has small bounded density. For example, a ball the radius of the solar system filled with matter at the same density as water would collapse to a black hole in a matter of hours. If that mass started out slightly charged, the resulting black hole would be charged.

Alternatively, if a spherically symmetric cloud of electrons were to contract into a preexisting Schwarzschild black hole, followed by a concentric spherically symmetric cloud of protons, and so on in alternation, we would get, effectively, a spherically symmetric oscillating charge.

All of these things would be a very dramatic violation of the integral form of Maxwell's equations at any length scale between kilometers and the size of solar system (and larger scales too). Even if the reader is not completely confident that such large scale violations of Maxwell's equations are impossible, then he must still agree that any experimental evidence proving charged black holes exist proves that the photon mass is *exactly* zero.

[2]This is a far larger ratio of charge-drift speed to RMS thermal speed than is achieved in all earthly conductors. A thin fully-ionized plasma with a drift speed comparable to or exceeding the ion thermal velocity would be unstable. That should have led to plasma shock waves from instabilities in the intergalactic medium; such are not seen.

[3]The magnetic field B on the surface of a radius-r cylinder filled with uniform axial current density j, assuming a photon Yukawa e-folding length ℓ, obeys $B \le \frac{\mu_0 j}{r} \int_0^r x e^{-|r-x|/\ell} \mathrm{d}x$. which becomes $B \le \mu_0 \ell j$ as $r \to \infty$. But $B \to \mu_0 r j/2$ in the massless photon limit $\ell \to \infty$.

Such evidence exists, although at present none of it seems completely convincing:

1. Gredin et al. [6] made polarimetric observations of the Cygnus-X1 black hole. From them they deduced that its accretion disk was generating at least a 3000 Tesla magnetic field. They then deduced a black hole charge equivalent in Planck units to about 1% of its mass. Assuming a 7 solar mass black hole, this is about 10^{19} Coulombs!

2. It has been speculated [14] that certain strong gamma ray bursts, especially the GRB 991216 burst, were caused by the formation of charged black holes with ultrastrong electric fields exceeding the threshhold necessary for e^+e^- pair creation.

4 Masslessness due to extra symmetries?

Is there any other reason we might expect the photon to be exactly massless? The gauge-invariance of QED would be lost with massless photons, but since it is certainly possible to build quantum field theories with massive vector bosons (indeed the weak force is thought to be so described) this by itself does not close the door.

The massless-photon Maxwell case of the Maxwell-Proca equations (in flat space) has additional symmetry. The Maxwell-Proca equations are invariant under Lorentz transformations and (in vacuum, or if we admit magnetic monopoles) also under a continuous 1-parameter group of transformations "rotating" electric and magnetic fields into each other (e.g. including the interchange of the two, in appropriate units).

But the Maxwell equations *also* are invariant under a change of scale (multiplying x, y, z, t by any nonzero constant). Much more generally, in vacuum they also are invariant under (what we shall call) "Lorentzian inversion" transformations. These nonlinear transformations were discovered by Cunningham in 1910 [3], and deserve to be better known. Therefore we will now reexpress them in modern language. Recall that the "inversions" of Euclidean n-space consist of the following transformation:

1. select a sphere and call its radius r_0;

2. replace each point at distance r from the sphere's center, by the point in the same direction, but at distance r_0/r.

It is a famous theorem of J.Liouville that, if $n \geq 3$, these (and their compositions and limiting cases) are the full set of angle-preserving self-mappings of Euclidean n-space. A remarkable property of inversions is that they always map spheres to other spheres (provided planes are regarded as the limiting case, of infinite radius, of a sphere). Now, the "Lorentzian inversions" are the same formulae as ordinary inversions, except that ict is employed instead of the time-coordinate t. Then a "sphere of radius r_0 and center x_0, y_0, z_0, t_0" becomes instead the "expanded light cone"

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 + (ict - ict_0)^2 = r_0^2 \tag{1}$$

i.e. a sphere, initially of radius r_0 when $t = t_0$, expanding at lightspeed thereafter. The corresponding inversion-map is

$$(x, y, z, t) \to (x_0, y_0, z_0, t_0) + r_0 \frac{(x - x_0, y - y_0, z - z_0, t - t_0)}{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 - c^2(t - t_0)^2}. \tag{2}$$

The remarkable sphere-mapped-to-sphere property of ordinary inversion maps now becomes the property that expanded light cones are mapped to expanded light cones.

It is curious why Cunningham's symmetry is there, since little or no use has been made of it by physicists so far. (As I.I.Rabi once said in a different context, "who ordered that?") Certainly most of the other equations of mathematical physics (e.g. the Dirac and Proca equations) are not invariant under these transformations; but Maxwell's are. One may verify that so are the scalar wave equation $F_{tt} = \nabla^2 F$, the equation that would be satisfied by *massless* neutrinos in vacuum, and the equations satisfied by the metric perturbation in linearized general relativity. So it appears that all vacuum wave equations satisfied by massless particles obey Cunningham's symmetries, but any mass, however small, breaks those symmetries.

If we now leave flat space and allow, more generally, an arbitrary $(3 + 1)$-manifold spacetime, then the corresponding statement is that the wave equations satisfied by massless spin-s particles, $s = 1/2, 1, 3/2, 2, \ldots$, are "conformally invariant" [12]. (This is also true in the spin-0 case, *provided* the "conformally coupled" massless scalar wave equation $\nabla_\mu \nabla^\mu F = \frac{n-2}{4(n-1)} RF$ is employed in n-dimensional spacetime where R is the scalar curvature. Physically $n = 4$.) That is, multiplying the metric tensor $g_{\alpha\beta}$ by any smooth positive-valued function of position (in differential-geometry lingo, a "conformal transformation"), changes the manifold but leaves the form of these wave equations unchanged.

This makes intuitive sense, because there is a physical length scale, the Compton wavelength $h/(mc)$, associated with any mass-m particle. So only massless particles can hope to behave the same if lengths are locally rescaled by a conformal transformation.

It is often said by particle physicists that at high energies there are additional symmetries of nature, causing a "unification." Here we have an example different from the usual ones. If all particles have very large kinetic energies compared to their rest masses, and if all length scales of interest are small compared to the Compton wavelengths $h/(mc)$ of the particles, then all conformal symmetries with a bounded conformal factor become approximately satisfied. Hence all physics should behave the same at high enough temperatures, regardless of the metric of spacetime, provided we restrict consideration to metrics that are conformally equivalent. Therefore, we might as well demand that our metric have *constant* scalar curvature R, since (at least [15][18] for Riemannian n-metrics with $n \geq 3$) every metric may be conformally distorted to one with constant scalar curvature.

Summary. If some reason were found why photons "should" have conformal symmetries, then that too would be a reason they must be exactly massless. But I know no such reason. Anybody trying to invent such a reason would have the onus of explaining why neutrinos *do* have mass. But more seriously, the metric itself, in general relativity, is not a static arena – thanks to gravity, it is dynamic. And there *is* a length scale, the Planck length $\ell_{\text{Planck}} = \sqrt{\hbar G_{\text{Newton}}/c^3} \approx 1.6 \times 10^{-35}$meter, associated with gravity even in the absence of all mass. So the true laws of physics, even in the absence of mass, if we include quantum gravity, cannot be invariant under changes of scale. Thus this additional symmetry seems a false god. If so, it ultimately cannot be the true justification for the masslessness of anything.

5 Conclusion

While none of the ingredients herein are original, nevertheless the way in which they are combined, and the conclusions that are reached, are new. We have described several possible arguments for the exact masslessness of the photon. The most convincing one is the vast violations of Maxwell's laws that would be caused by black holes. It seems very hard to believe such immense violations are possible.

Unfortunately the case lacks solidity until it is linked to *experimental* proof, and the astronomical observations so far remain incompletely convincing. We have argued that convincing astronomical observations already suffice to improve the best known upper bound on $|m_\gamma|$, and it now might be possible to get convincing astronomical evidence that $m_\gamma = 0$ (now that the incentive exists to seek it). For example, the 10^{19} Coulomb charge postulated by Gredin et al. [6] would generate an electric field of 4MV/meter even 1 AU away. Such large fields would cause spectral "Stark shifts" that perhaps would be detectable even from Earth. Finally, additional possible masslessness arguments based on symmetries seem ultimately to be incapable of being the real justification – leaving the question "who ordered conformal symmetry?" still a mystery.

References

[1] A.M.Atoyan & F.A.Aharonian: On the mechanisms of the gamma radiation of the Crab nebula, Monthly Notices Royal Astronomical Soc. 278 (1996) 525-541.

[2] J.D.Bekenstein: Nonexistence of Baryon number of static black holes, I: Phys.Rev.D 5 (1972) 1239-1246 and II: Phys.Rev.D 5 (1972) 2403-2412. Followup papers on this include J.D.Bekenstein: Transcendence of the law of baryon number conservation in black hole physics, Phys.Rev.Lett. 28 (1972) 452-455, S.A.Adler & R.P.Pearson: "No hair" theorems for abelian Higgs and Goldstone models, Phys.Rev. D 18 (1978) 2798-2803, and (probably most relevent for the purposes of the present paper) E.Ayón-Beato: Improving the "no-hair" theorem for the Proca field, `gr-qc/0210001`.

[3] E.Cunningham: The principle of relativity in electrodynamics and an extension thereof, Proc. London Math'l. Soc. 8 (1910) 77-98.

[4] L.Feretti, D.Dallacasa, G.Giovannini, A.Tagliani: The magnetic field in the Coma cluster, Astronomy & Astrophysics 302 (1995) 680-.

[5] L.Feretti & 5 others: The radio galaxies and the magnetic field in Abell 119, `astro-ph/9902019`, Astronomy & Astrophysics, in press.

[6] Yu.N.Gredin & 4 others: Magnetic and electric fields around the black hole in Cygnus X1, `astro-ph/0304158`.

[7] J.B.Hartle: Long range neutrino forces exerted by Kerr black holes, Phys.Rev. D 3 (1971) 2938-2940.

[8] S.W.Hawking: Breakdown of predictability in gravitational collapse, Phys.Rev. D 14,10 (1976) 2460-2473.

[9] J.D.Jackson: Classical electrodynamics, second edition, Wiley 1975.

[10] Jun Luo, L-C. Tu, Z-K. Hu, E-J. Luan: New experimental bound on the photon rest mass with a rotating Torsion balance, Phys.Rev.Lett. 90,8 (28 Feb 2003) 081801 1-4.

[11] C.W.Misner, K.S.Thorne, J.A.Wheeler: Gravitation, Freeman 1973.

[12] R.Penrose & W.Rindler: Spinors and Spacetime (2 vols), Cambridge university press (1984-1986).

[13] Wolfgang Reich: Observations of galactic magnetic fields, Acta Astronomica Sinica (Feb. 2003) 136-143 (review article).

[14] Remo J. Ruffini: Observational tests of the electromagnetic black hole theory in gamma ray bursts, `astro-ph/0202454`. Related matters are discussed in Ruffini et al. Astrophys. J. Lett. 555 (2001) L107, L113, L117 and Nuovo Cimento B 116 (2001) 99-108, and in B.Punsly: Black hole gravitohydromagnetics, Springer 2001.

[15] R.Schoen: Conformal deformation of a Riemannian metric to constant scalar curvature, J.Differential Geometry 20 (1984) 479-495; A report on some recent progress in nonlinear problems in geometry, Surveys in Differential Geometry 1 (1991) 201-241.

[16] C.Teitelboim: Nonmeasurability of the baryon number of a black hole, Lett.Nuovo Cimento II 3 (1972) 326-328 and 397.

[17] J.P.Vallee: A possible excess rotation measure and large scale magnetic field in the Virgo supercluster of galaxies, Astronomical J. 99 (1990) 459-469.

[18] Rugang Ye: Global existence and convergence of the Yamabe flow, J.Differential Geometry 39 (1994) 35-50.