The Heaviside's Experiment

Erik Margan
Experimental Particle Physics Department, Jožef Stefan Institute, Ljubljana, Slovenia

Abstract

An easy to repeat macroscopic experiment, leading to the invention of a method of fast short electromagnetic pulse generation, performed by Oliver Heaviside in 1887, leads also to a better understanding of the apparent discrepancy between the Coulomb ('static') and Maxwell ('dynamic') interaction between the electromagnetic fields and charged matter particles. It also leads to a better understanding of the internal field structure of elementary particles, as well as the electromagnetic properties of quantum vacuum fluctuations, and the cosmological problem of the vacuum energy density.

The Controversy

Every electronics technician knows that electromagnetic waves propagate at a finite velocity, as determined in the XIX century by *Helmholtz*, *Hertz*, *Faraday*, and *Maxwell*. Yet many electric phenomena are still being explained in terms of a *Coulomb*'s 'static' field. To illustrate the problem, let us consider two charged particles, each having a charge, q_1 and q_2, and each charge generating a field, E_1 and E_2. Conventionally their interaction is described in terms of a net force resulting from the action of fields on the charges. Let us concentrate on the action of the field E_1 on the charge q_2. The force vector F_2 is expressed as:

$$\vec{F}_2 = q_2 \vec{E}_1 \tag{1}$$

The strength of the field E_1 is proportional to the charge q_1 as the field source, and is inversely proportional to the distance squared (for a spherical field geometry) from q_1:

$$E_1 \propto \frac{q_1}{r^2} \tag{2}$$

In order to obtain an exact equality, we must account for a certain proportionality constant, which turns out to be equal to $1/4\pi\varepsilon_0$. Here the factor 4π stems from the spherical geometry of the field, and ε_0 is the dielectric constant or permittivity of free space (vacuum). Thus in the direction $\widehat{r}$ the field strength is:

$$\vec{E}_1 = \frac{1}{4\pi\varepsilon_0} \frac{q_1}{r^2} \widehat{r} \tag{3}$$

so the force in (1) is:

$$\vec{F}_2 = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{r^2} \widehat{r} \tag{4}$$

Equation (4) bears striking similarity to *Newton*'s gravitational force:

$$\vec{F} = G \frac{m_1 m_2}{r^2} \widehat{r} \tag{5}$$

A remarkable property of equation (4), and also (5), is the nonexistence of any propagator (time-dependent) terms. This has led some scientists to erroneously think that Coulomb's interaction (and Newton's gravitation) must be instantaneous[1].

In more modern XX century terms, we have on one hand the *Einstein*'s photon as a carrier of the dynamic electromagnetic field, whilst on the other hand we have the *Thompson*'s electron as the matter–based electric charge carrier acting as the source of the surrounding 'electrostatic' field, and any variation of this field is being interpreted as a consequence of electron kinematics. This mind boggling duality of one and the same phenomenon has been often the cause of hot debates, even before the discovery of the electron, and it persists to these days (the first decades of the XXI century).

By the development of quantum mechanics (QM), and later of quantum electrodynamics (QED), the attempt to 'solve' this discrepancy (and several others!) forced the introduction of 'virtual photons' (*Dirac*, *Feynman*) mediating the interactions between particles. In this way the 'static' field was explained in terms of QED, thus bridging the gap between the two views.

Nevertheless, many people regard such an explanation as unsatisfactory, and most of them quote certain experiments performed by *Tesla*, especially those leading to the construction of the tower at the Wardenclyffe plant on Long Island in 1904, as well as his own writings on the subject of wireless energy transmission.

To answer this question, let us return to the two charges we were dealing with above. We have the means to remove one of those charges in a very short time. If this charge is in form of an elementary particle, it can be removed by firing at it a complementary charged particle, so that the two annihilate into a pair of photons. If the charge is accumulated on the surface of a metal sphere, containing inside a battery and a relay switch, the switch can be configured so that it disconnects the battery and shorts the internal circuit. In both cases the charge is removed in time much shorter than required for the field propagation between the charges. Assume the two charges being one meter apart; we know that the propagation speed of electromagnetic phenomena is the same as the speed of light, about 3×10^8 m/s. This means that any field disturbance at one end will be felt by the other some 3.3 ns later.

The question we ask ourselves is: **when the field collapses, does the other charge sense it at the same instant, or 3.3 ns later?**

It is our intention to show that there has never been any static/dynamic 'duality' in the first place. Instead, all 'static' phenomena can be described more consistently by dynamic phenomena only. Nature always behaves consistently, of course; the problem is with our interpretation of natural phenomena, and with our limited ability to recognize serious flaws in our 'obvious' explanations.

In addition to offering a proper explanation, we are going to open a wider perspective on the problem of charge at the elementary particle level.

[1]To be completely fair, there is the *Lorentz*'s force: $\vec{F}_{q_2} = q_2\left(\vec{E} + \vec{v} \times \vec{B}\right)$, but here the velocity vector $\vec{v}$ is attributed to the movement of q_2 through the field $\vec{B}$, not to the propagation of the field $\vec{E}$.

The History

The first indication that Coulomb's and Maxwell's equations do not tell the whole story became apparent already in the second half of the XIX century with the development of the telegraph.

Since *Volta*'s discovery, electricity has been thought of as a 'fluid', 'flowing' as a 'current' (*Weber*) from the 'source' 'through' a conductor to the load and back 'through' another conductor. This water pipe analogy became quickly the prevailing model of electric energy transfer. To these days, children at primary school and university students alike are still educated by being encouraged to imagine and keep in mind this simple model. I call it sarcastically the "Plumber's Electricity" (please, don't get me wrong, I have nothing against plumbers, in fact I respect them deeply, especially when there is a leak in my bathroom; no, the sarcasm is all aimed at us physicists and electronics engineers!).

However, in the 1870s, when the telegraph developers were trying to analyse the behaviour of very long transmission lines under various pulse propagation modes, that simple picture faced serious challenges.

One of the first to realize the inadequacy of modeling electricity by a water pipe analogy was *Oliver Heaviside*. His long debates with various opponents, *Sprague* in particular, are well known. Instead of concentrating on the charge current **inside the wires**, Heaviside takes the field **between the wires** as the primary cause of electrical phenomena, referring to it as the 'energy current':

> *"Now in Maxwell's theory there is the potential energy of the displacement produced in the dielectric part by the electric force, and there is the kinetic or magnetic energy of the magnetic induction due to the magnetic force in all parts of the field, including the conducting parts. They are supposed to be set by the current in the wire.* ***We reverse this;*** *the current in the wire is set up by the energy transmitted through the medium around it ..."*

The importance of the phrase "We reverse this;" can not be overstated! The requirement for this reversal stems from the fact that the waves in free space and around the wires travel considerably faster than the waves inside conductors, and many orders of magnitude faster than the average speed of electrons. It is therefore reasonable to assume that the faster phenomena must be the cause, and the slower ones the effect.

Most people take for granted that energy (light, heat) from the Sun comes to the Earth in form of radiation through free space (seeing is believing!), but at the same time they dismiss as nonsense the idea of the electric energy coming into our homes 'between' the wires. In their opinion any field between the wires must be only a consequence of what is going on inside each wire. Not just ordinary people, but also most top scientists and engineers adhere to this view. Once a fundamentally wrong idea becomes familiar and 'obvious', it is very hard to get rid of it in favour of a more correct, but less obvious one.

In order to help us to become accustomed to this 'new' (130 years old!) idea, we are going to describe a simple experiment, easily performed by anybody who has access to an ordinary oscilloscope (preferably a digital memory instrument, to ease the view of fast but rarely repeated events), and who knows how to connect to it a simple network consisting of:

- a 9 V battery;
- a single pole, double throw switch;
- three pieces of type RG-58 coaxial cable (having a 50 Ω characteristic impedance), each a couple of meters long;
- and two resistors, one of 50 Ω, equal to the characteristic impedance of the cable, serving as a termination resistor at the end of the cable, and the other of a much higher value, say 50 kΩ, to allow charging of one cable section by the battery via a switch.

Preparations for the Experiment

In Heaviside's days there were no ready made pulse generators, nor oscilloscopes, and neither shops with shelves full of sophisticated electronics components. To generate well defined, short, and fast rising pulses for testing the telegraph lines one had to build his own equipment by himself. All that was at hand were chemical batteries, switches, wires, and a variety of insulating material. And, given the high propagation velocity of electrical signals, kilometers of wire were needed to produce pulses long enough to be recorded by such a primitive device as the magnetograph. Today we have oscilloscopes capable of capturing nanosecond pulses, so we are going to need only a few meters of cable.

Before we connect the components to form the required network, we need to examine the properties of coaxial cables (such cables were invented and patented by Heaviside in 1880), in particular their **propagation time delay** as a function of their **characteristic impedance** and their **length**. An important performance aspect of such cables is that all of the field between the inner conductor ('core') and the external conductor ('shield') is kept inside all along the length. **Fig.1** shows a typical example.

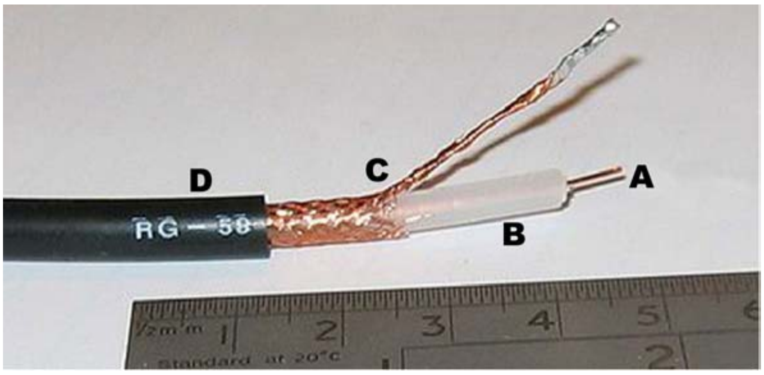

Fig.1: Coaxial cable (conductors sharing a common geometry axis): A – inner conductor, B – polyethylene insulator, C – outer conductor, D – outer jacket.

The characteristic impedance of a coaxial cable is governed by the geometry of its cross section and the permittivity of the dielectric material insulating the inner conductor from the outer shield. Following the definition of the geometry and the material constants of a coaxial cable as in **Fig.2**, it is possible to determine the cable's specific inductance and capacitance. We label the radius of the inner conductor as a, and the radius to the inner surface of the outer conductor as b. We also label the unit length along the cable as l.

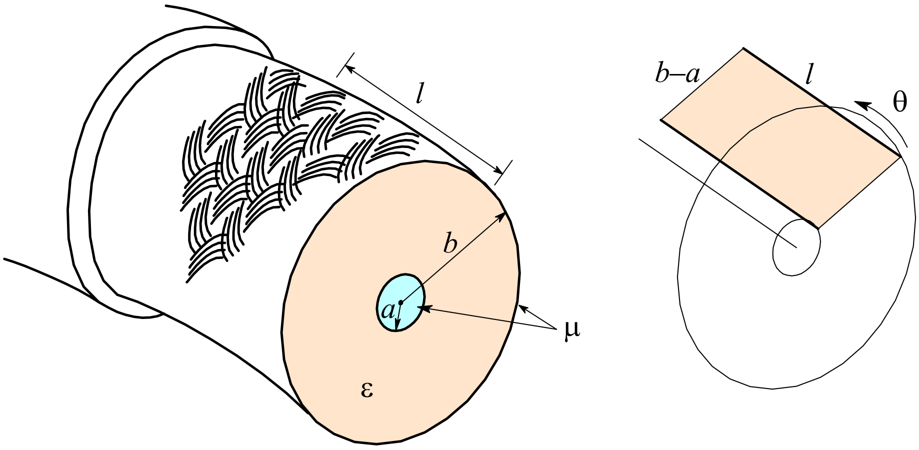

Fig.2: Geometry of field integration and material constants of a coaxial cable.

The magnetic permeability of the wire material is:

$$\mu = \mu_0 \mu_r \tag{6}$$

where μ_0 is the permeability of free space (exact by definition):

$$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{Vs/Am} \tag{7}$$

and μ_r is the relative permeability of the wire material. For copper, Cu, the permeability is $\mu_{rCu} = 0.999994 \approx 1$. Following the 1983 agreement on the units of length and time in SI, the speed of light in vacuum is exactly:

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 299\,792\,458\ \mathrm{m/s} \tag{8}$$

Consequently the permittivity (the dielectric constant) of free space is:

$$\varepsilon_0 = \frac{1}{c^2 \mu_0} \approx 8.8542 \times 10^{-12}\ \mathrm{As/Vm} \tag{9}$$

In most textbooks the value of permittivity "just happens to be such", and is accepted without any further consideration; in **Appendix 3** we show that this value has much deeper roots. The dielectric constant of the insulating material is:

$$\varepsilon = \varepsilon_0 \varepsilon_r \tag{10}$$

where ε_r is the relative dielectric constant of the material. The insulator in RG-58 coaxial cables is typically made of polyethylene, for which the relative dielectric constant is $\varepsilon_r \approx 2.25$.

As in (9), the propagation speed of the electromagnetic wavefront in a cable is:

$$v = \frac{1}{\sqrt{\mu_0 \mu_r \varepsilon_0 \varepsilon_r}} = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \cdot \frac{1}{\sqrt{\mu_r \varepsilon_r}} = \frac{c}{\sqrt{\mu_r \varepsilon_r}} \approx \frac{2}{3} c \approx 2 \times 10^8\ \mathrm{m/s} \tag{11}$$

From a known v we find the propagation time delay per unit length:

$$\tau = \frac{l}{v} \approx \frac{1\,[\mathrm{m}]}{2{\times}10^8\,[\mathrm{m/s}]} = 5{\times}10^{-9}\,\mathrm{s/m} = 5\,\mathrm{ns/m} \tag{12}$$

If we now make a cross-section of the cable along the unit length l, and we take the area defined by the radius difference of both conductors, $(b-a)\cdot l$, and integrate it for the angle θ from 0 to 2π radians, perpendicular to the cable axis, as indicated in **Fig.2**, we can express the cable capacitance per unit length:

$$\frac{C}{l} = \frac{2\pi\varepsilon}{\ln(b/a)} \tag{13}$$

As in (11), the propagation speed of the electromagnetic wave within a circuit is determined by a capacitance C and inductance L, and can be expressed as:

$$v = \frac{1}{\sqrt{LC}} \tag{14}$$

so we can use (13) and (14) to determine the inductance per unit length:

$$\frac{L}{l} = \frac{\mu}{2\pi}\ln\left(\frac{b}{a}\right) \tag{15}$$

Now that the inductance and the capacitance of a cable segment are known, we can draw the lumped circuit model of that segment, **Fig.3**. The complete cable is modeled as a distributed series of such segments. The core has a small serial resistance R and a serial inductance L, whilst its capacitance to the shield is represented by C. The insulation between the core and the shield is represented by a small conductance G.

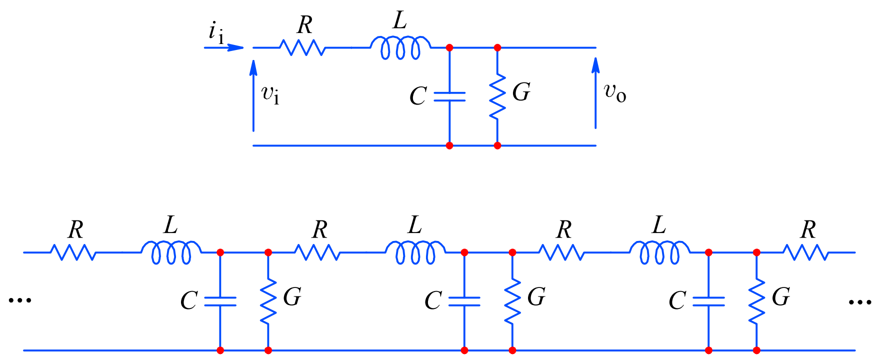

Fig.3: Above: Electrical lumped circuit model of a small segment of a coaxial cable. The core has a small serial resistance R and a serial inductance L. Its capacitance to the shield is represented by C and the conductance of the insulating material between the core and the shield is represented by G. **Below:** a long cable can be modeled by an infinite series of such small segments.

The input impedance the circuit in **Fig.3** as seen by a signal source can be found as a ratio of the input voltage v_i to the input current i_i:

$$Z_\mathrm{i} = \frac{v_\mathrm{i}}{i_\mathrm{i}} = R + j\omega L + \frac{1}{G + j\omega C} \tag{16}$$

From this we can express the characteristic impedance of the cable (see **Appendix 2** for a complete derivation):

$$Z_0 = \lim_{\delta \to 0} (\delta Z_{\mathrm{i}}) = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{17}$$

The core resistance R is very small, because the specific resistance of copper is $\rho = 1.68 \times 10^{-8}\,\Omega\mathrm{m}$ (at a temperature of 20°C). *Ohm*'s law states that the resistance is proportional to the specific resistance of the material, ρ, and to the wire length, l, and inversely proportional to the wire cross-section area, A, thus $R = \rho l / A$. In a typical coaxial cable the core diameter is $d = 2a = 1\,\mathrm{mm}$, so the cross-section is $A = \pi d^2/4 \approx 0.785 \times 10^{-6}\,\mathrm{m}^2$. With $l = 1\,\mathrm{m}$ the core resistance is about $0.021\,\Omega$.

Similarly, the specific resistance of polyethylene is very high, usually above $10^{16}\,\Omega\mathrm{m}$ for low voltages (in our case 9V); because conductance is the inverse of resistance, G will be very low.

Because $R \ll j\omega L$ and $G \ll j\omega C$ for frequencies $f = \omega/2\pi$ in the MHz range and above, we can assume a lossless circuit, with $R \to 0$ and $G \to 0$. This simplifies the characteristic impedance to:

$$Z_0 = \sqrt{\frac{L}{C}} \tag{18}$$

Because $j\omega$ cancels, Z_0 is frequency independent and purely real, so we may think of it as a resistor R_0, however this is true only for transient phenomena, as long as the wave propagates along the cable; once the wave propagation ends, the impedance changes back to Z_{in}. For example, if we measure the resistance between the core and the shield of an unterminated cable by an Ohm-meter (a very slow measuring device) we will see an essentially open circuit (actually, a very precise high resistance Ohm-meter may show $1/G$), whilst a capacitance meter would show C.

Why do we need to know the characteristic impedance for our experiment?

For the RG-58 type coaxial cable the geometry and the insulation material are chosen so that $Z_0 = 50\,\Omega$. The capacitance per unit length is $C/l = 100\,\mathrm{pF/m}$. Then, by using (18) the inductance must be $L/l = 250\,\mathrm{nH/m}$. Both C and L are reactive elements, they can temporarily store energy and return it back. The thermal losses can be neglected (for cables a couple of meters long and for low voltages). Thus any energy applied to the cable will be transferred without significant loss to the load at the end of the cable. However, if the load impedance is different from Z_0, some energy will be reflected back towards the source. **Only if the termination resistance $R_{\mathrm{T}} = Z_0$ will all the energy be absorbed by R_{T}** and converted to heat.

This is important because we want to see clean individual signal transitions, undisturbed by reflections, so that the signal recorded is very simple to interpret.

Let us verify the above findings by a simple measurement. We shall verify the propagation velocity (11) and the time delay (12) using a setup in **Fig.4**.

Here we need one cable of precise length, chosen so that the oscilloscope we use will be able to display and clearly discriminate the time delay between the input and the output pulse. Most oscilloscopes on the market have the fastest time base setting of 5 ns/division (or 50 ns across the full screen width). Our calculation shows the cable delay should be 5 ns/m. Thus 1 m of cable should be sufficient to display a one screen division delay between the pulses. But let us play it safe and take a 2 m long cable for a 10 ns delay; from now on we shall refer to it as the **section A**.

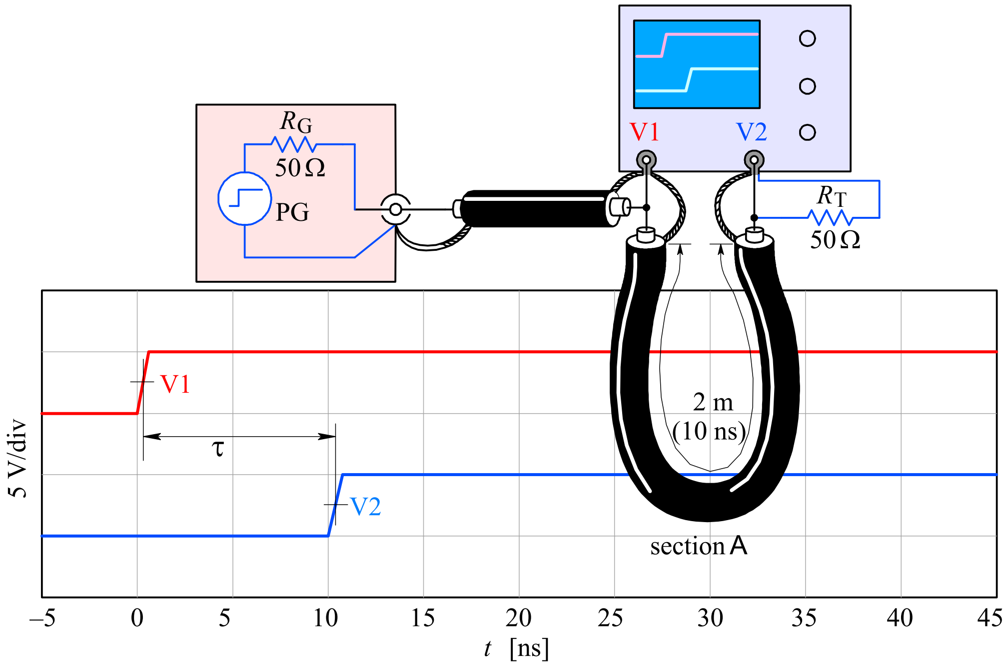

Fig.4: Measurement of the cable time delay. The pulse generator (PG) with an internal impedance of $R_G = 50\,\Omega$ is connected by a coaxial cable (of unimportant length) to the V1 oscilloscope input and to the 2 m long section **A**. Its other end is connected to the V2 input and a terminating resistor $R_T = 50\,\Omega$. R_T dissipates all the incoming energy, preventing signal reflections. Section **A** delays the wavefront by 10 ns. The propagation velocity of the EM signal in the cable must be 0.2 m/ns (~2/3 of the speed of light in vacuum).

We could now use the battery and the switch to generate the input pulse, but it is more convenient to use a pulse generator, as in **Fig.4**. With a square wave frequency set to 1 kHz and the amplitude to 5 V, the signal and its repetition rate will be high enough for finding quickly the proper trigger setup for a suitable display.

The 2 m long section **A** delays the input pulse by 10 ns, which gives a specific delay value of 5 ns/m. From this, the propagation velocity is 0.2 m/ns, or about 2/3 of the speed of light in vacuum, in accordance with the calculated value (11).

To measure the signal delay we should preferably have a step or a pulse signal with a clean sharp edge; signal distortion such as wavefront rounding, ringing, or reflections, would make the interpretation of the measurement needlessly difficult. Signal distortion is avoided by preserving the network impedance continuity. The connections between cables and the oscilloscope must be as short as practical. In professional RF work, special BNC connectors and T-pieces are used for interconnection continuity. The proper resistive cable termination converts all the incoming energy into heat, preventing reflections. However, because neither the waveform generator not the oscilloscope can have infinite bandwidth, the pulse edges will not be infinitely steep, instead the oscilloscope will display some transition slope of the order of 1–2 ns. But this is not a problem, since the pulse delay is by definition measured between the two half amplitude points.

The Experiment

Now that we know the delay as a function of cable length we can arrange the experimental setup by connecting the required components and cables as in **Fig. 5**.

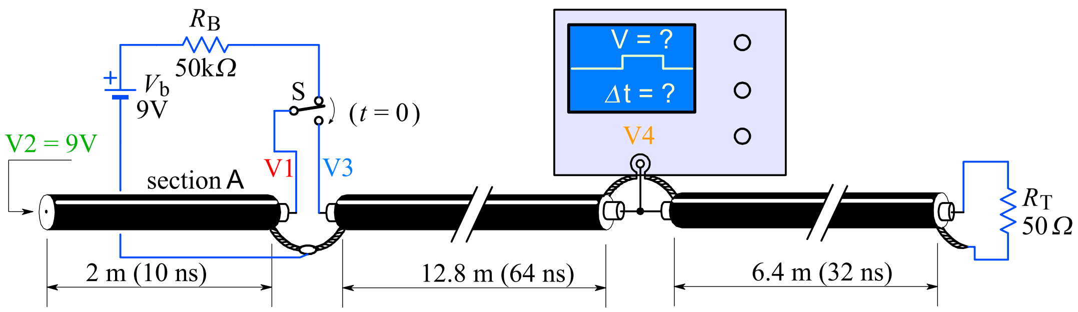

Fig.5: The 2 m long section **A** is connected via a switch and a 50 kΩ resistor R_B to a 9 V battery, the other switch pole is connected to the two serially connected cables. Their junction is connected to the oscilloscope input (V4), and the far end is terminated by a 50 Ω resistor. When we flip the switch, the energy present within the section **A** rushes out through the newly available path, reaching the terminating resistor, where it dissipates into heat, so there are no reflections. What is the amplitude and the duration of the pulse recorded by the oscilloscope?

In addition to the already used 2 m long section **A**, we shall need two other cables. Their lengths are not important, they only need to have a delay much longer than the delay of section **A**, so that the generated waveform is clearly distinguished by avoiding any possible mutual interaction. Form a bunch of standard laboratory values lets us take one cable 12.8 m long, which would delay the signal by 64 ns, and another 6.4 m long for a further 32 ns delay.

Initially one side (V1) of section **A** is connected by the switch via a 50 kΩ resistor (R_B) to the battery [+] pole, the [−] pole goes to the shield, whilst the other end of the cable (V2) is left open. The 64 ns cable is connected between the other pole of the switch and the oscilloscope input V4. The 32 ns cable connects the V4 point to the termination resistor $R_T = 50\,\Omega$. With the switch in the position shown in **Fig.5**, the cable's capacitance is slowly charged to the voltage of the battery, 9 V, within a time equal to about $5RC$ (~50 µs).

What do we expect to see on the oscilloscope screen after flipping the switch?

By naive reasoning, as the battery voltage is 9 V, the pulse amplitude should also be 9 V. We measured the delay of section **A** to be 10 ns; since all the energy leaves section **A** to be dissipated by R_T, we expect to see a 10 ns long pulse.

As this pulse travels by the oscilloscope input V4, it is recorded and displayed on the screen. The oscilloscope input has a very high resistance and low capacitance (1 MΩ, 12 pF), so the pulse edges suffer only minor aberrations. Likewise, the impedance discontinuity at the switch will not spoil the signal if the connections are kept short.

Now we flip the switch. As the switch travels between the two contacts, the section **A** is effectively open at both ends; the energy that was accumulated from the battery is preserved by the cable capacitance (~200 pF) until the new contact is made. Then the energy rushes out via the newly available path, eventually reaching R_T, where it dissipates into heat, so there are no reflections to worry about.

What are the pulse **amplitude** and the **duration** displayed by the oscilloscope?

Obviously, our expectations were completely **wrong!** What happens in reality is displayed on the oscillogram in **Fig.6**, notably on trace V4.

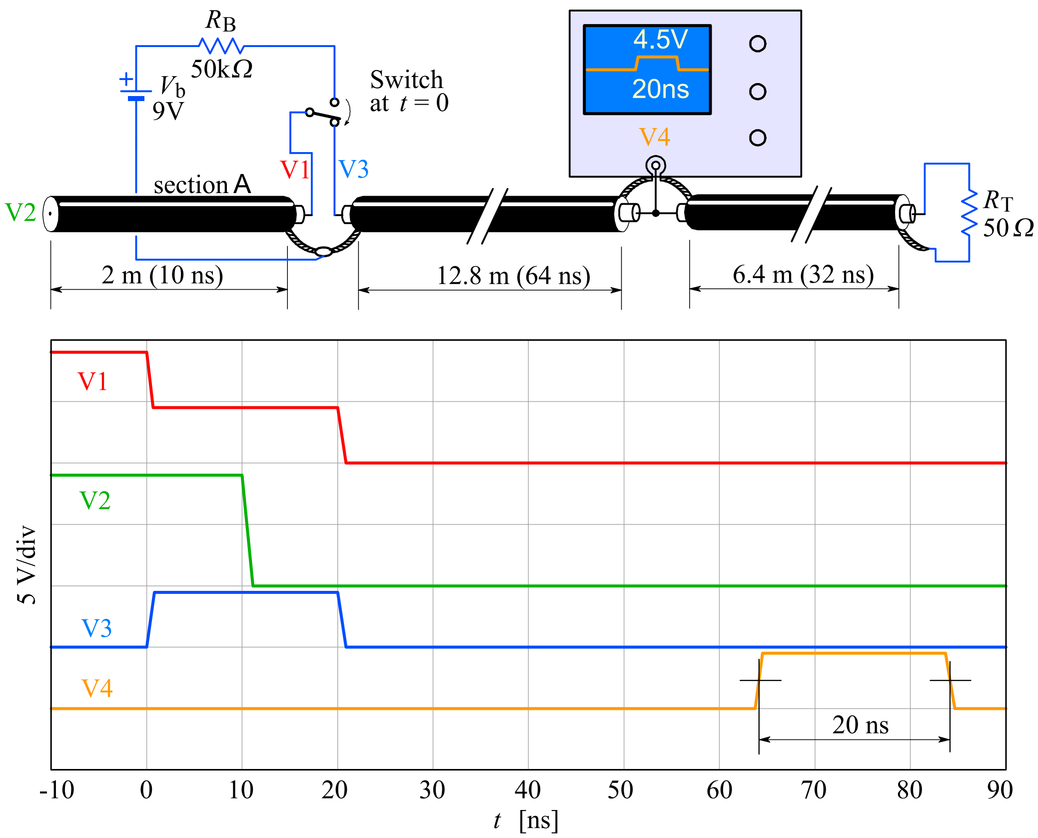

Fig.6: By flipping the switch at $t = 0$ the voltage V1 falls to $1/2$ of its original value and remains such for 20 ns (which is 2× the delay of the section **A**), then it falls to zero. But if we look at the voltage at the free end of section **A** (V2), we see that it remains at its full amplitude value (9 V) for 10 ns only, and then falls to zero. The generated pulse (V3) is of double length (20 ns) and of half the amplitude (4.5 V), and such is also the pulse recorded by the oscilloscope (V4) 64 ns later. The oscillogram proves that initially only half the energy runs out, followed by the other half after one time delay of the first cable section. Why does this happen?

Surprise, surprise: the **amplitude** of the recorded pulse (V4) **is only one half** of the original voltage value, 4.5 V instead of 9 V, but the **pulse length has doubled**, 20 ns instead of 10 ns.

How can we explain this?

NOTE - Switch problems: Switch contacts can bounce! Accounting fror the lever mass and the spring stiffness, the bouncing time is usually in the ~30 μs to ~3 ms range, so mostly there are no problems in handling <100 ns pulses. Occasionally though, oscillations may appear. Simply repeat the procedure.

Explaining the Results

By investigating all the available junctions of the network (V1 to V4 in **Fig.6**) we have no other option but to conclude that **initially only one half of the energy runs out** of section **A** for 10 ns, and is immediately **followed by the other half** in the next 10 ns.

This is where the water pipe analogy breaks down completely!

Just think of it: the battery initially transferred a number of free electrons from the core to the shield, leaving vacancies in the core; but, after the switch was flipped, why would half of those electrons decide to stay in place, patiently waiting for the other half to leave, and then follow them obediently?

This is really odd: if electrons were the cause of EM phenomena, one would expect all of them to behave in the same manner, otherwise the theory does not make sense at all!

OK, so we admit that the field between the conductors must have a role somehow. But exactly how does it behave?

By examining the oscillogram in **Fig.6** more closely we note that the voltage at the free end (V2) of section **A** remains at the full 9 V value for 10 ns, exactly the time delay of that cable section, and then falls to zero. Therefore **if half of the energy runs away from node V2, the other half must be running towards it, in order to keep the voltage unchanged**.

But why? Here we ask again the same question as for the electrons: why does not all the field behave in the same manner?

The conventional explanation is that the initial section of the cable represents a 50 Ω source, driving a 50 Ω load, therefore the voltage falls by half at the closure of the switch; this voltage change then propagates in both directions, so that the half propagating towards the free end is reflected there to follow the first half propagating already towards the load.

Really so? The EM pulse does not know what kind of a load it will encounter 96 ns later at the end of the cable, and it does not care until it reaches the load. The EM pulse 'feels' just the local μ and ε at the point where it momentarily is.

To prove this let us **remove the termination resistor** at the end and repeat the experiment. The result is exactly the same, only now we have yet another free end at which the pulse will reflect back to be recorded by the oscilloscope 32+32 ns (twice the delay of the last cable section) after the first pulse; we can see it by changing the oscilloscope time base to 20 ns/division. And it does not stop there: the pulse will continue to travel, bouncing back and forth at both open ends until the energy is eventually exhausted by the resistive loses in the circuit.

Alternatively, we could **replace R_T by a short** and obtain the same initial result, same double pulse length and same half amplitude, but now the reflection at the short would be inverted, and a **negative**, −4.5 V, 20 ns long pulse will be recorded 32+32 ns after the first one. Further reflections will now have alternating signs.

OK, so the termination does not matter until the pulse reaches it. But then, the cable itself has the same 50 Ω characteristic impedance; what if the half voltage drop at the closure of the switch is a consequence of the impedance divider formed by the two cable pieces at the switch. Well, is it really so?

No! If it were so, we would have the same situation at the junction of the two cables at the oscilloscope input, and we do not see any odd behavior there. In fact, at this point we have an impedance continuity (apart from the negligible influence of the oscilloscope input itself), thus no reflections there. And at the switch we also create a path of uniform impedance (if we keep the switch connections short), therefore this can not be the cause of the initial one half voltage drop.

Now how about this:

Photons cannot stand still. Photons carrying the **E** field inside section **A** travel at speed $v = 1/\sqrt{LC}$, but are being continuously reflected at both open ends. Before we flipped the switch, at any instant, we had two energy fields, each 10 ns long, traveling in opposite directions. At $t = 0$ the switch closed and a new path of equal impedance became available, so the energy half which was traveling forward **continued to travel forward** by the new path through the 64 ns cable and beyond. Likewise, the energy half which was traveling backward **continued to travel backwards because it was already traveling backwards** at the moment of switch closure, so no reason for it to reverse its direction; only upon reaching the open end (V2) it experienced a reflection, and therefore followed the first half back to V1 and out through the newly available path. See **reference [21]**.

This may sound strange, but is simple and logical, and there is no need to invoke any special conditions! But in order to confirm this hypothesis there are a few relations which need to be explained.

The first important relation is the so called *Poynting*’s vector (named after *John Henry Pointing*, 1852–1914, but Heaviside and *Nikolai Umov* both discovered the same relation independently):

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \tag{19}$$

Here **E** and **H** are the vectors of the electric and magnetic field strength, respectively. Their vector product, **S**, has a physical meaning, it is interpreted as the instantaneous power density flux through a unit area, measured in $\mathrm{W/m^2}$.

It is important to understand a peculiar property of the vector product. The ‘right hand rule’ of **Fig.7** applies here.

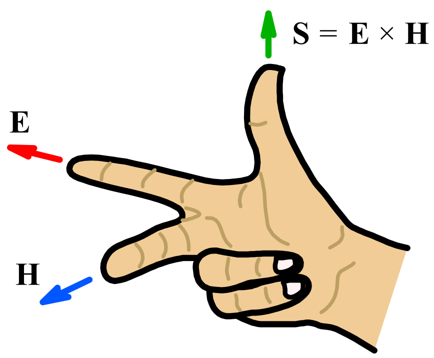

Fig.7: The right hand rule for vector products.

The result of the vector product points in a direction perpendicular to the plane in which the original **E** and **H** vectors reside, which is why both **E** and **H** point in the transversal direction in respect to the direction of energy propagation. Also, if either **E** or **H** reverses direction, the vector product **S** also reverses (to view this rotate the hand about either the **H** axis or the **E** axis); however, if both **E** and **H** reverse their directions, the vector product **S** does not reverse (the product of two negative values is a positive value; to view this rotate the hand about the **S** axis).

With this in mind we can show the linear **superposition** of our two energy halves traveling in opposite directions at the open end of the cable. Open circuit reflection occurs about the axis of the electric field, so it retains its direction, and consequently doubles, whilst the magnetic component is reversed and cancels (open circuit means zero current, thus also zero magnetic field). The Poynting's vector **S** cancels too, which means no net energy flow, resulting in an apparently 'static' condition, as shown in **Fig.8**:

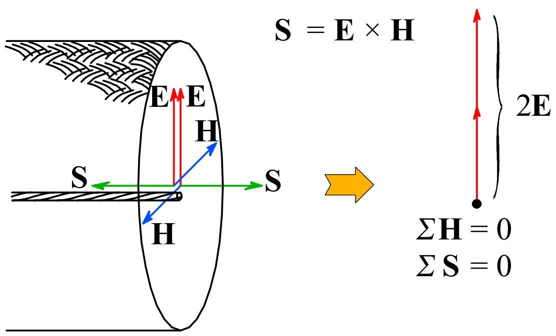

Fig.8: Vector sum of the outward and inward reflected components at the open end of the coaxial cable: the electric component **E** doubles, whilst the magnetic component **H** cancels, and so does the Poynting's vector **S**. The result is an apparently static electric field.

In case of a short circuit the opposite happens: because of the short, the resulting **E** field must be zero, and the short circuit current i causes the magnetic field **H**. The reflection occurs about the magnetic field axis, therefore the electric field cancels and the magnetic field doubles. Since the Poynting's vector **S** is again canceled, we have an apparently 'static' magnetic field, as shown in **Fig.9**:

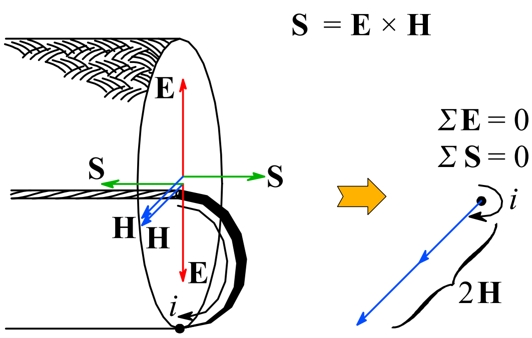

Fig.9: Vector sum of the outward and inward reflected components at a shorted end of a coaxial cable. The magnetic component doubles, caused by a loop current i, whilst the electric component cancels, and so is the Poynting's vector. The result is an apparently static magnetic field.

Actually, during the transition of the switch when section **A** is disconnected from the circuit, **Fig.7** is valid at any cross-section point of the cable, because the energy flow is continuous.

It is instructive to see what happens when the switch is reconnected to the battery. The cable capacitance is then refilled with energy. But instead of 50 kΩ, let us use a 330 Ω resistor for R_B, to have an impedance comparable to the 50 Ω of the cable, but still mismatched. **Fig.9** shows the voltages at both cable ends, V1 and V2.

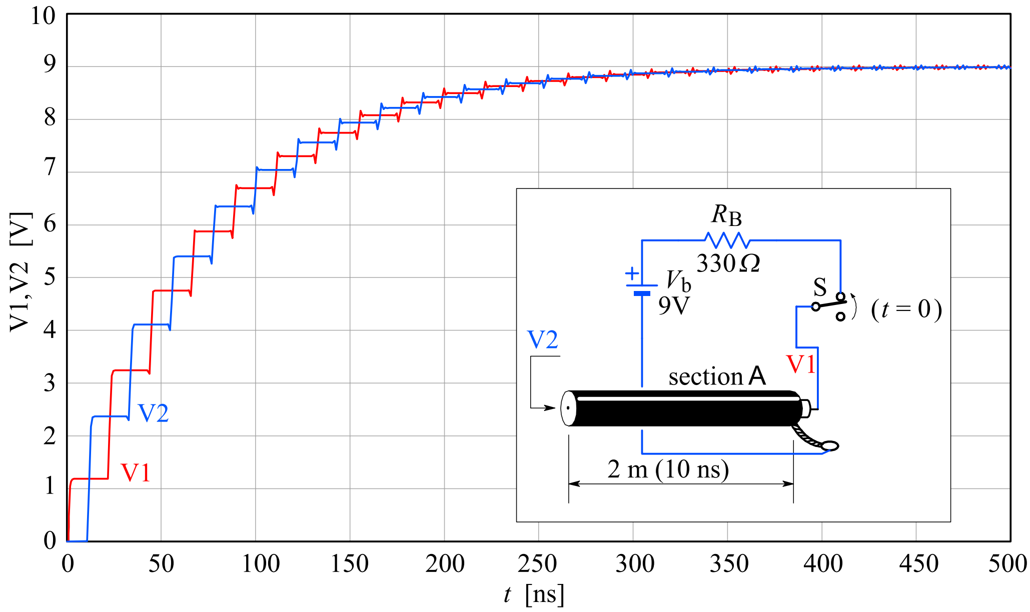

Fig.9: Refilling the cable with energy from the battery through a 330 Ω resistor. The mismatch of impedance Z_0/R_B and the time delay cause a stairstep like voltage build up. The small aberrations at each transition are the consequence of the oscilloscope input capacitance (~12 pF) loading of the line.

As the switch reconnects the cable core to the battery, the voltage V1 jumps to a value determined by the impedance mismatch, the Z_0/R_B ratio, namely:

$$\Delta V1_{(t=0)} = V_b \frac{Z_0}{Z_0 + R_B} = 9 \cdot \frac{50}{50 + 330} \approx 1.184\,\mathrm{V}$$

V1 remains at 1.184 V as the line is filled up. A 1.184 V wavefront propagates towards V2, where it arrives 10 ns later. V2 being an open end, the wavefront must reflect back, doubling the V2 voltage, and the cable now being filled backwards by another 1.184 V on top of the existing 1.184 V, maintaining the 2.368 V level at V2.

After another 10 ns the wavefront returns to V1. Because of the impedance mismatch the new reflection is now smaller by the available voltage difference, or:

$$\Delta V1_{(t=20\mathrm{ns})} = \left(V_b - 2V1_{(t=0)}\right) \cdot \frac{Z_0}{Z_0 + R_B} = (9 - 2.368) \cdot \frac{50}{50 + 330} \approx 0.873\,\mathrm{V}$$

This step is of course superimposed on the existing 2.368, so the line is being filled up to about 3.241 V.

The process now repeats and after each reflection the line voltage increases by a lower voltage difference multiplied by the impedance mismatch factor, reaching very close to the battery voltage V_b after some 500 ns. And it does not end there, it continues to infinity, but the observation of the steps is eventually limited by thermal noise of the circuit.

To see more clearly the fields, let us examine **Fig.10**, which shows the situation some 3 ns after the first reflection, or 13 ns after the closure of the switch.

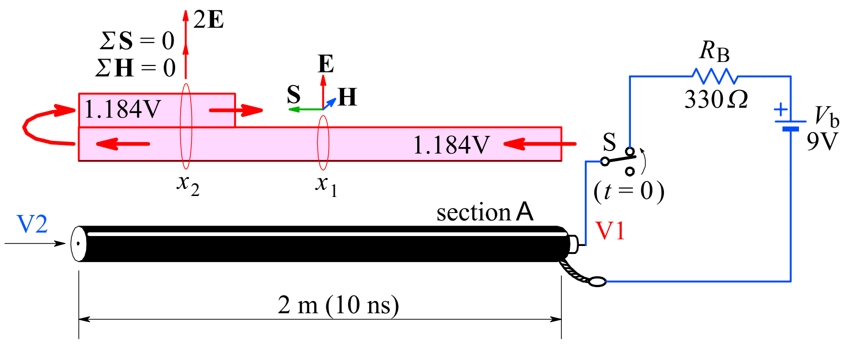

Fig.10: Superposition of fields ~13 ns after the closure of the switch. The battery is filling the line, the first step size being determined by the battery voltage V_b and the impedance ratio. After 10 ns the field reaches the open end at V2 and reflects back, doubling the voltage. Another 3 ns later the field has filled about 1/3 of the line from the back to twice the original step voltage. At a distance x_1 from V1 there is still a net energy flow forwards, as indicated by the presence of $\mathbf{S} = \mathbf{E} \times \mathbf{H}$, but at x_2 the field is apparently 'static', having 2**E**, but $\mathbf{S} = 0$ and $\mathbf{H} = 0$. Yet, the 1.184V wavefront is traveling back superimposed on the existing 1.184V. Of course, the battery is supplying the energy for filling the line backwards, but that energy must first travel towards V2 and be reflected at the open end.

The situation in **Fig.10** is interesting because here we have at the same time a net energy flow at x_1 which means that the battery is filling the line with energy, and yet at the open end we have an apparently 'static' condition, with twice the voltage and zero current!

As the line is filled, the battery supplies the current (less after each reflection at V1), until several µs later the current has dropped practically to zero, and the voltage is practically equal to the battery voltage. But the important question to ask is this:

Did the battery stop supplying energy to the circuit?

The correct answer is: no! Photons cannot stand still! The battery continues to supply the energy to the system, but the system now reflects all the energy back to the battery. The effective current is zero, so no power is dissipated, but the energy continues to flow back and forth throughout the system.

Another point to make is the following: if we put back the 50 kΩ resistor as R_B and charge again the cable, we will not see any stairsteps. This is because the impedance mismatch is very high, 50 Ω to 50 kΩ, so the steps are less than 9 mV, too small to see (at the input sensitivity of 2V/division). Because of the much smaller steps a much larger number of reflections is needed to charge up the line, leading to the classical capacitor charging equation $V(t) = V_b\left(1 - e^{-t/R_B C_c}\right)$. It is important to realize that this classical equation is only a mathematical approximation (for the case when $Z_0 \rightarrow 0$)! The actual charging always occurs in smaller or larger steps.

Discussion

What have we learned?

The experiment and the analysis of its unexpected result force us to acknowledge that there can be no such thing as 'static' electric or magnetic field. Photons (the field carriers of EM energy) cannot stand still, they can only travel at the speed determined by the local μ and ε, which in vacuum limits to the speed of light:

$$v = \frac{1}{\sqrt{\mu\,\varepsilon}} \leq c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \tag{20}$$

So every apparently 'static' field must be formed by at least two EM energy fields traveling in opposite directions. Depending on the type of impedance reflection at a boundary ($Z_1 > Z_2$ or $Z_1 < Z_2$), either the **E** field or the **H** field cancel, producing the 'magnetostatic' or the 'electrostatic' field.

The 'static' energy is kept trapped partially by the circuit shape, and partially by a sharp impedance discontinuity, forcing the outgoing field to be reflected back, thus preventing it from escaping. The impedance discontinuity in the case of an open circuit is determined by the vacuum impedance:

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 377\ \Omega \tag{21}$$

This is a real Ohmic value, and yet no power is dissipated by it. Every radio amateur who ever had to deal with the impedance of his antenna can tell you about it. The antenna impedance has to be adjusted in order to provide a suitable transition from the RF amplifier output and the antenna cable to free space to avoid standing wave reflections (SWR) within the frequency range of interest, otherwise there is a risk that the energy from the transmitter is reflected back and after twice the delay of the antenna cable enter the transmitter output stage out of phase with the generated RF wave, causing excessive power dissipation and eventually a destruction of the output transistors.

So far we were interested only in the numerical values of vacuum permeability and permittivity. But what are they in the physical sense? An insight may be acquired by looking at their units of measure, and the way they are defined.

Let us start with the vacuum permittivity, ε_0. In (9) it is expressed in units of [As/Vm]. Now [As] is the Coulomb [C], which is a measure of charge, but [As/V] is the Farad [F], which is the measure of capacitance. Then, [F/m] means simply the capacitance per unit length. A capacitance is defined in two ways, in terms of geometry (as the effective area of the plates and their distance), but also in terms of a variable current and voltage:

$$C = \frac{i}{dv/dt}\ \text{[F]} \tag{22}$$

which is fundamentally related to **a change of energy**.

A similar discourse is valid for vacuum permeability, μ_0. In (7) we expressed it in units of [Vs/Am]. Now [Vs/A] is the Henry, [H], a measure of inductance, and [H/m] means inductance per unit length. Like the capacitance, the inductance may be represented by the geometry of a coil with a certain number of turns, but also (more importantly) as:

$$L = \frac{v}{di/dt} \text{ [H]} \tag{23}$$

which also relates to a change of energy. Consequently, [F/m] and [H/m] denote the **ability of vacuum to store energy** per unit length.

Alternatively, we may look at the permeability and the permittivity of vacuum as the amount by which vacuum opposes any change in the magnetic and electric energy content, which fundamentally provides the reason for the constancy of the speed of light, $c = 1/\sqrt{\varepsilon_0 \mu_0}$.

The same is valid in coaxial cables: as long as the thermal (Ohmic) losses are low and the applied voltage and current are low, the materials (conductors and insulators) will behave linearly, so their ε and μ can be considered constant. Then, any energy applied to one side of the cable will be transferred essentially without loss to the load at the other end with a propagation velocity $v = 1/\sqrt{\varepsilon\mu}$.

A Broader Perspective

Quantum mechanics, quantum electrodynamics, and quantum field theory all predict that vacuum is not 'empty space', but is full of energy, and the quantum fluctuations of this energy result in constant particle-antiparticle pairs being created and almost immediately annihilated, returning the energy to the vacuum EM fields; during their short lifetime those pairs behave like an electric dipole, they react to any external field by adjusting their polarization, thus giving the vacuum its EM properties.

We have learned that EM energy cannot be static. This means that the instantaneous Coulomb action at a distance is a myth, and there are no longitudinal electric waves in the way envisaged by Tesla.

But there are many other interesting aspects of the problem. For example, the problem of the *skin effect* at high frequencies or sharp pulse edges, the field penetration depth depending on the type of conductive material and the field frequency. But for a step wave the penetration depth increases exponentially after the leading edge of the wave outside the conductor. This means that electrons inside the conductor react to the applied field after a finite time.

A similar problem is superconductivity, where the superconductor expels the field out. At low temperatures the phonon deformations of the material crystal lattice fall below a certain energy threshold; the free electrons can now sense each other's spin over a considerable distance, and those having opposite spin can couple, forming so called *Cooper*'s pairs. Such pairs behave like bosons, since the pair spin is zero; such electron pairs reject any incoming EM energy below a certain high threshold, consequently the field can propagate by the supeconductor's surface without losses.

Finally there is the problem of the electron itself and its own 'static' field.

A pair of photons with energy beyond a certain threshold can interact, giving birth to a pair of complementary charged particles. The lowest possible energy of interaction is when the photons have at least 2×511 keV; their encounter creates an electron–positron pair:

$$\gamma + \gamma \rightleftarrows \mathrm{e}^- + \mathrm{e}^+ \tag{24}$$

The bi-directional arrow indicates that the reaction is reversible, the electron–positron pair annihilates into a pair of photons. This is but one example of vacuum quantum fluctuations.

In the case that there is an excess energy, the created pair acquires a kinetic momentum, and can be separated before annihilation, so the particles remain stable. The energy trapped in the newly formed particles constantly tries to escape, emitting 'virtual photons' (virtual in the sense that there is nothing near enough to catch them), but they are being constantly reflected back by the impedance discontinuity between the 'inside' of the particle and the surrounding vacuum; since the energy cannot escape, the particles remain stable. But if they happen to encounter each other again, the impedance of the space between them is lowered, allowing their internal energy to escape, and we get a pair of photons again.

This effect is used in the Positron Emission Tomography (PET) scanner, where a radioactive marking chemical, bonded to a sugar molecule or other nutrient, is introduced into a patient's blood stream and there decays emitting a positron, which annihilates with the first electron encountered, allowing us to detect the coincidence of two photons hitting the detectors close to the body. By reconstructing the path of the photons we can identify the precise position of annihilation, and thus monitor the biological processes of interest by following the concentration of the radioactive marker in organs.

Such an understanding of elementary particles leads naturally to the question of an internal electron 'structure'. Based on (24) it is reasonable to assume that there must exist a particular geometrical configuration of the internal energy field, which could explain the observed particle charge, its spin, its anomalous magnetic momentum, and possibly also its inertial and relativistic mass in terms of a distortion of the internal field under external forcing. There are many interesting models in literature, ranging from rings to knots, structures more or less similar to what has been proposed by String Theory (all five versions of them). Unfortunately, there are as yet no experimental data which would allow us to prefer one model over any other, so this remains in the area of (educated) speculation.

Appendix 1: Field Wave Equations

We start from the Laplace equation for a potential function within a cylindrical coordinate system. The symmetry is circular and both boundaries are equipotential, so the distribution of the potential is independent of the angle θ. The potential ϕ depends only on the distance r from the center of symmetry, ranging from a to b. So we have:

$$\frac{\partial}{\partial r}\left(r\frac{\partial\phi(r)}{\partial r}\right)=0 \tag{A1.1}$$

A general solution for such a differential equation is:

$$\phi(r)=A\ln r+B \tag{A1.2}$$

where the constants A and B must conform to the boundary conditions. Let us assume that the shield is the zero reference potential, and the core is at a potential U:

$$\begin{aligned}\phi(a)&=U\\ \phi(b)&=0\end{aligned} \tag{A1.3}$$

The particular solution is then:

$$\phi(r)=\frac{U}{\ln(b/a)}\ln(b/r) \tag{A1.4}$$

The electrical field strength in the transversal direction equals the potential gradient:

$$\mathbf{E}(r)=-\nabla\phi(r)=\frac{U}{\ln(b/a)}\cdot\frac{1}{r}\mathbf{1}_r \tag{A1.5}$$

where $\mathbf{1}_r$ is the unit vector in the r direction, and the operator 'nabla' is:

$$\nabla=\mathbf{1}_x\frac{\partial}{\partial x}+\mathbf{1}_y\frac{\partial}{\partial y}+\mathbf{1}_z\frac{\partial}{\partial z} \tag{A1.6}$$

for which the domain of r is the (x,y) plane, and the wave propagates in the z direction. From this we can write the propagation in the z direction of the wave function:

$$\mathbf{E}(r,z)=\frac{U}{\ln(b/a)}\cdot\frac{1}{r}\,\mathrm{e}^{-j\beta_z z}\,\mathbf{1}_r \tag{A1.7}$$

where the phase number $\beta_z=\omega\sqrt{\mu\varepsilon}$ is the same as for the propagation of the planar wave in homogeneous medium or in the free space. The magnetic component is:

$$\mathbf{H}(r,z)=\frac{\mathbf{1}_z\times\mathbf{E}(r,z)}{Z} \tag{A1.8}$$

where $Z=\sqrt{\mu/\varepsilon}$ is the impedance of the medium. Thus:

$$\mathbf{H}(r,z)=\frac{U}{Z\ln(b/a)}\cdot\frac{1}{r}\,\mathrm{e}^{-j\beta_z z}\,\mathbf{1}_\theta \tag{A1.9}$$

where the unity vector is in the direction of the angle θ, around the core conductor.

Fig.A1.1 shows the field in the cable cross-section under the propagation conditions, and under the superposition with a reflected wave.

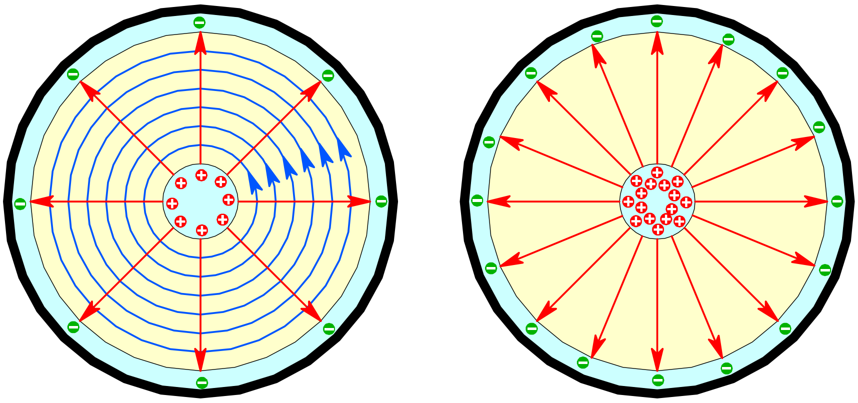

Fig.A1.1: Left: Field cross-section under the propagation condition, with the radial electric field and circular magnetic field. **Right:** under the superposition with the reflected wave the magnetic field cancels and the density of the electric field doubles.

In the case of a lossless line the wave equation is the general Laplace wave equation, relating the space and time derivatives of the second order:

$$\frac{\partial^2 u(x,t)}{\partial x^2} = LC\frac{\partial^2 u(x,t)}{\partial t^2} \tag{A1.10}$$

The product LC has a dimension of the inverse velocity squared:

$$LC = \frac{1}{v^2} \tag{A1.11}$$

The general solutions are the *D'Alambert* functions; for the voltage:

$$u(x,t) = u^+\left(t - \frac{x}{v}\right) + u^-\left(t + \frac{x}{v}\right) \tag{A1.12}$$

and for the current:

$$i(x,t) = i^+\left(t - \frac{x}{v}\right) + i^-\left(t + \frac{x}{v}\right) \tag{A1.13}$$

The positive and the negative superscripts denote the direction of propagation of the wave: the positive superscript is associated with a forward propagation (away from the source), and the negative superscript is associated with a backward propagation (a reflected wave).

The current is not independent of voltage:

$$i(x,t) = -K\int \frac{\partial u(x,t)}{\partial t}dx = Kv\left[u^+\left(t - \frac{x}{v}\right) - u^-\left(t + \frac{x}{v}\right)\right] \tag{A1.14}$$

The product of the constant K and the wave propagation velocity v has a dimension of conductance, so we may define the characteristic resistance of the line:

$$R_0 = \frac{1}{Kv} = \sqrt{\frac{L}{C}} \tag{A1.15}$$

It is of interest to review the Maxwell equations for a superposition of the forward ad reflected wave. For the open end reflection we have no magnetic field:

$$\begin{aligned}\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \cong 0 \\ \nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{A1.16}$$

This is called the 'electro-quasistatic' solution. Similarly, for a reflection at the shorted end we have a 'magneto-quasistatic' solution:

$$\begin{aligned}\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \cong \mathbf{J} \\ \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{A1.17}$$

However, the fact that in both 'static' solutions we do not have any time derivative, that does not mean that the appropriate fields are not functions of time: any change at the source must still propagate to the point x, and then to the end of line and back to x in order to reestablish the new quasistatic condition.

We can verify this by inspecting the Poynting vector. For the open end reflection the power transfer will be:

$$\mathbf{S} = \frac{\mathbf{E} \times \left(\mathbf{H}^{+*} + \mathbf{H}^{-*}\right)}{2} = \frac{1}{2Z}\left(\mathbf{E} \times \mathbf{1}_z \times \mathbf{E}^{+*} - \mathbf{E} \times \mathbf{1}_z \times \mathbf{E}^{-*}\right) \tag{A1.18}$$

By considering the vector product identity: $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$ we can write:

$$\begin{aligned}\mathbf{S} &= \frac{1}{2Z}\left[\left(\mathbf{E}_0^+ \mathrm{e}^{-j\beta\cdot\mathbf{r}} + \mathbf{E}_0^- \mathrm{e}^{j\beta\cdot\mathbf{r}}\right) \cdot \mathbf{E}_0^{+*} \mathrm{e}^{j\beta\cdot\mathbf{r}} - \left(\mathbf{E}_0^+ \mathrm{e}^{-j\beta\cdot\mathbf{r}} + \mathbf{E}_0^- \mathrm{e}^{j\beta\cdot\mathbf{r}}\right) \cdot \mathbf{E}_0^{+*} \mathrm{e}^{-j\beta\cdot\mathbf{r}}\right]\mathbf{1}_z \\ &= \frac{1}{2Z}\left[\left|\mathbf{E}_0^+\right|^2 - \left|\mathbf{E}_0^-\right|^2 + \mathbf{E}_0^- \cdot \mathbf{E}_0^{+*} \mathrm{e}^{2j\beta\cdot\mathbf{r}} - \mathbf{E}_0^+ \cdot \mathbf{E}_0^{-*} \mathrm{e}^{-2j\beta\cdot\mathbf{r}}\right]\mathbf{1}_z \end{aligned} \tag{A1.19}$$

The sum of the last two term in the brackets is imaginary, and does not contribute to the real (work) power density $\mathbf{P}$, so we can write:

$$\mathbf{P} = \Re\{\mathbf{S}\} = \frac{1}{2Z}\left[\left|\mathbf{E}_0^+\right|^2 - \left|\mathbf{E}_0^-\right|^2\right]\mathbf{1}_z = \mathbf{P}^+ + \mathbf{P}^- \tag{A1.20}$$

This means that the power of the two propagation directions are effectively independent and the result is simply a linear superposition (sum) of the two.

Appendix 2: Characteristic impedance

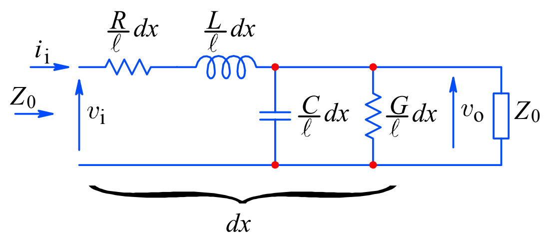

Fig.A2.1: Characteristic impedance calculation.

Fig.A2.1 represents a very short cable segment of a length dx. An infinitely long line is modeled by an infinite number of such segments connected serially. The component values, which are ordinarily given per unit length, say R/l [Ω/m], etc., are now reduced to $(R/l)dx$. The input impedance Z_{in} of the finite line element dx isolated from the rest of the line is equal to the ratio of the input voltage v_{i} to the input current i_{i} (after the input wave is reflected at the open end of the line element):

$$Z_{\mathrm{in}} = \frac{v_{\mathrm{i}}}{i_{\mathrm{i}}} = \frac{R}{l}dx + \frac{sL}{l}dx + \frac{1}{\dfrac{sC}{l}dx + \dfrac{G}{l}dx} \tag{A2.1}$$

With an infinite line composed of infinitely many equal line elements connected, the wave is never reflected back, and the effective input impedance of the line is that of the characteristic impedance Z_0. The first segment is thus loaded by the characteristic impedance. In such conditions the line can be replaced by a lumped impedance of the same value as the characteristic impedance Z_0, on which the wave dissipates completely (converted into heat), so that no energy is reflected. This means that Z_0 must have a completely real Ohmic value: $Z_0 \equiv R_0$.

However, Z_{in} is complex, because $s = j\omega$. If the angular frequency ω is high enough, $R \ll j\omega L$; also, $1/G \gg 1/j\omega C$. This means that for transient phenomena across a cable section a couple of meters in length we can neglect R and G. Nevertheless, we are going to derive the characteristic impedance for the general case. In order to simplify the equations, let us rename the series impedance of the dx section as Z_{s}:

$$dxZ_{\mathrm{s}} = dx\frac{R + sL}{l} \tag{A2.2}$$

and the parallel (shunting) admittance as Y_{p}:

$$dxY_{\mathrm{p}} = dx\frac{G + sC}{l} \tag{A2.3}$$

The input impedance of the dx section loaded by the rest of the line is:

$$Z_{\mathrm{in}} = Z_0 = dxZ_{\mathrm{s}} + \frac{1}{dxY_{\mathrm{p}} + \dfrac{1}{Z_0}} \tag{A2.4}$$

From this we can express the characteristic impedance of the dx section:

$$Z_0\left(dxY_\mathrm{p}+\frac{1}{Z_0}\right)=dxZ_\mathrm{s}\left(dxY_\mathrm{p}+\frac{1}{Z_0}\right)+1 \tag{A2.5}$$

By multiplying the terms in the parentheses we have:

$$Z_0dxY_\mathrm{p}+1=dx^2Z_\mathrm{s}Y_\mathrm{p}+dxZ_\mathrm{s}\frac{1}{Z_0}+1 \tag{A2.6}$$

The term $(+1)$ on both sides cancels:

$$Z_0dxY_\mathrm{p}=dx^2Z_\mathrm{s}Y_\mathrm{p}+dxZ_\mathrm{s}\frac{1}{Z_0} \tag{A2.7}$$

and by multiplying all by Z_0 we obtain:

$$Z_0^2dxY_\mathrm{p}=Z_0dx^2Z_\mathrm{s}Y_\mathrm{p}+dxZ_\mathrm{s} \tag{A2.8}$$

If we now let $dx\to 0$, the term with the second order derivative dx^2 will approach zero more rapidly than the first order terms, so it can be neglected:

$$Z_0^2dxY_\mathrm{p}=dxZ_\mathrm{s} \tag{A2.9}$$

From this we arrive at the value of Z_0:

$$Z_0^2=\frac{dxZ_\mathrm{s}}{dxY_\mathrm{p}}=\frac{Z_\mathrm{s}}{Y_\mathrm{p}}=\frac{\dfrac{R+sL}{l}}{\dfrac{G+sC}{l}}=\frac{R+sL}{G+sC} \tag{A2.10}$$

or:

$$Z_0=\sqrt{\frac{R+sL}{G+sC}} \tag{A2.11}$$

As already explained, for very high frequencies and relatively short cable lengths we can neglect the thermal losses of both dissipative terms, $R\to 0$ and $G\to 0$, so we finally obtain the characteristic impedance of a lossless transmission line:

$$Z_0=\sqrt{\frac{L}{C}} \tag{A2.12}$$

By expressing C and L by their appropriate geometrical functions of (13) and (15), we can write:

$$Z_0=\sqrt{\frac{L}{C}}=\sqrt{\frac{\dfrac{\mu}{2\pi}\ln(b/a)}{\dfrac{2\pi\varepsilon}{\ln(b/a)}}}=\sqrt{\frac{\mu}{\varepsilon}\left[\ln\left(\frac{b}{a}\right)\right]^2}=\ln\left(\frac{b}{a}\right)\sqrt{\frac{\mu}{\varepsilon}} \tag{A2.13}$$

In free space there are no geometrical boundaries, so the factor $\ln(b/a)$ does not apply, $\mu_\mathrm{r}=1$ and $\varepsilon_\mathrm{r}=1$, and the characteristic impedance of free space is:

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 377\ \Omega \tag{A2.14}$$

Therefore in homogeneous, isotropic, and linear (HIL) media the impedance is independent of frequency, the wave propagation is non-dispersive, and the waveform shape is preserved.

Appendix 3: Vacuum Polarization

In classical electrodynamics, a weak electric field influencing a dielectric medium slightly displaces the electrons from their ordinary distribution in respect with their atomic nuclei, thus creating small dipoles within the material. The classical formalism treats such effects in terms of a polarization vector **P**, which is proportional to the electric field **E**:

$$\mathbf{P} = \chi\varepsilon_0\mathbf{E} \tag{A3.1}$$

Usually we define the dielectric displacement as:

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P} = \varepsilon_\mathrm{r}\varepsilon_0\mathbf{E} \tag{A3.2}$$

so that the relative permittivity is understood as a function of the susceptibility of the optical medium in question:

$$\varepsilon_\mathrm{r} = 1 + \chi \tag{A3.3}$$

Assume that an external field interacts with a virtual electron-positron pair. The dipole momentum:

$$p = q_\mathrm{e}x \tag{A3.4}$$

is induced on that pair, with q_e as the elementary charge, and x as the displacement. We can compute its magnitude by assuming that a virtual pair behaves like a harmonic oscillator in the quasistatic limit:

$$m_\mathrm{e}\omega_0^2 x = q_\mathrm{e}E \tag{A3.5}$$

with m_e as the electron mass, and ω_0 as the natural resonant frequency. When the externally applied field is small, $E \ll E_s$ (E_s is the Schwinger limit, 1.3×10^{18} V/m), it is possible to find the resonant frequency from the energy associated with the quantum transition. If the virtual pair represents the ground energy state of the $\mathrm{e}^+\mathrm{e}^-$ pair, and the real positronium [17] 'atom' represents the excited state, we can assume the existence of an energy gap E_g such that a particular oscillation frequency can be associated with it:

$$\omega_0 = \frac{E_\mathrm{g}}{\hbar} \tag{A3.6}$$

This energy gap should be equal to the rest mass of the positronium:

$$\hbar\omega_0 = 2m_\mathrm{e}c^2 \tag{A3.7}$$

Thus by combining (A3.4) and (A3.5) we find the induced dipole momentum:

$$p = \frac{q_\mathrm{e}^2}{m_\mathrm{e}\omega_0^2}E \tag{A3.8}$$

Of course, the magnitude of polarization must depend on the effective volume per each dipole. Both theoretical and experimental analyses have shown that the *Compton*'s wavelength [18] is an appropriate size for a virtual $\mathrm{e}^+\mathrm{e}^-$ pair:

$$\lambda_{\mathrm{C}} = \frac{\hbar}{m_{\mathrm{e}} c} \tag{A3.9}$$

This brings us to the value of the vacuum polarization:

$$P_0 = \frac{q_{\mathrm{e}}^2}{m_{\mathrm{e}} \omega_0^2 \lambda_{\mathrm{C}}^3} E \tag{A3.10}$$

Because of the similarity with (A3.1), we can assign to the quantity multiplying E the role of an effective permittivity:

$$\widetilde{\varepsilon}_0 = \frac{q_{\mathrm{e}}^2}{m_{\mathrm{e}} \omega_0^2 \lambda_{\mathrm{C}}^3} \tag{A3.11}$$

Note that this would be the value of the vacuum permittivity if the particle pairs created by the vacuum energy quantum fluctuations were only electrons and positrons. However, it is equally possible that the vacuum energy density is much higher, thus permitting fluctuations high enough for the creation of a whole spectrum of more massive particle pairs, the average being probably close to pion pairs, $\pi^+\pi^-$. Indeed, the actuall permittivity value is more than an order of magnitude higher than what (A3.11) suggests, indicating that such a scenario is very probable.

More on the subject of vacuum energy and its quantum fluctuations can be found at [22] and beyond.

Further Reading

Visiting Wikipedia is like visiting a dentist: one does it only when necessary. Nevertheless, I recommend some pages here for a quick flyby, as well as to ease the search for related subjects; also, it is often possible to find historically important original articles on the problem of interest at the end of a page, and there are some nice GIF animations, too.

[1] <http://en.wikipedia.org/wiki/Coaxial_cable>

[2] <http://en.wikipedia.org/wiki/Transmission_line>

[3] <http://en.wikipedia.org/wiki/Heaviside_condition>

[4] <http://en.wikipedia.org/wiki/Permittivity>

[5] <http://en.wikipedia.org/wiki/Magnetic_permeability>

[6] <http://en.wikipedia.org/wiki/Wave_impedance>

[7] <http://en.wikipedia.org/wiki/Characteristic_impedance>

[8] <http://en.wikipedia.org/wiki/Wave_equation>

[9] <http://en.wikipedia.org/wiki/Superposition_principle>

[10] <http://en.wikipedia.org/wiki/Quantum_superposition>

[11] <http://en.wikipedia.org/wiki/Telegrapher%27s_equations>

[12] <http://en.wikipedia.org/wiki/Vector_product>

[13] <http://en.wikipedia.org/wiki/Longitudinal_wave>

[14] <http://en.wikipedia.org/wiki/Electromagnetic_wave>

[15] <http://en.wikipedia.org/wiki/Heaviside_step_function>

[16] <http://en.wikipedia.org/wiki/Virtual_particles>

[17] <http://en.wikipedia.org/wiki/Positronium>

[18] <http://en.wikipedia.org/wiki/Compton_wavelength>

More on transmission line theory:

[19] <http://alignment.hep.brandeis.edu/Lab/XLine/XLine.html>

[20] <http://www.allaboutcircuits.com/vol_2/chpt_14/3.html>

A numerical simulation movie showing the experiment field dynamics:

[21] <http://www-f9.ijs.si/~margan/Articles/TLMovie.mpg>

On the properties of vacuum as an electromagnetic medium:

[22] <http://www-f9.ijs.si/~margan/Articles/SomeConsequences.pdf>

[23] <http://www-f9.ijs.si/~margan/Articles/vacuum_energy_density.pdf>

[24] <http://en.wikipedia.org/wiki/Vacuum_energy>

[25] <http://en.wikipedia.org/wiki/Vacuum_state>

[26] <http://en.wikipedia.org/wiki/Zero-point_energy>