

Reasonable doubt: Cosmology's gift for physics

Cynthia Kolb Whitney
Rhinecliff Street, Arlington, MA, USA

Abstract. This paper is about some of the ways in which the study of Cosmology has been key, both in initially developing, and in probing retroactively into, the foundations of modern physics. The past role of cosmological observations in triggering the development of mathematical physics, and modern science in general, is well known. But the present role of cosmological observations in possibly triggering a review of some never-challenged assumptions is playing out only presently, and so is not completely known yet. The cosmological observations cited in this paper concern the apparently universal pattern displayed by barred spiral galaxies. The typical barred spiral galaxy follows a common 'body plan' consisting of two spiral arms, acutely curved, and a central bar, not exactly straight. The arm curvature seems about the same for all such galaxies, and the bar wandering, if it is visible, is similar in initial slope to the arms. The several issues in physical theory that this so-common pattern invites us to acknowledge and investigate first arose in Electrodynamics. They then propagated into Quantum Mechanics and into Special Relativity Theory, and then into General Relativity Theory. Quantum Mechanics and Relativity Theory are the main pillars of twentieth century physics. So better clarifying the issues in Electrodynamics is important. If that elucidation comes to pass, Cosmology will have driven modern science, yet again, to a new level of understanding.

Keywords: Galaxy evolution, galaxy morphology, Galilean transformations, Gaussian pulses, general relativity, generalized functions, Hermite polynomials, Liénard-Wiechert potentials and fields, Lorentz transformations, Maxwell's equations, quantum gravity, quantum mechanics, special relativity, tensor analysis

1. Introduction

Cosmology was certainly the starting point for the development of all our modern sciences. The short story goes: curious individuals invented telescopes, looked through them, and saw things that were not the way they were expected to be on the basis of prevailing dogmas; facts were documented, major hardships were inflicted on, and endured by, courageous individuals; eventually, many people saw what was to be seen, and developed a whole new narrative about our place in the cosmos, and what is happening around us. Along with that new narrative about Cosmology came a new attitude toward all of Nature, and the methods of modern science then developed to study absolutely everything in a non-dogmatic way. We are now deep into elementary particles, and it seems unlikely that we would ever have gotten even to atomic physics if we had not had the idea of planets revolving around the Sun!

But Cosmology itself remains not fully modern. Its limitation is that cosmologists cannot do what the rest of modern science does: experiments that involve controlled conditions, repetitions, and competing hypotheses. Observations, Yes, but Experiments, No. It is an ironic situation. But even so, Cosmology is still in a position to do again what it did before for the rest of modern science. We live in a time of ever improving telescopes and other observational technology, and with every advance, we see more things we cannot yet explain. So Cosmology remains still able to upset prevailing dogmas.

*Corresponding author: C. Whitney, 141 Rhinecliff Street, Arlington, MA 02476-7331, USA. E-mail: Galilean_Electrodynamics@comcast.net.



Fig. 1. A typical barred spiral galaxy: “A Barred Spiral Galaxy ngc 1300 hubble photo”.

This paper begins with the intriguing commonality seen in the images of so many barred spiral galaxies. The typical features of such galaxies are illustrated in Fig. 1, plucked off the world-wide web. We see two spiral arms, acutely curved, and a central bar, not exactly straight. Section 2 advances a mathematical characterization of the common pattern.

The analysis of this pattern triggers investigation of several theoretical issues discussed in Section 3. They all first emerged in Electrodynamics, and then propagated into Special Relativity Theory, and then into General Relativity Theory. The first issue comes from early ideas about field propagation that were assumed true in the development of relativity theory. The second issue comes from the late 19th century angst over a seeming incompatibility between Maxwell’s equations and Newtonian physics. The third issue comes from a newly revealed incompatibility between the physical reality of a finite-energy signal pulse, or a photon, and the mathematical idealization of an infinite plane wave, which has infinite energy. Section 4 reiterates the basic conclusion: physics still has dogmas to question and root out.

2. The barred spiral galaxy body plan

There are many barred spiral galaxies, with many individual personalities; see [5]. Some are so full of stars, and are so bright, that the central bar is hard to discern. Some have spiral arms that go around so many times that the number of arms is hard to discern. Some are rather depleted of stars, and seem to have something dark and dead scattered around the outer reaches. But just as all human beings, regardless of size or shape or color or age, are similar in following a basic human-species body plan – head, trunk, four limbs, so also all barred spiral galaxies are apparently similar in following a barred spiral galaxy body plan: two spiraling arms connected by a slightly sinuous central bar. Why are these features so consistently the way they are? There must be some kind of common cause behind the common body plan.

My first suspicion about the barred spiral galaxy body plan was that the ‘two-ness’ of it is caused by a massive two-body system at the center. Two-body systems are the physicist’s favorite kind of system to analyze (most tractable), and maybe Nature’s favorite kind of system to construct (most stable). A truly super massive two-body system is plausible, given the idea of ‘black holes’. With respect to two

black holes, the ordinary stars are insignificant little satellites. They would just obey whatever field of gravitational potential the two-body system created.

My second suspicion was that the two spiral arms and the central bar of the galaxy all link together at the locations of the two massive bodies. That is, they are all part of the same phenomenology. Such simplicity is always sought by physicists, and is often provided by Nature. What common phenomenology could prevail over such huge distances? Gravity certainly seemed the most likely candidate, inasmuch as no other phenomenology can perturb it, or shield against it.

My third suspicion was that the spiral shape of the arms and the slightly sinuous shape of the bar were both the result of the finite speed of propagation that gravity must have. In Newtonian theory, the speed of gravity is infinite, but in General Relativity Theory (GRT), that speed is finite, apparently the same as light speed c . The idea of gravitational signal delay is plausible by analogy to the behavior of all systems that have signal feedback with delay, a situation played out every day in engineering disciplines.

The value c for propagation speed for gravitational signals has never been confirmed, and is indeed suspected by some researchers. The value does not really matter though; only the ratio between that speed and the orbit speed of the two-body system actually matters. The example chosen for detailed investigation was orbit speed equal to half the propagation speed. This turned out to be a good guess because, for reasons described in Section 3 and detailed numerically in [8], it allows a steady state such that the massive two-body system can work on its satellite stars for eons and eons without itself changing in any way.

From this physical model for the workings of a barred spiral galaxy, it is possible to calculate all kinds of numerical data [8–11,13]. It is possible to evaluate the gravitational potential at any spatial point in the plane of the galaxy, figure the radial and tangential forces at any point, and thereby develop a mental picture of the whole galaxy. It is possible to plot just about anything of interest as a function of angle around the galactic disc. One can look at both the radial force component and the tangential force component, with both Newtonian instantaneous signal propagation and with the proposed signal propagation delay.

One interesting result from such exercises concerns the tangential force component. In neither case is it zero. When followed around a full circle, it does average to zero in the Newtonian case; however, with the proposed delay for signal propagation, it does not even average to zero. So with delay, there is a net tangential force working on the stars. It is, on average, pumping more energy into each star's orbit. As time goes by, such an effect has to push the stars further and further outward in the galactic plane. Whatever geometry the galaxy may have started out with, it ends up being a disc.

Does this really happen? Observe that the sky is full of disk-shaped galactic systems. 'Pennies from Heaven', one might say. Observe too that the darkest, and probably oldest, stars, and the so-called dark matter, maybe dead stars, are found in the outermost regions of galaxies. So, yes, there is much to indicate that outward migration of stars in a disc galaxy really does happen.

Supposing that such star migration does indeed happen, how does it keep on happening? Where does a continuing supply of stars come from? Consider that the two-body system at the center is incredibly massive. It can draw new matter into the system along the system polar axis, and then transform that matter variously, and then drive it outward in the system plane. The new matter may come into the galactic system as diffuse plasma, plentiful in space, and may exit the two-body neighborhood as new stars that then move ever outward.

A more synoptic type of plot produced by the computational exercises shows the loci of zero crossings of the tangential forces on stars. Those would be the loci of the angular maxima and minima in the two-body gravitational potential well. The minima are more interesting: satellite stars fall 'into' the lower

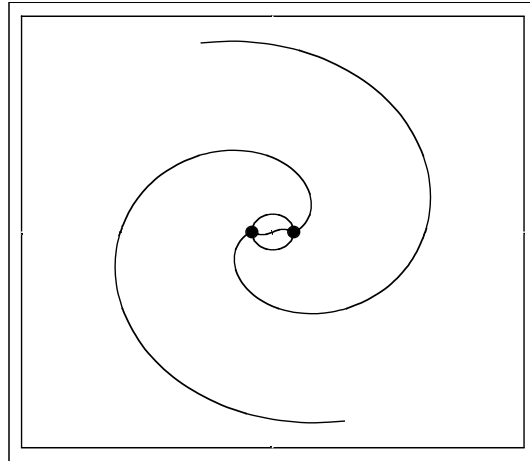


Fig. 2. The loci of angular minima in the potential around two massive bodies orbiting.

potential neighborhood, and somewhat ‘pile up’ there. So the locus of the minima predicts the higher density of satellite stars that defines the visual appearance of the galaxy.

The angular minima of gravitational potential follow the pattern shown on Fig. 2. The two black spots represent the two driving massive bodies. The circle represents the common orbit of the two massive bodies, which are orbiting in the clockwise sense. The exterior spiral arms and the interior slightly sinuous bar are the loci of angular minima in the gravitational potential from the massive two-body system.

Figure 2 clearly suggests the body plan of a barred spiral galaxy. This pattern is, of course, not static: it revolves with its driving two-body system. That revolution is much faster than the orbiting of the individual satellite stars that are driven by the two-body system. So the pattern is *not* defined by the satellite stars themselves, but rather by the two-body potential field guiding them. The pattern regularly passes through particular neighborhoods of stars, putting them into a condition of higher star density. That condition is followed by lower star density, and, on a cosmic time scale, perhaps rather quickly; remember, the barred spiral pattern is tied to the central two-body system, which, compared to the satellite stars, revolves quickly. In short, the blender blades move faster than the banana bits!

Figure 2 is the representation of a physical model that was first investigated via brute-force calculation of gravitational potentials with delay, and only much later abstracted into a mathematical formula. To express that formula, imagine superposed onto the picture polar coordinates with their origin at the center. Take the diameter of the circle as the unit of length for the radial coordinate. For each spiral arm, the radial coordinate increases exactly with the angle from its initial point, which is located at its originating massive body. The two spiral describing functions are defined for $r_1 > 0.5$ and $r_2 > 0.5$, where index 1 refers to the massive body on the right and index 2 refers to the massive body on the left. They are:

$$\theta_1 = r_1 - 0.5 \quad \text{and} \quad \theta_2 = r_2 - 0.5$$

The angles θ_1 and θ_2 are measured directly from their respective massive bodies in the counterclockwise sense, so that $\theta_2 = \theta_1 - \pi$.

The spiral arms look like each source completely dominates one arm, and the condition of maximum radial force, zero tangential force, just propagates radially out to the spiral while that one source moves

on. Near in, the dominance is real; it occurs because the other source is relatively far away. Far out, the dominance is illusory; from there, the sources are so relatively close together that they are indistinguishable. Also, from there, both of them attract very weakly.

The 'bar' of the barred spiral galaxy is defined for $0 < r_1 < 0.5$ and $0 < r_2 < 0.5$. The two parts are well approximated by the function

$$y = \frac{1}{4\pi} \sin(2\pi x)$$

which is defined for $x = -0.5$ to $x = +0.5$. Near the sources, this function provides slope magnitude $1/2$, the same as the slope magnitude of the spiral for the spiral arms outside the circular orbit circle. That is, near the sources, the interior central bar is just the reverse of the exterior spiral arms:

$$\theta_1 = 0.5 - r_1 \quad \text{and} \quad \theta_2 = 0.5 - r_2$$

Note that the formulae given here for arms and bar are really pretty simple. I was remiss in not finding them 25 years ago!

3. Electrodynamic issues implicated

We all like to think that the acquisition of knowledge is linear, and that knowledge once acquired is established for good. It isn't so. This Section discusses three examples of received wisdom where 'reasonable doubt' now looks like the prudent attitude to take. All of them originate in Electrodynamics, but relate to the barred spiral galaxy problem because our current understanding of gravity is based on our current understanding of Electrodynamics, so when we find a surprise in Cosmology, and suspect that gravity is its cause, that situation invites review of our current understanding of Electrodynamics. For reference, many aspects of that current understanding are very well covered in Jackson [3].

3.1. Beliefs about potentials and fields from moving charges

Many physicists believe that gravitational attraction cannot exhibit a significant retardation effect because gravitational attraction should be analogous to the Coulomb attraction in Electrodynamics, and everyone believes the Coulomb attraction exhibits no significant retardation effect. This belief traces to the turn of the 20th century, when Liénard [4] and Wiechert [15] developed models for the potentials and fields created by rapidly moving charges. Although Liénard and Wiechert worked independently, they made the same assumption, and they got the same results, and so confirmed each other. This Subsection looks at those results, and thereby develops a motivation to look back at their underlying assumption.

Expressed in Gaussian units, the Liénard-Wiechert scalar and vector potentials are

$$\Phi(\mathbf{r}_{\text{observer}}, t) = e[1/\kappa R]_{\text{retarded}} \quad \text{and} \quad \mathbf{A}(\mathbf{r}_{\text{observer}}, t) = e[\vec{\beta}/\kappa R]_{\text{retarded}}.$$

where \mathbf{r}, t are space, time coordinates, $\kappa = 1 - \mathbf{n} \cdot \vec{\beta}$, $\vec{\beta}$ is source velocity normalized by c , and $\mathbf{n} = \mathbf{R}/R$ (a unit vector), and $\mathbf{R} = \mathbf{r}_{\text{observer}}(t) - \mathbf{r}_{\text{source}}(t - R/c)$ (an implicit definition for the terminology 'retarded'). The Liénard-Wiechert fields obtained from those potentials are then

$$\mathbf{E}(\mathbf{x}, t) = e \left[\frac{(\mathbf{n} - \vec{\beta})(1 - \beta^2)}{\kappa^3 R^2} + \frac{\mathbf{n}}{c\kappa^3 R} \times \left((\mathbf{n} - \vec{\beta}) \times \frac{d\vec{\beta}}{dt} \right) \right]_{\text{retarded}} \quad \text{and}$$

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{n}_{\text{retarded}} \times \mathbf{E}(\mathbf{x}, t).$$

The Liénard-Wiechert fields have some interesting properties. The $1/R$ fields are radiation fields, and they make a Poynting vector that lies along $\mathbf{n}_{\text{retarded}}$:

$$\mathbf{P} = \mathbf{E}_{\text{radiative}} \times \mathbf{B}_{\text{radiative}} = \mathbf{E}_{\text{radiative}} \times (\mathbf{n}_{\text{retarded}} \times \mathbf{E}_{\text{radiative}}) = E_{\text{radiative}}^2 \mathbf{n}_{\text{retarded}}.$$

But the $1/R^2$ fields are Coulomb-Ampère fields, and the Coulomb field does *not* lie along $\mathbf{n}_{\text{retarded}}$ as one might naively expect; instead, it lies along $(\mathbf{n} - \vec{\beta})_{\text{retarded}}$, which is virtually indistinguishable from $\mathbf{n}_{\text{present}}$.

Anyone might feel moved to check this surprising result. Fortunately, anyone can look up the original sources, obtain translations if necessary, and verify the original algebra. There is no problem with the algebra. There are also numerous re-derivations that use more modern techniques involving the ‘Dirac delta function’ and the ‘Heaviside step function’. These are ‘generalized’ functions of some parameter that, when driven to infinity, produces an infinite pulse or a unit step. One can study these re-derivations too. One finds various re-orderings of the mathematical operators ‘differentiate’, ‘integrate’, and ‘go to parameter limit’. These re-orderings are dodgy because the generalized functions lack the mathematical property of uniform convergence, so these operations don’t necessarily commute; it is possible to change the result by changing operation order. But even so, such findings do not change the fact that the original derivations, though pedestrian, were correct.

Many people have just accepted that this is ‘the way things are’ with classical field theory. But what if one wanted to describe the same scenarios in a thoroughly modern way, with ‘photons’ instead of radiation fields, and ‘virtual photons’ instead of Coulomb-Ampère fields? Could anyone really accept the idea that the real photons and the virtual photons created by the same space-time event arrive at a detector from different directions?

The reader may sense that the Liénard-Wiechert situation is troubling. If a problem exists with this Liénard-Wiechert result, then there is really only one place where it can arise: in the initial assumption; namely, that electromagnetic fields propagate like bullets shot at speed c . This is the very *same* assumption that Einstein later formalized as his Second Postulate [1,2]. It is the foundation for Special Relativity Theory (SRT), which is in turn the foundation for General Relativity Theory (GRT). SRT and GRT together make one of the two great pillars of 20th century Physics: Relativity Theory (RT). Questioning the foundation threatens this pillar.

3.2. Beliefs about non-invariance of Maxwell's equations

The late 19th century was bedeviled by a seeming crisis in Physics: Maxwell’s Electromagnetic Theory (EMT) appeared at odds with Newtonian physics, as the latter was clearly form invariant under Galilean transformations between different frames of reference, whereas the former was thought not to be. Since EMT *was* demonstrably form-invariant under Lorentz transformations between frames of reference, the advent of Einstein’s SRT seemed to offer the way forward. Today, it is generally believed by physicists that Maxwell’s EMT is uniquely linked to Einstein’s SRT.

The ironic thing is that the development of SRT led to the widespread adoption of tensor analysis as a tool of Mathematical Physics. The equations of SRT, and of EMT as a part of SRT, are today recognized as tensor equations. But all tensor equations are supposed to be form-invariant under any invertible coordinate transformation. The word ‘any’ means ‘*any*’. It includes not only Lorentz transformations, but also Galilean transformations, and of course rotations, plus any change of units, or screwy deformation the user may wish to inflict. The implication is: Maxwell’s equations *cannot* be form invariant

only under Lorentz transformation. They *have to* be invariant under Galilean transformation too. So all demonstrations to the contrary have to have been defective.

But in contemporary Physics, this possibility has been neglected in favor of adopting a more restricted definition for the word 'any'. It being believed *a priori* that Maxwell's equations mandate the exclusive use of Lorentz transformations, the phrase 'any coordinate transformation' is taken to mean 'any coordinate transformation in the Lorentz group',

Faced with this disagreement between mathematicians and physicists as to the meaning of the word 'any', I have to choose to go with the mathematicians, and then try to convince the physicists. But how? The choices are: 1) to write out some equations in tensor format, which is a little terse, and so might be unconvincing, or 2) to write out the same equations in explicit 4×4 matrix format for a specific case of Galilean transformation. The latter should be plain enough to convince, but it takes a lot of space, so here I will go with the former.

One good thing about tensor analysis is that it imposes an organizational discipline. Its statements are all-inclusive, and prevent the user from focusing too much on one part of a problem, while ignoring another part of the problem. This is important when something in the ignored part of the problem cancels something strange that appeared in the selected part of the problem. This opportunity to 'make sense' by finding the cancellation of 'nonsense' was missed in the classical arguments contrived to demonstrate non-invariance of Maxwell's equations under Galilean transformation.

The disciplined approach to Maxwell's equations goes like this: Maxwell's equations in tensor notation read:

$$\partial_\alpha F^{\alpha\beta} = \frac{4\pi}{c} J^\beta \quad \text{and} \quad \partial_\alpha D^{\alpha\beta} = 0^\beta$$

All indices can have four values corresponding to four dimensions, three for space and one for time. The ∂_α is the differential operator $\partial/\partial x^\alpha$. The $F^{\alpha\beta}$ is the electromagnetic field tensor, whose entries are components of \mathbf{E} and \mathbf{B} fields, and $D^{\alpha\beta}$ is the dual to $F^{\alpha\beta}$, whose entries are components of \mathbf{B} and $-\mathbf{E}$. The J^β is the source charge-current tensor. The distinction expressed by 'index down' vs. 'index up' is expressed in words as 'covariant' vs. 'contravariant', and in numbers as opposite signs on the three space-like components. Indices repeated up and down mean 'sum the indicated products using all four possible index values'.

The transformations of the constituent parts of Maxwell's equations read:

$$\begin{aligned} \partial'_\beta &= \frac{\partial x^\alpha}{\partial x'^\beta} \partial_\alpha, & J'^\delta &= \frac{\partial x'^\delta}{\partial x^\beta} J^\beta. \\ F'^{\gamma\delta} &= \frac{\partial x'^\gamma}{\partial x^\alpha} F^{\alpha\beta} \frac{\partial x'^\delta}{\partial x^\beta}, & D'^{\gamma\delta} &= \frac{\partial x'^\gamma}{\partial x^\alpha} D^{\alpha\beta} \frac{\partial x'^\delta}{\partial x^\beta}, \end{aligned}$$

So the first tensor Maxwell equation transforms to

$$\partial'_\gamma F'^{\gamma\delta} = \left[\frac{\partial x^\epsilon}{\partial x'^\gamma} \partial_\epsilon \right] \left[\frac{\partial x'^\gamma}{\partial x^\alpha} F^{\alpha\beta} \frac{\partial x'^\delta}{\partial x^\beta} \right] = \frac{\partial x'^\delta}{\partial x^\beta} \frac{4\pi}{c} J^\beta = \frac{4\pi}{c} J'^\delta$$

Observe how the tensor contraction $\frac{\partial x^\epsilon}{\partial x'^\gamma} \frac{\partial x'^\gamma}{\partial x^\alpha}$ reduces to the Kronecker delta δ_α^ϵ , and so makes this equation form invariant.

The second tensor Maxwell equation transforms to

$$\partial'_\gamma D'^{\gamma\delta} = \left[\frac{\partial x^\epsilon}{\partial x'^\gamma} \partial_\epsilon \right] \left[\frac{\partial x'^\gamma}{\partial x^\alpha} D^{\alpha\beta} \frac{\partial x'^\delta}{\partial x^\beta} \right] = \partial_\alpha D^{\alpha\beta} \frac{\partial x'^\delta}{\partial x^\beta} = 0^\beta \frac{\partial x'^\delta}{\partial x^\beta} = 0'^\delta$$

Again the tensor contraction $\frac{\partial x^\varepsilon}{\partial x'^\gamma} \frac{\partial x'^\gamma}{\partial x^\alpha}$ reduces to $\delta_\alpha^\varepsilon$, and the equation ends up form invariant.

Nothing visible in these tensor transformation equations precludes Galilean transformation, or any other invertible transformation. If there is really a problem here, we need to create more visibility to be able to detect it.

Fortunately, tensor notation offers an opportunity not yet exploited. In tensor notation, the position of an index encodes the kind of transformation that the index requires. There are two index positions in use today: up-right, called 'contravariant', and down-right, called 'covariant'. The unexploited index positions are up-left and down-left. I believe we need to use them too. The reason why is that not just two, but actually four, transformation types can be written down.

The most familiar transformation is the contravariant one. The prefix 'contra' means these tensor merates change opposite to the way the basis unit vectors of the reference frame are changed. For an arbitrary input vector X^α , the transformation reads

$$\bar{X}^\beta = [\partial \bar{x}^\beta / \partial x^\alpha]_{\text{direct}} X^\alpha,$$

where we see the transformation as partial derivatives of coordinates, new with respect to old. We also see the specification 'direct', to further distinguish this one from the others that follow below.

The prefix 'contra' means reverse to the prefix 'co'. The covariant transformation goes the same way the basis unit vectors change. So the covariant transformation uses transformation equal to the reverse contravariant transformation. I emphasize the prefix in reverse to distinguish it from inverse. In the case of perfect space-time symmetry, *i.e.* the case of Lorentz transformation, reverse and inverse are the same, but they are not the same in general. After a century of commitment to Lorentz transformations exclusively, physics has lost the distinction between reverse and inverse.

The covariant transformation goes:

$$\bar{X}_\beta = [C_\beta^\alpha] X_\alpha = [\partial \bar{x}^\alpha / \partial x^\beta]_{\text{reverse}} X_\alpha.$$

The transformation that is actually inverse to the contravariant one has to look different from this, so we need to use the down-left index position to distinguish it. Then we can write it as

$${}_\beta \bar{X} = [{}_\beta C^\alpha]({}_\alpha X) = [\partial x^\alpha / \partial \bar{x}^\beta]_{\text{direct}}({}_\alpha X).$$

To recall the use of left index, we can call this one 'transcovariant'

The final case is the inverse to the covariant transformation:

$${}^\beta \bar{X} = [{}^\beta T_\alpha] X = [\partial x^\beta / \partial \bar{x}^\alpha]_{\text{reverse}}({}^\alpha X)$$

This one we can call 'transcontravariant'.

Using this new expanded tensor notation, we can see that the differential operator in Maxwell's equations should be identified as a transcovariant object. We should write:

$$({}_\alpha \partial) F^{\alpha\beta} = \frac{4\pi}{c} J^\beta \quad \text{and} \quad ({}_\alpha \partial) D^{\alpha\beta} = 0^\beta.$$

With this clarification of tensor notation, Maxwell's equations are manifestly invariant even under Galilean transformation [14]. They are in fact invariant under *any* transformation that leaves the measurement units unaltered. The invariance of Maxwell's equations under Lorentz transformation in particular is not a mandate to use Lorentz transformations in general, and hence accept Einstein's SRT. So if we are unhappy with the Liénard-Wiechert result discussed earlier, then Maxwell leaves us free to consider variations on its underlying assumption, and hence consider variations on Einstein's SRT.

3.3. Beliefs about EM signals and photons

It has always been presumed that a light signal, or a finite-energy pulse of light, propagates without distortion, just like an infinite plane wave would do. This presumption is based on the presumption that when we take Maxwell's four coupled field equations, which are first order in derivatives, insert one into another to produce two wave equations, which are second order in derivatives, we can be sure that the original first order equations and the derived second order equations have the same set of solutions.

But that presumption overlooks something. It is well known that light emanating from a finite aperture cannot be focused to a point; it is limited to a finite Airy spot, with rings around it. This is the well-known phenomenon of diffraction. Diffraction amounts to waveform spreading in directions transverse to the nominal propagation direction, arising from initial waveform limitation in the directions transverse to the propagation direction. The ubiquity of diffraction invites us to ask the so-far neglected complementary question: What about possible waveform spreading *along* the nominal propagation direction? That could be caused by waveform limitation in the longitudinal direction. And such limitation is absolutely necessary to define a light signal pulse, or a quantum photon, with finite total energy. Analysis of this problem is the second, and previously neglected, part of the light propagation story.

Like the barred spiral galaxy problem, this problem too can be studied with a bit of computer modeling. Let us begin a scenario with single pulses in E and B that have Gaussian profile along the propagation direction, say x , with $E, B \sim \exp(-x^2)$. We can apply Maxwell's equations, and watch what happens. The Gaussian is the so-called 'generating function' for the infinite set of Hermite polynomials, all of which have very regularly spaced zero crossings. What happens is that all the derivatives in play generate successively higher-order Hermite polynomials multiplying the Gaussian. Meanwhile, all the $\mathbf{E} \times \mathbf{B}$ Poynting vectors in play support general spreading of the Gaussian. With each step, the emergent functions look more and more like wavelets, and the individual peaks in the wavelet stay about the same width as more of them accrue, so the wavelength for the wavelet becomes more and more defined. Figure 3 illustrates this behavior. Series 1 is the original input Gaussian function, Series 2 is the Gaussian after some spreading has taken place, and Series 3 is the wavelet that has emerged in the process; *i.e.* the spread-out Gaussian times the Hermit polynomial generated.

What we have so far is only one eighth of the story needed to fully represent a photon: development from a pulse into a waveform. We have told the story for one pulse in E . If we would match that with another pulse in B , we would have, along with the waveform development, the overall propagation of energy. That would bring us to one quarter of the whole story of the photon. If we would match that with two more pulses, E and B pointing at 90° in space from the first pair and coming 'alive' a quarter cycle out of phase with the first pair, we would have the circular polarization characteristic of photons, but we would still have just half the story. So let us move on, and seek the other half.

The remaining half of the story of the photon is about waveform regression. How does this complex structure of four Hermite polynomials multiplied by their generating Gaussian unwind, and go back to being a set of four pulses, so that it can be absorbed into a receiver?

The answer to this question comes from the imposition of realistic boundary conditions. The Gaussian pulse that was used to describe the waveform development part of the scenario was unrealistic in that its tails extended to infinity. There is no way that a localized source could emit an energy pulse whose tails would extend to infinity. It is somewhat more realistic to imagine the equivalent of a mirror at the source, and another mirror at the eventual receiver, to confine the waveform like a wave in a box, with zero amplitude at the surface of each mirror and everywhere beyond.

With such boundary conditions imposed, the analytic functions involved in the model are no longer the simple Gaussian and the simple Hermite polynomials that it generates. Now we have not one, but three,

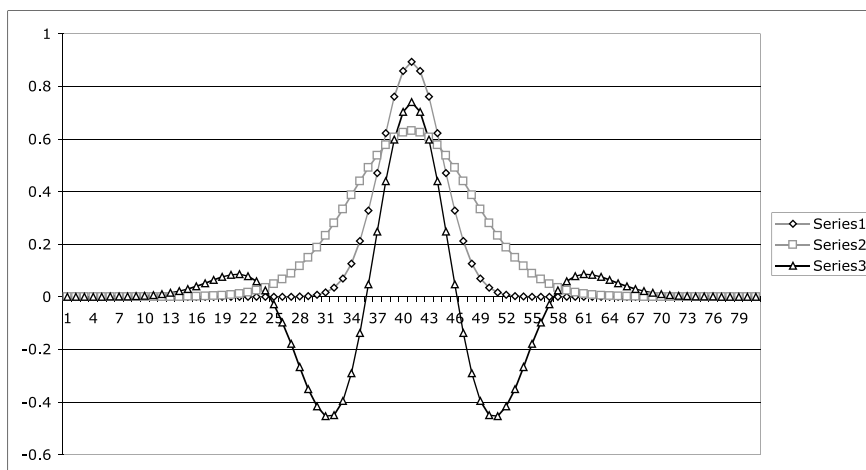


Fig. 3. A wavelet develops when a pulse is acted upon by Maxwell's equations.

Gaussians, the extra two being needed to cancel the first one at the two boundaries. Correspondingly, we always have at least three (actually six) Hermite polynomials alive at any given time. That is a loss of mathematical simplicity. But there is a gain of conceptual simplicity. It is easy to envision that the propagation scenario has some symmetry about its mid point. The waveform will spread until its central peak is halfway between the source and the receiver. After that, the mirror at the receiver will be more significant than the mirror at the source, causing the waveform to start 'piling up' near the receiver, and eventually end up as a pulse near the receiver, similar to the pulse originally launched near the source.

The idea of such a 'contracting' solution is somewhat reminiscent of 'advanced' solutions going backwards in time, which were introduced many times in the early 20th century, but particularly popularized in the mid 20th century by Wheeler and Feynman [6,7], who were looking to time symmetry as the basis for an electromagnetic generalization of instantaneous (Newtonian) gravitational interaction. There are important differences between the contracting waveforms introduced above and the Wheeler-Feynman advanced solutions: 1) Wheeler and Feynman were looking at interactions between essentially point sources and receivers, and so had to be looking at spherically expanding retarded solutions and spherically contracting advanced solutions, not at essentially one-dimensional expanding and contracting wavelets. 2) The Wheeler-Feynman expansion or contraction is related to the spherical area of a wave front, not the waveform in the radial propagation direction. 3) A lengthy discussion of the paradox of advanced actions is necessitated in the Wheeler-Feynman work, whereas the 'contracting' solutions introduced here are not in fact 'advanced' at all; they are just contracting, in real time, in the propagation direction.

With both expanding and contracting solutions in hand, it is possible to imagine a more realistic model for a light signal: it begins as E and B pulses, expands from its source into a wavelet, transitions mid-scenario to a contracting wavelet, and then collapses down to E and B pulses that can finally be absorbed by a receiver.

In this rather complex scenario, the characterization of light 'speed' is not just one simple number c . Consider that the expanding wavelet must move energy initially at speed c relative to its proximate boundary, the source, and that the contracting wavelet must move energy finally at speed c relative to its proximate boundary, the receiver. So there has to be a transition in between. The full description of this transition presently awaits further study.

Note that if the transition would be abrupt, then there would be a discontinuity in time derivatives: acceptable in Maxwell's four coupled field equations, because they are first order in derivatives, but unacceptable when we reduce Maxwell's equations to two wave equations, because those are second order in derivatives. So no such solution was considered before, even for the signal problem.

But as described here, the transition controlled by changing dominance of boundary conditions, is not even abrupt. So there is no foreseeable reason why wave equations, as well as field equations, should not be satisfied.

The presently articulated signal concept eliminates the central mystery of Einstein's SRT: one light speed with respect to however many observers there may be. With the new signal concept, one can develop a more benign form of SRT [12] that avoids all the familiar paradoxes, plus the Liénard-Wiechert problem mentioned earlier. That in turn allows solution to one big problem common to everything from galaxies to atoms: how any two-body system of any kind can ever persist over time, without 'running down' on account of radiation.

4. Conclusion

Since the Renaissance, the study of Cosmology has been the source of ever-surprising observations that can prompt us all to question our fundamental beliefs. This is very, very important. It is Cosmology's continuing gift for foundational Physics to receive, and appreciate. In particular, I believe the typical body plan of barred spiral galaxies has a big message for us all.

The first part of the message is that the commonality of underlying spiral pattern among the many barred spiral galaxies suggests that there must be a common cause. I think the common cause is a typical field of gravitational potential created by a super massive two-body system at the heart of such a galaxy. I could be wrong – maybe there is something else that can be so strikingly universal. We must all remember that having *an* explanation is not necessarily having *the* explanation. And we must all remember that, because of its restriction to observation without controlled experiment, Cosmology is even less able than the rest of Physics to endorse any particular candidate explanation for anything. So no one need endorse the first part of the message before proceeding to the second part:

The second part of the message from Cosmology is that we need to review our current understanding of Electrodynamics – especially everything fundamental that came *after* Maxwell. Review is important because Electrodynamics precedes and sets the stage for the two major pillars of twentieth century theoretical physics: Quantum Mechanics (QM) and Relativity Theory (RT: SRT and GRT). GRT is somewhat limited; for example, a simple two-body problem is beyond its comfort zone. And SRT has some structural issues to worry about; *e.g.*, the three problems discussed in Section 3. Furthermore, SRT and QM are somewhat incompatible: SRT is traditionally a rather reductionist theory, featuring point particles, communicating only by light-signal 'bullets', whereas QM is a rather holistic theory, featuring spatially extended systems, with instantaneous statistical correlations. And finally, QM just has to *postulate* the stability of atoms; it thus denies the possibility of asking questions about such subjects.

Despite these problems, QM and RT are our pillars, and we love them, and we are now trying to put a capstone over them: Quantum Gravity (QG). But the placement of a capstone may be premature, and may not succeed. And even if it does succeed, the resulting structure will surely resemble something built at Stonehenge – not a proper Gothic cathedral, much less anything modern!

If we want to develop a robust theory for 21st century physics, we will have to review the foundations of 20th century physics much more than we have done so far. This is where the 'reasonable doubt' that is Cosmology's continuing gift for foundational physics will yet again be truly indispensable.

Acknowledgments

My son Karl carried out the early brute-force calculations on the barred spiral galaxy model. Our thanks go to Dr. Gerd Herzinger for enabling us to start this work at the Bavarian DFVLR at Oberpfaffenhofen in 1983. Thanks go also to Dr. Ruggiero Santilli, who published all of that early work in *Hadronic Journal*, and to Dr. Anthony Peratt, who gave it subsequent further exposure. I thank my long-time friend Trudy Eyges for reading the current account from the viewpoint of a cultured non-specialist. I thank Dr. Rudy Schild for reading it with a view to making it better.

References

- [1] A. Einstein, On the electrodynamics of moving bodies, *Translated in The Collected Papers of Albert Einstein* **2** (1905), Princeton University Press, Princeton NJ (1989), pp. 140–171.
- [2] A. Einstein, On the relativity principle and the conclusions drawn from it, *Translated in The Collected Papers of Albert Einstein* **2** (1907), Princeton University Press, Princeton NJ (1989), pp. 252–311.
- [3] J.D. Jackson, Classical electrodynamics, *Second Edition John Wiley and Sons*, New York, US (1975).
- [4] A. Liénard, Champ électrique et magnétique produit par une charge électrique concentrée en un point et animée d'un mouvement quelconque, *L'Eclairage Electrique* **XVI**(5–14) (53–59) (1898), 106–112.
- [5] A. Sandage, The hubble atlas of galaxies, *Carnegie Institution of Washington, Publication* (1961), 618.
- [6] J.A. Wheeler and R.P. Feynman, Interaction with the absorber as the mechanism of radiation, *Revs Mod Phys* **17**(2–3) (1945), 157–181.
- [7] J.A. Wheeler and R.P. Feynman, Classical electrodynamics in terms of direct interparticle action, *Revs Mod Phys* **21**(3) (1949), 425–433.
- [8] C.K. Whitney, Steady state in the gravitational Two-Body problem, *Hadronic Journal* **11** (1988), 153–156.
- [9] C.K. Whitney, Barred spiral field pattern around a relativistic two-body system, *Hadronic Journal* **11** (1988), 157–162.
- [10] C.K. Whitney, Distribution of stars as test particles in a two-body background field, *Hadronic Journal* **11** (1988), 163–167.
- [11] C.K. Whitney, How can spirals persist? in: *Plasma Astrophysics and Cosmology*, A.L. Peratt, eds, Kluwer Academic Publishers, Dordrecht, Boston, London, (1995), pp. 175–186.
- [12] C.K. Whitney, Editor's essays, Galilean electrodynamics 16, *Special Issues* 3, whole issue, (2005).
- [13] C.K. Whitney, On the visual images that galaxies create, *Proceedings of the Natural Philosophy Alliance* 4 (2007), 290–295.
- [14] C.K. Whitney, Is our current relativity theory uniquely mandated? *Presented at The Ninth International Workshop Applied Category Theory Graph-Operad-Logic*, (2011), 14–19 March; view at <http://math.utsa.edu>.
- [15] E. Wiechert, Elektrodynamische elementargesetze, *Archives Néerlandaises des Sciences Exactes et Naturelles série II, Tome IV* (1901).