LA-UR- 08-5197

Approved for public release;
distribution is unlimited.

Title: Photon and Graviton Mass Limits

Author(s): Alfred Scharff Goldhaber, SUNY
Michael Martin Nieto, Los Alamos National Laboratory

Intended for: Reviews of Modern Physics

Los Alamos
NATIONAL LABORATORY
EST. 1943

August 2, 2008

Photon and Graviton Mass Limits

Alfred Scharff Goldhaber*,† and Michael Martin Nieto†

*C. N. Yang Institute for Theoretical Physics, SUNY Stony Brook, NY 11794-3840 USA
and
†Theoretical Division (MS B285), Los Alamos National Laboratory, Los Alamos, NM 87545 USA

We review past and current studies of possible long-distance, low-frequency deviations from Maxwell electrodynamics and Einstein gravity. Both have passed through three phases: (1) Testing the inverse-square laws of Newton and Coulomb, (2) Seeking a nonzero value for the rest mass of photon or graviton, and (3) Considering more degrees of freedom, allowing mass while preserving gauge or general-coordinate invariance. For electrodynamics there continues to be no sign of any deviation. Since our previous review the lower limit on the photon Compton wavelength (associated with weakening of electromagnetic fields in vacuum over large distance scales) has improved by four orders of magnitude, to about one astronomical unit. Rapid current progress in astronomical observations makes it likely that there will be further advances. These ultimately could yield a bound exceeding galactic dimensions, as has long been contemplated. Meanwhile, for gravity there have been strong arguments about even the concept of a graviton rest mass. At the same time there are striking observations, commonly labeled 'dark matter' and 'dark energy' that some argue imply modified gravity. This makes the questions for gravity much more interesting. For dark matter, which involves increased attraction at large distances, any explanation by modified gravity would be qualitatively different from graviton mass. Because dark energy is associated with reduced attraction at large distances, it *might* be explained by a graviton-mass-like effect.

Contents

I. INTRODUCTION

Photons and gravitons are the only known "free" particles whose rest masses may be exactly zero.[1] This bald statement covers a rich and complex history, from Newton and Gauss, through Maxwell and Einstein, and even up to the present. During that development, the matrix of interlocking concepts surrounding the notions of photon and graviton rest mass, or, more generally, long-distance, low-frequency deviations from Maxwell electrodynamics and Einstein gravity, has become increasingly elaborate.

There are many similarities between the photon and graviton cases, but also striking differences. We literally see photons all the time, as only a few photons of visible light are enough to activate one 'pixel' in a human retina. Besides that, conspicuous electromagnetic wave phenomena play an enormous role in modern physics.

For gravity the situation is radically different. Even gravitational waves are in a situation analogous to that of neutrinos during the first 25 years after their proposal by Pauli, when their emission could be inferred from loss of energy, momentum, and angular momentum in beta decay, but they hadn't yet been detected in an absorption experiment. Binary pulsar systems exhibit energy loss well accounted-for by radiation of gravitational waves, but experiments underway to detect absorption of such waves have not yet achieved positive results. Even if this were accomplished sometime soon, the chances of ever detecting individual quanta – gravitons – seem remote indeed, because graviton coupling to matter is so enormously weak.

These are not the only differences. For electrodynamics, the history has been one of increasingly sensitive null experiments giving increasingly stringent negative results for any possible mass. Nevertheless, theories describing such a mass seem well-developed and consistent, even if not esthetically appealing. On the other hand, for gravity, there *are* long-distance effects which some argue provide evidence supporting modification of Einstein's formulation. At the same time, the theoretical basis for a gravitational phenomenon analogous to photon mass has come under severe attack. All this means that currently there is much more dynamism in the issue of deviations for gravity than for electrodynamics.

Even though quantum physics gave shape to the concept of mass for electrodynamics and gravitation, the obvious implication of a dispersion in velocity with energy for field quanta, or even for waves, is beyond our capacity to detect with methods identified so far. This is thanks to the very strong limits already obtained based on essentially static fields. Thus, the domain of potentially interesting experiment and observation for mass or "mass-like" effects indeed is restricted to the long-distance and low-frequency scales already mentioned.

A. How to test a theory

Let us begin by seeking a broad perspective on what it means to probe, not merely the validity but also the accuracy of a theory. The canonical view of theory-testing is that one tries to falsify the theory: One compares its predictions with experiment and observation. The predictions use input data, for example initial values of certain parameters, which then are translated by the theory into predictions of new data. If these predicted data agree with observation within experimental uncertainty (and sometimes also uncertainties in application of the theory), then the theory has, for the moment, passed the test. One may continue to look for failures in new domains of application, even if the incentive for doing so declines with time.

Of course, without strong 'ground rules' it is impossible to falsify a theory, because one almost always can find explanations for a failure. So, in fact no scientific theory either may be disproved or proved in a completely rigorous way; everything always is provisional, and continual skepticism always is in order. However, based on a strong pattern of success a theory can earn trust at least as great as in any other aspect of human inquiry.

The above is an essential, but we believe only partial, view of how theories gain conviction. At least three important additional factors may help to achieve that result. First, a striking, even implausible, prediction is borne out by experiment or observation.

Examples of this include 'Poisson's spot', Poisson's devastating attack on the notion that light is a wave phenomenon, because this would require that the shadow of a circular obstacle have a bright spot in its center. The discovery of the spot by Arago provided the conceptual equivalent of a judo maneuver, using the opponent's own impulse to overcome him. Another example is the assertion by Appelquist and Politzer in the summer of 1974 that the existence of a heavy quark carrying the quantum number 'charm' would imply the existence of a positronium-like spectrum, meaning very sharp resonances in electron-positron scattering. When the J/ψ was discovered at BNL and SLAC in November of that year, the outlandish prediction of Appelquist and Politzer suddenly was the best explanation, in good part because it was the only one that had been stated boldly (though not yet in print) beforehand.

A second way in which a theory gains credence is by fecundity: People see ways to apply the idea in other contexts. If many such applications are fruitful, then by the time initial experimental verification is rechecked there may be little interest, because the theory already has become a foundation stone for a whole array of applications. An example from our subject here is the transfer of the

[1] Gluons, the gauge particles of quantum chromodynamics, are believed to have no bare mass. However, they are not seen in isolation, meaning they cannot be observed as free particles.

$1/R^2$ force law from gravity to electricity, in a speculative leap during the 1700s. Of course, it was not this transfer which gave Newton's gravity its great authority, but rather the enormous number of precise and successful predictions of his theory – fecundity in the original, literally astronomical domain.

Closely related to the above is a third feature, connectivity. If many closely neighboring subjects are described by connecting theoretical concepts, then the theoretical structure acquires a robustness which makes it increasingly hard – though certainly never impossible – to overturn.

The latter two concepts fit very well with Thomas Kuhn's epilogue to his magnum opus *The Structure of Scientific Revolutions* (1), in which he muses that the best description of scientific development may be as an evolutionary process. For biological evolution, both fecundity and reinforcing connections play decisive roles in how things happen. It seems to us, as it did to Kuhn, that the same is often true for *ideas* in all science, including physics.

To the extent that observational errors can be ruled out, whenever discrepancies appear between theory and experiment one is compelled to contemplate the possibility that new, or at least previously unaccounted-for, physics is contributing to the phenomena. A classic case is the famous solar-neutrino puzzle. At first, presumed errors in the actual measurements themselves were widely and caustically viewed as the problem. Later critiques focused more on consideration of errors in models of processes in the Sun and, perhaps more creatively, on potential new physics modifying the simplest picture of neutrino propagation from source to detector. By now there is overwhelming evidence that neutrino mixing, a modification of neutrino propagation, accounts precisely for the observed rate of neutrino observation on Earth, confirming the basic validity of early work on both solar modeling and neutrino detection.

A more refined question about confirming a theory is: How does one quantify limits on deviations from the theory? Now it becomes necessary to specify some form of deviation that depends on certain parameters. Then experimental uncertainties can be translated into quantitative limits on these parameters. As a theory evolves, the favored choice for an interesting form of deviation may change. Of course, at any point new observations contradicting even well-established theoretical predictions can reopen issues that might have seemed settled

B. Photons

Our earlier review (2) described the state of theory for accommodating a non-zero photon mass, and the state of observation and experiment giving limits on the mass at that time. Since then there have been significant theoretical developments (Section II), as well as advances in experimental approach and precision (Section III).

The issue of possible long-range deviations from the existing theory of electrodynamics long remained purely a matter of choosing, and then setting limits on, parameters of a possible deviation. Originally, in analogy to studies of Newton's Law for gravity, deviations in power of radius from that in the inverse square law were used. 20th-century relativistic wave equations led to discussion in terms of a finite rest mass of the photon, thus introducing a length scale. Today one can consider a more sophisticated approach incorporating gauge invariance even when describing nonzero photon mass, and allowing even more parameters.

We shall not reprise in detail the theoretical paradigms and experimental details discussed in Ref. (2) for the photon. Rather, we refer the reader to that work for an introduction, and focus here on elucidating more recent advances. Also, since the publication of (2), there have been other works which have summarized specific aspects of the photon mass (3)-(9). These can be consulted especially for experimental summaries. In particular, Byrne (3) concentrated on astrophysical limits, Tu and Luo (6) on laboratory limits based on tests of Coulomb's Law, and, joined by Gillies, on all experimental tests (7). Okun, in a compact review (8), gives some interesting early history on the concept of the "photon" in quantum mechanics. He also gives details on what the Russian school accomplished during the earlier period.

Of course the types of deviation we discuss in this review do not cover all possibilities. In particular, the Maxwell equations are linear in the electromagnetic field strengths. This does not mean that all phenomena are linear, because the coupling between fields and currents allows back-reaction and thus nonlinearity. Nevertheless, the linearity of the equations is an especially simple feature. Higher-order terms in the field strengths clearly are an interesting possibility, but they are intrinsically tied to short-distance modifications of the theory, rather than the long-distance deviations emphasized here. The reason is that the nonlinear terms become more important as the field strengths increase, meaning that the numbers of flux lines per unit area increase, clearly a phenomenon associated with short distance scales.

For completeness, we briefly mention here discussions in the literature that go in the direction of nonlinearity. Born and Infeld (10) introduced the notion of nonlinear damping of electromagnetic fields, precisely to cope with the short-distance singularities of the classical linear theory. Their approach was pursued by a number of investigators over many years, as reviewed by Plebanski (11). More recently their ideas have been revived because the kind of structure they discussed arises naturally in string theory, as reviewed by Tseytlin and Gibbons (12; 13). A second approach to nonlinearity was introduced by Heisenberg and Euler (14), who observed that what we now would call virtual creation of electron-positron pairs leads inevitably to an extra term in the Maxwell equations that is cubic in field strengths. This work also has a living legacy, as reviewed recently by Dunne (15).

C. Gravitons

The scientific question in gravity most naturally related to photon mass is the issue of a possible graviton mass. However, at least to our knowledge, there never has been a review on this topic. Its study progressed more slowly than the photon-mass issue, at least in part because gravity is so weak that even today classical gravitational waves have not been detected directly. Further, gravitons, regardless of their mass, seem beyond the possibility of detection in the foreseeable future.

We therefore proceed to discuss theoretical issues (Section IV) and observations (Section V) for the case of classical gravity, just as our discussion of electromagnetism is most germane for the classical theory. There are several important contrasts between the photon and graviton cases. First, from a theoretical point of view the possibility of nonzero graviton mass is open to question. This makes what is a relatively straightforward discussion for photons much more problematic for gravitons. Secondly, there is a highly developed formalism for seeking to measure deviations of gravity from Einstein's General Theory of Relativity – the parametrized post-Newtonian [PPN] expansion.

In this framework there have been many measurements, principally under weak-field conditions, both for low-velocity and high-velocity phenomena to test for deviations. As with photon mass, none of these measurements to date have produced "unexpected-physics" results, only increasingly stringent limits on departures from Einstein gravity.

However, there is another important distinction. Two sets of characteristic phenomena show significant departures from Einstein gravity with the matter sources being only familiar "visible" matter – stars, hot gases, and photons. The first, indicated already by observations in the 1930s, and much more definitely in the 1970s, has been labeled "dark matter." Trajectories of visible objects (including the most visible of all – light itself) seem to be bent more than would be expected if the only sources for gravity were pieces of visible matter. In principle, a possible explanation for this could be long-range modifications of Newtonian and Einsteinian gravity, but of a type very different from what would be called graviton mass. The second departure, discovered much more recently, is an accelerated expansion of the universe neatly described by the presence of another sort of invisible source, "dark energy." We discuss these issues also in Section IV.

D. Overview

This review, then, can be considered an evolution of our earlier review on the mass of the photon (2). Here we discuss current understanding and ideas on the masses of the photon and graviton, in light of the many developments in theoretical and experimental physics over the past decades.

Because in principle there is no end to the types of deviation that could be contemplated, we need some restrictions. We therefore consider only deviations that obey abstract symmetries, gauge invariance in the case of electrodynamics, and general-coordinate invariance in the case of gravity. It might seem that this excludes a mass for the Proca photon or the graviton. However, as we discuss, the Higgs mechanism can 'hide' a symmetry, which nevertheless remains unbroken. For the photon that becomes equivalent in a certain limit to the fixed Proca mass, which therefore needn't break gauge invariance after all.

In addition we survey possible phenomenologically implied modifications to gravity, which would give alternative views of the observations commonly ascribed to 'dark matter' and 'dark energy.' In other words, in principle these effects could be wholly or partly due to modifications of gravity, rather than previously unknown sources of gravity.

II. ELECTROMAGNETIC THEORIES

As indicated in the introduction, one can mark three stages in the search for long-range deviations from electrodynamics, of which the first two were described in our previous review (2). That article appeared just as a 'sea change' in the theoretical picture of physics was beginning to emerge, the notion that gauge theories and gauge invariance might underlie not only electromagnetism and gravity but also weak and strong interactions. Nevertheless, until quite recently there had been surprisingly little discussion of the new perspective in the context of photon mass.

A. Power-law deviation from Coulomb's form

The method of choosing parameters for a deviation from a law depends on a matrix of aesthetic considerations and theoretical patterns, which tend to grow more definite as the conceptual framework develops. We can see this happening in the history of electricity and magnetism.

The first stage, as recounted in (2), focused on the inverse-square force for the interactions of electric charges or magnetic poles. The guess was that the strength of the electric force along the line between two charges would be similar to Newton's Law,

$$F = \frac{kq_1q_2}{r^2}. \tag{1}$$

Early experimenters chose to parametrize possible deviations from this form in a scale-invariant manner, presumably because they had no framework to choose a dimensional parameter instead. Therefore they looked for modifications of the form

$$F = \frac{kq_1q_2}{r^{2+\alpha}}, \tag{2}$$

and sought limits for a possible shift in power α from the inverse-square. This early history, which started before Coulomb (although Coulomb eventually received credit for the law) is described in Ref (2) and even more completely in Ref. (16). Indeed, experimenters used this parametrization up to the mid-20th century (17).

Even at early times, any departure from the inverse-square law was seen to violate an appealing geometric principle: the conservation of the number of lines of force emanating from a charge. (The force, by definition, is proportional to the number of lines per unit area.) For nonzero α, the electric flux coming out of a charge is radius-dependent – there is no Gauss law relating charge and flux. (See Section II.C.)

Then, just around the time of the appearance of Ref. (17) a competing, scale-dependent form of deviation began to seem more appropriate. This more sophisticated reasoning, and its development, has governed the discussion of possible deviations in later formulations of the issue.

B. Photon mass from the Proca equation

The new stage arose after two break-throughs. The first was the electrodynamics of Maxwell and Lorenz[2] which, when fully articulated, included among its solutions freely moving electromagnetic waves naturally identified with light. The second was the realization, beginning with Einstein, that if there are particles of light that are exactly massless, as they are in Maxwell's scale-invariant theory, they travel at the ultimate speed $c = 1/[\epsilon_0\mu_0]^{1/2}$.

Alexandre Proca (20)-(24), under the influence of de Broglie, introduced a consistent modification of Maxwell's equations which would give a nonzero mass to the photon while preserving the invariance of electrodynamics under transformations of special relativity. In modern notation designed to make relativistic invariance manifest, with electric and magnetic field strengths measured in the same units, the Lagrangian density Proca wrote is

$$\mathcal{L} = -F_{\alpha\beta}F^{\alpha\beta}/4 - m^2c^4A_\alpha A^\alpha/2(\hbar c)^2 \ , \qquad (3)$$

with

$$F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha. \qquad (4)$$

Here the main notational changes from Proca's original form are to describe the photon vector potential by A_α and to include contravariant vectors (with the metric signature spacelike positive).[3]

This Lagrangian naturally elicits the notion that the photon might have a small but still nonzero rest mass. The obvious implication, and the only one discussed by Proca, is a dispersion of velocity with frequency. In fact, the classical field equations derived from Proca's start imply not only velocity dispersion, but also departures of electrostatic and magnetostatic fields from the forms given by Coulomb's law and Ampère's law. We shall see that these implications give more sensitive ways to detect a photon mass than the observation of velocity dispersion.

The Maxwell equations as modified to Proca form are, beginning with the Gauss law,[4]

$$\nabla\cdot\mathbf{E} = \rho/\epsilon_0 - \mu^2 V \ , \qquad (9)$$

where μ is the photon rest mass in units of the inverse reduced Compton wave length,

$$\mu = \frac{1}{\lambda\!\!\!^{-}_C} = \frac{mc}{\hbar} \ , \qquad (10)$$

and where $(\mathbf{A}, V)$ is the now observable 4-vector potential.

Equation (9) implies a 'Yukawa' form for the potential due to a point charge q at the origin of coordinates:

$$V(r) = \frac{q}{4\pi\epsilon_0}\frac{e^{-\mu r}}{r} \ . \qquad (11)$$

Note the exponentially decreasing factor, which gives a departure from the inverse-square law for the electric field, scaling with the length $\lambda\!\!\!^{-}_C = \mu^{-1}$.

A similar phenomenon occurs in magnetism, with

$$\nabla\times\mathbf{B} = \mu_0\left(\mathbf{J} + \epsilon_0\frac{\partial\mathbf{E}}{\partial t}\right) - \mu^2\mathbf{A} \ , \qquad (12)$$

[2] In the period 1862-1867 the Dane Ludwig Lorenz independently derived the "Maxwell equations" of 1865, but received relatively little credit for this work (18; 19).

[3] Neither the titles nor the detailed texts of Proca's papers indicate explicitly that this is an equation for the electromagnetic field. Indeed, from the context it is clear that he was thinking of a charged, massive spin-1 field. The idea that this could be identified with a massive photon came later. We have converted to modern sign conventions for $\mathcal{L}$, but of course this has no effect on the free-field equations of motion.

[4] Here we adopt SI units. This is not the usual fashion in modern particle physics but it simplifies calculation of photon mass limits from astrophysical data, which increasingly are the most pertinent sources of new and better values. For the record, the usual notation is

$$\nabla\cdot\mathbf{E} = 4\pi\rho - \mu^2 V, \qquad (5)$$

$$\nabla\cdot\mathbf{B} = 0, \qquad (6)$$

$$\nabla\times\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}, \qquad (7)$$

$$\nabla\times\mathbf{B} = \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j} - \mu^2\mathbf{A}, \qquad (8)$$

The remaining Maxwell equations,

$$\nabla\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}, \qquad (13)$$

$$\nabla\cdot\mathbf{B} = 0 \ , \qquad (14)$$

equivalent to the definitions of the field strengths in terms of the potentials,

$$\mathbf{B} = \nabla\times\mathbf{A} \ , \qquad (15)$$

$$\mathbf{E} = -\nabla V - \frac{\partial\mathbf{A}}{\partial t} \ , \qquad (16)$$

are unchanged by the introduction of a photon mass.

Interestingly, although the moment Proca wrote down his equations (21)-(24) the finite range of static electromagnetic forces was implied, as far as we can tell Proca himself never drew this inference. (He and de Broglie (25) focused on velocity dispersion.) In 1935 Hideki Yukawa did recognize this consequence for a scalar or Klein-Gordon particle (26). From the finite range of the nuclear force he predicted that a new massive particle should be found, a prediction eventually vindicated by discovery of the π meson. In a later paper (27), Yukawa referred to Proca (21). So did Kemmer (28), who observed the equivalence of his 10-dimensional spin-1 solution to the 4-vector and antisymmetric tensor of Proca (21).

Schrödinger, in a number of papers (29)-(32), emphasized the link between photon mass and a finite range of static forces. Interestingly, in 1943 Schrödinger (31) mentioned Yukawa in this connection, but said nothing about Proca, although earlier he had mentioned de Broglie (29). Further, Schrödinger appears to have been the first to write the two massive Maxwell equations (9) and (12) in modern format (31). Finally, as briefly mentioned above, an abstract but important symmetry appears to be violated by the mass terms in Proca's equations: gauge invariance. (See Sec. II.D.)

Two comments are in order: First, radiation effects labeled as coming from a nonzero photon mass, including dispersion in velocity of photons (or even of classical electromagnetic waves), were found to be too small for observation. Therefore, as already mentioned, effects in classical electrostatics and magnetostatics were focused on. This is despite the fact that early tests seem to have been inspired by the marriage of quantum physics (implying light-particles or photons) with the special theory of relativity,

Secondly, a consequence of this proposed (Proca) deviation from Maxwell theory, like the departure from the inverse-square power law discussed earlier, is a violation of the Gauss law. This time the term $-\mu^2 V$ in Eq. (9) may be interpreted as giving a density of 'pseudo-charge', compensating for the charges of ordinary electrified particles.

Before continuing, this is a good point to say something about choices of scale. Physicists constantly adjust scales to make them convenient, without needing very large or very small exponential factors. In the case of photon mass, the limit even at mid-twentieth century was so low that all familiar choices of mass, or energy, or even frequency scale required exponential factors. At that time, with a lower limit on $\lambdabar_C$ comparable with the radius of the earth, the corresponding period of oscillation for a photon at rest, assuming that the actual rest mass saturated the limit, would have been of order 0.1 s. At first sight that seems a very manageable number, but if one tries to imagine a process of producing or observing a massless photon of such a long period, it quickly becomes absurd.

Put differently, such low-frequency photons are essentially unobservable as single objects. At best, one might hope to use an ultra-low-frequency circuit to detect a classical wave corresponding to an enormous number of individual photons. Thus, then, and even more so now, the only meaningful measure of photon mass less than or equal to the limit is in terms of Compton wavelength, i.e., phenomena observable for classical, long-range, static, electric or magnetic fields. Even though we shall quote limits on mass expressed in other terms, those values will be so far from ones we could measure directly, or ones that have been measured for any other kind of object, that they have only formal interest. Nevertheless, for the record let us state the relations that determine those values, using $\hbar \equiv \lambdabar_C mc$:

$$\lambdabar_C[\mathrm{m}] \equiv \frac{1.97\times10^{-7}}{m[\mathrm{eV}]} \equiv \frac{3.52\times10^{-43}}{m[\mathrm{kg}]}, \qquad (17)$$

We have then in units of c^2,

$$1\ [\mathrm{kg}] \equiv 5.61\times10^{35}\ [\mathrm{eV}]. \qquad (18)$$

C. Conservation of electric charge

There is a deeper level in the aesthetic considerations supporting vanishing photon mass, arising from an elaboration of the Gauss law. If the electric flux out of any surface measures the total electric charge enclosed, then special relativity assures that charge must be locally conserved. This is because the only way charge can change is by changing the flux at the same time, and for a distant surface that flux could not change instantaneously if the charge changed. More specifically, manifest gauge invariance (ignoring the Stueckelberg-Higgs mechanism discussed in the following subsection) implies local charge conservation through the invariance of the integrated $\mathbf{J}\cdot\mathbf{A}$ term in the action. Thus this conservation law is consistent with vanishing photon mass, as both follow from manifest gauge invariance.

Weinberg (33), assuming only special relativity for the scattering matrix (taken to lowest order in coupling), demonstrated a stronger result, that vanishing photon (graviton) mass implies vanishing four-divergence of the electric current (energy-momentum tensor) density. This tells us two things. If the photon or graviton mass vanishes, we have an explanation for another accurately verified observation, local conservation of electric charge or of energy and momentum. If not, we are allowed to contemplate the possibility of processes violating local conservation.

For electrodynamics this question has been studied by Okun and collaborators (34–36), by Nussinov and collaborators (37–39), and Tsypin (40). Perhaps the most interesting aspect is that coupling of longitudinal photons to electric currents, which vanishes for conserved current in the zero mass limit, now becomes divergent as the photon mass goes to zero. In view of Weinberg's result this makes sense: If zero mass implies conserved current, then a term violating the conservation would be "resisted" by the electromagnetic field (i.e., its otherwise decoupled longitudinal part), which would radiate furiously to compensate.[5]

There is an important additional point. Charge nonconservation destroys the renormalizability of perturbative quantum electrodynamics. The theory begins to resemble gravity in that the latter theory also is not renormalizable (even with locally conserved energy and momentum). Also, in the case of a massive graviton, the lowest-order perturbative theory for interaction between two gravitational sources does not limit with decreasing graviton mass to the result for zero graviton mass, as we discuss later in this article.

Thus, electrodynamics with non-conserved charge prefigures many of the features found in quantum gravity. Nussinov (37) suggested that the regulator energy cutoff Λ in a theory where charge conservation is violated may be connected to the photon mass by the relation $\mu \approx \delta e \Lambda$, with δe the coefficient of an effective charge-violating coupling such as $\bar{\psi}_e \gamma_\alpha \psi_\nu A^\alpha$.

We finally mention here an intriguing work that at least opens the possibility for a future proof that the photon mass must be identically zero. Rosenstein and Kovner studied electrodynamics in 2+1 dimensions, and concluded that a magnetic flux condensate would form, assuring zero photon mass (41). If their method could be extended to 3+1 dimensions it might yield such a proof.

D. Gauge Invariance: Its violation and restoration

Already in pre-quantum physics, the significance of continuous symmetries such as translation invariance in space and time, as well as rotation invariance, and their links to conservation laws of momentum, energy, and angular momentum, had been recognized. On top of this, the Maxwell equations admit another symmetry, classical gauge invariance, or gauge invariance of the first kind. (See the review of Jackson and Okun on gauge invariance (19).) From the relations (15) and (16) one finds that the electromagnetic field strengths $\mathbf{E}$ and $\mathbf{B}$ are unchanged by the transformations

$$V \to V' = V - \frac{\partial \Lambda}{\partial t}; \qquad \mathbf{A} \to \mathbf{A}' = \mathbf{A} + \nabla \Lambda \ . \tag{19}$$

Without the explicit μ^2 terms added to two of the Maxwell equations, the equations and the corresponding action is invariant under the transformations (19). With the μ^2 terms, and if one also assumes that the electric charge and current densities obey the equation of continuity (also known as local charge conservation)

$$\nabla \cdot \mathbf{J} + \partial \rho / \partial t = 0 \ , \tag{20}$$

one finds that the potentials must obey a restrictive condition. This condition yields what is known as the Lorentz gauge,

$$\partial_\mu A^\mu = \nabla \cdot \mathbf{A} + \epsilon_0 \mu_0 \frac{\partial V}{\partial t} = \nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial V}{\partial t} = 0. \tag{21}$$

Thus, gauge invariance appears to be broken by introduction of a photon mass. The only allowed residual gauge transformations entail solely functions obeying the wave equation

$$[\nabla^2 - (1/c^2)(\partial/\partial t)^2]\Lambda = 0. \tag{22}$$

Indeed, both for the photon (and the graviton) there was long a feeling that gauge invariance (and its gravitational analogue general coordinate invariance) provides a fundamental basis for assuming exactly zero mass.

To examine this issue more fully, we need to remind ourselves of how the form taken by gauge invariance in the context of quantum mechanics came to be. Weyl introduced the term "gauge invariance" in 1918–19 (42; 43), before the appearance of modern quantum mechanics. He wanted the gravitational metric and the electromagnetic field to transform as

$$g_{\mu\nu}(x) \ \to \ e^{2\alpha(x)} g_{\mu\nu}(x), \tag{23}$$

$$A_\mu(x) \ \to \ A_\mu(x) - e\partial_\mu \alpha(x). \tag{24}$$

(Here gauge is used in the sense of scale, because α is real.) This type of change is now known as a conformal or scale transformation.

In 1929 Weyl revised his approach for electromagnetism in quantum mechanics (44–46), setting the stage

---

[5] If the electric current is conserved, then the $J_\mu A^\mu$ coupling for a longitudinal photon, which can be written $A_\mu = \partial_\mu \Lambda$, becomes zero because $\partial_\mu J^\mu = 0$, where one has used integration by parts. However, if the current is not conserved, then this zero isn't so. Instead, because the D'Alembertian on A is J, one gets a radiated longitudinal A field going like J divided by D'Alembertian plus μ^2 in the Proca case. In the $\mu \to 0$ limit, where the photons travel at the speed of light, this becomes divergent: For the longitudinal part going as four-gradient of Λ this gives rise to a singularity as long as the divergence of J does not vanish. So the longitudinal A field goes like four-gradient of divergence of J divided by D'Alembertian, which on-shell clearly is divergent.

for all future discussions. He allowed the transformation to be complex.[6] He considered the Schrödinger equation for a particle with electric charge q

$$\left[\frac{i\hbar\partial}{\partial t}\right]\psi = \left[\frac{(-i\hbar\nabla - q\mathbf{A})^2}{2m} + qV\right]\psi \quad . \tag{25}$$

Under the simultaneous transformations (19) and

$$\psi \rightarrow \psi' = e^{iq(\Lambda/\hbar)}\psi \quad , \tag{26}$$

we see that the Schrödinger equation is unchanged. This is known as gauge invariance of the second kind.

Two decades after Proca introduced his mass mechanism enforcing the Lorentz gauge, Stueckelberg found what initially may have seemed merely a formal way of restoring gauge invariance (52). He introduced a new scalar field, Φ, with fixed magnitude and carrying electric charge, q, whose 'kinetic', gauge-invariant contribution to the Lagrangian density is

$$\mathcal{L}_{\mathrm{S}} = \frac{1}{2}\left[(-\partial_t\Phi + iqV\Phi/\hbar)^2 - (\nabla\Phi - iq\mathbf{A}\Phi/\hbar)^2\right] \quad . \tag{27}$$

Here we are dealing with a Klein-Gordon equation, rather than a Schrödinger equation. Otherwise this is simply an example of the new gauge invariance required in quantum mechanics, even though we may treat Φ as a classical field. At this point we may choose a gauge by assuming that the phase of Φ is zero everywhere or, indeed, has any constant value. In that gauge, it is easy to see that the extra term in the action becomes

$$\mathcal{L}_{\mathrm{S}} = -\frac{1}{2}\mu^2(\mathbf{A}^2 - V^2) \,, \tag{28}$$

$$\mu \equiv q\Phi/\sqrt{\epsilon_0}\hbar c = q\Phi\sqrt{\mu_0}/\hbar \,. \tag{29}$$

This is just the Proca photon mass term we have seen before, again with mass expressed in units of inverse (reduced) Compton wavelength. Now, however, the restriction to Lorentz gauge comes only because we made a specific choice (zero) for the phase variation of Φ. With no such specification, full gauge invariance is restored, even though the photon now has a non-zero mass.

Therefore we may replace the earlier guess, that gauge invariance implies zero photon mass, by a new, more precise assertion: The minimal dynamics obeying gauge invariance (the Maxwell action) implies zero photon mass. However, by adding more dynamics, for example, another field Φ interacting with the photon field, we may keep gauge invariance and accommodate non-zero mass at the same time.

If we think of variation in the phase as a (spacetime) position-dependent rotation, then it is immediately clear that the corresponding symmetry must be unbreakable as well as unobservable: Observable arbitrary position-dependent rotations would put arbitrarily great stresses on any system, and thus could not be symmetries.

For reasons like this, gauge invariance and general coordinate invariance have (sometimes) been called "fake" symmetries. This term should be treated with care, since it could be taken to imply that the whole concept is useless. However, as we have seen, this abstract and unobservable symmetry, infinitely flexible and therefore intrinsically unbreakable, provides a powerful organizing principle for dynamics. It has an especially simple and aesthetic starting point, the minimal theory, namely Maxwell theory, for electrodynamics (and of course general relativity for gravity).[7]

There remains one unsatisfactory point in the Stueckelberg formulation. Because the magnitude of Φ is fixed this theory is not renormalizable. Stated differently, Φ is not a fully dynamical field. A simple way to address that defect is presented in the next subsection.

E. Higgs mechanism, or hidden gauge invariance

Though Stueckelberg's construction removed the formal gauge-invariance argument for zero photon mass, there still was little motive for going beyond the minimal theory. The physical interest in doing so began with the work of Yang and Mills (53), who proposed the idea of a more elaborate gauge symmetry, where the rotations are in a three-dimensional space, rather than the single phase or rotation angle (corresponding to a two-dimensional space) found in electrodynamics. In later years their proposal was generalized by many authors, leading to the conclusion that gauge symmetries can apply for arbitrary compact transformation groups. The immediate question arising when one contemplates such additional gauge symmetries is, "Where are the corresponding massless photon-like particles?"

One answer to this question had its intellectual beginnings in the early 1960s with the work of a number of authors (54)-(60). Schwinger (54; 55), and more explicitly Anderson (56), followed by Englert and Brout (57) by Higgs (58–60) and by Guralnik, Hagen, and Kibble (61)

[6] It is to be noted that Klein (47), Fock (48), and London (49; 50), shortly after the appearance of Schrödinger's wave mechanics, each came up with the pre-requisite idea that the full electromagnetic interaction in the Schrödinger equation entails the form $-i\hbar\nabla - q\mathbf{A}$ found in Eq. (25) Further, (50) cites an early related paper of Schrödinger, written before the wave equation was discovered (51).

[7] In condensed-matter physics, the compatibility of non-zero photon mass with gauge invariance is well-known. The simplest example is a plasma, where plasma longitudinal and transverse sound waves combine to provide the three degrees of freedom one expects for a massive spin-1 particle. Of course the plasma fixes a local rest frame, so that Lorentz invariance is broken. An insulator has electromagnetic excitations of arbitrarily low energy, which might make them seem massless, yet the excitations travel at subluminal speed compared to light in vacuum.

found increasingly clear ways to describe gauge particles possessing mass while the underlying gauge invariance remains unbroken.[8]

The big change from Stueckelberg's idea, in what has become known as the Higgs mechanism, is to allow the magnitude as well as the phase orientation of the 'mass-generating' field to become dynamical. In its simplest form, this corresponds to adding a term to the Lagrangian density of Eq. (27), yielding

$$\mathcal{L}_S = \frac{1}{2}\left[(-\partial_t\Phi + iqV\Phi/\hbar)^2 - (\nabla\Phi - iq\mathbf{A}\Phi/\hbar)^2\right] - \frac{\lambda}{4}(\Phi^2 - v^2)^2 \ , \tag{30}$$

where v is called the "vev" or "vacuum expectation value" of the field Φ: $v \equiv \langle\Phi\rangle$.

This "Higgs mechanism" is a relativistically invariant analogue of the behavior of a superconductor, where a collective wave function of many charged particles leads to damping of electric and magnetic fields.[9] The simplest form of this mechanism introduces a charged scalar field which in the ground state of the system has nonzero magnitude everywhere. Varying the action with respect to the four-vector potential, $A^\mu = (V, \mathbf{A})$, yields exponential damping of a static, electromagnetic field in space and so, of course, a dispersion (small though it might be!) of wave or photon velocity with frequency. This corresponds to the introduction of a finite mass for the gauge boson (the photon). As in the superconductor case, a sufficiently strong electromagnetic field, or sufficiently high temperature, can force Φ to vanish in some region (which was not possible for the Stueckelberg field), in which case the photon may exhibit zero effective mass in that region.

Despite their appeal, these ideas lay dormant for nearly a decade, until 't Hooft's proof (63; 64) that such a theory fits into the pattern established by quantum electrodynamics, a renormalizable perturbative quantum field theory. This means that, for phenomena where the gauge coupling can be treated as small, there is a well-defined, systematic expansion in powers of the coupling (and depending only on a finite number of parameters) to deduce precise values for cross sections and other observable quantities,

The original work mentioned at the beginning of this subsection was all for an Abelian theory, i.e., massive electrodynamics. However, the applications in the early 1970s were for more complicated, non-Abelian theories. The Higgs mechanism became an integral part of the highly successful standard model unifying weak and electromagnetic interactions. It should be noted that a more complicated form of this mechanism, in which (as is true for superconductivity) there is no particle corresponding to quantum fluctuations of the Higgs field, remains a logical possibility. Even more important than a possible Higgs particle as a validation of this view of electroweak interactions is the theoretically predicted and experimentally confirmed existence of the massive gauge bosons $W^\pm$ and Z. Another non-Abelian theory, quantum chromodynamics, though with a different mechanism (color confinement) for avoiding free massless gauge bosons, has been similarly successful in describing the strong interactions.

Surprisingly, until recently the option of using the Higgs mechanism to parametrize possible deviations from Maxwell theory remained relatively unexplored. Indeed, we know of only two attempts to apply these ideas to the question of a possible photon mass.

1. Temperature effect

Primack and Sher (65) focused on an effect familiar in superconductivity, that above a critical temperature the condensate disappears. Thus, they considered the possibility that at very low temperatures there might be a Higgs mechanism that would generate a small photon mass, but at higher temperatures photons would be massless. Though they did not view this as especially likely, it still is worthwhile to examine the notion a bit more closely.

For the condensate value Φ to disappear in a large region of space, the energy density corresponding to a given temperature in that region, $\sim (kT)^4/(\hbar c)^3$, should exceed the vacuum energy density $\frac{\lambda}{4}v^4$ associated with vanishing Φ. This may happen either because of a very small value of λ or a small value of v, or of course a combination. Also, λ and v may be temperature-dependent, yielding a zero value for v at sufficiently high T. For the condensate to be restored in a volume characterized by length L, the temperature in that region must fall below a critical value. Further, the gradient energy $\sim v^2 L$ must be smaller than the vacuum energy $\sim \frac{\lambda}{4}v^4 L^3$; that is, $\lambda v^2 L^2 \geq 1$. Thus, if the coupling λ were too small the effect would not occur, even if the temperature in the region were low enough.

At the same time, to find a detectable photon mass effect inside that region, one must be sensitive to a term of order

$$(\mu L)^2 = \frac{(qvL)^2}{(\hbar c)^2\epsilon_0} = (qvL/\hbar)^2\mu_0. \tag{31}$$

This illuminates the difficulty of implementing this mechanism: If q were appreciable but the effective mass of the scalar field were small, then one would expect to observe production of light charged "Higgs" particles, which have not been seen. Thus q, the charge of the Higgs field, must

[8] Possibly the first discussion of the non-Abelian version of the Higgs mechanism was in a remarkable paper representing an independent discovery of the mechanism by Migdal and Polyakov (62).

[9] In a superconductor the macroscopic electron-pair condensate wave function produces an effect like that of a non-zero photon mass, again without breaking gauge invariance.

be quite small, while the Higgs mass must be small[10] and the charge q sufficiently large, so that the Primack-Sher temperature effect would be observable. This leaves at best a very small region in the three-dimensional parameter space (q, v, λ) for which the effect would be possible (67; 68).

2. Large-scale magnetic fields and the photon mass

Recently Adelberger, Dvali, and Gruzinov (ADG) (69) proposed using this type of mechanism to parametrize possible deviations from Maxwell theory on large length scales. Like Primack and Sher they used an Abelian Higgs field (not related to the standard-model Higgs). ADG's most striking point is that a phenomenon like that of Abrikosov vortices in a superconductor could allow a substantial mean magnetic field $\mathbf{B}$ over galactic or even larger regions.

The Higgs field would have null lines parallel to the direction of $\mathbf{B}$, while the phase of the field would circulate with period 2π about each line.

If one did not happen to be sitting near a vortex line, extremely precise local measurements of *electric* fields could indicate patterns associated with a tiny nonzero photon mass. Even so, the implication from Proca theory – that a nonzero average field over a region (with dimensions transverse to the field direction characterized by a very large length scale L) requires an upper bound on the photon mass – would no longer hold. Provided the photon magnetic Compton wavelength were large compared to the typical separation between vortex lines, $\mathbf{B}$ would be essentially constant.

The basis for this effect goes back to Stueckelberg's observation (52), discussed above. By gauge-transforming to a gauge in which the Higgs field has constant phase, one obtains a vector potential

$$\mathbf{A} = \mathbf{A}_{\text{Maxwell}} - \boldsymbol{\nabla}\Lambda \quad . \tag{32}$$

Then the energy density contribution

$$\mathcal{E}_{\text{photon-mass}} = |q^2\Phi|^2\mathbf{A}^2/2 \tag{33}$$

is suppressed compared to the Proca case because the phase vortices make the average $\langle\mathbf{A}\rangle$ vanish. In fact the average "photon-mass" energy density is reduced to

$$\langle\mathcal{E}\rangle = \mathcal{O}(\mu^2\ell^2)B^2/\mu_0 \tag{34}$$

instead of $\mathcal{O}(\mu^2L^2)B^2/\mu_0$, where ℓ gives the typical separation between vortex lines and L, as before, is the typical length transverse to $\mathbf{B}$ over which $\mathbf{B}$ is roughly uniform.

For fixed λ and v, and $B < \sqrt{\lambda\mu_0}v^2$, as μ decreases ($\lambda\!\!\!^-_C$ increases) the Higgs field becomes increasingly 'stiff', and this more complicated theory reduces to the second-stage or Proca form. The possibility of achieving such a limit demonstrates that there exists a mathematical transformation which formally restores gauge invariance to the Proca theory, just as had been observed by Stueckelberg (52). Thus, previous limits on the Proca mass remain valid provided $\lambda\!\!\!^-_C$ is not too small, so that there is a smooth continuity between stages 2 and 3.

A novelty of the more general (Higgs) form, in addition to possible measurements of apparent photon mass, is that in the regime of moderate or small $\lambda\!\!\!^-_C$ one also has possible observations of critical field or critical temperature effects associated with extinction of the mass. In particular, one may consider a regime where the typical field strength $\mathbf{B}$ is so great that $\langle\Phi\rangle$ is brought to zero. Now one has a situation quite similar to that discussed by Primack and Sher for temperature (65), where much of space shows no photon mass. Still, in a sufficiently large region of true vacuum, with sufficiently small $\mathbf{B}$ or T, one possibly could detect a nonzero and perhaps even quite substantial mass, for example by repeating the WFH experiment (77) there. (See below.)

3. Empirical and formal considerations on the Abelian Higgs mechanism

From the viewpoint of testing this Abelian Higgs concept, there is a major change from the fixed Proca mass. This time there are three parameters, (i) the optimum, energy-minimizing magnitude of the Higgs field vev, v, (ii) a coefficient of assumed quartic self-coupling of the Higgs field, λ, and (iii) a parameter, q, representing the charge of the Higgs field which determines its coupling to the electromagnetic field. This increases the challenge of determining the parameters, or even limits on them.

At the same time this gives more observational tools for constraining the parameters. For example, if the particles had low mass, then their charge would have to be very small, because otherwise they would be created copiously, and easily detected, in any high-energy process involving collisions of ordinary charged particles. Clearly there is incentive for followup work, beyond the discussion for the "zero-temperature" case presented by ADG, to map out regions in the three-dimensional parameter space still allowed by existing measurements. This also could determine what further measurements might best improve the constraints on the allowed parameter domains.

While the massive gauge bosons of electroweak interactions show that gauge-invariant mass of gauge particles is possible, there may still be constraints of principle. First, extensive studies of self-coupled scalar fields indicate that such a system would only make sense if the dimensionless coupling $\lambda\hbar c$ were of order unity or less. Secondly, the dimensionless gauge coupling in elec-

[10] As pointed out in a previous subsection, limits on the magnitude of possible small electric charges carried by very light particles have been discussed before (38). Also see (66).

troweak interactions is comparable for the electric and the weak sectors. This makes the domain of possibilities opened up by the discussions in the previous two parts seem questionable, because they inevitably would entail an exponentially smaller electric charge for the Abelian Higgs field than for any other particle.[11] Meanwhile, the limitation on λ excludes a strictly fixed photon mass (although for large enough v the Proca-Stueckelberg limit could be an excellent approximation).

Thus, modern quantum field theory gives some arguments to suggest that there may be no photon mass at all (even of the "gauge-invariant" type), reinforcing older considerations such as the geometrical significance of the Gauss law and the appeal of the minimal gauge coupling hypothesis seen in Maxwell theory. While the Gauss law relating charge to electric flux is broken explicitly for the deviation from Coulomb's law considered in stage one, it can be argued that it still holds formally for stage two and physically for stage three. This is because the vector potential in stage two and the electrically charged Higgs field in stage three can be taken to contribute to the electromagnetic charge and current densities.

Thus, if one looks at things in a certain way the symmetries and conservation laws apparently broken if a photon mass effect were observed could be said merely to be hidden. In any case, despite the lack of positive observations up to now, the issue of a nonzero mass of course remains open, because an exact zero can never be established by experiment.

The evolution we have described entails increasing numbers of parameters for an assumed deviation of classical electrodynamics from a strictly Maxwellian form. At the same time, there are more phenomena which can be examined to test for the deviation. Thus, the process of testing becomes more demanding, but the accuracy of limits in principle can be maintained or possibly improved.

F. Zero-mass limit and sterile longitudinal photons

There is a profound conceptual discontinuity associated with the zero-mass limit of massive electrodynamics. For any nonzero mass, there are three degrees of freedom, corresponding to the three possible orthogonal polarizations of a photon in its rest frame. Nevertheless, all observable phenomena of electrodynamics are continuous in the limit. Part of the reason is that the coupling amplitude for radiation of longitudinal photons is suppressed by a factor $\mathcal{O}(\mu^2/k^2)$ for photons of wavenumber k. Thus for any fixed k the coupling vanishes as $\mu \to 0$. In the limit then, longitudinal photons exist, but are completely invisible, or "sterile."

[11] That small charge has as a possible consequence that the time for the field to come into equilibrium with a nonzero value at very low temperature would be too long for practical observation.

Bass (70) made use of this idea to examine a possible explanation for cooling of the Earth's core, by emission of slightly coupled longitudinal photons. However, using Schrödinger's and Schrödinger and Bass's earlier estimates (31),(32) of a limit on photon mass from the properties of the Earth's magnetic field, he could rule out cooling by longitudinal-photon emission – the coupling is far too weak.

Even in static or low-frequency phenomena, the relative deviations of electromagnetic fields from their values for small μ are small, $\mathcal{O}(\mu^2 L^2)$, where L is a characteristic spatial dimension of the region under study (2). Although one is not looking at radiation here, the root cause for the suppression factor is the same. This can be understood by asking what would be the typical wavenumbers of virtual photons associated with such a configuration.

As mentioned in II.C, if electric charge were not locally conserved then longitudinal photons would be superstrongly coupled. Thus the continuity of the zero-mass limit depends on delicate cancellations that could easily be upset. Nevertheless, as long as local charge conservation holds, the continuity applies not only for observable electromagnetic fields in vacuum but also for fields in all kinds of material backgrounds.

Interesting examples of this statement include: 1) The continuity of the index of refraction and other electromagnetic quantities in μ implies that the recently discovered phenomena of "fast" and "slow" light (71) should not be affected by a small Proca mass. 2) The same applies even to explicitly quantum phenomena, such as the well-known Casimir effect of attraction between two uncharged conducting plates (72; 73).

III. SECURE AND SPECULATIVE PHOTON MASS LIMITS

Quoted photon mass limits have at times been overly optimistic in the strengths of their characterizations. This is perhaps due to the temptation to assert too strongly something one "knows" to be true, A look at the summary of the Particle Data Group (9) hints at this. In such a spirit, we here give our understanding of both secure and speculative mass limits.

The key to intuitively understanding the new physics is to solve the time-independent Proca equations (9)-(12). In particular the electric potential is not the Coulomb potential but a Yukawa potential. Putting Eq. (11) in more common form, it is

$$V(r) = -\frac{e}{r}\exp[-\mu r], \tag{35}$$

where again μ is the mass in the inverse (reduced) Compton wavelength. A similar Yukawa fall-off occurs for the magnetic vector potential and field. By taking the gradient of Eq. (35) one finds that the first non-Coulombic term is of order $(\mu r)^2$. This size turns out to be general, and can be given by a theorem (74).

Therefore, as we (74) and others (75; 76) have emphasized, to measure a small photon mass you need either a very precise experiment or a very large apparatus. That is, a precise experiment can measure the very small deviation from unity in a slowly falling exponential and a very large apparatus has the advantage of having a large exponential fall-off vs. unity. Since the publication of Ref. (2) there have been extensions of previously introduced approaches to do this, and also two new ideas.

A. Local laboratory experiments

1. Electric ("Cavendish") experiment

Laboratory tests of Coulomb's Law are the cleanest one can perform. This is not surprising, as the experiments are small and local. They can be repeated and the systematics can be characterized and reduced,obviously important here. Since the apparatus is "small" a precise experiment is necessary. It is both a tribute to their ingenuity and also a comment on how the size of an experiment limits a photon mass measurement, that the 35-year-old result of Williams, Faller, and Hill (77) remains a landmark test of Coulomb's Law. Their limit of

$$\lambda\!\!\!^{-}_C \gtrsim 2 \times 10^7\,\mathrm{m}, \quad \text{or} \qquad \mu \lesssim 10^{-14}\,\mathrm{eV} \equiv 2 \times 10^{-50}\,\mathrm{kg}\,, \tag{36}$$

is unsurpassed in the substantiated (laboratory) literature.[12]

2. Magnetic (Aharonov-Bohm) experiment

Boulare and Deser (80) observed that another null experiment can be done with a magnetic field confined by a superconductor. The flux inside the superconductor must be an integer number of flux quanta, but with nonzero photon mass there will be an antiparallel flux outside in the vicinity of the superconductor suppressed by a factor of $\mathcal{O}(\mu^2\ell^2)$, where ℓ is a characteristic dimension of the apparatus. They estimated that an experiment of this sort could produce a limit $\lambda\!\!\!^{-}_C \gtrsim 10^5$m. To the best of our knowledge no such dedicated experiment yet has been performed. We suspect that with the help of a SQUID detector their proposed sensitivity could be improved, but perhaps not to the level of the result in (77).

3. Temperature effect

The ideas of Primack and Sher (65) on a photon phase transition at low temperature, even if incomplete (67; 68), inspired a low-precision ($\lambda\!\!\!^{-}_C \gtrsim 300$ m) experiment.by Ryan, Acceta, and Austin (81), performed at 1.36 K.[13] As we have mentioned already, this negative result need not be meaningful, because (1) gradient energy of the Higgs field could prevent its acquiring a nonzero value in a small region maintained at low temperature, and (2) a very small electric charge of the field could keep it from coming into thermal equilibrium during a time practical for observation.

4. Dispersion, radio waves, and the Kroll effect

For decades de Broglie hoped to find a photon mass, at first by the dispersion of optical light from stars (25). He performed a calculation in 1945 that claimed a limit of $\mu \le 10^{-47}$ kg, although there was a numerical error (2), meaning the correct limit was

$$\lambda\!\!\!^{-}_C \gtrsim 0.5\,\mathrm{m}, \quad \text{or} \qquad \mu \lesssim 4 \times 10^{-7}\,\mathrm{eV} \equiv 0.8 \times 10^{-42}\,\mathrm{kg}\,, \tag{37}$$

In our article (2) we discussed at length the dispersion in pulsar waves, which is an easily measurable effect. However, this dispersion is commonly accepted as a measure of the density of interstellar plasma. Interpreted as a photon mass it would give a value far above that excluded even by laboratory experiments (83). Because pulsar signals have such a long flight path, we incorrectly assumed that no better result could be found from velocity dispersion.

At about the same time as our previous review appeared, Kroll (84) discovered a way to do something we had thought impossible – obtain a reasonably competitive limit on photon mass from wave velocity dispersion. Kroll showed a way around this. The Schumann resonances are very low-frequency standing electromagnetic waves in the atmosphere between the earth's surface and the ionosphere, two conductive layers.

There are two important considerations here. For a wave propagating between between and parallel to two plane conducting layers, perhaps surprisingly there is a special mode whose speed is c, *even* if there be a nonzero photon mass (2). However, two concentric spherical conducting layers are not really parallel to each other. Kroll found that now the mass contributes to velocity dispersion of the special mode, but with $\mu^2_{eff} = g\mu^2$, where the dilution factor g for the modes that would travel at speed c between parallel conductors is of order $(R_> - R_<)/(R_> + R_<)$, which in this case would be slightly less than 1%.

---

[12] later reanalysis proposed a smaller number (78). Around the same period, a small improvement was claimed in (79), but the result was never published to our knowledge.

[13] An experiment considered by Clark was never completed to our knowledge. See Ref. (82). Such discussions also stimulated the late Henry Hill, who expressed strong interest in performing a Coulomb's Law test at very low temperatures (mK range) to search for a phase transition. (See Ref. (5).)

This means that the limit obtained on the photon mass would be only an order of magnitude worse than naive expectations (i.e., expectations in ignorance of behavior of the special mode) might have suggested. The second point is that the atmosphere in between the two conducting layers has a conductivity far smaller than interstellar plasma, despite the much higher mass density of the atmosphere. Thus, by looking at really low frequencies (where the lowest is about 8 Hz), one may obtain an interesting limit even for waves whose travel distance is no more than the circumference of the earth. Kroll deduced a limit

$$\lambdabar_C \gtrsim 8\times 10^5\,\mathrm{m}, \quad \text{or}$$
$$\mu \lesssim 3\times 10^{-13}\,\mathrm{eV} \equiv 4\times 10^{-49}\,\mathrm{kg}\ , \tag{38}$$

i.e., $\lambdabar_C$ about a tenth the radius of the earth.

Recently Füllekrug (85) adapted Kroll's method to new and more refined data on the Schumann resonances and the height of the ionosphere. He claimed a result about three orders of magnitude better than Kroll's. Füllekrug made the assumption that the frequency shift due to photon mass μ is linear rather than quadratic in μ. His assumption is contrary to the theorem discussed in the preamble of this section (74–76), and therefore leads us to strong reservations about the details of his approach.

A possible explanation for his assumption is that according to his analysis a fractional shift in circular frequency ω *is* equal to the ratio $A = (\Delta h_2)/(2\sqrt{h_1 h_2})$, where h_2 is the ionosphere height (about 100 km), Δh_2 is its possible fluctuation, and h_1 is the height of that point in the atmosphere where the displacement current and the electric current are equal in magnitude (about 50 km). Simply from dimensional analysis, he likely is right that this effect on wave phase velocity is linear in the quoted ratio, but because the Maxwell equations involve μ^2 we do not see how there can be a linear dependence of phase velocity on a very small photon mass.

In our view the proper way to obtain an optimum limit on photon mass from these data would be to fit deviations in the lowest frequencies to the formula

$$\delta\omega_i = A\omega_i + B/\omega_i\ , \tag{39}$$

with $B = g\mu^2c^2/2$. Unfortunately the data presented in the paper are insufficient to carry out this fit. We think that although it is likely there would be to a significant improvement over Kroll's result, it would not be by three orders of magnitude.

B. Solar system tests

1. Magnetic fields

The idea of Schrödinger to test for a photon mass by measuring the Earth's magnetic field (31; 32) took advantage of the other side of the above paradigm, it used a large apparatus! Over the years a number of improvements were made to Schrödinger's method for the Earth. The best current result of this type came from using an even bigger apparatus, Jupiter. A limit of

$$\lambdabar_C \gtrsim 5\times 10^8\,\mathrm{m}, \quad \text{or}$$
$$\mu \lesssim 4\times 10^{-16}\,\mathrm{eV} \equiv 7\times 10^{-52}\,\mathrm{kg} \tag{40}$$

came from the Pioneer 10 flyby of Jupiter (86).

We emphasize that because this limit is due to data from the first flyby of Jupiter, it was calculated in an extremely conservative manner, at least by a factor of 2. Furthermore, with modern data a more precise number could be obtained. However, once again, because of the $(\mu r)^2$ effect, an order of magnitude improvement basically calls for an order of magnitude larger magnet, say the Sun. Ideas on how a solar probe mission could do this were given in Ref. (87).

Finally, there is the largest magnetic field in the solar system, that associated with the solar wind. In principle this could yield the best directly measured limit. Using the MHD equations for a finite Proca mass and a generous upper bound for the μ^2A^2 energy of the solar wind magnetic field, Ryutov (88) found some time ago that a limit at "a factor of a few better" than the Jupiter limit should follow.

Recently Ryutov has been able to use fuller data on the plasma and magnetic field, extending to the edge of the solar system, to make a bf dramatic further improvement (89):

$$\lambdabar_C \gtrsim 2\times !0^{11}\ \mathrm{m} \quad \text{or}$$
$$\mu \lesssim 10^{-18}\ \mathrm{eV}\ \equiv 2\times 10^{-54}\ \mathrm{kg}, \tag{41}$$

or a minimum reduced Compton wavelength about 1.3 AU.

The key point in Ryutov's analysis is that the solar wind field does not exhibit any perceptible exponential decrease with radial distance from the sun. To prevent such a decrease if there were a photon mass would require a calculable actual current to cancel the Proca current $-\mu^2\mathbf{A}/\mu_0$ implied by the Proca equations.

There are satellite measurements of magnetic fields out to the orbit of Pluto and beyond. Using these measurements, one may deduce the currents corresponding to any particular photon mass, and the associated Lorentz forces and vector-potential energy densities. Putting in generous assumptions for the possible values of these, he obtains $\lambdabar_C \geq 1$ AU as a clearly conservative limit. To the best of our knowledge, this is the strongest documented limit in the research literature supported by well controlled and understood data.

C. Cosmic tests

1. Fields on galactic scales

Given the fact that large-scale magnetic fields in vacuum would be direct evidence for a limit on their exponential decay with distance (and hence a limit on the photon mass), large-scale magnetic fields in the galaxy or even in extra-galactic space have long been of interest. Yamaguchi (90) wrote the pioneering comment, arguing that turbulent cells in the Crab nebula of size 0.1 ly = 10^{15} m implied a Compton wave length of at least this size:

$$\lambda\!\!\!^{-}_C \gtrsim 10^{15}\ \mathrm{m} \quad (0.1\ \mathrm{ly}), \tag{42}$$

This principle can be investigated by looking at magnetic fields through measurements of frequency-dependent rotation in the plane of polarization of electromagnetic waves (Faraday rotation). The polarization rotation is sensitive to the product of plasma density and magnetic field strength, and in many cases the observations are consistent with uniform plasma and field distributions.

However, these observations also would be consistent with a volume-averaged value for the product, even if each individual factor varied substantially. For example, the density ρ, which is non-negative-definite, must have a nonzero average, but $\mathbf{B}$ might have zero average, even with $\langle \rho \mathbf{B} \rangle$ nonzero. Thus, as a matter of logic, the nonzero average of $\langle \rho \mathbf{B} \rangle$ has no implications for the magnitude of $\mathbf{A}$.

Besides Faraday rotation, an even more conspicuous signal of interstellar magnetic fields is synchrotron radiation. Because this radiation would look exactly the same if the direction of a magnetic field $\mathbf{B}$ were reversed, data on this phenomenon cannot discriminate against frequent reversals of the field, and thus are consistent with the zero average field suggested in the above paragraph (91).[14]

The same kinds of question apply even more to limits based on galactic-sized fields (92–94), because observations on such scales are less precise: Chibisov (94) claimed a limit

$$\lambda\!\!\!^{-}_C \gtrsim 10^{20}\ \mathrm{m} \quad (10^4\ \mathrm{ly}), \tag{43}$$

by following Yamaguchi (90) in extending the Crab Nebula analysis to the entire galaxy.

If the region of uniform $\mathbf{B}$ extends over a galactic arm, and is aligned parallel to the axis of the arm, then $\mu A \sim \mu R B$ (where R is the radius of the arm) arguably should be no bigger than B: This would follow from the virial assumption that plasma kinetic energy, ordinary magnetic field energy, and photon-mass-induced vector potential energy all should be in electromechanical equilibrium. Thus, the energy density associated with the magnetic vector potential should not vastly exceed the energy associated with the magnetic field.

The virial assumption recently has been asserted forcefully by ADG (69), who state that in the Proca case the Yamaguchi-Chibisov limit is valid. If one could confirm sufficiently detailed information about the plasma and the magnetic field, such a result might become well established. At present, though, there are at least two obstacles, besides those mentioned already. First, there could be significant time dependence of the fields on a scale as small as 1000 years. Secondly, there is good reason to believe that there are substantial inhomogeneities in the field and plasma, which could be reservoirs of much greater total energy than the average magnetic field energy.

Still, the Proca energy emphasized by ADG is so large that it would be tempting to dismiss all the above caveats, and at most use them to weaken somewhat the Yamaguchi-Chibisov limit associated with phenomena on a given scale. However, there is another issue already hinted at above which can change the calculus completely. If the photon mass were zero, then data consistent with uniform magnetic fields over large regions naturally should be interpreted as indicating that uniformity really is present. After all, there is no obvious mechanism for reversals, and no natural length scale for the reversals. The same kind of energy consideration championed by ADG changes this if the mass is nonzero.

With a given photon Compton wavelength $\lambda\!\!\!^{-}_C$, balance of energy among plasma, magnetic field, and photon mass contributions could occur if there were "pencils" or filaments of plasma with an average $\mathbf{B}$ aligned in one direction parallel to the filament axis, and outside each filament an exponentially decaying vector potential producing an equal and opposite flux to that contained in the filament. As explained above, such a configuration would be consistent with all observations to date relating to $\mathbf{B}$.

There is another relevant set of observations within our galaxy, the velocity dispersion of pulsar radio signals mentioned in Sec. III.A.4. It is proportional to the integrated plasma density along the path between each pulsar and the observation point (83). This clearly gives a constraint on the average plasma density, but given the relative paucity of pulsars may not provide enough information to determine whether there is or is not a filamentary structure on a particular scale.

We believe that something like the Yamaguchi-Chibisov limit might be verified in the not-too-distant future by additional observations (thanks to extraordinarily rapid progress in gathering astrophysical data beginning in the last decade or so). However, it is not established by present knowledge. There are several issues, including the poorly known magnetic fluctuations

[14] A similar comment about insensitivity to field reversals applies to signals from Zeeman splitting of OH and other molecules, as well as linear polarization of interstellar dust grains.

at short distance scales (tens of pc), [15] the role in the virial theorem of gravitational energies, and short-time phenomena that 'dump' energy into the medium, especially supernova explosions.[16]

When we come to galactic-cluster-sized magnetic fields, the same problems are even more challenging, because the detail available at greater distances of course is reduced.

2. The Lakes method

With perhaps the most creative observational method put forth in half a century for detecting photon mass, Lakes (95) proposed to measure the torque on a magnetic flux loop as it rotates with the Earth' surface. Lakes noted that if a magnetic field $\mathbf{B}$ is nearly uniform over a region of dimension L, then at a typical random point the vector potential is of order LB in magnitude. The $-\mu^2\mathbf{A}^2/2$ term in the Lagrangian then leads to a toroidal moment interaction between a toroidal solenoid of moment $\mathbf{a}$ and the ambient "vector potential field" $\mu^2\mathbf{A}_{\mathrm{amb}}$, analogous to the torque on a loop of electric current from an ambient magnetic field. In other words, nonzero photon mass makes the vector potential observable, and this technique allows its direct observation.

The torque

$$\boldsymbol{\tau} = \boldsymbol{\nu}\times\mu^2\mathbf{A}_{\mathrm{amb}} \tag{44}$$

acts on $\boldsymbol{\nu}$, the 'vector-potential dipole moment' of the flux loop. As one knows $\boldsymbol{\nu}$, measuring or limiting the value of the torque on the solenoid, $\boldsymbol{\tau}$, yields $(\mu^2\mathbf{A}_{\mathrm{amb}})$. Determining a lower bound on $\mathbf{A}_{\mathrm{amb}}$ then places a value on μ. A typical value of $\mathbf{A}_{\mathrm{amb}}$ might be very large in galactic and intergalactic space, when $|\mathbf{A}| \approx |\mathbf{B}|L$ with L the radius of a cross section transverse to a cylinder aligned parallel to a field $\mathbf{B} \approx$ constant.

In his original experiment, Lakes (95) studied a torque on the solenoid about one particular axis (the rotation axis of the earth), and hence had to assume that this axis was not parallel to $\mathbf{A}$. He also assumed, based on inferred values for galactic and intergalactic fields and the associated scales L, a magnitude for A, and thus obtained a limit.

A later experiment by Luo et al. (96), was both more precise and also allowed the axis about which the torque was measured to vary in time. This eliminated Lakes' angle problem, but still left the assumption that the magnitude of A is $\langle B\rangle L$. These experiments (95; 96) suggested that a lower limit on $\lambda\!\!\!^{-}_C$ as high as 3×10^{11} m could be obtained from fields in the Coma cluster. This would be even stronger than the solar wind limit.

Unfortunately, at the present the assumption $|\mathbf{A}| \approx |\mathbf{B}|L$ is not guaranteed for measurements on earth (97; 98). This is true not only because (as Lakes pointed out (95)) one may in principle be near a zero of A, but also because the evidence for uniformity of $\mathbf{B}$ is fragmentary. If there are holes in the distribution of the plasma supporting B, and if we are in such a hole, then, with a substantial μ, the linearly growing A envisioned by the experiments (95; 96) would be damped exponentially. Thus, A could indeed be small, making even a large μ^2 invisible in these experiments.

ADG (69) observed that in their vortex scenario the effective A would also be much smaller than LB, and again only a much weaker limit would hold. Thus, both in the Proca case and the vortex case it is not possible at this point to obtain a secure quantitative limit using the Lakes method.

There is another possible approach to seeking a value for $\mu^2\mathbf{A}$ (as mentioned already in the discussion of Ryutov's solar wind limit): In the presence of plasma, a static magnetic field may take exactly the form it would have in $\mu = 0$ magnetohydrodynamics, *provided* (2; 92; 97; 98) the plasma supports a current $\mathbf{J}$ that exactly cancels the 'pseudo-current' $-\mu^2\mathbf{A}/\mu_0$ induced by the photon mass. Thus, a uniform average $\mathbf{B}$ over a region large compared to $\lambda\!\!\!^{-}_C$ would require such a plasma current. This holds even if there are large fluctuating fields in addition to the average field.

By putting an upper limit on the true current one would put an upper limit on $\mu^2\mathbf{A}$. This limit would not be subject to the caveat that $\mu^2\mathbf{A}$ may be small at some particular point, because the plasma covers the same volume as the apparent volume over which $\mathbf{B}$ is spread. Following Lakes and Luo et al. in considering the Coma cluster, one may obtain a more conservative upper estimate of a possible plasma current, as follows.

From (99)-(101) there are estimates $L \lesssim 1.5\times10^{22}$ m, $B \gtrsim 10^{-10}$ T, the plasma free electron density in interstellar space satisfies $\rho \lesssim 10^4/\mathrm{m}^3$ and the plasma temperature $T \lesssim 10$ keV. Taking the generous view that the electron velocity in a coherent current could be as big as the r.m.s. thermal velocity yields a limit $\mu^2\langle A\rangle \lesssim 10^{-13}$ T/m, two orders of magnitude smaller than the laboratory experimental result. Furthermore it is unaffected by uncertainties about zeros in A at any particular location (such as the earth). Thus observations of volume-averaged properties of the cluster *could* yield a more conservative upper bound

$$\begin{aligned} \lambda\!\!\!^{-}_C &\gtrsim 3\times10^{12}\ \mathrm{m} \quad \textit{or} \\ \mu &\lesssim 7\times10^{-20}\ \mathrm{eV} \equiv 10^{-55}\ \ \mathrm{kg}\,. \end{aligned} \tag{45}$$

This ($\lambda\!\!\!^{-}_C \gtrsim 20$ AU) *would* be substantially better than the Lakes-method limits and the solar wind limit. We use the conditional forms *could* and *would* because, as discussed in III.C.1, there still are issues associated with

[15] These fluctuations are, however, certainly substantial compared to the uniform or slowly-varying field.

[16] From a logical viewpoint, an important issue may be the possibility of a filamentary structure as just described. At least some of the relevant factors are discussed by Beck (91).

the inference of large-scale uniform magnetic fields from observations.

One factor that could lead to an even more conservative form in this limit is, as pointed out by ADG, that the circulating current in the presence of **B** leads to a large Lorentz force, tending to 'explode' the plasma. A careful calculation of the rate of expansion might well force a reduced estimate of the current, and hence yield a reduced limit on $\mu^2\mathbf{A}$. This could imply

$$\begin{aligned} \lambda\!\!\!^{-}_C &\gtrsim 3\times 10^{9}\ \mathrm{m} \quad \mathit{or} \\ \mu &\lesssim 7\times 10^{-17}\ \mathrm{e}V \equiv 10^{-52}\ \mathrm{kg}\ . \end{aligned} \tag{46}$$

Albeit with all the same reservations mentioned before, this is a reduced Compton wave length of about 4 times the radius of the Sun, $R_\odot$ (97).

D. Transition from experiment to observation

With the new solar wind results (89) we may have arrived at the end of the era in which direct "local" laboratory experiments could contribute to limits on μ_γ. These results are based on experiments using apparatus in satellites to measure magnetic fields and plasma currents in large parts of the solar system. It is hard to imagine how such direct observations could be carried out to much larger distances. Thus, further research must rely on observations of radiation from more remote regions, as well, possibly, as observation of $\mu^2\mathbf{A}$.

1. Pursuing the Higgs effect

Among other points they discussed, ADG (69) noted that if μ arises from an Abelian Higgs mechanism in the regime analogous to that of a Type II superconductor, the existence of a non-zero photon mass implies generation of a primordial magnetic field in the early universe. This is quite interesting, because in the absence of any photon-mass considerations there has been substantial debate in the astrophysical community about whether the galactic field had a primordial 'seed' or is solely a consequence of a currently existing 'galactic dynamo'.

If, on the other hand, the galactic magnetic field is in the symmetry-breaking Proca regime, then according to ADG the very existence of a large-scale field would give

$$\begin{aligned} \lambda\!\!\!^{-}_C &\gtrsim 3\times 10^{19}\ \mathrm{m}\ (1\ \mathrm{kpc}) \quad \mathrm{or} \\ \mu &\lesssim 6\times 10^{-26}\ \mathrm{e}V \ \equiv 10^{-62}\ \ \mathrm{kg}\ . \end{aligned} \tag{47}$$

We already have seen that as of now the observations do not assure that there is a uniform **B**, but only that there is a uniform axis for a **B** that may, from place to place, reverse sign along that axis.

Thus, the ADG paper, with its insightful analysis of possibilities for the Abelian Higgs formulation of photon mass, also contains interesting assertions about the Proca limit. These assertions are not provable today, although they might be established in the future.

There remain significant issues for the Higgs scenario. A theoretical basis for the physical parameters $(q.\lambda, v)$ needed to make the vortex idea workable is lacking. Clearly the parameters would be enormously smaller than for the still unverified electroweak Higgs. In view of the many very small ratios of parameters found in particle physics already, this is not absurd, but it also is not compelling.

As discussed in Sec. III.C.1, given the complexities in the real astrophysical world, it may not be easy to distinguish effects of those complexities from effects of Higgs vortices. The flood of new data which we can confidently expect in the relatively near future may well shed light on these issues by clarifying the properties of astrophysical structures.

2. Photon dispersion as a lead into gravity

We explained earlier that limits on photon mass from static fields already are so stringent that any consequent observable dispersion in photon velocity likely is ruled out.[17] Nevertheless, there *is* an observed dispersion with frequency in arrival times of electromagnetic waves from a pulsar to a detector in our vicinity. If this is not due to a photon mass, one has to determine another cause, and the obvious one is interaction with the interstellar plasma (83). In fact, this "whistler effect" gives a way of detecting the mean plasma density along the path of the pulsar signal.

The phenomenon introduces a notion that will become even more important in the discussion of gravity to follow: When deviations are found from the implications of theory with known sources taken into account, one must look for modifications in the theory, or additional sources (or, of course, both). In the radio dispersion case the plasma explanation fits so many facts so well that there is no controversy about it, no suggestion that there is something missing in Maxwell theory.

This "non-mass" source of photon velocity dispersion has special interest for us, It was the effect (83) that first enticed us to study the photon-mass issue (5), at the time of the early pulsar discoveries. In the discussion of gravity to follow, the question of whether to ascribe anomalies to modification of gravity or to the addition of sources will become more interesting.

---

[17] Although as we have seen in III.A.4 not by as enormous a factor as holds for dispersion of pulsar wave velocities which we discuss now.

E. The primacy of length over all other measures

After people considered the old scale-free, but otherwise unmotivated, notion of a power deviating from that of the inverse square law, they came to a somewhat motivated idea, giving a nonzero mass to the photon. Already very early it became clear that the only likely observable effect along these lines would be a departure from Maxwellian structure for the very long-range behavior of static fields. By now the length scale in question is related to solar system dimensions, and there is every reason to expect that it will be extended much further still. In particular, there is no hope left of detecting directly a finite rest mass of the photon.

In principle, the Abelian Higgs formulation might accommodate (for true vacuum at zero temperature and zero ambient magnetic field) an actual, observable photon mass giving measurable dispersion of photon velocity with energy, but that is (literally) quite remote from anything we might hope to detect. ADG suggested that perhaps beyond galactic scales, where magnetic fields are somewhat weaker than in our galaxy, a finite, even directly observable photon mass might emerge. It could be interesting, and certainly would be challenging, to find types of observation that could be sensitive to such an effect.

IV. GRAVITATIONAL THEORIES

There are interesting parallels as well as divergences between developments in electromagnetic and in gravitational radiation theory. The most generic statement is that the latter has evolved more slowly. It began earlier, but even today it is less developed and also less well-tested. The realization that there are wave solutions of the equations of electromagnetism arose in the mid-nineteenth century, but the analogous realization for gravity came only with the advent of general relativity. Even after that there was wavering for at least half a century about the existence of these waves. By that time, quanta of electromagnetism, photons, were long established, so that wave and particle properties of light were on an equal footing.

Also, perturbative quantum electrodynamics [QED] had become a science still being refined today. The quanta or particles corresponding to gravitational waves are unlikely to be observed in the foreseeable future, simply because of the extraordinary weakness of gravity at scales accessible to humans. Indeed, even the existence of classical gravitational waves has been established only in the same sense as for neutrinos in the first half of the 20th century: Their radiation accounts quantitatively for energy loss observed in binary pulsar systems. Absorption of energy from gravitational waves, yet to be confirmed, is a target of current and planned large-scale gravitational-wave detectors.

Meanwhile, a quantum theory of gravity analogous to QED does not exist, in part because the most straightforward formulation is not renormalizable. String theory offers promise of providing a consistent quantum formulation including gravity, but still is far from complete.

Thus, it should not be surprising that, even more than in the case of electrodynamics, long-distance, low-frequency deviations from the preferred theory are more likely to be detected in the study of quasi-static phenomena than in an effect like dispersion of wave or graviton-particle velocity with frequency. Once again, let us review the stages in evolution of the subject.

A. Newton's law of gravity

According to Newton, the force between two masses acts along the line between them and takes the form

$$F = -\frac{Gm_1m_2}{r^2} \quad . \tag{48}$$

The success of this form was the basis for the later introduction of Coulomb's law. Here the negative sign indicates that the force between two masses is attractive, unlike the repulsive force between like-sign electric charges.

Of course, even at an early stage celestial mechanics gave a much higher precision in verifying the inverse-square law for gravity than the corresponding law for electricity.[18] Newton himself considered GM, which is much easier to measure than G, what we now call Newton's constant.[19]

Newton never reported an attempt to determine G, even though he had built pendulums of size 11 feet and had correctly calculated the average density of the Earth to be about 5-6 times the density of water (104; 105). The reason for his omission may be a surprising error that appeared in the Principia, stating that two spheres of Earth density and of size one foot placed 1/4 inch apart would take of order a month to come together, indicating that terrestrial experiments would be useless.

As discussed by Poynting (107), Newton's error produced an inhibition against performing terrestrial experiments until the work of Cavendish (106). Cavendish's purpose, the same as Newton's stated goal, was to determine the average density of the Earth. For this he needed

[18] The first quantitative test for the inverse-square law of electric force was done by John Robison in 1769, predating Coulomb (16)! It yielded an accuracy of $F \propto r^{-(2+q)}$, where q was found to be ~ 0.6 on a scale of a few inches. Contrariwise, at the end of the 1500's Tycho Brahe's naked eye observations were already good to 1 arc-sec or better (102), about the naked eye diffraction limit of $\sim \lambda_\nu/D \sim 10^{-4}$. Kepler used these observations to establish his laws of planetary motion, specifically the ellipse for the Mars orbit. Half a century later Newton quantified this in the inverse-square law, with the advent of telescopes bringing increasing accuracy (102).

[19] Even today, $GM_\odot$ for the Sun is known to a part in 10^{10} whereas G is only known to about a part in 10^4 (103).

only the ratio of the gravitational force between two test bodies of known mass to the gravitational force exerted on a test body by the Earth. He did not explicitly compute or even define G, which was introduced only much later.[20]

Over the following century, advances in mathematics allowed ever more precise calculations, and Newtonian theory always triumphed. Then, in 1781 Herschel discovered what he thought was a new parabolic-orbit comet, but which quickly turned out to be a new elliptical-orbit planet, Uranus. (The entire story is described in (109).) In 1784 Fixlmillner combined two years of then modern observations with two old sightings that had been mistaken for stars and calculated an orbit. By 1788 this elliptical orbit already did not work.

By 1820 there were 39 years of recent observations combined with 17 ancient observations (going back to 1690). Bouvard used these data to calculate a precise orbit but could not reconcile the entire data set. To resolve the dilemma he specifically attributed gross error to the ancient observations of eminent astronomers rather than allow for some unexplained cause of the irregularities. This all led to much disagreement, and over the succeeding decades the observed orbital deviations from calculated orbits got worse.

Into this situation came John Couch Adams and Urbain Jean Joseph le Verrier. In the time frame of 1843-1846 they independently used Newton's Law to predict the location of a new planet, Neptune, discovered in 1846 by Galle, on the first day he looked (109). They solved what we would call an inverse problem: What object was causing the not-understood perturbations of the planet Uranus?[21]

Clearly the Neptune solution (what in today's parlance would be called 'dark matter')[22] also had an alternative explanation, a modification of gravity. This same type of problem arose again soon after when le Verrier started a complete study of all the planets. When he returned to Mercury in 1859, he again found an earlier troubling problem (113), the precession of Mercury's perihelion was too large, by 33-38 arc-seconds/century (113–115). Later Simon Newcomb did a more precise calculation and found the "modern" value of 43 arc-seconds/century.[23]

The "obvious" most likely resolution was that there had to be a new planet, Vulcan, in the interior of the solar system. However, this time the answer was *not* missing dark matter. A hint in that direction was given by Asaph Hall. He followed on Newcomb's observation and Bertrand's work (which led to Bertrand's Theorem). Bertrand had shown that for small eccentricity the angle between successive radii vectors to the closest and furthest points in a bound orbit is (116)

$$\theta = \frac{\pi}{\sqrt{n+3}}, \tag{49}$$

where n is the power law of the force ($n = -2$ for Newton's law).

Using this, Hall (117) calculated that a force law with $2 \rightarrow 2.000\,000\,16$ would account for Mercury's precession.[24] Of course the accepted resolution today is the replacement of Newtonian gravity by general relativity, effectively leading to an added (small) r^{-3} term in the force law.

It is worth dwelling on this a bit. Hall's parametrization was a purely phenomenological one. It is hard to imagine that a phenomenological approach could ever have come close to evoking the complete general relativity. However, a discrepancy like that of the Mercury orbit was an alert for a possible need to modify the theory, and did give guidance for a possible (though in the end incorrect) form of the required modification.

B. Einstein's general theory of relativity and beyond?

While the evolution of electrodynamics entailed a harmonious progression fed both by experiment and by theory, the next stage in gravity was a theoretical accomplishment. General relativity [GR] immediately provided an accurate solution to the Mercury precession problem. Soon GR was vindicated by observations of the solar deflection of light, and more recently has been vetted by many other tests. Einstein's eight-year intellectual struggle, assisted by many colleagues, produced general relativity as a new version of Newton's gravity, now consistent with the principle of relativity.

Given the assumption that gravity is a metric theory, a systematic parametrization of such theories for phenomena depending on gravitational sources with velocities low compared to the velocity of light yields the PPN or parametrized post-Newtonian expansion for corrections to Newtonian gravity. Einstein's minimal theory, with no added gravitational fields besides the metric itself, gives definite values for these parameters, and many observations have provided increasingly stringent limits on deviations from the Einstein values. Because this subject has been reviewed extensively in the literature (118; 119), we

[20] An early reference to measuring "G"" was given by Cornu and Baille (108) (who called it "f"). In some folklore Cornu is given credit for popularizing the use of "G".

[21] An input into the solution (109) was what amounted to the Titius-Bode Law of Planetary Distances (110; 111).

[22] Another such problem was announced by Bessel in 1844 when he concluded that the observed wobble in Sirius' location must be due to a companion (112). (Hew made a similar observation for Procyon.) In 1862 Alvan Clark observed the very faint companion. We now know that it could cause the wobble because it is a high-density white dwarf.

[23] See. p. 136 of (115).

[24] Amusingly, this is precisely an example of the original way of parametrizing departures from the Coulomb law, thus very much in the spirit of Hall's time.

refer the reader there rather than sparingly touch on the same material here.

Although the PPN program began as a search for a certain class of deviations from Einstein gravity, as with scalar-tensor theories, it really has become more an increasingly extensive set of verifications for GR. As such, PPN so far has followed a similar trajectory to the search for photon mass described earlier in this paper – much interesting and creative theoretical work, many beautiful and ingenious experiments, but no evidence of any deviation from the simple starting point.

Finally, there is the school of quantum gravity,[25] whose most intensely studied formulation in recent decades is string theory (120), with its predicted extra dimensions (122). Besides the long-distance deviations on which we focus here this can also produce deviations at short distance scales and in strong gravitational fields. These both are not easy to detect.

C. Dark matter versus modified gravity

For more than eight decades it has been argued (123–126) that stars and globular clusters in galaxies, and galaxies themselves, move as if they are being deflected by bigger forces than the above assumptions would imply. Two possible explanations arise. 1) There is more (and different) matter present as a source for gravity than what we infer from both the visible radiation and also our knowledge of ordinary matter behavior. 2) The laws governing gravity are different from what Newton and Einstein would tell us. (Of course, a combination of both explanations might be needed.)

1. Dark matter

Today there are two widely-discussed proposals of phenomena which may be labeled new sources of gravity: "dark matter," whose implied effects are very well established (127), and "dark energy," which in a very short time has become strongly indicated by a variety of different classes of observation (128), (129), (130). Of course, either or both sets of phenomena in principle could result from modification of GR rather than from new sources.

Note that proposals of such modifications are directly motivated by observation (hence phenomenological in nature) and so far have not yielded a well-agreed upon conceptual basis in "new" theory. In this context, for the second phenomenon, "dark energy," it may be almost a matter of definition whether this is a new kind of matter or a modification of Einstein gravity. In particular, a constant cosmological term is consistent with all current data, and such a term put on the gravity side of the Einstein equations represents modified gravity, while put on the matter side it represents a new form of matter.[26]

"Dark matter," on the other hand has been and remains the most combative question. A whole array of different observations over a long period of time has established the following: If one assumes that we know the nature of matter in the universe, i.e., the standard-model particles – nucleons and nuclei, electrons, photons, and neutrinos, and one also assumes that we know how these elements combine to determine the structure of stars and the nature of interstellar gas, plasma, and dust, then Newton-Einstein gravity does not account correctly for the way that various objects are seen to move.

As the name implies, dark matter does not radiate anything we can see, either directly or through its interactions with other matter. Thus, this matter must be very weakly interacting, so that its (astronomically) observable effects come largely or entirely from its gravitational influence on ordinary matter. Because its interactions are presumed to be so weak, such matter could be very difficult to detect in the laboratory, and indeed there are no well-confirmed reports of such detection to date. We have no direct evidence for the existence of dark matter. Bertone, Hooper, and Silk (133) have given a recent, thorough review of the evidence for dark matter and hypotheses about its form, including possible discrepancies with observation.

A significant problem for the dark-matter hypothesis is that even today simulations of the evolution of ordinary galaxies do not account well for the precise patterns seen in the motions of the visible stars and globular clusters in such galaxies. This phenomenon, which was an original motive for introducing dark matter, thus remains a roadblock to easy acceptance of the idea. There is no intrinsic contradiction, but on the other hand no simple picture of how galaxies got to be the way they are through gravitational evolution of small initial density fluctuations, including the dark matter.

2. Proposed modifications of GR

Into this dilemma came Milgrom's idea of MOND (Modified Newtonian Dynamics) (134)-(137). Milgrom discovered that the rotational velocity vs. distance curves (velocity of stars in orbit at a given distance from the center of a spiral galaxy) could be generally explained

[25] A question that also comes from the school of quantum gravity is whether there are measurable vector and scalar partners of the graviton that have mass; so-called "fifth forces." These would die out after a finite distance, leaving only the effects of the zero-mass graviton behind. Repeated experiments, on scales from the laboratory to astronomical, have thus far found no evidence for such forces (121).

[26] Recent discussions of the difficulties that can arise in distinguishing by observation between dark energy and modified gravity can be found in (131) and (132).

by presuming that the Newtonian acceleration of a point mass M is changed to

$$a_N = -\frac{GM}{r^2} \rightarrow -\frac{\sqrt{GMa_0}}{r}, \tag{50}$$

where M is only composed of the visible mass in the galaxy and a_0 is a constant of size $\sim 10^{-8}$ cm/s^2. The transition occurs when the acceleration falls to a_0.

This simple phenomenology has had amazing success in describing a large class of these galactic-rotation curves, gives the Tully-Fisher relation between galaxy rotation rate and intrinsic luminosity (138) automatically, and avoids the need to calculate the amount of "dark matter" on a case by case basis. It was originally derived as a phenomenological fit, but it has been used to test many further galaxies.[27] Therefore, any successful dark matter solution has to explain the success of this phenomenology. The success is too great to be an accident.

MOND advocates have found difficulty in describing mass distributions on the scale of galactic clusters. Their preferred solution is that dark matter indeed is present in significant amounts, except they argue that this dark matter is ordinary baryonic matter, such as brown dwarfs, or at least standard-model matter, such as neutrinos with the maximum mass consistent with constraints from experiment (139).

Bekenstein (140) and others (140)-(143).have found a way to embed the MOND scheme in a fully relativistic version. However, this involves more phenomenological quantities, which in the language of the Higgs mechanism entail not only a scalar field but also a vector and a tensor field with nontrivial vacuum expectation values, all coupled directly to gravity.

While these are not traditional particle forms of matter, they can reasonably be called a new form of matter, so that logically this version of modified gravity is itself a (different) dark-matter theory. Thus we see classical discrete-particle matter giving a good account of extra contributions to gravity from the largest scale of the visible universe all the way down to clusters of galaxies. On the other hand, classical continuous fields acting as sources of gravity neatly describe extra contributions to gravity at galactic scales and below. In terms of their established domains of applicability, there is a conspicuous duality or complementarity between the two approaches.

If there is one thing on which advocates of (particle) dark matter and advocates of (field-induced) modified gravity seem to agree, it is that if one of these viewpoints is right the other must be wrong. Caution may be in order about this assertion. There is a striking precedent: From the 17th century on, Newton's prestige made the particle hypothesis for light dominant, but early in the 19th century examples of diffraction phenomena overthrew this picture, replacing it with the wave hypothesis. A hundred years later quantum mechanics showed that both descriptions are needed, each valid in answering appropriate questions.

We observe that there are other proposed modifications of gravity besides MOND. Mannheim (144) has written an accessible survey of the subject, beginning with reports in the 1930s of anomalies that could be taken as evidence for dark matter, and including a number of later observational and theoretical works. We have concentrated on MOND because it is the most discussed alternative.

Especially noteworthy is the work by Mannheim and colleagues (144) on "Weyl" or "conformal" gravity, which uses the symmetry of Weyl's original gauge (i.e., length rather than phase) invariance, alluded to around Eq. (23) (145). This is a genuine alternative gravity theory and makes interesting predictions on both galactic and longer scales.

In particular, Mannheim has made a prediction based on conformal gravity that when we learn about the expansion of the universe at still earlier epochs than have been explored up to now we shall find that the expansion already was accelerating. That statement appears to distinguish this approach from others being considered today, including the most popular "ΛCDM" model, with Einstein gravity exact and dark energy and dark matter sources present.

Because conformal gravity involves higher time derivatives, extra boundary conditions are required to make the theory well-defined. This has led to some debate. In particular, Flanagan (146) has argued that conformal gravity contradicts the original successful predictions of Einstein gravity for the effects of the sun. Mannheim (147) has presented a counter-argument. We look forward to an eventual consensus on the status of conformal gravity.[28]

3. Cluster collisions

From the viewpoint of testing theories, an unusual object called the "Bullet Cluster" added important input to the debate. The Bullet Cluster contains two subclusters which appear to have collided some time ago. Initially, each subcluster should have contained visible matter in the form of stars, and an order of magnitude more in the form of gas and plasma. During the collision, the stars should have gone through each other, but the gas clouds from the two subclusters should have experienced a great deal of friction, tending to coalesce and be left be-

[27] Observe that verification of this idea could require a new formulation for the theory of gravity. The status of such investigations, and their analogues for dark energy, may be the most interesting current aspect of what we started to describe with the label "graviton mass."

[28] A different approach mentioned briefly by Mannheim (144) has been put forward by Moffat and collaborators, e.g., in (148).

hind in the middle when the stars passed by each other. Dark matter, being weakly interacting, should have gone straight ahead with little friction or coalescence. Thus the dark-matter hypothesis leads to an unambiguous prediction about the distribution of matter in the cluster, with the bulk of the matter located near the two star subclusters.

The tool used to analyze this collision was gravitational lensing, that is, looking separately at strong and at week gravitational focusing of light from more distant sources passing by the cluster, and matching this with light images of the system itself in various frequency ranges. A painstaking analysis showed convincingly that the centers of lensing are located quite near to where the subclusters of galaxies are seen, rather than where the far more massive gas clouds were left behind (149). This success of a dark-matter prediction is a strong argument for dark matter on larger distance scales.

Already, certain advocates of modified gravity have acknowledged that they may need some dark matter (possibly in the forms of brown dwarfs and neutrinos of the maximum mass allowed by current limits) to account for the behavior of clusters of galaxies (139), Thus, the sharp line between the MOND phenomenology on the one hand, and the hypothesis that Einstein gravity is unmodified on the other, became blurred. This concession by MOND advocates had the consequence that their theory also could predict the kind of mass distribution observed in the Bullet cluster, so that the result need not distinguish between the two competing approaches.

A recent development holds the possibility of complicating the conclusions from the Bullet Cluster. In another such system, Abell 520, a weak lensing analysis indicated that the bulk of the dark matter is located in the core, close to the gas, rather than being associated with the subclusters of galaxies as in the Bullet case (150). It appears that the collision velocity is significantly lower here than for the Bullet, and there may even be more than two colliding systems. Evidently some clarification is needed to decide what really can be learned from colliding clusters.

4. Current status

In summary, the debate over dark matter versus modified gravity undoubtedly is good for the science, because it impels people on both sides to work hard at picking holes in the opposing picture, and repairing such holes in their own. There is no need to make a choice right now, and the fact that a majority of practitioners favor the pure dark-matter explanation need not be decisive.[29]

Of course the majority view faces challenges, of which the biggest is identifying this dark matter, which is four or five times bigger than ordinary matter in its contribution to the mass of the universe. Finding dark matter in the laboratory, perhaps in the form of interactions by dark-matter particles from space incident on sensitive laboratory apparatus, might well settle the issue.

However, it is conceivable that even though there are dark-matter particles, they are too weakly interacting to be detected by feasible apparatus. If so, the case for dark matter would require convincing simulations that reproduce the MOND phenomenology for ordinary-size galaxies.

The main arguments supporing the majority view at the moment are two:

1) Simulations on the largest scales, based on both a cosmological term in the Einstein equations ("dark energy") and cold dark matter, give excellent agreement with a whole set of phenomena. These include the current baryon-to-photon ratio as well as evidence for a spatially flat universe with accelerating expansion. That evidence comes from three complementary classes of information, data on i) distant supernovae, ii) on the structure of the cosmic microwave background, and on iii) large-scale distributions of galaxies.

2) Cold dark matter (CDM) gives a simple and successful foundation for the structure of galaxy clusters, of which the Bullet Cluster is only the most striking example

As we have seen, "fixes" already introduced by MOND advocates to explain the gross features of clusters appear to make MOND also compatible with the Bullet results. Of course the new report about Abell 520 leaves all of this in possible disarray. The structure of ordinary galaxies remains the chief open issue for CDM.[30]

D. Nonzero graviton mass: Is it possible?

1. Early considerations

A naive approach to modifying gravity at long distances would be imitate Proca and introduce a massive graviton analogous to the massive photon. This could be meaningful even though individual gravitons may never be observable. It turns out, however, that the intricate structure of GR makes introduction of a graviton mass a much more delicate exercise. The upshot is, as we will see in the following, that a graviton mass corresponding to a length scale much smaller than the radius of the visible universe appears to be excluded. Certainly any corresponding velocity dispersion would be unobservable.

---

[29] A well-balanced overview (even though from a dark matter advocate) can be found in Ref. (151).

[30] A recent analysis of motion of small "satellite" galaxies around a regular galaxy suggests good agreement with the CDM hypothesis. This does not necessarily rule out MOND for the region just outside a galaxy (152), but may extend down to even smaller scales than before the successful reach of CDM (153).

The study of long-range deviations from GR in this context began in 1939 with papers by Wolfgang Pauli and Markus Fierz [PF] (154; 155). Fierz was Pauli's assistant, and this was his "Habilitation" thesis. They considered particles with finite mass, which meant that in the rest frame of such a particle with spin s there must be $2s + 1$ degrees of freedom. This is in contrast to the two degrees of freedom for massless particles in a theory with parity conservation, or just one degree of freedom for such a particle in the absence of parity symmetry.

For an integer-spin particle with spin wave function represented by a contravariant tensor, one obtains the constraint $\partial_\alpha T^{\alpha\beta\cdots\cdots} = 0$, meaning that in the rest frame the spin wave function is described by a tensor with no time components. For a massive spin-one particle this is just the Lorentz gauge condition discussed earlier for a massive photon. This tensor should be symmetric under interchange of any pair of indices, and traceless in any pair. Simple counting shows that these conditions give $2s + 1$ degrees of freedom if the tensor has s indices.

The focus of PF was on coupling of these massive particles with spin to the electromagnetic field, *not* on speculating about a massive graviton.[31] Indeed, there are difficulties with the electromagnetic coupling to spins higher than $s = \frac{1}{2}$. In particular, Rarita and Schwinger (156) looked at these issues for $s = \frac{3}{2}$. It was not until the 1960s that such matters gained serious attention. This eventually led to a consistent theory for spin-one charged particles, identified first with $SU(2)$ and then with $SU(2) \times U(1)$ non-Abelian gauge theory for electroweak physics, as discussed in II.E. Even later, in the 70s, this kind of consideration was one of the routes that led to supergravity, and its relation to string theory.

The Pauli-Fierz approach to a massive graviton starts with the notion that space-time is approximately flat. Then one may consider small-amplitude deviations and describe them by a wave equation like the Proca equation for the electromagnetic field. The usual Einstein equation is modified by addition of a mass term:

$$G_{\mu\nu} - m^2(h_{\mu\nu} - \eta_{\mu\nu}h) = GT_{\mu\nu} \quad , \qquad (51)$$

where m is the graviton mass in inverse-length units, $\eta_{\mu\nu}$ is the Lorentz metric, $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$ is the departure of the metric from perfect flatness, $h = \eta^{\mu\nu}h_{\mu\nu}$ is the four-dimensional trace of $h_{\mu\nu}$, G is Newton's constant, and $G_{\mu\nu}$ is the Einstein tensor to linear order in $h_{\mu\nu}$:

$$\begin{aligned} G_{\mu\nu} &= \Box(h_{\mu\nu} - \eta_{\mu\nu}h) - \partial^\alpha\partial_\mu h_{\alpha\mu} \\ &\quad -\partial^\alpha\partial_\nu h_{\alpha\nu} + \eta_{\mu\nu}\partial^\alpha\partial^\beta h_{\alpha\beta} + \partial_\mu\partial_\nu h \ . \qquad (52) \end{aligned}$$

Once again, the five degrees of freedom for $h_{\mu\nu}$ with non-zero m must all be present for the limit $m = 0$, but this time the narrow escape for the photon discussed in II.F does not quite work. The helicity-± 1 states appear with a four-gradient factor, and integrating by parts in the coupling to $T_{\mu\nu}$ yields a four-divergence which vanishes because of local conservation of energy and momentum. However, the helicity-0 state multiplies the trace of $T_{\mu\nu}$, and this in general does not vanish. Thus even in the $m = 0$ limit we have a scalar-tensor theory of Fierz-Jordan-Brans-Dicke type (157),(158),(159).

[31] Thus they were following closely Proca's original approach in his papers on a massive spin-1 field.

2. Recent considerations

Surprisingly, the realization of the last paragraph took a long time coming, and then was announced in two almost simultaneous papers, by van Dam and Veltman (160) and by Zakharov (161) [vDV-Z]. Van dam and Veltman found the most striking aspect of their result in a comparison with non-Abelian gauge theories with no Higgs mechanism. Such theories exhibit a transition between behavior at short distances and the confinement at long distances which can only be understood in the context of quantum physics. The paradoxical discontinuity in a mass or inverse-length parameter found by vDV-Z occurs already in purely classical field theory. This was the stimulus for the ensuing theoretical study of graviton mass, and also the beginning of a debate continuing till quite recently over the viability of graviton mass as a meaningful concept.

The next stage in that contest was a paper by Vainshtein (162), who argued that the vDV-Z position, though clearly correct in linear gravity, could be overcome by the intrinsic nonlinearity of Einstein gravity. He started with an argument that, in the vicinity of a gravitational source, the corrections due to graviton mass should be suppressed by a factor of order μ^2L^2, where L is the dimension of the region around the source that is under examination. This argument is quite appealing to the present authors, because we had used exactly the same notion for photon-mass effects, where the continuity of electrodynamics at zero photon mass made our argument correct (2). However, in linearized massive gravity the vDV-Z discontinuity would imply that suppression by μ^2L^2 does not apply.

Very soon there was a riposte by Boulware and Deser [BD] (163; 164), who found a number of reasons to question Vainshtein's conclusion. They noted that graviton mass treated as a fixed constant seems to violate general-coordinate invariance, just as a photon mass seems to violate gauge invariance. As we have seen, there is a way around that through the Stueckelberg construction, and therefore this is not a compelling point. The analogue of the Stueckelberg approach for gravity was introduced by Siegel (165), and discussed more recently by Arkani-Hamed, Georgi, and Schwartz (166).

BD also said, and this *does* seem to us valid, that at best Vainshtein's case was not proved, because his assumptions about behavior near the source might imply exponential growth at large radius, rather than the required exponential decay. What makes this seem a potential obstacle to the Vainshtein construction is precisely the strong coupling to scalar gravitons, even in the zero-mass limit. Vainshtein's hope was that nonlinearity of gravity could heal this problem. One may express this differently as hoping one can continue in from infinity the allowed exponentially decaying behavior. In a purely linear theory this would lead to anomalous behavior near the source.

There the matter rested for about a quarter-century, when a new context of higher-dimensional theories inspired by string theory led to a concrete example with something like graviton mass, the Dvali-Gabadadze-Porrati [DGP] model (167). In this framework, our four (i.e., three plus one)-dimensional world is embedded in a five-dimensional spacetime, with the gravitational action having two pieces, one confined to our world, and the other uniformly defined over the entire five dimensions. The fifth (purely spacelike) dimension is perpendicular to what then is a 'brane' describing the three spatial dimensions of our world.

Interestingly, a group including Vainshtein (168) made the first study using this model for the gravitational field of a massive source, thus giving some vindication for Vainshtein's original position.[32] For a gravitational source sufficiently dilute that one may work to first-order in the mass density of the source, Gruzinov (169) obtained a perturbative solution for the gravitational field. This later was made exact by Gabadadze and Iglesias (170). The solution involves, instead of exponential decay with radius at spatial infinity, a power-law falloff including the scalar component mentioned earlier, but at sufficiently short distances it looks like the Newton-Einstein field.

Dvali (171) recently has explained from a very general perspective how nonlinear coupling can make the zero-mass limit continuous at finite radius. The crucial point, despite the nonlinear nature of Einstein gravity as modified to include something like a mass $\mu_g \equiv r_c^{-1}$ (or some other modification setting in at or above the length scale r_c), is that the source be treatable as linear. This means that in the neighborhood of the source the field is linear in the source strength, while the nonlinearity of gravity itself extinguishes the contributions of the three extra polarizations expected for finite mass. It is quite possible that for a strong source, and in particular a black hole, the conclusions would be different, but perhaps again only at very large length scales or very long times after formation of the black hole.

Even when this type of solution "works" it does so only out to a radius substantially smaller than the effective graviton Compton wavelength $\lambda\!\!\!^{-}_{Cg}$.[33] Because that smaller radius is related to $\lambda\!\!\!^{-}_{Cg}$, Gruzinov argued that knowledge about the solar gravitational field implies (169)

$$\lambda\!\!\!^{-}_{Cg} \gtrsim 10^8 \text{ pc}. \tag{53}$$

There is a potentially significant concern here: The solution is for a source embedded in flat space-time, but we live in a world that appears spatially flat, yet expanding and even accelerating with the passage of time.

The smallest (and also largest) Compton wavelength it would make sense to consider, if one accepts the objections by Boulware and Deser to Vainshtein's argument, would be about the size of the visible universe, of order

$$\begin{aligned} \lambda\!\!\!^{-}_{Cg} &\gtrsim 3\times 10^{26}\text{ m} \equiv 10^{10}\text{ pc} \qquad \text{or} \\ \mu_g &\lesssim 6\times 10^{-32}\text{ eV} \equiv 10^{-67}\text{ kg}, \end{aligned} \tag{54}$$

(meaning exponential growth would not be substantial). Even after all the considerations stemming from DGP one is not far from that value.

The conclusion is that nothing except quasi-static fields could be sensitive to a graviton mass, and quite possibly even such fields would not be able to signal such a mass. Indeed, the consensus even among advocates of the DGP model[34] is that the Compton wavelength may be less than infinity but, as remarked by Nicolis and Rattazzi (173), not appreciably less than the radius of the visible universe. The reason is quite simple: A significantly smaller Compton wavelength inevitably would modify drastically phenomena seen on the largest visible scales. As we discuss later, the DGP model is a possible way of accounting for the accelerating expansion of the universe, but *only* with the largest possible Compton wavelength.

V. PHENOMENOLOGY OF DEVIATIONS FROM GR

A. Inverse square law

1. Inverse square law on large scales

Some time ago it was pointed out that looking for the largest scales over which gravity is known to work is not only a test for dark matter but also, in the paradigm of a Yukawa-like fall off, a limit on a Proca rest mass (174). Even in 1974, when the Hubble constant stood at $H = (55 \pm 7)$ km/(s Mpc), a conservative bound for

[32] However, it is not at all clear that such vindication is possible for the fixed-mass case discussed in the early work. Once again, it appears (as in the Higgs mechanism for gauge theories) that additional degrees of freedom are needed to provide a consistent realization of mass.

[33] Actually, for the DGP model, there is effectively a continuum of masses contributing, which means that the graviton could at best be viewed as an unstable resonance, certainly not a fixed-mass particle.

[34] Gabadadze and Gruzinov (172) give a nice overview of the reasons to describe a mass-like effect using a higher-dimensional theory.

galaxy clusters of size 580 kpc (vs. even then known clusters of size 10 Mpc) yielded a bound of

$$\mu_g \lesssim 2 \times 10^{-65}\,\mathrm{kg} \equiv 10^{-29}\,\mathrm{eV}. \tag{55}$$

This corresponds to a reduced Compton wavelength of

$$\lambdabar_{Cg} \gtrsim 2 \times 10^{22}\ \mathrm{m} \qquad \mathcal{O}(10^{-4}\ R)\ , \tag{56}$$

where $R = c/H$ is the Hubble radius of the universe. Similar assertions, accompanying discussion of enormously less sensitive limits associated with massive graviton decay to two photons and dispersion of gravitational wave velocity, were made slightly earlier by Hare (175).

Indeed, leaving aside the modern finding of dark energy, one can postulate that the ultimate mass limit could be no smaller than the inverse of R itself, as primordial gravitons will be off the quantum mass shell by at least this amount.

2. Small extra dimensions and a "fat" graviton

Kapner et al. (176) recently conducted torsion-balance experiments to test the gravitational inverse-square law at separations between 9.53 mm and 55 μm. This probed distances smaller than the "dark-energy length scale" of

$$d = (\hbar c/\rho_d)^{1/4} \approx 85\ \mu\mathrm{m}. \tag{57}$$

Assuming the coupling is $\leq G$, they found with a 95% confidence level that the inverse-square law holds down to a length scale $\lambda = 56\ \mu$m. They also determined that an extra dimension must have a size $\hat{R} \leq 44\ \mu$. (Also see (177).)

Note that this extra dimension should not be confused with that in the DGP model, which is infinite in extent. The length scale in that model comes from the relative normalization between the five-dimensional and the four-dimensional contributions to the gravitational action. What we are talking about here actually would be a modification of gravity at small distance scales. As such it would be a departure from the main thrust of this paper, although related to the fifth force ideas of finite-sized new forces.

B. Speed of gravity

1. Dispersion in gravitational waves

Recently, a small industry has arisen based on the possibility of finding dispersion in gravitational waves. The starting point is the observation that, at least in some linearized theories, one can allow a massive graviton which would propagate freely via the Klein-Gordon equation of a particle with mass μ_g. If the graviton had a rest mass, the decay rate of an orbiting binary would be affected (178; 179). As the decay rates of binary pulsars agree very well with GR, the errors in their agreements provide a limit on a graviton mass.

Finn and Sutton (180) applied this idea in a reinvestigation of the data from the Hulse-Taylor binary pulsar and from the pulsar PSR B1534+12. From their analysis of the data they found a limit

$$\begin{aligned} \mu_g &\lesssim 7.6 \times 10^{-20}\ \mathrm{eV} \equiv 1.35 \times 10^{-55}\ \mathrm{kg} \\ \lambdabar_{Cg} &\gtrsim 2.6 \times 10^{12}\ \mathrm{m} \end{aligned} \tag{58}$$

to 90% confidence level.[35] This corresponds to a value of v_g whose deviation from c is limited to a value on the order of a part in a thousand (180), at a frequency comparable to the orbital frequency of the binary pulsar system.[36]

Many related ideas have been proposed to measure dispersion in gravitational waves using interferometers or by observing gravitational radiation from in-spiralling, orbiting (non-pulsar) binaries (181)-(184). These should lead to stronger limits if gravitational wave arrivals can be detected. It seems likely that, as in the photon case, such limits never will be as strong as those deduced from quasi-static fields.

2. Shapiro time delay and the speed of gravity

If there were a graviton mass, then there would be dispersion of gravitons of different energies. Intertwined with this is the fact that we tacitly assume that the "limiting velocity" of GR, c_g, is exactly the limiting velocity of light, c. This assumption is not just esthetically pleasing, it also is of fundamental importance.

As mentioned in the introduction, an important aspect of the robustness of scientific theories is the interconnections among different components. If we look at the development of general relativity, then it is clear that the only possible limiting speed for any kind of disturbance is the speed of light, c. It is always worth checking even the most strongly held claims, but one must bear in mind the cost associated with violations of those claims. In this case, the rupture resulting if the speed c_g for gravity turned out to be different from c would be dramatic indeed.

Even with this as background, Kopeikin boldly suggested that if the speed of gravity differed from the speed

---

35 Note that this value is dramatically lees restrictive than that found by looking for departures from the inverse-square law quoted in (55).

36 It should be noted that in the linearized theory, the vDV-Z discontinuity applies, meaning that for finite graviton mass there should be coupling to a scalar graviton. If the radius of the orbit changes appreciably during each cycle, then this would give a comparable contribution to the expected ($\mu_g = 0$) quadrupole radiation. All that could prevent this would be Vainshtein's strong self-coupling for scalar gravitons. It is not clear to us what, if any, effect this strong self-coupling would have on the "standard" graviton-mass effect considered by Finn and Sutton (180).

of light then it could be measured in the Shapiro time delay of the microwave light of a quasar passing close by the foreground of Jupiter (185). Kopeikin claimed that he effect would be a first order correction, caused by the retarded gravity signal due to Jupiter's velocity, v/c_g.

However, Will criticized this assertion (186). His first statement was that retarded-potential theory would yield an effect only to order $(v/c_g)^2$. Motivated by this he used the PPN expansion of GR to find that the first-order correction to the Shapiro time delay is (in the GR limit)

$$\Delta = -\frac{2Gm_J}{c^3}\ln\left(|\mathbf{x}_{\odot J}| - \mathbf{x}_{\odot J}\cdot\mathbf{k}\right) \quad \rightarrow \tag{59}$$

$$- \frac{2Gm_J}{c^3}\left[\ln\left(|\mathbf{x}_{\odot J}| - \mathbf{x}_{\odot J}\cdot\mathbf{K}\right)\left(1-\frac{\mathbf{K}\cdot\mathbf{v}_J}{c}\right)\right] \tag{60}$$

$$\mathbf{K} \equiv \mathbf{k} - [\mathbf{k}\times(\mathbf{v}_J\times\mathbf{k})]/c. \tag{61}$$

The difference between Eqs. (59) and (60) gives the first-order velocity correction, where m_J is the mass of Jupiter, c is the speed of light, $\mathbf{x}_{\odot J}$ is the distance vector from the observer on Earth to Jupiter's center, and $\mathbf{k}$ is the unit vector in the direction of the incoming light.

Note that "c" is to be found in Eq. (60) in two different places. This is where the disagreement is. Kopeikin would have the c's inside the square brackets be c_g's. Contrarily, Will calculates Eq.(60) from GR with $c =$ "c". These terms are thus found to be the next order GR time-delay.

Therefore, Will finds that agreement of this formula with experiment is a (not too precise (186)) test of GR rather than a test of c_g. Will finds that any $c_g \neq c$ effects would only appear in the next order (c_g^{-2}). Similar conclusions were drawn by others (187; 188). The consensus (189) agrees with this conclusion, despite the continuing disagreement of the Kopeikin school (190; 191).

There is a an appealing way to motivate this position. In first approximation, the Shapiro time delay is an effect on the propagation of light in an essentially static gravitational field, so that the speed of gravitational waves should not be immediately relevant. Furthermore, if a heavy source is moving with respect to an observer, to first order in the source velocity the only change in the field shows no effect of retardation. Simply on dimensional grounds, acceleration of the source at most would give an effect second-order in the inverse speed of gravity.

In any event, a measurement was done when the quasar J0842+ 1835 passed within 3.7′ of Jupiter on 8 Sept. 2002. Fomalont and Kopeikin (192), compared the $\sim (51 \pm 10)$ μas deflection observed with the higher-order term. They determined a value for "c_g" of

$$c_g = (1.06 \pm 0.19)\, c\,. \tag{62}$$

This result is consistent both with standard GR (where any effect of $c_g \neq c$ appears only in order $(v/c_g)^2$) and also with Kopeikin's theory, but with $c \rightarrow c_g$ inside the large brackets of Eq. (60). Therefore, experimentally no nonstandard result is found under either interpretation.

C. Meta-phenomenology: Looking at all questions together

Before even the VdV-Z discontinuity or the Vainshtein nonlinear-gravity effect attenuating this discontinuity comes the prime fact stressed repeatedly in this : any graviton-mass effect begins with the *weakening* of gravitational attraction at long distances. Galactic rotation curves and motions of objects in clusters of galaxies apparently exhibit exactly the opposite effect: as *strengthening* of gravity at increasing distances from the center compared to Newtonian expectations based on the visible matter in a system.

Thus, the one thing "dark matter" could *not* be is a graviton-mass effect, where we include in this category generalizations such as the DGP model. This fact may increase the attractiveness of supposing that some combination of discrete-particle dark matter and continuous-field dark matter might account for these phenomena; in other words, there may be new sources rather than modifications of Newton-Einstein gravity.

On the other hand, the accelerating expansion of the universe, indicated by numerous observations in the last decade, *is* what one might expect from a weakening of gravity at large distances. Dvali, Gruzinov and Zaldarriaga (193) have considered the possibility that the length scale corresponding to a DGP 'graviton-mass' effect may be comparable with the size of the visible universe, and so could explain quantitatively the observed acceleration. They then note that there should be small effects at distances within the solar system.[37]

In particular, the precession of the perigee[38] of the moon's orbit [39][40] should have a contribution about an order of magnitude smaller than the sensitivity of current measurements using laser lunar ranging. It is possible that observational sensitivity could increase suffi-

[37] This is because of Vainshtein's nonlinear suppression of graviton-mass effects at distances very short compared to the graviton Compton wavelength.

[38] The authors refer to 'perihelion', but in context it seems clear that 'perigee' is intended.

[39] In a way, this brings us back to the beginning. In the Principia, one thing Newton could not calculate satisfactorily was the precession of the moon's line of apsides. As described in Ref. (115), this necessitated two further advances. The first, was the development of equations of motion for the 3-body problem. The second was the inclusion of the effects of the sun's motion with respect to the earth-moon system, mainly caused by Jupiter. (Previously the sun had been treated as lying at a constant set point.) This was done in 1749 by the Frenchman Alexis-Claude Clairaut.

[40] In 1758 Clairaut applied his perturbation theory to the timing of the return of Halley's comet, with great success (115). Thereupon Clairaut became a scion of the Paris salons. (The French treated scientists well then, as Benjamen Franklin would discover.) "Engaged with suppers, late nights, and attractive women, desiring to combine pleasures with his ordinary work, he was deprived of his rest, his health, and finally at the age of only 52, of his life (115)."

ciently to detect such a precession. If so, that would tend to confirm a graviton-mass explanation for accelerating expansion, clearly an example of modified gravity. If such a shift in the perigee precession were ruled out, then that might be an indication instead favoring a modified source, i.e., some form of dark energy.

VI. DISCUSSION AND OUTLOOK

A. Conclusions

The subject of possible photon or graviton rest mass is appealing because there are so many levels of beautiful argument for the masses to vanish (and of course a counter-argument for each argument). Both of these examples (of the only long-range fields we know) reveal important strands of physics relevant to many different areas. Taken as a whole, their study illuminates the history, logic, and remarkably complex yet interlocking structure of physics.

In Table I we give a list of what we find to be the most significant and/or interesting mass limits so far proposed.

The first reason for vanishing mass is, of course, something Newton adopted instinctively for gravity and Gauss justified in a beautiful way for electrostatics: the inverse square law of force. For gravity this was confirmed with tremendous precision by the match with Kepler's laws. For electrostatics Gauss's statement that the number of lines of force coming out of a charge is, at any distance, a direct measure of that charge, gave a powerful geometric interpretation to the force law. (This argument applies equally to mass in gravity.) Later there came the notions of gauge invariance, discovered in the mathematical structure of classical electrodynamics, and general-coordinate invariance, invented by Einstein to constrain the possible structure of his emerging theory of gravity.

There is also a "backwards" connection: As Weinberg showed in S-matrix theory, zero mass for photon or graviton implies local conservation of electric charge in electrodynamics or energy and momentum in gravity. Like the sizes of the masses, violations of these conservation laws are strongly constrained by experiment. Thus, for electrodynamics as well as gravity, two effects known to be small are logically related in the limit where they both are zero.

Simplicity also favors these zeros. They represent the minimal structure consistent with all symmetry requirements. Anything different requires more parameters if not more fields.

Despite all these arguments,there is another side to the story. As Stueckelberg showed, gauge invariance can be satisfied at least formally even in the presence of a mass (and Siegel showed that the same holds for general-coordinate invariance). Perhaps even more powerful is the example of the $W^{\pm}$ and Z^0 mesons, which show every sign of being gauge-coupled, yet clearly have mass.

At least for the photon case, there is an objection to this argument. In the context of a Higgs mechanism one must introduce an electrically charged scalar field, where constraints from observation imply that the value of the charge is an extraordinarily tiny fraction of the charge of an electron. Such a charge, of course, would violate the pattern of all known charges, and also would contradict an appealing (though unproved) idea, that of grand unified theory. It would be a bizarre modification of electroweak theory, where the compact group $SU(2)_{\rm left}$ automatically leads to quantized left-handed charges.

In other words, 'mini-charged' Higgs particles, if they existed, would perforce be coupled only to the $U(1)_{\rm right}$. Because weak interactions are short-ranged, and the W and Z bosons are so massive, the effects of the weak charge of the new Higgs particle would be even more insignificant than those of the electric charge. The effect of the new Higgs coupling on the masses of the W and Z also would be unobservably small.[41]

As far as observation relevant to photon mass goes, the only debate is about how stringent a limit currently can be placed on that mass. To date no evidence at all has appeared for a nonzero value. Even with the generalization to a Higgs-mechanism framework, the "obvious" experiment of seeking to detect dispersion of velocity with frequency is guaranteed to give no useful information, because limits from static magnetic fields are so low that nothing could be detected by dispersion measurements (at least in regions identified so far where we could measure the velocity). Even the Schumann resonances do not give as strong a constraint as the magnetostatic limit. Thus, the only even potentially observable effect of photon mass would be found in the photon Compton wavelength, and that already is known to be at least comparable in dimension to the earth-sun distance. Therefore, any future improvements in the limit probably will come from astronomical observation rather than laboratory (even satellite-laboratory) experiment.

The story for graviton mass is different in two principal aspects. The argument that the mass must be zero (or at least no bigger than the inverse of the Hubble radius) is much stronger than even what we have just related for photon mass. On the other hand, there *are* observations that can not be reconciled with unadorned GR unless there are new forms of matter, described as dark matter and dark energy. Thus, if we generalize the notion of graviton mass to that of long-distance, low-frequency modifications of gravity, then there may indeed be modifications, even though (as we have seen) there are powerful arguments in favor of extra sources rather than changed gravity.[42]

---

[41] If instead of a Higgs particle, the associated Higgs field were a composite of other fields, then these too would have extraordinarily small, unquantized $U(1)_{\rm right}$ charges, or at least super-small offsets of the charges for different elements of the composite.

[42] Actually, for dark energy the choice between a new source and modification of gravity is not necessarily well-defined. Einstein's

TABLE I A list of the most significant mass limits of various types for the photon and graviton.

Description of method	$\lambda\!\!\!^{-}_C \gtrsim$ (m)	$\mu \lesssim$ (eV)	$\mu \lesssim$ (kg)	Comments
1 Secure photon mass limits:				
Dispersion in the ionosphere (84)	8×10^5	3×10^{-13}	10^{-49}	
Coulomb's law (77)	2×10^7	10^{-14}	2×10^{-50}	
Jupiter's magnetic field (86)	5×10^8	4×10^{-16}	7×10^{-52}	
Solar wind magnetic field (89)	$2\times$!0^{11} (1.3 AU)	10^{-18}	2×10^{-54}	
2 Speculative photon mass limits:				
Lake's method (95–97)	3×10^9 $\Leftrightarrow 3 \times 10^{12}$	7×10^{-7} $\Leftrightarrow 7 \times 10^{-20}$	10^{-52} $\Leftrightarrow 10^{-55}$	$\lambda\!\!\!^{-}_C \sim 4\, R_\odot$ to 20 AU, depending on **B** speculations.
Higgs photon (69)	3×10^{19} (1 kpc)	6×10^{-26}	10^{-62}	Needs constant **B** in galaxy regions.
Cosmic magnetic fields (90; 94)	10^{20} (10^4 ly)	2×10^{-27}	4×10^{-63}	Needs constant **B** in galaxy regions.
3 Graviton mass limits:				
Grav. wave dispersion (180)	3×10^{12}	8×10^{-20}	10^{-55}	
Gravity over cluster sizes (174)	2×10^{22}	10^{-29}	2×10^{-65}	
Gruzinov graviton (169)	3×10^{24} (10^8 pc)	6×10^{-32}	10^{-67}	
DGP graviton (167)	3×10^{26} (10^{10} pc)	6×10^{-32}	10^{-67}	

B. Prospects

Accepting that the primary tool for limiting or detecting a rest mass of the photon or graviton is to exploit the "Yukawa" effect – modification at long distances of essentially static electromagnetic or gravitational fields – possibilities for extending the range in the case of electromagnetism look very good. With rapidly evolving instruments and techniques, we are in an era of rapid expansion in the depth of exploration of the universe. As detailed knowledge of galactic and extragalactic magnetic fields accumulates, along with knowledge about the structure of the associated plasmas, there is every reason to suppose that a lower limit on the photon Compton wavelength limit of galactic or even larger dimensions could be attainable.

If a finite value for $\lambda\!\!\!^{-}_C$ were detected, almost certainly it would be so large and the corresponding photon mass so small that even in the Higgs framework the corresponding electric "mini-charges" would be too small to detect. Thus the continuity of electrodynamics in the zero-mass limit (unless electric charge is not locally conserved) already assures that the *only* mass effect still possible to observe would be long-distance modifications of static magnetic fields.

In other words, for all lab-scale purposes the mass already may be taken as zero. Still if a non-zero value were established by new astronomical observations, this small departure would have enormous conceptual implications, giving incentive for searching examination of the accepted foundations of electrodynamics.

For gravity the situation is much more dramatic. Even though the consensus on the evidence favors keeping Einstein gravity unchanged, and seeking new, massive, weakly interacting particles to account for "dark matter" effects, the case against the alternative, "modified gravity," is not closed. Note that in the Bekenstein formulation the modification of gravity again may be ascribed to dark matter, but this time in the form of classical fields at most weakly coupled to ordinary matter, rather than classical, weakly-interacting particles.

It is worth emphasizing the complementarity between the classical dark-matter particles, which describe very well phenomena at large scales down to galaxy-cluster size, and the classical dark-matter fields, which describe very well phenomena at and below galactic scales. The assumption of partisans favoring either view, that if it is correct the other must be wrong, might turn out to be as false here as it was in the dispute over wave versus particle pictures of light.

Unquestionably the competition between two different understandings of deviant observations has been healthy for progress in the field. Because dark matter and dark

cosmological constant was in his own eyes a modification of gravity. On the other hand, inflation models involve a scalar field whose vacuum energy drives exponentially rapid expansion, and such a dynamical cosmological term is placed most naturally on the 'matter' side of the Einstein equations.

energy were not widely anticipated before the observations which elicited these concepts, one cannot rule out the possibility of still further surprises. Thus, galactic and larger-scale dynamics involving gravity represent a most exciting frontier of modern physics and astronomy.

At the same time, from a theoretically-inspired point of view the great debate about the possibility of a massive graviton in GR seems pretty much complete. The most conservative lower bound on the graviton Compton wavelength, based on limits to deviations from Einstein-Newton gravity in the solar system, puts it at $\sim$1% of the radius [R] of the visible universe (169). Even that estimate is for an asymptotically flat space time, whereas the actual universe is expanding and even accelerating in its expansion.[43]

Taking all known now about photon and graviton mass (and other possible modifications of the two great classical theories) into account, and remembering that physics is an experimental science, it behooves the community to continue to search, keeping a watch for those possible additional surprises.

Acknowledgments

To begin, it is most appropriate to acknowledge T. Alexander Pond. It was his informal lecture at Stony Brook on 12 June 1968 (5), which was the original inspiration for our collaboration.

Further thanks are due to John Anderson, Nima Arkani-Hamed, Maruša Bradač, Bruce Draine, Georgi Dvali, Glennys Farrar, Éanna Flanagan, Alexander Friedland, Gregory Gabadadze, George Gillies, Donald Groom, Andrei Gruzinov, Marek Karliner, Philipp Kronberg, Philip Mannheim, Mordehai Milgrom, Lyman Page, Ari Pakman, Massimo Porrati, Joel Primack, Dmitri Ryutov, Warren Siegel, David Spergel, Volker Springel, Scott Tremaine, Virginia Trimble, Peter van Nieuwenhuizen, and Chen Ning Yang.

The work of ASG was supported in part by the National Science Foundation. That of MMN was supported by the United States Department of Energy.

References

[1] T. S. Kuhn, *The Structure of Scientific Revolutions*, 3rd ed. (Univ. of Chicago Press, Chicago, 1996).

[2] A. S. Goldhaber, and M. M. Nieto, "Terrestrial and Extraterrestrial Limits on the Photon Mass," Rev. Mod. Phys. **43**, 277-296 (1971).

[3] J. C. Byrne, "Cosmic Tests of Maxwell's Equations," Astrophys. and Space Sci. **46**, 115-132 (1977).

[4] J. D. Barrow and R. R. Burman, "New Light on Heavy Light," Nature **307**, 14-15 (1984).

[5] M. M. Nieto, "Past, Present, and Possible Future Limits on the Photon Rest Mass," in: *Gamma Ray - Neutrino Cosmology and Planck Scale Physics*, ed. by D. B. Cline (World Scientific, Singapore, 1993), pp. 291-296. ArXiv hep-ph/9212283.

[6] L.-C. Tu and J. Luo, "Experimental Tests of Coulomb's Law and the Photon Rest Mass," Metrologia **41**, S136-S146 (2004).

[7] L.-C. Tu and J. Luo, and G. T. Gillies, "The Mass of the Photon," Repts. Prog. Phys. **68**, 77-130 (2005).

[8] L. B. Okun, "Photon: History, Mass, Charge," Acta. Phys. Polonica B **37**, 565-573 (2006).

[9] C. Amsler et al., The Particle Data Group, "2009 Review of Particle Properties," Phys. Lett. B **667**, 1 (2008) W.-M. Yao *et al.*, The Particle Data Group, See pp. 359-360 for the photon and graviton masses.

[10] M. Born and L. Infeld, "Foundations of the New Field Theory," Proc. Roy. Soc. London A **144**, 425-451 (1934).

[11] J. Plebanski, *Lectures in Nonlinear Electrodynamics*, (Nordita, Copenhagen, 1970).

[12] A. A. Tseytlin, "Born-Infeld action, Supersymmetry and String Theory," in: *The Many Faces of the Superworld: Yuri Golfand Memorial Volume*," ed. M. A. Shifman (World Scientific, Singapore, 2000), pp. 417-452. Ar Xiv: hep-Th/9908105.

[13] G. W. Gibbons, "Aspects of Born-Infield Theory and String/M-Theory," Rev. Mex. Fis. **49-S1**, 19-29 (2003). ArXiv: hep-th/0106059.

[14] W. Heisenberg and H. Euler, "Folgerungen aus der Diracshen Theorie des Positrons," Zeit. Phys. **98**, 714-732 (1936).

[15] G. V. Dunne, "Heisenberg-Euler Effective Lagrangians: Basics and Extensions," in: *From Fields to Strings: Circumnavigating Theoretical Physics: Ian Kogan Memorial Collection*, Vol I, eds: M. Shifman, A. Vainshtein, and J. Wheater (World Scientific, Singapore, 2004), pp. 445-522. ArXiv: hep-th/0406216. Also see http://webusers.physics.umn.edu/ vainshte/Kogan/

[16] A. S. Goldhaber and M. M. Nieto, "The Photon Rest Mass," Sci. Am. **234**, No. 5, 86-96 (May, 1976).

[17] S. J. Plimpton and W. E. Lawton, "A Very Accurate Test of Coulomb's Law of Force Between Charges," Phys. Rev. **50**, 1066-1071 (1936).

[18] H. Kragh, "Ludvig Lorenz and Nineteenth Century Optical Theory: the Work of a Great Danish Scientist," Applied Optics **30**, 4688-4695 (1991).

[19] J. D. Jackson and L. B. Okun, "Historical Roots of Gauge Invariance," Rev. Mod. Phys. **73**, 663-680 (2001). ArXiv: hep-ph/0012061.

[20] A. Proca, "Sur la Théorie du Positron," Acad. Sci., Paris, Comptes Rendus **202**, 1366-1368 (1936).

[21] A. Proca, "Sur la Théorie Ondulatoire des Électrons Positifs et Négatifs," J. Physique Radium, Ser. VII, **7**, 347-353 (1936).

[22] A. Proca, "Sur les Photons et les Particules *charge*

[43] So, if spatially flat, the universe still has (negative) curvature in the relation between time and space variables. This curvature introduces a ghost in the DGP theory, making it suspect, and therefore making the limit (169) possibly too conservative. Perhaps more to the point, a $\lambda\!\!\!^{-}_C$ significantly smaller than R would produce dramatic (and certainly not seen) effects on the large-scale picture of the universe.

pure," Acad. Sci., Paris, Comptes Rendus **203**, 709-711 (1936).

[23] A. Proca, "Particles Libres: Photons et Particules 'charge pure'," J. Physique Radium, Ser. VII, **8**, 23-28 (1937).

[24] A. Proca, "Théorie Non Relativiste des Particles a Spin Entier," J. Physique Radium, Ser. VII, **9** 61-66 (1938).

[25] L. de Broglie, *La Méchanique Undulatoire du Photon, Une Novelle Théorie de la Lumière* (Hermann, Paris, 1940), pp. 39-40.

[26] H. Yukawa, "On the Interaction of Elementary Particles. I," Proc. Phys.-Math. Soc. Japan **17**, 48-57 (1935). Reprinted in Supl. Prog. Theor. Phys. **1**, 1-10 (1955).

[27] H. Yukawa, S. Sakata, and M. Taketani, "On the Interaction of Elementary Particles. III," Proc. Phys.-Math. Soc. Japan **20**, 319-340 (1938). Reprinted in Supl. Prog. Theor. Phys. **1**, 24-45 (1955).

[28] N. Kemmer, "The Particle Aspect of Meson theory," Proc. Roy. Soc. London A **173**, 91-116 (1939).

[29] E. Schrödinger, "On the Solutions of Wave Equations for Non-Vanishing Rest-Mass Including A Source Function," Proc. Roy. Irish Acad. A **47**, 1-23 (1941).

[30] E. Schrödinger, "The General Unitary Theory of the Physical Fields," Proc. Roy. Irish Acad. A **49**, 43-58 (1943).

[31] E. Schrödinger, "The Earth's and the Sun's Permanent Magnetic Fields in the Unitary Field Theory," Proc. Roy. Irish Acad. A **49**, 135-148 (1943).

[32] L. Bass and E. Schrödinger, "Must the Photon Mass be Zero?" Proc. Roy. Soc. London A **232**, 1-6 (1955).

[33] S. Weinberg, "Photons and Gravitons in S Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass," Phys. Rev. **135**, B1049-B1056 (1964).

[34] L. B. Okun and Y. B. Zeldovich, "Paradoxes of Unstable Electron," Phys. Lett. B **78**, 597-600 (1978).

[35] L. B. Okun and M. B. Voloshin, "On the Electric Charge Conservation," JETP Lett. **28**, 145-149 (1978) [Pisma Zh. Eksp. Teor. Fiz. **28**, 156-160 (1978)].

[36] L. B. Okun, "Tests of Electric Charge Conservation and the Pauli Principle," Sov. Phys. Usp. **32**, 543-547 (1989) [Usp. Fiz. Nauk **158**, 293-301 (1989)].

[37] S. Nussinov, "Charge Nonconserving Decays," Phys. Rev. Lett. **59**, 2401-2404 (1987).

[38] R. N. Mohapatra and S. Nussinov, "Electric Charge Nonconservation and Minicharged Particles: Phenomenological Implications," Int. J. Mod. Phys. A **7**, 3817-3834 (1992).

[39] Y. Aharonov, F. T. Avignone, III, R. L. Brodzinski, J. I. Collar, E. Garca, H. S. Miley, A. Morales, J. Morales, S. Nussinov, A. Ortiz de Solórzano, J. Puimedón, J. H. Reeves, C. Sáenz, A. Salinas, M. L. Sarsa, and J. A. Villar, "New Laboratory Bounds on the Stability of the Electron," Phys. Rev. D **52**, 3785-3792 (1995).

[40] M. M. Tsypin, "Longitudinal Photons and Conservation of Charge," Sov. J. Nucl. Phys. **50**, 269-274 (1989) [Yad. Fiz. **50**, 431-441 (1989)].

[41] B. Rosenstein and A. Kovner, "Masslessness of the Photon and Goldstone Theorem," Int. J. Mod. Phys. A **6**, 3559-3569 (1991).

[42] H. Weyl, "Reine Infinitesimalgeometrie," Math. Zeit. **2**, 384-411 (1918).

[43] H. Weyl, "Eine neue Erweiterung der Relativitätstheorie," Ann. Phys. (Leipzig) **59**, 101-133 (1919).

[44] H. Weyl, "Gravitation and the Electron," Proc. Natl. Acad. Sci. USA **15**, 323-333 (1929).

[45] H. Weyl, "The Problem of Symmetry in Quantum Mechanics," J. Franklin Insti. **207**, 509-518 (1929).

[46] H. Weyl "Electron und Gravitation," Zeit. Phys. **56**, 330-352 (1929).

[47] O. Klein, "Quantentheorie und funfdimensionale Relativitätstheorie," Zeit. Phys. **37**, 895-906 (1926).

[48] V. Fock, "Über die invariante Form der Wellen- und der Bewegungensgleichungen für einen geladenen Massenpunkt," Zeit. Phys. **38**, 226-232 (1926).

[49] F. London, "Die Theorie von Weyl und die Quantenmechanik," Naturwiss. **15**, 187 (1927).

[50] F. London, "Quantenmechanischke Deutung der Theorie von Weyl," Zeit. Phys. **42**, 325-389 (1927).

[51] E. Schrödinger, "Über eine bemerkenswerte Eigenschaft der Quantenbahnen eines einzelnen Elektrons," Zeit. Phys. **12**, 13-23 (1922).

[52] E. C. G. Stueckelberg, "Théorie de la Radiation de Photons de Masse Arbitrairement Petite," Helv. Phys. Acta **30**, 209-215 (1957).

[53] C. N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance," Phys. Rev. **96**, 191-195 (1954).

[54] J. S. Schwinger, "Gauge Invariance and Mass," Phys. Rev. **125**, 397-398 (1962).

[55] J. S. Schwinger, "Gauge Invariance And Mass. Part II," Phys. Rev. **128**, 2425-2429 (1962).

[56] P. W. Anderson, "Plasmons, Gauge Invariance, and Mass," Phys. Rev. **130**, 439-442 (1963).

[57] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," Phys. Rev. Lett. **13**, 321-323 (1964).

[58] P. W. Higgs, "Broken Symmetries, Massless Particles and Gauge Fields," Phys. Lett. **12**, 132-133 (1964).

[59] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," Phys. Rev. Lett. **13**, 508-509 (1964).

[60] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons," Phys. Rev. **145**, 1156-1163 (1966).

[61] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," Phys. Rev. Lett. **13**, 585-587 (1964).

[62] A. A. Migdal and A. M. Polyakov, "Spontaneous Breakdown of Strong Interaction Symmetry and the Absence of Massless Particles," Sov. Phys. JETP **24**, 91-98 (1967). [Zh. Eksp. Teor. Fiz. **51**, 135-146 (1966).]

[63] G. t'Hooft, "Renormalization of Massless Yang-Mills fields." Nucl. Phys. B **33**, 173-199 (1971).

[64] G. t'Hooft "Renormalizable Lagrangians for Massive Yang-Mills Fields," Nucl. Phys. B bf35, 167-188 (1971).

[65] J. R. Primack and M. A. Sher, "Photon Mass at Low Temperature?" Nature **288**, 680-681 (1980).

[66] S. Davidson, S. Hannestad, and G. Raffelt, "Updated Bounds on Milli-Charged Particles," JHEP **2000**, No. 5, 003/1-13 (2000). ArXiv: hep-ph/0001179.

[67] L. F. Abbott and M. B. Gavela, "Absence of Thermal Effect on Photon Mass Measurements," Nature **299**, 187 (1982).

[68] M. A. Sher and J. R. Primack, "Sher and Primack's Reply," Nature **299**, 187 (1982).

[69] E. Adelberger, G. Dvali, and A. Gruzinov, "Photon

Mass Bound Destroyed by Vortices," Phys. Rev. Lett. **98**, 010402/1-4 (2007) ArXiv hep-ph/0306245.

[70] L. Bass, "Radiation with a Finite Rest-Mass and the Heat Balance of the Earth," Nuovo Cimento **3**, 1204-1212 (1956).

[71] P. W. Milonni, *Fast Light, Slow Light, and Left-Handed Light* (IOP Pub., Bristol, 2005).

[72] G. Barton and N. Dombey, "Casimir Effect For Massive Photons," Nature **311**, 336-339 (1984).

[73] G. Barton and N. Dombey, "The Casimir Effect With Finite Mass Photons," Annals Phys. **162**, 231-272 (1985).

[74] A. S. Goldhaber and M. M. Nieto, "How to Catch a Photon and Measure its Mass," Phys. Rev. Lett. **26**, 1390-1392 (1971).

[75] D. Park and E. R. Williams, "Comments on a Proposal for Determining the Photon Mass," Phys. Rev. Lett. **26**, 1393-1394 (1971).

[76] N. M. Kroll, "Theoretical Interpretation of a Recent Experimental Investigation of the Photon Rest Mass," Phys. Rev. Lett. **26**, 1395-1398 (1971).

[77] E. R. Williams, J. E. Faller, and H. A. Hill, "New Experimental Test of Coulomb's Law: A Laboratory Upper Limit on the Photon Rest Mass," Phys. Rev. Lett. **26**, 721-724 (1971).

[78] L. P. Fulcher, "Improved Result for the Accuracy of Coulomb Law - A Review of the Williams, Faller, and Hill Experiment," Phys. Rev. A **33**, 759-761 (1986).

[79] R. E. Crandall, "Photon Mass Experiment," Am. J. Phys. **51**, 698-702 (1983).

[80] D. G. Boulware and S. Deser, "The Aharonov-Bohm Effect and the Mass of the Photon," Phys. Rev. Lett. **63**, 2319 (1989).

[81] J. J. Ryan, F. Accetta, and R. H. Austin, "Cryogenic Photon-Mass Experiment," Phys. Rev. D **32**, 802?05 (1985).

[82] N. Dombey, "How Heavy is a Cold Photon," Nature **288**, 643-644 (1980).

[83] G. Feinberg, "Pulsar Test of a Variation of the Speed of Light with Frequency (Optical and Radio Pulses from Pulsars as Test of Light Speed Variation with Frequency and Photon Mass Possible Existence)," Science **166**, 879-881 (1969).

[84] N. M. Kroll, "Concentric Spherical Cavities And Limits On The Photon Rest Mass," Phys. Rev. Lett. **27**, 340-343 (1971).

[85] M. Füllekrug, "Probing the Speed of Light with Radio Waves at Extremely Low Frequencies," Phys. Rev. Lett. **93**, 043901/1-3 (2004).

[86] L. Davis, Jr., A. S. Goldhaber, and M. M. Nieto, "Limit on the Photon Mass Deduced from Pioneer-10 Observations of Jupiter's Magnetic Field," Phys. Rev. Lett. **35**, 1402-1405 (1975).

[87] V. A. Kostelecký and M. M. Nieto, "Bounding the Photon Rest Mass from Solar Probes," Phys. Lett. B **317**, 223-226 (1993). ArXiv hep-ph/9304250.

[88] D. D. Ryutov, "The Role of Finite Photon Mass in Magnetohydrodynamics of Space Plasmas," Plasma Phys. Controlled. Fusion **39**, A73-A82 (1997).

[89] D. D. Ryutov, "Using Plasma Physics to Weigh the Photon," Plasma Phys. Controlled Fusion," **49**, B429-B438 (2007).

[90] Y. Yamaguchi, "A Composite Theory of Elementary Particles," Prog. Theoret. Phys. Supp. **11**, 1-36 (1959).

[91] R. Beck, "Magnetic Field Structure from Synchrotron Polarization," in: *Sky Polarisation at Far-Infrared to Radio Wavelengths: The Galactic Screen before the Cosmic Microwave Background*, eds. M. A. Miville-Deschenes and F. Boulanger, EAS Pub. Ser. **23**, 19-36 (2007).

[92] E. Williams and D. Park, "Photon Mass and the Galactic Magnetic Field," Phys. Rev. Lett. **26**, 1651-1652 (1971).

[93] J. C. Byrne and R. R. Burman, "On Fundamental Electrodynamics and Astrophysics," Nature **253**, 27 (1975).

[94] G. V. Chibisov, "Astrophysical Upper Limits on the Photon Rest Mass," Sov. Phys. Usp. **19**, 624-626 (1976). [Uspek. Fiz. Nauk **119**, 551-555 (1976).]

[95] R. Lakes "Experimental Limits on the Photon Mass and Cosmic Magnetic Vector Potential," Phys. Rev. Lett. **80**, 1826-1829 (1998).

[96] J. Luo, L.-C. Tu, Z.-K. Hu, and E.-J. Luan, "New Experimental Limit on the Photon Rest Mass with a Rotating Torsion Balance," Phys. Rev. Lett. **90**, 081801/1-4 (2003).

[97] A. S. Goldhaber and M. M. Nieto, "Problems of the Rotating-Torsion-Balance Limit on the Photon Mass," Phys. Rev. Lett. **91**, 149101 (2003). ArXiv hep-ph/0305241.

[98] J. Luo, L.-C. Tu, Z.-K. Hu, and E.-J. Luan, "Luo et al. Reply," Phys. Rev. Lett. **91**, 149102 (2003).

[99] C.L. Carilli, and G.B. Taylor, "Cluster magnetic fields," Ann. Rev. Astron. Astrophys. **40**, 319 (2002).

[100] H. Ohno, M. Takada, K. Dolag, M. Bartelman, and N. Sugiyama, "Probing Intercluster Magnetic Fields with Cosmic Microwave Background," Astrophys. J. **584**, 599 (2003).

[101] T. E. Clarke, P. P. Kronberg, and H. Boehringer, "A New Radio-X-Ray Probe of Galaxy Cluster Magnetic Field." Astrophys. J **547**, L111 (2001).

[102] W. B. Ashworth, Jr., in: J. Langford, ed., *History of Astronomy. An Encyclopedia,* (Garland Pub., New York, 1997), pp. 97-99.

[103] A. N. Cox, *Allen's Astrophysical Quantities,* 4th Ed., (AIP Press, New York, 2000).

[104] R. S. Westfall, *Never at Rest:A Biography of Isaac Newton* (Cambridge U. Press, 1980).

[105] B. E. Clotfelter, "The Cavendish Experiment as Cavendish Knew it," Am. J. Phys. **55**, 210-213 (1987).

[106] H. Cavendish, "Experiments to Determine the Density of the Earth," Proc. R. Soc. London **88**, Pt II, 469-526 (1798).

[107] J. H. Poynting, *The Mean Density of the Earth: an Essay to which the Adams Prize was Adjudged in 1893,* (Charles Griffen & Co., London 1894).

[108] A. Cornu and J. Baille, "Détermination nouvelle de la constante de L'attraction et de la densité moyenné de la Terre," Acad. Sci., Paris, Comptes Rendus **76**, 954-958 (1873).

[109] M. Grosser, *The Discovery of Neptune,* (Harvard U. Press, Cambridge, 1962).

[110] M. M. Nieto, *The Titius-Bode Law of Planetary Distances: Its History and Theory* (Pergamon Press, Oxford, 1972).

[111] M. M. Nieto, "The Letters between Titius and Bonnet and the Titius-Bode Law of Planetary Distances," Am. J. Phys. **53**, 22-25 (1985).

[112] W. Fricke, in: *Dictionary of Scientific Biography,* Vol.

II, ed. C. C. Gillispie, (Charles Scribner's Sons, NY, 1970), pp. 97-102.

[113] U. J. J. le Verrier, "Théorie du mouvement de Mercure," Ann. Observ. Imp. Paris **V**, 1-196 (1859). See p. 78 and 99.

[114] U. J. J. le Verrier, "Théorie nouvellede mouvement de la planète Neptune: Remarques sur l'ensemble des thóries huit planètes princepales: Mercure, Venus, la Terre, Mars, Jupiter, Saturn, Uranus, et Neptune; par M. Le Verrier," Acad. Sci., Paris, Comptes Rendus **79**, 1421-1427 (1874).

[115] R. Baum and W. Sheehan, *In Search of Planet Vulcan: The Ghost in Newton's Clockwork Universe*, (Basic, Cambridge, 1997). See. p. 136.

[116] J. Bertrand, "Mechanique analytique," Comptes Rendus Acad. Sci., Paris, **77**, 849-873 (1873).

[117] A. Hall, "A Suggestion in the Theory of Mercury," Astron. J. bf 14, 49-51 (1894).

[118] C. M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge U. Press, Cambridge, 1993).

[119] C. M. Will, "The Confrontation between General Relativity and Experiment," Astrophys. Space Sci. **283**, 543-552 (2003).

[120] V. A. Kostelecký and S. Samuel, "Photon and Graviton Masses in String Theory," Phys. Rev. Lett. **66**, 1811-1814 (1991).

[121] E. Fischbach, G. T. Gillies, D. E. Krause, J. G. Schwan, and C. Talmadge, "Non-Newtonian Gravity and New Weak Forces: an Index of Measurements and Theory," Metrologia **29**, 215-260 (1992).

[122] P. Callin and C. P. Burgess, "Deviations from Newton's Law in Supersymmetric Large Extra Dimensions," Nucl. Phys. B **752**, 60-79 (2006).

[123] J. H. Jeans, "Motions of the Stars in a Kapteyn Universe," MNRAS **82**, 122-132 (1922).

[124] J. H. Jeans, "Dynamics of Moving Clusters," MNRAS **82**, 132-139 (1922).

[125] J. C. Kapteyn, "First Attempt at a Theory of the Arrangement and Motion of the Sidereal System," Ap. J. **55**, 302-328 (1922).

[126] V. Trimble, "Looking Backward, Darkly," in: *Dark Matter*, AIP Conference Proceedings **336**, S. S. Holt and C. L. Bennett, eds. (AIP, New York, 1995) pp, 57-68. See. Sec. 3.

[127] L. Bergström, "Non-Baryonic Dark Matter: Observational Evidence and Detection Methods," Repts. Prog. Phys. **63**, 793-841 (2000).

[128] S. Perlmutter, "Studying Dark Energy with Supernovae: Now, Soon, and the Not-Too-Distant Future," Phys. Scrip. **2005-T117**, 17-28 (2005).

[129] P. J. E. Peebles, "Probing General Relativity on the Scales of Cosmology." in: , *General Relativity and Gravitation. Proceedings of GR17*, P. Florides, B. Nolan, and A. Ottewil (eds) (World Scientific, Singapore, 2005), pp. 106-117. ArXiv: astro-ph/0410284.

[130] M. Doran, G. Robbers and C. Wetterich, "Impact Of Three Years Of Data From The Wilkinson Microwave Anisotropy Probe on Cosmological Models With Dynamical Dark Energy," Phys. Rev. D **75**, 023003/1-5 (2007). ArXiv: astro-ph/0609814.

[131] E. Bertschinger and P. Zukin, "Distinguishing Modified Gravity from Dark Energy." Phys. Rev D (to be published). ArXiv:0801.2431 [astro-ph].

[132] H. Wei and S. N. Zhang, "How to Distinguish Dark Energy and Modified Gravity?." Phys. Rev. D **78**, 023011/1-7 (2008) ArXiv: 0803.3292 [astro-ph].

[133] G. Bertone, D. Hooper and J. Silk, "Particle Dark Matter: Evidence, Candidates and Constraints," Phys. Rept. **405**, 279-390 (2005). ArXiv: hep-ph/0404175.

[134] M. Milgrom, "A Modification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis," Astrophys. J. **270**, 365-370 (1983).

[135] M. Milgrom, "A Modification of the Newtonian Dynamics - Implications for Galaxies," Astrophys. J. **270**, 371-383 (1983).

[136] M. Milgrom, "A Modification of the Newtonian Dynamics - Implications for Galaxy Systems," Astrophys. J. **270**, 384-389 (1983).

[137] M. Milgrom, "MOND – A Pedagogical Review," Acta Phys. Polon. B **32**, 3613 (2001). ArXiv: astro-ph/0112069.

[138] R. B. Tully and J. R. Fisher, "A New Method of Determing Distances to Galaxies," Astron. Astrophys. **54**, 661 (1977).

[139] M. Milgrom, "The MOND Paradigm," talk presented at the XIX Rencontres de Blois "Matter and Energy in the Universe: from Nucleosynthesis to Cosmology", May 2007. ArXiv: 0801.3133 [astro-ph].

[140] J. D. Bekenstein, "Relativistic Gravitation Theory for the Modified Newtonian Dynamics Paradigm," Phys. Rev. D **70**, 083509/1-28 (2004); (E) **71**, 069901 (2005).

[141] R. H. Sanders, "A Tensor-Vector-Scalar Framework for Modified Dynamics and Cosmic Dark Matter," Mon. Not. Roy. Astron. Soc. **363**, 459-468 (2005). ArXiv: astro-ph/0502222.

[142] J. D. Bekenstein and R. H. Sanders, "A Primer to Relativistic MOND Theory," in: *Mass Profiles and Shapes of Cosmological Structures*, eds. G. Mamon, F. Combes, C. Deffayet, and B. Fort, EAS Pub. Ser. **20**, 225-230 (2006). ArXiv: astro-ph/0509519.

[143] J. D. Bekenstein, "The Modified Newtonian Dynamics – MOND and its Implications for New Physics," Cont. Phys. **47**, 387-403 (2006).

[144] P. D. Mannheim, "Alternatives to Dark Matter and Dark Energy," Prog. Part. Nucl. Phys. **56**, 340-445 (2006). ArXiv: astro-ph/0505266.

[145] P. D. Mannheim and D. Kazanas, "Newtonian Limit of Conformal Gravity and the Lack of Necessity of the Second Order Poisson Equation," Gen. Rel. Grav. **26**, 337-361 (1994).

[146] E. E. Flanagan, "Fourth Order Weyl Gravity," Phys. Rev. D **74**, 023002/1-4 (2006) ArXiv: astro-ph/0605504.

[147] P. D. Mannheim, "Schwarzschild Limit of Conformal Gravity in the Presence of Macroscopic Scalar Fields," Phys. Rev. D **75**, 124006/1-15 (2007) ArXiv: gr-qc/0703037.

[148] J. R. Brownstein and J. W. Moffat, "Gravitational Solution to the Pioneer 10/11 Anomaly," Class. Quantum Grav. **23**, 3427-3436 (2006). ArXiv: gr-qc/0511026.

[149] D. Clowe, M. Brada, A. H. Gonzalez, M. Markevitch, and S. W. Randall, "A Direct Empirical Proof of the Existence of Dark Matter." Astrophys. J. Lett. **648**, L109-L113 (2006).

[150] A. Mahdavi, H. y. Hoekstra, A. y. Babul, D. y. Balam and P. Capak, "A Dark Core in Abell 520," Ap. J. **668**, 806-814 (2007). ArXiv: 0706.3048 [astro-ph].

[151] S. Carroll, "Dark Matter vs Modified Gravity," (2006).

http://online.kitp.ucsb.edu/online/lens06/carroll/
[152] G. Angus, B. Famaey, O. Tiret, F. Combes and H. Zhao, "The Velocity Distribution of Sloan Digital Sky Survey Satellites in Modified Newtonian Dynamics," MNRAS **383**, L1-L4 (2008). ArXiv: 0709.1966 [astro-ph].
[153] A. Klypin and F. Prada, "Testing Gravity with Motion of Satellites around Galaxies: Newtonian Gravity against Modified Newtonian Dynamics," ArXiv: 0706.3554 [astro-ph].
[154] W. Pauli and M. Fierz, "Über relativistische Feldgleichungen von Teilchen mit beliebigen Spin im elektromagnetishcen Feld," Helv. Phys. Acta **12**, 297-300 (1939).
[155] M. Fierz and W. Pauli,"On Relativistic Wave Equations for Particles of Arbitrary Spin in an Electromagnetic Field," Proc. Roy. Soc. London A **173**, 211-232 (1939).
[156] W. Rarita and J. S. Schwinger, "On a Theory of Particles with Half-Integral Spin," Phys. Rev. **60**, 61 (1941).
[157] M. Fierz, "Uber die physikalishe Deutung der erweiterten Gravitätionstheorie P. Jordans," Helv. Phys. Acta **29**, 128-136 (1956).
[158] P. Jordan, "Zum Gegenwartigen Stand der Diracschen Kosmologischen Hyposthesen," Zeit. Phys. **157**, 112-121 (1959).
[159] C. Brans and R. H. Dicke, "Mach's Principle and a Relativistic Theory of Gravitation," Phys. Rev. **124**, 925-035 (1961).
[160] H. van Dam and M. Veltman, "Massive and Massless Yang-Mills and Gravitational Fields," Nucl. Phys. B **22**, 397-411 (1970).
[161] V. I. Zakharov, "Linearized Gravitation Theory and the Graviton Mass," JETP Lett. **12**, 312-314. [ZhETF Pis. Rad. **12**, 447-449 (1970).]
[162] A. I. Vainshtein, "To the Problem of Nonvanishing Graviton Mass," Phys. Lett. B **39**, 393-394 (1972).
[163] D. G. Boulware and S. Deser, "Inconsistency of Finite Range Gravitation," Phys. Lett. B **40**,227-229 (1972).
[164] D. G. Boulware and S. Deser, "Can Gravitation Have a Finite Range?" Phys. Rev. D **12**, 3368-3382 (1972).
[165] W. Siegel, "Hidden Gravity In Open-String Field Theory," Phys. Rev. D **49**, 4144-4153 (1994). ArXiv: hep-th/9312117.
[166] N. Arkani-Hamed, H. Georgi, and M. D. Schwartz, "Effective Field Theory for Massive Gravitons and Gravity in Theory Space," Annals Phys. **305**, 96-118 (2003). ArXiv: hep-th/0210184.
[167] G. R. Dvali, G. Gabadadze and M. Porrati, "4D Gravity on a Brane in 5D Minkowski Space," Phys. Lett. B **485**, 208-214 (2000). ArXiv:hep-th/0005016.
[168] C. Deffayet, G. R. Dvali, G. Gabadadze and A. I. Vainshtein, "Nonperturbative Continuity in Graviton Mass versus Perturbative Discontinuity," Phys. Rev. D **65**, 044026/1-10 (2002). arXiv:hep-th/0106001.
[169] A. Gruzinov, "On the Graviton Mass," New Astron. **10**, 311-314 (2005)
[170] G. Gabadadze and A. Iglesias, "Schwarzschild Solution in Brane Induced Gravity," Phys. Rev. D **72**, 084024/1-12 (2005). ArXiv: hep-th/0407049.
[171] G. Dvali, "Predictive Power of Strong Coupling in Theories with Large Distance Modified Gravity," New J. Phys. **8**, 326/1-9 (2006) arXiv: hep-th/0610013.
[172] G. Gabadadze and A. Gruzinov, "Graviton Mass or Cosmological Constant?" Phys. Rev. D **72**, 124007/1-9 (2005). ArXiv: hep-th/0312074.
[173] A. Nicolis and R. Rattazzi, "Classical and Quantum Consistency of the DGP Model," JHEP **2004**, 06059/1-28 (2004). ArXiv: hep-th/0404159.
[174] A. S. Goldhaber and M. M. Nieto, "Mass of the Graviton," Phys. Rev. D **9**, 1119-1121 (1974).
[175] M. G. Hare, "Mass of the Graviton," Canad. J. Phys. **51**, 431-433 (1973).
[176] D. J. Kapner, T. S. Cook, E. G. Adelberger, J. H. Gundlach, B. R. Heckel, C. D. Hoyle, and H. E. Swanson, "Tests of the Gravitational Inverse-Square Law below the Dark-Energy Length Scale," Phys. Rev. Lett. **98**, 021101/1-4 (2007).
[177] A. A. Geraci, S. J. Smullin, D. M. Weld, J. Chiaverini, and A. Kapitulnik, "Improved constraints on Non-Newtonian Forces at 10 Microns," Phys. Rev. D **78**, 022002/1-12 (2008).
[178] T. Damour and J. H. Taylor, "On the Orbital Period Change of the Binary Pulsar PSR 1913 + 16," Astrophys. J. **366**, 501-511 (1991).
[179] J. H. Taylor, A. Wolszzan, T. Damour, and J. M. Weisberg, "Experimental Constraints on Strong-Field Relativistic Gravity," Nature **355**, 132-136 (1992).
[180] L. S. Finn and P. J. Sutton, "Bounding the Mass of the Graviton Using Binary Pulsar Observations," Phys. Rev. D **65**, 044022/1-7 (2002).
[181] C. M. Will, "Bounding the Mass of the Graviton Using Gravitational-Wave Observations of Inspiralling Compact Binaries," Phys. Rev. D **57**, 2061 - 2068 (1998).
[182] S. L. Larson and W. A. Hiscock, "Using Binary Stars to Bound the Mass of the Graviton," Phys. Rev. D **61**, 104008/1-18 (2000).
[183] C. Cutler, W. A. Hiscock, and S. L. Larson, "LISA, Binary Stars, and the Mass of the Graviton," Phys. Rev. D **67**, 024015/1-5 (2003).
[184] D. I. Jones, "Bounding the Mass of the Graviton Using Eccentric Binaries," Astroph. J. **618**, L115-L118 (2005).
[185] S. M. Kopeikin, "Testing the Relativistic Effect of the Propagation of Gravity by Very Long Baseline Interferometry," Astrophys. J. **556**, L1-L5 (2001).
[186] C. M. Will, "Propagation Speed of Gravity and the Relativistic Time Delay," Astrophys. J. **590** 683-690 (2003).
[187] S. Samuel, "On the Speed of Gravity and the v/c Corrections to the Shapiro Time Delay," Phys. Rev. Lett. **90**, 231101/1-4 (2003).
[188] S. Carlip, "Model-Dependence of Shapiro Time Delay and the ?peed of Gravity/Speed of Light?Controversy," Class. Quant. Grav. **21**. 3803-3812 (2004).
[189] T. Damour, "100 Years of Relativity: Was Einstein 100% right?" AIP Conference Proc. **841**, 51-62 (2006), 28th Spanish Relativity Meeting; September 6-10, 2005; Oviedo, SPAIN
[190] S. M. Kopeikin, "The speed of Gravity in General Relativity and Theoretical Interpretation of the Jovian Deflection Experiment," Class. Quant. Grav. **21**. 3251-3286 (2004).
[191] S. M. Kopeikin, and E. B. Fomalont, . "On the Speed of Gravity and Relativistic v/c Corrections to the Shapiro Time Delay," Phys. Lett. A **355**, 163-166 (2006).
[192] E. B. Fomalont and S. M. Kopeikin, "The Measurement of the Light Deflection from Jupiter: Experimental Results," Astrophys. J. **598**, 704-711 (2003).
[193] G. Dvali, A. Gruzinov and M. Zaldarriaga, "The ac-

celerated universe and the Moon," Phys. Rev. D **68**, 024012/1-8 (2003). ArXiv: hep-ph/0212069.