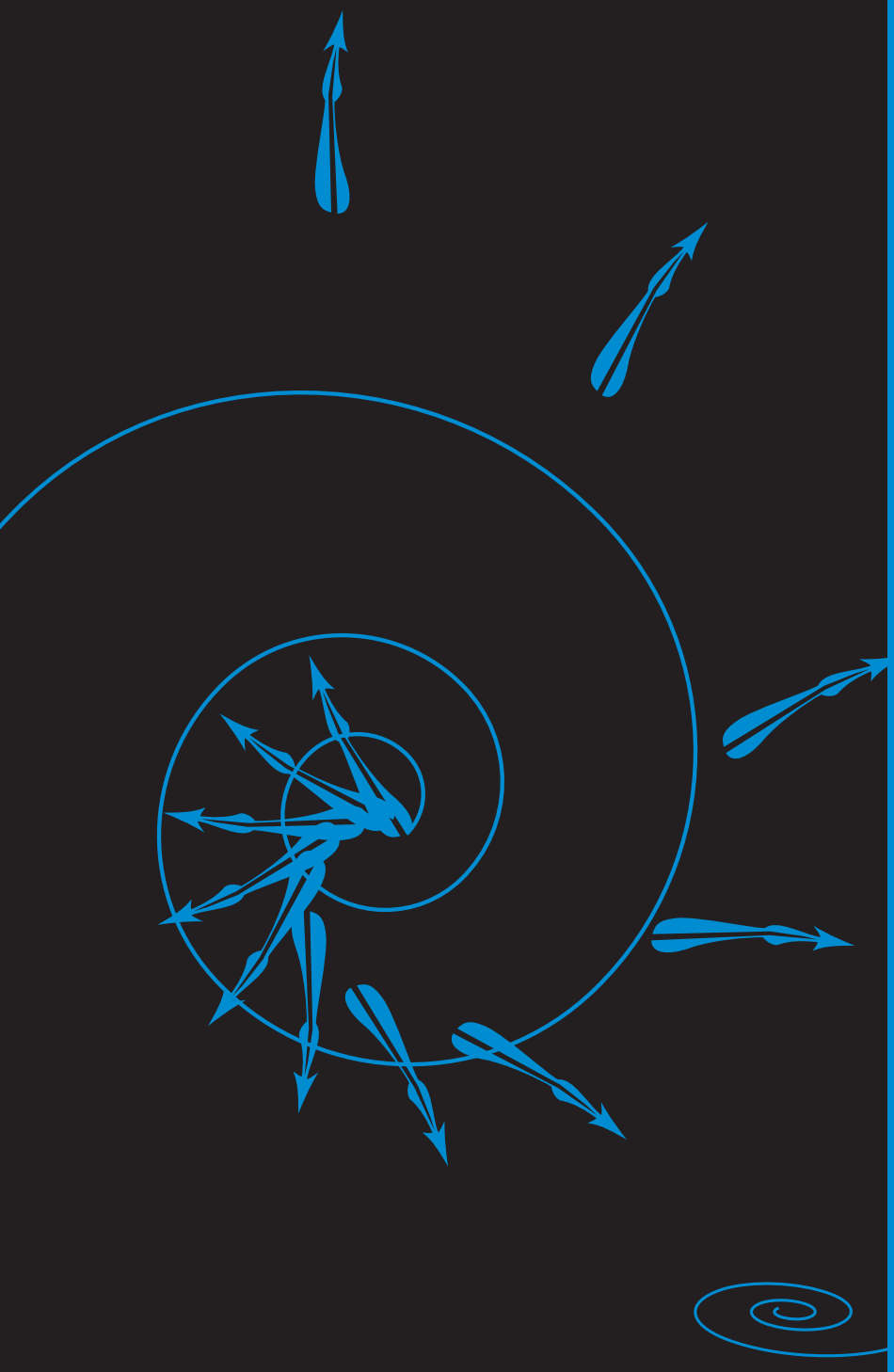
# Calculus 2000

**E. R. Huggins**
Dartmouth College



physics2000.com

## MKS Units

| | | |
|---|---|---|
| m = meters | kg = kilograms | s = seconds |
| N = newtons | J = joules | C = coulombs |
| T = tesla | F = farads | H = henrys |
| A = amperes | K = kelvins | mol = mole |

| | | |
|---|---|---|
| speed of light | $c$ | $3.00 \times 10^{8}\ \mathrm{m/s}$ |
| gravitational constant | $G$ | $6.67 \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$ |
| permittivity constant | $\varepsilon_0$ | $8.85 \times 10^{-12}\ \mathrm{F/m}$ |
| permeability constant | $\mu_0$ | $1.26 \times 10^{-6}\ \mathrm{H/m}$ |
| elementary charge | $e$ | $1.60 \times 10^{-19}\ \mathrm{C}$ |
| electron volt | eV | $1.60 \times 10^{-19}\ \mathrm{J}$ |
| electron rest mass | $m_e$ | $9.11 \times 10^{-31}\ \mathrm{kg}$ |
| proton rest mass | $m_p$ | $1.67 \times 10^{-27}\ \mathrm{kg}$ |
| Planck constant | $h$ | $6.63 \times 10^{-34}\ \mathrm{J \cdot s}$ |
| Planck constant / $2\pi$ | $\hbar$ | $1.06 \times 10^{-34}\ \mathrm{J \cdot s}$ |
| Bohr radius | $r_b$ | $5.29 \times 10^{-11}\ \mathrm{m}$ |
| Bohr magneton | $\mu_b$ | $9.27 \times 10^{-24}\ \mathrm{J/T}$ |
| Boltzmann constant | $k$ | $1.38 \times 10^{-23}\ \mathrm{J/K}$ |
| Avogadro constant | $N_A$ | $6.02 \times 10^{23}\ \mathrm{mol^{-1}}$ |
| universal gas constant | $R$ | $8.31\ \mathrm{J/mol \cdot K}$ |

## Powers of 10

| Power | Prefix | Symbol |
|---|---|---|
| $10^{12}$ | tera | T |
| $10^{9}$ | giga | G |
| $10^{6}$ | mega | M |
| $10^{3}$ | kilo | k |
| $10^{2}$ | hecto | h |
| $10^{-1}$ | deci | d |
| $10^{-2}$ | centi | c |
| $10^{-3}$ | milli | m |
| $10^{-6}$ | micro | μ |
| $10^{-9}$ | nano | n |
| $10^{-12}$ | pico | p |
| $10^{-15}$ | femto | f |

## Dimensions

| Quantity | Unit | | Equivalents | |
|---|---|---|---|---|
| Force | newton | N | J/m | $\mathrm{kg \cdot m/s^2}$ |
| Energy | joule | J | $\mathrm{N \cdot m}$ | $\mathrm{kg \cdot m^2/s^2}$ |
| Power | watt | W | J/s | $\mathrm{kg \cdot m^2/s^3}$ |
| Pressure | pascal | Pa | $\mathrm{N/m^2}$ | $\mathrm{kg/m \cdot s^2}$ |
| Frequency | hertz | Hz | cycle/s | $\mathrm{s^{-1}}$ |
| Electric charge | coulomb | C | | $\mathrm{A \cdot s}$ |
| Electric potential | volt | V | J/C | $\mathrm{kg \cdot m^2/A \cdot s^3}$ |
| Electric resistance | ohm | Ω | V/A | $\mathrm{kg \cdot m^2/A^2 \cdot s^3}$ |
| Capacitance | farad | F | C/V | $\mathrm{A^2 \cdot s^4/kg \cdot m^2}$ |
| Magnetic field | tesla | T | $\mathrm{N \cdot s/C \cdot m}$ | $\mathrm{kg/A \cdot s^2}$ |
| Magnetic flux | weber | Wb | $\mathrm{T \cdot m^2}$ | $\mathrm{kg \cdot m^2/A \cdot s^2}$ |
| Inductance | henry | H | $\mathrm{V \cdot s/A}$ | $\mathrm{kg \cdot m^2/A^2 \cdot s^2}$ |

# Calculus 2000

# A Physics-Based Calculus Text

*When developing a physics curriculum, a major concern is the mathematical background of the student. The **Physics 2000** text was developed teaching premedical students who were supposed to have had one semester of calculus. Because many of the students had taken calculus several years previously, and had forgotten much of it, the physics text used strobe photographs and the computer to carefully introduce the calculus concepts such as velocity, acceleration, and the limiting process. By the time we got to electricity and magnetism in **Part 2** of **Physics 2000** we relied on the student being familiar with the basic steps of differentiation and integration.*

*For students who have forgotten much of their calculus course, or those who have not had calculus but wish to study the Physics 2000 text, we have written Chapter 1 of **Calculus 2000**. This chapter not only covers all the calculus needed for the **Physics 2000** text, but is also carefully integrated with it. The chapter is much shorter than the typical introductory calculus text because the basic calculus concepts are discussed in the physics text and the calculus chapter only has to deal with the formalism.*

*After the introductory courses, the standard physics curriculum repeatedly goes over the same topics at successively higher mathematical levels. A typical example is the subject of electricity and magnetism which is taught using integral equations in the introductory course, using differential operators in an upper level undergraduate course, and then taught all over again in a graduate level course. In each of the courses it takes a while for the student to realize that this is just the same old subject dressed up in new math.*

*With Chapters 2 through 13 of the **Calculus 2000**, we introduce a different approach. We take the topics that we have already introduced in **Physics 2000**, and show how these topics can be handled in progressively more sophisticated mathematical ways. Once we have introduced the mathematical concepts of gradient, divergence and curl in the calculus text, we can turn the integral form of Maxwell's equation into a wave equation for electric and magnetic fields. With the introduction of the Laplacian and complex variables, we can study Schrödinger's equation and begin to solve for the hydrogen wave patterns discussed in Chapter 38 of the physics text.*

*Beyond seeing the same topics in a more sophisticated way, the student finds that new insights can result from the advanced mathematical approach. Chapter 10 of the calculus text is a short chapter less than two pages. But it is one of the most significant chapters in the text. For there we see that* ***Maxwell's equations for electric and magnetic fields require that electric charge be conserved****. This intimate connection between a conservation law and field theory becomes clear when we have sufficiently powerful mathematical tools to handle the theory.*

*The physics text began its discussion of vector fields in Chapter 23, using the velocity field as its first example. We did that because it is much easier to visualize the familiar flow of water than the abstract concept of an electric field. We saw that the streamlines in fluid flow went over to electric field lines, Gauss's law in fluid theory simply represented the incompressibility of the fluid, and Bernoulli's equation provided an introduction to the concept of voltage and potential.*

*However our discussion of electric and magnetic fields, particularly in this calculus text, go way beyond the simple fluid flow topics we introduced in the physics text. In the last two chapters of the calculus text, we turn the tables and apply to fluid theory the mathematical techniques we learned studying electricity and magnetism. In Chapter 12 we discuss the concept of vorticity which is the curl of the velocity field. The focus is to develop an intuitive understanding of the nature of vorticity and the role it plays in fluid flows, particularly vortices and vortex rings.*

*Chapter 13 is an introduction to fluid dynamics. The idea is to bring our discussion of the velocity field up to the same level as our treatment of electric and magnetic fields. We begin with a derivation of the Navier-Stokes equation which applies to constant density viscous fluids. This is then converted into an equation for vortex dynamics from which we derive an extended form of the famous Helmholtz equation. We then use that to derive the well known properties of vortex motion such as the so called Magnus force, and discuss the experiment Rayfield and Reif used to measure the circulation and core diameter of quantized vortices in superfluid helium.*

# Table of Contents

## CHAPTER 11 SCALAR AND VECTOR POTENTIALS



## CHAPTER 12 VORTICITY



## CHAPTER 13 INTRODUCTION TO FLUID DYNAMICS



## CH 13 APP 1 COMPONENT NOTATION



## CH 13 APP 2 VORTEX CURRENTS

***FORMULARY***



***INDEX***

***BACK COVER***

# Index

## Symbols



## A



## B



## C

## D

## *E*

## F



## G

## H



## I

## L



## M

## *N*



## *O*



## *P*

## Q



## R

## s

## T



## U



## V

## *w*

## x



## z

# Calculus Chapter 1

# Introduction to Calculus

**This first chapter covers all the calculus that is needed for the Physics 2000 text. The remaining chapters allow students to look at the physics from an advanced mathematical point of view.**

*This chapter, which replaces Chapter 4 in Physics 2000, is intended for students who have not had calculus, or as a calculus review for those whose calculus is not well remembered. If, after reading part way through this chapter, you feel your calculus background is not so bad after all, go back to Chapter 4 in Physics 2000, study the derivation of the constant acceleration formulas beginning on page 4-8, and work the projectile motion problems in the appendix to Chapter 4. Those who study all of this introduction to calculus should then proceed to the projectile motion problems in the appendix to Chapter 4 of the Physics text.*

## LIMITING PROCESS

In Chapter 3 of Physics 2000, we used strobe photographs to define velocity and acceleration vectors. The basic approach was to turn up the strobe flashing rate, as we did in going from Figure (3-3) to (3-4) shown below. We turned the rate up until all the kinks are clearly visible and the successive displacement vectors give a reasonable description of the motion. We did not turn the flashing rate too high, for the practical reason that the displacement vectors became too short for accurate work.

In our discussion of instantaneous velocity we conceptually turned the strobe all the way up as illustrated in Figures (2-22a) through (2-22d), redrawn here in Figure (1). In these figures, we initially see a fairly large change in $\vec{v}_0$ as the strobe rate is increased and $\Delta t$ reduced. But then the change becomes smaller, and it looks as if we are approaching some final value of $\vec{v}_0$ that does not depend on the size of $\Delta t$, provided $\Delta t$ is small enough. It looks as if we have come close to the final value in Figure (1c).

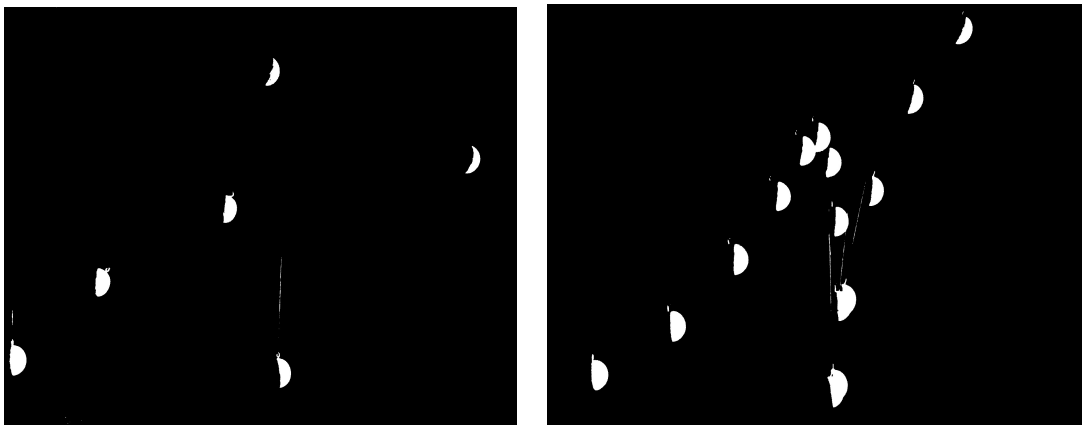
The progression seen in Figure (1) is called a ***limiting process***. The idea is that there really is some true value of $\vec{v}_0$ which we have called the ***instantaneous velocity***, and that we approach this true value for sufficiently small values of $\Delta t$. This is a calculus concept, and in the language of calculus, we are ***taking the limit as $\Delta t$ goes to zero***.

## THE UNCERTAINTY PRINCIPLE

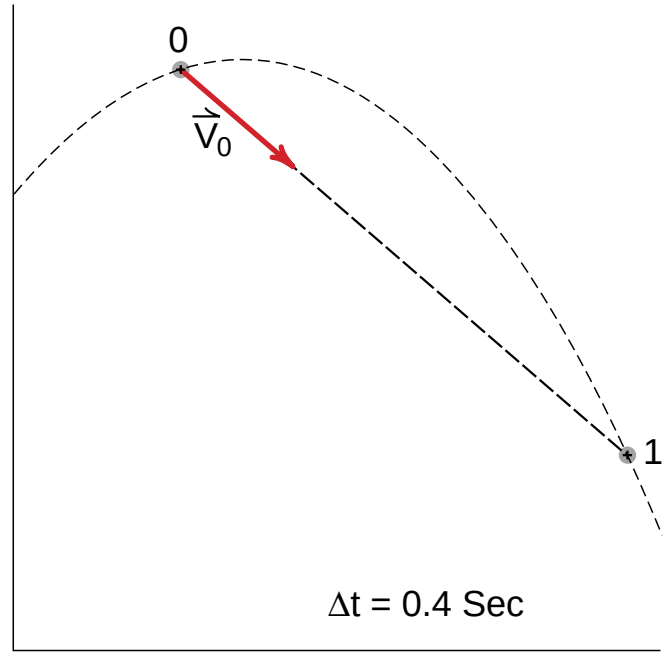
For over 200 years, from the invention of calculus by Newton and Leibnitz until 1924, the limiting process and the resulting concept of instantaneous velocity was one of the cornerstones of physics. Then in 1924 Werner Heisenberg discovered what he called the ***uncertainty principle*** which places a limit on the accuracy of experimental measurements.

Heisenberg discovered something very new and unexpected. He found that the act of making an experimental measurement unavoidably affects the results of an experiment. This had not been known previously because the effect on large objects like golf balls is undetectable. But on an atomic scale where we study small systems like electrons moving inside an atom, the effect is not only observable, it can dominate our study of the system.

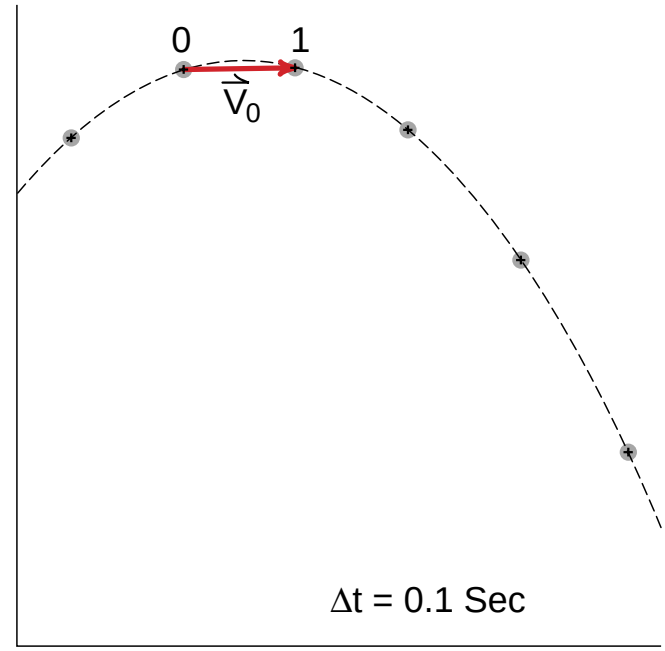
One particular consequence of the uncertainly principle is that the more accurately we measure the position of an object, the more we disturb the motion of the object. This has an immediate impact on the concept of instantaneous velocity. If we turn the strobe all the way up, reduce $\Delta t$ to zero, we are in effect trying to measure the position of the object with infinite precision. The consequence would be an infinitely big disturbance of the motion of the object we are studying. If we actually could turn the strobe all the way up, we would destroy the object we were trying to study.



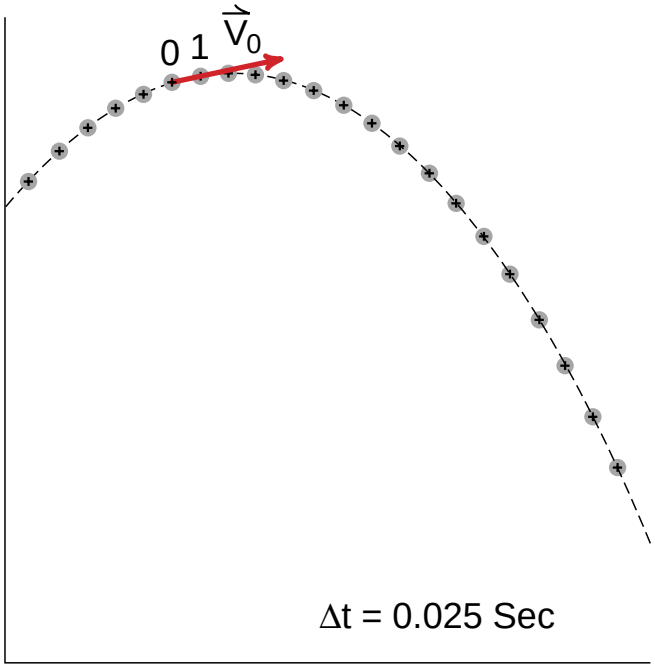
**Figures 3-3 and 3-4 from Physics 2000**
***Strobe photographs of a moving object. In the first photograph, the time between flashes is so long that the motion is difficult to understand. In the second, the time between flashes was reduced and the motion is more easily understood.***



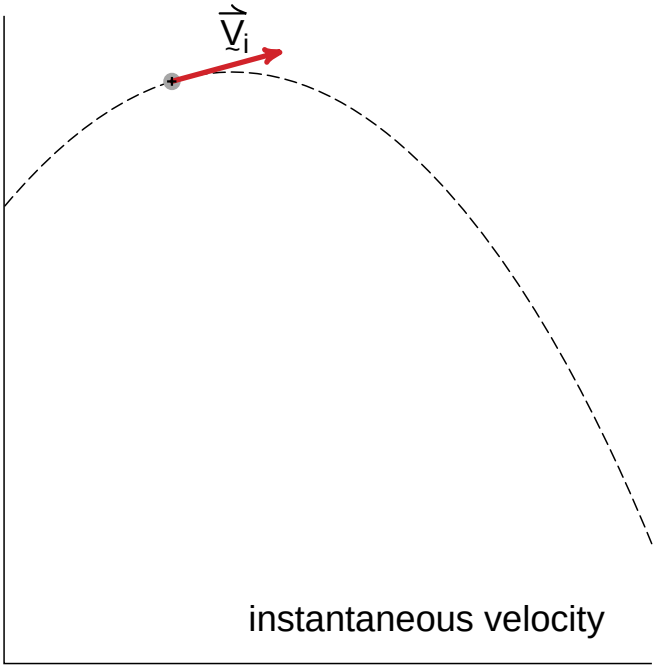

*(a)*




*(c)*




*(b)*




*(d)*

**Figure 1**
***Transition to instantaneous velocity. As we reduce $\Delta t$, there is less and less change in the vector $\vec{V}_0$. It looks as if we are approaching an exact final value.***

## Uncertainty Principle on a Larger Scale

It turns out that the uncertainty principle can have a significant impact on a larger scale of distance than the atomic scale. Suppose, for example, we constructed a chamber that is 1 cm on each side, and wished to study the projectile motion of an electron inside. Using Galileo's idea that objects of different mass fall at the same rate, we would expect that the motion of the electron projectile should be the same as more massive objects. If we took a strobe photograph of the electron's motion, we would expect to get results like those shown in Figure (2). This figure represents projectile motion with an acceleration $g = 980$ cm/sec$^2$ and $\Delta t = .01$sec, as the reader can easily check.

When we study the uncertainty principle in Chapter 40 of the Physics text, we will see that a measurement which is accurate enough to show that position (2) is below position (1), could disturb the electron enough to reverse its direction of motion. The next position measurement could find the electron over where we drew position (3), or back where we drew position (0), or anywhere in the region in between. As a result we could not even determine what direction the electron is moving. This uncertainty would not be the result of a sloppy experiment, it is the best we can do with the most accurate and delicate measurements possible.

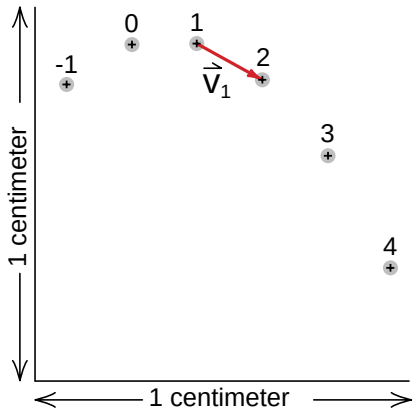



**Figure 2**
***Hypothetical electron projectile motion experiment. The uncertainty principle tells us that such an experiment cannot lead to predictable results.***

The uncertainty principle has had a significant impact on the way physicists think about motion. Because we now know that the measuring process affects the results of the measurement, we see that it is essential to provide experimental definitions to any physical quantity we wish to study. A conceptual definition, like turning the strobe all the way up to define instantaneous velocity, can lead to fundamental inconsistencies.

Even an experimental definition like our strobe definition of velocity can lead to inconsistent results when applied to something like the electron in Figure (2). But these inconsistencies are real. Their existence is telling us that the very concept of velocity is beginning to lose meaning for these small objects.

On the other hand, the idea of the limiting process and instantaneous velocity is very convenient when applied to larger objects where the effects of the uncertainty principle are not detectable. In this case we can apply all the mathematical tools of calculus developed over the past 250 years. The status of instantaneous velocity has changed from a basic concept to a useful mathematical tool. Those problems for which this mathematical tool works are called problems in ***classical physics***; those problems for which the uncertainty principle is important, are in the realm of what we call ***quantum physics***.

## CALCULUS DEFINITION OF VELOCITY

With the above perspective on the physical limitations of the limiting process, we can now return to the main topic of this chapter—the use of calculus in defining and working with velocity and acceleration.

In discussing the limiting process in calculus, one traditionally uses a special set of symbols which we can understand if we adopt the notation shown in Figure (3). In that figure we have drawn the coordinate vectors $\vec{R}_i$ and $\vec{R}_{i+1}$ for the i th and (i + 1) positions of the object. We are now using the symbol $\overrightarrow{\Delta R}_i$ to represent the displacement of the ball during the i to i+1 interval. The vector equation for $\overrightarrow{\Delta R}_i$ is

$$\overrightarrow{\Delta R}_i = \vec{R}_{i+1} - \vec{R}_i \qquad (1)$$

In words, Equation (1) tells us that $\overrightarrow{\Delta R}_i$ is the change, during the time $\Delta t$, of the position vector $\vec{R}$ describing the location of the ball.

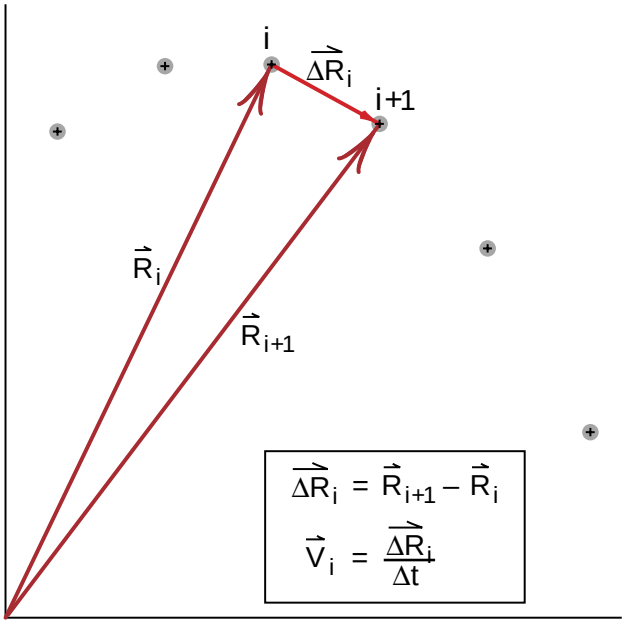



**Figure 3**
***Definitions of $\overrightarrow{\Delta R}_i$ and $\vec{v}_i$.***

The velocity vector $\vec{v}_i$ is now given by

$$\vec{v}_i = \frac{\overrightarrow{\Delta R}_i}{\Delta t} \qquad (2)$$

This is just our old strobe definition $\vec{v}_i = \vec{s}_i/\Delta t$, but using a notation which emphasizes that the displacement $\vec{s}_i = \overrightarrow{\Delta R}_i$ is the ***change in position*** that occurs during the time $\Delta t$. The Greek letter $\Delta$ (delta) is used both to represent the idea that the quantity $\overrightarrow{\Delta R}_i$ or $\Delta t$ is small, and to emphasize that both of these quantities change as we change the strobe rate.

The limiting process in Figure (1) can be written in the form

$$\underline{\vec{v}}_i \equiv \lim_{\Delta t \to 0} \frac{\overrightarrow{\Delta R}_i}{\Delta t} \qquad (3)$$

where the word "limit" with $\Delta t \to 0$ underneath, is to be read as "limit as $\Delta t$ goes to zero". For example we would read Equation (3) as "***the instantaneous velocity*** $\underline{\vec{v}}_i$ ***at position*** i ***is the limit, as*** $\Delta t$ ***goes to zero, of the ratio*** $\overrightarrow{\Delta R}_i/\Delta t$ ."

For two reasons, Equation (3) is not quite yet in standard calculus notation. One is that in calculus, only the limiting value, in this case, the instantaneous velocity, is considered to be important. Our strobe definition $\vec{v}_i = \overrightarrow{\Delta R}_i/\Delta t$ is only a step in the limiting process. Therefore when we see the vector $\vec{v}_i$, we should assume that it is the limiting value, and no special symbol like the underline is used. For this reason we will drop the underline and write

$$\vec{v}_i = \lim_{\Delta t \to 0} \frac{\overrightarrow{\Delta R}_i}{\Delta t} \qquad (3a)$$

The second change deals with the fact that when $\Delta t$ goes to zero we need an infinite number of time steps to get through our strobe photograph, and thus it is not possible to locate a position by counting time steps. Instead we measure the time t that has elapsed since the beginning of the photograph, and use that time to tell us where we are, as illustrated in Figure (4). Thus instead of using $\vec{v}_i$ to represent the velocity at position i, we write $\vec{v}(t)$ to represent the velocity at time t. Equation (3) now becomes

$$\vec{v}(t) = \lim_{\Delta t \to 0} \frac{\overrightarrow{\Delta R}(t)}{\Delta t} \qquad (3b)$$

where we also replaced $\overrightarrow{\Delta R}_i$ by its value $\overrightarrow{\Delta R}(t)$ at time t.

Although Equation (3b) is in more or less standard calculus notation, the notation is clumsy. It is a pain to keep writing the word "limit" with a $\Delta t \to 0$ underneath. To streamline the notation, we replace the Greek letter $\Delta$ with the English letter d as follows

$$\boxed{\lim_{\Delta t \to 0} \frac{\overrightarrow{\Delta R}(t)}{\Delta t} \equiv \frac{d\vec{R}(t)}{dt}} \qquad (4)$$

(The symbol $\equiv$ means *defined equal to*.) To a mathematician, the symbol $d\vec{R}(t)/dt$ is just shorthand notation for the limiting process we have been describing. But to a physicist, there is a different, more practical meaning. Think of dt as a short $\Delta t$, short enough so that the limiting process has essentially occurred, but not too short to see what is going on. In Figure (1), a value of dt less than .025 seconds is probably good enough.

If dt is small but finite, then we know exactly what the $d\vec{R}(t)$ is. It is the small but finite displacement vector at the time t. It is our old strobe definition of velocity, with the added condition that dt is such a short time interval that the limiting process has occurred. From this point of view, dt is a real time interval and $d\vec{R}(t)$ a real vector, which we can work with in a normal way. The only thing special about these quantities is that when we see the letter d instead of $\Delta$, we must remember that a limiting process is involved. In this notation, the calculus definition of velocity is

$$\boxed{\vec{v}(t) = \frac{d\vec{R}(t)}{dt}} \qquad (5)$$

where $\vec{R}(t)$ and $\vec{v}(t)$ are the particle's coordinate vector and velocity vector respectively, as shown in Figure (5). Remember that this is just fancy shorthand notation for the limiting process we have been describing.

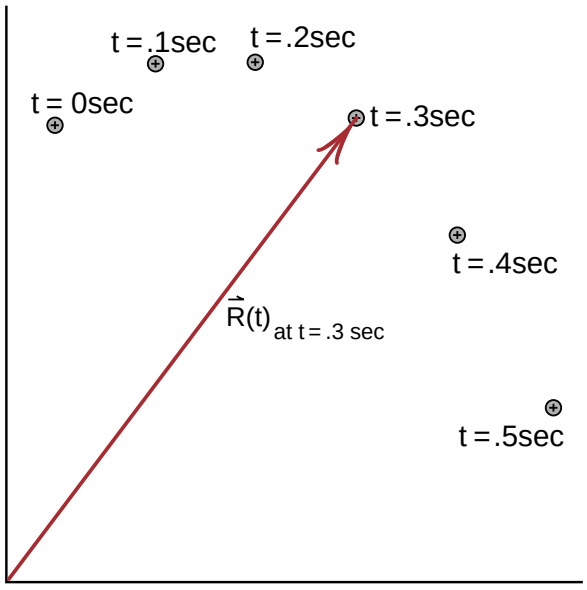



**Figure 4**
***Rather than counting individual images, we can locate a position by measuring the elapsed time t. In this figure, we have drawn the displacement vector $\vec{R}(t)$ at time t = .3 sec.***

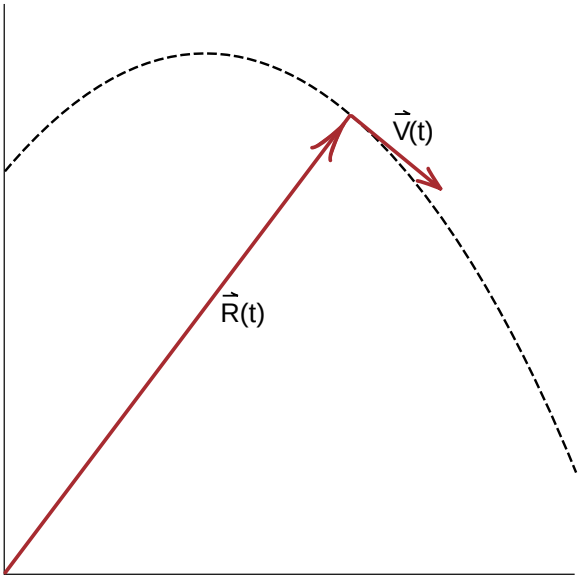



**Figure 5**
***Instantaneous position and velocity at time t.***

## ACCELERATION

In the analysis of strobe photographs, we defined both a velocity vector $\vec{v}$ and an acceleration vector $\vec{a}$. The definition of $\vec{a}$, shown in Figure (2-12) reproduced here in Figure (6), was

$$\vec{a}_i \equiv \frac{\vec{v}_{i+1} - \vec{v}_i}{\Delta t} \qquad (6)$$

In our graphical work we replaced $\vec{v}_i$ by $\vec{s}_i/\Delta t$ so that we could work directly with the displacement vectors $\vec{s}_i$ and experimentally determine the behavior of the acceleration vector for several kinds of motion.

Let us now change this graphical definition of acceleration over to a calculus definition, using the ideas just applied to the velocity vector. First, assume that the ball reached position i at time t as shown in Figure (6). Then we can write

$$\vec{v}_i = \vec{v}(t)$$

$$\vec{v}_{i+1} = \vec{v}(t+\Delta t)$$

to change the time dependence from a count of strobe flashes to the continuous variable t. Next, define the vector $\Delta\vec{v}(t)$ by

$$\Delta\vec{v}(t) \equiv \vec{v}(t+\Delta t) - \vec{v}(t) \quad \left( = \vec{v}_{i+1} - \vec{v}_i \right) \qquad (7)$$

We see that $\Delta\vec{v}(t)$ is the change in the velocity vector as the time advances from t to $t+\Delta t$. The strobe definition of $\vec{a}_i$ can now be written

$$\vec{a}(t)\left(\begin{matrix}strobe\\definition\end{matrix}\right) = \frac{\vec{v}(t+\Delta t) - \vec{v}(t)}{\Delta t} \equiv \frac{\Delta\vec{v}(t)}{\Delta t} \qquad (8)$$

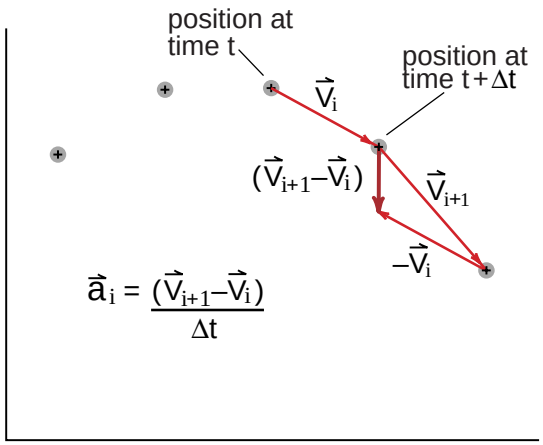



**Figure 6**
***Experimental definition of the acceleration vector.***

Now go through the limiting process, turning the strobe up, reducing $\Delta t$ until the value of $\vec{a}(t)$ settles down to its limiting value. We have

$$\vec{a}(t)\left(\begin{matrix}calculus\\definition\end{matrix}\right) = \lim_{\Delta t \to 0} \frac{\vec{v}(t+\Delta t) - \vec{v}(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta\vec{v}(t)}{\Delta t} \qquad (9)$$

Finally use the shorthand notation d/dt for the limiting process:

$$\boxed{\vec{a}(t) = \frac{d\vec{v}(t)}{dt}} \qquad (10)$$

Equation (10) does not make sense unless you remember that it is notation for all the ideas expressed above. Again, physicists think of dt as a short but finite time interval, and $d\vec{v}(t)$ as the small but finite change in the velocity vector during the time interval dt. It's our strobe definition of acceleration with the added requirement that $\Delta t$ is short enough that the limiting process has already occurred.

### Components

Even if you have studied calculus, you may not recall encountering formulas for the derivatives of vectors, like $d\vec{R}(t)/dt$ and $d\vec{v}(t)/dt$ which appear in Equations (5) and (10). To bring these equations into a more familiar form where you can apply standard calculus formulas, we will break the vector Equations (5) and (10) down into component equations.

In the chapter on vectors, we saw that any vector equation like

$$\vec{A} = \vec{B} + \vec{C} \qquad (11)$$

is equivalent to the three component equations

$$\begin{aligned} A_x &= B_x + C_x \\ A_y &= B_y + C_y \\ A_z &= B_z + C_z \end{aligned} \qquad (12)$$

The advantage of the component equations was that they are simply numerical equations and no graphical work or trigonometry is required.

The limiting process in calculus does not affect the decomposition of a vector into components, thus Equation (5) for $\vec{v}(t)$ and Equation (10) for $\vec{a}(t)$ become

$$\vec{v}(t) = d\vec{R}(t)/dt \qquad (5)$$

$$v_x(t) = dR_x(t)/dt \qquad (5a)$$

$$v_y(t) = dR_y(t)/dt \qquad (5b)$$

$$v_z(t) = dR_z(t)/dt \qquad (5c)$$

and

$$\vec{a}(t) = d\vec{v}(t)/dt \qquad (10)$$

$$a_x(t) = dv_x(t)/dt \qquad (10a)$$

$$a_y(t) = dv_y(t)/dt \qquad (10b)$$

$$a_z(t) = dv_z(t)/dt \qquad (10c)$$

Often we use the letter x for the x coordinate of the vector $\vec{R}$ and we use y for $R_y$ and z for $R_z$. With this notation, Equation (5) assumes the shorter and perhaps more familiar form

$$v_x(t) = dx(t)/dt \qquad (5a')$$

$$v_y(t) = dy(t)/dt \qquad (5b')$$

$$v_z(t) = dz(t)/dt \qquad (5c')$$

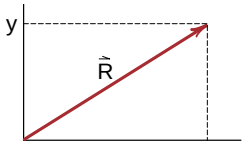


**Figure 7**

At this point the notation has become deceptively short. You now have to remember that x(t) stands for the x coordinate of the particle at a time t.

We have finally boiled the notation down to the point where it would be familiar in any calculus course. If we restrict our attention to one dimensional motion along the x axis, then all we have to concern ourselves with are the x component equations

$$v_x(t) = \frac{dx(t)}{dt}$$

$$a_x(t) = \frac{dv_x(t)}{dt} \qquad (10a)$$

## INTEGRATION

When we worked with strobe photographs, the photograph told us the position $\vec{R}(t)$ of the ball as time passed. Knowing the position, we can then use Equation (5) to calculate the ball's velocity $\vec{v}(t)$ and then Equation (10) to determine the acceleration $\vec{a}(t)$. In general, however, we want to go the other way, and predict the motion from a knowledge of the acceleration. For example, imagine that you were in Galileo's position, hired by a prince to predict the motion of cannonballs. You know that a cannonball should not be much affected by air resistance, thus the acceleration throughout its trajectory should be the constant gravitational acceleration $\vec{g}$. You know that $\vec{a}(t) = \vec{g}$. How then do you use that knowledge in Equations (5) and (10) to predict the motion of the ball?

The answer is that you cannot with the equations in their present form. The equations tell you how to go from $\vec{R}(t)$ to $\vec{a}(t)$, while to predict motion you need to go the other way, from $\vec{a}(t)$ to $\vec{R}(t)$. The topic of this section is to see how to reverse the directions in which we use our calculus equations. Equations (5) and (10) involve the process called ***differentiation***. We will see that when we go the other way the reverse of differentiation is a process called ***integration***. We will see that integration is a simple concept, but a process that is sometimes hard to perform without the aid of a computer.

## Prediction of Motion

In our earlier discussion, we have used strobe photographs to analyze motion. Let us see what we can learn from such a photograph for predicting motion. Figure (8) is our familiar projectile motion photograph showing the displacement $\vec{s}$ of a ball during the time the ball traveled from a position labeled (0) to the position labeled (4). If the ball is now at position (0) and each of the images is (.1) seconds apart, then the vector $\vec{s}$ tells us where the ball will be at a time of (.4) seconds from now. If we can predict $\vec{s}$, we can predict the motion of the ball. The general problem of predicting the motion of the ball is to be able to calculate $\vec{s}(t)$ for any time t.

From Figure (8) we see that $\vec{s}$ is the vector sum of the individual displacement vectors $\vec{s}_1$, $\vec{s}_2$, $\vec{s}_3$ and $\vec{s}_4$

$$\vec{s} = \vec{s}_1 + \vec{s}_2 + \vec{s}_3 + \vec{s}_4 \qquad (11)$$

We can then use the fact that $\vec{s}_1 = \vec{v}_1\Delta t$, $\vec{s}_2 = \vec{v}_2\Delta t$, etc. to get

$$\vec{s} = \vec{v}_1\Delta t + \vec{v}_2\Delta t + \vec{v}_3\Delta t + \vec{v}_4\Delta t \qquad (12)$$

Rather than writing out each term, we can use the ***summation sign*** $\Sigma$ to write

$$\vec{s} = \sum_{i=1}^{4} \vec{v}_i\Delta t \qquad (12a)$$

Equation (12) is approximate in that the $\vec{v}_i$ are approximate (strobe) velocities, not the instantaneous velocities we want for a calculus discussion. In Figure (9) we improved the situation by cutting $\Delta t$ to 1/4 of its previous value, giving us four times as many images and more accurate velocities $\vec{v}_i$.

We see that the displacement $\vec{s}$ is now the sum of 16 vectors

$$\vec{s} = \vec{s}_1 + \vec{s}_2 + \vec{s}_3 + \ldots + \vec{s}_{15} + \vec{s}_{16} \qquad (13)$$

Expressing this in terms of the velocity vectors $\vec{v}_1$ to $\vec{v}_{16}$ we have

$$\vec{s} = \vec{v}_1\Delta t + \vec{v}_2\Delta t + \vec{v}_3\Delta t + \ldots + \vec{v}_{15}\Delta t + \vec{v}_{16}\Delta t \qquad (14)$$

or using our more compact notation

$$\vec{s} = \sum_{i=1}^{16} \vec{v}_i\Delta t \qquad (14a)$$

While Equation (14) for $\vec{s}$ looks quite different than Equation (12)—the sum of sixteen vectors instead of four—the displacement vectors $\vec{s}$ in the two cases are exactly the same. Adding more intermediate images did not change where the ball was located at the time of t = .4 seconds. In going from Equation (12) to (14), what has changed as a result of shortening the time step $\Delta t$, is that ***the individual velocity vectors $\vec{v}_i$ become more nearly equal to the instantaneous velocity of the ball at each image.***

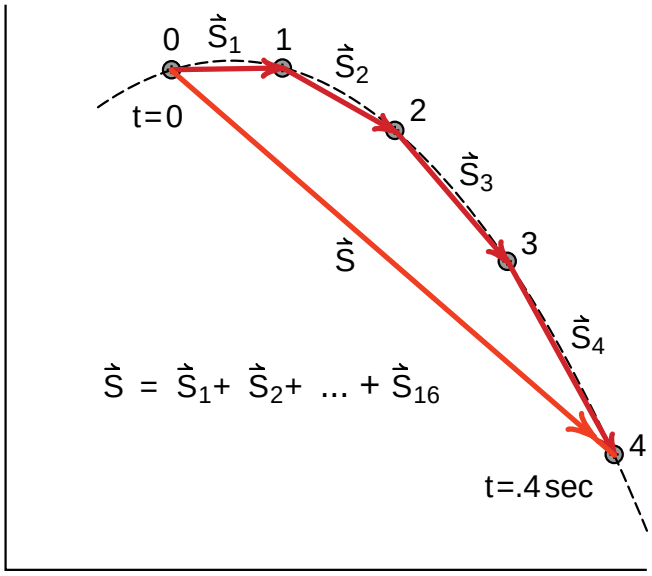



**Figure 8**
***To predict the total displacement $\vec{s}$, we add up the individual displacements $\vec{s}_i$.***

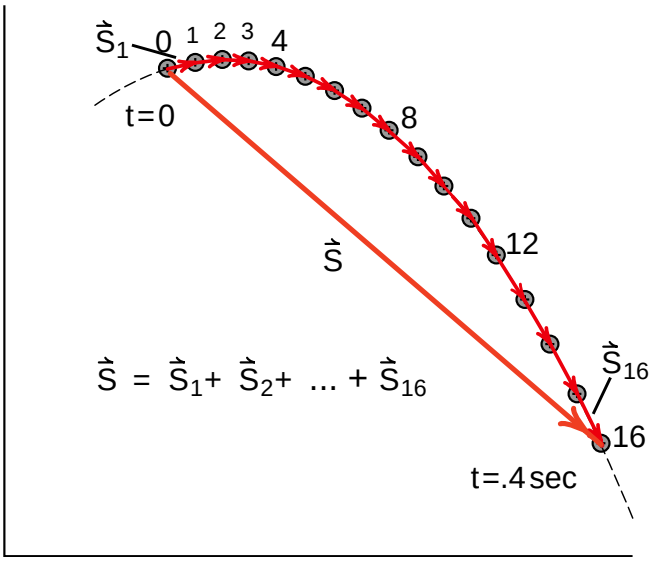



**Figure 9**
***With a shorter time interval, we add up more displacement vectors to get the total displacement $\vec{s}$.***

If we reduced $\Delta t$ again by another factor of 1/4, so that we had 64 images in the interval t = 0 to t = .4 sec, the formula for $\vec{s}$ would become

$$\vec{s} = \sum_{i=1}^{64} \vec{v}_i \Delta t \qquad (15a)$$

where now the $\vec{v}_i$ are still closer to representing the ball's instantaneous velocity. The more we reduce $\Delta t$, the more images we include, the closer each $\vec{v}_i$ comes to the instantaneous velocity $\vec{v}(t)$. While adding more images gives us more vectors that we have to add up to get the total displacement $\vec{s}$, there is very little change in our formula for $\vec{s}$. If we had a million images, we would simply write

$$\vec{s} = \sum_{i=1}^{1000000} \vec{v}_i \Delta t \qquad (16a)$$

In this case the $\vec{v}_i$ would be physically indistinguishable from the instantaneous velocity $\vec{v}(t)$. We have essentially reached a calculus limit, but we have problems with the notation. It is clearly inconvenient to label each $\vec{v}_i$ and then count the images. Instead we would like notation that involves the instantaneous velocity $\vec{v}(t)$ and expresses the beginning and end points in terms of the initial time $t_i$ and final time $t_f$, rather than the initial and final image numbers i.

In the calculus notation, we replace the summation sign $\Sigma$ by something that looks almost like the summation sign, namely the ***integral sign*** $\int$. (The French word for integration is the same as their word for summation.) Next we replaced the individual $\vec{v}_i$ by the continuous variable $\vec{v}(t)$ and finally express the end points by the initial time $t_i$ and the final time $t_f$. The result is

$$\vec{s} = \sum_{i=1}^{n} \vec{v}_i \Delta t \rightarrow \left( \begin{array}{l} \textit{as the number} \\ \textit{n becomes} \\ \textit{infinitely} \\ \textit{large} \end{array} \right) \int_{t_i}^{t_f} \vec{v}(t)dt \qquad (17)$$

Calculus notation is more easily handled, or is at least more familiar, if we break vector equations up into component equations. Assume that the ball started at position i which has components $x_i = x(t_i)$ [read $x(t_i)$ as "x at time $t_i$ "] and $y_i = y(t_i)$ as shown in Figure (10). The final position f is at $x_f = x(t_f)$ and $y_f = y(t_f)$.

Thus the displacement $\vec{s}$ has x and y components

$$s_x = x(t_f) - x(t_i)$$

$$s_y = y(t_f) - y(t_i)$$

Breaking Equation (17) into component equations gives

$$s_x = x(t_f) - x(t_i) = \int_{t_i}^{t_f} v_x(t)dt \qquad (18a)$$

$$s_y = y(t_f) - y(t_i) = \int_{t_i}^{t_f} v_y(t)dt \qquad (18b)$$

Here we will introduce one more piece of notation often used in calculus courses. On the left hand side of Equation (18a) we have $x(t_f) - x(t_i)$ which we can think of as the variable x(t) evaluated over the interval of time from $t_i$ to $t_f$. We will often deal with variables evaluated over some interval and have a special notation for that. We will write

$$x(t_f) - x(t_i) \equiv x(t)\Big|_{t_i}^{t_f} \qquad (19)$$

You are to read the symbol $x(t)\Big|_{t_i}^{t_f}$ as "x of t evaluated from $t_i$ to $t_f$". We write the initial time $t_i$ at the bottom of the vertical bar, the final time $t_f$ at the top.

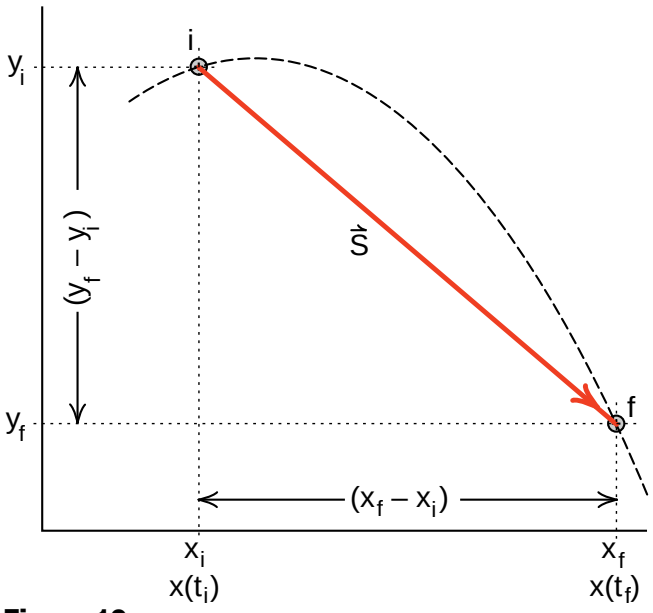



**Figure 10**
***Breaking the vector $\vec{s}$ into components.***

We use similar notation for any kind of variable, for example

$$f(x)\Big|_{x_1}^{x_2} \equiv f(x_2) - f(x_1) \qquad (19a)$$

(Remember to subtract when the variable is evaluated at the value at the bottom of the vertical bar.)

With this notation, our Equation (18) can be written

$$s_x = x(t)\Big|_{t_i}^{t_f} = \int_{t_i}^{t_f} v_x(t)dt \qquad (18a')$$

$$s_y = y(t)\Big|_{t_i}^{t_f} = \int_{t_i}^{t_f} v_y(t)dt \qquad (18b')$$

## Calculating Integrals

Equation (18) is nice and compact, but how do you use it? How do you calculate integrals? The key is to remember that an integral is just a fancy notation for a sum of terms, where we make the time step $\Delta t$ very small. Keeping this in mind, we will see that there is a very easy way to interpret an integral.

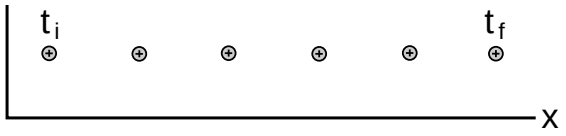



**Figure 11a**
***Strobe photograph of ball moving at constant velocity in x direction.***

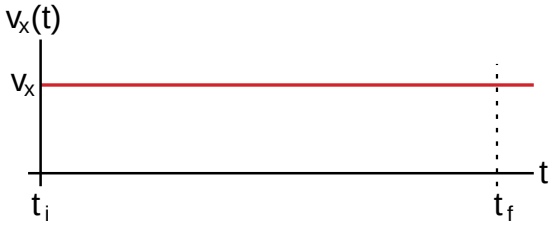



**Figure 11b**
***Graph of $v_x(t)$ versus t for the ball of Figure 11a.***

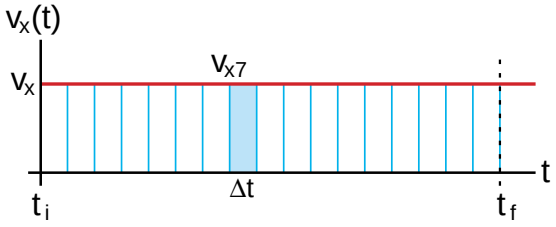



**Figure 11c**
***Each $v_x\Delta t$ is the area of a rectangle.***

To get this interpretation, let us start with the simple case of a ball moving in a straight line, for instance, the x direction, at a constant velocity $v_x$. A strobe picture of this motion would look like that shown in Figure (11a).

Figure (11b) is a graph of the ball's velocity $v_x(t)$ as a function of the time t. The vertical axis is the value of $v_x$, the horizontal axis is the time t. Since the ball is traveling at constant velocity, $v_x$ has a constant value and is thus represented by a straight horizontal line. In order to calculate the distance that the ball has traveled during the time interval from $t_i$ to $t_f$, we need to evaluate the integral

$$s_x = \int_{t_i}^{t_f} v_x(t)dt \quad \begin{matrix} \textit{distance ball} \\ \textit{travels in} \\ \textit{time interval} \\ t_i \textit{ to } t_f \end{matrix} \qquad (18a)$$

To actually evaluate the integral, we will go back to our summation notation

$$s_x = \sum_{i_{initial}}^{i_{final}} v_{xi}\Delta t \qquad (20)$$

and show individual time steps $\Delta t$ in the graph of $v_x$ versus t, as in Figure (11c).

We see that each term in Equation (20) is represented in Figure (11c) by a rectangle whose height is $v_x$ and whose width is $\Delta t$. We have shaded in the rectangle representing the 7th term $v_{x7}\Delta t$. We see that $v_{x7}\Delta t$ is just the ***area*** of the shaded rectangle, and it is clear that the sum of all the areas of the individual rectangles is the total area under the curve, starting at time $t_i$ and ending at time $t_f$. Here we are beginning to see that the process of integration is equivalent to finding the area under a curve.

With a simple curve like the constant velocity $v_x(t)$ in Figure (11c), we see by inspection that the total area from $t_i$ to $t_f$ is just the area of the complete rectangle of height $v_x$ and width $(t_f - t_i)$. Thus

$$s_x = v_x \times (t_f - t_i) \qquad (21)$$

This is the expected result for constant velocity, namely

$$\begin{matrix}\text{distance} \\ \text{traveled}\end{matrix} = \text{velocity} \times \text{time} \quad \begin{matrix}\textit{for} \\ \textit{constant} \\ \textit{velocity}\end{matrix} \qquad (21a)$$

To see that you are not restricted to the case of constant velocity, suppose you drove on a freeway due east (the x direction) starting at 9:00 AM and stopping for lunch at 12 noon. Every minute during your trip you wrote down the speedometer reading so that you had an accurate plot of $v_x(t)$ for the entire morning, a plot like that shown in Figure (12). From such a plot, could you determine the distance $s_x$ that you had traveled?

Your best answer is to multiply each value $v_i$ of your velocity by the time $\Delta t$ to calculate the average distance traveled each minute. Summing these up from the initial time $t_i$ = 9:00 AM to the final time $t_f$ = noon , you have as your estimate

$$s_x \approx \sum_i v_{xi}\Delta t$$

(The symbol ≈ means ***approximately equal***.)

To get a more accurate value for the distance traveled, you should measure your velocity at shorter time intervals $\Delta t$ and add up the larger number of smaller rectangles. The precise answer should be obtained in the limit as $\Delta t$ goes to zero

$$s_x = \lim_{\Delta t \to 0} \sum_i v_{xi}\Delta t = \int_{t_i}^{t_f} v_x(t)dt \qquad (22)$$

This limit is just the area under the curve that is supposed to represent the instantaneous velocity $v_x(t)$ .

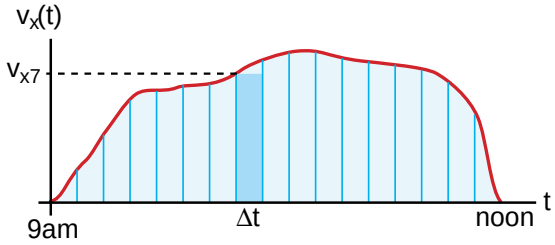



**Figure 12**
***Plot of $v_x(t)$ for a trip starting at 9:00 AM and finishing at noon. The distance traveled is the area under the curve.***

Thus we can interpret the integral of a curve as the area under the curve even when the curve is not constant or flat. Mathematicians concern themselves with curves that are so wild that it is difficult or impossible to determine the area under them. Such curves seldom appear in physics problems.

While the basic idea of integration is simple—just finding the area under a curve—in practice it can be quite difficult to calculate the area. Much of an introductory calculus course is devoted to finding the formulas for the areas under various curves. There are also books called ***tables of integrals*** where you look up the formula for a curve and the table tells you the formula for the area under that curve.

In Chapter 16 of the Physics text, we will discuss a mathematical technique called ***Fourier analysis***. This is a technique in which we can describe the shape of any continuous curve in terms of a sum of sine waves. (Why we want to do that will become clear then.) The process of Fourier analysis involves finding the area under some very complex curves, curves often involving experimental data for which we have no formula, only graphs. Such curves cannot be integrated by using a table of integrals, with the result that Fourier analysis was not widely used until the advent of the modern digital computer.

The computer made a difference, because we can find the area under almost any curve by breaking the curve into short pieces of length $\Delta t$ , calculating the area $v_i\Delta t$ of each narrow rectangle, and adding up the area of the rectangles to get the total area. If the curve is so wild that we have to break it into a million segments to get an accurate answer, that might be too hard to do by hand, but it usually a very simple and rapid job for a computer. Computers can be much more efficient than people at integration.

## The Process of Integrating

There is a language for the process of integration which we will now take you through. In each case we will check that the results are what we would expect from our summation definition, or the idea that an integral is the area under a curve.

The simplest integral we will encounter is the calculation of the area under a curve of unit height as shown in Figure (13). We have the area of a rectangle of height 1 and length $(t_f - t_i)$

$$\int_{t_i}^{t_f} 1\,dt = \int_{t_i}^{t_f} dt = (t_f - t_i) \qquad (22)$$

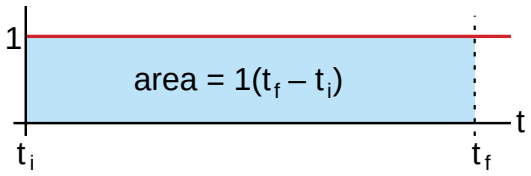



**Figure 13**
***Area under a curve of unit height.***

We will use some special language to describe this integration. We will say that the integral of dt is simply the time t, and that the integral of dt from $t_i$ to $t_f$ is equal to t evaluated from $t_i$ to $t_f$. In symbols this is written as

$$\int_{t_i}^{t_f} dt = t\Big|_{t_i}^{t_f} = (t_f - t_i) \qquad (23)$$

Recall that the vertical line after a variable means to evaluate that variable at the final position $t_f$ (upper value), minus that variable evaluated at the initial position $t_i$ (lower value). Notice that this prescription gives the correct answer.

The next simplest integral is the integral of a constant, like a constant velocity $v_x$ over the interval $t_i$ to $t_f$

$$\int_{t_i}^{t_f} v_x\,dt = v_x(t_f - t_i) \qquad (24)$$

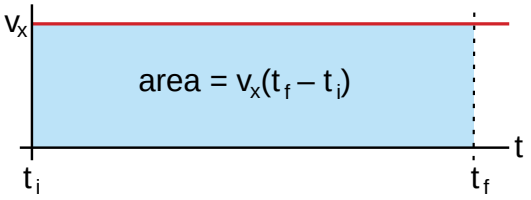



**Figure 14**
***Area under the constant $v_x$ curve.***

Since $(t_f - t_i) = \int_{t_i}^{t_f} dt$, we can replace $(t_f - t_i)$ in Equation (24) by the integral to get

$$\int_{t_i}^{t_f} v_x\,dt = v_x\int_{t_i}^{t_f} dt \qquad v_x \text{ a constant} \qquad (25)$$

and we see that a constant like $v_x$ can be taken outside the integral sign.

Let us try the simplest case we can think of where $v_x$ is not constant. Suppose $v_x$ starts at zero at time $t_i = 0$ and increases linearly according to the formula

$$v_x = at \qquad (26)$$

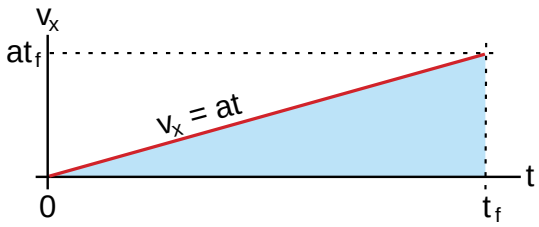



**Figure 15**

When we get up to the time $t_f$ the velocity will be $(at_f)$ as shown in Figure (15). The area under the curve $v_x = at$ is a triangle whose base is of length $t_f$ and height is $at_f$. The area of this triangle is one half the base times the height, thus we get for the distance $s_x$ traveled by an object moving with this velocity

$$s_x = \int_0^{t_f} v_x\,dt = \frac{1}{2}(\text{base}) \times (\text{height})$$
$$= \frac{1}{2}(t_f)(at_f) = \frac{1}{2}at_f^2 \qquad (27)$$

Now let us repeat the same calculation using the language one would find in a calculus book. We have

$$s_x = \int_0^{t_f} v_x\,dt = \int_0^{t_f} (at)\,dt \qquad (28)$$

The constant (a) can come outside, and we know that the answer is $1/2at_f^2$, thus we can write

$$s_x = a\int_0^{t_f} t\,dt = \frac{1}{2}at_f^2 \qquad (29)$$

In Equation (29) we can cancel the a's to get the result

$$\int_0^{t_x} t\,dt = \frac{1}{2}t_f^2 \qquad (30)$$

In a calculus text, you would find the statement that the integral $\int t\,dt$ is equal to $t^2/2$ and that the integral should be evaluated as follows

$$\int_0^{t_f} t\,dt = \frac{t^2}{2}\Big|_0^{t_f} = \frac{t_f^2}{2} - \frac{0}{2} = \frac{t_f^2}{2} \qquad (31)$$

## Indefinite Integrals

When we want to measure an actual area under a curve, we have to know where to start and stop. When we put these limits on the integral sign, like $t_i$ and $t_f$, we have what is called a ***definite integral***. However there are times where we just want to know what the form of the integral is, with the idea that we will put in the limits later. In this case we have what is called an ***indefinite integral***, such as

$$\int t\,dt = \frac{t^2}{2} \quad \textit{indefinite integral} \qquad (32)$$

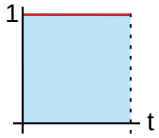
The difference between our definite integral in Equation (31) and the indefinite one in Equation (32) is that we have not chosen the limits yet in Equation (32). If possible, a table of integrals will give you a formula for the indefinite integral and let you put in whatever limits you want.

## Integration Formulas

For some sets of curves, there are simple formulas for the area under them. One example is the set of curves of the form $t^n$. We have already considered the cases where n = 0 and n = 1.

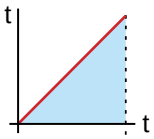
*n = 0*

$$\int t^0 dt = \int dt = t \qquad (33a)$$




*n = 1*

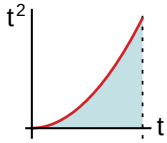
$$\int t^1 dt = \int t\,dt = \frac{t^2}{2} \qquad (33b)$$




Some results we will prove later are

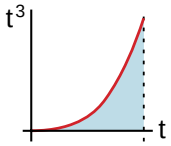
*n = 2*

$$\int t^2 dt = \frac{t^3}{3} \qquad (33c)$$




*n = 3*

$$\int t^3 dt = \frac{t^4}{4} \qquad (33d)$$




Looking at the way these integrals are turning out, we suspect that the general rule is

$$\boxed{\int t^n dt = \frac{t^{n+1}}{n+1}} \qquad (34)$$

It turns out that Equation (34) is a general result for any value of n except n = –1. If n = –1, then you would have division by zero, which cannot be the answer. (We will shortly discuss the special case where n = –1.)

As long as we stay away from the n = –1 case, the formula works for negative numbers. For example

$$\int t^{-2} dt = \int \frac{dt}{t^2} = \frac{t^{(-2+1)}}{-2+1} = \frac{t^{-1}}{(-1)}$$

$$\boxed{\int \frac{dt}{t^2} = -\frac{1}{t}} \qquad (35)$$

In our discussion of gravitational and electrical potential energy, we will encounter integrals of the form seen in Equation (35).

---

**Exercise 1**

Using Equation (34) and the fact that constants can come outside the integral, evaluate the following integrals:

(a) $\int x\,dx$ *it does not matter whether we call the variable t or x*

(b) $\int_{x=1}^{x=2} x^5 dx$ *also sketch the area being evaluated*

(c) $\int_{t=1}^{t=2} \frac{dt}{t^2}$ *show that you get a positive area*

(d) $\int \frac{GmM}{r^2} dr$ *where G, m, and M are constants*

(e) $\int \frac{a}{y^{3/2}} dy$ *"a" is a constant*

---

## NEW FUNCTIONS

We have seen that when we integrate a curve or function like $t^2$, we get a new function $t^3/3$. The functions $t^2$ and $t^3$ appear to be fairly similar; the integration did not create something radically different. However, the process of integration can lead to some curves with entirely different behavior. This happens, for example, in that special case n = –1 when we try to do the integral of $t^{-1}$.

### Logarithms

It is certainly not hard to plot $t^{-1}$, the result is shown in Figure (16). Also there is nothing fundamentally difficult or peculiar about measuring the area under the $t^{-1}$ curve from some $t_i$ to $t_f$, as long as we stay away from the origin t = 0 where $t^{-1}$ blows up. The formula for this area turns out, however, to be the new function called the ***natural logarithm***, abbreviated by the symbol *ln*. The area in Figure (16) is given by the formula

$$\int_{t_i}^{t_f} \frac{1}{t}\,dt \;=\; \ln(t_f) - \ln(t_i) \qquad (36)$$

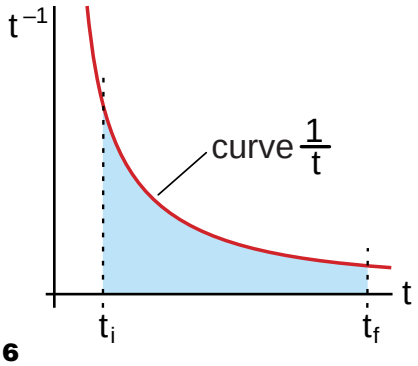



**Figure 16**
***Plot of $t^{-1}$. The area under this curve is the natural logarithm ln.***

Two of the important but peculiar features of the natural logarithm are

$$\ln(ab) \;=\; \ln(a) + \ln(b) \qquad (37)$$

$$\ln\left(\frac{1}{a}\right) \;=\; -\ln(a) \qquad (38)$$

Thus we get, for example

$$\ln(t_f) - \ln(t_i) \;=\; \ln(t_f) + \ln\left(\frac{1}{t_i}\right)$$

$$= \ln\left(\frac{t_f}{t_i}\right) \qquad (39)$$

Thus the area under the curve in Figure (16) is

$$\int_{t_i}^{t_f} \frac{dt}{t} \;=\; \ln\left(\frac{t_f}{t_i}\right) \qquad (40)$$

While the natural logarithm has some rather peculiar properties it is easy to evaluate because it is available on all scientific calculators. For example, if $t_i$ = .5 seconds and $t_f$ = 4 seconds, then we have

$$\ln\left(\frac{t_f}{t_i}\right) \;=\; \ln\left(\frac{4}{.5}\right) \;=\; \ln(8) \qquad (41)$$

Entering the number 8 on a scientific calculator and pressing the button labeled *ln*, gives

$$\ln(8) \;=\; 2.079 \qquad (42)$$

which is the answer.

---

**Exercise 2**

Evaluate the integrals

$$\int_{.001}^{1000} \frac{dx}{x} \qquad \int_{.000001}^{1} \frac{dx}{x}$$

Why are the answers the same?

---

## The Exponential Function

We have just seen that, while the logarithm function may have some peculiar properties, it is easy to evaluate using a scientific calculator. The question we now want to consider is whether there is some function that undoes the logarithm. When we enter the number 8 into the calculator and press *ln*, we get the number 2.079. Now we are asking if, when we enter the number 2.079, can we press some key and get back the number 8? The answer is, you press the key labeled $e^x$. The $e^x$ key performs the ***exponential function*** which undoes the logarithm function. We say that the exponential function $e^x$ is the ***inverse*** of the logarithm function *ln*.

## Exponents to the Base 10

You are already familiar with exponents to the base 10, as in the following examples

$$\begin{array}{ll} 10^0 = 1 & \\ 10^1 = 10 & 10^{-1} = 1/10 = .1 \\ 10^2 = 100 & 10^{-2} = 1/100 = .01 \\ \cdots\cdots & \cdots\cdots \\ 10^6 = 1{,}000{,}000 & 10^{-6} = .000001 \end{array} \qquad (43)$$

The exponent, the number written above the 10, tells us how many factors of 10 are involved. A minus sign means how many factors of 10 we divide by. From this alone we deduce the following rules for the exponent to the base 10.

$$\boxed{10^{-a} = \frac{1}{10^a}} \qquad (44)$$

$$\boxed{10^a \times 10^b = 10^{a+b}} \qquad (45)$$

(Example $10^2 \times 10^3 = 100 \times 1000 = 100{,}000$ .)

The inverse of the *exponent to the base 10* is the function called *logarithm to the base 10* which is denoted by the key labeled *log* on a scientific calculator. Formally this means that

$$\boxed{\log(10^y) = y} \qquad (46)$$

Check this out on your scientific calculator. For example, enter the number 1,000,000 and press the *log* button and see if you get the number 6. Try several examples so that you are confident of the result.

## The Exponential Function $y^x$

Another key on your scientific calculator is labeled $y^x$. This allows you to determine the value of any number y raised to the power (or exponent) x. For example, enter the number y = 10, and press the $y^x$ key. Then enter the number x = 6 and press the = key. You should see the answer

$$y^x = 10^6 = 1000000$$

It is quite clear that all exponents obey the same rules we saw for powers of 10, namely

$$\boxed{y^a \times y^b = y^{a+b}} \qquad (47)$$

[Example $y^2 \times y^3 = (y \times y)(y \times y \times y) = y^5$.]

And as before

$$\boxed{y^{-a} \equiv \frac{1}{y^a}} \qquad (48)$$

**Exercise 3**

Use your scientific calculator to evaluate the following quantities. (You should get the answers shown.)

(a) $10^6$ (1000000)
(b) $2^3$ (8)
(c) $23^0$ (1)
(d) $10^{-1}$ (.1)

(To do this calculation, enter 10, then press $y^x$ . Then enter 1, then press the +/– key to change it to –1, then press = to get the answer .1)

(e) $2^{-.5}$ ($1/\sqrt{2} = .707$)
(f) log (10) (1)
(g) ln (2.7183) (1) (very close to 1)

Try some other examples on your own to become completely familiar with the $y^x$ key. (You should note that any positive number raised to the 0 power is 1. Also, some calculators, in particular the one I am using, cannot handle any negative values of y, not even $(-2)^2$ which is +4)

## Euler's Number e = 2.7183. . .

We have seen that the function *log* on the scientific calculator undoes, is the inverse of, powers of 10. For example, we saw that

$$\log(10^x) = x \qquad \text{(46) repeated}$$

Example: $\log(10^6) = 6$

Earlier we saw that the exponential function $e^x$ was the inverse of the natural logarithm *ln*. This means that

$$\ln(e^x) = x \qquad (49)$$

The difference between the logarithm *log* and the natural logarithm *ln*, is that *log* undoes exponents of the number 10, while *ln* undoes exponents of the number *e*. This special number *e*, one of the fundamental mathematical constants like $\pi$, is known as ***Euler's number***, and is always denoted by the letter *e*.

You can find the numerical value of Euler's number *e* on your calculator by evaluating

$$e^1 = e \qquad (50)$$

To do this, enter 1 into your calculator, press the $e^x$ key, and you should see the result

$$e^1 = e = 2.718281828 \qquad (51)$$

We will run into this number throughout the course. You should remember that *e* is about 2.7, or you might even remember 2.718. (Only remembering *e* as 2.7 is as klutzy as remembering $\pi$ as 3.1)

The terminology in math courses is that the function *log*, which undoes exponents of the number 10, is the *logarithm to the base 10*. The function *ln*, what we have called the *natural logarithm*, which undoes exponents of the number *e*, is the *logarithm to the base e*. You can have logarithms to any base you want, but in practice we only use base 10 (because we have 10 fingers) and the base *e*. The base *e* is special, in part because that is the logarithm that naturally arises when we integrate the function 1/x. We will see shortly that the functions *ln* and $e^x$ have several more, very special features.

## DIFFERENTIATION AND INTEGRATION

The scientific calculator is a good tool for seeing how the functions like *ln* and $e^x$ are inverse of each other. Another example of inverse operations is integration and differentiation. We have seen that integration allows us to go the other way from differentiation [finding x(t) from v(t), rather than v(t) from x(t)]. However it is not so obvious that integration and differentiation are inverse operations when you think of integration as finding the area under a curve, and differentiation as finding limits of $\Delta x/\Delta t$ as $\Delta t$ goes to zero. It is time now to make this relationship clear.

First, let us review our concept of a derivative. Going back to our strobe photograph of Figure (3), replacing $\vec{R}_i$ by $\vec{R}(t)$ and $\vec{R}_{i+1}$ by $\vec{R}(t+\Delta t)$, as shown in Figure (3a), our strobe velocity was then given by

$$\vec{v}(t) = \frac{\vec{R}(t+\Delta t) - \vec{R}(t)}{\Delta t} \qquad (52)$$

The calculus definition of the velocity is obtained by reducing the strobe time interval $\Delta t$ until we obtain the instantaneous velocity $\vec{v}$.

$$\vec{v}_{calculus} = \lim_{\Delta t \to 0} \frac{\vec{R}(t + \Delta t) - \vec{R}(t)}{\Delta t} \qquad (53)$$

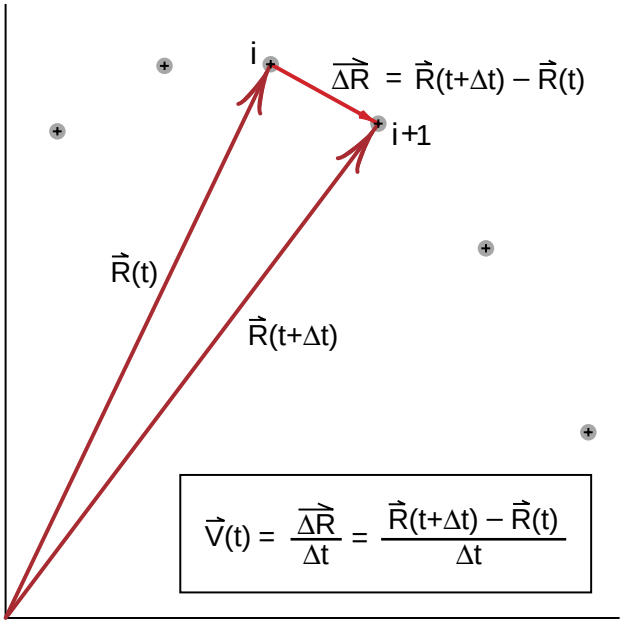



**Figure 3a**
***Defining the strobe velocity.***

While Equation (53) looks like it is applied to the explicit case of the strobe photograph of projectile motion, it is easily extended to cover any process of differentiation. Whatever function we have [we had $\vec{R}(t)$, suppose it is now f(t)], evaluate it at two closely spaced times, subtract the older value from the newer one, and divide by the time difference $\Delta t$. Taking the limit as $\Delta t$ becomes very small gives us the derivative

$$\frac{d\,f(t)}{dt} \equiv \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} \qquad (54)$$

The variable with which we are differentiating does not have to be time t. It can be any variable that we can divide into small segments, such as x

$$\frac{d}{dx} f(x) \equiv \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \qquad (55)$$

Let us see how the operation defined in Equation (55) is the inverse of finding the area under a curve.

Suppose we have a curve, like our old $v_x(t)$ graphed as a function of time, as shown in Figure (17). To find out how far we traveled in a time interval from $t_i$ to some later time T, we would do the integral

$$x(T) = \int_{t_i}^{T} v_x(t)\,dt \qquad (56)$$

The integral in Equation (56) tells us how far we have gone at any time T during the trip. The quantity x(T) is a function of this time T.

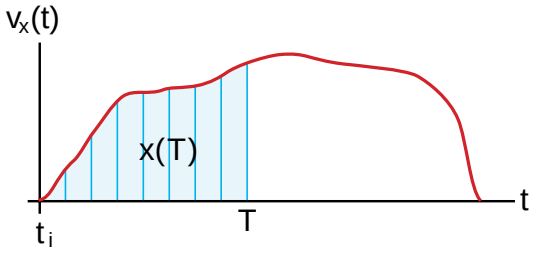



**Figure 17**
***The distance traveled by the time T is the area under the velocity curve up to the time T.***

Now let us differentiate the function x(T) with respect to the variable T. By our definition of differentiation we have

$$\frac{d}{dT}x(T) = \lim_{\Delta t \to 0} \frac{x(T+\Delta t) - x(T)}{\Delta t} \qquad (57)$$

Figure (17) shows us the function x(T). It is the area under the curve v(t) starting at $t_i$ and going up to time t = T. Figure (18) shows us the function $x(T+\Delta t)$. It is the area under the same curve, starting at $t_i$ but going up to $t = T + \Delta t$. When we subtract these two areas, all we have left is the area of the slender rectangle shown in Figure (19).

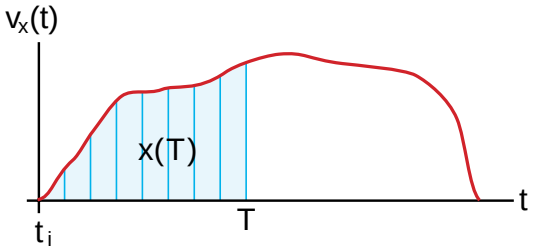



**Figure 17 repeated**
***The distance x(T) traveled by the time T***

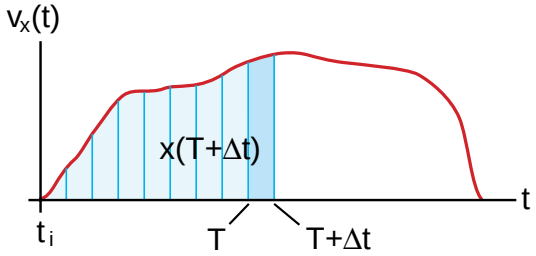



**Figure 18**
***The distance $x(T+\Delta t)$ traveled by the time $T+\Delta t$.***

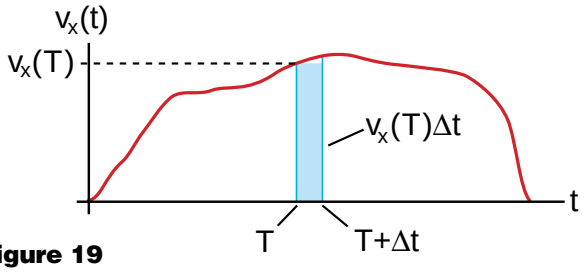



**Figure 19**
***The distance $x(T+\Delta t) - x(T)$ traveled during the time $\Delta t$.***

The rectangle has a height approximately $v_x(T)$ and a width $\Delta t$ for an area

$$x(T+\Delta t) - x(T) = v_x(T)\Delta t \qquad (58)$$

Dividing through by $\Delta t$ gives

$$v_x(T) = \frac{x(T+\Delta t) - x(T)}{\Delta t} \qquad (59)$$

The only approximation in Equation (59) is at the top of the rectangle. If the curve is not flat, $v_x(T+\Delta t)$ will be different from $v_x(T)$ and the area of the sliver will have a value somewhere between $v_x(T)\Delta t$ and $v_x(T+\Delta t)\Delta t$. But if we take the limit as $\Delta t$ goes to zero, the value of $v_x(T+\Delta t)$ must approach $v_x(T)$, and we end up with the exact result

$$v_x(T) = \lim_{\Delta t \to 0} \frac{x(T+\Delta t) - x(T)}{\Delta t} \qquad (60)$$

This is just the derivative dx(t)/dt evaluated at t = T.

$$\boxed{v_x(T) = \left.\frac{dx(t)}{dt}\right|_{t=T}} \qquad (61a)$$

where we started from

$$\boxed{x(T) = \int_{t_i}^{T} v_x(t)\,dt} \qquad (61b)$$

Equations (61a) and (61b) demonstrate explicitly how differentiation and integration are inverse operations. The derivative allowed us to go from x(t) to $v_x(t)$ while the integral took us from $v_x(t)$ to x(t). This inverse is not as simple as pushing a button on a calculator to go from *ln* to $e^x$. Here we have to deal with limits on the integration and a shift of variables from t to T. But these two processes do allow us to go back and forth.

## A Fast Way to go Back and Forth

We introduced our discussion of integration by pointing out that equations

$$v_x(t) = \frac{dx(t)}{dt}; \qquad a_x(t) = \frac{dv_x(t)}{dt} \qquad (62a,b)$$

went the wrong way in that we were more likely to know the acceleration $a_x(t)$ and from that want to calculate the velocity $v_x(t)$ and distance traveled x(t). After many steps, we found that integration was what we needed.

We do not want to repeat all those steps. Instead we would like a quick and simple way to go the other way around. Here is how you do it. Think of the dt in (62a) as a small but finite time interval. That means you can treat it like any other number and multiply both sides of Equation (62a) through by it.

$$v_x(t) = \frac{dx(t)}{dt}$$

$$dx(t) = v_x(t)dt \qquad (63)$$

Now integrate both sides of Equation (63) from some initial time $t_i$ to a final time T. (If you do the same thing to both sides of an equation, both sides should still be equal to each other.)

$$\int_{t_i}^{T} dx(t) = \int_{t_i}^{T} v_x(t)\,dt \qquad (64)$$

If dt is to be thought of as a small but finite time step, then dx(t) is the small but finite distance we moved in the time dt. The integral on the left side of Equation (64) is just the sum of all these short distances moved, which is just the total distance moved during the time from $t_i$ to T.

$$\int_{t_i}^{T} dx(t) = x(t)\Big|_{t_i}^{T} = x(T) - x(t_i) \qquad (65)$$

Thus we end up with the result

$$x(t)\Big|_{t_i}^{T} = \int_{t_i}^{T} v_x(t)dt \qquad (66)$$

Equation (66) is a little more general than (62b) for it allows for the fact that $x(t_i)$ might not be zero. If, however, we say that we started our trip at $x(t_i) = 0$, then we get the result

$$x(T) = \int_{t_i}^{T} v_x(t)dt \qquad (67)$$

representing the distance traveled since the start of the trip.

## Constant Acceleration Formulas

The constant acceleration formulas, so well known from high school physics courses, are an excellent application of the procedures we have just described.

We will begin with motion in one dimension. Suppose a car is traveling due east, in the x direction, and for a while has a constant acceleration $a_x$. The car passes us at a time $t_i = 0$, traveling at a speed $v_{x0}$. At some later time T, if the acceleration $a_x$ remains constant, how far away from us will the car be?

We start with the equation

$$a_x(t) = \frac{dv_x(t)}{dt} \qquad (68)$$

Multiplying through by dt to get

$$dv_x(t) = a_x(t)dt$$

then integrating from time $t_i = 0$ to time $t_f = T$, we get

$$\int_0^T dv_x(t) = \int_0^T a_x(t)dt \qquad (69)$$

Since the integral $\int dv_x(t) = v_x(t)$, we have

$$\int_0^T dv_x(t) = v_x(t)\Big|_0^T = v_x(T) - v_x(0) \qquad (70)$$

where $v_x(0)$ is the velocity $v_{x0}$ of the car when it passed us at time t = 0.

While we can always do the left hand integral in Equation (69), we cannot do the right hand integral until we know $a_x(t)$. For the constant acceleration problem, however, we know that $a_x(t) = a_x$ is constant, and we have

$$\int_0^T a_x(t)dt = \int_0^T a_x dt \qquad (71)$$

Since constants can come outside the integral sign, we get

$$\int_0^T a_x dt = a_x \int_0^T dt = a_x t \Big|_0^T = a_x T \qquad (72)$$

where we used $\int dt = t$. Substituting Equations (70) and (72) in (69) gives

$$v_x T - v_{x0} = a_x T \qquad (73)$$

Since Equation (73) applies for any time T, we can replace T by t to get the well known result

$$\boxed{v_x(t) = v_{x0} + a_x t} \qquad (a_x \textit{ constant}) \qquad (74)$$

Equation (74) tells us the speed of the car at any time t after it passed us, as long as the acceleration remains constant.

To find out how far away the car is, we start with the equation

$$v_x(t) = \frac{dx(t)}{dt} \qquad (62a)$$

Multiplying through by dt to get

$$dx(t) = v_x(t)\, dt$$

then integrating from time t = 0 to time t = T gives (as we saw earlier)

$$\int_0^T dx(t) = \int_0^T v_x(t)\, dt \qquad (75)$$

The left hand side is

$$\int_0^T dx(t) = x(t) \Big|_0^T = x(T) - x(0) \qquad (76)$$

If we measure along the x axis, starting from where we are (where the car was at t = 0) then x(0) = 0.

In order to do the right hand integral in Equation (75), we have to know what the function $v_x(t)$ is. But for constant acceleration, we have from Equation (74) $v_x(t) = v_{x0} + a_x t$ , thus

$$\int_0^T v_x(t)\, dt = \int_0^T (v_{x0} + a_x t)\, dt \qquad (77)$$

One of the results of integration that you should prove for yourself (just sketch the areas) is the rule

$$\int_i^f \big[a(x) + b(x)\big] dx = \int_i^f a(x) dx + \int_i^f b(x) dx \qquad (78)$$

thus we get

$$\int_0^T (v_{x0} + a_x t) dt = \int_0^T v_{x0} dt + \int_0^T a_x t\, dt \qquad (79)$$

Since constants can come outside the integrals, this is equal to

$$\int_0^T (v_{x0} + a_x t) dt = v_{x0} \int_0^T dt + a_x \int_0^T t\, dt \qquad (80)$$

Earlier we saw that

$$\int_0^T dt = t \Big|_0^T = T - 0 = T \qquad (23)$$

$$\int_0^T t dt = \frac{t^2}{2} \Big|_0^T = \frac{T^2}{2} - 0 = \frac{T^2}{2} \qquad (30)$$

Thus we get

$$\int_0^T (v_{x0} + a_x t) dt = v_{x0} T + \frac{1}{2} a_x T^2 \qquad (81)$$

Using Equations (76) and (81) in (75) gives

$$x(T) - x_0 = v_{x0} T + \frac{1}{2} a_x T^2$$

Taking $x_0 = 0$ and replacing T by t gives the other constant acceleration formula

$$\boxed{x(t) = v_{x0} t + \frac{1}{2} a_x t^2} \qquad (a_x \textit{ constant}) \qquad (82)$$

You can now see that the factor of $t^2/2$ in the constant acceleration formulas comes from the integral $\int t dt$.

**Exercise 4**

Find the formula for the velocity v(t) and position x(t) for a car moving with constant acceleration $a_x$, that was located at position $x_i$ at some initial time $t_i$.

Start your calculation from the equations

$$v_x(t) = \frac{dx(t)}{dt}$$

$$a_x(t) = \frac{dv_x(t)}{dt}$$

and go through all the steps that we did to get Equations (74) and (82). See if you can do this without looking at the text.

If you have to look back to see what some steps are, then finish the derivation looking at the text. Then a day or so later, clean off your desk, get out a blank sheet of paper, write down this problem, put the book away and do the derivation. Keep doing this until you can do the derivation of the constant acceleration formulas without looking at the text.

## Constant Acceleration Formulas in Three Dimensions

To handle the case of motion with constant acceleration in three dimensions, you start with the separate equations

$$\begin{aligned} v_x(t) &= \frac{dx(t)}{dt} \qquad a_x(t) = \frac{dv_x(t)}{dt} \\ v_y(t) &= \frac{dy(t)}{dt} \qquad a_y(t) = \frac{dv_y(t)}{dt} \\ v_z(t) &= \frac{dz(t)}{dt} \qquad a_z(t) = \frac{dv_z(t)}{dt} \end{aligned} \tag{83}$$

Then repeat, for each pair of equations, the steps that led to the constant acceleration formulas for motion in the x direction. The results will be

$$\begin{aligned} x(t) &= v_{x0}t + \frac{1}{2}a_x t^2 \qquad v_x(t) = v_{x0} + a_x t \\ y(t) &= v_{y0}t + \frac{1}{2}a_y t^2 \qquad v_y(t) = v_{y0} + a_y t \\ z(t) &= v_{z0}t + \frac{1}{2}a_z t^2 \qquad v_z(t) = v_{z0} + a_z t \end{aligned} \tag{84}$$

The final step is to combine these six equations into the two vector equations

$$\boxed{\vec{x}(t) = \vec{v}_0 t + \frac{1}{2}\vec{a}t^2 \;; \qquad \vec{v}(t) = \vec{v}_0 + \vec{a}t} \tag{85}$$

These are the equations we analyzed graphically in Chapter 3 of the Physics text, in Figure (3-34) and Exercise (3-9). (There we wrote $\vec{s}$ instead of $\vec{x}(t)$, and $\vec{v}_i$. rather than $v_0$.)

In many introductory physics courses, considerable emphasis is placed on solving constant acceleration problems. You can spend weeks practicing on solving these problems, and become very good at it. However, when you have done this, you have not learned very much physics because most forms of motion are not with constant acceleration, and thus the formulas do not apply. The formulas were important historically, for they were the first to allow the accurate prediction of motion (of cannonballs). But if too much emphasis is placed on these problems, students tend to use them where they do not apply. For this reason we have placed the exercises using the constant acceleration equations in an appendix at the end of Chapter 4 of the Physics text. There are plenty of problems there for all the practice you will need with these equations. Doing these exercises requires only algebra, there is no practice with calculus. To get some experience with calculus, be sure that you can confidently do Exercise 4.

## MORE ON DIFFERENTIATION

In our discussion of integration, we saw that the basic idea was that the integral of some curve or function f(t) was equal to the area under that curve. That is an easy enough concept. The problems arose when we actually tried to find the formulas for the areas under various curves. The only areas we actually calculated were the rectangular area under f(t) = constant and the triangular area under f(t) = at. It was perhaps a surprise that the area under the simple curve 1/t should turn out to be a logarithm.

For differentiation, the basic idea of the process is given by the formula

$$\frac{df(t)}{dt} = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} \qquad \text{(54) repeated}$$

Equation (54) is short hand notation for a whole series of steps which we introduced through the use of strobe photographs. The basic idea of differentiation is more complex than integration, but, as we will now see, it is often a lot easier to find the derivative of a curve than its integral.

### Series Expansions

An easy way to find the formula for the derivative of a curve is to use a series expansion. We will illustrate the process by using the binomial expansion to calculate the derivative of the function $x^n$ where n is any constant.

We used the binomial expansion, or at least the first two terms, in Chapter 1 of the Physics text. That was during our discussion of the approximation formulas that are useful in relativistic calculations. As we mentioned in Exercise (1-5), the binomial expansion is

$$(x + \alpha)^n = x^n + n\alpha x^{n-1} + \frac{n(n-1)}{2!}\alpha^2 x^{n-2} \ldots \qquad (86)$$

When $\alpha$ is a number much smaller than 1 ($\alpha << 1$), we can neglect $\alpha^2$ compared to $\alpha$ (if $\alpha = .01$, $\alpha^2 = .0001$), with the result that we can accurately approximate $(x + \alpha)^n$ by

$$\boxed{(x + \alpha)^n \approx x^n + n\alpha x^{n-1}} \qquad \alpha << 1 \qquad (87)$$

Equation (87) gives us all the approximation formulas found in Equations (1-20) through (1-25) on page 1-28 of the Physics text.

As an example of Equation (87), just to see that it works, let us take x = 5, n = 7 and $\alpha = .01$ to calculate $(5.01)^7$. From the calculator we get

$$(5.01)^7 = 79225.3344 \qquad (88)$$

(To do this enter 5.01, press the $y^x$ button, then enter 7 and press the = button.) Let us now see how this result compares with

$$(x + \alpha)^n \approx x^n + n\alpha x^{n-1}$$
$$(5 + .01)^7 \approx 5^7 + 7(.01)5^6 \qquad (89)$$

We have

$$5^7 = 78125 \qquad (90)$$

$$7 \times .01 \times 5^6 = 7 \times .01 \times 15625 = 1093.75 \qquad (91)$$

Adding the numbers in (90) and (91) together gives

$$5^7 + 7(.01)5^6 = 79218.75 \qquad (92)$$

Thus we end up with 79218 instead of 79225, which is not too bad a result. The smaller $\alpha$ is compared to one, the better the approximation.

## Derivative of the Function $x^n$

We are now ready to use our approximation formula (87) to calculate the derivative of the function $x^n$. From the definition of the derivative we have

$$\frac{d(x^n)}{dx} = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^n - x^n}{\Delta x} \qquad (93)$$

Since $\Delta x$ is to become infinitesimally small, we can use our approximation formula for $(x + \alpha)^n$. We get

$$(x + \alpha)^n \approx x^n + n(\alpha)x^{n-1} \qquad (\alpha << 1)$$

$$(x + \Delta x)^n \approx x^n + n(\Delta x)x^{n-1} \quad (\Delta x << 1) \qquad (94)$$

Using this in Equation (93) gives

$$\frac{d(x^n)}{dx} = \lim_{\Delta x \to 0} \left[ \frac{[x^n + n(\Delta x)x^{n-1}] - x^n}{\Delta x} \right] \qquad (95)$$

We used an equal sign rather than an approximately equal sign in Equation (95) because our approximation formula (94) becomes exact when $\Delta x$ becomes infinitesimally small.

In Equation (95), the terms $x^n$ cancel and we are left with

$$\frac{d(x^n)}{dx} = \lim_{\Delta x \to 0} \left[ \frac{n(\Delta x)x^{n-1}}{\Delta x} \right] \qquad (96)$$

At this point, the factors $\Delta x$ cancel and we have

$$\frac{d(x^n)}{dx} = \lim_{\Delta x \to 0} \left[ nx^{n-1} \right] \qquad (97)$$

Since no $\Delta x$'s remain in our formula, we end up with the exact result

$$\boxed{\frac{d(x^n)}{dx} = nx^{n-1}} \qquad (98)$$

Equation (98) is the general formula for the derivative of the function $x^n$.

In our discussion of integration, we saw that a constant could come outside the integral. The same thing happens with a derivative. Consider, for example,

$$\frac{d}{dx}\left[a\,f(x)\right] = \lim_{\Delta x \to 0} \left[ \frac{a\,f(x + \Delta x) - a\,f(x)}{\Delta x} \right]$$

Since the constant a has nothing to do with the limiting process, this can be written

$$\frac{d}{dx}\left[af(x)\right] = a \lim_{\Delta x \to 0} \left[ \frac{f(x + \Delta x) - f(x)}{\Delta x} \right] = a\,\frac{df(x)}{dx} \qquad (99)$$

---

**Exercise 5**

Calculate the derivative with respect to x (i.e., d/dx) of the following functions. (When negative powers of x are involved, assume x is not equal to zero.)

(a) $x$

(b) $x^2$

(c) $x^3$

(d) $5x^2 - 3x$

(Before you do part (d), use the definition of the derivative to prove that $\frac{d}{dx}\left[f(x) + g(x)\right] = \frac{df(x)}{dx} + \frac{dg(x)}{dx}$ )

(e) $x^{-1}$

(f) $x^{-2}$

(g) $\sqrt{x}$

(h) $1/\sqrt{x}$

(i) $3x^{.73}$

(j) $7x^{-.2}$

(k) 1

(In part (k) first show that this should be zero from the definition of the derivative. Then write $1 = x^0$ and show that Equation (98) also works, as long as x is not zero.)

(l) 5

---

## The Chain Rule

There is a simple trick called the ***chain rule*** that makes it easy to differentiate a wide variety of functions. The rule is

$$\boxed{\frac{df[y(x)]}{dx} = \frac{df(y)}{dy}\frac{dy}{dx}} \qquad \textit{chain rule} \quad (100)$$

To see how this rule works, consider the function

$$f(x) = (x^2)^n \qquad (101)$$

We know that this is just $f(x) = x^{2n}$, and the derivative is

$$\frac{df(x)}{dx} = \frac{d}{dx}(x^{2n}) = 2nx^{2n-1} \qquad (102)$$

But suppose that we did not know this trick, and therefore did not know how to differentiate $(x^2)^n$. We do, however, know how to differentiate powers like $x^2$ and $y^n$. The chain rule allows us to use this knowledge in order to figure out how to differentiate the more complex function $(x^2)^n$.

We begin by defining y(x) as

$$y(x) = x^2 \qquad (103)$$

Then our function $f(x) = (x^2)^n$ can be written in terms of y as follows

$$f(x) = (x^2)^n = [y(x)]^n = (y)^n = f(y)$$

$$f(y) = (y)^n \qquad (104)$$

Differentiating (103) and (104) gives

$$\frac{dy(x)}{dx} = \frac{d}{dx}(x^2) = 2x \qquad (105)$$

$$\frac{df(y)}{dy} = \frac{d}{dy}(y^n) = ny^{n-1} \qquad (106)$$

Using (104) and (105) in the chain rule (100) gives

$$\begin{aligned}\frac{df(y)}{dx} &= \frac{df}{dy}\times\frac{dy}{dx} = (ny^{n-1})\times(2x)\\ &= 2ny^{n-1}x\\ &= 2n(x^2)^{n-1}x\\ &= 2n(x^{2[n-1]})x\\ &= 2n(x^{[2n-2]})x\\ &= 2n(x^{[2n-2]+1})\\ &= 2nx^{2n-1}\end{aligned} \qquad (107)$$

which is the answer we expect.

In our example, using the chain rule was more difficult than differentiating directly because we already knew how to differentiate $x^{2n}$. But we will shortly encounter examples of new functions that we do not know how to differentiate directly, but which can be written in the form f[y(x)], where we know df/dy and dy/dx. We can then use the chain rule to evaluate the derivative df/dx. We will give you practice with the chain rule when we encounter these functions.

## Remembering the Chain Rule

The chain rule can be remembered by thinking of the dy's as cancelling as shown.

$$\boxed{\frac{df(y)}{\cancel{dy}}\frac{\cancel{dy}}{dx} = \frac{df(y)}{dx}} \qquad \textit{remembering the chain rule} \quad (108)$$

## Partial Proof of the Chain Rule (optional)

The proof of the chain rule is closely related to cancellation we showed in Equation (108). A partial proof of the rule proceeds as follows.

Suppose we have some function f(y) where y is a function of the variable x. As a result f[y(x)] is itself a function of x and can be differentiated with respect to x.

$$\frac{d}{dx} f\big[y(x)\big] = \lim_{\Delta x \to 0} \frac{f\big[y(x+\Delta x)\big] - f\big[y(x)\big]}{\Delta x} \qquad (123)$$

Now define the quantity $\Delta y$ by

$$\Delta y \equiv y(x+\Delta x) - y(x) \qquad (124)$$

so that

$$y(x+\Delta x) = y(x) + \Delta y$$

$$f[\,y(x+\Delta x)] = f(y+\Delta y)$$

and Equation (123) becomes

$$\frac{d}{dx} f\big[y(x)\big] = \lim_{\Delta x \to 0} \frac{f(y+\Delta y) - f(y)}{\Delta x} \qquad (125)$$

Now multiply (125) through by

$$1 = \frac{\Delta y}{\Delta y} = \frac{y(x+\Delta x) - y(x)}{\Delta y} \qquad (126)$$

to get

$$\frac{d}{dx} f\big[y(x)\big] = \lim_{\Delta x \to 0}\left[\frac{f(y+\Delta y) - f(y)}{\Delta x} \times \frac{y(x+\Delta x) - y(x)}{\Delta y}\right]$$

$$= \lim_{\Delta x \to 0}\left[\frac{f(y+\Delta y) - f(y)}{\Delta y} \times \frac{y(x+\Delta x) - y(x)}{\Delta x}\right] \qquad (127)$$

where we interchanged $\Delta x$ and $\Delta y$ in the denominator.

(We call this a partial proof for the following reason. For some functions y(x), the quantity $\Delta y = y(x+\Delta x) - y(x)$ may be identically zero for a small range of $\Delta x$. In that case we would be dividing by zero (the $1/\Delta y$) even before we took the limit as $\Delta x$ goes to zero. A more complete proof handles the special cases separately. The resulting chain rule still works however, even for these special cases.)

Since $\Delta y = y(x+\Delta x) - y(x)$ goes to zero as $\Delta x$ goes to zero, we can write Equation (127) as

$$\frac{d}{dx} f\big[y(x)\big] = \left[\lim_{\Delta y \to 0} \frac{f(y+\Delta y) - f(y)}{\Delta y}\right] \times \left[\lim_{\Delta x \to 0} \frac{y(x+\Delta x) - y(x)}{\Delta x}\right]$$

$$= \left[\frac{df(y)}{dy}\right]\left[\frac{dy}{dx}\right] \qquad (100) \text{ repeated}$$

This rule works as long as the derivatives df/dy and dy/dx are meaningful, i.e., we stay away from kinks or discontinuities in f and y.

## INTEGRATION FORMULAS

Knowing the formula for the derivative of the function $x^n$ , and knowing that integration undoes differentiation, we can now use Equation (98)

$$\frac{dx^n}{dx} = nx^{n-1} \qquad \text{(98) repeated}$$

to find the integral of the function $x^n$ . We will see that this trick works for all cases except the special case where n = –1, i.e., the special case where the integral is a natural logarithm.

To integrate $x^n$, let us go back to our calculation of the distance $s_x$ or x(t) traveled by an object moving in the x direction at a velocity $v_x$ . This was given by Equations (19) or (56) as

$$x(t)\Big|_{t_i}^{T} = \int_{t_i}^{T} v_x(t)\,dt \qquad (128)$$

where the instantaneous velocity $v_x(t)$ is defined as

$$v_x(t) = \frac{dx(t)}{dt} \qquad (129)$$

Suppose x(t) had the special form

$$x(t) = t^{n+1} \qquad (\textit{a special case}) \qquad (130)$$

then we know from our derivative formulas that

$$v(t) = \frac{dx(t)}{dt} = \frac{dt^{(n+1)}}{dt} = (n+1)t^n \qquad (131)$$

Substituting $x(t) = t^{n+1}$ and $v(t) = (n+1)t^n$ into Equation (128) gives

$$x(t)\Big|_{t_i}^{T} = \int_{t_i}^{T} v_x(t)\,dt \qquad (128)$$

$$t^{n+1}\Big|_{t_i}^{T} = \int_{t_i}^{T} (n+1)t^n dt = (n+1)\int_{t_i}^{T} t^n dt \qquad (132)$$

Dividing through by (n+1) gives

$$\boxed{\int_{t_i}^{T} t^n dt = \frac{1}{n+1} t^{n+1}\Big|_{t_i}^{T}} \qquad (133)$$

If we choose $t_i = 0$ , we get the simpler result

$$\int_{0}^{T} t^n dt = \frac{T^{n+1}}{n+1} \qquad (134)$$

and the indefinite integral can be written

$$\boxed{\int t^n dt = \frac{t^{n+1}}{n+1}} \qquad (135) \text{ (also 34)}$$

This is the general rule we stated without proof back in Equation (34). Note that this formula says nothing about the case n = –1, i.e., when we integrate $t^{-1} = 1/t$ , because n +1 = –1 +1 = 0 and we end up with division by zero. But for all other values of n, we now have derived a general formula for finding the area under any curve of the form $x^n$ (or $t^n$ ). This is a rather powerful result considering the problems one encounters actually finding areas under curves. (If you did not do Exercise 1, the integration exercises on page 14, or had difficulty with them, go back and do them now.)

## Derivative of the Exponential Function

The previous work shows us that if we have a series expansion for a function, it is easy to obtain a formula for the derivative of the function. We will now apply this technique to calculate the derivative and integral of the exponential function $e^x$.

There is a series expansion for the function $e^x$ that works for any value of x is but is most useful for small values of $x = \alpha << 1$, is

$$e^{\alpha} \approx 1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \cdots \qquad (136)$$

where $2! = 2 \times 1$, $3! = 3 \times 2 \times 1 = 6$, etc. (The quantities 2!, 3! are called ***factorials***. For example 3! is called *three factorial*.)

To see how well the series (136) works, consider the case $\alpha = .01$. From the series we have, up to the $\alpha^3$ term

$$\alpha = .01$$

$$\alpha^2 = .0001; \qquad \alpha^2/2 = .00005$$

$$\alpha^3 = .000001; \qquad \alpha^3/6 = .000000167$$

Giving us the approximate value

$$1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} = 1.010050167 \qquad (137)$$

When we enter .01 into a scientific calculator and press the $e^x$ button, we get exactly the same result. Thus the calculator is no more accurate than including the $\alpha^3$ term in the series, for values of $\alpha$ equal to .01 or less.

Let us now see how to use the series (136) for calculating the derivative of $e^x$. We have, from the definition of a derivative,

$$\frac{d}{dx}f(x) \equiv \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (56) \text{ repeat}$$

If $f(x) = e^x$, we get

$$\frac{d(e^x)}{dx} = \lim_{\Delta x \to 0}\left[\frac{e^{x + \Delta x} - e^x}{\Delta x}\right] \qquad (138)$$

To do this calculation, we have to evaluate the quantity $e^{x + \Delta x}$. First, we use the fact that for exponentials

$$e^{a + b} = e^a e^b$$

(Remember that $10^{2+3} = 10^2 \times 10^3 = 10^5$.) Thus

$$e^{x + \Delta x} = e^x e^{\Delta x} \qquad (139)$$

Now use the approximation formula (136), setting $\alpha = \Delta x$ and throwing out the $\alpha^2$ and $\alpha^3$ and higher terms because we are going to let $\Delta x$ go to zero

$$e^{\Delta x} \approx 1 + \Delta x \qquad (140)$$

Substituting (140) in (139) gives

$$e^{x+\Delta x} \approx e^x(1 + \Delta x)$$

$$= e^x + e^x \Delta x \qquad (141)$$

Next use (141) in (138) to get

$$\frac{d(e^x)}{dx} = \lim_{\Delta x \to 0}\left[\frac{(e^x + e^x \Delta x) - e^x}{\Delta x}\right] \qquad (142)$$

The $e^x$ terms cancel and we are left with

$$\frac{d(e^x)}{dx} = \lim_{\Delta x \to 0}\left[\frac{e^x \Delta x}{\Delta x}\right] = \lim_{\Delta x \to 0} e^x \qquad (143)$$

Since the $\Delta x$'s cancelled, we are left with the exact result

$$\boxed{\frac{d(e^x)}{dx} = e^x} \qquad (144)$$

We see that the exponential function $e^x$ has the special property that it is its own derivative.

We will often want to know the derivative, not just of the function $e^x$ but of the slightly more general result $e^{ax}$ where a is a constant. That is, we want to find

$$\frac{d}{dx}e^{ax} \qquad (a = constant) \tag{145}$$

Solving this problem provides us with our first meaningful application of the chain rule

$$\frac{df(y)}{dx} = \frac{df(y)}{dy}\frac{dy}{dx} \qquad \text{(100) repeated}$$

If we set

$$y = ax \tag{146}$$

then we have

$$\frac{de^{ax}}{dx} = \frac{de^{y}}{dy}\frac{dy}{dx} \tag{147}$$

Now

$$\frac{de^{y}}{dy} = e^{y} \tag{148}$$

$$\frac{dy}{dx} = \frac{d}{dx}(ax) = a\frac{dx}{dx} = a \times 1 = a \tag{149}$$

Using (148) and (149) in (147) gives

$$\frac{de^{ax}}{dx} = (e^{y})(a) = (e^{ax})(a) = ae^{ax}$$

Thus we have

$$\boxed{\frac{d}{dx}e^{ax} = ae^{ax}} \tag{150}$$

This result will be used so often it is worth memorizing.

---

**Exercise 6**

For further practice with the chain rule, show that

$$\frac{de^{ax^2}}{dx} = 2axe^{ax^2}$$

Do this by choosing $y = ax^2$, and then do it again by choosing $y = x^2$.

---

## Integral of the Exponential Function

To calculate the integral of $e^{ax}$, we will use the same trick as we used for the integral of $x^n$, but we will be a bit more formal this time. Let us start with Equation (128) relating position x(t) and velocity v(t) = dx(t)/dt go get

$$x(t)\Big|_{t_i}^{t_f} = \int_{t_i}^{t_f} v_x(t)\,dt = \int_{t_i}^{t_f} \frac{dx(t)}{dt}\,dt \tag{128}$$

Since Equation (128) holds for any function x(t) [we did not put any restrictions on x(t)], we can write Equation (128) in a more abstract way relating any function f(x) to its derivative df(x)/dx

$$\boxed{f(x)\Big|_{x_i}^{x_f} = \int_{x_i}^{x_f} \frac{df(x)}{dx}\,dx} \tag{151}$$

To calculate the integral of $e^{ax}$, we set $f(x) = e^{ax}$ and $df(x)/dx = ae^{ax}$ to get

$$e^{ax}\Big|_{x_i}^{x_f} = \int_{x_i}^{x_f} a\,e^{ax}dx \tag{152}$$

Dividing (157) through by (a) gives us the definite integral

$$\boxed{\int_{x_i}^{x_f} e^{ax}dx = \frac{1}{a}e^{ax}\Big|_{x_i}^{x_f}} \qquad (a = constant) \tag{153}$$

The corresponding indefinite integral is

$$\boxed{\int e^{ax}dx = \frac{e^{ax}}{a}} \qquad (a = constant) \tag{154}$$

---

**Exercise 7**

The natural logarithm is defined by the equation

$$\ln(x) = \int \left(\frac{1}{x}\right)dx \qquad \text{(see Equations 33-40)}$$

Use Equation (151) to show that

$$\boxed{\frac{d}{dx}(\ln x) = \frac{1}{x}} \tag{155}$$

(Hint—integrate both sides of Equation (155) with respect to x.)

---

## DERIVATIVE AS THE SLOPE OF A CURVE

Up to now, we have emphasized the idea that the derivative of a function f(x) is given by the limiting process

$$\frac{df(x)}{dx} = \lim_{\Delta x \to 0}\left[\frac{f(x+\Delta x)-f(x)}{\Delta x}\right] \quad (55)\ \text{repeat}$$

We saw that this form was convenient when we had an explicit way of calculating $f(x+\Delta x)$, as we did by using a series expansion. However, a lot of words are required to explain the steps involved in doing the limiting process indicated in Equation (55). In contrast, the idea of an integral as being the area under a curve is much easier to state and visualize. Now we will provide an easy way to state and interpret the derivative of a curve.

Consider the function f(x) graphed in Figure (20). At a distance x down the x axis, the curve had a height f(x) as shown. Slightly farther down the x axis, at $x+\Delta x$, the curve has risen to a height $f(x+\Delta x)$.

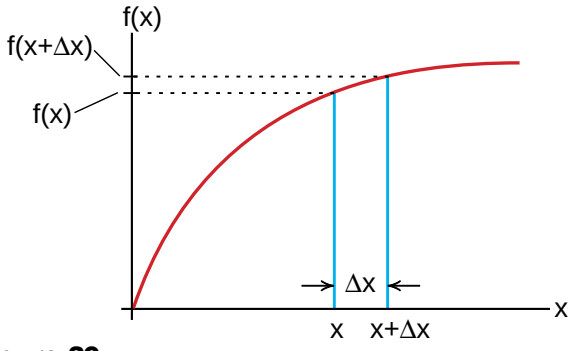



**Figure 20**

***Two points on a curve, a distance Δx apart.***

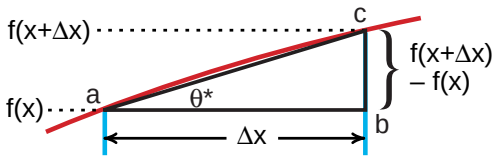



**Figure 20a**

***At this point, the curve is tilted by approximately an angle θ*.***

Figure (20a) is a blowup of the curve in the region between x and $x+\Delta x$. If the distance $\Delta x$ is sufficiently small, the curve between x and $x+\Delta x$ should be approximately a straight line and that part of the curve should be approximately the hypotenuse of the right triangle abc seen in Figure (20a). Since the side opposite to the angle $\theta^*$ is $f(x+\Delta x)-f(x)$, and the adjacent side is $\Delta x$, we have the result that the tangent of the angle $\theta^*$ is

$$\tan(\theta^*) = \frac{f(x+\Delta x)-f(x)}{\Delta x} \qquad (156)$$

When we make $\Delta x$ smaller and smaller, take the limit as $\Delta x \to 0$, we see that the angle $\theta^*$ becomes more nearly equal to the angle $\theta$ shown in Figure (21), the angle of the curve when it passes through the point x. Thus

$$\tan\theta = \lim_{\Delta x \to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x} \qquad (157)$$

The tangent of the angle at which the curve passes through the point x is called the ***slope of the curve at the point x.*** Thus from Equation (157) we see that the slope of the curve is equal to the derivative of the curve at that point. We now have the interpretation that the derivative of a curve at some point is equal to the slope of the curve at that point, while the integral of a curve is equal to the area under the curve up to that point.

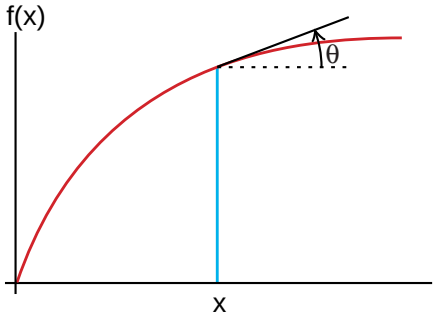



**Figure 21**

***The tangent of the angle θ at which the curve passes through the point x is called the slope of the curve at that point.***

## Negative Slope

In Figure (22) we compare the slopes of a rising and a falling curve. In (22a), where the curve is rising, the quantity $f(x + \Delta x)$ is greater than $f(x)$ and the derivative or slope

$$\frac{df(x)}{dx} = \underset{\Delta x \to 0}{\text{limit}} \left[ \frac{f(x + \Delta x) - f(x)}{\Delta x} \right]$$

is a positive number.

In contrast, for the downward curve of Figure (22b), $f(x + \Delta x)$ is less than $f(x)$ and the slope is negative. For a curve headed downward, we have

$$\frac{df(x)}{dx} = -\tan(\theta) \qquad \textit{downward heading curve} \qquad (158)$$

(For this case you can think of $\theta$ as a negative angle, so that $\tan(\theta)$ would automatically come out negative. However it is easier simply to remember that the slope of an upward directed curve is positive and that of a downward directed curve is negative.)

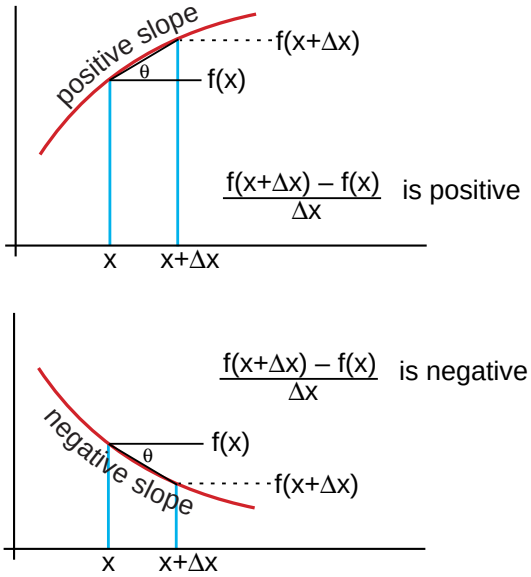



**Figure 22 a,b**

***Going uphill is a positive slope, downhill is a negative slope.***

### Exercise 8

Estimate the numerical value of the slope of the curve shown in Figure (23) at points (a), (b), (c), (d) and (e). In each case do a sketch of $\left[f(x + \Delta x) - f(x)\right]$ for a small $\Delta x$, and let the slope be the ratio of $\left[f(x + \Delta x) - f(x)\right]$ to $\Delta x$. Your answers should be roughly 1, 0, –1, $+\infty$, $-\infty$.

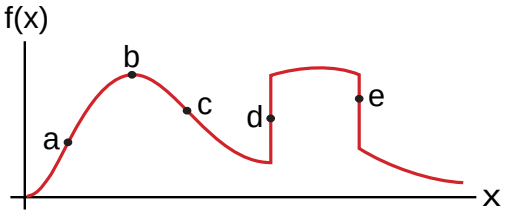



**Figure 23**

***Estimate the slope at the various points indicated.***

## THE EXPONENTIAL DECAY

A curve that we will encounter several times during the course is the function $e^{-ax}$ shown in Figure (24), which we call an exponential decay. Since exponents always have to be dimensionless numbers, we are writing the constant (a) in the form $1/x_0$ so that the exponent $x/x_0$ is more obviously dimensionless.

The function $e^{-x/x_0}$ has several very special properties. At $x = 0$, it has the numerical value 1 ($e^0 = 1$). When we get up to $x = x_0$, the curve has dropped to a value

$$e^{-x/x_0} = e^{-1} = \frac{1}{e} \quad (at\ x = x_0) \approx \frac{1}{2.7} \tag{159}$$

When we go out to $x = 2x_0$, the curve has dropped to

$$e^{-2x_0/x_0} = e^{-2} = \frac{1}{e^2} \tag{160}$$

Out at $x = 3x_0$, the curve has dropped by another factor of e to (1/e)(1/e)(1/e). This decrease continues indefinitely. It is the characteristic feature of an exponential decay.

### Muon Lifetime

In the muon lifetime experiment, we saw that the number of muons surviving decreased with time. At the end of two microseconds, more than half of the original 648 muons were still present. By 6 microseconds, only 27 remained. The decay of these muons is an example of an exponential decay of the form

$$\begin{matrix}\text{number of}\\ \text{surviving}\\ \text{muons}\end{matrix} = \left(\begin{matrix}\text{number of}\\ \text{muons at}\\ \text{time t = 0}\end{matrix}\right) \times e^{-t/t_0} \tag{161}$$

where $t_0$ is the time it takes for the number of muons remaining to drop by a factor of 1/e = 1/2.7. That time is called the muon ***lifetime***.

We can use Equation (161) to estimate the muon lifetime $t_0$. In the movie, the number of muons at the top of the graph, reproduced in Figure (25), is 648. That is at time t = 0. Down at time t = 6 microseconds, the number surviving is 27. Putting these numbers into Equation (161) gives

$$27 \begin{matrix}\text{surviving}\\ \text{muons}\end{matrix} = 648 \begin{matrix}\text{initial}\\ \text{muons}\end{matrix} \times e^{-6/t_0}$$

$$e^{-6/t_0} = \frac{27}{648} = .042 \tag{162}$$

Take the natural logarithm *ln* of both sides of Equation (162), [remembering that $ln\left(e^x\right) = x$ ] gives

$$ln\left(e^{-6/t_0}\right) = \frac{-6}{t_0} = ln(.042) = -3.17$$

where we entered .042 on a scientific calculator and pressed the *ln* key. Solving for $t_0$ we get

$$t_0 = \frac{6 \text{ microsec}}{3.17} = 1.9 \text{ microseconds} \tag{163}$$

This is close to the accepted value of $t_0 = 2.2\,0$ microseconds which has been determined from the study of many thousands of muon decays.

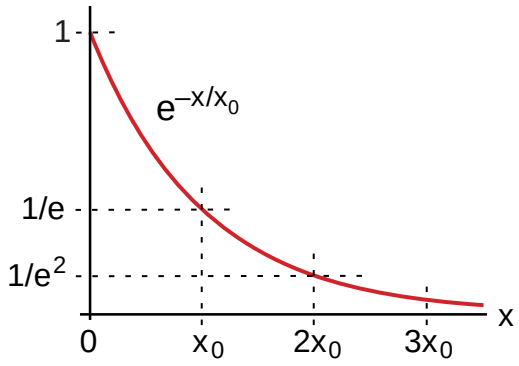



**Figure 24**

***As we go out an additional distance $x_0$, the exponential curve drops by another factor of 1/e.***

**Figure 25**

***The lifetime of each detected muon is represented by the length of a vertical line. We can see that many muons live as long as 2 microseconds (2μs), but few live as long as 6 microseconds.***

## Half Life

The exponential decay curve $e^{-t/t_0}$ decays to $1/e = 1/2.7$ of its value at time $t_0$. While $1/e$ is a very convenient number from a mathematical point of view, it is easier to think of the time $t_{1/2}$ it takes for half of the muons to decay. This time $t_{1/2}$ is called the ***half life*** of the particle.

From Figure (26) we can see that the half life $t_{1/2}$ is slightly shorter than the lifetime $t_0$. To calculate the half life from $t_0$, we have

$$e^{-t/t_0}\Big|_{t=t_{1/2}} = e^{-t_{1/2}/t_0} = \frac{1}{2} \qquad (164)$$

Again taking the natural logarithm of both sides of Equation (164) gives

$$ln\left(e^{-t_{1/2}/t_0}\right) = \frac{-t_{1/2}}{t_0} = ln\left(\frac{1}{2}\right) = -.693$$

$$\boxed{t_{1/2} = .693\, t_0} \qquad (165)$$

From Equation (165) you can see that a half life $t_{1/2}$ is about .7 of the lifetime $t_0$. If the muon lifetime is 2.2μsec (we will abbreviate microseconds as μsec), and you start with a large number of muons, you would expect about half to decay in a time of

$$(t_{1/2})_{muon} = .693 \times 2.2\mu sec = 1.5\ \mu sec$$

The basic feature of the exponential decay curve $e^{-t/t_0}$ is that for every time $t_0$ that passes, the curve decreases by another factor of 1/e. The same applies to the half life $t_{1/2}$. After one half life, $e^{-t/t_0}$ has decreased to half its value. After a second half life, the curve is down to $1/4 = 1/2 \times 1/2$. After 3 half lives it is down to $1/8 = 1/2 \times 1/2 \times 1/2$ as shown in Figure (27).

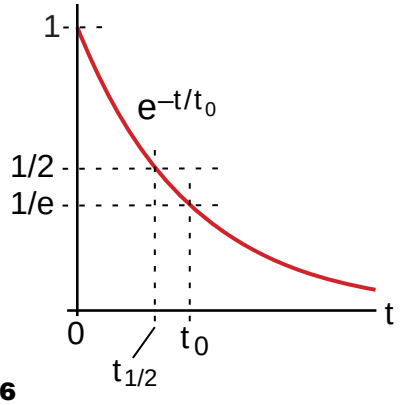



**Figure 26**

***Comparison of the lifetime $t_0$ and the half-life $t_{1/2}$.***

To help illustrate the nature of exponential decays, suppose that you started with a million muons. How long would you expect to wait before there was, on the average, only one left?

To solve this problem, you would want the number $e^{-t/t_0}$ to be down by a factor of 1 million

$$e^{-t/t_0} = 1 \times 10^{-6}$$

Taking the natural logarithm of both sides gives

$$ln\left(e^{-t/t_0}\right) = \frac{-t}{t_0} = ln\left(1\times10^{-6}\right) = -13.8 \qquad (166)$$

(To calculate $ln\left(1\times10^{-6}\right)$, enter 1, then press the *exp* key and enter 6, then press the +/– key to change it to –6. Finally press = to get the answer –13.8.)

Solving Equation (166) for t gives

$$t = 13.8\, t_0 = 13.8 \times 2.2\ \mu sec$$

$$\boxed{t = 30 \text{ microseconds}} \qquad (167)$$

That is the nature of an exponential decay. While you have nearly half a million left after around 2 microseconds, they are essentially all gone by 30 microseconds.

---

**Exercise 9**

How many factors of 1/2 do you have to multiply together to get approximately 1/1,000,000? Multiply this number by the muon half-life to see if you get about 30 microseconds.

---

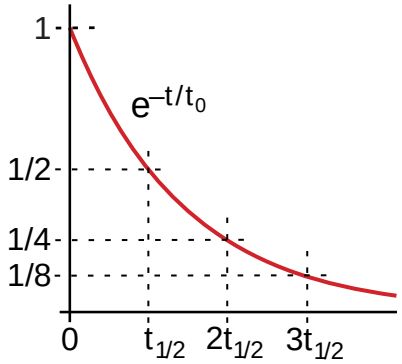



**Figure 27**

***After each half-life, the curve decreases by another factor of 1/2.***

## Measuring the Time Constant from a Graph

The idea that the derivative of a curve is the slope of the curve, leads to an easy way to estimate a lifetime $t_0$ from an exponential decay curve $e^{-t/t_0}$.

The formula for the derivative of an exponential curve is

$$\frac{de^{at}}{dt} = ae^{at} \qquad (150) \text{ repeated}$$

Setting $a = -1/t_0$ gives

$$\frac{d}{dt}\left(e^{-t/t_0}\right) = -\frac{1}{t_0}e^{-t/t_0} \qquad (168)$$

Since the derivative of a curve is the slope of the curve, we set the derivative equal to the tangent of the angle the curve makes with the horizontal axis.

$$\frac{d}{dt}\left(e^{-t/t_0}\right) = -\frac{1}{t_0}e^{-t/t_0} = \tan\theta \qquad (168a)$$

The minus sign tells us that the curve is headed down.

In Figure (28), we have drawn a line tangent to the curve at the point $t = T$. This line intersects the (t) axis (the axis where $e^{-t/t_0}$ goes to zero) at a distance (x) down the $t$ axis.

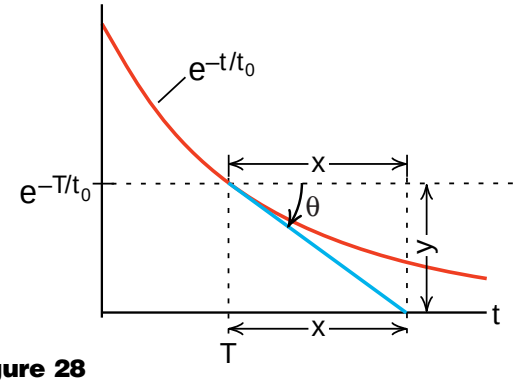



**Figure 28**

***A line, drawn tangent to the exponential decay curve at some point T, intersects the axis a distance x down the axis. We show that this distance x is equal to the time constant $t_0$. This is true no matter what point T we start with.***

The height (y) of the point where we drew the tangent curve is just the value of the function $e^{-T/t_0}$. The tangent of the angle $\theta$ is the opposite side (y) divided by the adjacent side (x)

$$\tan\theta = \frac{y}{x} = \frac{e^{-T/t_0}}{x} \qquad (169)$$

Equating the two magnitudes of $\tan\theta$ in Equations (169) in (168a) gives us

$$\frac{1}{t_0}e^{-T/t_0} = \frac{1}{x}e^{-T/t_0}$$

which requires that

$$x = t_0 \qquad (170)$$

Equation (170) tells us that the distance (x), the distance down the axis where the tangent lines intersect the axis, is simply the time constant $t_0$.

The result gives us a very quick way of determining the time constant $t_0$ of an exponential decay curve. As illustrated in Figure (29), choose any point on the curve, draw a tangent to the curve at that point and measure the distance down the axis where the tangent line intersects the axis. That distance will be the time constant $t_0$. We will use this technique in several laboratory exercises later in the course.

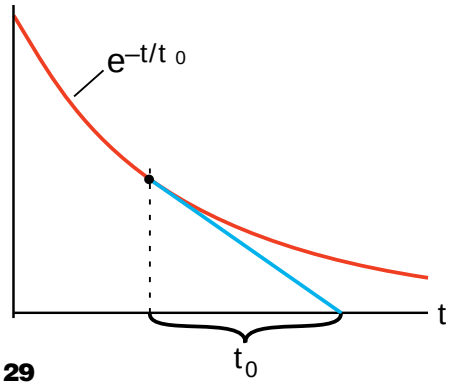



**Figure 29**

***A quick way to estimate the time constant $t_0$ for an exponential decay curve is to draw the tangent line as shown.***

## THE SINE AND COSINE FUNCTIONS

The final topic in our introduction to calculus will be the functions $\sin\theta$ and $\cos\theta$ and their derivatives and integrals. We will need these functions when we come to rotational motion and wave motion.

The definition of $\sin\theta$ and $\cos\theta$, which should be familiar from trigonometry, are

$$\sin\theta = \frac{a}{c} \qquad \left(\frac{\textit{opposite}}{\textit{hypotenuse}}\right) \qquad (171a)$$

$$\cos\theta = \frac{b}{c} \qquad \left(\frac{\textit{adjacent}}{\textit{hypotenuse}}\right) \qquad (171b)$$

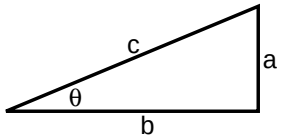



**Figure 30**

where $\theta$ is an angle of a right triangle as shown in Figure (30), (a) is the length of the side opposite to $\theta$, (b) the side adjacent to $\theta$ and (c) the hypotenuse.

The formulas are simplified if we consider a right triangle whose hypotenuse is of length $c = 1$ as in Figure (31). Then we have

$$\sin\theta = a \qquad (172a)$$

$$\cos\theta = b \qquad (172b)$$

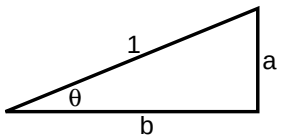



**Figure 31**

We can then fit our right triangle inside a circle of radius 1 as shown in Figure (32).

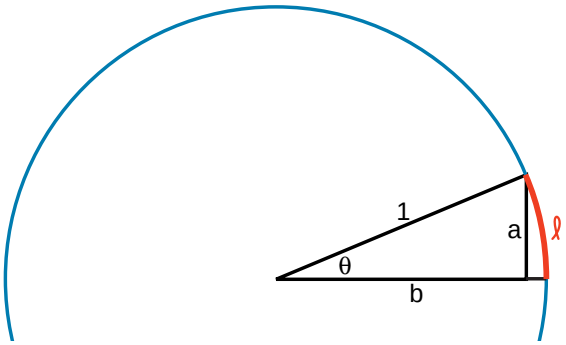



**Figure 32**
***Fitting our right triangle inside a unit radius circle.***

### Radian Measure

We are brought up to measure angles in degrees, but physicists and mathematicians usually measure angles in ***radians***. The angle $\theta$ measured in radians is defined as the ***arc length*** $\ell$ subtended by the angle $\theta$ on a circle of unit radius, as shown in Figure (32).

$$\theta_{\text{radians}} = \ell \qquad \begin{matrix}\textit{arc length subtended}\\ \textit{by } \theta \textit{ on a unit circle}\end{matrix} \qquad (173)$$

(If we had a circle of radius c, then we would define $\theta_{\text{radians}} = \ell/c$, a dimensionless ratio. In the special case $c = 1$, this reduces to $\theta_{\text{radians}} = \ell$.)

Since the circumference of a unit circle is $2\pi$, we see that $\theta$ for a complete circle is $2\pi$ radians, which is the same as 360 degrees. This tells us how to convert from degrees to radians. We have the conversion factor

$$\frac{360 \text{ degrees}}{2\pi \text{ radians}} = 57.3 \frac{\text{degrees}}{\text{radian}} \qquad (174)$$

As an example of using this conversion factor, suppose we want to convert 30 degrees to radians. We would have

$$\frac{30 \text{ degrees}}{57.3 \text{ degrees/radian}} = .52 \text{ radians} \qquad (175)$$

To decide whether to divide by or multiply by a conversion factor, use the dimensions of the conversion factor. For example, if we had multiplied 30 degrees by our conversion factor, we would have gotten

$$30 \text{ degrees} \times 57.3 \frac{\text{degrees}}{\text{radian}} = 1719 \frac{\text{degrees}^2}{\text{radian}}$$

This answer may be correct, but it is useless.

The numbers to remember in using radians are the following:

$$\begin{aligned} 90° &= \pi/2 \text{ radians} \\ 180° &= \pi \text{ radians} \\ 270° &= 3\pi/2 \text{ radians} \\ 360° &= 2\pi \text{ radians} \end{aligned} \qquad (176)$$

The other values you can work out as you need them.

## The Sine Function

In Figure (33) we have started with a circle of radius 1 and, in a somewhat random way, labeled 10 points around the circle. The arc length up to each of these points is equal to the angle, in radian measure, subtended by that point. The special values are:

$\theta_0 = 0$ radians
$\theta_4 = \pi/2$ radians (90°)
$\theta_6 = \pi$ radians (180°)
$\theta_8 = 3\pi/2$ radians (270°)
$\theta_{10} = 2\pi$ radians (360°)

In each case the $\sin\theta$ is equal to the height (a) at that point. For example

$\sin\theta_1 = a_1$
$\sin\theta_2 = a_2$
. . . . . .
$\sin\theta_{10} = a_{10}$

We see that the height (a) starts out at $a_0 = 0$ for $\theta_0$, increases up to $a_4 = 1$ at the top of the circle, drops back down to $a_6 = 0$ at $\theta_6 = \pi$, goes negative, down to $a_8 = -1$ at $\theta_8 = 3\pi/2$, and returns to $a_{10} = 0$ at $\theta_{10} = 2\pi$.

Our next step is to construct a graph in which $\theta$ is shown along the horizontal axis, and we plot the value of $\sin\theta$ = (a) on the vertical axis. The result is shown in Figure (34). The eleven points, representing the heights $a_0$ to $a_{10}$ at $\theta_0$ to $\theta_{10}$ are shown as large dots in Figure (34). We have also sketched in a smooth curve through these points, it is the curve we would get if we had plotted the value of (a) for every value of $\theta$ from $\theta = 0$ to $\theta = 2\pi$. The smooth curve is a graph of the function ***sin θ***.

---

**Exercise 10**

Using the fact that the cosine function is defined as

$\cos\theta = b$ (b is defined in Figures 31, 32)

plot the values of $b_0, b_1, \cdots, b_{10}$ on a graph similar to Figure (34), and show that the cosine function $\cos\theta$ looks like the curve shown in Figure (35).

---

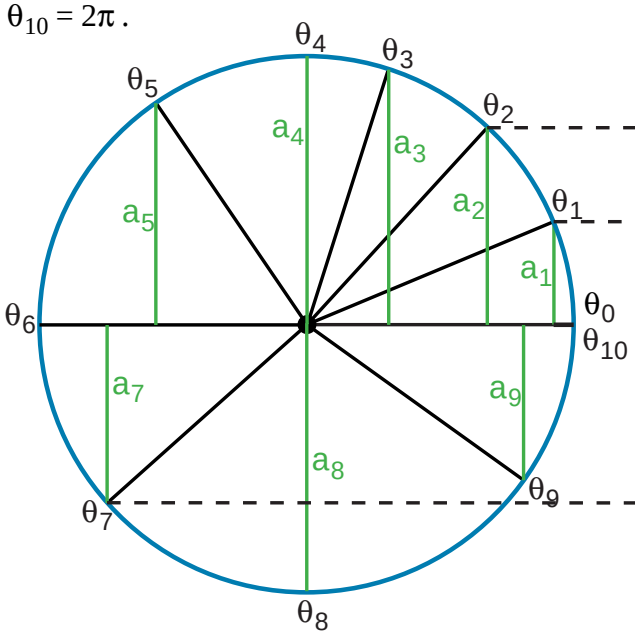



**Figure 33**
***The heights $a_i$ at various points around a unit circle.***

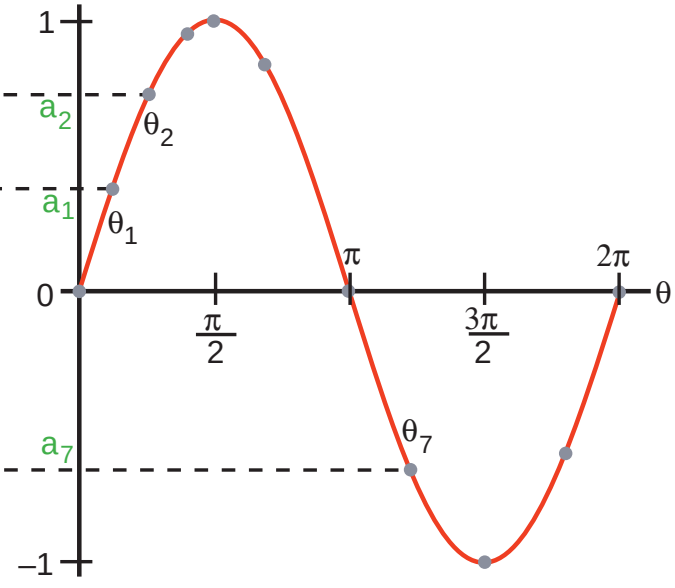



**Figure 34**
***Graph of the function sin(θ).***

There is nothing that says we have to stop measuring the angle $\theta$ after we have gone around once. On the second trip around, $\theta$ increases from $2\pi$ up to $4\pi$, and the curve $\sin\theta$ repeats itself. If we go around several times, we get a result like that shown in Figure (36). We often call that a ***sine wave***.

Several cycles of the curve $\cos\theta$ are shown in Figure (37). You can see that the only difference between a sine and a cosine curve is where you set $\theta = 0$. If you move the origin of the cosine axis back (to the left) 90° $(\pi/2)$, you get a sine wave.

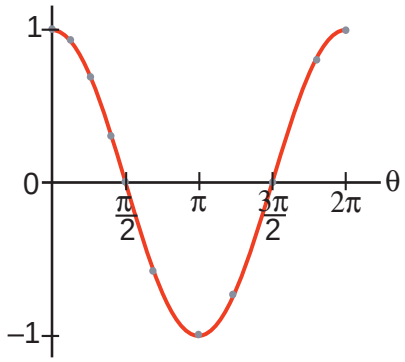



**Figure 35**
***The cosine function.***

## Amplitude of a Sine Wave

A graph of the function $y(\theta) = c\sin\theta$ looks just like the curve in Figure (36), except the curve goes up to a height c and down to –c as shown in Figure (38). We would get the curve of Figure (38) by plotting points around a circle as in Figure (33), but using a circle of radius c. We call this factor c the *amplitude* of the sine wave. The function $\sin\theta$ has an amplitude 1, while the sine wave in Figure (38) has an ***amplitude c*** (its values range from +c to –c).

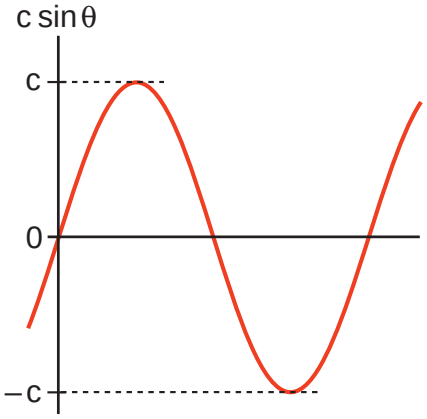



**Figure 38**
***A sine wave of amplitude c.***

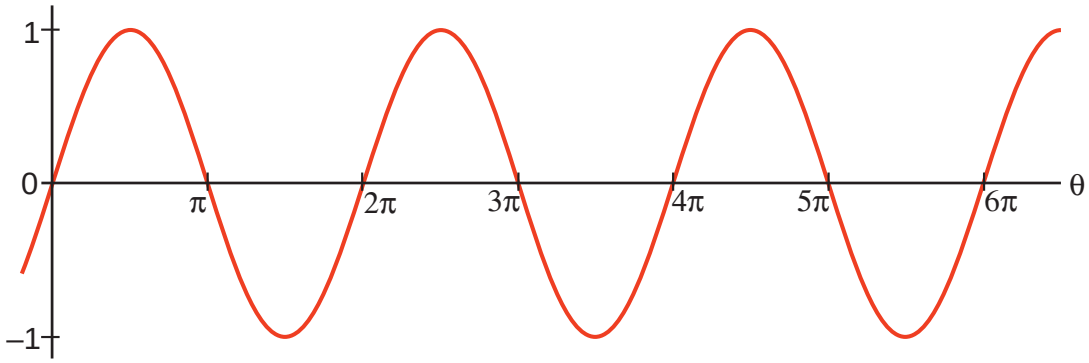



**Figure 36**
***Several cycles of the curve $\sin(\theta)$.***

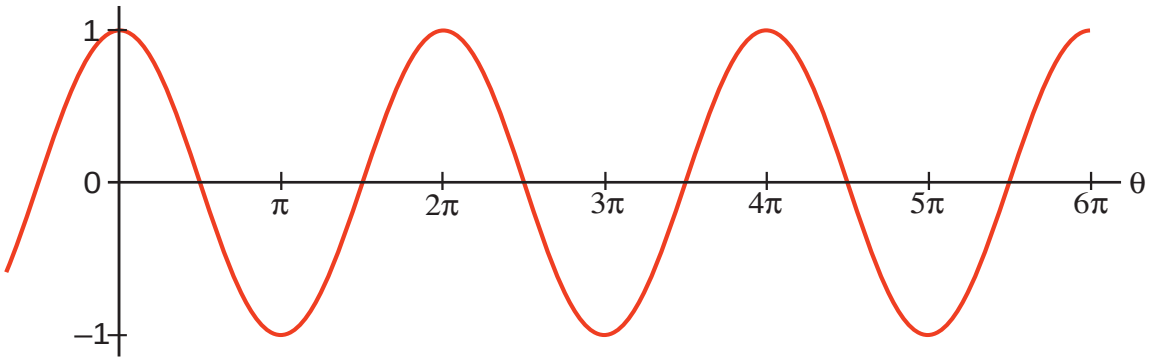



**Figure 37**
***Several cycles of the curve $\cos(\theta)$.***

## Derivative of the Sine Function

Since the sine and cosine functions are smooth curves, we should be able to calculate the derivatives and integrals of them. We will do this by first calculating the derivative, and then turning the process around to find the integral, just as we did for the functions $x^n$ and $e^x$.

The derivative of the function $\sin\theta$ is defined as usual by

$$\frac{d(\sin\theta)}{d\theta} = \lim_{\Delta\theta\to 0}\left[\frac{\sin(\theta+\Delta\theta)-\sin\theta}{\Delta\theta}\right] \qquad (177)$$

where $\Delta\theta$ is a small change in the angle $\theta$.

The easiest way to evaluate this limit is to go back to the unit circle of Figure (25) and construct both $\sin\theta$ and $\sin(\theta+\Delta\theta)$ as shown in Figure (39). We see that $\sin\theta$ is the height of the triangle with an angle $\theta$, while $\sin(\theta+\Delta\theta)$ is the height of the triangle whose center angle is $(\theta+\Delta\theta)$. What we have to do is calculate the difference in heights of these two triangles.

In Figure (40) we start by focusing our attention on the slender triangle abc with an angle $\Delta\theta$ at (a) and long sides of length 1 (since we have a unit circle). Since the angle $\Delta\theta$ is small, the short side of this triangle is essentially equal to the arc length along the circle from point (b) to point (c). And since we are using radian measure, this arc length is equal to the angle $\Delta\theta$.

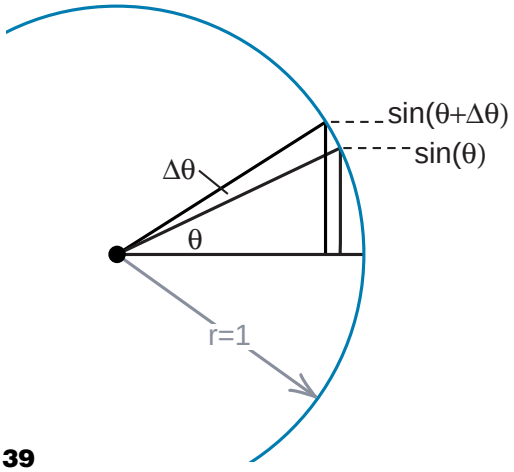



**Figure 39**
***Triangles for the $\sin\theta$ and the $\sin(\theta+\Delta\theta)$.***

Now draw a line vertically down from point (c) and horizontally over from point (b) to form the triangle bcd shown in Figure (40). The important point is that the angle at point (c) in this tiny triangle is the same as the angle $\theta$ at point (a). To prove this, consider the sketch in Figure (41). A line bf is drawn tangent to the circle at point (b), so that the angle abf is a right angle. That means the other two angles in the triangle add up to 90°, the total angle in any triangle being 180°

$$\theta + \varphi = 90° \qquad (178)$$

Since the angle at (e) in triangle bef is also a right angle, the other two angles in the triangle bef, must also add up to 90°.

$$\alpha + \varphi = 90° \qquad (179)$$

For both Equations (178) and (179) to be true, we must have $\alpha = \theta$.

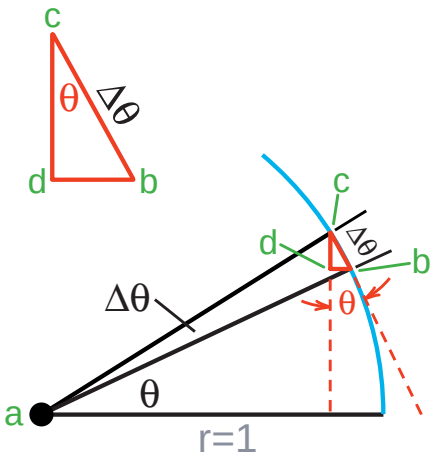



**Figure 40**
***The difference between $\sin\theta$ and $\sin(\theta+\Delta\theta)$ is equal to the height of the side cd of the triangle cdb.***

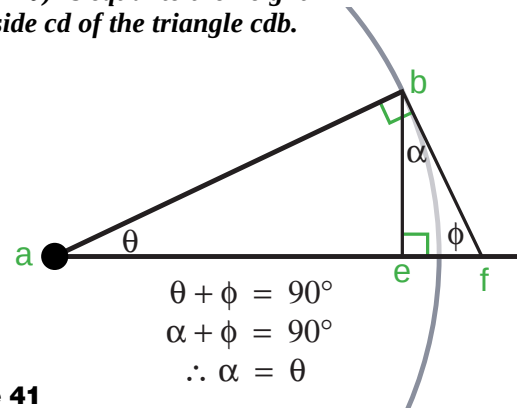



**Figure 41**
***Demonstration that the angle $\alpha$ equals the angle $\theta$.***

The final step is to note that when $\Delta\theta$ in Figure (40) is very small, the side cb of the very small triangle is essentially tangent to the circle, and thus parallel to the side bf in Figure (41). As a result the angle between cb and the vertical is also the same angle $\theta$.

Because the tiny triangle, shown again in Figure (42) has a hypotenuse $\Delta\theta$ and a top angle $\theta$, the vertical side, which is equal to the difference between $\sin\theta$ and $\sin(\theta+\Delta\theta)$ has a height $(\cos\theta)\Delta\theta$. Thus we have

$$\sin(\theta+\Delta\theta) - \sin\theta = (\cos\theta)\Delta\theta \qquad (180)$$

Equation (180) becomes exact when $\Delta\theta$ becomes an infinitesimal angle.

We can now evaluate the derivative

$$\frac{d(\sin\theta)}{d\theta} = \lim_{\Delta\theta\to 0}\left[\frac{\sin(\theta+\Delta\theta)-\sin\theta}{\Delta\theta}\right]$$

$$= \lim_{\Delta\theta\to 0}\left[\frac{(\cos\theta)\Delta\theta}{\Delta\theta}\right]$$

$$= \lim_{\Delta\theta\to 0}\cos\theta$$

Thus we get the exact result

$$\boxed{\frac{d}{d\theta}(\sin\theta) = \cos\theta} \qquad (181)$$

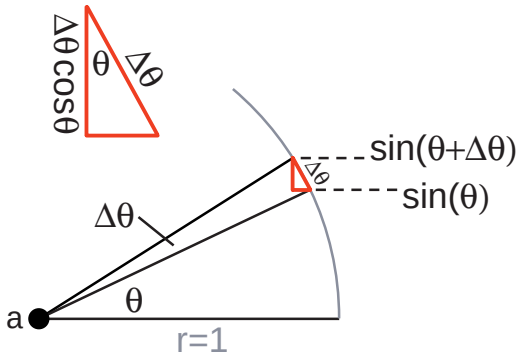



**Figure 42**
***The difference between $\sin\theta$ and $\sin(\theta+\Delta\theta)$ is equal to $\Delta\theta\cos\theta$. Check that this result is reasonable by considering the special cases $\theta = 0$ and $\theta = 90°$ $(\pi/2)$.***

**Exercise 11**

Using a similar derivation, show that

$$\boxed{\frac{d}{d\theta}(\cos\theta) = -\sin\theta} \qquad (182)$$

**Exercise 12**

Using the chain rule for differentiation, show that

$$\boxed{\begin{aligned}\frac{d}{d\theta}(\sin a\theta) &= a\cos a\theta\\ \frac{d}{d\theta}(\cos a\theta) &= -a\sin a\theta\end{aligned}} \quad (a = \textit{constant}) \qquad (183)$$

(Hint—if you need to, look at Equation (145) through (150).

**Exercise 13**

Using the fact that integration reverses differentiation, as we did in integrating the function $e^x$ (Equations (151) through (154), show that

$$\boxed{\begin{aligned}\int_{\theta_i}^{\theta_f}(\cos a\theta)d\theta &= \frac{1}{a}\sin a\theta\Big|_{\theta_i}^{\theta_f} \qquad (184a)\\ \int_{\theta_i}^{\theta_f}(\sin a\theta)d\theta &= -\frac{1}{a}\cos a\theta\Big|_{\theta_i}^{\theta_f} \qquad (184b)\end{aligned}} \quad (a = \text{constant})$$

Use sketches of the integrals from $\theta_i = 0$ to $\theta_f = \pi/2$ to show that Equations (184a) and (184b) have the correct numerical sign. (Explicitly explain the minus sign in (184b).

# Calculus 2000-Chapter 2

# Second Derivatives and the One Dimensional Wave Equation

*In our discussion of a wave pulse on a rope, in Chapter 15 of Physics 2000, we used a combination of physical observation and a somewhat tricky argument to show that the speed of the wave pulse was given by the formula* $v = \sqrt{T/\rho}$ . *The physical observation was noting that a pulse travels down the rope at an apparently uniform speed. The trick was to analyze the behavior of the rope from the point of view of someone moving along with the pulse (as on pages 15-4, 5).*

*Another way to handle the problem is to directly apply Newton's second law to a section of the rope. When we use this direct approach, we end up with an equation that involves second derivatives not only with respect in time, but also with respect to space. The resulting equation with its second derivatives is what is known as the* **wave equation**. *The aim of this chapter is to learn how to handle the wave equation, at least for waves moving in one dimension. (Handling three dimensional wave equations comes later.)*

*To use the wave equation with any real understanding, not just manipulating formulas, requires more familiarity with the properties of a second derivative than we have needed so far. Thus we will begin this chapter with a discussion of the second derivative, and then go on to the one dimensional wave equation.*

## THE SECOND DERIVATIVE

We have already encountered the idea of a *second derivative* in our discussion of velocity and acceleration. Consider a particle moving down the x axis, whose position is described by the function x(t). The particle's velocity $v_x(t)$ is given by

$$v_x(t) = \frac{dx(t)}{dt} \qquad \textit{first derivative} \tag{1}$$

which is the first derivative, with respect to time, of x(t). The particle's acceleration a(t) is given by

$$a_x(t) = \frac{dv_x(t)}{dt} \tag{2}$$

When we use (1) for v(t) in Equation (2) we get

$$a_x(t) = \frac{d}{dt}\left[\frac{dx(t)}{dt}\right] \qquad \textit{second derivative} \tag{3}$$

In Equation (3), we see that $a_x(t)$ is obtained from x(t) by differentiating twice with respect to time. We say that $a_x(t)$ is the second derivative of x(t) and use the simplified notation

$$\frac{d}{dt}\left[\frac{dx(t)}{dt}\right] \equiv \frac{d^2x(t)}{dt^2} \qquad \textit{simplified notation for second derivative} \tag{4}$$

With this notation, the position x(t), velocity $v_x(t)$, and acceleration $a_x(t)$ are related by

x(t)

$$v_x(t) = \frac{dx(t)}{dt}$$

$$a_x(t) = \frac{d^2x(t)}{dt^2} \tag{5}$$

There is nothing particularly difficult about carrying out a second derivative, just do the derivative operation twice as illustrated in the following example.

### Example 1

Calculate the second derivative, with respect to θ, of sin(aθ)

$$\frac{d^2\sin(a\theta)}{d\theta^2} = ? \tag{6}$$

**Solution**

Begin by taking the first derivative

$$\frac{d\sin(a\theta)}{d\theta} = a\cos(a\theta) \tag{7}$$

Now differentiate again

$$\frac{d}{d\theta}\left(\frac{d\sin(a\theta)}{d\theta}\right) = \frac{d}{d\theta}\left[a\cos(a\theta)\right]$$

$$= a\frac{d}{d\theta}\left[\cos(a\theta)\right]$$

$$= a\left[-a\sin(a\theta)\right] \tag{8}$$

Thus we get

$$\frac{d^2\sin(a\theta)}{d\theta^2} = -a^2\sin(a\theta) \tag{9}$$

We see that the second derivative of a sine curve is itself a sine curve, with a minus sign.

---

**Exercise 1**

Calculate the following second derivatives

(a) $\frac{d^2}{d\theta^2}\left[\cos(a\theta)\right]$

(b) $\frac{d^2}{dx^2}\left[e^{-ax}\right]$

(c) $\frac{d^2}{dx^2}\left[\ln(x)\right]$

(d) $\frac{d^2}{dx^2}\left[\ln(ax)\right]$

(e) $\frac{d^2}{dx^2}\left[x^n\right]$

(f) $\frac{d^2}{dx^2}\left[(ax)^n\right]$

(g) $\frac{d^2}{dx^2}\left[\frac{1}{(x)^n}\right]$

(h) $\frac{d^2}{dx^2}\left[\frac{1}{(ax)^n}\right]$

---

## Geometrical Interpretation of the Second Derivative

We have seen that the various calculus operations have a geometrical interpretation. Integration was equivalent to finding the area under a curve, while the first derivative represented the slope of a curve. We now want to obtain the geometrical interpretation of the second derivative. We will see that the second derivative is equal to what we will call the *curvature* of the curve. To see exactly what that is, consider the following derivation.

Let y(x) be the section of a circle as shown in Figure (1). Let us use notation found in a number of calculus texts, and denote the derivative of y(x) by $y'(x)$

$$y'(x) \equiv \frac{dy(x)}{dx} \qquad \begin{matrix}\textit{simplified}\\ \textit{notation}\end{matrix} \qquad (10)$$

In terms of $y'(x)$, the second derivative is

$$\frac{d^2y(x)}{dx^2} = \lim_{\Delta x \to 0} \frac{y'(x+\Delta x) - y'(x)}{\Delta x} \qquad (11)$$

Remember that $y'(x) = dy/dx$ is the slope of the curve at position x as shown in Figure (2) (For example, see Figure 21 of Chapter 1). Thus Equation (11) tells us that to find the second derivative of y(x) we have to find the change in slope as we move from x to $x + \Delta x$.

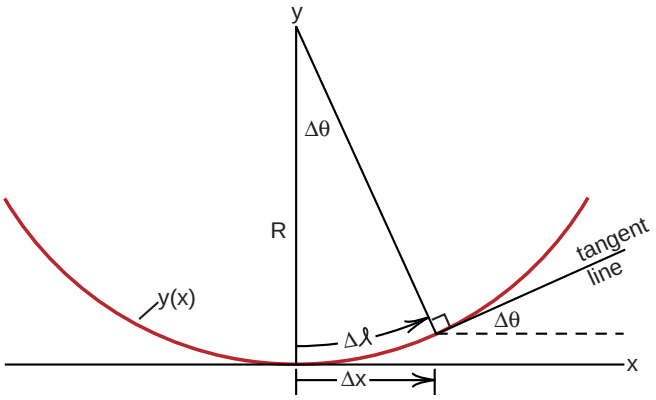



**Figure 1**
***Calculating the change in the slope of the circle, as we go from x = 0 to x = Δx.***

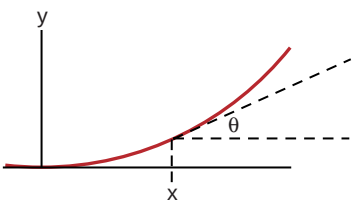



**Figure 2**
***The slope is the tangent of the angle.***

$$\textit{slope at } x \equiv \frac{dy(x)}{dx} = \tan\theta$$

We will evaluate the second derivative at the bottom of the circle, where the curve is horizontal and the slope is zero.

$$y'[x = 0] = 0 \qquad \begin{matrix}\textit{curve horizontal}\\ \textit{at } x = 0\end{matrix} \qquad (12)$$

Now move down the x axis a distance $\Delta x$ as shown in Figure (1). If $\Delta x$ is small, then $\Delta x$ is essentially equal to the arc length $\Delta\ell$ along the circle, and the angle $\Delta\theta$ in radian measure is the arc length divided by the radius R of the circle

$$\Delta\theta = \frac{\Delta\ell}{R} \approx \frac{\Delta x}{R} \qquad (13)$$

If we draw a line tangent to the circle at position $x = \Delta x$, this tangent line will make an angle $\Delta\theta$ to the horizontal as shown in Figure (1). (The two angles labeled $\Delta\theta$ in Figure 1 are equal no matter how big $\Delta\theta$ is.) Thus the slope of the tangent line at $x = \Delta x$ is

$$\begin{matrix}\text{slope of circle}\\ \text{at } x = \Delta x\end{matrix} = \tan(\Delta\theta) \qquad (14)$$

Now if $\Delta\theta$ is a small angle, which it will be if we take the limit as $\Delta x \to 0$, we can use the approximation

$$\tan(\Delta\theta) \approx \Delta\theta \qquad (15)$$

You can see why this approximation is good for small angles from Figure (2a).

Thus the slope of the tangent line at $x = \Delta x$ is given by

$$\begin{matrix}\text{slope of}\\ \text{tangent line}\\ \text{at } x = \Delta x\end{matrix} = y'[x = \Delta x] = \Delta\theta = \frac{\Delta x}{R} \qquad (16)$$

where we used Equation (13) for $\Delta\theta$.

Now we have values of $y'$ at $x = 0$ (Equation 12) and at $x = \Delta x$ (Equation 16), we can use these values in Equation (10) to get the value of $d^2y/dx^2$ at $x = 0$, i.e., at the bottom of the circle.

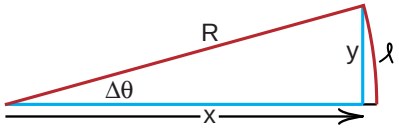



**Figure 2a**
***For small angles, the angle and the tangent of the angle are essentially equal.***

$$\tan(\Delta\theta) = \frac{y}{x}$$

$$\Delta\theta = \frac{\ell}{R} \approx \frac{y}{x}$$

Introducing the notation

$$\left.\frac{d^2y(x)}{dx^2}\right|_{x=0} \qquad \begin{array}{l}\textit{means } d^2y(x)/dx^2 \\ \textit{evaluated at the} \\ \textit{point } x=0\end{array}$$

We have from Equation (10)

$$\left.\frac{d^2y(x)}{dx^2}\right|_{x=0} = \lim_{\Delta x \to 0} \frac{y'[x=\Delta x] - y'[x=0]}{\Delta x} \tag{17}$$

With $y'[x=\Delta x] = \Delta x/R$ (Equation 16) and $y'[x=0] = 0$, we get

$$\left.\frac{d^2y(x)}{dx^2}\right|_{x=0} = \lim_{\Delta x \to 0} \frac{\Delta x/R - 0}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{1}{R}$$

Since the $\Delta x$'s canceled, we see that 1/R is the limiting value and we get

$$\left.\frac{d^2y(x)}{dx^2}\right|_{x=0} = \frac{1}{R} \tag{18}$$

With a slightly messier derivation we could calculate $d^2y/dx^2$ anywhere around the circle, not just at the bottom, and we get the same answer 1/R. Thus we have the more general result

$$\boxed{\frac{d^2y(x)}{dx^2} = \frac{1}{R}} \qquad \textit{anywhere around the circle} \tag{19}$$

## CURVATURE

Consider the curve shown in Figure (3) representing some function y(x). At point $x_0$ we have drawn a circle that just fits against the curve. The radius of the circle is adjusted to give the closest match possible between the curve $y(x_0)$ and the circle. When we get this closest fit, both the first and the second derivatives of the circle and y(x) are equal at $x = x_0$. In other words

$$\boxed{\left.\frac{d^2y(x)}{dx^2}\right|_{x=x_0} = \frac{1}{R}} \tag{20}$$

In Figure (3) the quantity R is called the ***radius of curvature*** of the curve y(x) at the point $x_0$, and 1/R is called the ***curvature***

$$\frac{1}{R} \equiv \textbf{\textit{curvature}} \text{ of the curve} \tag{21}$$

You can see intuitively why 1/R is called curvature. If R is very large, the curve is almost flat and we would say it has little curvature. As R becomes smaller, the curve bends in a tighter circle, and the curvature 1/R becomes greater.

This is the geometrical picture of the second derivative. While the first derivative was equal to the slope of the curve at some point, the ***second derivative is equal to the curvature*** of the curve at that point. The curvature is explicitly the reciprocal of the radius of curvature of the curve where the radius of curvature is found by fitting a circle to the curve as in Figure (3). [*Exercise: under what circumstances would the second derivative or curvature be negative?*]

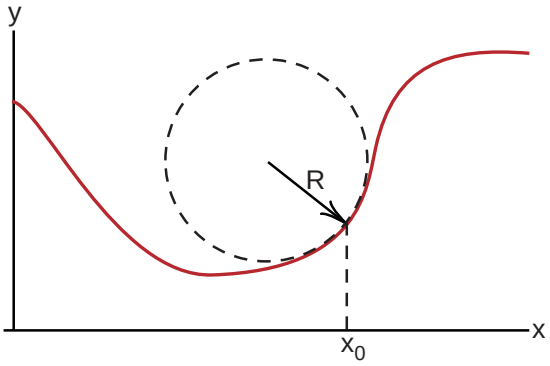



**Figure 3**
***At any point along a curve, the curvature is 1/R or –1/R, where R is the radius of the circle that just fits the curve as shown.***

## Curve Fitting and Boat Lofting

The problem of working with curves has a number of practical applications, one of the more interesting of which, at least to a sailor, is the *lofting* of boats.

It turns out that the eye is extremely good at judging the smoothness of a curve. We can, for example, easily spot the slightest wrinkle in what is supposed to be the smooth side of a boat. (It is an interesting question as to how the eye and brain can do this so well.)

Through the 16th century, boats were rather crude looking. Starting in the 17th century, better looking boats were built using the following steps. The first was to carve a model of the hull that was to be built. Then conceptually slice the model as you would slice a loaf of bread. Each of these cuts was called a *station*. Typically one used about 15 stations, each representing a cross section of the hull at different distances along the length of the boat. Then points were taken from the model to represent the shape of the hull at each station. Figure (4) is a typical example of a hull cross section at a middle station.

Since the points showing the shape of the hull were taken from a small model, any errors in measurement would be greatly magnified when the hull was laid out full scale. An error of a fraction of a millimeter in measuring the model would lead to a very obvious bump in the final hull shape.

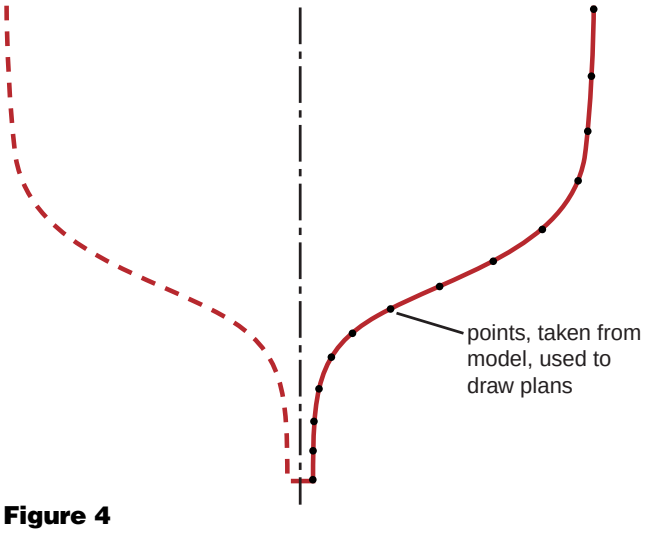



**Figure 4**
***Typical cross section. (Since boats are supposed to be symmetric, only one side is usually drawn.)***

To avoid these bumps, the plans were taken up into the loft of the boat shed (hence the name *lofting*), and drawn full scale. Wooden splines, typically thin strips of spruce, were bent along the points of the curve. Since the splines bent along smooth curves, any points that were out of place would not be fitted by the spline and the points would be moved to fit the smooth curve. This process is called ***spline fitting***. Once all the full scale curves were smoothed by spline fitting, then the boat hull was constructed using these smoothed plans and the result, if done correctly, was a smooth, good looking hull.

In the early 1970's, shortly after we had started using the computer in teaching introductory physics, we had lunch with a boat builder who described the rather tedious process of lofting a boat. He wondered if lofting could be done more easily on the computer. This was before the availability of inexpensive line plotters, so that the work would all have to be done numerically. We agreed to try, the incentive being a reduced price on a diesel engine for our boat if we successfully lofted the boat builder's new lobster boat design.

The most successful part of the project was finding an easy and very effective way to spot a smooth curve. Just print out a list of the third derivatives of the curve. Since the second derivative is the curvature of the curve, the third derivative is the rate at which the curvature is changing as you go along the curve. If the curvature changes slowly, then the curve looks smooth. A bump represents a sudden change in curvature and therefore has a large third derivative. What a spruce spline essentially does is to minimize the third derivative.

About the same time that we wrote the lofting program, a physicist, Peter Karos in Germany, also wrote a boat lofting program. As one does not make much of a living from a lofting program, Karos turned to the problem of using the computer to create letter forms. The letters of the alphabet are constructed from different curves that depend upon which font you are using. And just as in boat design, the eye is very sensitive to the smoothness of the curves, even for relatively small letters.

Karos based his boat lofting and letter design programs on what are called ***Besier*** curves. To construct a Besier curve through a series of points, at each point you specify the location y(x) of the point, the first derivative $y'(x) = dy/dx$, and the second derivative $y''(x) = d^2y/dx^2 = 1/R$. The section of curve between two adjacent points is then constructed to match the first and second derivative at the end points and minimize the third derivative in the region in between. This uniquely determines the line.

Karos's techniques using Besier curves was built into the Postscript™ language used for letter design. A way of graphically handling the construction of Besier curves was developed and became the basis of the *Adobe Illustrator*™ program.

Those of you who have used Adobe Illustrator, or any of the similar drawing programs, will be familiar with the constructing of Besier curves. You place the pen tool at a point and press the mouse button. That establishes the point y(x). Then you drag the pen tool in some direction. That direction establishes the slope of the curve $y'(x)$ at that point. How far out you drag the pen tool before you let up on the mouse button determines the radius of curvature R at that point, and thus establishes the second derivative $y''(x) = d^2y/dx^2 = 1/R$ there (see Figure 5). When you move the mouse to another point, press the mouse button and drag, you determine y(x), $y'(x)$ and $y''(x)$ at the new point, and then the computer draws the smooth Besier curve between the two points.

When you are using Adobe Illustrator, or other drawing programs, think of the fact that you are controlling the position, the first derivative, and the second derivative every time you place and drag the mouse.

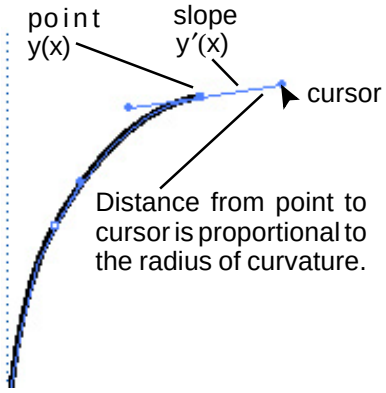



**Figure 5**
***Constructing Besier curves with Adobe Illustrator®. In that program, the radius of curvature is set to about 60% of the distance that the cursor is pulled out from the point.***

## THE BINOMIAL EXPANSION

We have seen, starting in Chapter 1 of the Physics text, the usefulness of the binomial expansion

$$(1 + \alpha)^n = 1 + n\alpha + \frac{n(n-1)}{2!}\alpha^2 + \cdots \qquad (22)$$

which is valid for any value of $\alpha$ less than one, but which gets better as $\alpha$ becomes smaller. For very small $\alpha$, we could neglect all terms involving $\alpha^2$ or higher powers of $\alpha$, giving us the approximation formula

$$(1 + \alpha)^n \approx 1 + n\alpha \qquad (\alpha << 1) \qquad (23)$$

which is good for any value of n.

With calculus, we can easily derive the formula for the various terms in the binomial expansion. We begin with the assumption that the quantity $(1 + \alpha)^n$ can be expanded in some kind of a series involving powers of $\alpha$. We will write the series in the form

$$(1 + \alpha)^n = A_0\alpha^0 + A_1\alpha^1 + A_2\alpha^2 + A_3\alpha^3 + \cdots \qquad (24)$$

where the $A_0$, $A_1$, $A_2$, etc. are unknown coefficients that we have to determine.

Equation (24) is supposed to be correct for small values of $\alpha$ including $\alpha = 0$. Setting $\alpha = 0$ gives

$$(1 + 0)^n = A_0 0^0 + A_1 0^1 + A_2 0^2 + A_3 0^3 + \cdots \qquad (25)$$

Here is a peculiar convention we use. We assume that any number $x^0 = 1$ no matter what x is, including $0^0$. Thus $A_0 0^0 = A_0$, all the other terms on the right side of Equation (25) are zero, and we get

$$1^n = 1 = A_0 \qquad (26)$$

which determines $A_0$.

(Writing $A_0\alpha^0$ instead of just $A_0$ for the first term in the series is formalism that makes the series look more consistent, but is unnecessary if you do not like the idea of $0^0 = 1$.)

To determine the value of $A_1$, differentiate Equation (24) with respect to $\alpha$. We get, using the chain rule

$$\frac{d}{d\alpha}(1+\alpha)^n = \left[\frac{d}{d(1+\alpha)}(1+\alpha)^n\right]\left[\frac{d(1+\alpha)}{d\alpha}\right]$$
$$= \left[n(1+\alpha)^{n-1}\right][1]$$
$$= n(1+\alpha)^{n-1} \qquad (27)$$

Differentiating the right hand side of Equation (24) gives

$$\frac{d}{d\alpha}\left(A_0\alpha^0 + A_1\alpha^1 + A_2\alpha^2 + A_3\alpha^3 + \cdots\right)$$
$$= 0 + A_1 + 2A_2\alpha + 3A_3\alpha^2 + \cdots \qquad (28)$$

Thus the first derivative of Equation (24), with respect to $\alpha$, is

$$n(1+\alpha)^{n-1} = A_1 + 2A_2\alpha + 3A_3\alpha^2 + \cdots \qquad (29)$$

Now set $\alpha = 0$ and we get

$$n(1+0)^{n-1} = A_1 + 2A_2\times 0 + 3A_3\times 0^2 + \cdots \qquad (30)$$

which gives us

$$n = A_1 \qquad (31)$$

and determines the coefficient $A_1$.

To determine $A_2$, differentiate Equation (29) with respect to $\alpha$. With

$$\frac{d}{d\alpha}(1+\alpha)^{n-1} = (n-1)(1+\alpha)^{n-2}$$

we get

$$n\left[(n-1)(1+\alpha)^{n-2}\right] = 2A_2 + 3(2\alpha) + \cdots \qquad (32)$$

Setting $\alpha = 0$ gives

$$A_2 = \frac{n(n-1)}{2} \qquad (33)$$

**Exercise 2**

Differentiate Equation (32) with respect to $\alpha$, set $\alpha = 0$, and show that $A_3$ is given by

$$A_3 = \frac{n(n-1)(n-2)}{3\times 2\times 1} \qquad (34)$$

From Equation (34) you can see the general formula emerging

$$A_n = \frac{n(n-1)(n-2)(n-3)\cdots}{n!} \qquad (35)$$

Thus by successive differentiation we can rather easily determine all the terms in the binomial expansion.

(One thing we have not worried about, but which is of major concern in calculus texts, is the range of values of $\alpha$ for which the series is valid. Such questions are important from a purely mathematical point of view, but are seldom of practical importance. From a practical point of view, you can usually evaluate a few terms, and if the last ones are negligibly small, the series is probably good enough.)

## The Taylor Series Expansion

The binomial expansion we have just discussed is a special case of the more general expansion called the *Taylor series expansion*. In Figure (6) we have sketched a curve representing some function

$$y = f(x) \qquad (36)$$

Suppose we know everything about the function at the point $x_0$ and would like to figure out where the curve is going as we move away from that point. By knowing everything about f(x) at the point $x_0$, we mean that we know $f(x_0)$ as well as all the derivatives of f(x) evaluated at $x = x_0$.

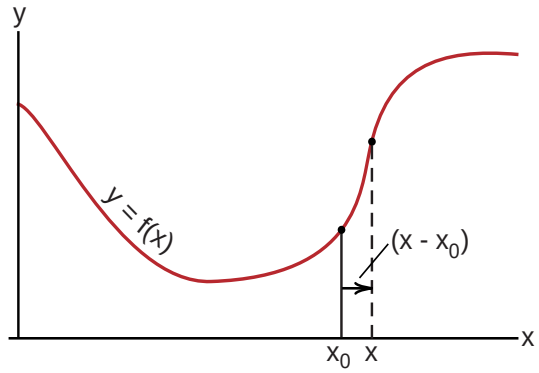



**Figure 6**
***If we know everything about the curve $y = f(x)$ at the point $x_0$, can we predict where the curve will be a short distance farther down the x axis?***

The derivation of the Taylor series expansion begins with the assumption that the function f(x) can be expanded, in the vicinity of the point $x_0$ by the so called ***power series***

$$f(x) = A_0(x-x_0)^0 + A_1(x-x_0)^1 + A_2(x-x_0)^2 + \cdots \tag{37}$$

If you think of $(x - x_0)$ as being some small distance $\alpha$, then the expansion in Equation (37) is the same form as the expansion of the function $(1 + \alpha)^n$ back in Equation (24). The difference is that for different functions f(x) we get different coefficients $A_n$.

To calculate the $A_n$, we do the same thing that we did in deriving the binomial expansion. We differentiate both sides of the equation and then set $x = x_0$ (which corresponds to setting $\alpha = (x - x_0)$ equal to zero).

First we set $x = x_0$ in Equation (37) to get

$$\begin{aligned} f(x_0) &= A_0(x-x_0)^0 + A_1(x-x_0)^1 + \cdots \\ &= A_0(0)^0 + A_1(0)^1 + \cdots \\ &= A_0 \end{aligned} \tag{38}$$

which determines the first coefficient $A_0$.

Differentiating both sides of Equation (37) with respect to x and then setting $x = x_0$ gives

$$f'(x) \equiv \frac{df(x)}{dx} = A_1 + 2A_2(x-x_0) + 3A_3(x-x_0)^3 + \cdots \tag{39}$$

where we used the chain rule to show that

$$\frac{d}{dx}(x-x_0)^n = n(x-x_0)^{n-1} \tag{40}$$

Setting $x = x_0$ in Equation (39) gives

$$f'(x_0) \equiv \left.\frac{df(x)}{dx}\right|_{x=x_0} = A_1 \tag{41}$$

all the other terms being zero.

---

**Exercise 3**

Show that

$$A_2 = \frac{1}{2}\left.\frac{d^2f(x)}{dx^2}\right|_{x=x_0} = \frac{f''(x_0)}{2} \tag{42}$$

$$A_3 = \frac{1}{3\times 2\times 1}\left.\frac{d^3f(x)}{dx^3}\right|_{x=x_0} = \frac{f'''(x_0)}{3!} \tag{43}$$

---

From Exercise 3 you can see that the general form of the Taylor series expansion is

$$\begin{aligned} f(x-x_0) &= f(x_0) + f'(x_0)(x-x_0)^1 \\ &\quad + \frac{1}{2!}f''(x_0)(x-x_0)^2 \\ &\quad + \frac{1}{3!}f'''(x_0)(x-x_0)^3 + \cdots \end{aligned}$$

This can be written in the compact form

$$\boxed{f(x-x_0) = \sum_{n=0}^{\infty}\frac{f^n(x_0)}{n!}(x-x_0)^n} \quad \textit{Taylor series expansion} \tag{44}$$

where we used the notation

$$f^n(x_0) \equiv \left.\frac{d^nf(x)}{dx^n}\right|_{x=x_0} \tag{45}$$

The tricky part of the mathematics of the Taylor series expansion is how far you can go, how far x can be away from $x_0$, and still have a valid expansion. Perhaps more important to the physicist is how far you can go before you have to include too many terms and the expansion is not useful.

---

**Exercise 4**

Apply the Taylor series expansion, Equation (44) to the function

$$f(x) = (x-x_0)^n$$

evaluated at $x_0 = 1$, and show that you get the binomial expansion. (Hint—set $\alpha = x - x_0$, i.e., substitute $x = x_0 + \alpha$ at the end.)

---

## The Constant Acceleration Formulas

While the Taylor series expansion, Equation (44), looks like a very new topic, we have been using a Taylor series expansion since the very beginning of our discussion of calculus. The constant acceleration formulas are a simple example of this expansion.

Figure (7) is a reproduction of our instantaneous velocity drawing, Figure (3-32d) from Chapter 3 of the Physics text and Figure (Cal 1-1d) of the Calculus text. At some instant of time, the ball is located at some position $(x_0, y_0)$ at time $t_0$, and we wish to predict the position of the ball at some later time t.

The location of the ball is described by two functions x(t) and y(t). We know $x(t_0)$, $y(t_0)$ and all the derivatives of these functions at time $t_0$, they are simply the velocity and acceleration

$$x'(t) = \frac{dx(t)}{dt} = v_x(t) \tag{46}$$

$$x''(t) = \frac{d^2x(t)}{dt^2} = a_x(t) \tag{47}$$

$$y'(t) = \frac{dy(t)}{dt} = v_y(t) \tag{48}$$

$$y''(t) = \frac{d^2y(t)}{dt^2} = a_y(t) \tag{49}$$

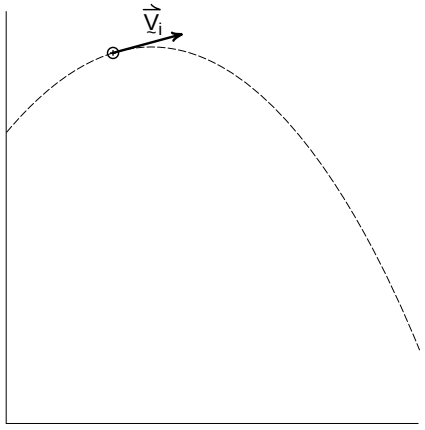



**Figure 7**
***Instantaneous velocity at time (t).***

If the particle is moving with constant acceleration, then all higher derivatives are zero. For example

$$y'''(t) \equiv \frac{d^3y(t)}{dt^3} = \frac{da_y(t)}{dt} = 0 \quad \begin{matrix}\textit{for}\\ \textit{constant}\\ \textit{acceleration}\end{matrix} \tag{50}$$

The Taylor series expansion for y directed motion y(t) is

$$y(t-t_0) = \sum_{n=0}^{\infty} y^n(t_0)\frac{(t-t_0)^n}{n!}$$

$$= y_0(t_0) + \left.\frac{dy(t)}{dt}\right|_{t_0} \frac{(t-t_0)}{1!} + \left.\frac{d^2y(t)}{dt^2}\right|_{t_0} \frac{(t-t_0)^2}{2!} + \cdots$$

With $dy/dt = v_y$ and $d^2y/dt^2 = a_y$, we get

$$y(t-t_0) = y_0(t_0) + v_y(t_0)(t-t_0) + \frac{1}{2}a_y(t_0)(t-t_0)^2 \tag{51}$$

with all higher powers of $(t-t_0)$ having zero coefficients.

If we set $t_0 = 0$ Equation (51), we get the very familiar result

$$y(t) = y_0 + v_{y0}t + \frac{1}{2}a_yt^2 \tag{52}$$

Here is an example of a Taylor series expansion that is valid for any range of values $(t-t_0)$. It is good for all times t because all derivatives of y(t) above the second derivative are zero.)

---

**Exercise 5**

Suppose a particle is moving in the y direction with a constantly increasing acceleration. I.e., assume that

$$a'_y(t) \equiv \frac{da_y(t)}{dt} = \text{constant}$$

Find the formula for y(t) for all future times t. (This is one step above the constant acceleration formulas.)

---

# THE WAVE EQUATION

In the Physics text, we calculated the speed of a wave pulse on a rope in Chapter 15, pages 15-4 and 5 . As we mentioned in the introduction, the calculation was relatively simple because of two tricks we were able to pull. One was to walk along with the pulse, so that it looked as if the pulse were standing still and the rope were passing through it. The second was to picture the top of the pulse as an arc of a circle, so that we would know the acceleration of the rope as it went around the arc. We got the right answer, but the process did not generate much confidence that we could handle more general cases, like calculating the speed of a sound wave pulse, or even of a compressional pulse on a Slinky. (Remember that we used dimensional analysis, an important but approximate tool, to estimate the wave speeds in these cases.)

What we will do now is the more direct approach of applying Newton's laws to a section of the wave pulse, get a differential equation, which happens to involve second derivatives in both space and time, and then solve the differential equation in the usual way. That is, we guess a solution, plug it into the equation, and see if we made the correct guess. We will use as much physical insight as we can to guide us in making the guess. The differential equation we will be working with is called the ***wave equation***.

Here we will be working with the wave equation for waves moving in one dimension. The three dimensional wave equation will be discussed later.

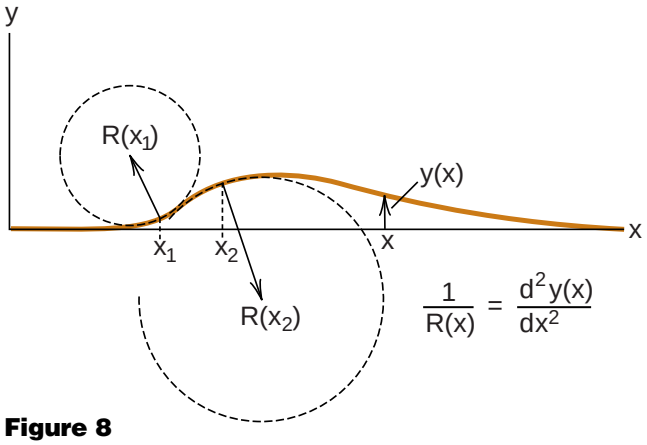



**Figure 8**
***Wave pulse on a rope. The curvature is positive (points up) at $x_1$, and negative at $x_2$.***

## Waves on a Rope

Our analysis of a wave pulse on a rope begins much as it did in Chapter 15. Figures (8) and (9) are similar to Figures (15-3c) and (15-3d), except that we are now standing still relative to the rope, and we are assuming the pulse is passing by us.

In our current analysis of the wave pulse, we will be somewhat more formal than we were in Chapter 15. We will say that the rope, at the present time, lies along a curve y(x) as shown in Figure (8). The quantity x is the distance down the rope (say from one end) and y(x) is the height of the pulse there, i.e., the distance the rope is displaced from its equilibrium position. From our various discussions of derivatives, we know that dy(x)/dx is the slope of the rope at position x, and $d^2y(x)/dx^2 = 1/R(x)$ is the curvature, which is equal to the reciprocal of the radius of curvature R(x) at that point. In Figure (8) we have sketched in circles to show the radius of curvature at the two points $x_1$ and $x_2$ along the curve. The curvature is positive at $x_1$ and negative at $x_2$ .

Let us consider a short section of rope of length $\Delta\ell$ located at position x as shown in Figure (9). For now assume that this section begins at the top of the pulse where the rope is horizontal. Shortly we will see that our results apply at any position along the rope.

The two ends of the section of rope are being pulled along the rope by the tension T. If the rope were straight, if there were no curvature at this point, the tension forces would cancel each other and there would be no net force on $\Delta\ell$ . Only because there is curvature is there a net force which we have labeled $T_y$ in Figure (9).

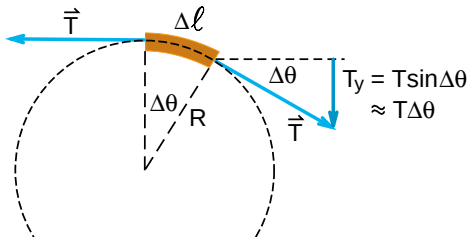



**Figure 9**
***Due to the tension pulling on both sides, this section of rope feels a net downward force $T_y \approx T\Delta\theta$.***

As long as $\Delta\ell$ is short enough, this section of rope will lie along the circle we have drawn to show the radius of curvature, and the two tension forces $\vec{T}$ will be tangent to the circle at the two ends. The result, from geometry we have now seen several times, is that the two angles labeled $\Delta\theta$ will be equal and the right hand tension force will have a downward pointing component $T_y$ given by

$$T_y = T \sin(\Delta\theta) \approx T\Delta\theta \qquad (53)$$

where for small angles we can replace the sine of an angle by the angle itself.

From Figure (9) we see that the angle $\Delta\theta$ is given by

$$\Delta\theta = \frac{\Delta\ell}{R} \qquad (54)$$

so that

$$T_y = T\Delta\theta = T\Delta\ell\left(\frac{1}{R}\right) \qquad (55)$$

Since $1/R$ is the curvature $d^2y(x)/dx^2$ at $\Delta\ell$, we get

$$T_y(x) = T\Delta\ell\,\frac{d^2y(x)}{dx^2} \qquad (56)$$

While Equation (56) was derived starting from the top of the pulse, we can see that as long as the sides of the pulse are not steep, as long as we are dealing with a shallow wave pulse, Equation (56) should apply all along the wave.

To see this, we have in Figure (10) analyzed the net force $\vec{T}_y$ acting slightly to the left side of the top of the pulse (at point $x_2$ in Figure (8)). Actually Figure (10) is the same as Figure (9), rotated by an angle $\varphi = dy(x)/dx$ which is the slope of the rope at point $x_2$. Here is where the shallow wave approximation comes in. As long as the wave is shallow and the sides of the pulse do not become steep, the angle $\varphi$ will be small, there will be very little rotation of Figure (9), and the net force $\vec{T}_y'$ will point nearly straight down and have a magnitude close to that given by Equation (56).

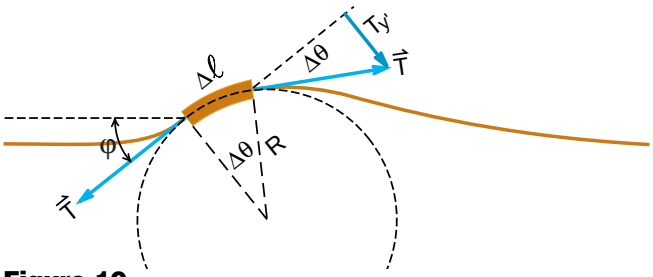



**Figure 10**
***If the section of rope slopes at an angle $\varphi$, then the net force $T_y'$ slopes at the same angle. That has little effect as long as the waves are shallow and $\varphi$ remains small.***

On the other hand, if the pulse becomes steep, the net force is no longer y directed and our current analysis will no longer apply. Whoever has watched ocean waves break as they approach the beach and become steeper and steeper, will recognize that steep waves behave very differently from shallow ones. Here we are working only with the theory of shallow waves.

Returning to Equation (56), which we have written here again

$$T_y(x) = T\Delta\ell\,\frac{d^2y(x)}{dx^2} \qquad \text{(56) repeated}$$

we want to point out that this equation gives us not only the magnitude but also the direction of the net force $T_y$. Where the curvature $d^2y(x)/dx^2$ is positive, as it is at point $x_1$ in Figure (8), the net force $T_y$ is directed upwards. Where the curvature is negative as at point $x_2$, the net force $T_y$ points down. Thus Equation (56) for $T_y(x)$ correctly changes sign when the direction of the net force changes.

Now that we have a reasonably general formula for the net force $T_y$ on a section $\Delta\ell$ of the rope (the only approximation being the shallow wave approximation), we are ready to apply Newton's second law, relating this net force to the mass $\Delta m$ and the acceleration $\vec{a}(t)$ of this section.

If the rope has a mass density $\mu$ kg/meter, then the mass of a section of length $\Delta\ell$ is simply

$$\Delta m = \mu\Delta\ell \qquad \begin{matrix}\textit{mass of}\\ \textit{section } \Delta\ell\end{matrix} \qquad (57)$$

We need to think a bit more about the situation to describe the acceleration of $\Delta m$. So far we have described the rope by the curve y(x), which is essentially a single snapshot of the rope at some special time t.

Another way to look at a wave pulse is to look at one point on the rope, and watch the point move up and down as the pulse comes by. We can describe this changing height by the function y(t). The acceleration $a_y(t)$ is then given by

$$a_y(t) = \frac{d^2y(t)}{dt^2} \qquad (58)$$

Equation (58) is limited in that it describes the motion of only one point of the rope. We can describe the motion of the whole rope for all times with a function y(x,t) that is a function of both space and time. If we look at the rope at some instant of time $t_0$, then the shape of the rope is given by

$$y(x) = y(x,t)\Big|_{t = t_0} \qquad (59)$$

while if we stand at one point $x_0$, the motion of the rope is given by

$$y(t) = y(x,t)\Big|_{x = x_0} \qquad (60)$$

An explicit example of such a function y(x,t) was our traveling wave formula of Equation (15-26) of the Physics text

$$y(x,t) = A \sin(kx - \omega t) \qquad (61) \text{ (also 15-26)}$$

which as we saw represented a sinusoidal wave traveling to the right at a speed

$$v_{wave} = \frac{\omega}{k} \qquad (62) \text{ (also 15-30)}$$

where the special frequency k is related to the wavelength $\lambda$ by $k = 2\pi/\lambda$, and the angular frequency $\omega$ is related to period T by $\omega = 2\pi/T$. (As a quick exercise, show that $\omega/k$ has the dimensions of a velocity).

With Equation (61) for y(x,t), you can easily see that if you look at the wave at one time, say t = 0, then

$$y(x,t)\Big|_{t = 0} = y(x) = \sin(kx) \qquad (63)$$

is a pure spacial sine wave. If you look at one particular point, for instance, x = 0, you get

$$y(x,t)\Big|_{x = 0} = y(t) = \sin(-\omega t) \qquad (64)$$

which is a pure sinusoidal oscillation.

## Partial Derivatives

When dealing with a function of two or more variables, like y(x,t), we have to be somewhat careful when we talk about derivatives. For now, we will always assume that if we are differentiating with respect to space, we will hold the time variable constant, i.e., consider the curve at one instant of time. Conversely, if we are differentiating with respect to time, we will consider only one point in space, i.e., hold x constant. There is a special notation for these so called ***partial derivatives***, where we differentiate with respect to one variable holding the other constant. In this notation we replace the d's, as in dx or dt by the symbol $\partial$. Thus

$$\frac{dy(x,t)}{dx}\Big|_{\text{holding t constant}} \equiv \frac{\partial y(x,t)}{\partial x} \qquad (65)$$

$$\frac{dy(x,t)}{dx}\Big|_{\text{holding x constant}} \equiv \frac{\partial y(x,t)}{\partial t} \qquad (66)$$

With this notation we get, for

$$y(x,t) = \sin(kx - \omega t) \qquad (67a)$$

$$\frac{\partial y(x,t)}{\partial x} = k \cos(kx - \omega t) \qquad (67b)$$

$$\frac{\partial y(x,t)}{\partial t} = -\omega \cos\left(kx - \omega t\right) \qquad (67c)$$

Using this new notation for partial derivatives, our Equations (56) for the net force $T_y$ on $\Delta\ell$, and (58) for the acceleration $a_y$ of $\Delta\ell$ becomes

$$T_y(x,t) = T\Delta\ell \frac{\partial^2 y(x,t)}{\partial x^2} \qquad (56a)$$

$$a_y(x,t) = \frac{\partial^2 y(x,t)}{\partial t^2} \qquad (58a)$$

To apply Newton's second law, we equate the net force $T_y(x,t)$ to the mass $\Delta m = \mu\Delta\ell$ times the acceleration $a_y(x,t)$ to get

$$T_y(x,t) = \Delta m\, a_y(x,t)$$

$$T\Delta\ell \frac{\partial^2 y(x,t)}{\partial x^2} = (\mu\Delta\ell) \frac{\partial^2 y(x,t)}{\partial t^2} \qquad (68)$$

The factors of $\Delta\ell$ cancel, and after dividing through by $\mu$ we get

$$\boxed{\frac{T}{\mu}\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2 y(x,t)}{\partial t^2}} \qquad (69)$$

as our final differential equation for the motion of the wave pulse on the rope.

How do you solve such a differential equation? As we have mentioned several times, we guess an answer for y(x,t), and plug the guess into the differential equation to see if we have made the correct guess. Also, we use whatever physics we have available to help us make a good guess.

Right now we do not have a formula for a single pulse that we can use as a guess for a solution to Equation (69). However we do have the formula in Equation (61) for a sine wave traveling to the right at a speed $v = \omega/k$

$$y(x,t) = A\sin(kx - \omega t) \qquad (61) \text{ repeated}$$

To see if this traveling wave is a solution to our differential Equation (69), we have to take a number of partial derivatives. They are

$$\frac{\partial y(x,t)}{\partial x} = \frac{\partial}{\partial x} A\sin(kx - \omega t)$$

$$= A\,k\cos(kx - \omega t) \qquad (70a)$$

$$\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial}{\partial x} Ak\cos(kx - \omega t)$$

$$= -A\,k^2\sin(kx - \omega t) \qquad (70b)$$

$$\frac{\partial y(x,t)}{\partial t} = \frac{\partial}{\partial t} A\sin(kx - \omega t)$$

$$= -\omega\,A\cos(kx - \omega t) \qquad (70c)$$

$$\frac{\partial^2 y(x,t)}{\partial t^2} = \frac{\partial}{\partial t}(-\omega\,A)\cos(kx - \omega t)$$

$$= -(-\omega\,A)(-\omega)\sin(kx - \omega t) \qquad (70d)$$

Using Equations (70b) and (70d) in Equation (69) gives

$$\frac{T}{\mu}(-Ak^2)\sin(kx - \omega t) \stackrel{?}{=} -A\omega^2\sin(kx - \omega t) \qquad (71)$$

The question mark in Equation (71) means that this is a guess, and we still have to see if the guess works.

First we notice that the functions $\sin(kx - \omega t)$ cancel. We had to have this cancellation or there was no chance of making the two sides equal for all times t and all positions x. We also note that the amplitudes A cancel, which means that the solution does not depend upon the amplitude A. After these cancellations we get

$$\frac{T}{\mu}(-k^2) = -\omega^2$$

$$\frac{T}{\mu} = \frac{\omega^2}{k^2} = v_{wave}^2 \qquad (72)$$

where we noted that $v_{wave} = \omega/k$ . Taking the square root of Equation (72) gives

$$\boxed{v_{wave} = \sqrt{\frac{T}{\mu}}} \qquad (73)$$

which is the answer we got in the Physics text, Equation (15-5), for the speed of a pulse on a rope.

## The One Dimensional Wave Equation

If we go back to Equation (69), and replace $T/\mu$ by $v_{wave}^2$, we get

$$\boxed{v_{wave}^2 \frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2 y(x,t)}{\partial t^2}} \quad \textit{one dimensional wave equation} \qquad (74)$$

This is a general form of what is called the ***one dimensional wave equation***. As we have just seen, a traveling sine wave, moving to the right at a speed $v_{wave}$ is a solution to this equation. The following exercises demonstrate that waves traveling to the left, and standing waves, are also solutions to this equation.

---

**Exercise 6**

(a) The formula for a sine wave moving to the left at a speed $v_{wave} = \omega/k$ was given in Equation (15-33) of the Physics text as

$$y(x,t)_{\text{wave moving left}} = A \sin(kx + \omega t) \qquad (15\text{-}33)$$

Show that this wave also obeys the wave Equation (73).

(b) Later in Chapter 15 we saw that a standing wave, which is the sum of a left moving and a right moving traveling wave, was given by the formula

$$y = A \sin kx \cos \omega t \qquad (15\text{-}35)$$

Show that this wave is also a solution to the wave Equation.

---

**Exercise 7**

Suppose you have two solutions $y_1(x,t)$ and $y_2(x,t)$, both of which are a solution to the wave equation with the same speed $v_{wave}$. Show that the sum wave

$$y(x,t) = y_1(x,t) + y_2(x,t) \qquad (75)$$

is also a solution of the same wave equation.

---

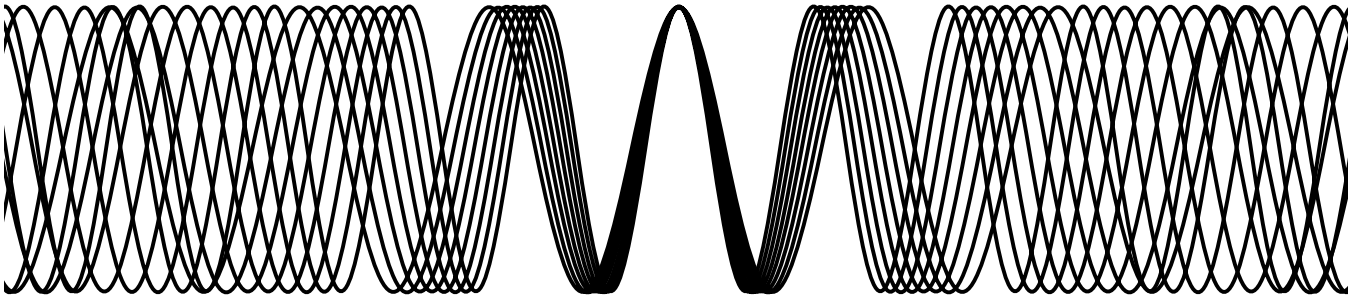
Exercise 7 gives us an important result. For our wave equation, which we got by considering wave pulses that were not too steep, the sum of two or more waves, each of which is a solution of the wave equation, is itself a solution.

In our discussion of Fourier analysis, introduced on page 16-6 of the Physics text, we saw that any continuous curve can be constructed from a sum of sine wave shapes. This suggests that we could construct a single wave pulse, moving to the left at a speed $v_{wave}$, by adding up a bunch of traveling sine waves of different wavelengths $\lambda_i = 2\pi/k_i$, but all with the same speed $v_{wave} = \omega_i/k_i$. The construction in Figure (11) suggests how we could add the sine (actually cosine) waves to get a pulse. Since each wave is a solution to the same wave Equation (73), the sum, i.e., the single pulse, is also a solution.

From Figure (11), it should be clear that we can construct a solution to the wave equation representing a pulse with very steep sides. However, in our analysis of the motion of the rope, we had to restrict ourselves to shallow waves in order to derive the wave equation for pulses on the rope. What this means is that the wave equation has solutions that we will not see on the rope. The shallow pulses on the rope will obey the wave equation, but we should expect that a steep pulse on the rope will behave differently. Not as differently as a breaking ocean wave, but differently.

**Figure 11**

***How to add cosine waves to get a pulse. At x = 0, all the waves add to give a big amplitude y. As we go out from x = 0, there is more and more cancellation until the sum wave adds to zero. If all these are traveling waves moving to the right at the same speed $v_{wave} = \omega_i k_i$, then the whole pulse must move at the same speed, maintaining its shape.***



## Compressional Waves on a Spring

When we came to the discussion of compressional waves on a spring, in particular the compressional Slinky wave we saw in Figure (Phys1-6) reproduced here, we resorted to dimensional analysis in Chapter 15 of the Physics text because there are no obvious tricks to calculate the speed of the pulse. Now we are in a position to set up a differential equation describing the motion of a short segment $\Delta\ell$ of the spring. We will get the wave equation, and from that we can immediately tell the speed of the pulse.

Suppose we have a stretched spring of length L as shown in Figure (12). The force required to stretch the spring, which is equal to the tension T in the spring, is given by Hook's law as

$$T = k(L - L_0) \qquad (76)$$

where $L_0$ is the unstretched length of the spring.

Now suppose that we stretch the spring an additional amount $\Delta L$. The tension will increase by an amount $\Delta T$ given by

$$T + \Delta T = k(L + \Delta L - L_0)$$
$$= k(L - L_0) + k\Delta L$$

Using Equation (76) to cancel the T and $k(L - L_0)$ terms, we are left with

$$\Delta T = k\Delta L = kL\left(\frac{\Delta L}{L}\right) \qquad (77)$$

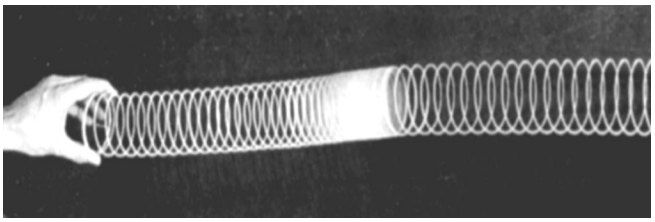


**Figure 1-6 (Physics 2000)**
***Compressional wave on a Slinky.***

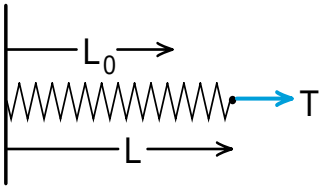



**Figure 12**
***A tension T stretches the spring from a length $L_0$ to a length L.***

There are two reasons why we have written $\Delta T$ as $kL(\Delta L/L)$ rather than just $k\Delta L$. The first is that $\Delta L/L$ is the amount of ***stretch per unit length***, a quantity engineers call ***strain***. It is a more inherent property of the spring than the total stretch $\Delta L$.

The second reason is that the product kL is also an inherent property of the spring. In Chapter 15, page 15-7 of the Physics text, we saw that if you had two identical springs of spring constant k, and attached them together, you got a spring twice as long but with half the spring constant. It is the product kL that does not change when you connect identical springs or cut a spring in half. Engineers would call this inherent property kL of the spring a ***spring modulus***.

To describe the stretched spring, we will introduce a function y(x) that represents the displacement of a point on the spring from its equilibrium (or initial) position. When we stretch a spring from a length L to a length $L + \Delta L$, as shown in Figure (13), every point on the spring moves to the right a distance y(x) given by the formula

$$y(x)\left(\begin{array}{l}\textit{displacement}\\ \textit{of a point}\\ \textit{on the spring}\end{array}\right) = \frac{x}{L}\Delta L \qquad (78)$$

where x is the distance down the spring, starting at the left end. You can see where we got Equation (78). If we are at the left end where x = 0, y(x) = 0 and there is no displacement. At the right end, where x = L, we get the full displacement $y(L) = (L/L)\Delta L = \Delta L$. In Equation (78) we are assuming that the displacement increases uniformly as we go down the spring.

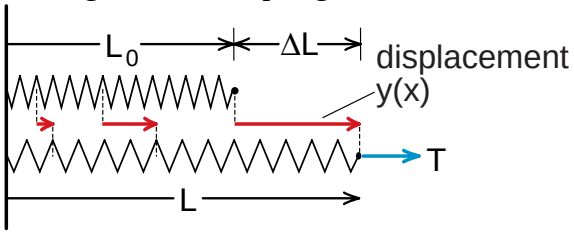



**Figure 13**
***The displacement y(x) increases as we go down the spring. With the formula $y(x) = (x/L)\Delta L$, we are assuming that the displacement is increasing uniformly.***

If we differentiate y(x) with respect to x we get

$$\frac{dy(x)}{dx} = \frac{d}{dx}\left(x\frac{\Delta L}{L}\right) = \frac{\Delta L}{L} \qquad (79)$$

Thus for a uniformly stretched spring, $y'(x) = dy(x)/dx$ is the amount of stretch per unit length, which we have called the *strain* of the spring.

If the strain is not uniform, if for example, we have a compressional wave on the spring, the strain is still given by

$$\text{strain} = \begin{matrix}\textit{local stretch}\\ \textit{per unit}\\ \textit{length}\end{matrix} = \frac{dy(x)}{dx} \qquad (80)$$

To help see that $y'(x) = dy(x)/dx$ is the local amount of stretching per unit length, note that when we integrate the local stretching per unit length over the total length of the spring, we get the total stretch $\Delta L$ .

$$\int_0^L \left(\frac{dy(x)}{dx}\right)dx = \int_0^L dy(x) = y(x)\Big|_0^L$$

$$= y(L) - y(0) \qquad (81)$$

$$= \Delta L - 0 = \Delta L$$

where $y(L) = \Delta L$ the total displacement at the end.

Now go back to Equation (77)

$$\Delta T = kL\left(\frac{\Delta L}{L}\right) \qquad (77) \text{ repeated}$$

which said that the change in tension in the spring is proportional to the strain $\Delta L/L$ . We proved this was true for a uniform strain $\Delta L/L$ . The obvious generalization when the strain is not uniform is to replace the average strain $\Delta L/L$ by the local strain $y'(x) = dy(x)/dx$ to get

$$\Delta T(x) = kL\frac{dy(x)}{dx} = kLy'(x) \qquad (82)$$

where $\Delta T(x)$ is the increase in the tension in a point x due to the local strain $y'(x)$ .

This gives us as the formula for the tension T(x) at point x

$$T(x) = T_0 + \Delta T(x)$$

$$T(x) = T_0 + kLy'(x) \qquad (82a)$$

where $T_0$ is the equilibrium tension, and $kLy'(x)$ is the change in tension caused by the displacement of parts of the spring from their equilibrium position.

Let us now apply Equation (82a) to a short section of spring of length $\Delta x$ , as shown in Figure (14). If the tension were uniform, the tension forces would cancel and there would be no net force on this section of the spring. A net force arises only if there is a change in tension as we go from x to $x + \Delta x$ . This net force will be

$$\begin{matrix}\textit{net}\\ \textit{force}\\ \textit{on } \Delta x\end{matrix} = T(x+\Delta x) - T(x)$$

$$= \left[T_0 + kLy'(x+\Delta x)\right] - \left[T_0 + kLy'(x)\right]$$

$$= kL\left[y'(x+\Delta x) - y'(x)\right]$$

$$= kL\Delta x\left[\frac{y'(x+\Delta x) - y'(x)}{\Delta x}\right] \qquad (83)$$

We immediately see that the last quantity in the square brackets is going to become, in the limit as $\Delta x \to 0$ , the second derivative of y(x) with respect to x. Thus our formula for the net force on a section of length $\Delta x$ is

$$\begin{matrix}\textit{net}\\ \textit{force}\\ \textit{on } \Delta x\end{matrix} = kL\Delta x\frac{d^2y(x)}{dx^2} \qquad (84)$$

If the spring has a mass per unit length of $\mu$ kg/meter , the mass $\Delta m$ of a length $\Delta x$ is

$$\Delta m = \mu\Delta x \qquad (85)$$

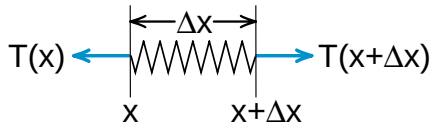



**Figure 14**
***There will be a net force on this short section of spring if the tension changes as we go from x to x+Δx.***

If we allow waves on the spring, the displacement y(x) from equilibrium depends not only on the position x down the spring, but also on the time (t) . Thus the displacement is described by the function y(x,t). The acceleration $a_y(x,t)$ at position x on the spring is

$$a_y(x,t) = \frac{\partial^2 y(x,t)}{\partial t^2} \qquad (86)$$

We are using the partial derivative symbol $\partial$ because we want to measure the change in y(x,t) with time at a fixed position x.

In terms of partial derivatives, Equation (84) for the net force on $\Delta x$ is

$$\begin{matrix}\textit{net}\\ \textit{force}\\ \textit{on } \Delta x\end{matrix} = kL\Delta x\frac{\partial^2 y(x,t)}{\partial x^2} \qquad (84a)$$

With Equations (84a), (85) and (86), Newton's second law applied to $\Delta m$ gives

$$\begin{matrix}\textit{net}\\ \textit{force}\\ \textit{on } \Delta m\end{matrix} = (\Delta m)\, a_y(x,t) \qquad \begin{matrix}\textit{Newton's law}\\ F = ma\end{matrix}$$

$$kL\Delta x\frac{\partial^2 y(x,t)}{\partial x^2} = (\mu\Delta x)\frac{\partial^2 y(x,t)}{\partial t^2} \qquad (87)$$

The factors of $\Delta x$ cancel and we are left with

$$\boxed{\left(\frac{kL}{\mu}\right)\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2 y(x,t)}{\partial t^2}} \qquad (88)$$

We recognize Equation (88) as the wave equation

$$v_{wave}^2\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2 y(x,t)}{\partial t^2}$$

where we can identify the wave speed as

$$\boxed{v\left(\begin{matrix}\textit{compressional}\\ \textit{Slinky wave}\end{matrix}\right) = \sqrt{\frac{kL}{\mu}}} \qquad (89)$$

We got this same answer on page (15-8) of the Physics text using dimensional analysis. However, with dimensional analysis we were not sure whether a factor of 2 or $\pi$ might be missing. Having derived the wave equation, we know that $\sqrt{kL/\mu}$ is the correct answer with no missing constant factors.

## The Speed of Sound

The analysis of compressional sound waves in air can be carried out along lines very similar to our analysis of a compressional wave on a spring. However to do this, we need to build on our discussion of the behavior of an ideal gas in Chapters 17 and 18 of the Physics text. Thus we will assume that the reader is familiar with this material, including the discussion on adiabatic expansion in the Chapter 18 appendix.

Consider a column of gas with a cross sectional area A and length L as shown in Figure (15). We can think of the gas as being in a cylinder with frictionless walls, but it could be a hypothetical column in a large volume of gas. Let the variable x measure the distance down the column, starting at the left end, and imagine that we have a frictionless piston at the right end.

If we pull the piston out a small distance $\Delta L$ , we change the volume of the gas by an amount

$$\Delta V = A\Delta L \qquad (90)$$

and in so doing, decrease the pressure p.

How much the pressure changes depends upon the way the gas is expanded. If we expand it very slowly so that heat has time to flow into the gas and the temperature remains constant (this is called an ***isothermal expansion)*** then we have, from the ideal gas law

$$pV = NRT = \text{constant} \qquad \begin{matrix}\textit{isothermal}\\ \textit{expansion}\end{matrix} \qquad (91)$$

where N is the number of moles of gas in the cylinder, R is the gas constant, and T the temperature in kelvins.

However in a sound wave, expansions and compressions happen so rapidly that there is not enough time for heat to flow in or out, and the temperature changes.

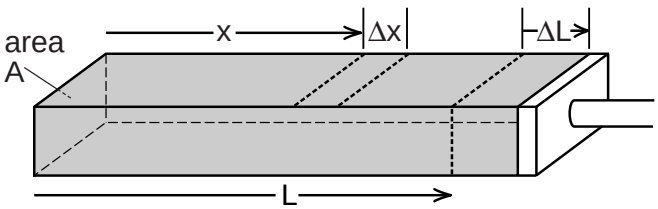



**Figure 15**
***Column of gas of cross-sectional area A and length L.***

When we expand a gas with no heat flow, we call this an ***adiabatic expansion***. As we saw in the appendix to Chapter 18, in an adiabatic expansion the gas obeys the equation

$$pV^\gamma = \text{constant} \qquad \begin{matrix}\textit{adiabatic}\\ \textit{expansion}\end{matrix} \qquad (92)$$

where

$$\gamma = \frac{c_p}{c_v} \qquad (93)$$

is the ratio of the specific heat $c_p$ at constant pressure to the specific heat $c_v$ at constant volume. It is Equation (92) for an adiabatic expansion rather than Equation (91) for an isothermal expansion that we need to use to describe the relationship between pressure and volume for a sound wave.

The quantity $\gamma = c_p/c_v$ depends, as we saw at the beginning of Chapter 18, on the number of effective degrees of freedom of the gas molecules. As you found if you did Exercise 2 of Chapter 18, for a monatomic gas like helium or argon with no rotational degrees of freedom, $\gamma = 1.66\ (5/3)$. For diatomic gases like oxygen, nitrogen, and of course air, that have two rotational degrees of freedom, $\gamma = 1.40$. When we get to more complex structures like $CO_2$ and $NH_4$, then $\gamma$ drops to 1.28.

We will now use Equation (92) for an adiabatic expansion to calculate the change $\Delta p$ in pressure when we change the volume of the gas in the cylinder by an amount $\Delta V$. Before we compress we have

$$pV^\gamma = p_0V_0^\gamma \qquad (94)$$

where $p_0$ and $V_0$ are our original pressure and volume. After the expansion, V goes to $V_0 + \Delta V$ and p goes to $p_0 + \Delta p$, where we know that $\Delta p$ is negative for an expansion. Thus after the expansion we have

$$pV^\gamma = (p_0 + \Delta p)(V_0 + \Delta V)^\gamma \qquad (95a)$$

With $pV^\gamma = p_0V_0^\gamma$ = constant, we get

$$p_0V_0^\gamma = (p_0 + \Delta p)(V_0 + \Delta V)^\gamma \qquad (95b)$$

We can use the fact that $\Delta V$ is very small compared to $V_0$ to get

$$(V_0 + \Delta V)^\gamma = \left[V_0\left(1 + \frac{\Delta V}{V_0}\right)\right]^\gamma = V_0^\gamma\left(1 + \frac{\Delta V}{V_0}\right)^\gamma$$

Using the approximation $(1 + \alpha)^\gamma \approx 1 + \gamma\alpha$ for a small $\alpha$, we have

$$\left(1 + \frac{\Delta V}{V_0}\right)^\gamma \approx 1 + \gamma\frac{\Delta V}{V_0} \qquad (96)$$

Using (96) in (95b), with $p_0 + \Delta p = p_0(1 + \Delta p/p_0)$, gives

$$\begin{aligned} p_0V_0^\gamma &= (p_0 + \Delta p)(V_0 + \Delta V)^\gamma \\ &= p_0\left(1 + \frac{\Delta p}{p_0}\right)V_0^\gamma\left(1 + \gamma\frac{\Delta V}{V_0}\right) \end{aligned} \qquad (97)$$

Multiplying this out gives

$$p_0V_0^\gamma = p_0V_0^\gamma\left[1 + \frac{\Delta p}{p_0} + \frac{\gamma\Delta V}{V_0} + \frac{\gamma}{p_0V_0}\Delta p\Delta V\right] \qquad (98)$$

The factors $p_0V_0^\gamma$ cancel, and we can neglect the second order term $\Delta p\Delta V$, giving

$$1 = 1 + \frac{\Delta p}{p_0} + \frac{\gamma\Delta V}{V_0}$$

After canceling the 1's and multiplying through by $p_0$ we get for the pressure change $\Delta p$

$$\Delta p = -\gamma p_0\left(\frac{\Delta V}{V_0}\right) \qquad (99)$$

If you look at the appendix to Chapter 18 in our discussion of the adiabatic expansion, you see that we started with the equation

$$\gamma p_0\Delta V + \Delta pV_0 = 0 \qquad (18\text{-}A8)$$

[which is Equation (99) if we solve for $\Delta p$] and went through a number of calculus steps to derive $pV^\gamma = \text{constant}$. What we have done in going from $pV^\gamma = \text{constant}$ to Equation (99) is to undo the calculus steps in that appendix. However one typically remembers the equation $pV^\gamma = \text{constant}$ for adiabatic expansions rather than Equation (18-A8), and it seemed worthwhile to show how to get from $pV^\gamma = \text{constant}$ to our formula for $\Delta p$.

Now that we have Equation (99) for $\Delta p$, we can follow essentially the same steps that we did earlier to calculate the speed of a compressional wave pulse on a spring.

If the cylinder in Figure (15) has a cross sectional area A, length L, and we move the piston out a distance $\Delta L$, we have

$$V_0 = AL$$
$$\Delta V = A\Delta L \qquad (100)$$

thus from Equation (99) we have

$$\Delta p = -\gamma p_0\left(\frac{\Delta V}{V}\right) = -\gamma p_0\left(\frac{A\Delta L}{AL}\right)$$
$$= -\gamma p_0\frac{\Delta L}{L} \qquad (101)$$

In moving the piston out, the average displacement of a molecule y(x) at position x will be

$$y(x) = \frac{x}{L}\Delta L \qquad (102)$$

which is the same as our Equation (78) for the average displacement of a piece of spring at position x. Differentiating Equation (102) with respect to x gives

$$y'(x) = \frac{dy(x)}{dx} = \frac{\Delta L}{L} \qquad (103)$$

Thus we see that for a uniform displacement of the gas molecules, the strain, the displacement per unit length, is $y'(x) = dy(x)/dx$. We will now assume that even for non uniform displacements such as the kind we would have in a pressure pulse, $y'(x)$ represents the local strain or displacement per unit length. In terms of this local strain, our formula (101) for $\Delta p(x)$ becomes

$$\Delta p(x) = -\gamma p_0 y'(x) \qquad \begin{matrix}\textit{local}\\ \textit{pressure}\\ \textit{change}\end{matrix} \qquad (104)$$

As in our discussion of springs, we can write this equation in the form

$$p(x) = p_0 + \Delta p(x)$$
$$p(x) = p_0 - \gamma p_0 y'(x) \qquad (105)$$

where we see that variations from the static pressure $p_0$ are caused by local strains $y'(x)$.

Now consider a section of the cylinder of length $\Delta x$ located at x as shown in Figure (16). The gas external to $\Delta x$ on the left, where the pressure is p(x), exerts a right directed force of magnitude

$$F(x) = Ap(x) \qquad (106)$$

while the gas on the right exerts a left directed force of magnitude

$$F(x + \Delta x) = Ap(x + \Delta x) \qquad (107)$$

where we have used the fact that the force is the pressure times the area. The net force on $\Delta x$ is thus

$$F_{\text{net on }\Delta x} = F(x) - F(x + \Delta x)$$
$$= A\Big[p(x) - p(x + \Delta x)\Big] \qquad (108)$$

Using Equation (105) for p(x) we get

$$F_{\text{net}} = A\Big\{\Big[p_0 - \gamma p_0 y'(x)\Big] - \Big[p_0 - \gamma p_0 y'(x + \Delta x)\Big]\Big\}$$

The $p_0$ terms cancel and we are left with

$$\vec{F}_{\text{net}} = A\gamma p_0\Big[y'(x + \Delta x) - y'(x)\Big]$$

We can multiply by $\Delta x/\Delta x$ to get

$$F_{\text{net}} = A\gamma p_0\Delta x\left[\frac{y'(x + \Delta x) - y'(x)}{\Delta x}\right] \qquad (109)$$

As in the case of the spring, we will end up taking the limit as $\Delta x$ goes to zero, so that the term in the square brackets in Equation (109) becomes the second derivative $d^2y(x)/dx^2$.

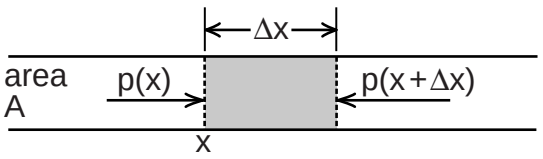



**Figure 16**
***Pressure forces acting on a small section of gas in our hypothetical cylinder.***

We will also let y(x) become a function of time y(x,t), so that second derivative becomes a partial derivative with respect to x only, and we get

$$F_{net} = \gamma p_0 (A\Delta x)\frac{d^2y(x,t)}{dx^2} \qquad (110)$$

as our final formula for the net force on the gas in $\Delta x$.

The next step is to calculate the mass $\Delta m$ of the gas in the region $\Delta x$ . If the density of the gas is $\rho$ kg/meter$^3$ and the volume inside $\Delta x$ is $(A\Delta x)$ meters$^3$, we have

$$\Delta m = \rho A\Delta x \qquad (111)$$

The acceleration of the gas in $\Delta x$ is

$$a_x(t) = \frac{\partial^2 y(x,t)}{\partial t^2} \qquad (112)$$

Using Equations (110), (111), and (112) in Newton's second law gives

$$F_{net\ on\ \Delta x} = \Delta m\, a_x(t)$$

$$\gamma p_0 (A\Delta x)\frac{\partial^2 y(x,t)}{\partial x^2} = \rho (A\Delta x)\frac{\partial^2 y(x,t)}{\partial t^2} \qquad (113)$$

The factor $A\Delta x$ cancels and we are left with

$$\frac{\gamma p_0}{\rho}\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2 y(x,t)}{\partial t^2} \qquad (114)$$

and we get the wave equation

$$v_{wave}^2 \frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2 y(x,t)}{\partial t^2} \qquad (74)\ \text{repeated}$$

where we immediately see that the speed of the sound wave is given by

$$\boxed{v_{sound} = \sqrt{\frac{\gamma p_0}{\rho}}} \qquad (115)$$

In our discussion of sound waves in Chapter 15 of the Physics text, where we used dimensional analysis to predict the speed of sound, we came up with the formula

$$v_{sound} = \sqrt{\frac{B}{\rho}} \qquad (116)$$

where

$$B \equiv \frac{\Delta p}{\Delta V/V}$$

was called the ***bulk modulus*** of the gas. Going back to Equation (101), we have

$$\Delta p = = -\gamma p_0 \frac{\Delta L}{L} \qquad (101)$$

$$= -\gamma p_0 \left(\frac{\Delta V}{V}\right) \qquad (117)$$

for an adiabatic expansion, and the same with a + sign for compression. Thus

$$\frac{\Delta p}{\Delta V/V} = \gamma p_0 = B \quad \textit{for adiabatic compression} \qquad (118)$$

and our old formula for the speed of sound can be written as

$$v_{sound} = \sqrt{\frac{B}{\rho}} = \sqrt{\frac{\gamma p_0}{\rho}} \qquad (119)$$

which is the same result we got from the wave equation.

Using the ideal gas law, we can re-express the quantity $p_0/\rho$ in our formula for the speed of sound in terms of the temperature T of the gas and some other constants. First we will write the density $\rho$ as

$$\rho \frac{kg}{meter^3} = \frac{M\ kg/mole \times N\ moles}{V\ meters^3} \qquad (120)$$

where M is the mass of one mole of the gas (an Avogadro's number of the gas molecules), N is the number of moles in our cylinder, and V the volume of the cylinder.

Next write the ideal gas law pV = NRT as

$$\frac{N}{V} = \frac{p}{RT} \qquad (121)$$

where R is the gas constant and T the temperature in kelvins. Combining Equations (120) and (121) gives

$$\rho = M\frac{N}{V} = \frac{Mp}{RT}$$

or we have

$$\frac{p}{\rho} = \frac{RT}{M} \qquad (122)$$

and our formula for the speed of sound becomes

$$v_{sound} = \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma RT}{M}} \qquad (123)$$

To interpret the physics of Equation (123), it is perhaps clearer to express the answer in terms of mass of the gas molecules involved. We have

$$m_{molecule} = \frac{M}{N_A}\frac{\text{kilograms/mole}}{\text{molecules/mole}}$$

$$= \frac{M}{N_A}\frac{\text{kilograms}}{\text{molecule}}$$

where $N_A$ is Avogadro's number, and

$$k = \frac{R}{N_A}\frac{\text{joules/mole kelvin}}{\text{molecules/mole}} = \frac{R}{N_A}\frac{\text{joules}}{\text{kelvin}}$$

is Boltzman's constant. Thus

$$\frac{R}{M} = \frac{N_A k}{N_A m_{molecule}} = \frac{k}{m_{molecule}} \qquad (124)$$

and in terms of the molecular mass $m_{molecule}$ we get

$$\boxed{v_{sound} = \sqrt{\frac{\gamma kT}{m_{molecule}}}} \qquad (125)$$

From Equation (125), we immediately see that for a gas like hydrogen consisting of light molecules, the speed of sound is considerably greater than in a gas with heavy molecules.

---

**Exercise 8**

Calculate the speed of sound at a temperature of 300 kelvin, in hydrogen, helium, nitrogen and $CO_2$. Use the fact that a hydrogen molecule has the mass of 2 protons, a helium atom the mass of 4 protons (with a nucleus of 2 protons and 2 neutrons), a nitrogen molecule the mass of 28 protons (each nucleus has 7 protons and usually 7 neutrons) and a $CO_2$ molecule has a mass of around 44 protons (carbon nucleus has 6 protons and 6 or 7 neutrons, oxygen has 8 protons and 8 neutrons, for a total of 12 + 16 + 16 = 44 nuclear particles).

---

Aside from its dependence on the mass of the gas molecules, the other important feature is that the speed of sound is proportional to the square root of temperature. Thus the warmer the gas the greater the speed. This dependence of the speed of sound on the square root of temperature leads to a close connection between the speed of sound and the average speed of the air molecules due to their thermal motion.

In our discussion of the ideal gas law, we used the fact that the temperature was a measure of the average thermal kinetic energy of the gas, the precise relationship being

$$\frac{1}{2}m_{molecule}\bar{v}^2 = \frac{3}{2}kT \qquad (126)$$

where $\bar{v}^2$ is the average of the square of the speed of the gas molecules ($v^2 = v_x^2 + v_y^2 + v_z^2$). Writing Equation (126) in the form

$$\frac{kT}{m_{molecule}} = \frac{\bar{v}^2}{3} \qquad (127)$$

and using this in Equation (125) gives

$$v_{sound} = \sqrt{\frac{\gamma kT}{m_{molecule}}} = \sqrt{\frac{\gamma \bar{v}^2}{3}}$$

$$\boxed{v_{sound} = \bar{v}\sqrt{\frac{\gamma}{3}}} \qquad (128)$$

Several times we mentioned that the speed of sound is closely related to the speed of the air molecules due to their thermal motion. Equation (128) gives us the precise relationship. For air, for example, where $\gamma = 1.28$ we get

$$v_{sound} = \bar{v}\sqrt{\frac{1.28}{3}} = .65\,\bar{v} \qquad (129)$$

Sound travels over half as fast as the average speed $\bar{v}$ of the air molecules.

# Calculus 2000-Chapter 3

# The Gradient

*The gradient operation represents the fundamental way that we go from a scalar field like the electric voltage V to a vector field like the electric field $\vec{E}$.*

*In this chapter, we present two distinct ways to introduce the gradient operation. One is to use the fact that electric fields are related to electric voltage the same way that forces are related to potential energy. The second, more geometrical way, is to picture the electric voltage as being described by a contour map, and that the electric field is described by the lines of steepest decent in the map. We present these two points of view as separate sections, View 1 and View 2, that can be read in either order.*

*We end the chapter with View 3, an application to fluids, where we see that the pressure force $\vec{f}_p$ acting on fluid particles is the gradient of the pressure field p. This represents a straightforward example of obtaining a vector field $\vec{f}_p$ from a scalar field p.*

## TWO VIEWS OF THE GRADIENT

In the Physics text, our first laboratory exercise on electric phenomena was the potential plotting experiment illustrated in Figure (25-10) reproduced here. Two small brass cylinders connected to a battery were placed in a shallow tray of slightly conducting water. In order to measure the distribution of voltages V(x,y) at various points (x,y) in the water, we had two probes of bent, stiff, wire attached to blocks of wood, adjusted so that the tips of the wire stuck down in the water. The other end of the wire probes were attached to a voltmeter as shown.

By leaving one probe fixed, and moving the other in a way that the reading on the voltmeter remained constant, we could map out lines of constant voltage in the water. The results from a student lab notebook are shown slightly cleaned up in Figure (25-11). These lines of constant voltage are also known by the name ***equipotential lines*** or lines of equal electric potential. We also pointed out that these lines were analogous to lines of equal height, the contour lines in a contour map of the countryside.

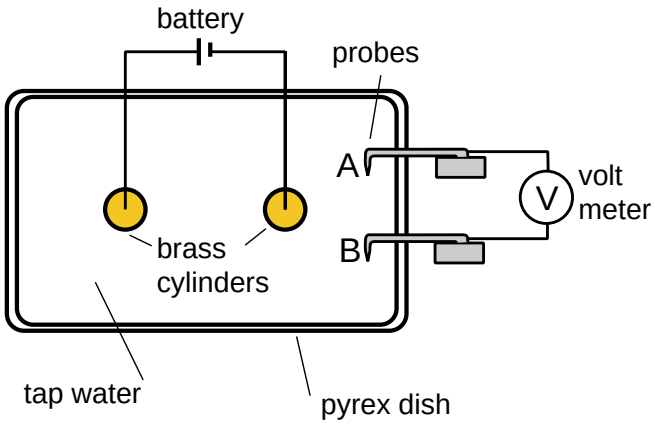



**Figure 25-10 (from Physics text)**
***Simple setup for plotting fields. You plot equipotentials by placing one probe (A) at a given position and moving the other (B) around. Whenever the voltage V on the voltmeter reads zero, the probes are at points of equal potential.***

While mapping the voltage V(x,y) at various points in the water was a straightforward process, our construction of the electric field lines $\vec{E}(x,y)$ was not so obvious. Our procedure was to map $\vec{E}(x,y)$ by drawing a set of lines perpendicular to the equipotential lines as shown in Figure (25-12). With this technique we were just barely able to tell whether the resulting field $\vec{E}(x,y)$ more closely resembled the field of line charges or point charges. Our technique was conceptually correct, but a very crude way to determine the electric field $\vec{E}(x,y)$ from a map of the voltage V(x,y).

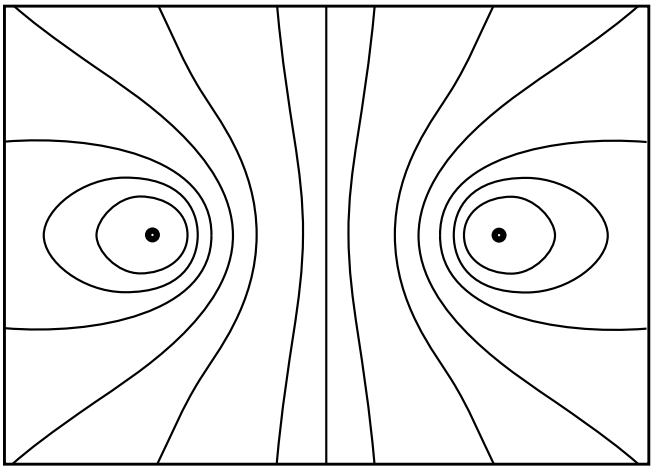


**Figure 25-11**
***Plot of the equipotential lines from a student project by B. J. Grattan. Instead of a tray of water, Grattan used a sheet of conductive paper, painting two circles with aluminum paint to replace the brass cylinders. We used the Adobe Illustrator® program to draw the lines through Grattan's data points.***

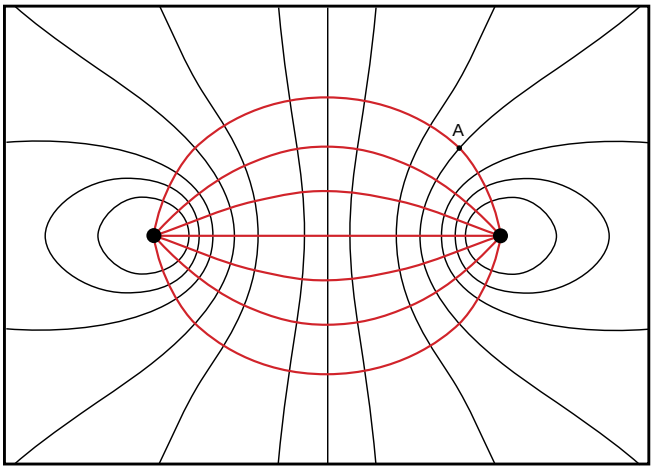


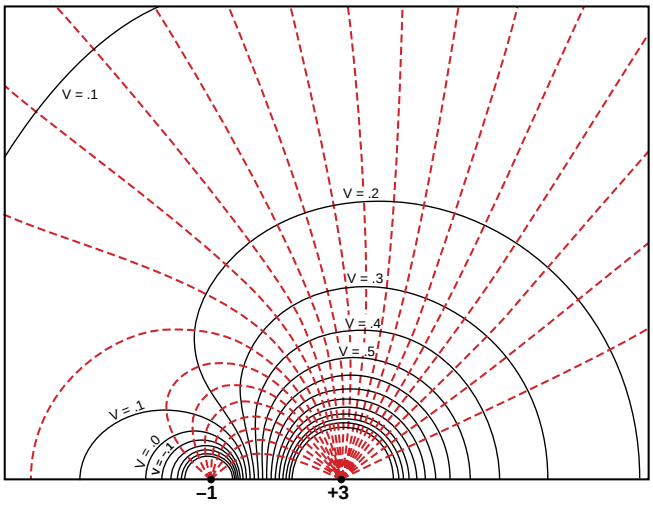

**Figure 25-12**
***To sketch the field lines, draw smooth lines, always perpendicular to the equipotential lines, and maintain any symmetry that should be there.***

After this initial experiment, we resorted to computer plots, like the one shown in Figure (25-15), to see the relationship between the electric field and a voltage map.

The computer plots, and the models we constructed from them, nicely illustrate the geometrical relationship between a voltage map and the electric field lines, but did not provide a convenient technique for actually calculating the field. The missing technique, which is the subject of this chapter, is the mathematical procedure called the ***gradient***, a procedure involving the partial derivatives of the voltage function V(x,y).

As Figure (25-15) illustrates, there is a complete analogy between the contour map of a hilly terrain and electric field plots from a voltage map. We can build our discussion of the gradient operation either upon our knowledge of the mathematics of the electric field, or by developing the ideas from a discussion of the nature of a hilly terrain. While both approaches are equivalent, we see the subject from two rather different points of view. The electric field approach is more efficient, while the hilly terrain approach develops some concepts that we will need later on.

As we mentioned in the introduction, we will begin this chapter with the electric field approach, and later discuss the hilly terrain viewpoint separately in **View 2**. You should study both approaches to see this important topic from two points of view. It does not really matter which one you study first.



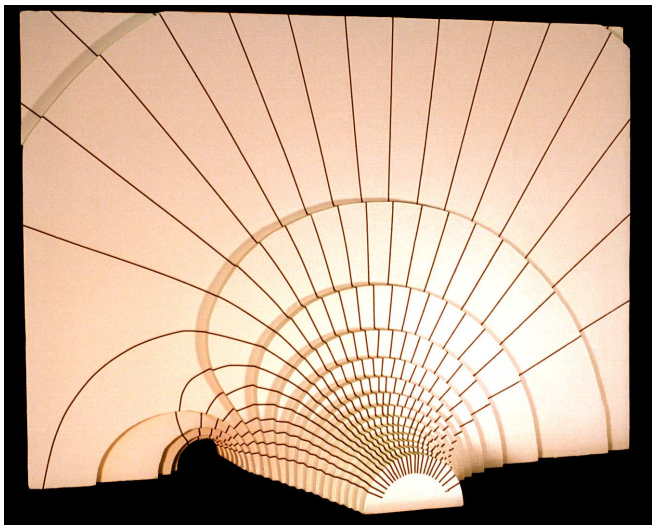



**Figure 25-15**
***Computer plot of the field lines and equipotentials for a charge distribution consisting of a positive charge + 3 and a negative charge – 1. These lines were then used to construct the plywood model.***

# View 1

# The Gradient from a Force – Energy Perspective

## CALCULATING THE ELECTRIC FIELD

Figure (1) shows a small section of the voltage map of Figure (25-15) on the previous page. The solid lines are equipotential lines, lines of constant voltage spaced .1 volts apart. We want to imagine that we actually have a detailed map of the voltage V(x,y) at every point (x,y), and want to mathematically determine, from that map, the electric field $\vec{E}(x,y)$ at every point.

In the Physics text, we emphasized the idea that the electric voltage V(x,y) was the electric potential energy of a unit test charge, while the electric field $\vec{E}(x,y)$ was the electric force on a unit test charge. Thus the connection between $\vec{E}$ and V is the relationship between force and potential energy.

To review this relationship, imagine that I place a unit test particle at point A in Figure (1), where the voltage is $V_A = .3$ volts. Since the voltage is the potential energy, in joules, of a unit test charge, our test particle at point A has a potential energy of .3 joules.

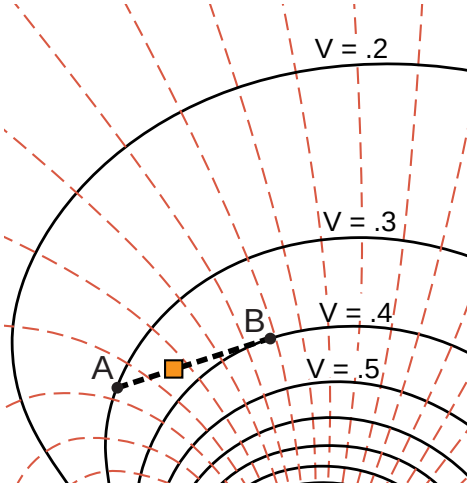



**Figure 1**
***A small section of the voltage map, showing equipotential lines spaced .1 volts apart. We will calculate the amount of work required to move a unit test charge from point A to point B.***

Now imagine that I move the test particle along the dashed line from point A at .3 volts over to point B at .4 volts. The potential energy of the particle has increased from .3 joules to .4 joules. Thus to move the particle, I must supply (.1) joules of energy to the particle.

Imagine that I move the test particle slowly, so that the force $\vec{F}_{me}(x,y)$ that I exert on the particle is just enough to oppose the force $\vec{E}(x,y)$ that the electric field is exerting on the particle. Thus for the entire trip from A to B we have

$$\vec{F}_{me}(x,y) = -\vec{E}(x,y) \qquad (1)$$

The amount of work I do in moving the particle is given by the formula first discussed in Chapter 10 of the Physics text (see page 10-15, Equation (10-25)).

$$\left.\begin{array}{l}\text{work I do in}\\ \text{moving the}\\ \text{test particle}\end{array}\right\} = \int_A^B \vec{F}_{me}\cdot d\vec{\ell} \qquad (2)$$

Because I am moving the particle slowly so that all the work I do is stored as electric potential energy, and because the increase of potential energy of the unit test charge is $V_B - V_A$, we have

$$\int_A^B \vec{F}_{me}\cdot d\vec{\ell} = V_B - V_A \qquad (3)$$

We can get *me* out of the equation by using Equation (1) to give

$$\boxed{-\int_A^B \vec{E}\cdot d\vec{\ell} = V_B - V_A} \qquad (4)$$

Equation (4) is the integral equation that relates the voltage V(x,y) to the electric field $\vec{E}(x,y)$. It is a relationship we used extensively in the Physics text. In the Calculus text, we will often translate from integral to differential equations, and this chapter on the gradient will be our first example of how this is done.

The first step in going to a differential equation is to focus in on a very small region of Figure (1), a region shown in Figure (2), centered at the point $(x_i,y_i)$ on the path from A to B. We have zoomed in so closely to the point $(x_i,y_i)$ in Figure (2), we have so greatly magnified the plot, that the equipotential lines and the field lines in this region are simply straight lines at right angles to each other.

Now suppose we move our test particle from point (1) at $(x_i,y_i)$ over a distance $\Delta\vec{\ell}$ along the path to point (2) as shown. Equation (4) applied to this short displacement is

$$V_2 - V_1 = -\int_1^2 \vec{E}(x,y)\cdot d\vec{\ell} \qquad (5)$$

For this short path, we can assume that $\vec{E}(x,y)$ is essentially constant and replace the integral by the product $\vec{E}(x_i,y_i)\cdot\Delta\vec{\ell}$, giving us

$$V_2 - V_1 = -\vec{E}(x_i,y_i)\cdot\Delta\vec{\ell} \qquad (6)$$

[You can see that in going from Equation (5) to (6) we are essentially undoing the step we took in Chapter (10) to derive the integral Equation (4).]

We are discussing the electric field of point charges. This is a conservative field, which is a fancy way of saying that the change in potential energy when we move a particle between two points does not depend upon the path we take. Thus if we first go a distance $\Delta x$ along the x axis to point (3), then up the y axis a distance $\Delta y$ to point (2), we should get the same change in voltage $V_2 - V_1$ that we got by going directly from point (1) to point (2) along $\Delta\vec{\ell}$.

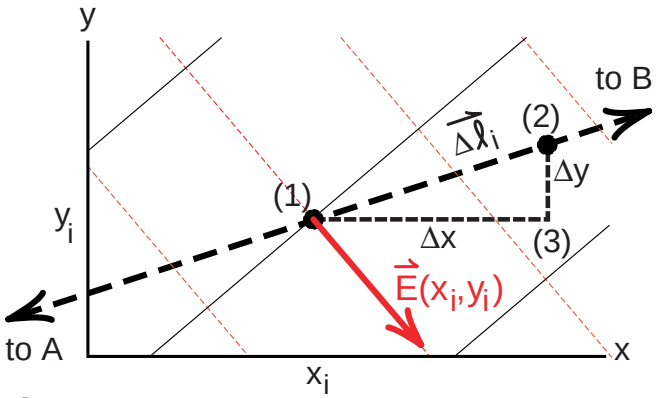



**Figure 2**
***If we zoom in far enough, we reach a point where the equipotential lines and contour lines are straight lines perpendicular to each other.***

In going along the x axis, we have

$$\begin{aligned} V_3 - V_1 &= -\vec{E}(x_i,y_i)\cdot\Delta\vec{x} \\ &= -E_x(x_i,y_i)\Delta x \end{aligned} \qquad (7)$$

where the dot product of $\vec{E}$ with the x directed displacement $\Delta\vec{x}$ leaves us with the x component $E_x$. Writing out $V_1$ and $V_3$ in the form

$$V_1 = V(x_i,y_i)$$

$$V_3 = V(x_i+\Delta x, y_i)$$

Equation (7) becomes

$$V(x_i+\Delta x, y_i) - V(x_i,y_i) = -E_x(x_i,y_i)\Delta x \qquad (8)$$

Dividing through by $-\Delta x$ gives

$$E_x(x_i,y_i) = -\frac{V(x_i+\Delta x, y_i) - V(x_i,y_i)}{\Delta x}$$

When we take the limit that $\Delta\vec{\ell}$ goes to zero, both $\Delta\vec{x}$ and $\Delta\vec{y}$ will go to zero, giving

$$E_x(x_i,y_i) = -\lim_{\Delta x\to 0}\left[\frac{V(x_i+\Delta x, y_i) - V(x_i,y_i)}{\Delta x}\right] \qquad (9)$$

By now you should recognize that the limit in Equation (9) is the partial x derivative of the function V(x,y) evaluated at the point $(x_i,y_i)$. Since this is true for any point (x,y), we get

$$\boxed{E_x(x,y) = -\frac{\partial V(x,y)}{\partial x}} \qquad (10)$$

where the symbol $\partial$ is used for partial derivatives.

---

**Exercise 1**

Use the above line of reasoning to show that

$$E_y(x,y) = -\frac{\partial V(x,y)}{\partial y} \qquad (11)$$

---

Introducing the unit vectors $\hat{x}$ and $\hat{y}$, we can combine Equations (10) and (11) into the single vector equation

$$\vec{E}(x,y) = \hat{x}\,E_x(x,y) + \hat{y}\,E_y(x,y)$$

$$\boxed{\vec{E}(x,y) = -\left[\hat{x}\frac{\partial V(x,y)}{\partial x} + \hat{y}\frac{\partial V(x,y)}{\partial y}\right]} \qquad (12)$$

Equation (12) is the differential equation we can use to calculate the electric field $\vec{E}(x,y)$ at every point from a knowledge of the voltage V(x,y).

## Interpretation

$$\vec{E}(x,y) = -\left[\hat{x}\frac{\partial V(x,y)}{\partial x} + \hat{y}\frac{\partial V(x,y)}{\partial y}\right] \qquad (12)$$

To help interpret Equation (12) repeated above, let us go back to Figure (25-15) where we started with a plot of the equipotential lines of the voltage V(x,y) and constructed a three dimensional plywood model of the voltage. The equipotential lines became the contour lines of this model, and the perpendicular electric field lines are the lines of steepest slope. If you were standing on terrain represented by this model, and the slope became slippery, the field line is the direction you would start to slide. Ski instructors call this direction of steepest slope the ***fall line***.

To simplify the job of interpreting Equation (12), imagine that we are standing at the point $A = (x_A, y_A)$ shown in Figure (3), where the contour line happens to be running in the y direction. If we move along a contour line there is no change in height, thus the partial derivative of V(x,y) with respect to y—the rate of change of V(x,y) in the y direction—is zero at point A.

$$\left.\frac{\partial V(x,y)}{\partial y}\right|_{\substack{x = x_A \\ y = y_A}} = 0 \qquad (13)$$

The formula for $\vec{E}(x,y)$ at point A becomes

$$\left.\vec{E}(x,y)\right|_{\substack{x = x_A \\ y = y_A}} = -\hat{x}\left.\frac{\partial V(x,y)}{\partial x}\right|_{\substack{x = x_A \\ y = y_A}} \qquad (14)$$

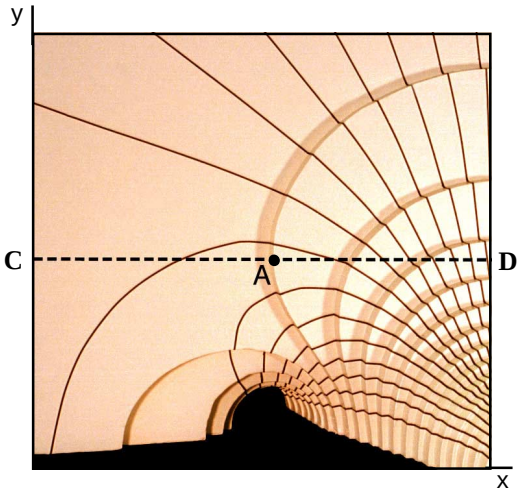



**Figure 3**
***The V = .2 volt contour line passes straight up through the point labeled A. Imagine that the surface is smoothed out and you walk along the dotted line.***

To interpret Equation (14), imagine that we smooth out our plywood model of the voltage surface, then saw the model in two, cutting through the point A with the saw blade oriented along the x axis, along the dotted line in Figure (3). A side view of the upper piece is shown at the bottom of Figure (4). You can see that the voltage at the beginning of the cut, point C, is somewhat greater than .1 volts, and rises to just over .4 volts at the end, point D. The mathematical formula for the curve we see in Figure (4) is $V(x,y_A)$, and the partial derivative with respect to x at point A is the slope of the curve $V(x,y_A)$ at $x = x_A$. This is just the tangent of the angle $\theta$ in Figure (4).

$$\left.\begin{array}{l}\text{slope at point}\\ \text{A going in}\\ \text{x direction}\end{array}\right\} = \left.\frac{\partial V(x,y_A)}{\partial x}\right|_{x = x_A} = \tan\theta \qquad (15)$$

This is the maximum slope at point A. If we sawed through point A, orienting the saw blade in any other direction, the slop at point A would be less. In particular the slope would be zero if we oriented the saw in the y direction.

From this discussion we see that the vector $\vec{E}(x,y)$ points in the direction of the maximum slope and has a magnitude equal to that slope. The minus sign results from the fact that the force $\vec{E}$ is in the downward direction toward lower energy, while the positive slope, or ***gradient*** as we will call it, is in the upward direction.

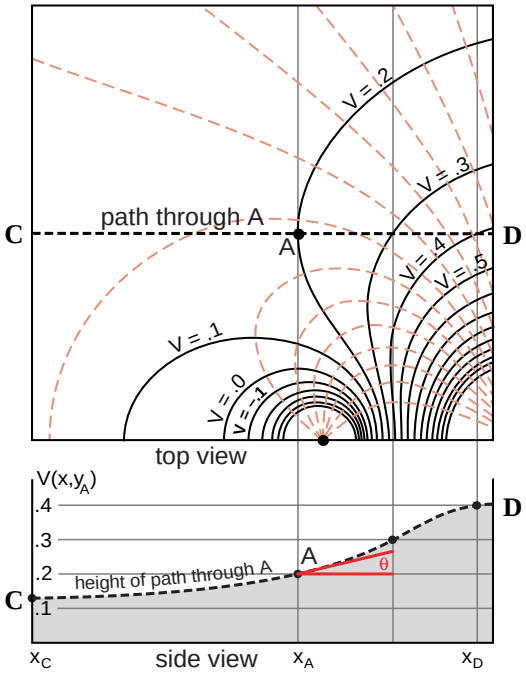



**Figure 4**
***The top view shows the point A and the horizontal path through that point. The side view shows the path we would have to climb if the surface were smooth. The steepest slope at the point A is in the +x direction and is the tangent of the angle labeled θ.***

## THE GRADIENT OPERATOR

The extension of Equation (12) to the case where the voltage varies in three dimensions, where V = V(x,y,z) is fairly obvious. It is

$$\vec{E}(x,y,z) = -\left[\hat{x}\,\frac{\partial V(x,y,z)}{\partial x} + \hat{y}\,\frac{\partial V(x,y,z)}{\partial y} + \hat{z}\,\frac{\partial V(x,y,z)}{\partial z}\right] \tag{16}$$

Until the beginning of the 20th century, research papers and textbooks dealing with partial derivatives used notation similar to Equation (16), and the formulas could become cumbersome and difficult to read. It was Willard Gibbs who introduced the ***gradient operation*** $\vec{\nabla}$ defined by the equation

$$\begin{aligned}\vec{\nabla} &\equiv \hat{x}\,\frac{\partial}{\partial x} + \hat{y}\,\frac{\partial}{\partial y} + \hat{z}\,\frac{\partial}{\partial z} \\ &\equiv \hat{x}\,\nabla_x + \hat{y}\,\nabla_y + \hat{z}\,\nabla_z\end{aligned} \tag{17}$$

where $\nabla_x = \partial/\partial x$ , etc.

We call $\vec{\nabla}$ an ***operator*** because it does not have an explicit meaning until it operates on something like the voltage function V(x,y,z).

$$\begin{aligned}\vec{\nabla}V(x,y,z) &= \hat{x}\,\frac{\partial V}{\partial x} + \hat{y}\,\frac{\partial V}{\partial y} + \hat{z}\,\frac{\partial V}{\partial z} \\ &= \hat{x}\,\nabla_x V + \hat{y}\,\nabla_y V + \hat{z}\,\nabla_z V\end{aligned} \tag{18}$$

With this notation, the formula for the electric field $\vec{E}(x,y,z)$ in terms of the voltage V(x,y,z) is

$$\boxed{\vec{E}(x,y,z) = -\vec{\nabla}V(x,y,z)} \tag{19}$$

We say that the ***electric field $\vec{E}$ is minus the gradient of the voltage V***.

In the Physics text, we defined a vector field as a quantity with a vector value at every point in space. We began our discussion of vector fields in Chapter 23 with the velocity field rather than the electric field because the velocity field is easier to visualize. At any point in space the vector is simply the velocity vector of the fluid particle located there. For the electric field we first have to invent the concept of a tiny unit test charge before we can visualize the force vector at each point in space.

Another mathematical concept, which we did not bother naming in the Physics text, is the ***scalar field***. It is a quantity that has a scalar or numerical value at every point in space. An example of a scalar field is voltage, the potential energy of a unit test charge. At every point in space that we place the unit test charge, we get a voltage reading. Since energy has a magnitude but does not point anywhere, this reading has a ***scalar*** or numerical value only.

From Equation (19), we see that the gradient operator $\vec{\nabla}$ , operating on a scalar field V creates the vector field $\vec{E} = -\vec{\nabla}V$ . The vector $\vec{\nabla}V$ has a numerical value equal to the maximum slope of V(x,y,z), and points opposite to the direction where the slope is greatest.

In the remainder of this part of the chapter, we will give examples of using the gradient operation to calculate the electric field from the voltage. In only a few cases, like the example of the parallel plate capacitor, is a Cartesian coordinate system (x,y,z) the most convenient coordinate system to use. In our study of electric and magnetic phenomena, we often dealt with point charges where there is spherical symmetry or line charges with cylindrical symmetry. We will see that to handle problems with spherical or cylindrical symmetry, it is much easier to work with the gradient $\vec{\nabla}V$ expressed in spherical or cylindrical coordinate systems. Much of the detailed work for the remainder of the chapter will be to work out the formulas for the gradient in these coordinate systems. (You do these derivations once, and then use the results for the remainder of your scientific career.)

As we mentioned, we have View 2 later in the chapter, where we look at the gradient from a more geometrical and mathematical point of view. We end up with Equation (16) as the formula for the gradient, but explicitly demonstrate that the components $\nabla_x V$ and $\nabla_y V$ of the gradient transform (change) the same way the components of a displacement vector change when we go to a rotated coordinate system. Such discussions will become very useful later on.

## THE PARALLEL PLATE CAPACITOR

We introduced the parallel plate capacitor in Chapter 26, page 26-14 of the Physics text. We dealt with an idealized situation where we assumed that the plate diameters were much greater than the separation. Then we could neglect edge effects and assume that the electric field was uniform between the plates, as shown in Figure (26-27) reproduced here.

Since $\vec{E}$ is the force on a unit test charge, and the voltage V is its potential energy, we can calculate the voltage V between the plates by calculating the amount of work required to lift the unit charge a distance y above the bottom plate. Since the force $\vec{E}$ we have to work against is constant, the work we do is simply the force of magnitude E times the height y. If we say that the bottom plate is grounded, i.e., define the potential energy or voltage as being zero at the bottom plate, then the formula for the voltage between the plates is simply

$$V(x,y,z) = E\,y \tag{20}$$

To evaluate E, we note that when we get to the top plate where y = d, the voltage is up to $V_0$, the voltage to which we charged the capacitor

$$V_0 = Ed \tag{21}$$

Thus $E = V_0/d$, and the voltage between the plates is given by

$$V(x,y,z) = \frac{V_0}{d}\,y \tag{22}$$

Let us now turn the problem around and use the gradient formula $\vec{E} = -\vec{\nabla}V$ to calculate the electric field $\vec{E}$ from our voltage formula Equation (22). Writing out all the components of $-\vec{\nabla}V$ as partial derivatives, we have from Equation (16)

$$\vec{E}(x,y,z) = -\left[\hat{x}\frac{\partial V}{\partial x} + \hat{y}\frac{\partial V}{\partial y} + \hat{z}\frac{\partial V}{\partial z}\right] \tag{16}$$

The x partial derivative is

$$\frac{\partial V(x,y,z)}{\partial x} = \frac{\partial}{\partial x}\left(\frac{V_0\,y}{d}\right) = 0 \tag{23}$$

This is zero because there is no x dependence in our formula for V. When we take the partial derivative with respect to x, we hold y and z constant. Thus nothing in the formula $V_0 y/d$ changes when we change x, and this partial derivative is zero.

The other partial derivatives are

$$\frac{\partial V(x,y,z)}{\partial y} = \frac{\partial}{\partial y}\left[\frac{V_0\,y}{d}\right] = \frac{V_0}{d} \tag{24}$$

$$\frac{\partial V(x,y,z)}{\partial z} = \frac{\partial}{\partial z}\left[\frac{V_0\,y}{d}\right] = 0 \tag{25}$$

Using Equations (23), (24), and (25) in (22) gives us

$$\vec{E} = -\hat{y}\frac{V_0}{d} \tag{26}$$

which says that $\vec{E}$ points down in the $-\hat{y}$ direction, and has a magnitude $V_0/d$ which we already know from Equation (21). We see that the calculation of $\vec{E}$ from V using $\vec{E} = -\vec{\nabla}V$ is a fairly straightforward process.

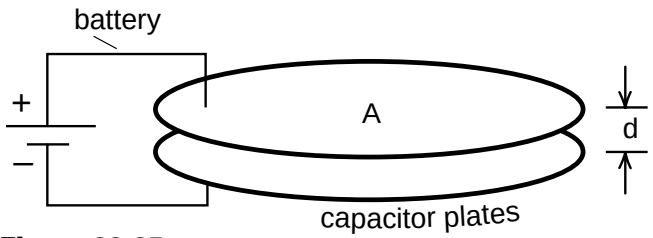



**Figure 26-25**
***The parallel plate capacitor. The capacitor is charged up by connecting a battery across the plates as shown.***

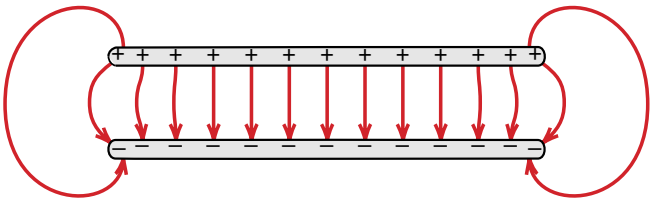


**Figure 26-26**
***The electric field between and around the edge of the capacitor plates.***

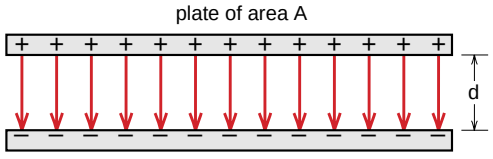



**Figure 26-27**
***In our idealized parallel plate capacitor the field lines go straight from the positive to the negative plate, and the field is uniform between the plates.***

## Voltage Inside a Conductor

The main idea of Chapter 26 of the Physics text was that you cannot have a static electric field inside a conductor if there is no flow of charge. The equivalent statement in terms of electric voltage is that the voltage is constant inside a conductor

$$V(x,y,z)_{\text{inside a conductor}} = \text{constant} \qquad (27)$$

To see that this gives a zero electric field, we have

$$\vec{E} = -\vec{\nabla} V_{\text{inside a conductor}} = 0 \qquad (28)$$

All the components are zero because the partial derivative of a constant is zero.

To provide an explicit example, suppose we turn our parallel plate capacitor on its side and assume that it is constructed from thick metal plates as shown in Figure (5). The voltage as a function of distance is shown below the drawing of the plates. Inside the left plate the voltage has the constant value $V_0$, which gives zero field inside. Between the plates the voltage drops uniformly. It has a constant gradient, which gives us a constant electric field $\vec{E} = -\vec{\nabla} V$ pointing in the direction of the downward slope. The voltage is again constant (V = 0) in the left hand plate.

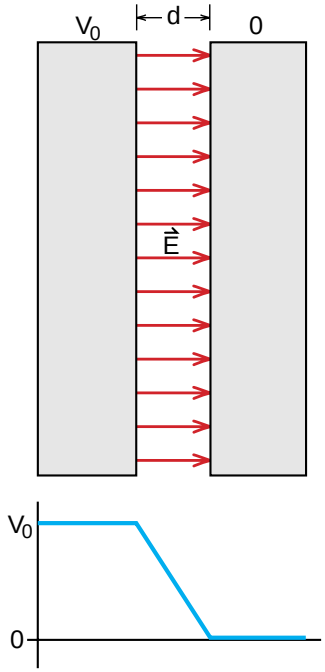



**Figure 5**
***Voltage in a parallel plate capacitor. The voltage is constant inside the plates and, for the assumed uniform field structure, drops uniformly between the plates.***

## ELECTRIC FIELD OF A POINT CHARGE

Our first example of an electric field in the Physics text was the field of a point charge. If we have a charge Q located at the origin of our coordinate system, then the electric field at a position $\vec{r} = (x,y,z)$ as shown in Figure (6) is given by

$$\vec{E}(r) = \hat{r}\frac{Q}{4\pi\varepsilon_0 r^2} = \hat{r}\frac{kQ}{r^2} \qquad (29)$$

where $\hat{r}$ is a unit vector in the $\vec{r}$ direction and $k = 1/4\pi\varepsilon_0$ .

In the Physics text, we mentioned, but never accurately derived, that the voltage V(r) of a point charge was

$$V(r) = \frac{Q}{4\pi\varepsilon_0 r} = \frac{kQ}{r} \qquad (30)$$

when we chose the zero of potential energy at r = infinity. What we want to do now is to show that the formula for $\vec{E}(r)$ follows directly from Equation (30) for V(r) when we use the relationship

$$\vec{E} = -\vec{\nabla}V \qquad (14) \text{ repeated}$$

The work is a bit messy, because we will be using a Cartesian coordinate system to solve a problem with spherical symmetry. Later we will find the formula for the gradient in spherical coordinates, and then see that it is very easy to evaluate $\vec{E} = -\vec{\nabla}V$ for a point charge.

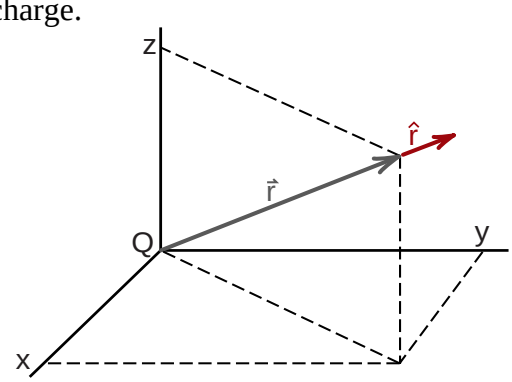



**Figure 6**
***Out at a point given by the coordinate vector $\vec{r}$, we have the unit vector $\hat{r}$.***

Our first step will be to write out the vector equation $\vec{E} = -\vec{\nabla}V$ as three component equations

$$E_x = -\frac{\partial V}{\partial x} \; ; \; E_y = -\frac{\partial V}{\partial y} \; ; \; E_z = -\frac{\partial V}{\partial z} \qquad (31)$$

Focusing on the x component equation we have

$$E_x = -\frac{\partial V}{\partial x} = -\frac{\partial}{\partial x}\left(\frac{kQ}{r}\right)$$

Taking the constant kQ outside the derivative we have

$$E_x = -kQ\frac{\partial}{\partial x}\left(\frac{1}{r}\right) \qquad (32)$$

To go any farther, we have to express the distance r as a function of the coordinate x. This is done by the three dimensional Pythagorean theorem

$$r = \sqrt{x^2 + y^2 + z^2}$$

To calculate the derivative of (1/r) with respect to x now becomes an exercise in the use of the chain rule for differentiation. Let us start with

$$r^2 = x^2 + y^2 + z^2$$

which is easy to differentiate. We get

$$\frac{\partial r^2}{\partial x} = \frac{\partial}{\partial x}\left(x^2 + y^2 + z^2\right) = 2x \qquad (33)$$

Next look at

$$\frac{\partial r}{\partial x} = \frac{\partial}{\partial x}\sqrt{r^2} = \frac{\partial\sqrt{r^2}}{\partial r^2}\frac{\partial r^2}{\partial x} \qquad (34)$$

To evaluate $\partial\sqrt{r^2}/\partial r^2$ , set $y = r^2$ so that we have, using $\partial y^n/\partial y = ny^{n-1}$

$$\frac{\partial\sqrt{r^2}}{\partial r^2} = \frac{\partial\sqrt{y}}{\partial y} = \frac{\partial}{\partial y}y^{.5} = \frac{1}{2}y^{-.5} = \frac{1}{2r} \qquad (35)$$

Thus using Equation (33) and (35) in (34) gives

$$\frac{\partial r}{\partial x} = \frac{\partial\sqrt{r^2}}{\partial r^2}\frac{\partial r^2}{\partial x^2} = \left(\frac{1}{2r}\right)(2x)$$

$$\boxed{\frac{\partial r}{\partial x} = \frac{x}{r}} \qquad (36)$$

which is a fairly simple result considering what we went through.

Finally we have

$$\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = \frac{\partial r^{-1}}{\partial x} = \frac{\partial r^{-1}}{\partial r}\frac{\partial r}{\partial x}$$

$$= -r^{-2}\frac{\partial r}{\partial x} = -\frac{1}{r^2}\left(\frac{x}{r}\right)$$

$$\boxed{\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = -\frac{x}{r^3}} \qquad (37)$$

and our formula for $E_x$ becomes

$$E_x = -kQ\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = -kQ\left(-\frac{x}{r^3}\right)$$

$$\boxed{E_x = kQ\frac{x}{r^3}} \qquad (38a)$$

Clearly the y and z components are

$$\boxed{E_y = kQ\frac{y}{r^3}} \qquad (38b)$$

$$\boxed{E_z = kQ\frac{z}{r^3}} \qquad (38c)$$

To check that we got the right answer, we can go back to Equation (29)

$$\vec{E}(r) = \hat{r}\frac{kQ}{r^2} \qquad (29) \text{ repeated}$$

and replace the unit vector $\hat{r}$ with its definition $\vec{r}/r$ giving

$$\vec{r} = (r_x, r_y, r_z) = (x,y,z)$$

$$\hat{r} = \frac{\vec{r}}{r} = \frac{1}{r}(x,y,z)$$

$$\hat{r}_x = \frac{x}{r}\ ;\ \hat{r}_y = \frac{y}{r}\ ;\ \hat{r}_z = \frac{z}{r} \qquad (39)$$

Equation (39) says, for example, that the x component of the unit vector $\hat{r}$ has a length x/r. Thus the x component of $\vec{E}$ in Equation (29) is

$$E_x = \hat{r}_x\frac{kQ}{r^2} = \frac{x}{r}\frac{kQ}{r^2} = kQ\frac{x}{r^3} \qquad (40)$$

with similar equations for $E_y$ and $E_z$. Since Equations (38) and (40) are the same, we have verified that $\vec{E} = -\vec{\nabla}V$ gives the correct result for V = kQ/r.

The messiness we encountered calculating the field of a point charge from V = kQ/r resulted from our calculating x, y, and z components of $\vec{E}$ when we knew that $\vec{E}$ pointed in the radial direction. If we use what is called a ***spherical coordinate system***, we will find that the formula for the radial component of the electric field is simply

$$E_r = \frac{-\partial V(r)}{\partial r} \qquad (41)$$

With V(r) = kQ/r we get

$$E_r = -kQ\frac{\partial}{\partial r}\left(\frac{1}{r}\right) = -kQ\left(-\frac{1}{r^2}\right) = \frac{kQ}{r^2} \qquad (42)$$

and we get the final answer in a one line calculation.

To get this simple result requires, however, a fair amount of work deriving the formula for the gradient in spherical coordinates. First we have to define precisely what a spherical coordinate system is, show what the unit vectors are, and then calculate the components of the gradient when we move in the directions defined by the unit vectors. When this is all done, when we have the formula for the gradient in spherical coordinates, we can use the formula without ever going through the derivation again.

In the Physics text we encountered problems with plane symmetry, like the parallel plate capacitor, cylindrical symmetry, like the field of a line charge, and spherical symmetry like the field of a point charge we have just discussed. The plane symmetry problems are most easily handled in a Cartesian coordinate system, the cylindrical problems in what is called a ***cylindrical coordinate system***, and spherical problems in a spherical coordinate system. We will now discuss these three coordinate systems and develop the formulas for the components of the gradient vector in each coordinate system. Since we have already done this for the Cartesian coordinate system, that discussion will serve as a review of the procedure we will use.

## GRADIENT IN THE CARTESIAN COORDINATE SYSTEM

An example of a right handed Cartesian coordinate system is shown in Figure (7). Out at some point $\vec{r} = (x,y,z)$ the unit vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$ are parallel to the x, y, and z axis as shown. It is called a right handed coordinate system because the unit vectors obey the relationship

$$\hat{x} \times \hat{y} = \hat{z} \qquad (43)$$

when we use the right hand rule for the cross product. (If we used a left hand rule, the z axis would point the other way.)

---

**Exercise 1**

Show that $\hat{y} \times \hat{z} = \hat{x}$ and $\hat{z} \times \hat{x} = \hat{y}$.

---

We will use the force/potential energy relationship to define the gradient vector. If I move a unit test charge a short distance $\Delta\vec{\ell}$, exerting a force $\vec{F}_{me} = -\vec{E}$ to just overcome the electric field $\vec{E}$, the work $\Delta W$ I do is

$$\Delta W = \vec{F}_{me} \cdot \Delta\vec{\ell} = -\vec{E} \cdot \Delta\vec{\ell} \qquad (44)$$

Since this work is the change $\Delta V$ in the potential energy of the unit test charge, we have

$$\Delta V = -\vec{E} \cdot \Delta\vec{\ell} \qquad (45)$$

But the voltage V is related to the field $\vec{E}$ by the gradient

$$\vec{E} = -\vec{\nabla}V \qquad (14) \text{ repeated}$$

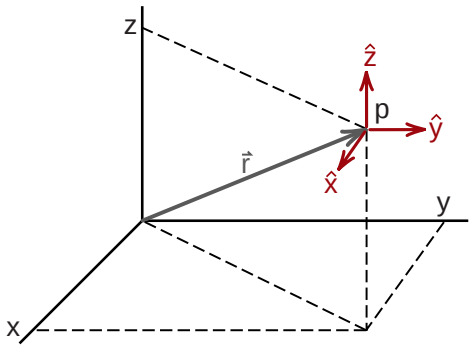



**Figure 7**

***The unit vectors $\hat{x}, \hat{y}, \hat{z}$ out at the point $\vec{r}$.***

Using Equation (14) in (45), we can eliminate $\vec{E}$ and get the relationship between the small change in voltage $\Delta V$ and the voltage gradient $\vec{\nabla}V$

$$\boxed{\Delta V = (\vec{\nabla}V) \cdot \Delta\vec{\ell}} \qquad (46)$$

Equation (46) will allow us to find the formula for the gradient in the various coordinate systems.

To see how we are going to use Equation (46), we will start with the Cartesian coordinate system and choose $\Delta\vec{\ell}$ to be a short step $\Delta x$ in the x direction. Explicitly we will start at a point (x,y,z) and move to the point $(x + \Delta x, y, z)$ so that $\Delta V$, $\Delta\vec{\ell}$ and $(\vec{\nabla}V) \cdot \Delta\vec{\ell}$ become

$$\Delta V = V(x + \Delta x, y, z) - V(x, y, z) \qquad (47)$$

$$\Delta\vec{\ell} = \hat{x}\Delta x \qquad (48)$$

$$(\vec{\nabla}V) \cdot \Delta\vec{\ell} = (\vec{\nabla}V)_x \, \Delta x \qquad (49)$$

Using (47) and (49) in (46) gives

$$V(x + \Delta x, y, z) - V(x, y, z) = (\vec{\nabla}V)_x \, \Delta x \qquad (50)$$

Dividing through by $\Delta x$ and taking the limit as $\Delta x$ goes to zero gives

$$(\vec{\nabla}V)_x = \lim_{\Delta x \to 0} \left[ \frac{V(x + \Delta x, y, z) - V(x, y, z)}{\Delta x} \right] \qquad (51)$$

which is the definition of the partial derivative. Thus

$$(\vec{\nabla}V)_x = \frac{\partial V(x, y, z)}{\partial x} \qquad (52)$$

which is our earlier result. This procedure does not give us anything new for a Cartesian coordinate system, but will give us new results for other coordinate systems.

*(On the next page you will find two pictures of our model of the electric field of two point charges. We put the pictures there so that the discussion of the gradient in cylindrical and spherical coordinates would each be completed on facing pages.)*

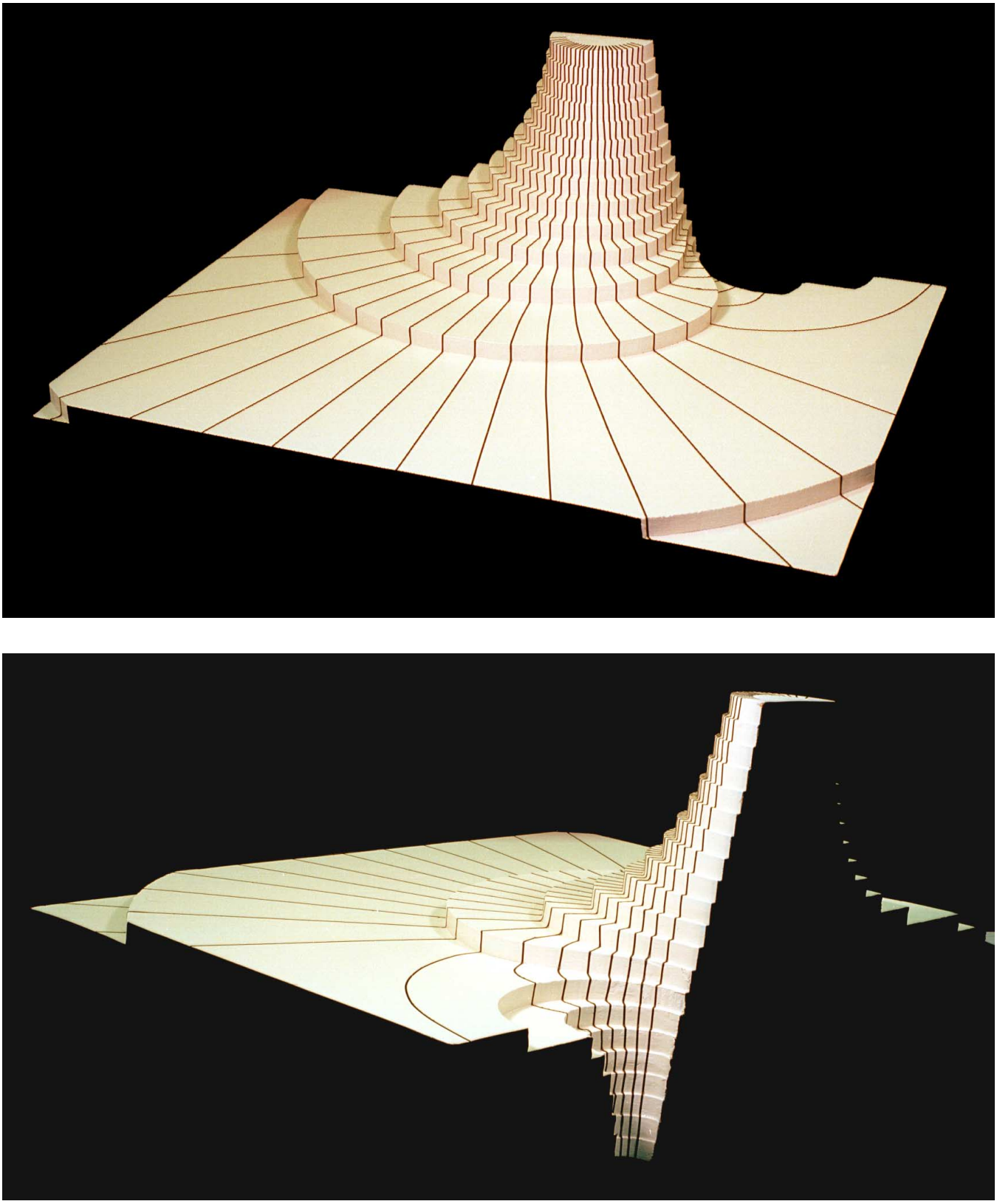


**Figure 25-14 (from Physics text)**
***Different views of the model of the electric field of two point charges $Q_+ = +3$ and $Q_- = -1$.***

## GRADIENT IN CYLINDRICAL COORDINATES

In a cylindrical coordinate system, we define the location of a point p by giving the distance $\vec{r}$ out from the z axis, the angle $\theta$ over from the x axis, and the height z above the xy plane as shown in Figure (8). The unit vectors are $\hat{r}$ which points radially out from the z axis, $\hat{z}$ which points in the z direction, and $\hat{\theta}$ which is perpendicular to the $\hat{r}\,\hat{z}$ plane. The direction of $\hat{\theta}$ is the direction we move when increasing the angle $\theta$. This gives us a right handed coordinate system where the unit vectors are related by

$$\hat{r} \times \hat{\theta} = \hat{z} \tag{53}$$

You should check for yourself that Equation (53) works for the unit vectors shown in Figure (8), and that $\hat{\theta} \times \hat{z} = \hat{r}$ and $\hat{z} \times \hat{r} = \hat{\theta}$.

We will assume that in cylindrical coordinates, the gradient vector at point p is given by the equation

$$\vec{\nabla}V = \hat{r}\,(\vec{\nabla}V)_r + \hat{\theta}\,(\vec{\nabla}V)_\theta + \hat{z}\,(\vec{\nabla}V)_z \tag{54}$$

where $(\vec{\nabla}V)_r$, $(\vec{\nabla}V)_\theta$ and $(\vec{\nabla}V)_z$ are the components of the gradient vector that we want to determine.

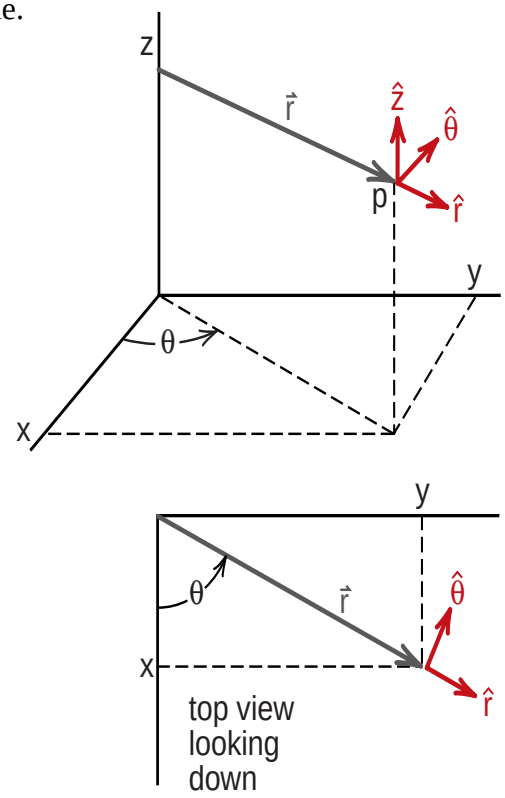



**Figure 8**
***The unit vectors $\hat{r}$, $\hat{\theta}$, $\hat{z}$ in cylindrical coordinates.***

To calculate the first component $(\vec{\nabla}V)_r$, we will start at the point p at (r, θ, z) and move a short distance $\Delta r$ in the $\hat{r}$ direction, to the point (r+Δr, θ, z). Our change in voltage $\Delta V$, displacement $\vec{\Delta\ell}$ and the dot product $(\vec{\nabla}V) \cdot \vec{\Delta\ell}$ are for this move

$$\Delta V = V(r + \Delta r, \theta, z) - V(r, \theta, z) \tag{55}$$

$$\vec{\Delta\ell} = \hat{r}\,\Delta r \tag{56}$$

$$(\vec{\nabla}V)\cdot\vec{\Delta\ell} = \left[\hat{r}\,(\vec{\nabla}V)_r + \hat{\theta}\,(\vec{\nabla}V)_\theta + \hat{z}\,(\vec{\nabla}V)_z\right]\cdot \hat{r}\,\Delta r \tag{57}$$

Since the unit vectors are all at right angles to each other, $\hat{r} \cdot \hat{r} = 1$, $\hat{\theta} \cdot \hat{r} = 0$ and $\hat{z} \cdot \hat{r} = 0$, giving us

$$\Delta V = (\vec{\nabla}V) \cdot \vec{\Delta\ell} = (\vec{\nabla}V)_r \Delta r \tag{58}$$

Dividing (58) through by $\Delta r$, using (55) for $\Delta V$ and taking the limit as $\Delta r$ goes to zero gives

$$(\vec{\nabla}V)_r = \lim_{\Delta r \to 0}\left[\frac{V(r + \Delta r, \theta, z) - V(r, \theta, z)}{\Delta r}\right] \tag{59}$$

The right side of Equation (59) is what we will define to be the partial derivative of V(r, θ, z) with respect to r in cylindrical coordinates

$$\frac{\partial V(r, \theta, z)}{\partial r} \equiv \lim_{\Delta r \to 0}\left[\frac{V(r + \Delta r, \theta, z) - V(r, \theta, z)}{\Delta r}\right] \tag{60}$$

This is the rate of change of the function V(r, θ, z) as we change the r coordinate. With this definition, we get

$$\boxed{(\vec{\nabla}V)_r = \frac{\partial V(r, \theta, z)}{\partial r}} \tag{61}$$

So far, our results look very much like what we had for Cartesian coordinates. However, we get something new when our step $\Delta\vec{\ell}$ is in the $\hat{\theta}$ direction. Suppose we are at the position (r, θ, z) , and move to the new point (r, θ+Δθ, z) where we increased the θ coordinate angle by Δθ as shown in Figure (9). Since the angle θ+Δθ is measured in radians, the arc length $\Delta\vec{\ell}$ that we move when going from θ to θ+Δθ is

$$\Delta\ell = r\Delta\theta$$

You will notice that the vector displacement $\Delta\vec{\ell}$ is in the same direction as the $\hat{\theta}$ unit vector, thus

$$\Delta\vec{\ell} = \hat{\theta} r\Delta\theta \qquad (62)$$

The change in voltage ΔV and the dot product $(\vec{\nabla}V)\cdot\Delta\vec{\ell}$ are thus

$$\Delta V = V(r, \theta+\Delta\theta, z) - V(r, \theta, z) \qquad (63)$$

$$\vec{\nabla}V\cdot\Delta\vec{\ell} = \left[\hat{r}(\vec{\nabla}V)_r + \hat{\theta}(\vec{\nabla}V)_\theta + \hat{z}(\vec{\nabla}V)_z\right]\cdot\hat{\theta} r\Delta\theta$$

$$= (\vec{\nabla}V)_\theta r\Delta\theta \qquad (64)$$

where we used $\hat{\theta}\cdot\hat{\theta} = 1$, $\hat{r}\cdot\hat{\theta} = \hat{z}\cdot\hat{\theta} = 0$.

Using (63) and (64) in our equation $\Delta V = \vec{\nabla}V\cdot\Delta\vec{\ell}$, we get

$$V(r, \theta+\Delta\theta, z) - V(r, \theta, z) = (\vec{\nabla}V)_\theta r\Delta\theta \qquad (65)$$

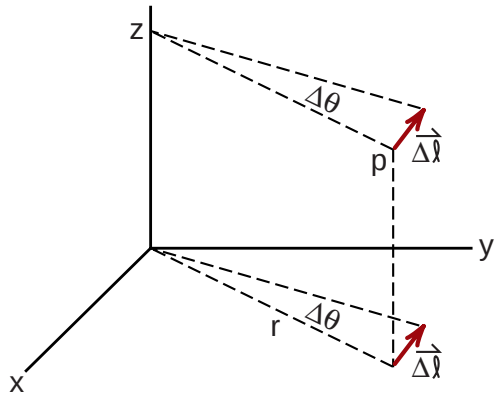



**Figure 9**
***The displacement $\Delta\vec{\ell}$ when we increase the angle θ by Δθ.***

Dividing Equation (65) through by rΔθ and then taking the limit as Δθ goes to zero gives

$$(\vec{\nabla}V)_\theta = \frac{1}{r}\left\{\lim_{\Delta\theta\to 0}\left[\frac{V(r, \theta+\Delta\theta, z) - V(r, \theta, z)}{\Delta\theta}\right]\right\} \qquad (66)$$

We define the quantity in curly brackets to be the partial derivative of V(r, θ, z) with respect to the variable θ

$$\frac{\partial V(r, \theta, z)}{\partial\theta} \equiv \lim_{\Delta\theta\to 0}\left[\frac{V(r, \theta+\Delta\theta, z) - V(r, \theta, z)}{\Delta\theta}\right] \qquad (67)$$

Thus we end up with the equation

$$\boxed{(\vec{\nabla}V)_\theta = \frac{1}{r}\frac{\partial V(r, \theta, z)}{\partial\theta}} \qquad (68)$$

and we get a factor of 1/r in our formula for the θ component of the gradient in cylindrical coordinates. The factor of 1/r appears because the partial derivative with respect to θ measures the rate of change of V for a given change Δθ in angle, while the gradient measures the rate of change of V with respect to a given step in distance. When we make a change Δθ in angle, the distance we move is rΔθ which increases with r. The factor of r has to be divided out to get the rate of change of V with distance.

---

**Exercise 2**

Following the above steps, show that

$$\boxed{(\vec{\nabla}V)_z = \frac{\partial V(r, \theta, z)}{\partial z}} \qquad (69)$$

This should look the same as our derivation for the Cartesian coordinate system.

---

## GRADIENT IN SPHERICAL COORDINATES

While the steps are fresh, let us derive the formulas for the components of the gradient vector in spherical coordinates. We will then return to various applications of the new gradient formulas.

In the spherical coordinate system shown in Figure (10), a point p is located by the displacement $\vec{r}$ from the origin, the angle $\theta$ that the coordinate vector $\vec{r}$ makes with the z axis, and the angle $\phi$ that the projection of $\vec{r}$ on the x,y plane makes with the x axis. The unit vectors are $\hat{r}$ pointing out in the $\vec{r}$ direction, $\hat{\theta}$ which lies in the $\hat{r}\,\hat{z}$ plane pointing in the direction of increasing $\theta$, and $\hat{\phi}$ which is perpendicular to the $\hat{r}\,\hat{z}$ plane, in the direction of increasing $\phi$. This gives us a right handed coordinate system where

$$\hat{r} \times \hat{\theta} = \hat{\phi} \tag{70}$$

(Again, show for yourself that $\hat{\theta} \times \hat{\phi} = \hat{r}$ and $\hat{\phi} \times \hat{r} = \hat{\theta}$.)

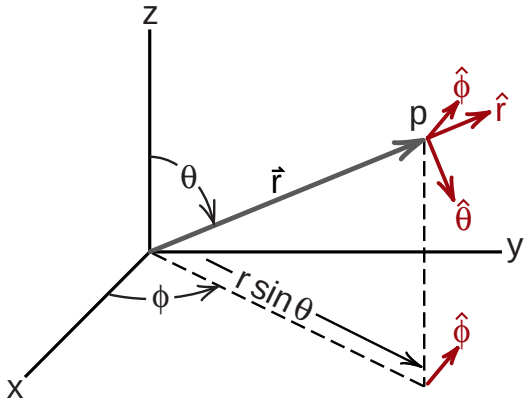



**Figure 10**
***The unit vectors $\hat{r}$, $\hat{\theta}$, $\hat{\phi}$ for a spherical coordinate system.***

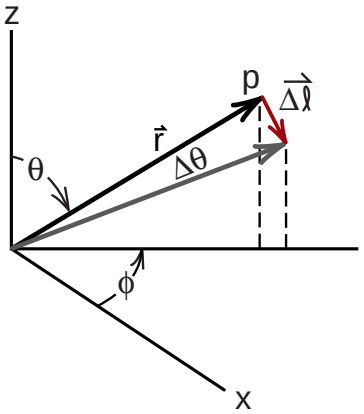



**Figure 11**
***The step $\Delta\vec{\ell}$ when we increase $\theta$ by $\Delta\theta$. We are directly facing the rz plane.***

---

**Exercise 3**

Start at the point (r, θ, ϕ) and move a distance $\Delta\vec{\ell}$ to the point V(r + Δr, θ, ϕ) and show that the r component of the gradient in spherical coordinates is

$$\boxed{(\vec{\nabla}V)_r = \frac{\partial V(r,\theta,\phi)}{\partial r}} \tag{71}$$

where

$$\frac{\partial V(r,\theta,\phi)}{\partial r} = \lim_{\Delta r \to 0}\left[\frac{V(r+\Delta r,\theta,\phi) - V(r,\theta,\phi)}{\Delta r}\right] \tag{72}$$

It was Equation (71) that we used to show in one line that the voltage V = kQ/r leads to the field $\vec{E} = \hat{r}\,kQ/r^2$.

---

In spherical coordinates, the radial component of the gradient is simply the partial derivative, as we asked you to show in Exercise 3. We get new results when we look at the $\theta$ and $\phi$ components, where the change in distance $\Delta\vec{\ell}$ is not equal to $\Delta\theta$ or $\Delta\phi$ alone.

First let $\Delta\vec{\ell}$ be in the $\hat{\theta}$ direction, so that we go from the point (r, θ, ϕ) to (r, θ+Δθ, ϕ). The distance $\Delta\vec{\ell}$ is shown in Figure (11) where we are looking squarely at the rz plane. You can see that $\Delta\vec{\ell}$ is in the $\hat{\theta}$ direction and has a magnitude $\Delta\ell = r\Delta\theta$ so that

$$\Delta\vec{\ell} = \hat{\theta}\, r\Delta\theta \tag{73}$$

The change in voltage $\Delta V$ and the dot product $\vec{\nabla}V \cdot \Delta\vec{\ell}$ are

$$\Delta V = V(r, \theta+\Delta\theta, \phi) - V(r, \theta, \phi) \tag{74}$$

$$\vec{\nabla}V \cdot \Delta\vec{\ell} = \left[\hat{r}(\vec{\nabla}V)_r + \hat{\theta}(\vec{\nabla}V)_\theta + \hat{\phi}(\vec{\nabla}V)_\phi\right] \cdot \hat{\theta}\, r\Delta\theta$$

$$= (\vec{\nabla}V)_\theta\, r\Delta\theta \tag{75}$$

where $\hat{\theta}\cdot\hat{\theta} = 1$, $\hat{r}\cdot\hat{\theta} = \hat{\phi}\cdot\hat{\theta} = 0$.

Equating $\Delta V$ from (74) with $\vec{\nabla}V \cdot \Delta\vec{\ell}$ in (75), then dividing through by $r\Delta\theta$ and taking the limit as $\Delta\theta$ goes to zero, gives

$$(\vec{\nabla}V)_\theta = \frac{1}{r}\lim_{\Delta\theta \to 0}\left[\frac{V(r, \theta+\Delta\theta, \phi) - V(r, \theta, \phi)}{\Delta\theta}\right] \tag{76}$$

We define the partial derivative of $V(r, \theta, \phi)$ with respect to $\theta$ in spherical coordinates as

$$\frac{\partial V(r, \theta, \phi)}{\partial \theta} \equiv \lim_{\Delta\theta \to 0} \left[ \frac{V(r, \theta + \Delta\theta, \phi) - V(r, \theta, \phi)}{\Delta\theta} \right] \tag{77}$$

so that we get

$$\boxed{(\vec{\nabla}V)_\theta = \frac{1}{r} \frac{\partial V(r, \theta, \phi)}{\partial \theta}} \tag{78}$$

as the formula for the $\theta$ component of the gradient vector in spherical coordinates.

Finally we will derive the $\phi$ component of $\vec{\nabla}V$ by taking a step $\Delta\vec{\ell}$ in the $\phi$ direction. The geometry is shown in Figure (12). The first thing to note is that the projection of the coordinate vector $\vec{r}$ down on the xy plane has a length $(r \sin\theta)$. This is the distance the point p is out from the z axis. When we rotate an angle $\Delta\phi$ about the z axis, the arc length $\Delta\ell$ out a distance $(r \sin\theta)$ is $(r \sin\theta)\Delta\phi$. This distance is in the direction of the unit vector $\hat{\phi}$, thus

$$\Delta\vec{\ell} = \hat{\phi}\,(r \sin\theta)\Delta\phi \tag{79}$$

The change in voltage, going from $(r, \theta, \phi)$ to $(r, \theta, \phi + \Delta\phi)$ is

$$\Delta V = V(r, \theta, \phi + \Delta\phi) - V(r, \theta, \phi) \tag{80}$$

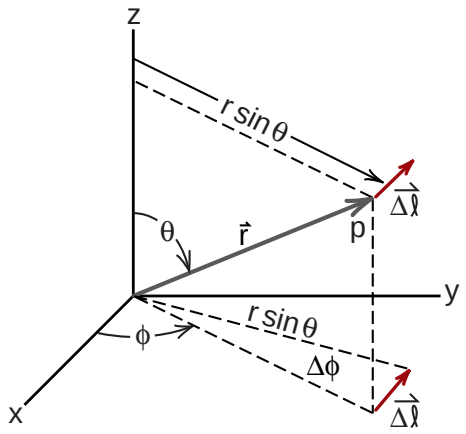



**Figure 12**
***The step $\Delta\vec{\ell}$ when we increase $\phi$ by $\Delta\phi$. Note that we are out a distance $r \sin\theta$ from the z axis.***

The quantity $\vec{\nabla}V \cdot \Delta\vec{\ell}$ is

$$\begin{aligned} \vec{\nabla}V \cdot \Delta\vec{\ell} &= \left[ \hat{r}(\vec{\nabla}V)_r + \hat{\theta}(\vec{\nabla}V)_\theta + \hat{\phi}(\vec{\nabla}V)_\phi \right] \cdot \hat{\phi}\,(r \sin\theta)\Delta\phi \\ &= (\vec{\nabla}V)_\phi\,(r \sin\theta)\Delta\phi \end{aligned} \tag{81}$$

because $\hat{\phi} \cdot \hat{\phi} = 1$ and $\hat{r} \cdot \hat{\phi} = \hat{\theta} \cdot \hat{\phi} = 0$.

Equating $\vec{\nabla}V \cdot \Delta\vec{\ell}$ in Equation (81) to $\Delta V$ in (80) gives

$$V(r, \theta, \phi + \Delta\phi) - V(r, \theta, \phi) = (\vec{\nabla}V)_\phi\,(r \sin\theta)\Delta\phi \tag{82}$$

Dividing (82) through by $(r \sin\theta)\Delta\phi$ and taking the limit at $\Delta\phi$ goes to zero gives

$$(\vec{\nabla}V)_\phi = \frac{1}{r \sin\theta} \lim_{\Delta\phi \to 0} \left[ \frac{V(r, \theta, \phi + \Delta\phi) - V(r, \theta, \phi)}{\Delta\phi} \right] \tag{83}$$

We define the partial derivative with respect to $\phi$ in spherical coordinates as

$$\frac{\partial V(r, \theta, \phi)}{\partial \phi} = \lim_{\Delta\phi \to 0} \left[ \frac{V(r, \theta, \phi + \Delta\phi) - V(r, \theta, \phi)}{\Delta\phi} \right] \tag{84}$$

to get the result

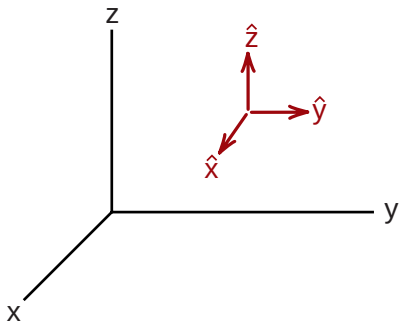
$$\boxed{(\vec{\nabla}V)_\phi = \frac{1}{r \sin\theta} \frac{\partial V(r, \theta, \phi)}{\partial \phi}} \tag{85}$$

## SUMMARY OF GRADIENT FORMULAS

We collect in one place the formulas for the gradient in Cartesian, cylindrical and spherical coordinates.

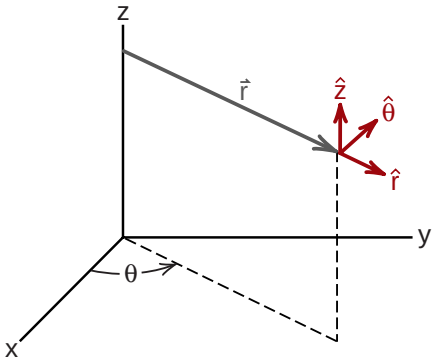
### *Cartesian Coordinates*

$$\vec{\nabla}V(x,y,z) = \hat{x}\frac{\partial V}{\partial x} + \hat{y}\frac{\partial V}{\partial y} + \hat{z}\frac{\partial V}{\partial z} \qquad (86)$$



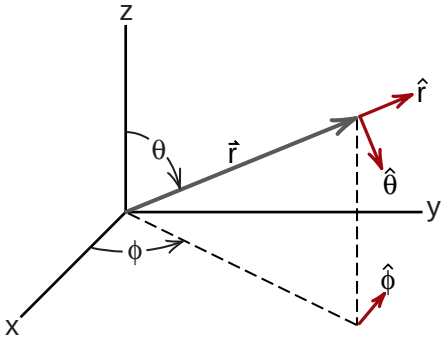

### *Cylindrical Coordinates*

$$\vec{\nabla}V(r,\phi,z) = \hat{r}\frac{\partial V}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial V}{\partial \theta} + \hat{z}\frac{\partial V}{\partial z} \qquad (87)$$




### *Spherical Coordinates*

$$\vec{\nabla}V(r,\theta,\phi) = \hat{r}\frac{\partial V}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial V}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial V}{\partial \phi} \qquad (88)$$




## EXAMPLES

### Electric Field of a Point Charge

Let us now see explicitly how the formula for the gradient in spherical coordinates, Equation (88), makes it easy to calculate the electric field of a point charge, starting from the voltage formula

$$V(r) = \frac{kQ}{r} \qquad (27) \text{ repeated}$$

The formula for the gradient in spherical coordinates is

$$\vec{\nabla}V = \hat{r}\frac{\partial V}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial V}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial V}{\partial \phi} \quad (88) \text{ repeat}$$

While Equation (88) looks somewhat messy, the thing to note is that V(r) has no dependence on the variables $\theta$ and $\phi$, thus the partial derivatives with respect to these variables are zero

$$\frac{\partial V(r)}{\partial \theta} = 0 \; ; \quad \frac{\partial V(r)}{\partial \phi} = 0 \qquad (89)$$

and all we are left with is

$$\vec{\nabla}V = \hat{r}\frac{\partial V(r)}{\partial r} \qquad (90)$$

We have for $\partial V(r)/\partial r$

$$\frac{\partial}{\partial r}\left(\frac{kQ}{r}\right) = kQ\frac{\partial}{\partial r}(r^{-1}) = -1\frac{kQ}{r^2} \qquad (91)$$

thus we get

$$\vec{E} = -\vec{\nabla}V = -\left(-\hat{r}\frac{kQ}{r^2}\right) = \hat{r}\frac{kQ}{r^2} \qquad (92)$$

which is the correct answer.

The advantage of using spherical coordinates to calculate the field of a point charge was that, two out of three of the components of the gradient were zero, and we had only a simple derivative for the remaining component. This is the kind of simplification you get when you use a coordinate system that matches the symmetry of the problem at hand. Our next example will be the calculation of the electric field of a line charge. That problem has cylindrical symmetry, and is most easily handled using a cylindrical coordinate system.

## Electric Field of a Line Charge

In the Physics text, our first calculation of the electric field of an extended object was to show that the radially directed electric field of a charged wire, shown in Figure (24-27) repeated here, had a magnitude

$$E(r) = \frac{\lambda}{2\pi\varepsilon_0 r} \qquad \text{(24-43) repeated}$$

where $\lambda$ is the amount of charge per meter on the wire and r is the radial distance out from the wire. To simplify the constants, we will set $k = 1/2\pi\varepsilon_0$ so that the vector formula for $\vec{E}$ is

$$\vec{E}(r) = \hat{r}\,\frac{k\lambda}{r}\ ; \qquad k = \frac{1}{2\pi\varepsilon_0} \qquad (93)$$

In the Physics text we never did say what the voltage was in the vicinity of a charged wire. You will see why shortly.

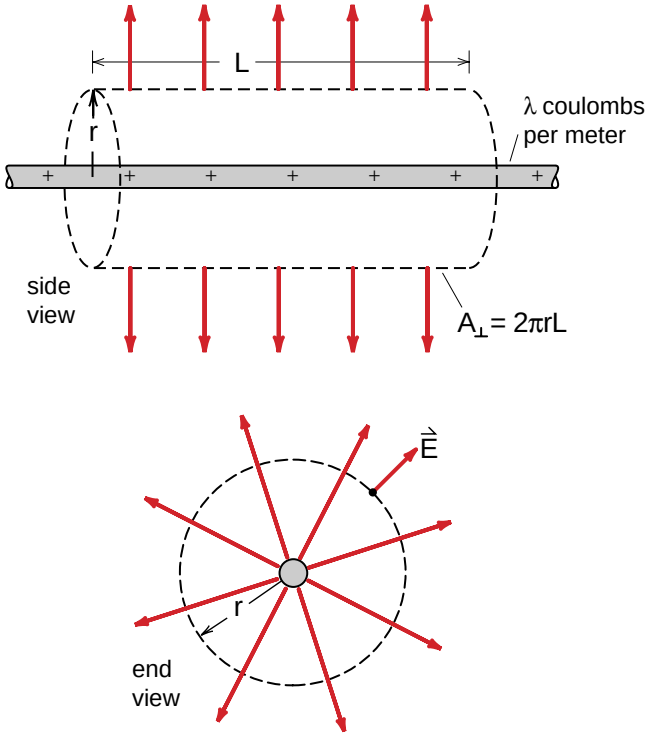



**Figure 24-27 (repeated)**
***Using Gauss' law to calculate the electric field of a line charge. Draw the Gaussian surface around a section of the rod. The flux all flows out through the cylindrical surface.***

We can assume, because of the cylindrical symmetry of the problem, that the voltage V depends only on the radial distance r out from the wire. That is, that V = V(r). Thus the partial derivatives with respect to the variables $\theta$ and z (using cylindrical coordinates) should be zero and we should be left with

$$\vec{E} = -\vec{\nabla}V = -\left(\hat{r}\,\frac{\partial V(r)}{\partial r} + \frac{\hat{\theta}}{r}\,\frac{\partial V(r)}{\partial \theta} + \hat{z}\,\frac{\partial V(r)}{\partial z}\right)$$

$$= -\hat{r}\,\frac{\partial V(r)}{\partial r} \qquad (94)$$

where we used Equation (87) for the gradient in cylindrical coordinates.

Comparing Equations (93) and (94) for $\vec{E}$ we get

$$\vec{E} = \hat{r}\,\frac{k\lambda}{r} = \hat{r}\left(-\frac{\partial V(r)}{\partial r}\right) \qquad (95)$$

As a result, the voltage V(r) should obey the equation

$$\frac{\partial V(r)}{\partial r} = -k\lambda\left(\frac{1}{r}\right) \qquad (96)$$

The question we have now is, what function of r, when differentiated with respect to r, gives 1/r? The answer, you may recall from Chapter 1 of the Calculus text, is the *natural logarithm*. Explicitly

$$\frac{d}{dr}(\ln r) = \frac{1}{r} \qquad (97)$$

Thus the appropriate voltage V(r) is

$$\boxed{V(r) = -k\lambda \ln r} \qquad (98)$$

Going back from this V(r) to $\vec{E}$ we have

$$\vec{\nabla}V(r) = \hat{r}\,\frac{\partial}{\partial r}(-k\,\lambda\,\ln r)$$

$$= \hat{r}\left(-k\,\lambda\,\frac{\partial \ln r}{\partial r}\right) = \hat{r}\left(\frac{-k\,\lambda}{r}\right) \qquad (99)$$

and

$$\vec{E}(r) = -\vec{\nabla}V(r) = \hat{r}\left(\frac{k\lambda}{r}\right) \qquad (100)$$

This explicitly checks that the voltage $-(k\lambda \ln r)$ leads to the electric field of a line charge.

The logarithm ln(r) that appears in Equation (100) is an interesting function in that it is zero at r = 1, goes to $-\infty$ at r = 0 and $+\infty$ at $r = \infty$ as shown in Figure (13). Thus, for example, at r = 0 we get

$$V(r)\Big|_{r=0} = -kr\ln(0) = -kr(-\infty) = +\infty \quad (101)$$

and the voltage becomes infinite. This tells us that it is not physically reasonable to put a finite charge density $\lambda$ on an infinitely thin wire. We had the same problem with a point charge. The formula V = kQ/r also goes to infinity at r = 0 which tells us we have a problem with the potential energy of a point charge of zero radius. (The modern theory of quantum electrodynamics treats the electron as a point charge of zero radius. The tricky part of the theory is to get around the infinities that result from this.)

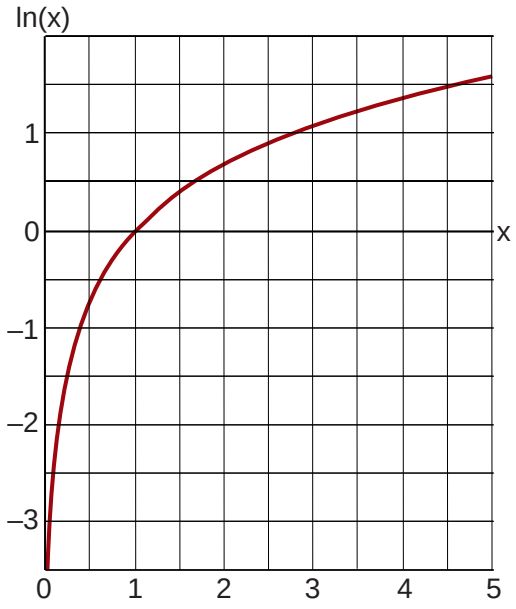



**Figure 13**
***The function ln(x) starts out at minus infinity at x = 0, goes through zero at x = 1, and slowly goes to plus infinity at x = infinity.***

At large distances, there is no problem with the formula for the voltage of a *point charge*. At $r = \infty$, the voltage V = kQ/r goes to zero, which is what we wanted for the potential energy of a test charge infinitely far away. But for a *line charge*, Equation (94) gives

$$V(r)\Big|_{r=+\infty} = -k\lambda\ln(+\infty) \quad (102)$$

This predicts a voltage or potential energy of ***minus infinity*** when we are infinitely far away from a line charge! How did this happen?

Either the mathematics is wrong, or our physical interpretation is wrong. The answer lies with the physical interpretation. What is wrong is that ***you cannot get infinitely far away from a line charge***. Any real physical piece of wire must have a finite length. The wire may look infinitely long when you are close to it, but as you move away, you will eventually be able to see both ends. The farther away you move, the shorter the wire looks. Move infinitely far from the wire and the wire looks like a point charge and the voltage it produces goes to zero. Thus physically we will not encounter the infinity that appears at large distances in the formula for the voltage of a line charge.

As we have often mentioned, in any formula for potential energy, we can arbitrarily choose the zero of potential energy (the floor) wherever we want. For point charges, we usually choose the zero of potential energy out at $r = \infty$. We have seen that we cannot make the same choice for a line charge. What we have to do is write the formula for the potential energy in the more general form

$$V(r) = -k\lambda\ln(r) + \text{constant} \quad (98a)$$

and adjust the constant so that V(r) is zero at some convenient place. We can see how this works in the following discussion of a coaxial cable.

## The Coaxial Cable

A physical example where our voltage formula (98a) makes sense is the coaxial cable. Suppose we have a cable whose inner conductor has a radius $r_i$ and the outer shield has an inside radius $r_0$ as shown in Figure (14). Assume that the inner conductor has a charge density $\lambda$ coulombs per meter, and the outer conductor is grounded (i.e., we say that the voltage V(r) is zero at $r = r_0$.) What is the voltage throughout the cable?

First of all, we know that the voltage inside a conductor must be constant so that the field $\vec{E} = -\vec{\nabla}V$ inside is zero. Since the outer conductor is grounded, the voltage throughout the shield (for $r > r_0$) will be zero as shown in Figure (15). The voltage on the inner conductor will have some constant value $V_i$ (for $r < r_0$).

Between the conductors, in the region between $r_i$ and $r_0$, the voltage must have the logarithmic dependence given by Equation (98a)

$$V(r) = -k\lambda \ln r + \text{constant} \qquad (103)$$

We can evaluate the constant by setting the voltage equal to zero out at the grounded shield, at $r = r_0$. This gives

$$V(r_0) = -k\lambda \ln r_0 + \text{constant} = 0$$

$$\text{constant} = k\lambda \ln r_0 \qquad (104)$$

and V(r) becomes

$$V(r) = -k\lambda \ln r + k\lambda \ln r_0 \qquad (105)$$

Logarithms have the peculiar property

$$\ln a - \ln b = \ln\left(\frac{a}{b}\right) \qquad (106)$$

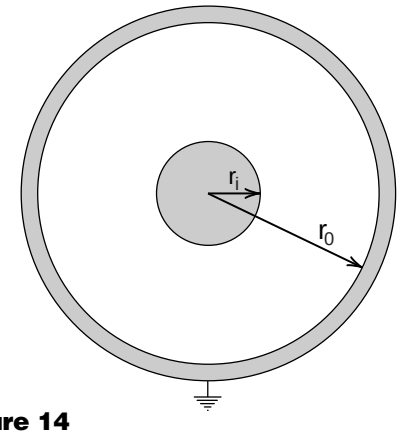



**Figure 14**
***A coaxial cable, where the inner wire has a radius $r_i$ and the outer grounded shield an inner radius $r_0$.***

Thus V(r) in Equation (99) can be more compactly written

$$V(r) = k\lambda \ln\left(\frac{r_0}{r}\right) \qquad (107)$$

With the constant k written out as $1/2\pi\varepsilon_0$ (see Equation 93), we get

$$\boxed{V(r) = \frac{\lambda}{2\pi\varepsilon_0}\ln\left(\frac{r_0}{r}\right)} \qquad (108)$$

At the outer shield, at $r = r_0$, we have

$$\ln(r_0/r) = \ln(1) = 0$$

and the voltage goes to zero. This is what we wanted for a grounded shield.

As demonstrated in Exercise 4 below, Equation (108) allows us to calculate the charge density $\lambda$ on the inner conductor of a coaxial cable when the outer conductor is grounded and the inner conductor is raised to some voltage $V_i$.

---

**Exercise 4**

(a) For the coaxial cable of Figure (14), find the formula for the charge density $\lambda$ when the inner conductor is at a voltage $V_i$ volts.

(b) Suppose $V_i = 100$ volts, $r_i = .5$ mm, $r_0 = 2$ mm and recall that $\varepsilon_0 \approx 9 \times 10^{-12}$. Then what is $\lambda$ in coulombs per meter?

(c) What is the general formula for the capacitance per meter of the coaxial cable in Figure (14)?

---

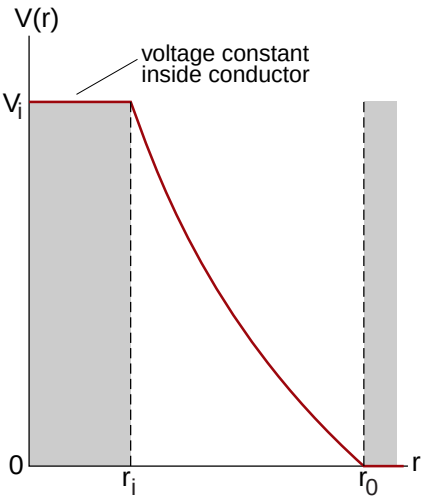



**Figure 15**
***Voltage in the coaxial cable.***

# View 2

# The Gradient from a Geometrical Perspective

*In the first part of this chapter, we used the relationship between force and potential energy to define what we meant by the gradient vector. We then used that relationship to derive the formulas for the gradient in cylindrical and spherical coordinates.*

*What we want to do now is to approach the gradient from a geometrical point of view. This is the point of view we began to develop when we constructed the physical models of electric voltage like the one shown in Figure (25-15) reproduced again here. Once we have developed a geometrical definition of the gradient we will check that the gradient behaves like a vector. To do that, we show that the components of the gradient change or transform the same way that the components of a displacement vector when we rotate the coordinate system. This idea of testing the vector nature of a new quantity will become particularly important when we get to a mathematically advanced discussion of special relativity.*

*This discussion of the gradient is designed to be independent of the first part of the chapter, so that you can start from either approach. This leads to some repetition of definitions, but the points of view are sufficiently different that some duplication should not be a problem. We, of course, end up with the same definition of the gradient vector from the two points of view.*

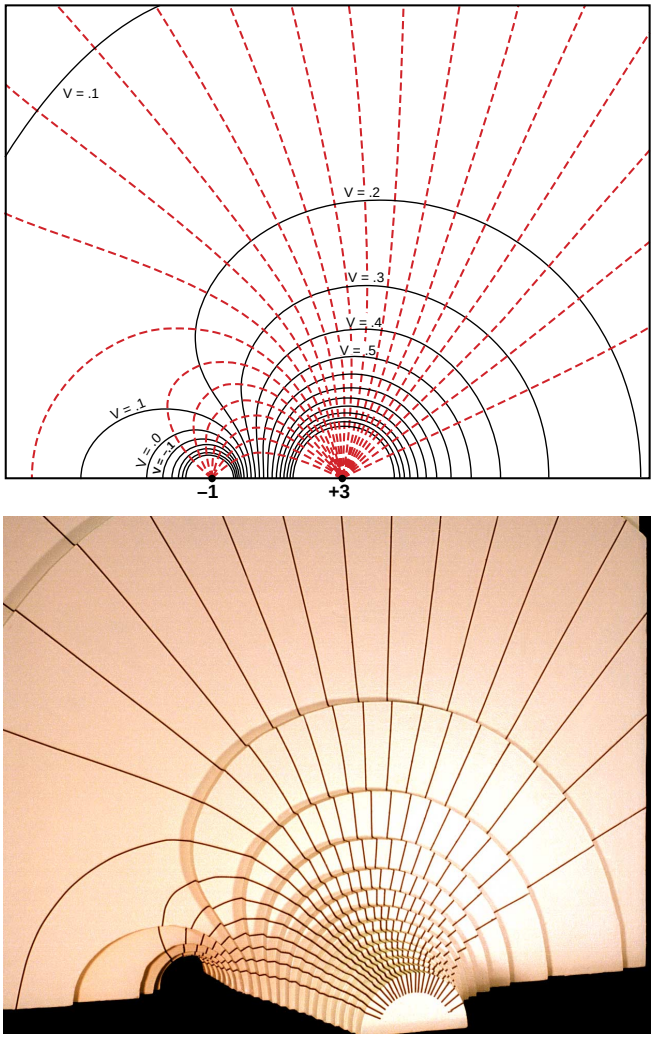



**Figure 25-15 (repeated)**
***Computer plot of the field lines and equipotentials for a charge distribution consisting of a positive charge + 3 and a negative charge – 1. These lines were then used to construct the plywood model.***

## SLOPE IN TWO DIMENSIONS

Imagine that you are planning a trip in a desert with hills and valleys. One possibility is to follow a path that heads due east through the desert. If you draw the path on a contour map, and note where the path crosses different contours, you can create a plot of the height (h) of the path as a function of the distance (x) of the path. The result might look like a plot of h(x) shown in Figure (1). This should at least represent a smoothed version of the terrain you will encounter.

Your curve h(x) tells you roughly how steep the path should be at any point $x_0$ . Mathematically, you can define the steepness as the slope of the tangent line at the point $x_0$ , which is equal to the first derivative of h(x).

$$\text{slope at } x_0 = h'(x_0)$$

$$= \left.\frac{dh(x)}{dx}\right|_{x = x_0}$$

$$= \tan\theta \qquad (1)$$

As long as you stay on the path, the slope at any point is uniquely determined by Equation (1).

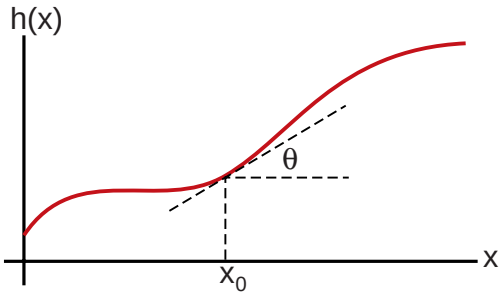



**Figure 1**
***Imagine that you are walking due east (x direction) in the desert. We will call h(x) the height of your path. At some point $x_0$, the slope of your path is dh(x)/dx evaluated at $x_0$, which is the tangent of the angle θ.***

However the interesting part about going out in the desert is that you do not have to follow any particular path. If you do not want to climb very much, you can walk along a contour line. If you are anxious to get to the top of a hill and want the steepest climb possible, you walk at right angles to a contour line, along what we have called a ***field line***, or what ski instructors call the ***fall line***. At any point you can choose a path whose slope ranges from zero along a contour line to the maximum along the field line. To define the slope at some point, you have to state the direction you are traveling.

To handle this new feature mathematically, we first introduce a coordinate system (x,y), where the x direction, for example, could be east-west and the y direction north-south. The terrain is then described by a function h(x,y) giving the height of the land at any point (x,y).

To describe the slope of a one dimensional curve h(x) at some point $x_0$ , we drew a ***tangent line*** at $x_0$ as shown in Figure (1). To describe slopes for a two dimensional function h(x,y), at some point $(x_0, y_0)$ we look at the ***tangent plane*** at that point. This assumes that the function h(x,y) is smooth enough that, as we get closer and closer to the point $(x_0, y_0)$ the landscape looks smoother and smoother. It assumes that when we get very close, the landscape looks flat and we are looking at the tangent plane.

Not all functions h(x,y) are necessarily that smooth. Curves describing real landscapes, like the shape of a coastline, look just as rough no matter how close we look. Such curves are described by what is called ***fractal geometry***. What we will be discussing are curves, or surfaces that become smooth when we look close enough. A sufficient mathematical criteria for such smoothness is that all derivatives with respect to any variable are finite.

If the terrain h(x,y) is smooth enough to have a unique tangent plane at every point, then our discussion of the nature of slopes on a curved surface can begin with a study of how slopes behave in a tangent plane. What we learn from the study of one tangent plane can then be applied to all tangent planes in the terrain.

To visualize a tangent plane at some point $(x_0, y_0)$, start by imagining that the point is on the surface of a table, and construct a coordinate axis (x,y,z) whose origin is at $(x_0, y_0)$ as shown in Figure (2). The xy plane is the table surface and the z axis points straight up. Let us assume that the x axis faces east and the y axis north.

To represent a tangent plane, take a thin flat object like a piece of cardboard, and place it on the table surface, tilted at an angle $\theta$ as shown in Figure (3). Orient the cardboard so that the line of contact with the table is the x axis.

It is easy to see that in our flat tilted surface, all lines parallel to the x axis are contour lines, and that all lines parallel to the y axis headed north are field lines with a maximum slope. It is also clear that the field lines are perpendicular to the contour lines.

These features carry over to a smooth curved surface h(x,y). At any point $(x_0, y_0)$ construct a tangent plane. Unless this tangent plane happens to be horizontal, there will be a unique horizontal line in the plane that passes through the point $(x_0, y_0)$. This horizontal line corresponds to the x axis in Figure (3). In a region very close to the point $(x_0, y_0)$ this horizontal line will coincide with the contour line of h(x,y) that passes through that point.

Perpendicular to the x axis in the tangent plane will be a line of maximum slope heading in the y direction of Figure (3). The field line of our curved surface h(x,y) that passes through the point $(x_0, y_0)$ will be y oriented for a small region around $(x_0, y_0)$. As a result, in this small region the contour lines and the field lines of the curved surface have the same properties as the contour and field lines in the tangent plane. In particular, even for curved surfaces, contour lines and field lines will always be perpendicular to each other where the contour lines are in the direction of zero slope and the field lines in the direction of maximum slope.

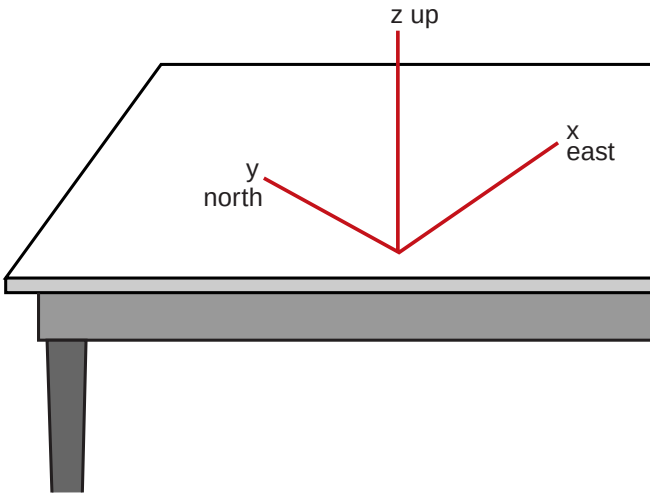



**Figure 2**
***Our coordinate system.***

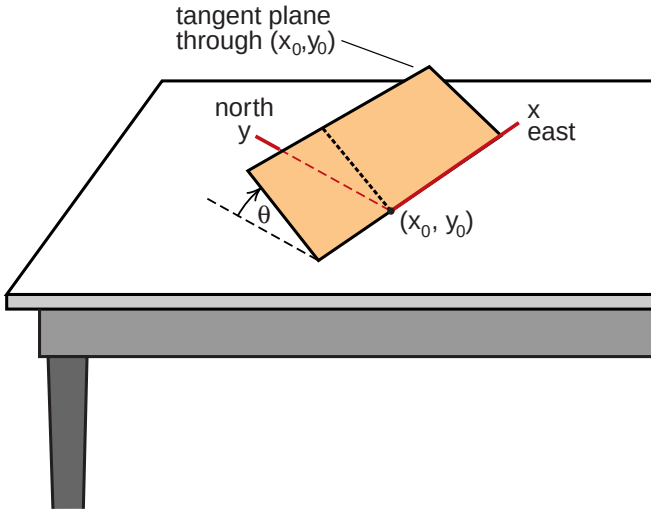



**Figure 3**
***The tangent plane. All lines in the tangent plane that are parallel to the x axis are lines of equal height, or contour lines. Lines in the perpendicular y direction are lines of maximum slope, or field lines.***

## THE GRADIENT

When you have a mathematical function h(x,y) that describes a surface, the slope of that surface in some direction is given by the partial derivative in that direction. Explicitly the slope in the x direction at the point $(x_0, y_0)$ is given by

$$\left.\begin{array}{l}\text{slope in x}\\ \text{direction at}\\ (x_0, y_0)\end{array}\right\} = \left.\frac{\partial h(x,y)}{\partial x}\right|_{x = x_0,\, y = y_0} \qquad (2a)$$

and the slope in the y direction is

$$\left.\begin{array}{l}\text{slope in y}\\ \text{direction at}\\ (x_0, y_0)\end{array}\right\} = \left.\frac{\partial h(x,y)}{\partial y}\right|_{x = x_0,\, y = y_0} \qquad (2b)$$

What we will do now is to define a quantity we will call the ***gradient***, and represent it by the symbol $\vec{\nabla}h(x,y)$. Explicitly we define $\vec{\nabla}h(x,y)$ by the equation

$$\vec{\nabla}h(x,y) = \hat{x}\frac{\partial h(x,y)}{\partial x} + \hat{y}\frac{\partial h(x,y)}{\partial y} \qquad (3)$$

where $\hat{x}$ and $\hat{y}$ are unit vectors in the x and y directions respectively.

The gradient $\vec{\nabla}h(x,y)$ looks like a vector with x and y components equal to the slope of h(x,y) in the x and y directions. However a vector is more than a quantity with some components. We saw in Chapter 2 of the Physics text that a vector has a basic physical significance that does not depend upon the coordinate system used to define the vector. What we need to do for our gradient is to find the basic significance of the quantity $\vec{\nabla}h(x,y)$ and then show that the physical picture does not change when the gradient is evaluated in a different coordinate system.

To see the physical significance of the gradient, we will evaluate $\vec{\nabla}h(x,y)$ at some point $(x_0, y_0)$, using a coordinate system where the x axis is parallel to the contour line passing through that point. That is the same coordinate system we used in our discussion of the tangent plane in Figure (3). Since the x axis lies along a contour line at the point of interest, there is no change in height as we move a short distance in the x direction, and thus the partial derivative in the x direction is zero.

$$\left.\frac{\partial h(x,y)}{\partial x}\right|_{\substack{x = x_0\\ y = y_0}} = 0 \qquad \begin{array}{l}\text{for an x axis}\\ \text{lying along}\\ \text{a contour line}\end{array} \qquad (4)$$

What remains of the gradient is

$$\left.\vec{\nabla}h(x,y)\right|_{\substack{x = x_0\\ y = y_0}} = \hat{y}\left.\frac{\partial h(x,y)}{\partial y}\right|_{\substack{x = x_0\\ y = y_0}} \qquad \begin{array}{l}\text{for an x axis}\\ \text{lying along}\\ \text{a contour line}\end{array} \qquad (5)$$

For this coordinate system, the gradient is purely y oriented, which is the direction of the field line through $(x_0, y_0)$. Also the magnitude of the gradient is equal to the magnitude of the steepest slope at $(x_0, y_0)$. As a result, physical significance of the gradient, at least in this special coordinate system, is that it ***describes both the direction and magnitude of the steepest slope***.

Thus the gradient has both a magnitude and a direction like the displacement vectors we discussed in Chapter 2 of the Physics text. ***If the components of the gradient change (transform) in the same way as the components of a displacement vector, then the magnitude and direction will be preserved when we go to a new (rotated) coordinate system.*** The components will look different, but the magnitude and direction will be unchanged.

To see whether the components of the gradient transform (change) like the components of a displacement vector, let us first review what happens to a purely y oriented displacement vector $\vec{B}$ when we go to a new coordinate system $(x',y')$ that is rotated by an angle $\phi$ about the z axis as shown in Figure (4). You can easily see that in the $x',y'$ coordinate system, the components of $\vec{B}$ are

$$\begin{aligned} B_{x'} &= B \sin\phi \\ B_{y'} &= B \cos\phi \end{aligned} \qquad (6)$$

---

**Exercise 1**

(a) Show that for a purely x oriented vector $\vec{A}$ the components of $\vec{A}$ in the rotated $(x', y')$ coordinate system are

$$\begin{aligned} A_{x'} &= A \cos\phi \\ A_{y'} &= -A \sin\phi \end{aligned} \qquad (7)$$

(b) Now show that if you start with a vector

$$\vec{C} = \hat{x}A + \hat{y}B \equiv \hat{x}C_x + \hat{y}C_y$$

which has components $C_x = A$ in the x direction and $C_y = B$ in the y direction, then in the rotated coordinate system, the components of $\vec{C}$ are

$$\boxed{\begin{aligned} C_{x'} &= +C_x \cos\phi + C_y \sin\phi \\ C_{y'} &= -C_x \sin\phi + C_y \cos\phi \end{aligned}} \qquad (8)$$

(Equations (8) are the general formula for the transformation of the x and y components of a vector when we rotate the coordinate system by an angle $\phi$ about the z axis.)

---

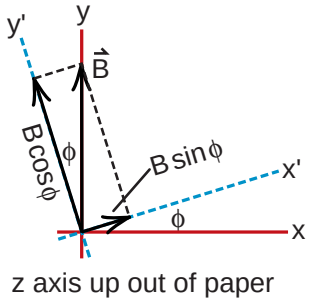



**Figure 4**
***When we rotate the coordinate system about the z axis, the y directed vector $\vec{B}$ gets components in both the $x'$ and $y'$ directions.***

When we go from the coordinate system (x,y) to the rotated coordinate system $(x',y')$, the gradient

$$\vec{\nabla}h(x,y) = \hat{x}\frac{\partial h(x,y)}{\partial x} + \hat{y}\frac{\partial h(x,y)}{\partial y} \qquad (3) \text{ repeat}$$

becomes

$$\vec{\nabla}h(x',y') = \hat{x}'\frac{\partial h(x',y')}{\partial x'} + \hat{y}'\frac{\partial h(x',y')}{\partial y'} \qquad (9)$$

To calculate the new components $\partial h(x',y')/\partial x'$ and $\partial h(x',y')/\partial y'$ at some arbitrary point (x,y) we will use our familiar tangent plane of Figure (3) reproduced here as Figure (5). We have also drawn in the rotated coordinate system $(x',y')$ seen in the top view of Figure (5). The coordinate axes x, y and $x',y'$ all lie in the table top surface, what we can call the z = 0 plane.

The partial derivative, for example $\partial h(x,y)/\partial y$ , represents the rate of change of the height h as we go out along the y axis. For the rotated coordinate system, the partial derivative $\partial h(x',y')/\partial x'$ represents the rate of change of the height h as we go out along the $x'$ axis. We will use these ideas to calculate the height $\Delta h$ of the point A shown in Figure (5), a point that is a distance $\Delta x'$ down the $x'$ axis.

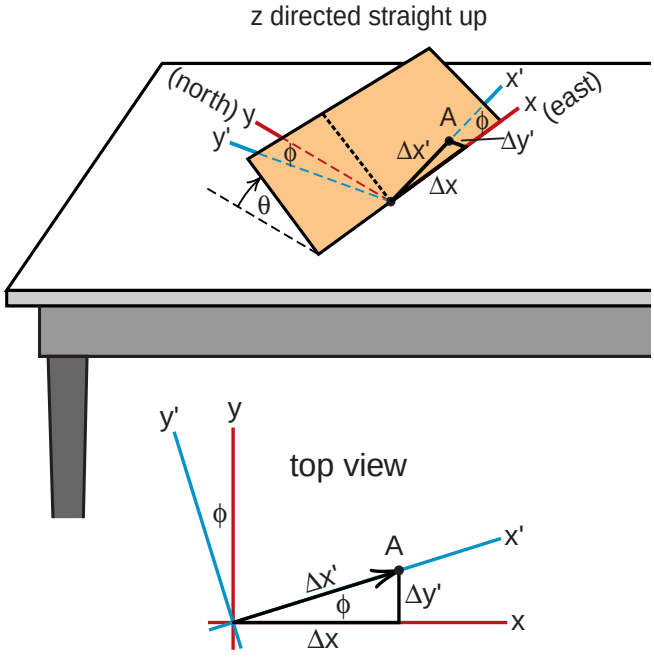



**Figure 5**
***Our tangent plane of Figure (3) showing the rotated coordinate system x',y', and the point A, a distance $\Delta x'$ down the x' axis.***

There are two distinct ways to get to the point A. One is to go down the x′ axis directly, a distance $\Delta x'$ . For this route we get as the formula for $\Delta h$

$$\Delta h = \begin{pmatrix} \text{slope in} \\ \text{the } x' \\ \text{direction} \end{pmatrix} \times \begin{pmatrix} \text{distance we} \\ \text{go in the} \\ x' \text{ direction} \end{pmatrix}$$

$$\Delta h = \frac{\partial h(x',y')}{\partial x'} \times \Delta x' \qquad (10)$$

The other way to get to point A is to go down the x axis a distance $\Delta x$ , gaining a height $\Delta h_x$ given by

$$\Delta h_x = \frac{\partial h(x,y)}{\partial x} \Delta x \qquad (11)$$

and then go out a distance $\Delta y$ in the old y direction, giving us an additional height $\Delta h_y$ given by

$$\Delta h_y = \frac{\partial h(x,y)}{\partial y} \Delta y \qquad (12)$$

The height $\Delta h$ at point A will be the sum of these two heights

$$\Delta h = \Delta h_x + \Delta h_y$$

$$= \frac{\partial h(x,y)}{\partial x}\Delta x + \frac{\partial h(x,y)}{\partial y}\Delta y \qquad (13)$$

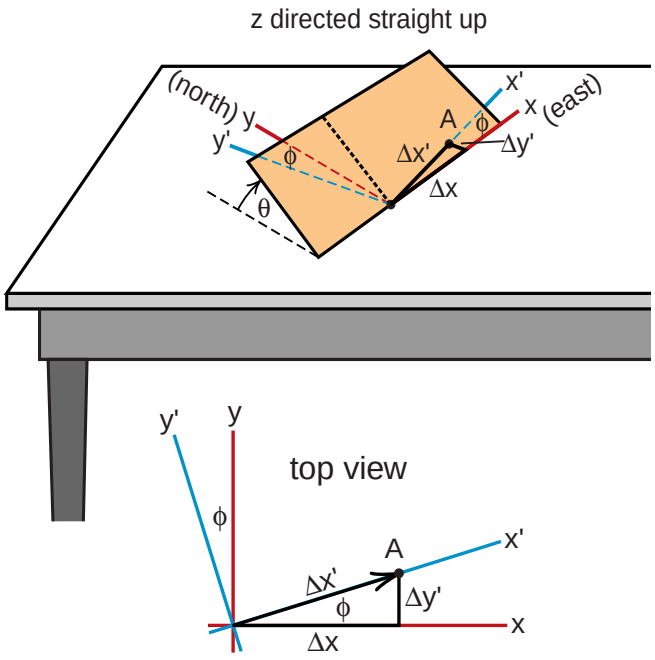



**Figure 5 repeated**
***Our tangent plane of Figure (3) showing the rotated coordinate system x', y', and the point A, a distance $\Delta x$ down the x' axis.***

(In our drawing of Figure (5), we have shown the x axis as being horizontal, so that the slope $\partial h(x,y)/\partial x$ would be zero. This makes the drawing easier to interpret, but we do not need to assume the x slope is zero for the current discussion.)

The final step in calculating the height $\Delta h$ of point A from the second route is to relate $\Delta x$ and $\Delta y$ to the distance $\Delta x'$ traveled along the x′ axis. From the top view of Figure (5) it is clear that

$$\Delta x = \Delta x' \cos\phi$$

$$\Delta y = \Delta x' \sin\phi \qquad (14)$$

Using these values in Equation (13) give us

$$\Delta h = \frac{\partial h(x,y)}{\partial x}\Delta x' \cos\phi + \frac{\partial h(x,y)}{\partial y}\Delta x' \sin\phi \qquad (15)$$

We can now equate our two formulas, Equation (10) and Equation (15) for the height $\Delta h$ at point A. The factors of $\Delta x'$ cancel and we are left with

$$\frac{\partial h(x',y')}{\partial x'} = \frac{\partial h(x,y)}{\partial x}(\cos\phi) + \frac{\partial h(x,y)}{\partial y}(\sin\phi) \qquad (16)$$

Comparing Equation (15) with Equation (8) for the transformation of the x component of the displacement vector $\vec{C}$

$$C_{x'} = C_x \cos\phi + C_y \sin\phi \qquad \text{(8a) repeated}$$

we see that the x component of the gradient transforms (changes) in the same way as a displacement vector when we rotate the coordinate system by an angle $\phi$ .

---

**Exercise 2**

Using similar arguments, show that the y′ slope $\partial h(x',y')/\partial y'$ is given by

$$\frac{\partial h(x',y')}{\partial y'} = \frac{\partial h(x,y)}{\partial x}(-\sin\phi) + \frac{\partial h(x,y)}{\partial y}(\cos\phi) \qquad (17)$$

which is the same as the transformation of the y component of a displacement vector.

---

## Gradient as a Vector Field

What is the significance of our demonstration that the quantity $\vec{\nabla}h(x,y)$, defined by

$$\vec{\nabla}h(x,y) = \hat{x}\frac{\partial h(x,y)}{\partial x} + \hat{y}\frac{\partial h(x,y)}{\partial y} \quad (3) \text{ Repeat}$$

transforms like a vector at each point (x,y) in space? As we pointed out in Chapter 29 of the Physics text, a vector field, which is a vector at every point in space, is uniquely determined if we have general formulas for the surface integral and the line integral of the field. There were four Maxwell's equations because we needed formulas for the surface and the line integrals of both the electric and magnetic fields.

In the Physics text and the first part of this chapter, we knew that the electric field was a vector field because of its definition as the force vector acting on a unit test charge. The knowledge that forces transform as vectors was sufficient to tell us that any correct formula for $\vec{E}$ gave us a vector field. In this section with the definition of Equation (3), the gradient is given a geometrical definition, which at first sight might or might not make $\vec{\nabla}h(x,y)$ behave as a vector field. The demonstration that $\vec{\nabla}h(x,y)$ transforms as a vector means that concepts like line and surface integrals can be applied to any gradient fields.

As we saw in the first part of this chapter, the extension of Equation (3) to the gradient of a three dimensional function is

$$\vec{\nabla}h(x,y,z) = \hat{x}\nabla_x h + \hat{y}\nabla_y h + \hat{z}\nabla_z h \qquad (18)$$

where $\nabla_x$, $\nabla_y$ and $\nabla_z$ are the partial derivatives $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$. Equation (18) here is equivalent to Equation (16) in the first part of the chapter relating $\vec{E}$ to $\vec{\nabla}V(x,y,z)$.

This completes our discussion of the gradient vector $\vec{\nabla}h(x,y,z)$ from a geometrical point of view. If you have not done so already, now is the time to look at applications of the gradient vector to electric field problems, starting with the discussion of the gradient vector just before Equation (16) of the first part of the chapter.

# View 3
# Pressure Force as a Gradient

*We end the chapter with View 3, an application to fluids, where we see that the pressure force $\vec{f}_p$ acting on the fluid particles is the gradient of the pressure field p. This represents a straightforward example of obtaining a vector field $\vec{f}_p$ from a scalar field p.*

## PRESSURE FORCE AS A GRADIENT

In the Physics text, there were two main places where we dealt with the concept of pressure. The first was in Chapters 17 and 18 on the ideal gas law, and the second was in Chapter 23 during our discussion of Bernoulli's equation. In both cases we mentioned that pressure had the dimensions of a force per unit area, but was itself a scalar field p(x,y, z) that did not point anywhere. We pointed out that the pressure force acting on an area $\Delta A$ was directed perpendicular to the area and had a magnitude

$$\Delta F = p\Delta A \qquad (1)$$

We will now use the concept of a gradient to show that the pressure force per unit volume $\vec{f}_p$, acting on the fluid particles, is equal to minus the gradient of the pressure p(x,y,z)

$$\vec{f}_p = -\vec{\nabla}p(x,y,z) \qquad (2)$$

This is analogous to the electric field being equal to minus the gradient of the electric voltage

$$\vec{E} = -\vec{\nabla}V(x,y,z) \qquad (3\text{-}19)$$

which we saw back in Equation (3-19) of this chapter.

To calculate the pressure force, we start with a small volume $\Delta V = \Delta x \Delta y \Delta z$ shown in Figure (1). This volume element has a left face located at z and a right face at $z + \Delta z$. The center of the faces are located at (x, y) where the pressures are p(x,y,z) and $p(x,y,z+\Delta z)$ respectively.

The pressure force $\Delta\vec{F}_1$ exerted on the left face of $\Delta V$ is equal to the force per unit area $p_1(x,y,z)$ times the area $\Delta A_1 = \Delta x \Delta y$ of that face. The pressure force is directed into the volume, toward the right in the $\hat{z}$ direction, as shown

$$\Delta\vec{F}_1 = \hat{z}p(x,y,z)\Delta x\Delta y \qquad (3)$$

On the right side, the force is directed back into $\Delta V$, in the –z direction, and has a value

$$\Delta\vec{F}_2 = -\hat{z}p(x,y,z+\Delta z)\Delta x\Delta y \qquad (4)$$

The net force on these two sides is

$$\begin{aligned}\left(\Delta\vec{F}_1 + \Delta\vec{F}_2\right) &= -\hat{z}\left[p(x,y,z+\Delta z) - p(x,y,z)\right]\Delta x\Delta y \\ &= -\hat{z}\left[\frac{p(x,y,z+\Delta z) - p(x,y,z)}{\Delta z}\right]\Delta x\Delta y\Delta z \end{aligned} \qquad (5)$$

You can immediately see that when we take the limit that $\Delta V$ is an infinitesimal volume and $\Delta z$ goes to zero, the quantity in the square brackets in Equation (5) becomes the partial derivative of p(x,y,z) with respect to z.

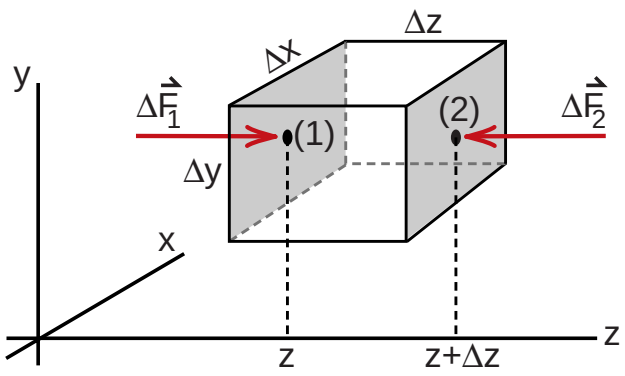



**Figure 1**
***The volume element Δx Δy Δz.***

$$\lim_{\Delta z \to 0}\left[\frac{p(x,y,z+\Delta z) - p(x,y,z)}{\Delta z}\right] = \frac{\partial p(x,y,z)}{\partial z} \qquad (6)$$

Thus Equation (5) can be written in the somewhat mixed form

$$\left(\Delta\vec{F}_1 + \Delta\vec{F}_2\right) = -\hat{z}\frac{\partial p(x,y,z)}{\partial z}\Delta x\Delta y\Delta z \qquad (7)$$

where we will shortly think in terms of the limit that $\Delta V = \Delta x \Delta y \Delta z$ goes to zero.

Before we do, let us add in the pressure forces $\vec{F}_3$ and $\vec{F}_4$ acting on the bottom and top faces respectively, and $\vec{F}_5$ and $\vec{F}_6$ acting on the back and front faces to get the total pressure force $\Delta\vec{F}_p$ acting on $\Delta V$. Following the same steps used to derive Equation (7), we get

$$\begin{aligned}\Delta\vec{F}_p &= \left(\Delta\vec{F}_1 + \Delta\vec{F}_2 + \Delta\vec{F}_3 + \Delta\vec{F}_4 + \Delta\vec{F}_5 + \Delta\vec{F}_6\right) \\ &= -\left[\hat{z}\frac{\partial p(x,y,z)}{\partial z} + \hat{y}\frac{\partial p(x,y,z)}{\partial y} + \hat{x}\frac{\partial p(x,y,z)}{\partial x}\right]\Delta x\Delta y\Delta z \end{aligned} \qquad (8)$$

The quantity in the square brackets in Equation (8) is the gradient $\vec{\nabla}p$ of the pressure field. Thus we have, after dividing both sides by $\Delta V = \Delta x \Delta y \Delta z$

$$\frac{\Delta\vec{F}_p}{\Delta V} = -\vec{\nabla}p(x,y,z) \qquad (9)$$

We recognize the left side of Equation (9) as the total pressure force acting on $\Delta V$ divided by the volume $\Delta V$. It is therefore the ***pressure force per unit volume*** $\vec{f}_p(x,y,z)$ acting in that region of the fluid, and we get our advertized result

$$\boxed{\vec{f}_p(x,y,z) = -\vec{\nabla}p(x,y,z)}$$

*pressure force per unit volume*

(2) repeated

With Equation (2) we have a powerful way of calculating pressure forces, since we can evaluate the gradient in any of the coordinate systems we have been discussing, such as cylindrical or spherical polar coordinates.

# Calculus 2000-Chapter 4

# The Operator $\nabla^2 \equiv \vec{\nabla} \cdot \vec{\nabla}$ (The Laplacian)

*In our earliest discussion of vectors in Chapter 2 of the Physics text, we were introduced to the vector dot product*

$$\vec{A} \cdot \vec{B} \equiv A_x B_x + A_y B_y + A_z B_z \qquad (1)$$

*as having the special property of being a scalar quantity. That is, the quantity $\vec{A} \cdot \vec{B}$ had the same value no matter what coordinate system we used to evaluate it. Having just seen that the gradient operator $\vec{\nabla}$ operating on a scalar field h(x,y,z) produces a vector field, one might wonder what we get when we take the dot product of two gradient operations acting on a scalar field. The answer is that we get another scalar field.*

*The standard name for this dot product of two gradient operators is* ***del squared****,* written as

$$\nabla^2 \equiv \vec{\nabla} \cdot \vec{\nabla} \qquad (2)$$

*It is often called the* **Laplacian** *operator after the French mathematician Laplace. This operator is essentially an extension to three dimensions of the second derivative we encountered in Calculus Chapter 2, during our discussion of the one dimensional wave equation. Thus we should expect $\nabla^2$ to appear when we begin to discuss three dimensional wave equations in the next few chapters.*

### *Fluid theory*

*Another area of physics where the operator $\nabla^2$ plays a prominent role is in fluid dynamics. For common fluids like water and air, the viscous force acting on the fluid particles turns out to be proportional to the Laplacian of the velocity field, namely $\nabla^2 \vec{v}$. We will derive that result starting from an assumption that Issac Newton made about the nature of viscous forces.*

*As an application of the theory of viscous forces, we will look at the steady flow of a viscous fluid in a pipe. This example provides a way to measure the so called* **coefficient of viscosity** *that appears in Newton's theory. It also provides an example of the use of the operator $\nabla^2$ acting on a vector field.*

## Schrödinger's Equation

*One of the glaring omissions in the Physics text resulted from our inability to calculate the electron wave patterns in the hydrogen atom. All we were able to do is show drawings of a few of the lowest energy wave patterns, describe the electron's energy and angular momentum in these wave patterns, and then state that these patterns came from a wave equation called Schrödinger's equation. We were neither able to write down or solve the equation itself.*

*To handle Schrödinger's equation as applied to the hydrogen atom, we needed two mathematical concepts we did not then have. One is the operator* $\nabla^2$ *which we are introducing in this chapter, the other is the concept of a* **complex variable** *which we will introduce in the next chapter, Chapter 5. Once we develop these two mathematical tools, we will be ready to approach Schrödinger's equation in Chapter 6.*

*When we apply Schrödinger's equation to the hydrogen atom, we are dealing with a system that has spherical symmetry. As a result it is much easier to deal with the theory using a coordinate system that has the same symmetry. The problem is that the operator* $\nabla^2$*, which in Cartesian coordinates is a straightforward extension of the second derivative, becomes quite complex when we work in other coordinate systems like spherical polar coordinates. The reason for the complexity is that in any coordinate system except Cartesian coordinates, the unit vectors may change direction as we move from one point in space to another. This change in the direction of the unit vectors complicates the formulas for* $\nabla^2$*.*

## The Formulary

*In the main part of this chapter we will simply state the formula, in spherical polar coordinates, for* $\nabla^2$ *acting on a scalar field* $\psi$*. This is the formula we will use in Chapter 6 in our discussion of the hydrogen atom. In the appendix, however, we will derive the formula, showing you exactly how the changing unit vectors affect the results. We have placed this derivation in an appendix because it is the kind of derivation you probably want to observe only once in your life, to find out where the rather messy results come from.*

*When you are actually working problems involving quantities like* $\nabla^2$ *in cylindrical or spherical coordinates, you do not want to derive the formulas yourself because the chances of your getting the right answer are too small. You are not likely to memorize them correctly either, unless you use a particular formula often. Instead, the best procedure is to look up the result in a table of formulas, sometimes called a* **formulary**. *We provide a formulary at the end of this text, one adapted from a formulary developed by David Book of the Naval Research Laboratory.*

*In our discussion of viscous forces in this chapter, we use the formulary to find the formula in cylindrical coordinates for* $\nabla^2$ *acting on the vector field* $\vec{v}$*.*

## $\nabla^2$ IN CARTESIAN COORDINATES

We will first take a careful look at $\nabla^2 = \vec{\nabla}\cdot\vec{\nabla}$ in Cartesian coordinates before we approach the spherical case. Using the unit vector notation for $\vec{\nabla}$ we have

$$\vec{\nabla} = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z} \tag{3}$$

where $\hat{x}$, $\hat{y}$ and $\hat{z}$ are unit vectors pointing in the x, y, and z directions respectively. The dot product $\vec{\nabla}\cdot\vec{\nabla}$ acting on some function f(x,y,z) should be given by

$$\begin{aligned}
&\vec{\nabla}\cdot\vec{\nabla}f(x,y,z)\\
&= \left(\hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}\right)\cdot\left(\hat{x}\frac{\partial f}{\partial x} + \hat{y}\frac{\partial f}{\partial y} + \hat{z}\frac{\partial f}{\partial z}\right)\\
&= \left(\hat{x}\frac{\partial}{\partial x}\right)\cdot\left(\hat{x}\frac{\partial f}{\partial x}\right) + \left(\hat{x}\frac{\partial}{\partial x}\right)\cdot\left(\hat{y}\frac{\partial f}{\partial y}\right) + \cdots\\
&\quad + \cdots + \left(\hat{z}\frac{\partial}{\partial z}\right)\cdot\left(\hat{z}\frac{\partial f}{\partial z}\right)
\end{aligned} \tag{4}$$

Being very careful with our differentiation, we have, for example,

$$\left(\hat{x}\frac{\partial}{\partial x}\right)\cdot\left(\hat{x}\frac{\partial f}{\partial x}\right) = \hat{x}\cdot\left[\frac{\partial\hat{x}}{\partial x}\frac{\partial f}{\partial x} + \hat{x}\frac{\partial^2 f}{\partial x^2}\right] \tag{5}$$

We have been overly careful because the unit vectors $\hat{x}$, $\hat{y}$ and $\hat{z}$ are constant in both magnitude and direction, thus

$$\frac{\partial\hat{x}}{\partial x} = 0 \tag{6}$$

and we are left with

$$\left(\hat{x}\frac{\partial}{\partial x}\right)\cdot\left(\hat{x}\frac{\partial f}{\partial x}\right) = \hat{x}\cdot\hat{x}\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial x^2} \tag{7}$$

Similarly

$$\begin{aligned}
\left(\hat{x}\frac{\partial}{\partial x}\right)\cdot\left(\hat{y}\frac{\partial f}{\partial y}\right) &= \hat{x}\cdot\left(\frac{\partial\hat{y}}{\partial x}\frac{\partial f}{\partial y} + \hat{y}\frac{\partial^2 f}{\partial x\partial y}\right)\\
&= \hat{x}\cdot\hat{y}\frac{\partial^2 f}{\partial x\partial y} = 0
\end{aligned} \tag{8}$$

because $\partial\hat{y}/\partial x = 0$ and $\hat{x}\cdot\hat{y} = 0$.

As a result, all we are left with, when we evaluate $\nabla^2 f$ in Cartesian coordinates is

$$\boxed{\nabla^2 f(x,y,z) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}} \tag{9}$$

which is an obvious extension to three dimensions of the second derivative $\partial^2 f/\partial x^2$ that appeared in our one dimensional wave equation in Chapter 2 of the Calculus text.

## $\nabla^2$ in Spherical Polar Coordinates

As we mentioned, the results are not so simple when we are working in other coordinate systems. In spherical polar coordinates, when $\nabla^2$ is acting on a scalar function, we get the following result which is derived in the appendix to this chapter.

$$\boxed{\begin{aligned}
\nabla^2 f &= \frac{1}{r}\frac{\partial^2}{\partial r^2}(rf)\\
&\quad + \frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}\right]
\end{aligned}} \tag{10}$$

where r, $\theta$, and $\varphi$ are the polar coordinates shown in Figure (1). Much of this complexity comes from the fact that the unit vectors are not constant, and have to be differentiated. You will see how this works by going to the appendix.

(We should note that, in non Cartesian coordinates, $\nabla^2$ acting on a vector, e.g. $\nabla^2\vec{E}$, has an even more complex formula, which is given in the formulary at the end of the text.)

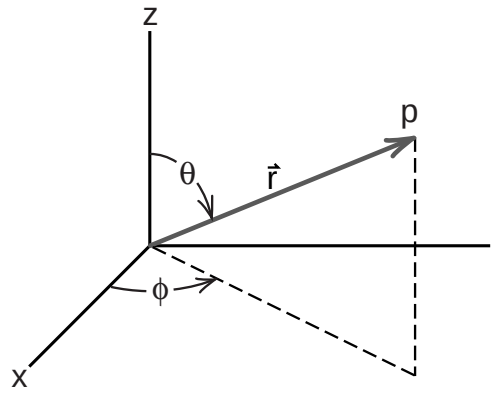



**Figure 1**
***Spherical polar coordinates.***

## NEWTONIAN FLUIDS

We now move on to our example of the use of the Laplacian operator to describe viscosity in a Newtonian fluid.

Newton proposed that viscous effects in a fluid resulted from the shearing motion of one layer of fluid over another. This shearing force can be introduced as follows.

Suppose we have a simple flow where all the fluid is moving in the x direction, and the velocity is increasing in the y direction as shown in Figure (2).

To analyze the forces involved, consider a horizontal plane indicated by the dashed line labeled by A----B. The fluid above the plane, which is travelling faster, drags the fluid below forward. The fluid below, which is going slower, drags the upper fluid back. Let $\vec{\tau}_+$ be the force per unit area exerted by the upper fluid on the lower fluid, and $\vec{\tau}_-$, the force exerted by the lower fluid on the upper. In Figure (2) we have drawn the forces $\vec{\tau}_+$ and $\vec{\tau}_-$ inside the fluids upon which they act.

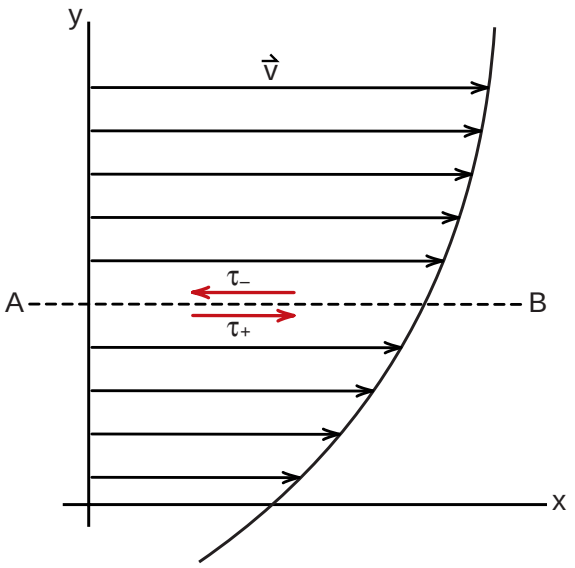



**Figure 2**
***Diagram of a simple flow where the velocity field $\vec{v}$ is x directed and increasing in the y direction.***

This combination of oppositely directed forces on opposite sides of the plane is called a ***stress***, in this case a stress generated by the action of viscosity. For a so called ***Newtonian fluid, the stress $\tau$ is assumed to be directly proportional to the rate at which the velocity field is changing as we move up***, which for our x directed flow is

$$\boxed{\tau = \mu \frac{\partial v_x(y)}{\partial y}} \qquad (11)$$

The quantity $\mu$ is called the ***coefficient of viscosity***

$$\boxed{\mu = \text{coefficient of viscosity}} \qquad (12)$$

For a Newtonian fluid, $\mu$ is assumed to be a constant throughout the fluid. In many situations, both water and air behave as Newtonian fluids.

## VISCOUS FORCE ON A FLUID ELEMENT

Suppose again that we have a simple x directed velocity field whose velocity profile is shown in Figure (3). Now consider a small volume element with sides $\Delta x$, $\Delta y$ and $\Delta z$, the bottom of which is located at (y) and the top at $(y + \Delta y)$ is shown. The fluid below the plane A----B at y is dragging the fluid above, back with a force per unit area $\tau_-(y)$

$$\tau_-(y) = -\mu\frac{\partial v_x(y)}{\partial y} \quad \begin{array}{l}\textit{force per unit}\\ \textit{area at the bottom}\\ \textit{of volume element}\end{array} \qquad (13)$$

The total force at the bottom is the force per unit area $\tau_-(y)$ times the area $\Delta x\Delta z$ upon which it is acting

$$\begin{aligned}\Delta F_-(y) &= \tau_-(y)\Delta x\Delta z \\ &= -\mu\frac{\partial v_x(y)}{\partial y}\Delta x\Delta z\end{aligned} \qquad (14)$$

Up at the top of the volume element, the faster fluid above the C----D plane at $(y + \Delta y)$, is pulling forward the slower fluid below with a total force

$$\begin{aligned}\Delta F_+(y+\Delta y) &= \tau_+(y+\Delta y)\Delta x\Delta z \\ &= +\mu\frac{\partial v_x(y+\Delta y)}{\partial y}\Delta x\Delta z\end{aligned} \qquad (15)$$

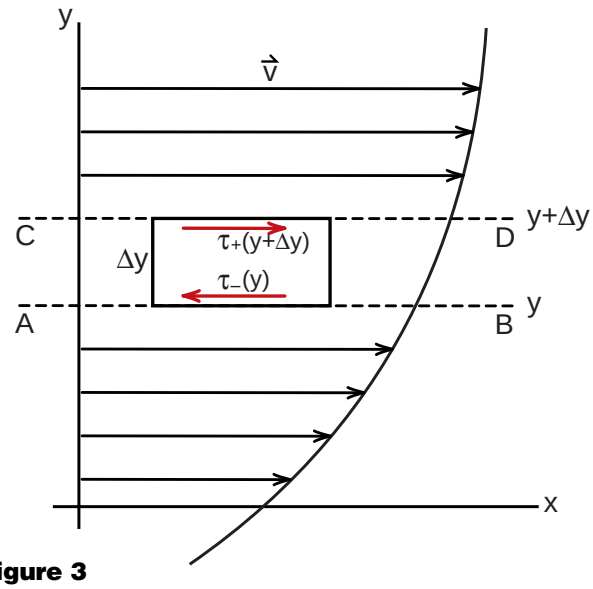



**Figure 3**
***Calculating the viscous force on a fluid element.***

With Equations (14) and (15) we see that the total viscous force on the fluid in our volume element can be written

$$\begin{aligned}\Delta F_x &= \Delta F_-(y) + \Delta F_+(y+\Delta y) \\ &= \left[-\mu\frac{\partial v_x(y)}{\partial y} + \mu\frac{\partial v_x(y+\Delta y)}{\partial y}\right]\Delta x\Delta z\end{aligned} \qquad (16)$$

Multiplying the right side by $\Delta y/\Delta y$ gives

$$\Delta F_x = \mu\left[\frac{\dfrac{\partial v_x(y+\Delta y)}{\partial y} - \dfrac{\partial v_x(y)}{\partial y}}{\Delta y}\right]\Delta x\Delta y\Delta z \qquad (17)$$

The quantity in the square brackets should be recognized as the second derivative of $v_x(y)$ with respect to y. Dividing through both sides by the volume $\Delta x\Delta y\Delta z$ gives us the viscous force per unit volume

$$\frac{\Delta F_x}{\Delta x\Delta y\Delta z} = f_{vx} \quad \begin{array}{l}\textit{viscous force per unit}\\ \textit{volume acting on}\\ \textit{the fluid element}\end{array}$$

$$\boxed{f_{vx} = \mu\frac{\partial^2 v_x(y)}{\partial y^2}} \qquad (18)$$

This is the formula for the viscous force per unit volume acting on the fluid particles when we have a purely x directed flow of a Newtonian fluid whose speed varies only in the y direction. In the next section we generalize the result to three dimensional flows.

## VISCOUS FORCE FOR THREE DIMENSIONAL FLOWS

At first sight, there seems to be a rather obvious extension of Equation (18) to three dimensional flows. In a chapter devoted to discussing the operator $\nabla^2$, we might expect that the generalization of our formula for the viscous force $\vec{f}_v$ per unit volume should be

$$\vec{f}_v = \mu\nabla^2\vec{v} \qquad (19)$$

To check that Equation (19) reduces to our result in Equation (18), when $\vec{v}$ is the one dimensional flow $v_x(y)$, we have

$$f_{vx} = \mu\nabla^2 v_x(y) = \mu\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)v_x(y)$$

$$= \mu\frac{\partial^2 v_x(y)}{\partial y^2} \qquad (20)$$

Thus we get the desired result for one dimensional flows.

However, complications arise in three dimensional flows that we did not consider in our analysis of the simple one dimensional flow pattern. In three dimensions, fluids flow around corners and x directed flows can become y or z directed. The definition of viscous stress we gave in Equation (11) simply cannot handle changes in the direction of the flow.

An effective way to deal with viscous forces in three dimensional flows is to note that the resulting force $\vec{f}_v$ per unit volume must be a vector field. That is, $\vec{f}_v$ must transform like a vector field when we rotate the coordinate system. (See the discussion of the transformation of vector fields at the end of the geometrical discussion of the gradient in Chapter 3.)

We will also require that $\vec{f}_v$ ***be made up of some combination of constants and second derivatives of the velocity field.*** These requirements on $\vec{f}_v$ are essentially what we mean by a ***Newtonian fluid with constant coefficients***. If the viscous forces are more complex, which they can be for something like a liquid crystal, then we say that the fluid is ***non Newtonian***.

What we want is the most general combination we can make out of constants, two derivatives $\vec{\nabla}$, and a velocity field $\vec{v}$. Basically we have three vectors $\vec{\nabla}$, $\vec{\nabla}$, $\vec{v}$, and we must multiply them together to get a single vector. To do this, we have to take the dot product of two of them. The possibilities are $(\vec{\nabla}\cdot\vec{\nabla})\vec{v}$ and $\vec{\nabla}(\vec{\nabla}\cdot\vec{v})$.* As a result, our most general formula for a Newtonian fluid with constant coefficients is

$$\vec{f}_v = \mu_1(\vec{\nabla}\cdot\vec{\nabla})\vec{v} + \mu_2\vec{\nabla}(\vec{\nabla}\cdot\vec{v}) \qquad (21)$$

where $\mu_1$ and $\mu_2$ are constants. There is no other combination of constants and second derivatives of the velocity field that transforms as a vector when we rotate the coordinate system.

If we are dealing with a constant density fluid, $\vec{\nabla}\cdot\vec{v} = 0$ and we are left with

$$\vec{f}_v = \mu_1(\vec{\nabla}\cdot\vec{\nabla})\vec{v} = \mu\nabla^2\vec{v} \qquad (19a)$$

which is the result we guessed back in Equation (19), with $\mu_1 = \mu$.

Equation (21) suggests that it is possible to have a second kind of viscosity when the fluid is compressible and $\vec{\nabla}\cdot\vec{v}$ is not zero. This has in fact been observed, and $\mu_2$ is sometimes called the ***second viscosity coefficient***. (Some texts use a second viscosity coefficient defined as $\lambda = \mu_2 - \mu$.) In this text we will only deal with incompressible fluids where there is no second viscosity, and $\vec{f}_v$ is simply given by the Laplacian operator $\nabla^2$ acting on $\vec{v}$, namely $\vec{f}_v = \mu\nabla^2\vec{v}$.

* *(You might also consider vector cross products involving $\vec{\nabla}$, $\vec{\nabla}$, and $\vec{v}$. The possibilities are $\vec{\nabla}\times(\vec{\nabla}\cdot\vec{v})$, $\vec{\nabla}\cdot(\vec{\nabla}\times\vec{v})$, and $\vec{\nabla}\times(\vec{\nabla}\times\vec{v})$. At the beginning of Chapter 9, we find that the first two of these are identically zero, and the third turns out to be*

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{v}) = \vec{\nabla}(\vec{\nabla}\cdot\vec{v}) - (\vec{\nabla}\cdot\vec{\nabla})\vec{v}$$

*which involves only the two terms we got from dot products. Thus we get nothing new by considering cross products. )*

## Viscous Force in Cylindrical Coordinates

Now that we have the formula for the viscous force $\vec{f}_v = \mu\nabla^2\vec{v}$, which applies to any fluid that we will consider in this text, we are free to use general formulas we have in the formulary for $\nabla^2$ in various coordinate systems. We are about to study the flow of a viscous fluid in a pipe, a problem that obviously has cylindrical symmetry. Thus to analyze the viscous forces, we should work with $\nabla^2\vec{v}$ in cylindrical coordinates.

We mentioned earlier that $\nabla^2$ acting on a vector field is more complex than $\nabla^2$ acting on a scalar field in anything except Cartesian coordinates. Thus evaluating $\nabla^2\vec{v}$ in cylindrical coordinates will give us some practice in correctly using the formulary.

From the formulary we find the following formula for $\nabla^2$ acting on a scalar field f and a vector field $\vec{A}$.

$$\nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial\theta^2} + \frac{\partial^2 f}{\partial z^2} \qquad (22)$$

where the coordinates r, $\theta$, z are the unit vectors $\hat{r}$, $\hat{\theta}$, $\hat{z}$ shown in Figure (4).

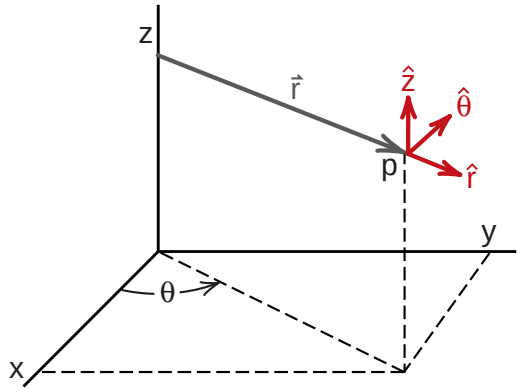



**Figure 4**
***Cylindrical coordinates.***

Looking farther down in the formulary we find for the components of $\nabla^2\vec{A}$

$$(\nabla^2\vec{A})_r = \nabla^2 A_r - \frac{2}{r^2}\frac{\partial A_\theta}{\partial\theta} - \frac{A_r}{r^2} \qquad (23a)$$

$$(\nabla^2\vec{A})_\theta = \nabla^2 A_\theta + \frac{2}{r^2}\frac{\partial A_r}{\partial\theta} - \frac{A_\theta}{r^2} \qquad (23b)$$

$$(\nabla^2\vec{A})_z = \nabla^2 A_z \qquad (23c)$$

where, for example, $\nabla^2 A_z$ means apply Equation (22) to $A_z$

$$\nabla^2 A_z(r,\theta,z) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial A_z}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 A_z}{\partial\theta^2} + \frac{\partial^2 A_z}{\partial z^2} \qquad (24)$$

All this looks like a terrible mess. But suppose we have a fluid flowing smoothly along a pipe as shown in Figure (5). Taking the $\hat{z}$ direction down the pipe and r the distance out from the axis of the pipe, we can assume, for cylindrical symmetry, that $\vec{v}(r,\theta,z)$ is purely $\hat{z}$ directed and depends only on the radius r.

$$\vec{v}(r,\theta,z) = \hat{z}v_z(r) \qquad (25)$$

Now let us work out $\nabla^2\vec{v}$ for this simple case using Equations (23) for $\nabla^2$ in cylindrical coordinates. Because $v_r$ and $v_\theta$ are zero, we do not worry about Equations (23a) and (23b). From (23c) we have

$$(\nabla^2\vec{v})_z = \nabla^2 v_z \qquad (26)$$

Thus for this case we do not have to worry about the extra stuff that comes in when we take $\nabla^2$ of a vector.

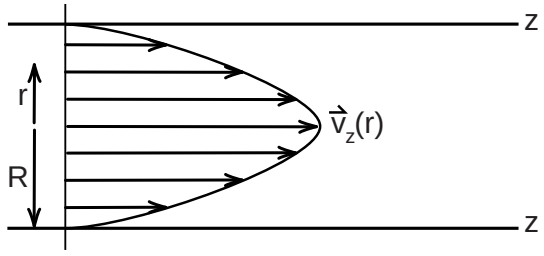



**Figure 5**
***Velocity profile for the uniform flow in a pipe.***

Next we note that $v_z = v_z(r)$, thus we can ignore the $\partial v_z/\partial\theta$ and $\partial v_z/\partial z$ terms in (13a) and we are left with

$$(\nabla^2\vec{v})_z = \frac{1}{r}\frac{\partial}{\partial r}\left[r\frac{\partial v_z(r)}{\partial r}\right] \qquad (27)$$

which is not such a difficult thing to work with after all.

To get a feeling for what the viscous force looks like for pipe flow, we look up in a fluids text what the so called ***laminar*** (i.e., non turbulent) velocity profile is in a pipe. The result they give is

$$v_z(r) = \frac{V_0}{R^2}(R^2 - r^2) \qquad \begin{matrix}\textit{parabolic}\\ \textit{velocity}\\ \textit{profile}\end{matrix} \qquad (28)$$

where R is the radius of the pipe, $V_0$ the flow speed at the center, and r the radial distance from the axis. This is the parabolic profile shown in Figure (5). You can see that at the edge of the pipe, where r = R, the velocity goes to zero. At the center where r = 0, $v_z = V_0$ is a maximum.

To calculate the viscous force per unit volume for this parabolic profile, we have

$$\vec{f}_v = \mu\nabla^2\vec{v} \qquad (29)$$

$$(f_v)_z = \mu(\nabla^2\vec{v})_z = \mu\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial v_z(r)}{\partial r}\right) \qquad (30)$$

With Equation (28) written as

$$v_z(r) = -\frac{V_0}{R^2}r^2 + V_0 \qquad (28a)$$

we easily get

$$\frac{\partial v_z(r)}{\partial r} = -\frac{2V_0}{R^2}r$$

$$r\frac{\partial v_z(r)}{\partial r} = -\frac{2V_0}{R^2}r^2 \qquad (31)$$

Thus $(\nabla^2\vec{v})_z$ becomes

$$\begin{aligned}(\nabla^2\vec{v})_z &= \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial v_z(r)}{\partial r}\right)\\ &= \frac{1}{r}\frac{\partial}{\partial r}\left(-\frac{2V_0}{R^2}r^2\right)\\ &= \frac{1}{r}\left(-\frac{2V_0}{R^2}\frac{\partial r^2}{\partial r}\right)\\ &= \frac{1}{r}\left(-\frac{2V_0}{R^2}2r\right)\end{aligned} \qquad (32)$$

The r's cancel and we are left with

$$(\nabla^2\vec{v})_z = -\frac{4V_0}{R^2} \qquad (33)$$

The viscous force $\vec{f}_v = \mu\nabla^2\vec{v}$ becomes

$$\boxed{f_{vz} = -\mu\left(\frac{4V_0}{R^2}\right)} \qquad (34)$$

We end up with the result that $\vec{f}_v$ points in the $-\hat{z}$ direction (it has only a negative z component) and is constant in magnitude throughout the pipe. This is a wonderfully simple result considering the staggering mess of terms we faced in Equation (23).

We will see that the physics of the parabolic laminar flow is that this uniform $-\hat{z}$ oriented viscous force is balanced by a uniform $+\hat{z}$ oriented pressure gradient down the tube. Thus there is no net force on each fluid element and the fluid moves down the pipe without acceleration, i.e., at constant velocity.

## Measuring the Viscosity Coefficient

If we have an apparatus where we know the pressure gradient, we can use that to measure the viscosity coefficient $\mu$ of the fluid. Such an apparatus is sketched in Figure (6), a sketch taken from the excellent fluid dynamics text by Tritton.

Since there is essentially no viscosity acting in the region between points (1) at the top of the fluid in the container, and point (2) near the entrance to the pipe, we can use Bernoulli's equation to get

$$p_1 + \frac{\rho v_1^2}{2} + \rho g h_1 = p_2 + \frac{\rho v_2^2}{2} + \rho g h_2 \qquad (35)$$

With $v_1 = 0$ and $h_1 - h_2 = h$, we get

$$p_2 - p_1 = \rho g h - \frac{\rho v_2^2}{2} \qquad (36)$$

If we use a sufficiently long and small diameter pipe, the pipe flow velocity will be sufficiently small that we can neglect $v_2^2$ compared to gh. Noting that both $p_3$ and $p_1$ are both atmospheric pressure and thus equal, we get for the pressure difference $(p_2 - p_3)$ at the ends of the pipe

$$(p_2 - p_3) = \rho g h \qquad \textit{pressure difference between ends of the pipe} \qquad (37)$$

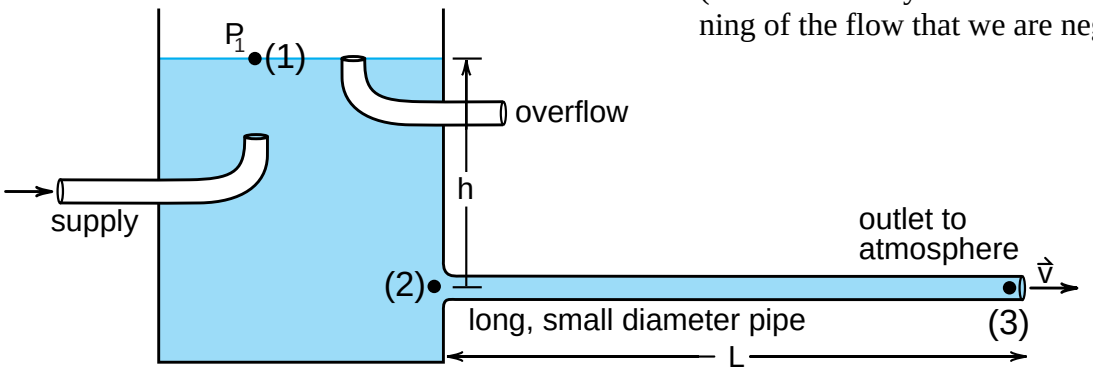



**Figure 6**
***Apparatus to measure the viscosity coefficient.***

The pressure force on the fluid at the front end of the pipe is $p_2 A_2 = p_2 A$ where A is the cross sectional area of the pipe. At the far end it is $-p_3 A$, the minus sign is used because the pressure force is in the –z direction. Thus the net pressure force $\vec{F}_p$ is

$$\begin{aligned} \vec{F}_p &= \hat{z}(p_2 A - p_3 A) \\ &= \hat{z}(p_2 - p_3)A \\ &= \hat{z}\rho g h A \end{aligned} \qquad (38)$$

If we divide $\vec{F}_p$ by the volume AL of the pipe, we get the average pressure force per unit volume $\vec{f}_p$.

$$\vec{f}_p = \frac{\vec{F}_p}{AL} = \frac{\hat{z}}{AL}\rho g h A$$

$$\vec{f}_p = \hat{z}\frac{\rho g h}{L} \qquad \textit{average pressure force per unit area} \qquad (39)$$

As we mentioned, for steady laminar flow, the viscous force should be exactly opposed by the pressure force so that there is no acceleration of the fluid. Since the viscous force per unit volume is uniform throughout the fluid for parabolic pipe flow, the pressure force per unit volume should also be uniform, with the result that Equation (39) for $\vec{f}_P$ should apply at all points in the fluid in the pipe. (There will always be some disturbance at the beginning of the flow that we are neglecting.)

Saying that the viscous and pressure forces oppose each other throughout the pipe flow gives us from Equations (34) for $\vec{f}_\nu$ and (39) for $\vec{f}_p$

$$\vec{f}_p = -\vec{f}_\nu$$

$$\hat{z}\,\frac{\rho g h}{L} = -\left[-\hat{z}\,\frac{4V_0}{R^2}\mu\right]$$

$$\frac{\rho g h}{L} = \frac{4V_0\mu}{R^2} \qquad (40)$$

We are left with an equation involving measurable constants and the viscosity coefficient $\mu$.

Later in the text, we will see that the ratio $\mu/\rho$, which is called the ***kinematic viscosity*** coefficient $\nu$, is more convenient for theoretical work. Equation (40) gives us for this ratio

$$\boxed{\nu \equiv \frac{\mu}{\rho} = \frac{gh}{L}\times\frac{R^2}{4V_0}} \qquad (41)$$

*kinematic viscosity determined from parabolic pipe flow*

The only constant that may be a bit difficult to measure directly is the stream velocity $V_0$ at the center. This can be accurately determined by measuring the flow rate which we will call $\Phi$ (phi), and then express $V_0$ in terms of $\Phi$. We have called the flow rate $\Phi$ because it is simply the flux $\Phi$ of the fluid through the pipe, given by our old flux formula

$$\Phi = \int_{\text{area of tube}} \vec{v}\cdot d\vec{A} \qquad (42)$$

and is measured, in the MKS system, in cubic meters per second.

To calculate $\Phi$, we divide the cross sectional area into circular bands of radius r, thickness dr, as shown in Figure (7). The area of a band is $2\pi r\,dr$ and the flux $d\Phi$ through the band is

$$d\Phi = 2\pi r v(r)\,dr \qquad (43)$$

With v(r) given by the parabolic profile $(V_0/R^2)(R^2 - r^2)$, we get for the total flux

$$\Phi = \int_0^R d\Phi = \int_0^R \frac{V_0}{R^2}\left(R^2 - r^2\right)2\pi r\,dr$$

$$= 2\pi V_0\left[\int_0^R r\,dr - \frac{1}{R^2}\int_0^R r^3 dr\right]$$

$$= 2\pi V_0\left[\frac{r^2}{2}\Big|_0^R - \frac{1}{R^2}\frac{r^4}{4}\Big|_0^R\right]$$

$$\Phi = 2\pi V_0\left[\frac{R^2}{2} - \frac{R^2}{4}\right] = \frac{V_0}{2}\,\pi R^2 \qquad (44)$$

Since $V_0(\pi R^2)$ is the flux we would get if the velocity were a uniform $V_0$ across the pipe, we see that the flow rate for a parabolic profile is half that for a uniform flow.

With Equations (41) and (44) we can now express the kinematic viscosity $\nu$ in terms of the easily measured volume flux $\Phi$. From Equation (44) we get

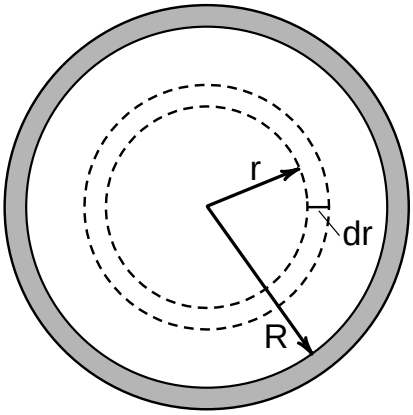



**Figure 7**
***The integration area is the area $2\pi R\,dr$ of the band.***

$$V_0 = \frac{2\Phi}{\pi R^2} \quad ; \quad \frac{1}{V_0} = \frac{\pi R^2}{2\Phi}$$

and from Equation (41) we get

$$\nu = \frac{\mu}{\rho} = \frac{gh}{L}\frac{R^2}{4}\frac{1}{V_0}$$

$$= \frac{gh}{L}\frac{R^2}{4}\frac{\pi R^2}{2\Phi}$$

$$\boxed{\nu = \frac{\pi g h R^4}{8 L \Phi}} \qquad \textit{formula for measuring kinematic viscosity} \qquad (45)$$

Although rather a mess of constants appears in our formula for the kinematic viscosity $\nu$, all are quite easily measured. Note that by going to the kinematic viscosity, the result is independent of the density of the fluid.

---

**Exercise 1**

Show that the kinematic viscosity $\nu$ has the dimensions of $\text{meters}^2/\text{second}$.

---

The two fluids that we will most often use in any discussion of fluid dynamics are water and air. At room temperature and pressure, the kinematic viscosity $\nu$ of these two fluids are approximately

$$\nu_{water} = 1.0 \times 10^{-6} \text{meter}^2/\text{second}$$

$$\nu_{air} = 1.5 \times 10^{-5} \text{meter}^2/\text{second} \qquad (46)$$

Intuitively you would think that air would be much less viscous than water, but the two coefficients $\nu_{air}$ and $\nu_{water}$ are quite close, with air having the greater value. What has happened is that we have divided by the density, which brings the viscosity coefficients much closer together.

# Appendix: The Operator $\nabla^2$ in Spherical Polar Coordinates

## SPHERICAL POLAR COORDINATES

We will begin with a review of spherical coordinates discussed in Chapter 3. In spherical polar coordinates, the three unit vectors are $\hat{r}$, $\hat{\theta}$, $\hat{\phi}$ are shown in Figure (A1) which is Figure (3-10) repeated. We have a complication in evaluating $\nabla^2 f$ in spherical polar coordinates because these unit vectors change direction as we move about, and we can no longer set the derivatives of the unit vectors to zero. Thus we have to evaluate derivatives of the unit vectors as well as use the rather messy formula for $\vec{\nabla} f$ we derived in Equation (3-88)

$$\vec{\nabla} f(r,\theta,\phi) = \hat{r}\frac{\partial f}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial f}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial f}{\partial \phi} \qquad (3\text{-}88)$$

What we have to evaluate is the complete expression

$$\begin{aligned}\nabla^2 f(r,\theta,\phi) &\equiv \vec{\nabla}\cdot\vec{\nabla} f(r,\theta,\phi)\\ &= \left(\hat{r}\frac{\partial}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial}{\partial \phi}\right)\\ &\cdot\left(\hat{r}\frac{\partial f}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial f}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial f}{\partial \phi}\right)\end{aligned} \qquad (A1)$$

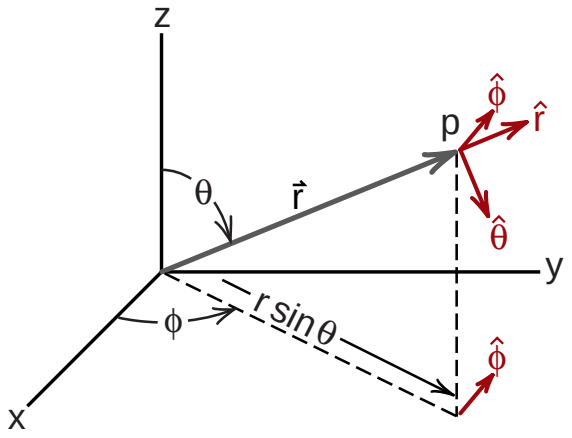



**Figure A1 (3-10 repeated)**
***Unit vectors in spherical polar coordinates.***

This product involves terms like

$$\left(\frac{\hat{\theta}}{r}\frac{\partial}{\partial \theta}\right)\cdot\left(\hat{r}\frac{\partial f}{\partial r}\right) = \frac{\hat{\theta}}{r}\cdot\left[\frac{\partial \hat{r}}{\partial \theta}\frac{\partial f}{\partial r} + \hat{r}\frac{\partial^2 f}{\partial \theta \partial r}\right]$$

$$= \frac{1}{r}\left(\hat{\theta}\cdot\frac{\partial \hat{r}}{\partial \theta}\right)\frac{\partial f}{\partial r} + (\hat{\theta}\cdot\hat{r})\frac{1}{r}\frac{\partial^2 f}{\partial \theta \partial r} \qquad (A2)$$

Because the unit vectors always remain perpendicular to each other as we move around in space, $\hat{\theta}\cdot\hat{r} = 0$ and the second term in Equation (A2) is zero. However, when we change the angle $\theta$, the unit vector $\hat{r}$ changes direction. For example, at $\theta = 0$, $\hat{r}$ points straight up, but at $\theta = 90°$, $\hat{r}$ is horizontal. Thus $\partial\hat{r}/\partial\theta$ is not zero and has to be evaluated.

In order to evaluate Equation (A1) for $\nabla^2 f$, we will first calculate all the derivatives of all the unit vectors, and then plug the whole mess together. We find derivatives like $\partial\hat{r}/\partial\theta$ by evaluating the change $\Delta\hat{r}$ as we make a change $\Delta\theta$ and then taking the limit as $\Delta\theta$ goes to zero. The nine derivatives are as follows.

## Derivatives of $\hat{r}$

### 1) Change of $\hat{r}$ with r

$$\frac{\partial \hat{r}}{\partial r} = 0$$

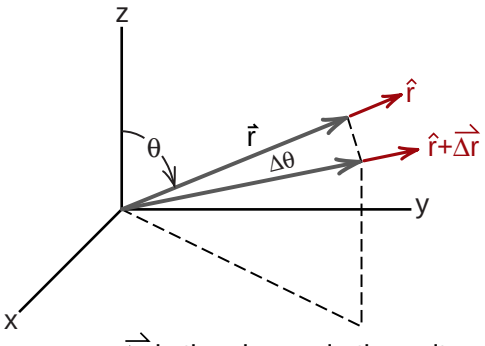
*because $\hat{r}$ does not change direction as we go out along a radius*

### 2) Change of $\hat{r}$ with $\theta$

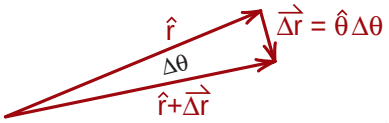
Figure (A2) shows $\Delta\vec{r}$ that we get when $\theta$ increases by $\Delta\theta$. We see that $\Delta\vec{r}$ points in the $\hat{\theta}$ direction and has a length $|\Delta\vec{r}| = |\hat{r}|\Delta\theta = 1 \times \Delta\theta$. Thus we get

$$\Delta\vec{r} = \hat{\theta}(\Delta\theta)$$

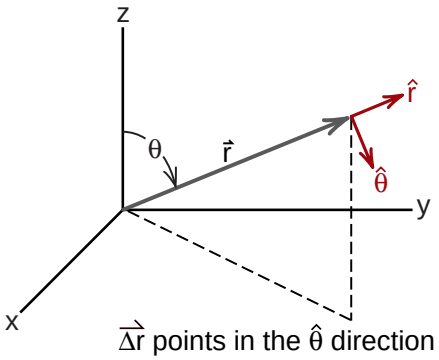
$$\frac{\Delta\vec{r}}{\Delta\theta} = \hat{\theta} \Rightarrow \boxed{\frac{\partial \hat{r}}{\partial \theta} = \hat{\theta}} \qquad \text{(A3)}$$




$\Delta\vec{r}$ is the change in the unit vector $\hat{r}$ when we increase $\theta$ by $\Delta\theta$.




Unit vectors enlarged. $|\hat{r}| = |\hat{r}+\Delta\vec{r}| = 1$




$\Delta\vec{r}$ points in the $\hat{\theta}$ direction

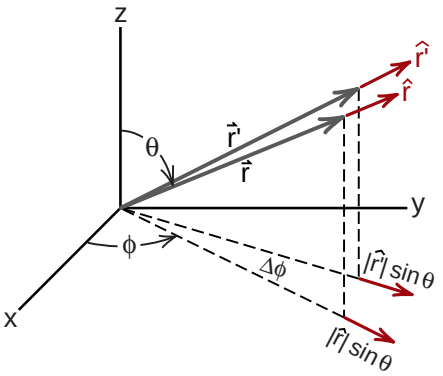
**Figure A2**
***Evaluation of $\partial\hat{r}/\partial\theta$.***

### 3) Change of $\hat{r}$ with $\phi$

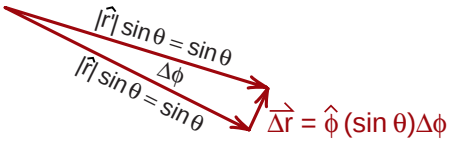
In Figure (A3), when we go from $\phi$ to $\phi + \Delta\phi$, the unit vector $\hat{r}$ goes to the unit vector $\hat{r}'$. The projections of $\hat{r}$ and $\hat{r}'$ in the horizontal plane have a length $|\hat{r}|\sin\phi = 1\sin\phi$, and differ in direction by an angle $\Delta\phi$. The change $\Delta\vec{r} = \hat{r}' - \hat{r}$ points in the $\hat{\phi}$ direction, and has the same length as the change in the horizontal projections of $\hat{r}$ and $\hat{r}'$, which from the small triangle is seen to be $(\sin\phi)(\Delta\phi)$. Thus

$$\Delta\vec{r} = \hat{\phi}(\sin\phi)\Delta\phi$$

$$\frac{\Delta\vec{r}}{\Delta\phi} = \boxed{\hat{\phi}(\sin\phi) = \frac{\partial \hat{r}}{\partial \phi}} \qquad \text{(A4)}$$




$\Delta\vec{r}$ is the change in the unit vector $\hat{r}$ when we increase $\phi$ by $\Delta\phi$.




Unit vectors enlarged.

**Figure A3**
***Evaluation of $\partial\hat{r}/\partial\phi$.***

## Derivatives of $\hat{\theta}$

### 4) Change of $\hat{\theta}$ with r

None of the unit vectors change direction as we go out along the radius, thus

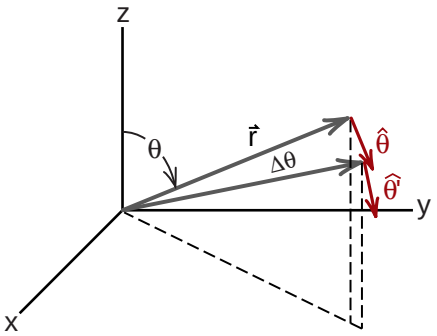
$$\boxed{\frac{\partial\hat{\theta}}{\partial r} = 0} \qquad \text{(A5)}$$

### 5) Change of $\hat{\theta}$ with $\theta$

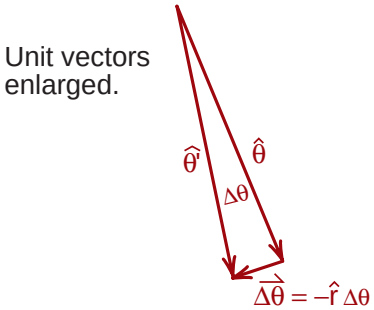
From Figure (A4) we see that as we increase $\theta$ to $\theta + \Delta\theta$, the unit vector $\hat{\theta}$ goes to $\hat{\theta}' = \hat{\theta} + \Delta\vec{\hat{\theta}}$. From the small triangle, we see that the change $\Delta\vec{\hat{\theta}}$ points in the $-\hat{r}$ direction, and has a magnitude $\Delta\theta$. Thus we have

$$\Delta\vec{\hat{\theta}} = (-\hat{r}\,)\Delta\theta$$

$$\frac{\Delta\vec{\hat{\theta}}}{\Delta\theta} = \boxed{-\hat{r} = \frac{\partial\hat{\theta}}{\partial\theta}} \qquad \text{(A6)}$$




$\Delta\vec{\hat{\theta}}$ is the change in the unit vector $\hat{\theta}$ when we increase $\theta$ by $\Delta\theta$.




$\Delta\vec{\hat{\theta}}$ points in the $-\hat{r}$ direction

**Figure A4**
***Evaluation of $\partial\hat{\theta}/\partial\theta$.***

### 6) Change of $\hat{\theta}$ with $\phi$

From Figure (A5), we see that $\hat{\theta}$ changes to $\hat{\theta}'$ as $\phi$ goes to $\phi + \Delta\phi$. The change $\Delta\vec{\hat{\theta}}$ points in the $\hat{\phi}$ direction. To determine the magnitude of $\Delta\vec{\hat{\theta}}$, note that $\Delta\vec{\hat{\theta}}$ and its projection in the horizontal plane are the same. Since the projections of $\hat{\theta}$ and $\hat{\theta}'$ have a length of $\cos\theta$, and an angle $\Delta\phi$ between them, the length of $\Delta\vec{\hat{\theta}}$ is $\cos\theta\Delta\phi$ as seen in the small horizontal triangle. Thus

$$\Delta\vec{\hat{\theta}} = \hat{\phi}(\cos\phi)\Delta\phi$$

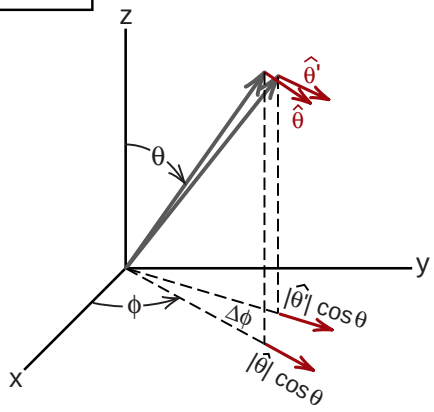
$$\frac{\Delta\vec{\hat{\theta}}}{\Delta\phi} = \boxed{\hat{\phi}\cos\theta = \frac{\partial\hat{\theta}}{\partial\phi}} \qquad \text{(A7)}$$

## Derivatives of $\hat{\phi}$

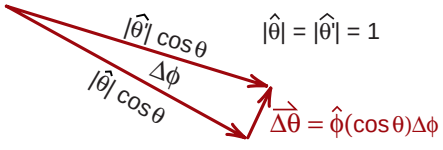
### 7) Change of $\hat{\phi}$ with r

As we noted earlier, the unit vectors do not change with r, thus

$$\boxed{\frac{\partial\hat{\phi}}{\partial r} = 0} \qquad \text{(A8)}$$




$\Delta\vec{\hat{\theta}}$ is the change in the unit vector $\hat{\theta}$ when we increase $\phi$ by $\Delta\phi$.




$\Delta\vec{\hat{\theta}}$ points in the $\hat{\phi}$ direction

**Figure A5**
***Evaluation of $\partial\hat{\theta}/\partial\phi$.***

### 8) Change of $\hat{\phi}$ with $\theta$

As we can see from Figure (A6), the unit vector $\hat{\phi}$ does not change direction when we change the angle $\theta$.

For example, when $\vec{r}$ is in the xz plane, $\hat{\phi}$ points in the $+\hat{y}$ direction for all angles $\theta$. Thus

$$\boxed{\frac{\partial \hat{\phi}}{\partial \theta} = 0} \qquad \text{(A9)}$$

### 9) Change of $\hat{\phi}$ with $\phi$

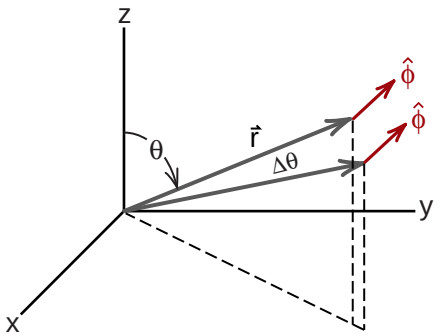
Finally, we have to figure out how the unit vector $\hat{\phi}$ changes with the angle $\phi$. This time we will take a top down view as shown in Figure (A7). When we change $\phi$ to $\phi + \Delta\phi$, the unit vector $\hat{\phi}$ goes to $\hat{\phi}'$. From the small triangle we see that the change $\Delta\hat{\phi}$ points *toward* the $\hat{z}$ axis and has a magnitude $\Delta\phi$.

In Figure (A8), we see that a unit vector $\hat{u}$ pointing toward the z axis is given by

$$\left.\begin{array}{l}\text{unit vector}\\ \text{pointing}\\ \text{toward}\\ \text{z axis}\end{array}\right\} = -\hat{r}\sin\theta - \hat{\theta}\cos\theta$$

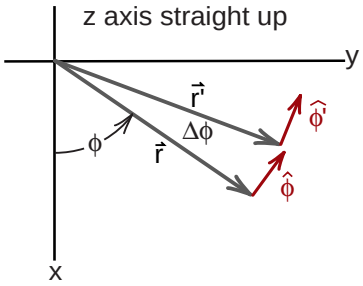
Thus $\Delta\hat{\phi} = \Delta\phi(-\hat{r}\sin\theta - \hat{\theta}\cos\theta)$ and we get

$$\frac{\Delta\hat{\phi}}{\Delta\phi} = \boxed{-\hat{r}\sin\theta - \hat{\theta}\cos\theta = \frac{\partial\hat{\phi}}{\partial\phi}} \qquad \text{(A10)}$$

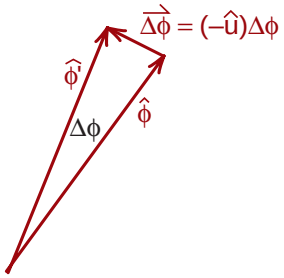



The unit vector $\hat{\phi}$ does not change when we increase $\theta$ by $\Delta\theta$.

**Figure A6**
***Evaluation of $\partial\hat{\phi}/\partial\theta$.***

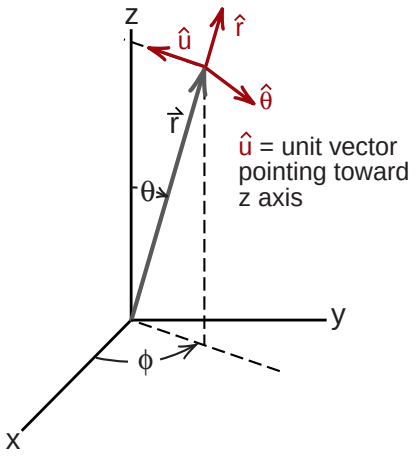


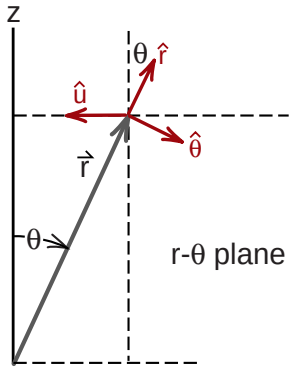

$\Delta\vec{\phi}$ is the change in the unit vector $\hat{\phi}$ when we increase $\phi$ by $\Delta\phi$.




Unit vectors enlarged. The unit vector –û points toward the z axis.

**Figure A7**
***Evaluation of $\partial\hat{\theta}/\partial\phi$.***




$\hat{u} = (-\hat{r})\sin\theta + (-\hat{\theta})\cos\theta$

**Figure A8**
***The unit vector we call $\hat{u}$ that points toward the z axis.***

### Summary of Derivatives of Unit Vectors

In summary, we get

$$\frac{\partial \hat{r}}{\partial r} = 0\,; \quad \frac{\partial \hat{r}}{\partial \theta} = \hat{\theta}\,; \quad \frac{\partial \hat{r}}{\partial \phi} = \hat{\phi}\sin\theta$$

$$\frac{\partial \hat{\theta}}{\partial r} = 0\,; \quad \frac{\partial \hat{\theta}}{\partial \theta} = -\hat{r}\,; \quad \frac{\partial \hat{\theta}}{\partial \phi} = \hat{\theta}\cos\theta$$

$$\frac{\partial \hat{\phi}}{\partial r} = 0\,; \quad \frac{\partial \hat{\phi}}{\partial \theta} = 0\,; \quad \frac{\partial \hat{\phi}}{\partial \phi} = -\hat{r}\sin\theta - \hat{\theta}\cos\theta \qquad \text{(A11)}$$

### Calculation of $\nabla^2 f$

We are now ready to calculate $\nabla^2 f$ given again by Equation (A1)

$$\nabla^2 f(r,\theta,\phi) \equiv \vec{\nabla}\cdot\vec{\nabla} f(r,\theta,\phi)$$

$$= \left(\hat{r}\frac{\partial}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial}{\partial \phi}\right)$$

$$\cdot\left(\hat{r}\frac{\partial f}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial f}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial f}{\partial \phi}\right)$$

$$= \hat{r}\cdot\left[\cancel{\frac{\partial \hat{r}}{\partial r}\frac{\partial f}{\partial r}} + \hat{r}\frac{\partial^2 f}{\partial r^2} + \frac{\partial \hat{\theta}}{\partial r}\cancel{\frac{1}{r}\frac{\partial f}{\partial \theta}} + \hat{\theta}\cancel{\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial f}{\partial \theta}\right)}\right.$$
$$\left. + \cancel{\frac{\partial \hat{\phi}}{\partial r}\frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}} + \hat{\phi}\cancel{\frac{\partial}{\partial r}\left(\frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\right)}\right]$$

$$+ \frac{\hat{\theta}}{r}\cdot\left[\frac{\partial \hat{r}}{\partial \theta}\frac{\partial f}{\partial r} + \cancel{\hat{r}\frac{\partial^2 f}{\partial \theta \partial r}} + \cancel{\frac{\partial \hat{\theta}}{\partial \theta}\frac{1}{r}\frac{\partial f}{\partial \theta}} + \frac{\hat{\theta}}{r}\frac{\partial^2 f}{\partial \theta^2}\right.$$
$$\left. + \cancel{\frac{\partial \hat{\phi}}{\partial \theta}\frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}} + \hat{\phi}\cancel{\frac{\partial}{\partial \theta}\left(\frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\right)}\right]$$

$$+ \frac{\hat{\phi}}{r\sin\theta}\cdot\left[\frac{\partial \hat{r}}{\partial \phi}\frac{\partial f}{\partial r} + \cancel{\hat{r}\frac{\partial^2 f}{\partial \phi \partial r}} + \frac{\partial \hat{\theta}}{\partial \phi}\frac{1}{r}\frac{\partial f}{\partial \theta}\right.$$
$$+ \cancel{\frac{\hat{\theta}}{r}\frac{\partial^2 f}{\partial \phi \partial \theta}} + \cancel{\frac{\partial \hat{\phi}}{\partial \phi}\frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}}$$
$$\left. + \hat{\phi}\left(\frac{1}{r\sin\theta}\frac{\partial f^2}{\partial \phi^2}\right)\right] \qquad \text{(A12)}$$

The terms in Equation (58) with a single line through them are zero because the unit vectors are orthogonal: i.e., $\hat{r}\cdot\hat{\theta} = 0$, $\hat{\theta}\cdot\hat{\phi} = 0$ , etc. Next we use our summary, Equation (57) to evaluate the following terms.

$$\hat{r}\cdot\frac{\partial \hat{r}}{\partial r} = 0 \quad \text{because} \quad \frac{\partial \hat{r}}{\partial r} = 0 \qquad \text{(A13a)}$$

$$\hat{r}\cdot\frac{\partial \hat{\theta}}{\partial r} = 0 \quad \text{because} \quad \frac{\partial \hat{\theta}}{\partial r} = 0 \qquad \text{(A13b)}$$

$$\hat{r}\cdot\frac{\partial \hat{\phi}}{\partial r} = 0 \quad \text{because} \quad \frac{\partial \hat{\phi}}{\partial r} = 0 \qquad \text{(A13c)}$$

$$\hat{\theta}\cdot\frac{\partial \hat{r}}{\partial \theta} = \hat{\theta}\cdot\hat{\theta} = 1 \qquad \text{(A13d)}$$

$$\hat{\theta}\cdot\frac{\partial \hat{\theta}}{\partial \theta} = \hat{\theta}\cdot(-\hat{r}) = 0 \qquad \text{(A13e)}$$

$$\hat{\theta}\cdot\frac{\partial \hat{\phi}}{\partial \theta} = 0 \quad \text{because} \quad \frac{\partial \hat{\phi}}{\partial \theta} = 0 \qquad \text{(A13f)}$$

$$\hat{\phi}\cdot\frac{\partial \hat{r}}{\partial \phi} = \hat{\phi}\cdot(\hat{\phi}\sin\theta) = \sin\theta \qquad \text{(A13g)}$$

$$\hat{\phi}\cdot\frac{\partial \hat{\theta}}{\partial \phi} = \hat{\phi}\cdot(\hat{\phi}\cos\theta) = \cos\theta \qquad \text{(A13h)}$$

$$\hat{\phi}\cdot\frac{\partial \hat{\phi}}{\partial \phi} = \hat{\phi}\cdot(-\hat{r}\sin\theta - \hat{\theta}\cos\theta) = 0 \qquad \text{(A13i)}$$

The terms in Equation (A12) for $\nabla^2 f$, that are zero because of Equations (A13), have a double line through them. We are left with

$$\nabla^2 f = \hat{r}\cdot\hat{r}\frac{\partial^2 f}{\partial r^2} + \frac{\hat{\theta}}{r}\cdot\frac{\partial \hat{r}}{\partial \theta}\frac{\partial f}{\partial r}$$
$$+ \frac{\hat{\theta}}{r}\cdot\frac{\hat{\theta}}{r}\frac{\partial^2 f}{\partial \theta^2} + \frac{\hat{\phi}}{r\sin\theta}\cdot\frac{\partial \hat{r}}{\partial \phi}\frac{\partial f}{\partial r}$$
$$+ \frac{\hat{\phi}}{r\sin\theta}\cdot\frac{\partial \hat{\theta}}{\partial \phi}\frac{1}{r}\frac{\partial f}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\cdot\frac{\hat{\phi}}{r\sin\theta}\frac{\partial^2 f}{\partial \phi^2} \qquad \text{(A14)}$$

Using Equations (A13), Equation (A14) becomes

$$\nabla^2 f = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{1}{r \sin\theta}\sin\theta\frac{\partial f}{\partial r} + \frac{1}{r \sin\theta}\cos\theta\frac{1}{r}\frac{\partial f}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 f}{\partial \phi^2} \qquad \text{(A15)}$$

This becomes

$$\nabla^2 f = \frac{\partial^2 f}{\partial r^2} + \frac{2}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\left(\frac{\cos\theta}{\sin\theta}\frac{\partial f}{\partial \theta} + \frac{\partial^2 f}{\partial \theta^2}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 f}{\partial \phi^2} \qquad \text{(A16)}$$

In most textbooks, you will find the equivalent formula

$$\nabla^2 f = \frac{1}{r}\frac{\partial^2}{\partial r^2}(rf) + \frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial f}{\partial \theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial \phi^2}\right] \qquad (10)$$

which is the result we stated earlier in the chapter.

---

**Exercise 2**

Show that Equation (A15) follows from Equations (A13) and (A14).

---

**Exercise 3**

Show that Equations (A16) and (10) are equivalent.

---

# Calculus 2000-Chapter 5

# Introduction to Complex Variables

## A ROAD MAP

*In this chapter you will see that the use of complex variables greatly simplifies the analysis of RLC circuits and other forms of sinusoidal behavior. This chapter does not depend on previous chapters of the Calculus text and may be studied directly in connection with the related material in Chapters 27 and 31 of the Physics text.*

*This chapter is also background material for the next chapter, Chapter 6, on the Schrödinger's equation. The wave equations we have discussed so far can be solved using either real variables or complex variables. Schrödinger's wave equation is different in that the equation itself involves complex numbers and cannot be handled by real variables alone. That is why this chapter is a prerequisite. Also, to solve Schrödinger's equation for the hydrogen atom requires the use of* $\nabla^2$ *in spherical polar coordinates, which we discussed in the last chapter.*

*Once we finish Chapters 5 and 6 on complex variables and Schrödinger's equation, we return to basic calculus operations, discussing divergence in Chapter 7 and curl in Chapter 8. We then apply divergence and curl to electromagnetism in Chapters 9, 10, and 11, and to fluid dynamics in Chapters 12 and 13.*

## INTRODUCTION

*After introducing the concepts of imaginary and complex numbers, we find that an important feature of a complex number is that it can be expressed as a complex exponential. We then go on to two major applications of complex variables that we just mentioned. One is the analysis of RLC circuits, which can be handled using real variables only, but where there is an enormous simplification if we use complex variables. Then in the next chapter, we discuss the Schrödinger's equation where the equation itself involves complex variables.*

*There are other topics involving the theory of complex variables that we will not discuss in these introductory chapters. It is possible to construct fascinating maps of complex functions and distort these maps in intriguing ways (not completely unlike the distortion of images one can create on the computer). Complex variables are also useful in finding the formulas for various integrals. These advanced topics are usually covered in a graduate level mathematical physics course.*

## IMAGINARY NUMBERS

What number, when multiplied by itself gives (-1)? The answer is none of the ordinary numbers. This number, $\sqrt{-1}$ is not one of the real numbers like 5, –2, 3, etc. It belongs to a completely different system of numbers which we call ***imaginary*** numbers. The number $\sqrt{-1}$ is denoted by the letter $i$, and the square root of any negative number can be written as a real number times $i$. For example

$$\sqrt{-7} = \sqrt{7\times(-1)} = \sqrt{7}\times\sqrt{-1} = (\sqrt{7})i \qquad \begin{matrix}\textit{example}\\ \textit{of an}\\ \textit{imaginary}\\ \textit{number}\end{matrix} \qquad (1)$$

All numbers with one factor of $i$ are imaginary.

## COMPLEX NUMBERS

We can make things a bit more complicated by adding together a real number and an imaginary number, such as $(4 + 3i)$. Such a mixture with both a ***real part*** (4) and an ***imaginary part*** $(3i)$ is called a ***complex number***. These two parts are distinct; there is no way we can confuse the real and imaginary parts because imaginary numbers are not part of the real number system.

This is not the first time we have encountered a quantity that has two distinct parts. In our strobe photographs showing the motion of a ball, we noted that the position of the ball could be described by the ***coordinate vector*** $\vec{r}$, as shown in Figure (1). For the strobe photographs, which only show two dimensions, the coordinate vector $\vec{r}$ was completely specified by its (x) and (y) components. Thus two dimensional coordinate vectors and complex numbers are similar in that they both consist of two independent components.

This similarity suggests that we can treat a complex number in the same way we handle a two dimensional coordinate vector, plotting the real and imaginary parts along different axes. It is traditional to plot the real part along the x axis and the imaginary part along the y axis. Thus, for example, the complex number $(4 + 3i)$ can be represented by a point whose coordinate vector has an x component of 4 and a y component of 3 as shown in Figure (2).

In this chapter you will see that in some cases there is considerable simplification of the mathematics and much greater insight when we use complex numbers. This is illustrated in our analysis of the RLC circuit where we will see that a sinusoidal oscillation and an exponential decay can both be handled by one simple complex function.

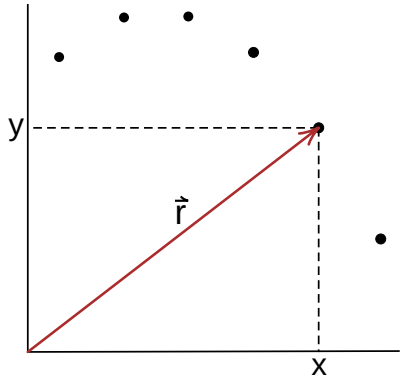



**Figure 1**
***The coordinate vector for a two dimensional strobe photograph.***

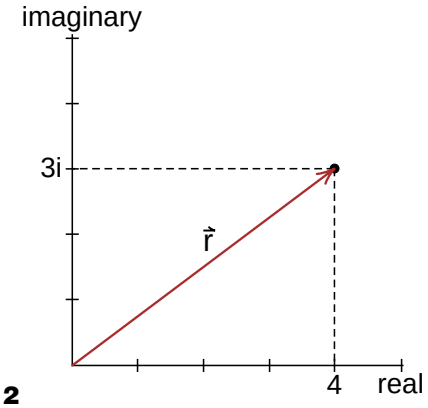



**Figure 2**
***Plot of the complex number (4 + 3i), where the real part is plotted along the x axis and the imaginary part long the y axis.***

## EXPONENTIAL FORM OF THE COMPLEX NUMBER

Once we start plotting complex numbers on x and y axes, we will find that any complex number can be expressed in the exponential form $re^{i\theta}$. How we get to this rather remarkable result can be seen in the following way.

Let us go back to Figure (2), showing the plot of the complex number $(4 + 3i)$. One way to describe that point is to give its x and y coordinates $(x = 4, y = 3i)$. An equally good description, shown in Figure (2a), is to give the distance r from the origin to the point, and the angle $\theta$ that $\vec{r}$ makes with the x or real axis. From the Pythagorean theorem we have

$$r^2 = x^2 + y^2 = 4^2 + 3^2 = 16 + 9 = 25$$

$$r = 5 \qquad (2)$$

The tangent of the angle $\theta$ is the opposite side y divided by the adjacent side x

$$\tan\theta = \frac{y}{x} = \frac{3}{4} = .75 \qquad (3)$$

Entering .75 in our calculator and pressing the $\tan^{-1}$ button gives 36.9°. Thus the point is located at a distance r = 5 from the origin at an angle $\theta = 36.9°$.

It is traditional to use the letter (z) to describe a complex number. Thus if a complex number (z) has a real part (x) and an imaginary part (iy), we can write

$$z = x + iy \qquad (4)$$

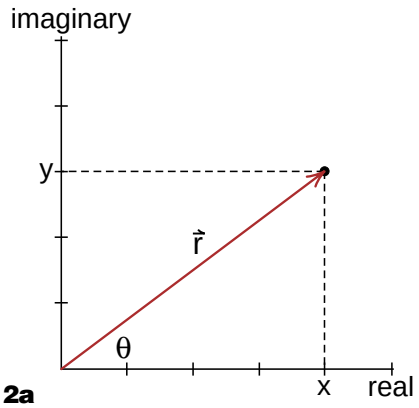



**Figure 2a**
***Plot of the complex number (4 + 3i), showing the angle θ.***

Now let us express z in terms of the variables r and $\theta$ rather than x and y. Since from Figure (2a) we see that

$$x = r\cos\theta$$

$$y = r\sin\theta \qquad (5)$$

we can write (z) as

$$z = x + iy = r\cos\theta + ir\sin\theta$$

$$z = r(\cos\theta + i\sin\theta) \qquad (6)$$

It is the function $(\cos\theta + i\sin\theta)$ that we wish to study in detail.

Let us first look at the derivative of $(\cos\theta + i\sin\theta)$. Since

$$\frac{d}{d\theta}\cos\theta = -\sin\theta; \quad \frac{d}{d\theta}\sin\theta = \cos\theta \qquad (7)$$

we get

$$\frac{d}{d\theta}(\cos\theta + i\sin\theta) = -\sin\theta + i\cos\theta$$

Since $(-1) = i^2$, this can be written

$$\frac{d}{d\theta}(\cos\theta + i\sin\theta) = i^2\sin\theta + i\cos\theta = i(\cos\theta + i\sin\theta) \qquad (8)$$

To express this result more formally, let us write

$$f(\theta) = (\cos\theta + i\sin\theta) \qquad (9)$$

Then Equation (8) becomes

$$\frac{d}{d\theta}f(\theta) = if(\theta) \qquad (8a)$$

To within a constant (i), the function $f(\theta)$ is equal to its own derivative. What function that you are already familiar with, behaves this way? The ***exponential function***! Recall that

$$\frac{d}{dx}e^{ax} = ae^{ax} \qquad (10)$$

Thus if we replace (x) by $\theta$ and (a) by (i), we get

$$\frac{d}{d\theta}e^{i\theta} = ie^{i\theta} \qquad (11)$$

Comparing Equations (8) and (11), we see that the function $(\cos\theta + i\sin\theta)$ and the function $e^{i\theta}$ obey the same rule for differentiation.

When two functions $(\cos\theta + i\sin\theta)$ and $e^{i\theta}$ have the same derivatives, does that mean that they are the same functions? It will if we show that both functions start off with the same value for small values of $\theta$. Then as we increase $\theta$, if both functions have the same derivative or slope, they must continue to be the same function for all values of $\theta$.

## Small Angle Approximations

We can show that $(\cos\theta + i\sin\theta)$ and $e^{i\theta}$ have the same values for small $\theta$ by using the small angle approximations for $\sin\theta$, $\cos\theta$ and $e^{i\theta}$. In our discussion of the exponential in Chapter 1 of the Calculus text, (Cal 1, Eq. 136), we had

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \qquad (1\text{-}136)$$

While this expansion is true for any value of x, it is most useful for small values of x where we do not have to keep many terms to get an accurate answer.

Setting $x = i\theta$ gives

$$e^{i\theta} = 1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \cdots \qquad (12)$$

***(Since our previous discussion of exponents only dealt with real numbers, we can consider Equation (12) as the definition of what we mean when the exponent is a complex number)***.

What we did not discuss earlier were the expansions for $\cos\theta$ and $\sin\theta$. Let us state them and check their accuracy now. They are

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \cdots \qquad (13)$$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots \qquad (14)$$

where $\theta$ is in radians. Again these expansions are valid for any value of $\theta$, but most useful for small values where we do not have to keep many terms.

Let us check the accuracy of these expansions for $\theta = .1$ radians. We have, keeping three terms,

$$\begin{aligned}\cos(.1) &= 1 - \frac{(.1)^2}{2!} + \frac{(.1)^4}{4!} \\ &= 1 - \frac{.01}{2} + \frac{.0001}{4\times 3\times 2} \\ &= 1 - .005 + .0000004166 \\ &= .995004166\end{aligned} \qquad (15)$$

Changing our calculator from degrees to radians and taking the cos(.1) gives

$$\cos(.1) = .995004165 \qquad (16)$$

We see that we get almost a nine place accuracy by keeping the first three terms of the expansion.

For sin(.1), keeping the first three terms, we have

$$\begin{aligned}\sin(.1) &= .1 - \frac{(.1)^3}{3!} + \frac{(.1)^5}{5!} \\ &= .1 - \frac{.001}{3\times 2} + \frac{.0001}{5\times 4\times 3\times 2} \\ &= .1 - .000166666 + .000000083 \\ &= .009833417\end{aligned} \qquad (17)$$

The calculator gives

$$\sin(.1) = .099833416 \qquad (18)$$

which is again accurate to almost nine places.

*If you can't figure out how to get your calculator to work in radians, you can convert .1 radians to degrees by using the conversion factor*

$$\frac{360 \text{ degrees/cycle}}{2\pi \text{ radians/cycle}} = 57.29577951 \frac{\text{degrees}}{\text{radian}} \qquad (19)$$

Now that we have checked the expansions (13) and (14) let us see what the expansion for $(\cos\theta + i\sin\theta)$ is. We get, replacing – signs by $i^2$, and using $i^4 = +1$,

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} = 1 + \frac{i^2\theta^2}{2!} + \frac{i^4\theta^4}{4!}$$

$$i\sin\theta = i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!}\right) = i\theta + \frac{i^3\theta^3}{3!} + \frac{i^5\theta^5}{5!} \qquad (20)$$

Adding these gives

$$(\cos\theta + i\sin\theta) = 1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \frac{i^4\theta^4}{4!} + \frac{i^5\theta^5}{5!} + \cdots \qquad (21)$$

which is just the expansion we had for $e^{i\theta}$ in Equation (12).

Comparing Equations (20) and (21), you can see that the expansions for $\cos\theta$ and $i\sin\theta$ fit together to produce the much more regular expansion of $e^{i\theta}$. We will also see that it is often much easier to work with the complete function $e^{i\theta}$ than with $\cos\theta$ and $\sin\theta$ separately.

In summary, if we define a complex exponential by the series expansion of Equation (12), then we have shown that

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta} \qquad (22)$$

Even though we checked the sin and cosine expansions for a small value of $\theta$, the fact that $e^{i\theta}$ and $(\cos\theta + i\sin\theta)$ have the same derivative properties means that Equation (22) holds for all values of $\theta$.

If we replace $\theta$ by $-\theta$ in Equation (22) we get

$$e^{-i\theta} = \cos(-\theta) + i\sin(-\theta)$$

Since $\cos(-\theta) = \cos\theta$, $\sin(-\theta) = -\sin\theta$, this gives

$$\boxed{e^{-i\theta} = \cos\theta - i\sin\theta} \qquad (23)$$

If we add Equations (22) and (23), the $\sin\theta$ terms cancel, and we are have, after dividing through by 2,

$$\boxed{\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}} \qquad (24)$$

Subtracting Equation (23) from (22) cancels the $\cos\theta$ terms, leaving, after dividing through by 2i,

$$\boxed{\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}} \qquad (25)$$

If we note that

$$\frac{1}{i} = \frac{1}{i}\times\frac{i}{i} = \frac{i}{i^2} = \frac{i}{-1} = -i$$

we can write Equation (25) as

$$\sin\theta = (-i)\frac{e^{i\theta} - e^{-i\theta}}{2}$$

$$\boxed{\sin\theta = i\frac{e^{-i\theta} - e^{i\theta}}{2}} \qquad (25a)$$

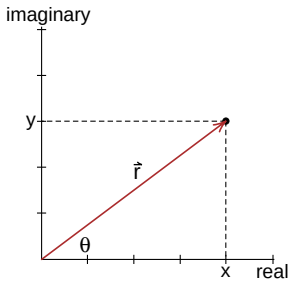
Equations (22) through (25) give a complete prescription of how to go back and forth between $\cos\theta$, $\sin\theta$, $e^{i\theta}$ and $e^{-i\theta}$.

Finally returning to our complex function

$$\boxed{z = x + iy}$$

$$= r\cos\theta + ir\sin\theta$$

$$= r(\cos\theta + i\sin\theta)$$




we now have

$$\boxed{z = re^{i\theta}} \qquad (26)$$

as the other way of expressing a complex number, where r is the distance from the origin and $\theta$ the angle the coordinate vector $\vec{r}$ makes with the x or real axis.

---

**Exercise 1**

(a) Construct a series expansion for $e^{-i\theta}$.

(b) Using the series expansions for $e^{i\theta}$ and $e^{-i\theta}$ in Equation (25a), show that you end up with the series expansion for $\sin(\theta)$.

---

## The Complex Conjugate Z*

The ***complex conjugate*** of a complex number is defined as the number we get by replacing all factors of ($i$) by ($-i$) in the formula for the number. We generally denote the complex conjugate by placing an asterisk after the number. For example, if

$$z = x + iy$$

then

$$z^* = x - iy \qquad (26a)$$

If we start with

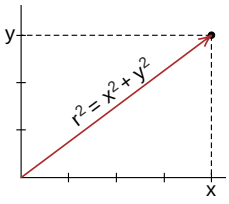
$$z = re^{i\theta}$$

then

$$z^* = re^{-i\theta} \qquad (26b)$$

The main reason for defining a complex conjugate is that the product of a complex number z with its complex conjugate z* is ***always a real positive number***, equal to the square of the distance r that the complex point is from the origin. For our two examples above, we have

$$\begin{aligned} z^*z &= (x - iy)(x + iy) \\ &= x^2 - ixy + iyx - i^2y^2 \\ &= x^2 + y^2 = r^2 \end{aligned}$$




and

$$z^*z = (re^{-i\theta})(re^{i\theta}) = r^2 \qquad (26c)$$

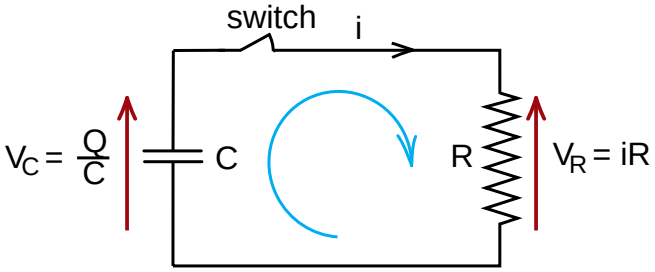



**Figure 3**
***The RC circuit. When we walk around in the direction shown by the circular arrow, we go with $V_C$ but against $V_R$, giving $V_C - V_R$ as the sum of the voltage rises.***

## DIFFERENTIAL EQUATIONS FOR R, L, C CIRCUITS

One of the most convenient uses of complex variables is in the analysis of electric circuits involving resistors, capacitors and inductors. We will see that using complex variables unifies the analysis and greatly simplifies the work involved.

### The RC Circuit

Let us begin with the RC circuit shown in Figure (3). If we charge up the capacitor to a voltage $V_0$, and close the switch, a current flows out of the capacitor through the resistor, and the voltage $V_C$ on the capacitor decays exponentially.

The formulas for the capacitor voltage $V_C$ and resistor voltage $V_R$ are

$$V_C = \frac{Q}{C} \; ; \quad V_R = iR \qquad (27)$$

where Q is the charge on the capacitor, C the capacitor's capacitance, (i) the current through the resistor and R the resistor's resistance in ohms. It is assumed that C and R are constants and that (i) is the rate at which charge Q is leaving the capacitor. That is,

$$i = -\frac{dQ}{dt} \qquad (28a)$$

Setting the sum of the voltage rises around the circuit equal to zero (see Equation 27-41 in the Physics text) and using (28a) for i, gives us

$$V_C - V_R = 0$$

$$\frac{Q}{C} - iR = \frac{Q}{C} + \frac{dQ}{dt}R = 0 \qquad (28b)$$

Dividing through by R, we get

$$\boxed{\frac{dQ}{dt} + \frac{Q}{RC} = 0} \qquad (29)$$

as the differential equation for the amount of charge Q remaining in the capacitor.

## An Aside on Labeling Voltages

To avoid worrying about minus signs like the $i = -dQ/dt$ for the discharging capacitor, we will obtain the differential equations for our L, R, and C circuits by sketching the voltages *when the rate of change of charge in the capacitor and change of the current in an inductor are all* ***positive***. If we had an ***increasing*** current running down through three circuit elements R, L, and C, all the voltages would point up as shown in Figure (4). The resistance voltage $V_R$ is always directed opposite to the current. If the downward current is increasing, then the inductor opposes the increase and points upward. With a positive current flowing into the capacitor, the current is equal to $+dQ/dt$. If the capacitor started off with zero charge, then the upper plate is becoming positively charged by the positive current flowing into it.

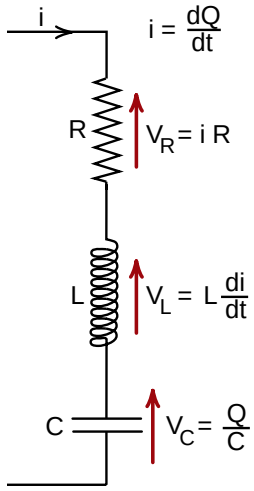



**Figure 4**
***Direction of the voltages for an increasing downward current.***

Using these conventions for the current and voltages, we can construct an RC circuit from Figure (4) by pulling out the inductor and connecting the back side of the circuit as shown in Figure (5). Setting the sum of the voltages to zero around Figure (5) gives (walking around the circuit counterclockwise as shown by the circular arrow)

$$V_R + V_C = 0$$

$$iR + \frac{Q}{C} = 0 \qquad (30)$$

With $i = +dQ/dt$, this gives

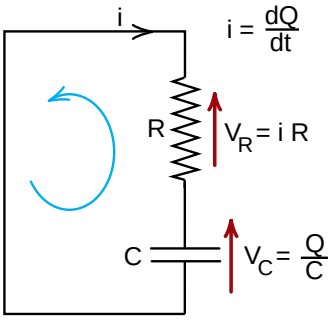
$$R\frac{dQ}{dt} + \frac{Q}{C} = 0 \; ; \qquad \boxed{\frac{dQ}{dt} + \frac{Q}{RC} = 0} \qquad (29a)$$

This is just the same as Equation (29) for the discharge of the capacitor.

Figure (5) appears less intuitive than Figure (3) because we have drawn a current flowing into the capacitor, while we know that the current actually flows out. But the fact that we analyzed the circuit in Figure (5), assuming the wrong direction for the current, does not affect the resulting differential equation for the circuit. When using Kirchoff's laws to analyze a circuit, you do not have to know the correct direction for the currents ahead of time. If you make the wrong guess, the resulting equations will fix things up by giving you a minus sign.

While Figure (5) is less intuitive than Figure (3), it is much more straightforward to stick with all positive quantities and always label our circuit element voltages and currents as shown in Figure (4). With more complex circuits it is the only way to maintain sanity and get the right differential equation.

**Figure 5**
***The RC circuit for a positive increasing current i.***



## Solving the RC Circuit Equation

Solving the differential equation

$$\frac{dQ}{dt} + \frac{Q}{RC} = 0 \qquad (29) \text{ repeated}$$

for the capacitor discharge was quite straightforward. We first looked at the circuit experimentally and saw that the voltage Q/C appeared to decay exponentially as shown in Figure (27-44c) from the Physics text, reproduced here. This suggested that we try, as a guess, a solution of the form

$$Q = Q_0 e^{-\alpha t} \qquad (31)$$

$$\frac{dQ}{dt} = -\alpha Q_0 e^{-\alpha t} \qquad (32)$$

Plugging our guess into Equation (30) gives

$$-\alpha Q_0 e^{-\alpha t} + \frac{Q_0 e^{-\alpha t}}{RC} \doteq 0$$

The constants $Q_0$ and the functions $e^{-\alpha t}$ cancel, and we are left with

$$-\alpha + \frac{1}{RC} \doteq 0$$

We can satisfy the differential equation if $\alpha$ has the value

$$\boxed{\alpha = \frac{1}{RC}} \qquad (33)$$

The formula for Q becomes

$$\boxed{Q = Q_0 e^{-t/RC}} \qquad (34)$$

We see the time constant for the decay of the charge Q is T = RC. I.e., when t gets up to T = RC, the value of the charge has decreased to $e^{-1} = 1/e$ of its original value.

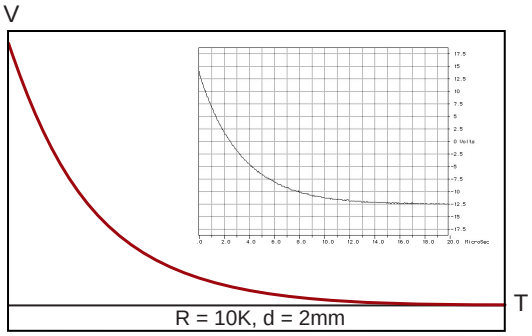



**Figure 27-44c**
***Discharge of our aluminum plate capacitor (separation 2mm) through a 10KΩ resistor. The inset is the experimental data and the solid curve is drawn from that data.***

## The LC Circuit

We will construct an LC circuit from Figure (4) by taking out the resistor and connecting the back side as shown in Figure (6). Setting the sum of the voltage rises around this circuit equal to zero gives

$$V_L + V_C = 0 \qquad (35)$$

$$L\frac{di}{dt} + \frac{Q}{C} = 0 \;; \quad i = +\frac{dQ}{dt} \qquad (36)$$

Writing $di/dt = d^2Q/dt^2$ , and dividing through by L gives

$$\boxed{\frac{d^2Q}{dt^2} + \frac{Q}{LC} = 0} \qquad (37)$$

Now suppose we naïvely try the same exponential decay solution we had for the RC circuit

$$Q = Q_0 e^{-\alpha t} \;; \quad \frac{dQ}{dt} = -\alpha Q_0 e^{-\alpha t} \qquad \text{(guess)}$$

$$\frac{d^2Q}{dt^2} = -\alpha Q_0\left(-\alpha e^{-\alpha t}\right) = \alpha^2 Q_0 e^{-\alpha t} \qquad (38)$$

Plugging our guess (38) into the differential Equation (37) gives

$$\alpha^2 Q_0 e^{-\alpha t} + \frac{Q_0 e^{-\alpha t}}{LC} \doteq 0$$

Again the $Q_0$'s and $e^{-\alpha t}$'s cancel and we are left with

$$\alpha^2 + \frac{1}{LC} \doteq 0 \qquad (39)$$

The differential equation will be solved if we can set

$$\alpha^2 = -\frac{1}{LC} \qquad (40)$$

$$\boxed{\alpha = -\frac{1}{\sqrt{LC}}\sqrt{-1}} \qquad (41)$$

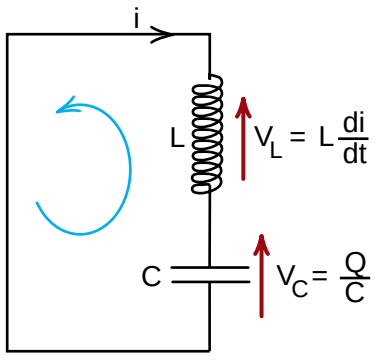



**Figure 6**
***The LC circuit.***

When we tried this in the Physics text, we noted that $\alpha$ comes out imaginary. We also noted that the LC circuit oscillated rather than decayed. Thus we concluded that we had guessed the wrong function, and tried a sine wave

$$Q = Q_0 \sin\omega_0 t \qquad (42)$$

instead. When you plug the guess (42) into the differential Equation (37) you end up with

$$\omega_0{}^2 = \frac{1}{\sqrt{LC}} \; ; \quad \omega_0 = \frac{1}{\sqrt{LC}} \qquad (43)$$

which avoids imaginary numbers and gives a result in agreement with experiment. The quantity $\omega_0 = 1/\sqrt{LC}$ is the resonant frequency of the LC circuit. (If you do not remember plugging the guess $Q = Q_0 \sin(\omega_0 t)$ into the LC differential equation, do so now.)

Knowing more about handling imaginary numbers, let us see what happens if we take our guess $Q = e^{-\alpha t}$ seriously for the LC circuit. We still have to satisfy Equation (40),

$$\alpha^2 = -1/LC$$

Writing $1/LC = \omega_0{}^2$, we get

$$\alpha^2 = -\omega_0{}^2 \qquad (44)$$

which has two solutions, namely

$$\alpha = i\omega_0 \qquad (45a)$$

$$\alpha = -i\omega_0 \qquad (45b)$$

You can see this by noting that both $i^2 = -1$ and $(-i)^2 = -1$. Thus the possible solutions for Q are

$$Q_1 = Q_0 e^{i\omega_0 t} \qquad (46a)$$

$$Q_2 = Q_0 e^{-i\omega_0 t} \qquad (46b)$$

While Equations (46a, b) are both mathematical solutions to the differential equation for the LC circuit, both are complex functions. But the amount of charge Q in the capacitor must be described by a real number. No imaginary charge resides there.

However in Equations (24) and (25) we saw that we could construct the real functions $\cos\theta$ and $\sin\theta$ from the complex functions $e^{i\theta}$ and $e^{-i\theta}$. Replacing $\theta$ by $\omega_0 t$, we have

$$\cos\omega_0 t = \frac{1}{2}\left(e^{i\omega_0 t} + e^{-i\omega_0 t}\right) \qquad (47a)$$

$$\sin\omega_0 t = \frac{i}{2}\left(e^{-i\omega_0 t} - e^{i\omega_0 t}\right) \qquad (47b)$$

One of the features of differential equations like the one for the LC circuit* is that if the equation has more than one solution, any combination of the solutions is also a solution of the equation. In our case the two solutions are $Q_1 = Q_0 e^{i\omega_0 t}$ and $Q_2 = Q_0 e^{-i\omega_0 t}$. Thus the combination

$$Q = aQ_1 + bQ_2 \qquad a,b,\text{constants} \qquad (48)$$

must also be a solution, as you will check for yourself in Exercise 2. Choosing the constants a = 1/2, b = +1/2 gives

$$Q = Q_0 \cos\omega_0 t \qquad (49a)$$

and choosing $a = -i/2$, $b = i/2$ gives

$$Q = Q_0 \sin\omega_0 t \qquad (49b)$$

These are both real functions which can describe the electric charge in the capacitor.

Thus we see that for both the RC and the LC circuit, we can use the same trial function $Q = Q_0 e^{-\alpha t}$. For the RC circuit, $\alpha$ was a real number, which gave us the exponential decay $Q = Q_0 e^{-t/RC}$. For the LC circuit, $\alpha$ turned out to be imaginary which gave us real oscillating solutions like $Q = Q_0 \cos\omega_0 t$. By using complex numbers, we are able to handle both the RC and the LC circuits with the same trial function. Whether $\alpha$ turns out to be real or imaginary tells us whether the circuit decays or oscillates.

*This is an example of what is called a ***homogenous differential equation***. We will have more to say about them shortly.

In the next section we will consider the RLC circuit, which is an LC circuit with resistance included. Experimentally we saw that such a circuit could have a decaying oscillation. When we plug the guess $Q = Q_0 e^{-\alpha t}$ into the equation for the RLC circuit, $\alpha$ will turn out in some cases to be complex, i.e., have both a real and an imaginary part. The imaginary part will describe the oscillation of the circuit while the real part will tell us how the oscillation decays. But before we get to the RLC circuit, we need to discuss a simpler way to get real solutions from complex solutions of differential equations. Before that, do Exercise 2 to see that $Q = aQ_1 + bQ_2$ is a solution of our LC equation.

---

**Exercise 2**

The differential equation for an LC circuit is

$$\frac{d^2Q}{dt^2} + \frac{Q}{LC} = 0 \qquad \text{(37) repeated}$$

This is called a *homogenous differential equation* because it contains only terms involving Q or its derivatives. An example of a *non homogeneous differential equation* will be

$$\frac{d^2Q}{dt^2} + \frac{Q}{LC} = a \sin \omega_1 t \qquad (50)$$

This will represent an LC circuit that is being forced to oscillate at some frequency $\omega_1$. The appearance of the term $(a \sin \omega_1 t)$ with no factor of Q makes this a *non homogeneous* equation. We will discuss this equation shortly, to show what effect the non homogeneous term has. For now we will limit our discussion to homogeneous equations.

You have seen that $Q_1 = ae^{i\omega_0 t}$ and $Q_2 = be^{-i\omega_0 t}$ are both solutions to Equation (37) when a and b are constants and $\omega_0 = 1/\sqrt{LC}$. Now explicitly plug in

$$Q = ae^{i\omega_0 t} + be^{-i\omega_0 t} \qquad (51)$$

into Equation (37), and show that this is a solution for any constant values of a and b. This demonstrates that any linear combination of $e^{i\omega_0 t}$ and $e^{-i\omega_0 t}$ is also a solution.

---

## A FASTER WAY TO FIND REAL SOLUTIONS

When we got the complex solutions $e^{+i\omega_0 t}$ and $e^{-i\omega_0 t}$ for the LC circuit differential equation, we were careful to construct real combinations of these complex solutions. You might think that it was lucky that we just happened to know that the combination $1/2(e^{i\omega_0 t} - e^{-i\omega_0 t})$ was the real function $\cos\omega_0 t$. You might be concerned that for some other differential equations you would not be so lucky.

Don't worry. If you find a complex solution for a homogeneous differential equation, you can simply take the real part of the complex solution and throw away the imaginary part. This works because both the real part and the imaginary part must separately be solutions of the differential equation. (You could also keep the imaginary part without the ($i$) and throw away the real part.)

To see why both the real and imaginary parts are solutions, let us write the complex solution for Q in the form

$$Q = Q_{\text{real}} + iQ_{\text{imaginary}} \qquad (52)$$

where both $Q_{\text{real}}$ and $Q_{\text{imaginary}}$ are real functions. Plugging Equation (52) into the LC differential equation gives

$$\frac{d^2Q}{dt^2} + \frac{Q}{LC} = \left[\frac{d^2Q_{\text{real}}}{dt^2} + \frac{Q_{\text{real}}}{LC}\right] + i\left[\frac{d^2Q_{\text{imaginary}}}{dt^2} + \frac{Q_{\text{imaginary}}}{LC}\right] = 0 \qquad (53)$$

Since both $Q_{\text{real}}$ and $Q_{\text{imaginary}}$ are real functions, their derivatives must also be real functions, and the quantities inside both square brackets in Equation (53) must be real. As a result the first square bracket is purely real, and the second square bracket with its factor of ($i$) must be purely imaginary. The only way you can add purely real and purely imaginary functions together to get zero is for both functions to be separately equal to zero.

That is, we must have

$$\frac{d^2Q_{real}}{dt^2} + \frac{Q_{real}}{LC} = 0 \qquad (54a)$$

$$\frac{d^2Q_{imaginary}}{dt^2} + \frac{Q_{imaginary}}{LC} = 0 \qquad (54b)$$

Equations (54) tell us that both functions $Q_{real}$ and $Q_{imaginary}$ must be solutions of the LC differential equation. If we want a real solution, we can use either $Q_{real}$ , $Q_{imaginary}$ or any linear combination of the two. A similar argument applies to the solution of any homogeneous differential equation.

As an explicit example for our LC equation, suppose we had come up with the solution

$$Q = Q_0 e^{i\omega_0 t} \qquad (55)$$

and had not noticed that $Q = Q_0 e^{-i\omega_0 t}$ was also a solution. Instead of hunting for another complex solution and then trying to find real combinations, we could just break $e^{i\omega_0 t}$ into its real and imaginary parts using $e^{i\theta} = \cos\theta + i\sin\theta$ to get

$$Q = Q_0 e^{i\omega_0 t} = Q_0 \cos\omega_0 t + iQ_0 \sin\omega_0 t \qquad (55a)$$

Then we immediately know that the real functions $Q_0\cos\omega_0 t$ and $Q_0\sin\omega_0 t$ are solutions of the LC differential equation. We can use either one or some linear combinations of the two. (Using a linear combination is equivalent to using an arbitrary phase angle, like $Q = Q_0\sin(\omega_0 t + \phi)$. See the Physics text, pages 15-17 or 16-31.)

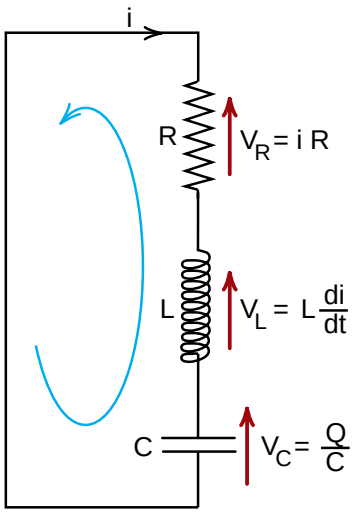



**Figure 7**
***The RLC circuit.***

## THE RLC CIRCUIT

Adding a resistor to an LC circuit gives us the RLC circuit shown in Figure (7). If the resistance R is not too large, we get a decaying oscillation like that shown in Figure (31-A9) taken from the Physics text.

The equation for the RLC circuit is obtained by setting to zero the sum of the voltage rises around the circuit, giving

$$V_R + V_L + V_C = 0 \qquad (56)$$

$$iR + L\frac{di}{dt} + \frac{Q}{C} = 0 \qquad (57)$$

Setting

$$i = \frac{dQ}{dt} \; ; \quad \frac{di}{dt} = \frac{d^2Q}{dt^2} \qquad (58)$$

and dividing through by L gives

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{Q}{LC} = 0 \qquad \textit{the LRC equation} \qquad (59)$$

As a trial function, suggested by the decaying oscillation of Figure (31-A9), we could try the solution

$$Q = Q_0 e^{-\alpha t}\cos\omega t \qquad \textit{guess} \qquad (60)$$

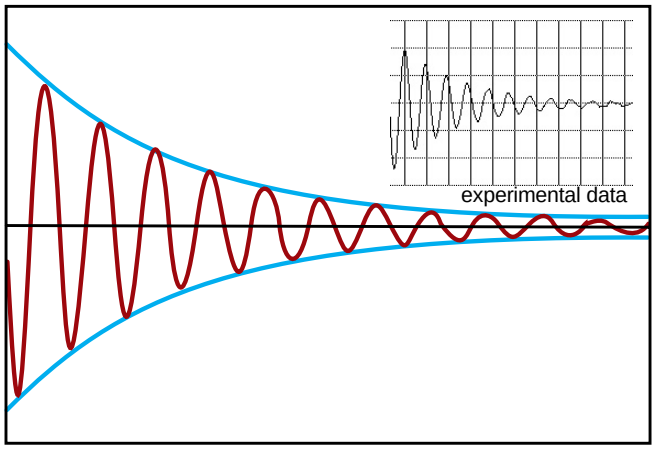



**Figure 31-A9 – Ringing like a bell**
***We hit the RLC circuit with a square wave and the circuit responded like a bell struck by a hammer. We are looking at the voltage across the capacitor.***

If you plug the guess (60) into Equation (59), you get many terms involving both $\sin\omega t$ and $\cos\omega t$.

$$Q = Q_0 e^{-\alpha t}\cos\omega t \qquad \textit{guess} \qquad (60)\text{ repeated}$$

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{Q}{LC} = 0 \qquad (59)\text{ repeated}$$

To see where the terms come from, consider

$$\frac{dQ}{dt} = Q_0(-\alpha)e^{-\alpha t}\cos\omega t + Q_0 e^{-\alpha t}(-\sin\omega t) \qquad (61)$$

where we had to differentiate the two terms $e^{-\alpha t}$ and $\cos\omega t$ separately. Differentiating again we get four terms for $d^2Q/dt^2$, two with a $\cos\omega t$ and two with a $\sin\omega t$. When we plug this all back into Equation (59), we end up with seven terms, four with $\cos\omega t$ and three with $\sin\omega t$. In order for all this to be equal to zero, you have to separately set the $\sin\omega t$ and the $\cos\omega t$ terms equal to zero. This leads to two equations, from which you can determine both the constants $\alpha$ and $\omega$. If you are careful, your chances of getting the answer without making a mistake may be as high as 50%. In other words this is the hard way to solve the problem.

---

**Exercise 3**

Try finding the coefficients $\alpha$ and $\omega$ by using Equation (60) as a trial solution for Equation (59). Then check your answer with the one we get in the next section.

---

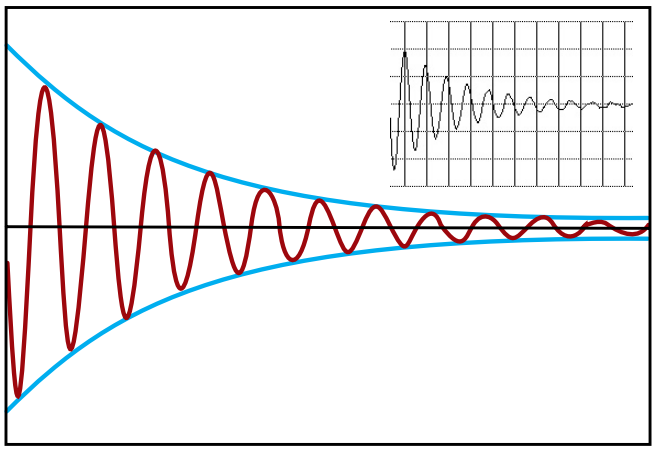


**Figure 31-A9 (repeated)**
***We are looking at the voltage across the capacitor in an RLC circuit.***

## The Easy Way

Working with separate sines and cosines is the difficult way to handle the RLC circuit. Using complex variables which provide a unified treatment of both decay and oscillation is the easy way.

For a trial solution, let us use

$$Q = Q_0 e^{-at}\ ; \quad \frac{dQ}{dt} = -aQ_0 e^{-at}$$

$$\frac{d^2Q}{dt^2} = a^2 Q_0 e^{-at} \qquad (62)$$

It looks much easier already. Substituting this trial solution into the LCR differential equation gives

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{Q}{LC} = 0 \qquad (59)\text{ repeated}$$

$$a^2Q_0e^{-at} - \frac{aR}{L}Q_0e^{-at} + \frac{Q_0e^{-at}}{LC} \doteq 0 \qquad (63)$$

The function $e^{-at}$ and constant $Q_0$ cancel and we are left with

$$a^2 - \frac{aR}{L} + \frac{1}{LC} \doteq 0 \qquad (64)$$

This is a standard quadratic equation of the form

$$x^2 + bx + c = 0 \qquad (65)$$

whose solution is

$$x = \frac{-b \pm \sqrt{b^2 - 4c}}{2} \qquad (66)$$

For our case, $-b = R/L$, $c = 1/LC$, thus (a) is given by

$$a = \frac{1}{2}\left(\frac{R}{L} \pm \sqrt{\frac{R^2}{L^2} - \frac{4}{LC}}\right)$$

$$= \frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \qquad (67a)$$

$$= \frac{R}{2L} \pm \sqrt{(-1)\left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)} \qquad (67b)$$

Setting $1/LC = \omega_0{}^2$, where $\omega_0$ is the resonant frequency of the undamped ($R = 0$) circuit, and taking $\sqrt{-1}$ outside the square root as a factor of ($i$) gives

$$a = \frac{R}{2L} \pm i\sqrt{\left(\omega_0{}^2 - \frac{R^2}{4L^2}\right)} \qquad (68)$$

We now introduce the notation

$$\alpha = \frac{R}{2L}$$
$$\omega = \sqrt{\omega_0^2 - \frac{R^2}{4L^2}} = \sqrt{\omega_0^2 - \alpha^2} \qquad (69)$$

So that

$$a = \alpha \pm i\omega \qquad (70)$$

and our trial solution $Q = Q_0 e^{-at}$ becomes

$$Q = Q_0 e^{-(\alpha \pm i\omega)t}$$

$$\boxed{Q = Q_0 e^{-\alpha t} e^{\pm i\omega t}} \qquad \textit{solution for the RLC circuit} \qquad (71)$$

where

$$\boxed{\alpha = \frac{R}{2L} \; ; \quad \omega = \sqrt{\omega_0^2 - \alpha^2}} \quad (69) \text{ repeated}$$

For the case where $\omega_0$ is bigger than $\alpha$, $\omega = \sqrt{\omega_0^2 - \alpha^2}$ is a real number, the real part of $e^{i\omega t}$ is $\cos\omega t$ and we get the real solution

$$Q_1 = Q_0 e^{-\alpha t}\cos\omega t \qquad (72)$$

The imaginary part of $e^{i\omega t}$ is proportional to $\sin\omega t$, which gives us the other real solution

$$Q_2 = Q_0 e^{-\alpha t}\sin\omega t \qquad (73)$$

As in the case of the LC circuit, the sine and cosine waves can be combined as a sine wave with an arbitrary phase angle $\phi$ to give the general solution

$$\boxed{Q = Q_0 e^{-\alpha t}\sin(\omega t + \phi)} \qquad \textit{damped oscillation of an RLC circuit} \qquad (74)$$

Equation (74) represents a damped oscillation of frequency $\omega = \sqrt{\omega_0^2 - \alpha^2}$ and a damping time constant T given by

$$\boxed{T = \frac{1}{\alpha} = \frac{2L}{R}} \qquad \textit{damping time constant} \qquad (75)$$

Imagine that we start with an RLC circuit that initially has negligible resistance, and that we gradually increase the resistance. When $R = 0$, then $\alpha = 0$ and the oscillation frequency is $\omega = \sqrt{\omega_0^2} = \omega_0$, where $\omega_0$ is the undamped frequency.

As R and $\alpha = R/2L$ are increased, the oscillation frequency $\omega = \sqrt{\omega_0^2 - \alpha^2}$ decreases until we reach $\alpha = \omega_0$. At that point, $\omega = \sqrt{\omega_0^2 - \alpha^2} = 0$, oscillation ceases, and we have what is called ***critical damping***. The time constant for decay at critical damping is just the length of time it takes the undamped circuit to go through one radian of oscillation, or $1/2\pi$ of a complete cycle. You can see that result from dimensions. We have

$$\frac{1}{\omega_0 \frac{\text{radians}}{\text{second}}} = \frac{1}{\omega_0}\frac{\text{seconds}}{\text{radian}} \qquad (76)$$

and at critical damping, where $\alpha = \omega_0$,

$$T = \frac{1}{\alpha} = \frac{1}{\omega_0}\frac{\text{seconds}}{\text{radian}} \qquad (77)$$

At critical damping, there is only one unique solution for the RLC circuit. As we increase the resistance beyond critical damping, when $\alpha = R/2L$ becomes larger than $\omega_0$, the solution becomes ***overdamped***. For $\alpha > \omega_0$, it is easiest to go back to writing the solution in the form

$$Q = Q_0 e^{-at} \qquad \text{(from Eq.62)}$$

$$a = \frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} = \alpha \pm \sqrt{\alpha^2 - \omega_0^2} \qquad \text{(from Eq.67a)}$$

and we see that we now have two exponential decay solutions

$$Q_1 = Q_0 e^{-\left(\alpha + \sqrt{\alpha^2 - \omega_0^2}\right)t} \qquad (78a)$$

$$Q_2 = Q_0 e^{-\left(\alpha - \sqrt{\alpha^2 - \omega_0^2}\right)t} \qquad (78b)$$

If we increase the resistance so much that $\omega_0^2$ is completely negligible compared to $\alpha^2$, then the two solutions become

$$Q_1 \rightarrow Q_0 e^{-2\alpha t} \qquad \alpha^2 >> \omega_0^2 \qquad (79a)$$

$$Q_2 \rightarrow Q_0 e^{0} = Q_0 \qquad (79b)$$

In this limit we easily see that the solution $Q_1$ damps more rapidly than $Q_2$. For the $Q_2$ solution, we have increased the resistance so much that no charge leaves the capacitor and the charge remains at $Q_0$.

We can get a better insight into the solution $Q_2$ by assuming that $\omega_0{}^2$ is small but not quite zero. In this case we can write $Q_2$ as

$$Q_2 = Q_0\, e^{-\alpha\left(1-\sqrt{1-\omega_0^2/\alpha^2}\right)t} \qquad (80)$$

Since $\omega_0^2/\alpha^2 << 1$, we can use the approximation formula

$$\sqrt{1-x} \approx 1 - \frac{x}{2} \qquad x << 1$$

We get, for $x = \omega_0^2/\alpha^2$,

$$\alpha\left[1-\sqrt{1-\frac{\omega_0^2}{\alpha^2}}\right] \approx \alpha\left[1-\left(1-\frac{\omega_0^2}{2\alpha^2}\right)\right]$$

$$= \alpha\left[\frac{\omega_0^2}{2\alpha^2}\right] = \frac{1}{2}\,\omega_0^2\,\frac{1}{\alpha}$$

With $\omega_0^2 = 1/LC$ and $\alpha = R/2L$, we get

$$= \frac{1}{2} \times \frac{1}{LC} \times \frac{2L}{R}$$

$$= \frac{1}{RC} \qquad (81)$$

Thus for $\alpha^2 >> \omega_0^2$ we have

$$\boxed{Q_2 = Q_0\, e^{-t/RC}} \qquad \alpha^2 >> \omega_0^2 \qquad (82)$$

This is just the solution for the decay of an RC circuit with a time constant T = RC.

The condition $\alpha^2 >> \omega_0^2$ can be written as

$$\frac{R^2}{4L^2} >> \frac{1}{LC} \quad \text{or} \quad \frac{R^2C}{4L} >> 1 \qquad (83)$$

We can meet this condition for finite values of R and C by making L small enough.

---

**Exercise 4**

To make our study of the RLC circuit more concrete, suppose that in the circuit you use a 0.10 microfarad capacitor and one millihenry inductor, so that

$L = 10^{-3}$ hy

$C = 10^{-5}$ farads

(a) What is the resonant frequency $\omega_0$ radians/second and $f_0$ cycles/second, when R = 0?

(b) What is the length of time it takes the R = 0 circuit to go through one radian of its oscillation?

(c) What value of resistance $R_C$ should you use for critical damping?

(d) What is the time constant for the decay at critical damping?

(e) Suppose you raise R from its critical value $R_C$ up to $\sqrt{2}R_C$. What are the time constants $T_1$ and $T_2$ for the decay of the solutions $Q_1$ and $Q_2$ respectively? (Partial answer: $Q_2$ takes twice as long as $Q_1$ to decay when $R = \sqrt{2}R_C$.)

---

## IMPEDANCE

Circuits commonly encountered are ***AC circuits*** where the current has a sinusoidal form

$$i = i_0 \sin \omega t \qquad (84)$$

For standard American households, the household current has a frequency of 60 cycles/second, or $\omega = 2\pi \times 60$ radians/second. In much of the rest of the world the standard household frequency is 50 cycles per second. World War II aircraft used a standard frequency of 400 cycles per second which resulted in smaller and lighter transformers.

The concept of impedance, which involves complex variables, provides an easy way to handle the voltages across R, L, and C circuit elements in an AC circuit. To demonstrate the advantage of the complex variable approach, we will first analyze these voltages using our standard real variables, and then see how much the calculations are simplified by complex variables.

Suppose we have three circuit elements, an R, L, and C, connected in series as shown in Figure (8), and run an AC current through them. In the diagram we show the formula for the voltage across each circuit element. What we wish to calculate is the total voltage V across all three elements.

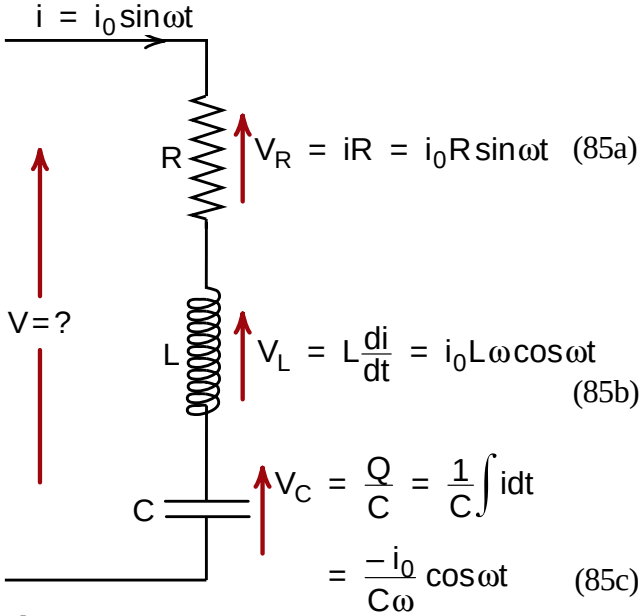



**Figure 8**
***AC voltages in the R, L, and C circuit elements.***

The individual voltages were calculated noting that

$$\frac{d \sin \omega t}{dt} = \omega \cos \omega t$$

$$\int \sin \omega t \, dt = -\frac{1}{\omega} \cos \omega t .$$

The voltage V across all three elements is just the sum of the individual voltages

$$V = i_0 \left( R \sin \omega t + L\omega \cos \omega t - \frac{1}{C\omega} \cos \omega t \right)$$

$$= i_0 \left[ R \sin \omega t + \left( L\omega - \frac{1}{\omega C} \right) \cos \omega t \right]$$

$$\boxed{V = i_0 \left[ A \sin \omega t + B \cos \omega t \right]} \qquad (86)$$

where

$$A = R; \qquad B = \left( L\omega - \frac{1}{\omega C} \right) \qquad (87)$$

We want to express the term $[A \sin \omega t + B \cos \omega t]$ as a single sine wave with an amplitude which we will call $Z_0$, and a phase angle $\phi$

$$[A \sin \omega t + B \cos \omega t] = Z_0 \sin(\omega t + \phi) \qquad (88)$$

To do this we use the trigonometric identity

$$\sin(a + b) = \cos b \sin a + \sin b \cos a$$

to write

$$\sin(\omega t + \phi) = \cos\phi \sin \omega t + \sin\phi \cos \omega t \qquad (89)$$

Multiplying through by $Z_0$ gives

$$Z_0 \sin(\omega t + \phi) = (Z_0 \cos\phi) \sin \omega t + (Z_0 \sin\phi) \cos \omega t$$

$$= A \sin \omega t + B \cos \omega t \qquad (90)$$

where

$$A = Z_0 \cos\phi ; \qquad B = Z_0 \sin\phi \qquad (91)$$

$$\frac{B}{A} = \frac{\sin\phi}{\cos\phi} = \tan\phi \qquad (92)$$

$$A^2 + B^2 = Z_0{}^2 (\cos^2\phi + \sin^2\phi) = Z_0{}^2 \qquad (93)$$

Applying Equations (91), (92), and (93) to our formula

$$V = i_0 [A\sin\omega t + B\cos\omega t] \qquad (86)$$ repeated

gives

$$\boxed{V = i_0 Z_0 \sin(\omega t + \phi)} \qquad (94)$$

where from Equations (92) and (87)

$$\boxed{\tan\phi = \frac{B}{A} = \frac{L\omega - 1/\omega C}{R}} \qquad (95)$$

and from Equation (93)

$$\boxed{Z_0^{\,2} = R^2 + \left(L\omega - \frac{1}{\omega C}\right)^2} \qquad (96)$$

After a fair amount of calculation, we see that the voltage across all three circuit elements is still proportional to $\sin\omega t$. Its amplitude $Z_0$ is given by Equation (96) and there is a phase shift by an angle $\phi$ that is given by Equation (95).

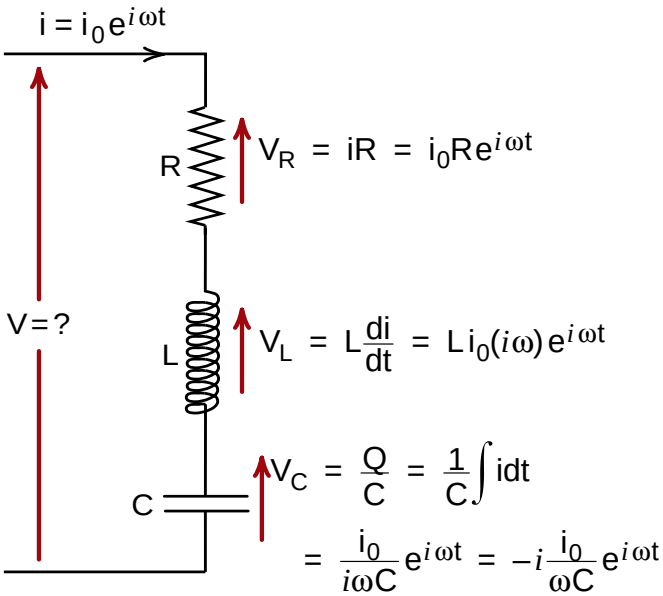



**Figure 9**
***AC voltages in the R, L, and C circuit elements, using complex notation.***

Now let us see how much more quickly we can arrive at the amplitude $Z_0$ and phase shift $\phi$ using the complex variables shown in Figure (9).

In Figure (9) we have a current $i$ given by the formula

$$i = i_0 e^{i\omega t} \qquad (97)$$

and the resulting voltage across the three circuit elements is the sum of the individual voltages which can easily be written in the form

$$V = i_0\left[R + i\left(L\omega - \frac{1}{\omega C}\right)\right]e^{i\omega t} \qquad (98)$$

The quantity in square brackets is the complex number $R + i(L\omega - 1/\omega C)$ graphed in Figure (10). It can be represented by an arrow whose length is $Z_0$ given by the Pythagorean theorem as

$$Z_0^{\,2} = R^2 + \left(L\omega - \frac{1}{\omega C}\right)^2 \qquad (99)$$

and is oriented at an angle $\phi$ whose tangent is

$$\tan\phi = \frac{L\omega - 1/\omega C}{R} \qquad (100)$$

Notice that the formulas for $Z_0$ and $\tan\phi$ are the same as in Equations (96) and (95), which we got after so much more work.

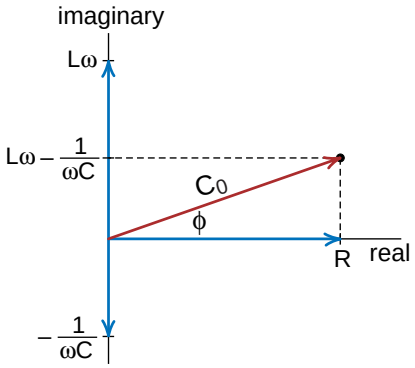



**Figure 10**
***Graph of the complex number $R + i(L\omega - 1/\omega C)$.***

From our earliest work with complex variables we saw that the complex number

$$z = x + iy \qquad (4) \text{ repeated}$$

could be written as the exponential

$$z = re^{i\theta} \qquad (26) \text{ repeated}$$

where z is graphed in Figure (2a) repeated here.

Thus the complex number $R + i(L\omega - 1/\omega C)$, graphed in Figure (10) can also be written in the exponential form

$$R + i\left(L\omega - \frac{1}{\omega C}\right) = Z_0 e^{i\theta} \qquad (101)$$

where $Z_0$ is the distance from the origin and $\phi$ the angle above the real axis.

Using Equation (100) for the square brackets in Equation (98) for the voltage V gives

$$V = i_0\left[R + i\left(L\omega - \frac{1}{\omega C}\right)\right]e^{i\omega t}$$

$$= i_0\left[Z_0 e^{i\phi}\right]e^{i\omega t}$$

$$\boxed{V = i_0 Z_0 e^{i(\omega t + \phi)}} \qquad (102)$$

$$Z_0^{\,2} = R^2 + \left(L\omega - \frac{1}{\omega C}\right)^2 \qquad (99) \text{ repeated}$$

$$\tan\phi = \frac{L\omega - 1/\omega C}{R} \qquad (100) \text{ repeated}$$

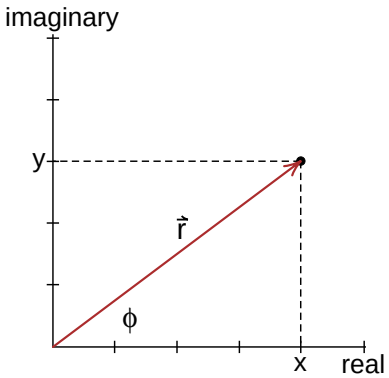



**Figure 2a (repeated)**
***Plot of the complex number (4 + 3i), showing the angle ϕ.***

Equation (102) is our complex formula for the voltage across the three circuit elements.

To find the real voltage, we simply take the real (or imaginary) part of the complex voltage. Choosing the imaginary part (without the *i*) to get a sine wave, we get

$$\boxed{V = i_0 Z_0 \sin(\omega t + \phi)} \qquad (103)$$

which is the same answer, Equation (94), that we got from the real analysis.

The main advantage of the complex analysis is that all the voltages had the same factor $e^{i\omega t}$, so that we could simply add the voltages without using the fairly messy trigonometric identities. Also note that the main result of all the work of the real analysis was to calculate the amplitude $Z_0$ and the phase angle $\phi$. We got $Z_0$ and $\phi$ immediately in the complex analysis, as soon as we graphed the complex coefficient of $e^{i\omega t}$ in Figure (10).

## Impedance Formulas

The concept of a ***complex impedance*** which we will now introduce, allows you to determine the amplitude $Z_0$ and phase angle $\phi$ by inspection, without doing hardly any calculation at all.

In Figure (11), we have redrawn our three circuit elements, introduced a complex current $i = i_0 e^{i\omega t}$, and expressed voltage in terms of i and the complex impedances $Z_R$, $Z_L$, $Z_C$ defined by

$$Z_R \equiv R \tag{104a}$$

$$Z_L \equiv i\omega L \tag{104b}$$

$$Z_C \equiv -\frac{i}{\omega C} \tag{104c}$$

In terms of these Z′s, the voltages are

$$\begin{aligned} V_R &= iZ_R \\ V_L &= iZ_L \\ V_C &= iZ_C \end{aligned} \tag{105}$$

The sum of the three voltages V becomes

$$\begin{aligned} V &= V_R + V_L + V_C \\ &= i(Z_R + Z_L + Z_C) \end{aligned} \tag{106}$$

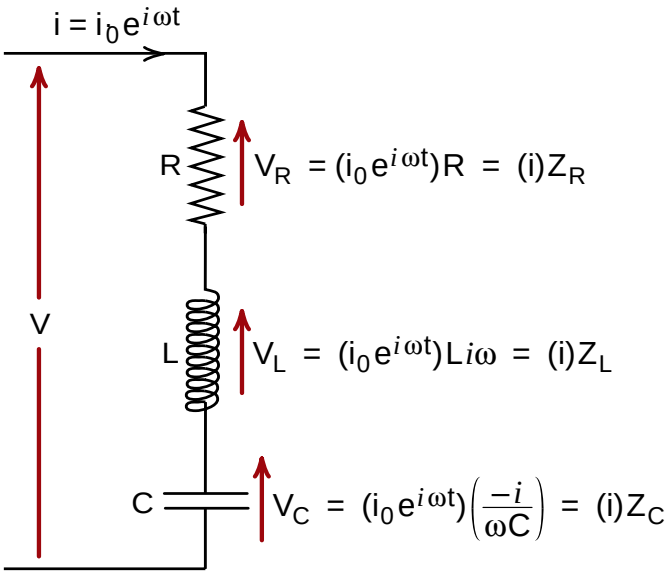



**Figure 11**

***The voltages $V_R$, $V_L$, and $V_C$ expressed in terms of impedances Z.***

If we define the total impedance Z of the three circuit elements connected in series by the equation

$$\boxed{Z = Z_R + Z_L + Z_C} \tag{107}$$

then our formula for the complex voltage is

$$\boxed{V = iZ} \tag{108}$$

Comparing this with Ohm's law for a single resistor

$$V_R = iR \qquad \textit{Ohm's law} \tag{Physics 27-1}$$

we see that we can think of Equation (108) as simply a complex form of Ohm's law.

When we graph the complex impedance Z we can immediately read off the amplitude $Z_0$ and phase angle $\phi$, as shown in Figure (12). We have

$$\boxed{Z = R + i\left(L\omega - \frac{1}{\omega C}\right) = Z_0 e^{i\phi} \quad \textit{complex impedance}} \tag{109}$$

where

$$Z_0^2 = R^2 + \left(L\omega - \frac{1}{\omega C}\right)^2 \quad \textit{magnitude of impedance} \tag{110a}$$

$$\tan\phi = \frac{L\omega - 1/\omega C}{R} \quad \textit{phase of impedance} \tag{110b}$$

In Equation (109), we introduced the exponential form $Z_0 e^{i\phi}$ for the complex variable Z.

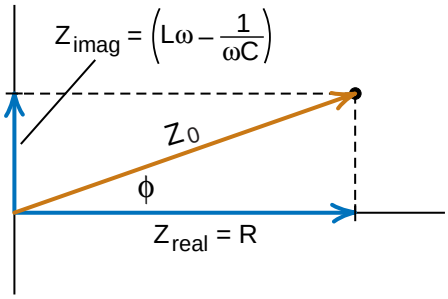



**Figure 12**

***The complex impedance can be pictured as an arrow of length $Z_0 = \sqrt{Z_{real}^2 + Z_{imag}^2}$ oriented at an angle $\phi$.***

## The Driven RLC Circuit

Our first demonstration in the physics course was the driven RLC circuit, which could be used to measure the speed of light without looking at light. (This was a crucial point in our discussion of special relativity.) In Chapter 31 we calculated the resonant frequency of an LC circuit and wrote down some formulas for the driven RLC circuit. But we did not derive the formulas because the work is messy when we have to use real functions. However with the complex analysis we have developed in this chapter, we get, almost by inspection, not only the formulas but considerable insight into the behavior of the circuit.

In the lecture demonstration, we drove the LRC circuit by wrapping a couple of turns of wire around the outside of the inductor and attaching the wire to an oscillator. The oscillating magnetic flux produced by these few turns induces a voltage $V_{ind}$ in the coil and drives the circuit to oscillate.

The important thing is that we did not put the oscillator directly in the circuit, for the oscillator has its own internal resistance, capacitance and inductance that could completely alter the behavior of the circuit. The idea is to give the circuit a gentle voltage shove of the form

$$V_{ind} = V_0 e^{i\omega t} \qquad (111)$$

as indicated in Figure (13), and see how the circuit responds.

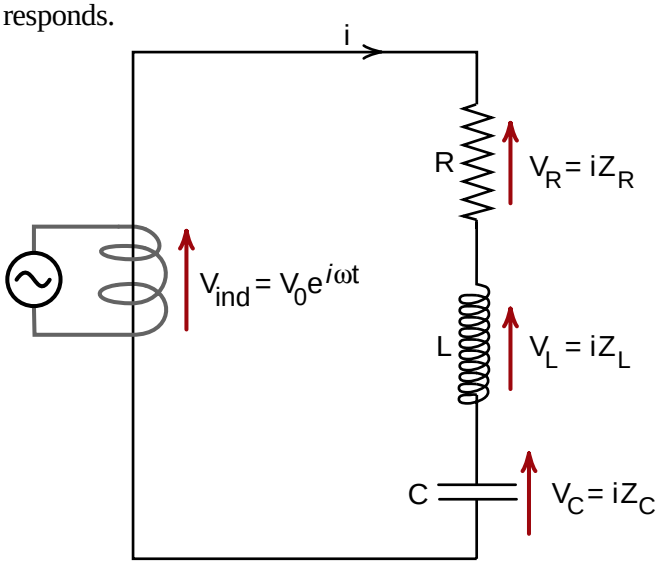
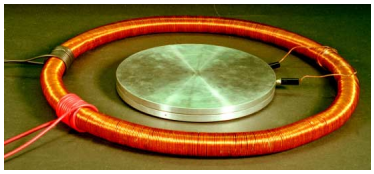



**Figure 13**
***The driven RLC circuit. Photo is Figure (1-10) from the Physics text.***



Setting the sum of the voltage rises to zero around the circuit in Figure (13) gives, (walking counter clockwise),

$$V_C + V_L + V_R - V_{ind} = 0 \qquad (112a)$$

$$iZ_R + iZ_L + iZ_C = V_0 e^{i\omega t} \qquad (112b)$$

Solving for the current $i$ in the circuit gives

$$\boxed{i = \frac{V_0 e^{i\omega t}}{Z}} \qquad (113)$$

where $Z = Z_R + Z_L + Z_C = Z_0 e^{i\phi}$ is the total impedance of the circuit.

Using the exponential form for Z in Equation (113) for the current $i$ gives

$$i = \frac{V_0 e^{i\omega t}}{Z_0 e^{i\phi}}; \qquad \boxed{i = \left(\frac{V_0}{Z_0}\right) e^{i(\omega t - \phi)}} \qquad (114)$$

Equation (114) tells us that if we drive an RLC circuit with an induced voltage $V_{ind} = V_0 e^{i\omega t}$ the circuit will respond with a current $i$ that has an amplitude $(V_0/Z_0)$ and a phase $(-\phi)$ relative to the driving voltage. We get this result almost without doing any calculation. To get the same result using real functions $\sin\omega t$ and $\cos\omega t$ would have taken several pages of algebra and trigonometric identities.

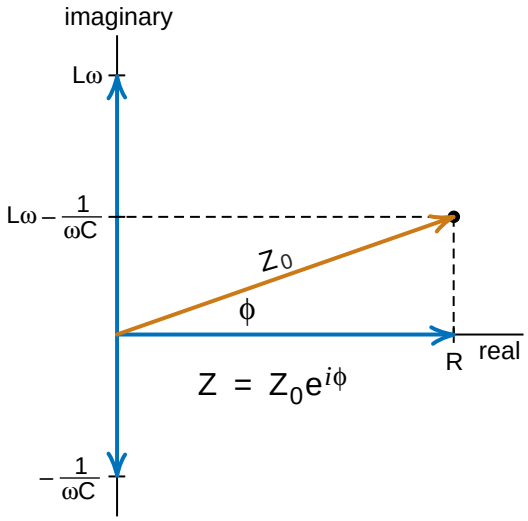



**Figure 14**
***Complex impedance for an RLC circuit.***

Let us look at the physics contained in Equation (114).

$$i = \left(\frac{V_0}{Z_0}\right) e^{i(\omega t - \phi)} \qquad \text{(114) repeated}$$

For very low frequencies, for sufficiently small $\omega$, the quantity $1/\omega C$ is much larger than either $L\omega$ or R, the impedance is essentially all capacitive as indicated in Figure (15). For this case,

$$Z_0 \approx \frac{1}{\omega C}; \quad \phi \approx -90° = -\frac{\pi}{2} \qquad (115)$$

and the formula for the current in the circuit caused by the induced voltage $V_{ind}$ is

$$i = V_0 \omega C e^{i(\omega t + \pi/2)} \qquad \textit{current at low frequencies} \qquad (116a)$$

$$V_{ind} = V_0 e^{i\omega t} \qquad \textit{complex induced voltage} \qquad (116b)$$

Taking the real part of Equations (116) gives us the real current for a real induced voltage

$$\boxed{\begin{aligned} i &= V_0 \omega C \cos(\omega t + \pi/2) \\ V_{ind} &= V_0 \cos \omega t \end{aligned}} \quad \textit{small } \omega \qquad (117)$$

From Equations (117), we see that at low frequencies, the phase of the current is $\pi/2$ ahead of the induced voltage, and the amplitude goes to zero as $\omega$ goes to zero.

The other extreme, at high frequencies where $\omega L$ is much bigger than R or $1/\omega C$, we have

$$Z_0 \approx L\omega \qquad (118)$$

$$\phi \approx +90° \quad (\pi/2) \qquad (119)$$

And we get

$$i = \frac{V_0}{L\omega} e^{i(\omega t - \pi/2)} \qquad \textit{current at high frequencies} \qquad (120)$$

Taking the real part gives us the real current

$$\boxed{\begin{aligned} i &= \frac{V_0}{L\omega} \cos(\omega t - \pi/2) \\ V_{ind} &= V_0 \cos \omega t \end{aligned}} \quad \textit{large } \omega \qquad (121)$$

We see that at high frequencies the phase of the current is $\pi/2$ behind of the induced voltage, and the amplitude goes to zero as $\omega$ goes to infinity.

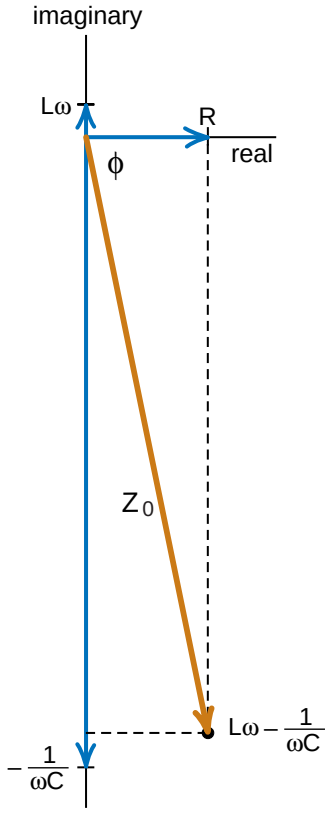



**Figure 15**
***Z for small $\omega$.***

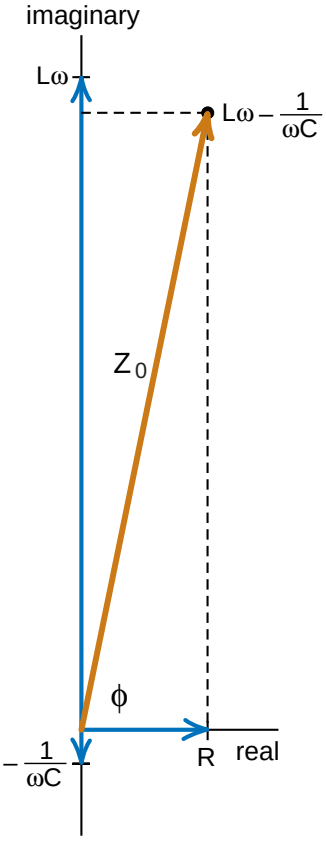



**Figure 16**
***Z for large $\omega$.***

There is a special frequency, call it $\omega_0$, where the capacitive impedance $Z_C = -1/\omega_0 C$ just cancels the inductive impedance $Z_L = L\omega_0$, leaving us with a pure resistive impedance $Z_R = R$, as shown in Figure (17).

This happens when

$$Z_L = -Z_C$$

$$\omega_0 L = +\frac{1}{\omega_0 C} \qquad (122)$$

$$\boxed{\omega_0^2 = \frac{1}{LC}} \qquad (123)$$

This special frequency is the resonant frequency $\omega_0 = 1/\sqrt{LC}$ of the RLC circuit. We now see that the resonance occurs when the capacitive and inductive impedances cancel, leaving only the resistance to dampen the current in the circuit. Also note that at this frequency the phase angle $\phi$ is zero, and the current i is given by

$$i = \frac{V_0}{R} e^{i(\omega_0 t)} \qquad \textit{current at resonance} \qquad (124)$$

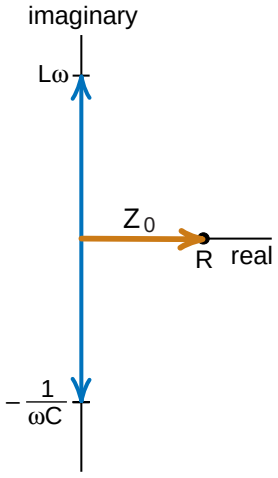



**Figure 17**
***At resonance, the capacitive and inductive impedances cancel, and we are left with only the resistive impedance.***

Taking the real part of Equation 24 gives

$$\boxed{\begin{aligned} i &= \frac{V_0}{R}\cos\omega_0 t \\ V_{ind} &= V_0 \cos\omega_0 t \end{aligned}} \qquad \textit{at resonance} \qquad (125)$$

We see that, at resonance, the current and the induced voltage are in phase with each other, and the only thing that limits the current is the actual resistance R in the circuit.

Comparing Equations (117, 121, and 125), we see that the phase of the current shifts by 180 degrees ($\pi$) as we go from well below to well above the resonance. The smaller the value of R, the sharper the resonance, and the faster this phase shift occurs. The shape of the resonance curves, for three different values of R were shown in the Physics text, Figure (14-31) repeated here.

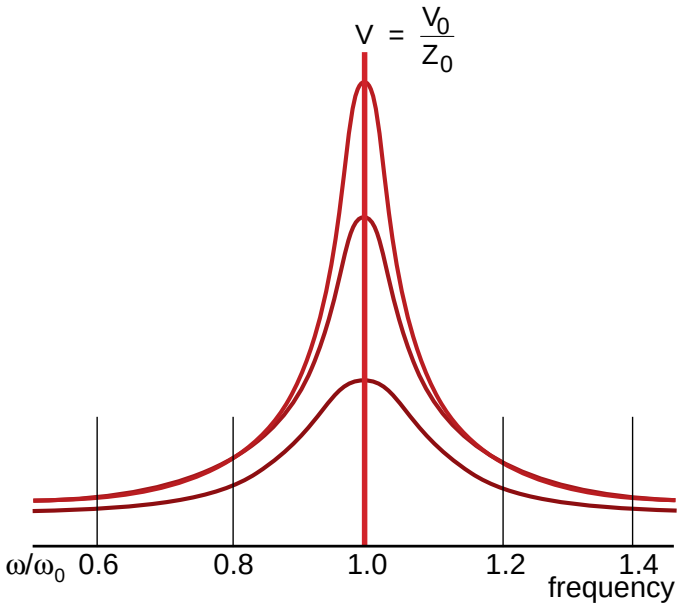



**Figure 14-31**
***Amplitude of the oscillation for various values of the resistance R. The peak occurs at $\omega = \omega_0$ because the inductive and capacitive impedances cancel at the resonant frequency $\omega_0$.***

## TRANSIENTS

While the above discussion of the driven RLC circuit describes what you most likely will see when you study the circuit in the lab, it is not the whole story. There are other solutions for the circuit, solutions which die out as time goes on, and thus are called ***transient*** solutions. To see where the transients come from, we need to go back to the differential equation for the driven circuit. We get the equation from Figure (18) which is simply Figure (13) with some labels changed. To make the circuit more nearly what we deal with in the lab, we are writing the induced voltage as a real function $V_0 \cos \omega_d t$, where we are now calling the driving frequency $\omega_d$.

### Particular Solution

Setting the sum of the voltages around the circuit equal to zero gives

$$V_R + V_L + V_C = V_{ind} \qquad (110) \text{ repeated}$$

$$iR + L\frac{di}{dt} + \frac{Q}{C} = V_0 \cos \omega_d t \qquad (126)$$

This time, let us express everything in terms of the current i rather than the charge Q, by differentiating Equation (126) with respect to time and using i = dQ/dt. We get, after dividing through by L

$$\boxed{\frac{d^2i}{dt^2} + \frac{R}{L}\frac{di}{dt} + \frac{i}{LC} = \left(\frac{-V_0\omega_d}{L}\right)\sin \omega_d t} \qquad (127)$$

where we used $d(\cos \omega_d t)/dt = -\omega_d \sin \omega_d t$.

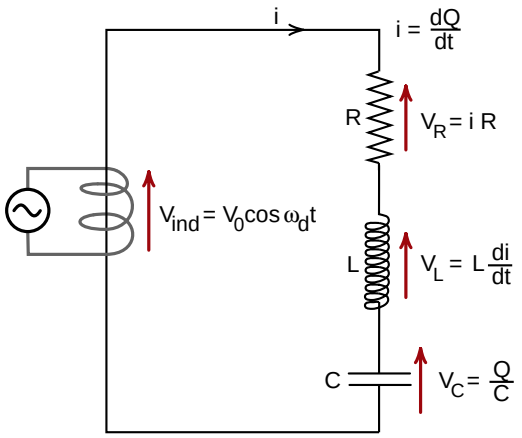



**Figure 18**
***The driven RLC circuit again.***

Equation (127) is an example of a ***non-homogeneous differential equation***. It is non-homogeneous because of the driving term $\left(-V_0\omega_d/L\right)\sin \omega_d t$ which does not have a factor of the variable (i) or a derivative of (i). This is called the ***inhomogeneous term***.

In the previous section, we found that Equation (127) has the solution

$$\boxed{i_p = \frac{V_0}{Z_0} e^{i(\omega t - \phi)}} \quad \textit{particular solution} \qquad (114) \text{ repeat}$$

where

$$Z_0^2 = R^2 + \left(L\omega - \frac{1}{\omega C}\right)^2 \qquad (99) \text{ repeated}$$

$$\tan \phi = \frac{L\omega - 1/\omega C}{R} \qquad (100) \text{ repeated}$$

The value of $i_p$ from Equation (113) is called the ***particular solution*** of the differential equation (127).

### Transient Solutions

To see what the other solutions are, let us look at the ***homogeneous differential equation***

$$\frac{d^2i}{dt^2} + \frac{R}{L}\frac{di}{dt} + \frac{i}{LC} = 0 \qquad (128)$$

which represents an RLC circuit with no driving term. I.e., it is Equation (127) without the *inhomogeneous* term.

As a review, let us see how quickly we can solve Equation (128). Using the trial solution

$$i = i_0 e^{-at}; \quad \frac{di}{dt} = -ae^{-at}; \quad \frac{d^2i}{dt^2} = a^2 e^{-at}$$

gives

$$a^2 - \frac{R}{L}a + \frac{1}{LC} = 0$$

This is a quadratic equation in a, of the form $a^2+ba+c = 0$ which has the solution

$$a = \frac{-b \pm \sqrt{b^2 - 4c}}{2} = \frac{-b}{2} \pm \sqrt{\frac{b^2}{4} - c}$$

With b = R/L and c = 1/LC we get

$$a = \frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

$$= \frac{R}{2L} \pm i\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

Thus the solution to Equation (128) is

$$\boxed{i_T = i_0\, e^{-\alpha t}\, e^{\pm i\omega t}} \qquad \begin{matrix}\textit{so called}\\ \textit{transient}\\ \textit{solution}\end{matrix} \qquad (129)$$

Where

$$\alpha = \frac{R}{2L} \;;\quad \omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

We can write $\omega$ in the form

$$\omega^2 = \omega_0{}^2 - \alpha^2$$

where

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

is the resonant frequency. Equation (129) is just Equation (71) expressed in terms of the current i rather than the charge Q. We are calling this a *transient solution* $i_T$. The reason for the name will become apparent shortly.

## Combined Solutions

Let us now go back to Equation (127) for the driven circuit, and write $i_d$ for the constant $(-V_0\omega_d/L)$ in order to simplify the equation's appearence

$$\frac{d^2i}{dt^2} + \frac{R}{L}\frac{di}{dt} + \frac{i}{LC} = i_d \sin\omega_d t \qquad (127a)$$

Now try the solution

$$i_{new} = i_p + a i_T \qquad (130)$$

where $i_p$ is the particular solution (113), $i_T$ is the transient solution of Equation (129), and (a) is an arbitrary constant. We know that

$$\frac{d^2(i_p)}{dt^2} + \frac{R}{L}\frac{d(i_p)}{dt} + \frac{(i_p)}{LC} = i_d \sin\omega_d t \qquad (131)$$

*non-homogeneous equation*

$$\frac{d^2(ai_T)}{dt^2} + \frac{R}{L}\frac{d(ai_T)}{dt} + \frac{(ai_T)}{LC} = 0 \qquad (132)$$

*homogeneous equation*

Adding Equations (131) and (132) together gives

$$\frac{d^2(i_p+ai_T)}{dt^2} + \frac{R}{L}\frac{d(i_p+ai_T)}{dt} + \frac{(i_p+ai_T)}{LC}$$

$$= i_d \sin\omega_d t \qquad (133)$$

and we see that $i_{new} = (i_p + ai_T)$ obeys the same equation as $i_p$ alone. Thus $i_{new}$ is a solution of the equation of the driven RLC circuit, for any value of the constant (a).

This result tells us that to the driven or particular solution $i_p$, we can add any amount of the homogeneous solution $i_T$, and we still have a solution for the driven RLC circuit.

The solutions $i_T$ for the homogeneous equation are fundamentally different from the particular solution $i_p$. The driven solution

$$i_p = \frac{V_0}{Z_0}\, e^{i(\omega t - \phi)} \qquad (113) \text{ repeated}$$

goes on at a constant amplitude $V_0/Z_0$ for as long as the driving voltage is attached. The transient solution

$$i_T = i_0\, e^{-\alpha t}\, e^{i\omega t} \qquad (129) \text{ repeated}$$

dies out exponentially with a time constant $T = 1/\alpha$. Because such solutions do not last, they are called *transient solutions.*

What you will observe in the lab is the following. When you first turn on or suddenly change the driving voltage $V_0 \cos\omega_d t$, you will see not only the particular solution $i_p$, but also some transients mixed in. If you wait for several time constants $T = 1/\alpha$, and keep the driving voltage amplitude $V_0$ constant, the transients will die out and the pure driven solution will appear on your oscilloscope. If you want to see the transient solutions, you have to look within a time constant $1/\alpha$ of the time you changed the driving voltage.

This finishes our discussion of the application of complex variables to the analysis of circuits. We now move on to the use of complex variables to describe wave motion.

## SOLUTIONS OF THE ONE DIMENSIONAL WAVE EQUATION

In Chapter 2 of the Calculus text we discussed the one dimensional wave equation applied to both waves on a rope and sound waves. Applied to waves on a rope, the equation was

$$\frac{\partial^2 y(x,t)}{\partial t^2} = v_{wave}\frac{\partial^2 y}{\partial x^2} \qquad (134)$$

(Calculus 2-73)

where y(x,t) represented the height of the rope above its equilibrium position at some point x along the rope at some time t. (For a sound wave, replace y(x,t) by p(x,t) where p(x,t) is the change in pressure due to the sound wave at some point x and time t.)

(Recall that when we are working with more than one variable, like x and t, we use the notation $\partial f(x,t)/\partial t$ to mean the derivative of f(x,t) with respect to t, holding x constant. This is called a ***partial derivative*** with respect to time).

We solved Equation (134) with a trial function of the form

$$y(x,t) = A\sin(kx - \omega t) \qquad (135)$$

$$\frac{\partial^2 y}{\partial x^2} = -k^2 y \; ; \quad \frac{\partial^2 y}{\partial t^2} = -\omega^2 y \qquad (136)$$

to get

$$-\omega^2 y = -v_{wave}^2 k^2 y$$

$$v_{wave}^2 = \frac{\omega^2}{k^2}$$

$$\boxed{v_{wave} = \frac{\omega}{k}} \qquad (137)$$

In the solution $\sin(kx - \omega t)$, $\omega$ is, as we have noted many times, the ***angular frequency***, of the number of radians per second. The quantity k, which is called by the rather bland name ***wave number*** is actually the ***spacial frequency*** or the number of radians per centimeter. When we take the ratio $\omega/k$ we get

$$\frac{\omega}{k}\frac{\text{radians/second}}{\text{radians/centimeter}} = \frac{\omega}{k}\frac{\text{centimeters}}{\text{second}} \qquad (138)$$

which is clearly a velocity.

As we saw in Chapter 15 of the Physics text and Chapter 2 of the Calculus text,

$$y_1 = A\sin(kx - \omega t) \quad \begin{matrix}\textit{sine wave moving}\\ \textit{to the right at a}\\ \textit{speed } v_{wave} = \omega/k\end{matrix} \qquad (139)$$

$$y_2 = A\sin(kx + \omega t) \quad \begin{matrix}\textit{sine wave moving}\\ \textit{to the left at a}\\ \textit{speed } v_{wave} = \omega/k\end{matrix} \qquad (140)$$

If we add $y_1$ and $y_2$ we get the standing wave

$$y_1 + y_2 = 2A\sin kx\cos\omega t \quad \begin{matrix}\textit{standing}\\ \textit{wave}\end{matrix} \qquad (141)$$

You can use the trigonometric identity $\sin(a + b) = \sin a\cos b + \cos a\sin b$, noting that $\sin(-b) = -\sin b$, and $\cos(-b) = \cos b$ to check Equation (141).

Rather than use the real function $\sin(kx - \omega t)$, we can, as a trial solution to the wave equation, use the complex function

$$y = Ae^{i(kx - \omega t)} \qquad (142)$$

$$\frac{\partial y}{\partial x} = ikAe^{i(kx - \omega t)}; \quad \frac{\partial y}{\partial t} = (-i\omega)Ae^{i(kx - \omega t)}$$

$$\frac{\partial^2 y}{\partial x^2} = (ik)^2 Ae^{i(kx - \omega t)} = -k^2 y$$

$$\frac{\partial^2 y}{\partial t^2} = (-i\omega)^2 Ae^{i(kx - \omega t)} = -\omega^2 y \qquad (143)$$

where $(-i)^2 = -1$.

We are now right back to Equation (136) and get the same solution $v_{wave}^2 = \omega^2/k^2$. In this case it is actually easier to work with the real function $\sin(kx - \omega t)$ rather than the complex function $e^{i(kx - \omega t)}$ because you do not have to take the real part of the complex function at the end. Working with the real variables was not difficult in this case because the wave equation did not mix up sine and cosine functions as the RLC equation did.

For completeness we have

$$y_1 = A\, e^{i(kx - \omega t)} = \begin{array}{l}\textit{complex sine wave}\\ \textit{moving to the right}\\ \textit{at a speed } \omega/k\end{array} \qquad (142)$$

$$y_2 = A\, e^{i(kx + \omega t)} = \begin{array}{l}\textit{complex sine wave}\\ \textit{moving to the left}\\ \textit{at a speed } \omega/k\end{array} \qquad (143)$$

The standing wave solution is

$$\begin{aligned} y_{standing} = y_1 + y_2 &= A\left[e^{i\,(kx - \omega t)} + e^{i\,(kx + \omega t)}\right] \\ &= A\left[e^{i\,kx}\, e^{-i\,\omega t} + e^{i\,kx}\, e^{i\,\omega t}\right] \\ &= 2A e^{i\,kx}\left[\frac{e^{-i\,\omega t} + e^{i\,\omega t}}{2}\right] \\ &= 2A e^{i\,kx} \cos\,\omega t \\ &= 2A(\cos kx + i \sin kx)\cos\,\omega t \\ &= 2A\cos kx\, \cos\omega t + i\left[2A\, \sin kx\, \cos\omega t\right] \end{aligned} \qquad (144)$$

The imaginary part of $y_{standing}$ is

$$(y_{standing})_{imag} = 2A\, \sin kx\, \cos\omega t \qquad (145)$$

which is the standing wave solution we got using real variables. Using complex variables to get the standing wave solution was not easier than using real variables.

# Calculus 2000-Chapter 6

# Introduction to the Schrödinger Wave Equation

*In the introduction to Chapter 37 of the Physics text, we quoted the following story from an address by Felix Block to the American Physical Society in 1976.*

*"Once at the end of a colloquium I heard Debye saying something like: 'Schrödinger, you are not working right now on very important problems...why don't you tell us some time about that thesis of de Broglie, which seems to have attracted some attention?' So in one of the next colloquia, Schrödinger gave a beautifully clear account of how de Broglie associated a wave with a particle, and how he could obtain the quantization rules ... by demanding that an integer number of waves should be fitted along a stationary orbit. When he had finished, Debye casually remarked that he thought this way of talking was rather childish ... To deal properly with waves, one had to have a wave equation."*

*As we mentioned, Schrödinger took Debye's advice, and in the following months devised a wave equation for the electron wave, an equation from which one could calculate the electron energy levels. That wave equation is now the foundation of chemistry.*

*In this chapter we sketch the ideas that led Schrödinger to formulate an equation involving complex variables to describe the electron. We then go on to solve that equation for the lowest energy spherically symmetric wave functions for the electron in a hydrogen atom. This is enough to show that the Schrödinger equation, without any extra assumptions, is enough to explain the quantized energy levels of hydrogen.*

## SCHRÖDINGER'S WAVE EQUATION

Schrödinger's approach to finding a wave equation for the electron was roughly as follows.

De Broglie, suspecting that the electron, like the photon, had a wave nature as well as a particle nature, went back to Einstein's formula for the energy of a photon

$$\boxed{E = hf} \qquad (1)$$

where h is Planck's constant, ( $f = c/\lambda$ ) the frequency of the photon and $\lambda$ its wavelength. Setting $E = mc^2$ where m is the mass of the photon gives

$$mc^2 = hf = h\frac{c}{\lambda} \; ; \quad m = \frac{h}{\lambda c}$$

Since photons travel at the speed c, the photon's momentum p should be its mass m times its speed c, or

$$p = mc = \left(\frac{h}{\lambda c}\right)c$$

$$\boxed{p = \frac{h}{\lambda}} \qquad (2)$$

Equation (2) is the famous de Broglie formula for the relationship between the wavelength and momentum of any particle. De Broglie explained the quantization of angular momentum in the Bohr theory by assuming that the allowed Bohr orbits were those in which exactly an integral number of wavelengths fit around the orbit.

Schrödinger's job was to find a wave equation based on the two fundamental relationships $E = hf$ for the particle energy and $p = h/\lambda$ for the particle wavelength. Because we have been writing wave equations in terms of the angular frequency $\omega$ radians/second rather than the regular frequency f cycles/second, and the wave number (spacial frequency) k radians/cm rather than the wavelength $\lambda$ cm/cycle, let us first re-express E and p in terms of $\omega$ and k rather than f and $\lambda$. Using dimensions we have

$$f\frac{\text{cycles}}{\text{second}} = \frac{\omega \text{ radians/sec}}{2\pi \text{ radians/cycle}} = \frac{\omega \text{ cycles}}{2\pi \text{ second}}$$

$$\frac{1}{\lambda \text{ cm/cycle}} = \frac{1}{\lambda}\frac{\text{cycles}}{\text{cm}} = \frac{k}{2\pi}\frac{\text{radians/cm}}{\text{radians/cycle}}$$

$$= \frac{k}{2\pi}\frac{\text{cycles}}{\text{cm}} \qquad (3)$$

Using the standard notation

$$\hbar \equiv \frac{h}{2\pi} \qquad h \text{ "bar"} \qquad (4)$$

we get

$$E = hf = h\frac{\omega}{2\pi} = \hbar\omega$$

$$p = \frac{h}{\lambda} = h\frac{k}{2\pi} = \hbar k$$

Thus we get the very simple formulas

$$\boxed{E = \hbar\omega \; ; \quad p = \hbar k} \qquad (5)$$

as the relationship between a particle's energy E and momentum p, and its wave's frequency $\omega$ and wave number k.

Schrödinger's first attempt at finding a wave equation was to start with the relativistic relationship between the energy and momentum of a particle. That relationship, as we saw in the section on particle accelerators, page 28-24 of the Physics text, is

$$E^2 = p^2c^2 + m_0{}^2c^4 \qquad \textit{relativistic relationship between E and p} \qquad (6)$$

where $m_0$ is the rest mass of the particle.

To see how to construct a wave equation, let us start with the simple case of a zero rest mass particle, namely the photon. For the photon, we have simply

$$E^2 = p^2c^2 \qquad \textit{zero rest mass particle} \qquad (7)$$

We will see that the one dimensional wave equation that leads to Equation (7) is

$$\frac{\partial^2\psi}{\partial t^2} = c^2\frac{\partial^2\psi}{\partial x^2} \qquad (8)$$

where $\psi$ (psi) is a Greek letter to represent the wave amplitude. (For rubber rope waves $\psi = y$ , the wave height. For sound waves $\psi = p$ , the excess pressure.) To check that Equation (8) is the correct equation, use the trial function

$$\psi = \psi_0 e^{i(kx - \omega t)} \qquad (9)$$

which, as we saw at the end of the last chapter (see Equation 5-142), represents a wave travelling to the right at a speed $\omega/k$.

We have

$$\psi = \psi_0 e^{i(kx - \omega t)}$$

$$\frac{\partial \psi}{\partial t} = -i\omega\psi \; ; \quad \frac{\partial^2 \psi}{\partial t^2} = (-i\omega)^2\psi = -\omega^2\psi$$

$$\frac{\partial \psi}{\partial x} = -ik\psi \; ; \quad \frac{\partial^2 \psi}{\partial x^2} = (-ik)^2\psi = -k^2\psi$$

Plugging these values into Equation (8) gives

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \frac{\partial^2 \psi}{\partial x^2} \qquad \text{(8) repeated}$$

$$-\omega^2\psi = c^2(-k^2)\psi$$

The factor $-\psi$ cancels and we get

$$\omega^2 = c^2k^2 \qquad (10)$$

Multiply through by $\hbar^2$ and noting that $E = \hbar\omega$ and $p = \hbar k$ we get

$$\hbar^2\omega^2 = c^2(\hbar^2k^2)$$

$$E^2 = c^2p^2 \qquad (11)$$

which is the result we wanted.

---

**Exercise 1**

For a traveling wave, use the trial function

$$\psi = \psi_0 \sin(kx - \omega t)$$

and show that you get the same result.

---

You can see that the process is quite straightforward. For each factor of $\omega$ you want from your differential equation, you put a $\partial/\partial t$ into the equation. For each factor of (k), you include a $\partial/\partial x$ .

If we set $\psi = \vec{E}$ or $\vec{B}$ in Equation (8) we get the wave equations

$$\frac{\partial^2 \vec{E}}{\partial t^2} = c^2 \frac{\partial^2 \vec{E}}{\partial x^2} \qquad (12a)$$

$$\frac{\partial^2 \vec{B}}{\partial t^2} = c^2 \frac{\partial^2 \vec{B}}{\partial x^2} \qquad (12b)$$

These turn out to be the differential form (in one dimension) of the electromagnetic wave we discussed in Chapter 32 in the Physics text. (These are Equations (24a) and (24b) of Chapter 9 of the Calculus text, if we set $c^2 = 1/\mu_0\varepsilon_0$ .) This should not be surprising, because an electromagnetic wave just represents the wave nature for the zero rest mass photon.

Now that we have some experience constructing wave equations, let us go for the equation for a particle with rest mass. This time let us first convert the relationship between the particle energy E and momentum p into a relationship between $\omega$ and k. We have

$$E^2 = p^2c^2 + m_0{}^2c^4$$

Setting $E = \hbar\omega$ and $p = \hbar k$ gives

$$\hbar^2\omega^2 = \hbar^2k^2c^2 + m_0{}^2c^4$$

Dividing through by $\hbar^2$ gives

$$\omega^2 = c^2k^2 + \frac{m_0{}^2c^4}{\hbar^2} \qquad (13)$$

Using a $\partial/\partial t$ for each $\omega$ and a $\partial/\partial x$ for each k suggests the wave equation

$$\boxed{\frac{\partial^2 \psi}{\partial t^2} = c^2 \frac{\partial^2 \psi}{\partial x^2} - \frac{m_0{}^2c^4}{\hbar^2}\psi} \qquad (14)$$

Plugging in the trial solution

$$\psi = \psi_0 e^{i(kx - \omega t)}$$

$$\frac{\partial^2 \psi}{\partial t^2} = -\omega^2\psi \; ; \quad \frac{\partial^2 \psi}{\partial x^2} = -k^2\psi$$

gives

$$-\omega^2\psi = -c^2k^2\psi - \frac{m_0{}^2c^4}{\hbar^2}\psi \qquad (15)$$

cancelling the factor of $-\psi$ gives

$$\omega^2 = c^2k^2 + \frac{m_0{}^2c^4}{\hbar^2} \qquad (16)$$

which is the result we wanted.

Equation (14) is the one dimensional form of ***Schrödinger's relativistic wave equation***. This is the first wave equation Schrödinger found, but he ran into trouble with it.

Consider the case of a particle at rest, or nearly at rest, so that we can neglect $p^2c^2$ compared to $m_0{}^2c^4$. Then the square of the energy E is approximately equal to the square of the rest energy $m_0c^2$

$$E^2 \approx m_0{}^2c^4 \qquad \textit{for small p} \tag{17}$$

This equation has two solutions

$$\begin{aligned} E_1 &= m_0c^2 \\ E_2 &= -m_0c^2 \end{aligned} \tag{18}$$

Solution (2) appears to represent a particle with a negative rest energy, a very un-physical thing. The corresponding wave solutions are

$$\psi_1 = \psi_0 e^{i(kx-\omega_1 t)} \; ; \quad \hbar\omega_1 = E_1 \tag{19}$$

$$\psi_2 = \psi_0 e^{i(kx-\omega_2 t)} \; ; \quad \hbar\omega_2 = E_2 \tag{20}$$

When you encounter two solutions to a physical problem, and one is nonsense, you usually throw the bad solution out. For example, the hypotenuse of a right triangle is given by the equation

$$c^2 = a^2 + b^2 \tag{21}$$

which has two solutions

$$c_1 = +\sqrt{a^2+b^2} \tag{22}$$

$$c_2 = -\sqrt{a^2+b^2} \tag{23}$$

Since you know that you cannot have a negative hypotenuse, you just throw out the un-physical solution $c_2$.

Schrödinger tried to throw out the un-physical solution $\psi_2$ of his relativistic wave equation, but ran into the following problem. If he started with pure $\psi_1$ waves for the electrons, and let the electrons interact, $\psi_2$ waves were generated. In other words, if he threw out the un-physical $\psi_2$ waves, the equations put them back in. We did not have this problem with the Pythagorean theorem.

Schrödinger gave up on the relativistic wave equation and decided to use the nonrelativistic relationship between the kinetic energy E and momentum p of a slowly moving particle.

That relationship is

$$\underset{\text{energy}}{\text{kinetic}}\ E = \frac{1}{2}mv^2 = \frac{1}{2m}(m^2v^2) \tag{24}$$

where v is the speed of the particle, m the rest mass, and mv = p is the momentum. Thus E and p are related nonrelativistically by

$$E = \frac{(mv)^2}{2m} = \frac{p^2}{2m} \tag{25}$$

Writing $E = \hbar\omega$, $p = \hbar k$, the nonrelativistic relationship between $\omega$ and k is

$$\boxed{\hbar\omega = \frac{\hbar^2k^2}{2m}} \qquad \textit{nonrelativistic relationship between } \omega \textit{ and } k \tag{26}$$

Schrödinger went to the nonrelativistic form because the relationship $E = p^2/2m$ does not involve negative rest masses.

To construct a wave equation that gives this nonrelativistic relationship between $\omega$ and k, we need one time derivative to give the one factor of $\omega$, and two x derivatives to give the factor of $k^2$. What works, as we will check, is

$$\boxed{i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}} \qquad \textit{one dimensional Schrödinger's equation for a free electron} \tag{27}$$

With the trial solution

$$\psi = \psi_0 e^{i(kx-\omega t)}$$

$$\frac{\partial\psi}{\partial t} = -i\omega\psi \; ; \quad \frac{\partial^2\psi}{\partial x^2} = -k^2\psi \tag{28}$$

we get

$$i\hbar(-i\omega\psi) = \frac{\hbar^2}{2m}k^2\psi$$

$$-i^2\omega\hbar\psi = \frac{\hbar^2}{2m}k^2\psi \tag{29}$$

The $\psi$'s cancel, and with $-i^2 = 1$, we are left with the desired result

$$\hbar\omega = \frac{\hbar^2k^2}{2m} \qquad \text{(26) repeated}$$

Equation (27) is the one dimensional form of Schrödinger's equation for a free particle.

In Chapter 2 of the Calculus text, we saw that the equations for rope waves, sound waves, and electromagnetic waves all had second derivatives of both space and time. That is how we got the oscillating solutions. In our study of the RLC circuit, we saw that the presence of a first derivative, the R term in

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{Q}{LC} = 0 \qquad \text{(5-59) repeated}$$

led to an exponential decay.

One might wonder, since there is only a first derivative with respect to time in Schrödinger's equation, shouldn't that lead to an exponential decay with time, of the wave amplitude $\psi$? It did not do so because of the explicit factor of ($i$) in Schrödinger's equation. With the trial solution $\psi = \psi_0 e^{i(kx - \omega t)}$ the ($-i$) from the first derivative with respect to time was turned into a 1 by the $i$ in the $\partial/\partial t$ term. Thus by having an ($i$) in Schrödinger's equation itself, we can get an oscillating solution with a first time derivative.

The reason we have introduced Schrödinger's equation after a chapter on complex variables is that factor of ($i$) in the equation itself. With the other differential equations we have discussed so far, we had the choice of using real or complex variables. But we cannot write, let alone solve, Schrödinger's equation without the use of complex variables.

---

**Exercise 2**

In three dimensions, the momentum vector $\vec{p} = (p_x, p_y, p_z)$ has a magnitude p given by the Pythagorean theorem as

$$p^2 = (p_x^2 + p_y^2 + p_z^2) \qquad (30)$$

With $\vec{p} = \hbar\vec{k}$, we have

$$p^2 = \hbar^2(k_x^2 + k_y^2 + k_z^2) \qquad (31)$$

We got the one dimensional wave equation by replacing $k_x^2$ by $\partial^2/\partial x^2$. This suggests that the extension of Equation (27) to describe three dimensional plane waves should be

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right) \qquad (32)$$

As a trial solution, try the guess

$$\psi = e^{i(\vec{k}\cdot\vec{x} - \omega t)} = e^{i(k_x x + k_y y + k_z z - \omega t)} \qquad (33)$$

and show that the guess implies

$$\hbar\omega = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2) \qquad (34)$$

and

$$E = \frac{p^2}{2m}$$

---

## POTENTIAL ENERGY & SCHRÖDINGER'S EQUATION

The relationship $E = p^2/2m = mv^2/2$ is for a free particle traveling at a constant speed v. If the particle has a potential energy $V(\vec{x})$, like spring potential energy

$$\boxed{V(x) = -\frac{1}{2}Kx^2} \qquad \textit{spring potential energy} \qquad (35)$$

where K is the spring constant, then the formula for the total nonrelativistic energy E is

$$E = \frac{1}{2}mv^2 + V(x) = \frac{p^2}{2m} + V(x) \qquad (36)$$

In terms of $\omega$ and k we have

$$\hbar\omega = \frac{\hbar^2k^2}{2m} + V(x) \qquad (37)$$

and the corresponding one dimensional wave equation should be

$$\boxed{i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi} \qquad \textit{one dimensional Schrödinger equation} \qquad (38)$$

If you did Exercise (2), it is clear that the three dimensional form of Schrödinger's equation is expected to be

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right) + V(x,y,z)\psi \qquad (39)$$

In Chapter 4 of the Calculus text, we discussed the combination of derivatives $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ and gave them the special name

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \qquad \textit{definition of } \nabla^2 \qquad (40)$$

With this notation, the three dimensional form of Schrödinger's equation can be written in the more compact and familiar form

$$\boxed{i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + V(x,y,z)\psi} \qquad \textit{full Schrödinger equation} \qquad (41)$$

We can immediately get back to the one dimensional Schrödinger's equation by replacing $\nabla^2$ by $\partial^2/\partial x^2$ .

## THE HYDROGEN ATOM

The reason Schrödinger developed his wave equation was to handle the electron waves in hydrogen in a mathematically rigorous way. To apply Schrödinger's equation of the hydrogen atom, you use the fact that the electron is bound to the proton nucleus by a Coulomb force of magnitude $e^2/r^2$ whose potential V(r) is

$$V(r) = \frac{-e^2}{r} \qquad \textit{Coulomb potential energy} \qquad (42)$$

With this potential energy, Schrödinger's equation (41) for the hydrogen atom becomes

$$\boxed{\frac{\hbar}{i}\frac{\partial\psi}{\partial t} = \frac{\hbar^2}{2m}\nabla^2\psi - \frac{e^2}{r}\psi} \qquad \textit{Schrödinger's equation for hydrogen atom} \qquad (43)$$

Solving Equation (43) is not easy. The first problem we encounter is the fact that we have been writing $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ using Cartesian coordinates x, y, z, while the Coulomb potential $-e^2/r$ has spherical symmetry. The best way to handle the situation is to use a coordinate system that has the same symmetry as the potential energy.

The coordinate system of choice is the *spherical polar coordinate system* that has an inherent spherical symmetry. This coordinate system is described in Chapter 4 of the Calculus text and indicated in Figure (1). Instead of locating a point by giving its x, y, and z coordinates, we locate it by the r, $\theta$ and $\phi$ coordinates. The quantity r is the distance from the origin, $\theta$ the angle down from the z axis, and $\phi$ the angle over from the x axis, as shown.

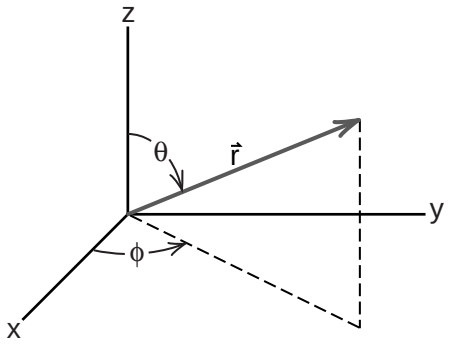



**Figure 1**
***Spherical polar coordinates.***

In the appendix to Chapter 4 of the Calculus text, we calculated $\nabla^2$ in spherical polar coordinates. The result was

$$\nabla^2\psi = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin\theta}\frac{\partial^2\psi}{\partial\phi^2} \qquad (44)$$

This surely does not look simpler than $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$, but it does allow you to find solutions to Schrödinger's equation for the hydrogen atom.

In the appendix to this chapter, we calculate some spherically symmetric solutions to Schrödinger's equations. These are solutions that depend only on r, namely $\psi = \psi(r)$, so that $\partial\psi/\partial\theta = 0$ and $\partial\psi/\partial\phi = 0$, which eliminates the second and third terms in Equation (44). The solutions we get, (we solve one and leave the second as a homework exercise) are

$$\boxed{\psi_1 = e^{-r/a_0}\, e^{-i\omega_1 t}} \qquad (45)$$

$$\boxed{\psi_2 = \left(1 - \frac{r}{2a_0}\right) e^{-r/2a_0}\, e^{-i\omega_2 t}} \qquad (46)$$

where $a_0$ has the value

$$a_0 = \frac{\hbar^2}{me^2} \qquad \textit{Bohr radius} \qquad (47)$$

This quantity $a_0$ is the ***Bohr radius***, the radius of the smallest orbit in the Bohr theory of hydrogen. (See Exercise 7 in Chapter 35 of the Physics text.)

---

**Exercise 3**

Go to Appendix II of this chapter (page 6-14) and study the steps that led to the solution $\psi_1$. Then work Exercise 5 to find the solution $\psi_2$. After that return here and continue reading.

---

A special feature we discover when we solve Schrödinger's equation in Appendix II, is that in order for $\psi_1$ and $\psi_2$ to be solutions of Schrödinger's equation (43), ***the frequencies*** $\omega_1$ ***and*** $\omega_2$ ***have to have the following values***

$$\hbar\omega_1 = \frac{-e^4 m}{2\hbar^2} = -13.6\,\text{eV} \qquad (48)$$

$$\hbar\omega_2 = \frac{-e^4 m}{8\hbar^2} = -3.60\,\text{eV} \qquad (49)$$

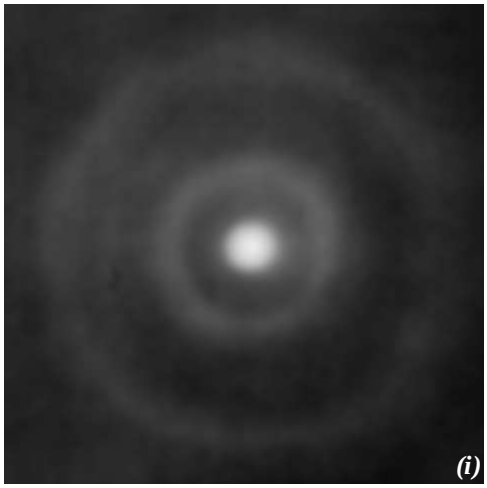
You can immediately see that $\hbar\omega_1$ is the energy of the electron in the lowest hydrogen energy level, and $\hbar\omega_2$ is the electron energy in the second energy level. Just looking at the spherically symmetric solutions begins to tell us that Schrödinger's equation is going to explain, in a natural way, the hydrogen energy levels.

As we mentioned in our discussion of the hydrogen atom in Chapter 38 of the Physics text, there are many allowed standing wave patterns for the electron in hydrogen. In Figure (38-1), reproduced on the next page, we show sketches of the six lowest energy patterns $\psi_{n,\ell,m}$ labeled by their energy quantum number (n), angular momentum quantum number ($\ell$) and z projection of angular momentum quantum number (m). We noted that all the zero angular momentum patterns ($\ell = 0$) are spherically symmetric. By solving Schrödinger's equation for spherically symmetric standing waves, we began to generate the $\ell = 0$ patterns. Explicitly, the waves we got are

$$\psi_{1,0,0} = \psi_1 \text{ (}\textit{of Equation}\text{ 45)}$$

$$\psi_{2,0,0} = \psi_2 \text{ (}\textit{of Equation}\text{ 46)}$$

To solve for the non symmetric patterns like $\psi_{2,1,1}$ that have angular momentum, you have to be able to handle angular terms involving $\theta$ and $\phi$ in the formula (44) for $\nabla^2$. Differential equations involving $\nabla^2$ have been studied for well over a century, and the angular terms, which are common to many of these equations, have been carefully worked out with standardized notation. The angular dependence of the non spherical standing waves involve what are called ***spherical harmonics*** which are briefly discussed in Appendix II of this chapter.



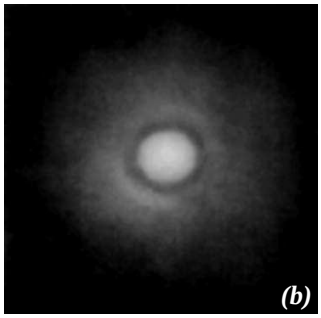

*E = –1.51eV*

***n = 3, ℓ = 0, m = 0***

There are 8 more n = 3 patterns in addition to the one shown. The ℓ and m quantum numbers are

ℓ = 1; m = 1, 0, –1

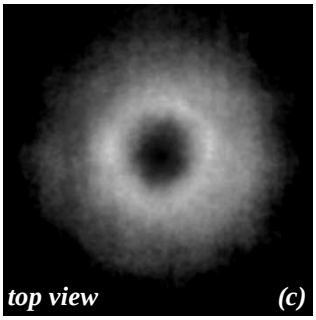
ℓ = 2; m = 2, 1, 0, –1, –2.

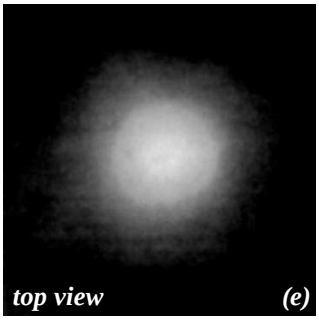


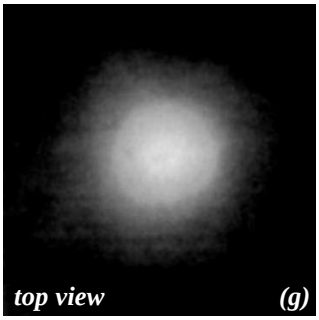

*E = –3.40eV*

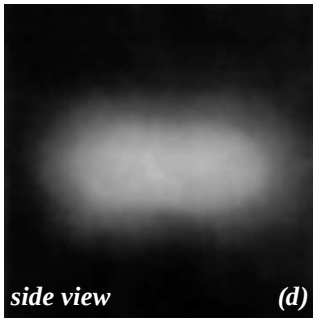
***n = 2, ℓ = 0, m = 0***

$$\psi_2(r) = \left(1 - \frac{r}{2a_0}\right)e^{-r/2a_0}$$

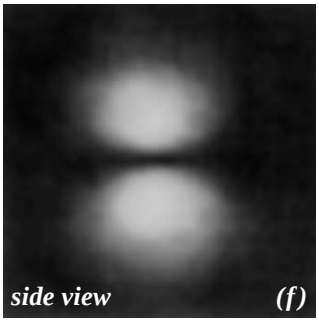


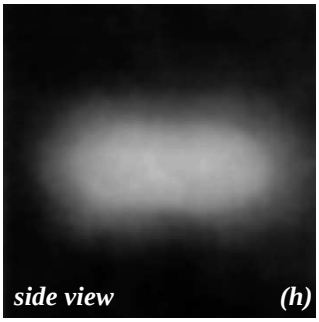



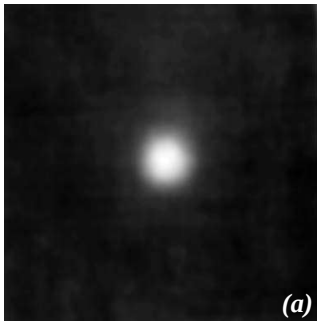

***n = 2, ℓ = 1, m = 1***




***n = 2, ℓ = 1, m = 0***




***n = 2, ℓ = 1, m = –1***




*E = –13.6eV*

***n = 1, ℓ = 0, m = 0***

$$\psi_1(r) = e^{-r/a_0}$$

**Figure 38-1 (page 38-3 of the Physics text)**
***The lowest energy standing wave patterns in hydrogen. The intensity is what you would see looking through the wave. We have labeled $\psi_1$ and $\psi_2$ on the diagram.***

## INTERPRETATION OF SOLUTIONS TO SCHRÖDINGER'S EQUATION

Bohr's theory of the hydrogen atom, although quite successful, was based on Newtonian mechanics with the ad hoc assumption that angular momentum was quantized in units of $\hbar$. De Broglie's theory suggested that the reason for the quantization of angular momentum was due to the wave nature of the electron, but he also treated the electron wave in a rather ad hoc manner. If one assumes that Schrödinger's equation rather than Newtonian mechanics provides the basic theory for the electron in hydrogen, then all the quantized energy levels follow a direct consequence of the theory. No extra assumptions have to be fed in. Schrödinger had found the theory to replace Newtonian mechanics in describing atoms.

But questions remained. The electron's wave nature was well established, but what was the meaning of the electron wave? The answer to that was provided a couple of years later by Max Born, who was calculating how electron waves would be scattered by atoms. The calculations suggested to him that the electron wave should be interpreted as a probability wave, as we discussed in Chapter 40 of the Physics text.

One of the main features of a probability wave is that it has to be represented by a real, positive number. You cannot have negative probabilities or imaginary probabilities. But so far, our electron waves are described by a complex variable $\psi$, obtained from an equation that was itself complex. How do we get real positive numbers from the complex $\psi$?

We ran into a somewhat similar problem in our discussion of electromagnetic radiation. Maxwell's equations predict that light waves consist of electric and magnetic fields $\vec{E}$ and $\vec{B}$. Yet most of the time we are concerned with the intensity or energy density of a light wave. To predict the intensity from Maxwell's theory, we have to know how to calculate the intensity from the vectors $\vec{E}$ and $\vec{B}$. The answer is that the intensity is proportional to the square of $\vec{E}$ and $\vec{B}$. If we use the correct units, the intensity is proportional to $(\vec{E} \cdot \vec{E} + \vec{B} \cdot \vec{B})$. These dot products $\vec{E} \cdot \vec{E}$ and $\vec{B} \cdot \vec{B}$ are always positive numbers and therefore can represent an energy density or intensity.

If we can get a positive number for a vector field by taking the dot product of the vector with itself, what do we do to get a positive number from a complex $\psi$? The answer, as we mentioned at the beginning of Chapter 5 (see Equation 5-26), is that we get a real positive number from a complex number by multiplying by the ***complex conjugate***. To remind you how this works, suppose that we have separated $\psi$ into its real and imaginary parts

$$\psi = \psi_{\text{real}} + i\psi_{\text{imag}} \tag{50}$$

where both $\psi_{\text{real}}$ and $\psi_{\text{imag}}$ are real numbers. Then the complex conjugate, which we designate by $\psi^*$, is defined by changing ($i$) to ($-i$)

$$\psi^* = \psi_{\text{real}} - i\psi_{\text{imag}} \tag{51}$$

To calculate the complex conjugate $\psi^*$ you do not have to separate the function into real and imaginary parts ahead of time. You get the same result by replacing all ($i$) by ($-i$) in the complex formula.

When you multiply a complex number $\psi$ by its complex conjugate $\psi^*$, the result is a real positive number, as you can see below

$$\begin{aligned}\psi^*\psi &= (\psi_{\text{real}} - i\psi_{\text{imag}})(\psi_{\text{real}} + i\psi_{\text{imag}}) \\ &= \psi_{\text{real}}\psi_{\text{real}} + i\psi_{\text{real}}\psi_{\text{imag}} \\ &\quad - i\psi_{\text{imag}}\psi_{\text{real}} - i^2\psi_{\text{imag}}\psi_{\text{imag}}\end{aligned}$$

The $i\psi_{\text{real}}\psi_{\text{imag}}$ terms cancel, and with $-i^2 = 1$ we get

$$\boxed{\psi^*\psi = \psi_{\text{real}}^2 + \psi_{\text{imag}}^2} \tag{52}$$

and thus $\psi^*\psi$ is a real, positive number.

For electron waves, the positive number $\psi^*\psi$ represents the intensity of the wave in much the same way that $(\vec{E} \cdot \vec{E} + \vec{B} \cdot \vec{B})$ represented the intensity of the electromagnetic wave.

## Normalization

In describing probabilities, one usually represents a probability of 1 as being certainty, and that the probability of an event as being allowed to range from zero to one. If the wave function $\psi$ is to represent a probability wave for an electron, we have to include the idea that the probability of something ranges from zero to one.

The intensity $\psi^*\psi$ is a density that varies over space. If you have an energy density, call it $E$, then the total energy E is the integral over all of space of the energy density $E$. We can write this symbolically as

$$E = \int_{\text{all space}} E(x,y,z)\,d^3V \qquad (53)$$

where, if we are using Cartesian coordinates, the volume element $d^3V$ would be $(dx \times dy \times dz)$.

If we are to interpret $\psi^*\psi$ as a probability density, then the total probability should be the integral of the probability density over all space. We can write this as

$$\text{total probability} = \int_{\text{all space}} \psi^*\psi\, d^3V \qquad (54)$$

The question is, this is the total probability of what? If we are talking about the electron wave in hydrogen, and we think of $\psi^*\psi\, d^3V$ as the probability of finding the electron in some small volume element $d^3V$, then if we sum these probabilities over all space, we should end up with the total probability of finding the electron somewhere in space. If the hydrogen atom has one electron, and you look everywhere, you should eventually find the electron with a probability (1). Thus the total probability should be given by the formula

$$\boxed{1 = \int_{\text{all space}} \psi^*\psi\, d^3V} \qquad (55)$$

The wave functions $\psi_1$ and $\psi_2$ that we presented you in Equations (45) and (46) do not have this property.

Let us see what the integral of $\psi_1^*\psi_1$ over all space is. We have

$$\psi_1 = e^{-r/a_0}\, e^{-i\omega_1 t} \qquad (56a)$$

$$\psi_1^* = e^{-r/a_0}\, e^{+i\omega_1 t} \qquad (\textit{change } -i \textit{ to } i) \qquad (56b)$$

so that

$$\psi_1^*\psi_1 = e^{-r/a_0}\, e^{+i\omega_1 t}\; e^{-r/a_0}\, e^{-i\omega_1 t}$$

$$\boxed{\psi_1^*\psi_1 = e^{-2r/a_0}} \qquad (57)$$

The $e^{i\omega_1 t}$'s cancelled and we end up with a real positive density.

To integrate $\psi^*\psi$ over all space, we notice that since $\psi^*\psi$ is spherically symmetric, we can take $d^3V$ as the volume of the spherical shell shown in Figure (2), a shell of radius r and thickness dr. That volume is

$$d^3V = (4\pi r^2)dr \qquad (58)$$

because $4\pi r^2$ is the area of a sphere of radius r. Throughout the shell, $\psi^*\psi$ has the same value $e^{-2r/a_0}$, thus our volume integral is simply

$$\int_{\text{all space}} \psi^*\psi\, d^3V = \int_{r=0}^{\infty} e^{-2r/a_0}(4\pi r^2)dr \qquad (59)$$

Being somewhat lazy, we look up in our short table of integrals, the integral of $r^2 e^{-\alpha r}$. After some manipulation shown in Appendix 1, we get

$$4\pi \int_0^{\infty} r^2 e^{-2r/a_0} dr = \pi (a_0)^3 \qquad (60)$$

The result is that the integral of $\psi^*\psi$ over all space is $\pi(a_0)^3$ instead of the desired value of 1.

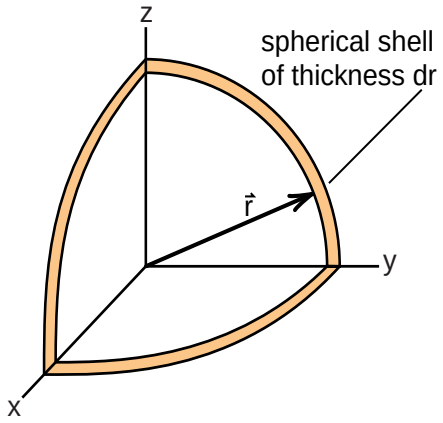



**Figure 2**
***We can use as the volume element $d^3V$ the spherical shell of radius r and thickness dr.***

To fix this problem, we use a so-called normalized wave function $(\psi_1)_{\text{normalized}}$, which is simply $\psi_1$ multiplied by an appropriate normalization constant C. To find out what C should be, write

$$(\psi_1)_{\text{normalized}} = C\psi_1 \tag{61a}$$

$$(\psi_1^*)_{\text{normalized}} = C^*\psi_1^* \tag{61b}$$

where, if we want, the normalization constant can be complex. Then we have

$$(\psi_1^*)_{\text{normalized}}(\psi_1)_{\text{normalized}} = (C^*C)\,\psi_1^*\psi_1$$

$$1 = \int_{\text{all space}} (\psi_1^*)_{\text{normalized}}(\psi_1)_{\text{normalized}}\, d^3V$$

$$= C^*C\int_{\text{all space}} \psi^*\psi\, d^3V \tag{62}$$

$$= C^*C\pi(a_0)^3$$

Thus

$$C^*C = \frac{1}{\pi(a_0)^3} \tag{63}$$

The simplest choice is to take C real, giving

$$\boxed{C = \frac{1}{\sqrt{\pi(a_0)^3}}} \quad \textit{normalization constant for } \psi_1 \tag{64}$$

As a result our normalized wave function becomes

$$\boxed{(\psi_1)_{\text{normalized}} = \frac{1}{\sqrt{\pi(a_0)^3}}\, e^{-2r/a_0}\, e^{-i\omega t}} \tag{65}$$

When you look at tables of wave functions, you will see factors like $1/\sqrt{\pi(a_0)^3}$ or $\sqrt{3/8\pi}$. They are merely the normalization constants. In one sense, the normalization constants just make the formulas look complicated. Most of the physics in our equation for $\psi_1$ is contained in the factor $e^{-r/a_0}$. It tells us that the electron wave decays exponentially as we go out from the proton, decaying by a factor of 1/e when we go out one Bohr radius $a_0$. The intensity, or probability $\psi^*\psi$ is proportional to $e^{-2r/a_0}$ and thus drops off by a factor $1/e^2$ when we are a Bohr radius from the proton. We also calculated the energy levels $E_1$ and $E_2$ without worrying about the normalization constants. It is nice to have a table that gives you the normalization constants, but you get a better insight into the shape of the standing wave patterns if you have another table without them.

---

**Exercise 4**

At what finite radius is there zero probability of finding an electron when the electron is in the n = 2, $\ell$ = 0, m = 0 standing wave pattern? Explain why and sketch the intensity $\psi^*_{2,0,0}\psi_{2,0,0}$

---

## THE DIRAC EQUATION

Our story is incomplete if we stop our discussion of particle wave equations with Schrödinger's equation. As successful as that equation is, it still does not handle relativistic effects. As we saw, Schrödinger could avoid the negative rest mass solutions by starting with the nonrelativistic formula $E = p^2/2m$ rather than the relativistic one $E^2 = p^2c^2 + m_0{}^2c^4$ .

It appeared to Dirac that the reason Schrödinger could avoid the nonphysical solutions is because the nonrelativistic equation involves only the first derivative with respect to time $\partial\psi/\partial t$ , rather than the second derivative $\partial^2\psi/\partial t^2$ that appeared in the relativistic equation (see Equation (14). Dirac thought that if he could develop a relativistic wave equation that avoided second time derivatives, then perhaps he could avoid the un-physical negative mass solutions.

By 1929, when Dirac was working on the problem, it was known that the electron had two spin states, spin up and spin down. It was these two spin states, along with the Pauli exclusion principle, that led to an understanding of the structure of the periodic table. These spin states are not included in or explained by Schrödinger's equation.

Slightly earlier, Wolfgang Pauli had introduced a new mathematical quantity called a ***spinor*** to describe the spin state of the electron. Spinors are quantities, involving complex numbers, that are in a sense half way between a scalar number and a vector. The existence of such a mathematical quantity was unknown until its invention was required to explain the electron. Pauli was able to modify Schrödinger's equation with the use of spinors to include the effects of electron spin.

Dirac found that by using a certain combination of spinors, he could write a relativistic wave equation for the electron that had only a first order time derivative $\partial\psi/\partial t$ . He hoped that this equation would avoid the un-physical negative mass solutions.

Dirac's equation was successful in that it not only included all the results of Schrödinger's and Pauli's equations, but it also correctly predicted tiny relativistic effects that could be detected in the spectra of hydrogen. However, Dirac soon found that his equation also led to the apparently negative mass solutions.

Dirac could not throw his equation away because it successfully predicted relativistic effects that were observed by experiment. Instead he found a new interpretation of the previously undesirable solutions. He found that these solutions could be reinterpreted as the wave for a particle whose mass was positive but whose electric charge was of the opposite sign. The equation led to the prediction that there should exist a particle with the same rest mass as the electron but with a positive electric charge. That particle was observed four years later in Carl Anderson's cloud chamber in the basement of the physics building at Caltech. It became known as the ***positron***.

We now know that any relativistic wave equation for a particle has two kinds of waves for a solution. One represents matter particles, and the other, like the wave for the positron represents ***antimatter***. If you have a relativistic wave equation, even if you start only with matter particles, the equation contains the mechanism for particle-antiparticle pair creation. You let the matter particles interact, and antimatter has a finite probability of being created. That is why Schrödinger and Dirac could not suppress the antimatter waves in the relativistic equations. However, by going to a nonrelativistic equation, representing situations where not enough energy is available to create electron positron pairs, Schrödinger could avoid the antimatter waves.

# Appendix I – Evaluation of a Normalization Integral

Our normalization integral is

$$\int_{\text{all space}} \psi^*\psi \, d^3V = 4\pi \int_{r=0}^{\infty} r^2 e^{-2r/a_0} \, dr \quad (59) \text{ repeat}$$

Looking for the integral of $r^2e^{-\alpha r}$ in our short table of integrals in the formulary, we find instead

$$\int x^2 e^{-ax} dx = \frac{1}{a^3}(a^2x^2 + 2ax + 2)e^{-ax} \quad (66)$$

If we set x = r and integrate from 0 to infinity, we have

$$\begin{aligned} \int_0^{\infty} r^2 e^{-ar} dr &= \frac{1}{a^3}(a^2r^2 + 2ar + 2)e^{-ar}\Big|_0^{\infty} \\ &= \frac{1}{a^3}(a^2R^2 + 2aR + 2)e^{-aR}\Big|_{R=\infty} \\ &\quad - \frac{1}{a^3}(a^20^2 + 2a\times 0 + 2)e^{-a\times 0} \end{aligned} \quad (67)$$

The exponential decay is so powerful that in the limit of large R, a term of the form $R^ne^{-aR}$ goes to zero for any value of n for positive (a). Thus all terms with a $e^{-aR}$ go to 0 as R goes to infinity. With $e^0 = 1$, we are left with

$$\int_0^{\infty} r^2 e^{-ar} dr = \frac{2}{a^3} \quad (68)$$

Now set $a = 1/\ 2a_0$ and we get

$$\begin{aligned} 4\pi \int_0^{\infty} r^2 e^{-2r/a_0} dr &= 4\pi \frac{2}{(2/a_0)^3} \\ &= \pi(a_0)^3 \end{aligned} \quad (69)$$

# APPENDIX II - An introduction to Schrödinger's Equation Applied to the Hydrogen Atom

## The Hydrogen Atom

Schrödinger's first major success with his wave equation was to solve for the electron standing waves in hydrogen, and to determine the electron energies in each of the standing wave patterns. For an electron in hydrogen, the potential energy is given by Coulomb's law as

$$V(r) = -\frac{e^2}{r} \qquad \text{(42) repeated}$$

where –e is the charge on the electron and r is the separation of the electron and proton. Thus the equation Schrödinger had to solve for hydrogen is the three dimensional equation

$$\boxed{i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi - \frac{e^2}{r}\psi}$$

*Schrödinger's equation for hydrogen atom*

(43) repeated

Quite a few steps are required to obtain solutions to Equation (43). The first is to look for solutions of definite frequency $\omega$ or energy $E = \hbar\omega$ by using the trial function

$$\psi = \psi(x,y,z)\,e^{-i\omega t} = \psi(\mathbf{x})\,e^{-i\omega t} \qquad (70)$$

where we will use the bold face **x** to stand for (x,y,z). Plugging this guess into Equation (43) gives

$$i\hbar(-i\omega)\psi(\mathbf{x})e^{-i\omega t} = -\frac{\hbar^2}{2m}\left[\nabla^2\psi(\mathbf{x})\right]e^{-i\omega t} - \frac{e^2}{r}\psi(\mathbf{x})e^{-i\omega t}$$

The factor $e^{-i\omega t}$ cancels and we are left with

$$\hbar\omega\,\psi(\mathbf{x}) = -\frac{\hbar^2}{2m}\left[\nabla^2\psi(\mathbf{x})\right] - \frac{e^2}{r}\psi(\mathbf{x}) \qquad (71)$$

With $\hbar\omega = E$, this becomes

$$E\,\psi(\mathbf{x}) = -\frac{\hbar^2}{2m}\left[\nabla^2\psi(\mathbf{x})\right] - \frac{e^2}{r}\psi(\mathbf{x}) \qquad (72)$$

The next step is to note that it is not convenient to handle a spherically symmetric potential $V(r) = -e^2/r$ using Cartesian coordinates x, y, and z. In the Chapter 4 of the Calculus text we derived the formula for $\nabla^2$ in *spherical polar coordinates* r, $\theta$, $\phi$ which are shown in Figure (1) reproduced here. In these spherical coordinates we show, after considerable work, that $\nabla^2\psi$ is given by Equation (4-10) as

$$\nabla^2\psi = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2} \qquad (4\text{-}10)$$

(Note: many texts write the first term as $1/r^2\,\partial/\partial r\,(r^2\partial\psi/\partial r)$ which is an equivalent but usually less convenient form.)

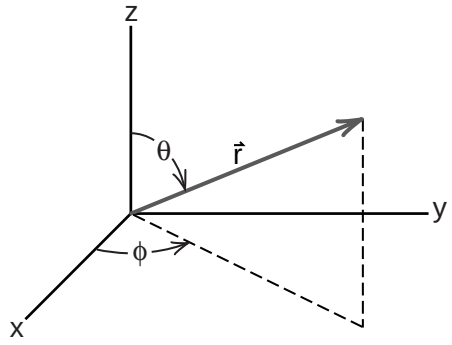



**Figure 1 (repeated)**
***Spherical polar coordinates.***

If we look at only the spherically symmetric solutions where

$$\psi(x,y,z) = \psi(r) \qquad \begin{matrix}\textit{spherically}\\ \textit{symmetric}\\ \textit{wave}\end{matrix} \tag{73}$$

then $\partial\psi(r)/\partial\theta = 0$, $\partial\psi/\partial\phi = 0$, and only the radial part of $\nabla^2\psi(r)$ survives. Schrödinger's equation for the spherically symmetric waves of energy E becomes

$$E\psi = -\frac{\hbar^2}{2m}\frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi) - \frac{e^2}{r}\psi \tag{74}$$

Multiplying through by $2mr/\hbar^2$, Equation (74) can be written in the form

$$\frac{\partial^2}{\partial r^2}(r\psi) + \left(\frac{a}{r} + b\right)r\psi = 0 \tag{75}$$

where

$$a = \frac{2me^2}{\hbar^2} \; ; \quad b = \frac{2mE}{\hbar^2} \tag{76}$$

If we define the variable u(r) by

$$u = r\psi \; ; \quad \psi = \frac{u}{r} \tag{77}$$

our equation for u becomes

$$\frac{\partial^2 u}{\partial r^2} + \left(\frac{a}{r} + b\right)u = 0 \tag{78}$$

---

**Exercise 5**

Derive Equation (78) starting from Equation (74).

---

Equation (78) is a differential equation we have not encountered before. Neither of our familiar guesses for a solution, like $u = e^{-\alpha r}$ or $u = \sin\omega r$, will work, as you can check for yourself. What does work is the function we will call $u_1$, which is

$$u_1(r) = re^{-\alpha r} \qquad \textit{guess} \tag{79}$$

Plugging our guess into Equation (79) gives

$$\frac{du_1}{dr} = e^{-\alpha r} - \alpha re^{-\alpha r}$$

$$\frac{d^2u_1}{dr^2} = -\alpha e^{-\alpha r} - \alpha e^{-\alpha r} + \alpha^2 re^{-\alpha r}$$

Thus $\dfrac{d^2u_1}{dr^2} + \left(\dfrac{a}{r} + b\right)u = 0$ becomes

$$-2\alpha e^{-\alpha r} + \alpha^2 re^{-\alpha r} + \frac{a}{r}re^{-\alpha r} + bre^{-\alpha r} = 0$$

The common factor $e^{-\alpha r}$ cancels and we are left with

$$\left[-2\alpha + a\right] + r\left[\alpha^2 + b\right] = 0 \tag{80}$$

The only way we can satisfy Equation (80) for arbitrary values of r is to set both square brackets separately equal to zero, giving

$$2\alpha = a \; ; \quad \alpha = a/2 \tag{81a}$$

$$\alpha^2 = -b \tag{81b}$$

Squaring Equation (81a) gives

$$\alpha^2 = \frac{a^2}{4} \tag{81c}$$

For Equations (81b) and (81c) to be consistent, the constants (a) and (b) must satisfy the relationship

$$-b = \frac{a^2}{4} \tag{82}$$

To see what Equation (82) implies, let us put back in the values of (a) and (b)

$$a = \frac{2me^2}{\hbar^2} \; ; \quad \frac{a^2}{4} = \frac{1}{4}\times\frac{4m^2e^4}{\hbar^4} \tag{83a}$$

$$-b = \frac{-2mE}{\hbar^2} \tag{83b}$$

Thus Equation (82) requires

$$-\frac{2mE}{\hbar^2} = \frac{m^2e^4}{\hbar^4}$$

or

$$\boxed{E = -\frac{me^4}{2\hbar^2} = -13.6\text{ eV}} \tag{84}$$

In our study of the Bohr theory, we found that the lowest energy level of the hydrogen atom was $E_1 = -me^4/2\hbar^2$ which turns out to be –13.6 electron volts. We now see that if the hydrogen wave amplitude is given by the solution $u_1$, or $\psi_1 = u_1 r$, then the energy of the electron in this wave pattern must be the same as the lowest energy level of the Bohr theory. This is a prediction of Schrödinger's wave equation without any arbitrary added assumptions like assuming angular momentum is quantized.

To see what the wave pattern is that corresponds to the energy level $E_1$, note that the Bohr radius $a_0$, the radius of the smallest Bohr orbit in the Bohr theory, is given by

$$a_0 = \frac{\hbar^2}{me^2} \qquad \textit{Bohr radius} \qquad (85)$$

Thus our constant (a) in Equation (77) can be written

$$a = \frac{2me^2}{\hbar^2} = \frac{2}{a_0} \qquad (86)$$

Thus Equation (81a) requires that

$$a = 2\alpha = \frac{2}{a_0} \; ; \qquad \alpha = \frac{1}{a_0} \qquad (87)$$

and the wave function $\psi_1(r)$ is given by

$$\psi_1(r) = \frac{u_1(r)}{r} = \frac{re^{-\alpha r}}{r} = e^{-\alpha r}$$

$$\boxed{\psi_1(r) = e^{-r/a_0}} \qquad (88)$$

The electron wave decays exponentially as we go out from the nucleus, decaying by a factor of 1/e when we go out one Bohr radius. We have just used Schrödinger's equation to solve for the ground state wave function, the lowest energy level standing wave pattern in hydrogen.

## The Second Energy Level

In the following exercise you will find another spherically symmetric solution for the hydrogen atom.

---

**Exercise 6**

Try the guess

$$u_2(r) = (r + cr^2)e^{-\alpha r} \; , \quad u_2 = r\psi_2 \qquad (89)$$

as a possible solution to Equation (78) where (c) is an unknown constant. Show that for (89) to be a solution, you have to satisfy the conditions

$$-2\alpha + 2c + a = 0 \qquad (90a)$$

$$\alpha^2 - 4c\alpha + ac + b = 0 \qquad (90b)$$

$$\alpha^2 c + bc = 0 \qquad (90c)$$

Then show that this requires $\alpha^2 = -b$ as before, and that

$$-b = \frac{a^2}{16} \Rightarrow \frac{-2mE_2}{\hbar^2} = \frac{1}{16} \times \frac{4m^2e^4}{\hbar^4} \qquad (91)$$

or

$$E_2 = -\frac{1}{4}\left(\frac{me^4}{2\hbar^2}\right) = \frac{-13.6\,eV}{4} = -3.60\,eV \qquad (92)$$

Then show that $\psi_2(r)$ is given by

$$\boxed{\psi_2 = (1 - \frac{r}{2a_0})e^{-r/2a_0}}$$

$$\boxed{E_2 = -3.6\,eV} \qquad (93)$$

---

In the Bohr theory, the energy levels $E_n$ are given by

$$E_n = \frac{E_1}{n^2} = -\frac{13.6\ eV}{n^2} \qquad (94)$$

The second energy level $E_2$ is thus

$$E_2 = \frac{E_1}{(2)^2} = \frac{E_1}{4} = -3.6\ eV$$

Thus the wave pattern you solved for in Exercise (8) is the spherically symmetric standing wave pattern in the second energy level. It is what we have called the $n = 2, \ell = 0$ wave pattern. Note that in the solution

$$\psi_2(r) = \left(1 - \frac{r}{2a_0}\right)e^{-r/2a_0} \qquad (93)\ \text{repeated}$$

when we are at a distance

$$\frac{r}{2a_0} = 1\ ; \quad r = 2a_0 \qquad (95)$$

the wave pattern in Equation (93) goes to zero. This means that the standing wave $\psi_2(r)$ has a spherical node out at a distance $r = 2a_0$. This is the spherical node we saw in the $\psi(n = 2, \ell = 0)$ pattern shown in the Physics text, Figure (38-1) repeated here.

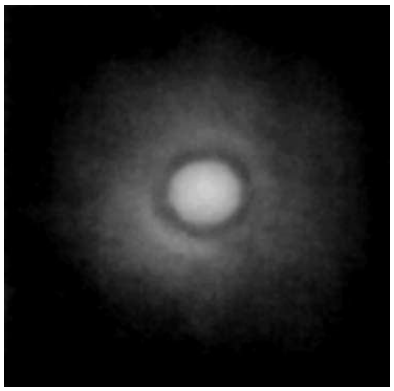


**Figure 38-1a**
***Hydrogen atom standing wave pattern for $n = 2, \ell = 0$.***

**Figure 3**
***Tacoma Narrows bridge in an n = 2 second harmonic standing wave pattern.***

(Movie. Press esc to stop)

If you try a guess of the form

$$u_3(r) = (1 + c_2 r + c_3 r^2)e^{-\alpha r} \qquad (96)$$

you end up with a spherical wave pattern $\psi_3(r)$ that has two spherical nodes, and has an energy

$$E_3 = \frac{E_1}{3^2} \qquad (97)$$

which is the third energy level.

You can now see the pattern. We can generate all the spherically symmetric $\ell = 0$ wave patterns by adding terms like $c_4 r^3, c_5 r^4, \cdots c_n r^{n-1}$ to our guess for $u_n(r)$. Solving for all the constants, we end up with

$$E_n = \frac{E_1}{n^2} \qquad (98)$$

which is the energy level structure Bohr discovered.

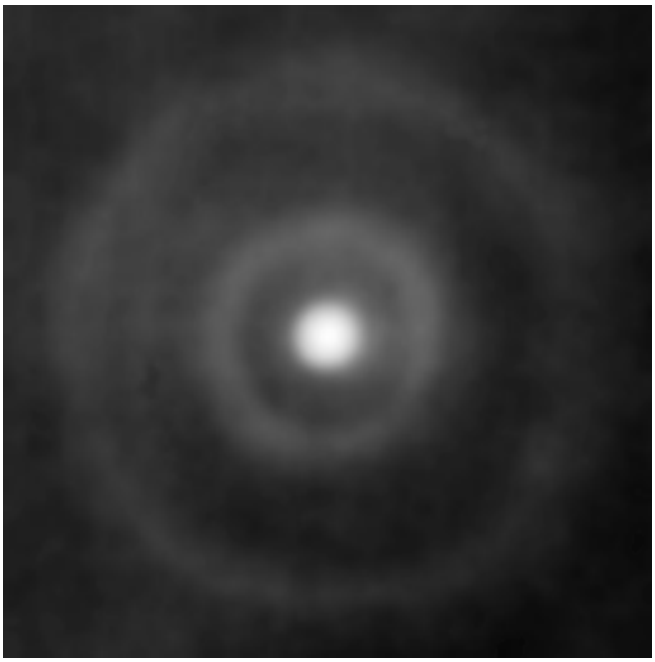


**Figure 38-1i**
***Wave pattern for $n = 3, \ell = 0$.***

## Non Spherically Symmetric Solutions

It was fairly easy to handle the spherically symmetric solutions to Schrödinger's equation for hydrogen, because we did not have to deal with the angular terms involving $\theta$ and $\phi$ in Equation (4-10) for $\nabla^2$. To find non spherically symmetric solutions, we have to work with the complete equation

$$\begin{aligned} E\psi &= -\frac{\hbar^2}{2m}\nabla^2\psi + V(r)\psi \\ \nabla^2\psi &= \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi) \\ &\quad + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) \\ &\quad + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2} \end{aligned} \qquad (99)$$

Differential equations involving $\nabla^2$ in spherical coordinates have been studied for a long time and standard procedures have been carefully worked out to handle the angular dependence of the solutions of these equations. As long as the equation has no other angular terms except those that appear in $\nabla^2$, then the solutions are of the form

$$f(r,\theta,\phi) = R_{n\ell m}(r)Y_{\ell m}(\theta,\phi) \qquad (100)$$

where $R_{n\ell m}(r)$ are functions that depend only on the variable (r), and the $Y_{\ell m}(\theta,\phi)$ are functions only of the angles $\theta$ and $\phi$. The subscripts n, $\ell$ and m can take on only integer values.

When we are dealing with Schrödinger's equation, the solutions are of the form

$$\psi(r,\theta,\phi) = \psi_{n\ell m}(r)Y_{\ell m}(\theta,\phi) \qquad (101)$$

where each different allowed integer value of the subscripts n, $\ell$, and m corresponds to a different allowed standing wave pattern for the electron.

The functions $Y_{\ell m}(\theta,\phi)$, which are called ***spherical harmonics***, start off quite simply for small $\ell$, m, n, but become more complex as $\ell$ and m increase. The simplest are

$$\begin{aligned} Y_{0,0}(\theta,\phi) &= 1 \qquad \left(\textit{no angular dependence}\right) \\ Y_{1,0} &= \cos\theta \\ Y_{1,1} &= \frac{-1}{\sqrt{2}}\sin\theta\, e^{i\phi} \\ Y_{1,-1} &= \frac{1}{\sqrt{2}}\sin\theta\, e^{-i\phi} \end{aligned} \qquad (102)$$

Since $Y_{0,0}$ has no angular dependence, all solutions of the form

$$\psi_{n,0,0} = \psi_n(r)Y_{0,0} = \psi_n(r) \qquad (103)$$

are the spherically symmetric solutions we have already been studying. We calculated $\psi_1(r)$ and had you calculate $\psi_2(r)$, which corresponds to the values n = 1 and n = 2 respectively.

When we worked out the solution $\psi_1(r)$ we found that it represented an electron in the lowest, n = 1, energy level. You were to show that $\psi_2(r)$ represented an electron in the second, n = 2, energy level. We can see that for the symmetric solutions, the integer subscript n is the energy quantum number for the electron.

It turns out that the integer subscripts $\ell$ and m define the amount of angular momentum the electron has in a particular wave pattern. When $\ell$ = 0, m = 0, the electron has no angular momentum. Thus the symmetric solutions represent an electron with no angular momentum.

The quantum number $\ell$ is related to the total orbital angular momentum of the electron, and m is proportional to the z component $L_z$ of orbital angular momentum. Explicitly

$$L_z = m\hbar \qquad (104)$$

The fact that the numbers $\ell$, m and n have to have integer values is simply a consequence that for any confined wave, there is an explicit set of allowed standing wave patterns. The electron in the hydrogen atom is confined by the Coulomb force of the proton. When you work out the mathematics to handle $\nabla^2$ in spherical coordinates, you find that the allowed standing wave patterns can be identified by the integers $\ell$, m and n.

There are certain rules for the possible values of $\ell$, m and n. When n = 1, there is only one solution which we found. It corresponds to $\ell$ = m = 0. For n = 2, the possible solutions are:

| n | $\ell$ | m | |
|---|---|---|---|
| 2 | 0 | 0 | *possible values of $\ell$ and m for n = 2* |
| 2 | 1 | 0 | |
| 2 | 1 | 1 | |
| 2 | 1 | –1 | |

In general, n ranges from 1 to infinity, $\ell$ can have values from 0 up to n - 1, and m can range in integer steps from $+\ell$ down to $-\ell$. These are the rules that define the possible standing wave patterns of the electron in hydrogen.

# Calculus 2000-Chapter 7

# Divergence

*In the Physics text we pointed out that a vector field was uniquely determined by formulas for the surface integral and the line integral. As we have mentioned several times, that is why there are four Maxwell equations, since we need equations for the surface and line integral of both the electric and magnetic fields. The divergence and curl are the surface and line integrals shrunk down to an infinitesimal or differential scale. We will discuss divergence in this chapter and curl in the next.*

## THE DIVERGENCE

As we mentioned, the divergence is a surface integral shrunk down to an infinitesimal or differential scale. To see how this shrinking takes place, we will start with the concept of the surface integral as expressed by Gauss' law and see how we can apply it on a very small scale.

We begin with Equation (29-5) of the Physics text

$$\int_{\substack{\text{closed}\\ \text{surface}}} \vec{E}\cdot d\vec{A} = \frac{Q_{in}}{\varepsilon_0} \qquad (29\text{-}5)$$

Equation (29-4) says that for any closed surface, the integral of $\vec{E}\cdot d\vec{A}$ over the surface is equal to $1/\varepsilon_0$ times the total charge $Q_{in}$ inside the volume bounded by the surface.

The interpretation we gave to this equation was to call $\vec{E}\cdot d\vec{A}$ the ***flux*** of the field $\vec{E}$ out through the area element $d\vec{A}$. The integral over the closed surface is the total flux flowing out through the surface. We said that this net flux out was created by the electric charge inside. By calculating the flux of $\vec{E}$ out through a spherical surface centered on a point charge, we found that the amount of flux created by a charge Q was $Q/\varepsilon_0$.

The fact that Equation (29-4) applies to a surface of arbitrary shape follows from the fact that the electric field of a point charge is mathematically similar to the velocity field of a point source in an incompressible fluid like water. We described a point source of a velocity field as some sort of "magic" device that created water molecules. The physical content of Gauss' law applied to water was that the total flux of water out through any closed surface had to be equal to the rate at which water molecules were being created inside.

Of course for a real situation there are no "magic" sources creating water molecules, with the result that there is no net flux of water out through any closed surface, and the velocity field of water obeys the equation

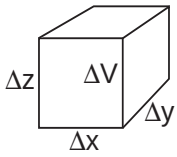
$$\int_{\text{closed surface}} \vec{v}\cdot d\vec{A} = 0 \qquad (1)$$

Equation (1) is the condition that the velocity field is a purely solenoidal field like the magnetic field.

Back to Gauss' law, Equation (29-5). Before we shrink the law to an infinitesmal scale, we would like to change the right hand side, expressng the total charge $Q_{in}$ in terms of the charge density $\rho(x,y,z)$ that is within the volume bounded by the closed surface.

We do this by considering a small volume element $\Delta V_i = (\Delta x\,\Delta y\,\Delta z)_i$. If the charge density at point (i) is $\rho(x_i,y_i,z_i)$ then the amount of charge $\Delta Q_i$ at $\Delta V_i$ is

$$\Delta Q_i = \rho(x_i,y_i,z_i)\Delta V_i \qquad (2)$$




Adding up all the $\Delta Q_i$ that reside inside the surface gives us

$$\begin{aligned} Q_{in} &= \sum_i \Delta Q_i = \sum_i \rho_i \Delta V_i \\ &= \sum_i \rho(x_i,y_i,z_i)\,\Delta x_i\,\Delta y_i\,\Delta z_i \end{aligned} \qquad (3)$$

Taking the limit as the $\Delta x$, $\Delta y$ and $\Delta z$ go to zero gives us the integral

$$Q_{in} = \int_{\substack{\text{volume}\\ \text{bounded by}\\ \text{closed surface}}} \rho(x,y,z)\,dx\,dy\,dz \qquad (4)$$

To shorten the notation, let V be the volume bounded by the closed surface S, and introduce the notation

$$d^3V \equiv dx\,dy\,dz \qquad (5)$$

Then Equation (4) can be written

$$\boxed{Q_{in} = \int_V \rho(x,y,z)\,d^3V} \qquad (6)$$

Using Equation (6) in Gauss' law (29-5) gives us

$$\int_S \vec{E}\cdot d\vec{A} = \frac{1}{\varepsilon_0}\int_V \rho(x,y,z)\,d^3V \qquad (7)$$

Equation (7) is a more general integral form of Gauss' law, relating the surface integral of $\vec{E}$ over a closed surface S to the volume integral of $\rho$ over the volume bounded by S. It is Equation (7) that we would now like to shrink down to an infinitesmal scale.

We know how to go to the small scale version of the volume integral of $\rho$, just undo the steps (2) through (6) that we used to derive the volume integral. In particular we will focus our attention on one small volume element $\Delta V_i = \Delta x_i\, \Delta y_i\, \Delta z_i$ and apply Gauss' law to this volume

$$\int_{\substack{\text{surface}\\ \text{bounding}\\ \Delta V_i}} \vec{E}\cdot d\vec{A} = \frac{\Delta Q_i}{\varepsilon_0} = \frac{1}{\varepsilon_0}\rho(x_i,y_i,z_i)\Delta V_i \qquad (8)$$

It is clear how we got the total charge $Q_{in}$ when we added up all the $\Delta Q_i$ inside the volume V. But how do we handle the surface integral of $\vec{E}$ ? How do we interpret adding a bunch of surface integrals over the small volume elements $\Delta V_i$ to get the surface integral over the entire surface S?

The way to picture it is to remember that the surface integral over the surface of $\Delta V_i$ is equal to the flux of $\vec{E}$ created inside $\Delta V_i$. From this point of view, ***the total flux flowing out through the surface of the entire volume will be the sum of the fluxes created within each volume element.*** To calculate this sum, we first have to calculate the flux flowing out of the volume element $\Delta V_i$.

In Figure (1), we show the volume element $\Delta V_i$ located at $(x_i,y_i,z_i)$, with sides $\Delta x$, $\Delta y$ and $\Delta z$. Flowing through this volume element is the electric field $\vec{E}(x,y,z)$.

Also in Figure (1) we have drawn the surface area vectors $\overrightarrow{\Delta A}_1$, and $\overrightarrow{\Delta A}_2$ for the left and right vertical faces. Recall that for a surface integral, the area vector $\overrightarrow{\Delta A}$ or $d\vec{A}$ is perpendicular to the surface, pointing out of the surface. Thus $\overrightarrow{\Delta A}_2$ is x directed with a magnitude equal to the area $\Delta y\Delta z$ of that side, while $\overrightarrow{\Delta A}_1$ points in the –x direction and has the same magnitude.

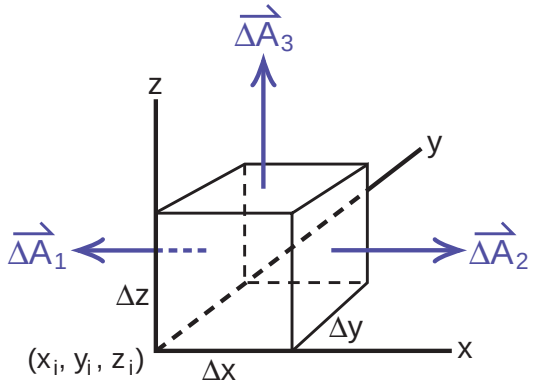



**Figure 1**
***The volume element $\Delta V_i$.***

We can formally write

$$\overrightarrow{\Delta A}_1 = -\hat{x}\Delta y\Delta z \;;\quad \overrightarrow{\Delta A}_2 = \hat{x}\Delta y\Delta z \qquad (9)$$

where $\hat{x}$ is the unit vector in the x direction. Similar formulas hold for the area vectors for the other four faces of $\Delta V$. For example, on the top face we have $\overrightarrow{\Delta A}_3 = \hat{z}\Delta x\Delta y$.

To calculate the total flux of $\vec{E}$ out of $\Delta V$, we have to calculate the flux out through each of the six faces. For the two x oriented areas $\overrightarrow{\Delta A}_1$, and $\overrightarrow{\Delta A}_2$, only the x component of $\vec{E}$ will contribute to the dot products $\vec{E}\cdot\overrightarrow{\Delta A}$. Let $\bar{E}_x(x,y,z)$ be the average value of $E_x$ at face 1, and $\bar{E}_x(x+\Delta x,y,z)$ be the average value of $E_x$ at face 2, which is a distance $\Delta x$ down the x axis from face 1. The flux out of face 2 will be

$$\begin{aligned}\text{flux out of face 2} &= \bar{E}_x(x+\Delta x,y,z)\Delta A_2 \\ &= \bar{E}_x(x+\Delta x,y,z)\Delta y\Delta z\end{aligned} \qquad (10)$$

At face 1, where $\overrightarrow{\Delta A}_1 = -\hat{x}\Delta y\Delta z$, the dot product $\vec{E}\cdot\overrightarrow{\Delta A}$ can be written

$$\begin{aligned}\vec{E}\cdot\overrightarrow{\Delta A}_1 &= (\hat{x}\bar{E}_x+\hat{y}\bar{E}_y+\hat{z}\bar{E}_z)\cdot(-\hat{x}\Delta y\Delta z) \\ &= -\bar{E}_x(x,y,z)\Delta y\Delta z\end{aligned} \qquad (11)$$

where $\hat{x}\cdot\hat{x} = 1$, $\hat{y}\cdot\hat{x} = \hat{z}\cdot\hat{x} = 0$. We wrote the full dot product in Equation (11) so that you could see explicitly where the minus sign came from.

Combining Equations (10) and (11) for the total flux out of the two x directed faces of $\Delta V$, we get

$$\text{flux out of x directed faces} = \left[\,\bar{E}_x(x+\Delta x,y,z) - \bar{E}_x(xy,z)\,\right]\Delta y\Delta z \qquad (12)$$

If we multiply Equation (12) by $\Delta x/\Delta x = 1$ we get

$$\begin{matrix}\text{flux out}\\ \text{of x}\\ \text{directed}\\ \text{faces of } \Delta V\end{matrix} = \left[\frac{\bar{E}_x(x+\Delta x,y,z)-\bar{E}_x(x,y,z)}{\Delta x}\right]\Delta x\Delta y\Delta z \qquad (13)$$

At this point, $\bar{E}_x(x+\Delta x,y,z)$ and $\bar{E}_x(x,y,z)$ are the average values of $\bar{E}_x$, averaged over the x directed faces at $x + \Delta x$ and x respectively, while the functions without averaging, namely $E_x(x+\Delta x,y,z)$ and $E_x(x,y,z)$ are just the values of $E_x$ at the lower front corners of the x oriented faces as shown in Figure (2). Any difference between the average values of $\bar{E}_x$ and the corner values $E_x$ will be due to y and z variations of $E_x$ over the area $\Delta y\Delta z$.

In Equation (13) we see that the change of $E_x$, as we move in the x direction, is going to become very important. It should be clear that we are going to get a partial derivative of $E_x$ with respect to x. What we are going to do now is say that variations of $E_x$ in the x direction are important but variations of $E_x$ in the y and z direction are not, and as a result we can replace the average values of $\bar{E}_x$ with the corner values $E_x$.

The above paragraph was intended to sound like a questionable procedure. If we do it, Equation (13) immediately simplifies, as we will see shortly. But how do we justify such a step? The answer, which we work out in detail in the appendix to this chapter, is that when we take the limit as $\Delta V$ goes to zero, contributions due to y and z variations of $E_x$ go to zero faster than the contribution from the x variation. Neglecting the y and z variations turns out to be similar to neglecting $\alpha^2$ terms compared to $\alpha$ terms in an expansion of $(1+\alpha)^n$ when $\alpha$ is a small number.

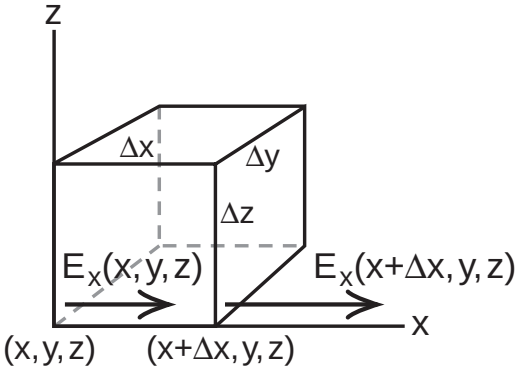



**Figure 2**
***Electric field at the lower front corners.***

We put this discussion in the appendix because it takes some effort which distracts from our goal of reducing Gauss' law to a differential equation. However it is important to know how to figure out when certain terms or dependencies can be neglected when we take calculus limits. Thus the appendix should not be skipped.

Assuming that we can replace $\bar{E}_x$ by $E_x$ in Equation (13), noting that $\Delta x\Delta y\Delta z = \Delta V$, and taking the limit as $\Delta x$ goes to zero gives us

$$\begin{matrix}\text{flux out of}\\ \text{x directed}\\ \text{faces of } \Delta V\end{matrix} = \lim_{\Delta x\to 0}\left[\frac{E_x(x+\Delta x,y,z)-E_x(x,y,z)}{\Delta x}\right]\Delta V \qquad (14)$$

The limit is clearly the partial derivative $\partial E_x(x,y,z)/\partial x$ and we get

$$\begin{matrix}\text{flux out of}\\ \text{x directed}\\ \text{faces of } \Delta V\end{matrix} = \frac{\partial E_x(x,y,z)}{\partial x}\Delta V \qquad (15a)$$

Similar equations should apply to the y and z faces, giving us

$$\begin{matrix}\text{flux out of}\\ \text{y directed}\\ \text{faces of } \Delta V\end{matrix} = \frac{\partial E_y(x,y,z)}{\partial y}\Delta V \qquad (15b)$$

$$\begin{matrix}\text{flux out of}\\ \text{z directed}\\ \text{faces of } \Delta V\end{matrix} = \frac{\partial E_z(x,y,z)}{\partial z}\Delta V \qquad (15c)$$

**Exercise 1**

Draw the appropriate sketches and reproduce the arguments needed to derive Equation (15b) or (15c).

When we add up the flux out of all six faces, we get the total flux out of $\Delta V$

$$\begin{matrix}\text{total flux}\\ \text{out of } \Delta V\end{matrix} = \left[\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right]\Delta V \qquad (16)$$

You should spot immediately that the notation in Equation (16) can be simplified by introducing the partial derivative operator

$$\begin{aligned}\vec{\nabla} &= \left(\hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}\right)\\ &\equiv (\hat{x}\nabla_x + \hat{y}\nabla_y + \hat{z}\nabla_z)\end{aligned} \qquad (17)$$

From the definition of the vector dot product we have

$$\begin{aligned}&\vec{\nabla}\cdot\vec{E}\\ &= \left(\hat{x}\frac{\partial}{\partial x}+\hat{y}\frac{\partial}{\partial y}+\hat{z}\frac{\partial}{\partial z}\right)\cdot(\hat{x}E_x+\hat{y}E_y+\hat{z}E_z)\\ &= \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\end{aligned} \qquad (18)$$

where we used $\hat{x}\cdot\hat{x} = 1$, $\hat{x}\cdot\hat{y} = 0$, etc., and noted that the unit vectors are constants that can be taken outside the derivative. For example,

$$\frac{\partial}{\partial x}(\hat{x}E_x) = \hat{x}\frac{\partial E_x}{\partial x} \qquad (18a)$$

Using the notation of Equation (18), we get for the total flux out of $\Delta V$

$$\begin{matrix}\text{total flux}\\ \text{out of } \Delta V\end{matrix} = (\vec{\nabla}\cdot\vec{E})\Delta V \qquad (19)$$

Equation (19) applies to each $\Delta V_i$ at each point $(x_i, y_i, z_i)$ within any volume V bounded by a closed surface S. The total flux out through the surface S, which is the surface integral of $\vec{E}$, will be equal to the sum of all the flux created inside in all the $\Delta V_i$. Thus we get

$$\int_{\substack{\text{surface}\\ \text{bounding V}}} \vec{E}\cdot d\vec{A} = \sum_i (\vec{\nabla}\cdot\vec{E})\Delta V_i \qquad (20)$$

As we take the limit at $\Delta V_i$ goes to zero size, the sum becomes an integral, and we end up with

$$\boxed{\int_{\substack{\text{closed surface}\\ \text{bounding}\\ \text{volume V}}} \vec{E}\cdot d\vec{A} = \int_V \vec{\nabla}\cdot\vec{E}\, d^3V} \quad \begin{matrix}\textit{divergence}\\ \textit{theorem}\end{matrix} \qquad (21)$$

where we are using the notation of Equation (5) that $d^3V \equiv dx\,dy\,dz$.

Equation (21) is known as the ***divergence theorem***, and the quantity $\vec{\nabla}\cdot\vec{E}$ is known as the ***divergence*** of the vector field $\vec{E}$. We saw the same operator $\vec{\nabla}$ in the Chapter 3 when it acted on a scalar field f(x,y,z). Then we had what was called a ***gradient***

$$\begin{aligned}&\vec{\nabla}f && \textit{gradient of a scalar field}\\ &\vec{\nabla}\cdot\vec{E} && \textit{divergence of a vector field}\end{aligned} \qquad (22)$$

You can see that $\vec{\nabla}$ operating on a scalar field f(x,y,z) creates a vector field $\vec{\nabla}f$. In contrast, the dot product of $\vec{\nabla}$ with a vector field $\vec{E}$ creates a scalar field $\vec{\nabla}\cdot\vec{E}$ that has a value at every point in space but does not point anywhere.

Equation (21), the divergence theorem, is an extremely useful result for it allows us to go back and forth between a surface integral and a volume integral. In Equation (7) reproduced here,

$$\int_S \vec{E}\cdot d\vec{A} = \frac{1}{\varepsilon_0}\int_V \rho(x,y,z)\,d^3V \qquad (7)\ \text{repeated}$$

we had a mixed bag with a surface integral over a closed surface on the left and a volume integral over the enclosed volume V on the right. Back then, there was not much more we could do with that equation.

But now we can replace the surface integral of $\vec{E}$ with a volume integral of $\vec{\nabla}\cdot\vec{E}$ to get

$$\int_S \vec{E}\cdot d\vec{A} = \frac{1}{\varepsilon_0}\int_V \rho(x,y,z)\,d^3V \qquad (7)\ \text{repeated}$$

$$\int_V \vec{\nabla}\cdot\vec{E}\,d^3V = \frac{1}{\varepsilon_0}\int_V \rho(x,y,z)\,d^3V \qquad (23)$$

Since we are integrating over the same volume V for both integrals, we can write (23) as

$$\int_V \left[\vec{\nabla}\cdot\vec{E}(x,y,z) - \frac{\rho(x,y,z)}{\varepsilon_0}\right] d^3V = 0 \qquad (24)$$

The next argument is one often used in physics. ***Since the integral in Equation (24) has to be zero for any volume V we choose, the only way that can happen is if the integrand, the stuff in the square brackets, is zero.*** This gives us the differential equation

$$\boxed{\vec{\nabla}\cdot E(x,y,z) = \frac{\rho(x,y,z)}{\varepsilon_0}} \quad \textit{Gauss' law in differential form} \qquad (25)$$

Equation (25) is the differential equation representing Gauss' law. When Maxwell's equations are written as differential equations, this will be one of the four.

---

**Exercise 2**

Another of Maxwell's equations in integral form is

$$\int_{\text{closed surface}} \vec{B}\cdot d\vec{A} = 0$$

What is the corresponding differential equation?

---

## Electric Field of a Point Charge

Until now, in both the Physics and Calculus texts, when we obtained a new differential equation, we illustrated its use with explicit examples. This time we do not yet have a good example for our new Equation (25) $\vec{\nabla}\cdot\vec{E} = \rho/\varepsilon_0$. This is the differential form of Gauss' law, and our best example for the use of Gauss' law was in calculating the electric field of a point charge. The problem is that, at the point charge itself, the field $\vec{E}$ and its partial derivatives are infinite and the assumptions we made in deriving Equation (25) do not apply.

When we are dealing with the electric field of a point charge, the field $\vec{E}$ is well behaved and all partial derivatives are finite, except at the charge. The way we can handle point charges is to use Equation (25) $\vec{\nabla}\cdot\vec{E} = \rho/\varepsilon_0$ everywhere except in a small region around the charge. In that region we revert to the integral form of Gauss' law which allows us to work just outside the point charge and avoid the infinities.

Here is an outline of the way we handle the problem of a point charge. We are working with Equation (25)

$$\vec{\nabla}\cdot E(x,y,z) = \frac{\rho(x,y,z)}{\varepsilon_0} \qquad (25)\ \text{repeated}$$

and everything is going well until we come up to a point charge located at the point $(x_0,y_0,z_0)$. In a small region surrounding the point charge, we integrate Equation (25) over the volume, getting

$$\int_{\substack{\text{volume}\\ \text{surrounding}\\ \text{charge}}} \vec{\nabla}\cdot\vec{E}\,d^3V = \int_{\substack{\text{volume}\\ \text{surrounding}\\ \text{charge}}} \frac{\rho}{\varepsilon_0}\,d^3V \qquad (26)$$

The volume integral of the charge density $\rho$ over the region of the point charge is simply the charge Q itself, thus we can immediately do that volume integral, giving us

$$\int_{\substack{\text{volume}\\ \text{surrounding}\\ \text{charge}}} \vec{\nabla}\cdot\vec{E}\,d^3V = \frac{Q}{\varepsilon_0} \qquad (27)$$

We still have the problem that $\vec{\nabla}\cdot\vec{E}$ is infinite at the charge itself. But we can avoid this problem by converting the volume integral of $\vec{\nabla}\cdot\vec{E}$ to a surface integral of $\vec{E}$ using the divergence theorem, Equation (21)

$$\int_{\substack{\text{volume}\\ \text{surrounding}\\ \text{charge}}} \vec{\nabla}\cdot\vec{E}\, d^3V = \int_{\substack{\text{surface}\\ \text{enclosing}\\ \text{charge}}} \vec{E}\cdot d\vec{A} \qquad \text{(21) repeated}$$

to get

$$\int_{\substack{\text{surface}\\ \text{surrounding}\\ \text{charge}}} \vec{E}\cdot d\vec{A} = \frac{Q}{\varepsilon_0} \qquad (28)$$

In Equation (28), which we recognize as the form of Gauss' law we started with in the Physics text, the electric field is evaluated only at the surface surrounding the point charge, and not at the charge itself. Away from the charge, the field is finite and we have no problem with Equation (28).

There is a mathematical problem with the concept of a point charge, where a finite amount of charge is crammed into a region of zero volume, giving us infinite charge densities and infinite fields there. We have just shown how these infinities can be avoided mathematically, at least for Gauss' law, by converting the volume integral of $\vec{\nabla}\cdot\vec{E}$ at the charge to a surface integral of $\vec{E}$ out from the charge. Was this just a mathematical exercise, or in physics do we really have to deal with point charges?

The theory of quantum electrodynamics, which describes the interaction of electrons with light (with photons), is the most precisely verified theory in science. It explains, for example, the very smallest relativistic corrections observed in the spectrum of the hydrogen atom. This theory treats the electron as an actual ***point particle*** with a finite amount of mass and charge confined to a region of zero volume. The trick we just pulled to handle the electric field of a point charge was quite simple compared to the tricks that the inventors of quantum electrodynamics, Feynman, Schwinger, and Tomonaga, had to pull to handle the infinite mass and energy densities they encountered. The remarkable accomplishment was that they succeeded in constructing a theory of point particles, a theory that gave finite and correct, answers.

The question that remains unanswered, is whether the electron is truly a point particle, or does it have some size that is so small that we have not been able to see the structure yet? The important feature of quantum electrodynamics is that it makes testable predictions without any reference to the electron's structure. We get the same predictions whether the electron has no size, or is some structure that is too small to see. Our handling of the electric field of a point charge is your first example of how such a theory can be constructed. By converting to a surface integral surrounding the charge, it makes no difference whether the charge is truly a point, or confined to some region too small to see.

By the way, in the current picture of elementary particles, in what is often called the ***standard model***, the true elementary particles are all point particles. These elementary particles are the six electron type particles called ***leptons*** (they are the electron, the muon, the tau particle, and three kinds of neutrinos) and six kinds of ***quarks***. The standard model makes many successful predictions but appears to have one critical flaw. The problem is that no one has yet succeeded in constructing a theory for the interaction of point particles with gravity, the so called ***quantum theory of gravity***. Every attempt to do so has thus far led to infinities that could not be gotten rid of by any known mathematical technique.

This failure to develop a quantum theory of gravity in which gravity interacts with point particles, has led to theories such as ***string theory*** where the elementary particles have a finite, but tiny size. String theory appears to avoid the infinities in the gravitational interaction, but the strings, from which particles are assumed to be made, are predicted to be so small that no way has been found to test whether they actually exist or not. It is interesting that so far our only evidence that elementary particles actually have structure is our failure to construct a theory of gravity.

## THE δ FUNCTION

When we applied the differential form of Gauss' law $\vec{\nabla}\cdot\vec{E} = \rho/\varepsilon_0$ to the field of a point charge, we avoided the problem of mathematical infinities by integrating the equation over a small volume surrounding the charge. We never did say what the charge density $\rho(x,y,z)$ was for a point charge Q, because we knew that if we integrated $\rho(x,y,z)$ over the region of the charge, the answer would be simply Q itself.

In physics we often run into quantities like the charge density of a point charge where the density at the charge looks infinite, but when we integrate the density over the region of the charge, we get a finite, reasonable answer. There is a convenient way to handle such problems by using what is called the ***delta (δ) function***.

The one dimensional δ function is a curve with a unit area under it, but all the area is confined to a region of zero width. We obtain such a curve mathematically through the use of a limiting process.

Consider the curve shown in Figure (3) that is zero everywhere except in the region around the point $x_0$. In that region it is a rectangle of width $\Delta x$ and height $1/\Delta x$. The area under this curve is

$$\text{area under rectangle} = (\Delta x)\left(\frac{1}{\Delta x}\right) = 1 \qquad (29)$$

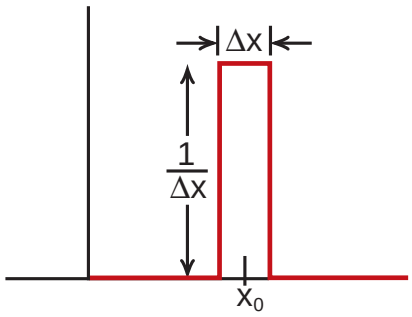



**Figure 3**
***When we take the limit as Δx goes to zero, we get a one dimensional delta function.***

Now take the limit as $\Delta x \to 0$, and we end up with a curve, whose total area remains 1, but whose width goes to zero and height goes to infinity. We will call this curve $\delta(x_0)$

$$\delta(x_0) \equiv \lim_{\Delta x \to 0}\left(\begin{array}{l}\text{of the curve of width}\\ \Delta x \text{ and height } 1/\Delta x,\\ \text{centered at } x_0\end{array}\right) \qquad (30)$$

Even though $\delta(x_0)$ is infinitely high at the point $x_0$, its integral over any region that includes the point $x_0$ is just the number 1

$$\int_{\text{x less than } x_0}^{\text{x greater than } x_0} \delta(x_0)dx = 1 \qquad (31)$$

Actually the only important property of the δ function is Equation (31). The curve does not have to be a rectangle, it could be the limit of some smooth curve like that shown in Figure (4). As long as, in the limit that $\Delta x \to 0$, the curve becomes infinitely high, infinitely narrow, and has a unit area under it, it is a δ function.

In three dimensions, the δ function $\delta(x_0,y_0,z_0)$ is a quantity that is zero everywhere except at the point $(x_0,y_0,z_0)$, but whose integral over that region is 1

$$\int_{\substack{\text{any volume}\\ \text{including the}\\ \text{point } (x_0,y_0,z_0)}} \delta(x_0,y_0,z_0)dV = 1 \qquad (32)$$

An example of such a δ function is the function whose value is zero everywhere except within a distance $\Delta x$ of $x_0$, $\Delta y$ of $y_0$, and $\Delta z$ of $z_0$. In that region the value is $(1/\Delta x)(1/\Delta y)(1/\Delta z)$, so that the total volume is 1. Then take the limit as $\Delta x \to 0$, $\Delta y \to 0$, and $\Delta z \to 0$.

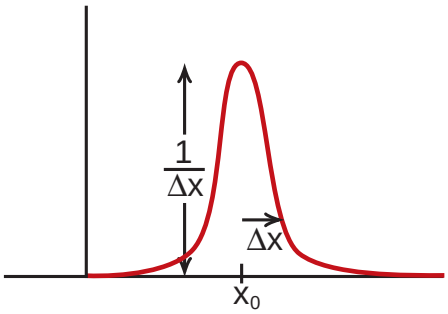



**Figure 4**
***We have a delta function as long as the area remains 1, and the width goes to zero.***

We can now use the $\delta$ function to describe the charge density of a point charge. If a point charge has a total charge Q and is located at the point $(x_0,y_0,z_0)$, then the charge density $\rho(x,y,z)$ is

$$\rho(x,y,z) = Q\delta(x_0,y_0,z_0) \quad \begin{matrix}\textit{charge density}\\ \textit{of point charge}\\ \textit{at } x_0,y_0,z_0\end{matrix} \qquad (33)$$

The differential form of Gauss' law applied to this charge density is

$$\begin{aligned}\vec{\nabla}\cdot\vec{E} &= \frac{\rho(x,y,z)}{\varepsilon_0}\\ \vec{\nabla}\cdot\vec{E} &= \frac{Q}{\varepsilon_0}\delta(x_0,y_0,z_0)\end{aligned} \qquad (34)$$

To handle Equation (34), we use our old trick of going back to the integral form by first integrating over a volume that includes the charge

$$\int_{\substack{\text{volume}\\ \text{including}\\ \text{charge}}} \vec{\nabla}\cdot\vec{E}\,dV = \int_{\substack{\text{volume}\\ \text{including point}\\ x_0,\,y_0,\,z_0}} \frac{Q}{\varepsilon_0}\delta(x_0,y_0,z_0)dV \qquad (35)$$

Since $Q/\varepsilon_0$ is a constant, it can be taken outside the integral on the right side of Equation (35), giving

$$\frac{Q}{\varepsilon_0}\int_{\substack{\text{volume}\\ \text{including point}\\ x_0,\,y_0,\,z_0}} \delta(x_0,y_0,z_0)dV = \frac{Q}{\varepsilon_0}\times 1 \qquad (36)$$

where we used the fact that the integral of the $\delta$ function was 1. Now convert the volume integral of $\vec{\nabla}\cdot\vec{E}$ to a surface integral

$$\int_{\substack{\text{volume}\\ \text{including point}\\ x_0,\,y_0,\,z_0}} \vec{\nabla}\cdot\vec{E}\,dV = \int_{\substack{\text{surface}\\ \text{surrounding}\\ x_0,\,y_0,\,z_0}} \vec{E}\cdot d\vec{A} \qquad (37)$$

Using (36) and (37) gives

$$\int_{\substack{\text{closed surface}\\ \text{including Q}}} \vec{E}\cdot d\vec{A} = \frac{Q}{\varepsilon_0}$$

which is our integral form of Gauss' law.

From this example, you can see that the $\delta$ function allows us to write an explicit formula for the charge density of a point charge, and you can see that the only things we have to know about a $\delta$ function is that $\delta(x_0,y_0,z_0)$ is zero except at $(x_0,y_0,z_0)$ and that its volume integral around that point is 1. As you go farther in physics, you will encounter the $\delta$ function more and more often. It is rather nice in that there is no function easier to integrate.

---

**Exercise 3**

Explain why the following mathematical relationship is true for any continuous function f(x,y,z)

$$\int_{\substack{\text{any volume}\\ \text{including the}\\ \text{point } (x_0,y_0,z_0)}} f(x,y,z)\,\delta(x_0,y_0,z_0)d^3V = f(x_0,y_0,z_0) \qquad (38)$$

---

## DIVERGENCE FREE FIELDS

It may seem a bit discouraging that we did all this work to derive the differential form of Gauss' law $\vec{\nabla} \cdot \vec{E} = \rho/\varepsilon_0$ , and then end up, when we want to actually solve a problem, going back to the integral form of the equation. At this point, that is about all we can do to solve for explicit field patterns $\vec{E}$. However, the differential form begins to tell us about some general features of a vector field as we shall now see. With a lot more practice with the differential form of the field equations, and perhaps a computer thrown in, one can begin to solve for complex field shapes. In this text we will focus on what we can learn about general features and leave the solution of complex field shapes to a later course.

To see what we can learn about general features of a field, suppose that we have a velocity field $\vec{v}(x,y,z)$ , whose divergence is zero, i.e., it obeys the equation

$$\vec{\nabla} \cdot \vec{v}(x,y,z) = 0 \qquad (39)$$

We say that such a field is ***divergence free***. What can we say about the properties of such a field?

To answer that question, we will again go back to the integral form, by integrating Equation (1) over some volume V to get

$$\int_{\text{volume V}} \vec{\nabla} \cdot \vec{v}\, d^3V = 0 \qquad (40)$$

Now use the divergence theorem to convert this volume integral to a surface integral, giving

$$\int_{\substack{\text{closed}\\ \text{surface}}} \vec{v} \cdot d\vec{A} = 0 \qquad (41)$$

Equation (41) is our old equation for a vector field that has no sources or sinks. It is the equation for an incompressible, constant density fluid, a real one like water where water molecules are not being created or destroyed. Thus the condition that a vector field be divergence free, i.e., $\vec{\nabla} \cdot \vec{v} = 0$ or $\vec{\nabla} \cdot \vec{E} = 0$ or $\vec{\nabla} \cdot \vec{B} = 0$ , is that the field behaves like the velocity field of an incompressible fluid.

What kind of solutions are possible for a divergence free field? What are the solutions to the equation $\vec{\nabla} \cdot \vec{v} = 0$ ?

The answer is at least as complex as the behavior of water. You have seen water flow smoothly in a lazy river. That is called ***laminar flow***. Such laminar flow is one solution to $\vec{\nabla} \cdot \vec{v} = 0$ . But in a fast flowing stream there can be complex eddies called ***turbulence***. Turbulent flow is also a solution to the equation $\vec{\nabla} \cdot \vec{v} = 0$ .

You can now see that the equation $\vec{\nabla} \cdot \vec{v} = 0$ puts a restriction on the field $\vec{v}$ , but still allows an enormous range of solutions. Because of your familiarity with the flow of water you have some insight into what these solutions can be.

# APPENDIX — DERIVATION OF FLUX EQUATION (14)

Earlier in the chapter we had the following formula for the flux out of the x directed faces of the small cube $\Delta V = \Delta x \Delta y \Delta z$

$$\begin{matrix}\text{flux out}\\ \text{of x directed}\\ \text{faces of } \Delta V\end{matrix} = \left[\frac{\bar{E}_x(x+\Delta x,y,z) - \bar{E}_x(x,y,z)}{\Delta x}\right]\Delta x\Delta y\Delta z$$

(13) repeated

where $\bar{E}_x(x+\Delta x,y,z)$ and $\bar{E}_x(x,y,z)$ were the average values of $E_x$ on the two x directed faces of the cube.

In Equation (14) we replaced the average values $\bar{E}_x$ by the values $E_x(x+\Delta x,y,z)$ and $E_x(x,y,z)$ at the lower front corners as shown in Figure (2), repeated here giving

$$\begin{matrix}\text{flux out}\\ \text{of x directed}\\ \text{faces of } \Delta V\end{matrix} = \underset{\Delta x\to 0}{\text{limit}}\left[\frac{E_x(x+\Delta x,y,z) - E_x(x,y,z)}{\Delta x}\right]\Delta V$$

(14) repeated

What we are doing is in going from Equation (13) to (14) is to neglect the y and z dependence of $E_x$ while developing an equation for the x dependence. This step needs justification.

To see what effect the y and z dependence has, let us start by approximating the average value of $E_x$ over the entire x faces by the average of the top and bottom values of the front side of $\Delta A_X$, i.e., the average of $E_x$ at points (1) and (3) on the left and points (2) and (4) on the right as shown in Figure (5).

This is a rather crude approximation for the average over the face, but begins to show us what the effect of the y and z dependence of $E_x$ is.

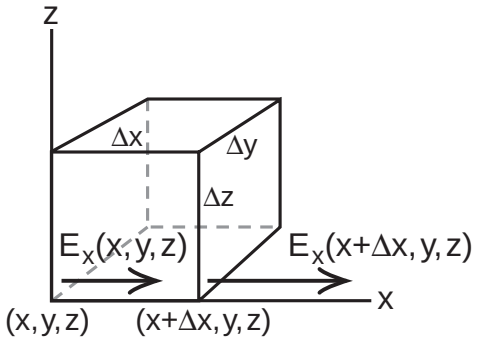



**Figure 2 (repeated)**
***Electric field at the lower front corners.***

To evaluate $E_x$ at $(x, y, z + \Delta z)$, up at point (3), we can use a Taylor series expansion. So far we have discussed a Taylor series expansion only of a function of a single variable f(x). The expansion was, from Equation (2-44 of Calculus Chapter 2)

$$f(x-x_0) = f(x_0) + \frac{\partial f}{\partial x}(x-x_0) + \frac{1}{2!}\frac{\partial^2 f}{\partial x^2}(x-x_0)^2 + \cdots$$

(2-44) repeated

which is good for small steps $(x-x_0)$.

What we are doing when we go from point (1) to point (3) in Figure (2), is keeping the values of x and y constant, and looking at the change in $E_x$ as we vary z. Thus in going up, we have a function $E_x(z)$ that is only a function of z, and we can use our old Taylor series expansion to get

$$\begin{aligned} E_x(x,y,z+\Delta z) &= E_x(x,y,z) \\ &\quad + \frac{\partial E_x(x,y,z)}{\partial z}(\Delta z) \\ &\quad + \frac{1}{2}\frac{\partial^2 E_x(x,y,z)}{\partial z^2}(\Delta z)^2 \\ &\quad + \cdots \end{aligned} \qquad (42)$$

where $\Delta z$ is analogous to the step $(x-x_0)$ in the Taylor series formula.

Because we are eventually going to take the limit as $\Delta z$ goes to zero, we will be able to neglect terms of order $(\Delta z)^2$ compared to $\Delta z$. Because of that, it is sufficient to write

$$\begin{aligned} E_x(x,y,z+\Delta z) &= E_x(x,y,z) \\ &\quad + \frac{\partial E_x}{\partial z}\Delta z \\ &\quad + \text{terms of order } \Delta z^2 \end{aligned} \qquad (42a)$$

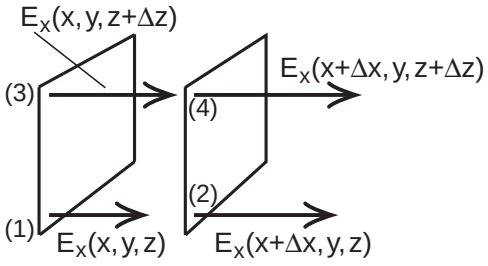



**Figure 5**
***Electric field at four positions.***

When we take the average of $E_x$ at points (1) and (3), a result we will call $\bar{E}_x(x)_{1,3}$, we get

$$\bar{E}_x(x)_{1,3} = \frac{E_x(x,y,z) + E_x(x,y,z+\Delta z)}{2}$$
$$= E_x(x,y,z) + \frac{1}{2}\frac{\partial E_x(x,y,z)}{\partial z}\Delta z + O(\Delta z^2) \quad (43)$$

where $O(\Delta z^2)$ means terms of order $(\Delta z^2)$.

A similar argument gives the average $\bar{E}_x(x+\Delta x)_{2,4}$ at points (2) and (4)

$$\bar{E}_x(x+\Delta x)_{2,4} = \frac{E_x(x+\Delta x,y,z) + E_x(x+\Delta x,y,z+\Delta z)}{2}$$
$$= E_x(x+\Delta x,y,z) + \frac{1}{2}\frac{\partial E_x(x+\Delta x,y,z)}{\partial z}\Delta z + O(\Delta z^2) \quad (44)$$

Using our 2 point averages in Equation (13) for the flux out of $\Delta V$ gives us

$$\begin{matrix}\text{flux out of}\\ \text{x directed face}\\ \text{of } \Delta V \text{ for 2}\\ \text{point average}\end{matrix} = \left[\frac{\bar{E}_x(x+\Delta x)_{2,4} - \bar{E}_x(x)_{1,3}}{\Delta x}\right]\Delta V$$

$$= \frac{\Delta V}{\Delta x}\left[E_x(x+\Delta x,y,z) + \frac{1}{2}\frac{\partial E_x(x+\Delta x,y,z)}{\partial z}\Delta z \right.$$
$$\left. - E_x(x,y,z) - \frac{1}{2}\frac{\partial E_x(x,y,z)}{\partial z}\Delta z + O(\Delta z^2)\right]$$

$$= \left[\frac{E_x(x+\Delta x,y,z) - E_x(x,y,z)}{\Delta x}\right]\Delta V$$
$$+ \frac{1}{2}\left[\frac{\dfrac{\partial E_x(x+\Delta x,y,z)}{\partial z} - \dfrac{\partial E_x(x,y,z)}{\partial z}}{\Delta x}\right]\Delta z \Delta V \quad (45)$$
$$+ O(\Delta z^2)\Delta V$$

When we go to the limit that $\Delta x$ goes to zero, we see that we get the partial derivatives

$$\lim_{\Delta x \to 0}\left[\frac{E_x(x+\Delta x,y,z) - E_x(x,y,z)}{\Delta x}\right] = \frac{\partial E_x(x,y,z)}{\partial x} \quad (46)$$

$$\lim_{\Delta x \to 0}\left[\frac{\dfrac{\partial E_x(x+\Delta x,y,z)}{\partial z} - \dfrac{\partial E_x(x,y,z)}{\partial z}}{\Delta x}\right] = \frac{\partial^2 E_x(x,y,z)}{\partial x \partial z} \quad (47)$$

Thus Equation (45) is taking on the form

$$\begin{matrix}\text{flux out of}\\ \text{x faces of}\\ \Delta V \text{ for 2}\\ \text{point average}\end{matrix} = \Delta V\left[\frac{\partial E_x}{\partial x} + \frac{\partial^2 E_x}{\partial x \partial z}\Delta z + O(\Delta z^2)\right] \quad (48)$$

We see that corrections due to the z dependence of $E_x$ are of magnitude $\Delta z$ times the partial second derivative $\partial^2 E_x/\partial x \partial z$. As long as all derivatives of $E_x$ are bounded, stay finite as we take the limit as $\Delta x$, $\Delta y$, and $\Delta z$ go to zero, then the $\Delta z$ term in Equation (48) becomes negligently small, which means that in the limit we can neglect the z dependence of $E_x$, at least in this two point approximation.

Our 2 point approximation to the average of $E_x$ can be improved by using more points. If we included the back points at $(y+\Delta y)$, we would add terms to Equation (48) of the form

$$\frac{\partial^2 E_x}{\partial x \partial y}\Delta y + O(\Delta z^2) \quad (49)$$

terms which would go to zero in the limit $\Delta y \to 0$. All points we add in to the average will give terms proportional to $\Delta x$ or $\Delta y$ or some combination, and all these terms will go to zero when we take the limit as $\Delta x$, $\Delta y$, and $\Delta z$ goes to zero. Thus, it is an exact result that, in the limit that $\Delta v \to 0$, only the x dependence of $E_x$ has to be taken into account, provided all derivatives of $E_x$ are finite.

# Calculus 2000-Chapter 8

# Curl

## ABOUT THE CURL

*In the Physics text, we saw that a vector field was uniquely determined by formulas for the surface integral and the line integral. In the last chapter, we saw that the divergence, such as* $\vec{\nabla} \cdot \vec{E}$ *, represented the surface integral shrunk down to an infinitesimal scale. In this chapter, we study the curl, which is the line integral shrunk down to an infinitesimal scale. Here our emphasis will be on the application of the curl to electric and magnetic fields. In the final chapters of this text, Chapters 12 and 13, we develop an intuitive picture of the curl applied to the velocity field of fluids such as water and superfluid helium. The curl of the velocity field is called* ***vorticity****, a concept that plays a fundamental role in understanding such phenomena as quantum vortices and turbulence.*

## INTRODUCTION TO THE CURL

The partial derivative operator

$$\vec{\nabla} = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}$$

has now appeared in our formulas for the ***gradient*** of a scalar field f(x,y,z)

$$\vec{\nabla} f(x,y,z) = \hat{x}\frac{\partial f}{\partial x} + \hat{y}\frac{\partial f}{\partial y} + \hat{z}\frac{\partial f}{\partial z} \qquad (1)$$

in the ***divergence*** of a vector field $\vec{E}(x,y,z)$

$$\vec{\nabla}\cdot\vec{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \qquad (2)$$

and in the *Laplacian*

$$\vec{\nabla}\cdot\vec{\nabla} f = \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \qquad (3)$$

While $\vec{\nabla}$ is an operator in the sense that it only has a value when operating on some field, we see that it acts very much like a vector. This suggests that we may encounter other vector like operations involving $\vec{\nabla}$.

In our discussion of vectors in Chapter 2 of the Physics text, we saw that there were two kinds of vector products, the scalar or dot product

$$C = \vec{A}\cdot\vec{B} = (A_xB_x + A_yB_y + A_zB_z) \quad \begin{matrix}\textit{scalar}\\ \textit{product}\end{matrix} \qquad (4)$$

and the vector cross product

$$\vec{C} = \vec{A}\times\vec{B} \qquad \textit{vector cross product} \qquad (5)$$

where the formulas for the components of $\vec{C}$ were

$$\begin{aligned} C_x &= A_yB_z - A_zB_y \\ C_y &= A_zB_x - A_xB_z \\ C_z &= A_xB_y - A_yB_x \end{aligned} \qquad (6)$$

We saw that the vector $\vec{C} = \vec{A}\times\vec{B}$ was oriented perpendicular to the plane of the vectors $\vec{A}$ and $\vec{B}$, the choice of which direction being given by the right hand rule as shown in Figure (1). The magnitude was $C = AB\sin\theta$ which is maximum when $\vec{A}$ and $\vec{B}$ are perpendicular and zero when parallel.

The vector cross product seems like a rather peculiar mathematical construct, but it plays an important role in physics, particularly in describing rotational motion. You will recall that the angular analogy to Newton's second law was

$$\vec{\tau} = \frac{d\vec{L}}{dt} \qquad (7)$$

where the torque, $\vec{\tau} = \vec{r}\times\vec{F}$, is what we called the angular force, and $\vec{L} = \vec{r}\times\vec{p}$ is the angular momentum. Despite the appearance of two cross products in Equation (7), the equation led to a very successful prediction of the motion of a gyroscope at the end of Chapter 12 in the Physics text (see page12-18).

With this background, we see that there is one more natural vector product involving the operator $\vec{\nabla}$. It is the cross product of $\vec{\nabla}$ with some vector field like $\vec{E}$, $\vec{B}$, or $\vec{v}$. The cross product, for example with $\vec{B}$, is called the ***curl*** of $\vec{B}$.

$$\begin{aligned} \vec{\nabla}\times\vec{B} = {} & \hat{x}(\nabla_yB_z - \nabla_zB_y) \\ & + \hat{y}(\nabla_zB_x - \nabla_xB_z) \qquad \textit{curl} \\ & + \hat{z}(\nabla_xB_y - \nabla_yB_x) \end{aligned} \qquad (8)$$

With all these derivatives in the formula for $\vec{\nabla}\times\vec{B}$, the concept of the curl looks rather formidable. Later in this chapter we will discuss the formula for the curl in cylindrical coordinates. That formula looks even worse than Equation (8). However when we apply the curl in cylindrical coordinates to a problem with cylindrical symmetry, we end up with a simple, easily applied formula (which we will see in Equation 58).

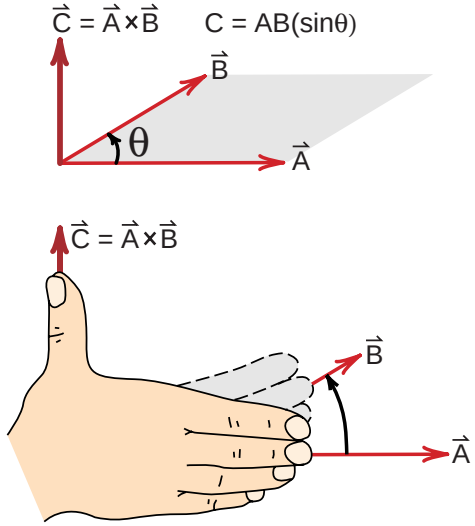



**Figure 1**
***Right hand rule for the cross product.***
***(Discussed in Physics 2000, page 2-15.)***

As we have mentioned several times now, to determine a vector field we need formulas for the surface integral and the line integral. In the last chapter we saw that when we go to the small scale limit, the volume integral becomes a divergence. An example was Gauss' law which in the integral form was

$$\int \vec{E} \cdot d\vec{A} = \frac{Q_{in}}{\varepsilon_0} \qquad (9)$$

It became the differential equation

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0} \qquad (10)$$

In this chapter we will see that the differential limit of the line integral is the curl. We will see, for example, that the old form of Ampere's law (when $\partial\vec{E}/\partial t = 0$)

$$\oint \vec{B} \cdot d\vec{\ell} = \mu_0 i_{in} \qquad (11)$$

becomes the differential equation

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{i}(x,y,z) \qquad (12)$$

where $\vec{i}(x,y,z)$ is the current density.

In our discussion of divergence, one of the important results was the divergence theorem

$$\int_S \vec{E} \cdot d\vec{A} = \int_V \vec{\nabla} \cdot \vec{E}\, d^3V \qquad \textit{divergence theorem} \qquad (13)$$

where V is the volume bounded by a closed surface S and $d^3V = dx\,dy\,dz$. The divergence theorem allowed us to immediately go back and forth between surface integrals and volume integrals.

An important result of this chapter is what one could call the ***curl theorem***, but which is known as ***Stokes' law***. It is

$$\oint_{\substack{\text{around}\\ \text{closed path}}} \vec{B} \cdot d\vec{\ell} = \int_{\substack{\text{area of}\\ \text{closed path}}} (\vec{\nabla} \times \vec{B}) \cdot d\vec{A} \qquad \textit{Stokes' law} \qquad (14)$$

which relates the line integral of $\vec{B}$ around a closed path to an integral of the curl of $\vec{B}$ over any area bounded by the closed path. An example of a closed path is the wire loop shown in Figure (2). One of the areas bounded by this closed path is that of the soap film.

Our discussion of the curl will proceed through the remaining chapters of the text. In this chapter we will focus on deriving Stokes' theorem and applying that theorem to the theory of electricity and magnetism. This allows us to finish translating Maxwell's equations from the integral to the differential form.

In Chapter 9 we derive a set of equations called ***vector identities*** that simplify working with formulas involving the curl. We will use the vector identities to show that Maxwell's equations in empty space become the wave equations for electromagnetic fields.

In Chapter 11 we find that the wave equation for electromagnetic fields in the presence of electric charge and current is considerably simplified by expressing the magnetic field as the curl of a new kind of a vector field called the ***vector potential*** $\vec{A}$. This is a rather technical subject, the study of which can be put off for a while. We placed this material where we did so that you could see what happens to the electromagnetic wave equation when sources are present.

In Chapter 12 we apply the curl to the velocity field $\vec{v}$. It is in that chapter where you can develop the best intuitive picture of the curl. If you want to put off for a while studying the wave equation for electromagnetic fields, you can go directly from this chapter to Chapter 12 and build your intuition for curl.

In case you were wondering about Chapter 10, it deals with the extension of the continuity equation to handle compressible conserved flows, like the flow of electric charge. We discover from this work a rather remarkable result, namely that ***Maxwell's equations require that electric charge be conserved***. This is one of the first completely new physical predictions we get by going to the differential form of Maxwell's equations.

**Figure 2**
***Example of a surface bounded by a closed path (wire loop).***

## STOKES' LAW

As we noted, Stokes' law, Equation (14) allows us to convert from a line integral around a closed path to a surface integral over the area bounded by the path. Once we have derived Stokes' law, it will be quite easy to use it to convert to differential equations the two Maxwell equations involving path integrals.

To derive Stokes' law, we begin by calculating the path integral of some vector field $\vec{B}$ around a small rectangular path of sides $\Delta x$ and $\Delta y$ shown in Figure (3). Our arguments will be somewhat similar to those we used to derive the divergence theorem.

The line integral around the rectangle $\Delta x \Delta y$ can be written as the four integrals

$$\oint_{\substack{\text{around}\\ \Delta x\Delta y}} \vec{B}\cdot d\vec{\ell} = \int_1^2 \vec{B}\cdot d\vec{\ell} + \int_2^3 \vec{B}\cdot d\vec{\ell} + \int_3^4 \vec{B}\cdot d\vec{\ell} + \int_4^1 \vec{B}\cdot d\vec{\ell} \tag{15}$$

Along the path from point (1) to point (2), along the bottom of the rectangle, we are integrating in the x direction, thus

$$\int_1^2 \vec{B}\cdot d\vec{\ell} = \int_1^2 B_x d\ell_x \tag{16}$$

The integral of $B_x d\ell_x$ over the bottom side can be written as

$$\int_1^2 B_x d\ell_x = \bar{B}_x(x,y-\Delta y/2)\Delta x \tag{17}$$

where $\bar{B}_x(x,y-\Delta y/2)$ is the average value of $B_x$ along the lower edge, a distance $\Delta y/2$ below the center (x,y) of the rectangle.

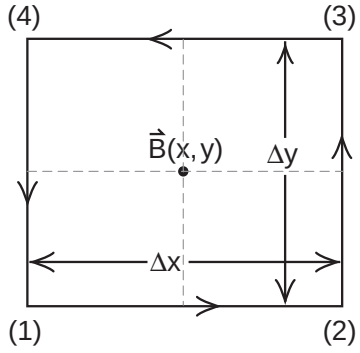



**Figure 3**
***Calculating the integral of $\vec{B}\cdot d\vec{l}$ around a small rectangular path centered at the point (x,y).***

The integral up the right hand side becomes

$$\int_2^3 \vec{B}\cdot d\vec{\ell} = \int_2^3 B_y d\ell_y = \bar{B}_y(x+\Delta x/2,y)\Delta y \tag{18}$$

where $\bar{B}_y(x+\Delta x/2,y)$ is the average value of $B_y$ along the right side, out at a distance $\Delta x/2$ from the center.

On the top side, we are integrating in the –x direction, the dot product $\vec{B}\cdot d\vec{\ell}$ is negative, and we get

$$\int_3^4 \vec{B}\cdot d\vec{\ell} = \int_3^4 -B_x d\ell_x = -\bar{B}_x(x,y+\Delta y/2)\Delta x \tag{19}$$

where $\bar{B}_x(x,y+\Delta y/2)$ is the average value of $B_x$ on the top edge.

Going back down from point (4) to point (1) we are going in the –y direction, $\vec{B}\cdot d\vec{\ell} = -B_y d\ell_y$ and we get

$$\int_4^1 \vec{B}\cdot d\vec{\ell} = \int_4^1 -B_y d\ell_y = -\bar{B}_y(x-\Delta x/2,y)\Delta y \tag{20}$$

Using Equations (17) through (20) in (15) gives, after some rearranging

$$\oint_{\substack{\text{around}\\ \Delta x\Delta y}} \vec{B}\cdot d\vec{\ell} = \left[\frac{\bar{B}_y(x+\Delta x/2,y)-\bar{B}_y(x-\Delta x/2,y)}{\Delta x}\right]\Delta x\Delta y - \left[\frac{\bar{B}_x(x,y+\Delta y/2)-\bar{B}_x(x,y-\Delta y/2)}{\Delta y}\right]\Delta x\Delta y \tag{21}$$

As a first approximation to Equation (21), we could replace the average values of $\bar{B}_x, \bar{B}_y$ on the four sides by the actual values of $B_x$, $B_y$ at the center of each side. For example, since the center of the side from (2) to (3) is at the point $(x+\Delta x/2,y)$, we would be making the substitution for that side of

$$\bar{B}_y(x+\Delta x/2,y) \rightarrow B_y(x+\Delta x/2,y) \tag{22}$$

I.e., we would be removing the bars over the values of B in Equation (21).

When we remove the bars and then take the limit as $\Delta x \to 0$ and $\Delta y \to 0$, the first square bracket in Equation (21) becomes the partial derivative of $B_y$ with respect to x

$$\lim_{\Delta x \to 0}\left[\frac{B_y(x+\Delta x/2,y) - B_y(x-\Delta x/2,y)}{\Delta x}\right] = \frac{\partial B_y}{\partial x} \tag{23}$$

and the second square bracket in Equation (21) becomes $\partial B_x/\partial y$. In this approximation, Equation (21) becomes

$$\oint_{\substack{\text{around}\\ \Delta x\Delta y}} \vec{B}\cdot d\vec{\ell} \;=\; \left[\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right](\Delta x \Delta y) \tag{24}$$

The approximation we made to get Equation (24), which was replacing the average value of B along a line by the value at the center of the line, assumes that variations along the line (e.g. changes in $B_x$ in the x direction) are not as important as variations perpendicular to the line (e.g. changes in $B_x$ in the y direction). This is somewhat similar to the situation we had in our derivation of the divergence theorem where changes in the field were important in one direction and not in the other.

In the appendix to Chapter 7 we used a Taylor series expansion to show that as $\Delta x$, $\Delta y$ or $\Delta z$ went to zero, the variations we ignored went to zero faster than the variations we kept. They were proportional to a higher power of $\Delta x$, $\Delta y$ or $\Delta z$, and therefore did not contribute in the calculus limit.

We leave it as an exercise for the ambitious reader to show, using arguments similar to those made in the appendix to Chapter 7, that by replacing average values $\bar{B}_x$ and $\bar{B}_y$ by center values $B_x$ and $B_y$, we are making errors that go to zero faster than the terms we keep. I.e., show that the errors are of the order $\Delta x$, $\Delta y$ or $\Delta z$ smaller than the terms we keep.

With Equation (24), we have the formula for the line integral around one small rectangle lying in the xy plane. We can generalize this result by turning the area element $(\Delta x \Delta y)$ into a vector $\overrightarrow{\Delta A}$. An area vector $\overrightarrow{\Delta A}$ is perpendicular to the surface as shown in Figure (4). In this case, where the surface is in the xy plane, we see that $\overrightarrow{\Delta A}$ is purely z directed, and we can write $\Delta x \Delta y = (\Delta A)_z$. With this notation Equation (24) becomes

$$\oint_{\substack{\text{around}\\ \Delta A}} \vec{B}\cdot d\vec{\ell} \;=\; \left[\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right](\Delta A)_z \tag{25}$$

Next, we notice that the z component of the curl of $\vec{B}$ is given by Equation (8) as

$$(\vec{\nabla}\times\vec{B})_z = (\nabla_x B_y - \nabla_y B_x) = \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \tag{26}$$

so that Equation (25) becomes

$$\oint_{\substack{\text{around}\\ \Delta A}} \vec{B}\cdot d\vec{\ell} \;=\; (\vec{\nabla}\times\vec{B})_z (\Delta A)_z \tag{27}$$

The obvious extension of Equation (27) to the case where our area $\Delta A$ does not happen to lie in the xy plane, where the vector $\overrightarrow{\Delta A}$ has components other than $(\Delta A)_z$, is to recognize that in Equation (27) we are looking at one term in the vector dot product

$$\oint_{\substack{\text{around}\\ \Delta A}} \vec{B}\cdot d\vec{\ell} \;=\; (\vec{\nabla}\times\vec{B})\cdot\overrightarrow{\Delta A} \tag{28}$$

---

**Exercise 1**

Suppose we have an area $\Delta y \Delta z$ as shown in Figure (5). Write out the formula for $\oint \vec{B}\cdot d\vec{\ell}$ around this area (i.e., repeat the steps in Equations 15-27 for this area).

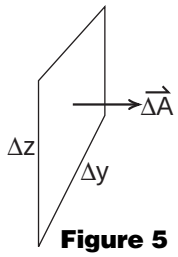



**Figure 5**

---

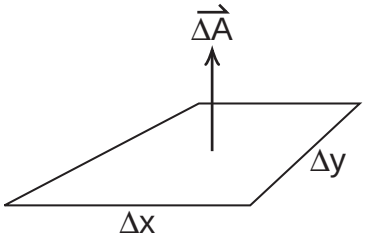



**Figure 4**
***Turning the area element (ΔxΔy) into a vector $\overrightarrow{\Delta A}$***

With Equation (28), we have the formula for the line integral around a small rectangular area $\overrightarrow{\Delta A}$ of any orientation. The final step is to determine the line integral around a finite loop like the wire loop with the soap film across it, shown in Figure (1).

The way we can do this is to conceptually cut the soap film up into many tiny rectangles as shown in Figure (6). Think of the soap film as being replaced by a window screen, with the rectangles being the holes in the window screen.

At each hole, each rectangle, we have a vector $\overrightarrow{\Delta A}_i$ that is oriented perpendicular to the surface as shown in Figure (7). The positive direction is determined by noting which way we are going around the loop, and then using the right hand rule. For Figure (6), the positive direction is up out of the paper.

Next we note that when two rectangles touch each other, the part of the line integrals on the touching sides cancel, and we are left with a line integral around the perimeter of the two rectangles as shown in Figure (8).

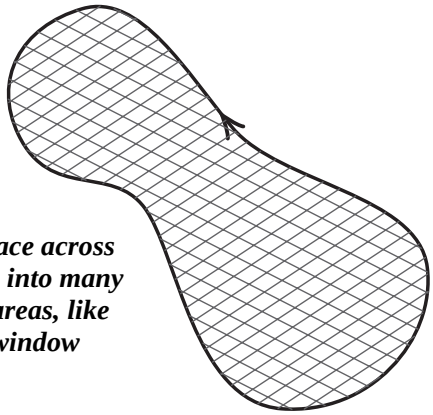


**Figure 6**
***Break the surface across the closed loop into many small surface areas, like the holes in a window screen.***

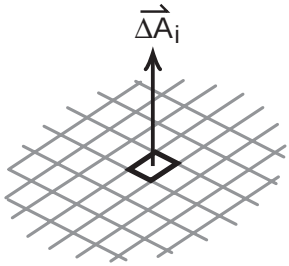



**Figure 7**
***Each small surface area is described by an area vector $\overrightarrow{\Delta A}_i$***

Applying this argument to all rectangles in Figure (6), we see that when we add up the line integrals for all the rectangles, we end up with the line integral around the outside perimeter of the surface. Mathematically we can write this as

$$\oint_{\substack{\text{around}\\ \text{whole}\\ \text{surface}}} \vec{B}\cdot d\vec{\ell} \;=\; \begin{array}{l}\text{sum of the line}\\ \text{integrals around}\\ \text{each small}\\ \text{area } \overrightarrow{\Delta A}_i\end{array} \qquad (29)$$

Using Equation (28) for the line integral around $\overrightarrow{\Delta A}_i$ we get

$$\oint_{\substack{\text{around}\\ \text{whole}\\ \text{surface}}} \vec{B}\cdot d\vec{\ell} \;=\; \sum_i (\vec{\nabla}\times\vec{B})\cdot\overrightarrow{\Delta A}_i \qquad (30)$$

Taking the limit as the $\overrightarrow{\Delta A}_i$ goes to zero turns this sum into an integral, giving

$$\oint_{\substack{\text{around}\\ \text{perimeter of}\\ \text{a surface S}}} \vec{B}\cdot d\vec{\ell} \;=\; \int_{\substack{\text{over the}\\ \text{surface S}}} (\vec{\nabla}\times\vec{B})\cdot d\vec{A} \qquad \textit{Stokes' law}$$

(31)

which is Stokes' law. It says that we get the line integral of any vector field $\vec{B}$ around the perimeter of a surface S by integrating the flux of $(\vec{\nabla}\times\vec{B})$ out through the surface.

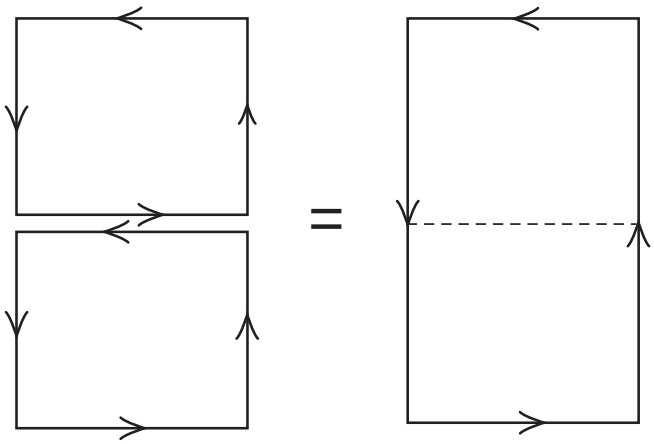


**Figure 8**
***When two rectangles touch, the line integrals on the paths between them cancel, leaving a line integral around the perimeter of the two rectangles.***

In the future we will shorten our notation by letting C be some closed path, and the surface S be a surface like our soap film, that is bounded by the path. Then we simply write

$$\oint_C \vec{B}\cdot d\vec{\ell} = \int_S (\vec{\nabla}\times\vec{B})\cdot d\vec{A} \qquad \textit{Stokes' law}$$

(31a)

Our use of the soap film analogy for the surface S is important for it emphasizes the fact that there is no one correct surface. Just as you can change the shape of a soap film by gently blowing on it (don't blow a bubble), you can use different surfaces S as long as they are bounded by the same circuit C.

We also want to emphasize that the quantity $(\vec{\nabla}\times\vec{B})$ is itself a vector field, and that the integral of $(\vec{\nabla}\times\vec{B})\cdot d\vec{A}$ over a surface is the flux of $(\vec{\nabla}\times\vec{B})$ through that surface. Thus, we should remember Stokes' law as telling us that ***the line integral of $\vec{B}$ around the circuit C is equal to the flux of $(\vec{\nabla}\times\vec{B})$ through the circuit C***.

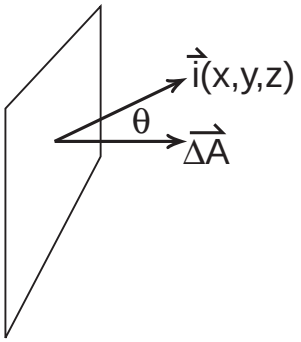



**Figure 9**
***When the current flows at an angle θ as shown, the total current through ΔA is i(x,y,z) ΔA cos θ.***

## AMPERE'S LAW

The original form of Ampere's law, before Maxwell's addition of the $\partial\Phi_E/\partial t$ term, was given in Chapter 29 of the Physics text as

$$\oint_{\text{any closed path}} \vec{B}\cdot d\vec{\ell} = \mu_0 I_{\text{enclosed}} \qquad (29\text{-}26)$$

It says that the line integral of $\vec{B}$ around any closed path is equal to $\mu_0$ times the total current flowing through that path. Since Stokes' law tells us that the line integral of $\vec{B}$ around any closed path is equal to the total flux of $(\vec{\nabla}\times\vec{B})$ through that path, there must be a close relationship between the vector field $(\vec{\nabla}\times\vec{B})$ and the electric current. That is the relationship we want to establish.

The first step is to express the total current i through a closed path in terms of the current density $\vec{i}(x,y,z)$. The current density $\vec{i}(x,y,z)$ is a vector field whose direction at each point in space is the direction of flow on the electric current $\vec{i}$ there, and whose magnitude is equal to the density of current, which has the dimensions of the number of amperes per square meter.

Calculating the electric current through a small area element $\Delta\vec{A}$ is analogous to calculating the flux of water through an area element $\Delta\vec{A}$, a calculation we did in Equation (3) of Chapter 29 of the Physics text. From Figure (9), you can see that the current through $\Delta A$ will be a maximum, will have the value $\vec{i}(x,y,z)\Delta A$ when the area $\Delta A$ is perpendicular to the flow. This is when the vector $\Delta\vec{A}$ is parallel to $\vec{i}(x,y,z)$. For any other orientation of $\Delta\vec{A}$, the current $\Delta I$ through $\Delta\vec{A}$ will be equal to $\vec{i}(x,y,z)\Delta A\cos\theta$ which is equal to the dot product of the vectors $\vec{i}(x,y,z)$ and $\Delta\vec{A}$. Thus

$$\Delta I = \vec{i}(x,y,z)\cdot\Delta\vec{A} = \text{current through an area element } \Delta\vec{A} \qquad (32)$$

To calculate the total current $I_{enclosed}$ through an entire surface S, we break the surface up into small areas $\Delta\vec{A}_i$ as we did in Figure (6), calculate the current $\Delta I_i$ through each $\Delta\vec{A}_i$, and add up all the $\Delta I_i$ to get the total.

$$I_{enclosed} = \sum_i \Delta I_i = \sum_i \vec{i}(x_i,y_i,z_i)\cdot\Delta\vec{A}_i \qquad (33)$$

Taking the limit as the $\Delta\vec{A}_i$ go to zero size gives us the surface integral

$$I_{enclosed} = \int_{\substack{\text{surface}\\ \text{bounded}\\ \text{by path C}}} \vec{i}(x,y,z)\cdot d\vec{A} \qquad \begin{matrix}\textit{total current}\\ \textit{through a}\\ \textit{closed path C}\end{matrix} \qquad (34)$$

Using our new formula for $I_{enclosed}$ in Ampere's law, Equation (29-26), gives

$$\oint_{\substack{\text{any}\\ \text{closed}\\ \text{path}}} \vec{B}\cdot d\vec{\ell} = \mu_0 \int_{\substack{\text{over the area}\\ \text{bounded by}\\ \text{the closed path}}} \vec{i}(x,y,z)\cdot d\vec{A} \qquad (35)$$

Following a procedure similar to the one we used in our discussion of Gauss' law in Chapter 7, we will use Stokes' law to convert the line integral of $\vec{B}$ to a surface integral, so that both terms in Ampere's law are surface integrals. With

$$\oint_C \vec{B}\cdot d\vec{\ell} = \int_S (\vec{\nabla}\times\vec{B})\cdot d\vec{A} \qquad (31) \text{ repeated}$$

Equation (35) becomes

$$\int_{\substack{\text{surface}\\ S}} (\vec{\nabla}\times\vec{B})\cdot d\vec{A} = \int_{\substack{\text{surface}\\ S}} \mu_0\vec{i}(x,y,z)\cdot d\vec{A} \qquad (36)$$

where we took the constant $\mu_0$ inside the integral. The surfaces for the two integrals only have to have the same perimeter C, but we are free to choose identical surfaces, and thus combine the two integrals into one giving

$$\int_{\substack{\text{any}\\ \text{surface}\\ S}} \left[(\vec{\nabla}\times\vec{B}) - \mu_0\vec{i}(x,y,z)\right]\cdot d\vec{A} = 0 \qquad (37)$$

We then argue that if Equation (37) is to hold for any surface S, the only way for that to happen is to set the integrand, the stuff in the square brackets, equal to zero, giving

$$\boxed{\vec{\nabla}\times\vec{B} = \mu_0\vec{i}(x,y,z)} \qquad (38)$$

Equation (38) is the differential form of the original Ampere's law

$$\oint \vec{B}\cdot d\vec{\ell} = \mu_0 I_{enclosed} \qquad (29\text{-}26) \text{ repeated}$$

In Chapter 32 of the Physics text we explained why Maxwell added a term to Ampere's law to get

$$\oint_{\substack{\text{around a}\\ \text{closed}\\ \text{circuit C}}} \vec{B}\cdot d\vec{\ell} = \mu_0 I_{enclosed} + \mu_0\varepsilon_0\frac{d\Phi_E}{dt} \qquad (32\text{-}11)$$

where $\Phi_E$, the electric flux through the closed circuit is given by

$$\Phi_E = \int_S \vec{E}\cdot d\vec{A} \qquad (39)$$

and S is any surface bounded by the closed circuit C.

To include the $d\Phi_E/dt$ term in our differential form of Ampere's law, we need to evaluate

$$\frac{d}{dt}\Phi_E(t) = \frac{d}{dt}\int_S \vec{E}(x,y,z,t)\cdot d\vec{A} \qquad (40)$$

where the field E is not only a function of space (x,y,z) but also of time (t).

On the left side of Equation (40) we have $d\Phi_E(t)/dt$ which is simply the time derivative of some function $\Phi_E(t)$ of time. That is a straightforward derivative. On the right, we have the derivative of the integral of a quantity $\vec{E}(x,y,z,t)$ which is a function of four variables. What we are going to do this one time, is to be very careful about how we bring the time derivative inside the integral, and see what we get when we do.

Our first step will be to write the integral over the surface as the sum over many small but finite areas $\Delta\vec{A}_i$

$$\frac{d\Phi_E}{dt} = \frac{d}{dt}\sum_i \vec{E}(x_i,y_i,z_i,t)\cdot\Delta\vec{A}_i \qquad (41)$$

where $(x_i,y_i,z_i)$ is the coordinate of the area element $\Delta\vec{A}_i$. By working with a sum of finite terms, we can see that the change in time of the sum will be the sum of the changes in each term

$$\frac{d\Phi_E}{dt} = \sum_i \frac{d}{dt}\left[\vec{E}(x_i,y_i,z_i,t)\cdot\Delta\vec{A}_i\right] \qquad (42)$$

During this calculation, we are keeping the surface S and all the $\Delta\vec{A}_i$ fixed. At any given $\Delta\vec{A}_i$ the only thing that is allowed to change is the field $\vec{E}$ at the point $(x_i,y_i,z_i)$. Thus we have

$$\frac{d\Phi_E}{dt} = \sum_i \left[\frac{d\vec{E}(x_i,y_i,z_i,t)}{dt}\right]\cdot\Delta\vec{A}_i \qquad (43)$$

The term in the square brackets is the change in the variable $\vec{E}(x,y,z,t)$ as we change the time (t) while holding the other three variables constant at $x = x_i$, $y = y_i$, $z = z_i$. This is precisely what we mean by the ***partial derivative*** of $\vec{E}(x,y,z,t)$ with respect to (t).

$$\frac{d}{dt}\vec{E}(x_i,y_i,z_i,t) = \left.\frac{\partial\vec{E}(x,y,z,t)}{\partial t}\right|_{\substack{x=x_i\\y=y_i\\z=z_i}} \qquad (44)$$

Thus we have

$$\frac{d\Phi_E}{dt} = \sum_i \left.\frac{\partial\vec{E}(x,y,z,t)}{\partial t}\right|_{\substack{x=x_i\\y=y_i\\z=z_i}}\cdot\Delta\vec{A}_i \qquad (45)$$

We can now go back to the limit as $\Delta\vec{A}_i$ goes to zero, giving

$$\frac{d\Phi_E}{dt} = \int_S \frac{\partial\vec{E}(x,y,z,t)}{\partial t}\cdot d\vec{A} \qquad (46)$$

Writing $d\Phi_E/dt$ in Equation (46) as an integral gives

$$\frac{d}{dt}\int_{\substack{\text{fixed}\\\text{surface}\\S}} \vec{E}(x,y,z,t)\cdot d\vec{A} = \int_{\substack{\text{fixed}\\\text{surface}\\S}} \frac{\partial\vec{E}(x,y,z,t)}{\partial t}\cdot d\vec{A} \qquad (47)$$

In writing Equation (47) we placed special emphasis on the fact that the surface S (and also the $\Delta\vec{A}_i$'s) were fixed, did not change with time. Later, in the first fluid dynamics chapter, we will want to calculate the rate of change of flux through a moving surface. (In that case it will be a surface that moves with the fluid particles.) When we allow the surface S to move, then in going from Equation (42) to (43), we get more terms representing changes in the $\Delta\vec{A}_i$.

But with the fixed surface, Equation (47) tells us that we can bring the time derivative inside the integral if we change the derivative to a partial derivative with respect to time.

---

**Exercise 2**

Start from the integral form of Ampere's law

$$\oint\vec{B}\cdot d\vec{\ell} = \mu_0 I_{enclosed} + \mu_0\varepsilon_0\frac{d\Phi_E}{dt} \qquad (32\text{-}11)$$

Using Equation (39) for $\Phi_E$, and using Equation (47), show that the corresponding differential equation is

$$\boxed{\vec{\nabla}\times\vec{B} = \mu_0\vec{i} + \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t}} \qquad (48)$$

---

**Exercise 3**

As a review, start with all of Maxwell's equations in integral form, as summarized in Equation (32-19) of the Physics text

$$\int_{\text{closed surface}} \vec{E}\cdot d\vec{A} = \frac{Q_{in}}{\varepsilon_0} \qquad \text{Gauss' law}$$

$$\int_{\text{closed surface}} \vec{B}\cdot d\vec{A} = 0 \qquad \text{no monopole}$$

$$\oint\vec{B}\cdot d\vec{\ell} = \mu_0 I + \mu_0\varepsilon_0\frac{d\Phi_E}{dt} \qquad \text{Ampere's law} \qquad (32\text{-}19)$$

$$\oint\vec{E}\cdot d\vec{\ell} = \frac{-d\Phi_B}{dt} \qquad \text{Faraday's law}$$

and show that in differential form, the equations are

$$\vec{\nabla}\cdot\vec{E} = \frac{\rho}{\varepsilon_0} \qquad \text{Gauss' law}$$

$$\vec{\nabla}\cdot\vec{B} = 0 \qquad \text{no monopole}$$

$$\vec{\nabla}\times\vec{B} = \mu_0\vec{i} + \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t} \qquad \text{Ampere's law} \qquad (49)$$

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \qquad \text{Faraday's law}$$

---

## CURL OF THE MAGNETIC FIELD OF A WIRE

In the section after this, we will discuss the formula for the curl in cylindrical coordinates, a rather formidable looking formula. We will then apply it to the calculation of the curl $\vec{\nabla}\times\vec{B}$ of the magnetic field of a straight wire. A lot of terms are involved but, most of them go to zero and we are left with what appears to be a surprisingly simple result. The result should be no surprise however, if we first look at Ampere's law in differential form, as applied to the field of a wire.

The magnetic field produced by a steady current in a wire was shown in Figure (28-14) in Chapter 28 of the Physics text. The current (i) is confined to the wire, and the magnetic field travels in circles around the wire. If the current density is more or less uniform in the wire, then we have a circular magnetic field inside the wire also (a field you calculated in Exercise 4 of Chapter 29). The result is sketched in Figure (10).

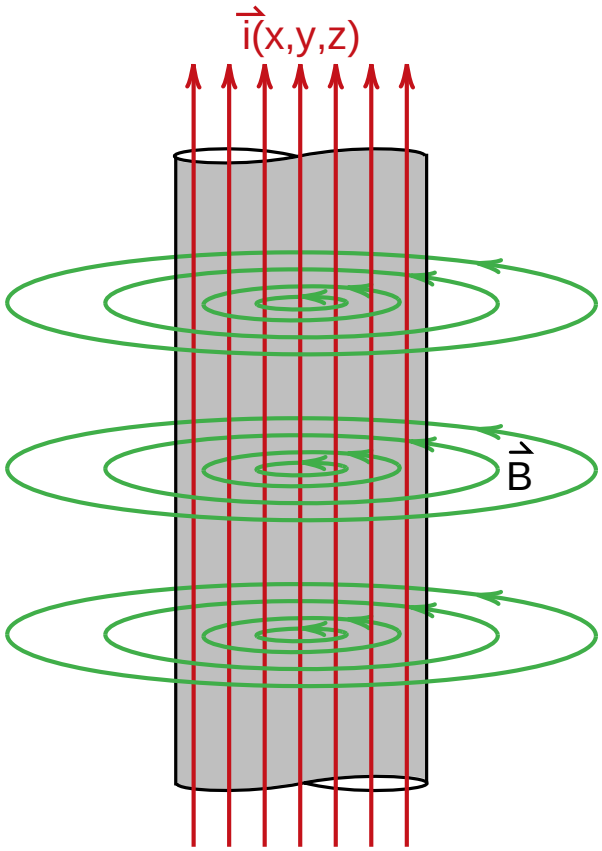



**Figure 10**
***The magnetic field inside and outside a wire carrying a uniform current.***

For a steady current, where $\partial\vec{E}/\partial t = 0$, Ampere's law in differential form is simply

$$\vec{\nabla}\times\vec{B} = \mu_0\vec{i}(x,y,z) \qquad \text{(38) repeated}$$

The first thing to note about Equation (38) is that in all places where the current density $\vec{i}(x,y,z)$ is zero, the curl $\vec{\nabla}\times\vec{B}$ must also be zero. Since the current is confined to the wire, $\vec{\nabla}\times\vec{B}$ must be confined there, and ***the curl of the magnetic field outside the wire must be zero***. It will take us several pages to obtain the same result using the formulas for the curl.

Next we note that the current density $\vec{i}(x,y,z)$ is not only confined to the wire, but also directed along the wire. Thus $\vec{\nabla}\times\vec{B}$ must not only be confined to the wire, ***but also directed along the wire*** as shown in Figure (11). As a result we know what $\vec{\nabla}\times\vec{B}$ must look like before we do any calculations.

In the next sections we will go through the calculation of the curl of this magnetic field. When we finally get the simple results described above, you can look upon that as a check that the formulas for curl are correct after all.

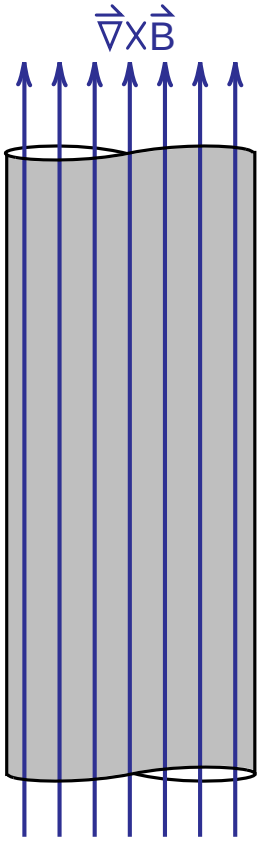



**Figure 11**
***The curl of that magnetic field, determined by*** $\vec{\nabla}\times\vec{B} = \mu_0\vec{i}(x,y,z)$.

## CURL IN CYLINDRICAL COORDINATES

In our study of the gradient in Chapter 3 and of Schrödinger's equation in Chapter 6, we saw that when a problem had cylindrical or spherical symmetry, there was a considerable advantage to using the formulas in cylindrical or spherical coordinates. Very often problems involving the curl, like the magnetic field of the current in a straight wire, have a cylindrical symmetry. For such problems it is much easier to work with the curl in cylindrical coordinates.

Deriving formulas for curl $\vec{\nabla}\times\vec{B}$ and divergence $\vec{\nabla}\cdot\vec{E}$ in cylindrical or spherical coordinates is made difficult because of the unit vectors. In Cartesian coordinates, the unit vectors are constant. But in other coordinate systems the unit vectors change as we move around in space. When we take the partial derivative of a vector, we also have to include the effects of changes in the unit vectors.

In the appendix to Chapter 4, where we calculated $\vec{\nabla}\cdot(\vec{\nabla}f) = \nabla^2 f$ in spherical polar coordinates, most of the calculation dealt with the changing unit vectors. In a more closely related example, suppose we have the vector $\vec{B}$ expressed in cylindrical coordinates as

$$\vec{B} = \hat{r}B_r + \hat{\theta}B_\theta + \hat{z}B_z$$

where the unit vectors $\hat{r}$, $\hat{\theta}$, and $\hat{z}$ are shown in Figure (12). If we make a change in the angle $\theta$ from $\theta$ to $\theta+\Delta\theta$, the unit vectors $\hat{r}$ and $\hat{\theta}$ change directions by an angle $\Delta\theta$ as shown in Figure (13).

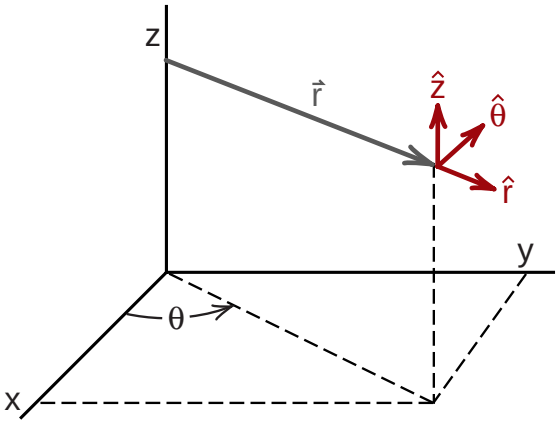



**Figure 12**
***The unit vectors in cylindrical coordinates.***

When we calculate the partial derivative of the vector $\vec{B}$, as we change the angle $\theta$ from $\theta$ to $\theta+\Delta\theta$, we not only have to include the change in the value of $\vec{B}$ as we move from points (1) to (2) in Figure (11), we also have to account for the fact that the unit vectors $\hat{r}$ and $\hat{\theta}$ have also changed. This change mixes up the components of $\vec{B}$.

It is not impossible to work out the formulas for the divergence or curl of a vector in cylindrical or spherical coordinates, but one is not likely to do it on the back of an envelope and get the right answer. Any practicing physicist or engineer, who needs to use these formulas, looks them up in a reliable reference.

What we will do is simply state the formula for curl in cylindrical coordinates, and then check that the formula gives the simple results we discussed in the last section for the case of the magnetic field of a wire. At the end of this text, in the ***Formulary***, we summarize all the formulas for gradient, divergence and curl, in Cartesian, cylindrical and spherical coordinates. Such a summary can be a very useful thing to have.

Given a field $\vec{B}$ expressed in cylindrical coordinates as

$$\vec{B} = \hat{r}B_r + \hat{\theta}B_\theta + \hat{z}B_z \qquad (50)$$

the formula for the curl is

$$\boxed{\begin{aligned}(\vec{\nabla}\times\vec{B})_r &= \frac{1}{r}\frac{\partial B_z}{\partial\theta} - \frac{\partial B_\theta}{\partial z}\\ (\vec{\nabla}\times\vec{B})_\theta &= \frac{\partial B_r}{\partial z} - \frac{\partial B_z}{\partial r}\\ (\vec{\nabla}\times\vec{B})_z &= \frac{1}{r}\frac{\partial}{\partial r}(rB_\theta) - \frac{1}{r}\frac{\partial B_r}{\partial\theta}\end{aligned}} \qquad (51)$$

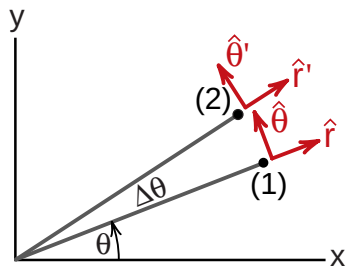



**Figure 13**
***We see that the unit vectors $\hat{r}$ and $\hat{\phi}$ change direction when we change the angle $\theta$ by $\Delta\theta$.***

## CALCULATING THE CURL OF THE MAGNETIC FIELD OF A WIRE

While Equation (51) for $\vec{\nabla}\times\vec{B}$ in cylindrical coordinates looks worse than the curl in Cartesian coordinates, you will see a major simplification when applied to a problem with cylindrical symmetry. The magnetic field of a wire travels in circles about the wire as shown in Figure (14). We see that $\vec{B}$ has only a $\hat{\theta}$ component $B_\theta$. In addition, the value of $B_\theta$ does not depend on, i.e., change with, the height $z$ or the angle $\theta$. Thus we can write $\vec{B}$ as

$$\vec{B} = \hat{\theta}B_\theta(r) \qquad \textit{magnetic field of a straight wire} \qquad (52)$$

where the only variable $B_\theta$ depends upon is the radius.

### Outside the wire

We will first calculate $\vec{B}$ using the integral form of Gauss' law, and then see what happens when we apply the curl formula, Equation (51) to $\vec{B}$. Integrating $\vec{B}$ around the circular path of radius r, shown by the dotted circle in Figure (12) gives

$$\oint \vec{B}\cdot d\vec{\ell} = \mu_0 i_{\text{enclosed}}$$

$$B_\theta(r)\times 2\pi r = \mu_0 i_{\text{tot}}$$

$$\boxed{B_\theta(r) = \frac{\mu_0 i_{\text{tot}}}{2\pi r}} \qquad (53) \text{ also } (28\text{-}18)$$

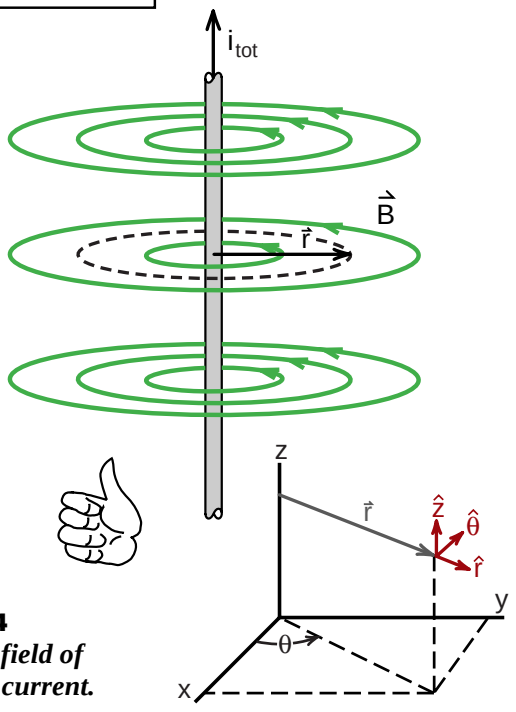



**Figure 14**
***Magnetic field of a straight current.***

This is the result we saw in Chapter 28 of the Physics text. Here $i_{\text{enclosed}}$ is equal to the total current $i_{\text{tot}}$ because our path goes around the wire.

We are now ready to plug in the values

$$B_r = 0$$

$$B_\theta = \frac{\mu_0 i_{\text{tot}}}{2\pi r}$$

$$B_z = 0 \qquad (54)$$

into Equation (51) to get the value of the curl

$$\vec{\nabla}\times\vec{B} = \hat{r}(\vec{\nabla}\times\vec{B})_r + \hat{\theta}(\vec{\nabla}\times\vec{B})_\theta + \hat{z}(\vec{\nabla}\times\vec{B})_z \qquad (55)$$

Because $B_r$ and $B_z$ are zero, a lot of the terms in the formula for $\vec{\nabla}\times\vec{B}$ vanish, and we are left with

$$(\vec{\nabla}\times\vec{B})_r = -\frac{\partial B_\theta}{\partial z}$$

$$(\vec{\nabla}\times\vec{B})_\theta = 0$$

$$(\vec{\nabla}\times\vec{B})_z = \frac{1}{r}\frac{\partial}{\partial r}(rB_\theta) \qquad (56)$$

You should check for yourself that this is all that is left of $\vec{\nabla}\times\vec{B}$ for the $\vec{B}$ of Equation (54).

We now note that $B_\theta(r) = \mu_0 i_{\text{tot}}/2\pi r$ depends only on the variable r and has no z dependence. Thus

$$\frac{\partial B_\theta(r)}{\partial z} = 0 \qquad (57)$$

and all we are left with for the curl is

$$\boxed{(\vec{\nabla}\times\vec{B})_z = \frac{1}{r}\frac{\partial}{\partial r}(rB_\theta)} \qquad (58)$$

Equation (58) applies to any vector field that looks like the magnetic field in Figure (12). It applies to any vector field of the form

$$\vec{B} = \hat{\theta}f(r) \qquad (59)$$

where f(r) is any function of r. These are the kinds of fields we are most likely to deal with in a discussion of the curl, in which case we can use the much simpler Equation (58).

Applying Equation (58) to our special value $B_\phi = \mu_0 i_{tot}/2\pi r$ , we get

$$\left(\vec{\nabla}\times\vec{B}\right)_z = \frac{1}{r}\frac{\partial}{\partial r}\left(rB_\theta\right)$$

$$= \frac{1}{r}\frac{\partial}{\partial r}\left[r\left(\frac{\mu_0 i_{tot}}{2\pi r}\right)\right] \qquad (60)$$

Notice that the r's in the square bracket cancel, leaving us with

$$\left(\vec{\nabla}\times\vec{B}\right)_z = \frac{1}{r}\frac{\partial}{\partial r}\left(\frac{\mu_0 i_{tot}}{2\pi}\right) \qquad (61)$$

We see that $\mu_0 i_{tot}/2\pi$ is a constant and the derivative of a constant is zero

$$\frac{\partial}{\partial r}\left(\frac{\mu_0 i_{tot}}{2\pi}\right) = 0 \qquad (62)$$

Thus we end up with the simple result

$$\boxed{\vec{\nabla}\times\vec{B} = 0 \qquad \left(\text{for } B_\theta = \frac{\mu_0 i_{tot}}{2\pi r}\right)} \qquad (63)$$

This is what we expected from our earlier discussion of Ampere's law in differential form. Neglecting the $\partial\vec{E}/\partial t$ term, the law is

$$\vec{\nabla}\times\vec{B} = \mu_0\vec{i}(x,y,z) \qquad \text{(38) repeated}$$

where the vector $\vec{i}(x,y,z)$ is the current density. Since the current is confined to the wire, the curl $\vec{\nabla}\times\vec{B}$ must also be confined to the wire, and be zero outside.

## Inside the Wire

What about inside the wire where the current density is not zero? Equation (53) does not apply there because the formula $B_\theta = \mu_0 i_{tot}/2\pi r$ applies only outside the wire.

To calculate the magnetic field inside the wire, we have to know something about the current density. Let us assume that we have a uniform current inside a wire of radius R. We will apply Ampere's law to a circular path of radius r as shown in the end view of the wire in Figure (15).

The amount of current enclosed by our path of radius r is, for a uniform current, simply the total current $i_{tot}$ times the ratio of the area $\pi r^2$ of the path, to the area $\pi R^2$ of the wire

$$i_{enclosed} = i_{total}\frac{\pi r^2}{\pi R^2} = i_{tot}\frac{r^2}{R^2} \qquad (64)$$

Using this value in Ampere's law, we get for the magnetic field inside the wire

$$\oint\vec{B}\cdot d\vec{\ell} = \mu_0 i_{enclosed}$$

$$\left(B_\theta \times 2\pi r\right) = \mu_0 i_{tot}\frac{r^2}{R^2} \qquad (65)$$

One of the r's cancels, and we are left with

$$B_\theta(r) = \left[\frac{\mu_0 i_{total}}{2\pi R^2}\right] r \qquad (66)$$

where everything in the square brackets is a constant. You derived this result in Exercise (29-4) of the Physics text.

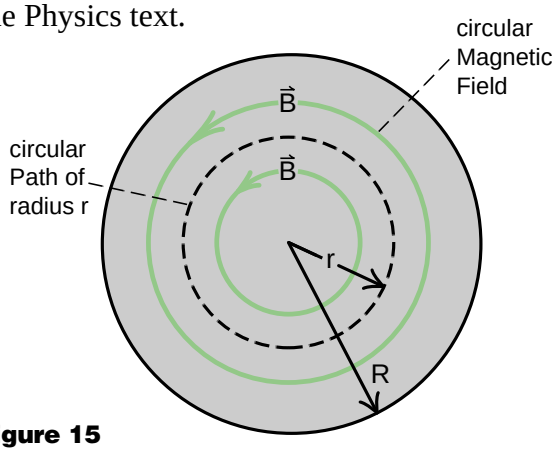



**Figure 15**
***Calculating the magnetic field inside the wire, assuming a uniform current density.***

Repeating Equation (66), we had for the field inside the wire

$$B_\theta(r) = \left[ \frac{\mu_0 i_{total}}{2\pi R^2} \right] r \qquad (66)$$

We see that $B_\theta$ increases linearly with r until we reach the surface of the wire at r = R, as shown in Figure (16). Then outside the wire, $B_\theta$ drops off as 1/r.

To simplify the formulas, let us write $B_\theta$ inside the wire as

$$B_\theta(r) = kr \qquad \begin{matrix} inside \\ wire \end{matrix} \qquad (66a)$$

where

$$k = \frac{\mu_0 i_{total}}{2\pi R^2} \qquad (66b)$$

The curl of this value of $B_\theta$ is given by Equation (58) as

$$\begin{aligned} (\vec{\nabla}\times\vec{B})_z &= \frac{1}{r}\frac{\partial}{\partial r}(rB_\theta) \\ &= \frac{1}{r}\frac{\partial}{\partial r}(rkr) \\ &= \frac{k}{r}\frac{\partial}{\partial r}(r^2) \end{aligned} \qquad (67)$$

Since $\partial(r^2)/\partial r = 2r$ , we get

$$(\vec{\nabla}\times\vec{B})_z = \frac{k}{r}(2r) = 2k \qquad (68)$$

Putting back our value for $k = \mu_0 i_{tot}/2\pi R^2$ we get

$$(\vec{\nabla}\times\vec{B})_z = \mu_0\left(\frac{i_{tot}}{\pi R^2}\right) \qquad (69)$$

Now $i_{tot}/\pi R^2$ is the total current in the wire divided by the area of the wire, which is the current density i(x,y,z). Since the current is $\hat{z}$ directed, we can write the current density as

$$\vec{i}(x,y,z) = \hat{z}\frac{i_{tot}}{\pi R^2} \qquad (70)$$

and Equation (70) can be written as the vector equation

$$\vec{\nabla}\times\vec{B} = \mu_0\vec{i}(x,y,z) \qquad (38) \text{ repeated}$$

which is the differential form of Ampere's law (for $\partial\vec{E}/\partial t = 0$ ).

This is the result we expected in the first place. The fact that we got back to Ampere's law serves as a check that the formulas for the curl in cylindrical coordinates are working.

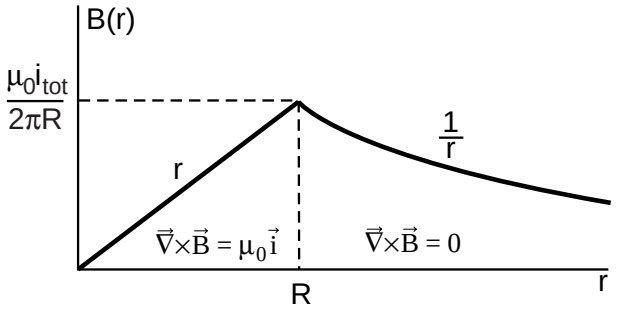



**Figure 16**
***The magnetic field inside and outside the wire, for a uniform current density inside the wire.***

# Calculus 2000-Chapter 9

# Electromagnetic Waves

*In the Physics text we had some difficulty showing that Maxwell's equations led to the prediction of the existence of electromagnetic radiation. The problem was that the integral form of Maxwell's equations are not particularly well suited for the derivation. The best we could do was to show that the wave pulse, shown in Figure (32-16) reproduced here, travels out at a speed* $v = 1/\sqrt{\mu_0\varepsilon_0}$ *which turns out to be the speed of light.*

*In discussing light waves, we made the argument that if we started with a series of wave pulses shown in Figure (32-23a) and smoothed them out, we could get the sinusoidal pulse shown in (32-23b). We never did show that the smoothed out version was actually a solution of Maxwell's equations, or that the sinusoidal structure traveled at a speed* $c = 1/\sqrt{\mu_0\varepsilon_0}$ *. With the differential form of Maxwell's equations, we can now do that.*

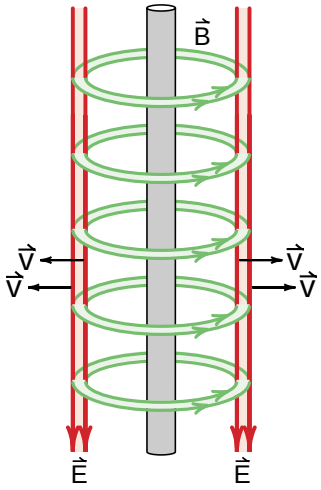
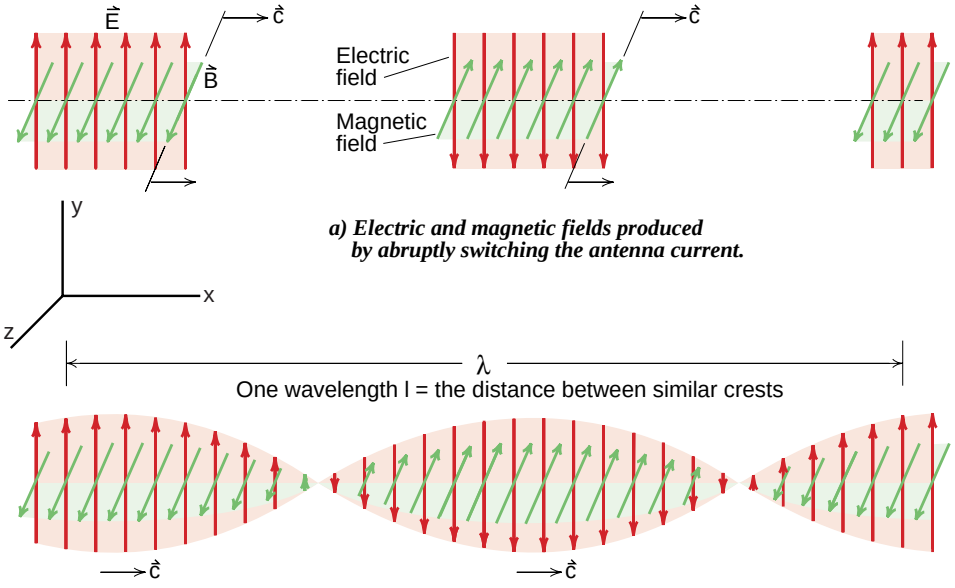



**Figure 32-16**
***Electromagnetic pulse produced by turning the current on and then quickly off. We will see that this structure agrees with Maxwell's equations.***




***a) Electric and magnetic fields produced by abruptly switching the antenna current.***

***b) Electric and magnetic fields produced by smoothly switching the antenna current.***

**Figure 32-23**
***Structure of electric and magnetic fields in light and radio waves.***

## VECTOR IDENTITIES

To use the differential forms of Maxwell's equations, it is convenient to first develop three formulas known as ***vector identities***. These are mathematical relationships involving curls that apply to any vector field. We will state these identities first and then spend the rest of the section deriving them. You should go through these derivations at least once to get a feeling for how they work and how general they are.

### Identity 1

The curl of a gradient $\vec{\nabla}f$ is zero for any scalar field f(x,y,z).

$$\boxed{\vec{\nabla}\times(\vec{\nabla}f) = 0} \qquad (1)$$

### Identity 2

The divergence of a curl is zero. That is, for any vector field $\vec{A}(x,y,z)$

$$\boxed{\vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) = 0} \qquad (2)$$

### Identity 3

This identity gives us a formula for the curl of a curl. The formula is

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{A}) = -(\vec{\nabla}\cdot\vec{\nabla})\vec{A} + \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) \qquad (3)$$

where $\vec{\nabla}\cdot\vec{\nabla} = \nabla_x\nabla_x + \nabla_y\nabla_y + \nabla_z\nabla_z$ is the Laplacian operator discussed in Chapter 4. We will often use the notation

$$\vec{\nabla}\cdot\vec{\nabla} \equiv \nabla^2 = \nabla_x\nabla_x + \nabla_y\nabla_y + \nabla_z\nabla_z \qquad (4)$$

so that the vector identity can be written as

$$\boxed{\vec{\nabla}\times(\vec{\nabla}\times\vec{A}) = -\nabla^2\vec{A} + \vec{\nabla}(\vec{\nabla}\cdot\vec{A})} \qquad (5)$$

In the special case that $\vec{A}$ has zero divergence, if $\vec{\nabla}\cdot\vec{A} = 0$, then we get

$$\boxed{\vec{\nabla}\times(\vec{\nabla}\times\vec{A}) = -\nabla^2\vec{A} \quad \textit{if } \vec{\nabla}\cdot\vec{A} \textit{ is zero}} \qquad (5a)$$

### Proof of Identity 1

The proof of these identities relies on the fact that we can interchange the order of partial differentiation, a result we prove in the appendix to this chapter. As an example of how this is used, consider one component of the first identity. Using the cross product formula

$$(\vec{A}\times\vec{B})_x = A_yB_z - A_zB_y \qquad (6)$$

we get

$$\begin{aligned}\left[\vec{\nabla}\times(\vec{\nabla}f)\right]_x &= \nabla_y(\nabla_z f) - \nabla_z(\nabla_y f)\\ &= \nabla_y\nabla_z f - \nabla_z\nabla_y f\end{aligned} \qquad (7)$$

Interchanging $\nabla_y\nabla_z$ to get $\nabla_y\nabla_z f = \nabla_z\nabla_y f$ immediately makes this component zero. The same thing happens to the y and z components of $\vec{\nabla}\times(\vec{\nabla}f)$, thus the entire expression is zero.

**Proof of Identity 2**

To prove the second identity $\vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) = 0$, we start with the components of $\vec{\nabla}\times\vec{A}$, which are

$$(\vec{\nabla}\times\vec{A})_x = \nabla_y A_z - \nabla_z A_y$$

$$(\vec{\nabla}\times\vec{A})_y = \nabla_z A_x - \nabla_x A_z$$

$$(\vec{\nabla}\times\vec{A})_z = \nabla_x A_y - \nabla_y A_x \qquad (8)$$

Note that to get all three components of $\vec{\nabla}\times\vec{A}$, you do not have to memorize all three equations. If you memorize only the first $(\vec{\nabla}\times\vec{A})_x = \nabla_y A_z - \nabla_z A_y$ you can get the other two by using ***cyclic permutations***. That means, start with $(\vec{\nabla}\times\vec{A})_x = \nabla_y A_z - \nabla_z A_y$, and replace the subscripts cyclically, letting $x \to y$, $y \to z$, and $z \to x$. That gives you $(\vec{\nabla}\times\vec{A})_y = \nabla_z A_x - \nabla_x A_z$. Do the cyclic permutation again and you get $(\vec{\nabla}\times\vec{A})_z = \nabla_x A_y - \nabla_y A_x$ which is the third equation.)

Now take the dot product of $\vec{\nabla}$ with $\vec{\nabla}\times\vec{A}$ to get

$$\begin{aligned}
&\vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) \\
&\quad = \nabla_x(\vec{\nabla}\times\vec{A})_x + \nabla_y(\vec{\nabla}\times\vec{A})_y + \nabla_z(\vec{\nabla}\times\vec{A})_z \\
&\quad = \nabla_x\nabla_y A_z - \nabla_x\nabla_z A_y \\
&\qquad + \nabla_y\nabla_z A_x - \nabla_y\nabla_x A_z \\
&\qquad + \nabla_z\nabla_x A_y - \nabla_z\nabla_y A_x \qquad (9)
\end{aligned}$$

**Exercise 1**

Show that all the terms in Equation (9) cancel, giving $\vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) = 0$ for any $\vec{A}$.

**Proof of Identity 3**

The third vector identity

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{A}) = -\nabla^2 A + \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) \quad (5)\ \text{repeat}$$

looks worse but is not that hard to prove. We will start with the $\vec{x}$ component of $\vec{\nabla}\times(\vec{\nabla}\times\vec{A})$ which is

$$\begin{aligned}
\left[\vec{\nabla}\times(\vec{\nabla}\times\vec{A})\right]_x &= \nabla_y(\vec{\nabla}\times\vec{A})_z - \nabla_z(\vec{\nabla}\times\vec{A})_y \\
&= \nabla_y(\nabla_x A_y - \nabla_y A_x) - \nabla_z(\nabla_z A_x - \nabla_x A_z) \\
&= -\nabla_y\nabla_y A_x - \nabla_z\nabla_z A_x + \nabla_x\nabla_y A_y + \nabla_x\nabla_z A_z
\end{aligned} \qquad (10)$$

where we changed the order of differentiation in the last two terms. The trick is to add and then subtract $\nabla_x\nabla_x A_x$ to Equation (10), giving

$$\begin{aligned}
&\left[\vec{\nabla}\times(\vec{\nabla}\times\vec{A})\right]_x \\
&= -\nabla_x\nabla_x A_x - \nabla_y\nabla_y A_x - \nabla_z\nabla_z A_x \\
&\qquad + \nabla_x\nabla_x A_x + \nabla_x\nabla_y A_y + \nabla_x\nabla_z A_z \\
&= -(\nabla_x\nabla_x + \nabla_y\nabla_y + \nabla_z\nabla_z)A_x \\
&\qquad + \nabla_x(\nabla_x A_x + \nabla_y A_y + \nabla_z A_z) \\
&= -\nabla^2 A_x + \nabla_x(\vec{\nabla}\cdot\vec{A})
\end{aligned} \qquad (11)$$

This is just the x component of Equation (5). Similar derivations verify the y and z components of that vector identity.

## DERIVATION OF THE WAVE EQUATION

We are now in a position to derive the wave equation for electromagnetic waves, starting from Maxwell's equations. We will use Maxwell's equations for empty space, because Maxwell's major discovery was that electric and magnetic fields could propagate through empty space in a wavelike manner, and that these waves were light waves.

Maxwell's equations in differential form are, from Equations (8-49) of Chapter 8

$$\begin{aligned}
&\vec{\nabla}\cdot\vec{E} = \frac{\rho}{\varepsilon_0} && \textit{Gauss' law}\\
&\vec{\nabla}\cdot\vec{B} = 0 && \textit{no monopole}\\
&\vec{\nabla}\times\vec{B} = \mu_0\vec{i} + \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t} && \textit{Ampere's law}\\
&\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} && \textit{Faraday's law}
\end{aligned} \qquad (12)$$

Maxwell's Equations

where $\rho(x,y,z)$ is the electric charge density in coulombs per cubic meter, and $\vec{i}(x,y,z)$ is the electric current density in amperes per square meter.

In empty space, where the charge density $\rho(x,y,z)$ and the current density $\vec{i}(x,y,z)$ are zero, we get

$$\vec{\nabla}\cdot\vec{E} = 0 \qquad \textit{Gauss' law} \qquad (13a)$$

$$\vec{\nabla}\cdot\vec{B} = 0 \qquad \textit{no monopole} \qquad (13b)$$

$$\vec{\nabla}\times\vec{B} = \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t} \qquad \textit{Ampere's law} \qquad (13c)$$

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \qquad \textit{Faraday's law} \qquad (13d)$$

Maxwell's Equations in Empty Space

In our discussion of vector fields in the Physics text, we pointed out that a vector field is uniquely determined if we have general formulas for the volume and line integrals of that field. Now, working with differential equations, that statement becomes the rule that a vector field like $\vec{E}$ is determined if we know the divergence $\vec{\nabla}\cdot\vec{E}$ and the curl $\vec{\nabla}\times\vec{E}$ at every point in space*. There are four Maxwell equations because we have to specify both the divergence and the curl of both $\vec{E}$ and $\vec{B}$.

Equation (10) tells us that in empty space, neither $\vec{E}$ nor $\vec{B}$ have a divergence ($\vec{\nabla}\cdot\vec{E} = \vec{\nabla}\cdot\vec{B} = 0$), and we only have to deal with the curls of these fields.

The trick we use to get a wave equation from Equations (13) is to take the curl of Equations (13c) and (13d). This gives us

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{B}) = \mu_0\varepsilon_0\vec{\nabla}\times\left(\frac{\partial\vec{E}}{\partial t}\right) \qquad (14a)$$

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{E}) = -\vec{\nabla}\times\left(\frac{\partial\vec{B}}{\partial t}\right) \qquad (14b)$$

where we took the constants $\mu_0$ and $\varepsilon_0$ outside the derivative in Equation (14a).

*(If we have a field known only in some region of space, like the velocity field of a fluid in a section of pipe, we can uniquely determine the field if we know the divergence and curl within that region, and also the normal components of the field at the region's surface.)

The next step is to use the fact that we can interchange the order of partial differentiation to get

$$\vec{\nabla}\times\left[\frac{\partial \mathrm{E}(x,y,z,t)}{\partial t}\right] = \frac{\partial}{\partial t}\left[\vec{\nabla}\times \mathrm{E}(x,y,z,t)\right] \quad (15)$$

and a similar result for $\vec{\nabla}\times(\partial\vec{B}/\partial t)$ to give

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{B}) = \mu_0\varepsilon_0\frac{\partial}{\partial t}(\vec{\nabla}\times\vec{E}) \quad (16a)$$

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{E}) = -\frac{\partial}{\partial t}(\vec{\nabla}\times\vec{B}) \quad (16b)$$

Notice that the right hand sides of Equations (16) involve $(\vec{\nabla}\times\vec{E})$ and $(\vec{\nabla}\times\vec{B})$ which are given by Maxwell's Equations (13c) and (13d) as

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \quad \text{(13d) repeated}$$

$$\vec{\nabla}\times\vec{B} = \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t} \quad \text{(13c) repeated}$$

Thus Equations (16) can be written as

$$\begin{aligned}\vec{\nabla}\times(\vec{\nabla}\times\vec{B}) &= \mu_0\varepsilon_0\frac{\partial}{\partial t}\left(-\frac{\partial\vec{B}}{\partial t}\right)\\ &= -\mu_0\varepsilon_0\frac{\partial^2 B}{\partial t^2}\end{aligned} \quad (17a)$$

$$\begin{aligned}\vec{\nabla}\times(\vec{\nabla}\times\vec{E}) &= -\frac{\partial}{\partial t}\left(\mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t}\right)\\ &= -\mu_0\varepsilon_0\frac{\partial^2 E}{\partial t^2}\end{aligned} \quad (17b)$$

Notice that at this point $\vec{E}$ and $\vec{B}$ obey exactly the same differential equation.

The final step is to use the vector identity

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{A}) = -\nabla^2 A + \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) \quad \text{(5) repeat}$$

Since both $\vec{\nabla}\cdot\vec{E}$ and $\vec{\nabla}\cdot\vec{B}$ are zero in empty space, we have

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{B}) = -\nabla^2 B \quad (18)$$

and the same for $\vec{\nabla}\times(\vec{\nabla}\times\vec{E})$ to give us

$$-\nabla^2\vec{E} = -\mu_0\varepsilon_0\frac{\partial^2\vec{E}}{\partial t^2} \quad (19a)$$

$$-\nabla^2\vec{B} = -\mu_0\varepsilon_0\frac{\partial^2\vec{B}}{\partial t^2} \quad (19b)$$

Dividing through by $\mu_0\varepsilon_0$ gives

$$\boxed{\frac{1}{\mu_0\varepsilon_0}\nabla^2\vec{E} = \frac{\partial^2\vec{E}}{\partial t^2}} \quad (20a)$$

$$\boxed{\frac{1}{\mu_0\varepsilon_0}\nabla^2\vec{B} = \frac{\partial^2\vec{B}}{\partial t^2}} \quad (20b)$$

## PLANE WAVE SOLUTION

Repeating Equations (20), we have

$$\frac{1}{\mu_0\varepsilon_0}\nabla^2\vec{E} = \frac{\partial^2\vec{E}}{\partial t^2} \qquad (20a)$$

$$\frac{1}{\mu_0\varepsilon_0}\nabla^2\vec{B} = \frac{\partial^2\vec{B}}{\partial t^2} \qquad (20b)$$

To interpret these equations, let us assume that $\vec{E}$ and $\vec{B}$ have the shape more or less like that shown in Figure (32-23b) reproduced here again. All we need from that picture is that both $\vec{E}$ and $\vec{B}$ vary only in the direction of motion (call this the $\hat{x}$ direction) and in time. There is no change of $\vec{E}$ and $\vec{B}$ in the $\hat{y}$ and $\hat{z}$ directions. Such a wave is called a ***plane wave***, because there are no variations within a plane.

Using the coordinate system added to Figure (32-23b), we see that $\vec{E}$ is $\hat{y}$ directed (we would call this $\hat{y}$ polarized radiation) and $\vec{B}$ is $\hat{z}$ directed. The formulas for $\vec{E}$ and $\vec{B}$ can thus be written for this $\hat{z}$ directed plane wave

$$\vec{E} = \hat{y}E_y(x,t) \qquad (21a)$$

$$\vec{B} = \hat{z}B_z(x,t) \qquad (21b)$$

where Equations (21a) and (21b) remind us that we are dealing with a plane wave with no x or y dependence.

As a result

$$\nabla_y\vec{E} = \hat{y}\frac{\partial E(x,t)}{\partial y} = 0$$

and the same for $\nabla_z\vec{E}$, $\nabla_y\vec{B}$ and $\nabla_z\vec{B}$. Thus

$$\nabla^2\vec{E} = (\nabla_x\nabla_x\vec{E} + \nabla_y\nabla_y\vec{E} + \nabla_z\nabla_z\vec{E})$$

$$= \nabla_x\nabla_x\vec{E} = \hat{y}\frac{\partial^2 E_y}{\partial x^2} \qquad (22a)$$

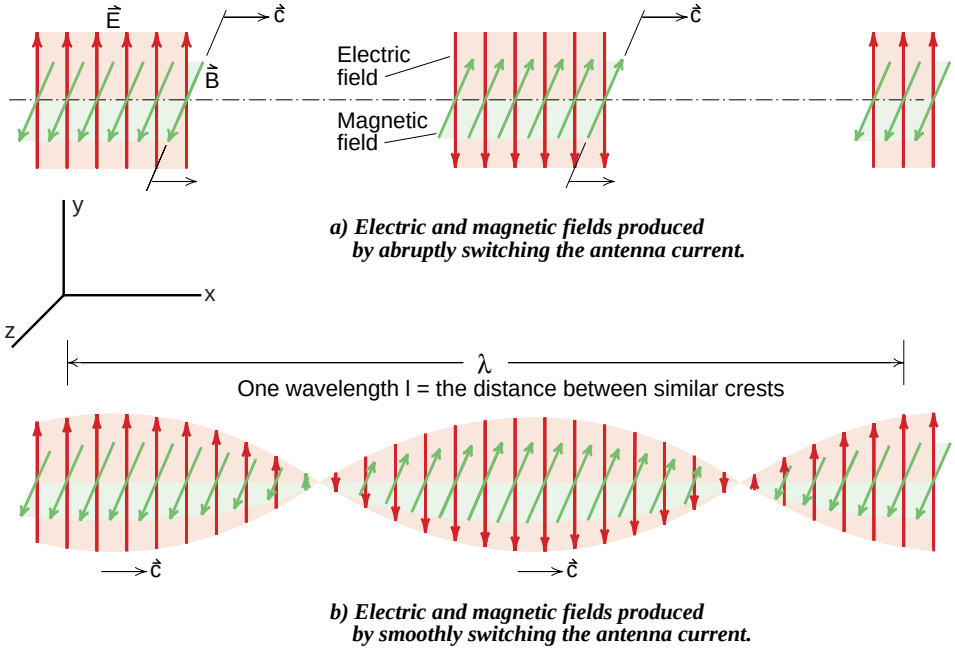
and

$$\nabla^2\vec{B} = \hat{z}\frac{\partial^2 B_z}{\partial x^2} \qquad (22b)$$

The time derivatives of the plane wave fields of Equations (21) are

$$\frac{\partial^2\vec{E}}{\partial t^2} = \hat{y}\frac{\partial^2 E_y(x,t)}{\partial t^2} \qquad (23a)$$

$$\frac{\partial^2\vec{B}}{\partial t^2} = \hat{z}\frac{\partial^2 B_z(x,t)}{\partial t^2} \qquad (23b)$$




*a) Electric and magnetic fields produced by abruptly switching the antenna current.*

*b) Electric and magnetic fields produced by smoothly switching the antenna current.*

**Figure 32-23**
***Structure of electric and magnetic fields in light and radio waves.***

When we use Equation (22a) for $\nabla^2\vec{E}$ and (23a) for $\partial^2\vec{E}/\partial t^2$ in Equation (20a), the unit vectors $\hat{y}$ cancel and we are left with

$$\frac{1}{\mu_0\varepsilon_0}\frac{\partial^2 E_y(x,t)}{\partial x^2} = \frac{\partial^2 E_y(x,t)}{\partial t^2} \qquad (24a)$$

We get a similar equation for $B_z$, namely

$$\frac{1}{\mu_0\varepsilon_0}\frac{\partial^2 B_z(x,t)}{\partial x^2} = \frac{\partial^2 B_z(x,t)}{\partial t^2} \qquad (24b)$$

In our discussion of the one dimensional wave equation in Chapter 2 of this text we had as the formula for the wave equation

$$v_{wave}^2\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2 y(x,t)}{\partial t^2} \quad \textit{one dimensional wave equation} \qquad (2\text{-}73)$$

Comparing this wave equation with Equation (24), we see that the plane wave of Figure (32-23b) obeys the one dimensional wave equation with

$$v_{wave}^2 = \frac{1}{\mu_0\varepsilon_0}$$

$$\boxed{v_{wave} = \frac{1}{\sqrt{\mu_0\varepsilon_0}}} \qquad (25)$$

From the wave equation alone we immediately find that the speed of the wave is $1/\sqrt{\mu_0\varepsilon_0}$ which is the speed of light. We get this result without going through all the calculations we did in the Physics text to derive the speed of the electromagnetic pulse.

What we have shown in addition is that the speed of the wave does not depend on its shape. All we used was that E = E(x,t) without saying what the x dependence was. Thus both the series of pulses in Figure (32-23a) and the sinusoidal wave in (32-23b) should have the same speed $1/\sqrt{\mu_0\varepsilon_0}$. This we were not able to show using the integral form of Maxwell's equations.

## THE THREE DIMENSIONAL WAVE EQUATION

We have seen that if $\vec{E}$ and $\vec{B}$ are plane waves, i.e., vector fields that vary in time and only one dimension, then Equations (20a) and (20b) become the one dimensional wave equation for $\vec{E}$ and $\vec{B}$. Since Equations (20) do not single out any one direction as being special, we would get a wave equation for a plane wave moving in any direction, and we see that Equations (20) are three dimensional wave equations for waves traveling at a speed $v_{wave}^2 = 1/\sqrt{\mu_0\varepsilon_0}$. Rewriting these equations in terms of $v_{wave}$ rather than $\mu_0\varepsilon_0$ gives us the general form of the three dimensional wave equation

$$v_{wave}^2\nabla^2\vec{E} = \frac{\partial^2 E}{\partial t^2} \qquad (26)$$

and the same for $\vec{B}$.

The form we will generally recognize as being the three dimensional wave equation is the trivial rearrangement of Equation (26),

$$\boxed{\frac{1}{v_{wave}^2}\frac{\partial^2\vec{E}}{\partial t^2} - \nabla^2\vec{E} = 0} \quad \textit{three dimensional wave equation applied to } \vec{E} \qquad (27)$$

Equation (27) is the way the wave equation is usually written in textbooks.

So far we have only shown that plane waves are a solution to the three dimensional wave equation. For now that is enough. Solutions to the wave equation can become quite complex in three dimensions, and we do not yet have to deal with these complications.

# APPENDIX: ORDER OF PARTIAL DIFFERENTIATION

It is worth while to show once and for all that you can interchange the order of partial differentiation. We do this by going back to the limiting process, where

$$\frac{\partial f(x,y)}{\partial x} = \underset{\Delta x \to 0}{\text{limit}} \left[ \frac{f(x+\Delta x,y) - f(x,y)}{\Delta x} \right] \quad \text{(A-1)}$$

and a similar formula for $\partial f/\partial y$. For the second derivative we have

$$\nabla_x \nabla_y f(x,y) = \frac{\partial}{\partial x}\left[ \frac{\partial f(x,y)}{\partial y} \right] \quad \text{(A-2)}$$

Let us temporarily introduce the notation

$$f_y'(x,y) = \frac{\partial f(x,y)}{\partial y} \quad \text{(A-3)}$$

so that Equation (A-2) becomes

$$\nabla_x \nabla_y f(x,y) = \frac{\partial}{\partial x} f_y'(x,y)$$

$$= \underset{\Delta x \to 0}{\text{limit}} \left[ \frac{f_y'(x+\Delta x,y) - f_y'(x,y)}{\Delta x} \right] \quad \text{(A-4)}$$

Now in Equation (A-4) make the substitution

$$f_y'(x,y) = \underset{\Delta y \to 0}{\text{limit}} \left[ \frac{f(x,y+\Delta y) - f(x,y)}{\Delta y} \right] \quad \text{(A-5)}$$

$$f_y'(x+\Delta x,y) = \underset{\Delta y \to 0}{\text{limit}} \left[ \frac{f(x+\Delta x,y+\Delta y) - f(x+\Delta x,y)}{\Delta y} \right] \quad \text{(A-6)}$$

Using Equations (A-5) and (A-6) in (A-4) gives

$$\nabla_x \nabla_y f(x,y) = \underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{\text{limit}} \left[ \frac{f(x+\Delta x,y+\Delta y) + f(x,y) - f(x+\Delta x,y) - f(x,y+\Delta y)}{\Delta x \Delta y} \right] \quad \text{(A-7)}$$

**Exercise 1**

Show that you get exactly the same result for $\nabla_y \nabla_x f(x,y)$.

You can see that our result, Equation (A-7) is completely symmetric between x and y, thus it should be obvious that we should get the same result by reversing the order of differentiation.

The only possible fly in the ointment is the order in which we take the limits as $\Delta x \to 0$ and $\Delta y \to 0$. As long as f(x,y) is smooth enough so that f(x,y) and its first and second derivatives are continuous, then the order in which we take the limit makes no difference.

# Calculus 2000-Chapter 10

# Conservation of Electric Charge

*In this short chapter, we obtain a very important result. We will see that Maxwell's equations themselves imply that electric charge is conserved. In our development of Maxwell's equations, our attention was on the kind of electric and magnetic fields that were produced by electric charges and currents. We said, for example, that given some electric charge, Gauss' law would tell us what electric field it would produce. Or given an electric current, Ampere's law would tell us what magnetic field would result.*

*Then later on, we found out that for mathematical consistency, a changing electric field would create a magnetic field and vice versa. All this was summarized in Maxwell's equations, which we repeat here*

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \rho/\varepsilon_0 \\ \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{E} &= -\partial B/\partial t \\ \vec{\nabla} \times \vec{B} &= \mu_0 \vec{i} + \mu_0 \varepsilon_0 \partial \vec{E}/\partial t \end{aligned} \qquad (1)$$

*What we did not notice in this development of the equations for $\vec{E}$ and $\vec{B}$ is that the equations place a fundamental restriction on the sources $\rho$ and $\vec{i}$ of the fields. As we will now see, the restriction is that the electric charge, which is responsible for the charge density $\rho$ and current $\vec{i}$, must be conserved.*

## THE CONTINUITY EQUATION

We began our discussion of fluid dynamics in Chapter 23 of the Physics text, by introducing the continuity equation for an incompressible fluid. For a tube with an entrance cross sectional area $A_1$ and exit area $A_2$, the equation was

$$v_1 A_1 = v_2 A_2 \qquad \textit{continuity equation} \qquad (23\text{-}3)$$

which says that the same volume of fluid per second flowing into the entrance flows out of the exit. Later this statement that the fluid is incompressible (or does not get lost or created) became

$$\int_{\text{closed surface}} \vec{v}\cdot d\vec{A} = 0 \qquad \textit{incompressible fluid} \qquad (2)$$

The differential form of Equation (2) is

$$\vec{\nabla}\cdot\vec{v} = 0 \qquad \textit{incompressible fluid} \qquad (3)$$

as we showed in our initial discussion of divergence. All three equations, (23-3), (2) and (3) are saying the same thing in a progressively more detailed way.

Equation (3) is not the most general statement of a continuity equation. It is the statement of the conservation of an incompressible fluid, but you can have flows of a compressible nature where something like mass or charge is still conserved. A more general form of the continuity equation allows for the conservation of these quantities. We will now see that this more general form of the continuity equation naturally arises from Maxwell's equations.

## CONTINUITY EQUATION FROM MAXWELL'S EQUATIONS

To derive the continuity equation for electric charge, we start by taking the divergence of the generalized form of Ampere's law

$$\vec{\nabla}\cdot\left[\ \vec{\nabla}\times\vec{B} = \mu_0\vec{i} + \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t}\ \right] \qquad (4)$$

which becomes

$$\vec{\nabla}\cdot(\vec{\nabla}\times\vec{B}) = \mu_0\vec{\nabla}\cdot\vec{i} + \mu_0\varepsilon_0\vec{\nabla}\cdot\left(\frac{\partial\vec{E}}{\partial t}\right) \qquad (5)$$

Using the fact that the divergence of a curl is identically zero, $\vec{\nabla}\cdot(\vec{\nabla}\times\vec{B}) = 0$, and the fact that we can interchange the order of differentiation, we get

$$0 = \mu_0\vec{\nabla}\cdot\vec{i} + \mu_0\varepsilon_0\frac{\partial}{\partial t}(\vec{\nabla}\cdot\vec{E}) \qquad (6)$$

Divide Equation (6) through by $\mu_0$, and use Gauss' law

$$\vec{\nabla}\cdot\vec{E} = \frac{\rho}{\varepsilon_0}$$

to get

$$\vec{\nabla}\cdot i + \varepsilon_0\frac{\partial}{\partial t}\left(\frac{\rho}{\varepsilon_0}\right) = 0 \qquad (7)$$

The $\varepsilon_0$'s cancel and we are left with

$$\boxed{\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot\vec{i} = 0} \qquad \textit{continuity equation for electric charge} \qquad (8)$$

Equation (8) is the continuity equation for electric charge.

You can immediately see from Equation (8) that if the electric charge density $\rho$ were unchanging in time, if $\partial\rho/\partial t = 0$, then we would have $\vec{\nabla}\cdot i = 0$ and the electric current would flow as an incompressible fluid. The fact that a $\partial\rho/\partial t$ term appears in Equation (8) is telling us what happens when $\rho$ changes, for example, if we compress the charge into a smaller region.

## Integral Form of Continuity Equation

The way to interpret Equation (8) is to convert the equation to its integral form. We do this by integrating the equation over some volume V bounded by a closed surface S. We have

$$\int_V \frac{\partial \rho}{\partial t} dV + \int_V \vec{\nabla} \cdot \vec{i}\, dV = 0 \qquad (9)$$

Using the divergence theorem to convert the volume integral of $\vec{\nabla} \cdot \vec{i}$ to a surface integral gives

$$\int_{\substack{\text{volume} \\ V}} \vec{\nabla} \cdot \vec{i}\, dV = \int_{\substack{S \text{ (surface} \\ \text{of V)}}} \vec{i} \cdot d\vec{A} \qquad (10)$$

Using Equation (10) in (9) we get

$$\int_{\substack{\text{closed} \\ \text{surface S}}} \vec{i} \cdot d\vec{A} = -\int_{\substack{\text{volume} \\ \text{V inside S}}} \frac{\partial \rho}{\partial t} dV \qquad (11)$$

*integral form of continuity equation*

On the left side of Equation (11) we have the term representing the net flow of electric current out through the surface S. It represents the total amount of electric charge per second leaving through the surface. On the right side we have an integral representing the rate at which the amount of charge remaining inside the volume V is decreasing (the – sign). Thus Equation (38) is telling us that the rate at which charge is flowing out through any closed surface S is equal to the rate at which the amount of charge remaining inside the surface is decreasing. This can be true for any surface S only if electric charge is everywhere conserved.

The fact that the continuity equation was a consequence of Maxwell's equation tells us that if we do have the correct equations for electric and magnetic fields, then the source of these fields, which is electric charge and current, must be a conserved source. Later, when we discuss the process of constructing theories of fields, we will see in more detail how conservation laws and theories of fields are closely related. Basically for every fundamental conservation law there is a field associated with the law. In this case the law is the conservation of electric charge and the associated field is the electromagnetic field. It turns out that the law of conservation of energy is associated with the gravitational field.

# Calculus 2000-Chapter 11

# Scalar And Vector Potentials

*In our first experiment on electricity in the Physics text we studied the relationship between voltage on electric fields. We constructed the lines of constant voltage, the equipotential lines, and then constructed the perpendicular electric field lines. In Chapter 3 of the Calculus text we developed the more detailed relationship that the electric field* $\vec{E}$ *was equal to minus the gradient of the voltage*

$$\vec{E}(x,y,z) = -\vec{\nabla}V(x,y,z) \qquad (3\text{-}19)$$

*As you study more advanced topics in science, you sometimes encounter situations where the name or symbol used to describe some quantity is different in the advanced texts than in the introductory ones. Various historical accidents are often responsible for this change.*

*In introductory texts and in the laboratory we talk about the voltage V which we measure with a voltmeter. The first hint that we would use a different name for voltage was when we called the lines of constant voltage* **equipotential** *lines, or lines of constant* **potential**. *Advanced texts, particularly those with a theoretical emphasis, use the name potential rather than voltage, and typically use the symbol* $\phi(x,y,z)$ *rather than* $V(x,y,z)$. *In this notation, Equation (3-19) becomes*

$$\vec{E}(x,y,z) = -\vec{\nabla}\phi(x,y,z) \qquad (1)$$

*This is how we left the relationship between* $\vec{E}$ *and* $\phi$ *in Chapter 3 on gradients.*

*From our discussion of divergence and curl, it does not take long to see that there is a problem with Equation (1) . If we take the curl of both sides of this equation, we get*

$$\vec{\nabla}\times\vec{E} = -\vec{\nabla}\times(\vec{\nabla}\phi) \qquad (2)$$

*However our first vector identity, Equation (9-1) was that the curl of a divergence was identically zero.*

$$\vec{\nabla}\times(\vec{\nabla}\phi) = 0 \qquad (3)$$

*Thus Equation (1) implies that the field* $\vec{E}$ *has zero curl*

$$\vec{\nabla}\times\vec{E} = 0 \qquad \text{as a consequence of Equation (1)} \qquad (4)$$

*which is not consistent with Maxwell's equations. In particular, Faraday's law says that*

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \qquad \text{Faraday's law} \qquad (5)$$

*Thus Equation (1) cannot be true, or at least cannot be the whole story, when changing magnetic fields are present, when* $\partial\vec{B}/\partial t$ *is not zero. If we only have static charges, or even stationary currents so that* $\vec{B}$ *is zero or constant in time, then Faraday's law becomes*

$$\vec{\nabla}\times\vec{E} = 0 \qquad \text{when } \partial\vec{B}/dt = 0 \qquad (6)$$

*and then* $\vec{E}$ *can be described completely as the gradient of a voltage V or potential* $\phi$.

*Since the curl is the line integral on an infinitesimal scale, Equation (6) is equivalent to the statement that the line integral of $\vec{E}$ is zero everywhere*

$$\oint \vec{E}\cdot d\vec{\ell} \;=\; 0 \qquad \begin{matrix}\text{when}\\ \partial\vec{B}/dt = 0\end{matrix} \qquad (6a)$$

*In our initial discussion of the line integral in Chapter 28 of the Physics text (pages 28-5,6), we pointed out that Equation (6a) was the condition for what we called a* **conservative force**, *a force that could be described in terms of potential energy. The equation $\vec{E} = -\vec{\nabla}\phi$ (or $-\vec{\nabla}V$ ) does exactly that, since V or $\phi$ is the potential energy of a unit test charge.*

*What we are seeing now is that for static fields, where $\partial\vec{B}/\partial t$ is zero, $\vec{E}$ is a conservative field that can be described as the gradient of a potential energy $\phi$. However when changing magnetic fields are present, the curl of $\vec{E}$ is no longer zero and $\vec{E}$ has a component that cannot be described as the gradient of a potential energy.*

*We will see in this chapter that $\vec{E}$ and $\vec{B}$ can both be described in terms of potentials by introducing a new kind of potential called the* **vector potential** $\vec{A}$**(x,y,z)**. *When combined with what we will now call the* **scalar potential** $\phi$**(x,y,z)**, *we not only have complete formulas for $\vec{E}$ and $\vec{B}$, but also end up simplifying the electromagnetic wave equation for the case that sources like charge density $\rho$ and current density $\vec{i}$ are present.*

*The topic of the vector potential $\vec{A}$(x,y,z) is often left to later advanced physics courses, sometimes introduced at the graduate course level. There is no need to wait; the introduction of the vector potential provides good practice with curl and divergence. What we will not cover in this chapter are the ways the vector potential is used to solve complex radiation problems. That can wait. What we will focus on is how the vector potential can be used to simplify the structure of Maxwell's equations. In addition we need the vector potential to handle the concept of voltage when changing magnetic fields are present.*

## THE VECTOR POTENTIAL

It seems to be becoming a tradition in this text to begin each chapter with a repeat of Maxwell's equations. In order not to break the tradition, we do it again.

$$\begin{aligned}
\vec{\nabla}\cdot\vec{E} &= \frac{\rho}{\varepsilon_0} && \textit{Gauss' law}\\
\vec{\nabla}\cdot\vec{B} &= 0 && \textit{no monopole}\\
\vec{\nabla}\times\vec{B} &= \mu_0\vec{i} + \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t} && \textit{Ampere's law}\\
\vec{\nabla}\times\vec{E} &= -\frac{\partial\vec{B}}{\partial t} && \textit{Faraday's law}
\end{aligned} \qquad (7)$$

Let us now set the magnetic field $\vec{B}$(x,y,z) equal to the curl of some new vector field $\vec{A}$(x,y,z) . That is,

$$\boxed{\vec{B}(x,y,z) \equiv \vec{\nabla}\times\vec{A}(x,y,z)} \qquad \begin{matrix}\textit{introducing}\\ \textit{the vector}\\ \textit{potential } \vec{A}\end{matrix} \qquad (8)$$

Equation (7) is the beginning of our definition of what we will call the *vector potential* $\vec{A}(x,y,z)$. To begin to see why we introduced the vector potential, take the divergence of both sides of Equation (8). We get

$$\vec{\nabla}\cdot\vec{B} \;=\; \vec{\nabla}\cdot(\vec{\nabla}\times\vec{A}) \;=\; 0 \qquad (9)$$

This is zero because of the second vector identity studied in Chapter 9, Equation (9-2). There we showed that the divergence of the curl $\vec{\nabla}\cdot(\vec{\nabla}\times\vec{A})$ was identically zero for any vector field $\vec{A}$.

Thus if we define $\vec{B}$ as the curl of some new vector field $\vec{A}$, then one of Maxwell's equations, $\vec{\nabla}\cdot\vec{B} = 0$ is automatically satisfied.

Our next step is to see what happens when we introduce the vector potential into the other Maxwell equations. Let us start with Faraday's law

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \qquad (10)$$

If we replace $\vec{B}$ with $\vec{\nabla}\times\vec{A}$ we get

$$\vec{\nabla}\times\vec{E} = -\frac{\partial}{\partial t}(\vec{\nabla}\times\vec{A}) \qquad (11)$$

Using the fact that we can change the order of partial differentiation, and remembering that the curl is just a lot of partial derivatives, we get

$$\vec{\nabla}\times\vec{E} = \vec{\nabla}\times\left(-\frac{\partial\vec{A}}{\partial t}\right) \qquad \begin{matrix}\textit{Faraday's law}\\ \textit{in terms of } \vec{A}\end{matrix} \qquad (12)$$

We see that Equation (12) would be satisfied if we could set $\vec{E} = -\partial\vec{A}/\partial t$ on the left side.

We cannot do that, however, because we already know that for static charges, $\vec{E} = -\vec{\nabla}\phi$. But see what happens if we try the combination

$$\boxed{\vec{E} = -\vec{\nabla}\phi - \frac{\partial\vec{A}}{\partial t}} \qquad \begin{matrix}\textit{electric field}\\ \textit{in terms of}\\ \textit{potentials}\\ \phi \textit{ and } \vec{A}\end{matrix} \qquad (13)$$

Taking the curl of Equation (7) gives

$$\vec{\nabla}\times\vec{E} = -\vec{\nabla}\times(\vec{\nabla}\phi) - \vec{\nabla}\times\frac{\partial\vec{A}}{\partial t} \qquad (14)$$

Since $\vec{\nabla}\times(\vec{\nabla}\phi) = 0$ because the curl of a gradient is identically zero, we get

$$\vec{\nabla}\times\vec{E} = -\vec{\nabla}\times\frac{\partial\vec{A}}{\partial t} \qquad (15)$$

Next interchange the order of partial differentiation to get

$$\vec{\nabla}\times\vec{E} = -\frac{\partial}{\partial t}(\vec{\nabla}\times\vec{A}) = -\frac{\partial\vec{B}}{\partial t} \qquad (16)$$

which is Faraday's law.

Thus when we define the electric and magnetic fields $\vec{E}$ and $\vec{B}$ in terms of the potentials $\phi$ and $\vec{A}$ by

$$\vec{B} = \vec{\nabla}\times\vec{A} \qquad (8) \text{ repeated}$$

$$\vec{E} = -\vec{\nabla}\phi - \partial\vec{A}/\partial t \qquad (13) \text{ repeated}$$

then two of Maxwell's equations

$$\vec{\nabla}\cdot\vec{B} = 0 \qquad \textit{no monopole}$$

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \qquad \textit{Faraday's law}$$

are automatically satisfied.

You can now see how we handle potentials or voltages when changing magnetic fields are present. For the field of static charges, we have $\vec{E} = -\vec{\nabla}\phi$ as before. When changing magnetic fields are present, we get an additional contribution to $\vec{E}$ due to the $-\partial\vec{A}/\partial t$ term.

In Maxwell's theory of electric and magnetic fields, in what is often called the ***classical theory of electromagnetism***, you can solve all problems by using Maxwell's equations as shown in Equation (7) and never bother with introducing the vector potential $\vec{A}$. In the classical theory, the potentials are more of a mathematical convenience, trimming the number of Maxwell's equations from four to two because two of them are automatically handled by the definition of the potentials.

Things are different in quantum theory. There are experiments involving the wave nature of the electron that detect the vector potential $\vec{A}$ directly. These experiments cannot be explained by the fields $\vec{E}$ and $\vec{B}$ alone. It turns out in quantum mechanics that the potentials $\phi$ and $\vec{A}$ are the fundamental quantities and $\vec{E}$ and $\vec{B}$ are derived concepts, concepts derived from the equations $\vec{B} = \vec{\nabla}\times\vec{A}$ and $\vec{E} = -\vec{\nabla}\phi - \partial\vec{A}/\partial t$ .

## WAVE EQUATIONS FOR $\phi$ AND $\vec{A}$

The other two Maxwell's equations turn out to be wave equations for $\phi$ and $\vec{A}$. There is one surprise in store. So far we have defined only the curl of $\vec{A}$ through the equation $\vec{B} = \vec{\nabla} \times \vec{A}$. In general a vector field like $\vec{A}$ can have both a divergent part $\vec{A}_{div}$ and a solenoidal part $\vec{A}_{sol}$ where

$$\vec{A} = \vec{A}_{div} + \vec{A}_{sol} \qquad (17)$$

where the divergent part has no curl and the solenoidal part has no divergence

$$\vec{\nabla} \times \vec{A}_{div} = 0 \qquad (18a)$$

$$\vec{\nabla} \cdot \vec{A}_{sol} = 0 \qquad (18b)$$

We saw this kind of separation in the case of electric fields. When the electric field was created by static electric charges it was purely divergent, i.e., had zero curl. An electric field created by a changing magnetic field is purely solenoidal, with zero divergence.

As a result our equation $\vec{B} = \vec{\nabla} \times \vec{A}$ defines only the solenoidal part of $\vec{A}$, namely $\vec{A}_{sol}$. We are still free to choose $\vec{A}_{div}$, which has not been specified yet. We will see that we can choose $\vec{A}_{div}$ or $\vec{\nabla} \cdot \vec{A}$ in such a way that considerably simplifies the wave equations for $\phi$ and $\vec{A}$. This choice is not essential, only convenient. Sometimes, in fact, it is more convenient not to specify any choice for $\vec{A}_{div}$, and to work with the more general but messier wave equations.

For very obscure historical reasons, the choice of a special value for $\vec{\nabla} \cdot \vec{A}$ is called a ***choice of gauge***. In a later chapter we will look very carefully at what it means to make different choices for $\vec{\nabla} \cdot \vec{A}$. We will see that there are no physical predictions affected in any way by changing our choice for $\vec{\nabla} \cdot \vec{A}$. As a result the theory of electromagnetism is said to be ***invariant*** under different choices of gauge, or ***gauge invariant***. This feature of electromagnetism will turn out to have extremely important implications, particularly in the quantum theory. For now, however, we will simply make a special choice of $\vec{\nabla} \cdot \vec{A}$ that simplifies the form of Maxwell's equations for $\phi$ and $\vec{A}$.

The two Maxwell's equations that are not automatically satisfied by $\vec{B} = \vec{\nabla} \times \vec{A}$ and $\vec{E} = -\vec{\nabla}\phi - \partial\vec{A}/\partial t$ are

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0} \qquad \textit{Gauss' law}$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{i} + \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \qquad \textit{Ampere's law}$$

Making the substitutions $\vec{E} = -\vec{\nabla}\phi - \partial\vec{A}/\partial t$ in Gauss's law gives

$$\vec{\nabla} \cdot \vec{E} = \vec{\nabla} \cdot \left( -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \right) = \frac{\rho}{\varepsilon_0} \qquad (19)$$

Noting that $\vec{\nabla} \cdot \partial\vec{A}/\partial t = \partial(\vec{\nabla} \cdot \vec{A})/\partial t$ because we can change the order of partial differentiation, and that $\vec{\nabla} \cdot (\vec{\nabla}\phi) = \nabla^2\phi$, we get

$$-\nabla^2\phi - \frac{\partial(\vec{\nabla} \cdot \vec{A})}{\partial t} = \frac{\rho}{\varepsilon_0}$$

$$-\nabla^2\phi = \frac{\rho}{\varepsilon_0} + \frac{\partial(\vec{\nabla} \cdot \vec{A})}{\partial t} \qquad (20)$$

You can see the divergence of $\vec{A}$, namely $\vec{\nabla} \cdot \vec{A}$ appearing in the equation for $\phi$.

Making the substitutions in Ampere's law gives

$$\vec{\nabla} \times \vec{B} = \vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \mu_0 \vec{i} + \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t}$$

$$= \mu_0 \vec{i} + \mu_0 \varepsilon_0 \frac{\partial}{\partial t} \left( -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \right) \qquad (21)$$

Using the third vector identity of Chapter 9, namely

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = -\nabla^2\vec{A} + \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) \qquad (9\text{-}3)$$

Equation 21 becomes

$$-\nabla^2\vec{A} + \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) = \mu_0 \vec{i} - \mu_0 \varepsilon_0 \frac{\partial(\vec{\nabla}\phi)}{\partial t} - \mu_0 \varepsilon_0 \frac{\partial^2 \vec{A}}{\partial t^2} \qquad (22)$$

Writing $\partial(\vec{\nabla}\phi)/\partial t = \vec{\nabla}(\partial\phi/\partial t)$ and moving the $\partial^2(\vec{A})/\partial t^2$ term to the left and $\vec{\nabla}(\vec{\nabla}\cdot\vec{A})$ to the right gives

$$-\nabla^2\vec{A} + \mu_0\varepsilon_0\frac{\partial^2(\vec{A})}{\partial t^2} = \mu_0\vec{i} - \vec{\nabla}\left(\mu_0\varepsilon_0\frac{\partial\phi}{\partial t}\right) - \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) \qquad (23)$$

In Equation (23) we see the wave equation for $\vec{A}$ appearing on the left side, but we have some weird stuff involving $\vec{\nabla}\cdot\vec{A}$ and $\partial\phi/\partial t$ on the right. We can simplify things a bit by noting that both of these terms have a factor of $\vec{\nabla}$ and writing

$$\boxed{-\nabla^2\vec{A} + \frac{1}{c^2}\frac{\partial^2(\vec{A})}{\partial t^2} = \mu_0\vec{i} - \vec{\nabla}\left(\vec{\nabla}\cdot\vec{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t}\right)} \quad \textit{Ampere's law} \qquad (24)$$

where we have replaced $\mu_0\varepsilon_0$ by $1/c^2$, c being the speed of light.

Equation (24) is beginning to look like a wave equation with some peculiar terms on the right hand side. Equation (20) for $\phi$ does not, at least now, look like a wave equation. However we can make it look like a wave equation by adding the term $(1/c^2)(\partial^2\phi/\partial t^2)$ to both sides, giving

$$-\nabla^2\phi + \frac{1}{c^2}\frac{\partial^2(\phi)}{\partial t^2} = \frac{\rho}{\varepsilon_0} + \frac{1}{c^2}\frac{\partial^2(\phi)}{\partial t^2} + \frac{\partial(\vec{\nabla}\cdot\vec{A})}{\partial t} \qquad (25)$$

We can factor out a $\partial/\partial t$ in the last two terms on the right side of Equation (25) giving us

$$\boxed{-\nabla^2\phi + \frac{1}{c^2}\frac{\partial^2(\phi)}{\partial t^2} = \frac{\rho}{\varepsilon_0} + \frac{\partial}{\partial t}\left(\vec{\nabla}\cdot\vec{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t}\right)} \quad \textit{Gauss' law} \qquad (26)$$

The rather messy looking Equations (24) and (26) are Ampere's law and Gauss' law written in terms of the scalar and vector potentials $\phi$ and $\vec{A}$.

On the left side of each we have the beginning of a wave equation, but somewhat of a mess on the right. However we see that the term

$$\vec{\nabla}\cdot\vec{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t} \qquad (27)$$

is common to both equations. If we could find some way to get rid of this term, there would be a considerable simplification.

We have, however, not yet specified what the value of $\vec{\nabla}\cdot\vec{A}$ should be. We have only specified $\vec{\nabla}\times\vec{A} = \vec{B}$. If we make the choice

$$\boxed{\vec{\nabla}\cdot\vec{A} = -\frac{1}{c^2}\frac{\partial\phi}{\partial t}} \quad \textit{special choice of gauge} \qquad (28)$$

then the term (27) goes to zero. Making a choice for $\vec{\nabla}\cdot\vec{A}$ is called ***making a choice of gauge***, and this particular choice leads to the much simpler equations

$$\boxed{-\nabla^2\phi + \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = \frac{\rho}{\varepsilon_0}} \quad \textit{Gauss' law} \qquad (29)$$

$$\boxed{-\nabla^2\vec{A} + \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} = \mu_0\vec{i}} \quad \textit{Ampere's law} \qquad (30)$$

We get the rather elegant result that both potentials, the scalar potential $\phi$ and vector potential $\vec{A}$, obey wave equations with source terms on the right hand side. The source for the scalar potential is the charge density $\rho/\varepsilon_0$, and the source for the vector potential is the current density $\mu_0\vec{i}$.

**Exercise 1**

The choice of gauge we made to get Equations (29) and (30) was $\vec{\nabla}\cdot\vec{A} = -(1/c^2)\partial\phi/\partial t$ . This gave us simple wave equations which are convenient if we are working with electromagnetic waves. Sometimes another choice of gauge is more convenient. Derive Gauss' law and Ampere's law in terms of $\phi$ and $\vec{A}$, using the choice of gauge

$$\vec{\nabla}\cdot\vec{A} = 0 \qquad \textit{Coulomb gauge} \qquad (31)$$

which is called the ***Coulomb gauge***.

Do this derivation two ways. One by starting from Maxwell's equations in terms of $\vec{E}$ and $\vec{B}$, and secondly, starting from Equations (24) and (26) where we made no special choice of gauge.

**Exercise 2**

This exercise is optional, but should give some very good practice with Maxwell's equations. In Chapter 9 we derived the wave equation for electromagnetic waves in empty space by first writing Maxwell's equations for empty space, Equations (9-12), and then taking the curl of Ampere's and Faraday's law. The results were

$$-\nabla^2\vec{E} + \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} = 0$$

$$-\nabla^2\vec{B} + \frac{1}{c^2}\frac{\partial^2\vec{B}}{\partial t^2} = 0 \qquad \textit{wave equations in empty space} \qquad (9\text{-}20)$$

Now repeat these calculations for the case that the charge and current densities $\rho$ and i are not zero. Show that you get the following wave equations for $\vec{E}$ and $\vec{B}$

$$-\nabla^2\vec{E} + \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} = -\frac{\vec{\nabla}\rho}{\varepsilon_0} - \mu_0\frac{\partial\vec{i}}{\partial t} \qquad (32)$$

$$-\nabla^2\vec{B} + \frac{1}{c^2}\frac{\partial^2\vec{B}}{\partial t^2} = \mu_0\vec{\nabla}\times\vec{i} \qquad (33)$$

You can see that we still get wave equations for $\vec{E}$ and $\vec{B}$, but the source terms, the stuff on the right hand side, are much more complex than the source terms for the wave equations for $\phi$ and $\vec{A}$. For example, the source term for the $\vec{A}$ wave is simply $\mu_0\vec{i}$, while the source term for a $\vec{B}$ wave is the $\mu_0\vec{\nabla}\times\vec{i}$ . It is even worse for the $\vec{E}$ field. Instead of the source term $\rho/\varepsilon_0$ for the $\phi$ field, we have $(-\nabla\rho/\varepsilon_0 - \mu_0\partial\vec{i}/\partial t)$ as a source for the $\vec{E}$ wave.

## Summary

Here we collect in one place, all the forms of Maxwell's equations.

(a) Maxwell's equations in terms of $\vec{E}$ and $\vec{B}$

$$\vec{\nabla}\cdot\vec{E} = \frac{\rho}{\varepsilon_0} \qquad \textit{Gauss' law}$$

$$\vec{\nabla}\cdot\vec{B} = 0 \qquad \textit{no monopole}$$

$$\vec{\nabla}\times\vec{B} = \mu_0\vec{i} + \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t} \qquad \textit{Ampere's law}$$

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \qquad \textit{Faraday's law}$$

(b) Wave equations for $\vec{E}$ and $\vec{B}$

$$-\nabla^2\vec{E} + \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} = -\frac{\vec{\nabla}\rho}{\varepsilon_0} - \mu_0\frac{\partial\vec{i}}{\partial t}$$

$$-\nabla^2\vec{B} + \frac{1}{c^2}\frac{\partial^2\vec{B}}{\partial t^2} = \mu_0\vec{\nabla}\times\vec{i}$$

For the wave equations in empty space, set $\rho = 0$ and $\vec{i} = 0$.

(c) Scalar and vector potentials $\phi$ and $\vec{A}$

$$\vec{B} = \vec{\nabla}\times\vec{A}$$

$$\vec{E} = -\vec{\nabla}\phi - \partial\vec{A}/\partial t$$

These automatically satisfy

$$\vec{\nabla}\cdot\vec{B} = 0$$

$$\vec{\nabla}\times\vec{E} = -\partial\vec{B}/\partial t$$

The remaining two Maxwell's equations become

$$-\nabla^2\phi + \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = \frac{\rho}{\varepsilon_0} + \frac{\partial}{\partial t}\left[\vec{\nabla}\cdot\vec{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t}\right]$$

$$-\nabla^2\vec{A} + \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} = \mu_0\vec{i} - \vec{\nabla}\left[\vec{\nabla}\cdot\vec{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t}\right]$$

The terms in the square brackets can be set to zero with the choice of gauge

$$\vec{\nabla}\cdot\vec{A} = -\frac{1}{c^2}\frac{\partial\phi}{\partial t} \qquad \textit{special choice of } \vec{\nabla}\cdot\vec{A}$$

With this choice of gauge, Maxwell's equations reduce to

$$-\nabla^2\phi + \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = \frac{\rho}{\varepsilon_0}$$

$$-\nabla^2\vec{A} + \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} = \mu_0\vec{i}$$

*all that is left of Maxwell's equations*

# Calculus 2000-Chapter 12

# Vorticity

*At the beginning of Part II of the Physics text, we used the velocity field to introduce the concept of a vector field. It is easier to picture velocity vectors attached to water molecules in a flowing stream than to visualize a vector at each point in space. We could introduce Gauss' law as a conservation law for an incompressible fluid, and then show that the electric field behaved in a similar way.*

*Since that early introduction, we have come a long way in our study of the mathematical behavior of vector fields. In this and the next chapter, we will turn the tables on our earlier approach and apply to the velocity field the techniques and insights we have gained in our study of electric and magnetic fields. This will lead to a much deeper understanding of the behavior of fluids than we got in our old discussion of Bernoulli's equation.*

*The most important concept that carries us beyond Bernoulli's equation is* ***vorticity****, which is the curl of the velocity field. Vorticity is important not only in the study of vortex structures like vortex rings and tornadoes, it plays a fundamental role in all aspects of fluid motion. In this chapter, we will develop an intuitive picture of vorticity. In the next chapter, we focus on its dynamic behavior.*

*These two chapters are designed to be an introduction to the basic concepts of* ***fluid dynamics****. For most of the past century, this subject has been eliminated from the undergraduate physics curriculum, despite exciting advances in the understanding of the behavior of superfluids. One of our aims with these chapters is to bring this subject back.*

## DIVERGENCE FREE FIELDS

In the Physics text, we have often noted the similarity between the magnetic field and the velocity field. The fact that there are no magnetic charges led to the equation

$$\int_S \vec{B}\cdot d\vec{A} = 0 \qquad \textit{for any closed surface S} \qquad (1)$$

For an incompressible fluid like water, the continuity equation, i.e., the fact that we cannot create or destroy water molecules, leads to the equation

$$\int_S \vec{v}\cdot d\vec{A} = 0 \qquad \textit{for any closed surface S} \qquad (2)$$

With the introduction of our differential notation, we saw that Equation (1) for the magnetic field became

$$\vec{\nabla}\cdot\vec{B} = 0 \qquad (1a)$$

The same mathematics leads to the equation for the velocity field

$$\vec{\nabla}\cdot\vec{v} = 0 \qquad \textit{continuity equation for an incompressible fluid} \qquad (2a)$$

Thus we see that both the magnetic field, and the velocity field of an incompressible fluid, are divergence free fields.

Another way to see the same result is to look at the form of the continuity equation we discussed a short while ago in Chapter 10. We saw how Maxwell's equations automatically led to a continuity equation for electric charge. That equation was

$$\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot\vec{i} = 0 \qquad \textit{continuity equation for electric charge} \qquad \text{(Cal 10-8)}$$

When applied to a fluid of mass density $\rho$ and mass current density $\rho\vec{v}$ the continuity equation for mass becomes

$$\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{v}) = 0 \qquad \textit{continuity equation for a fluid of mass density } \rho \qquad (3)$$

If the fluid density $\rho$ is constant, then $\partial\rho/\partial t = 0$ and $\vec{\nabla}\rho = 0$ .This leads to $\vec{\nabla}\cdot(\rho\vec{v}) = \rho\vec{\nabla}\cdot\vec{v} = 0$ and we are left with

$$\vec{\nabla}\cdot\vec{v} = 0 \qquad \text{(2a) repeated}$$

as the continuity equation for a constant density fluid.

## THE VORTICITY FIELD

When we were discussing electric and magnetic fields in the Physics text, we found that we needed equations for both the surface integral and the line integral in order to specify the field. That is why we ended up with four Maxwell's equations in order to describe the two fields $\vec{E}$ and $\vec{B}$. In the Calculus text, we have shrunk the surface and line integrals down to infinitesimal size where they become the divergence and the curl. Thus to specify a field, we now need equations for both the divergence and curl of the field.

As we mentioned in Chapter 9, if we have a field known only in some limited volume of space, like the velocity field of a fluid within a section of pipe, then ***in order to uniquely determine the field, we must know not only the divergence and curl within that volume, but also the perpendicular components of the field at the volume's surface***. It is the perpendicular components of the velocity field at the volume's surface that tell us how the fluid is flowing in and out.

For a constant density or incompressible fluid, we already know that the divergence is zero. Thus if we know how the fluid is flowing into and out of a volume, the only other thing we need to specify is its curl $\vec{\nabla}\times\vec{v}$ inside. From this point of view we see that the curl $\vec{\nabla}\times\vec{v}$ plays a key role in determining the nature of fluid flows. It should thus not be too surprising that most of this chapter is devoted to understanding the nature and behavior of the curl $\vec{\nabla}\times\vec{v}$.

Our first step will be to give the curl $\vec{\nabla}\times\vec{v}$ a name. We will call it ***vorticity*** and designate it by the Greek letter $\vec{\omega}$ (omega).

$$\boxed{\vec{\omega} \equiv \vec{\nabla}\times\vec{v}} \qquad \textit{vorticity} \qquad (4)$$

At this point, we have a slight problem with notation. In the Physics text we used the symbol $\omega$ to designate angular velocity $d\theta/dt$. While there is some relationship between angular velocity $d\theta/dt$ and vorticity $\vec{\omega} = \vec{\nabla}\times\vec{v}$, they are different quantities. Worse yet, in one important example, namely the rotation of a solid body, they differ by exactly a factor of 2. To avoid ambiguity, we will in this chapter use $\vec{\omega}$ for vorticity $\vec{\nabla}\times\vec{v}$, and the symbol $\omega_{rot}$ for angular velocity.

$$\boxed{\omega_{rot} \equiv \frac{d\theta}{dt}} \qquad \textit{angular velocity} \qquad (5)$$

## POTENTIAL FLOW

In the next few sections, we will develop an intuition for the concept of vorticity by considering various examples. We will start with the simplest example, namely flow with no vorticity, i.e., when $\vec{\nabla} \times \vec{v} = 0$. Such flows are called ***potential flows***. The reason for the name is as follows.

In our early discussion of electric fields, we pointed out that both the gravitational field, and the electric field of stationary point charges were conservative fields. A conservative field was defined as one where the total work done by the field acting on a mass or charge was zero if we carried the particle around and came back to the original starting point. (See page 25-5 of the Physics text.) For the work done by an electric field on a unit test charge, this statement took the form

$$\oint \vec{E} \cdot d\vec{\ell} = 0 \qquad \begin{matrix} \textit{condition that } \vec{E} \\ \textit{is a conservative field} \end{matrix} \qquad (6)$$

In our differential notation, Equation (6) becomes

$$\vec{\nabla} \times \vec{E} = 0 \qquad \begin{matrix} \textit{condition that } \vec{E} \\ \textit{is a conservative field} \end{matrix} \qquad (7)$$

You will recall that when $\vec{E}$ was a conservative field, we could introduce a unique potential energy provided we defined the zero of potential energy. We called the potential energy of a unit test charge ***electric voltage*** or ***electric potential***.

When we got to Faraday's law, we had some problems with the concept of electric voltage. In our discussion of the betatron where electrons are circling a region of changing magnetic flux, the electrons gained voltage each time they went around the circle. When a changing magnetic field or magnetic flux $\Phi_B$ is present, the voltage or electric potential is not unique because the electric field is no longer a conservative field. Faraday's law in integral and differential form is

$$\oint \vec{E} \cdot d\vec{\ell} = -\frac{d\Phi_B}{dt} \qquad \text{(Physics 32-19)}$$

$$\vec{\nabla} \times \vec{E} = -\frac{dB}{dt} \qquad (8\text{-}49)$$

and we see that $\vec{\nabla} \times \vec{E}$ is no longer zero.

When $\vec{\nabla} \times \vec{E}$ is zero we have a unique electric voltage (once we have defined the zero of voltage), and we can use the concept of the gradient, discussed in the Calculus Chapter 3, to calculate the electric field from the voltage. The formula we had was

$$\vec{E} = -\vec{\nabla} V(x, y, z) \qquad (3\text{-}19)$$

where $V(x, y, z)$ is the voltage.

By similar arguments, if we have a conservative velocity field $\vec{v}$, one obeying the condition

$$\vec{\nabla} \times \vec{v} = 0 \qquad \begin{matrix} \textit{conservative} \\ \textit{velocity field} \end{matrix} \qquad (8)$$

then we can introduce potential $\varphi(x, y, z)$ that is analogous to the voltage $V(x, y, z)$ for the electric field. In terms of the potential $\varphi$, the velocity field $\vec{v}$ would be given by

$$\vec{v} = -\vec{\nabla}\varphi \qquad \begin{matrix} \textit{velocity field derived} \\ \textit{from a potential} \end{matrix} \qquad (9)$$

Because such a velocity field is derived from a potential $\varphi$, the flow field is called *potential flow*.

As a quick check that our formulas are working correctly, suppose we start with some potential flow $\vec{v} = -\vec{\nabla}\varphi$ and ask what its curl is. We have

$$\vec{\nabla} \times \vec{v} = \vec{\nabla} \times (-\nabla\varphi) \qquad (10)$$

One of the vector identities, from Calculus Chapter 9 was

$$\vec{\nabla} \times (\vec{\nabla} f) = 0 \qquad (9\text{-}1)$$

where f is any scalar function. Thus $\vec{\nabla} \times (\vec{\nabla}\varphi)$ is identically zero, and any flow derived from a potential $\varphi$ has to have zero curl, or no vorticity.

## Examples of Potential Flow

If we combine the equation $\vec{v} = -\vec{\nabla}\varphi$ for potential flow with the divergence free condition $\vec{\nabla}\cdot\vec{v} = 0$ we get

$$\vec{\nabla}\cdot\vec{v} = \vec{\nabla}\cdot(-\vec{\nabla}\varphi) = 0$$

or

$$\nabla^2\varphi = 0 \qquad (11)$$

The operator $\nabla^2$ is the Laplacian operator we discussed in detail in Chapter 4. Equation (11) itself is known as ***Poisson's equation***.

To find examples of potential flow, one can use Equation (11) subject to the boundary conditions on the velocity field at the walls of the container. A number of techniques have been developed to solve this problem, both approximation techniques for analytical solutions and numerical techniques for computer solutions. We are not going to discuss these techniques because the work is hard and the results are not particularly applicable to real fluid flows. We will see that almost all fluid flows involve vorticity, and our interest in this chapter will be the behavior of the vorticity. When we need a potential flow solution, we will either choose one simple enough to guess the shape or rely on someone else's solution.

## Potential Flow in a Sealed Container

As our first example, suppose we have a constant density fluid in a completely sealed container. That means that no fluid is flowing in or out. Now suppose the fluid has no vorticity, that $\vec{\nabla}\times\vec{v} = 0$ inside. The resulting flow then must be a potential flow.

One possible solution for $\vec{\nabla}\times\vec{v} = 0$ is that the fluid inside is at rest (assuming that the container walls are at rest). That is,

$$\vec{v} = 0 \qquad \textit{a potential flow solution for a sealed container} \qquad (12)$$

This solution clearly obeys the condition $\vec{\nabla}\times\vec{v} = 0$ and $\vec{\nabla}\cdot\vec{v} = 0$, and has no normal flow at the boundary walls.

What other potential flow solutions are there? **NONE**. Our mathematical theorem given at the beginning of the chapter states that the vector field $\vec{v}$ is uniquely determined if we specify $\vec{\nabla}\cdot\vec{v}$ and $\vec{\nabla}\times\vec{v}$ within a closed volume V and the normal components of $\vec{v}$ at the surface of V. We have done that. Thus the solution $\vec{v} = 0$ is unique, and there is no other potential flow solution.

This solution emphasizes the importance of vorticity in the study of fluid flows. If we have a sealed container filled with a constant density fluid, there can be no flow without vorticity. ***In this case, the source of all fluid motion must be vorticity.*** This is why it is so important in the study of fluid behavior to understand the role and behavior of vorticity.

## Potential Flow in a Straight Pipe

We began our discussion of fluid motion in Chapter 23 of the Physics text, with the example of a fluid entering a pipe at a velocity $\vec{v}_1$ and exiting at a velocity $\vec{v}_2$ as shown in Figure (1). We assumed that $\vec{v}_1$ was uniform over the entire inlet and $\vec{v}_2$ over the entire exit. The continuity equation gave $v_1 A_1 = v_2 A_2$. If the pipe is uniform, so that $A_1 = A_2$, we get $v_1 = v_2$.

What is the potential flow solution for the uniform pipe of Figure (1)? One possible answer is shown in Figure (2), namely that the velocity field is a constant throughout the pipe.

$$\vec{v} = \vec{v}_1 = \text{constant} \quad \begin{matrix}\textit{potential}\\ \textit{flow solution}\end{matrix} \qquad (13)$$

Let us check that $\vec{v} = \vec{v}_1 =$ constant is a potential flow solution. It is clear that the divergence $\vec{\nabla}\cdot\vec{v}_1$ and the curl $\vec{\nabla}\times\vec{v}_1$ are both zero for a constant vector field $\vec{v}_1$. Thus the flow $\vec{v} = \vec{v}_1$ is potential flow. The solution $\vec{v} = \vec{v}_1$ also has the correct normal components, being $\vec{v}_1$ at the entrance and exit, and no normal flow at the pipe walls. Thus Figure (2), with $\vec{v} = \vec{v}_1 =$ constant, is our unique solution for potential flow in a straight pipe with uniform entrance and exit velocities. As we said, in some cases we can guess the potential flow solutions.

The problem with the potential flow solution of Figure (2) is that a fluid like water cannot flow that way. In Figure (2), the fluid is slipping at the pipe walls. The first layer of atoms next to the walls is moving just as fast as the atoms in the center of the flow. For all normal fluids the first layer of atoms is stuck to the wall by molecular forces, and due to viscous effects, the fluid velocity has to increase gradually as we go into the fluid. There is no potential flow solution for pipe flow that has this property, thus all flows of normal fluids in a pipe must involve vorticity.

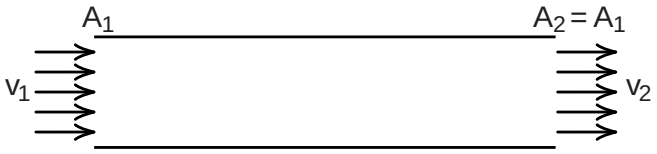



**Figure 1**

***A fluid enters a uniform pipe at a velocity $\vec{v}_1$.***

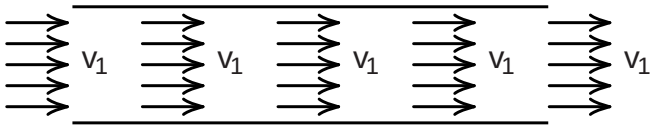



**Figure 2**

***One possible solution to the potential flow problem. If we have a uniform pipe, with a uniform inlet and outflow velocities as shown in Figure (1), then this is the only solution.***

## SUPERFLUIDS

Normal fluids like water cannot slip along the surface of a pipe, but superfluids, which have zero viscosity, can. As a result a superfluid can have a potential flow pattern like that shown in Figure (2). We have good experimental evidence that in a number of examples superfluid helium does flow that way.

In the 1940s, the Russian physicist Lev Landau made the prediction, based on his wave equation for the atoms in a superfluid, that superfluid helium had to flow without vorticity, that $\vec{\nabla} \times \vec{v} = 0$ and only potential flow solutions would be possible. This was a prediction that was fairly easy to check by the following experiment.

If you place a glass of water on a spinning turntable and wait until the water rotates with the glass, the surface of the water will be slightly curved, as the water is pushed to the outside by "centrifugal forces". (If you choose a coordinate system that is rotating with the glass, then in this rotating coordinate system there is an outward centrifugal pseudo force.) The shape of the surface of the water turns out to be a parabola. In fact, large modern telescopes are now made by cooling the molten glass in a rotating container so that the rough parabolic shape is already there when the glass hardens.

Now consider how superfluid helium should behave when in such a rotating container. If the container is circular, like a drinking glass, and centered on the axis of rotation, the container can rotate without forcing the fluid to have any sideways motion. Also no fluid is flowing into or out of the bottom or top. Thus the normal or perpendicular component of flow is zero all around the fluid.

Superfluid helium is essentially a constant density fluid, thus $\vec{\nabla} \cdot \vec{v} = 0$ within the fluid. If Landau were right, then $\vec{\nabla} \times \vec{v}$ should also be zero inside the fluid, and we would have to have potential flow.

We have already discussed the potential flow solution for this case. If there is no normal flow through the fixed boundaries of the fluid, the unique potential flow solution for a constant density fluid is $\vec{v} = 0$. The fluid cannot rotate with the bucket. It cannot move at all! We get the unique prediction that the fluid must be at rest, and as a result the surface of the fluid must be flat. This prediction is easy to test; rotate a bucket of superfluid helium and see if the surface is flat or parabolic.

There are a few complications to the experiment. Above a temperature of 2.17 kelvins, liquid helium is a normal fluid with viscosity like other fluids with which we are familiar. When helium is cooled to just below 2.17 kelvins, superfluidity sets in, but in a rather peculiar way. The best way to understand the properties of liquid helium below 2.17 k is to think of it as a mixture of two fluids, a normal fluid with viscosity and a superfluid with no viscosity. At the temperature 2.17 k, the helium is almost all normal fluid. As we cool further, we get more superfluid and less normal fluid. Down at a temperature of 1 kelvin, which is quite easy to reach experimentally, almost all the normal fluid is gone and we have essentially pure superfluid.

In Landau's picture, the normal fluid below 2.17 k has viscosity, is not bound by the condition $\vec{\nabla} \times \vec{v} = 0$, and thus can rotate. Only the superfluid component must have $\vec{\nabla} \times \vec{v} = 0$ and undergo only potential flow. Thus if we have a rotating bucket of superfluid helium at just below 2.17 k, it should be mostly normal fluid and eventually start rotating with the bucket. We should expect to see a parabolic surface, and that is what is seen experimentally.

However, as we cool the helium from just below 2.17 k down to 1 k, the normal fluid turns to superfluid. If Landau were right, the flow should go over to a potential flow and the surface of the liquid should become flat even though the container keeps rotating. This does not happen, and something has to be wrong with Landau's prediction. The curved surface at 1 k indicates that the superfluid is moving, and thus must contain some vorticity. In a later section we will see how Feynman was able to explain the parabolic surface, while still obeying Landau's condition $\vec{\nabla} \times \vec{v} = 0$ almost everywhere in the fluid.

## VORTICITY AS A SOURCE OF FLUID MOTION

In our discussion of potential flow of a constant density fluid in a sealed container, we saw that there could be no flow without vorticity. Vorticity must be the source of any flow found there. In this section, we will illustrate the idea that vorticity is the source of fluid motion by comparing the velocity field with the magnetic field of electric currents. We will see that vorticity is a source of the velocity field in much the same way that an electric current is a source of the magnetic field.

In our discussion of magnetic fields, it was clear that magnetic fields are created by electric currents. Before we learned about Maxwell's correction to Ampere's law, the relationship between the magnetic field $\vec{B}$ and the current i was

$$\oint \vec{B} \cdot d\vec{\ell} = \mu_0 i \qquad \textit{old Ampere's law} \qquad (29\text{-}18)$$

where i was the total electric current flowing through the closed integration loop. Shrinking the integration loop down to infinitesimal size, i.e., going to our differential notation, we get

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{i} \qquad (14)$$

where $\vec{i}$ is the electric current density. Equation (14), which is missing the $\partial\vec{E}/\partial t$ term of Maxwell's equation, applies if we can neglect changing electric flux.

In the Physics text, we used the old form of Ampere's law to calculate the magnetic field of a straight wire and of a solenoid. In these examples it was clear that the current i in the wire was the source of the magnetic field.

Let us now compare the equations we have for the magnetic field $\vec{B}$ (neglecting $\partial\vec{E}/\partial t$ terms) and for the velocity field $\vec{v}$ of a constant density fluid. We have

| *Magnetic field* | *Velocity Field of Constant Density Fluid* | |
|---|---|---|
| $\vec{\nabla} \cdot \vec{B} = 0$ | $\vec{\nabla} \cdot \vec{v} = 0$ | |
| $\vec{\nabla} \times \vec{B} = \mu_0 \vec{i}$ | $\vec{\nabla} \times \vec{v} = \vec{\omega}$ | (15) |

where $\vec{\omega}$ is the vorticity field of the fluid. If we can interpret $\mu_0 \vec{i}$ as the source of the magnetic field in the equation $\vec{\nabla} \times \vec{B} = \mu_0 \vec{i}$, then by analogy we should be able to interpret the vorticity $\vec{\omega}$ as the source of the velocity field in the equation $\vec{\nabla} \times \vec{v} = \vec{\omega}$.

To be more precise, we will see that the vorticity $\vec{\omega}$ can be interpreted as the source of any additional velocity beyond the simple potential flow we discussed earlier. If boundary layers, vortices, turbulence, or other derivations from potential flow are present, we can say that vorticity is responsible.

## Picturing Vorticity

When we discussed the magnetic field of a current, the current itself was quite easy to picture. It was the flow of electrons along the wire, and for a straight wire this flow of charge produced a circular magnetic field around the wire as shown in Figure (3). We also found from Ampere's law that the strength of the circular magnetic field dropped off as 1/r as we went out from the wire.

In Figure (4) we have drawn a picture of the velocity field of a straight vortex like the one pictured in Figure (23-25) of the Physics text. We observed that the fluid travels in circles around the vortex core. In our funnel vortex we made the core hollow by letting fluid flow out of the funnel, but initially the core contained fluid. We also saw that the fluid flowed faster near the core than far away. The tendency for a fluid vortex is for the velocity field to drop off as 1/r out from the core.

Since the circular velocity field of a straight vortex is similar to the circular magnetic field of a current in a straight wire, we should expect that both fields have similar sources. In Figure (3) the source of the magnetic field is an upward directed current density $\vec{i}$ in the wire. We therefore expect that the source of the vortex velocity field in Figure (4) should be an upward directed vorticity $\vec{\omega}$ in the center of the vortex.

Outside the wire, the circular magnetic field drops off as 1/r and has zero curl. If the circular velocity field of the vortex drops off as 1/r outside the core, it must have zero curl there also. Thus a vortex with a 1/r velocity field outside the core must have all the vorticity $\vec{\omega}$ concentrated inside the core, just as the current producing the magnetic field is confined to the wire. The vorticity must run up the core as shown in Figure (5). We are beginning to see how the vorticity acts as a source of the velocity field in the same way currents are the source of magnetic fields.

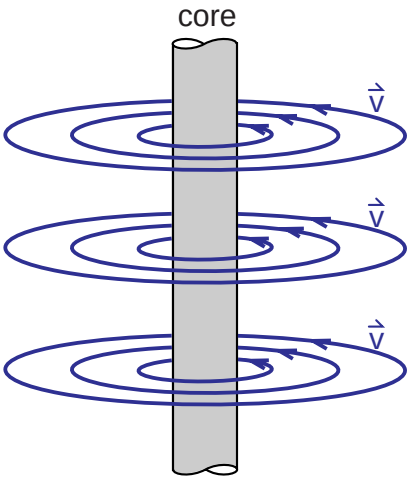



**Figure 4**
***Circular velocity field around a vortex core.***

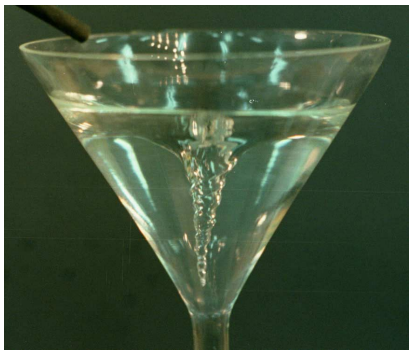


**Figure 23-25**
***Hollow core vortex in a funnel.***

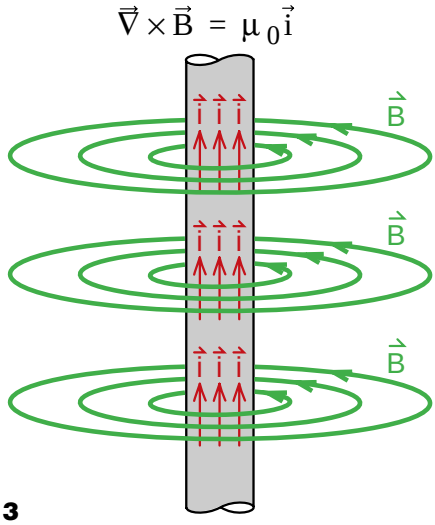



**Figure 3**
***A current in a straight wire produces a circular magnetic field around the wire.***

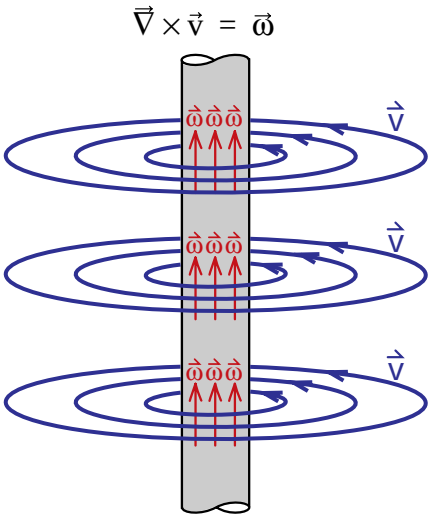



**Figure 5**
***Vorticity field $\vec{\omega}$ producing a circular velocity field.***

## SOLID BODY ROTATION

Enough of analogies, it is now time to actually calculate the vorticity field $\vec{\omega} = \vec{\nabla} \times \vec{v}$ of a flow pattern. Our example will be to calculate $\vec{\omega}$ when $\vec{v}$ is the velocity field of a solid rotating object.

As an explicit example, imagine that you are looking at the end of a rotating shaft shown in Figure (6). If the shaft has an angular velocity $\omega_{rot}$, so that

$$\frac{d\theta}{dt} = \omega_{rot} \tag{16}$$

then at a point p, out at a distance r from the axis of rotation, the velocity is in the $\hat{\theta}$ direction and given by the formula

$$\vec{v} = \hat{\theta} r \omega_{rot} \tag{17}$$

where the unit vectors $\hat{r}$, $\hat{\theta}$ and $\hat{z}$ are for a cylindrical coordinate system are shown in Figure (7).

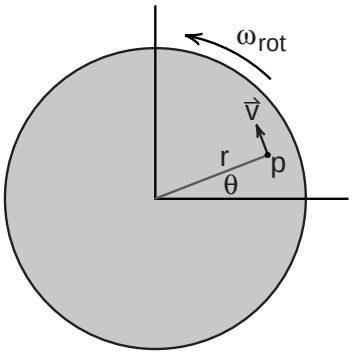



**Figure 6**

***End of a shaft rotating with an angular velocity $\omega_{rot}$.***

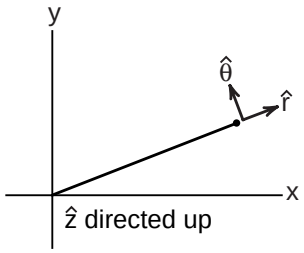



**Figure 7**

***Unit vectors for a cylindrical coordinate system.***

In Chapter 8 of the Calculus text, we wrote down the formula for the curl in cylindrical coordinates. (It can also be found in the Formulary at the end of this text.) Applied to the velocity field $\vec{v}$, given by

$$\vec{v} = \hat{r} v_r + \hat{\theta} v_\theta + \hat{z} v_z \tag{18}$$

the result is

$$(\vec{\nabla} \times \vec{v})_r = \frac{1}{r}\frac{\partial v_z}{\partial \theta} - \frac{\partial v_\theta}{\partial z} \tag{19a}$$

$$(\vec{\nabla} \times \vec{v})_\theta = \frac{\partial v_r}{\partial z} - \frac{\partial v_z}{\partial r} \tag{19b}$$

$$(\vec{\nabla} \times \vec{v})_z = \frac{1}{r}\frac{\partial}{\partial r}(r v_\theta) - \frac{1}{r}\frac{\partial v_r}{\partial \theta} \tag{19c}$$

In our example of solid body rotation, $\vec{v}$ has only a $\hat{\theta}$ component, and this component $v_\theta(r)$ depends only upon the distance r out from the axis of rotation. Thus $v_r$, $v_z$, and $\partial v_\theta/\partial\theta$ and $\partial v_\theta/\partial z$ are all zero and we are left with only the term

$$(\vec{\nabla} \times \vec{v})_z = \frac{1}{r}\frac{\partial}{\partial r}(r v_\theta) \tag{20}$$

You can see that the use of cylindrical coordinates when we have cylindrical symmetry eliminates many terms in the formula for the curl.

---

**Exercise 1**

In the last section, we noted that the circular velocity field of a vortex had zero curl if the velocity drops off as 1/r. This corresponds to a velocity

$$v_\theta = \frac{\text{constant}}{r} \; ; \quad v_r = v_z = 0 \tag{21}$$

Use Equation (19) or (20) to show that $\vec{\nabla} \times \vec{v} = 0$ for this vortex velocity field.

---

For solid body rotation, we use $v_\theta = r\omega_{rot}$ to get

$$(\vec{\nabla}\times\vec{v}_{\text{solid body}})_z = \frac{1}{r}\frac{\partial}{\partial r}(rv_\theta)$$
$$= \frac{1}{r}\frac{\partial}{\partial r}(r^2\omega_{rot})$$
$$= \frac{1}{r}(2r\omega_{rot})$$

$$\boxed{(\vec{\nabla}\times\vec{v}_{\text{solid body}})_z = 2\omega_{rot}} \qquad (22)$$

Using our notation $\vec{\nabla}\times\vec{v}_{\text{solid body}} \equiv \vec{\omega}_{\text{solid body}}$, we get

$$(\vec{\omega}_{\text{solid body}})_z = 2\omega_{rot} \qquad (22a)$$

This is the example we mentioned earlier where the vorticity $\vec{\omega}$ has a magnitude of exactly twice the rotational velocity $\omega_{rot}$.

(It is a challenge to find an intuitive explanation for the factor of 2 difference between the vorticity $\vec{\omega} = \vec{\nabla}\times\vec{v}$ and the rotational velocity $\omega_{rot}$. The analogy is even closer, because when we turned $\omega_{rot}$ into the vector $\vec{\omega}_{rot}$ in our discussion of gyroscopes, $\vec{\omega}_{rot}$ pointed down the rotational axis just as $\vec{\omega} = \vec{\nabla}\times\vec{v}$ does. I have not met this challenge. After much thought, I have found no satisfactory *intuitive* explanation for the factor of 2. It came in when we differentiated $r^2$, but that is not good enough.)

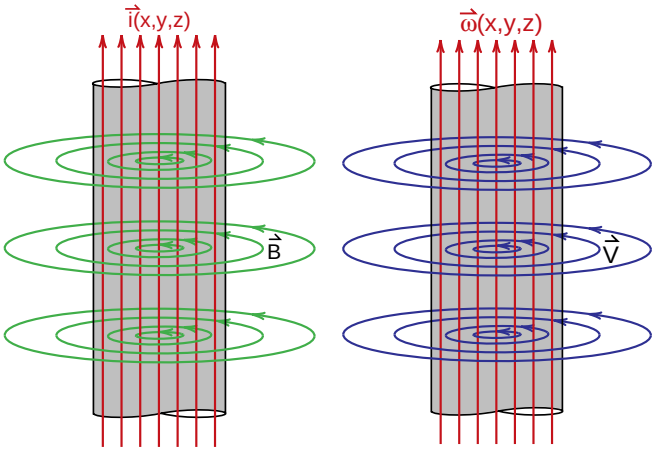



**Figure 8**
***Comparison of the magnetic field of a current in a wire with the velocity field of a fluid core vortex.***

The main result from our calculation of the curl for solid body rotation is that the ***curl points along the axis of rotation***, and has the ***constant magnitude 2 $\omega_{rot}$*** across the entire rotating surface.

## Vortex Core

With our results for the vorticity of solid body rotation, we can see an even closer analogy between the magnetic field of a wire and the vorticity field of a fluid core vortex. The corresponding formulas and field diagrams are shown again in Figure (8).

At the end of Calculus Chapter 8 we studied the magnetic field produced by a uniform current in a wire. We got as the formula for the field inside the wire

$$\vec{B}(r) = \hat{\theta}kr \qquad \begin{matrix}\textit{inside}\\ \textit{wire}\end{matrix} \qquad (8\text{-}66a)$$

where k was the collection of constants given by

$$k = \frac{\mu_0 i_{total}}{2\pi R^2} \qquad (8\text{-}66b)$$

**Exercise 2**

Show that $\vec{B}$ in Equation (8-66) above obeys the relationship $\vec{\nabla}\times\vec{B} = \mu_0 i$ .

The magnetic field in Equation (8-66) has the same form as the velocity field for solid body rotation, $v_\theta = r\omega_{rot}$ or

$$\vec{v}_{\text{solid body rotation}} = \hat{\theta}(\omega_{rot})r \qquad (23)$$

Thus there will be a complete analogy between the magnetic field of a wire, and a fluid core vortex, if the wire carries a uniform current density i and the vortex core consists of fluid undergoing solid body rotation. In the magnetic field case, the source of the magnetic field is the uniform current in the wire. For the fluid core vortex, the source of the velocity field is the uniform vorticity in the solid body rotating core. Outside the wire and outside the core, both the magnetic field and the velocity field are $\hat{\theta}$ directed and drop off as 1/r, a field pattern that has zero curl.

## STOKES' LAW REVISITED

For quite a while now we have seen that there are basically two kinds of vector fields. There is what we can call the divergent kind like the electric field of stationary charges that has zero curl. And then there is the rotational kind like the magnetic field and the velocity field of a constant density fluid that has zero divergence. Just as ***Gauss' law*** played an important role in determining the behavior of divergent fields, we will see that ***Stokes' law*** has an equally important role in determining the shape and behavior of the rotational kind of vector field. In this section we will take a closer look at Stokes' law, giving it a more physical interpretation than you will find in the mathematics textbooks.

We introduced Stokes' law in Chapter 8 of this text, writing it essentially in the form

$$\oint_C \vec{v}\cdot d\vec{\ell} = \int_S (\vec{\nabla}\times\vec{v})\cdot dA \quad \textit{Stokes' law} \qquad (8\text{-}14)$$

where $\vec{v}$ is a vector field, C is some closed contour, and S is the surface bounded by the contour C. We asked you to picture the contour C as being made up of a wire loop, and S the surface of a soap film stretched across the loop. The point was that if you gently blow on a soap film, it can take on various shapes, and Stokes' law applies no matter which shape you consider.

**Figure 8-2 (repeated)**
***Example of a surface bounded by a closed path (wire loop).***

### Total Circulation and Density of Circulation

Because we are going to make extensive use of Stokes' law, we will give special names to the terms in the law. The names are chosen to particularly apply to a velocity field, but can be used in general. First, we will call the line integral of $\vec{v}$ around a closed path the ***total circulation*** for the path.

$$\boxed{\text{total circulation} \equiv \oint_C \vec{v}\cdot d\vec{\ell}} \qquad (24)$$

In addition, we will refer to the vorticity $\vec{\nabla}\times\vec{v}$ as the ***density of circulation***

$$\boxed{\text{density of circulation} \equiv \vec{\nabla}\times\vec{v}} \qquad (25)$$

Then Stokes' law

$$\oint_C \vec{v}\cdot d\vec{\ell} = \int_S (\vec{\nabla}\times\vec{v})\cdot dA$$

can be stated in words that the ***total circulation of the fluid around a closed path C is equal to the density of circulation integrated over any surface bounded by the path***.

We are using the same terminology one would use in describing a current in a wire. You would say that the total current carried by a wire is equal to the current density integrated over some cross-sectional area of the wire. Why we have introduced this terminology for the velocity field will become clear as we discuss a few examples.

## Velocity Field of a Rotating Shaft, Again

As our first example, let us apply Stokes' law to the velocity field of a rotating shaft, shown in Figure (6) repeated here. Over the area of the end of the shaft we have solid body rotation where the velocity field is $\hat{\theta}$ directed

$$\vec{v} = \hat{\theta} r \omega_{rot} \qquad (17) \text{ repeated}$$

and the vorticity $\vec{\omega} \equiv \vec{\nabla} \times \vec{v}$ is directed up the axis of the shaft and of magnitude $2\omega_{rot}$

$$\vec{\omega} = \vec{\nabla} \times \vec{v} = \hat{z}\, 2\omega_{rot} \qquad (22) \text{ repeated}$$

To apply Stokes' theorem, let the circuit C be the circuit of radius R around the perimeter of the shaft. We then get

$$\oint_C \vec{v} \cdot d\vec{\ell} = \oint v_\theta (d\ell)_\theta \qquad (26)$$

At the perimeter, $v_\theta = R\omega_{rot}$, and $(d\ell)_\theta = R\,d\theta$, to give

$$\oint \vec{v} \cdot d\vec{\ell} = \int_0^{2\pi} (R\omega_{rot})(R\,d\theta)$$

$$= R^2 \omega_{rot} \int_0^{2\pi} d\theta = 2\pi R^2 \omega_{rot}$$

Thus the total circulation of the shaft is given by

$$\boxed{\text{total circulation of the shaft} = \pi R^2 (2\omega_{rot})} \qquad (27)$$

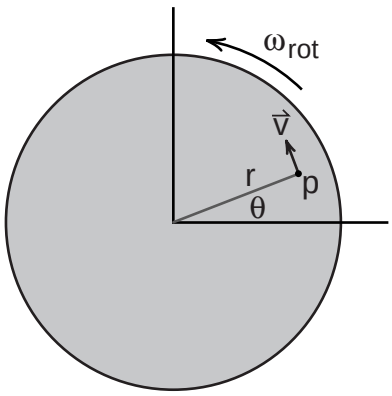



**Figure 6 (repeated)**
***End of a shaft rotating with an angular velocity $\omega_{rot}$.***

Stokes' theorem states that this total circulation should be equal to the density of circulation $\vec{\nabla} \times \vec{v}$ integrated over the area of the shaft. We know that for solid body rotation

$$\text{density of circulation} = \vec{\nabla} \times \vec{v} = \vec{\omega} = \hat{z}\, 2\omega_{rot} \qquad (28)$$

This density, of magnitude $\omega_z = 2\omega_{rot}$, is constant over the area of the shaft, thus the integral of the density is simply

$$\int_S (\vec{\nabla} \times \vec{v}) \cdot d\vec{A} = \int_S \omega_z dA_z$$

$$= \omega_z \int_S dA_z = \omega_z \pi R^2$$

$$\boxed{= \pi R^2 (2\omega_{rot})} \qquad (29)$$

Comparing Equations (27) and (29), we see that the total circulation is, as expected, equal to the density integrated over the area of the shaft.

## Wheel on Fixed Axle

Before you think everything is too obvious, let us consider a more challenging example. Suppose we have a wheel of radius R, rotating on a fixed axle of radius $R_{axle}$, as shown in Figure (9). The velocity field for this example is

$$\begin{aligned} \vec{v} &= 0 && r < R_{axle} \\ \vec{v} &= \hat{\theta} r \omega_{rot} && R_{axle} < r < R \end{aligned} \qquad (30)$$

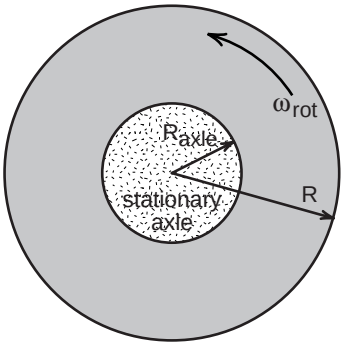



**Figure 9**
***Wheel rotating on a stationary axle.***

To apply Stokes' law again, let C be a circuit of radius R about the perimeter of the wheel. The total circulation is the same as before, namely

$$\begin{aligned}\text{total circulation} &= \int_C \vec{v}\cdot d\vec{\ell} = (R\omega_{rot})(2\pi R) \\ &= \pi R^2(2\omega_{rot})\end{aligned} \qquad (31)$$

When we measure the total circulation around the wheel, the result is uniquely determined by the value of $\vec{v}$ out at the circuit C. It makes no difference whatever whether the axle inside is turning or not.

But when we integrate the density of circulation $\vec{\nabla}\times\vec{v}$ over the area of the wheel, we have a problem. Over the wheel $\vec{\nabla}\times\vec{v} = \hat{z}\,2\omega_{rot}$ as before, but $\vec{\nabla}\times\vec{v} = 0$ over the axle. It appears that we have lost an amount of circulation $(2\omega_{rot})(\pi R_{axle}^2)$, and that Stokes' law fails.

Mathematics textbooks would say that we did not apply Stokes' law correctly. You will find statements like "*Stokes' law applies only to singly connected surfaces*" or "*you have to add a cut*". ***Don't believe it!*** Stokes' law applies quite generally, and you do not need so called cuts. What went wrong in this example is not Stokes' law, it is that we did not look carefully enough.

Suppose Figure (9) represented the wheel on a railroad car. Look carefully at the boundary between the wheel and the axle and what do you find? ***Roller bearings!*** As the wheel rotates on the axle, the roller bearings really spin. The circulation that we lost in the axle is now located in the roller bearings, and in the velocity field of the oil lubricating the bearings.

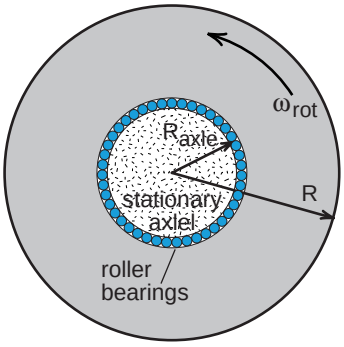



**Figure 9a**
***Wheel with roller bearings rotating on a stationary axle.***

You might be a bit worried about this explanation. After all, a fixed amount of circulation, namely $(2\omega_{rot})(\pi R_{axle}^2)$ was lost when we stopped the axle from rotating. But the space where the roller bearings reside, between the axle and the wheel can be made as thin as we want, reducing the area of the bearings that we integrate $\vec{\nabla}\times\vec{v}$ over. If we make the area of the bearings go to zero, can we still get a finite amount of circulation $(2\omega_{rot})(\pi R_{axle}^2)$ when we integrate over this vanishing area?

The answer is yes. Look what happens to roller bearings as we make the diameter of the bearings smaller and smaller. They have to spin faster and faster so that they roll smoothly between the axle and the wheel. As we decrease the thickness of the bearings, we increase the vorticity $\vec{\nabla}\times\vec{v}$ in the bearings in just such a way that the integral of $\vec{\nabla}\times\vec{v}$ over the bearings remains constant. In the mathematical limit that the thickness of the bearings goes to zero, we end up with a delta function of vorticity spread around the perimeter of the axle. This delta function of vorticity is called a ***vortex sheet***. When you correctly account for vortex sheets, you can always make sense of Stokes' law without caveats relating to singly connected surfaces or cuts.

## A Conservation Law for Vorticity

Imagine that our solid shaft of Figure (6) represented a wheel and axle where the axle was rotating with the wheel. Then the axle would have vorticity of magnitude $2\omega_{rot}$ just like the wheel. Now suppose we grab hold of the axle to stop it from rotating, giving us the velocity field shown in Figure (9). By stopping the axle from rotating, we did not destroy the vorticity, we just moved it out to the roller bearings or vortex sheet. ***For a given total circulation around the rim of the wheel, we cannot create or destroy vorticity within, only move it around.*** With a given total circulation, we have a conserved amount of vorticity within. In this sense, Stokes' law provides us with a conservation law for vorticity. (In Appendix 2 of Chapter 13, we show you a more general, three dimensional law for the conservation of vorticity.)

## CIRCULATION OF A VORTEX

In an ideal straight vortex like the one we pictured in Figure (8) more or less redrawn here as Figure (10), the vorticity is concentrated in the core and we have a curl free 1/r velocity field outside the core. It is traditional to use the Greek letter $\kappa$ (kappa) to designate the total circulation of the vortex.

$$\oint_{\text{over any area that includes the vortex core}} \vec{v} \cdot d\vec{\ell} = \kappa \qquad \textit{total circulation or strength of a vortex} \qquad (32)$$

Evaluating the integral around a circle outside the core gives

$$\oint \vec{v} \cdot d\vec{\ell} = 2\pi r v_\theta = \kappa$$

$$\boxed{v_\theta = \frac{\kappa}{2\pi r}} \qquad \textit{velocity field of a straight vortex} \qquad (33)$$

This is the formula for the velocity field of a straight vortex, outside the core. For shorthand, we sometimes use $\bar{\kappa} = \kappa/2\pi$ just as we used $\hbar = h/2\pi$ in quantum mechanics, giving

$$\boxed{v_\theta = \frac{\bar{\kappa}}{r}} \qquad \textit{velocity field of a straight vortex} \qquad (33a)$$

Note that talking about the total circulation $\kappa$ of a vortex, we know that when there is cylindrical symmetry, the velocity field $v_\theta$ outside the core is $\kappa/r$ independent of the structure of the core. The core can be a fluid core with solid body rotating fluid inside, or be a hollow core vortex like the funnel vortex of Figure (23-25). With a solid body rotating core the vorticity $\vec{\omega}$ is spread uniformly across the core. With a hollow core vortex, we can think of the vorticity as being in a vortex sheet around the core.

We have a similar situation for the magnetic field of a straight wire. In a normal wire, there is a more or less uniform current density in the wire which produces a magnetic field of strength $B_\theta = \mu_0 I_{total}/2\pi r$ outside. In some superconducting wires, those made from the so called type 1 superconductors like lead and tin, the electric current flows very near the surface of the wire with no current farther inside. This surface sheet of current still produces the same magnetic field $B_\theta = \mu_0 I_{total}/2\pi r$ outside.

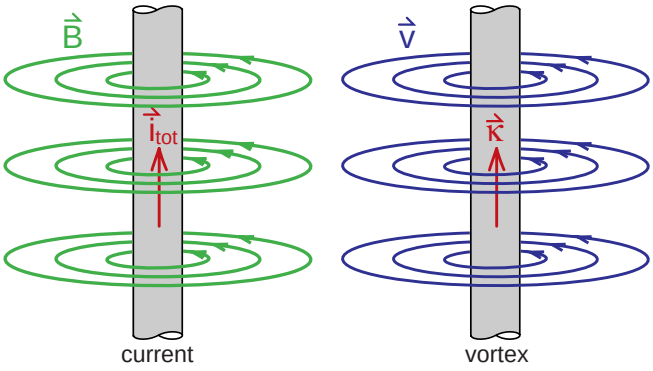



**Figure 10**
***The total circulation $\bar{\kappa}$ of the vortex is related to the velocity field $\vec{v}$ the same way the total current $\vec{i}_{tot}$ is related to the magnetic field $\vec{B}$. (For straight vortices, we often think of $\bar{\kappa}$ as a vector pointing in the direction of $\vec{\omega}$, as shown above.)***

## QUANTUM VORTICES

We are now ready to deal with the failure of Landau's prediction that superfluid helium could only undergo potential flow, with the consequence that helium in a bucket could not rotate. The appearance of a parabolic surface on a rotating bucket of superfluid helium is experimental evidence that vorticity is present in the fluid despite Landau's prediction.

Feynman solved the problem by proposing that most of the fluid in a rotating bucket of superfluid helium was in fact undergoing potential flow, and that all the vorticity that was responsible for the curved surface was contained in little quantized vortices.

As we have mentioned in the Physics text, a single quantized vortex can be pictured as a giant Bohr atom where all the superfluid atoms taking part in the vortex flow have one unit of angular momentum $\hbar$ about the vortex core. The angular momentum of an atom out at a distance r from the core, moving at a speed $v_\theta$, is

$$L_{\substack{\text{angular}\\\text{momentum}}} = m_{He} v_\theta r \qquad (34)$$

where $m_{He}$ is the mass of a helium atom. If we set the angular momentum L equal to Planck's constant $\hbar$, and solve for $v_\theta$, we get

$$L = \hbar = m_{He} v_\theta r$$

$$\boxed{v_\theta = \frac{\hbar}{m_{He} r}} \qquad (35)$$

We immediately see that the velocity field outside the core drops off as 1/r which is potential flow.

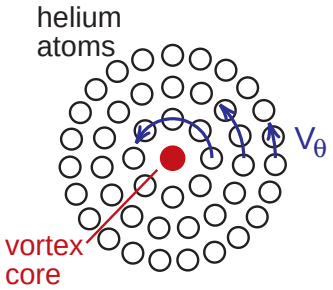



**Figure 11**
***Each atom in a quantum vortex has one unit of angular momentum about the vortex core.***

The 1/r velocity field cannot continue in to r = 0; there has to be a core that is not potential flow. There are two questions that need to be settled by experiment. One is how big is the core radius $r_{core}$, and the second is whether the core is hollow, or filled with rotating fluid. The answer to the first question is rather amazing. Under most circumstances the core is about as small as it can get, ***about one atomic diameter***. That makes it difficult to answer the second question; it is hard to tell what is inside a tube only one atomic diameter across.

### Circulation of a Quantum Vortex

One thing we can do immediately from Equation (35) is to calculate the total circulation $\kappa$ of a quantum vortex. Remembering that $\hbar = h/2\pi$ we have

$$v_\theta = \frac{\hbar}{m_{He} r} = \frac{h}{(2\pi r) m_{He}}$$

$$(2\pi r) v_\theta = \frac{h}{m_{He}}$$

But $2\pi r v_\theta$ is simply the integral of $v_\theta$ around a circle centered on the core. Thus we have

$$\oint \vec{v} \cdot d\vec{\ell} = 2\pi r v_\theta = \kappa$$

$$\boxed{\kappa = \frac{h}{m_{He}}} \qquad \textit{circulation of a quantum vortex in superfluid helium} \qquad (36)$$

## Rotating Bucket of Superfluid Helium

If you have a rotating bucket of normal fluid, the fluid will end up rotating with solid body rotation with constant vorticity $\vec{\omega} = \hat{z}2\omega_{rot}$. The total circulation $\kappa_{total}$ of all the fluid in the bucket will be

$$\begin{matrix}\text{total circulation of}\\ \text{fluid in rotating}\\ \text{bucket}\end{matrix} = \int_{\substack{\text{bucket}\\ \text{surface}}} (\vec{\nabla}\times\vec{v})\cdot dA$$

$$\kappa_{total} = (2\omega_{rot})(\pi R^2{}_{bucket}) \qquad (37)$$

For solid body rotation, this vorticity is spread uniformly across the bucket.

Feynman proposed that a rotating bucket of superfluid helium would have the same total circulation $\kappa_{total}$, but that the vorticity, instead of being spread throughout the fluid, would be contained in a bundle of quantized vortex cores. This difference between the classical and quantum picture is indicated in Figure (12).

Because the core of a quantum vortex is so small, and because all the fluid between the cores is undergoing potential flow, you can see that Landau was almost right. But the quantum cores allow vorticity to be spread throughout the bucket, roughly imitating solid body rotation, and give rise to a nearly parabolic surface.

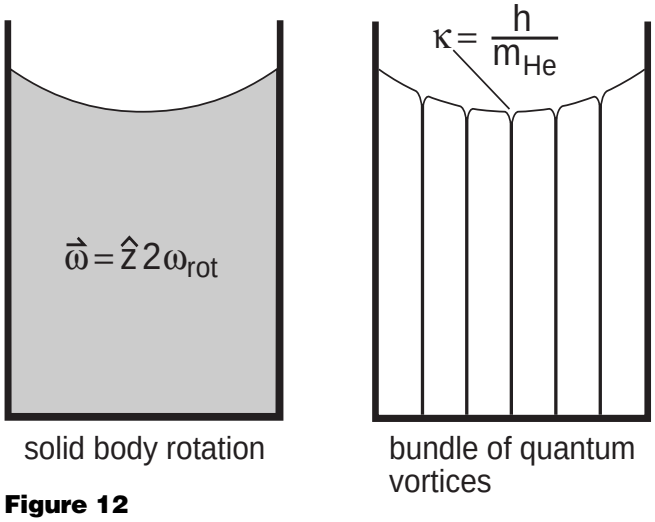



**Figure 12**
***Comparison of solid body rotation with a bundle of quantized vortices. (We have not tried to reproduce the exact shape of the surface when vortices are present.) Between the vortices the flow is potential, but the rough shape of the surface is parabolic.***

We can easily calculate the number of quantized vortices required to imitate solid body rotation. From Equation (37), we saw that the total circulation of the bucket was $\kappa_{total} = (2\omega_{rot})(\pi R^2{}_{bucket})$. Each quantum vortex supplies a circulation $h/m_{He}$. If we have N quantum vortices, their total circulation will be $Nh/m_{He}$. Equating these two numbers gives

$$\kappa_{total} = (2\omega_{rot})\pi R^2{}_{bucket} = N\frac{h}{m_{He}}$$

Solving for N, and then dividing by the area of the bucket, gives us the number n of quantized vortices per unit area.

$$\boxed{n = \frac{N}{\pi R^2{}_{bucket}} = \frac{2\omega_{rot}m_{He}}{h}} \qquad (38)$$

To see what the density is of quantized vortices needed to imitate solid body rotation, let us use CGS units where the unit area is $1\text{cm}^2$, and solve for an angular velocity $\omega_{rot}$ of one radian/second which is about 1/6 of a revolution per second. We have

$$\omega_{rot} = 1$$

$$m_{He} = 4\times 1.67\times 10^{-24}\text{gm} \quad (4 \text{ proton masses})$$

$$h = 6.62\times 10^{-27}$$

We get for the vortex density n

$$n = \frac{2\omega_{rot}m_{He}}{h}$$

$$= \frac{2\times 4\times 1.67\times 10^{-24}}{6.62\times 10^{-27}}$$

$$= 2020 \text{ lines/cm}^2$$

If these lines were in a rectangular array, there would be $\sqrt{n}$ lines on each side of a square centimeter

$$\sqrt{n} = 45 \text{ lines/cm}$$

The spacing between lines would be $1\sqrt{n}$

$$1\sqrt{n} = .022 \text{ cm/line}$$

$$= .22 \text{ millimeters/line} \qquad (39)$$

Thus to imitate solid body rotation with an array of quantized vortices in superfluid helium, the quantum vortices have to be .22 millimeters apart when the rotational velocity is 1 radian per second.

For a number of years after Feynman's explanation of the curved surface on a bucket of superfluid helium, there was a considerable effort to see if quantum vortices really exist in the superfluid. The most conclusive evidence for their existence, with the predicted circulation $\kappa = h/m_{He}$, came from experiments by Rayfield and Reif using charged vortex rings. A few years later Richard Packard at Berkeley succeeded in actually photographing the vortices in a rotating bucket of helium. He did this by loading up the vortex lines with electrons, and then firing the electrons into a film placed at the surface of the liquid. The result is shown in Figure (13) for various rotational speeds.

What Feynman and others have shown is that the flow pattern with quantized vortices is a wave pattern for the helium atoms in the bucket. It is the lowest energy solution of a wave equation, subject to the boundary condition that the atoms near the surface of the bucket are moving with a velocity nearly equal to the velocity of the bucket. Although we have used the terminology of classical fluid dynamics, we are describing a quantum mechanical phenomenon. What is remarkable is that we are seeing quantum mechanical phenomena on a large human scale, not just an atomic scale. You can see a separation of .22 millimeters without the use of a microscope.

---

**Exercise 3 - A Superfluid Gyroscope**

Counting vortices in a bucket of superfluid helium can be a sensitive way of detecting rotation. Suppose a bucket of helium were placed at the North Pole. How many vortices per $cm^2$ would there be in the bucket due to the rotation of the earth?

---

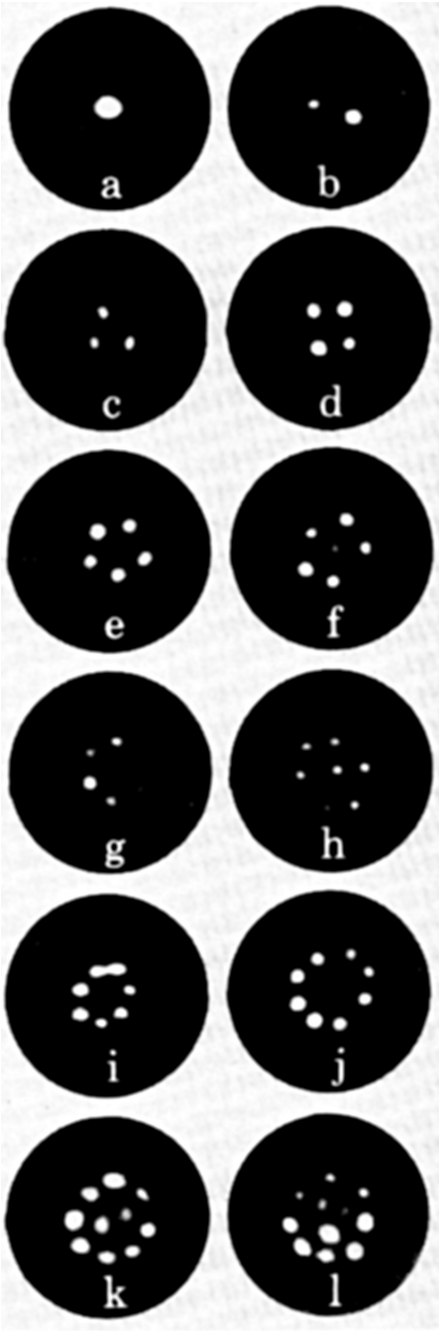



**Figure 13**
***Packard's photograph of vortex lines in rotating superfluid helium. As the rotational speed is increased, more quantum vortices appear. Angular velocities range up to half a radian per second. (The camera was rotated with the helium and many exposures were taken to build up the image. The slight jiggling of the vortices between exposures spread the vortex images out a bit.)***

## Bose-Einstein Condensates

Since 1995, it has been possible to create a new kind of superfluid, consisting of a small drop of gas cooled to temperatures in the range of a millionth of a kelvin. What happens to the gas atoms at these temperatures is that they can come together and "condense" into a single quantum mechanical wave pattern. The process is not unlike photons condensing into a single wave pattern in a laser beam. For the gas atoms the result is a liquid-like drop with superfluid properties.

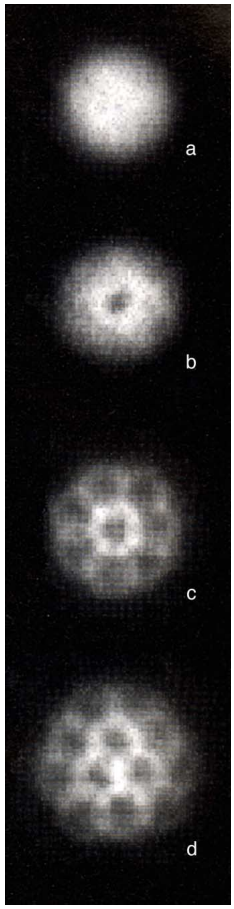



**Figure 14**

It is called Bose-Einstein condensation because back in the 1920's, Einstein predicted this effect, basing his ideas on the work of the Indian physicist Nath Bose. It turns out that atoms or objects that have integer spin like to congregate into a single quantum wave pattern if the temperature is low enough, i.e., if the pattern is not disturbed by thermal effects. Examples of integer or zero spin objects that do this are photons that form laser beams, Helium 4 atoms that form superfluid helium, and electron pairs that become a superconductor.

In 1999, a group at the École Normale Supérieure in Paris succeeded in rotating a drop of rubidium atoms and photographing the quantized vortices as they appeared. Due to the weak attraction between the rubidium atoms, the vortex cores are some 5000 times bigger than the core of a superfluid helium vortex, but have the same circulation $h/m_{atom}$. Photographs of the drop, with 0, 1, 8, and 13 vortices are seen in Figure (14). Figure (15) is a computer simulation of the vortex core structure of a drop with four vortices passing through the drop, and two forming at the edge.

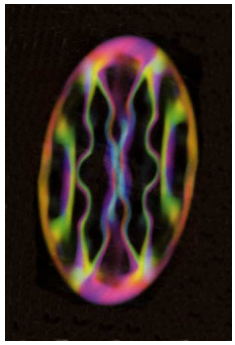


**Figure 15**

## THE VORTICITY FIELD

So far we have described vorticity as something we look for in a vortex core or something that characterizes solid body rotation. In this section we will treat the vorticity $\vec{\omega} = \vec{\nabla} \times \vec{v}$ as a dynamic field that has field lines and can behave much like the other vector fields we have been discussing.

The singular property of vorticity is that it always has identically zero divergence

$$\vec{\nabla} \cdot \vec{\omega} = \vec{\nabla} \cdot (\vec{\nabla} \times \vec{v}) \equiv 0 \qquad (40)$$

because the divergence of a curl is identically zero. (See the vector identities.) This means that vorticity is always a solenoidal field without sources or sinks.

We defined a field line of the velocity field as a small flow tube, like those seen in Figure (23-3) reproduced below. Similarly, we define a vortex line as a small flow tube of vorticity. The total flux of vorticity in the flow tube is by definition, the circulation $\kappa$ of that tube. As a reminder, this comes from Stokes' law

$$\begin{aligned} \text{flux of } \vec{\omega} \text{ in a vortex tube} &= \int_{\substack{\text{surface} \\ \text{across} \\ \text{tube}}} \vec{\omega} \cdot dA = \int_{S} (\vec{\nabla} \times \vec{v}) \cdot dA \\ &= \oint_{\substack{\text{around} \\ \text{tube}}} \vec{v} \cdot d\ell = \kappa_{tube} \end{aligned} \qquad (41)$$

Because the vorticity $\vec{\omega}$ is solenoidal, the flux tubes or lines of $\vec{\omega}$ cannot start or stop inside the fluid. Vortex lines can only start or stop on the fluid boundaries, or close on themselves within the fluid. Two examples are the straight vortices we have been discussing which run from the bottom of a container to the top , and a vortex ring where the vortex lines go around and close on themselves like the magnetic field lines around a wire. A smoke ring is the classic example of a vortex ring.

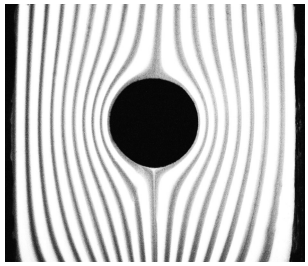


**Figure 23-3**
***Flow tubes bounded by streamlines. We define a field line as a small flow tube.***

## HELMHOLTZ THEOREM

In 1858 Heinrich Helmholtz discovered a remarkable theorem related to vortex motion. He discovered that ***when all the forces acting on fluid particles are conservative forces, i.e., force fields that have zero curl, vortex lines move with the fluid particles***. Gravity is an example of a conservative force, viscous forces are not. If viscosity can be neglected and only gravity is acting on the fluid, vortex lines and fluid particles move together.

To emphasize this point, in the absence of non conservative forces, we can say that the fluid particles become trapped on vortex lines, or we can say that vortex lines become stuck on and have to move with the fluid particles. ***To move vorticity onto or off a fluid particle requires a non conservative force like viscosity***.

### The Two Dimensional "Vortex Ring"

The simplest illustration of Helmholtz's theorem is the behavior of a vortex ring where the vortex lines go around a circle and close on themselves. The most well known example of a vortex ring is the smoke ring.

Before we discuss circular vortex rings, we will consider the simpler example of two oppositely oriented straight vortices which form what is often called a ***two dimensional (2D) vortex ring***. A view down upon the two vortices, showing their independent velocity fields, is shown in Figure (16). The total velocity field of these two vortices is the vector sum of the fields from each vortex.

Notice that the upper vortex has a forward velocity field at the lower vortex core. If Helmholtz's theorem is obeyed, then this upper velocity field must be moving the vortex lines in the lower core forward. Likewise the velocity field of the lower vortex must move the core of the upper vortex forward. As a result this two dimensional vortex configuration is a self propelled, forward moving object.

We can easily calculate the forward speed of our 2D vortex ring. The velocity field of a vortex of circulation $\kappa$ was given by Equation (33a) as

$$v_\theta = \frac{\bar{\kappa}}{r} \; ; \quad \bar{\kappa} = \frac{\kappa}{2\pi} \qquad \text{(33) repeated}$$

If the separation of the vortices is d, then the speed of the fluid at the opposite core, and therefore the speed of the ring will be

$$\boxed{v_{2d\ ring} = \frac{\bar{\kappa}}{d}} \qquad \begin{array}{l}\textit{speed of a pair}\\ \textit{of oppositely}\\ \textit{oriented vortices}\end{array} \qquad (40)$$

You can see that the ring moves faster (1) if the circulation $\bar{\kappa}$ is increased, or (2) if the vortices are closer together.

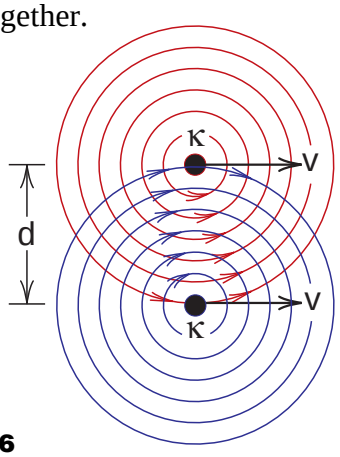



**Figure 16**
***Velocity fields of two oppositely oriented straight vortices.***

## The Circular Vortex Ring

For a circular, or 3D vortex ring, the vortex core has the shape of a doughnut. If we look at the velocity field in a plane that slices through the doughnut, as shown in Figure (17), the result is in many ways similar to the velocity field of the 2D vortex in Figure (16). In particular the velocity field of the top part of the ring moves the bottom part of the ring forward, while the field of the bottom of the ring moves the top forward.

In addition, the smaller the ring, the faster it moves. If the ring has a circulation $\kappa$ and diameter d, the speed of the ring is approximately given by the same equation $v_{ring} = \bar{\kappa}/d$ that applied to the 2D vortex.

The actual velocity field of a vortex ring has the same shape as the magnetic field of a circular current loop, (provided the current density in the wire has the same shape as the vorticity in the vortex core). It is a classic and rather nasty problem to calculate the precise shape of this field. When we get a more accurate answer for the speed of the ring, we end up with additional terms, one of which involves the logarithm of the core radius. This logarithm would go to infinity if we tried to make the core radius zero, but the term becomes small for reasonable core radii. We do not need to worry about these small additional terms now. The analogy to the behavior of the two dimensional ring is good enough.

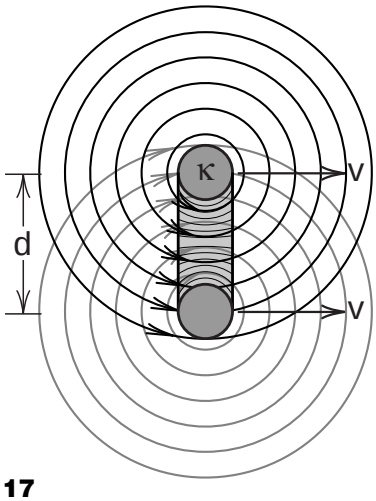



**Figure 17**
***Velocity field in a slice through a vortex ring.***

## Smoke Rings

In several ways the smoke ring provides a superb illustration of Helmholtz's theorem. In the days when smoking was popular and thought to be harmless, it was a common stunt to blow a smoke ring. Today we would rather create smoke rings using the apparatus shown in Figure (18). The apparatus is simple, and the rings are better.

Start with a cardboard box, cut a fairly large hole in the front as shown, and replace the back side with a rubber sheet. Fill the box with smoke, and hit the rubber sheet with your hand. A beautiful ring will emerge, like the one shown in Figure (19).

(If titanium tetrachloride solution available, you can get a denser smoke ring by squirting this liquid around the perimeter of the hole in the box. The titanium tetrachloride quickly turns to titanium dioxide smoke and hydrochloric acid. The titanium dioxide is a coloring agent for white paint, and the hydrochloric acid is obnoxious to deal with, but the resulting rings are quite good.)

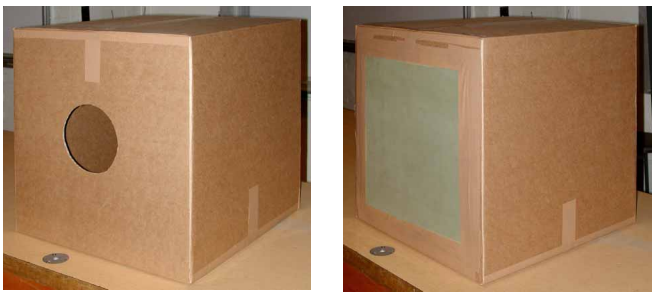


**Figure 18 a,b**
***Front and back of apparatus for creating smoke rings.***

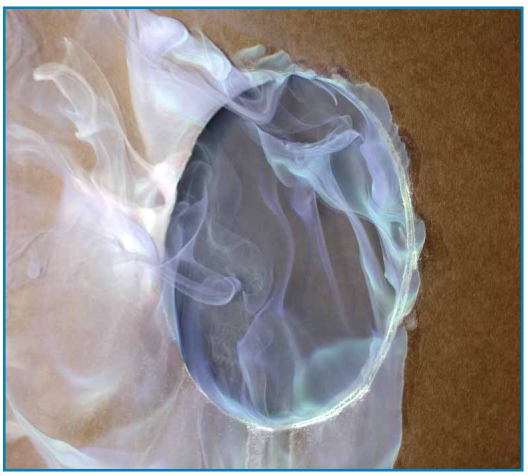


**Figure 18 c**
***Smoke at hole due to titanium tetrachloride.***

The most impressive feature of the smoke rings created by our box is how stable they are. They move in a straight line, at constant speed, without changing their shape, just as predicted by our analysis of the two and three dimensional vortex rings. If you hit the rubber sheet harder, you add more circulation $\kappa$ to the rings, and they travel faster. You can experiment with different size holes in the box, seeing that smaller rings travel faster than larger ones.

One of the interesting predictions that you can think about and try to observe is the following. If a faster ring approaches a slower one in front of it, the velocity field of the front ring will tend to make the back ring smaller and thus move still faster. Conversely, the velocity fields of the back ring should expand the front ring making it move more slowly. (Sketch the velocity fields yourself to check this prediction.) As a result, if the back ring is aimed right at the front one, the smaller back ring should shoot through the larger front ring, becoming itself the front ring. If the rings have not bumped into each other, tangled and destroyed themselves (the usual case), then the new back ring will be squeezed in size, the front ring expanded, and the process repeated. This is a famous prediction, but I have not seen it carried out very well.

While the motion of a smoke ring represents a successful prediction of Helmholtz's theorem, the fact that the smoke ring is so sharply defined, escaping from the amorphous cloud of smoke around the cardboard box, is an even more dramatic prediction of the theorem. When we hit the back of the box to create the ring, air was expelled out through the hole in the front. The vortex ring was created at the perimeter of the hole from air that contained smoke particles. ***These smoke particles in the vortex core become attached to the vortex lines in the core*** and have to move with the core. As the vortex ring moves out of the box, it carries the trapped smoke particles in its core and leaves the rest of the smoke behind.

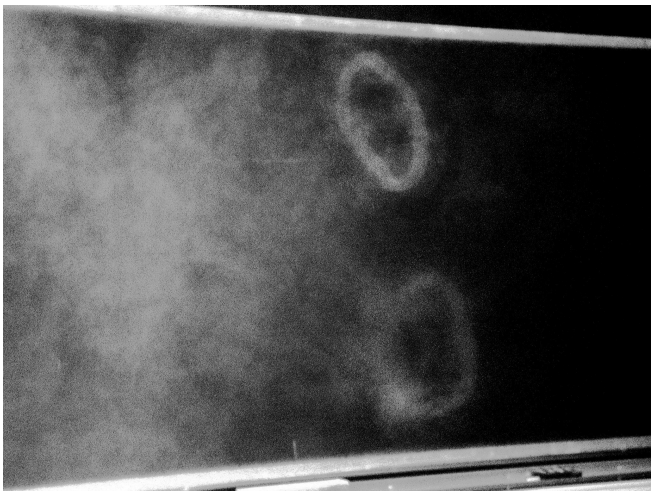


**Figure 19**
***Two smoke rings after they have collided.***

## Creating the Smoke Ring

The reason why is as follows. Before we hit the rubber sheet at the back of the box, all the air in the box was at rest and contained no vorticity. If Helmholtz's theorem strictly applied, then a vortex line could not move onto fluid particle that initially had no vorticity.

As we mentioned earlier, Helmholtz's theorem applied ***if only conservative forces*** (like gravity) were acting on the fluid. But gravity is not the only force acting on the particles of air in our smoke ring apparatus. Air is a slightly viscous fluid, and viscous forces in a fluid are not curl free conservative forces. Viscous forces move a vortex line onto fluid particle and create a vortex core.

# Calculus 2000-Chapter 13

# Introduction to Fluid Dynamics

*One should think of this chapter as an introduction to fluid dynamics. In it we derive the basic equations for the behavior of the velocity field* $\vec{v}$ *and the vorticity field* $\vec{\omega}$ *in a constant density fluid. We begin by applying Newton's second law to a fluid particle to obtain what is known as the Navier-Stokes equation. This equation for the velocity field* $\vec{v}$ *serves as the fundamental equation of fluid dynamics.*

*Taking the curl of the Navier-Stokes equation gives us the basic equation for the dynamics of the vorticity field* $\vec{\omega}$*. From that equation we derive the Helmholtz theorem, and an extension of the Helmholtz theorem that deals with the effect of non potential forces acting on fluid cores. The extended Helmholtz theorem is used in the analysis of the experiments of Rayfield and Reif who first measured the circulation* $\kappa$ *and core radius (a) of a quantized vortex in superfluid helium. We end the regular part of the chapter with a discussion of the* ***Magnus effect*** *and the pseudo force called the* ***Magnus force*** *that appears in all the vortex dynamics literature.*

*There are two major appendices to this chapter. Appendix 1 deals with the use of component notation in vector equations. This includes the Einstein summation convention, and emphasizes the use of the permutation tensor* $\varepsilon_{ijk}$ *for calculating vector cross products. There we show you an easy way to derive vector identities involving cross products.*

*The second appendix shows how you can interpret the dynamical behavior of the vorticity field as a conserved two dimensional flow of vorticity. Appendix 2 begins with an intuitive derivation of that result, a derivation that requires little mathematical background. (It can be explained at dinner parties.) However deriving the formula for the conserved vortex current requires the use of the permutation tensor* $\varepsilon_{ijk}$*, which is why we delayed this discussion until after Appendix 1.*

*The use of vortex currents turns out to be a particularly effective way to handle vortex motion. We use it, for example, to derive the Magnus force equation for curved fluid core vortices, a result that has not been obtained any other way.*

# THE NAVIER-STOKES EQUATION

When we apply Newton's second law $\vec{F} = d\vec{p}/dt$ to a particle like a baseball, the analysis is fairly simple. With $\vec{p} = m\vec{v}$ for the baseball, if m is constant, the result is $\vec{F} = m d\vec{v}/dt$. In particular, if $\vec{v} = \text{constant}$, then $d\vec{v}/dt = 0$ and $\vec{F} = 0$.

Applying Newton's second law to a fluid is more complicated. Even if we have a steady flow where $\vec{v} = \text{constant}$, the fluid particles themselves will be accelerating when the streamlines go around a corner or the flow tubes become narrower or wider. Some net force acting on the fluid particles is required to produce this acceleration. If the flow is not steady, if $\partial\vec{v}/\partial t$ is not zero, an additional force is required to produce this change in the velocity field. The first problem you encounter in the study of fluid mechanics is to correctly evaluate the acceleration of the fluid particles taking both of these effects into account.

What we will do is to consider a volume V of fluid bounded by a closed surface S′. The surface S′ is special in that it moves with the fluid particles. As a result the same fluid particles remain inside V as the fluid moves about. We will then calculate the rate of change of the total momentum of these fluid particles and equate that to the total force acting on the particles within V. Following this procedure we will end up with a differential equation called the ***Navier-Stokes*** equation which is very successful in describing the behavior of fluids.

(*In most textbooks you will find what looks to be a simpler derivation of the Navier-Stokes equation. Our derivation involves volume and surface integrals, while the textbooks make what looks like simpler arguments using what is called a* ***substantive derivative****. When the textbook arguments are applied to non constant density fluids, you also find some talk about what should be included inside the substantive derivative and what should not. It almost seems that one includes only those terms that give the right answer.*

*By using surface and volume integrals, our focus remains on the application of Newton's second law to the fluid particles with no ambiguities of interpretation.*)

## Rate of Change of Momentum

As we mentioned, we will consider a volume V of fluid whose surface S′ moves with the fluid particles. As a result the same particles remain inside the volume V. We then equate the rate of change of the total momentum of these particles to the total force acting on them. The main problem involves calculating the rate of change of the momentum of the particles in a volume whose surface is moving.

Suppose we have a volume V(t) that is now, at time (t), bounded by a surface S′(t) ( shown in Figure 1). If the fluid has a density $\rho$ and the velocity field of the fluid is $\vec{v}$ then the total momentum $\vec{P}_V(t)$ of the fluid in V(t) is

$$\vec{P}_V(t) = \int_{V(t)} \vec{p}(\vec{x},t)\,d^3V \quad ; \quad \vec{p} = \rho\vec{v} \qquad (1)$$

At this point we are even allowing the density to vary, so that both $\rho$ and $\vec{v}$ can be functions of space and time.

A short time $\delta t$ later, the surface will have moved to $S'(t+\delta t)$ and the volume becomes $V(t+\delta t)$ as shown in Figure (2).

At this later time, the momentum of the fluid particles will be

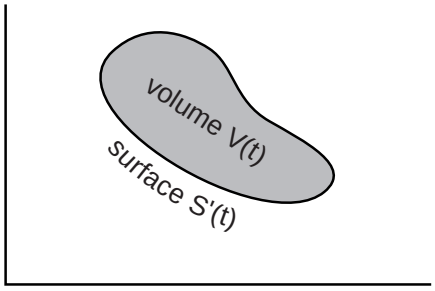



**Figure 1**
***The volume V bounded by the surface S′ at time (t).***

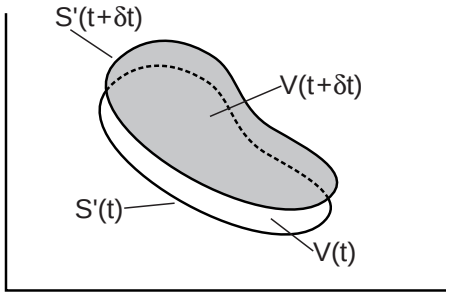



**Figure 2**
***The volume V a short time δt later.***

$$\vec{P}_V(t+\delta t) = \int_{V(t+\delta t)} \vec{p}(t+\delta t)d^3V \qquad (2)$$

The change $\delta\vec{P}_V$ in momentum of the fluid particles as time goes from (t) to $(t+\delta t)$ is

$$\begin{aligned}\delta\vec{P}_V &= \vec{P}_V(t+\delta t) - \vec{P}_V(t) \\ &= \int_{V(t+\delta t)} \vec{p}(t+\delta t)d^3V - \int_{V(t)} \vec{p}(t)d^3V\end{aligned} \qquad (3)$$

We can do a Taylor series expansion of $\vec{p}(t+\delta t)$ to get

$$\vec{p}(t+\delta t) = \vec{p}(t) + \frac{\partial\vec{p}}{\partial t}\delta t + 0(\delta t^2) \qquad (4)$$

This gives

$$\begin{aligned}\delta\vec{P}_V &= \left[\int_{V(t+\delta t)} \vec{p}(t)d^3V - \int_{V(t)} \vec{p}(t)d^3V\right] \\ &\quad + \delta t\int_{V(t+\delta t)} \frac{\partial\vec{p}}{\partial t}d^3V + 0(\delta t^2)\end{aligned} \qquad (5)$$

From Figure (2), we see that much of the same volume is included in both $V(t+\delta t)$ and V(t). Thus, in the square brackets in Equation (5), the integral of $\vec{p}(t)$ over the common volume cancels, and what we want is an integral of $\vec{p}(t)$ over the volume that the fluid has entered during the time $\delta t$, minus the integral of $\vec{p}(t)$ over the volume the fluid has left during $\delta t$.

In Figure (3a) we show part of the region between $S'(t)$ and $S'(t+\delta t)$ where the fluid has entered during $\delta t$. Consider a particle at point (1) at time t, moving at a velocity $\vec{v}_1$. In the short time $\delta t$ it moves a distance $\vec{v}_1\delta t$ as shown.

Now let $d\vec{A}_1$ be an element of the surface $S'(t)$ at point (1). The standard convention is that a surface element $d\vec{A}$ points perpendicularly ***out*** of a closed surface. Thus $d\vec{A}_1$ points out of surface $S'(t)$ as shown.

A time $\delta t$ later, the surface element $d\vec{A}_1$ will have moved out to the surface $S'(t+\delta t)$, sweeping out a volume $\delta V_1$ given by

$$\delta V_1 = (\vec{v}_1\delta t)\cdot d\vec{A}_1 \qquad (6)$$

You can see that the dot product is appropriate, for if $\vec{v}_1$ and $d\vec{A}_1$ are parallel, we have a right circular cylinder of volume $(v_1\delta t\, dA_1)$. The volume is zero if $\vec{v}_1$ and $d\vec{A}_1$ are perpendicular, and negative if oppositely oriented.

In Figure (3b) we show part of the region between $S'(t)$ and $S'(t+\delta t)$ where the fluid in $S'(t)$ has left during the time $\delta t$. The diagram is the same as Figure (3a) except that the vector $d\vec{A}_2$ pointing out of $S'(t)$ is pointing essentially opposite to the vector $\vec{v}_2$. In the formula $\delta V_2 = (\vec{v}_2\delta t)\cdot d\vec{A}_2$, the dot product $\vec{v}_2\cdot d\vec{A}_2$ and therefore $\delta V_2$ is negative in the region where the fluid is leaving.

As a result, if we calculate the integral of $\vec{p}(t)\delta V$ over both the volumes in Figures (3a) and (3b), we get an integral of $\vec{p}(t)$ over the region the fluid is entering, minus the integral of $\vec{p}(t)$ over the region the fluid is leaving. This just gives us the quantity in the square brackets in Equation (5)

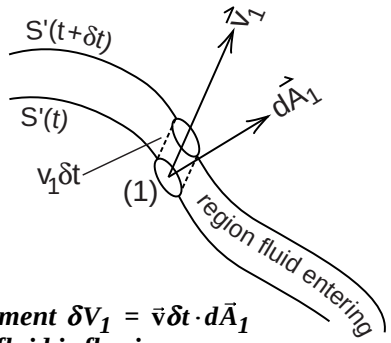



**Figure 3a**
***The volume element $\delta V_1 = \vec{v}\delta t\cdot d\vec{A}_1$ into which the fluid is flowing.***

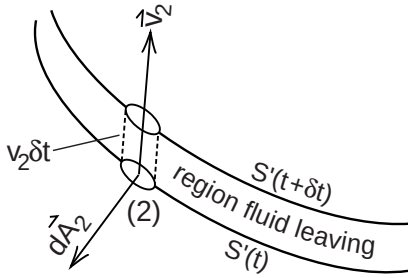



**Figure 3b**
***The volume element $\delta V_2 = \vec{v}\delta t\cdot d\vec{A}_2$ out of which the fluid is flowing.***

We get

$$\left[ \int_{V(t+\delta t)} \vec{p}(t) d^3V - \int_{V(t)} \vec{p}(t) d^3V \right]$$

$$= \int_{\substack{\text{over entering} \\ \text{and leaving} \\ \text{regions}}} \vec{p}(t)(\delta V) = \int_{S'(t)} \vec{p}(t)(\delta t \vec{v} \cdot d\vec{A}) \qquad (6)$$

By integrating over the entire area $S'(t)$ we have included both the entering and leaving regions.

Using Equation (6) for the square brackets in Equation (5) gives

$$\delta \vec{P}_V = \delta t \int_{S'(t)} \vec{p}(t)(\vec{v} \cdot d\vec{A}) + \delta t \int_{V(t+\delta t)} \frac{\partial \vec{p}(t)}{\partial t} d^3V \qquad (7)$$

plus terms of the order $\delta t^2$. At this point, we have everything expressed at the time (t) except the volume of integration in the $\partial \vec{p}/\partial t$ term. If we integrated over the volume V(t) instead of $V(t + \delta t)$, we would be incorrectly handling the integral of $\partial \vec{p}/\partial t$ over the narrow difference volume of thickness $\vec{v}\delta t$. Since the $\partial \vec{p}/\partial t$ term already has a factor $\delta t$, this would lead to an error of order $\delta t^2$ which we can ignore.

Replacing $V(t + \delta t)$ by V(t) in the volume integral, and dividing through by $\delta t$ gives

$$\frac{\delta \vec{P}_V}{\delta t} = \int_{V(t)} \frac{\partial \vec{p}}{\partial t} d^3V + \int_{S'(t)} \vec{p}(t)(\vec{v} \cdot d\vec{A}) \qquad (8)$$

We now have all quantities in our formula for $\delta \vec{P}_V/\delta t$ expressed at the time (t).

We have one more step before we are finished with the $\delta \vec{P}_V/\delta t$ term. We want to convert the surface integral to a volume integral.

We have already had some experience converting surface to volume integrals back in Chapter 7 on divergence. There we derived the ***divergence theorem***

$$\int_S \vec{E} \cdot d\vec{A} = \int_V \vec{\nabla} \cdot \vec{E} d^3V \qquad (7\text{-}21)$$

where $\vec{E}$ is any vector field, and the surface S bounds the volume V.

In Equation (8), we have something that looks more complex than the surface integral in (7-21), because of the presence of the extra vector $\vec{p}$. To handle this let us define three fields $\vec{E}_1$, $\vec{E}_2$ and $\vec{E}_3$ by

$$\vec{E}_1 = p_x \vec{v} \; ; \; \vec{E}_2 = p_y \vec{v} \; ; \; \vec{E}_3 = p_z \vec{v} \qquad (9)$$

Then we get

$$\int_S \vec{p}(\vec{v} \cdot d\vec{A})$$

$$= \hat{x} \int_S p_x \vec{v} \cdot d\vec{A} + \hat{y} \int_S p_y \vec{v} \cdot d\vec{A} + \hat{z} \int_S p_z \vec{v} \cdot d\vec{A}$$

$$= \hat{x} \int_S \vec{E}_1 \cdot d\vec{A} + \hat{y} \int_S \vec{E}_2 \cdot d\vec{A} + \hat{z} \int_S \vec{E}_3 \cdot d\vec{A} \qquad (10)$$

Now we can use the divergence theorem on the three quantities $\vec{E}_1$, $\vec{E}_2$ and $\vec{E}_3$ to get

$$\int_S \vec{p}(\vec{v} \cdot d\vec{A})$$

$$= \hat{x} \int_V \vec{\nabla} \cdot \vec{E}_1 d^3V + \hat{y} \int_V \vec{\nabla} \cdot \vec{E}_2 d^3V + \hat{z} \int_V \vec{\nabla} \cdot \vec{E}_3 d^3V$$

$$= \hat{x} \int_V \vec{\nabla} \cdot (p_x \vec{v}) d^3V + \hat{y} \int_V \vec{\nabla} \cdot (p_y \vec{v}) d^3V$$

$$+ \hat{z} \int_V \vec{\nabla} \cdot (p_z \vec{v}) d^3V \qquad (11)$$

(A quantity like $\vec{E}_1 = p_x \vec{v}$ is not really a vector field because it does not transform like a vector when we rotate the coordinate system. But if no rotations are involved, $p_x$ acts like a scalar field p, and $p_x \vec{v}$ acts like a vector field $\vec{j} = p\vec{v}$ in the divergence theorem.)

## Einstein Summation Convention

In Equation (11) we have some fairly mixed up vector components like

$$\hat{x}\vec{\nabla}\cdot(p_x\vec{v}) = \hat{x}[\nabla_x(p_xv_x) + \nabla_y(p_xv_y) + \nabla_z(p_xv_z)] \tag{12}$$

There is a notation, credited to Einstein, that makes it easy to handle such terms. In Equations (13), we write the dot product of two vectors in three different ways.

$$\vec{a}\cdot\vec{b} = a_xb_x + a_yb_y + a_zb_z \tag{13a}$$

$$= \sum_{i=x,y,z} a_ib_i \tag{13b}$$

$$= a_ib_i \tag{13c}$$

In (13a) we see the usual definition of the dot product of two vectors. In (13b), we used the index (i) to represent the subscripts x, y, z and included a summation sign to show we are adding up the three terms. Supposedly Einstein got tired of writing summation signs and introduced the notation in (13c). He said that if the index appears twice, then automatically take a sum. As an example, if you encounter $a_ib_j\,c_i$ you would sum over the repeated index (i) to get

$$a_ib_j\,c_i = \sum_{i=x,y,z} a_ib_jc_i = a_xb_jc_x + a_yb_jc_y + a_zb_jc_z \tag{14}$$

Since the index (j) is not summed over, it remains the same index throughout. We would say that $a_ib_j\,c_i$ is the (j)th component of the vector $a_i\vec{b}\,c_i$.

Using this notation in Equation (12), we have

$$\hat{x}[\nabla_x(p_xv_x) + \nabla_y(p_xv_y) + \nabla_z(p_xv_z)] = \hat{x}[\nabla_i(p_xv_i)] = \nabla_i([\hat{x}p_x]v_i) \tag{15}$$

and Equation (11) can be written as

$$\int_S \vec{p}(\vec{v}\cdot d\vec{A}) = \int_V \nabla_i([\hat{x}p_x + \hat{y}p_y + \hat{z}p_z]v_i)d^3V = \int_V \nabla_i(\vec{p}v_i)d^3V \tag{16}$$

Using Equation (16) in Equation (8) gives

$$\frac{\delta\vec{P}_V}{\delta t} = \int_{V(t)} \left[\frac{\partial\vec{p}}{\partial t} + \nabla_i(\vec{p}v_i)\right]d^3V \tag{17}$$

This is the formula for the rate of change of the momentum of the fluid particles inside the volume V that moves with the particles. It is all expressed in terms of variables at the time (t).

## Mass Continuity Equation

When we substitute $\vec{p} = \rho\vec{v}$ into Equation (17) we end up with quite a few terms. The result can be simplified by using the equation for the conservation of mass during the flow. The derivation, which is worth repeating, is similar to our derivation in Chapter 10 of the conservation of electric charge.

Consider a volume V bounded by a fixed surface S in a fluid of density $\rho$. The rate at which mass is flowing out of V (the mass flux) is given by the integral over S

$$-\frac{dM}{dt} = \int_S (\rho\vec{v})\cdot d\vec{A} \quad \begin{matrix}\textit{rate at which}\\ \textit{mass is flowing}\\ \textit{out across S}\end{matrix} \tag{18}$$

where $\rho\vec{v}$ is the mass current. We can use the divergence theorem to convert this surface integral to a volume integral, giving

$$-\frac{dM}{dt} = \int_V \vec{\nabla}\cdot(\rho\vec{v})d^3V \tag{19}$$

If mass is flowing out of V, there must be a decrease in the density $\rho$ inside. The rate at which the total mass inside is decreasing is related to the change in density $\rho$ by

$$-\frac{dM}{dt} = -\int_V \frac{\partial\rho}{dt}d^3V \tag{20}$$

Equating our two formulas for – dM/dt gives

$$\int_V \vec{\nabla}\cdot(\rho\vec{v})d^3V = -\int_V \frac{\partial\rho}{\partial t}d^3V \tag{21}$$

The two volume integrals can be combined to give

$$\int\left[\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{v})\right]d^3V = 0 \tag{22}$$

Since Equation (22) must hold for any volume V or fixed surface S we can construct, the terms in the square brackets must be zero, giving

$$\boxed{\frac{\partial \rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{v}) = 0} \qquad \textit{mass continuity equation} \qquad (23)$$

## Rate of Change of Momentum when Mass is Conserved

With the continuity equation written down, let us return to our formula for the rate of change of the momentum of the fluid particles, replacing the momentum density $\vec{p}$ by $\rho\vec{v}$ to get

$$\frac{\delta \vec{P}_V}{\delta t} = \int_V \left[ \frac{\partial(\rho\vec{v})}{\partial t} + \nabla_i(\rho\vec{v}v_i) \right] d^3V \qquad (24)$$

The terms in the square bracket become

$$\begin{aligned} [\ ] &= \frac{\partial \rho}{\partial t}\vec{v} + \rho\frac{\partial \vec{v}}{\partial t} + \vec{v}\nabla_i(\rho v_i) + \rho v_i \nabla_i \vec{v} \\ &= \rho\left[ \frac{\partial \vec{v}}{\partial t} + \vec{v}_i\nabla_i\vec{v} \right] + \vec{v}\left[ \frac{\partial \rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{v}) \right] \end{aligned} \qquad (25)$$

where we wrote $\nabla_i(\rho v_i) = \vec{\nabla}\cdot(\rho\vec{v})$ .

We immediately see that the second bracket is zero by the mass continuity equation, and we are left with our final result

$$\boxed{\frac{\delta \vec{P}_V}{\delta t} = \int_V \rho\left[ \frac{\partial v}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v} \right] d^3V} \qquad (26)$$

Equation (26) holds even when the density of the fluid is changing.

## Newton's Second Law

We are now in a position to apply Newton's second law to the fluid in our volume V. Equation (26) gives us the total rate of change of the momentum of the particles within V. We now want to equate that to the total force $\vec{F}_{tot}$ acting on the particles. We will calculate that by adding up the individual forces per unit volume, which are the pressure force, the viscous force, and the other forces. Then we integrate the sum over the volume V.

In View 3 of Chapter 3 on divergence, we found that the pressure force per unit volume was

$$\vec{f}_p = -\vec{\nabla}p \qquad (3.3\text{-}2)$$

In Chapter 4 we found that the viscous force per unit volume for a constant density Newtonian fluid was

$$\vec{f}_v = \mu\nabla^2\vec{v} \qquad (4\text{-}19)$$

Letting $\vec{f}_{other}$ represent all other forces per unit volume, we get for the total force $\vec{F}_{tot}$ acting on the fluid within V

$$\vec{F}_{tot} = \int_V \left[ -\nabla p + \mu\nabla^2\vec{v} + \vec{f}_{other} \right] d^2V \qquad (27)$$

Equating the total force $\vec{F}_{tot}$ to the rate of change of momentum $\delta\vec{P}_V/\delta t$, Equations (27) and (26), gives

$$\vec{F}_{tot} = \frac{\delta \vec{P}_V}{\delta t}$$

$$\begin{aligned} \vec{F}_{tot} &= \int_V \left[ -\nabla p + \mu\nabla^2\vec{v} + \vec{f}_{other} \right] d^3V \\ &= \int_V \rho\left[ \frac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v} \right] d^3V \end{aligned} \qquad (28)$$

Putting everything under a single integral sign gives us

$$\begin{aligned} &\int_V \left[ \rho\frac{\partial \vec{v}}{\partial t} + \rho(\vec{v}\cdot\vec{\nabla})\vec{v} + \vec{\nabla}p - \mu\nabla^2\vec{v} - \vec{f}_{other} \right] d^3V \\ &= 0 \end{aligned} \qquad (29)$$

Next we have our usual argument that Equation (29) must hold for any volume V. The only way we can always get the answer zero for the integral is for the integrand, the stuff in the square brackets, to be zero. Thus we end up with the equation

$$\rho\left[\frac{\partial\vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v}\right] = -\vec{\nabla}p + \mu\nabla^2\vec{v} + \vec{f}_{other} \tag{30}$$

This is one form of the Navier-Stokes equation.

It is usually more convenient to divide through by $\rho$, using

$$\nu = \frac{\mu}{\rho} \qquad \begin{matrix}\textit{kinematic}\\ \textit{viscosity}\\ \textit{coeffienct}\end{matrix} \tag{4-41}$$

where $\nu$ is the so called kinematic viscosity described in the pipe flow experiment of Chapter 4 (page Cal 4-9). We will also define $\vec{g}_{other}$ by

$$\vec{g}_{other} = \frac{\vec{f}_{other}}{\rho} \qquad \begin{matrix}\textit{other forces}\\ \textit{per unit mass}\end{matrix} \tag{31}$$

which represents all other forces, but now as force per unit mass, since we have divided by mass per unit volume $\rho$. We get

$$\boxed{\frac{\partial\vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v} = \frac{-\vec{\nabla}p}{\rho} + \nu\nabla^2\vec{v} + \vec{g}_{other}}$$

*Navier-Stokes Equation* (32)

Equation (32) is the form of the Navier-Stokes equation you are likely to find in the textbooks. It represents the basic starting point for fluid dynamics theory.

Equation (32) is quite general. Only in the formula $\nu\nabla^2\vec{v}$ for the viscous force have we made any assumptions about the density being constant (i.e., $\vec{\nabla}\cdot\vec{v} = 0$), and that the coefficient of viscosity $\nu$ is constant. If we have a non constant density fluid, or non constant coefficient of viscosity, all we have to do is correct the viscosity term.

In Chapter 23 of the Physics text, we began our discussion of vector fields with the velocity field. We made this choice because it is easier to picture a velocity field than an electric field, and we could immediately derive Bernoulli's equation from some simple energy arguments. How things have changed in this chapter! The derivation of the Navier-Stokes equation for the velocity field was harder to do than deriving the wave equations for $\vec{E}$ and $\vec{B}$, and the result is more complex. We have seen terms that resemble $\nu\nabla^2\vec{v}$ and $\partial\vec{v}/\partial t$ in our discussion of wave equations, but we have not encountered a term that looks anything like $(\vec{v}\cdot\vec{\nabla})\vec{v}$.

Not only does $(\vec{v}\cdot\vec{\nabla})\vec{v}$ have a peculiar combination of components, it is essentially proportional to the square of the velocity field, which makes the Navier-Stokes equation a ***non linear*** equation. What that means is as follows. The equations we have studied so far, the wave equations for $\vec{E}$ and $\vec{B}$, and Schrödinger's equation for $\psi$, are ***linear*** equations. This means that there are no terms involving the square of $\vec{E}$, $\vec{B}$ or $\psi$, and as a result we have the rule that ***waves add***. What this implies is that if you have two solutions to a wave equation, the sum of these two solutions is also a solution. For a non linear equation, the sum of two solutions is not necessarily a solution.

In the case of water waves, if the amplitudes of the waves are small, the $(\vec{v}\cdot\vec{\nabla})\vec{v}$ term is not important and waves add, as we saw in the ripple tank experiments. However, if the amplitudes become large, the $(\vec{v}\cdot\vec{\nabla})\vec{v}$ term, being proportional to $v^2$, becomes large and we get non linear effects like the breakers we see when ocean waves come up to the beach. There is no way you can get the solution describing a breaking wave from adding up the solutions for many small amplitude waves. The non linear term brings in completely new physics.

Despite the apparent complexity of the Navier-Stokes equation, some fairly simple results can be derived from it. One is Bernoulli's equation which we will discuss in the next section, the other is a generalized Helmholtz theorem which we will derive after that. In our discussion of Bernoulli's equation we learn more than we did in the Physics text. Here we will determine the conditions when Bernoulli's equation applies, and when it does not.

## BERNOULLI'S EQUATION

There is a vector identity which allows us to change the form of the Navier-Stokes equation so that the terms in Bernoulli's equation begin to appear. The vector identity is

$$(\vec{v}\cdot\vec{\nabla})\vec{v} = \vec{\nabla}\left(\frac{v^2}{2}\right) - \vec{v}\times(\vec{\nabla}\times\vec{v}) \qquad (33)$$

In Appendix 1 of this chapter we show you a relatively easy way to derive vector identities involving the curl. Equation (33) is the explicit example we use.

Noting that $\vec{\nabla}\times\vec{v}$ is the vorticity $\vec{\omega}$, we can write Equation (33) as

$$(\vec{v}\cdot\vec{\nabla})\vec{v} = \vec{\nabla}\left(\frac{v^2}{2}\right) - \vec{v}\times\vec{\omega} \qquad (33a)$$

Using Equation (33a) for the $(\vec{v}\cdot\vec{\nabla})\vec{v}$ term in the Navier-Stokes equation (32) gives

$$\frac{\partial\vec{v}}{\partial t} - \vec{v}\times\vec{\omega} = -\vec{\nabla}\left(\frac{v^2}{2}\right) - \frac{\vec{\nabla}p}{\rho} + \nu\nabla^2\vec{v} + \vec{g}_{other} \qquad (34)$$

Our next step is to extract the gravitational force from $\vec{g}_{other}$ and display it explicitly. The gravitational force per unit volume of fluid $\vec{f}_g$ is

$$\vec{f}_g = \rho\vec{g} = \rho(-\vec{\nabla}gy) \qquad (35)$$

where y is the upward directed coordinate and g the acceleration due to gravity. (Take a break and show that $-\vec{\nabla}(gy)$ is equal to $\vec{g}$, a vector of magnitude g pointing down.)

The force terms in Equation (32) are forces per unit mass. We get the gravitational force per unit mass, $\vec{g}_{gravity}$ by dividing $\vec{f}_g$ by the density $\rho$.

$$\vec{g}_{gravity} = \frac{\vec{f}_g}{\rho} = -\vec{\nabla}(gy) \qquad (36)$$

The force $\vec{g}_{other}$ becomes

$$\vec{g}_{other} = -\vec{\nabla}(gy) + \vec{g}'_{other} \qquad (36a)$$

where $\vec{g}'_{other}$ represents other forces not including gravity.

Using Equation (36a) in Equation (34) gives

$$\frac{\partial\vec{v}}{\partial t} - \vec{v}\times\vec{\omega} = -\vec{\nabla}\left(\frac{v^2}{2} + gy\right) - \frac{\vec{\nabla}p}{\rho} + \nu\nabla^2\vec{v} + \vec{g}'_{other} \qquad (37)$$

Up to this point the only place we assumed that $\rho$ was constant was in the viscosity term $\nu\nabla^2\vec{v}$. But for the remainder of this chapter we will assume that $\rho$ is constant and use that to simplify other terms. For example, we can pull a constant $\rho$ inside the gradient, giving

$$-\frac{\vec{\nabla}p}{\rho} = -\vec{\nabla}\left(\frac{p}{\rho}\right) \qquad \textit{if } \rho \textit{ is constant} \qquad (38)$$

Using Equation (38) in Equation (37) gives

$$\boxed{\frac{\partial\vec{v}}{\partial t} - \vec{v}\times\vec{\omega} = -\vec{\nabla}\left(\frac{p}{\rho} + \frac{v^2}{2} + gy\right) + \nu\nabla^2\vec{v} + \vec{g}'_{other}}$$

*constant density fluids* (39)

It is in Equation (39) we see the Bernoulli terms $(p/\rho + v^2/2 + gy)$. We can now use the equation both to derive Bernoulli's equation and to state the conditions under which it applies.

Suppose we have the following four conditions: *(1)* constant density, *(2)* a steady flow so that $\partial\vec{v}/\partial t = 0$, *(3)* that viscosity is not important so that we can neglect the viscosity term $\nu\nabla^2\vec{v}$, and *(4)* that there are no forces other than pressure and gravity acting on the fluid so that we can set $\vec{g}'_{other} = 0$. These conditions are

$$\boxed{\begin{aligned} \rho &= \text{constant} \\ \frac{\partial\vec{v}}{\partial t} &= 0 \quad \textit{steady flow} \\ \nu\nabla^2\vec{v} &= 0 \quad \textit{neglect viscosity} \\ \vec{g}'_{other} &= 0 \quad \textit{no other forces} \end{aligned}} \qquad (40)$$

Under conditions (40) the Navier-Stokes equation becomes

$$\boxed{\vec{v}\times\vec{\omega} = \vec{\nabla}\left(\frac{p}{\rho} + \frac{v^2}{2} + gy\right)} \qquad (41)$$

## Applies Along a Streamline

In Chapter 23 of the Physics text, we called the collection of Bernoulli terms the ***hydrodynamic voltage***. Labeling their sum by $\phi_H$, we have

$$\phi_H \equiv \frac{p}{\rho} + \frac{v^2}{2} + gy \qquad \begin{matrix}\textit{hydrodynamic}\\ \textit{voltage}\end{matrix} \qquad (42)$$

With this notation, Equation (41) becomes

$$\vec{v}\times\vec{\omega} = \vec{\nabla}\phi_H \qquad (43)$$

We used the name hydrodynamic voltage for $\phi_H$ to stress the similarity between hydrodynamic voltage-drops in a fluid circuit and electric voltage-drops in an electric circuit.

Later in the Physics text, in our discussion of electric voltage in Chapter 25, we changed the name from voltage to potential, and started constructing contour maps of the potential $\phi$. Our main example was the map of the electric potential produced by charges +3 and –1 shown in Figure (25-15) reproduced again here. The lines of constant potential are the contour lines, and the lines of steepest descent are the field lines.

In our discussion of gradient in this text, we saw that the gradient vector $\vec{\nabla}\phi$ pointed along the field lines. Or to say it another way, the gradient $\vec{\nabla}\phi$ was a maximum in the direction where the slope is the steepest, and was zero in the direction of a contour line where the value of $\phi$ remains constant.

Our Equation (43), $\vec{v}\times\vec{\omega} = \vec{\nabla}\phi_H$, is an equation relating the gradient of the potential $\phi_H$ to what at first looks like a rather complicated term $\vec{v}\times\vec{\omega} = \vec{v}\times(\vec{\nabla}\times\vec{v})$. But there is one thing that is simple about $\vec{v}\times\vec{\omega}$. Because of the cross product, $\vec{v}\times\vec{\omega}$ ***is always perpendicular to*** $\vec{v}$***, i.e., always zero in the direction of*** $\vec{v}$.

In a fluid flow, the streamlines follow in the direction of the velocity field $\vec{v}$. Thus if we move in the direction of a streamline, we are moving in a direction where $\vec{v}\times\vec{\omega}$ and thus $\vec{\nabla}\phi_H$ is zero. But if we move in a direction where the gradient of $\phi_H$ is zero, we must be moving along a contour line of $\phi_H$, and the value of $\phi_H$ must be constant. Thus the physical content of the equation $\vec{\nabla}\phi_H = \vec{v}\times\vec{\omega}$ is that $\phi_H$ ***is constant along a streamline***. Re-expressing $\phi_H$ as $p/\rho + v^2/2 + gy$, we get the result

$$\boxed{\frac{p}{\rho} + \frac{v^2}{2} + gy = \begin{matrix}\text{constant along}\\ \text{a streamline}\end{matrix}} \qquad (44)$$

when conditions (40) are obeyed.

Equation (44), with the associated conditions, is our precise statement of Bernoulli's equation. It tells us both when Bernoulli's equation can be used, and why it should be applied along a streamline. In the special case of potential flow where $\vec{\omega} = \vec{\nabla}\times\vec{v}$ is zero everywhere, then Equation (41) becomes $\vec{\nabla}\phi_H = 0$, which implies $\vec{\nabla}\phi_H = p/\rho + v^2/2 + gy =$ *constant throughout the fluid*. For potential flow we do not have to apply Bernoulli's equation only along a streamline.

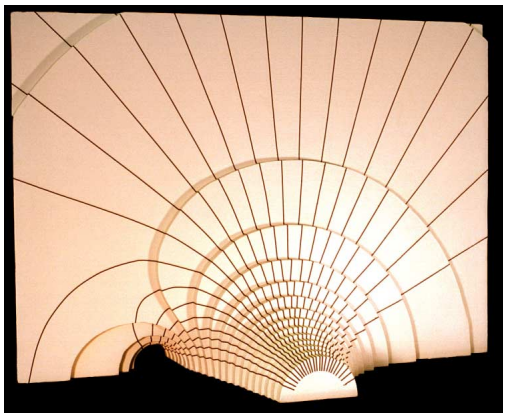


**Figure 25-15 (repeated)**
***The lines of equal height, the contour lines, are the lines along which the potential $\phi$ is constant.***

## The Viscosity Term

Although the Navier-Stokes equation is a rather formidable equation, we are beginning to see some fairly simple or recognizable results emerge. A lot can be learned by studying the nature of the terms in the equation. Here we will see that the viscous force term $\nu\nabla^2\vec{v}$ can be re-expressed in a form that gives one a better understanding of the nature of vortices.

Back in Chapter 8 on the curl, we proved the vector identity

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{A}) = -\nabla^2\vec{A} + \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) \qquad (8\text{-}5)$$

If we apply this to the velocity field $\vec{v}$ of a constant density fluid where $\vec{\nabla}\cdot\vec{v} = 0$, we get

$$\nabla^2\vec{v} = -\vec{\nabla}\times(\vec{\nabla}\times\vec{v}) = -\vec{\nabla}\times\vec{\omega} \qquad (45)$$

Where $\vec{\omega} = \vec{\nabla}\times\vec{v}$. Thus the viscous force term in the Navier-Stokes equation can be written as

$$\boxed{\nu\nabla^2\vec{v} = -\nu\vec{\nabla}\times\vec{\omega}} \quad \textit{viscous force per unit mass} \qquad (46)$$

From Equation (46) we see that there are no viscous forces where the vorticity $\vec{\omega}$ is zero, or even when $\vec{\omega}$ is constant as in solid body rotation.

In our discussion of vortices in the last chapter, we pictured an ideal vortex as one whose velocity field $\vec{v}$ was analogous to the magnetic field of a current in a straight wire. If the current in the wire is uniform, then $\vec{\nabla}\times\vec{B} = \mu_0\vec{i}$ is a constant inside the wire and zero outside. Thus in our ideal vortex, $\vec{\omega} = \vec{\nabla}\times\vec{v}$ is uniform inside the core (representing a solid body rotation of the fluid there), and $\vec{\omega} = 0$ outside where we have the $\hat{\theta}$ directed 1/r velocity field.

With our new formula $-\nu\vec{\nabla}\times\vec{\omega}$ for the viscous force, we see that there is no viscous force acting inside the core where $\vec{\omega}$ = constant. What is surprising is that there is also no viscous force acting outside the core in the 1/r circular velocity field. The only place where viscous forces act in an ideal vortex is at the boundary between the core and the fluid outside. The fact that viscous forces do not act either inside or outside the core of an ideal vortex is one reason for the permanence of the vortex structure.

Because the velocity field of a vortex ring is analogous to the magnetic field of a current loop, the fact that $\vec{\nabla}\times\vec{B} = \mu_0\vec{i}$ is zero outside the wire loop, implies that the vorticity $\vec{\omega} = \vec{\nabla}\times\vec{v}$ is zero outside the core of a vortex ring. Thus in a vortex ring or a smoke ring, viscous forces do not act on the fluid outside the core.

$\vec{F}_{magnus} = -\rho\vec{V}_{rel}\times\vec{\kappa}$ acting on that vortex. But there is no extra mass associated with a fluid core vortex, so one must treat the vortex as a massless object, with the result that the net force on the vortex must be zero. That means that there must be an external force $\vec{F}_{external}$ acting on the vortex to cancel the Magnus lift force. That is, one must have

$$\vec{F}_{external} + \vec{F}_{magnus} = 0 \qquad (108)$$

## THE HELMHOLTZ THEOREM

While Bernoulli's theorem may be the most famous theorem of fluid dynamics, Helmholtz's theorem is perhaps the most dramatic. To see a smoke ring emerge from an amorphous cloud of smoke and travel across a room in a straight line has to be one of the impressive phenomena of physics. Yet we saw that it was explained by Helmholtz's theorem that in the absence of non potential forces, the fluid particles become trapped on, and move with, the vortex lines.

In this section we will derive Helmholtz's theorem from the Navier-Stokes equation. As a result, all the phenomena we have seen that are explained by Helmholtz's theorem can be viewed as being a consequence of the Navier-Stokes equation.

### Equation for Vorticity

The first step in deriving Helmholtz's theorem is to turn the Navier-Stokes equation into an equation for the vorticity field $\vec{\omega}$. We do this by taking the curl of both sides of Equation (39). We have

$$\vec{\nabla}\times\left[\frac{\partial\vec{v}}{\partial t}-\vec{v}\times\vec{\omega}\right] = \vec{\nabla}\times\left[-\vec{\nabla}\left(\frac{p}{\rho}+\frac{v^2}{2}+gy\right)-\nu\vec{\nabla}\times\vec{\omega}+\vec{g}'_{other}\right] \quad (47)$$

where we used Equation (46) to replace $\nu\nabla^2\vec{v}$ by $-\nu\vec{\nabla}\times\vec{\omega}$.

At this point you might be discouraged by the number of cross products that appear in Equation (47). But immediately there is noticeable simplification. Recall that the curl of a gradient is identically zero,

$$\vec{\nabla}\times\vec{\nabla}\phi \equiv 0 \qquad \textit{any } \phi \quad (48)$$

Thus the Bernoulli terms all go out in Equation (47)

$$\vec{\nabla}\times\vec{\nabla}\left(\frac{p}{\rho}+\frac{v^2}{2}+gy\right) = 0 \quad (49)$$

which considerably shortens the equation.

Next, we note that because we can interchange the order of partial differentiation, we get

$$\vec{\nabla}\times\frac{\partial\vec{v}}{\partial t} = \frac{\partial}{\partial t}(\vec{\nabla}\times\vec{v}) = \frac{\partial\vec{\omega}}{\partial t} \quad (50)$$

Thus Equation (47), the curl of the Navier-Stokes equation, becomes

$$\boxed{\frac{\partial\vec{\omega}}{\partial t}-\vec{\nabla}\times(\vec{v}\times\vec{\omega}) = \vec{\nabla}\times\vec{g}} \quad (51)$$

where $\vec{g}$, given by

$$\boxed{\vec{g} = -\nu\vec{\nabla}\times\vec{\omega}+\vec{g}'_{other}} \quad (52)$$

represents all forces per unit mass acting on the fluid, except pressure and gravity. Equation (51) is the differential equation for the dynamical behavior of the vorticity field $\vec{\omega}$. The only restriction is that it applies to constant density fluids. If we wish to work with non constant density fluids we have to go back and work with Equation (39) and perhaps use a more general formula for the viscous force.

### Non Potential Forces

An important simplification we obtained in going to an equation for the vorticity field $\vec{\omega}$ was the elimination of the Bernoulli terms. This removes the pressure and gravitational forces from the equation for $\vec{\omega}$, implying that pressure and gravity have no direct effect on the behavior of vorticity. We saw this result in the case of the motion of a smoke ring. The ring moved in a straight line across the room completely unaffected by gravity. (Pressure and gravity can have an indirect effect in that they affect the velocity field $\vec{v}$ which appears in the $\vec{\nabla}\times(\vec{v}\times\vec{\omega})$ term.)

In Equation (51),

$$\frac{\partial\vec{\omega}}{\partial t} - \vec{\nabla}\times(\vec{v}\times\vec{\omega}) = \vec{\nabla}\times\vec{g} \qquad \text{(51) repeated}$$

the only force terms that survive are those with a non zero curl like the viscosity term. Let us introduce the terminology ***potential force*** $\vec{g}_\phi$ and a ***non potential force*** $\vec{g}_{np}$. Potential forces are those that can be expressed as the gradient of a potential $\phi$, and thus have a zero curl

$$\vec{g}_\phi = -\vec{\nabla}\phi \;\; ; \;\; \vec{\nabla}\times\vec{g}_\phi = 0 \qquad (53)$$

while non potential forces $\vec{g}_{np}$ have non zero curl

$$\vec{\nabla}\times\vec{g}_{np} \neq 0 \qquad (54)$$

and thus survive the curl in Equation (51). As a result we can write Equation (51) in the form

$$\boxed{\frac{\partial\vec{\omega}}{\partial t} - \vec{\nabla}\times(\vec{v}\times\vec{\omega}) = \vec{\nabla}\times\vec{g}_{np}} \quad \textit{vortex dynamics equation} \quad (55)$$

We will call Equation (55) the vortex dynamics equation.

To be quite general, one might like to separate an arbitrary force field $\vec{g}$ into its potential part $\vec{g}_\phi$ and its non potential part $\vec{g}_{np}$, writing

$$\vec{g} = \vec{g}_\phi + \vec{g}_{np} \qquad (56)$$

The problem is that there is no unique separation of an arbitrary vector field into potential and non potential parts. The only thing that is unique is the curl

$$\vec{\nabla}\times\vec{g} = \vec{\nabla}\times\vec{g}_{np} \qquad (57)$$

Physically, Equation (57) is telling us that if we accidentally included some potential terms in our formula for $\vec{g}_{np}$, they would disappear when we took the curl in Equation (57).

For a practical matter, the best thing to do is to include all obviously potential forces like pressure and gravity in $\vec{g}_\phi$, and leave all others that are not obviously potential forces, like the viscous force $-\nu\vec{\nabla}\times\vec{\omega}$, in the non potential category $\vec{g}_{np}$.

## A VECTOR IDENTITY FOR A MOVING CIRCUIT

Before we obtain a really clear interpretation of the vortex dynamics equation (55), we need a way of understanding the impact of the rather complex looking term $-\vec{\nabla}\times(\vec{v}\times\vec{\omega})$. In this section, we will derive a vector identity that will lead to a strikingly simple interpretation of the combination of terms $\partial\vec{\omega}/\partial t - \vec{\nabla}\times(\vec{v}\times\vec{\omega})$. The vector identity involves the rate of change of flux of a solenoidal field like $\vec{\omega}$ through a circuit that moves with the fluid particles.

It takes a considerable effort to derive this vector identity, an effort involving steps somewhat similar to those we used to calculate the rate of linear momentum in a moving volume. But the resulting simplification in the interpretation of the vortex dynamics equation is more than worth the effort.

To emphasize the general nature of the vector identity, we will calculate the rate of change of the flux of a vector field $\vec{A}$ through the circuit $C'$ that moves with the fluid particles. The restriction on $\vec{A}$ will be that it is a solenoidal field with $\vec{\nabla}\cdot\vec{A} = 0$.

Let the circuit $C'(t)$ shown in Figure (4) be attached to the fluid particles through which it passes. As time progresses from (t) to $(t+\delta t)$, the fluid motion will carry the circuit from position $C'(t)$ to the position $C'(t+\delta t)$ as shown. We will also assume that there is a divergence free vector field $\vec{A}(t)$ in the fluid at time (t). At time $(t+\delta t)$ the vector field will have changed to $\vec{A}(t+\delta t)$. What we wish to calculate is the change in the flux of $\vec{A}$ through the circuit $C'$ as we go from (t) to $(t+\delta t)$. We will do the

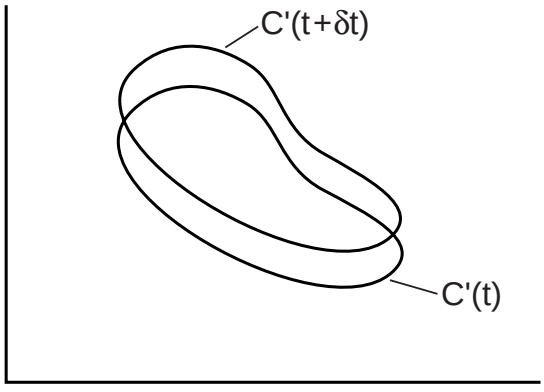



**Figure 4**
***The circuit C' moves with the fluid particles.***

calculation throwing out terms of order $\delta t^2$ compared to $\delta t$.

At time t, the flux $\Phi(t)$ of $\vec{A}$ through $C'(t)$ is

$$\Phi(t) = \int_{S'(t)} \vec{A}(t)\cdot d\vec{S} \qquad (58)$$

where $S'$is a surface bounded by $C'(t)$. At time $(t+\delta t)$ the flux has become

$$\Phi(t+\delta t) = \int_{S'(t+\delta t)} \vec{A}(t+\delta t)\cdot d\vec{S} \qquad (59)$$

The change in flux $\delta\Phi$ during the time $\delta t$ is

$$\delta\Phi = \int_{S'(t+\delta t)} \vec{A}(t+\delta t)\cdot d\vec{S} - \int_{S'(t)} \vec{A}(t)\cdot d\vec{S} \qquad (60)$$

Using a Taylor series expansion we can write

$$\vec{A}(t+\delta t) = \vec{A}(t) + \frac{\partial A}{\partial t}\delta t + 0(\delta t^2) \qquad (61)$$

Thus

$$\delta\Phi = \int_{S'(t+\delta t)} \vec{A}(t)\cdot d\vec{S} - \int_{S'(t)} \vec{A}(t)\cdot d\vec{S} + \delta t\int_{S'(t+\delta t)} \frac{\partial\vec{A}}{\partial t}\cdot d\vec{S} \qquad (62)$$

To calculate the effect of the first two terms in Equation (62), consider the guitar shaped volume shown in Figure (5). The top of the volume is bounded by the curve $C'(t+\delta t)$, while the bottom by $C'(t)$. A certain amount of flux $\Phi_1$

$$\Phi_1 = \int_{S'(t)} \vec{A}(t)\cdot d\vec{S} \qquad (63)$$

enters up through the bottom of the volume. Some more flux, $\Phi_2$ flows in through the sides, and an amount $\Phi_3$

$$\Phi_3 = \int_{S'(t+\delta t)} \vec{A}(t)\cdot d\vec{S} \qquad ((64)$$

flows out through the top.

Because $\vec{A}(t)$ is a divergence free field $[\vec{\nabla}\cdot\vec{A}(t) = 0]$, all the flux flowing in through the bottom, $\Phi_1$, and the sides, $\Phi_2$, must flow out through the top, $\Phi_3$, giving

$$\Phi_3 = \Phi_1 + \Phi_2 \qquad (65)$$

(Any of these fluxes could be negative, indicating $\vec{A}$ pointing in other directions, but all signs are correctly handled by the formalism.)

Using Equations (63) and (64), our formula (62) for $\delta\Phi$ becomes

$$\delta\Phi = \Phi_3 - \Phi_1 + \delta t\int_{S'(t+\delta t)} \frac{\partial\vec{A}}{\partial t}\cdot d\vec{S}$$

With $\Phi_3 = \Phi_1 + \Phi_2$ we get

$$\delta\Phi = \Phi_2 + \delta t\int_{S'(t+\delta t)} \frac{\partial\vec{A}}{\partial t}\cdot d\vec{S} \qquad (66)$$

Equation (66) tells us that the change in the flux of $\vec{A}(t)$ through the moving circuit $C'(t)$ is made of two parts. One is due to the change $\partial\vec{A}(t)/\partial t$ of the field itself, the other to flux coming in from the sides.

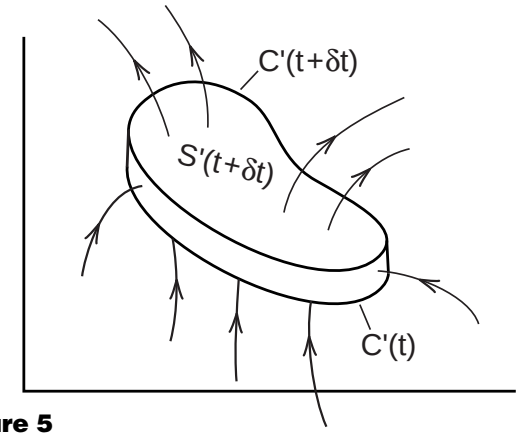



**Figure 5**
***Volume bounded by the curves $C'(t+\delta t)$ and $C'(t)$. The drawing shows flux entering through the bottom and sides, and flowing out through the top.***

Our problem now is to calculate the flux $\Phi_2$ flowing in through the sides of our volume shown in Figure (5). The calculation of $\Phi_2$ turns out not to be so hard. In Figure (6) we show a small piece of the side of our volume. A fluid particle that is located at position (1) in that diagram at time (t), moves to position (2) during the time $\delta t$. The distance from (1) to (2) is described by the displacement vector $\vec{v}\delta t$ as shown.

We also mark a short length $d\vec{\ell}$ of the path $C'(t)$ starting at position (1). If we take the cross product of $\vec{v}\delta t$ with $d\vec{\ell}$, we get a vector $d\vec{S}$ that points into the volume, perpendicular to both $\vec{v}\delta t$ and $d\vec{\ell}$. The length of $d\vec{S}$ is equal to the area of the parallelogram defined by $\vec{v}\delta t$ and $d\vec{\ell}$. Thus $d\vec{S}$ represents the inward area vector for the shaded area in Figure (6). The flux $d\Phi_2$ of $\vec{A}(t)$ in through this side area $d\vec{S}$ is

$$\begin{aligned} d\Phi_2 &= \vec{A}(t)\cdot d\vec{S} = \vec{A}(t)\cdot[(\vec{v}\delta t)\times d\vec{\ell}] \\ &= \delta t\left[\vec{A}(t)\cdot(\vec{v}\times d\vec{\ell})\right] \end{aligned} \tag{67}$$

In the appendix to this chapter, where we show you an easy way to handle vector identities involving cross products, we derive the identity

$$\vec{A}\cdot(\vec{B}\times\vec{C}) = (\vec{A}\times\vec{B})\cdot\vec{C} \tag{68}$$

Using this identity, we can write Equation (67) in the form

$$d\Phi_2 = \delta t[\vec{A}(t)\times\vec{v}]\cdot d\vec{\ell} \tag{69}$$

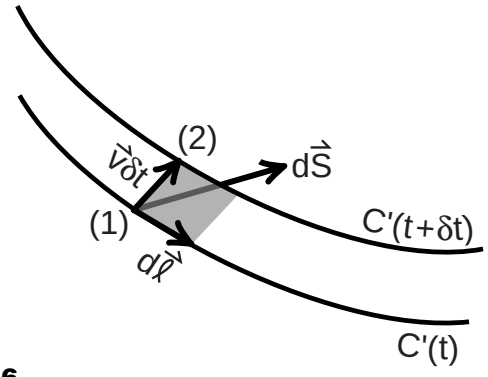



**Figure 6**
***The area element $d\vec{S}$ on the side of our volume.***

To calculate the total flux $\Phi_2$ in through the sides of our volume, all we have to do is integrate the contributions $\Phi_2$ around the circuit $C'(t)$. We get

$$\Phi_2 = \delta t\oint_{C'(t)}[\vec{A}(t)\times\vec{v}]\cdot d\vec{\ell} \tag{70}$$

Stokes' law, derived in Chapter 8 relates the integral of a vector field $\vec{B}$ around a closed path to the flux of $\vec{\nabla}\times\vec{B}$ through the path. We had

$$\oint_C \vec{B}\cdot d\vec{\ell} = \int_S \vec{\nabla}\times\vec{B}\cdot d\vec{S} \tag{8-31}$$

where S is the surface bounded by the closed curve C. If we set $\vec{B} = \vec{A}(t)\times\vec{v}$, $C = C'(t)$ and $S = S'(t)$, Equation (8-31) becomes

$$\oint_{C'(t)}[\vec{A}(t)\times\vec{v}]\cdot d\vec{\ell} = \int_{S'(t)}\vec{\nabla}\times\left[\vec{A}(t)\times\vec{v}\right]\cdot d\vec{S} \tag{71}$$

As a result, the flux $\Phi_2$ of $\vec{A}(t)$ flowing in through the sides of our volume is

$$\Phi_2 = \delta t\int_{S'(t)}\vec{\nabla}\times\left[\vec{A}(t)\times\vec{v}\right]\cdot d\vec{S} \tag{72}$$

Using this result in Equation (66) for the change in flux $\delta\phi$ through our moving circuit gives

$$\delta\Phi = \Phi_2 + \delta t\int_{S'(t+\delta t)}\frac{\partial\vec{A}}{\partial t}\cdot d\vec{S} \qquad \text{(66) repeated}$$

$$= \delta t\int_{S'(t)}\vec{\nabla}\times\left[\vec{A}(t)\times\vec{v}\right]\cdot d\vec{S} + \delta t\int_{S'(t+\delta t)}\frac{\partial\vec{A}}{\partial t}\cdot d\vec{S} \tag{73}$$

At this point everything is evaluated at the time (t) except for the integral of the flux of $\partial\vec{A}(t)/\partial t$ at the surface $S'(t+\delta t)$.

As we have just seen, the flux of any vector field through $S'(t+\delta t)$ is equal to the flux through the end $S'(t)$ plus a term like $\Phi_2$ representing a flow in through the sides. Because the flux in through the sides is of the order $\delta t$ smaller than the flow in through the end, and because the $\partial\vec{A}/\partial t$ term already has a factor of $\delta t$, our neglect of the flux of $\partial\vec{A}/\partial t$ in through the sides will be an error of order $\delta t^2$ which may be ignored. Thus we can replace $S'(t+\delta t)$ by $S'(t)$ in Equation (73). Dividing through by $\delta t$, and for later convenience replacing $\vec{A}(t)\times\vec{v}$ by $-\vec{v}\times\vec{A}(t)$, we get

$$\frac{\delta\Phi(\vec{A})}{\delta t} = \int_{S'(t)} \left\{ \frac{\partial\vec{A}(t)}{\partial t} - \vec{\nabla}\times\left[ \vec{v}\times\vec{A}(t) \right] \right\} \cdot d\vec{S} \qquad (74)$$

Equation (74) is the general formula for the rate of change of flux of the vector $\vec{A}(t)$ through a circuit $C'(t)$ that moves with the fluid particles. The circuit $C'(t)$ bounds the surface $S'(t)$, and it is assumed that $\vec{A}$ is a solenoidal field ($\vec{\nabla}\cdot\vec{A}=0$).

## The Integral Form of the Vortex Dynamics Equation

Although the derivation of Equation (74) was rather lengthy, the result can be immediately applied to our vortex dynamics Equation (55). If we integrate Equation (55) over a surface $S'(t)$ bounded by a circuit $C'(t)$ we get

$$\int_{S'(t)} \left[ \frac{\partial\vec{\omega}(t)}{\partial t} - \vec{\nabla}\times(\vec{v}\times\vec{\omega}) \right] \cdot d\vec{S} = \int_{S'(t)} [\vec{\nabla}\times\vec{g}_{np}]\cdot d\vec{S} \qquad (75)$$

Because the vorticity $\vec{\omega}$ is always a solenoidal field, we can replace $\vec{A}(t)$ by $\vec{\omega}(t)$ in Equation (74) and immediately recognize the left side of Equation (75) as the rate of change of the flux of $\vec{\omega}$ through the moving circuit $C'(t)$. Calling this rate $\delta\Phi(\vec{\omega})/\delta t$, we have

$$\int_{S'(t)} \left[ \frac{\partial\vec{\omega}(t)}{\partial t} - \vec{\nabla}\times(\vec{v}\times\vec{\omega}) \right] \cdot d\vec{S} = \frac{\delta\Phi(\vec{\omega})}{\delta t} \qquad (76)$$

On the right side of Equation (75), we can use Stokes' theorem to replace the surface integral of $\vec{\nabla}\times\vec{g}_{np}$ over $S'(t)$ by the line integral of $\vec{g}_{np}$ around $C'(t)$ giving

$$\int_{S'(t)} [\vec{\nabla}\times\vec{g}_{np}]\cdot d\vec{S} = \oint_{C'(t)} \vec{g}_{np}\cdot d\vec{\ell} \qquad (77)$$

Combining Equations (76) and (77) gives us the general vortex dynamics Equation (78), a result which assumes only that $\rho$ is constant.

the rate of change of the flux of $\vec{\omega}$ through a circuit $C'(t)$ that moves with the fluid particles

$$\frac{\delta\Phi_\omega}{\delta t} = \oint_{C'(t)} \vec{g}_{np}\cdot d\vec{\ell}$$

*extended Helmholtz equation*

(78)

It seems rather remarkable that an equation as complex looking as the Navier-Stokes equation can be converted, by taking the curl, to something simple enough to be described almost completely in words. In a sense the only calculation we have to do to apply Equation (78), is to calculate the line integral of a non potential force $\vec{g}_{np}$ around a closed path. For reasons that will become clear shortly, we will call Equation (78) the ***extended Helmholtz equation***.

## The Helmholtz Theorem

It is an immediate step to go from Equation (78) to Helmholtz's famous theorem of 1858. If there are no non potential forces acting on the fluid, i.e., if $\vec{g}_{np} = 0$, then we get the simple statement

> If there are no non potential forces acting on the fluid, then there is no change in the flux of $\vec{\omega}$ through any closed circuit $C'(t)$ that moves with the fluid particles

*Helmholtz theorem* (79)

At this point we have reduced much of fluid dynamics to a simple word equation.

Equation (79) is perhaps the most precise statement of Helmholtz's theorem, but equivalent statements are also enlightening. Suppose, for example, we define a ***vortex line*** as a ***small unit flux tube of*** $\vec{\omega}$. Because $\vec{\omega}$ is solenoidal, the flux tubes or vortex lines cannot stop or start in the fluid. Equation (79) tells us that, in the absence of non potential forces, the number of vortex lines threading any circuit $C'(t)$, i.e., the total flux of $\vec{\omega}$, remains constant as the circuit moves with the fluid particles. This clearly will happen if the lines themselves move with the fluid.

Equation (79) does not actually require, in all cases, that the vortex lines must move with the fluid particles. As we saw back in Chapter 12, the vorticity $\vec{\omega}$ is uniform for solid body rotation. Thus the flux of $\vec{\omega}$ will remain constant through any circuit $C'(t)$ moving with the fluid, whether or not we think of the vortex lines themselves as moving with the fluid. With a uniform $\vec{\omega}$, we cannot tell if the vortex lines are moving or not.

We saw, however, that the situation is very different when dealing with a quantum fluid where the vorticity $\vec{\omega}$, although roughly imitating solid body rotation, is lumped up in the vortex cores. In this case Equation (79) clearly requires that the separate vortex cores move around with the fluid. We can easily tell whether lumped up vorticity is moving.

There is, however, no harm in assuming that the vortex lines move with the fluid for solid body rotation. This interpretation has the advantage that if a slight perturbation is introduced into the vorticity field, we can follow the perturbation and see that the associated lines do move.

## EXTENDED HELMHOLTZ THEOREM

If the Helmholtz theorem tells us that in the absence of non potential forces, vortex lines move with the fluid particles, then what happens when non potential forces are present? What is the effect on vorticity of a force $\vec{g}_{np} \neq 0$? The answer, which we obtain from our vortex dynamics Equation (78) is quite simple. It is that ***the non potential forces*** $\vec{g}_{np}$ ***cause a relative motion of the vortex lines and the fluid particles.***

It was the study of the behavior of quantized vortices in superfluid helium and superconductors that led to a more complete understanding of the effect of non potential forces on vortex motion. One experiment in particular, an experiment by Rayfield and Reif involving charged vortex rings in superfluid helium, is what initiated this detailed study. We will use a discussion of the Rayfield-Reif experiment to develop the ideas contained in the extended Helmholtz theorem.

## The Rayfield-Reif Experiment

Rayfield and Reif were able to create their charged vortex rings by placing a radioactive substance in a container of superfluid helium. The radioactive substance emitted charged particles, either electrons or protons, depending on the substance. What they found was that the charged particle, moving through the superfluid, would create quantized vortex rings in the superfluid, and then in a process still not perfectly understood, the charged particle would become trapped in the core of the ring it created, producing an electrically charged vortex ring.

The interesting part about having an electrically charged vortex ring, is that you can apply an electric field and exert an electric force on the core of the ring. We will see that this electric force acting on the core represents a non potential force acting on the fluid in the region of the core. As a result, Rayfield and Reif were able to study, in detail, the effects of non potential forces acting on vortex lines. Their experiments provided a superb verification of Equation (78) and the interpretation that non potential forces cause a relative motion of the vortex lines and the fluid particles.

To apply Equation (78) to the Rayfield-Reif experiment, consider Figure (7) where we show the cross section of a vortex core with a force density $\vec{g}$ acting on the fluid in the core. The force $\vec{g}$ represents the electric force acting on the charged fluid in the core. Outside the core there is no force where the fluid is electrically neutral.

On Figure (7) we have drawn three contours labeled $C'_1$, $C'_2$, and $C'_3$. The primes indicate that these paths are moving with the fluid particles, and that we are looking at the paths now at time (t). If we integrate $\vec{g}$ around contour $C'_1$, we get a positive contribution along the bottom section of the path, and no contribution from the other sections that lie outside the core. Thus we get

$$\oint_{C'_1} \vec{g} \cdot d\vec{\ell} \;=\; \text{positive number} \qquad (80)$$

For the force density $\vec{g}$ to be a conservative potential force, we would have to have $\oint \vec{g} \cdot d\vec{\ell} = 0$ for any possible path. Because the integral is not zero for circuit $C'_1$, Equation (80) shows that ***$\vec{g}$ is a non potential force***.

To see what a localized force like $\vec{g}$ cannot do, look at the path $C'_3$ that goes completely around the core and lies completely in a region where $\vec{g} = 0$. For this path we get

$$\oint_{C'_3} \vec{g} \cdot d\vec{\ell} \;=\; 0 \qquad (81)$$

Thus from Equation (78) we find that there is no change in the flux of $\vec{\omega}$ through the path $C'_3$. Since $C'_3$ goes around the entire core, the flux of $\vec{\omega}$ through $C'_3$ is the total circulation *k* of the vortex. Thus a ***localized non potential force***, (one where we can draw a circuit like $C'_3$ that is in the fluid but outside the force) ***cannot change the circulation $\kappa$ of the vortex line***.

If $\vec{g}$ cannot change the circulation $\kappa$, what does it do? To find out we look more closely at the paths $C'_1$ and $C'_2$ lying above and below the line. We saw in Equation (80) that $\oint \vec{g} \cdot d\vec{\ell}$ was a positive number for the upper path $C'_1$. Thus $\vec{g}$ must be causing an increase in the flux of $\vec{\omega}$ through the upper path.

When we integrate $\vec{g}$ around the lower path $C'_2$, we get zero except where the path comes back through the core, in a direction opposite to $\vec{g}$, making $\oint \vec{g} \cdot d\vec{\ell}$ negative there. As a result

$$\oint_{C'_2} \vec{g} \cdot d\vec{\ell} \;=\; \text{negative number} \qquad (82)$$

and we find that $\vec{g}$ is causing a decrease in the flux of $\vec{\omega}$ through the lower path.

What does it mean when we see that $\vec{g}$ is causing the flux of $\vec{\omega}$ to decrease in the lower path, increase in the upper path, but not change the total flux of the core? It means that $\vec{g}$ is *causing the vortex line to move upward*. Since the paths $C'_1$ and $C'_2$ are attached to the fluid particles, the flow of $\vec{\omega}$ from the lower path to the upper path represents an upward motion of the vortex line ***relative to the fluid particles***. Thus the non potential force $\vec{g}$ causes a relative motion of the vortex lines and the fluid particles, a relative motion that is absent if there are no non potential forces acting on the fluid.

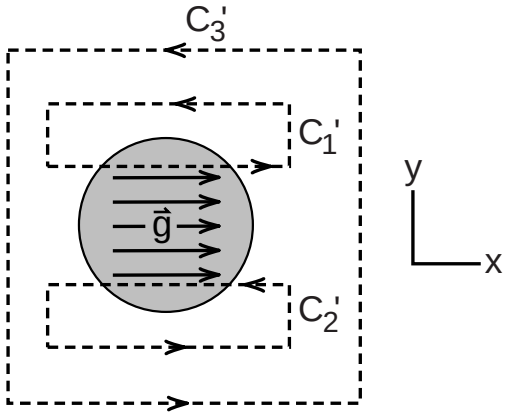



**Figure 7**
***An external force $\vec{g}$ is applied to the fluid in the core of a vortex. We see that the $\oint \vec{g} \cdot d\vec{\ell}$ is positive around the upper path $C'_1$, meaning that flux of $\vec{\omega}$ is increasing through that path. The integral is negative through the lower path $C'_2$ meaning that flux of $\vec{\omega}$ is decreasing there. This results in an upward flow of vorticity. Since $\oint \vec{g} \cdot d\vec{\ell} = 0$ for the big path surrounding the entire core, the total flux, or total circulation $\kappa$, is unchanged.***

This relative motion of the vortex line is sketched in Figure (8), where we designate the relative velocity by the vector $\vec{v}_{rel}$. Note that the motion is gyroscope like; when we push in the $\hat{x}$ direction on a $\hat{z}$ oriented vortex line, the line moves, not in the direction we push, but up in the $\hat{y}$ direction.

---

**Exercise 1**

Use Equation (78) and Figure (9) to show that the vortex line has no relative velocity in the direction that $\vec{g}$ pushes on the fluid.

---

**Exercise 2**

What is the direction of the relative velocity $\vec{v}_{rel}$ if $\vec{g}$ is $\hat{x}$ directed as in Figure (8), but $\vec{\omega}$ points in the $-\hat{z}$ direction? (I.e., what happens if we reverse $\vec{\omega}$?) Explain using Equation (78).

---

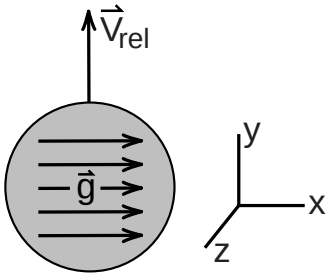



**Figure 8**
***The relative velocity $\vec{v}_{rel}$ of the vortex caused by the non potential force $\vec{g}$.***

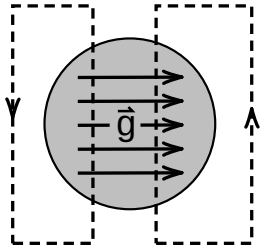



**Figure 9**
***Paths for determining the relative motion of the line in the direction of the force $\vec{g}$.***

## Motion of Charged Vortex Rings

Now that we have some idea of the effect of a localized force acting on a vortex line, let us return to our discussion of the Rayfield-Reif experiment.

As we mentioned, Rayfield and Reif created charged vortex rings in superfluid helium by placing a radioactive substance in the superfluid that emitted charged particles, either an electron or a proton depending on the substance. They ended up with charged objects in the superfluid, objects whose motion they could control using electric fields, and whose speed they could measure by timing a pulse of the particles moving between two grids.

But how could they know that the charged objects in the superfluid were actually vortex rings? The objects were tiny, carrying the charge of only one proton or one electron. In addition the core of a quantum vortex is of the order of an atomic diameter, so that the rings they were dealing with could be as small as only a few tens of atomic diameters. How could they be sure that these objects, that were much too small to be seen, were actually vortex rings?

The answer was in the peculiar behavior of these objects, a behavior only exhibited by vortex rings. The more they accelerated these objects, ***the harder they pushed on them,*** with an electric field, ***the slower they went!*** The reason for this behavior follows directly from the extended Helmholtz equation, Equation (78).

In Figure (10) we show the cross section of a vortex ring moving to the right, down the x axis. This is essentially Figure (12-15) of the last chapter, which shows how the velocity field of the top half of the ring pushes the bottom half forward, while the velocity field of the bottom half pushes the top half forward. Because the velocity decreases as we go away from the core, the bigger the ring becomes, the farther the halves are apart, the slower the ring moves.

In Figure (11), we show the same vortex ring, but now we are assuming that there is a charged fluid in the core, and an external $\hat{x}$ directed electric field is pushing on this charged fluid. It looks like we are attempting to accelerate the ring by pushing on it in the direction of its motion.

To see what this force does, we go back to Figure (8) and see that the $\hat{x}$ directed force $\vec{g}$ acting on the fluid in a $+\hat{z}$ oriented core causes the core to move up in the $+\hat{y}$ direction. At the bottom of the ring where the vorticity points in the opposite direction the same $\hat{x}$ directed force causes the core to move down (see Exercise 2). Overall the force $\vec{g}$ is causing the entire ring to grow in size, which results in the ring moving more slowly.

Thus we have the peculiar phenomenon that when we push on a ring in the direction the ring is moving, we make the ring bigger and slow it down. In Exercise (3), you show that if you push opposite to the direction of motion of the ring, you make the ring smaller and faster.

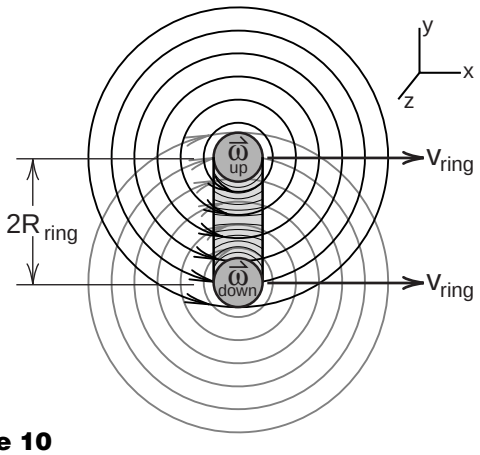



**Figure 10**
***Cross section of a vortex ring. Each side of the ring moves the other side forward. The smaller the ring, the greater the velocity field, and the faster the ring moves.***

**Exercise 3**
Using Equation (78), show that when you push opposite to the direction of motion of the ring you speed it up.

## Conservation of Energy

At first sight you might think you have a problem with the law of conservation of energy when it comes to the behavior of vortex rings. When we push on an object in the direction that it is moving, we are doing positive work on the object, and expect that, in the absence of friction, the energy of the object would increase. But for a vortex ring, when we push in the direction of the ring's motion the ring slows down. Does the ring loose energy as a result?

No. Unlike baseballs and other objects we are familiar with, a vortex ring's kinetic energy increases when it slows down. That is because its diameter increases and thus there is more length of vortex line. The kinetic energy of the ring is the kinetic energy $1/2\ mv^2$ of the fluid particles whose motion is caused by the ring. The larger the ring, the more fluid involved in the vortex motion, and the more kinetic energy associated with the ring. Thus pushing on a ring in the direction of motion increases its energy, as it should.

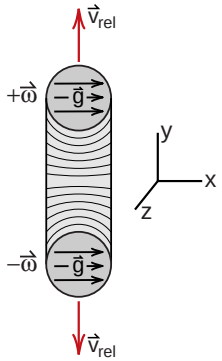



**Figure 11**
***An $\vec{x}$ directed force acting on a ring moving in the $\vec{x}$ direction causes the ring to expand.***

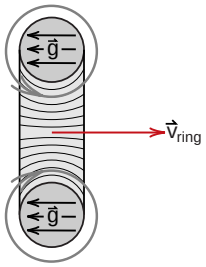



**Figure 12**
***Pushing opposite to the direction of motion of the ring.***

## Measurement of the Quantized Circulation $\kappa = h/m_{He}$

We have mentioned that Rayfield and Reif could control and measure the behavior of their charged vortex rings by sending pulses of the rings between grids in the superfluid. By timing the pulse, they could measure the speed of the rings. By applying a voltage difference to the grids, they could change the energy of the rings. A voltage difference $V_{voltage}$ would cause an energy change of magnitude $(eV_{voltage})$ for each ring because each ring carried either one proton of charge (+e) or one electron of charge (–e). We will give a rough argument as to how these two kinds of measurements allowed Rayfield and Reif to accurately measure the quantized circulation $\kappa = h/m_{He}$ of the ring.

We have noted that the energy of a ring is the kinetic energy $1/2\, mv^2$ of the fluid particles. Since the velocity field of a vortex is proportional to the vortex's circulation $\kappa$ ($\vec{v} = \kappa/2\pi r$ for a straight vortex), the fluid kinetic energy is proportional to $\kappa^2$. The fluid energy in a vortex ring is also proportional to the length $2\pi R$ of line in the ring. As a result the fluid kinetic energy is proportional to $\kappa^2 R_{ring}$

$$E_{ring} \propto \kappa^2 R_{ring} \qquad \begin{matrix}\textit{kinetic energy}\\ \textit{of a}\\ \textit{vortex ring}\end{matrix} \qquad (83)$$

---

**Exercise 4**

Show that $\rho\, \kappa^2 R_{ring}$ has the dimensions of kinetic energy.

---

We have seen that the velocity of a pair of oppositely oriented vortices is given by the formula

$$V_{2D\ ring} = \frac{\kappa}{4\pi R_{ring}} \qquad (12\text{-}40)$$

and have noted that the speed of a circular ring is roughly the same but more complex. In any case it is proportional to $\kappa/R_{ring}$

$$V_{ring} \propto \frac{\kappa}{R_{ring}} \qquad \begin{matrix}\textit{speed of}\\ \textit{vortex ring}\end{matrix} \qquad (84)$$

Neither Equation (83) or (84), or an accurate calculation of these quantities, can be used to measure the circulation $\kappa$ of the ring because you cannot see the rings to measure their radius $R_{ring}$. But in the product of the two terms, the unmeasurable term $R_{ring}$ cancels and we are left with the formula

$$E_{ring} \times V_{ring} \propto \kappa^3 \qquad (85)$$

Equation (85) suggests that an experimental measurement of $E_{ring} \times V_{ring}$ will give an experimental value of $\kappa^3$. A careful (and messy) calculation shows that both $E_{ring}$ and $V_{ring}$ have factors of the logarithm of the ring radius $R_{ring}$ divided by the core diameter (a). As a result there are factors of $\ln(R_{ring}/a)$ in a more accurate formula for the product $E_{ring} \times V_{ring}$. However this logarithm is quite insensitive to the actual value of $R_{ring}/a$ (increase the ring radius by 1000 and the logarithm $\ln(R_{ring}/a)$ increases only by an additional amount of 6.9). By making a number of measurements of $E_{ring} \times V_{ring}$, Rayfield and Reif were not only able to determine $\kappa$, but also the core diameter (a). That is when they found that the core diameter was roughly the diameter of a helium atom.

## The Magnus Equation

In Figure (8) repeated here we show a $\hat{z}$ directed vortex line, subjected to an $\hat{x}$ directed force, moving in the $\hat{y}$ direction. This motion labeled $\vec{V}_{rel}$ is the motion of the line relative to the fluid particles due to the non potential force $\vec{g}$. For the special case of a straight vortex, it is fairly easy to calculate the magnitude of this relative velocity $\vec{V}_{rel}$. The result we will call the ***Magnus equation***, named after a person who first studied sideways motion due to vortex effects.

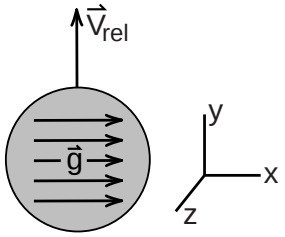



**Figure 8 (repeated)**
***The y directed motion of a z oriented vortex line subject to an x directed force.***

For this calculation, assume that we have a core of diameter D, with a uniform $\hat{z}$ directed vorticity $\vec{\omega}$ and an $\hat{x}$ directed force inside, as shown in Figure (13). We have drawn two paths $C'_1(t)$ and $C'_2(t)$ attached to the fluid particles. The circuits nearly touch each other so that half of the flux of $\vec{\omega}$ goes through $C'_1$ and half through $C'_2$ at the time (t).

A little time $\delta t$ later, the core has moved upward a distance $\delta y$ relative to the fluid particles as shown in Figure (13b). To keep the calculation simple, we will assume that the force $\vec{g}$ is strong enough to move the core up a reasonable distance $\delta y$ before the fluid has moved the circuits $C'_1$ and $C'_2$ noticeably. (The more accurate calculation in Appendix 2 does not make this assumption, but gets the same answer.)

Because the vorticity is moving up relative to the fluid particles, and thus up relative to the circuits $C'_1$ and $C'_2$, by the time $(t+\delta t)$ we have an additional band of flux of area $(D\,\delta y)$ through circuit $C'_1$. Thus the increase $\delta\Phi_1$ of flux in circuit $C'_1$, as we go from (t) to $(t+\delta t)$, is

$$\delta\Phi_1 = \omega(D\,\delta y) \tag{86}$$

Applying our vortex dynamics Equation (78) to the upper circuit $C'_1$, we have

$$\frac{\delta\Phi_1}{\delta t} = \oint_{C'_1} \vec{g}\cdot d\vec{\ell} \qquad \begin{matrix}\text{rate of increase}\\ \text{of flux of } \vec{\omega}\\ \text{through } C'_1\end{matrix} \tag{87}$$

Looking at Figure (13a) we see that the only contribution we get to $\oint \vec{g}\cdot d\vec{\ell}$ around $C'_1$ is through the center of the core, where $\vec{g}$ acts for a distance D, giving

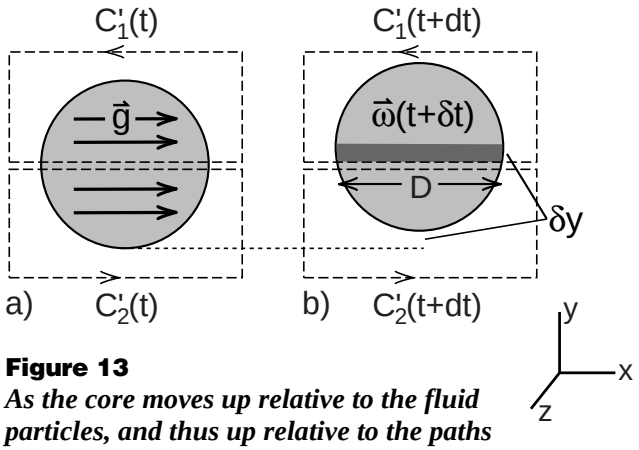



**Figure 13**
***As the core moves up relative to the fluid particles, and thus up relative to the paths $C'_1$ and $C'_2$ attached to the fluid particles, we get at time $(t+\delta t)$ an additional band of flux of area $(D\,\delta y)$ in circuit $C'_1$.***

$$\oint_{C'_1} \vec{g}\cdot d\vec{\ell} = gD \tag{88}$$

Thus

$$\frac{\delta\Phi_1}{\delta t} = gD \;;\quad \delta\Phi_1 = gD\delta t \tag{89}$$

Equating the values of $\delta\Phi_1$ from Equations (86) and (89) gives

$$\delta\Phi_1 = \omega D\,\delta y = gD\,\delta t \tag{90}$$

The D's cancel, and we are left with

$$g = \omega\frac{\delta y}{\delta t} = \omega V_{rel} \tag{91}$$

where $V_{rel}$ is the relative velocity of the vortex core and the fluid particles.

Equation (91) can be put in a more useful form if we multiply both sides by $\rho$, converting the force g per unit mass to $\rho g = f$, the force per unit volume. Then integrate f over the area of the core, giving us the ***force per unit length*** acting on the core. We get, using Equation (91) $g = \omega V_{rel}$,

$$F_e = \int_{\text{area of core}} \rho g\,dA = \rho\int_{\text{area of core}} (\omega V_{rel})\,dA$$

$$= \rho V_{rel}\int_{\text{area of core}} \omega\,dA \tag{92}$$

But the integral of $\omega$ over the area of the core is $\kappa$, the total circulation of the core. Thus Equation (92) becomes

$$F_e = \rho\kappa V_{rel} \tag{93}$$

The final step is to turn Equation (93) into a vector equation. We let the vector $\vec{\kappa} = \hat{z}\kappa$ point in the direction of the vorticity $\vec{\omega}$. The force $\vec{F}_e$ points in the $\hat{x}$ direction and $\vec{V}_{rel}$ is $\hat{y}$ directed. Using the right hand rule, we see that the cross product $\vec{V}_{rel}\times\vec{\kappa}$ points in the $\hat{x}$ direction like $\vec{F}_e$. Thus we have the vector equation

$$\boxed{\vec{F}_e = \rho\vec{V}_{rel}\times\vec{\kappa}} \qquad \textit{Magnus equation} \tag{94}$$

which is a remarkably simple result for what looked like a complex situation.

In Appendix 2 to this chapter, we derive an equation for the effect of non potential forces on curved fluid core vortices. The result looks exactly like Equation (94), but it tells us how to define $\vec{V}_{rel}$ when we have a curved vortex.

$$\vec{F}_e = \rho \vec{V}_{rel} \times \vec{\kappa} \qquad \text{(94) repeated}$$

When the exact formula is applied to a straight vortex in a two dimensional flow, the terms in Equation (94) have the following meaning. If $\hat{z}$ is the direction perpendicular to the flow, then $\vec{F}_e$ is the x-y component of the total force per unit length acting on the fluid in the core region. The component $(F_e)_z$ parallel to the vortex has no effect. The circulation $\vec{\kappa}$ is the total flux of $\vec{\omega}$ in the core, and is $\hat{z}$ oriented.

The relative velocity $\vec{V}_{rel}$ is given by the formula

$$\vec{V}_{rel} = \vec{V}_{vortex} - \vec{V}_{fluid} \qquad (95)$$

where the vortex velocity $\vec{V}_{vortex}$ is the velocity of the center of mass of the vorticity $\omega_z$, and the fluid velocity $\vec{V}_{fluid}$ is the weighted average of the fluid velocity $\vec{v}$ in the core region, given by the integral

$$\vec{V}_{fluid} = \frac{1}{\kappa}\int \omega_z \vec{v}\, dx\, dy \qquad (96)$$

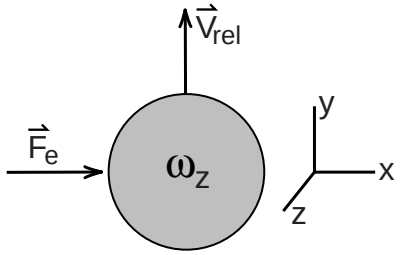



**Figure 14**

***Relative directions of $\vec{\omega}, \vec{F}_e$, and $\vec{V}_{rel}$.***

With these definitions, Equation (94) is an exact equation for a straight fluid core vortex. The result is independent of the shape of the core or the force density $\vec{g}$, as long as both are confined to a localized region.

The derivation of the exact Magnus equation, which we do in Appendix 2, is obtained by going back to Equation (55) and rewriting that equation as a continuity equation for the flow of vorticity. In some ways the continuity equation is simpler to derive and use than the Helmholtz theorem approach. But the continuity equation involves the quantity $\varepsilon_{ijk}$ which we introduce and use in Appendix 1 to derive various vector identities. Thus it seemed appropriate to delay a discussion of the continuity equation until after the reader has studied Appendix 1.

(The beginning of Appendix 2 gives a complete physical explanation of the continuity equation approach with virtually no mathematics and can be read at any time.)

## IMPULSE OF A VORTEX RING

Although we have discussed the Magnus equation $\vec{F} = \rho \vec{V}_{rel} \times \vec{\kappa}$ as applied to a straight vortex, the same ideas can be used for a curved vortex as long as the radius of curvature of the vortex is large compared to the core radius. When we apply the Magnus equation to a vortex ring, we get a simple formula relating the total force on the ring to the rate of change of the area of the ring. Introducing the concept of the ***impulse of a vortex ring***, we can write this formula so that looks a lot like Newton's law for vortex rings.

In Figure (15) we again show the cross section of a vortex ring, now showing the force $\vec{F}_e$ per unit length acting on each section of the core, and the relative velocity $\vec{V}_{rel}$ causing the ring to expand. For simplicity let $\vec{F}_e$ be in the direction of the motion of the ring, so that the Magnus equation implies

$$F_e = \rho \kappa V_{rel} \qquad (94a)$$

The velocity $V_{rel}$ is just the rate $dR_{ring}/dt$ that the ring radius is increasing. Thus Equation (94a) becomes

$$F_e = \rho \kappa \frac{dR_{ring}}{dt} \qquad (97)$$

The $F_e$ in Equation (97) is the force per unit length of the ring. The total length of the ring is its circumference $2\pi R_{ring}$, thus the total force $F_{total}$ is $2\pi R_{ring} F_e$, giving

$$F_{total} = 2\pi R_{ring} F_e = 2\pi \rho \kappa \left[ R_{ring} \frac{dR_{ring}}{dt} \right] \qquad (98)$$

However

$$R \frac{dR}{dt} = \frac{1}{2} \frac{d}{dt}(R^2) \qquad (99)$$

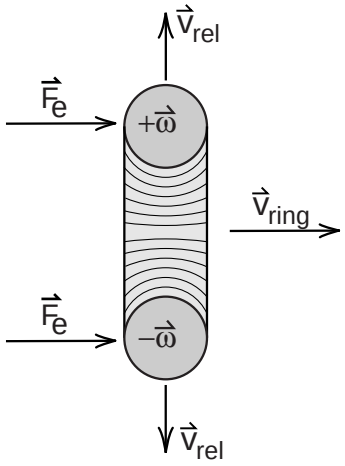



**Figure 15**
***An external force pushing on the ring in the direction of motion causes the ring to expand.***

Thus Equation (98) can be written in the form

$$F_{total} = \rho \kappa \frac{d}{dt}(\pi R_{ring}^2) \qquad (100)$$

But $\pi R_{ring}^2$ is just the area $A_{ring}$ of the ring, thus we get

$$F_{total} = \frac{d}{dt}(\rho \kappa A_{ring}) \qquad (101)$$

Let us define the vector $\vec{A}_{ring}$ as a vector of magnitude $\pi R_{ring}^2$, pointing in the direction of the motion of the ring. Then since the total force $\vec{F}_{total}$ also points in the same direction, we can write Equation (101) as the vector equation

$$\vec{F}_{total} = \frac{d}{dt}(\rho \kappa \vec{A}_{ring}) \qquad (102)$$

Of course we have derived Equation (102) only for the special case that $\vec{F}_{total}$ points in the direction the ring is moving. It becomes an interesting exercise with the vector form of the Magnus equation to show that Equation (102) applies for any direction of $\vec{F}_{total}$.

Equation (102) seems to look a lot like Newton's second law relating the total force $\vec{F}$ acting on a particle to the particle's momentum $\vec{p}$

$$\vec{F} = \frac{d\vec{p}}{dt} \qquad \textit{Newton's second law}$$

Equation (102) suggests that the quantity $\rho \kappa \vec{A}_{ring}$ plays a role for vortex rings similar to the role of momentum for particles. As a result it has become traditional to give $\rho \kappa \vec{A}_{ring}$ a special name, the ***impulse*** $\vec{I}$ of the ring

$$\vec{I} \equiv \rho \kappa \vec{A}_{ring} \qquad \begin{matrix} \textit{impulse of a} \\ \textit{vortex ring} \end{matrix} \qquad (103)$$

With Equation (103) the formula for $\vec{F}_{total}$ becomes

$$\vec{F}_{total} = \frac{d\vec{I}}{dt} \qquad \begin{matrix} \textit{impulse} \\ \textit{equation} \end{matrix} \qquad (104)$$

A common error one can make is to associate the impulse $\vec{I}$ of a vortex ring with an actual fluid momentum. Suppose, for example, you have a vortex ring in a sealed container. If you integrate $\rho \vec{v}$ for that ring over the entire fluid, the answer is ***zero***! In other words ***vortex rings do not carry linear momentum***. The impulse $\vec{I}$ is a separate quantity with its own special properties. One important property is that it makes it easy to predict the behavior of a ring subject to external forces. But it is not the momentum of the ring.

## THE AIRPLANE WING

In the fluid dynamics Chapter 23 of the Physics text, we used Bernoulli's equation to provide a qualitative view of why airplanes fly and sailboats can sail into the wind. In this section we will first look at the flow pattern of the fluid past an airplane wing, and see that for there to be lift, there has to be a net circulation of the fluid around the ring. This means that there is a vortex surrounding the wing. We then use the Magnus equation (95) to obtain a formula relating the weight of the airplane to the forward speed of the airplane and the circulation $\kappa$ of the vortex about the wing.

Figure (16) is a sketch of the streamlines we might expect for the flow of a fluid past an airplane wing. Our Bernoulli equation argument was that because the fluid was flowing faster over the top of the wing (where the streamlines are closer together) and slower under the wing, the pressure must be higher under the wing than on top so that the sum of the terms $(p + \rho v^2/2)$ be constant. (The $\rho g y$ term is too small to worry about for a fluid like air.) This higher pressure below suggests that the fluid is exerting a lift force on the wing. In Figure (16) we have drawn a circuit $C'$ around the wing. When we calculate the integral $\oint \vec{v}\cdot d\vec{\ell}$ around this circuit, we get a big positive contribution from the high speed fluid at the top, and a smaller negative contribution from the slow fluid at the bottom. Thus there is a net positive circulation $\vec{\kappa}$ surrounding the wing. In Figure (16), the circulation $\vec{\kappa}$ points in the $+\hat{z}$ direction. If there were no net circulation, if the fluid had the same speeds above and below the wing, there would be no lift.

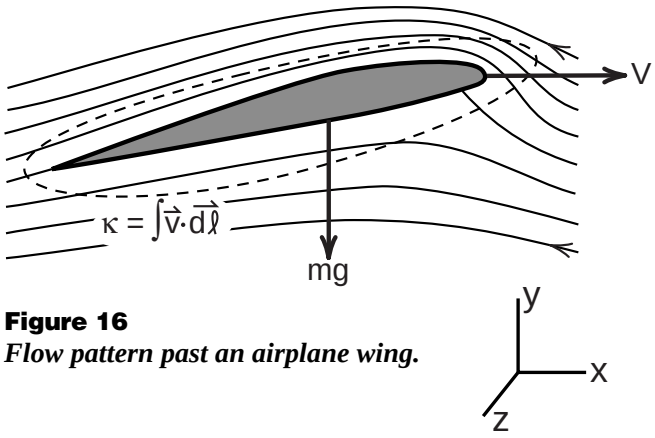



**Figure 16**
***Flow pattern past an airplane wing.***

Here is where we will adopt a rather unconventional view in order to directly apply the Magnus equation (94) to the airplane wing. We will picture the wing as being made of frozen fluid of the same density as the air flowing over it. This way we can think of the wing itself as part of the fluid, giving us a constant density, fluid core vortex to which we can apply Equation (95). Because the Magnus equation involves only the total circulation $\vec{\kappa}$ and not the details of the structure of the core, it makes no difference that our core now consists of a vortex sheet around the surface of the wing rather than the solid-body like rotation we assumed in our other vortex cores.

The purpose of the wing is to support the weight $m\vec{g}$ of the airplane. If we divide $m\vec{g}$ by the total length L of the wings, we get the downward, $-\hat{y}$ directed force $\vec{F}_g$ per unit length acting on the wings, and thus on the core of the wing vortex.

Here is the unconventional part of the argument. If you exert a downward, $-\hat{y}$ directed force on a $\hat{z}$ oriented vortex, you will get an $\hat{x}$ directed relative velocity of the core as shown in Figure (17). (Figure (17) is just Figure (8) rotated 90°.) Comparing Figures (16) and (17), we can say that the ***downward gravitational force*** on the wing, i.e., on the core of the vortex around the wing, is ***causing the wing vortex to move forward relative to the fluid*** through which the airplane is flying. The Magnus equation, with $\vec{F}_e = \vec{F}_g$ is

$$\vec{F}_g = \rho \vec{V}_{rel} \times \vec{\kappa} \qquad (104)$$

This gives us an explicit formula relating the downward gravitational force $\vec{F}_g$ per unit length, the circulation $\vec{\kappa}$ of the wing vortex, and the forward speed $\vec{V}_{rel}$ of the airplane.

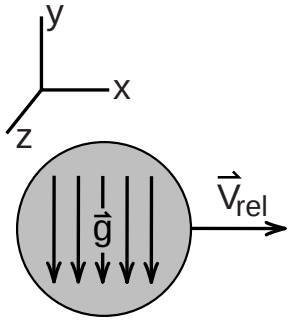



**Figure 17**
***Motion of a vortex subject to a localized force $\vec{g}$.***

The first thing this equation tells you is that there must be a vortex around the wing of an airplane for the airplane to fly. In addition, the vortex cannot stop at the end of the wing because vortex lines, being solenoidal ($\vec{\nabla}\cdot\vec{\omega} = 0$), cannot stop in the fluid. Instead the vortices trail back behind the airplane and are sometimes very visible during takeoff on a misty morning.

Equation (94) also tells us that for a given speed $\vec{V}_{rel}$, the heavier the airplane, i.e., the greater $\vec{F}_g$ is, the greater the circulation $\vec{\kappa}$ has to be. To lift the airplane, the circulation has to be particularly strong during takeoff where the forward velocity $\vec{V}_{rel}$ of the airplane is small. As a result the massive jumbo jets have strong wing tip vortices trailing after them, strong enough to flip small airplanes taking off behind them. Pilots of small aircraft are warned to stay clear of the jumbo jets.

We have just presented the rather different picture that the forward motion of an airplane is caused by the gravitational force acting down on the core of the wing vortex. When this point of view was presented in a science journal article, a reviewer replied that it was the airplane motors which pulled the airplane forward. Our response to that was—what about a glider that flies without motors? The main role of the motors in level flight is to overcome the viscous drag on the wings and fuselage.

Although it works well, our picture is still unconventional. When we used the Bernoulli argument in Chapter 23 of the Physics text, we were using the conventional picture that the fluid is exerting a lift force on the wing. The conventional derivation of the lift force involves calculating the momentum transfer between the fluid and the solid object. This is a somewhat messy calculation involving integration of pressure forces over the surface of the object. When you finish, you find that the lift force is proportional to the total circulation $\vec{\kappa}$ about the wing and the velocity $\vec{V}_{rel}$ of the wing relative to the fluid through which it is moving. Such a lift force on a moving vortex is called the ***Magnus Force***.

## The Magnus Lift Force

We are in a position to write down the formula for the lift force on an airplane wing without doing any pressure force integrations. Start with Equation (104)

$$\vec{F}_g = \rho\vec{V}_{rel}\times\vec{\kappa} \qquad \text{(104) repeated}$$

which relates the gravitational force $\vec{F}_g$ per unit length to the circulation $\vec{\kappa}$ and the relative velocity $\vec{V}_{rel}$ of the vortex. If the plane is in level flight, then the downward gravitational force $\vec{F}_g$ must be exactly balanced by the upward lift force $\vec{F}_{lift}$ for the plane not to rise or fall. Thus we have

$$\vec{F}_{lift} = -\vec{F}_g \qquad (105)$$

which gives us

$$\vec{F}_{lift} = -\rho\vec{V}_{rel}\times\vec{\kappa} \qquad (106)$$

In addition to airplane wings, spinning objects generally have a vortex around them. If the object is moving through the fluid at a velocity $\vec{V}_{rel}$, it will experience a sideways lift force given by Equation (106). This sideways lift force on a spinning object is called the Magnus force $\vec{F}_{magnus}$ after G. Magnus who studied the sideways motion of spinning objects in 1852*. The Magnus lift force formula found in textbooks is

$$\boxed{\vec{F}_{magnus} = -\rho\vec{V}_{rel}\times\vec{\kappa}} \qquad \textit{Magnus lift force formula} \qquad (107)$$

* "***On the deviation of projectiles; and on a remarkable phenomenon of rotating bodies.***" G. Magnus, Memoirs of the Royal Academy, Berlin(1852). English translation in Scientific Memoirs, London(1853)., p.210. Edited by John Tyndall and William Francis.

## The Magnus Force and Fluid Vortices

The extended Helmholtz theorem, Equation (78) and its application to the motion of vortex lines through a fluid, was developed in the 1960s to help understand vortex behavior in the Rayfield-Reif experiment. Before that, and still in most textbooks, the motion of vortices through a fluid is explained in the following way.

The Magnus force formula $\vec{F}_{magnus} = -\rho\vec{V}_{rel}\times\vec{\kappa}$ tells us the lift force on a solid object moving through a fluid at a velocity $\vec{V}_{rel}$, when there is a circulation $\kappa$ about the object.

If one has a fluid core vortex moving relative to the fluid, one says that there must be a lift force $\vec{F}_{magnus} = -\rho\vec{V}_{rel}\times\vec{\kappa}$ acting on that vortex. But there is no extra mass associated with a fluid core vortex, so one must treat the vortex as a massless object, with the result that the net force on the vortex must be zero. That means that there must be an external force $\vec{F}_{external}$ acting on the vortex to cancel the Magnus lift force. That is, one must have

$$\vec{F}_{external} + \vec{F}_{magnus} = 0 \qquad (108)$$

Using the Magnus formula (107) in (108) gives

$$\vec{F}_{external} = -\vec{F}_{magnus} = \rho\vec{V}_{rel}\times\vec{\kappa} \qquad (109)$$

This is just our Equation (95) relating the relative motion of a vortex to the localized, non potential force on the core of the vortex.

What we have shown, by deriving Equation (109) directly from the Navier-Stokes equation, which itself came from Newton's second law, is that we can describe vortex motion without any reference whatsoever to a Magnus lift force. The Magnus force is a ***pseudo force***, which like the centrifugal force, may be very useful for calculation, but which has no place in a basic description of the motion of the fluid itself.

# Appendix for Chapter 13

# Part 1

# Component Notation and the Functions $\delta_{ij}$ and $\varepsilon_{ijk}$

*In our derivation of the Navier-Stokes equation we ran into the term* $\nabla_i(\vec{p}\,v_i)$ *which we could not handle very well with vector notation like* $\vec{\nabla}\cdot\vec{v}$ *or* $\vec{\nabla}\phi$. *To handle this term we resorted to component notation* $\nabla_i$ *and* $v_i$, *and introduced the Einstein summation convention. Here we will briefly review the summation convention, and then discuss two quantities* $\delta_{ij}$ *and* $\varepsilon_{ijk}$ *that play basic roles when we work with dot and cross products in component notation. These quantities also become extremely useful when we are working out vector identities, like the relationship*

$$(\vec{v}\cdot\vec{\nabla})\,\vec{v} \;=\; \vec{\nabla}\left(\frac{v^2}{2}\right) - \vec{v}\times(\vec{\nabla}\times\vec{v}) \qquad (13\text{-}33)$$

*which we used to get the* $v^2/2$ *term in Bernoulli's equation.*

## THE SUMMATION CONVENTION

In Equation (12) of this chapter we wrote the dot product of two vectors $\vec{a}\cdot\vec{b}$ in the following three forms

$$\begin{aligned}\vec{a}\cdot\vec{b} &= a_x b_x + a_y b_y + a_z b_z \\ &= \sum_{i=x,y,z} a_i b_i \\ &= a_i b_i\end{aligned} \qquad (13\text{-}12)$$

With the summation convention, when we have repeated indices like $a_i b_i$, it is understood that we are to sum over all values of the repeated index i. We gave as an example

$$a_i b_j c_i = a_x b_j c_x + a_y b_j c_y + a_z b_j c_z$$

where we summed over the repeated index i, but the single index j was not summed. In mixed index-vector notation, $a_i b_j c_i$ could be written

$$(a_i \vec{b} c_i)_j = a_i b_j c_i \qquad (1)$$

## THE DOT PRODUCT AND $\delta_{ij}$

We will see that the quantity $\delta_{ij}$, defined by the simple relationship

$$\begin{aligned}\delta_{ij} &= 1 \quad \text{if } i = j \\ &= 0 \quad \text{if } i \neq j\end{aligned} \qquad (2)$$

is closelly related to the dot product in component notation. Consider the term

$$\delta_{ij} a_i b_j \qquad (3)$$

Here both indices i and j are repeated, so that we have to sum over both to get

$$\begin{aligned}\delta_{ij} a_i b_j = {} & \delta_{xx} a_x b_x + \delta_{xy} a_x b_y + \delta_{xz} a_x b_z \\ & + \delta_{yx} a_y b_x + \delta_{yy} a_y b_y + \delta_{yz} a_y b_z \\ & + \delta_{zx} a_z b_x + \delta_{zy} a_z b_y + \delta_{zz} a_z b_z\end{aligned} \qquad (4)$$

In Equation (4), the only non zero $\delta_{ij}$ terms are $\delta_{xx}$, $\delta_{yy}$ and $\delta_{zz}$, leaving

$$\delta_{ij} a_i b_j = \delta_{xx} a_x b_x + \delta_{yy} a_y b_y + \delta_{zz} a_z b_z \qquad (5)$$

Since $\delta_{xx} = \delta_{yy} = \delta_{zz} = 1$, we get

$$\delta_{ij} a_i a_j = a_x b_x + a_y b_y + a_z b_z = \vec{a}\cdot\vec{b} \qquad (6)$$

In component notation this can be written

$$\boxed{\delta_{ij} a_i b_j = a_j b_j = \vec{a}\cdot\vec{b}} \qquad (7)$$

You can see that the function $\delta_{ij}$ turns the product of two vectors $a_i$ and $b_j$ into a dot product.

Another way of handling $\delta_{ij} a_i b_j$ is to first work out the effect of $\delta_{ij}$ acting on $a_i$. Setting the index j to x we have

$$\delta_{ix} a_i = \delta_{xx} a_x + \delta_{xy} a_y + \delta_{xz} a_z = a_x$$

Similarly we get

$$\delta_{iy} a_i = a_y$$

$$\delta_{iz} a_i = a_z$$

Thus for any value of j, $\delta_{ij} a_i$ is equal to $a_j$

$$\boxed{\delta_{ij} a_i = a_j} \qquad (8)$$

Then when we want to evaluate the product $\delta_{ij} a_i b_j$ we can write

$$(\delta_{ij} a_i) b_j = (a_j) b_j = \vec{a}\cdot\vec{b} \qquad (9)$$

## THE CROSS PRODUCT AND $\varepsilon_{ijk}$

We just saw that $\delta_{ij}$ turned the product of two vectors $a_i$ and $b_j$ into a dot product $\vec{a}\cdot\vec{b}$. We will now see that a slightly more complex function $\varepsilon_{ijk}$ turns the product of two vectors $a_j b_k$ into a cross product $\vec{a}\times\vec{b}$

The cross product $\vec{a}\times\vec{b}$ of two vectors is given by

$$\begin{aligned}(\vec{a}\times\vec{b})_x &= a_y b_z - a_z b_y \\ (\vec{a}\times\vec{b})_y &= a_z b_x - a_x b_z \\ (\vec{a}\times\vec{b})_z &= a_x b_y - a_y b_x\end{aligned} \qquad (10)$$

We will see that this can all be written as the one equation

$$\boxed{(\vec{a}\times\vec{b})_i = \varepsilon_{ijk} a_j b_k} \qquad (11)$$

where the function $\varepsilon_{ijk}$ has the values

$$\begin{aligned}\varepsilon_{ijk} &= 0 \quad \textit{if any two indices are equal} \\ \varepsilon_{xyz} &= 1 \\ \varepsilon_{xzy} &= -1 \\ \varepsilon_{zxy} &= +1 \\ &\ldots\end{aligned} \qquad (12)$$

What we are indicating by the dots is that if you permute (interchange) any two neighboring indices, you change the sign.

For example, what is the sign of $\varepsilon_{zyx}$? To find out we do the following permutations starting with $\varepsilon_{xyz} = +1$

$$\begin{aligned}\varepsilon_{xyz} &= +1 \\ \varepsilon_{xzy} &= -1 \\ \varepsilon_{zxy} &= +1 \\ \varepsilon_{zyx} &= -1\end{aligned} \qquad (13)$$

It does not matter how you do the permutation you always come out with the same answer. For example

$$\begin{aligned}\varepsilon_{xyz} &= +1 \\ \varepsilon_{yxz} &= -1 \\ \varepsilon_{yzx} &= +1 \\ \varepsilon_{zyx} &= -1\end{aligned} \qquad (13a)$$

Because of this permutation property, $\varepsilon_{ijk}$ is often called the ***permutation tensor***. (A ***tensor*** is a vector like object with more than one index.)

Now we have to check that Equation (11), using $\varepsilon_{ijk}$ for the cross product, gives the correct result. Using the summation convention and crossing out terms like $\varepsilon_{xxk}$ which are zero, we have

$$\begin{aligned}(\vec{a}\times\vec{b})_x &= \varepsilon_{xjk} a_j b_k \\ &= \cancel{\varepsilon_{xxk} a_x b_k} + \varepsilon_{xyk} a_y b_k + \varepsilon_{xzk} a_z b_k \\ &= \cancel{\varepsilon_{xyx} a_y b_x} + \cancel{\varepsilon_{xyy} a_y b_y} + \varepsilon_{xyz} a_y b_z \\ &\quad + \cancel{\varepsilon_{xzx} a_z b_x} + \varepsilon_{xzy} a_z b_y + \cancel{\varepsilon_{xzz} a_z b_z}\end{aligned}$$

$$(\vec{a}\times\vec{b})_x = \varepsilon_{xyz} a_y b_z + \varepsilon_{xzy} a_z b_y \qquad (14)$$

With $\varepsilon_{xyz} = +1$, $\varepsilon_{xzy} = -1$ (one permutation), we get

$$(\vec{a}\times\vec{b})_x = a_y b_z - a_z b_y \qquad (15)$$

Which is the correct answer.

---

**Exercise 1**

Check that

$$(\vec{a}\times\vec{b})_y = e_{yjk} a_j b_k$$

---

As an example of the use of the $\varepsilon_{ijk}$, let us prove the vector identity

$$\vec{A}\cdot(\vec{B}\times\vec{C}) = (\vec{A}\times\vec{B})\cdot\vec{C} \qquad (13\text{-}68)$$

which we used in the derivation of the Helmholtz theorem. We have

$$\vec{A}\cdot(\vec{B}\times\vec{C}) = A_i(\vec{B}\times\vec{C})_i$$
$$= A_i\varepsilon_{ijk}B_jC_k$$

$$\boxed{\vec{A}\cdot(\vec{B}\times\vec{C}) = \varepsilon_{ijk}A_iB_jC_k} \qquad (16)$$

$$(\vec{A}\times\vec{B})\cdot\vec{C} = (\vec{A}\times\vec{B})_iC_i$$
$$= \varepsilon_{ijk}A_jB_kC_i$$

$$\boxed{(\vec{A}\times\vec{B})\cdot\vec{C} = \varepsilon_{ijk}A_jB_kC_i} \qquad (17)$$

To show that Equation (17) is equivalent to (16), we will first rename the indices in Equation (16). We will do this in two steps to avoid any possible errors. Changing $i \to r$, $j \to s$, $k \to t$ in Equation (16) gives

$$\varepsilon_{ijk}A_iB_jC_k = \varepsilon_{rst}A_rB_sC_t \qquad (16a)$$

We can do this because it does not matter what letter we use for a repeated index. Now we wish to rename the indices again so that the vector components in Equation (16) match those in (17). If we substitute $r \to j$, $s \to k$, $t \to i$, Equation (16a) becomes

$$\varepsilon_{rst}A_rB_sC_t = \varepsilon_{jki}A_jB_kC_i \qquad (16b)$$

which when combined with (17a) gives

$$\varepsilon_{ijk}A_iB_jC_k = \varepsilon_{jki}A_jB_kC_i \qquad (16c)$$

With some practice, you will not bother going through steps (16a) and write (16b) directly.

We now have

$$\vec{A}\cdot(\vec{B}\times\vec{C}) = \varepsilon_{jki}A_jB_kC_i \qquad (16d)$$

$$(\vec{A}\times\vec{B})\cdot\vec{C} = \varepsilon_{ijk}A_jB_kC_i \qquad (17) \text{ repeated}$$

The vector components now match, and what we now have to do is see how $\varepsilon_{jki}$ compares with $\varepsilon_{ijk}$. We will start with $\varepsilon_{ijk}$ and see how many permutations it takes to get to $\varepsilon_{jki}$. We have

$$\varepsilon_{jik} = -\varepsilon_{ijk}$$

$$\varepsilon_{jki} = -\varepsilon_{jik} = -(-\varepsilon_{ijk})$$

Two permutations are required, we have $\varepsilon_{jki} = \varepsilon_{ijk}$, and thus the terms in (16) and (17) are equal, which proves the identity.

While these steps may have looked a bit complex the first time through, with some practice they are much easier, faster, and more accurate than writing out all the x, y, and z components of the cross products.

## Handling Multiple Cross Products

To work out vector identities involving more than one cross product, there is a special identity that is worth memorizing. It is

$$\boxed{\varepsilon_{ijk}\varepsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}} \qquad (18)$$

First of all, note that Equation (18) has the correct symmetry. It must change sign on the right if you permute (interchange) i and j or l and m, because that is what $\varepsilon_{ijk}$ and $\varepsilon_{klm}$ do on the left side. This combination of $\delta$ functions has that property.

Before we try to prove Equation (18), we will give an example of how useful it is. Consider the rather messy set of cross products $\vec{a}\times(\vec{b}\times\vec{c})$. Using the $\varepsilon_{ijk}$ notation for cross products, we have

$$\begin{aligned}\left[\vec{a}\times(\vec{b}\times\vec{c})\right]_i &= \varepsilon_{ijk}a_j(\vec{b}\times\vec{c})_k \\ &= \varepsilon_{ijk}a_j\varepsilon_{klm}b_l c_m \\ &= (\varepsilon_{ijk}\varepsilon_{klm})a_j b_l c_m\end{aligned} \qquad (19)$$

Using Equation (18) we get

$$\left[\vec{a}\times(\vec{b}\times\vec{c})\right]_i = (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})\,a_j b_l c_m \qquad (20)$$

We will get some practice with the use of the $\delta$ functions $\delta_{ij}$. We have for example

$$\delta_{il}b_l = b_i \;\;;\;\; \delta_{jm}c_m = c_j \qquad (21)$$

So that

$$\delta_{il}\delta_{jm}a_j b_l c_m = a_j b_i c_j$$

and

$$-\delta_{im}\delta_{jl}a_j b_l c_m = -a_j b_j c_i$$

We get the result

$$\boxed{\left[\vec{a}\times(\vec{b}\times\vec{c})\right]_i = a_j b_i c_j - a_j b_j c_i} \qquad (22)$$

To apply Equation (22) to the problem we had with the Navier-Stokes equation, let

$$\vec{a} = \vec{v} \;\;;\;\; \vec{b} = \vec{\nabla} \;\;;\;\; \vec{c} = \vec{v} \qquad (23)$$

giving

$$\left[\vec{v}\times(\vec{\nabla}\times\vec{v})\right]_i = v_j\nabla_i v_j - v_j\nabla_j v_i \qquad (24)$$

By not changing the order of the vectors in Equation (22), the equation can be used when one or more of the vectors are the gradient vector $\vec{\nabla}$.

To get Equation (24) into the form we want, consider

$$\begin{aligned}\frac{1}{2}\nabla_i v^2 &= \frac{1}{2}\nabla_i(v_x^2 + v_y^2 + v_z^2) \\ &= \frac{1}{2}(2v_x\nabla_i v_x + 2v_y\nabla_i v_y + 2v_z\nabla_i v_z) \\ &= v_j\nabla_i v_j\end{aligned} \qquad (25)$$

Thus Equation (24) can be written

$$\left[\vec{v}\times(\vec{\nabla}\times\vec{v})\right]_i = \nabla_i\left(\frac{v^2}{2}\right) + v_j\nabla_j v_i \qquad (26)$$

To put this in pure vector notation, notice that Equation (26) is the (i)th component of the vector equation

$$\boxed{\vec{v}\times(\vec{\nabla}\times\vec{v}) = \vec{\nabla}\left(\frac{v^2}{2}\right) + (\vec{v}\cdot\vec{\nabla})\vec{v}} \qquad (27)$$

Equation (27) is equivalent to

$$(\vec{v}\cdot\vec{\nabla})\vec{v} = -\vec{\nabla}\left(\frac{v^2}{2}\right) + \vec{v}\times\vec{\omega} \qquad (28)$$

which we used to get the Bernoulli term $-\vec{\nabla}(v^2/2)$ into the Navier-Stokes equation.

## Proof of the ε Identity

We will use a rather brute force method to prove the identity

$$\varepsilon_{ijk}\varepsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \qquad \text{(18) repeated}$$

Let us consider the special case i = x and j = y. Then for the δ functions we get

$$\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} = \delta_{xl}\delta_{ym} - \delta_{xm}\delta_{yl} \qquad (29)$$

If l = x, m = y, get +1 from $\delta_{xl}\delta_{ym}$ (30a)

If l = y, m = x, get –1 from $-\delta_{xm}\delta_{yl}$ (30b)

All other values of l and m give zero (30c)

For this case i = x and j = y, the product of ε's, becomes

$$\varepsilon_{ijk}\varepsilon_{klm} = \varepsilon_{xyk}\varepsilon_{klm} \qquad (31)$$

The only non zero value for k is z giving

$$\varepsilon_{xyk}\varepsilon_{klm} = \varepsilon_{xyz}\varepsilon_{zlm} \qquad (32)$$

The only value of l and m that give a non zero result are l = x, m = y and l = y, m = x. For l = x , m = y, we get $\varepsilon_{xyz}\varepsilon_{zxy}$. Two permutations give

$$\varepsilon_{zxy} \Rightarrow -\varepsilon_{xzy} \Rightarrow \varepsilon_{xyz} = +1$$

Thus

$$\varepsilon_{xyk}\varepsilon_{klm} = +1 \qquad \textit{for } l = x, m = y \qquad (33a)$$

which agrees with Equation (30a).

For the case l = y, m = x, we get

$$\varepsilon_{xyk}\varepsilon_{klm} = \varepsilon_{xyz}\varepsilon_{zyx}$$

Now

$$\varepsilon_{zyx} \Rightarrow -\varepsilon_{zxy} \Rightarrow \varepsilon_{xzy} \Rightarrow -\varepsilon_{xyz} = -1$$

thus

$$\varepsilon_{xyz}\varepsilon_{zyx} = (+1)(-1) = (-1)$$

and we have

$$\varepsilon_{xyk}\varepsilon_{klm} = -1 \qquad \textit{for } l = y, m = x \qquad (33b)$$

which agrees with Equation (30b). All other values of l and m give zero, in agreement with (30c).

You can see that Equation (18) is correct for the special case i = x, j = y. In a few more pages of essentially identical work you can, if you want, show that Equation (18) works for any values of i and j. For practice, perhaps you might try a case like i = z, j = y.

# Appendix for Chapter 13

# Part 2

# **Vortex Currents**

*In the main part of Chapter 13, we derived the following equation that describes the behavior of vorticity in a constant density fluid.*

$$\frac{\partial\vec{\omega}}{\partial t} - \vec{\nabla}\times(\vec{v}\times\vec{\omega}) = \vec{\nabla}\times\vec{g}_{np} \qquad (13\text{-}55)$$

*It turns out that there are two rather different ways to handle this equation. The one we used in the main part of the chapter was to show that*

$$\frac{\delta\Phi(\vec{\omega})}{\delta t} = \int_{S'}\left[\frac{\partial\vec{\omega}}{\partial t} - \vec{\nabla}\times(\vec{v}\times\vec{\omega})\right]\cdot d\vec{A}$$

$$= \begin{array}{l}\textit{rate of change of the flux}\\ \textit{of } \vec{\omega} \textit{ through a circuit } S'\\ \textit{that moves with the fluid}\end{array} \qquad (13\text{-}74)$$

*Thus if $\vec{g}_{np} = 0$, there is no change in the flux and we have Helmholtz's theorem. If there is a change in flux, we have the relative motion of the vortex lines and fluid particles that we discussed in detail.*

*The other approach, which we discuss in this appendix, is to turn Equation (13-55) into a continuity equation for the flow of the vorticity field $\vec{\omega}$. The physical idea of how we get a continuity equation is very straightforward. The mathematics requires a fairly extensive use of the tensor $\varepsilon_{ijk}$ that we discussed in Appendix 1. That is why we have delayed the discussion of the flow of vorticity and vortex currents until this appendix.*

*Of the two approaches, the continuity equation approach is the more powerful. As we mentioned, it leads to an exact Magnus formula for curved fluid core vortices, a result that had not been obtained any other way. And the flow of vorticity, in the form of a vortex current tensor, appears to be playing a role in recent approaches to string theory.*

## CONSERVED TWO DIMENSIONAL CURRENTS

Before we go through any mathematical steps, let us look at the physical ideas of why we should expect to find a conserved flow of vorticity, and why working with a conserved flow might give us a simple way to handle the dynamics of the vorticity field.

In Figure (1a) we have sketched several vortices of rather arbitrary shape that we imagine are moving around in a constant density fluid. When we originally drew this diagram, we were thinking of quantized vortex lines moving around in superfluid helium. But it turns out that our analysis applies to ***tubes of flux*** for any solenoidal field, i.e., any field like $\vec{\omega}$ that has zero divergence. The significance of a solenoidal field is that the flux tubes cannot stop or start in the fluid. The tubes have no free ends in the fluid.

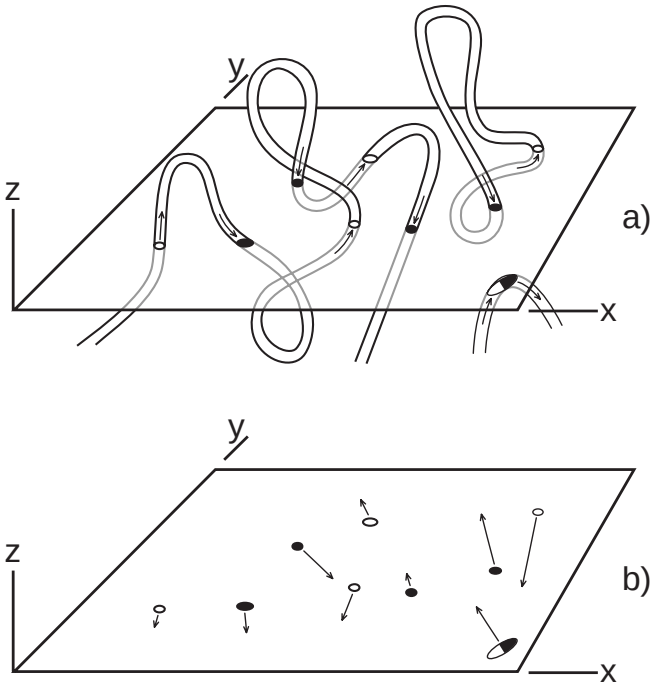



**Figure 1**

***If you slice the solenoidal vortex lines with an arbitrary xy plane, the circles, representing the intersection of the lines and the plane, form the objects of a conserved two dimensional current. When a loop pulls out of the plane, as in the lower right corner, two circles of opposite orientation annihilate each other. Circles can be created or annihilated only in pairs, or come in through the edges.***

In Figure (1a) we have also drawn a plane that cuts through these vortices. This is an arbitrary plane, slicing the fluid in any way we want. After drawing the plane, we then align the axis of our coordinate system so that the z axis is perpendicular to the plane. Thus we call this the $\hat{z}$ plane.

Where a vortex tube or line comes up through the plane, we have drawn a white circle, and where it goes down through—a black circle. Because the flux tubes of a solenoidal field cannot start or stop in the fluid, the circles in the $\hat{z}$ plane cannot appear or disappear one at a time. What can occur is that a loop may pull out of the plane as may be happening in the lower right hand corner. When this happens, a white circle and a black circle annihilate each other. If a loop enters the plane, we have the creation of a white circle-black circle pair.

If the plane extends well out beyond the region of the vortex lines, then we have a conservation law. The number of white circles minus the number of black circles is a constant.

We can go a step farther, and note that the circulation $\kappa$ of each vortex tube is given by the formula

$$\kappa = \int_{S'} \vec{\omega}\cdot d\vec{A} = \int_{\substack{\text{area of}\\ \text{intersection}}} \omega_z dA_z \qquad (1)$$

We get the same result for $\kappa$ no matter what $\hat{z}$ plane we use for integrating $\omega_z$, as long as the $\hat{z}$ plane cuts through the entire tube. As a result the white circles in Figure (1a) represent a net circulation $+\kappa$ and the black circles $-\kappa$. If all the flux tubes of $\vec{\omega}$ have the same circulation $\kappa$, then the total flux of $\vec{\omega}$ through the plane is simply $\kappa$ times the net number of circles, i.e., the number of white circles minus the number of black circles.

If the fluid is bounded, or the plane does not extend out beyond the region of the vortex lines, then the net number of circles can change by having vortex lines move in or out across the edges. Thus the more general conservation law is that ***the rate of change of the net number of circles in a given region of the plane is equal to the rate at which circles are flowing in or out across the edges of the region***. This is a verbal statement of a continuity equation for the flow of the black and white circles.

## CONTINUITY EQUATION FOR VORTICITY

To obtain the mathematical continuity equation for the flow of $\omega_z$, we start with the dynamic equation for vorticity, given by Equation (55) of Chapter 13 as

$$\frac{\partial\vec{\omega}}{\partial t} - \vec{\nabla}\times(\vec{v}\times\vec{\omega}) = \vec{\nabla}\times(\vec{g}_{np}) \qquad (13\text{-}55)$$

which obviously is equivalent to

$$\frac{\partial\vec{\omega}}{\partial t} = \vec{\nabla}\times(\vec{v}\times\vec{\omega} + \vec{g}_{np}) \qquad (13\text{-}55a)$$

In component notation this can be written as

$$\frac{\partial\omega_j}{\partial t} = \varepsilon_{jik}\nabla_i(\vec{v}\times\vec{\omega} + \vec{g}_{np})_k \qquad (2)$$

where $\varepsilon_{ijk}$ is the permutation tensor used in Appendix 1 to handle cross products. Using the fact that $\varepsilon_{jik} = -\varepsilon_{ijk}$, we get

$$\frac{\partial\omega_j}{\partial t} = -\nabla_i\Big[\,\varepsilon_{ijk}(\vec{v}\times\vec{\omega} + \vec{g}_{np})_k\,\Big] \qquad (3)$$

Rather than try to deal with all the components in Equation (3), let us look at the z component of the equation, which becomes

$$\frac{\partial\omega_z}{\partial t} = -\nabla_i\Big[\varepsilon_{izk}(\vec{v}\times\vec{\omega} + \vec{g}_{np})_k\Big] \qquad (4)$$

Defining the vector $\vec{j}(\omega_z)$ by the equation

$$\boxed{j(\omega_z)_i = \varepsilon_{izk}(\vec{v}\times\vec{\omega} + \vec{g}_{np})_k} \qquad (5)$$

we get the equation

$$\boxed{\frac{\partial\omega_z}{\partial t} = -\vec{\nabla}\cdot\vec{j}(\omega_z)} \qquad (6)$$

which has the form of a continuity equation if we interpret $\vec{j}(\omega_z)$ as the current vector for $\omega_z$.

This current vector $\vec{j}(\omega_z)$ has the very special property that it is ***two dimensional***; it has no $\hat{z}$ component. The formula for the z component is

$$j(\omega_z)_z = \varepsilon_{zzk}(\vec{v}\times\vec{\omega} + \vec{g}_{np})_k = 0 \qquad (7)$$

This is zero because $\varepsilon_{zzk} = 0$. Thus Equation (6) is the continuity equation for the two dimensional flow of $\omega_z$, which is exactly what we expected from our discussion of Figure (1b).

The formula for $\vec{j}(\omega_z)$ still needs some simplification. The first step is to write $\vec{v}\times\vec{\omega}$ in component notation to get

$$\varepsilon_{izk}(\vec{v}\times\vec{\omega})_k = \varepsilon_{izk}\varepsilon_{klm}v_l\omega_m$$

Next, use the relationship we proved in Appendix 1

$$\varepsilon_{ijk}\varepsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \qquad (A1\text{-}18)$$

to get

$$\begin{aligned}\varepsilon_{izk}(\vec{v}\times\vec{\omega})_k &= (\delta_{il}\delta_{zm} - \delta_{im}\delta_{zl})v_l\omega_m \\ &= v_i\omega_z - v_z\omega_i\end{aligned} \qquad (8)$$

The other simplification comes from noting that

$$(\hat{z}\times\vec{g}_{np})_i = \varepsilon_{ijk}\hat{z}_j(g_{np})_k = \varepsilon_{izk}(g_{np})_k \qquad (9)$$

where we set $\varepsilon_{ijk}\hat{z}_j = \varepsilon_{izk}$ because the unit vector $\hat{z}$ has only a z component.

In Equation (5) using Equation (8) for $\varepsilon_{izk}(\vec{v}\times\vec{\omega})_k$ and Equation (9) for $\varepsilon_{izk}(g_{np})_k$ to get

$$j(\omega_z)_i = v_i\omega_z - v_z\omega_i + (\hat{z}\times\vec{g}_{np})_i \qquad (10)$$

We can simplify the interpretation by introducing the notation

$$\vec{v} = (v_z\,,\,\vec{v}_{||}); \qquad \vec{\omega} = (\omega_z\,,\,\vec{\omega}_{||}) \qquad (11a)$$

where the vectors $\vec{v}_{||}$ and $\vec{\omega}_{||}$ are vectors representing the components of $\vec{v}$ and $\vec{\omega}$ parallel to the flow of $\omega_z$, i.e., components that lie in the $\hat{z}$ plane.

Since the current vector $\vec{j}(\omega_z)$ has no $\hat{z}$ component, it has only a parallel component

$$\vec{j}(\omega_z) = \vec{j}_{||}(\omega_z) \qquad (11b)$$

With this notation, we can let the index i be the parallel component in Equation (10), giving

$$\boxed{\vec{j}(\omega_z) = \vec{v}_{||}\omega_z - v_z\vec{\omega}_{||} + \hat{z}\times\vec{g}_{np}} \qquad (13)$$

Equation (13) is our final equation for the two dimensional current or $\omega_z$ in the $\hat{z}$ plane.

Roughly speaking, the terms in Equation (13), repeated below, have the following interpretation.

$$\vec{j}(\omega_z) = \vec{v}_{||}\omega_z - v_z\vec{\omega}_{||} + \hat{z}\times\vec{g}_{np} \quad \text{(13) repeated}$$

The $\vec{v}_{||}\omega_z$ term clearly represents the convection of $\omega_z$ due to the fluid motion $\vec{v}_{||}$ in the plane. The $\hat{z}\times\vec{g}_{np}$ term which we call the ***Magnus term***, gives us the sideways motion of the vortex when a non potential force is acting on the fluid. For example, if we have an $\hat{x}$ directed force $\vec{g}$ acting on the core of a $\hat{z}$ directed vortex, we end up with a $\hat{y}$ directed flow of vorticity as indicated in Figure (2), a diagram we have seen before.

The $-v_z\vec{\omega}_{||}$ term is more of a problem to interpret. We note, however, that for a two dimensional flow with straight vortices, we can orient the $\hat{z}$ plane to cut the vortex perpendicular to the core so that $\vec{\omega}_{||}$ is zero and the term vanishes. We will see that for three dimensional fluid flow with a curved vortex, this term can be made to go away by choosing a properly oriented $\hat{z}$ plane. From this point of view, the $-v_z\vec{\omega}_{||}$ term tells us which $\hat{z}$ plane to use.

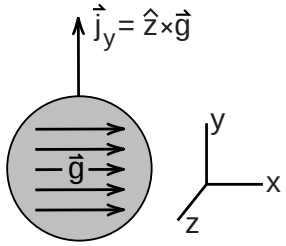



**Figure 2**
***Motion of a vortex line subject to an x directed force.***

## A SINGLE VORTEX LINE

To help interpret the equations for vortex motion, we will apply Equation (13) to the motion of a single vortex line. We cut the line with a $\hat{z}$ plane as shown in Figure (3a) and look at the behavior of $\omega_z$ in that plane, as seen in Figure (3b). The main result is that we end up with a formula for the motion of the center of mass of $\omega_z$. This result is a consequence of their being a conserved two dimensional current of $\omega_z$.

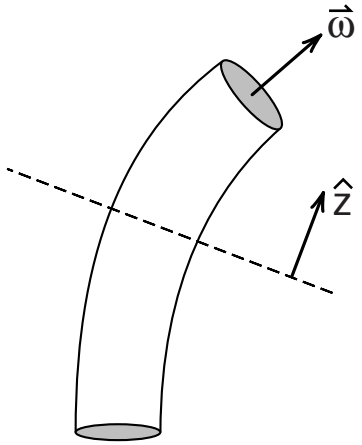



**Figure 3a**
***Cut the vortex line with a $\hat{z}$ plane.***

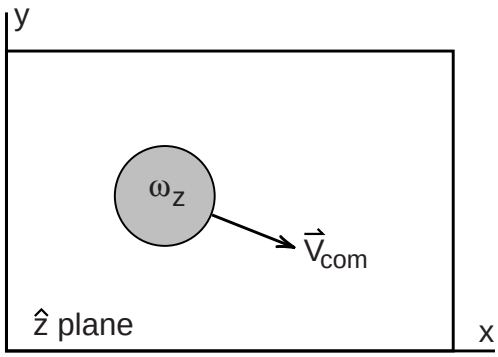



**Figure 3b**
***We will study the motion of $\omega_z$ in the $\hat{z}$ plane.***

## Center of Mass Motion

Our first step is to show that if we have an isolated vortex where both $\omega_z$ and non potential forces $\vec{g}_{np}$ are confined to a core region, then the vortex velocity $\vec{V}_{vortex}$, defined by

$$\vec{V}_{vortex} \equiv \frac{1}{\kappa}\int_{\substack{core\\area}} \vec{j}(\omega_z)dA_z = \vec{V}_{COM} \qquad (14)$$

is the velocity of the center of mass of $\omega_z$ in the $\hat{z}$ plane.

To show this, we begin with Figure (4) where we show the localized core area of a vortex as it passes through the $\hat{z}$ plane. We are assuming that the dotted rectangle from $x_a$ to $x_b$, and $y_a$ to $y_b$ lies outside the core area where both $\omega_z$ and $\vec{j}(\omega_z)$ are zero.

We define the area $\Delta A(y_i)$, seen in Figure (4), as a band of thickness $\Delta y$ that goes from $x_a$ to $x_b$, and from $y_i$ to $y_i+\Delta y$. The total vorticity $\Delta\kappa_i$ in this band is

$$\Delta\kappa_i = \Delta y\int_{x_a}^{x_b} \omega_z(x,y_i)dx \qquad (15)$$

The formula for the center of mass coordinate $\vec{R}_{COM}$ of a collection of masses $m_i$ is (see page 11-3 of the Physics text)

$$M\vec{R}_{COM} = \sum_i \vec{r}_i m_i \qquad (16)$$

where M is the total mass.

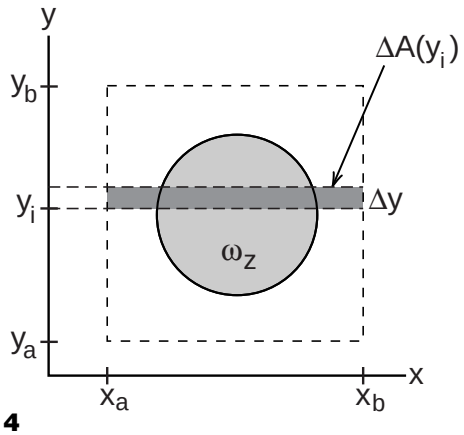



**Figure 4**
***Calculating the center of mass of $\omega_z$.***

Replacing M by the vortex total circulation $\kappa$, and $m_i$ by $\Delta\kappa_i$, the equation for the y component of the center of mass of the vorticity, $Y_{COM}$, becomes

$$\kappa Y_{COM} = \sum_i y_i\Delta\kappa_i \qquad (17)$$

Differentiating Equation (17) with respect to time, noting that the total circulation $\kappa$ does not change with time, gives

$$\kappa\frac{\partial Y_{COM}}{\partial t} = \kappa V_{yCOM} = \sum_i y_i\frac{\partial\Delta\kappa_i}{\partial t} \qquad (18)$$

Our problem now is to calculate the rate of change of the circulation $\Delta\kappa_i$ in our $\Delta y$ band. We do this by calculating the net rate of flow of vorticity into the band due to the vortex current $\vec{j}(\omega_z)$, indicated in Figure (5).

Along the line $y = y_i$, the net current into the band is

$$J_y(y_i) = \int_{x_a}^{x_b} j_y(x,y_i)dx \quad \textit{current in from below} \qquad (19)$$

where $j_y = j_y(\omega_z)$.

Up at $y_i+\Delta y$, the component $j_y(\omega_z)$ flows up out of the band, so that the net *inward* current up there has a minus sign

$$J_y(y_i+\Delta y)_{inward} = -J_y(y_i+\Delta y) = -\int_{x_a}^{x_b} j_y(x,y_i+\Delta y)dx \qquad (20)$$

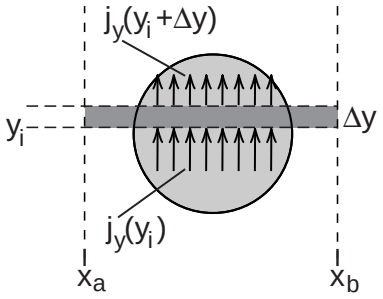



**Figure 5**
***Flow of vorticity into band.***

The total rate $\partial\Delta\kappa_i/\partial t$ at which vorticity is flowing into the band is thus

$$\frac{\partial\Delta\kappa_i}{\partial t} = -J_y(y_i+\Delta y) + J_y(y_i) \qquad (21)$$

Using Equation (21) in Equation (18) for $V_{yCOM}$ gives

$$\kappa V_{yCOM} = -\sum_i y_i \Delta y \left[\frac{J_y(y_i+\Delta y) - J_y(y_i)}{\Delta y}\right] \qquad (22)$$

where we multiplied the right hand side by $\Delta y/\Delta y$ .

In the limit $\Delta y \to 0$ , the square brackets become the derivative $\partial J_y(y)/\partial y$ , evaluated at $y = y_i$

$$\kappa V_{yCOM} = -\sum_i y_i \left[\left.\frac{\partial J_y(y)}{\partial y}\right|_{y=y_i}\right]\Delta y \qquad (23)$$

This sum $\sum_i \Delta y$ then becomes an integral from $y_a$ to $y_b$ , giving

$$\kappa V_{yCOM} = -\int_{y_a}^{y_b} y\frac{\partial J_y(y)}{\partial y}\,dy \qquad (24)$$

The next step, which is called ***integration by parts***, is a simple way to handle the factor y that appears in Equation (24). We note that, by the rules of differentiation

$$\frac{\partial}{\partial y}\left[yJ(y)\right] = \frac{\partial y}{\partial y}J(y) + y\frac{\partial J(y)}{\partial y} \qquad (25)$$

With $\partial y/\partial y = 1$ we get

$$y\frac{\partial J(y)}{\partial y} = \frac{\partial}{\partial y}\left[yJ(y)\right] - J(y) \qquad (26)$$

Substituting (26) into (24) gives

$$\kappa V_{yCOM} = -\int_{y_a}^{y_b} \frac{\partial}{\partial y}\left[yJ_y(y)\right]dy + \int_{y_a}^{y_b} J_y(y)dy \qquad (27)$$

We can explicitly carry out the first integral because the integral of a derivative is simply the function itself

$$\begin{aligned} -\int_{y_a}^{y_b} \frac{\partial}{\partial y}\left[yJ_y(y)\right]dy &= yJ_y(y)\Big|_{y_a}^{y_b} \\ &= y_bJ_y(y_b) - y_aJ_y(y_a) \\ &= 0 \end{aligned} \qquad (28)$$

We get zero because both $y_a$ and $y_b$ lie outside the core region, where $J_y$ is zero.

Thus we are left with

$$\begin{aligned} \kappa V_{yCOM} &= \int_{y_a}^{y_b} J_y(y)dy \\ &= \int_{y_a}^{y_b}\int_{x_a}^{x_b} j_y(x,y)dxdy \end{aligned} \qquad (29)$$

where we used Equation (19) to express $J_y(y)$ in terms of the vortex current density $j_y(x,y)$ .

Because we are assuming that $j_y(x,y)$ is non zero only over the core area, Equation (29) can be written in the more compact form

$$\kappa V_{yCOM} = \int_{\substack{\text{core}\\\text{area}}} j_y(\omega_z)dA_z \qquad (30)$$

where $dA_z = dxdy$ .

Similar arguments give

$$\kappa V_{xCOM} = \int_{\substack{\text{core}\\\text{area}}} j_x(\omega_z)dA_z \qquad (31)$$

Combining Equation (30) and (31), and dividing through by $\kappa$ gives

$$\vec{V}_{COM} = \frac{1}{\kappa}\int_{\substack{\text{core}\\\text{area}}} \vec{j}(\omega_z)dA_z \equiv \vec{V}_{vortex}$$

(14) repeated

which is the result we wanted to show.

## MAGNUS FORMULA FOR CURVED VORTICES

We are now ready to use Equation (13) to derive the Magnus effect formula for curved fluid core vortices. As a reminder, Equation (13) was

$$\vec{j}(\omega_z) = \vec{v}_{\|}\omega_z - v_z\vec{\omega}_{\|} + \hat{z}\times\vec{g}_{np} \quad \text{(13) repeated}$$

Slicing a curved vortex with a $\hat{z}$ plane as shown in Figure (3), integrating Equation (13) over the area of the core, and dividing through by $\kappa$ gives

$$\frac{1}{\kappa}\int_{\substack{\text{core}\\\text{area}}} \vec{j}(\omega_z)dA_z = \frac{1}{\kappa}\int \omega_z\vec{v}_{\|}dA_z \qquad (32a)$$

$$+ \frac{1}{\kappa}\int -v_z\vec{\omega}_{\|}dA_z \qquad (32b)$$

$$+ \frac{1}{\kappa}\int \hat{z}\times\vec{g}_{np}dA_z \qquad (32c)$$

We already know that the left side of Equation (32) is the vortex velocity $\vec{V}_{vortex}$. The first term on the right, which we will call $\vec{V}_{fluid}$

$$\vec{V}_{fluid} = \frac{1}{\kappa}\int \omega_z\vec{v}_{\|}dA_z \qquad (33)$$

is the weighted average of the velocity field $\vec{v}_{\|}$ in the core region.

As we mentioned earlier, the third term, the integral of $v_z\vec{\omega}_{\|}$ tells what $\hat{z}$ plane to use for the calculation. There will be some plane, more or less perpendicular to the core, which gives a zero value for the integral of $v_z\vec{\omega}_{\|}$ over the core. We will assume that we are using that $\hat{z}$ plane.

For this example, let us assume that $\vec{g}_{np}$ is an external force $\vec{g}_e$ acting on the fluid in the core, as sketched in Figure (2) repeated below. Multiplying this force per unit mass by $\rho$ gives $\vec{f}_e = \rho\vec{g}_e$ as the force per unit volume acting on the core. When $\vec{f}_e$ is integrated over the core, we get $\vec{F}_e$, the external force per unit length acting on the vortex.

With this notation the last term in Equation (32) becomes

$$\begin{aligned}\frac{1}{\kappa}\int_{\substack{\text{core}\\\text{area}}} \hat{z}\times\vec{g}_{np}dA_z &= \frac{1}{\rho\kappa}\hat{z}\times\int_{\substack{\text{core}\\\text{area}}} \rho\,\vec{g}_{np}dA_z \\ &= \frac{1}{\rho\kappa}\hat{z}\times\int_{\substack{\text{core}\\\text{area}}} \vec{f}_e dA_z \\ &= \frac{\hat{z}\times\vec{F}_e}{\rho\kappa}\end{aligned} \qquad (34)$$

Assuming we have chosen the correct $\hat{z}$ plane to eliminate the integral of $v_z\vec{\omega}_{\|}$, we get using Equations (14), (33) and (34) in Equation (32)

$$\vec{V}_{vortex} = \vec{V}_{fluid} + \frac{1}{\rho\kappa}\hat{z}\times\vec{F}_e \qquad (35)$$

The Helmholtz equation is now obtained by setting $\vec{F}_e = 0$ giving

$$\vec{V}_{vortex} = \vec{V}_{fluid} \qquad \begin{matrix}\textit{Helmholtz}\\ \textit{equation for}\\ \vec{F}_e = 0\end{matrix} \qquad (36)$$

In detail, Equation (36) says that when we choose the $\hat{z}$ plane correctly, the center of mass motion of the vortex core is equal to the weighted average of the fluid velocity in the core region.

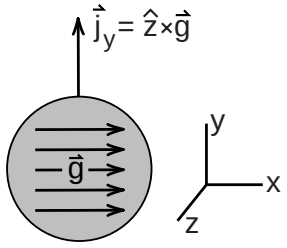



**Figure 2 (repeated)**
***Motion of a vortex line subject to an x directed force.***

When $\vec{F}_e$ is not zero and we have a relative motion of the vortex line and the fluid, we can define the relative motion vector $\vec{V}_{rel}$ as

$$\vec{V}_{rel} \equiv \vec{V}_{vortex} - \vec{V}_{fluid} \qquad (37)$$

and Equation (35) can be written

$$\vec{V}_{vortex} = \vec{V}_{fluid} + \frac{1}{\rho\kappa}\hat{z} \times \vec{F}_e \qquad \text{(35) repeated}$$

$$\hat{z} \times \vec{F}_e = \rho\kappa\vec{V}_{rel} \qquad (38)$$

We can get further insight from Equation (38) by writing $\vec{F}_e$ as

$$\vec{F}_e = (\vec{F}_{ez} + \vec{F}_{e\perp}) \qquad (39)$$

where $\vec{F}_{ez}$ is the component of $\vec{F}_e$ parallel to the $\hat{z}$ axis, and $\vec{F}_{e\perp}$ perpendicular to the z axis. Because $\hat{z}$ cross a vector parallel to $\hat{z}$ is zero, $\hat{z} \times \vec{F}_{ez} = 0$ and we get

$$\hat{z} \times \vec{F}_e = \hat{z} \times \vec{F}_{e\perp} \qquad (40)$$

Thus our final result for the Magnus equation is

$$\boxed{\hat{z} \times \vec{F}_{e\perp} = \rho\kappa\vec{V}_{rel}} \qquad \textit{Magnus equation} \qquad (41)$$

and we see that only the component of the external force perpendicular to the z axis, has an effect on the vortex motion. This reminds us why it is important, for a curved vortex, to find the correct $\hat{z}$ plane using the condition that the integral of $v_z\vec{\omega}_{||}$ be zero.

If we apply Equation (41) to a two dimensional flow in the xy plane, then the vorticity is automatically $\hat{z}$ directed and we can turn $\kappa$ into a $\hat{z}$ directed vector $\hat{\kappa}$. If the flow is to remain two dimensional, then the external force $\vec{F}_e$ must be in the xy plane, because a z component of $\vec{F}_e$ would create a $\hat{z}$ directed flow. Thus $\vec{F}_e$ must be $\vec{F}_{e\perp}$. With these restrictions, Equation (33) is equivalent to

$$\vec{F}_e = \rho\vec{V}_{rel} \times \vec{\kappa} \qquad (13\text{-}95)$$

which is our Equation (13-95) discussed in the regular part of the chapter. (Check for yourself that both Equations (41) and (13-95) predict that an $\hat{x}$ directed force $\vec{F}_e$ acting on a $\hat{z}$ directed vortex causes a $\hat{y}$ directed relative motion of the vortex.)

What we have learned from deriving the exact Magnus equation for curved vortices, that we cannot predict from a two dimensional derivation, is what component of $\vec{F}_e$ is important and exactly how $\vec{V}_{rel}$ is defined.

## CREATION OF VORTICITY

So far our emphasis has been on how non potential forces cause a relative motion of vortex lines and the fluid particles. But the vorticity we find in a fluid has to have been created somehow. Non potential forces do that, and we want to end this appendix with a brief discussion of how. The discussion is brief, because it is very incomplete. The creation of vorticity, which leads to turbulence, is not only a subject for an entire fluid dynamics textbook, it is also an active subject of current research. Here we will just indicate how the topic begins.

Non potential forces, at least in a constant density fluid like water, can create vorticity in two ways. One way is to pull it out of the walls of the container. Near the wall, where the velocity field rapidly goes to zero, we get a boundary layer where the non potential viscous forces are important. These viscous forces, if they are acting at the wall, will move vorticity out of the wall into the fluid. For example, this is how the vorticity in the smoke ring demonstration was created. Viscous forces acting on the high speed fluid at the perimeter of the hole in the box pulled a ring of vorticity in from the perimeter.

It turns out to be a tricky question of how viscous forces behave in a boundary layer. For laminar pipe flow, there are viscous forces acting at the wall continually pulling vorticity into the stream. In contrast, for a boundary layer solution called the ***Blasius profile***, the viscous forces act in the boundary layer but not at the wall. In that theory, the vorticity is all created upstream and all the viscous forces do is move the vorticity farther into the fluid, thickening the boundary layer. The velocity profiles near the wall look nearly the same for both laminar pipe flow and the Blasius profile, but the viscous effects are quite different. This indicates the kind of problem one has to deal with when working with boundary layers and the effects of viscosity.

Non potential forces can also create vorticity in the fluid away from the walls by creating vortex rings. In a sense, this is the way vorticity is created in the Rayfield-Reif experiment. To give you a rough classical picture of how a charged particle moving through a fluid could create a vortex ring, imagine that the charged particle, moving in what we will call the $\hat{z}$ direction exerts a local, more or less spherical shaped external force $\vec{g}$ on the fluid as shown in Figure (5). This looks much like the figure we have drawn so many times, except that there is no vortex line for $\vec{g}$ to push on. Thus $\vec{g}$ cannot be causing a relative motion of the line and the fluid. What it is doing instead is creating a vortex ring around the region.

We can see the ring creation by applying the extended Helmholtz equation (12-78) to the circuits $C'_1$, $C'_2$ and $C'_3$ shown in Figure (6). These circuits are moving with the fluid particles, and Equation (78) tells us that the rate of change of flux of $\omega$ through any of them is equal to $\oint \vec{g} \cdot d\vec{\ell}$ around the circuit. With this in mind, we see that the flux of $\omega$ through $C'_1$ is increasing because $\oint \vec{g} \cdot d\vec{\ell}$ is positive there, and it is decreasing through $C'_2$ where $\oint \vec{g} \cdot d\vec{\ell}$ is negative. Since $\oint \vec{g} \cdot d\vec{\ell}$ is zero for $C'_3$, there is no change in the flux of $\omega$ there.

What does it mean that $\vec{g}$ is decreasing the flux through the lower circuit $C'_2$ when there is no flux there to decrease? It means that $\vec{g}$ is creating negative flux of $\vec{\omega}$ through $C'_2$ while at the same time it is creating positive flux through $C'_1$. What it is doing is creating a ***band of flux*** of $\vec{\omega}$ around the spherical region, a band of flux that is becoming the core of a vortex ring.

Once vorticity has been introduced into the fluid, an effective method of introducing more vorticity is the stretching of existing vortex lines. How vortex line stretching affects fluid flows is a topic that has been studied for a long time by fluid engineers.

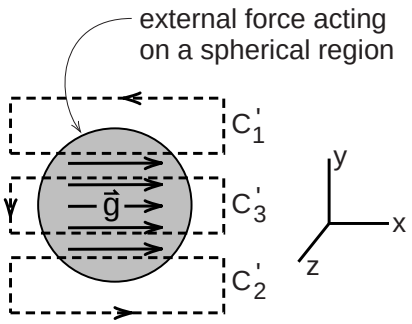



**Figure 6**
***External force creating a vortex ring.***

## ENERGY DISSIPATION IN FLUID FLOW

While a derivation of the Magnus formula for curved vortices demonstrates how mathematically effective the concept of a vortex current $\vec{j}(\omega_z)$ is, (the result has not been obtained any other way), the most important use so far of the concept is in studying the relationship between energy dissipation in a stream and the flow of vorticity across the stream. This relationship, discovered by Phillip Anderson in 1966, applies to such diverse situations as turbulent flow in a channel, and the motion of quantized vortices in both superfluids and superconductors. In the case of superconductors, the phenomenon is now involved in the legal definition of the electric volt.

We leave this topic for a later text, because one of the most interesting parts is to show how similar the vortex dynamics equations are for charged and neutral fluids. One can make the equations look identical by incorporating the magnetic field $\vec{B}$ in the definition of $\vec{\omega}$, and including the electric field $\vec{E}$ in $\vec{g}_{np}$. If you want to see this topic now, look at the article "Vortex Currents in Turbulent Superfluid and Classical Fluid Channel Flow . . .", Huggins, E.R., *Journal of Low Temperature Physics*, Vol. 96, 1994.

The 1852 article by Magnus is "***On the deviation of projectiles; and on a remarkable phenomenon of rotating bodies.***" G. Magnus, Memoirs of the Royal Academy, Berlin (1852). English translation in Scientific Memoirs, London (1853), p.210. Edited by John Tyndall and William Francis.

# Formulary

# For Vector Operations

## Contents



*When you are working problems involving quantities like $\nabla^2$ in cylindrical or spherical coordinates, you do not want to derive the formulas yourself because the chances of your getting the right answer are too small. You are not likely to memorize them correctly either, unless you use a particular formula often. Instead, the best procedure is to look up the result in a table of formulas, sometimes called a* ***formulary****.*

*In this formulary we summarize all the formulas for gradient, divergence and curl, in Cartesian, cylindrical and spherical coordinates. We also include integral formulas, formulas for working with cross products, and with tensors. The formulary was adapted from one developed by David Book of the Naval Research Laboratory.*

*We have also added a short table of integrals, and summarize some of the series expansions we discussed in the text.*

# CYLINDRICAL COORDINATES

## Divergence

$$\vec{\nabla}\cdot\vec{A} = \frac{1}{r}\frac{\partial}{\partial r}(rA_r) + \frac{1}{r}\frac{\partial A_\theta}{\partial\theta} + \frac{\partial A_z}{\partial z}$$

## Gradient

$$(\vec{\nabla}f)_r = \frac{\partial f}{\partial r}$$

$$(\vec{\nabla}f)_\theta = \frac{1}{r}\frac{\partial f}{\partial\theta}$$

$$(\vec{\nabla}f)_z = \frac{\partial f}{\partial z}$$

## Curl

$$(\vec{\nabla}\times\vec{A})_r = \frac{1}{r}\frac{\partial A_z}{\partial\theta} - \frac{\partial A_\theta}{\partial z}$$

$$(\vec{\nabla}\times\vec{A})_\theta = \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}$$

$$(\vec{\nabla}\times\vec{A})_z = \frac{1}{r}\frac{\partial}{\partial r}(rA_\theta) - \frac{1}{r}\frac{\partial A_r}{\partial\theta}$$

## Laplacian

$$\nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial\theta^2} + \frac{\partial^2 f}{\partial z^2}$$

## Laplacian of a vector

$$(\nabla^2\vec{A})_r = \nabla^2 A_r - \frac{2}{r^2}\frac{\partial A_\theta}{\partial\theta} - \frac{A_r}{r^2}$$

$$(\nabla^2\vec{A})_\theta = \nabla^2 A_\theta + \frac{2}{r^2}\frac{\partial A_r}{\partial\theta} - \frac{A_\theta}{r^2}$$

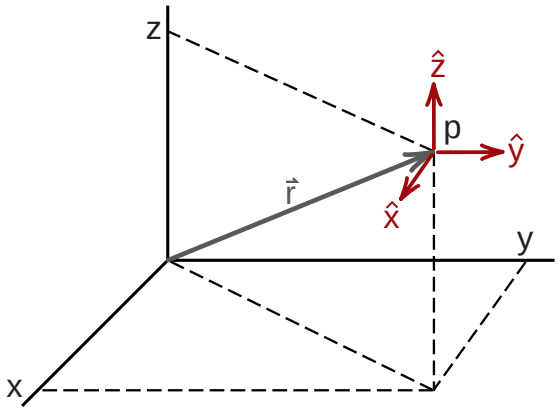
$$(\nabla^2\vec{A})_z = \nabla^2 A_z$$

## Components of $(\vec{A}\cdot\vec{\nabla})\vec{B}$

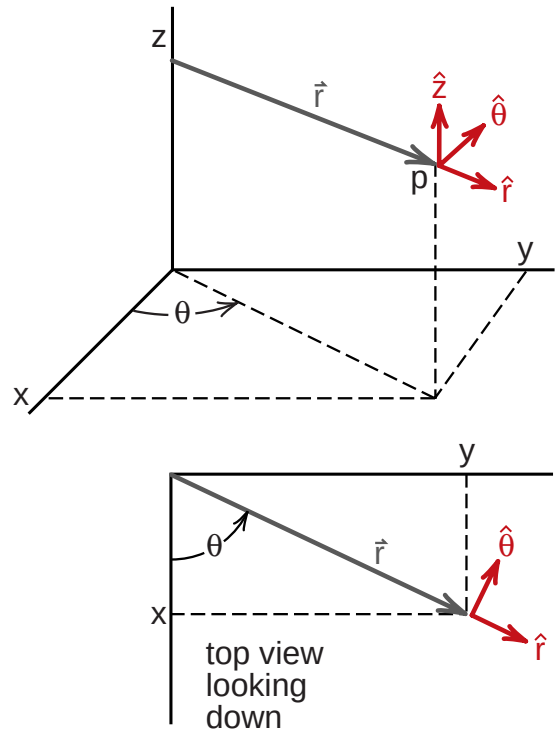
$$[(\vec{A}\cdot\vec{\nabla})\vec{B}]_r = A_r\frac{\partial B_r}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_r}{\partial\theta} + A_z\frac{\partial B_r}{\partial z} - \frac{A_\theta B_\theta}{r}$$

$$[(\vec{A}\cdot\vec{\nabla})\vec{B}]_\theta = A_r\frac{\partial B_\theta}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_\theta}{\partial\theta} + A_z\frac{\partial B_\theta}{\partial z} + \frac{A_\theta B_r}{r}$$

$$[(\vec{A}\cdot\vec{\nabla})\vec{B}]_z = A_r\frac{\partial B_z}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_z}{\partial\theta} + A_z\frac{\partial B_z}{\partial z}$$

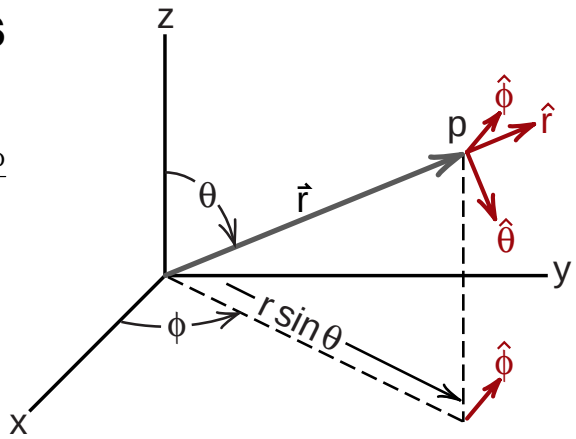



**Cartesian Coordinates**




**Cylindrical Coordinates**

# SPHERICAL POLAR COORDINATES




**Spherical Polar Coordinates**

## Divergence

$$\vec{\nabla}\cdot\vec{A} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(A_\theta\sin\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi}$$

## Gradient

$$(\vec{\nabla}f)_r = \frac{\partial f}{\partial r}$$

$$(\vec{\nabla}f)_\theta = \frac{1}{r}\frac{\partial f}{\partial\theta}$$

$$(\vec{\nabla}f)_\phi = \frac{1}{r\sin\theta}\frac{\partial f}{\partial\phi}$$

## Curl

$$(\vec{\nabla}\times\vec{A})_r = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(A_\phi\sin\theta) - \frac{1}{r\sin\theta}\frac{\partial A_\theta}{\partial\phi}$$

$$(\vec{\nabla}\times\vec{A})_\theta = \frac{1}{r\sin\theta}\frac{\partial A_r}{\partial\phi} - \frac{1}{r}\frac{\partial}{\partial r}(rA_\phi)$$

$$(\vec{\nabla}\times\vec{A})_\phi = \frac{1}{r}\frac{\partial}{\partial r}(rA_\theta) - \frac{1}{r}\frac{\partial A_r}{\partial\theta}$$

## Laplacian

$$\nabla^2 f = \frac{1}{r}\frac{\partial^2}{\partial r^2}(rf) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}$$

## Laplacian of a vector

$$(\nabla^2\vec{A})_r = \nabla^2A_r - \frac{2}{r^2}\frac{\partial A_\theta}{\partial\theta} - \frac{2A_\theta\cot\theta}{r^2} - \frac{2}{r^2\sin\theta}\frac{\partial A_\phi}{\partial\phi}$$

$$(\nabla^2\vec{A})_\theta = \nabla^2A_\theta + \frac{2}{r^2}\frac{\partial A_r}{\partial\theta} - \frac{A_\theta}{r^2\sin^2\theta} - \frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial A_\phi}{\partial\phi}$$

$$(\nabla^2\vec{A})_\phi = \nabla^2A_\phi - \frac{A_\phi}{r^2\sin^2\theta} + \frac{2}{r^2\sin\theta}\frac{\partial A_r}{\partial\phi} + \frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial A_\phi}{\partial\phi}$$

## Components of $(\vec{A}\cdot\vec{\nabla})\vec{B}$

$$[(\vec{A}\cdot\vec{\nabla})\vec{B}]_r = A_r\frac{\partial B_r}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_r}{\partial\theta} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_r}{\partial\phi} - \frac{A_\theta B_\theta + A_\phi B_\phi}{r}$$

$$[(\vec{A}\cdot\vec{\nabla})\vec{B}]_\theta = A_r\frac{\partial B_\theta}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_\theta}{\partial\theta} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_\theta}{\partial\phi} + \frac{A_\theta B_r}{r} - \frac{A_\phi B_\phi\cot\theta}{r}$$

$$[(\vec{A}\cdot\vec{\nabla})\vec{B}]_\phi = A_r\frac{\partial B_\phi}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_\phi}{\partial\theta} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_\phi}{\partial\phi} + \frac{A_\phi B_r}{r} + \frac{A_\phi B_\theta\cot\theta}{r}$$

# VECTOR IDENTITIES

Notation: f, g, etc., are scalars; $\vec{A}$ and $\vec{B}$, etc. are vectors

(1) $\vec{A}\cdot\vec{B}\times\vec{C} = \vec{A}\times\vec{B}\cdot\vec{C} = \vec{B}\cdot\vec{C}\times\vec{A} = \vec{B}\times\vec{C}\cdot\vec{A} = \vec{C}\cdot\vec{A}\times\vec{B} = \vec{C}\times\vec{A}\cdot\vec{B}$

(2) $\vec{A}\times(\vec{B}\times\vec{C}) = (\vec{A}\cdot\vec{C})\,\vec{B} - (\vec{A}\cdot\vec{B})\,\vec{C}$

(3) $\vec{A}\times(\vec{B}\times\vec{C}) + \vec{B}\times(\vec{C}\times\vec{A}) + \vec{C}\times(\vec{A}\times\vec{B}) = 0$

(4) $(\vec{A}\times\vec{B})\cdot(\vec{C}\times\vec{D}) = (\vec{A}\cdot\vec{C})(\vec{B}\cdot\vec{D}) - (\vec{A}\cdot\vec{D})(\vec{B}\cdot\vec{C})$

(5) $(\vec{A}\times\vec{B})\times(\vec{C}\times\vec{D}) = (\vec{A}\times\vec{B}\cdot\vec{D})\,\vec{C} - (\vec{A}\times\vec{B}\cdot\vec{C})\,\vec{D}$

(6) $\vec{\nabla}(fg) = \vec{\nabla}(gf) = f\vec{\nabla}(g) + g\vec{\nabla}(f)$

(7) $\vec{\nabla}\cdot(f\vec{A}) = f\vec{\nabla}\cdot\vec{A} + \vec{A}\cdot\vec{\nabla}f$

(8) $\vec{\nabla}\times(f\vec{A}) = f\vec{\nabla}\times\vec{A} + \vec{\nabla}f\times\vec{A}$

(9) $\vec{\nabla}\cdot(\vec{A}\times\vec{B}) = \vec{B}\cdot\vec{\nabla}\times\vec{A} - \vec{A}\cdot\vec{\nabla}\times\vec{B}$

(10) $\vec{\nabla}\times(\vec{A}\times\vec{B}) = \vec{A}\,(\vec{\nabla}\cdot\vec{B}) - \vec{B}\,(\vec{\nabla}\cdot\vec{A}) + (\vec{B}\cdot\vec{\nabla})\,\vec{A} - (\vec{A}\cdot\vec{\nabla})\,\vec{B}$

(11) $\vec{\nabla}\,(\vec{A}\cdot\vec{B}) = \vec{A}\times(\vec{\nabla}\times\vec{B}) + \vec{B}\times(\vec{\nabla}\times\vec{A}) + (\vec{A}\cdot\vec{\nabla})\,\vec{B} + (\vec{B}\cdot\vec{\nabla})\,\vec{A}$

(12) $\nabla^2 f = \vec{\nabla}\cdot\vec{\nabla}f$

(13) $\nabla^2\vec{A} = \vec{\nabla}\,(\vec{\nabla}\cdot\vec{A}) - \vec{\nabla}\times\vec{\nabla}\times\vec{A}$

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

(14) $\vec{\nabla}\times\vec{\nabla}f = 0$

(15) $\vec{\nabla}\cdot\vec{\nabla}\times\vec{A} = 0$

Let $\vec{r} = \hat{i}\,x + \hat{j}\,y + \hat{k}\,z$ be the radius vector of magnitude r, from the origin to the point x, y, z. Then

(16) $\vec{\nabla}\cdot\vec{r} = 3$

(17) $\vec{\nabla}\times\vec{r} = 0$

(18) $\vec{\nabla}r = \vec{r}/r$

(19) $\vec{\nabla}(1/r) = -\vec{r}/r^3$

(20) $\vec{\nabla}\cdot(\vec{r}/r^3) = 4\pi\delta(\vec{r})$

# INTEGRAL FORMULAS

If V is the volume enclosed by a surface S and $\vec{dS} = \hat{n}dS$ where $\hat{n}$ is the unit normal outward from V

(22) $$\int_V \vec{\nabla} f \, d^3V = \int_S f \, \vec{dS}$$

(23) $$\int_V \vec{\nabla} \cdot \vec{A} \, d^3V = \int_S \vec{A} \cdot \vec{dS}$$

(24) $$\int_V \vec{\nabla} \times \vec{A} \, d^3V = \int_S \vec{dS} \times \vec{A}$$

(25) $$\int_V (f\nabla^2 g - g\nabla^2 f) \, d^3V = \int_S (f\vec{\nabla} g - g\vec{\nabla} f) \cdot \vec{dS}$$

(26) $$\int_V \left[ \vec{A} \cdot \vec{\nabla} \times (\vec{\nabla} \times \vec{B}) - \vec{B} \cdot \vec{\nabla} \times (\vec{\nabla} \times \vec{A}) \right] d^3V = \int_S \left[ \vec{B} \times (\vec{\nabla} \times \vec{A}) - \vec{A} \times (\vec{\nabla} \times \vec{B}) \right] \cdot \vec{dS}$$

If S is an open surface bounded by the contour C of which the line element is $\vec{d\ell}$

(27) $$\int_S \vec{dS} \times \vec{\nabla} f = \oint_C f \vec{d\ell}$$

(28) $$\int_S (\vec{\nabla} \times \vec{A}) \cdot \vec{dS} = \oint_C \vec{A} \cdot \vec{d\ell}$$ *Stokes' law*

(29) $$\int_S (\vec{dS} \times \vec{\nabla}) \times \vec{A} = \oint_C \vec{d\ell} \times \vec{A}$$

(30) $$\int_S (\vec{\nabla} f \times \vec{\nabla} g) \cdot \vec{dS} = \oint_C f \vec{dg} = -\oint_C \vec{g} df$$

# WORKING WITH CROSS PRODUCTS

Use of the permutation tensor $\varepsilon_{ijk}$ to work effectively with the cross products.
(Reference: Appendix I in Chapter 13.)

## The cross product

$$(\vec{A}\times\vec{B})_i = \varepsilon_{ijk}A_jB_k$$

## Product of ε's

$$\varepsilon_{ijk}\varepsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}$$

## Example of use

$$\begin{aligned}
\left[\vec{\nabla}\times(\vec{\nabla}\times\vec{A})\right]_i &= \varepsilon_{ijk}\nabla_j(\vec{\nabla}\times\vec{A})_k \\
&= \varepsilon_{ijk}\varepsilon_{klm}\nabla_j\nabla_l A_m \\
&= (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})\,\nabla_j\nabla_l A_m \\
&= \nabla_j\nabla_i A_j - \nabla_j\nabla_j A_i \\
&= \nabla_i\nabla_j A_j - \nabla_j\nabla_j A_i \\
&= \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) - \nabla^2\vec{A}
\end{aligned}$$

# TENSOR FORMULAS

Notation: f, g, etc., are scalars; $\vec{A}$ and $\vec{B}$, etc. are vectors; $\vec{\vec{T}}$ is a tensor

## Definition

If $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ are orthonormal unit vectors, a second-order tensor $\vec{\vec{T}}$ can be written in the dyadic form

$$\vec{\vec{T}} = \sum_{i,j} T_{ij}\vec{e}_i\vec{e}_j$$

In Cartesian coordinates the divergence of a tensor is a vector with components

$$(\nabla \cdot \vec{\vec{T}})_i = \sum_j (\partial T_{ji}/\partial x_j)$$

## Formulas

$$\vec{\nabla}\cdot(\vec{A}\vec{B}) = (\vec{\nabla}\cdot\vec{A})\vec{B} + (\vec{A}\cdot\vec{\nabla})\vec{B}$$

$$\vec{\nabla}\cdot(f\vec{\vec{T}}) = \vec{\nabla}f\cdot\vec{\vec{T}} + f\vec{\nabla}\cdot\vec{\vec{T}}$$

$$\int_V \vec{\nabla}\cdot\vec{\vec{T}}\, d^3V = \int_S d\vec{S}\cdot\vec{\vec{T}}$$

## Divergence of a tensor (cylindrical coordinates)

$$(\vec{\nabla}\cdot\vec{\vec{T}})_r = \frac{1}{r}\frac{\partial}{\partial r}(rT_{rr}) + \frac{1}{r}\frac{\partial}{\partial\theta}(T_{\theta r}) + \frac{\partial T_{zr}}{\partial z} - \frac{1}{r}T_{\theta\theta}$$

$$(\vec{\nabla}\cdot\vec{\vec{T}})_\theta = \frac{1}{r}\frac{\partial}{\partial r}(rT_{r\theta}) + \frac{1}{r}\frac{\partial T_{\theta\theta}}{\partial\theta} + \frac{\partial T_{z\theta}}{\partial z} + \frac{1}{r}T_{\theta r}$$

$$(\vec{\nabla}\cdot\vec{\vec{T}})_z = \frac{1}{r}\frac{\partial}{\partial r}(rT_{rz}) + \frac{1}{r}\frac{\partial T_{\theta z}}{\partial\theta} + \frac{\partial T_{zz}}{\partial z}$$

## Divergence of a tensor (spherical coordinates)

$$(\vec{\nabla}\cdot\vec{\vec{T}})_r = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2T_{rr}) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(T_{\theta r}\sin\theta) + \frac{1}{r\sin\theta}\frac{\partial T_{\phi r}}{\partial\phi} - \frac{T_{\theta\theta}+T_{\phi\phi}}{r}$$

$$(\vec{\nabla}\cdot\vec{\vec{T}})_\theta = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2T_{r\theta}) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(T_{\theta\theta}\sin\theta) + \frac{1}{r\sin\theta}\frac{\partial T_{\phi\theta}}{\partial\phi} + \frac{T_{\theta r}}{r} - \frac{\cot\theta}{r}T_{\phi\phi}$$

$$(\vec{\nabla}\cdot\vec{\vec{T}})_\phi = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2T_{r\phi}) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(T_{\theta\phi}\sin\theta) + \frac{1}{r\sin\theta}\frac{\partial T_{\phi\phi}}{\partial\phi} + \frac{T_{\phi r}}{r} + \frac{\cot\theta}{r}T_{\phi\theta}$$

## SHORT TABLE OF INTEGRALS

In these integrals, (a) is a constant, and (u) and (v) are any functions of x.

1. $\int dx = x$
2. $\int au\,dx = a\int u\,dx$
3. $\int (u + v)\,dx = \int u\,dx + \int v\,dx$
4. $\int x^m\,dx = \frac{x^{m+1}}{m+1} (m \neq -1)$
5. $\int \frac{dx}{x} = \ln|x|$
6. $\int u\,\frac{dv}{dx}dx = uv - \int v\,\frac{du}{dx}dx$
7. $\int e^x dx = e^x$
8. $\int \sin x\,dx = -\cos x$
9. $\int \cos x\,dx = \sin x$
10. $\int \sin^2 x\,dx = \frac{1}{2}x - \frac{1}{4}\sin 2x$
11. $\int e^{-ax}dx = -\frac{1}{a}e^{-ax}$
12. $\int xe^{-ax}dx = -\frac{1}{a^2}(ax + 1)e^{-ax}$
13. $\int x^2e^{-ax}dx = -\frac{1}{a^3}(a^2x^2 + 2ax + 2)e^{-ax}$
14. $\int_0^\infty x^n e^{-ax}dx = \frac{n!}{a^{n+1}}$
15. $\int_0^\infty x^{2n}e^{-ax^2}dx = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^{n+1}a^n}\sqrt{\frac{\pi}{a}}$
16. $\int \frac{dx}{(x^2 + a^2)^{3/2}} = \frac{x}{a^2\sqrt{x^2 + a^2}}$

# SERIES EXPANSIONS

## The binomial expansion

(Ch 2, page 6)

$$(1+\alpha)^n = 1 + n\alpha + \frac{n(n-1)}{2!}\alpha^2 + \cdots \quad (2\text{-}22)$$

which is valid for any value of $\alpha$ less than one, but which gets better as $\alpha$ becomes smaller.

## Taylor series expansion

(Ch 2, page 8)

$$\begin{aligned} f(x-x_0) &= f(x_0) + f'(x_0)(x-x_0)^1 \\ &\quad + \frac{1}{2!}f''(x_0)(x-x_0)^2 \\ &\quad + \frac{1}{3!}f'''(x_0)(x-x_0)^3 + \cdots \end{aligned}$$

This can be written in the compact form

$$\boxed{f(x-x_0) = \sum_{n=0}^{\infty} \frac{f^n(x_0)}{n!}(x-x_0)^n} \quad \textit{Taylor series expansion} \quad (2\text{-}44)$$

where we used the notation

$$f^n(x_0) \equiv \left.\frac{d^n f(x)}{dx^n}\right|_{x=x_0} \quad (2\text{-}45)$$

## Sine and cosine

(Ch 5, page 4)

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \cdots \quad (13)$$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots \quad (14)$$

where $\theta$ is in radians. These expansions are valid for any value of $\theta$, but most useful for small values where we do not have to keep many terms.

## Exponential

(Ch 1, page 28 and Ch 5, page 4)

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad (1\text{-}136)$$

While this expansion is true for any value of x, it is most useful for small values of x where we do not have to keep many terms to get an accurate answer.

Setting $x = i\theta$ gives

$$e^{i\theta} = 1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \cdots \quad (5\text{-}12)$$

***(Since our previous discussion of exponents only dealt with real numbers, we can consider Equation (12) as the definition of what we mean when the exponent is a complex number)***.

## Physical Constants in CGS Units

| | |
|---|---|
| speed of light | $c = 3\times10^{10}\text{cm/sec} = 1000\,\text{ft}/\mu\text{sec} = 1\,\text{ft/nanosecond}$ |
| acceleration due to gravity at the surface of the earth | $g = 980\,\text{cm/sec}^2 = 32\,\text{ft/sec}^2$ |
| gravitational constant | $G = 6.67\times10^{-8}\text{cm}^3/(\text{gm sec}^2)$ |
| charge on an electron | $e = 4.8\times10^{-10}\text{esu}$ |
| Planck's constant | $h = 6.62\times10^{-27}\text{erg sec}\ (\text{gm cm}^2/\text{sec})$ |
| Planck constant / $2\pi$ | $\hbar = 1.06\times10^{-27}\text{erg sec}\ (\text{gm cm}^2/\text{sec})$ |
| Bohr radius | $a_0 = .529\times10^{-8}\text{cm}$ |
| rest mass of electron | $m_e = 0.911\times10^{-27}\text{gm}$ |
| rest mass of proton | $M_p = 1.67\times10^{-24}\text{gm}$ |
| rest energy of electron | $m_ec^2 = 0.51\,\text{MeV}\ (\approx 1/2\,\text{MeV})$ |
| rest energy of proton | $M_pc^2 = 0.938\,\text{BeV}\ (\approx 1\,\text{BeV})$ |
| proton radius | $r_p = 1.0\times10^{-13}\text{cm}$ |
| Boltzmann's constant | $k = 1.38\times10^{-16}\text{ergs/kelvin}$ |
| Avogadro's number | $N_0 = 6.02\times10^{23}\text{molecules/mole}$ |

absolute zero = $0°\text{K} = -273°\text{C}$
density of mercury = $13.6\,\text{gm}/\text{cm}^3$
mass of earth = $5.98\times10^{27}\text{gm}$
mass of the moon = $7.35\times10^{25}\text{gm}$
mass of the sun = $1.97\times10^{33}\text{gm}$
earth radius = $6.38\times10^{8}\text{cm} = 3960\,\text{mi}$
moon radius = $1.74\times10^{8}\text{cm} = 1080\,\text{mi}$
mean distance to moon = $3.84\times10^{10}\text{cm}$
mean distance to sun = $1.50\times10^{13}\text{cm}$
mean earth velocity in orbit about sun = $29.77\,\text{km}/\text{sec}$

## Conversion Factors

1 meter = 100 cm (100 cm/meter)
1 in. = 2.54 cm (2.54 cm/in.)
1 mi = 5280 ft (5280 ft/mi)
1 km (kilometer) = $10^5\text{cm}$ ($10^5\text{cm}/\text{km}$)
1 mi = 1.61 km = $1.61\times10^5\text{cm}$ ($1.61\times10^5\text{cm/mi}$)
1 Å (angstrom) = $10^{-8}\text{cm}$ ($10^{-8}\text{cm}/\text{Å}$)
1 day = 86,000 sec ($8.6\times10^4\text{sec}/\text{day}$)
1 year = $3.16\times10^7\text{sec}$ ($3.16\times10^7\text{sec/year}$)
1 $\mu$sec (microsecond) = $10^{-6}\text{sec}$ ($10^{-6}\text{sec}/\mu\text{sec}$)
1 nanosecond = $10^{-9}\text{sec}$ ($10^{-9}\text{sec}$ /nanosecond)
1 mi/hr = 44.7 cm/sec
60 mi/hr = 88 ft/sec
1 kg (kilogram) = $10^3\text{gm}$ ($10^3\text{gm}/\text{kg}$)
1 coulomb = $3\times10^9\text{esu}$ ($3\times10^9\text{esu/coulomb}$)
1 ampere = $3\times10^9\text{statamps}$ ($3\times10^9\text{statamps/ampere}$)
1 statvolt = 300 volts (300 volts/statvolt)
1 joule = $10^7\text{ergs}$ ($10^7\text{ergs}/\text{joule}$)
1 W (watt) = $10^7\text{ergs/sec}$ ($10^7\text{erg}/\text{W}$)
1 eV = $1.6\times10^{-12}\text{ergs}$ ($1.6\times10^{-12}\text{ergs/eV}$)
1 MeV = $10^6\text{eV}$ ($10^6\text{eV}/\text{MeV}$)
1 BeV = $10^9\text{eV}$ ($10^9\text{eV}/\text{BeV}$)
1 $\mu$ (micron) pressure = $1.33\,\text{dynes}/\text{cm}^2$
1 cm Hg pressure = $10^4\,\mu$
1 atm = 76 cm Hg = $1.01\times10^6\text{dynes/cm}^2$

**Moose Mountain**
**Digital Press**