

Appendix B

Introduction to Field Theory

B.1 Action at a Distance and Field Dynamics

The previous construction of Fresnel/Young/Huygens tell us how to construct an amplitude for light at any point in space given the amplitude at some other point in space. This is the first part of the construction of a field. A field is something, generally a measured quantity, that is defined at every point in space. At each point in space you can measure the entity. In addition, as you move from one point to a nearby point the value of the something changes smoothly; it varies as you change places. There will even be a rule on how the change as you move from point to point is manifest. To appreciate these rather abstract comments let's look at several examples.

There are numerous examples of fields. The temperature in a room is a field. Temperature is measured for instance by a mercury bulb thermometer. As you move the thermometer from point to point, you will get different values for the temperature. If the room is not too drafty, the temperature at nearby points will be similar; the temperature varies smoothly as you move to nearby points. You can even intuit certain rules for how the temperature changes as you move from point to point. For instance, you can guess that a point at the center of a surrounding group of points, the temperature will be the average of the temperatures of the surrounding points. It is because of rules like this that you expect that the temperature varies smoothly as you go among nearby points. Other obvious examples of fields are air pressure in a room, height above or below the normal height of water in a pool, or the transverse displacement of a stretched string. With some amount of smoothing you can make a field from such things as population density on the earth. Any system that is defined over a continuous manifold is a field.

The discussion of the previous examples generally did not deal with the time variation. It is not until we endow something with a time dependence that the something becomes interesting. In fact, as we will see, Section A.5.4, we cannot really talk about energy until we have temporal evolution. In the Fresnel/Young/Huygens construction of the amplitude for light, we eliminated the effect of the time variation by “seeing” only the brightness, the amplitude squared, and averaging for long times so that the short time oscillations of the phasers cancelled out, Section ???. Thus although the brightness as a field can be interpreted as slowly varying there is an intrinsic time variation that makes light especially interesting.

In other words, a field is something that is defined over some manifold, usually space, that has a temporal evolution. The rules for the behavior of the field are usually local in the sense that its variation in space and time is determined by what is going on at those points of space at those times. This is the meaning of local causality. It is one of the bedrock principles of modern physics. It ranks with reductionism as one of our formulating rules. The basic idea is that what happens to an entity happens because of what is going on at the place at which the entity is or the immediate neighborhood. This is in sharp contrast to the situation in theories that are based on action at a distance dynamics. Newton’s Laws of gravitation are an example of an action at a distance theories. To a large extent, it was the attempt to remove these action at a distance formulation and replace them with locally causal theories that motivated the development of field theories.

B.1.1 Action at a Distance

My former colleague, Johnny Wheeler calls it ”spooky” action at a distance. Newton, its inventor, was not comfortable with the concept but could not come up with something better. In a letter to the theologian Robert Bentley, he wrote:

that gravity should be innate, inherent and essential to Matter, so that one body may act upon another at a Distance thro’ a Vacuum, without the Mediation of any thing else, by and through which their Action and Force may be conveyed from one to another, is to me so great an Absurdity that I believe no Man who has in philosophical Matters a competent Faculty of thinking, can ever fall into it. Gravity must be caused by an Agent acting constantly according to certain laws; but whether

this Agent be material or immaterial, I have left the consideration of my Readers.

Regardless of his own reservations and because of the success of the Newtonian approach, physicists became accepting of the anomalous nature of action at a distance and the early formulations of most laws were all in the pattern of action at a distance. Fortunately, Maxwell could not believe these and, for the case of electricity and magnetism, this led him to the development of the first first-principle field theory. Prior to Maxwell's work there were field theories but these were derivative of an underlying structure. For example, the rules of fluid flow were formulated in a field theory vocabulary. But this was understood to be a consequence of the underlying structure of the fluid. Maxwell's formulation of the nature of the electric and magnetic systems was actually a statement on the intrinsic properties of these entities. In order to understand this important idea let's review the situation with action at a distance theories and the contrast to field theories.

All the satisfactory theories prior to the 19th century were not what we now call locally causal theories but instead were based on action at a distance theories, actions resulted from situations that were at a distance from the object of interest. Newton's theory of the gravitational force is a perfect example. In Newton's approach to gravitation, a body's motion is determined by the separation from a remote other body at the instant under consideration. The moon determined its acceleration from knowledge of the earth's position which is at a distance at that instant. It is hard to accept that, if the earth suddenly ceased to exist that, at instant, the moon would instantaneously react by traveling off in a straight line, no longer in orbit. There are two issues here. First the idea that somehow that moon is influenced not by things going on where it is and the fact that the earth's disappearance should be realized by the moon instantaneously; it should take some time. Consider the case that I am standing in the front of the lecture hall and announce that I am going to make the clock at the back of the room run differently. If I could do that, you would infer that I had a wire or used sound waves or some other mechanism to communicate the change to the immediate vicinity of the clock. Whatever ultimately changed the clocks running was at the place of the clock not at a distance.

Coulomb's law and all the other laws of electromagnetism that were formulated before the 19th Century were action at distance laws. A charge here effected a charge there.

The solution to this basic philosophical conundrum is in the idea of strict locality for all phenomena and the vehicle is the concept called the

field. Of course, in physics, a philosophic problem is not a good reason for doing something. The idea must be tested experimentally. The proof of the construction is in the testing. Through his treatment of electromagnetic phenomena as a field theory, Maxwell was lead to predict that light was a disturbance of the electromagnetic field. When this prediction was verified by Heinrich Hertz in 1887, there was a general acceptance of Maxwell's approach. Since that time, we have found that all fundamental theories are field theories; the ultimate modern expression of the nature of matter and energy being through the machinery of quantum field theory. For this reason, it is important to understand the idea of the field. For now we will develop the classical field, we will add the complications of quantum mechanics, see Chaper ??.

B.1.2 Local Field Theory

Maxwell developed a local field theory to describe the phenomena associated with what is called electricity and magnetism. He reduced all the known laws of electricity and magnetism into four reasonably simple equations. In so doing, he unified the electric and magnetic forces and predicted the fundamental nature of light. These are considerable accomplishments in their own right but also he somewhat inadvertently clarified the idea of the field and the idea of causality. His was not the first field theory; it was the first field theory of a fundamental force system. The first local field theory and the easiest to appreciate was the description of fluid flow. It was the success of a field theory of fluid flow that motivated him to attempt to write the rules of the electricity and magnetism in this field theory form.

How fluids move through space is very complex. At any point in the fluid there are several variables that are necessary to describe the state of the fluid. These variables such as density, velocity, and temperature are all fields, defined at each point in space and subject to change by some set of rules that are determined by the values of these variables at that point and nearby points and by the nature of the fluid. For example, if the temperature at a point is higher than its neighbors, that temperature will tend to decrease because of heat flow from the neighbors. Also depending on the nature of the fluid, the density may increase and this will cause flow away from the point. How much effect each variable has on the magnitude of the the other variables and how fast these variables respond will depend on the fluid. The parameters such as the thermal conductivity and compressibility of the fluid which will control the rates at which these effects can take place are measured phenomenologically for each fluid. It is not hard to understand

that the properties of a fluid in motion are controlled by local effects; flow at a point depends on the temperature and pressure and flow at the point and neighboring points not on what is going on some distance away. The rules for the fluid flow are thus local. The difference with the results of Maxwell is that we know there is an underlying structure, the atoms. In the case of the electromagnetic field, it is not made of anything but itself. The inability to associate a reality to the field independent of an underlying structure is the basis for the famous search for an ether, see Section ??.

In fact, Maxwell suffered from that same problem. He discovered his equations by trying to fill space with a hypothetical something that exhibited reasonable mechanical properties and attributing the electric and magnetic forces to whirling vortices in the pervasive medium. The idea was that charges produced vortices in this medium and that the whirling of the vortices close to the charge then produced other vortices etc. until space was filled with whirling vortices and the amount of whirling at any place was the electric force. In other words, in order to understand his own equations, he needed an ether, the famous ether that Einstein disposed of later. He also needed to have the vortices properties be determined by the charge or the whirlyness locally. To the modern physicist, the idea of an underlying mechanical system seems out of place and a little weird. In fact, several years back, there was a collection of articles published that were “lighthearted” musings by well known scientists, [Weber 1973]. These articles were written as joke. Among the collected articles was the original paper by Maxwell justifying his vortices in the ether as a mechanism for the electromagnetic field. At the time of the writing, there was nothing lighthearted about it.

B.2 The Stretched String

Since the concept of the field and its dynamical rules are rather hard to grasp in the abstract, let’s look at a particularly simple mechanical field system – the transverse displacement of a stretched string. I have to emphasize that this is a field with an obvious underlying mechanical structure – the string, a system with mass and an internal force, the tension. This is in contrast with the fields that we will deal with later. These fields are themselves the fundamental entities. The other thing to realize is that the string that we deal with is an idealized element. It has zero thickness and bends with no resistance. Its only possible displacement is transverse to its alignment.

The displacement of the string in a direction transverse to its direction is a field defined on all the points along the string. This field is much

simpler than the electromagnetic field which is a field composed of two vector quantities, the electric and magnetic forces. The string field also obeys a simple mechanical rule for its dynamics.

Like most mechanically based systems the dynamics of the string has two simple sources, energy of the motion of its masses and a potential energy that is due to its configuration. For the case of a string held tightly with a tension T_e and with only transverse displacements, the potential energy is the work associated with making the string longer. The displacement of the string in the transverse direction is the field that we will consider and any non-zero displacement causes the string to be longer and thus changes the potential energy. These are global approaches to the behavior of the string and will be useful to us later when we use a more universal approach to dynamics based on a concept called action, see Section A.2. For now because our goal will be an understanding of the electromagnetic field, we will use a more local approach and find that the electromagnetic field has many of the same properties as this the simplest of fields. In this approach the electromagnetic field is just a more complex field and the complications do not add any to the understanding of the field nature of the system. For example, the stretched string is a one dimensional field defined on a one dimensional manifold, the distance along the direction of the string. The field variable, $y(x, t)$, is also simple in that it is the transverse displacement of the string from its equilibrium position where x is the position along string. Both y and x range over a one dimensional range of values. The electromagnetic field is a pair of vectors in its field variable and it ranges over a three dimensional manifold, space.

You may also be perplexed by the idea of a stretched string under tension. Our experience is that a string has to be fastened to be under tension. If that is the case, think of the string as tightly stretched between fixed walls. The problem with this is that the walls add complications of their own and for the first pass are not necessary. Here we deal with an infinite string under tension. Later, we will deal with the walls, see Section ??.

The local statement of the dynamics of the string are easy to understand; the rule is very simple and intuitive: The force on a segment of the string caused by the transverse displacement of that piece of the string is proportional to the negative of the average of the displacement of that segment of the string from the displacement of its neighbors.

In order to implement this algorithm, divide the string into small segments of length Δl and concentrate the mass in the segment at a point, see Figure B.1. In the example shown, the segment of string labeled i is above the position of the average of its two neighbors. Thus there is a force to

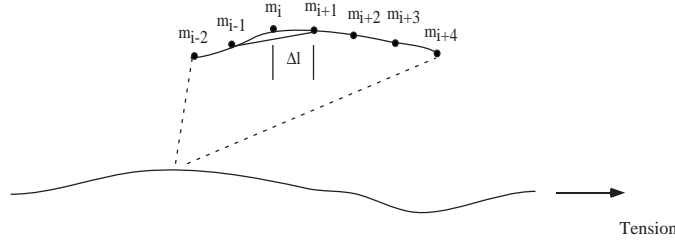


Figure B.1: **The Stretched String** A string that can move in the transverse direction under tension is a simple example of a local field. In the figure, a section is magnified. In this section, the string is divided into small segments of length Δl and the mass of each segment is concentrated at a point. The dynamic of the string is that the mass at segment at location i has a force on it if its transverse displacement is different from the average of its two neighbors. Thus in the case shown, by drawing a straight line between masses at $i - 1$ and $i + 1$, we can see that at the place of segment i , the neighbors' average is below i 's current position. Thus i has a downward force on it.

bring it to the position of the average. The proportionality constant for this force has the dimensions of a force per unit length and is thus the twice the tension in the string divided by the length of the segment of string; twice since both neighbors pull. ρ is the mass per unit length of the string and thus the mass of each segment is $\rho\Delta l$. Using $\vec{F} = m_i\vec{a}_i$ and using the position along the string x as label for the piece of string, the transverse displacement of the string at x is $y(x, t)$, the average of the two neighbors of x is $\frac{\{y(x+\Delta l, t)+y(x-\Delta l, t)\}}{2}$, the force equation for the segment at x is

$$\rho\Delta l a_{x,t} = -\frac{2T_e}{\Delta l} \left[y(x, t) - \frac{\{y(x + \Delta l, t) + y(x - \Delta l, t)\}}{2} \right], \quad (\text{B.1})$$

where T_e is the tension in the string.

Another way to organize the right side of Equation B.1, is to note that

$$\begin{aligned} \frac{2}{\Delta l^2} \left[y(x, t) - \frac{\{y(x + \Delta l, t) + y(x - \Delta l, t)\}}{2} \right] = \\ - \left\{ \frac{\Delta y}{\Delta l} \left(x + \frac{\Delta l}{2}, t \right) - \frac{\Delta y}{\Delta l} \left(x - \frac{\Delta l}{2}, t \right) \right\}. \end{aligned} \quad (\text{B.2})$$

This last term on the right is the negative of the definition of the second derivative of $y(x, t)$. Note also that the acceleration is the second derivative

with respect to time. In the limit that Δl is zero and using partial derivatives because we have both x and t dependence, this force equation becomes

$$\rho \frac{\partial^2 y}{\partial t^2}(x, t) = T_e \frac{\partial^2 y}{\partial x^2}(x, t). \quad (\text{B.3})$$

This is an excellent example of the general form in which the dynamics of fields are expressed. They are generally partial differential equations because we are interested in how the field changes for changes in position and time. Equation B.3 is second order in the time derivatives because that is how the dynamic operates; it emerged from a mechanical force law. Other orders of time derivatives are possible and it is not uncommon to have laws that are first order in time. In fact, it is preferable because the interpretation of the evolution is simpler. Maxwell's Equations are an example. The stretched string or any higher order temporal evolution can be reduced to a first order temporal evolution by defining new fields. Defining a new field, the velocity field, $v(x, t) \equiv \frac{\partial y}{\partial t}$, we can get an evolution that has only first time derivatives.

$$\begin{aligned} \frac{\partial y}{\partial t}(x, t) &= v(x, t) \\ \rho \frac{\partial v}{\partial t}(x, t) &= T_e \frac{\partial^2 y}{\partial x^2}(x, t). \end{aligned} \quad (\text{B.4})$$

In a very real sense, you could say the the magnetic part of the electromagnetic system is a manifestation of this kind of substitution. More on this later, see Section 1.3.

The fact that there are only values of the field and spatial derivatives of the field on the right side of the Equation B.3 is the expression of the locality of the dynamic. How the field evolves at a place depends only on what is going on at that point. Also note that the only parameters in the field equation are ρ and T_e . These express the intrinsic properties of the medium in which the field operates. By dividing Equation B.3 by ρ , we can reduce the effective number of parameters to one, $\frac{T_e}{\rho} \stackrel{\text{dim}}{=} \frac{L^2}{T^2}$. This has the dimensions of a velocity squared. The fact that there is only this parameter in the dynamic says a great deal about the nature of the evolution of the fields. There are not enough parameters to construct a length or a time. Thus for this field there is no intrinsic size except as it is put in by the starting conditions or put into the problem by boundaries like walls. Thus this particular field system, the stretched rope, is characterized by movement of field configurations. Since the parameter of the medium is a velocity squared, the movement is in both directions with a characteristic

speed, $\pm\sqrt{\frac{T_e}{\rho}}$. It is important to remember that the movement of a piece of string is only in the transverse direction whereas the movement of the field configurations is along the direction in which the string is aligned. This is a difficult situation to describe. If you attribute all reality to the hunk of string the only motion is up and down in the transverse direction. Yet the configuration of the string moves along the string. We will find that there is energy and momentum associated with the configuration of the string and that this thus moves with the configuration along the string. Thus we have the problem of the ‘string’ only moving up and down but energy and momentum flowing along the string.

The converse of the above result that the parameters of the system are not sufficient to determine a size or time scale is that the medium, in the case of the stretched string are ρ and T_e , implies that the disturbances in the string travel with a speed set by the medium, $\sqrt{\frac{T_e}{\rho}}$ and that this speed is independent of the form of the disturbance. In other words disturbances travel with speed $\pm\sqrt{\frac{T_e}{\rho}}$ without distortion. For this reason, systems with this field dynamic are called wavelike. This is the definition of a wavelike medium. Although many systems are wavelike such as sound and light, other field systems may not be. For instance the dynamic for temperature flow in one spatial dimension is

$$\frac{\partial T_{emp}}{\partial t}(x, t) = a^2 \frac{\partial^2 T_{emp}}{\partial x^2}(x, t). \quad (\text{B.5})$$

where a^2 is called the diffusion constant and is the ratio of the heat conductivity to the heat capacity of the material. Notice that $a^2 \stackrel{\text{dim}}{=} \frac{L^2}{T}$ and thus there is no special speed or length or time that is characteristic of the field.

In order to better understand the operation of field dynamics let’s work though the example of the string under tension. Consider our case of a stretched string with mass per unit length ρ and tension T_e . At $t = 0$, we put a distortion in the string as shown in Figure B.2. Note that at $t = 0$, the string is displaced but no part of the string is moving. It is simplest to interpret the operation of the dynamic in the first order time derivative form, Equation B.4. In this form, it is clear that a complete description of the initial configuration of the string involves the specification of two fields, the initial velocity field and the initial displacement field. In other words for the case in Figure B.2 at $t = 0$, the velocity of all parts of the string is zero and there is a simple pulse of displacement in the string. Other starting configurations are possible. You could have the situation in which the string has no displacement and the sting has a distribution of transverse velocity.

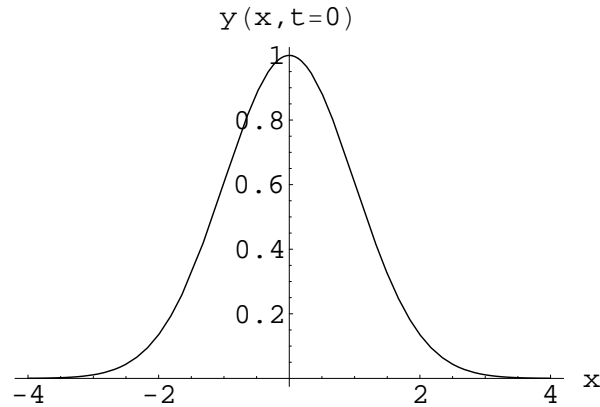


Figure B.2: **A Simple Displacement Pulse in a String** A simple pulse in a stretched string under tension. At $t = 0$, the string is distorted but no part of the string moving.

The difference in the operation of a harpsichord and piano is the the strings are plucked or distorted in the harpsichord and hammered in a piano. You can also have situations with both an initial displacement and velocity.

The dynamic of the string requires that all points on the string be at the average of its neighbors. An easy way to compute the average is to pick two neighbors, points on the string close to the point of interest and equidistant from it, and connect the points by a straight line. At the point of interest, x , the point on the line is the average of the two neighbors. Thus from Figure B.3, we see that the center of the string is pulled strongly down and the edges are pulled up. The points of steepest drop are not pulled at all. This last point is interesting to note. The string is not pulled to the neutral position. Each segment is pulled only by its neighbors. If the string were pulled to the neutral position there would be a force for the entire time of descent and then the string would still have a velocity when it reached the neutral position and thus would overshoot and there would be oscillation at each disturbed point on the string. As we know, the disturbance in the stretched string is removed by the dynamic with the string returning gently to its neutral position.

To make this discussion more quantitative, we look at what goes on in a few small time increments. In a small time, Δt , since the velocity field is initially zero everywhere, we find that the string has not moved.

$$y(x, \Delta t) = v(x, 0)\Delta t + y(x, 0)$$

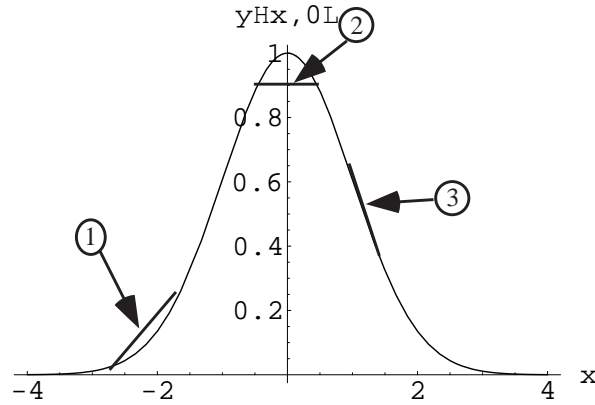


Figure B.3: **Forces on a Pulsed String** The dynamic of the stretched string require that all points in the string be at the average of its neighbors. A simple rule for finding the force and thus the acceleration of a place on the string is to connect the neighbors with a straight line. If the string at that place is above the line, there will be a downward acceleration with magnitude proportional to the distance above. There are three examples shown. At a point on the edge of the pulse, 1, the string is accelerating upward. At the center, 2, the string is accelerating down. At a point at the midpoint of the side of the pulse, 3, the string has no acceleration.

$$= y(x, 0), \tag{B.6}$$

where $v(x, t)$ is the velocity of the string at the point labeled x at time t . At $t = 0$, the string is not moving and $y(x, 0)$ is known.

We will need the velocity of the string at all times and, even in a small time, because of the forces from Figure B.3, the velocity changes.

$$\begin{aligned} v(x, \Delta t) &= a_{t=0}(x)\Delta t + v(x, 0) \\ &= a_{t=0}(x)\Delta t \end{aligned} \tag{B.7}$$

where we find $a_{t=0}(x)$ from an analysis such as that shown in Figures B.3 for each point on the string. Thus we see that after a time Δt the velocities will have the same pattern as a function of position as the initial accelerations.

Repeating the process for a second Δt using Equations B.6 and B.7 but with the time shifted another increment,

$$\begin{aligned} v(x, 2\Delta t) &= a_{t=\Delta t}(x)\Delta t + v(x, \Delta t) \\ &= a_{t=0}(x)\Delta t + a_{t=0}(x)\Delta t \\ &= 2a_{t=0}(x)\Delta t \end{aligned} \tag{B.8}$$

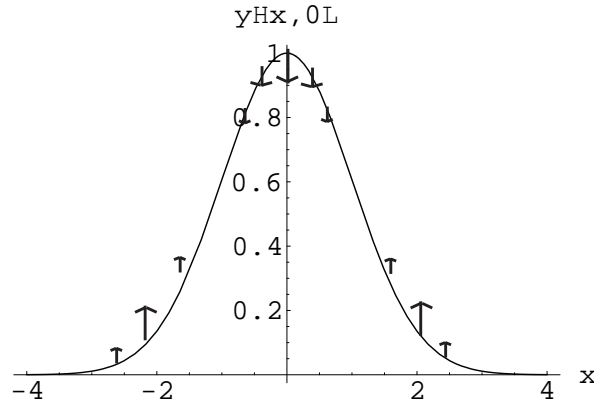


Figure B.4: **Accelerations on a Pulsed String** Using a technique such as shown in Figure B.4 for the forces on the string, the algorithm in Equation B.1 can be applied at each point, x , and find the accelerations shown as arrows above.

where in the second line, I used the fact that since $y(x, \Delta t) = y(x, 0)$ and the accelerations depend only on $y(x, t)$, then $a_{t=\Delta t}(x) = a_{t=0}(x)$.

The second dynamic is handled similarly,

$$\begin{aligned} y(x, 2\Delta t) &= v(x, \Delta t)\Delta t + y(x, \Delta t) \\ &= a_{t=0}(x)\Delta t^2 + y(x, 0). \end{aligned} \quad (\text{B.9})$$

We now begin to see the string moving.

We can intuit that the pattern shown in Figure B.5 develops. The region where there is a strong bend at the edge is pulled up and so has an upward velocity and begins to lift. The middle section is unchanged at first. The center is forced down and has a downward velocity. Because of the pattern of the upward velocity at the bends and the downward velocity at the center, the two separating pulses appear to be moving along the string away from each other. We have to remember that the all the motion of the string is transverse to its direction.

The general pattern then develops of two distinct pulses of half the original amplitude one moving to the left and one to the right, see figure B.6. This transverse velocity is patterned so that the two emergent pulses are one moving to the left with speed $-\sqrt{\frac{T_e}{\rho}}$ and one moving to the right with speed $\sqrt{\frac{T_e}{\rho}}$. Each of these are called traveling waves, one to the left and one to the right. It is the pattern of traveling waves that there is both a transverse

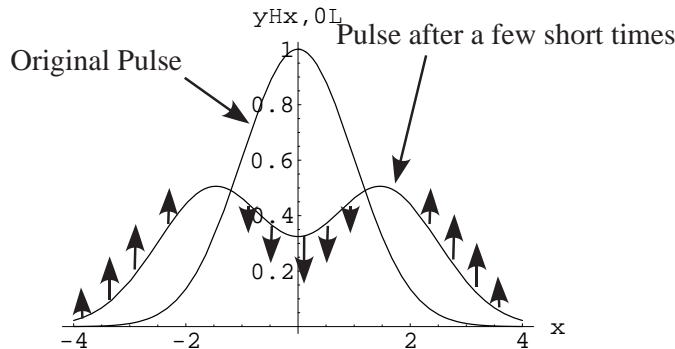


Figure B.5: **Pulsed String after a Few Short Times** Using appropriate versions of Equations B.6 and B.7 to evolve the system, we can see the development of two pulses. Also shown are the velocities by scaled arrows. Remember the parts of the string are only free to move up and down but the pattern of up and down motion conspires to produce the effect that the pulse at negative x is moving toward greater negative x and the pulse at positive x is moving toward greater positive x . The original pulse is shown for comparison.

displacement field and an associated transverse velocity field with the velocity field rising in front of the motion of the traveler and falling behind the traveler. This is a typical pattern for wavelike media. There are two fields that support each other and form the traveling configuration. For sound it is the density of the air and the pressure of the air. For electromagnetic waves, it is the electric and magnetic force fields.

It is worthwhile to also note that our original configuration of the displacement pulse with no velocity, Figure B.2 can be considered as the sum of two travelers, one going to the left and one going to the right, each of half amplitude. The addition of the displacement field gives the correct shape for the pulse and, at the instant of complete overlap, the initial instant, the two transverse velocity fields add to zero. The ability to treat the original distortion as a sum of two independent distortions is an example of superposition. This will be an important principle in many future discussions, see Section ??.

In addition, the travelers have an interesting relationship between the displacement field and the velocity field. For a traveler that moves to increasing x , the argument of the displacement field is a single variable, $x - \sqrt{\frac{T_0}{\rho}}t$, instead of x and t as independent variables. This traveler is called a right

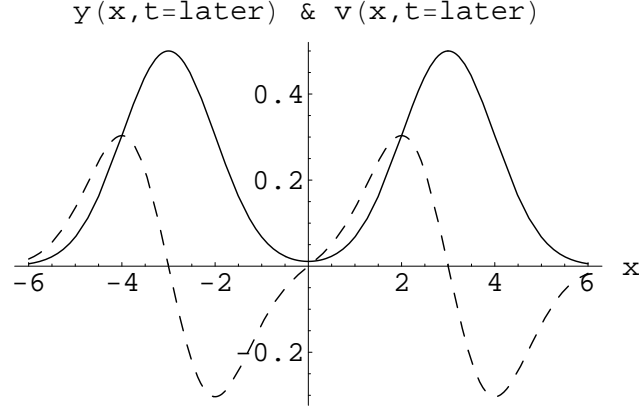


Figure B.6: **Pulses in String Separating** After a time, the pulse initially placed on a stretched string, see Figure B.2, separates into two half amplitude pulses. One travels to the left with velocity $v = -\sqrt{\frac{T_e}{\rho}}$ and one travels to the right with velocity $v = \sqrt{\frac{T_e}{\rho}}$. There is also a transverse velocity field that travels along with each pulse shown as the dashed curve instead of using arrows as in Figure B.5.

traveler. For waves that move to decreasing x , called left travelers, the argument is $x + \sqrt{\frac{T_e}{\rho}}t$. This is what makes them travelers; they move to increasing x or decreasing x uniformly without the shape of the disturbance changing. This is a general result and true for all one dimensional wavelike systems. We worked this out for the particular disturbance of Figure B.2, a simple pulse. It should be clear that this pattern of two separate travelers superposing to produce an initial distortion with no velocity field will hold for any form of distortion for the displacement field. Figure B.7, shows a more general initial configuration and the subsequent travelers. Because to the nature of the relationship between the x and t variables in the travelers, $x - \sqrt{\frac{T_e}{\rho}}t$ for the right traveler and $x + \sqrt{\frac{T_e}{\rho}}t$ for the left traveler, the time evolution of the displacement field which is the velocity field in this dynamic is related to the slope of the displacement of the traveler at that point.

$$\frac{\partial y_{rt}}{\partial t}(x, t) \equiv v_{rt}(x, t) = -\sqrt{\frac{T_e}{\rho}} \frac{\partial y_{rt}}{\partial x}(x, t) \quad (\text{B.10})$$

where $y_{rt}(x, t)$ and $v_{rt}(x, t)$ are the right traveling waves displacement field

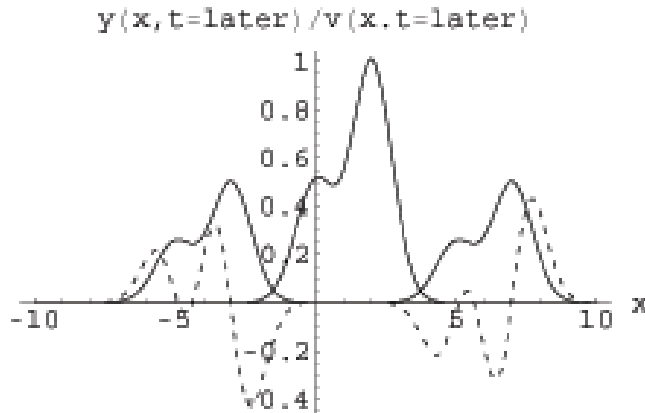


Figure B.7: **Arbitrary Traveling Waves** Using a more general form for the initial distortion of the string, shown at the center for reference, we see at a later time the two traveling distortions, one moving to increasing x called the right traveler and one moving to decreasing x called the left traveler. The associated velocity profile for each is shown dotted. Because of the special form of the argument of the travelers, the velocity profile for the right traveler is proportional to the negative of the slope of the displacement profile of the right traveler at that instant and the velocity profile of the left traveler is proportional to the slope of the displacement profile of the left traveler at that instant.

and velocity field. The relationship of the velocity field and the displacement field for the left traveler is similarly:

$$\frac{\partial y_{lt}}{\partial t}(x, t) \equiv v_{lt}(x, t) = -\sqrt{\frac{T_e}{\rho}} \frac{\partial y_{lt}}{\partial x}(x, t). \quad (\text{B.11})$$

Another feature of the travelers is that they carry energy and momentum. It takes a certain amount of work to distort the string; it has to become longer. This distortion energy is then distributed into the travelers and these then carry it off to remote regions of the string. Similarly there is momentum associated with the travelers that is transported down the string by the travelers. In a later section, we will develop a more nuanced identification for momentum and energy, see Section A.2 but for now our intuitive ideas will suffice. Notice that this is energy and momentum that moves down the string even though the string itself can only move in a transverse direction. Thus the traveler wave configurations act like a thing that moves along the

string even though nothing moves down the string. Note that a superposition of the travelers constitute the original disturbance. Here we begin to see the development of a thing, something that carries energy and momentum, in the context of a field. The electromagnetic field is a wave field and will have travelers also. These are more complex and constituted differently in the dynamic than these string travelers but they behave similarly. Since they generally operate in three spatial dimensions there is a geometric fall off in strength as they travel but they still carry energy and momentum to remote parts of the system.

In Section ??, The Stretched String Revisited, we will return to the dynamics of the string. For now we are content to use it as a simple example of a field system and to have it express the basic ideas of a field theory, a construction that is a local causal dynamical system. In the next section, we will discuss Maxwell's Equations, the first fundamental theory based on a field construction.

B.3 Maxwell's Theory of Electromagnetism

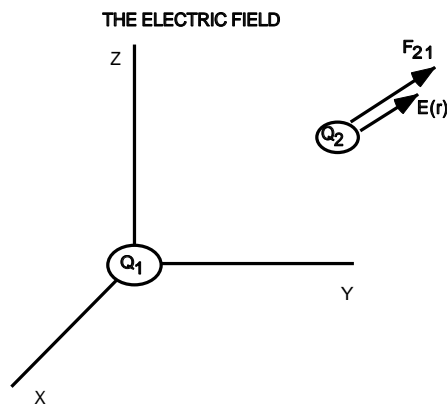


Figure B.8: **The Electric Field and Electric Forces** Maxwell said that electric and magnetic forces were due to the presence of the electric and magnetic field. In this figure, the electric force on Q_2 is due to the presence of the field at its location, $\vec{F}_{21} = Q_2 \vec{E}(\vec{r})$. There is a similar relationship for the magnetic force.

Maxwell was interested in developing a unified description of electric and magnetic phenomena. In his time, many of the basic ideas of the electric and magnetic force systems were known. The law for the electric interaction

between charged particles had been articulated in the period 1785 and 1791 by Coulomb. The force law between magnets and the force between moving charges and magnets was known and even Faraday's Law about the relationship of changing magnetic environments and electric currents was known. In fact, Faraday had already begun to describe magnetic and electric phenomena in a field like language. What Maxwell sought was an underlying mechanical basis for all the phenomena associated with electricity and magnetism. Reducing electricity and magnetism to a mechanical basis meant that he was looking for something to push or pull but it had to do so locally. He could not believe that fundamental phenomena could take place as an action at distance phenomena like gravity was thought at the time. In order to have a thing which could push or pull locally, he hypothesized the existence of a rather rich structure for the vacuum of space, whirling vortices in an ether that produced the electric and magnetic force. Thus not only did he seek a mechanical source for electric and magnetic phenomena, he developed a field theory basis for it. His picture of electric and magnetic forces was that they were mediated by fields, the electric, \vec{E} , and the magnetic, \vec{B} , fields. It was his basic idea that the correct description of electromagnetic phenomena required a locally causal dynamic. The idea was that not only did the charges generate the fields but the fields themselves responded to the local environment of the fields themselves. In addition, the forces experienced by the charges were because of the values of the fields at the place occupied by the charges, $\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$, where q is the charge in question and \vec{v} is its velocity.

In order to create the mechanical basis for the fields, Maxwell was forced to endow the ether with the correct mechanical properties of inertia and size to replicate the success of the earlier laws but now in context of a local mechanical model. The underlying idea was simple. Let's look at the simplest of the cases, Coulomb's Law. The situation is shown in Figures B.9 and B.10. A force on a charged particle took place as a two step process. A charge Q_1 is placed in empty unexcited space. This charge excites the ether next to it by creating vortices at its location. These vortices in turn excite neighboring vortices until space is full of whirling vortices. Each vortex is in dynamic equilibrium with its neighbors. There is a 'thing', the whirliness, which is a measure of the electric field at that point.

When a new charge, Q_2 , is located at some distance, \vec{r} , from the first charge, it detects the level of excitement of the local vortices and thus feels a corresponding force. The force is proportional to the charge Q_2 at that place and the amount of whirliness or electric field at that point.

The mechanical properties of the ether and its vortices determine how

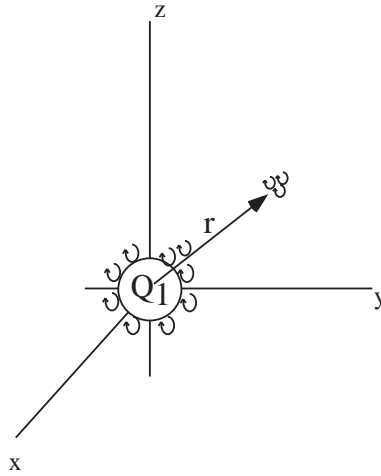


Figure B.9: **Maxwell's Vortices** Maxwell pictured the electric force as emerging in two steps. First any charged particle would excite vortices in the ether at its location. These vortices would excite other vortices nearby and so forth until all of space would fill with whirling vortices. In a sense, the whirliness of the vortices at any place was a measure of the strength of the electric field at that point.

the whirliness develops. This is set by the vortices inertia and size. These parameters for the mechanical properties of the vortices are then adjusted to accommodate Coulomb's Law.

In other words, Maxwell introduced local fields – a continuous quantity defined at all points in space and for all times – with a rule of dynamics to produce the electromagnetic forces. If an object experiences a force, there must be something at that place, the whirliness. In addition, the whirliness itself must be determined locally in both space and time. Let's go through the example of Coulomb's Law in a little more detail to see how this idea works.

The first problem is to reproduce the well known Coulomb's law of force for static situations. Coulomb's Law is an action at a distance description of interaction,

$$\vec{F}_{21} = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r_{12}^2} \frac{\vec{r}_{21}}{r_{12}} \quad (\text{B.12})$$

where \vec{r}_{12} is the separation between the charges. In order to simplify the discussion, let's place charge Q_1 at the origin. Since the force on the charge Q_2 is supposed to be $Q_2 \vec{E}(\vec{r})$, where \vec{r} is now the position at which Q_2 is

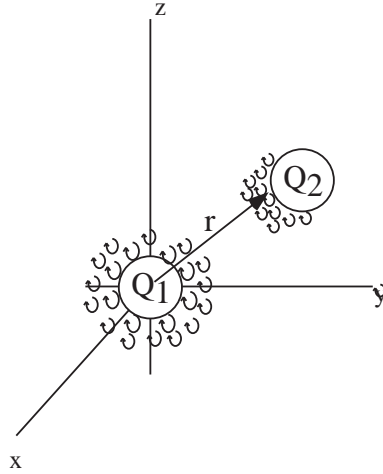


Figure B.10: **Vortices and the Electric Force** When a charged particle, Q_2 , is positioned, the particle detects the local amount of whirliness in the vortices of the ether. This generates the electric force in proportion to the charge and amount of whirliness at its location. The local whirliness is \vec{E} at \vec{r} .

located. For this case, we can identify the electric field as

$$\vec{E}(\vec{r}) = \frac{1}{4\pi\epsilon_0} \frac{Q_1 \vec{r}}{r^2 r} \quad (\text{B.13})$$

around a spherically symmetric charge placed at the origin. You will reproduce the static Coulomb's Law results with the electric field if you can make a local rule about how $\vec{E}(\vec{r})$ develops that reproduces this result. It should be clear that the hard part will be to reproduce the inverse square fall off with distance in the strength of the field.

In some sense, it is really not correct to say that Q_1 is the source of this field. The field is not attached to the charge. At any point, there is a field only if there is a field or a charge in the neighborhood. The field at some point, like all things, is to be determined locally. Maxwell used his whirling vortices of the ether to discover a rule for whirliness and how whirliness effected whirliness that recovers the characteristic the inverse square fall off with distance of Coulomb's Law. Like the stretched string, Section B.2, in which the transverse position of a place on the string is determined by the transverse position of the neighbors to that place, similarly here, the idea is to find the rule on how the field arranges itself and forget about the whirlies.

The following analysis reviews the process and becomes somewhat technical but the struggle to follow it is worth the effort.

Since the electric field is meant to produce a force, it must be a vector field, a directed quantity defined at every point in space and with a local rule for its construction. Basically you ask how much does the field change at a place because of what is there. For now, we are looking at a static case – no time change. But we can still ask about how the field varies as we change positions in space.

For a vector field such as the electric field since it is a vector field, you have a directed strength at each point in space and around each point you have directed strengths. At any point you can ask how much more “out pointy” these directed strengths become as you go from place to place. The analogy for our stretched string is that, at any place on the string, you can ask how “bendy” is the string. On a string, “bendiness” happens when that place differs from its neighbors. The string bends up when the place is lower than its neighbors and it bends down when it is higher. When there is no bend, that place on the string is at the average of its neighbors. In the static string it takes a force to maintain a bend in the string. Our case for the vector field case using “out pointiness” works in the same fashion. You can have “out pointiness” only if there are charges that are placed there, i. e. charge causes an outward directed field. Of course, we have to develop a definition, a measure, of “out pointy” and test it.

The measure of “out pointiness” is called the divergence and it is what you would have thought to define it as if you spent some time playing with the ideas of a vector field. At any point, find out how much the neighboring fields point away from where you are. That should indicate the “out pointiness”.

Since fluid flow is also a vector field it is worthwhile to think in terms of it. The vector field in this case is the velocity of the fluid. If at a point all the flow is uniform about you, you would not think of the field as becoming “out pointy”. On the other hand, if you were at a place like the drain, you would consider the surrounding flow to be “in pointy”, the opposite of “out pointy”. To be more quantitative, think of surrounding the place that you are interested in and measuring how much stuff flows in or out. By enclosing the point of interest with a surface, we can measure the incoming fluid by assessing how much stuff comes into any element of area on the surrounding surface and then adding the contribution to each part. In other words, surround the point with a surface. Cover it with elements of area, postage stamps. Each element of area has a normal vector that points either outward or inward, see Figure B.11. Choosing the outward normal, we are

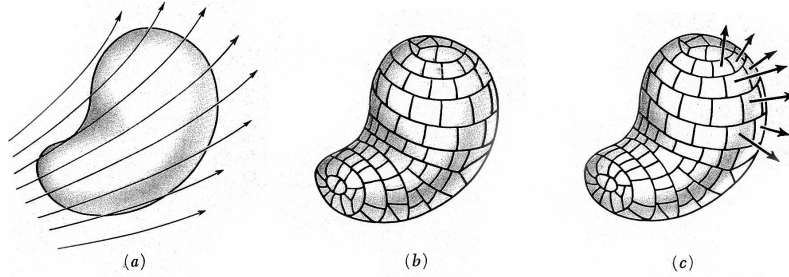


Figure B.11: **Construction of the Divergence** To find the divergence or “out pointiness” of a vector field at a point, surround the point with a surface, step (a). Cover the surface with small elements of area so that to all intents and purposes they can be considered flat. Each element of surface will now have a normal vector. Find the magnitude of the vector field at the surface to is along the surface. Add these magnitudes for each element of surface and the total is the divergence or “out pointiness” of the vector field at the point surrounded by the surface. Then shrink the volume surrounded to a point. For a fluid, applied to the velocity field, this tells the amount of fluid that goes into a point. This series of steps is encoded in the first part of Equation B.14 for the case of the electric field.

defining “out pointiness”, the amount of the velocity along the normal, is the flow through that area. Now do this for the each element of the entire surface and add up all the contribution from all the pieces. To reduce this analysis to a point, shrink the volume enclosed by the surrounding surface to zero. This same analysis holds for all vector fields. This construction at each point assesses the “out pointiness” of the neighborhood of the point and is called the divergence. Thus,

$$\text{Div}(\vec{\mathbf{E}}(\vec{\mathbf{r}})) = \lim_{V \rightarrow 0} \frac{\sum_{S \supset V} \vec{\mathbf{E}}(\vec{\mathbf{r}}') \cdot \Delta^2 \vec{\mathbf{S}}}{V} = \lim_{V \rightarrow 0} \frac{1}{\epsilon_0} \frac{Q_{\text{inside } V}}{V} = \frac{1}{\epsilon_0} \rho(\vec{\mathbf{r}}) \quad (\text{B.14})$$

where the first part is a mathematical statement of what is stated above but for the case of the electric field and the subsequent parts are the relationship with charge that is necessary to recover Coulomb’s Law, i. e. electric charge is the source of divergence.

Notice that this law, Equation B.14, says that for a static electric field there is divergence of the field only where there is charge. Yet the picture that we all have of the static electric field around an isolated point charge is a

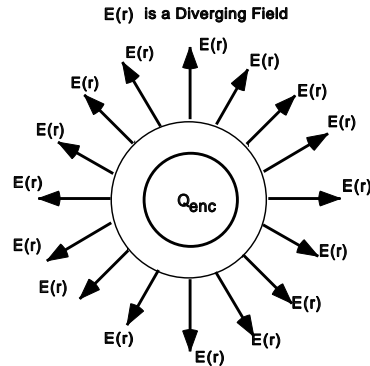


Figure B.12: ”**Outpointiness of the Electric Field**” A characteristic property of the electric field is that charge is the source of “outpointiness”. This is the idea that the electric field points away from nearby positive charges and toward nearby negative charges. This last example being negative outpointyness.

diverging field, the electric field points outward from the origin everywhere, see Figure B.12. How do we reconcile this?

Consider a point away from the isolated point charge. If a surface such as that shown in Figure B.11 is constructed area at nearer the charge is smaller whereas the area more distant is larger. In fact, the areas are in the ratio of the distances squared. Thus the field strength and the areas combine so that the net “out pointiness”, actually in pointiness, of the nearer surface balances the out pointiness of the far surface and the net is zero. Thus it is because the divergence is zero at places other than the charge that the field strength falls off with distance as $\frac{1}{r^2}$.

Another property that a vector field can manifest is rotation or curl. Again you develop a definition and test it. Here the idea is to follow a closed path around the point and see how much of the vector field follow the path. The electric field does not curl.

$$\text{Curl}(\vec{\mathbf{E}}(\vec{\mathbf{r}})) = \lim_{S \rightarrow 0} \frac{\sum_{p \in S} \vec{\mathbf{E}}(\vec{\mathbf{r}}') \cdot \Delta \vec{\mathbf{r}}'}{S} = 0 \quad (\text{B.15})$$

On the other hand, the magnetic field does curl. The magnetic field is the force experienced by a moving charged particle.

$$\vec{\mathbf{F}}_{\text{mag}} = Q\vec{\mathbf{v}} \times \vec{\mathbf{B}}(\vec{\mathbf{r}}) \quad (\text{B.16})$$

The magnetic field lines tend to wrap around their sources, the currents.

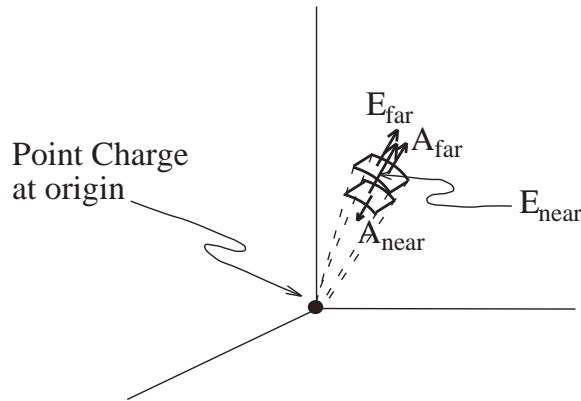


Figure B.13: **Divergence outside Charge** A characteristic property of the electric field is that charge is the source of “outpointiness”. This is the idea that the electric field points away from nearby positive charges and toward nearby negative charges. This last example being negative outpointiness.

$$\text{Curl}(\vec{\mathbf{B}}(\vec{\mathbf{r}})) = \lim_{S \rightarrow 0} \frac{\sum_{p \in S} \vec{\mathbf{B}}(\vec{\mathbf{r}}') \cdot \Delta \vec{\mathbf{r}}'}{S} = \lim_{S \rightarrow 0} \frac{1}{\mu_0} \frac{\mathbf{i}_{enc} p}{S} = \frac{1}{\mu_0} \vec{\mathbf{j}} \quad (\text{B.17})$$

and does not diverge

$$\text{Div}(\vec{\mathbf{B}}(\vec{\mathbf{r}})) = 0 \quad (\text{B.18})$$

Note that we have not added a time dependence. These are all static situations.

Maxwell insisted that the field was not established everywhere at once. It was made up of whirling vortices that pushed on each other. The rate at which the vortices could push was set by the parameters of the static theory. By endowing these whirling vortices with the correct properties to reproduce the laws of static electricity and magnetism, he found how to add a local set of rules for the time evolution of the fields. These are the full set of Maxwell’s equations including time dependence:

$$\text{Div}(\vec{\mathbf{E}}(\vec{\mathbf{r}}, t)) = \frac{1}{\epsilon_0} \rho(\vec{\mathbf{r}}, t) \quad (\text{B.19})$$

$$\text{Curl}(\vec{\mathbf{E}}(\vec{\mathbf{r}}, t)) = \frac{\partial \vec{\mathbf{B}}}{\partial t}(\vec{\mathbf{r}}, t) \quad (\text{B.20})$$

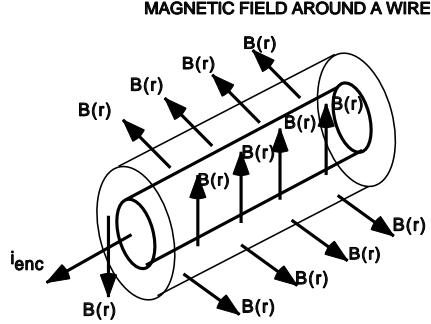


Figure B.14: **The Curl of the Magnetic Field** In contrast to the electric field, the magnetic field wraps around or curls around its sources, the currents in the problem.

$$\text{Div}(\vec{\mathbf{B}}(\vec{r}, t)) = 0 \quad (\text{B.21})$$

$$\text{Curl}(\vec{\mathbf{B}}(\vec{r}, t)) = \mu_0 \vec{\mathbf{j}}(\vec{r}, t) - \mu_0 \epsilon_0 \frac{\partial \vec{\mathbf{E}}}{\partial t}(\vec{r}, t) \quad (\text{B.22})$$

This is the standard format for these equations. For a discussion of the field dynamics, it is important to realize that only two of these equations are a dynamic, Equations B.20, and B.22. The other two equations, Equations B.19 and B.21, are what are called constraint equations; they control the pattern of the field but not the temporal evolution. It is apparent that the electromagnetic field is a much more complex field than the stretched string whose dynamic is Equation B.4. The vector nature of the field, the existence of constraints, and the sources, $\rho(\vec{r}, t)$ and $\vec{\mathbf{j}}(\vec{r}, t)$, obviously complicate the situation. We could have added external forces to the dynamic of the string but that would not have clarified the field nature of the string. Similarly, here we can discuss the electromagnetic field without the presence of $\rho(\vec{r}, t)$ and $\vec{\mathbf{j}}(\vec{r}, t)$. Rearranging and omitting the sources, the dynamical equations for the evolution of the electromagnetic field become

$$\frac{\partial \vec{\mathbf{E}}}{\partial t}(\vec{r}, t) = -\frac{1}{\mu_0 \epsilon_0} \text{Curl}(\vec{\mathbf{B}}(\vec{r}, t)) \quad (\text{B.23})$$

$$\frac{\partial \vec{\mathbf{B}}}{\partial t}(\vec{r}, t) = \text{Curl}(\vec{\mathbf{E}}(\vec{r}, t)) \quad (\text{B.24})$$

Identifying $\vec{\mathbf{E}}(\vec{r}, t)$ with the displacement field of the string, $y(x, t)$, and $\vec{\mathbf{B}}(\vec{r}, t)$ with the velocity field of the string, $v(x, t)$, we see that the electromagnetic dynamic is more complex but similar in structure.

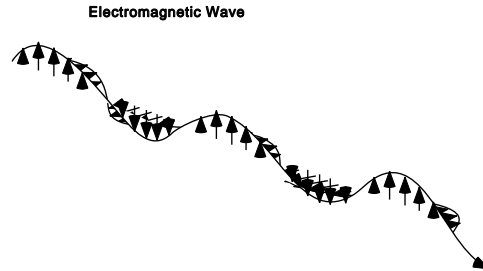


Figure B.15: **The Field Configuration for Light** Light is a traveling wave solution of Maxwell's Equations and is composed of propagating combination of electric and magnetic fields. The direction of flow of energy and momentum is along the normal to the plane of the oscillating electric and magnetic field vectors. In the figure the upward arrows represent the electric field and the perpendicular arrows are the magnetic field.

An important feature of the electromagnetic field that can be seen from the equations above is that, if you have an electric field in a localized region of space, finite somewhere but zero elsewhere like the pulse in the stretched string, the electric field will have a curl. Thus even if there are no charges or currents, this curl is the source of a developing magnetic field, Equation B.24. This is like the case in the string of the displacement producing a velocity field. As the new magnetic field grows which will also be localized and thus curled, it produces a reduction in the original electric field, Equation B.23. Thus the original field will start to reduce and there will be a growing magnetic field. This magnetic field will in turn change and produce a electric field. The relationship of the magnetic and electric fields is much like that of the velocity and displacement of the stretched string which produces traveling pulses, Section B.2. In fact, using Equations B.23 and B.24, in a region without charges or currents, the vacuum, you find that the electric and magnetic fields are a wavelike system and that a field configuration such as that shown in Figure B.15 produces a traveling wave that travels in the plane perpendicular to the plane of $\vec{\mathbf{E}}(\vec{\mathbf{r}}, t)$ and $\vec{\mathbf{B}}(\vec{\mathbf{r}}, t)$ with a speed

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

which dimensionally is a speed and the only dimensional factor in the dynamic. This is the same result that Maxwell discovered with his whirlies. Putting the values of μ_0 and ϵ_0 this is the speed of light. If it walks like a duck and quacks like a duck, it is a duck and thus Maxwell concluded that

light is the traveling wave solutions to the equations of electromagnetism.

It is important to realize that like in the stretched string which has only a transverse displacement and transverse velocity, the $\vec{\mathbf{E}}(\vec{\mathbf{r}}, t)$ and $\vec{\mathbf{B}}(\vec{\mathbf{r}}, t)$ fields are not traveling but only the disturbance – changes in the field configuration. It is also important to realize that the velocity of the disturbance does not depend on the field configuration. It only depends on the dynamic of the field. Another way that this is often said is that the velocity of propagation is a function only of the medium. Since the electromagnetic field operates in the vacuum of space, it is the properties of the vacuum that determine the speed with which light propagates. A difference for the electromagnetic travels with the travelers of the field of the stretched string is that in the string any distortion will produce simply related travelers but for the electromagnetic field there are configurations of the field that do not have simply related travelers.

We now understand the amplitude that was invented by Young and Fresnel, see Section ???. It is the electric field. The Fresnel construction is the general rule for the computation of the propagation of the light and holds for traveling waves of the electric and magnetic fields.