The Need For Relativity
Version A: Full Math

W. Blaine Dowler

April 21, 2012

Contents

1 The Success of 19th Century Physics

By the end of the 19th Century, physicists were standing proud. It seemed as though every unanswered question in physics was on the verge of being answered, and that physics would be the first field of science to be described in totality. The list of unanswered questions was short, and the answers seemed to be a matter of applying basic mathematics to several ideas and compiling the results. We shall now explore some of those specific questions in detail.

2 Mercury's Orbit

Most of us imagine planetary orbits as circles, or even ellipses, which wrap around a star. The planet then travels along this fixed path, over and over, repeating its exact positions again and again.

Most of us imagine this incorrectly. As it turns out, these orbits precess, meaning the orbit itself rotates around a star. In other words, if you were to take a piece of paper and plot the orbit over time, instead of going over the same line over and over again, it would shift ever so slightly, and produce more of a "floral" pattern as the planet's orbit itself revolves around the Sun. Newtonian mechanics predict such a shift. In the case of Mercury, however, the shifting and precession predicted by theory were not enough to explain the degree of shift found by observation. The orbit spins more rapidly than anyone had predicted.

3 Speed of Earth and the Ether

In the physics of the 19th century, it was thought that light was a wave with no particle properties at all, and that it couldn't travel without some sort of material or medium to travel through. As it was clear that light reached Earth from the Sun, and as the finite extent of the atmosphere had been established, that meant there was something out there in space that wasn't air between the Earth and the Sun. Thus was born an idea in physics that made so much sense at the time that some people still won't let it go today: the ether.

The ether was a material thought to be of exceptionally low density that spread throughout the heavens, and it was this ether that light traveled through. As it seemed to have no measurable impact on orbital mechanics, it was also suggested that all of the ether moved or flowed at a steady speed, and that planets sailed through it like a gentle ocean current. This begged an obvious question: how fast is the current?

The speed of the current could be measured with a relatively simple apparatus. Light waves have peaks and valleys, technically called crests and troughs, just as water waves do. As with water, when crest meets crest, they combine to form a higher crest. Similarly, two troughs combine to form a deeper trough. This is called constructive interference. When two waves are “in phase,” it means their crests and troughs always combine in this manner. When the crest of one wave combines with the trough from another, they cancel each other out through a process known as destructive interference, and are said to be “out of phase.”

This principle was applied to measure the speed of the Earth relative to the ether. Imagine two swimmers in a lake, starting at a floating platform. Now imagine that they are in a race, and have to swim from the floating platform to a landmark in the lake and then back to the platform. Further imagine that each swimmer has his or her own landmark to swim to, and that the landmarks are the same distance from the platform, but that they are in different directions, separated by a right angle. If the swimmers swim at the same speed, they would be expected to swim to the landmarks and back in the same amount of time. If the swimmers were waves, they’d arrive “in phase.”

Now take this same setup with our swimmers and put them in a river. Even traveling at the same speed, if one swimmer swims with and against the current, and the other swims at a right angle to the current, they would not be expected to return to the floating platform (which also drifts with the current) at the same time. This is because the flow of the current would ensure that the swimmers were moving at different speeds. If the swimmers were waves, they’d arrive “out of phase.”

Albert Michelson and Edward Morley decided to use this principle to measure the speed of Earth relative to the ether. They set up an apparatus using one spinning mirror and two stationary mirrors to split a single beam of light into two beams that traveled at right angles to each other before reforming. They could then measure how “out of phase” the light beams were upon their return and recombination, and determine the speed of the apparatus with respect to the ether along a particular direction at that time. By repeating the experiment several times in the course of a year, the Earth’s motion relative to the Sun could be measured and eliminated, and the motion of our solar system relative to the ether could be quantified.

The problem they had was this; no matter what time of year they did the experiment or what size apparatus they used, the light returned in phase every time without exception. For some reason, the speed of light wasn’t changing, with or without an ether. It appeared that the entire theory of the ether was fundamentally flawed and that there was no ether. That’s not the conclusion most scientists reached, however. The majority of the thinking was not in looking for an alternative to ether, but rather in looking for a flaw in the Michelson-

Morley experimental design, despite the number of independent confirmations that were coming in from those who reproduced the experiment. Others looked at creating a more complicated picture of the ether which would have the same result.

3.1 The Ether Today

Every reproducible experiment conducted over the past 125 years indicates that the ether does not exist. Despite this complete lack of supporting evidence, there are still those who cling to the theory. Some seem to do so for the novelty, as it has almost become a game to discern how complicated ether must be in order to remain consistent with all of these experiments. Others do so for other reasons. For example, ether is a fundamental piece of most theories which attempt to include existence of ghosts and other paranormal activities in modern science.

In short, people still talk about theories involving ether despite a complete lack of supporting evidence. If you encounter a theory involving ether, take it not with a grain of salt, but with an industrial size salt lick.

4 Compatibility of Newton and Maxwell

In the late nineteenth century, the world of physics' most legendary figure was Isaac Newton. Newton's laws of motion had been formulated in the 17th century, and stood unchallenged over 200 years later, having ushered in a new era in predictive, quantitative science. They agreed with experiments to an amazing degree.

In the latter half of the 19th century, the theories of electricity and magnetism started to mature and formal structures started to appear. Through a combination of deriving some relationships and compiling the work of others, James Clerk Maxwell assembled the so-called "Maxwell's Equations," which described electricity and magnetism effectively.

Now, Maxwell's equations had been amazingly successful. Not only did they accurately describe the behaviours of anything involving both electric charges and magnetic fields, they also allowed Maxwell to make the first (reliable) prediction of the speed of light. Light is a combination of electric and magnetic fields that feed off of and propagate each other. Using only some physical constants related to these waves, Maxwell calculated the speed of light.[1] Before

[1] To be pedantic, as any good mathematician or theoretical physicist must be, Maxwell actually calculated the speeds of electric and magnetic fields independently and then proved

that time, the speed of light was a quantity known only by experimental measurements, and had a considerable margin for error. Maxwell's calculation came very close to the now accepted value for the speed of light of $c = 299792458\,\frac{\mathrm{m}}{\mathrm{s}}$. Not only was it the first theoretical value for the speed of light, it was a value high enough to explain why the experimentalists had a hard time computing the value correctly. In these early days, the work of Newton and Maxwell appeared to join together like a marriage made in heaven.

In spite of these triumphs, there was a serious problem. To illustrate this problem, we must first discuss reference frames.

4.1 Reference Frames

In science, all experiments require a well defined reference frame.[2] It is, basically, a set of standard positions and directions that are used to measure things against.

For example, imagine you are doing the first physics experiment in a high school physics course. This usually involves dropping something from a given height, and measuring how much time it takes to fall.[3] The distance fallen is typically measured relative to the ground, and the analysis of the accelerations and speeds involved are all relative to the laboratory bench and floor.

This typical and effective setup completely ignores the fact that the Earth rotates during the day about its axis, and the entire planet is revolving around the Sun. The entire solar system also orbits around the centre of the galaxy, and our galaxy may be in motion with respect to some other as yet undetermined body. These motions can be safely ignored in the high school laboratory.[4] We have chosen that laboratory as our frame of reference; we measure with respect to a point in the lab, and can continue unrestrained from that perspective. This

---

they were not only identical, but that the two calculations were closely related. It was this consistency in both result and procedure that helped cement his findings in the minds of his peers.

[2]The term "reference frame" is interchangeable with "frame of reference." The latter is somewhat old fashioned, and is not seen as often these days.

[3]Physics experiments at the high school level can be painfully boring. I can assure you, they do get better. Eventually. We need to start students with the most basic experiments to make sure they have their experimentation skills down well enough to move on to the exciting, interesting experiments, which may or may not involve violent explosions and supersonic projectiles.

[4]The Earth's rotation does have an impact measurable by sufficiently advanced equipment, but we'll leave that discussion for later. If it bothers you now to ignore this piece, rest assured that the Earth's motion could be completely ignored if it were linear instead of rotating, and that high school equipment budgets make it extremely difficult to design high school experiments capable of detecting the difference.

is a valid assumption, provided that all frames of reference are equivalent from a scientific standpoint.

Note that we use the word "equivalent" instead of "equal" here, and with good reason. If two reference frames are equal, the quantities measured in one will be an exact match to quantities measured in the other. If they are equivalent, then observers in the two frames would be able to accurately predict future events, even though they measure different numbers for the quantities involved. For example, an observer in the high school laboratory might say that an object has zero speed, velocity, and energy of motion before it is dropped, while an observer on the Moon (which is moving relative to the Earth's surface) would disagree. Both observers would be able to accurately predict what would happen when the object is dropped, so the frames are equivalent, but not equal. Specifically, they both predict that the falling object would take the most direct path possible when moving towards a collision with the laboratory floor.

In Newton's world, these frames are equivalent. When Maxwell's equations are combined with Newton's theories, we find that two different frames of reference are not equivalent.

4.2 How Charged Objects Interact

Experiments had confirmed a number of behaviours of electrically charged objects. Maxwell's equations formalized these behaviours, but that's all they did; interactions between magnetic fields and electrically charged objects were jammed into the theory because they were observed in the lab, not because they were predicted by existing theory.

When charged objects are stationary, the basic phenomena are explained in a very simple fashion: unlike charges attract, while like charges repel. If you arrange a number of electrically charged marbles on an insulated surface and prevent them from moving, Maxwell's equations will accurately predict the attractive and repulsive forces they all experience.

When charges move, things get complicated. Moving charges produce magnetic fields[5] which can also attract or repel. These behaviours depend on the directions the charged objects are moving as well the charges themselves. Specifically:

- Like charges moving in the same direction produce attractive magnetic

[5] Why do moving charges produce magnetic fields? In Maxwell's time: "because they do." This was discovered in the lab, and subsequently jammed into the theory because it needed to be there, and not because any theorist could predict or explain the phenomenon.

fields.

- Like charges moving in opposite directions produce repulsive magnetic fields.
- Unlike charges moving in opposite directions produce attractive magnetic fields.
- Unlike charges moving in the same direction produce repulsive magnetic fields.

The faster the charges move, the stronger the magnetic fields they produce, and the more pronounced the forces become. The same is not true of the electric forces: they are entirely independent of the relative motions of the particles.

4.3 Newton to Maxwell: "It's Not Me, It's You"

Imagine two wires, running parallel to each other for as far as the eye can see. In fact, imagine them to be running farther than the eye can see; imagine they are infinitely long. Now, imagine that they carry identical electrical charges, and that these charges are spread uniformly across the wires. No one point on either wire is distinguishable from any other point on the same wire.

An observer standing next to the stationary wires would predict a repulsive force, and the two wires would move apart. That is the result of the laws of physics as measured in his or her reference frame.

Now imagine a second observer flying along parallel to the wires, exactly half way between them. From this observer's reference frame,[6] the wires are in motion, and moving in the same direction, so the wires experience *two* forces. In addition to the repulsive electrical force, there is also an *attractive* magnetic force. The electric force is independent of the relative motions of the wires. The magnetic force, however, increases with speed. As our flying friend accelerates, the attractive magnetic force gets stronger and stronger, until it eventually overpowers the repulsive force and the wires approach each other instead.

This is a significant discrepancy. Do the wires move apart, as predicted by our standing observer, or do they collide, as predicted by our flying observer? These predictions are *not* equivalent, and so neither are the two frames of reference.

Newton's theories were those that not only governed motion, but which also described how to work in moving frames of reference. His theories had stood up

[6] An observer's reference frame is one centered on that observer in which the observer is always treated as stationary.

to every conceivable experimental test for more than 250 years. As a physicist, he was revered by those working in the fields of research, and seemed more than human through the lens of history.[7] Maxwell, on the other hand, was very human, and his equations were relatively recent discoveries, having been collected and studied for less than 20 years. Thus, the overwhelming majority of physicists turned to Maxwell, looking for his mistake, and thereby fixing the problem with incompatibility.

Now, physicists are human, and when a theory that stands as long as Newton's is shown to be incompatible with something new and exciting, there are two natural reactions. The first, and by far most common, is to assume that the older Newtonian theories are correct and the new ones are wrong. However, in the world of faceless science, there are young upstarts excited about the prospect of becoming famous by rewriting the "old and busted" theory to force it to conform to the "new hotness" theory. One such upstart was Hendrik Antoon Lorentz, who, at the tender age of 51, decided to determine exactly how to rewrite Newton to make his work compatible with Maxwell's. His work resulted in an absolute absurdity: to reconcile the two theories, time had to depend on the speed of the observer! To most, this constituted further proof that the problem lay in Maxwell's work.

Unfortunately, no problem could be found. Many scientists started connecting the problem to theology, deciding that somewhere out there was a reference frame that worked. In other words, they believed that detailed explorations and experimentations would reveal situations that were not consistent phenomenologically with a laboratory based experimental reference frame. By determining which reference frame *was* consistent with such experiments, they would be able to determine God's preferred frame of reference. They intended to then search the heavens for such a body, in hopes of discovering the location of Heaven. The early results of this search were inconsistent and fruitless.

5 Enter Einstein

The world of professional physicists scoured the experimental and theoretical worlds looking for solutions to these problems. It took a self-taught physicist employed as a patent clerk to see something nobody else had seen in the work. Most physicists were excited because "Maxwell had calculated the speed of light using only some *physical constants!*" It was Einstein who got excited that "Maxwell had calculated the speed of light using *only* some physical constants!" Einstein realized that certain expected information was absent from

[7] The lens of history is about the only lens Newton didn't completely describe up to that point, which was likely a contributing factor.

the calculation. The full implications of this realization and the birth of the special theory of relativity are the subject of the next lesson.

The remainder of this lesson adds mathematical details relevant to the above discussion. Familiarity with vector calculus is assumed.

6 Mercury's Orbital Precession

The full calculations of Mercury's precession will be omitted. The basic argument is this: Mercury rotates on its axis as it orbits around the Sun, as do the other planets in our solar system. Thus, it carries angular momentum. This angular momentum experiences a torque as the gravity of the other planets in the solar system and the Sun play off of each other. This causes all planets, including Mercury, to shift their orbits. The orbits rotate around the Sun in the direction of the Sun's rotation, and the magnitude of this orbital rotation ("precession") depends on the gravitational interactions of all local bodies and the eccentricity of the precessing orbit. Quantitatively, Mercury's orbit was precessing about 41" more than it should have each century.[8]

7 Calculating the Speed of Light

In Maxwell's day, the general form of the wave equation was well known. If a wave described by function f propagated with speed v, then it satisfied the wave equation

$$\nabla^2 f = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}$$

where

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}$$

is the usual vector differential operator.[9]

[8] For those unfamiliar with this angular measure, $60''$ ("sixty seconds of arc") is equal to $1'$ ("one minute of arc"), and in turn, $60'$ is equal to $1°$. In other words, Mercury's orbit was wrong by $\frac{41}{3600}^{\circ}$ per *century* and the astronomers could measure accurately enough to know that was a problem. Observational astronomers have a history of being very, very good at their jobs.

[9] Granted, in Maxwell's day, vectors were not common mathematical tools in physics. Instead, each vector component had its own equation, which is why some of the variables chosen for electromagnetic quantities are so odd. Maxwell denoted the three components of the magnetic field as A, B and C, and the electric field was D, E and F, and so forth. When the usefulness of vectors was recognized, it was noticed that, by complete and total coincidence,

In general, Maxwell's equations appear as follows:

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= \frac{1}{\epsilon_0}\rho \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{B} &= \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}
\end{aligned}$$

where ρ is electric charge density, $\mathbf{J}$ is the electric current vector, ϵ_0 is the electric permittivity of vacuum and μ_0 is the magnetic permeability of free space. If one is working in a vacuum with no excess charge or current, then $\rho = 0$ and $\mathbf{J} = \mathbf{0}$, so that the equations reduce to

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= 0 \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{B} &= \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}
\end{aligned}$$

Maxwell noticed that the two fields can be decoupled, or separated into equations with only $\mathbf{E}$ or $\mathbf{B}$ and not both, through use of the identity

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

for any possible $\mathbf{A}$. For example, starting with

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

and "taking the curl" by calculating the cross product with ∇ positioned on the left gives

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t} \right)$$

Working with the left hand side first, recognizing that $\nabla \cdot \mathbf{E} = 0$, we get

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E}$$

the y component of the electric field was E, the y component of the magnetization (not the magnetic field) was M and the y component of the polarization was P. To make the transition to vectors as easy as possible for those comfortable with the old notation, it was decided that the symbol for the y component of every vector would become the symbol for the entire vector, since that allowed for three intuitive labels, which was typically three more than any other system that was compatible with both notations.

Meanwhile, noting that ∇ and $\frac{\partial}{\partial t}$ are operators that commute with each other, we find that the right hand side becomes

$$\begin{aligned} \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right) &= -\frac{\partial}{\partial t}\left(\nabla \times \mathbf{B}\right) \\ &= -\frac{\partial}{\partial t}\left(\mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right) \\ &= -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \end{aligned}$$

Thus,

$$-\nabla^2 \mathbf{E} = -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

or

$$\nabla^2 \mathbf{E} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

This fits the general form of the wave equation perfectly, provided

$$\frac{1}{v^2} = \mu_0 \epsilon_0$$

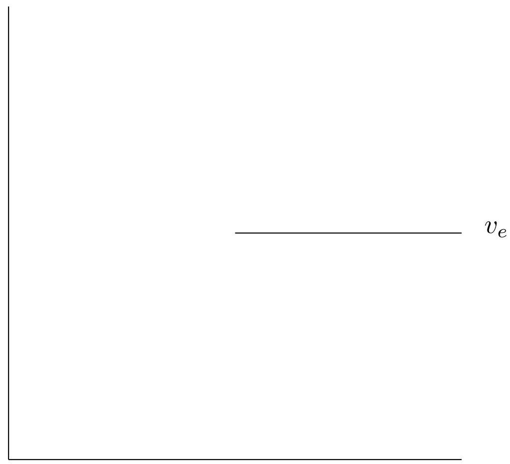
or

$$v = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

Therefore, it appeared that the electric field would propagate through vacuum with a speed $v = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$. A similar calculation can be performed for $\mathbf{B}$ to show that field also propagates with this speed. Maxwell performed this calculation, and found that this was a perfect fit for the experimentally determined speed of light! This was how physicists first realized the nature of light was that of an electromagnetic wave.

8 The Michelson-Morley Experiment

The Michelson-Morley experiment comes down to the difference in transit time for the light going through two paths. We start by imagining our apparatus is moving with one arm parallel to the motion of the ether, and the other arm perpendicular to the motion of the ether:



The light that travels parallel to the ether will move at two different speeds. With the speed of light c and the speed of the ether v_e, it will move with total speed $c + v_e$ when it moves with the current of the ether and with total speed $c - v_e$ when it moves against the current. Thus, the transit time there and back for the light moving parallel to the ether (given an arm length d in the apparatus) is

$$t_{\parallel} = \frac{d}{c + v_e} + \frac{d}{c - v_e} = \frac{2dc}{c^2 - v_e^2}$$

The light that moves perpendicular to the ether, on the other hand, takes the same amount of time to travel each trip, and is not affected by the speed of the ether. As a result, the time for that light to make the round trip is

$$t_{\perp} = \frac{2d}{c}$$

The difference in times can be calculated as well:

$$t_{\parallel} - t_{\perp} = \Delta t = \frac{2dc}{c^2 - v_e^2} - \frac{2d}{c} \tag{1}$$

With sufficient algebraic manipulation, we find that

$$v_e = \pm c\sqrt{1 - \frac{2d}{c\Delta t + 2d}} \tag{2}$$

The difficult part is pinning down Δt. This is an incredibly small time difference to measure. Thankfully, none of this motion is expected to alter the frequency of the light in question. Thus, the period of the light is the same. If the period of the light is 1.50×10^{-15}s, as it would be for the shortest wavelength of light the average human eye can detect, then counting the difference in periods can be extremely difficult. If the period of the light is T, then we can break the time interval Δt into the number of periods that have elapsed n with the simple formula

$$n = \frac{\Delta t}{T}$$

Unfortunately, since Δt is being measured by a wave interference pattern, there is a difficulty. The interference pattern is not unique for a particular n. In terms of the mixed numbers we all saw in our early exposure to fractions, the whole number portions are ignored. Thus,

$$n = 1\frac{1}{3} = 2\frac{1}{3} = 3\frac{1}{3} = 4\frac{1}{3} = 5\frac{1}{3} = 6\frac{1}{3}$$

are all indistinguishable cases. This is why the experiment takes so long to do; measurements must be taken year round as Earth's annual orbit around the Sun change its direction of motion with respect to the ether, so that sufficient data can be collected to determine precisely which value of n is the correct one to use.[10] In the end, the difficulties were somewhat moot, as every attempt to collect data from the experiment was consistent with $n = 0$.

9 Force Between Two Wires

Imagine the two wires spoken of before, parallel and separated by distance d. When they are stationary and carrying electrical charge, each with charge density λ, the electric force per unit length they experience is given by

$$\frac{F_E}{l} = \frac{\lambda^2}{2\pi\epsilon_0 d}$$

which is a repulsive force. When they are stationary, there is no magnetic field. From the perspective of a moving observer, however, they do produce magnetic fields, and the magnetic force per unit length they experience is the attractive force

$$\frac{F_B}{l} = \frac{\mu_0 \lambda^2 v^2}{2\pi d}$$

The issue comes in because of the difference in signs. The force between the two is completely balanced when

$$\frac{\lambda^2}{2\pi\epsilon_0 d} = \frac{\mu_0 \lambda^2 v^2}{2\pi d}$$

[10] The exact calculations Michelson and Morley did differed from this in details, but this illustrates the concept effectively and is a bit easier to grasp when one's skills are rusty.

or

$$v^2 = \frac{1}{\epsilon_0 \mu_0}$$

So, any observer traveling at $v < \sqrt{\frac{1}{\epsilon_0 \mu_0}}$ would observe a repulsive force, while any observer traveling at $v > \sqrt{\frac{1}{\epsilon_0 \mu_0}}$ would observe an attractive force. Those traveling at $v = \sqrt{\frac{1}{\epsilon_0 \mu_0}}$ would observe no net force at all.