Chapter 6

Maxwell’s Equations for Electromagnetic Waves

6.1 Vector Operations

Any physical or mathematical quantity whose amplitude may be decomposed into “directional” components often is represented conveniently as a vector. In this discussion, vectors are denoted by bold-faced underscored lower-case letters, e.g., $\mathbf{x}$. The usual notation for a vector with $N$ elements is a column of $N$ individual numerical scalars, where $N$ is the dimensionality of the vector. For example, the 3-D vector $\mathbf{x}$ is specified by a vertical column of the three ordered numerical components:

$$\mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Both real- and complex-valued scalars will be used as the components $x_n$ with the same notation. If the $x_n$ are real, then the vector $\mathbf{x}$ specifies a location in 3-D Cartesian space. The individual scalar components $x_1$, $x_2$, and $x_3$ are equivalent to the distances along the three axial directions (commonly labeled $x$, $y$, and $z$, respectively, in the space domain). In common situations, the components of the vector $\mathbf{x}$ have dimensions of length, but other representations are possible. For example, we shall often use a convenient representation of a sinusoid in the $x-y$ plane that is specified by a vector whose components have the dimensions of spatial frequency (e.g., cycles per mm).

To minimize any confusion resulting from the use of the symbol “$x$” to represent both a vector and a particular component of a vector, a normal-faced “$x_i$” with a subscript will be used to indicate the $i^{th}$ component of the vector $\mathbf{x}$, while the bold-faced subscripted symbol “$\mathbf{x}_i$” denotes the $i^{th}$ member of a set of vectors. Other notations also will be employed during certain aspects of the discussion, but these cases will be explicitly noted.
Definitions of the algebraic operations of vectors will be essential to this discussion. For example, the sum of two N-D vectors \( \mathbf{x} \) and \( \mathbf{y} \) is generated by summing the pairs of corresponding components:

\[
\mathbf{x} + \mathbf{y} = \begin{bmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_N 
\end{bmatrix} + \begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_N 
\end{bmatrix} = \begin{bmatrix}
    x_1 + y_1 \\
    x_2 + y_2 \\
    \vdots \\
    x_N + y_N 
\end{bmatrix}
\]

The notation “\( x \)” and “\( y \)” used here merely distinguish between the two vectors and their components; they are not references to the \( x \)- and \( y \)-coordinates of 2-D or 3-D space. Note that this definition implies that two vectors must have the same dimension for their sum to exist.

The definition of the difference of two vectors is evident from the equation for the sum:

\[
\mathbf{x} - \mathbf{y} = \begin{bmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_N 
\end{bmatrix} - \begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_N 
\end{bmatrix} = \begin{bmatrix}
    x_1 - y_1 \\
    x_2 - y_2 \\
    \vdots \\
    x_N - y_N 
\end{bmatrix}
\]

Obviously, if the number of dimensions \( N \) of the vector is 1, 2, or 3, then the corresponding vector \( \mathbf{x} \) specifies a location on a line, on a plane, or within a volume, respectively. This interpretation of a vector as the location of a point in space is so pervasive and intuitive that it may obscure other useful and perhaps more general interpretations of vectors and vector components. For example, we can use the vector notation to represent a two-dimensional (2-D) sampled object. Such an object formed from an \( N \times N \) array of samples or by “stacking” the \( N \) columns to create a 1-D vector with \( N^2 \) components. This stacking process is known as lexicographic ordering of the matrix. Such a representation often is used when constructing computer algorithms for processing digital images, but will not be considered further here.

The transpose of the column vector \( \mathbf{x} \) is the same set of scalar components arrayed as a horizontal row, and is denoted in this discussion by a superscript \( T \); another common notation uses an overscored tilde:

\[
\mathbf{x}^T = \begin{bmatrix}
    x_1 \\
    x_2 \\
    x_3 
\end{bmatrix} = \tilde{\mathbf{x}}
\]

By analogy with the usual interpretation of a vector in Cartesian space, the length of a vector with real-valued components is a real-valued scalar computed from the 2-D or 3-D “Pythagorean” sum of the components:

\[
\sum_{n=1}^{N} (x_n)^2 = |\mathbf{x}|^2
\]
The result is the *squared magnitude* of the vector. The vector’s length, or *norm*, is the square root of Eq.(3.5), as shown in the figure and thus also is real valued.

![Diagram of a 2-D vector with real-valued components](image)

\[ |\mathbf{x}| = \sqrt{x_1^2 + x_2^2} \]

Length, or “norm”, of 2-D vector with real-valued components.

From this definition, it is evident that the norm of a vector must be nonnegative (\(|\mathbf{x}| \geq 0\)) and that it is zero only if all scalar components of the vector are zero.

Vectors with unit length will be essential in the discussion of transformations into alternate representations. Such a *unit vector* often is indicated by an overscored caret. The unit vector pointing in the direction of any vector \( \mathbf{x} \) may be generated by dividing each component of \( \mathbf{x} \) by the scalar length \(|\mathbf{x}|\) of the vector:

\[
\hat{\mathbf{x}} = \frac{\mathbf{x}}{|\mathbf{x}|} = \left[ \begin{array}{c} \frac{x_1}{|\mathbf{x}|} \\ \frac{x_2}{|\mathbf{x}|} \\ \vdots \\ \frac{x_N}{|\mathbf{x}|} \end{array} \right]
\]

The squared-magnitude operation is the first example of the vector *scalar product* (also called the *dot product*), which defines a “product” of two vectors of the same dimension that generates a scalar. Following common mathematical notation, the scalar-product operation will be denoted by a “dot” (\(\cdot\)) between the symbols for the vectors. The process also may be written as the transpose of \( \mathbf{x} \) multiplied from the right by \( \mathbf{x} \). Therefore, the scalar product of a vector \( \hat{\mathbf{x}} \) with itself may be written in
6.1.1 Scalar Product of Two Vectors

It is easy to generalize the squared magnitude operation to apply to distinct vectors \( \mathbf{a} \) and \( \mathbf{x} \) that have real-valued components and that have the same dimension \( N \):

\[
\mathbf{a} \cdot \mathbf{x} \equiv \mathbf{a}^T \mathbf{x} = \begin{bmatrix} a_1 & a_2 & \cdots & a_N \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = a_1 x_1 + a_2 x_2 + \cdots + a_N x_N = \sum_{n=1}^{N} a_n x_n
\]

In words, the scalar product of two vectors is obtained by multiplying pairs of vector components with the same indices and summing these products. Note that the scalar product of two distinct vectors may be positive, negative, or zero, whereas that the squared magnitude of a vector must be nonnegative. From these equivalent mathematical expressions, it is apparent that the scalar product of vectors with real-valued components in either order are identical:

\[
\mathbf{a} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{a}
\]

Any process that performs an action between two entities and that may be performed in either order is commutative. The simple concept of the scalar product is the basis (future pun intended) for some very powerful tools for describing vectors and, after appropriate generalization, for functions of continuous variables. The features of the various forms of scalar product are the subject of much of the remainder of this chapter.

The scalar product of an arbitrary “input” vector \( \mathbf{x} \) with a “reference” vector \( \mathbf{a} \) has the form of an operator acting on \( \mathbf{x} \) to produce a scalar \( g \): The appropriate process was just defined:

\[
\mathcal{O} \{ \mathbf{x} \} = \mathbf{a} \cdot \mathbf{x} = \sum_{n=1}^{N} a_n x_n = g
\]

It is apparent that a multiplicative scale factor \( k \) applied to each component of the
real-valued input vector \( \mathbf{x} \) results in the same scaling of the output scalar:

\[
O\{k \, \mathbf{x}\} = \sum_{n=1}^{N} a_n \, (k \, x_n) = k \sum_{n=1}^{N} a_n \, x_n = k \, g
\]

which demonstrates that the scalar product “operator” satisfies the linearity condition.

The geometrical interpretation of a 2-D vector as the endpoint of a line drawn from the origin on the 2-D plane leads to an alternate expression for the scalar product of two vectors. It is convenient to use 2-D vectors denoted by \( \mathbf{f}_n \) with Cartesian components \([x_n, y_n]\), or represented in polar coordinates by the length \(|\mathbf{f}_n|\) and the azimuth angles \(\theta_n\). The geometric picture of the vector establishes the relationship between the polar and Cartesian representations to be:

\[
\mathbf{f}_n = [x_n, y_n] = [|\mathbf{f}_n| \cos [\theta_n], |\mathbf{f}_n| \sin [\theta_n]]
\]

where, in this case, \(x_n\) and \(y_n\) represent \(x\)- and \(y\)-coordinates of the vector \(\mathbf{f}_n\). The scalar product of two such vectors \(\mathbf{f}_1\) and \(\mathbf{f}_2\) is obtained by applying the definition and casting into a different form by using the well-known trigonometric identity for the cosine of the difference of two angles:

\[
\mathbf{f}_1 \cdot \mathbf{f}_2 = x_1 x_2 + y_1 y_2
\]

\[
= (|\mathbf{f}_1| \cos [\theta_1]) (|\mathbf{f}_2| \cos [\theta_2]) + (|\mathbf{f}_1| \sin [\theta_1]) (|\mathbf{f}_2| \sin [\theta_2])
\]

\[
= |\mathbf{f}_1| \, |\mathbf{f}_2| \, \cos [\theta_1 - \theta_2]
\]

where the symmetry of the cosine function has been used in the last step. In words, the scalar product of two 2-D vectors is equal to the product of the lengths of the vectors and the cosine of the included angle \(\theta_1 - \theta_2\). The knowledgeable reader is aware that this result has been obtained by circular reasoning; we are defining the scalar product form by using the Cartesian components of polar vectors, which were themselves determined by scalar products with the Cartesian basis vectors. This quandary is due in part to the familiarity of these concepts. Rather than resolve the issue from first principles, we will instead “sweep it under the rug” while continuing to use our existing intuition as a springboard to generalize these concepts to other applications. For example, it is easy now to generalize the scalar product to real-valued vectors \(\mathbf{a}\) and \(\mathbf{x}\) with arbitrary dimension \(N\):

\[
\mathbf{a} \cdot \mathbf{x} = |\mathbf{a}| \, |\mathbf{x}| \cos [\theta_a - \theta_x] = |\mathbf{a}| \, |\mathbf{x}| \cos [\theta]
\]

where \(\theta\) represents the “included” angle between the two N-D vectors. This angle is measured in the 2-D plane defined by the two vectors. If we consider the 3-D analogy of two vectors from the origin to the surface of a sphere, then the angle \(\theta\) represents the angle along the “great circle” that connects the two vector tips.

This last definition for the scalar product may be used to derive the Schwarz
inequality for vectors by recognizing that \( \cos [\theta] \leq 1 \):

\[
\mathbf{a} \cdot \mathbf{x} \leq |\mathbf{a}| |\mathbf{x}|
\]

The equality is satisfied only for vectors \( \mathbf{a} \) and \( \mathbf{x} \) that “point” in the same direction, which means that the ratios of the corresponding components of \( \mathbf{a} \) and \( \mathbf{x} \) are equal, and that the included angle \( \theta = 0 \) radians, which means that the vectors are scaled replicas. Note both the similarity and difference between the Schwarz inequality and triangle inequality for vectors:

\[
|\mathbf{a} + \mathbf{x}| \leq |\mathbf{a}| + |\mathbf{x}|
\]

In words, the Schwarz inequality says that the scalar product of two vectors can be no larger than the product of their lengths, while the triangle inequality establishes that one side of a triangle can be no longer than the sum of the other two sides. Both relations are illustrated in the figure.

Graphical comparison of Schwarz’ and the triangle inequalities for the same pair of 2-D vectors \( \mathbf{x} \) and \( \mathbf{a} \).

The Schwarz inequality may be combined with the definition of the unit vector to obtain an expression for the included angle between two unit vectors:

\[
\frac{\mathbf{a}}{|\mathbf{a}|} \cdot \frac{\mathbf{x}}{|\mathbf{x}|} = \hat{\mathbf{a}} \cdot \hat{\mathbf{x}} = \cos [\theta] \leq 1
\]

6.1.2 Cross Product

Consider the area of the parallelogram formed by two vectors \( \mathbf{A} \) and \( \mathbf{B} \), as shown:

The area of \( |\mathbf{A}| |\mathbf{B}| \sin [\theta] \) may be computed as a 3-D vector that points perpendicular to the two component vectors with length equal to the area; the calculation
6.1 VECTOR OPERATIONS

is the “cross product” of the two 3-D vectors. Given the two component vectors:

\[ \mathbf{A} = \hat{x}A_x + \hat{y}A_y + \hat{z}A_z \]
\[ \mathbf{B} = \hat{x}B_x + \hat{y}B_y + \hat{z}B_z \]

the cross product is defined:

\[
\mathbf{A} \times \mathbf{B} = \det \begin{bmatrix}
\hat{x} & \hat{y} & \hat{z} \\
A_x & A_y & A_z \\
B_x & B_y & B_z
\end{bmatrix}
= \hat{x}(A_yB_z - A_zB_y) + \hat{y}(A_zB_x - A_xB_z) + \hat{z}(A_xB_y - A_yB_x)
\]

In the example given, \( \mathbf{A} = \hat{x}|\mathbf{A}| \), \( \mathbf{B} = \hat{x}(|\mathbf{B}| \cos \theta) + \hat{y}(|\mathbf{B}| \sin \theta) \), so that \( A_z = B_z = 0 \)

\[
\det \begin{bmatrix}
\hat{x} & \hat{y} & \hat{z} \\
|\mathbf{A}| & 0 & 0 \\
|\mathbf{B}| \cos \theta & |\mathbf{B}| \sin \theta & 0
\end{bmatrix} = \hat{z} (|\mathbf{A}| \ |\mathbf{B}| \sin \theta)
\]

It is easy to see that:

\[ \mathbf{B} \times \mathbf{A} = -\mathbf{A} \times \mathbf{B} \]

Note that the cross product is defined for 3-D vectors ONLY (though we can apply the definition to 2-D vectors by considering their component in the third direction to be zero, see example for curl that follows later).

6.1.3 Triple Vector Product

The “triple vector product” is the cross product of two 3-D vectors (call them \( \mathbf{A} \) and \( \mathbf{B} \)) crossed with a third vector (\( \mathbf{C} \)). The result may be evaluated by straightforward
(yet tedious!) calculation and produces the result:

$$A \times B \times C = B (C \cdot A) - A (B \cdot C)$$

$$= B (C \cdot A) - A (C \cdot B)$$

where the fact that the scalar product commutes for vectors with real-valued components has been used. The triple vector product yields the difference of two scaled replicas of the first two vectors, where the scaling factors are the scalar products of \(C\) with \(A\) and \(B\). The “output” is a vector, as it must be.

We will use this expression for the triple vector product to evaluate the “curl of the curl” shortly.

### 6.2 Vector Calculus

Feynman, *Lectures on Physics II*, Chapter 2,3

In 1864, James Clerk Maxwell published a paper on the dynamics of electromagnetic fields, in which he collected four previously described equations which relate electric and magnetic forces, modified one (by adding a term to remove an inconsistency), and combined them to demonstrate the true nature of light waves. He demonstrated that the amplitudes of the electric and magnetic fields would decrease as the reciprocal of the distance (rather than the square of the reciprocal of the distance, as is true for static electric fields). In this way, an electric current in one location has a much larger effect on a distant electric charge than a static electric charge at the same location as the current.

The four equations are now collected into a group that bears his name. To interpret the four Maxwell equations, we must first understand some concepts of *differential vector calculus*, which may seem intimidating but is really just an extension of normal differentiation applied to scalar and vector fields. For our purposes, a *scalar field* is a description of scalar values in space (one or more spatial dimensions). One example of a scalar field is the temperature distribution in the air throughout the atmosphere. Obviously, a single number is assigned to each point in the space. On the other hand, a *vector field* defines the values of a vector quantity throughout a volume. For example, the vector field of wind velocity in the atmosphere assigns a three-dimensional vector to each point in space. Scalar quantities are denoted by normal-face type and vectors (usually) by underscored bold-face characters, e.g., \(f[x,y,z]\) and \(g[x,y,z]\) describe scalar and vector fields, respectively. Unit vectors (vectors with unit magnitude, also called unit length) are indicated by bold-faced characters topped by a caret, e.g., \(\hat{x}\), \(\hat{y}\), and \(\hat{z}\).

In preparation of the discussion of vector calculus, we’ll review a few concepts of classical mechanics. Consider a force described by the vector \(\mathbf{F} = \hat{x}F_x + \hat{y}F_y + \hat{z}F_z\). The force performs “work” if it acts to create a displacement (described by the vector \(\mathbf{s}\)).

$$\mathbf{F} \cdot \mathbf{s} = W$$

If the displacement is the differential element \(d\mathbf{s} = \hat{x}dx + \hat{y}dy + \hat{z}dz\), then the scalar
product yields a differential element of work

\[ dW = \mathbf{F} \cdot d\mathbf{s} \]

and the work resulting by the action of the force from point \( a \) to point \( b \) is:

\[ W = \int_a^b \mathbf{F} \cdot d\mathbf{s} \]

Note that no work is performed if the force acts at right angles to the displacement; the work is “positive” if the force acts in the direction of the displacement (e.g., a weight dropping in a gravitational field); the work is “negative” if the force acts in opposition to the displacement.

The work can be evaluated via:

\[ W = \int \mathbf{F} \cdot d\mathbf{s} = \int (\mathbf{\hat{x}} F_x + \mathbf{\hat{y}} F_y + \mathbf{\hat{z}} F_z) \cdot (\mathbf{\hat{x}} dx + \mathbf{\hat{y}} dy + \mathbf{\hat{z}} dz) \]
\[ = \int F_x \, dx + \int F_y \, dy + \int F_z \, dz = T + c \]

where \( T \) is the kinetic energy and \( c \) is a constant.

It the vector force is a function only of the distance from some reference point, it may be written in terms of a scalar function of that distance, called the 3-D “potential” (or “potential energy”) \( V \) that satisfies the conditions:

\[ F_x = -\frac{\partial V}{\partial x} \]
\[ F_y = -\frac{\partial V}{\partial y} \]
\[ F_z = -\frac{\partial V}{\partial z} \]

We can substitute these differential expressions into the integral equation for the work:

\[ \int \mathbf{F} \cdot d\mathbf{s} = \int (\mathbf{\hat{x}} F_x + \mathbf{\hat{y}} F_y + \mathbf{\hat{z}} F_z) \cdot (\mathbf{\hat{x}} dx + \mathbf{\hat{y}} dy + \mathbf{\hat{z}} dz) \]
\[ = \int \left( -\frac{\partial V}{\partial x} \right) \, dx + \int \left( -\frac{\partial V}{\partial y} \right) \, dy + \int \left( -\frac{\partial V}{\partial z} \right) \, dz \]
\[ = -\int dV = -V = T + c \]
\[ \implies T + V \equiv E = \text{constant} \]

The sum of the potential and kinetic energies is the total energy, a constant under these conditions of a “conservative system.”

For a simple illustration, consider the force of gravity near the earth’s surface; the
vector force is:

\[ \mathbf{F} = \hat{x}F_x + \hat{y}F_y + \hat{z}F_z \]

\[ = 0\hat{x} + 0\hat{y} + \hat{z}(-mg) \]

so that:

\[ -\frac{\partial V}{\partial x} = 0 \implies V = c_1 \]
\[ -\frac{\partial V}{\partial y} = 0 \implies V = c_2 \]
\[ -\frac{\partial V}{\partial z} = -mg \implies V = mg \int dz = mgz + c_3 \]

\[ \implies V[x, y, z] = mgz + (c_1 + c_2 + c_3) = mgz + \text{constant} \]

\[ E = mgz + \frac{1}{2}m v^2 \]

Under the conditions of a conservative force, we can write differentiate the first two expressions with respect to the “other” variable and equate them:

\[ \frac{\partial}{\partial y} F_x = \frac{\partial}{\partial y} \left( -\frac{\partial V}{\partial x} \right) = -\frac{\partial^2 V}{\partial x \partial y} \]
\[ \frac{\partial}{\partial x} F_y = \frac{\partial}{\partial x} \left( -\frac{\partial V}{\partial y} \right) = -\frac{\partial^2 V}{\partial y \partial x} = \frac{\partial}{\partial y} F_x \]

\[ \implies \frac{\partial}{\partial y} F_x = \frac{\partial}{\partial x} F_y \]

The same pattern of operations leads to two other relations:

\[ \frac{\partial}{\partial z} F_x = \frac{\partial}{\partial x} F_z \]
\[ \frac{\partial}{\partial z} F_y = \frac{\partial}{\partial y} F_z \]

These three are necessary and sufficient conditions that a force is conservative.

We can then write:

\[ \mathbf{F} = -\left( \hat{x} \frac{\partial V}{\partial x} + \hat{y} \frac{\partial V}{\partial y} + \hat{z} \frac{\partial V}{\partial z} \right) \]

which can be written in a shorthand form by defining the first-order differential vector operator \( \nabla \) (called “del”) with three components:

\[ \nabla = \left[ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right] \]
It also may be written in explicit vector form as:

$$\nabla = \hat{x} \frac{\partial}{\partial x} + \hat{y} \frac{\partial}{\partial y} + \hat{z} \frac{\partial}{\partial z}$$

where $\hat{x}$, $\hat{y}$, and $\hat{z}$ are the unit vectors along the $x$, $y$, and $z$ axes respectively. Thus we can write:

$$\mathbf{F} = -\nabla V$$

It is easy to show that $\nabla$ satisfies the requirements for a linear operator:

$$\nabla (a + b) = \nabla a + \nabla b$$

$$\nabla (\alpha a) = \alpha \nabla a$$

where $a, b$ are scalar fields (functions) and $\alpha$ is a numerical constant.

The del operator $\nabla$ may be applied in the same manner as a 3-D vector, though the result is a description of the rate of change of the entity to which it is applied. The operator may be applied to a 3-D “field” of scalars (such as $f(x, y, z)$, where $f$ is a scalar “weight”); an example is the measurement of temperature at each point in $[x, y, z]$. The result $\nabla f(x, y, z)$ assigns a 3-D vector to each point in space, (the gradient). The operator may be applied to a field of vectors (e.g. $\mathbf{g}(x, y, z)$) via a scalar product to create a scalar field $\nabla \cdot \mathbf{g}(x, y, z)$; this is the divergence of the vector field. Finally, it may be applied to a field of 3-D vectors to create a different 3-D vector field $\nabla \times \mathbf{g}(x, y, z)$ (the curl of the vector field). The first two operations may be generalized to operate on or generate 2-D vectors, whereas the curl is defined only for 3-D vector fields (though one or two of the components of this field may be zero).

### 6.3 Gradient

*Derives a Vector Field $\nabla f$ from a Scalar Field $f$*

Applying $\nabla$ to a scalar field $f(x, y, z)$ with three dimensions (such as the temperature of air at all points in the atmosphere) generates a field of 3-D vectors which describes the spatial rate-of-change of the scalar field, i.e., the gradient of the temperature at each point in the atmosphere is a vector that describes the direction and magnitude of the change in air temperature over the 3-D volume. In the 2-D case where the scalar field describes the altitude of landform topography, the gradient vector is the size and direction of the maximum slope of the landform. The 3-D gradient

$$\nabla f(x, y, z) \equiv \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix} = \hat{x} \frac{\partial f}{\partial x} + \hat{y} \frac{\partial f}{\partial y} + \hat{z} \frac{\partial f}{\partial z} = \text{a vector}$$

As implied by its name, the gradient vector at $[x, y, z]$ points “uphill” in the direction of maximum rate-of-change of the field; the magnitude of the gradient $|\nabla f|$ is the slope of the scalar field $x$: 
Scalar field represented as contour map and as 3-D display.

Gradient of the scalar field is a vector field. At each coordinate \([x, y]\), the vector points “uphill” and its length is equal to the slope.

As an aside, the operator \(\nabla\) is only defined in Cartesian coordinates; general expressions in cylindrical and spherical coordinates do not exist.

### 6.4 Divergence

Derives a Scalar Field \(\nabla \cdot \mathbf{g}\) from a Vector Field \(\mathbf{g}\)
\[ \nabla \cdot \mathbf{g}[x, y, z] \equiv \begin{bmatrix} \frac{\partial}{\partial x} \frac{\partial}{\partial y} \frac{\partial}{\partial z} \end{bmatrix} \cdot [g_x, g_y, g_z] = \frac{\partial g_x}{\partial x} + \frac{\partial g_y}{\partial y} + \frac{\partial g_z}{\partial z} = \text{a scalar} \]

The divergence at each point in a vector field is a number that describes the total spatial rate-of-change, such as the total outgoing vector flux per unit volume (flux = net outward flow), and thus is equal to:

\[
\text{flux} = (\text{average normal vector component}) \times (\text{surface area})
\]

For a vector field \( \mathbf{g}[x, y, z] \) and an infinitesimal surface “element” described by its normal differential vector element \( d\mathbf{a} \) directed outward from the volume, the differential element of the “flux” \( F \) (a scalar) of \( \mathbf{g} \) through the surface element \( d\mathbf{a} \) is the scalar (“dot”) product of the vector that describes the field with the vector normal to the surface. Thus the total flux is the integral over the surface:

\[
dF = \mathbf{g} \cdot d\mathbf{a} \\
\implies F = \iint_{\text{surface area } A} \mathbf{g} \cdot d\mathbf{a}
\]

The divergence of the vector field describes the total flux out of a volume \( V \), which is equivalent to the flux through the macroscopic surface area \( A \) of the volume \( V \), which is in turn built up from all of the differential surface elements \( d\mathbf{a} \) enclosing the volume.

\[
F = \iiint_{\text{volume } V} (\nabla \cdot \mathbf{g}) \, dV = \iint_{\text{surface area } A} \mathbf{g} \cdot \mathbf{\hat{n}} \, da
\]

where \( \mathbf{\hat{n}} \) is the unit vector normal to the differential element of area \( d\mathbf{a} \). If there are no net “sources” or “sinks” of flux within the volume (points from which the flux in the field “diverges out of” or “converges into”), then the divergence of flux through that surface must be zero:

\[
F = \iiint_{\text{volume } V} (\nabla \cdot \mathbf{g}) \, dV = \iint_{\text{surface area } A} \mathbf{g} \cdot d\mathbf{a} = 0
\]

if no “source” or “sink” of vector field within \( V \)
If there is a source of flux in the volume, then the divergence of the vector field is proportional to that source “strength”:

\[ \int \int_{\text{surface area } A} \mathbf{E} \cdot d\mathbf{a} > 0 \]

A vector field with nonzero divergence has a disparity between input and output flux.

Of course, the flux of an electric field is not made up of a substance that “moves” through the surface, since the electric field is not the “velocity of anything” (in Feynman’s words).

A vector field whose divergence is zero used to be called “solenoidal;” magnetic fields are solenoidal.

### 6.4.1 Gauss’ Theorem for Divergence

Consider the 6-sided cube with one face (#1) located in the y-z plane at \( x = 0 \), and the “opposite” face (#2) parallel to the y-z plane at \( x = \Delta x \), as shown in the figure.
We want to evaluate the flux of the vector field $g[x, y, z]$ through each of the 6 faces. The “inward” flux through face #1 is the area integral of $g_x[0, y, z]$: 

\[ g_1 = \iint g_x[0, y, z] \, dy \, dz \]

\[ \approx g_x[0, y, z] \cdot \Delta y \cdot \Delta z \]

The “inward flux” through face #2 (opposite #1) is the area integral:

\[ g_2 = \iint -g_x[\Delta x, y, z] \, dy \, dz = -\iint g_x[\Delta x, y, z] \, dy \, dz \]

\[ \approx \left(-g_1 + \frac{\partial g_1}{\partial x} \cdot \Delta x\right) \cdot \Delta y \cdot \Delta z \]

The sum of the two fluxes is:

\[ g_1 + g_2 = \left[\frac{\partial g_x}{\partial x} \cdot \Delta x\right] \cdot \Delta y \cdot \Delta z \]

Similarly, the sum of the fluxes through opposite faces are:

\[ g_3 + g_4 = \left[\frac{\partial g_y}{\partial y} \cdot \Delta y\right] \cdot \Delta x \cdot \Delta z \]

\[ g_5 + g_6 = \left[\frac{\partial g_z}{\partial z} \cdot \Delta z\right] \cdot \Delta x \cdot \Delta y \]
CHAPTER 6 MAXWELL’S EQUATIONS FOR ELECTROMAGNETIC WAVES

The sum of the 6 fluxes is easy to evaluate:

\[
\sum_{n=1}^{6} g_n = \frac{\partial g_x}{\partial x} \Delta x \cdot \Delta y \cdot \Delta z + \frac{\partial g_y}{\partial y} \Delta x \cdot \Delta y \cdot \Delta z + \frac{\partial g_z}{\partial z} \Delta x \cdot \Delta y \cdot \Delta z
\]

\[
= \left( \frac{\partial g_x}{\partial x} + \frac{\partial g_y}{\partial y} + \frac{\partial g_z}{\partial z} \right) \Delta x \cdot \Delta y \cdot \Delta z
\]

\[
= (\nabla \cdot \mathbf{g}) \Delta V
\]

This is the sum of the normal vectors of the surface fluxes through the sums of the surfaces:

\[
(\nabla \cdot \mathbf{g}) \Delta V = \iint_{\text{surface of differential volume}} (\mathbf{g}[x, y, z] \cdot \mathbf{n}) \, da
\]

If we construct a macroscopic volume by summing up these differential volumes, we obtain the integral formula:

\[
\iiint_{\text{macroscopic volume}} (\nabla \cdot \mathbf{g}) \, dV = \iint_{\text{surface of macroscopic volume}} (\mathbf{g}[x, y, z] \cdot \mathbf{n}) \, da
\]

6.5 Curl

Derives a 3-D Vector Field from a 3-D Vector Field \( \mathbf{g} \)

The curl of a vector field describes a spatial nonuniformity of the 3-D vector field \( \mathbf{g}[x, y, z] \). If the field describes the flow of a liquid (matter moving with a velocity), the curl determines whether the liquid is “circulating,” i.e. whether there is a net rotational motion of the vector field about some location. The word equation for the “circulation” of the vector is:

\[
\text{circulation} = (\text{average tangential component}) \times (\text{circumference})
\]

Rather than develop the measure from this equation, we again define an operator (the “curl”) and show that it measures the quantity in question. The “curl” of a 3-D vector field is the cross product of the differential operator \( \nabla \) with the field:

\[
\nabla \times \mathbf{g}[x, y, z] = \det \begin{bmatrix}
\mathbf{x} & \mathbf{y} & \mathbf{z} \\
\frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\
g_x & g_y & g_z
\end{bmatrix}
\]

\[
= \mathbf{x} \left( \frac{\partial g_z}{\partial y} - \frac{\partial g_y}{\partial z} \right) + \mathbf{y} \left( \frac{\partial g_z}{\partial x} - \frac{\partial g_x}{\partial z} \right) + \mathbf{z} \left( \frac{\partial g_y}{\partial x} - \frac{\partial g_x}{\partial y} \right) = \text{a vector}
\]

To visualize curl, imagine a vector field that describes motion of a fluid (e.g., water or wind). If a paddle-wheel placed at some location in the fluid does not revolve, the field at that location has no curl. If the wheel does revolve about some axis, then the
curl is nonzero. The direction of the curl vector is that of the axis of the paddlewheel for the maximum rotation in the counterclockwise direction; the magnitude of the curl vector is that rotation “rate.” The algebraic sign of the curl is determined by the direction of rotation (counterclockwise \(\Rightarrow\) positive curl). The paddle will rotate only if the vector field is spatially nonuniform, e.g., it will not rotate in a vector field where all vectors have the same length and point in the same direction. A vector field with zero curl used to be called “irrotational;” electrostatic fields (i.e., NOT traveling electric waves) are irrotational.

Note that some points in the field can have zero curl while others have nonvanishing curl. Both vector fields shown in the examples of divergence have zero curl, since a paddle wheel placed at any point in either field will not rotate.

\[
\begin{align*}
\text{Curl } f_1 &\neq 0 \\
\text{Curl } f_2 &\neq 0
\end{align*}
\]

Two vector fields with nonzero curl. The “paddlewheel” rotates in both cases.

### 6.5.1 Example of Function with Large Curl

Consider the 3-D field composed of vectors that satisfy:

\[
g [x, y, z] = (-y) \hat{x} + (+x) \hat{y} + 0 \hat{z}
\]

The vectors in this field lie in the \(x - y\) plane and those located on the \(x\) or \(y\) axes are oriented perpendicular to the axes and get longer with increasing distance from the origin.
The vector field \( g[x, y, z] = [-y, x, 0] \)

This drawing is not very complete and so might be misleading; consider the vector located at \([x, y] = [1, 1]\):

\[
g[1, 1, 0] = -\hat{x} + \hat{y} + 0\hat{z}
\]

and so “points” in the diagonal direction (towards \(\theta = +\frac{3\pi}{4}\) or towards about 10:30 if you remember what an analog clock face looks like!). Therefore the vectors in this field define a counterclockwise “flow” where the velocity of the flow increases with radial distance. It is a 1-D analogue of the “bathtub drain vortex.” The magnitudes and azimuth angles of the vectors in this field may be evaluated:

\[
|g[x, y, z]| = \sqrt{(-y)^2 + (+x)^2 + 0^2} = \sqrt{x^2 + y^2}
\]

\[
\phi[x, y, 0] = \tan^{-1}\left(\frac{x}{-y}\right) = -\theta
\]

Now evaluate the partial derivatives of the vectors:

\[
\frac{\partial g_x}{\partial x} = 0, \quad \frac{\partial g_x}{\partial y} = -1, \quad \frac{\partial g_x}{\partial z} = 0
\]

\[
\frac{\partial g_y}{\partial x} = +1, \quad \frac{\partial g_y}{\partial y} = 0, \quad \frac{\partial g_y}{\partial z} = 0
\]

\[
\frac{\partial g_z}{\partial x} = \frac{\partial g_z}{\partial y} = \frac{\partial g_z}{\partial z} = 0
\]

The curl of the field is obtained by substitution:

\[
\nabla \times g[x, y, z] = \hat{x}\left(\frac{\partial g_y}{\partial z} - \frac{\partial g_z}{\partial y}\right) + \hat{y}\left(\frac{\partial g_z}{\partial x} - \frac{\partial g_x}{\partial z}\right) + \hat{z}\left(\frac{\partial g_x}{\partial y} - \frac{\partial g_y}{\partial x}\right)
\]

\[
= \hat{x}(0 - 0) + \hat{y}(0 - 0) + \hat{z}(+1 - (-1))
\]

\[
= 0 \cdot \hat{x} + 0 \cdot \hat{y} + 2 \cdot \hat{z}
\]

The curl vector points in the direction of the \(+z\) axis because the flow is counterclockwise (in the direction of \(+\theta\)), i.e., out of the plane of the flow. The direction of the curl determines that the flow is in the \(x-y\) plane, and the magnitude of the curl is related to the “speed” of the flow, if the vectors describe a motion.

If the divergence

**6.6 Laplacian**

The divergence of the gradient often appears in problems in electromagnetic theory and in imaging. It may be applied to scalar functions via:
6.6 LAPLACIAN

\[ \nabla \cdot \nabla f = \nabla^2 f \]

\[ = \left( \hat{x} \frac{\partial}{\partial x} + \hat{y} \frac{\partial}{\partial y} + \hat{z} \frac{\partial}{\partial z} \right) \cdot \left( \hat{x} \frac{\partial f}{\partial x} + \hat{y} \frac{\partial f}{\partial y} + \hat{z} \frac{\partial f}{\partial z} \right) \]

\[ = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \]

where the scalar operator \( \nabla^2 \) is:

\[ \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \]

Since the Laplacian is a scalar operator, it also may be applied to a vector field to produce the sum of the Laplacians of the three component functions along the three directions:

\[
(\nabla \cdot \nabla) \mathbf{g} = \nabla^2 \mathbf{g} \\
= \hat{x} \frac{\partial^2 g}{\partial x^2} + \hat{y} \frac{\partial^2 g}{\partial y^2} + \hat{z} \frac{\partial^2 g}{\partial z^2}
\]

The Laplacian in other coordinate systems (cylindrical or spherical) also may be defined, but produces more complicated expressions (that we probably don’t need!)

\[
\nabla^2 f [x, y, z] = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \quad \text{(Cartesian)}
\]

\[
\nabla^2 f (\rho, \phi, z) = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left( \rho \frac{\partial f}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 f}{\partial \phi^2} + \frac{\partial^2 f}{\partial z^2} \quad \text{(cylindrical)}
\]

\[
\nabla^2 f (r, \theta, \phi) = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} \quad \text{(spherical)}
\]

The Laplacian is the spatial derivative in the 3-D wave equation, which will be considered in more detail shortly:

\[
\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \mu \frac{\partial^2 f}{\partial t^2}
\]

\[ \implies \nabla^2 f = \frac{1}{\nu^2} \frac{\partial^2 f}{\partial t^2} \]

6.6.1 Curl of Curl

The curl of the curl may be evaluated via the vector triple product that was presented earlier:
\[ \nabla \times (\nabla \times \mathbf{g}) = \nabla (\nabla \cdot \mathbf{g}) - (\nabla \cdot \nabla) \mathbf{g} \]
\[ = \nabla (\nabla \cdot \mathbf{g}) - \nabla^2 \mathbf{g} \]

In words, it is the difference of the gradient of the divergence and the Laplacian of the vector field.

### 6.7 Electric and Magnetic Fields

By 1864, much was known about electric and magnetic effects on materials. Faraday had discovered that a time-varying magnetic field (such as from a moving magnet) can generate an electric field, and Ampere demonstrated the corresponding effect that a time-varying electric field (as from a moving electric charge) produces a magnetic field. Both electric and magnetic fields were known to be vectors that could vary in time and space: the amplitudes of the electric and magnetic fields as functions of position. Both quantities are spatial 3-D vectors that vary over time, and may be denoted by \( \mathbf{E}[x, y, z, t] \) and \( \mathbf{B}[x, y, z, t] \), respectively. The electric field \( \mathbf{E} \) is measured by the force it exerts on a “test” electric charge \( Q \) (measured in coulombs). The force is determined by:

\[ \mathbf{F} \propto Q \cdot \mathbf{E} \Rightarrow \frac{\mathbf{F}}{Q} \propto \mathbf{E} \text{ measured in } \left[ \frac{\text{kg} - \text{m}}{\text{s}^2 - \text{C}} \right] \]

where the force is measured in newtons \( \left[ \frac{\text{kg} - \text{m}}{\text{s}^2} \right] \) as the product of the charge \( Q \) and the electric field \( \mathbf{E} \); it has dimensions of volts per meter (equivalent to joules per coulomb).

#### 6.7.1 A Note on Units

If you consult other books, you will likely see many differences in the equations due to the different systems of units used in electromagnetics (and thus in optics); many students (including the author!) find it difficult to cut through the seeming morass of differences. For example, two of the well-known physics texts on the subject, by Lorrain and Corson and by Jackson, use different systems; the former uses the rationalized MKS system (meter, kilogram, second), the latter uses CGS units (centimeter, gram, second), which includes many factors of \( 4\pi \). The systems evolved from Coulomb’s law that evaluates the force between two electrical charges \( Q_1 \) and \( Q_2 \):

\[ \mathbf{F} \propto \frac{Q_1 Q_2}{r_{12}^2} \hat{r}_{12} \]

The constant of proportionality may be called \( k \):

\[ \mathbf{F} = k \frac{Q_1 Q_2}{r_{12}^2} \hat{r}_{12} \]
If the charges are measured in electrostatic units (esu) (also called statcoulombs), the distance in centimeters, and the force in dynes \((g \cdot \text{cm}^2 \cdot s^{-2})\), then \(k = 1\). This means that two charges of 1 esu separated by 1 cm produces a force of 1 dyn. But what if the charges are measured in coulombs [C], the distance in meters [m], and the force in newtons \((1 \text{ N} = 1 \text{ kg} \cdot \text{m} \cdot \text{s}^{-2})\)? The value of \(k\) is determined from the knowledge that there are \(10^5\) dyn per N, 2.998 \(\times\) \(10^9\) esu per C, and \(10^{-2}\) cm per m:

\[
k = \frac{2.998 \cdot 10^9 \text{ cm}^2}{(10^2 \text{ cm})^2 \cdot 10^5 \text{ dyn}} = 8.988 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}
\]

The force between two charges of 1 C separated by 1 m is nearly \(10^{10}\) N \(\approx\) 4.5 \(\times\) \(10^{10}\) pounds of force [lbf], or about 1,100,000 tons (of force)! The constant \(k\) generally is normalized by a factor of \(4\pi\):

\[
k \equiv \frac{1}{4\pi\epsilon_0} \quad \Rightarrow \quad F = \frac{1}{4\pi\epsilon_0} \frac{Q_1Q_2}{r_{12}^2} \hat{r}_{12}
\]

where \(\epsilon_0 = \frac{1}{4\pi k} \cong 8.854 \times 10^{-12} \frac{\text{C}^2}{\text{N} \cdot \text{m}^2} = 8.854 \times 10^{-12} \frac{\text{F}}{\text{m}}\)

where 1 Farad [F] is equivalent to:

\[
1 \text{ F} = \frac{1}{N \cdot \text{m}} = \frac{1}{C \cdot \text{V}}
\]

(one coulomb per volt), so that 1 volt is equivalent to:

\[
1 \text{ V} = \frac{N \cdot \text{m}}{C}
\]

This “new” normalization constant is called the “dielectric constant” or “permittivity” of free space.

A similar procedure for the magnetic force between two current-carrying wires leads to the exact value of a proportionality constant \(\mu_0\):

\[
\mu_0 \equiv 4\pi \cdot 10^7 \frac{\text{N}}{\text{A}^2}
\]

where 1 ampere is one coulomb per second. The magnetic field in free space \(\mathbf{B}\) (the so-called magnetic induction, measured in tesla) is then related to the magnetic field intensity \(\mathbf{H}\) (also called the auxiliary field) in free space (measured in amperes per meter) by

\[
\mathbf{B} = \mu_0 \mathbf{H}
\]

### 6.7.2 Magnetic Fields

The concept of a magnetic field is seemingly somewhat less intuitive, so we’ll consider it in somewhat more detail. The magnetic field is measured in terms of the “flux”
(often labelled by $\phi$), which is a term arising from the original concept of “lines” of magnetic flux emanating from the magnetic “poles.” In fact, the original CGS unit for magnetic flux was called the “line” (now called the “maxwell,” Mx). The flux emanating from a unit field of 1 gauss is $4\pi$ lines because the area of the sphere is $4\pi r^2$. The MKS unit of magnetic flux is the “weber” (1 Wb = $10^8$ Mx), which was defined as the amount of flux which, when changing uniformly in one second, induces 1 volt in 1 turn of a conductor. In electromagnetism, the more important quantity is the magnetic flux density, labeled $B$ and measured in gauss (CGS) or tesla (MKS). One gauss is one line (maxwell) through an area of 1 cm$^2$ and one tesla is 1 Wb per m$^2$:

$$1 \text{ G} = 1 \frac{\text{Mx}}{\text{cm}^2}$$
$$1 \text{ T} = 1 \frac{\text{Wb}}{\text{m}^2} = 1 \frac{\text{N}}{\text{A} \cdot \text{m}} = 10^4 \text{ G}$$

Two other vector fields are required when describing propagation of electromagnetism through matter (rather than through vacuum): the electric displacement $\mathbf{D}$ and the magnetic field intensity $\mathbf{H}$ (also called the “magnetizing force” or the “auxiliary field”). We assume that any material is linear, isotropic, and homogeneous. “Linearity” means that the response of the medium to an incident field varies in proportion to the field. The response of “isotropic” media does not change with orientation of the field, while the characteristics of a “homogeneous” medium do not vary with position in the medium. The electric displacement $\mathbf{D}$ defines the total electric field within a material due to an external field $\mathbf{E}$. It is the sum of $\mathbf{E}$ and any local field $\mathbf{P}$ generated within the matter due to the changes in positions of electric charges within the material due to that field; this induced field $\mathbf{P}$ is called the “polarization” of the material (not to be confused with the “polarization” of the electric field vector that we will mention later). $\mathbf{H}$ is a similar construct for magnetic fields. $\mathbf{E}$ and $\mathbf{D}$, and $\mathbf{B}$ and $\mathbf{H}$, are related by the so-called constitutive equations that are determined by constants of the medium:

$$\mathbf{D} = \varepsilon \mathbf{E}$$
$$\mathbf{B} = \mu \mathbf{H}$$

where $\varepsilon$ and $\mu$ are the electric permittivity and magnetic permeability of the material, respectively. These are measures of the ability of the electric and magnetic fields to “permeate” the medium; if $\varepsilon$ is increased, then a larger electric field exists within the material, if $\mu$ is larger, then the magnetic field $\mathbf{H}$ does not penetrate as far into the medium.

Since we will consider propagation of light only in vacuum, $\mathbf{D} = \mathbf{E}$ and $\mathbf{B} = \mathbf{H}$. In vacuum, $\mu$ and $\varepsilon$ are denoted $\mu_0$ and $\varepsilon_0$ and both are set to unity in CGS units. In MKS units, the quantities are:
\[ \mu_0 = 4\pi \cdot 10^7 \frac{N}{A^2} \text{ (newton per square ampere)} \]

\[ \epsilon_0 = 8.85 \cdot 10^{-12} \frac{F}{m} \text{ (farads per meter).} \]

As is true for the refractive index \( n \), the permittivity and permeability in matter are larger than in vacuum, \( \epsilon > \epsilon_0 \), and \( \mu > \mu_0 \). In fact (though we won’t discuss it in detail), \( \epsilon \) and \( \mu \) determine the phase velocity \( v \) and the refractive index \( n \) via:

\[ v = \frac{1}{\sqrt{\mu \epsilon}} \]

\[ n = \frac{c}{v} = \frac{\sqrt{\mu \epsilon}}{\sqrt{\mu_0 \epsilon_0}} \]

### 6.8 Maxwell’s Equations

Maxwell collected the four differential equations relating the electric vector field \( \mathbf{E} \) and the magnetic vector field \( \mathbf{B} \) listed below and solved them to derive the character of electromagnetic waves. The equations may be written in equivalent differential and integral forms.

### 6.8.1 Gauss’ Law for Electric Fields

Gauss’ law relates the flux of the electric field over a closed surface to the total charge enclosed by the surface. In its simplest terms, Gauss’ law states that the existence of electrical charges within a volume produces electric fields that pass through the surface of the volume. The flux of the field through the surface is proportional to the “amount” of charge within the volume. If the volume is enlarged, then so is the surface area, so the flux density through the surface must decrease at the same rate that the surface area increases. Also note that if there is no charge within the volume, there still can be flux through the enclosing surface, but the ingoing and outgoing parts of the flux cancel out.

Consider an element of the closed surface defined by its normal vector \( d\mathbf{a} \). The flux of the electric field through this surface element is:

\[ d\Phi = \mathbf{E} \cdot d\mathbf{a} \]

where the symbol “\( \cdot \)” denotes the scalar product of the two vector quantities. According to Gauss’ law, the integral of this quantity over the entire closed surface
CHAPTER 6 MAXWELL’S EQUATIONS FOR ELECTROMAGNETIC WAVES

is:

$$\iint_{\text{surface}} d\Phi = \iint_{\text{surface}} \mathbf{E} \cdot d\mathbf{a}$$

$$= \frac{Q}{\varepsilon} = \frac{1}{\varepsilon} \iiint_{\text{volume}} \rho [x, y, z] \, dV$$

where \(\rho [x, y, z]\) is the volume density of charges (measured coulombs per unit volume).

If the surface encloses no charges, then this integral evaluates to zero. states that the divergence of the vector electric field is proportional to density of electric charges.

$$\iint_{\text{surface}} \mathbf{E} \cdot d\mathbf{a} = \iiint_{\text{volume}} \left( \frac{\partial}{\partial x} \mathbf{E} [x, y, z, t] + \frac{\partial}{\partial y} \mathbf{E} [x, y, z, t] + \frac{\partial}{\partial z} \mathbf{E} [x, y, z, t] \right) dV$$

$$= \iiint_{\text{volume}} (\nabla \cdot \mathbf{E} [x, y, z, t]) \, dV$$

$$\Rightarrow \iiint_{\text{volume}} (\nabla \cdot \mathbf{E} [x, y, z, t]) \, dV = \frac{1}{\varepsilon} \iiint_{\text{volume}} \rho [x, y, z] \, dV$$

$$\Rightarrow \nabla \cdot \mathbf{E} [x, y, z, t] = \frac{Q}{\varepsilon}$$

In words, the divergence of the vector electric field is a scalar that is proportional to the total charge within the volume.

### 6.8.2 Gauss’ Law for Magnetic Fields

Since there are no magnetic analogues for “charges”, the volume cannot enclose a magnetic analogue of \(\rho\) or \(Q\), which leads to the particularly simple forms for Gauss’s law for the magnetic flux density:

$$\iint_{\text{surface}} \mathbf{B} \cdot d\mathbf{a} = 0$$

$$\nabla \cdot \mathbf{B} [x, y, z, t] = 0$$

In other words, the flux of the magnetic field through any enclosed surface ALWAYS is zero. This is often interpreted by the statement that there are no magnetic analogues of charges (called magnetic “monopoles”). An “electric monopole” is an electron or proton that carries a net charge of one “sign.”

### 6.8.3 Faraday’s Law of Magnetic Induction

Michael Faraday observed in 1831 the phenomenon that he called “electromagnetic induction,” that generates (“induces”) electricity in a wire from a current in another wire. In other words, he discovered the basis for the electric transformer. Shortly thereafter, Faraday discovered magneto-electric induction: how to produce a steady
electric current in a wire by physically manipulating a magnet near the wire. Faraday attached two wires to a copper disc through a sliding contact. He rotated the disc between the poles of a horseshoe magnet and generated a continuous direct current; in short, this was the first generator (the “dynamo”).

The mathematical formulation of Faraday’s magneto-electric induction is called Faraday’s law, which states that the rate of change of a magnetic field through a 2-D surface is proportional to the circulation of the electric field around the 1-D perimeter of the 2-D surface. In mathematical terms, the time derivative of the magnetic field is proportional to the particular spatial derivative (the curl) of the electric field:

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

Thus a time-varying magnetic field produces a spatially varying electric field, and vice versa.

### 6.8.4 Ampere’s Law

The analogue of Faraday’s law relates the rate of change of the flux of an electric field through a surface to the circulation of the magnetic field around the perimeter of the surface. Maxwell added a “correction term” due to the flux of electric current (due to moving electric charges) through the surface. The corrected form of Ampere’s law is:

$$+\epsilon \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J} = \nabla \times \frac{\mathbf{B}}{\mu}$$

where the additional source term $\mathbf{J}$ is the “current density” of the electric field (measured in amperes per unit volume, equivalent to coulombs per second per unit volume). Note the difference in the algebraic sign in the analogous expressions of Faraday’s law and Ampere’s law.

We have already seen that:

$$\mu \epsilon = \frac{1}{c^2}$$

where $c$ is the velocity of light, $c = 2.99792458 \times 10^8$ m s$^{-1}$, which shows that the effect of the spatial variation of the magnetic field produces a much smaller temporal change in the electric field than vice versa.

There are two “source” terms in Maxwell’s equations: the “static” charge density $\rho$ and the “dynamic” current density $\mathbf{J}$. These can only be nonzero within media (such as copper wire) and thus vanish in vacuum. If we consider the propagation of light only in a vacuum, neither electric charges nor conductors are present and both source terms vanish.

### 6.8.5 Maxwell’s Equations

(Jackson, *Classical Electrodynamics*, §6)

In 1864, James Clerk Maxwell collected these four equations and derived the form...
of the fields that simultaneously satisfy them in some simple cases. In rationalized MKS units, the differential forms of the equations (after adding the correction term to Ampere’s law) are:

\[
\begin{align*}
\nabla \cdot \mathbf{E} &= \frac{\rho}{\varepsilon} \quad \text{Gauss’ Law for Electric Fields, Coulomb’s Law} \\
\nabla \cdot \mathbf{B} &= 0 \quad \text{Gauss’ Law for Magnetic Fields} \\
-\frac{\partial \mathbf{B}}{\partial t} &= \nabla \times \mathbf{E} \quad \text{Faraday’s Law of Magnetic Induction} \\
+\frac{\mu \varepsilon}{\partial t} &= \nabla \times \frac{\mathbf{B}}{\mu} - \frac{\mathbf{J}}{\mu} \quad \text{Ampere’s Law}
\end{align*}
\]

These four coupled first-order differential equations can be solved directly in many cases; we will do a simple one now that leads to propagation of electromagnetic plane waves.

The definition of curl may be used to rewrite the four first-order differential vector equations of Maxwell as eight first-order scalar differential equations:

\[
\begin{align*}
\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} &= 0 \quad \text{Gauss’ Law for } \mathbf{E} \text{ if no sources present} \\
\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} &= 0 \quad \text{Gauss’ Law for } \mathbf{B} \\
-\frac{\partial B_x}{\partial t} &= \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \quad \text{Faraday’s Law} \\
-\frac{\partial B_y}{\partial t} &= \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \\
-\frac{\partial B_z}{\partial t} &= \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \\
+\mu \frac{\partial E_x}{\partial t} &= \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} \quad \text{Ampere’s Law} \\
+\mu \frac{\partial E_y}{\partial t} &= \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \\
+\mu \frac{\partial E_z}{\partial t} &= \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}
\end{align*}
\]

### 6.9 Wave Equation

Take the curl of both sides of Faraday’s law. We can use the expression for the “curl of the curl” previously mentioned (though not derived) to evaluate the curl of the
curl of the electric field:

\[
\nabla \times (\nabla \times \mathbf{E}) = \nabla (\nabla \cdot \mathbf{E}) - (\nabla \cdot \nabla) \mathbf{E}
\]

\[
= \nabla (\nabla \cdot \mathbf{E}) - \nabla \cdot \nabla \mathbf{E}
\]

\[
= 0 - \nabla^2 \mathbf{E}
\]

where Gauss’ law applicable in “sourcefree” regions has been used in the last step (the enclosed charge is zero). The right side of the equation may be rewritten by applying Ampere’s law:

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

\[
= -\epsilon \mu \frac{\partial}{\partial t} \left( \frac{\partial \mathbf{E}}{\partial t} \right)
\]

\[
= -\frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}
\]

After equating the two sides of the equation:

\[
\nabla^2 \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}
\]

where \( c \) is the velocity of light. This expression relates the spatial and temporal second derivatives of the electric field and is called the wave equation. It was first introduced by d’Alembert in 1747. It assumes that no energy of the wave is lost as it propagates, as would occur if friction or damping forces are present.

The general wave equation may be written:

\[
\nabla^2 \psi [x, y, z, t] = \frac{1}{\sqrt{v^2 \partial t^2}} \psi [x, y, z, t]
\]

where \( v \) is the velocity of a point of constant phase: our old friend the phase velocity. The wave equation may be rigorously derived for a transverse wave on a string – you probably saw this in a classical mechanics course.

The wave equation for electric fields confirms our earlier observation:

\[
\epsilon \mu = \frac{1}{c^2} \implies c = \sqrt{\frac{1}{\epsilon \mu}}
\]

Think of this result for a second; the phase velocity of the wave in a medium is related to two measureable properties of the medium; the permittivity and the permeability.

The 1-D equation may be written in the form of a “second-order homogeneous” differential equation:

\[
\left( \frac{\partial^2}{\partial z^2} - \frac{1}{v^2 \partial t^2} \right) \psi [z, t] = 0
\]

Any differential equation is linear, so that if \( \psi_1 [z, t] \) and \( \psi_2 [z, t] \) are solutions to the
equation, so is $a\psi_1[z,t] + b\psi_2[z,t]$. The linearity property means that light beams can pass “through” each other and that waves can constructively or destructively interfere.

The wave equation has the simple solution:

$$\psi[z,t] = f[z \pm vt]$$

where $f[u]$ is any function that may be differentiated twice.

**Proof.**

Define $u \equiv z \pm vt \implies \frac{\partial u}{\partial z} = 1, \frac{\partial u}{\partial t} = \pm v$

Apply the chain rule:

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial z} \quad \text{and} \quad \frac{\partial f}{\partial t} = \frac{\partial f}{\partial u} \cdot (\pm v)$$

Substitute into wave equation:

$$\frac{\partial^2 f}{\partial u^2} = \frac{1}{v^2} \frac{\partial^2 f}{\partial t^2} \implies \frac{\partial^2 f}{\partial u^2} = \frac{1}{v^2} \left( \frac{\partial^2 f}{\partial u^2} \cdot v^2 \right) = \frac{\partial^2 f}{\partial u^2}$$

The expressions for sinusoidal waves derived in the last section satisfy the wave equation:

$$\frac{\partial^2}{\partial z^2} (\hat{x} E_0 \cos [kz - \omega t]) = \hat{x} E_0 (-k^2) \cos [kz - \omega t]$$

$$\frac{1}{v^2} \frac{\partial^2}{\partial t^2} (\hat{x} E_0 \cos [kz - \omega t]) = \frac{1}{v^2} \hat{x} E_0 (-\omega^2 \cos [kz - \omega t])$$

$$= -\hat{x} E_0 \left( \frac{\omega^2}{v^2} \right) \cos [kz - \omega t]$$

$$\implies v^2 = \frac{\omega^2}{k^2} \implies v = \frac{\omega}{k}$$

thus agreeing with our previous expression for phase velocity.

If the general solution to the wave equation has the form:

$$\psi[z,t] = f[z - vt]$$

where the form of the function $f$ is arbitrary, then the argument of the function $[z - vt]$ (the “phase”) remains constant if $x$ increases with increasing time. The “shape” $f$ moves towards $z = +\infty$ with increasing time without changing its shape (i.e., without “dispersion”). A second solution to this equation is:

$$\psi[z,t] = g[z + vt]$$

which moves towards $z = -\infty$. 
The spatial derivative of the corresponding 3-D wave equation is the sum of the three second partial derivatives:

\[
\frac{1}{v^2} \frac{\partial^2}{\partial t^2} \psi [x, y, z, t] = \frac{\partial^2}{\partial x^2} \psi [x, y, z, t] + \frac{\partial^2}{\partial y^2} \psi [x, y, z, t] + \frac{\partial^2}{\partial z^2} \psi [x, y, z, t] = \nabla^2 \psi [x, y, z, t]
\]

The 3-D wave may still be a sinusoid with argument in radians, so we must be more careful about how the 3-D function becomes a 1-D function. The \(x, y, \) and \(z\) dependencies all have associated “wavelengths” that may be defined by their corresponding “wavenumber” \(k_x, k_y, k_z\) that may be written as a “wavevector” \(\mathbf{k}_0\):

\[
\psi [x, y, z, t] = A \cos [\Phi [x, y, z, t]] = A \cos [k_x x + k_y y + k_z z \pm \omega_0 t - \phi_0]
\]

\[
= A \cos \left[ \begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right] \pm \omega_0 t - \phi_0
\]

\[
= A \cos [\mathbf{k}_0 \cdot \mathbf{r} \pm \omega_0 t - \phi_0]
\]

Note that the components of the electric and magnetic fields \((E_x, E_y, E_z, B_x, B_y, \) and \(B_z\)) all satisfy the wave equation.

### 6.9.1 Electromagnetic Waves from Maxwell’s Equations

In the general case, the electric field and magnetic fields can have the form:

\[
\mathbf{E} [x, y, z, t] = \hat{x} E_x [x, y, z, t] + \hat{y} E_y [x, y, z, t] + \hat{z} E_z [x, y, z, t]
\]

\[
\mathbf{B} [x, y, z, t] = \hat{x} B_x [x, y, z, t] + \hat{y} B_y [x, y, z, t] + \hat{z} B_z [x, y, z, t]
\]

We will now solve these equations for a single specific case: an infinite plane electric field wave propagating in vacuum toward \(z = +\infty\). The locus of points of constant phase (often called a wavefront) of a plane wave is (obviously) a plane. The vector electric field \(\mathbf{E}\) is assumed to be constant as a function of \(x\) and/or \(y\) at a particular value of \(z\), but its vector amplitude can vary with \(z\); this variation will be shown to be sinusoidal. This constraint means that derivatives of the components “transverse” to the direction of travel must vanish:

\[
\frac{\partial E_x}{\partial x} = \frac{\partial E_y}{\partial y} = \frac{\partial E_y}{\partial x} = \frac{\partial E_z}{\partial x} = \frac{\partial E_z}{\partial y} = 0
\]

and thus the expression for the 4-D vector field \(\mathbf{E} [x, y, z, t]\) can be simplified; the components are no longer functions of \(x\) or \(y\):

\[
\mathbf{E} [x, y, z, t] = \mathbf{E} [z, t] = \hat{x} E_x [z, t] + \hat{y} E_y [z, t] + \hat{z} E_z [z, t]
\]
From (9) and Gauss’ law for electric fields (1), we find that:

\[
\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0
\]

\[\Rightarrow 0 + 0 + \frac{\partial E_z}{\partial z} = 0 \Rightarrow \frac{\partial E_z}{\partial z} = 0 \quad (11)\]

Since the derivative of \(E_z\) with respect to \(z\) vanishes, then the \(z\)-component of the electric field \(E_z\) must be constant, and its amplitude is arbitrary so we select it to be 0:

\[E_z [x, y, z] = \text{constant} \rightarrow 0 \quad (12)\]

Therefore, the electric field is now expressable in a much simpler form:

\[\mathbf{E} [x, y, z] = E_x [z, t] + E_y [z, t] \quad (13)\]

i.e., the only existing electric field is perpendicular (transverse) to \(z\)! This is a significant result in and of itself. We can simplify eq.(13) by rotating the coordinate system about the \(z\) axis to align \(\mathbf{E}\) with the \(x\)-axis, so that:

\[E_y [z, t] = 0 \quad \text{by assumption} \quad (14)\]

The resulting expression for the electric field is now very simple:

\[\mathbf{E} [x, y, z] = \hat{x}E_x [z, t] \quad (15)\]

We can substitute this expression into Faraday’s Law (eqs. 3,4,5) to find the magnetic field:

\[-\frac{\partial B_x}{\partial t} = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = 0 - \frac{\partial E_y}{\partial z} \Rightarrow \frac{\partial B_x}{\partial t} = \frac{\partial E_y}{\partial z} = 0 \Rightarrow B_x [t] \text{ is constant with time} \quad (3)\]

\[-\frac{\partial B_y}{\partial t} = \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = \frac{\partial E_x}{\partial z} \Rightarrow \frac{\partial B_y}{\partial t} = -\frac{\partial E_x}{\partial z} \quad (4)\]

\[-\frac{\partial B_z}{\partial t} = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \Rightarrow B_z [t] \text{ is constant with time} \quad (5)\]

We can arbitrarily set the constant terms \(B_x [t] = B_z [t] = 0\), so the only remaining equation is:

\[-\frac{\partial B_y}{\partial t} = \frac{\partial E_x}{\partial z} \quad (4)\]

which says that the \(time\) derivative of the magnetic field \(B_y\) is equal to the negative of the \(space\) derivative of \(E_x\). We can now find a relation between \(B_y\) and \(E_x\) by standard solution techniques of differential equations. For example, we can use the
wave equation:

\[ \nabla^2 E = 1 \frac{\partial^2 E}{c^2 \partial t^2} \to \frac{\partial^2 E_x [z, t]}{\partial z^2} = 1 \frac{\partial^2 E_x [z, t]}{c^2 \partial t^2} \]

Assume that \( E_x [z, t] \) varies sinusoidally with \( z \):

\[ E [x, y, z, t] = \hat{x} E_x [z, t] = \hat{x} E_0 \cos [k_0 z - \omega_0 t] \]

It is easy to show that this satisfies the wave equation if the phase velocity of the wave is what we would predict:

\[
\begin{align*}
\frac{\partial^2 E_x [z, t]}{\partial x^2} &= \frac{\partial^2 E_x [z, t]}{\partial y^2} = 0 \\
\frac{\partial^2 E_x [z, t]}{\partial z^2} &= - (k_0^2) E_0 \cos [k_0 z - \omega_0 t] \\
\frac{\partial^2 E_x [z, t]}{\partial t^2} &= - (-\omega_0)^2 E_0 \cos [k_0 z - \omega_0 t]
\end{align*}
\]

\[ \implies v_0 = \frac{\omega_0}{k_0} \]

Now substitute this to solve for \( B_y \) by integration:

\[
\begin{align*}
\implies -\frac{\partial B_y}{\partial t} &= \frac{\partial E_x}{\partial z} = -k_0 E_0 \sin [k_0 z - \omega_0 t] \\
\implies -B_y [z, t] &= \int \frac{\partial E_x}{\partial z} dt = -(-k_0 E_0) \int \sin [k_0 z - \omega_0 t] dt \\
-B_y [z, t] &= +k_0 E_0 \cdot \left[ \frac{-\cos [k_0 z - \omega_0 t]}{-\omega_0} \right] \\
&= k_0 E_0 \cos [k_0 z - \omega_0 t] \\
&= \frac{E_0}{\omega_0} \cos [k_0 z - \omega_0 t] \\
&= \frac{E_0}{v_\phi} \cos [k_0 z - \omega_0 t] \\
\implies B [z, t] &= \hat{y} \left( -\frac{E_0}{v_\phi} \cos [k_0 z - \omega_0 t] \right)
\end{align*}
\]

where \( v_\phi \) is the phase velocity of the electromagnetic wave that we have already defined.

\[ B_y = \frac{E_0}{v_\phi} \cos [k_0 z - \omega_0 t] = \frac{E_x}{v_\phi} \implies E_x = v_\phi B_y \]
Note that the only existing component of \( \mathbf{B} \) is \( B_y \), which is \textit{perpendicular} to the component \( E_x \) of \( \mathbf{E} \), and that both \( \mathbf{E} \) and \( \mathbf{B} \) are transverse to the direction of propagation (in the \( +z \) direction). Also note that the sinusoidal variations of \( \mathbf{E} \) and \( \mathbf{B} \) have the same arguments, which means that they oscillate “in phase”. The amplitude of the magnetic field is smaller by the factor of the phase velocity \( v_\phi = c \), so the effect of the magnetic field on observations is generally much smaller and often may be ignored in physical situations, though it is essential for the electric field to propagate.

### 6.9.2 Poynting Vector

Given transverse electric and magnetic fields \( \mathbf{E} \) and \( \mathbf{B} \), the light propagates in the mutually orthogonal direction \( \mathbf{k} \):

\[
\mathbf{s} \equiv c^2 \mathbf{E} \times \mathbf{B}
\]

This is “Poynting’s vector” and has dimensions of power per unit area (e.g., \( \frac{W}{m^2} \)). It measures the total flow of energy through unit area in unit time.

In this case with \( \mathbf{E} [x, y, z, t] = \hat{x} E_x [z, t] \) and \( \mathbf{B} [x, y, z, t] = \hat{y} B_y [z, t] \), the propagation direction is:

\[
\mathbf{E} [x, y, z, t] = \hat{x} E_x [z, t] \\
\mathbf{B} [x, y, z, t] = \hat{y} B_y [z, t] = \hat{y} \frac{E_x}{v} [z, t]
\]

\[
\mathbf{s} = (\hat{x} \times \hat{y}) \cdot c^2 \epsilon E_x^2 = \hat{z} \left( v^2 \epsilon E_x^2 \right)
\]

along the \( +z \) direction, as predicted.

The average power of the light wave per unit area is the “irradiance,” and is determined from the Poynting vector

\[
I [x, y, z, t] = \langle \mathbf{s} [x, y, z, t] \rangle \equiv \frac{1}{\Delta T} \int_{t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} \mathbf{s} [x, y, z, t'] \ dt'
\]

\[
= \frac{1}{\Delta T} \int_{t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} c^2 \epsilon \left( \hat{x} E_0 \cos [k_0 z - \omega_0 t] \right) \times \hat{y} \left( \frac{E_0}{c} \cos [k_0 z - \omega_0 t] \right) \ dt'
\]

\[
= \frac{1}{\Delta T} \left( \hat{x} \times \hat{y} \right) \int_{t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} c^2 \epsilon E_0^2 \cos^2 [k_0 z - \omega_0 t] \ dt'
\]

\[
= \hat{z} \left( \epsilon E_0^2 \right) \cdot \frac{1}{\Delta T} \int_{t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} \cos^2 [k_0 z - \omega_0 t] \ dt'
\]

\[
= \hat{z} \left( \epsilon E_0^2 \right) \cdot \frac{1}{2}
\]

\[
|\mathbf{s}|_{\text{average}} = \frac{1}{2} \frac{E_0^2}{v}
\]
6.9.3 Redux on Phase Velocity of Electromagnetic Waves

Given the form for the plane electromagnetic wave in a vacuum, we can now use the three Ampere relations to find something else useful:

\[ +\mu \frac{\partial E_x}{\partial t} = \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} \]  
\[ +\mu \frac{\partial E_y}{\partial t} = \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \]  
\[ +\mu \frac{\partial E_z}{\partial t} = \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \]  

Because \( E = \hat{x}E_0 \), only (6) does not vanish:

\[ \mu \frac{\partial}{\partial t} (E_0 \cos [k_0z - \omega_0t]) = -\frac{\partial}{\partial z} \left( \frac{E_0}{v_\phi} \cos [k_0z - \omega_0t] \right) \]
\[ \Rightarrow \mu \varepsilon E_0 (\omega_0 \sin [k_0z - \omega_0t]) = -\frac{E_0}{v_\phi} (-k_0 \sin [k_0z - \omega_0t]) \]
\[ \Rightarrow \mu \varepsilon \omega_0 E_0 = \frac{E_0 k_0}{v_\phi} \]
\[ \Rightarrow \mu \varepsilon = \frac{k_0}{\omega_0 v_\phi} = \frac{1}{v_\phi^2} \Rightarrow v_\phi^2 = \left( \frac{\omega_0}{k_0} \right)^2 = \frac{1}{\mu \varepsilon} \]

which we already knew from the wave equation. In vacuum, \( \mu \equiv \mu_0, \varepsilon \equiv \varepsilon_0, v_\phi \equiv c \), and \( c = \sqrt{\frac{1}{\mu_0 \varepsilon_0}} \). The permittivity and permeability of free space (vacuum) can be measured in laboratory experiments, thus allowing a calculation of the phase velocity of electromagnetic waves. The permeability in vacuum is:

\[ \mu \gtrsim \mu_0 \equiv 4\pi \times 10^{-7} \text{ newton/ampere}^2 \approx 1.26 \times 10^{-6} \text{ newton/ampere}^2 \]

and the permittivity is:

\[ \varepsilon_0 \approx 8.85 \times 10^{-12} \text{ farads/m} \]
These values produce the result:

\[
\mu_0\varepsilon_0 = \left(8.85 \cdot 10^{-12}\text{coul/v-m}\right) \left(1.26 \cdot 10^{-6}\frac{\text{J}}{\text{amp}^2 \cdot \text{m}}\right)
\]
\[
= 1.11 \cdot 10^{-17}\text{coul/v-m} \cdot \frac{\text{J}}{\text{s}^2}
\]
\[
= 1.11 \cdot 10^{-17} \frac{1}{\text{v-m}} \cdot \frac{\text{J-s}^2}{\text{coul-m}}
\]
\[
= 1.11 \cdot 10^{-17}\frac{s^2}{\text{m}^2}
\]
\[
\implies c = \sqrt{\frac{1}{\mu_0\varepsilon_0}} = 2.99 \cdot 10^8\frac{\text{m}}{\text{s}}, \text{ which agrees with experiment.}
\]

In media (i.e., if not in vacuum), the phase velocity is different (we know it to be slower). The same relation may be written using the permittivity and permeability of the medium. The permeability of most optical materials is close to that of vacuum, while the permittivity \(\varepsilon\) of optical materials is larger than in vacuum:

\[
\epsilon > \varepsilon_0 \approx 8.85 \times 10^{-12}\frac{\text{farads}}{\text{m}}
\]

So therefore

\[
v = \sqrt{\frac{1}{\mu\varepsilon}} \approx \sqrt{\frac{1}{\epsilon}} \leq \sqrt{\frac{1}{\varepsilon_0}} \approx \sqrt{\frac{1}{\mu_0\varepsilon_0}} = c
\]

**Index of Refraction**

Of course, the dimensionless ratio of the velocity of light in vacuum to that in the medium is the index of refraction \(n\):

\[
n \equiv \frac{c}{v} \geq 1
\]

\[
n = \sqrt{\frac{1}{\mu_0\varepsilon_0}} = \sqrt{\frac{\mu\varepsilon}{\mu_0\varepsilon_0}} \approx \sqrt{\frac{\varepsilon}{\varepsilon_0}}
\]

\[
\implies n^2 = \frac{\varepsilon}{\varepsilon_0}
\]

For metals and absorptive materials, the index of refraction is complex valued and the permeabilities may not be equal. The complex refractive index often is denoted by \(\tilde{n}\) and its imaginary part by \(\kappa\):

\[
\tilde{n}^2 \equiv (n + i\kappa)^2 = \frac{\mu\varepsilon}{\mu_0\varepsilon_0}
\]
so that
\[ |k| = k = \frac{n}{c} \omega_0 = (n + i\kappa) \frac{\omega_0}{c} \]
which implies that the wavevector \( k \) is complex-valued as well. In this situation, the propagating electric field is written:
\[
\mathbf{E}[x, y, z, t] = E_0 \exp \left[ i \left( k \cdot \mathbf{r} - \omega_0 t \right) \right] = E_0 \exp \left[ i \left( k \cdot (\mathbf{\hat{s}} \cdot \mathbf{r}) - \omega_0 t \right) \right] = E_0 \exp \left[ i \omega_0 \left( \frac{n}{c} \mathbf{\hat{s}} \cdot \mathbf{r} - t \right) \right] \exp \left[ -\frac{\kappa}{c} \mathbf{\hat{s}} \cdot \mathbf{r} \right]
\]
If we assume that the direction of propagation \( \mathbf{\hat{s}} \) is in the direction of \( \mathbf{r} \) (as in a plane wave), then
\[ \mathbf{\hat{s}} \cdot \mathbf{r} = |\mathbf{\hat{s}}| |\mathbf{r}| \cos(0) = |\mathbf{r}| = r \]
so that the electric field may be simplified to:
\[
E_0 \exp \left[ i\omega_0 \left( \frac{n}{c} r - t \right) \right] \exp \left[ -\frac{\kappa}{c} r \right] \equiv E_0 \exp \left[ i\omega_0 \left( \frac{n}{c} r - t \right) \right] \exp \left[ -\frac{r}{\delta} \right]
\]
where the second term decays with increasing \( r \); it is attenuated. The amplitude decreases by the factor of \( e^{-1} \approx 0.368 \) of the incident value when \( r \) is equal to the skin depth \( \delta \):
\[ \delta \equiv \left( \frac{\kappa \omega_0}{c} \right)^{-1} = \frac{c}{\kappa \omega_0} = \frac{\lambda_0}{2\pi \kappa} \]
where \( \lambda_0 \) is the wavelength measured in vacuum. The skin depth \( \delta \) is a measure of the distance that light will penetrate through the attenuating medium. In a metal, the imaginary part \( \kappa \) of the index of refraction can be large, which means that the skin depth is small and the electric field “lies” on the surface of the metal; little field penetrates to the interior.

If \( n \) (and thus \( k \)) is real valued (i.e., \( \kappa = 0 \), as is true for optically transparent media), then the electric field is
\[
E_0 \exp \left[ i\omega_0 \left( \frac{n}{c} r - t \right) \right] = E_0 \exp \left[ i\omega_0 \left( \frac{r}{v} - t \right) \right] = E_0 \exp \left[ i (n k_0 r - \omega_0 t) \right]
\]
which confirms that the velocity is
\[ v = \frac{c}{n} \]
and that the wavelength in the medium is:
\[ \lambda'_0 = \frac{\lambda_0}{n} \]
In the case of complex-valued refractive index, the magnetic field is obtained from the electric field via:

\[ \mathbf{B} = \frac{k \times \mathbf{E}}{\omega} = \frac{\tilde{n}}{c} (\hat{s} \times \mathbf{E}) \]

where \( \hat{s} \) is the unit vector in the direction of propagation of the light (the Poynting vector). Note that some authors write \( \tilde{n} = n (1 + i\kappa) \), where \( \kappa \) is the attenuation index.

Values of the refractive index for common materials include:

<table>
<thead>
<tr>
<th>Medium</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>vacuum</td>
<td>1.0 (by definition)</td>
</tr>
<tr>
<td>air</td>
<td>( \approx 1.00027 )</td>
</tr>
<tr>
<td>water</td>
<td>1.33</td>
</tr>
<tr>
<td>“crown” glass</td>
<td>1.5</td>
</tr>
<tr>
<td>“flint” glass</td>
<td>1.7</td>
</tr>
<tr>
<td>diamond</td>
<td>2.417</td>
</tr>
<tr>
<td>germanium</td>
<td>( \approx 4.0 ) (only transparent for ( \lambda \gtrsim 2\mu m ))</td>
</tr>
</tbody>
</table>

### 6.10 Consequences of Maxwell’s Equations

1. **Copropagation of \( \mathbf{E} \) and \( \mathbf{B} \):** The wave travels in a direction mutually perpendicular to both \( \mathbf{E} \) and \( \mathbf{B} \), and in fact the propagation direction is defined by the direction:

\[ \hat{s} \propto \mathbf{E} \times \mathbf{B} \text{ – the Poynting vector} \]

In other words, the wave requires both electric and magnetic fields to propagate, and they copropagate. I like to interpret this result as meaning that the magnetic field provides the medium for propagation of the electric field and vice versa.

2. The electric and magnetic fields of an electromagnetic wave are mutually perpendicular.

3. In vacuum, \( \mathbf{E} \) and \( \mathbf{B} \) are in-phase, which means that the phases of the sinusoidal variation of \( \mathbf{E} \) and \( \mathbf{B} \) are identical (the phases of the fields often are out of phase in some types of matter).

4. Both \( \mathbf{E} \) and \( \mathbf{B} \) travel at \( c \), the phase velocity of the wave.

5. Energy is carried by both the electric and magnetic fields, and the magnitude of the energy \( \mathcal{E} \propto \mathbf{E}^2 \).
6. There is no limitation on the possible frequencies of the waves, i.e., \([0 \leq \omega \leq \infty]\), which implies the allowed wavelengths are in the interval \([\infty \geq \lambda \geq 0]\).

7. The average power of the light wave per unit area is the “irradiance,” and is determined from the Poynting vector

\[
I[x, y, z, t] = \frac{1}{2} (cE_0^2)
\]

Relationship between \(\mathbf{E}\) and \(\mathbf{B}\) for a linearly polarized wave traveling from left to right; the fields are in phase.

### 6.11 Dispersion Redux

Earlier we considered the effects of dispersion on traveling waves from a simple point of view where we just assumed that waves with different temporal frequencies might travel at different speeds. We called the dispersion “normal” if the velocity of waves with longer wavelengths exceeds that of waves with shorter wavelengths, which also means that the index of refraction decreases with increasing wavelength. At this point, we will try to understand why this is so and also determine the conditions that are necessary for anomalous dispersion (where the index of refraction increases with increasing wavelength).

The first theory of dispersion, based on the understanding of elastic solids, was put forth by Cauchy in 1836. He observed a relationship between the phase velocity of light in a medium and the \textit{elasticity} \(\varepsilon\) of the solid (the restoring force exerted upon a displaced particle by a neighboring particle) and the density \(\rho\) of the medium:

\[
v_\phi \propto \frac{\sqrt{\varepsilon}}{\rho}
\]

if measured at wavelengths much longer than the scale of the vibrating particles in the medium. Cauchy deduced a formula for the frequency dependence of the index of refraction that bears his name:

\[
n \cong A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}
\]
where $A$, $B$, and $C$ are constants determined from measurements of $n$ at three wavelengths. This expression gives a pretty good approximation of $n$. For example, the Cauchy formula for the index of refraction of air is:

$$n \approx 1.000287566 + \frac{1.3412 \text{nm}^2}{\lambda^2} + \frac{0}{\lambda^4}$$

which decreases with increasing wavelength.

The phenomenon of anomalous dispersion may have been observed (but not pursued) by William Fox Talbot. The first significant study was performed by Le Roux in 1862, who discovered that the index of refraction of a prism containing iodine vapor was 1.020 for red light and 1.019 for blue light, so that $n_{\text{blue}} < n_{\text{red}}$. Within 10 years, Christiansen noted similar behavior in an analine dye that exhibited a strong absorption of green light, normal refraction for red, orange, and yellow light, but a smaller refraction angle for blue light. The refractive index of a medium that exhibited strong absorption was seen to increase rapidly as the wavelength was decreased approaching the absorption band. It took some time to produce a theory of matter and light that explained this effect.

The reason for dispersion is due to the interaction of light with matter. Light can be absorbed by matter, where the energy of the light wave is converted to energy of some form in the matter (e.g., it may increase the thermal energy in the matter. Light also can be scattered, where the electric charges in the matter (protons in the nuclei or, more usually, electrons in the atomic shells) absorb and then re-emit electromagnetic waves. If the scattering is elastic, then no energy is transferred to the medium, all of it stays in the electromagnetic wave. If inelastic, some energy is transferred to the medium. Scattering generally occurs when the waves encounter a structure or obstacle whose dimension is smaller than a wavelength, and the wave is re-emitted into a new direction. Electromagnetic waves are scattered by electric charges that may be bound in an atom or free (unbound). A material like glass contains many charges, and the macroscopic effect is the sum of the effects from each individual charge.

The interactions of light with matter are characterized by two numerical factors: the “absorption coefficient” $\alpha$ and the (possibly complex valued) “refractive index” $n$, which both may be functions of the frequency of the incident light. The absorption is due to transfer of energy from the light to the medium; at frequencies where the absorption coefficient is small, the light can penetrate the matter to a significant depth, and thus the matter is “transparent.”

Light impinging on a medium causes the charged particles (protons in atomic nuclei or electrons in atomic shells) to vibrate at the oscillation frequency of the light. These accelerated charges emit light of that same oscillation frequency in turn (this is the scattered light). The relative phase of the incident light and the re-emitted light determines much of the effect of the medium on the incident light. For example, if the incident and scattered waves are out of phase by $\approx 180^\circ$ in some direction, then the light beam propagating in that direction will be attenuated.

The oscillations damp out because the electrons are influenced by other forces,
including the electric forces induced by neighboring charges. Thus the interaction of light with the medium acts like a damped harmonic oscillator that is driven by the sinusoidal force induced by the light wave. We can think of the interaction of light with charged particles in matter as a so-called “driven” or “forced” harmonic oscillator whose amplitude decreases with time due to the damping. The charges are “bound” to fixed equilibrium locations and can oscillate with one or more resonant frequencies determined by the internal forces due to neighboring atoms. The electric charges can absorb and re-emit the light (i.e., scatter it) in ANY direction. If the frequency of the light is close to the oscillation frequency of one of the resonant states, then some of the electromagnetic energy is retained by the charge and not scattered; it instead increases the energy of the charge. The wave then loses some amplitude when scattered. If the frequency of the light is far from a resonant frequency of the charges, then the scattered light constructively interferes along the same direction as the incident light and it can pass through the medium; in other words, the medium is transparent for light with those frequencies. Even in transparent media, the phase of the re-emitted light generally is shifted by the interaction of the medium.

The first question to consider is the reason why the light is “forward scattered” in transparent media. We’ve actually already given the answer; the light emitted in this direction by all atoms interferes constructively and light emitted in other directions interferes destructively. The charges in matter that scatter the light may be viewed as uniformly distributed and quite close together (separated by fractions of nanometers, significantly less than visible wavelengths). The phase “lag” of the forward-scattered light is equivalent to a “slowing down” of the light wave, hence the index of refraction.

6.11.1 Feynman’s Model for Refractive Index

From my point of view, the best discussion of refraction and dispersion was given by Richard Feynman in his famous Lectures on Physics, Volume I, Chapter 31. This discussion is derived from his treatment.

Consider an electromagnetic wave incident on a thin plate of glass. The source point is assumed to be a large distance away (to the left) and the observation point also is a large distance to the right (thus the figure is “not to scale!”).
Electric waves interacting with a layer of transparent glass: some of the field is reflected and the original field plus a “correction” term are transmitted.

The electric field at the measurement point is the sum of the original electric field plus a correction term $E_2$ due to emission by the electric charges in the glass plate. We can assume that the charges are electrons, since the protons are much more massive. The electrons in the glass oscillate under the influence of the incoming electric field from the source and thus emit their own electric fields. The observed field includes contributions from these charges. These modifications occur in a way that makes the field inside the glass “appear” to be moving at a different phase velocity – this is the reason why the index of refraction of glass is larger than one.

Mathematical Picture of Phase Change of Light Due to Glass

Consider one electron in the glass; it “feels” the effect of the incident field and the fields generated by all of the other charges in the glass, and the motions of all of these other charges are influenced by that one electron that we are observing. To simplify the problem, we assume that the influences of the other atoms are small relative to the effect of the source, so that the total field at the observation point is little affected by the motions of the other charges. In effect, we are assuming that the index of refraction of the glass is very small (close to one). The calculation will produce a field that travels in the same direction as the incident field ($E_2$) and a field that travels in the opposite direction ($E_1$ - the “reflected” field), but the latter is small because $n \approx 1$.

Because its source is far away, the incident electric field is a traveling plane wave that may be written in complex notation:

$$E_s = E_0 \exp \left[ +i \left( k_0 z - \omega t \right) \right] = E_0 \exp \left[ +i\omega \left( \frac{z}{c} - t \right) \right]$$

where the factoring conveniently leaves the distance $z$. Here $\omega$ is the angular temporal frequency of the driving force, and thus is a variable in the problem (rather than a
6.11 DISPERSION REDUX

parameter of the wave), hence it does not have a subscript. Assume that \( z = 0 \) at the “front” (input) side of the plate and \( z = \Delta z \) at the back side, so that the phase of the electric field at the front of the plate is:

\[
\phi [0, t] = -\omega t
\]

If there were no glass, then the phase at the back of the plate observed would be identical to that at the front of the plate observed at an earlier time \( t' \):

\[
\phi [\Delta z, t'; \text{vacuum}] = \omega \left( \frac{\Delta z}{c} - t \right) = -\omega t + \frac{\Delta z}{c}
\]

where the phase increment is due to the delay during the transfer of the extra distance \( \Delta z \). If the glass “slows down” the light so that the velocity in the glass is \( v < c \), then the phase at the rear of the glass will include an additional factor

\[
\phi [\Delta z, t; \text{glass}] = \omega \left( \frac{\Delta z}{v} - t \right) > \omega \left( \frac{\Delta z}{c} - t \right)
\]

If we substitute the index of refraction, then the distance parts of the phase are proportional

\[
n = \frac{c}{v}
\]

\[
\implies \phi [\Delta z, t; \text{glass}] = \omega \left( \frac{n \cdot \Delta z}{c} - t \right) > \omega \left( \frac{\Delta z}{c} - t \right)
\]

Thus the “additional” phase due to the extra time to travel through the glass is:

\[
\omega \left( \frac{n \cdot \Delta z}{c} - t \right) - \omega \left( \frac{\Delta z}{c} - t \right) = \omega \left( n - \frac{1}{n} \right) \cdot \Delta z
\]

The electric field at the back of the glass is:

\[
E_{after} = E_0 \exp \left[ +i\omega \left( \frac{\Delta z}{c} - t \right) \right] \cdot \exp \left[ +i\omega \left( \frac{n - 1}{c} \cdot \Delta z \right) \right]
\]

Thus the contribution to the electric field due to the glass plate may be interpreted as an additive contribution to the phase instead of an additive contribution to the amplitude. The term with \( (n - 1) \) is the change due to the glass and will be evaluated
by physical arguments in a moment. It may be expanded in a Taylor series:

\[
\exp[+i\theta] = \frac{(+i\theta)}{0!} + \frac{(+i\theta)^1}{1!} + \frac{(+i\theta)^2}{2!} + \cdots + \frac{(+i\theta)^N}{N!} + \cdots
\]

\[
\exp\left[+i\omega \left(\frac{n-1}{c} \cdot \Delta z\right)\right] = 1 + i\omega \left(\frac{n-1}{c} \cdot \Delta z\right) + \frac{(+i)^2}{2!} \left(\omega \left(\frac{n-1}{c} \cdot \Delta z\right)\right)^2 + \cdots + \frac{(+i)^N}{N!} \left(\omega \left(\frac{n-1}{c} \cdot \Delta z\right)\right)^N + \cdots
\]

If we assume that the glass is thin, so that \(\Delta z \gg 0\), then we can ignore all terms of order two or larger:

\[
\exp\left[+i\omega \left(\frac{n-1}{c} \cdot \Delta z\right)\right] \approx 1 + i\omega \left(\frac{n-1}{c} \cdot \Delta z\right)
\]

and thus the electric field at a large distance “behind” the glass plate is approximately:

\[
E_{after} = E_s + E_2 \approx E_0 \exp\left[+i\omega \left(\frac{z}{c} - t\right)\right] \cdot \left(1 + i\omega \left(\frac{n-1}{c} \cdot \Delta z\right)\right)
\]

\[
= \left(E_0 \exp\left[+i\omega \left(\frac{z}{c} - t\right)\right]\right) + i\omega \left(\frac{n-1}{c} \cdot \Delta z\right) \left(E_0 \exp\left[+i\omega \left(\frac{z}{c} - t\right)\right]\right)
\]

The first term is just the source field at the front of the plate:

\[
E_0 \exp\left[+i\omega \left(\frac{z}{c} - t\right)\right] = E_s
\]

and the second term is identified as the contribution from the charges within the glass, what is labeled in the Figure as \(E_2\):

\[
+i\omega \left(\frac{(n-1) \cdot \Delta z}{c}\right) \left(E_0 \exp\left[+i\omega \left(\frac{z}{c} - t\right)\right]\right) \approx E_2
\]

We know that the leading factor \(+i = \exp[+i \frac{\pi}{2}]\), which indicates that the electric field from the charges in the glass is out of phase with the original electric field by \(+\frac{\pi}{2}\) radians, as shown on the vector (phasor) diagram of the contributions of the two fields:

![Argand diagram of phasor contributions from incident field \(E_0\) and field due to charges in the glass \(E_2\), which is oriented approximately perpendicular to \(E_0\) and](image-url)
“delays” the phase of the electric field.

### Physical Picture of Electron Oscillations

If the field $E_2$ thus evaluated can be expressed in terms of the oscillating charges in the glass, then we will have explained the behavior of the refractive index. Again, we assume that the incident field has the form of a plane wave:

$$E_s [z, t] = E_0 \exp \left[ +i (k_0 z - \omega t) \right]$$

At the “front edge” of the glass ($z = 0$), the field is the same as given before:

$$E_s [0, t] = E_0 \exp \left[ -i \omega t \right]$$

The electrons in the glass “feel” this electric field and are driven in the same direction by the force:

$$F = e E_s [0, t] = e E_0 \exp \left[ -i \omega t \right]$$

(there also is a magnetic field, but its effect on the electrons is so much smaller that it can be ignored). The electrons have mass $m$ and act as though bound to the atoms by little springs that exert restoring forces proportional to the distance of the electron from its equilibrium position. The restoring force on the electron position has the form:

$$F = -k (x - x_0)$$

The electron oscillates at its “normal frequency,” which is:

$$\sqrt{\frac{k}{m}} \equiv \omega_0$$

this is a parameter of the electron + spring system (and thus has a subscript). The equation of motion of the electrons in the glass is:

$$m \frac{d^2 x}{dt^2} + m \omega_0^2 x = F = e E_0 \exp \left[ -i \omega t \right]$$

where the last term is the “driving force” due to the electric field with the variable angular temporal frequency $\omega$. We solve this equation by standard methods of differential equations; we assume that the position $x$ of each electron also oscillates about its equilibrium point at the same rate. The amplitude of the oscillation is assumed to be $x_0$:

$$x = x_0 \exp \left[ -i \omega t \right]$$

The derivatives needed in the equation of motion are

$$\frac{dx}{dt} = -i \omega x_0 \exp \left[ -i \omega t \right] = -i \omega x$$

$$\frac{d^2 x}{dt^2} = (-i \omega)^2 x_0 \exp \left[ -i \omega t \right] = -\omega^2 x$$
So the equation of motion is simplified to:

\[ m \left( -\omega^2 x_0 \right) \exp \left[ -i\omega t \right] + m\omega_0^2 x_0 \exp \left[ -i\omega t \right] = e E_0 \exp \left[ -i\omega t \right] \]

\[ \Rightarrow \left( -m\omega^2 + m\omega_0^2 \right) x_0 = e E_0 \]

And thus the amplitude of the electron displacement from equilibrium may be expressed in terms of the “normal” oscillation frequency, mass, electric charge, and the amplitude and oscillation frequency of the incident electric field:

\[ x_0 = \frac{e E_0}{m (\omega_0^2 - \omega^2)} \]

Therefore the motions of EACH individual charge in the thin glass plate due to the incident electric field is the simple expression:

\[ x \left[ t \right] = x_0 \exp \left[ -i\omega t \right] = \frac{e E_0}{m (\omega_0^2 - \omega^2)} \exp \left[ -i\omega t \right] \]

Note that this does not include the initial positions of the charges, which are (obviously) different for each.

We now must calculate the field at the observation point (distant from the plate) due to a “thin plane” of charges that all move with this same motion \( x \left[ t \right] \). We find the field at the observation point by adding the contributions from each of the charges in the glass. The electric field radiated by each electron in the glass is proportional to the acceleration just evaluated:

\[ \frac{d^2 x}{dt^2} = (-i\omega)^2 x_0 \exp \left[ -i\omega t \right] = -\omega^2 x \]

The electric field at large distances from the charge that oscillates perpendicularly to the distance decreases approximately as the reciprocal of the distance and includes the time delay for the field to arrive:

\[ E_e \left[ r, t \right] \approx \frac{e}{r} \left( -\omega^2 x_0 \right) \exp \left[ -i\omega \left( t - \frac{r}{c} \right) \right] \]

We assume that the observation point is so far away that the field oscillates approximately perpendicular to the “line of sight.”

The total field at the observation point is the vector sum of the contributions from the individual electrons in the thin glass plate. These contributions may be integrated in polar coordinates. If \( \eta \) is the number density of the electrons per unit area in the glass, then the electric field is:

\[ E_{all} = \int_{\rho=0}^{\rho=+\infty} \frac{e}{r} \left( -\omega^2 x_0 \right) \exp \left[ -i\omega \left( t - \frac{r}{c} \right) \right] \cdot \eta \cdot 2\pi \rho \cdot d\rho \]

\[ = 2\pi \eta e \left( -\omega^2 x_0 \right) \exp \left[ -i\omega t \right] \int_{\rho=0}^{\rho=+\infty} \exp \left[ +i\omega \frac{r}{c} \right] \cdot \frac{\rho}{r} \cdot d\rho \]
where 
\[ r^2 = \rho^2 + z^2 \implies r \, dr = \rho \, d\rho \]

Therefore:
\[
\int_{\rho=0}^{\rho=+\infty} \exp \left[ +i \omega \frac{r}{c} \right] \cdot \frac{\rho}{r} \cdot d\rho = \int_{r=z}^{r=+\infty} \exp \left[ +i \omega \frac{r}{c} \right] \cdot \frac{r}{r} \cdot dr
\]
\[
= \int_{r=z}^{r=+\infty} \exp \left[ +i \omega \frac{r}{c} \right] \cdot dr
\]
\[
= \frac{c}{i\omega} \left( \exp \left[ +i \cdot \infty \right] - \exp \left[ +i \omega \frac{z}{c} \right] \right)
\]

Note that the term \( \exp \left[ +i \cdot \infty \right] \) oscillates, and rapidly, so we can assume it to be zero:

\[
E_{\text{all}} = 2\pi \eta e \left( -\omega^2 x_0 \right) \exp \left[ -i\omega t \right] \cdot -\frac{c}{i\omega} \exp \left[ +i \omega \frac{z}{c} \right]
\]
\[
= \left( 2\pi \eta e x_0 \right) \cdot (-i\omega) \exp \left[ -i\omega \left( t - \frac{z}{c} \right) \right]
\]
\[
= \left( 2\pi \eta e \right) \cdot (-i\omega) x_0 \exp \left[ -i\omega t \right] \exp \left[ +i \omega \frac{z}{c} \right]
\]

which shows that the field measured at the observation point due to all of the oscillating charges is out of phase by \( \frac{\pi}{2} \) radians and delayed.

All we need to do now is substitute the formula for \( x_0 \) from the driven harmonic oscillator that was derived in the previous section:

\[
E_2 = \left( 2\pi \eta e \right) \cdot (-i\omega) \frac{eE_0}{m \left( \omega_0^2 - \omega^2 \right)} \exp \left[ -i\omega t \right] \exp \left[ +i \omega \frac{z}{c} \right]
\]

In words, the electrons in the glass that oscillate due to the incident field emit a wave that travels in the same direction (towards \( z = +\infty \)). The amplitude of the wave is proportional to the number of atoms \( \eta \) and to the strength of the source field \( E_0 \). This field resembles that evaluated before (in the earlier box)

\[
E_2 \propto -i\omega \left( \frac{n-1}{c} \right) \cdot \Delta z \left( E_0 \exp \left[ +i \omega \left( \frac{z}{c} - t \right) \right] \right)
\]

These two are equal if we identify that some of the factors are identical:

\[
\frac{(n-1) \cdot \Delta z}{c} = \frac{\left( 2\pi \eta e c^2 \right)}{m \left( \omega_0^2 - \omega^2 \right)}
\]
\[
\implies (n-1) \Delta z = \frac{\left( 2\pi \eta c^2 e^2 \right)}{m \left( \omega_0^2 - \omega^2 \right)}
\]
\[
\implies n = 1 + \frac{\left( 2\pi c^2 e^2 \right)}{m \left( \omega_0^2 - \omega^2 \right)} \cdot \frac{\eta}{\Delta z}
\]
We now define $N$ to be the number density of electrons per unit volume in the glass, which can be expressed as the product of the number density $\eta$ per unit area in the thin sheet and the thickness $\Delta z$:

$$N = \eta \cdot \Delta z \implies \frac{\eta}{\Delta z} = N$$

Thus the index of refraction can be written as:

$$n = 1 + \frac{2\pi N e^2}{m(\omega_0^2 - \omega^2)} = 1 + \frac{\alpha}{\omega_0^2 - \omega^2}$$

where the constant $\alpha$ includes the contributions from the number density, the mass, the charge on the electron, and the velocity of light. This is the frequency dependent index of refraction in the simple model of a thin sheet of glass with oscillating bound electrons. This expression is graphed for the case $\omega_0 = 2$ and $\alpha = 1$, showing that the index is largest for light with temporal frequency $\omega \equiv \omega_0$, i.e., in the vicinity of the frequency of the “resonant” oscillations of the electrons due to the restoring forces. In this simple model, the index of refraction can be less than unity, which indicates that this picture is incomplete.

In thick glass, the fields from different thin sheets interact which increases the mean refractive index. Also the oscillations of the electrons are actually “damped” out by other forces in the glass. The damping pushes the extrema of $n$ towards its mean value. To derive this behavior, we could continue the discussion in the same vein (most books do, including Feynman and Hecht). We could follow this line of reasoning, but we can demonstrate a useful connection to linear systems theory and thus simplify the derivation.
We can think of the action of light on matter as a linear shift-invariant system with a “causal” impulse response, i.e., a temporal system whose impulse response is zero for \( t < 0 \) and thus cannot “respond” until stimulated. An appropriate impulse response for a “causal” damped oscillating system is:

\[
    h[t] = A_0 \exp[-\gamma_0 t] \cdot \text{STEP}[t] \cdot \sin[2\pi \nu_0 t]
    = A_0 \exp[-\gamma_0 t] \cdot \text{STEP}[t] \cdot \cos[2\pi \nu_0 t - \frac{\pi}{2}]
\]

where \( \gamma_0 \) is the damping coefficient (i.e., the reciprocal of the time required for the output response to decrease by \( e^{-1} \approx 0.368 \)), \( \nu_0 \) is the natural oscillating frequency of the charged particle (the electron), and \( \phi_0 \) is the initial phase. This function \( h[t] \) measures the “response” of the system, which in our case is the position of the electron relative to its equilibrium position. The STEP function ensures that \( h[t < 0] = 0 \), meaning that the electron “sits” at its equilibrium position until disturbed; this forces the system to be causal. When stimulated by a “pulse of light,” (modeled by a Dirac delta function \( \delta[t] \)), the position of the electron increases from zero following a sinusoidal curve, but the amplitude of the sine wave decreases with time due to the decaying exponential, as shown:

The response of this system is characterized by the transfer function of the impulse response, which is its 1-D temporal Fourier transform. We use the known Fourier
transforms:

\[
\mathcal{F}_1 \{ e^{-t} \cdot \text{STEP} [t] \} = \frac{1}{1 + 2\pi i \nu} = \frac{1 + i(-2\pi \nu)}{1 + (2\pi \nu)^2}
\]

\[
\mathcal{F}_1 \{ \sin [2\pi \nu_0 t] \} = i \cdot \frac{1}{2 |\nu_0|} (\delta [\nu + \nu_0] - \delta [\nu - \nu_0])
\]

and the scaling theorem:

\[
\mathcal{F}_1 \{ f [t] \} = F [\nu] \implies \mathcal{F}_1 \{ f [\gamma_0 t] \} = \frac{1}{|\gamma_0|} F \left[ \frac{\nu}{\gamma_0} \right]
\]

to derive the transfer function of this system:

\[
H [\nu] = A_0 \mathcal{F}_1 \{ \sin [2\pi \nu_0 t] \cdot (\exp [-\gamma_0 t] \cdot \text{STEP} [t]) \}
\]

\[
= A_0 \mathcal{F}_1 \{ \sin [2\pi \nu_0 t] \cdot (\exp [-\gamma_0 t] \cdot \text{STEP} [\gamma_0 t]) \}
\]

\[
= i \left( \frac{A_0}{2 |\gamma_0|} \right) (\delta [\nu + \nu_0] - \delta [\nu - \nu_0]) \ast \frac{1}{1 + 2\pi i \left( \frac{\nu}{\gamma_0} \right)}
\]

\[
= i \left( \frac{A_0}{2 |\gamma_0|} \right) (\delta [\nu + \nu_0] - \delta [\nu - \nu_0]) \ast \frac{1 - 2\pi i \left( \frac{\nu}{\gamma_0} \right)}{1 + \left( \frac{2\pi \nu}{\gamma_0} \right)^2}
\]

\[
\text{Re} \{ H [\nu] \} = \frac{A_0 \pi}{\gamma_0^2} \left( \frac{\nu + \nu_0}{1 + \left( 2\pi \frac{\nu + \nu_0}{\gamma_0} \right)^2} - \frac{\nu - \nu_0}{1 + \left( 2\pi \frac{\nu - \nu_0}{\gamma_0} \right)^2} \right)
\]

\[
= \frac{A_0}{2 \gamma_0^2} \left( \frac{\omega + \omega_0}{\gamma_0^2 + (\omega + \omega_0)^2} - \frac{\omega - \omega_0}{\gamma_0^2 + (\omega - \omega_0)^2} \right)
\]

\[
\text{Im} \{ H [\nu] \} = \frac{A_0}{2 \gamma_0} \left( \frac{1}{\gamma_0^2 + (\omega + \omega_0)^2} - \frac{1}{\gamma_0^2 + (\omega - \omega_0)^2} \right)
\]

The transfer function acts as a frequency-dependent “scale factor” applied to the amplitude of the electron oscillation. The graphs of the real part, imaginary part, magnitude, and phase are shown below, where the domain is assumed to include negative temporal frequencies. In this example, \( \nu_0 = 2 \) and \( \gamma_0 = 0.5 \).
In words, the transfer function measures the “response” of the system, i.e., the amplitude of the oscillation of the charged particle, which we called $x[t]$ before.

$$x[t] \propto x_0 \cdot \left( \text{Re} \{H[\nu]\} + i \text{Im} \{H[\nu]\} \right) \cdot \exp[-i\omega t]$$

$$= \frac{A_0 x_0}{2\gamma_0^2} \exp[-i\omega t] \left[ \left( \frac{\omega + \omega_0 + i\gamma}{\gamma_0^2 + (\omega + \omega_0)^2} - \frac{\omega - \omega_0 - i\gamma}{\gamma_0^2 + (\omega - \omega_0)^2} \right) \right]$$

The amplitude is again a function of the temporal frequency of the incident light. Note that the phase of the transfer function is approximately 0 radians for $|\nu| < |\nu_0|$, approximately $-\pi$ radians for $\nu > \nu_0$, and $-\frac{\pi}{2}$ radians if the frequency of the incident light is $+\nu_0$. This means that the system response (the oscillation of the charged particle) is “in phase” if the frequency of the incident light is less than the “resonant frequency” of the oscillating charge. The oscillation of the charged particle is “out of phase” if the frequency of the incident light is larger than the resonant frequency. Also note from the magnitude that the system response is quite large near resonance, which means that the oscillation of the charged particle is large. Of course the meaning of light with a negative temporal frequency is not very clear in this context, and can be rectified by recognizing that the real and imaginary part of the response of the system are related due to causality, a fact reflected in the Kramers-Kronig equations, which are beyond the scope of this discussion.

The index of refraction is proportional to this amplitude plus a bias, so we can use the graphs of $H[\nu]$ to understand the frequency behavior of $n$. As mentioned previously, the index of refraction decreases with increasing wavelength (increases
with increasing temporal frequency), thus the phase velocity in a medium of light with longer wavelengths is larger and the dispersion is normal. The phase velocity of the modulation wave (the group velocity) is less than the phase velocity of the average wave, and messages travel more slowly than the carrier wave that conveys the message.

In the vicinity of an absorption due to the resonance of charged particles in the medium, the index of refraction increases with increasing wavelength over a small range, which means that shorter wavelengths travel faster and the dispersion is anomalous. In this region, the phase velocity of the modulation wave is larger than the phase velocity of the average wave. This implies that messages can travel faster than the velocity of light. HOWEVER, since this only happens where light is absorbed, the message cannot propagate.

Real and imaginary parts of refractive index in vicinity of a “weak” absorption.

Real and imaginary parts of refractive index for multiple absorptions. Note that anomalous dispersion only occurs in vicinity of absorptions, so that light cannot propagate (from Fowles, Introduction to Modern Optics, Dover, 1975).
6.12 Dual Nature of Light: Photons

In many contexts, the “particle” picture of light is more appropriate. In imaging, for example, consider images created with different exposure times. Photographs taken in with shorter exposures generally look grainier:

Images created from increasing numbers of photons, showing increase in signal-to-noise ratio.

A “particle of light” is the photon, whose energy is proportional to the temporal frequency:

\[ E = h\nu = h\frac{c}{\lambda} = \hbar\omega \]

where \( h \) is Planck’s constant, which is often normalized by a factor of \( 2\pi \), called “h-bar:”

\[ h \approx 6.625 \times 10^{-34} \text{ J} - \text{s} = 6.625 \times 10^{-27} \text{ erg} - \text{s} \]

\[ \hbar = \frac{h}{2\pi} \approx 1.054 \times 10^{-34} \text{ J} - \text{s} \]

If \( \lambda = 550 \text{ nm} \), the energy per photon is only:

\[ E = (6.625 \times 10^{-34} \text{ J} - \text{s}) \cdot \frac{3 \times 10^8 \text{ m/s}}{550 \text{ nm}} \approx 3.6 \times 10^{-19} \text{ J} \]

The “photon flux” is the number of photons per second in a light beam:

\[ \Phi = \frac{P}{h\nu}, \text{ where } P \text{ is the power} \]

Typical fluxes per unit area for some sources are shown in the table.
### Light Source \( \Phi/A \) in \( \text{photons/sec-m}^2 \)

<table>
<thead>
<tr>
<th>Light Source</th>
<th>( \Phi/A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>focused laser</td>
<td>( 10^{26} )</td>
</tr>
<tr>
<td>unfocused laser</td>
<td>( 10^{21} )</td>
</tr>
<tr>
<td>bright sunlight</td>
<td>( 10^{18} )</td>
</tr>
<tr>
<td>indoor light</td>
<td>( 10^{16} )</td>
</tr>
<tr>
<td>twilight</td>
<td>( 10^{14} )</td>
</tr>
<tr>
<td>moonlight</td>
<td>( 10^{12} )</td>
</tr>
<tr>
<td>starlight</td>
<td>( 10^{10} )</td>
</tr>
</tbody>
</table>

The pattern of photon arrivals tells something about the source. Random (incoherent) light sources (such as light bulbs) emit photons with random arrival times and a Bose-Einstein distribution. Coherent light sources, on the other hand, emit photons with a Poisson distribution, which is more uniform but still random.

#### 6.12.1 Momentum of Photons

Atoms that emit photons “recoil” in the opposite direction, and surfaces that absorb photons also recoil. The momentum of a single photon is

\[
p = \frac{h}{\lambda} = \hbar k
\]

The pressure due to radiation is the force per unit area, which is equal to the energy per unit volume, or the energy density. Radiation pressures are often neglected, but cannot be if the mass is small or the flux is large, e.g., in the motion of comet tails or spacecraft, in stellar interiors, and in the light of lasers.
6.13 Optical Frequencies – Detector Response

The general equation for a traveling electromagnetic wave is:

\[ y(z, t) = A_0 \cos (kz \mp \omega t) = A_0 \cos \left( \frac{2\pi (z - \nu t)}{\lambda} \right) = A_0 \text{Re} \{ e^{i(kz \mp \omega t)} \} \]

We see electromagnetic radiation with detectors, i.e., devices which respond in some way to incident electromagnetic radiation. The human eye is sensitive only to visible light, i.e., light with wavelengths in the range \(400 \text{ nm} \leq \lambda \leq 700 \text{ nm}\). This is not the case for all life, however. The pit viper can see radiation emitted by humans at a wavelength of about \(10 \mu m\); it needs special receptors on the sides of its head to do this.

As shown in the plot of the electromagnetic spectrum, the frequencies of visible wavelengths are quite large: \(\nu \simeq 10^{15} \text{ Hz}\). The temporal period of an optical wave is therefore \(T = \nu^{-1} \simeq 10^{-15} \text{ s}\). Human visual receptors cannot respond fast enough to detect the periodic oscillation of the wave amplitude; we see an invariant brightness. Note that this limitation exists for all detectors of visible radiation (e.g., photographic film, light meters, etc.); they all respond to the average brightness. The same is true for hearing; your ear cannot detect the variation of sound pressure due to the oscillation at frequencies above a few Hz. Because water waves have a much lower frequency, the amplitude and phase of the wave can be measured. Similarly, the phase can be measured of electromagnetic waves that have a much smaller temporal frequency, e.g., radio waves.

The average amplitude of a sinusoidal wave is:

\[
\langle y[z, t] \rangle = \frac{1}{T_d} \int_0^{T_d} y[z, t] \, dt = \frac{1}{T_d} \int_0^{T_d} A_0 \cos [kz - \omega t] \, dt = -\frac{A_0}{\omega T_d} \sin [kz - \omega t] \bigg|_{t=0}^{t=T_d}
\]

Since \(y[z, t]\) is sinusoidal, the average value of the wave will tend to zero unless \(T_d\) is smaller than the wave’s temporal period. However, the intensity (squared-magnitude)
of the wave does not average to zero:

$$\mathcal{E} \propto \langle y^2 [z, t] \rangle = \frac{1}{T_d} \int_0^{T_d} y^2 [z, t] \, dt$$

$$= \frac{1}{T_d} \int_0^{T_d} A_0^2 \cos^2 [kz - \omega t] \, dt$$

$$= \frac{A_0^2}{T_d} \int_0^{T_d} \cos^2 [kz - \omega t] \, dt$$

$$= \frac{A_0^2}{T_d} \cdot \frac{T_d}{2} = \frac{A_0^2}{2}$$

$$\Rightarrow \mathcal{E} \propto \langle y^2 [z, t] \rangle \Rightarrow \frac{A_0^2}{2} \text{ if } T_d >> \nu^{-1}$$

because the average value of $\cos^2 [x] = \frac{1}{2}$.

*Detectors of visible light are sensitive to time-averaged intensity, not amplitude.*