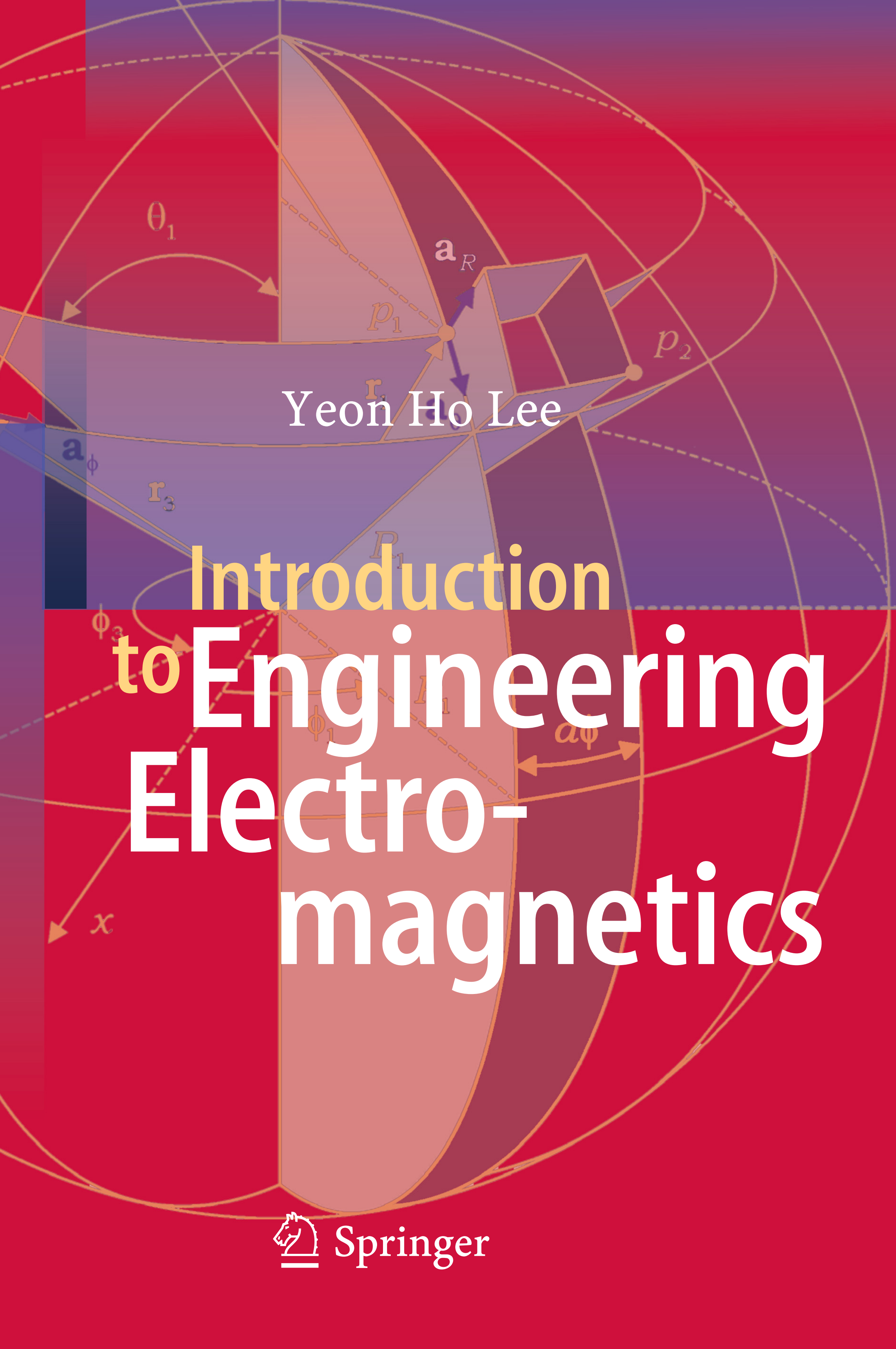
Yeon Ho Lee

Introduction to Engineering Electromagnetics



Springer

Introduction to Engineering Electromagnetics

Yeon Ho Lee

Introduction to Engineering Electromagnetics

Springer

Author
Prof. Yeon Ho Lee
School of Information and Communication Engineering
Sungkyunkwan University
Kyongkido
Republic of Korea

ISBN 978-3-642-36117-3 e-ISBN 978-3-642-36118-0
DOI 10.1007/978-3-642-36118-0
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012955857

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my parents

Preface

This is a text book on engineering electromagnetics, designed for an undergraduate course at the sophomore or junior level. The book can be covered in two semesters. The first part begins with vector algebra and coordinate systems covered in Chapter 1, and vector calculus covered in Chapter 2. The two chapters can take about half a semester for a full coverage. Although the instructor may skip some materials in those chapters, the students may use them as references. Chapter 3 discusses electrostatics, and Chapter 4 reviews currents. Chapter 5 deals with magnetostatics. The second part of the book begins with time-varying fields and Maxwell's equations covered in Chapter 6, and wave motion in general is covered in Chapter 7. In the two chapters, students learn about the interrelationship between time-varying electric and magnetic fields, and the concept of plane waves. Chapter 8 discusses the propagation of electromagnetic waves in material media. Chapter 9 discusses transmission lines, and Chapter 10 is on waveguides.

Although electromagnetics is one of the most fundamental subjects in electrical engineering and it attracts many students to the discipline, some students feel that it is not easy to master electromagnetics. Electromagnetics covers a wide range of topics that not only deal with various physical concepts, but also involve many different mathematical concepts, such as vector functions, coordinate systems, integrals, derivatives, complex numbers, and phasors. Confusion arises not from the amount of mathematical theorems and formulas, but from the lack of a thorough knowledge of the mathematical rules and lack of a rigorous application of the rules to electromagnetic problems. Moreover, the confusion becomes worse with a lack of consistency in the notations that are used to denote various physical quantities and constants in electromagnetics.

The main objective of the book is to present electromagnetic concepts in a more consistent and rigorous manner. This is achieved through elaborate reasoning and the strict application of mathematical concepts. This does not necessarily mean lengthy mathematical steps. On the contrary, I encourage students to obtain the solutions to electromagnetic problems in an intuitive way by considering the symmetry of configurations and applying the uniqueness theorem. The book contains detailed accounts of the following:

1. Students run into difficulties with the concept of vector fields at the beginning of the class, since they have been familiar only with vectors representing, for example, the force acting on a rigid body. Such vectors are closely related to the displacement of the body. However, a vector in a vector field does not necessarily imply a displacement of an object in space; it is a quantity specific to a point in space, and in most cases, is not allowed to move to another point in space.

2. Cylindrical and spherical coordinate systems are meaningful only if the geometry under consideration has cylindrical or spherical symmetry. When a position vector is expressed as $R\mathbf{a}_R$ in spherical coordinates, the unit vector in the radial direction $\mathbf{a}_R$ is treated in different ways: as a constant in the presence of spherical symmetry, or otherwise as a function of position. Base vectors in those coordinates are generally functions of position, and are therefore differentiable and integrable.

3. Symmetry is an integral part of Gauss's law and Ampere's law. The final form of electric flux density or magnetic field intensity of a given problem should be predicted from symmetry configurations so that a Gaussian surface or an Amperian path may be constructed. Typical symmetries in electromagnetics are discussed in detail in the text, including cylindrical, spherical, translational, and two-fold rotational symmetries. Symmetry considerations are useful for intuitively solving boundary value problems and problems of solenoidal and toroidal coils.

4. The inconsistency in notation among different books is a less attractive aspect of electromagnetics. For some authors, the meaning of the notation V_{12} is the potential difference between point 1 and point 2 (or $V_1 - V_2$), yet for others, it signifies the work done in moving a unit charge from point 1 to point 2(or $V_2 - V_1$). This is very confusing for expressing the electric force acting on a charge q_2 due to a charge q_1 as $\mathbf{F}_{21}$ along the direction of a unit vector $\mathbf{a}_{12}$ pointing from q_1 to q_2 such that $\mathbf{F}_{21} = |\mathbf{F}_{21}|\mathbf{a}_{12}$. This book adopts a new notation to avoid the confusion. In our notation, the potential difference is denoted as $V_{1-2} = V_1 - V_2$, in which the subscript 1-2 mimics the subtraction on the right-hand side, while the hyphen implies a sense of backward direction, such as "from 2 to 1," or the effect at point 1 due to a cause at point 2. Accordingly, the electric force on q_1 due to q_2 that is in the direction of a unit vector pointing from q_2 to q_1 is expressed as $\mathbf{F}_{1-2} = |\mathbf{F}_{1-2}|\mathbf{a}_{1-2}$ in our notation. Subscript 12 always represents something "from 1 to 2." For example, Ψ_{12} represents the magnetic flux through loop 2 due to the current in loop 1.

5. An electromagnetic quantity may take on different forms. Static field quantities are denoted by a boldface letter, such as **E** for a static electric field, while time-varying fields are denoted by a script letter, such as $\mathscr{E}$ for a time-varying electric field. Scalars are denoted by a regular letter, such as *E* for the magnitude of electric field intensity. Complex quantities are denoted by a caret on top, such as $\hat{E}_o$ for the complex amplitude of electric field intensity. Since an electric field phasor is independent of time, it is also denoted as **E**.

The book contains 300 figures in which real data are drawn to scale; many figures provide three-dimensional views. Each subsection includes a number of examples that are elaborately worked out by putting aforementioned concepts and relations into use and illustrating rigorous approaches in steps. Each subsection ends with exercises and answers that can be solved in a few simple steps and used to check if students correctly understood the concepts. At the end of each section, several review questions are provided to point out key concepts and relations discussed in the section. Since it has been found that open-ended questions are simply ignored by many students, the review questions are given with hints referring to related equations and figures. The book contains a total of 280 end-of-chapter problems.

I would like to thank the professors and students who provided valuable comments and suggestions, and corrected errors in the examples and exercises. I also wish to thank my wife, Hyunjoo, for her patience, inspiration, and confidence in me in the course of writing this book.

Contents

Chapter 1
Vector Algebra and Coordinate Systems

Electromagnetics entails the study of electric and magnetic phenomena in free space and material media. Electromagnetics comprises three branches: electrostatics concerning static electric fields, magnetostatics concerning static magnetic fields, and electrodynamics concerning time-varying electric and magnetic fields. Electromagnetic theories are based on electromagnetic models that consist of sources such as electric charges and currents, basic quantities such as electric and magnetic field intensities, rules of operations such as vector algebra and coordinate systems, and fundamental laws such as Coulomb's law and Maxwell's equations. The first two chapters of the text discuss the rules of operation, including vector algebra, coordinates systems, and vector calculus.

The basic quantities of electromagnetics are either scalars or vectors. A scalar is completely specified by a magnitude, as with electric potentials, whereas a vector is specified by a magnitude and a direction, as with electric forces. In sinusoidal steady-state conditions in which all electromagnetic quantities vary sinusoidally with time, a complex number can be conveniently used to represent the sinusoidal time-dependence, regardless of whether or not the quantity is a vector. In general, electromagnetic quantities are functions of space and time, whereas the quantities of electrostatics or magnetostatics vary with space coordinates only. Therefore, we usually deal with scalar and vector fields in electromagnetics.

Mathematical rules and techniques are essential for not only expressing electromagnetic concepts in compact forms, but also for constructing electromagnetic models of problems for solutions. Vector algebra, vector calculus, and coordinate systems are three basic mathematical tools for electromagnetics. Vector algebra concerns vector operations such as vector addition, scaling, and scalar and vector products. Vector calculus deals with derivatives and integrals of a scalar or a vector field, some of which are formulated into vector operators called the gradient, divergence, and curl. A coordinate system allows us to express geometric elements such as points, lines, surfaces, and volumes in terms of mathematical equations. Since physical quantities and laws are independent of the coordinate system, we choose a coordinate system that best fits the geometry under consideration and thus facilitates the electromagnetic model of a problem.

Although the definition of a vector is straightforward, we should be careful when we deal with a spatial distribution of vectors, called a vector field, because different types of vectors involved in a vector field may easily draw us into confusion. The position vector always starts at the origin and ends at a point in space, defining the point located at the terminal point of the vector. A distance vector starts at a point and ends at another, defining the distance between the two points and the direction from the initial to the terminal point. The distance and the direction are the quantities belonging to the terminal point of the vector. A vector in a vector field is specific to a point in space. Thus it would be meaningless if the vector is displaced to another point, because it represents the physical quantity observed at the original point. Its magnitude and direction are the quantities belonging to the initial point of the vector. In this case, the space coordinates of the terminal point have no significance. Base vectors are a set of three mutually orthogonal unit vectors whose directions generally vary with position.

1.1 Vectors and Vector Field

A vector is a quantity having a magnitude and a direction. An arrow can be conveniently used to represent the vector. The length of the arrow represents the magnitude of the vector, while its arrowhead points in the direction of the vector. The tail of the arrow is called the initial point of the vector, whereas the tip of the arrow is called the terminal point of the vector. It is customary to denote a vector by a letter in boldface type such as **A** and **B**, or a letter with an arrow on top such as $\vec{A}$ and $\vec{B}$. A vector **A** takes on different forms such as

$$\mathbf{A} = A\,\mathbf{a}_A = |\mathbf{A}|\,\mathbf{a}_A = A\frac{\mathbf{A}}{|\mathbf{A}|} \tag{1-1}$$

Here, both A and $|\mathbf{A}|$ represent the magnitude, while both $\mathbf{a}_A$ and $\mathbf{A}/|\mathbf{A}|$ represent the unit vector of **A**. The magnitude is a positive real number, and carries the unit of the vector. The unit vector has a unity magnitude, $|\mathbf{a}_A| = 1$, but no dimension; it only shows the direction.

If a vector is defined at each and every point in a given region of space, it is said that a vector field exists in the region. A vector function is a mathematical expression for a vector field, which is generally a function of space and time. Consider Fig. 1.1 depicting a vector field defined in the three-dimensional space. It shows that the vector quantity observed at point p_1 is represented by vector **A** with the initial point coincident with point p_1, while that observed at point p_2 is represented by vector $\mathbf{A}'$ starting at point p_2.

In Cartesian coordinate system, point p_1 shown in Fig. 1.1 is defined by three coordinates, x_1, y_1, and z_1. Thus the point is expressed as $p_1{:}(x_1, y_1, z_1)$ in our notation, in which colon is to distinguish the point from a scalar function usually expressed as $p_1(x, y, z)$. The point p_1 can also be defined by position vector $\mathbf{r}_1$, which is a vector drawn from the origin to point p_1. It should be noted that a position vector always starts at the origin; its magnitude represents the distance between the origin and the point, and its unit vector represents the direction from the origin to the point.

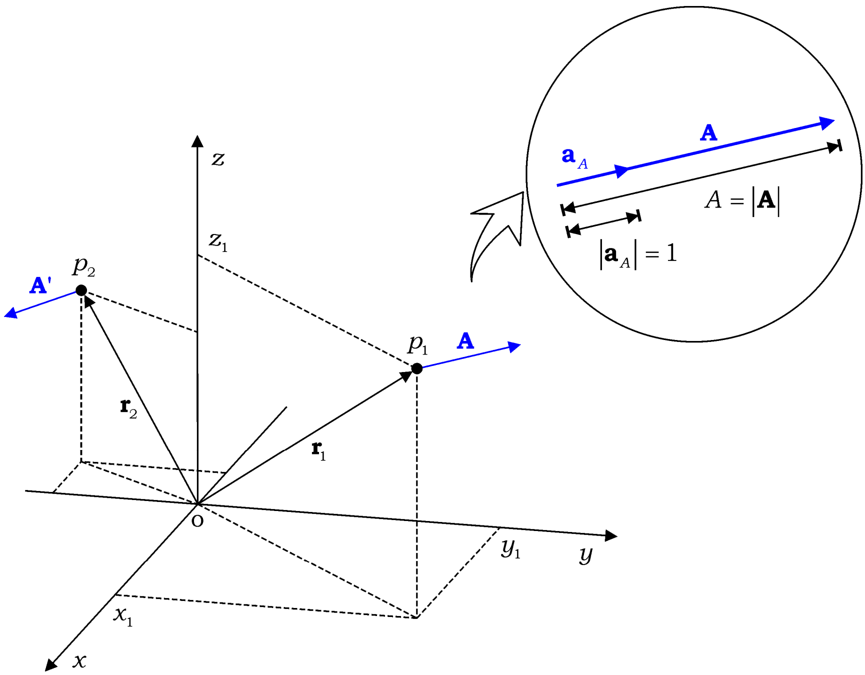


Fig. 1.1 A vector field in three-dimensional space.

As an example, let us suppose the vector field shown in Fig. 1.1 is the wind velocity in the air measured at a time $t = t_o$. In view of vector **A**, we see that wind blows at a speed A[m/s] at point p_1 in the direction of $\mathbf{a}_A$. It is important to note that the wind velocity **A** is specific to a point defined by the position vector $\mathbf{r}_1$ in the air; the point happens to be the initial point of **A**. By the same token, vector $\mathbf{A}'$ signifies that wind blows at a speed A'[m/s] in the direction of $\mathbf{a}_{A'}$ at point p_2 with position vector $\mathbf{r}_2$. If the wind velocity is described by a vector function $\mathbf{F}(\mathbf{r}, t)$, where $\mathbf{r}$ is position vector and t is time, the vector **A** is mathematically expressed as

$$\mathbf{F}(\mathbf{r}_1, t_o) = \mathbf{F}(x_1, y_1, z_1; t_o) = \mathbf{A} \tag{1-2}$$

Here, $\mathbf{r}_1$ is the position vector of point p_1 with Cartesian coordinates x_1, y_1, and z_1. Similarly, expressed mathematically,

$$\mathbf{F}(\mathbf{r}_2, t_o) = \mathbf{A}' \tag{1-3}$$

The constant t_o in the above equations means that the wind velocities are measured simultaneously at a time $t = t_o$.

If vector $\mathbf{A}'$ at point p_2 has the same magnitude and direction as vector **A** at another point p_1, the two vectors are equal, but not identical because they are at different points in space.

Exercise 1.1
Determine whether the following expressions represent a vector field or not:
(a) $\mathbf{A}(x)$, (b) $B(\mathbf{r})$, (c) $\mathbf{C}(\mathbf{r}_1)$, and (d) $\mathbf{D}(\mathbf{r}-\mathbf{r}_1)$.
Ans. (a) Yes, (b) No, (c) No, (d) Yes.

1.2 Vector Algebra

1.2.1 Vector Addition and Subtraction

The sum of two vectors **A** and **B** results in a third vector **C**. The vector addition of **A** and **B** is expressed as

$$\mathbf{A}+\mathbf{B}=\mathbf{C} \tag{1-4}$$

The vector addition is in general defined for two vectors that share a common initial point; they are said to be at the same point in space. Accordingly, the resultant vector **C** is located at the same point as **A** and **B**.

The parallelogram rule and head-to-tail rule are graphical methods of performing the vector addition. According to the parallelogram rule as shown in Fig. 1.2(a), the resultant vector **C** corresponds to the diagonal of the parallelogram formed by **A** and **B**. On the other hand, according to the head-to-tail rule as shown in Figs. 1.2(b)(c), vector **B** (or **A**) is first displaced such that its initial point touches the terminal point of **A** (or **B**). Then the resultant vector **C** corresponds to an arrow drawn from the initial point of **A** (or **B**) to the terminal point of **B** (or **A**). It should be noted that a vector is displaced in Figs. 1.2(b)(c) only for the purpose of conducting the vector addition graphically.

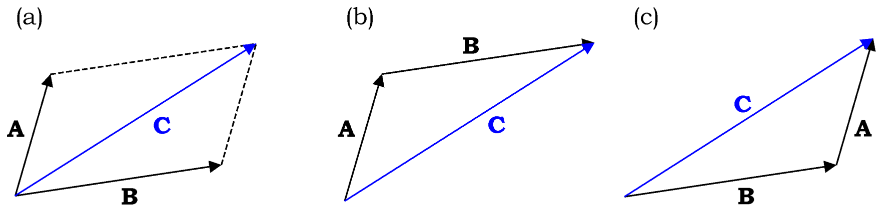


Fig. 1.2 Vector addition, $\mathbf{A}+\mathbf{B}$: (a) parallelogram rule, (b) and (c) head-to-tail rule.

Vector addition obeys the associative law and the commutative law:

$$\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C} \qquad \text{(associative)} \qquad (1\text{-}5a)$$

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A} \qquad \text{(commutative)} \qquad (1\text{-}5b)$$

Although a vector may be displaced from a point in space to another and added to the vector at that point, such a vector addition has no physical significance in most cases. As evident in Fig. 1.1, the sum of two wind velocities, $\mathbf{A}+\mathbf{A}'$, would have no consequence. ***A vector in a vector field is specific to a point in space***; it may not be moved to a new location in space.

Subtraction of vector **B** from vector **A** is equal to the vector addition of **A** and the negative of **B**, that is,

$$\mathbf{A}-\mathbf{B}=\mathbf{A}+(-\mathbf{B})=\mathbf{A}+B(-\mathbf{a}_B) \qquad (1\text{-}6)$$

The negative of **B**, or $-\mathbf{B}$, has the same magnitude as **B** but points in the direction opposite to that of **B**. Note that vector $-\mathbf{B}$ is at the same point as vector **B**. The vector subtraction can be done graphically following the same rule as was used for the vector addition(see Fig. 1.3).

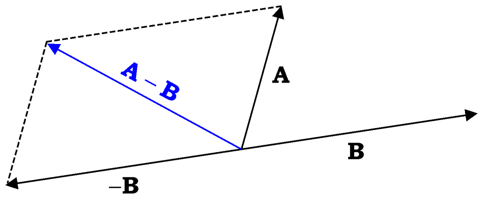


Fig. 1.3 Vector subtraction, $\mathbf{A}-\mathbf{B}$.

Vector subtraction of two position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ is of course expressed as $\mathcal{R}=\mathbf{r}_1-\mathbf{r}_2$, and called the distance vector. As we see in Fig. 1.4, the distance vector $\mathcal{R}$ is directed from point p_2 to point p_1. We assume the magnitude and the unit vector of $\mathcal{R}$ belong to point p_1, which happens to be the terminal point of $\mathcal{R}$. In other words, distance vector $\mathcal{R}$ is defined solely for the physical quantity observed at point p_1. Here, let us introduce a new notation, which will be useful for symbols with double indices. The distance vector is expressed as $\mathcal{R}_{1-2}=\mathbf{r}_1-\mathbf{r}_2$ in our notation. The subscript 1-2 mimics the subtraction on the right-hand side of the equality, while the hyphen in 1-2 signals that the subscript is read backwards such as “from 2 to 1”. The conventional subscript 12, meanwhile, signifies something “from 1 to 2” in our notation.

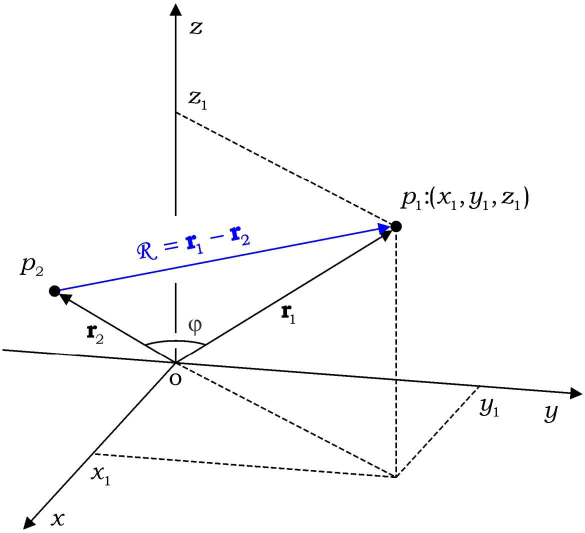


Fig. 1.4 Distance vector $\mathcal{R}$ is written as $\mathcal{R}_{1-2}$ in our notation.

Exercise 1.2
Are the following expressions true?
(a) $-(-\mathbf{A}) = \mathbf{A}$, (b) $-(\mathbf{A}+\mathbf{B}) = -\mathbf{A}-\mathbf{B}$, and (c) $\mathbf{A}-(\mathbf{B}-\mathbf{C}) = (\mathbf{A}-\mathbf{B})-\mathbf{C}$.
Ans. (a) Yes, (b) Yes, (c) No.

1.2.2 Vector Scaling

The multiplication of a vector by a scalar is called the vector scaling. It is equivalent to lengthening or shortening the vector without changing its direction. The scaling of a vector $\mathbf{A}$ by a scalar k is written as

$$k\mathbf{A} = kA\,\mathbf{a}_A \tag{1-7}$$

A negative scalar k reverses the direction of the vector in addition to modifying the magnitude.

Vector scaling obeys the associative, commutative, and distributive laws:

$$k(l\mathbf{A}) = l(k\mathbf{A}) \quad \text{(associative law)} \tag{1-8a}$$
$$k\mathbf{A} = \mathbf{A}k \quad \text{(commutative law)} \tag{1-8b}$$
$$(k+l)\mathbf{A} = k\mathbf{A} + l\mathbf{A} \quad \text{(distributive law)} \tag{1-8c}$$

Exercise 1.3
Do the following expressions make sense?
(a) $k(\mathbf{A}+\mathbf{B}) = k\mathbf{A}+k\mathbf{B}$, (b) $-k\mathbf{A} = k(-\mathbf{A})$, and (c) $k(\mathbf{A}+c) = k\mathbf{A}+kc$.
Ans. (a) Yes, (b) Yes, (c) No.

1.2.3 Scalar or Dot Product

Scalar product and vector product are two unique operations defined in vector algebra. They are very different from the simple multiplication of a vector by a scalar. As implied by the name, the scalar product results in a scalar, whereas the vector product results in a vector. The scalar product involves cosine of an angle between two vectors, whereas the vector product involves sine of the angle. They allow us to write an equation containing a cosine or a sine in vector notation. In general, the vector notation is extremely compact and able to conjure up the geometry of configuration.

The scalar product or dot product of **A** and **B** is denoted as $\mathbf{A}\bullet\mathbf{B}$ ("A dot B") and defined as

$$\boxed{\mathbf{A}\bullet\mathbf{B} = AB\cos\theta_{AB}} \tag{1-9}$$

where A and B are the magnitudes of **A** and **B**, respectively, and θ_{AB} is the smaller angle between the two vectors. The dot product results in a scalar: a positive real number for $0 \le \theta_{AB} < 90^o$, a negative real number for $90^o < \theta_{AB} \le 180^o$, and zero for $\theta_{AB} = 90^o$, implying that **A** and **B** are mutually orthogonal(see Fig. 1.6). If two vectors are at different points in space, one vector should be displaced to the point of the other, and expressed following the same way as for the other vector before the dot product of the two may be performed(see Section 1-3).

The term $B\cos\theta_{AB}$ on the right-hand side of Eq. (1-9) is called the projection of **B** in the direction of **A**, or the projection of **B** onto **A**(see Fig. 1.5). It is also called the scalar component of **B** in the direction of **A**. Accordingly, the dot product $\mathbf{A}\bullet\mathbf{B}$ can be viewed as the product of A and the projection of **B** onto **A**, or the product of B and the projection of **A** onto **B**.

The projection of a vector onto another is useful for decomposing the vector into vector components: the magnitude of a vector component is equal to the scalar component, and its unit vector is the same as that used for obtaining the scalar component. For instance, the vector component of **C** in the direction of **A** is expressed as $(\mathbf{C}\bullet\mathbf{a}_A)\mathbf{a}_A$, where $(\mathbf{C}\bullet\mathbf{a}_A)$ is the scalar component, and $\mathbf{a}_A$ is the unit vector in the direction of **A**. Decomposition of a vector is the reverse of the vector addition, and vice versa. Accordingly, a vector addition $\mathbf{A}+\mathbf{B}=\mathbf{C}$ may be viewed as the decomposition of **C** into two component vectors **A** and **B**. A vector can also be projected onto a plane. Such a projection gives us the vector component tangential to the plane.

From the definition given in Eq. (1-9) we see that the dot product of **A** and **B** cannot be larger than the product of the magnitudes of the two vectors, that is,

$$\mathbf{A}\bullet\mathbf{B} \le AB \tag{1-10}$$

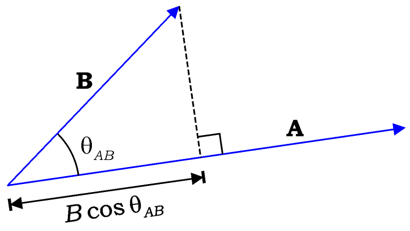


Fig. 1.5 Projection of **B** onto **A** .

The dot product of a vector with itself yields the magnitude squared, that is,

$$\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2 = A^2 \tag{1-11}$$

This can be used to find the magnitude of **A** as follows:

$$A = |\mathbf{A}| = \sqrt{\mathbf{A}\cdot\mathbf{A}} \tag{1-12}$$

The magnitude of **A** is equal to the positive square root of the product of **A** with itself.

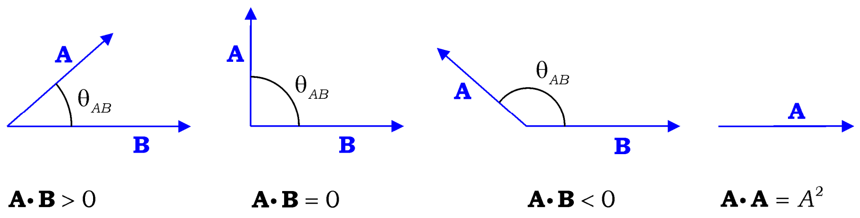


Fig. 1.6 Dot product of **A** and **B**

Dot product obeys the commutative and distributive law's such that

$$\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A} \qquad \text{(commutative)} \tag{1-13a}$$

$$\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C} \qquad \text{(distributive)} \tag{1-13b}$$

The commutative law of dot product directly follows from the obvious relation, $\theta_{AB} = \theta_{BA}$. The proof of the distributive law is also straightforward as shown in **Example 1-2**.

We can obtain the projection of a vector onto another by the method called the two-step projection, even if we have no knowledge of the angle between two vectors. With reference to Fig. 1.7, we recognize the following:

(1) Projection of **A** onto plane S containing **B** gives the tangential component $\mathbf{A}_T$.

(2) Projection of $\mathbf{A}_T$ onto **B** give the projection of **A** onto **B**.

It is apparent from Fig. 1.7 that the vector $\mathbf{A}$ can be decomposed into the normal component $\mathbf{A}_N$ and the tangential component $\mathbf{A}_T$. Expressed mathematically, the projection of $\mathbf{A}$ onto $\mathbf{B}$ is

$$\mathbf{A} \bullet \mathbf{a}_B = \left(\mathbf{A}_T + \mathbf{A}_N\right) \bullet \mathbf{a}_B = \mathbf{A}_T \bullet \mathbf{a}_B \tag{1-14}$$

where we used the relation $\mathbf{A}_N \bullet \mathbf{a}_B = 0$. The two-step projection is thus verified.

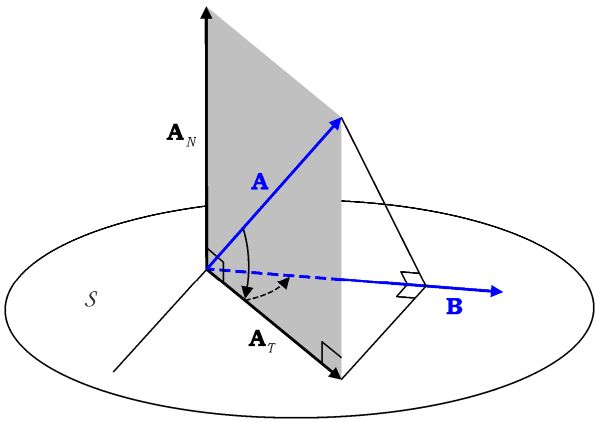


Fig. 1.7 Two-step projection.

Example 1-1
With reference to Fig. 1.4, find an expression for the magnitude of $\mathcal{R}$ in terms of $|\mathbf{r}_1|$, $|\mathbf{r}_2|$, and the angle φ between $\mathbf{r}_1$ and $\mathbf{r}_2$.

Solution
Dot product of the distance vector $\mathcal{R}$ with itself gives

$$\begin{aligned} \mathcal{R} \bullet \mathcal{R} = \mathcal{R}^2 &= \left(\mathbf{r}_1 - \mathbf{r}_2\right) \bullet \left(\mathbf{r}_1 - \mathbf{r}_2\right) \\ &= r_1^2 + r_2^2 - \mathbf{r}_1 \bullet \mathbf{r}_2 - \mathbf{r}_2 \bullet \mathbf{r}_1 \end{aligned} \tag{1-15}$$

Dot product of the two position vectors is

$$\mathbf{r}_1 \bullet \mathbf{r}_2 = \mathbf{r}_2 \bullet \mathbf{r}_1 = r_1 r_2 \cos\varphi \tag{1-16}$$

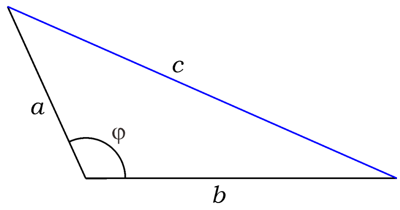
Substituting Eq. (1-16) into Eq. (1-15) we obtain

$$\mathcal{R}^2 = r_1^2 + r_2^2 - 2 r_1 r_2 \cos\varphi \tag{1-17}$$

The magnitude of the distance vector is therefore

$$\mathcal{R} = \left[r_1^2 + r_2^2 - 2r_1r_2\cos\varphi\right]^{1/2}$$

Equation (1-17) is known as the law of cosines. For future reference, the law of cosines is rewritten as shown in Fig. 1.8.



$$c^2 = a^2 + b^2 - 2ab\cos\varphi$$

Fig. 1.8 Law of cosines.

Example 1-2
Verify the distributive law of dot product as given in Eq. (1-13b) by assuming that the three vectors are contained in the same plane.

Solution
In Fig. 1.9, straight line segments b, c, and d are projections of $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{B}+\mathbf{C}$ onto the same vector $\mathbf{A}$. From Fig. 1.9 we immediately recognize

$$d = b + c \tag{1-18}$$

Multiplying both sides of Eq. (1-18) by A we have

$$Ad = Ab + Ac \tag{1-19}$$

The vector notation of Eq. (1-19) is

$$\mathbf{A}\bullet(\mathbf{B}+\mathbf{C}) = \mathbf{A}\bullet\mathbf{B} + \mathbf{A}\bullet\mathbf{C} \tag{1-20}$$

The distributive law of dot product is thus verified for three vectors lying in the same plane.

Exercise 1.4
Expand the dot product $(\mathbf{A}+\mathbf{B})\bullet(\mathbf{C}+\mathbf{D})$ and name the law that you apply.

Ans. $\mathbf{A}\bullet\mathbf{C} + \mathbf{A}\bullet\mathbf{D} + \mathbf{B}\bullet\mathbf{C} + \mathbf{B}\bullet\mathbf{D}$, distributive law of dot product.

Exercise 1.5
Do the following expressions make sense?
(a) $(-\mathbf{A})\bullet(-\mathbf{B}) = \mathbf{A}\bullet\mathbf{B}$, (b) $(-\mathbf{A})\bullet\mathbf{B} = -(\mathbf{A}\bullet\mathbf{B})$, and (c) $(\mathbf{A}\bullet\mathbf{B})\bullet\mathbf{C} = \mathbf{A}\bullet(\mathbf{B}\bullet\mathbf{C})$.

Ans. (a) Yes, (b) Yes, (c) No.

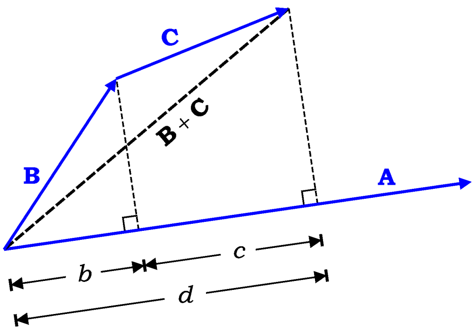


Fig. 1.9 Distributive law of dot product.

Exercise 1.6
What is the meaning of $\mathbf{A} \bullet \mathbf{B} < 0$ in terms of the projection of **A** onto **B**.
Ans. The projection falls onto $-\mathbf{B}$.

1.2.4 Vector or Cross Product

The vector product or cross product of **A** and **B** is denoted as $\mathbf{A} \times \mathbf{B}$ ("A cross B") and defined as

$$\boxed{\mathbf{A} \times \mathbf{B} = AB \sin \theta_{AB}\, \mathbf{a}_N} \tag{1-21}$$

where θ_{AB} is the smaller angle between **A** and **B**, and $\mathbf{a}_N$ is a unit vector that is normal to the plane formed by **A** and **B**. The direction of $\mathbf{a}_N$ is governed by the right-hand rule: the right thumb points in the direction of $\mathbf{a}_N$ when the fingers rotate from **A** to **B** through the angle θ_{AB} (see Fig. 1.10). The cross product is defined for two vectors that are at the same point in space. Otherwise, one vector should be displaced to the point of the other, and expressed following the same way as for the other vector before the cross product of the two may be performed(see Section 1-3).

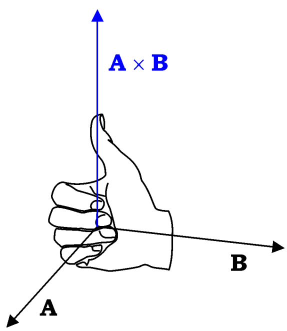


Fig. 1.10 Right-hand rule.

Figure 1.11 shows that there is no change in the direction of $\mathbf{A}\times\mathbf{B}$ even though the angle θ_{AB} increases from zero to 180^o.

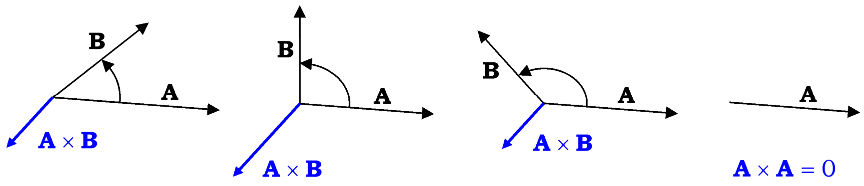


Fig. 1.11 Cross product of $\mathbf{A}$ and $\mathbf{B}$.

The magnitude of the cross product, $|\mathbf{A}\times\mathbf{B}|$, numerically represents the area of the parallelogram formed by $\mathbf{A}$ and $\mathbf{B}$: A is the base, and $B\sin\theta_{AB}$ is the height of the parallelogram as shown in Fig. 1.12. The cross product of two vectors is also useful for determining the perpendicular distance from a point in space to a straight line. It is apparent from Fig. 1.12 that the perpendicular distance from the tip of $\mathbf{B}$ to vector $\mathbf{A}$ is equal to $|\mathbf{B}\times\mathbf{a}_A|$.

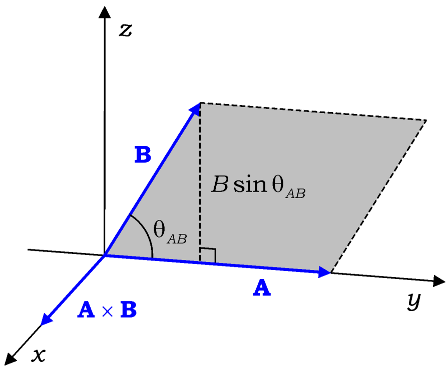


Fig. 1.12 $|\mathbf{A}\times\mathbf{B}|$ is the area of the parallelogram formed by $\mathbf{A}$ and $\mathbf{B}$.

Cross product obeys the distributive and anticommutative laws. However, it does not satisfy the associative law:

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) \neq (\mathbf{A}\times\mathbf{B})\times\mathbf{C} \quad \text{(not associative)} \tag{1-22a}$$

$$\mathbf{A}\times\mathbf{B} = -(\mathbf{B}\times\mathbf{A}) \quad \text{(anticommutative)} \tag{1-22b}$$

$$\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C} \quad \text{(distributive)} \tag{1-22c}$$

The anticommutative law directly follows from the right-hand rule: the right thumb points in the opposite direction when the fingers rotate in the reverse direction.

Example 1-3

Verify the distributive law of cross product, $\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}$, by assuming that **A** is perpendicular to the plane formed by **B** and **C**.

Solution

Figure 1.13 shows that **B** and **C** form a parallelogram *P*, while $\mathbf{A}\times\mathbf{B}$ and $\mathbf{A}\times\mathbf{C}$ form another parallelogram P'.

Angle between **B** and **C** is

$$\theta_1 = \cos^{-1}\frac{\mathbf{B}\bullet\mathbf{C}}{BC} \tag{1-23}$$

The angle between $\mathbf{A}\times\mathbf{B}$ and $\mathbf{A}\times\mathbf{C}$ is equal to θ_1. To prove this, noting that **A** is assumed to be perpendicular to both **B** and **C**, we write

$$\begin{aligned}\theta_2 &= \cos^{-1}\frac{(\mathbf{A}\times\mathbf{B})\bullet(\mathbf{A}\times\mathbf{C})}{(AB)(AC)} = \cos^{-1}\frac{[(\mathbf{A}\times\mathbf{B})\times\mathbf{A}]\bullet\mathbf{C}}{(AB)(AC)} \\ &= \cos^{-1}\frac{A^2\mathbf{B}\bullet\mathbf{C}}{(AB)(AC)} = \theta_1\end{aligned} \tag{1-24a}$$

where we used the vector identity $\mathbf{U}\bullet(\mathbf{V}\times\mathbf{W})=(\mathbf{U}\times\mathbf{V})\bullet\mathbf{W}$ (see Eq. (1-25)), and the following relations:

$$|\mathbf{A}\times\mathbf{B}| = AB \tag{1-24b}$$

$$|\mathbf{A}\times\mathbf{C}| = AC \tag{1-24c}$$

In view of Eq. (1-24), parallelogram P' is similar to parallelogram *P*: P' is the magnified, by a factor of *A*, and rotated, by an angle of 90^o, version of *P*. Following the same way as for the two sides of *P*, the diagonal of *P* is magnified and rotated, and comes to be the diagonal of P', that is,

$$\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = (\mathbf{A}\times\mathbf{B})+(\mathbf{A}\times\mathbf{C}).$$

The distributive law for the three vectors is therefore verified.

Exercise 1.7

Do the following expressions make sense? (a) $-(\mathbf{A}\times\mathbf{B})=(-\mathbf{A})\times(-\mathbf{B})$, (b) $-(\mathbf{A}\times\mathbf{B})=(-\mathbf{A})\times\mathbf{B}$, and (c) $\mathbf{C}\times(\mathbf{A}\times\mathbf{B})=(\mathbf{C}\times\mathbf{A})\times(\mathbf{C}\times\mathbf{B})$.

Ans. (a) No, (b) Yes, (c) No.

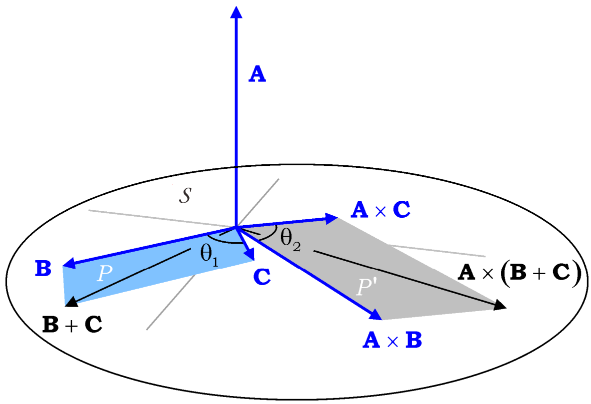


Fig. 1.13 Distributive law of cross product.

Exercise 1.8
Given **A** and **B** at a point in space, find the vector components of **A** in the direction (a) parallel to **B**, and (b) perpendicular to **B**.
Ans. (a) $(\mathbf{A}\cdot\mathbf{B})\mathbf{B}/B^2$, (b) $(\mathbf{B}\times\mathbf{A})\times\mathbf{B}/B^2$.

1.2.5 Scalar and Vector Triple Products

We can extend the dot or cross product of two vectors to multiple products by applying dot or cross product repeatedly, if allowed to do so. There are two types of triple products frequently encountered in electromagnetics: scalar triple product and vector triple product. They are so named because of the type of the results.

(1) Scalar triple product

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) \tag{1-25}$$

A scalar triple product results in a scalar. Eq. (1-25) follows cyclic permutations of the three vectors: ABC-BCA-CAB. We note that an interchange of the dot and cross symbols gives the same result; that is, $\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} = \mathbf{A}\times\mathbf{B}\cdot\mathbf{C}$. Note that the cross product always precedes the dot product in a scalar triple product; otherwise it would not make sense.

The scalar triple product $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ numerically represents the volume of the parallelepiped defined by three vectors **A**, **B**, and **C** as shown in Fig. 1.14. The magnitude of the cross product, $|\mathbf{B}\times\mathbf{C}|$, represents the area of the base, and the projection of **A** onto $\mathbf{B}\times\mathbf{C}$ represents the height of the parallelepiped.

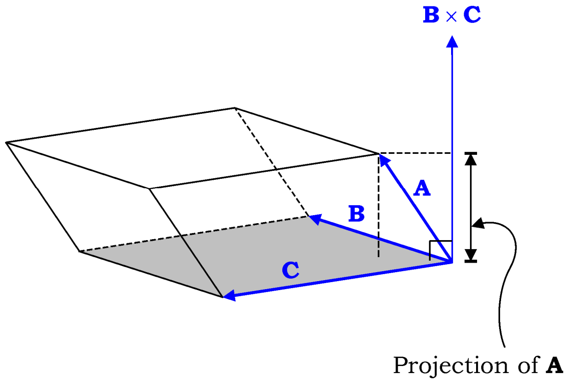


Fig. 1.14 Scalar triple product $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$.

(2) Vector triple product

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \tag{1-26}$$

It is also known as "BAC-CAB" rule. The vector triple product results in a vector. Eq. (1-26) can be verified by expanding the three vectors in component form and directly evaluating the dot and cross products(see Section 1-3).

The vector triple product does not obey the associative law, namely,

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) \neq (\mathbf{A} \times \mathbf{B}) \times \mathbf{C} \tag{1-27}$$

A parenthesis is required in the vector triple product to signify which cross product is performed first.

Exercise 1.9

Under what conditions does $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$ become zero?

Ans. $\mathbf{B} \parallel \mathbf{C}$, or all vectors lie in the same plane.

Exercise 1.10

Under what conditions is $(\mathbf{A} \times \mathbf{B}) \times \mathbf{A}$ equal to **B**?

Ans. $\mathbf{A} \perp \mathbf{B}$ and $|\mathbf{A}| = 1$.

Exercise 1.11

Are the following expressions true?

(a) $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{A} \cdot (-\mathbf{B} \times -\mathbf{C})$, and (b) $-\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{A} \cdot (-\mathbf{B} \times \mathbf{C})$.

Ans. (a) Yes, (b) Yes only if $\mathbf{B} \perp \mathbf{C}$.

Review Questions with Hints

RQ 1.1 Locate $\mathbf{A}+\mathbf{B}$ in space if $\mathbf{A}$ and $\mathbf{B}$ are defined in two different vector fields such as $\mathbf{A}=\mathbf{F}(\mathbf{r}_1)$ and $\mathbf{B}=\mathbf{G}(\mathbf{r}_1)$. [Fig.1.1]

RQ 1.2 Distinguish between position and distance vectors. [Fig.1.4]

RQ 1.3 May a scalar $ab\cos 30^{\circ}$ be rewritten in vector notation? [Eq.(1-9)]

RQ 1.4 Does $\mathbf{A}\bullet\mathbf{B}=\mathbf{A}\bullet\mathbf{C}$ lead directly to $\mathbf{B}=\mathbf{C}$? [Eq.(1-9)]

RQ 1.5 Does $\mathbf{A}\times\mathbf{B}=\mathbf{A}\times\mathbf{C}$ lead directly to $\mathbf{B}=\mathbf{C}$? [Eq.(1-21)]

RQ 1.6 Find the scalar component of $\mathbf{B}$ in the direction of $\mathbf{A}$. [Fig.1.5]

RQ 1.7 State the right-hand rule implicit in the cross product. [Fig.1.10]

RQ 1.8 What is the rule underlying the distributive law of cross product? [Eq.(1-21)]

1.3 Orthogonal Coordinate Systems

A coordinate system makes it possible to describe geometric elements such as point, line, surface, and volume in terms of three numbers called coordinates. Therefore, mathematical formulations of those geometric elements are possible in a coordinate system. A coordinate may be the distance from the origin measured along an axis or the angle measured with respect to an axis. An orthogonal coordinate system is one in which the three surfaces of constant coordinate are mutually perpendicular at every point in the system. A right-handed coordinate system enumerates the coordinates of a point in an order compatible with the right-hand rule: when the right fingers rotate from the direction of increase of the first coordinate to that of the second coordinate, the thumb points in the direction of increase of the third coordinate. Cartesian(or rectangular), cylindrical, and spherical coordinate systems are the three most commonly used orthogonal coordinate systems in electromagnetics.

Whereas physical quantities and laws are independent of the coordinate system, any coordinate system may be chosen for an electromagnetic problem. Nevertheless, a particular coordinate system may be advantageous over the others, if it can describe the geometry under consideration in a much simpler way than the others. Cylindrical coordinate system is useful specifically for problems having cylindrical symmetry, whereas spherical coordinate system is useful specifically for problems having spherical symmetry. Electromagnetic problems usually involve a source and an observer. Because the locations of the source and observer are independent of each other, it will is convenient to adopt two independent coordinate systems for a given problem; the spatial distribution of the source is described in one system, while the position of the observer is described in the other system. Such system is called a mixed coordinate system. In view of these considerations, it is important that we have a means of transforming the coordinates of a point or the components of a vector from one system to another, which we call the coordinate transformation.

1.3.1 Cartesian Coordinate System

Cartesian coordinate system specifies a point $p_1{:}(x_1, y_1, z_1)$ by the intersection of the three planes $x = x_1$, $y = y_1$, and $z = z_1$ planes. At point p_1, three unit vectors $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$, which are called the base vectors, are defined in such a way that they are perpendicular to the $x = x_1$, $y = y_1$, and $z = z_1$ planes, respectively, and point in the direction of increasing coordinate. Although the base vectors are constant in Cartesian coordinate system, they are generally functions of position in other coordinate systems.

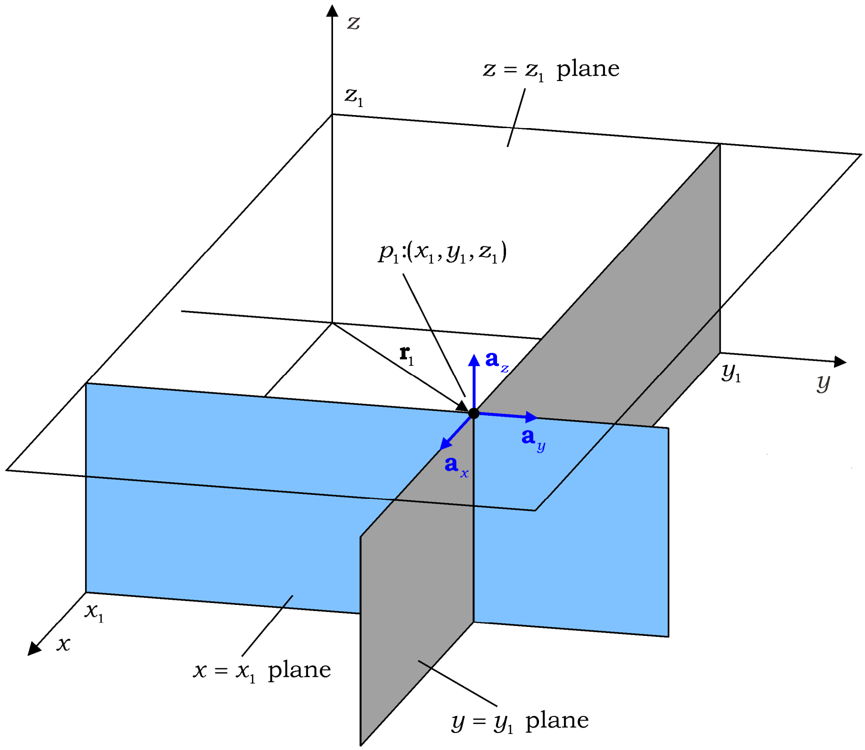


Fig. 1.15 Cartesian coordinate system. Base vectors at a point p_1.

The base vectors for Cartesian coordinates obey the orthonormality relations:

$$\mathbf{a}_x \cdot \mathbf{a}_y = \mathbf{a}_y \cdot \mathbf{a}_z = \mathbf{a}_z \cdot \mathbf{a}_x = 0 \tag{1-28a}$$

$$\mathbf{a}_x \cdot \mathbf{a}_x = \mathbf{a}_y \cdot \mathbf{a}_y = \mathbf{a}_z \cdot \mathbf{a}_z = 1 \tag{1-28b}$$

Equation (1-28a) shows that the base vectors are mutually perpendicular, and Eq. (1-28b) shows that the base vectors are unit vectors.

In the right-handed Cartesian system, the base vectors satisfy the cyclic relations:

$$\mathbf{a}_x \times \mathbf{a}_y = \mathbf{a}_z \tag{1-29a}$$

$$\mathbf{a}_y \times \mathbf{a}_z = \mathbf{a}_x \tag{1-29b}$$

$$\mathbf{a}_z \times \mathbf{a}_x = \mathbf{a}_y \tag{1-29c}$$

The three vectors in the cyclic permutations obey the right-hand rule: when the right fingers rotate from the first to the second vector, the thumb points in the direction of the third vector.

As can be seen in Fig. 1.15, position vector $\mathbf{r}_1$ is a vector drawn from the origin to point $p_1{:}(x_1, y_1, z_1)$. Its magnitude represents the distance between the origin and p_1, and its unit vector shows the direction from the origin toward p_1. For this reason, the magnitude and direction of $\mathbf{r}_1$ are expressed in terms of the coordinates and base vectors at p_1. Following the same procedure as was previously discussed, we obtain the vector components of $\mathbf{r}_1$ by projecting the position vector onto the base vectors at p_1 and appending corresponding unit vectors; that is, $x_1 \mathbf{a}_x$, $y_1 \mathbf{a}_y$, and $z_1 \mathbf{a}_z$. Omitting subscript 1 for generalization, the component form of position vector in Cartesian coordinates is therefore

$$\boxed{\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z} \tag{1-30}$$

Here, the scalar components x, y, and z are the Cartesian coordinates of a given point with position vector $\mathbf{r}$, and the unit vectors $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$ are the base vectors at the point, or the terminal point of $\mathbf{r}$.

Consider Fig. 1.16, in which vector $\mathbf{A}$ is defined at point p_1 with position vector $\mathbf{r}_1$ in Cartesian coordinates. It is important to note that vector $\mathbf{A}$ is specific to point p_1, because it, for instance, represents the wind velocity measured at p_1. For this reason, the magnitude and direction of $\mathbf{A}$ are expressed in terms of the coordinates and base vectors at p_1. Following the same procedure used for $\mathbf{r}_1$, we obtain the vector components of $\mathbf{A}$ by projecting the vector onto the base vectors at p_1 and appending the corresponding unit vectors to the scalar components. The component form of vector $\mathbf{A}$ in Cartesian coordinates is therefore,

$$\boxed{\mathbf{A} = A_x\mathbf{a}_x + A_y\mathbf{a}_y + A_z\mathbf{a}_z} \tag{1-31}$$

where the scalar components A_x, A_y, and A_z generally depend on x_1, y_1, and z_1. Although the terminal point of position vector $\mathbf{r}_1$ is a real point in space,

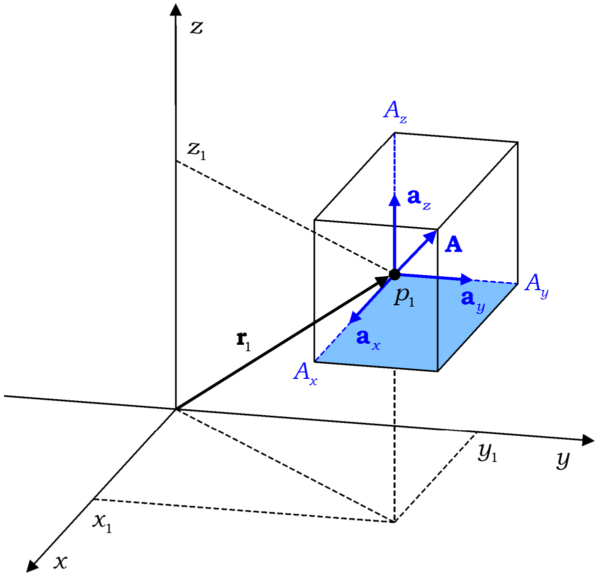


Fig. 1.16 Vector **A** defined at point p_1:(x_1, y_1, z_1) in Cartesian coordinates.

which can be specified by space coordinates, the terminal point of vector **A** is not a spatial point. This can be justified by the fact that vector **A** represents the physical quantity observed only at point p_1, such as wind velocity measured at p_1.

To summarize, we have discussed the different types of vectors: position vector, distance vector, base vectors, and vector field. The position and distance vectors are drawn from a point in space to another, representing the distance between the two points and the direction from the initial toward the terminal point, in spite of the fact that position vectors always start at the origin. The magnitude and direction of these vectors are used for the quantity observed at the terminal point of the vector. Base vectors are three mutually orthogonal unit vectors that are directed along the direction of increase of a coordinate. Base vectors are essential for expanding a vector in component form. A vector **A** in a vector field is specific to an observation point that is specified by a given position vector. The magnitude and direction of **A** represent the vector quantity observed at the point. For this reason, the tip of the vector **A** is not a spatial point and cannot be specified by space coordinates.

So far, we have discussed the component forms of position vector **r** and vector field **A**(**r**). We are now ready to express all the rules of vector algebra in terms of the vector components in Cartesian coordinates. If two vectors **A** and **B** are at the same point p_1 in Cartesian coordinates, they can be expanded in component form as follows:

$$\mathbf{A} = A_x \mathbf{a}_x + A_y \mathbf{a}_y + A_z \mathbf{a}_z \tag{1-32a}$$

$$\mathbf{B} = B_x \mathbf{a}_x + B_y \mathbf{a}_y + B_z \mathbf{a}_z \tag{1-32b}$$

where $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$ are the base vectors at point p_1.

Vector addition of **A** and **B** is simply given by the sum of Eq. (1-32a) and Eq. (1-32b), that is,

$$\boxed{\mathbf{A} + \mathbf{B} = (A_x + B_x)\mathbf{a}_x + (A_y + B_y)\mathbf{a}_y + (A_z + B_z)\mathbf{a}_z} \tag{1-33}$$

To verify the x-component on the right-hand side of Eq. (1-33), let us take the dot product of $(\mathbf{A} + \mathbf{B})$ with a unit vector $\mathbf{a}_x$. Upon applying the distributive law expressed by Eq. (1-13b) and the orthonormality relation expressed by Eq. (1-28), we have

$$(\mathbf{A} + \mathbf{B}) \bullet \mathbf{a}_x = \mathbf{A} \bullet \mathbf{a}_x + \mathbf{B} \bullet \mathbf{a}_x = A_x + B_x$$

which verifies the x-component on the right-hand side of Eq. (1-33). By following the same procedure, the y- and z-components can also be verified.

The dot product of **A** and **B** is literally written, by use of Eq. (1-32), as

$$\mathbf{A} \bullet \mathbf{B} = \left(A_x \mathbf{a}_x + A_y \mathbf{a}_y + A_z \mathbf{a}_z\right) \bullet \left(B_x \mathbf{a}_x + B_y \mathbf{a}_y + B_z \mathbf{a}_z\right) \tag{1-34}$$

The distributive law expressed by Eq. (1-13b) and the orthonormality relation expressed Eq. (1-28) can reduce Eq. (1-34) to a much simpler form. The dot product of **A** and **B** in Cartesian coordinates is

$$\boxed{\mathbf{A} \bullet \mathbf{B} = A_x B_x + A_y B_y + A_z B_z} \tag{1-35}$$

It should be noted that the dot product $\mathbf{A} \bullet \mathbf{B}$ expressed by Eq. (1-35) is exactly the same as that expressed by Eq. (1-9).

The cross product of **A** and **B** is literally written, by use of Eq. (1-32), as

$$\mathbf{A} \times \mathbf{B} = \left(A_x \mathbf{a}_x + A_y \mathbf{a}_y + A_z \mathbf{a}_z\right) \times \left(B_x \mathbf{a}_x + B_y \mathbf{a}_y + B_z \mathbf{a}_z\right) \tag{1-36}$$

With the help of the distributive law expressed by Eq. (1-22c) and the cyclic relations expressed by Eq. (1-29), we can rewrite the right-hand side of Eq. (1-36) in component form. The cross product of **A** and **B** in Cartesian coordinates is therefore

$$\boxed{\mathbf{A} \times \mathbf{B} = \left(A_y B_z - A_z B_y\right)\mathbf{a}_x + \left(A_z B_x - A_x B_z\right)\mathbf{a}_y + \left(A_x B_y - A_y B_x\right)\mathbf{a}_z} \tag{1-37}$$

Alternatively, we may express Eq. (1-37) in the form of a determinant:

$$\mathbf{A}\times\mathbf{B}=\begin{vmatrix}\mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z\\ A_x & A_y & A_z\\ B_x & B_y & B_z\end{vmatrix} \tag{1-38}$$

Although the right-hand side of Eq. (1-38) has nothing to do with the determinant of a matrix, it is computed by following the determinant rule.

Up to this point, we saw that a point in space may be specified either by coordinates or by a position vector. We next consider the case in which we desire to locate a second point p_2 in the close vicinity of a given point $p_1{:}(x_1, y_1, z_1)$. If we need to specify the location of p_2 relative to point p_1, it would be more convenient to define p_2 by an infinitesimal vector drawn from p_1 to p_2 rather than the position vector drawn from the origin to p_2. By use of the differential coordinates dx, dy, and dz, the nearby point can be specified as $p_2{:}(x_1+dx, y_1+dy, z_1+dz)$. Here we define the differential length vector $d\mathbf{l}$ as an infinitesimal vector drawn from a given point to a nearby point, which is measured in meters. The differential length vector in Cartesian coordinates is expressed, regardless of the interrelationship between the differential coordinates, as follows:

$$d\mathbf{l} = dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z \qquad \text{[m]} \tag{1-39}$$

The differential coordinates dx, dy, and dz may be independent of each other or closely related to each other, depending on the specific application. Although distance vector $\mathcal{R}$ and differential length vector $d\mathbf{l}$ both start at a point in space and end at another, the magnitude and direction of $d\mathbf{l}$ are used for the quantity observed at the initial point of $d\mathbf{l}$, whereas those of $\mathcal{R}$ are used for the quantity at the terminal point of $\mathcal{R}$.

As we see in Fig. 1.17, two end points of $d\mathbf{l}$ involve a total of six surfaces of constant coordinate: $x = x_1$, $y = y_1$, $z = z_1$, $x = x_1 + dx$, $y = y_1 + dy$, and $z = z_1 + dz$ planes. Those surfaces define a rectangular parallelepiped of sides dx, dy, and dz, and a volume $dxdydz$, which we call a differential volume. In Cartesian coordinate system, it is convenient to define the differential volume as

$$dv = dx\,dy\,dz \qquad [\mathrm{m}^3] \tag{1-40}$$

The unit for dv is the cubic meter. The differential volume provides a convenient way of subdividing a finite volume in Cartesian coordinates into many differential elements of volume, and vice versa.

The differential volume dv has six boundary surfaces of a differential area $dxdy$, $dydz$, or $dzdx$ as shown in Fig. 1.17. It is convenient to represent each surface in terms of a differential area vector $d\mathbf{s}$, which is a vector whose magnitude is equal to the area of the surface, and whose unit vector is normal to the surface,

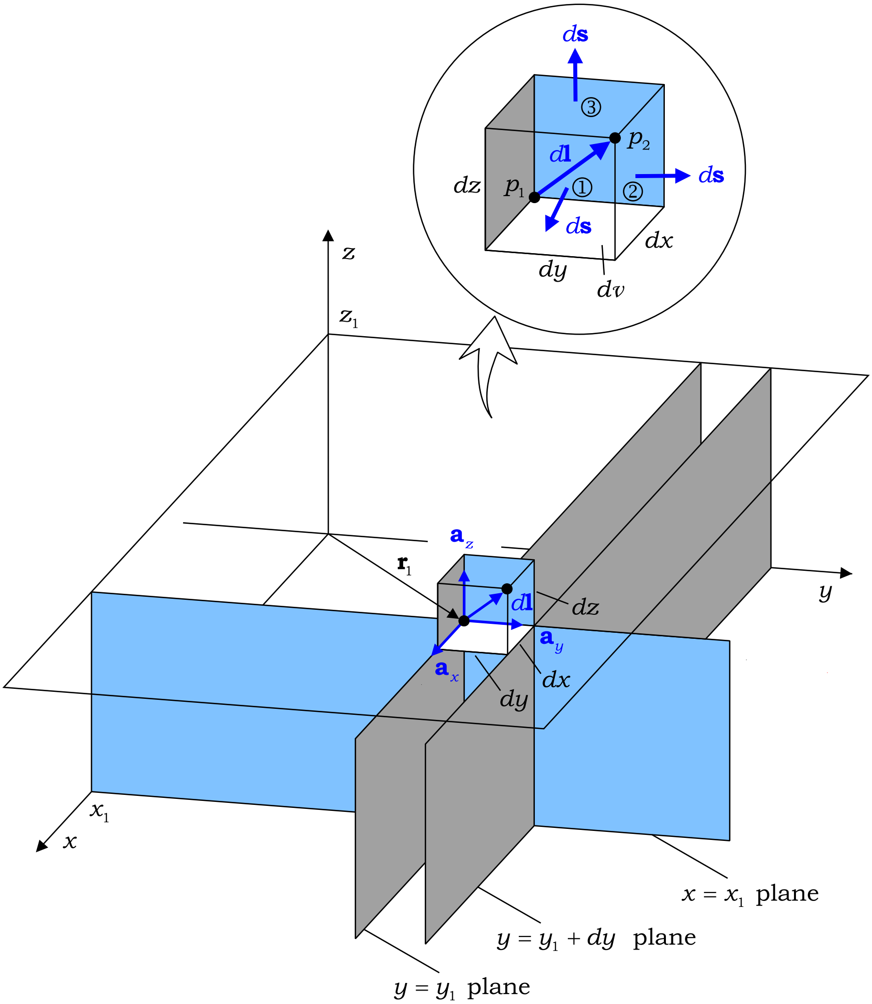


Fig. 1.17 Differential length vector $d\mathbf{l}$, differentia area vector $d\mathbf{s}$ and differential volume dv in Cartesian coordinates.

pointing outward away from the enclosed volume. The differential area vectors for the six surfaces of dv are as follows:

$$d\mathbf{s} = dydz\,\mathbf{a}_x \qquad \text{(face ①)} \tag{1-41a}$$

$$d\mathbf{s} = dxdz\,\mathbf{a}_y \qquad \text{(face ②)} \tag{1-41b}$$

$$d\mathbf{s} = dxdy\,\mathbf{a}_z \qquad \text{(face ③)} \tag{1-41c}$$

$$d\mathbf{s} = -dydz\,\mathbf{a}_x \tag{1-41d}$$

$$d\mathbf{s} = -dxdz\,\mathbf{a}_y \tag{1-41e}$$

$$d\mathbf{s} = -dxdy\,\mathbf{a}_z \tag{1-41f}$$

where the last three are for those hidden from sight behind the front faces. The unit of $d\mathbf{s}$ is the square meter. The differential area vectors given in Eq. (1-41) are useful for subdividing a surface of constant coordinate into many differential elements of surface. For an arbitrarily oriented surface, the differential area vector can be defined by following the more general method discussed in Chapter 2.

Example 1-4

Given a point $p{:}\left(2,-3,\sqrt{3}\right)$ in Cartesian coordinates, find

(a) position vector **r**, and
(b) magnitude of **r**.

Solution

(a) The scalar components of **r** are equal to the Cartesian coordinates of p

$$\mathbf{r} = 2\mathbf{a}_x - 3\mathbf{a}_y + \sqrt{3}\mathbf{a}_z\,.$$

(b) Using Eq. (1-11), we write

$$\mathbf{r}\bullet\mathbf{r} = \left(2\mathbf{a}_x - 3\mathbf{a}_y + \sqrt{3}\mathbf{a}_z\right)\bullet\left(2\mathbf{a}_x - 3\mathbf{a}_y + \sqrt{3}\mathbf{a}_z\right)$$

By applying the distributive law and orthonormality relations, we have

$$\begin{aligned}\mathbf{r}\bullet\mathbf{r} &= (2)(2) + (-3)(-3) + \left(\sqrt{3}\right)\left(\sqrt{3}\right)\\ &= 4 + 9 + 3 = 16\end{aligned}$$

The answer is

$$r = \sqrt{\mathbf{r}\bullet\mathbf{r}} = \sqrt{16} = 4\,.$$

Example 1-5

Two vectors $\mathbf{A} = 4\,\mathbf{a}_x + 3\,\mathbf{a}_y$ and $\mathbf{B} = \mathbf{a}_x + 2\,\mathbf{a}_y + 2\,\mathbf{a}_z$ given at a point with position vector $\mathbf{r} = -\mathbf{a}_x + 2\,\mathbf{a}_y + 4\mathbf{a}_z$ in Cartesian coordinates, find

(a) $\mathbf{A}\bullet\mathbf{B}$
(b) $\mathbf{A}\times\mathbf{B}$
(c) θ_{AB} between **A** and **B**

Solution

(a) Using Eq. (1-35), we write

$$\mathbf{A}\bullet\mathbf{B} = (4)(1) + (3)(2) + (0)(2) = 10\,.$$

(b) Using Eq. (1-38), we write

$$\mathbf{A}\times\mathbf{B} = \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ 4 & 3 & 0 \\ 1 & 2 & 2 \end{vmatrix}$$
$$= \mathbf{a}_x(3\times2-0\times2)-\mathbf{a}_y(4\times2-0\times1)+\mathbf{a}_z(4\times2-3\times1)$$
$$= 6\mathbf{a}_x - 8\mathbf{a}_y + 5\mathbf{a}_z.$$

(c) Magnitudes of **A** and **B** are

$$A = |\mathbf{A}| = \sqrt{(4)^2+(3)^2} = 5$$
$$B = |\mathbf{B}| = \sqrt{(1)^2+(2)^2+(2)^2} = 3$$

From Eq. (1-9), we have

$$\cos\theta_{AB} = \frac{\mathbf{A}\bullet\mathbf{B}}{AB} = \left(\frac{10}{5\times3}\right)$$

The answer is

$$\theta_{AB} = \cos^{-1}\left(\frac{2}{3}\right) = 48.2^o$$

Note that position vector **r** has nothing to do with the calculation of $\mathbf{A}\bullet\mathbf{B}$, $\mathbf{A}\times\mathbf{B}$, and θ_{AB} in this problem.

Example 1-6

Given a vector field $\mathbf{F}(\mathbf{r}) = yz\,\mathbf{a}_x - x^2\,\mathbf{a}_y + y\,\mathbf{a}_z$ in Cartesian coordinates, find

(a) vector **A** defined at the point with position vector $\mathbf{r}_1 = \mathbf{a}_x + 2\,\mathbf{a}_y + 2\,\mathbf{a}_z$,

(b) vector component of **A** in the direction parallel to $\mathbf{r}_1$, and

(c) vector component of **A** in the direction perpendicular to $\mathbf{r}_1$.

Solution

(a) Substituting $\mathbf{r} = \mathbf{r}_1$ into $\mathbf{F}(\mathbf{r})$ we obtain

$$\mathbf{A} = \mathbf{F}(\mathbf{r}_1) = (2)(2)\,\mathbf{a}_x - (1)^2\,\mathbf{a}_y + (2)\,\mathbf{a}_z$$
$$= 4\mathbf{a}_x - \mathbf{a}_y + 2\mathbf{a}_z. \qquad (1\text{-}42)$$

(b) Magnitude of $\mathbf{r}_1$ is

$$r_1 = \sqrt{(1)^2+(2)^2+(2)^2} = 3$$

Unit vector of $\mathbf{r}_1$ is

$$\mathbf{a}_{r_1} = \frac{\mathbf{r}_1}{r_1} = \frac{1}{3}\left(\mathbf{a}_x + 2\mathbf{a}_y + 2\mathbf{a}_z\right)$$

The projection of **A** onto $\mathbf{a}_{r_1}$ gives the parallel component:

$$A_{\parallel} \equiv \mathbf{A} \bullet \mathbf{a}_{r_1} = \left(4\mathbf{a}_x - \mathbf{a}_y + 2\mathbf{a}_z\right) \bullet \frac{1}{3}\left(\mathbf{a}_x + 2\mathbf{a}_y + 2\mathbf{a}_z\right) = 2$$

The vector component of **A** parallel to $\mathbf{r}_1$ is therefore

$$\mathbf{A}_{\parallel} = A_{\parallel}\,\mathbf{a}_{r_1} = \frac{2}{3}\left(\mathbf{a}_x + 2\mathbf{a}_y + 2\mathbf{a}_z\right). \tag{1-43}$$

(c) As can be seen in Fig. 1.18, the vector component of **A** perpendicular to $\mathbf{r}_1$ is given as

$$\mathbf{A}_{\perp} = \mathbf{A} - \mathbf{A}_{\parallel}$$

Inserting Eqs. (1-42) and (1-43) into the above equation we have

$$\mathbf{A}_{\perp} = \frac{1}{3}\left(10\mathbf{a}_x - 7\mathbf{a}_y + 2\mathbf{a}_z\right) \tag{1-44}$$

Alternatively, we can express $\mathbf{A}_{\perp}$ as

$$\mathbf{A}_{\perp} = \mathbf{a}_{r_1} \times \left(\mathbf{A} \times \mathbf{a}_{r_1}\right) \tag{1-45}$$

The term in parenthesis has the same magnitude as $\mathbf{A}_{\perp}$, but is rotated by 90^o with respect to the direction of $\mathbf{A}_{\perp}$. The cross product outside the parenthesis is to rotate the vector.

The term in parenthesis in Eq. (1-45) is evaluated as

$$\mathbf{A} \times \mathbf{a}_{r_1} = \frac{1}{3}\begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ 4 & -1 & 2 \\ 1 & 2 & 2 \end{vmatrix}$$

$$= \frac{1}{3}\left[\mathbf{a}_x\left\{(-1)\times 2 - 2\times 2\right\} - \mathbf{a}_y\left(4\times 2 - 2\times 1\right) + \mathbf{a}_z\left\{4\times 2 - (-1)\times 1\right\}\right]$$

$$= -2\,\mathbf{a}_x - 2\,\mathbf{a}_y + 3\,\mathbf{a}_z$$

The perpendicular component is therefore

$$\mathbf{A}_{\perp} = \mathbf{a}_{r_1} \times \left(\mathbf{A} \times \mathbf{a}_{r_1}\right) = \frac{1}{3}\begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ 1 & 2 & 2 \\ -2 & -2 & 3 \end{vmatrix} = \frac{1}{3}\left(10\,\mathbf{a}_x - 7\mathbf{a}_y + 2\,\mathbf{a}_z\right) \tag{1-46}$$

Equation (1-46) is the same as Eq. (1-44).

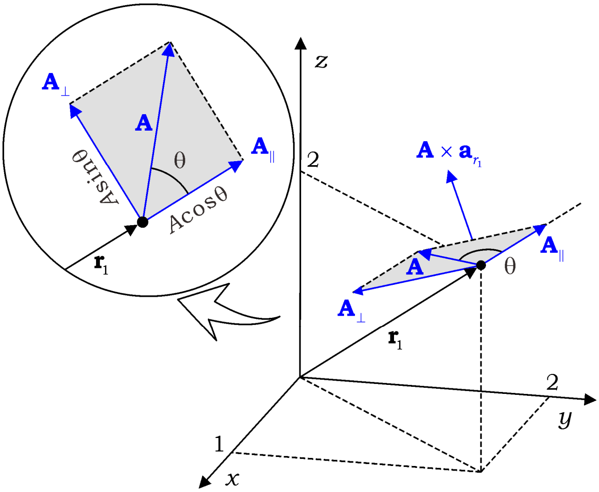


Fig. 1.18 Decomposition of a vector **A** .

Exercise 1.12
Given $\mathbf{A} = 2\mathbf{a}_x + \mathbf{a}_y + 3\mathbf{a}_z$, find (a) $|\mathbf{A}|$, (b) $\mathbf{a}_A$, and (c) angles between **A** and the x-, y-, and z-axes. (d) Is it a position vector or a vector in a vector field?
Ans. (a) $\sqrt{14}$, (b) $\mathbf{a}_A = 0.535\mathbf{a}_x + 0.267\mathbf{a}_y + 0.802\mathbf{a}_z$, (c) 57.7^o, 74.5^o, and 36.7^o, (d) not clearly defined.

Exercise 1.13
Find the projection of $\mathbf{A} = 4\,\mathbf{a}_x + 3\,\mathbf{a}_y + 8\,\mathbf{a}_z$ on the xy-plane.
Ans. $\sqrt{4^2 + 3^2} = 5$.

Exercise 1.14
Which faces of dv in Fig. 1.17 are referred to by Eqs. (1-41)(d)(e)(f)?
Ans. Rear, left, and bottom.

Exercise 1.15
Locate the point at which the magnitude and direction of the following vectors are most useful? (a) a vector in a vector field, (b) **r**, (c) $\mathcal{R}$, (d) $d\mathbf{l}$, and (f) $d\mathbf{s}$.
Ans. (a) Initial, (b) Terminal, (c) Terminal, (d) Initial, (f) Initial point of the vector.

Exercise 1.16
What are (a) the similarity and (b) the dissimilarity between the position vector $\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z$ and a vector field $\mathbf{A} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z$.
Ans. (a) Magnitude and direction, (b) Location.

1.3.2 Cylindrical Coordinate System

Cylindrical coordinate system specifies a point $p_1:(\rho_1, \phi_1, z_1)$ by the intersection of three surfaces; that is, a cylindrical surface of radius ρ_1 centered on the z-axis, a half-plane rotated about the z-axis by an angle ϕ_1, and the $z = z_1$ plane. The coordinate ρ_1 is the radial distance from the z-axis, and ϕ_1 is the azimuth measured from the positive x-axis in the xy-plane. The ranges of ρ, ϕ, and z are $0 \le \rho < \infty$ [m], $0 \le \phi \le 2\pi$ [rad], and $-\infty < z < \infty$ [m], respectively. At a point p_1, the base vectors $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ are defined in such a way that they are perpendicular to the aforementioned surfaces of constant coordinate, and point in the direction of increasing coordinate. It should be noted that unit vectors $\mathbf{a}_\rho$ and $\mathbf{a}_\phi$ vary with ϕ, whereas $\mathbf{a}_z$ is constant in space.

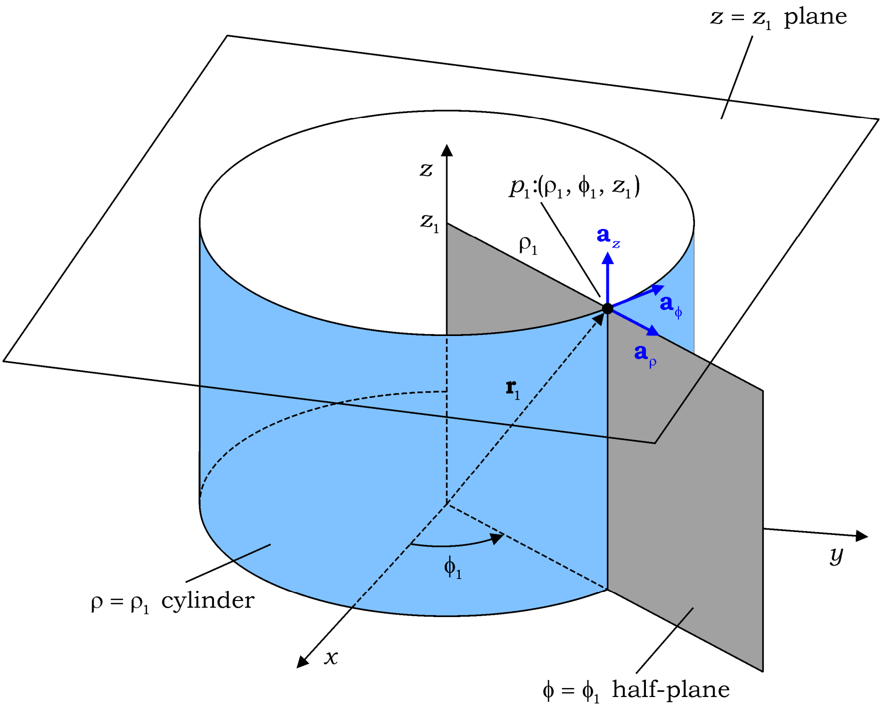


Fig. 1.19 Cylindrical coordinate system.

Here we digress briefly and consider cylindrical symmetry. An object is said to have cylindrical symmetry, if ***it appears the same as we rotate it about the z-axis, or as we move around it varying*** ϕ ***while keeping*** ρ ***and*** z ***constant***. As examples, consider three vector fields $\mathbf{U}(\mathbf{r}) = \mathbf{a}_\rho / \rho$, $\mathbf{V}(\mathbf{r}) = \mathbf{a}_\phi / \rho$, and

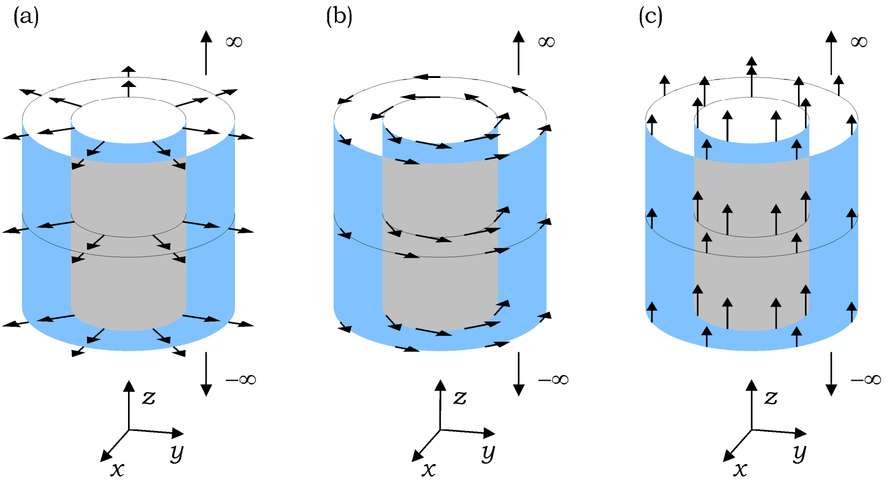


Fig. 1.20 Three vector fields with cylindrical symmetry. (a) $\mathbf{U(r)} = \mathbf{a}_\rho / \rho$ (b) $\mathbf{V(r)} = \mathbf{a}_\phi / \rho$ (c) $\mathbf{W(r)} = \mathbf{a}_z / \rho$.

$\mathbf{W(r)} = \mathbf{a}_z / \rho$ as shown in Fig. 1.20. They all have cylindrical symmetry, since each appears the same as we move around it in the direction of $\mathbf{a}_\phi$.

The base vectors for cylindrical coordinates obey the orthonormality relations:

$$\mathbf{a}_\rho \cdot \mathbf{a}_\phi = \mathbf{a}_\phi \cdot \mathbf{a}_z = \mathbf{a}_z \cdot \mathbf{a}_\rho = 0 \tag{1-47a}$$

$$\mathbf{a}_\rho \cdot \mathbf{a}_\rho = \mathbf{a}_\phi \cdot \mathbf{a}_\phi = \mathbf{a}_z \cdot \mathbf{a}_z = 1 \tag{1-47b}$$

Equation (1-47a) shows that the base vectors are mutually perpendicular, even though $\mathbf{a}_\rho$ and $\mathbf{a}_\phi$ are functions of ϕ, and Eq. (1-47b) shows that the base vectors are unit vectors.

In the right-handed cylindrical system, the base vectors satisfy the cyclic relations:

$$\mathbf{a}_\rho \times \mathbf{a}_\phi = \mathbf{a}_z \tag{1-48a}$$

$$\mathbf{a}_\phi \times \mathbf{a}_z = \mathbf{a}_\rho \tag{1-48b}$$

$$\mathbf{a}_z \times \mathbf{a}_\rho = \mathbf{a}_\phi \tag{1-48c}$$

The three vectors in the cyclic permutations obey the right-hand rule: when the right fingers rotate from the first to the second vector, the thumb points in the direction of the third vector.

Regardless of the coordinate system, position vector is always drawn from the origin to a point in space. Fig. 1.19 shows that position vector $\mathbf{r}_1$ uniquely specifies a point p_1:(ρ_1, ϕ_1, z_1) in cylindrical coordinates. Following the same procedure as was used for position vector in Cartesian coordinates, we obtain the vector components of $\mathbf{r}_1$ by projecting the position vector onto the base vectors $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ at point p_1 and attaching corresponding unit vectors. It is apparent from Fig. 1.19 that $\mathbf{r}_1$ is always perpendicular to $\mathbf{a}_\phi$, and thus has no ϕ-component. Omitting 1 for generalization, position vector in cylindrical coordinates is

$$\boxed{\mathbf{r} = \rho\mathbf{a}_\rho + z\mathbf{a}_z} \tag{1-49}$$

In the literal sense in which $\mathbf{a}_\rho$ is a function of ϕ, the position vector $\mathbf{r}$ in Eq. (1-49) can uniquely specify a given point by means of ρ, z, and $\mathbf{a}_\rho$, which all vary with position. If the geometry of a given configuration has cylindrical symmetry, $\mathbf{a}_\rho$ is considered to be a vector field with cylindrical symmetry and merged into the given configuration with the same symmetry. Under this condition, the position vector in Eq. (1-49) appears to vary with ρ and z only.

When vector $\mathbf{A}$ is defined at a point p_1:(ρ_1, ϕ_1, z_1) in cylindrical coordinates as shown in Fig. 1.21, its scalar components are equal to its projections onto the base vectors $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ at point p_1. The component form of $\mathbf{A}$ in cylindrical coordinates is in general expressed as

$$\boxed{\mathbf{A} = A_\rho\mathbf{a}_\rho + A_\phi\mathbf{a}_\phi + A_z\mathbf{a}_z} \tag{1-50}$$

where A_ρ, A_ϕ, and A_z are the scalar components, which generally vary with position. It is important to note that the unit vectors $\mathbf{a}_\rho$ and $\mathbf{a}_\phi$ vary as functions of ϕ. Since $\mathbf{A}$ is a vector in a vector field, its magnitude and direction are specific to point p_1, and its terminal point is a non-spatial point.

Symmetry considerations may allow us to predict the functional form of the final solution, which will greatly facilitate the solution of a given problem. Some electromagnetic problems actually require the functional form of the solution even before the problem can be solved: Gauss's and Ampere's laws. Let us consider the case in which the source of an electromagnetic quantity is distributed in a region of space in such a way as to have cylindrical symmetry. Then the resultant field quantity should also have cylindrical symmetry. When cylindrical symmetry exists, the geometry of a given configuration appears the same even if it is rotated about the z-axis. Therefore the resulting field should be independent of ϕ such as $G(\mathbf{r}) = G(\rho, z)$ for a scalar field and $\mathbf{H}(\mathbf{r}) = H_\rho(\rho, z)\,\mathbf{a}_\rho + H_\phi(\rho, z)\,\mathbf{a}_\phi + H_z(\rho, z)\,\mathbf{a}_z$ for a vector field. In many cases, the cylindrical symmetry is accompanied by other symmetries such as a translational symmetry in the z-direction and a 2-fold rotational symmetry about any horizontal axis. We can simply identify these symmetries by observing whether the geometry appears the same when it is displaced in the z-direction or turned upside

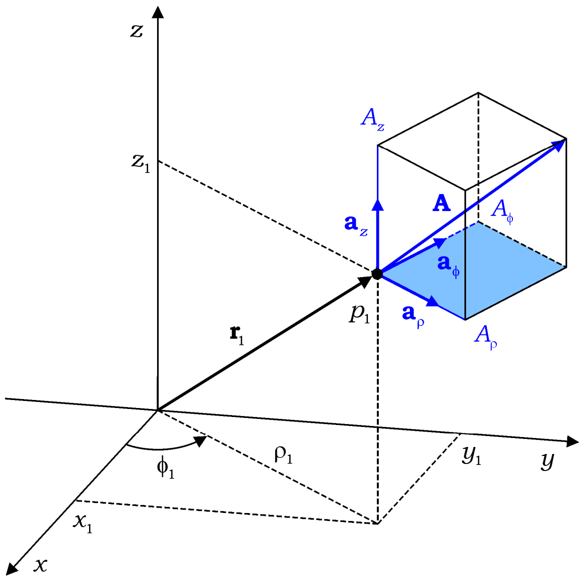


Fig. 1.21 Vector **A** at point p_1:(ρ_1, ϕ_1, z_1) in cylindrical coordinate system.

down. For instance, an infinitely long filament lying along the z-axis has the cylindrical, translational, and 2-fold rotational symmetries. The translational symmetry assures that the resultant field is independent of z as well, namely, $G(\mathbf{r}) = G(\rho)$ and $\mathbf{H(r)} = H_\rho(\rho)\,\mathbf{a}_\rho + H_\phi(\rho)\,\mathbf{a}_\phi + H_z(\rho)\,\mathbf{a}_z$. Moreover, the 2-fold rotational symmetry guarantees that the resulting vector field has neither the ϕ-component nor the z-component. This is because the vector components $H_\phi(\rho)\,\mathbf{a}_\phi$ and $H_z(\rho)\,\mathbf{a}_z$ do not have the 2-fold rotational symmetry: they reverse the sign when turned upside down(see Fig. 1.20). Consequently, in the presence of the cylindrical, translational, and 2-fold rotational symmetries, the resultant vector field must be of the form $\mathbf{H(r)} = H_\rho\left(\rho\right)\mathbf{a}_\rho$.

As an example, the vector field **U**(**r**) shown in Fig. 1.20 has cylindrical, translational, and 2-fold rotational symmetries, whereas the vector fields **V**(**r**) and **W**(**r**) have cylindrical and translational symmetries.

When vectors **A** and **B** are defined simultaneously at a point in cylindrical coordinates, they can be expanded in component form, according to Eq. (1-50), as

$$\mathbf{A} = A_\rho\,\mathbf{a}_\rho + A_\phi\,\mathbf{a}_\phi + A_z\,\mathbf{a}_z \tag{1-51a}$$

$$\mathbf{B} = B_\rho\,\mathbf{a}_\rho + B_\phi\,\mathbf{a}_\phi + B_z\,\mathbf{a}_z \tag{1-51b}$$

where $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ are the base vectors at the given point.

The vector addition of **A** and **B** in cylindrical coordinates is obtained from Eq. (1-51) as

$$\boxed{\mathbf{A}+\mathbf{B}=\left(A_\rho+B_\rho\right)\mathbf{a}_\rho+\left(A_\phi+B_\phi\right)\mathbf{a}_\phi+\left(A_z+B_z\right)\mathbf{a}_z} \tag{1-52}$$

The dot product of **A** and **B** in cylindrical coordinates is obtained from Eq. (1-51), by applying the distributive law of dot product given in Eq. (1-13b) and the orthonormality relations given in Eq. (1-47), as

$$\boxed{\mathbf{A}\bullet\mathbf{B}=A_\rho B_\rho+A_\phi B_\phi+A_z B_z} \tag{1-53}$$

The cross product of **A** and **B** in cylindrical coordinates is obtained from Eq. (1-51), by applying the distributive law of cross product given in Eq. (1-22c) and the cyclic relations given in Eq. (1-48), and written in determinant form as

$$\boxed{\mathbf{A}\times\mathbf{B}=\begin{vmatrix}\mathbf{a}_\rho & \mathbf{a}_\phi & \mathbf{a}_z\\ A_\rho & A_\phi & A_z\\ B_\rho & B_\phi & B_z\end{vmatrix}} \tag{1-54}$$

It is important to note that the rules of vector algebra expressed by Eqs. (1-52)-(1-54) are based on the assumption that **A** and **B** are defined at the same point in cylindrical coordinates. As an example, suppose that vector **A** is defined at a point $(\rho,\phi,z)=(1,0,0)$ such that $\mathbf{A}=\mathbf{a}_\rho$, while vector **B** is defined at another point $(\rho,\phi,z)=(1,\pi,0)$ such that $\mathbf{B}=\mathbf{a}_\rho$. Under these conditions, the dot product of **A** and **B** is equal to -1, instead of 1, because they are actually directed in the opposite directions on the x-axis; Eq. (1-53) is not applicable in this case. ***Vector algebra assumes that vectors have the same location***.

The differential length vector $d\mathbf{l}$ is an infinitesimal vector drawn from a given point $p_1{:}(\rho_1,\phi_1,z_1)$ to a nearby point p_2 whose cylindrical coordinates deviate only very slightly from those of p_1, namely $p_2{:}(\rho_1+d\rho,\phi_1+d\phi,z_1+dz)$. As was previously mentioned, the magnitude and direction of $d\mathbf{l}$ are used for the quantity defined at point p_1, and thus $d\mathbf{l}$ is expanded by the base vectors at point p_1. It is apparent from Fig. 1.22 that the projections of $d\mathbf{l}$ onto the base vectors $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ are equal to the differential lengths $d\rho$, $\rho_1 d\phi$, and dz, respectively. Here the differential angle $d\phi$ is converted to the differential length $\rho_1 d\phi$, because $d\mathbf{l}$ is measured in meters. Omitting 1 for generalization, the differential length vector in cylindrical coordinates is expressed as

$$\boxed{d\mathbf{l}=d\rho\,\mathbf{a}_\rho+\rho d\phi\,\mathbf{a}_\phi+dz\,\mathbf{a}_z} \quad [\text{m}] \tag{1-55}$$

In cylindrical coordinates, $d\mathbf{l}$ is a function of position through ρ, $\mathbf{a}_\rho$, and $\mathbf{a}_\phi$, which all vary with position. It is important to remember that $d\mathbf{l}$ is always given as Eq. (1-55) in cylindrical coordinates, regardless of the interrelationship between $d\rho$, $d\phi$, and dz, if the differential coordinates are nonzero.

Two end points of $d\mathbf{l}$ involve a total of six surfaces of constant coordinate as shown in Fig. 1.22. Those six surfaces define a differential volume dv in cylindrical coordinates. Since $|d\mathbf{l}|$ is infinitesimally small, the differential volume may be considered to be a rectangular parallelepiped of sides $d\rho$, $\rho_1 d\phi$, and dz, having a volume of $\rho_1\, d\rho\, d\phi\, dz$. The differential volume in cylindrical coordinates is in general defined as

$$\boxed{dv = \rho\, d\rho\, d\phi\, dz} \qquad [\mathrm{m}^3] \qquad (1\text{-}56)$$

Although dv may vary with position, the differential volume is always defined as Eq. (1-56) in cylindrical coordinates. The differential volume is useful for subdividing a volume in cylindrical coordinates into many differential elements of volume, and vice versa.

The differential volume dv is bounded by six infinitesimal surfaces as shown in Fig. 1.22. Each surface is represented by a differential area vector $d\mathbf{s}$, whose magnitude is equal to the area of the surface, and whose unit vector is normal to the surface, pointing outward from the enclosed volume. The differential area vectors for the six surfaces of dv shown in Fig. 1.22 are expressed as follows:

$$d\mathbf{s} = \rho_1\, d\phi\, dz\, \mathbf{a}_\rho \qquad \text{(face ①)} \qquad (1\text{-}57a)$$

$$d\mathbf{s} = -d\rho\, dz\, \mathbf{a}_\phi \qquad \text{(face ②)} \qquad (1\text{-}57b)$$

$$d\mathbf{s} = \rho_1\, d\phi\, d\rho\, \mathbf{a}_z \qquad \text{(face ③)} \qquad (1\text{-}57c)$$

$$d\mathbf{s} = -\rho_1\, d\phi\, dz\, \mathbf{a}_\rho \qquad (1\text{-}57d)$$

$$d\mathbf{s} = d\rho\, dz\, \mathbf{a}_\phi \qquad (1\text{-}57e)$$

$$d\mathbf{s} = -\rho_1\, d\phi\, d\rho\, \mathbf{a}_z \qquad (1\text{-}57f)$$

where the last three are for those hidden from sight behind the front faces. The differential area vectors expressed by Eq. (1-57) are useful for subdividing a surface of constant coordinate into many differential elements of surface in cylindrical coordinates, and vice versa. However, Eq. (1-57) cannot be used for other surfaces, such as $x = 1$ plane or a cylindrical surface lying along the x-axis.

Example 1-7

With reference to the three points $p_1{:}(\rho_1, \phi_1, z_1)$, p_2, and p_3 in Fig. 1.22, find

(a) expression for $\mathbf{r}_3$,

(b) cylindrical coordinates of p_2, and

(c) expression for $\mathbf{r}_2$.

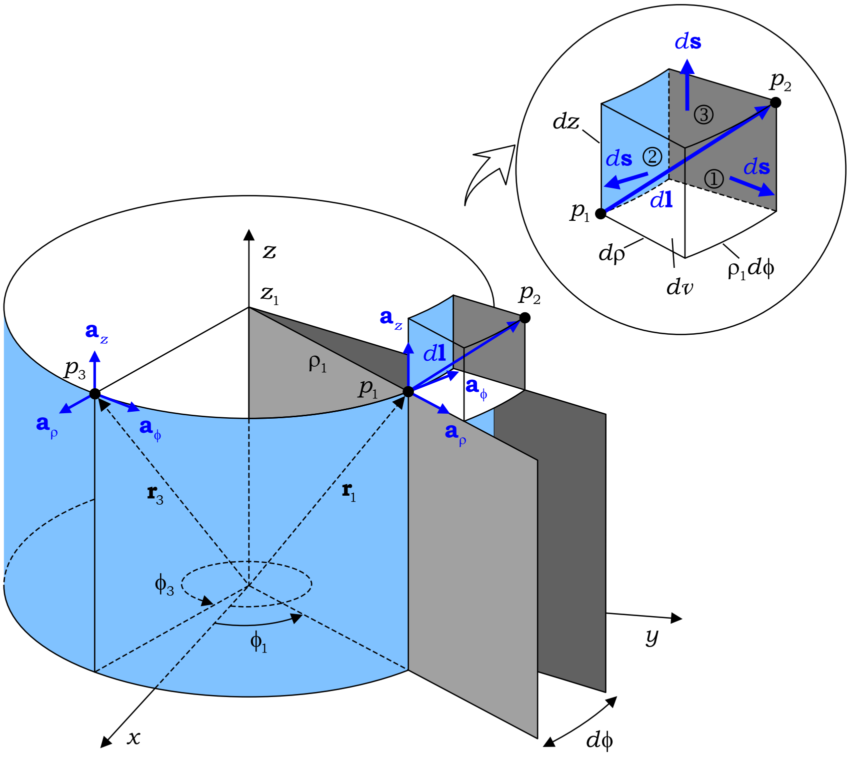


Fig. 1.22 Differential length vector $d\mathbf{l}$, differential area vector $d\mathbf{s}$ and differential volume dv in cylindrical coordinates.

Solution

(a) Projections of $\mathbf{r}_3$ onto $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ at point p_3 lead to

$$\mathbf{r}_3 = \rho_1\, \mathbf{a}_\rho + z_1\, \mathbf{a}_z$$

This has the same form as the position vector $\mathbf{r}_1$ of point p_1. In the presence of cylindrical symmetry, the vector function $\mathbf{a}_\rho$ with cylindrical symmetry merges into the geometry of configuration having the same symmetry. In this case, $\mathbf{r}_1$ and $\mathbf{r}_3$ appear to vary with ρ and z only, implying that two points p_1 and p_3 are indistinguishable.

(b) The coordinate differences between p_1 and p_2 are $d\rho$, $d\phi$, and dz. Thus the cylindrical coordinates of p_2 are

$$(\rho_1 + d\rho,\ \phi_1 + d\phi,\ z_1 + dz)\,. \tag{1-58}$$

(c) Projections of $\mathbf{r}_2$ onto $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ at point p_2, with the help of Eq. (1-58), lead to

$$\mathbf{r}_2 = (\rho_1 + d\rho)\,\mathbf{a}_\rho + (z_1 + dz)\,\mathbf{a}_z \tag{1-59}$$

Alternatively, by using $\mathbf{r}_1 = \rho_1\,\mathbf{a}_\rho + z_1\,\mathbf{a}_z$ and $d\mathbf{l} = d\rho\,\mathbf{a}_\rho + \rho_1 d\phi\,\mathbf{a}_\phi + dz\,\mathbf{a}_z$ defined at point p_1, one might be tempted to express $\mathbf{r}_2$ as

$$\mathbf{r}_2 = \mathbf{r}_1 + d\mathbf{l} = (\rho_1 + d\rho)\mathbf{a}_\rho + \rho_1 d\phi\,\mathbf{a}_\phi + (z_1 + dz)\mathbf{a}_z \tag{1-60}$$

It should be noted that the unit vectors in Eq. (1-59) are the base vectors at point p_2, whereas those in Eq. (1-60) are the base vectors at p_1. Eq. (1-60) is wrong in view of the fact that ***position vector should be expanded by the base vectors defined at its terminal point.***

Example 1-8

Both $\mathbf{a}_\rho$ and $\mathbf{a}_\phi$ form vector fields, which are smooth functions of position, and thus have continuous partial derivatives. Verify

(a) $\dfrac{\partial \mathbf{a}_\rho}{\partial \phi} = \mathbf{a}_\phi$, and (b) $\dfrac{\partial \mathbf{a}_\phi}{\partial \phi} = -\mathbf{a}_\rho$.

Solution

(a) With reference to Fig. 1.23(a), from calculus we write the rate of change of $\mathbf{a}_\rho$ with ϕ as

$$\left.\frac{\partial \mathbf{a}_\rho}{\partial \phi}\right|_{\phi=\phi_1} = \lim_{\Delta\phi \to 0} \frac{\mathbf{a}_\rho^2 - \mathbf{a}_\rho^1}{\phi_2 - \phi_1} \equiv \lim_{\Delta\phi \to 0} \frac{\Delta \mathbf{a}'}{\Delta \phi} \tag{1-61}$$

Here, $\mathbf{a}_\rho^1$ and $\mathbf{a}_\rho^2$ are unit vectors at $\phi = \phi_1$ and $\phi = \phi_2$, respectively.

In the figure, $\bar{\mathbf{a}}_\rho^1$ is a displaced version of $\mathbf{a}_\rho^1$; they are equal. Consider a triangle formed by three vectors: $\bar{\mathbf{a}}_\rho^1$, $\mathbf{a}_\rho^2$, and $\Delta\mathbf{a}' = \mathbf{a}_\rho^2 - \bar{\mathbf{a}}_\rho^1$. Noting that $|\bar{\mathbf{a}}_\rho^1| = |\mathbf{a}_\rho^2| = 1$, we have

$$|\Delta\mathbf{a}'| = |\mathbf{a}_\rho^2|\,\Delta\phi = \Delta\phi \tag{1-62}$$

As $\Delta\phi \to 0$, we see that $\mathbf{r}_2 \to \mathbf{r}_1$, $\mathbf{a}_\rho^2 \to \bar{\mathbf{a}}_\rho^1$, and $\Delta\mathbf{a}'$ tends to be a vector perpendicular to $\mathbf{a}_\rho^1$, pointing in the direction of $\mathbf{a}_\phi$. In view of these, we have

$$\Delta\mathbf{a}' = \Delta\phi\,\mathbf{a}_\phi \tag{1-63}$$

Substitution of Eq. (1-63) into Eq. (1-61) gives

$$\boxed{\frac{\partial \mathbf{a}_\rho}{\partial \phi} = \mathbf{a}_\phi.} \tag{1-64}$$

(b) With reference to Fig. 1.23(b), the rate of change of $\mathbf{a}_\phi$ with ϕ is written as

$$\left.\frac{\partial \mathbf{a}_\phi}{\partial \phi}\right|_{\phi=\phi_1} = \lim_{\Delta\phi \to 0} \frac{\mathbf{a}_\phi^2 - \mathbf{a}_\phi^1}{\phi_2 - \phi_1} \equiv \lim_{\Delta\phi \to 0} \frac{\Delta \mathbf{a}}{\Delta \phi} \tag{1-65}$$

Here, $\mathbf{a}_\phi^1$ and $\mathbf{a}_\phi^2$ are unit vectors at $\phi = \phi_1$ and $\phi = \phi_2$, respectively.

In the figure, $\bar{\mathbf{a}}_\phi^1$ is a displaced version of $\mathbf{a}_\phi^1$; they are equal. Consider a triangle formed by three vectors: $\bar{\mathbf{a}}_\phi^1$, $\mathbf{a}_\phi^2$, and $\Delta\mathbf{a} = \mathbf{a}_\phi^2 - \bar{\mathbf{a}}_\phi^1$. Noting that $|\bar{\mathbf{a}}_\phi^1| = |\mathbf{a}_\phi^2| = 1$, we have

$$|\Delta \mathbf{a}| = |\mathbf{a}_\phi^2| \Delta\phi = \Delta\phi \tag{1-66}$$

As $\Delta\phi \to 0$, we see that $\mathbf{r}_2 \to \mathbf{r}_1$, $\mathbf{a}_\phi^2 \to \bar{\mathbf{a}}_\phi^1$, and $\Delta\mathbf{a}$ tends to be a vector perpendicular to $\mathbf{a}_\phi^1$, pointing in the direction of $-\mathbf{a}_\rho$. In view of these, we have

$$\Delta \mathbf{a} = -\Delta\phi\, \mathbf{a}_\rho \tag{1-67}$$

Substitution of Eq. (1-67) into Eq. (1-65) gives

$$\boxed{\frac{\partial \mathbf{a}_\phi}{\partial \phi} = -\mathbf{a}_\rho} \tag{1-68}$$

The results given in Eqs. (1-64) and (1-68) can also be verified by using the coordinate transformation discussed in the next section.

Exercise 1.17

Given a point $p_1{:}\left(1, \sqrt{3}, 5\right)$ in Cartesian coordinates, find (a) cylindrical coordinates of p_1, and (b) position vector of p_1 in cylindrical coordinates.

Ans. (a) $p_1{:}(2, 60^o, 5)$, (b) $\mathbf{r}_1 = 2\,\mathbf{a}_\rho + 5\,\mathbf{a}_z$.

Exercise 1.18

Do the following vector fields have cylindrical symmetry?

(a) $\rho\,\mathbf{a}_z$, (b) $(\sin\phi)\,\mathbf{a}_\rho$, (c) $3\,\mathbf{a}_\phi$, (d) $(z-1)\,\mathbf{a}_\rho$, and (e) $\mathbf{V(r)} = \mathbf{a}_z \times \mathbf{r}$.

Ans. (a) Yes, (b) No, (c) Yes, (d) Yes, (e) Yes.

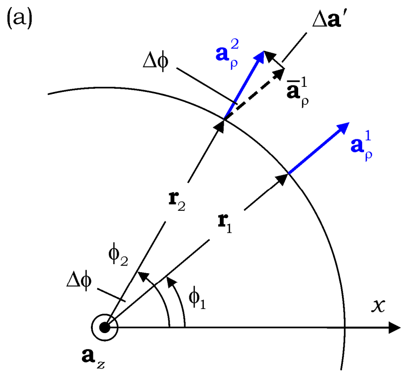


Fig. 1.23 (a) $\partial \mathbf{a}_\rho / \partial \phi$ (b) $\partial \mathbf{a}_\phi / \partial \phi$.

Exercise 1.19
Given two points p_1:$(2\sqrt{3}, 60^o, 4)$ and p_2:$(2, 90^o, 1)$ in cylindrical coordinates, find an expression for the distance vector from p_2 to p_1 in Cartesian coordinates.

Ans. $\mathcal{R} = \sqrt{3}\,\mathbf{a}_x + \mathbf{a}_y + 3\,\mathbf{a}_z$.

1.3.3 Spherical Coordinate System

Spherical coordinate system specifies a point p_1:(R_1, θ_1, ϕ_1) by the intersection of three surfaces; that is, a spherical surface of radius R_1 centered at the origin, a conical surface of half angle θ_1 with the apex at the origin, and a half-plane rotated about the z-axis by an angle ϕ_1. The coordinate R_1 is the radial distance from the origin. The coordinate θ_1 is called the polar angle measured from the $+z$-axis, and the coordinate ϕ_1 is called the azimuth measured from the $+x$-axis in the xy-plane. The ranges of R, θ, and ϕ are $0 \le R < \infty$ [m], $0 \le \theta \le \pi$ [rad], and $0 \le \phi \le 2\pi$ [rad], respectively. At a point p_1, the base vectors $\mathbf{a}_R$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ are defined in such a way that they are perpendicular to the aforementioned surfaces of constant coordinate, pointing in the direction of increasing coordinate. All three vectors $\mathbf{a}_R$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ are functions of position in spherical coordinates.

An object is said to have spherical symmetry, if ***it appears the same as we rotate it about any axis passing through the center, or as we move around it varying*** θ ***and*** ϕ ***while keeping*** R ***constant***. For instance, a vector field $\mathbf{U}(\mathbf{r}) = \mathbf{a}_R$ appears the same as we move around it, and thus has spherical symmetry(see Fig. 1.25). In contrast, the vector fields $\mathbf{V}(\mathbf{r}) = \mathbf{a}_\theta$ and $\mathbf{W}(\mathbf{r}) = \mathbf{a}_\phi$ have no spherical symmetry because they appear to reverse the sign when they are rotated about the x-axis by 180^o .

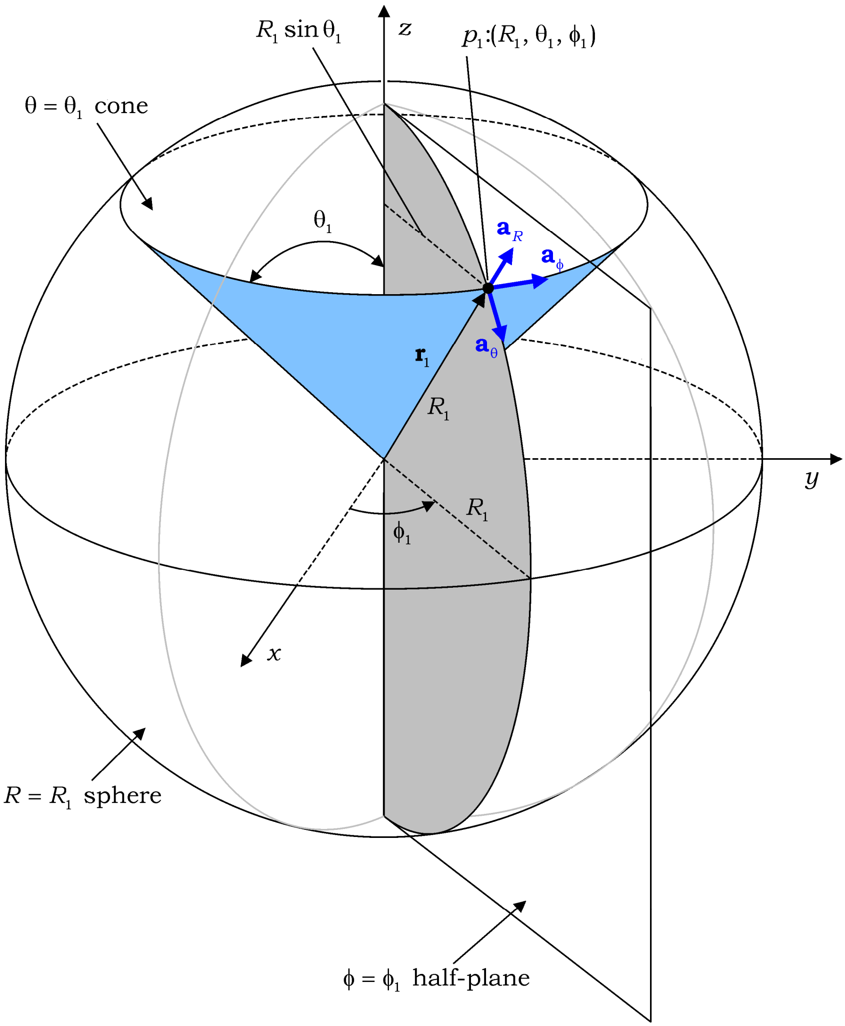


Fig. 1.24 Spherical coordinate system.

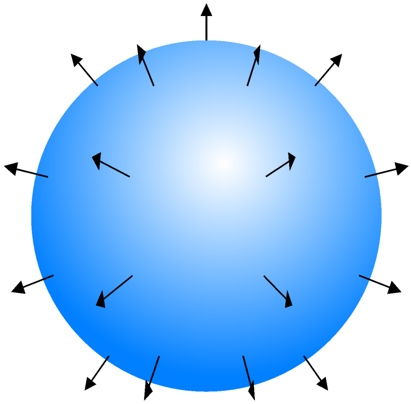


Fig. 1.25 Vector field $\mathbf{U}(\mathbf{r}) = \mathbf{a}_R$ with spherical symmetry.

The base vectors for spherical coordinates obey the orthonormality relations:

$$\mathbf{a}_R \cdot \mathbf{a}_\theta = \mathbf{a}_\theta \cdot \mathbf{a}_\phi = \mathbf{a}_\phi \cdot \mathbf{a}_R = 0 \tag{1-69a}$$

$$\mathbf{a}_R \cdot \mathbf{a}_R = \mathbf{a}_\theta \cdot \mathbf{a}_\theta = \mathbf{a}_\phi \cdot \mathbf{a}_\phi = 1 \tag{1-69b}$$

Equation (1-69a) shows that the base vectors are mutually perpendicular, even though they all vary from point to point, and Eq. (1-69b) shows that the base vectors are unit vectors.

In the right-handed spherical system, the base vectors satisfy the cyclic relations:

$$\mathbf{a}_R \times \mathbf{a}_\theta = \mathbf{a}_\phi \tag{1-70a}$$

$$\mathbf{a}_\theta \times \mathbf{a}_\phi = \mathbf{a}_R \tag{1-70b}$$

$$\mathbf{a}_\phi \times \mathbf{a}_R = \mathbf{a}_\theta \tag{1-70c}$$

The three vectors in the cyclic permutations obey the right-hand rule: when the right fingers rotate from the first to the second vector, the thumb points in the direction of the third vector.

A point p_1:(R_1, θ_1, ϕ_1) in spherical coordinates is uniquely specified by position vector $\mathbf{r}_1$ drawn from the origin to the point, as shown in Fig. 1.24. In view of the role of $\mathbf{r}_1$, the position vector is expanded by the base vectors defined at point p_1. It is evident from Fig. 1.24 that $\mathbf{r}_1$ is always perpendicular to $\mathbf{a}_\theta$ and $\mathbf{a}_\phi$ at point p_1, meaning that $\mathbf{r}_1$ only has a vector component along the radial direction, that is, $R_1 \mathbf{a}_R$. Position vector in spherical coordinates is in general given as

$$\boxed{\mathbf{r} = R\,\mathbf{a}_R} \tag{1-71}$$

In the literal sense in which $\mathbf{a}_R$ is a function of θ and ϕ, the position vector $\mathbf{r}$ in Eq. (1-71) can uniquely specify a point, with coordinates R, θ, and ϕ, by means of R and $\mathbf{a}_R$. If the geometry of a given configuration has spherical symmetry, $\mathbf{a}_R$ is considered to be a vector field with spherical symmetry and merged into the given configuration with the same symmetry. Under this condition, the position vector expressed by Eq. (1-71) appears to vary with R only, which is in accord with the fact that two points at the same radial distance from the origin are indistinguishable in the presence of the spherical symmetry.

When vector **A** is defined at a point $p_1{:}(R_1, \theta_1, \phi_1)$ in spherical coordinates as shown in Fig. 1.26, its scalar components are equal to its projections onto the base vectors $\mathbf{a}_R$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ at point p_1. The component form of **A** in spherical coordinates is in general expressed as

$$\boxed{\mathbf{A} = A_R\mathbf{a}_R + A_\theta\mathbf{a}_\theta + A_\phi\mathbf{a}_\phi} \tag{1-72}$$

where A_R, A_θ, and A_ϕ are the scalar components, which generally vary with position. It is important to note that the three base vectors vary from point or point in space, and that the base vectors are specific to the point where **A** is defined.

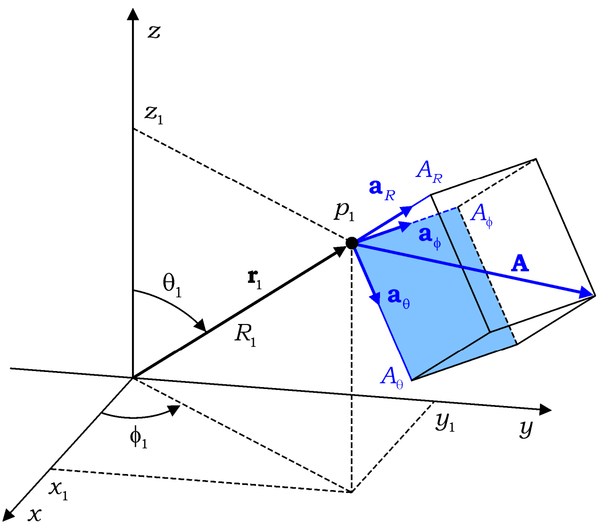


Fig. 1.26 Vector **A** defined at point $p_1{:}(R_1, \theta_1, \phi_1)$ in spherical coordinates.

Symmetry considerations can make it possible to predict the general form of the final solution. If the source is distributed in a region of space in such a way as to have spherical symmetry, the resultant field should also have spherical symmetry. When spherical symmetry exists, the geometry of a given configuration appears the same as even if we move around it along the direction of $\mathbf{a}_\theta$ or $\mathbf{a}_\phi$. Therefore the resulting field should be independent of θ and ϕ such that $V(\mathbf{r}) = V(R)$ for a scalar field and $\mathbf{D(r)} = D_R(R)\,\mathbf{a}_R + D_\theta(R)\,\mathbf{a}_\theta + D_\phi(R)\,\mathbf{a}_\phi$ for a vector field. Furthermore, the spherical symmetry assures that the vector field has neither the θ- nor the ϕ-component. This is because the vector components $D_\theta(R)\,\mathbf{a}_\theta$ and $D_\phi(R)\,\mathbf{a}_\phi$ brake the symmetry: they reverse the sign when rotated about the *x*-axis by 180^o. Consequently, when spherical symmetry is present, the resultant vector field must be of the form $\mathbf{D(r)} = D_R(R)\,\mathbf{a}_R$.

When vectors **A** and **B** are specified simultaneously at a point in spherical coordinates, the component form of the vectors is given, according to Eq. (1-72), as

$$\mathbf{A} = A_R\,\mathbf{a}_R + A_\theta\,\mathbf{a}_\theta + A_\phi\,\mathbf{a}_\phi \tag{1-73a}$$

$$\mathbf{B} = B_R\,\mathbf{a}_R + B_\theta\,\mathbf{a}_\theta + B_\phi\,\mathbf{a}_\phi \tag{1-73b}$$

where $\mathbf{a}_\rho$, $\mathbf{a}_\phi$ and $\mathbf{a}_z$ are the base vectors at the given point

The vector addition of **A** and **B** in spherical coordinates is obtained from Eq. (1-73) as

$$\boxed{\mathbf{A} + \mathbf{B} = (A_R + B_R)\,\mathbf{a}_R + (A_\theta + B_\theta)\,\mathbf{a}_\theta + (A_\phi + B_\phi)\,\mathbf{a}_\phi} \tag{1-74}$$

The dot product of **A** and **B** in spherical coordinates is obtained from Eq. (1-73), by applying the distributive law of dot product given in Eq. (1-13b) and the orthonormality relations given in Eq. (1-69), as

$$\boxed{\mathbf{A} \bullet \mathbf{B} = A_R B_R + A_\theta B_\theta + A_\phi B_\phi} \tag{1-75}$$

The cross product of **A** and **B** in spherical coordinates is obtained from Eq. (1-72), by applying the distributive law of cross product given in Eq. (1-22c) and the cyclic relations given in Eq. (1-70), and written in determinant form as

$$\boxed{\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{a}_R & \mathbf{a}_\theta & \mathbf{a}_\phi \\ A_R & A_\theta & A_\phi \\ B_R & B_\theta & B_\phi \end{vmatrix}} \tag{1-76}$$

It is should be noted that the rules expressed by Eqs. (1-74)-(1-76) are based on the assumption that **A** and **B** are given at the same point in spherical coordinates. Otherwise, one is moved to the point of the other, and expanded by the base vectors at that point, before a vector operation on the two vectors can be conducted.

The differential length vector $d\mathbf{l}$ in spherical coordinates is defined at a point $p_1 : (R_1, \theta_1, \phi_1)$ to represent the directed distance from the given point to a nearby point $p_2 :(R_1 + dR, \theta_1 + d\theta, \phi_1 + d\phi)$. In view of this, $d\mathbf{l}$ is expanded by the base vectors at point p_1. It is apparent from Fig. 1.27 that the projections of $d\mathbf{l}$ onto the base vectors $\mathbf{a}_R$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ are equal to the differential lengths dR, $R_1 d\theta$, and $R_1 \sin\theta_1 d\phi$, respectively. Here we note that the differential coordinates dR, $d\theta$, and $d\phi$ are converted to the differential lengths by the multiplicative factors 1, R_1, and $R_1 \sin\theta_1$, respectively, which are called the metric coefficients. The differential length vector in spherical coordinates is in general expressed as

$$\boxed{d\mathbf{l} = dR\,\mathbf{a}_R + R\,d\theta\,\mathbf{a}_\theta + R\sin\theta\,d\phi\,\mathbf{a}_\phi} \tag{1-77}$$

In spherical coordinates, $d\mathbf{l}$ is a function of position through R, $\mathbf{a}_R$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$. It is important to remember that $d\mathbf{l}$ is always expressed as Eq. (1-77) in spherical coordinates, no matter what the relation between dR, $d\theta$, and $d\phi$ is, only if the differential coordinates are nonzero.

Two end points of $d\mathbf{l}$ are linked with a total of six surfaces of constant coordinate. Those surfaces define a differential volume dv in spherical coordinates as shown in Fig. 1.27. In view of an infinitesimally small $|d\mathbf{l}|$, the differential volume may be considered to a rectangular parallelepiped of sides dR, $R_1 d\theta$, and $R_1 \sin\theta_1 d\phi$, having a volume of $R_1^2 \sin\theta_1 dR d\theta d\phi$. The differential volume in spherical coordinates is in general defined as

$$\boxed{dv = R^2 \sin\theta\, dR d\theta d\phi} \qquad [\mathrm{m}^3] \tag{1-78}$$

Although dv depends on the radial distance R and polar angle θ, we always define the differential volume according to Eq. (1-78) in spherical coordinates. The differential volume provides a way of subdividing a volume in spherical coordinates into many differential elements of volume, and vice versa.

The bounding surface of the differential volume comprises six infinitesimal surfaces as shown in Fig. 1.27. Each surface can be conveniently represented by a differential area vector $d\mathbf{s}$: the magnitude is equal to the infinitesimal area of the surface, while the unit vector is normal to the surface, pointing outward from the enclosed volume. In view of infinitesimally small $|d\mathbf{l}|$, we can assume each surface of dv to be a rectangle. The differential area vectors for the six surfaces of dv shown in Fig. 1.27 are expressed as follows:

$$d\mathbf{s} = R_1^2 \sin\theta_1\, d\theta\, d\phi\, \mathbf{a}_R \qquad \text{(face ①)} \tag{1-79a}$$
$$d\mathbf{s} = R_1 \sin\theta_1\, dR\, d\phi\, \mathbf{a}_\theta \qquad \text{(face ②)} \tag{1-79b}$$
$$d\mathbf{s} = -R_1\, dR\, d\theta\, \mathbf{a}_\phi \qquad \text{(face ③)} \tag{1-79c}$$
$$d\mathbf{s} = -R_1^2 \sin\theta_1\, d\theta\, d\phi\, \mathbf{a}_R \tag{1-79d}$$
$$d\mathbf{s} = -R_1 \sin\theta_1\, dR\, d\phi\, \mathbf{a}_\theta \tag{1-79e}$$
$$d\mathbf{s} = R_1\, dR\, d\theta\, \mathbf{a}_\phi \tag{1-79f}$$

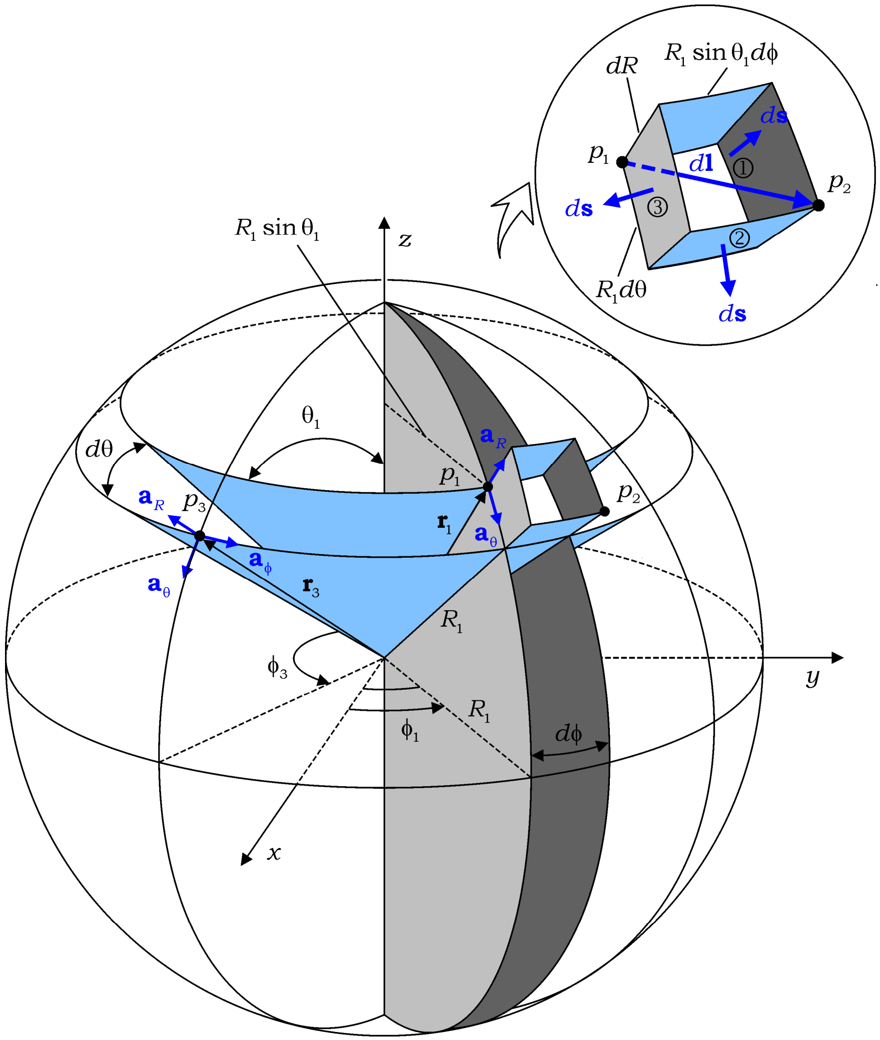


Fig. 1.27 Differential length vector $d\mathbf{l}$, differential area vector $d\mathbf{s}$ and differential volume dv in spherical coordinates.

where the last three are for those hidden from sight behind the front faces. According to Eq. (1-79), we can subdivide a surface of constant coordinate in spherical coordinates into many differential elements of surface, and vice versa. Note that Eq. (1-79) cannot be used for other surfaces such as $z = 1$ plane or a spherical surface with the center at a point other than the origin.

Example 1-9

With reference to a point p_1 in spherical coordinates as shown in Fig. 1.26, determine the projections of (a) $\mathbf{a}_R$ (b) $\mathbf{a}_\theta$, and (c) $\mathbf{a}_\phi$ onto three unit vectors $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$.

Solution

Projection of $\mathbf{a}_R$ onto the $z = 0$ plane : $\sin\theta_1$ (1-80)

By using the two-step projection, we obtain

projection of $\mathbf{a}_R$ onto $\mathbf{a}_x$: $\sin\theta_1 \cos\phi_1$ (1-81a)

projection of $\mathbf{a}_R$ onto $\mathbf{a}_y$: $\sin\theta_1 \sin\phi_1$ (1-81b)

Projection of $\mathbf{a}_R$ onto $\mathbf{a}_z$: $\cos\theta_1$ (1-81c)

Combining the results given in Eq. (1-81), we have

$$\boxed{\mathbf{a}_R = \sin\theta_1 \cos\phi_1\, \mathbf{a}_x + \sin\theta_1 \sin\phi_1\, \mathbf{a}_y + \cos\theta_1\, \mathbf{a}_z} \tag{1-82}$$

Projection of $\mathbf{a}_\theta$ onto the $z = 0$ plane : $\cos\theta_1$ (1-83)

By using the two-step projection, we obtain

projection of $\mathbf{a}_\theta$ onto $\mathbf{a}_x$: $\cos\theta_1 \cos\phi_1$ (1-84a)

projection of $\mathbf{a}_\theta$ onto $\mathbf{a}_y$: $\cos\theta_1 \sin\phi_1$ (1-84b)

projection of $\mathbf{a}_\theta$ onto $\mathbf{a}_z$: $-\sin\theta_1$ (1-84c)

Combining the results given in Eq. (1-84), we have

$$\boxed{\mathbf{a}_\theta = \cos\theta_1 \cos\phi_1\, \mathbf{a}_x + \cos\theta_1 \sin\phi_1\, \mathbf{a}_y - \sin\theta_1\, \mathbf{a}_z} \tag{1-85}$$

Next, we obtain

projection of $\mathbf{a}_\phi$ onto $\mathbf{a}_x$: $-\sin\phi_1$ (1-86a)

projection of $\mathbf{a}_\phi$ onto $\mathbf{a}_y$: $\cos\phi_1$ (1-86b)

projection of $\mathbf{a}_\phi$ onto $\mathbf{a}_z$: 0 (1-86c)

Combining the results given in Eq. (1-86), we have

$$\boxed{\mathbf{a}_\phi = -\sin\phi_1\, \mathbf{a}_x + \cos\phi_1\, \mathbf{a}_y} \tag{1-87}$$

The results given in Eqs. (1-82), (1-85), and (1-87) are useful for the coordinate transformation discussed in the following section.

Example 1-10
Given two vector fields $\mathbf{A}(\mathbf{r}) = (\cos\phi)\mathbf{a}_z$ and $\mathbf{B}(\mathbf{r}) = R^2\,\mathbf{a}_R$ in mixed coordinates, find, on a spherical surface of radius $R = 2$,
(a) $\mathbf{A} \bullet \mathbf{B}$.
(b) $\mathbf{A} \times \mathbf{B}$.

Solution
On the sphere, $(R, \theta, \phi) = (2, \theta, \phi)$, the vector fields become

$$\mathbf{A}(2, \theta, \phi) = (\cos\phi)\mathbf{a}_z$$
$$\mathbf{B}(2, \theta, \phi) = 4\,\mathbf{a}_R$$

(a) $\mathbf{A} \bullet \mathbf{B} = 4\cos\phi\left(\mathbf{a}_z \bullet \mathbf{a}_R\right) = 4\cos\phi\cos\theta$

(b) $\mathbf{A} \times \mathbf{B} = 4\cos\phi\left(\mathbf{a}_z \times \mathbf{a}_R\right) = 4\cos\phi\sin\theta\,\mathbf{a}_\phi$

Note that the dot and cross product are conducted in mixed coordinates.

Exercise 1.20
At a point on a sphere of radius $R = 3$, centered at the origin, find
(a) $d\mathbf{s}$, and (b) $d\mathbf{l}$ in a plane tangential to the sphere.
Ans. (a) $d\mathbf{s} = (9\sin\theta)d\theta d\phi\,\mathbf{a}_R$, (b) $d\mathbf{l} = 3d\theta\,\mathbf{a}_\theta + 3\sin\theta\,d\phi\,\mathbf{a}_\phi$.

Exercise 1.21
Repeat part (b) of **Exercise 1.20** in Cartesian coordinates.
Ans. $d\mathbf{l} = dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z$.

Exercise 1.22
Do the following vector fields have spherical symmetry? (a) $\mathbf{U}(\mathbf{r}) = (1/R^2)\mathbf{a}_R$, (b) $\mathbf{V}(\mathbf{r}) = (\cos\theta)\mathbf{a}_R$, (c) $\mathbf{W}(\mathbf{r}) = \mathbf{a}_\phi \times \mathbf{a}_\theta$, and (d) $\mathbf{X}(\mathbf{r}) = (\mathbf{r} \bullet \mathbf{a}_z)\mathbf{a}_R$.
Ans. (a) Yes, (b) No, (c) Yes, (d) No.

Review Questions with Hints

RQ 1.9 How do you characterize an orthogonal and a right-handed coordinate system? [Eqs.(1-28)(1-29)]
RQ 1.10 State the orthonormality and the cyclic relations. [Eqs.(1-28)(1-29)]
RQ 1.11 Explain how the definition of $\mathbf{A} \bullet \mathbf{B}$ given in Eq. (1-9) is expressed in component form as in Eq. (1-35). [Eqs.(1-32)(1-13b)(1-28)]
RQ 1.12 Explain how the definition of $\mathbf{A} \times \mathbf{B}$ given in Eq. (1-21) is expressed in component form as in Eq. (1-37). [Eqs.(1-32)(1-22c)(1-29)]
RQ 1.13 Define $d\mathbf{l}$ in the three coordinate systems. [Eqs.(1-39)(1-55)(1-77)]
RQ 1.14 Define dv in the three coordinate systems. [Eqs.(1-40)(1-56)(1-78)]
RQ 1.15 In which direction is $d\mathbf{s}$ directed on a closed surface? [Fig.1.17]
RQ 1.16 Can you still use Eqs. (1-54) and (1-76) even if **A** and **B** are not at the same location? [Eqs.(1-48)(1-70)]

1.4 Coordinate Transformation

The location of a point in space is independent of the coordinate system that is chosen to specify it, although different coordinate systems have different ways of describing the point, as was discussed in Section 1-3. Since one coordinate system may be advantageous over the others in solving a problem, we should be able to transform the coordinates of the point from one coordinates to another. Such an operation is called a coordinate transformation. The same is true for a vector in that the magnitude and direction of a vector are independent of the coordinate system, although different definitions of base vectors lead to different expressions for the vector components in different coordinate systems. We can transform the vector components according to the procedure called the coordinate transformation of the component of a vector.

1.4.1 Cartesian-Cartesian Transformation

We first examine the coordinate transformation between two Cartesian coordinate systems. Fig. 1.28 shows that the primed system is rotated about the z-axis by an angle φ relative to the unprimed system. Let us consider a point p_1 that is specified by the coordinates (x_1, y_1, z_1) in the unprimed system and by the coordinates (x_1', y_1', z_1') in the primed system. At the this point, the base vectors are given by either $(\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z)$ or $(\mathbf{a}_{x'}, \mathbf{a}_{y'}, \mathbf{a}_{z'})$, depending on the system. If vector $\mathbf{A}$ is defined at point p_1, it is generally expressed as

$$\mathbf{A} = A_x\mathbf{a}_x + A_y\mathbf{a}_y + A_z\mathbf{a}_z \tag{1-88a}$$

$$\mathbf{A} = A_{x'}\mathbf{a}_{x'} + A_{y'}\mathbf{a}_{y'} + A_{z'}\mathbf{a}_{z'} \equiv \mathbf{A}' \tag{1-88b}$$

Here, A_x, A_y, and A_z are the scalar components of $\mathbf{A}$ in unprimed coordinates, while $A_{x'}$, $A_{y'}$, and $A_{z'}$ are those in primed coordinates.

The transformation of the components of $\mathbf{A}$ is simply done by projecting the vector onto the base vectors for a new coordinate system. The scalar component of $\mathbf{A}$ in the x'-direction is obtained from the projection of $\mathbf{A}$ onto $\mathbf{a}_{x'}$ as

$$\begin{aligned} \mathbf{A}\bullet\mathbf{a}_{x'} &= A_{x'} \\ &= (A_x\mathbf{a}_x + A_y\mathbf{a}_y + A_z\mathbf{a}_z)\bullet\mathbf{a}_{x'} = A_x\cos\varphi + A_y\sin\varphi \end{aligned} \tag{1-89}$$

where we used $\mathbf{a}_x\bullet\mathbf{a}_{x'} = \cos\varphi$, $\mathbf{a}_y\bullet\mathbf{a}_{x'} = \sin\varphi$, and $\mathbf{a}_z\bullet\mathbf{a}_{x'} = 0$, which are evident from Fig. 1.28. Following the same procedure as for $A_{x'}$, we obtain the scalar components $A_{y'}$ and $A_{z'}$ from $\mathbf{A}\bullet\mathbf{a}_{y'}$ and $\mathbf{A}\bullet\mathbf{a}_{z'}$ as well. Transformation of the components of **A** from the unprimed to the primed system is expressed in matrix form as

$$\begin{bmatrix} A_{x'} \\ A_{y'} \\ A_{z'} \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} \tag{1-90}$$

where φ is the rotation angle of the primed system with respect to the unprimed system, which is measured in the counterclockwise direction. The 3X3 matrix in Eq. (1-90) is called the transformation matrix, denoted by **T**. Thus Eq. (1-90) can be simply written as $\mathbf{A}' = \mathbf{TA}$.

The transformation matrix **T** in Eq. (1-90) is independent of the location of **A**. This implies that **T** may be used for any vector, including the position vector. We acknowledge that the scalar components of position vector are equal to the coordinates of a given point such that $\mathbf{r}_1 = x_1\,\mathbf{a}_x + y_1\,\mathbf{a}_y + z_1\,\mathbf{a}_z$ in the unprimed system, and $\mathbf{r}_1 = x_1'\,\mathbf{a}_{x'} + y_1'\,\mathbf{a}_{y'} + z_1'\,\mathbf{a}_{z'}$ in the primed system. Substitution of $\mathbf{r}_1$ for **A** in Eq. (1-90) leads to the transformation of the coordinates of a point from the unprimed to the primed system:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1-91}$$

Here, subscript 1 is omitted for generalization.

The direction of the transformations expressed by Eqs. (1-90) and (1-91) can be reversed by replacing φ with −φ in the transformation matrix **T**, meaning the transformation from the primed to the unprimed system.

If vector **A** forms a vector field, its scalar components are functions of position such as $A_x(x,y,z)$, $A_y(x,y,z)$, and $A_z(x,y,z)$. In this case, the transformation for **A** is done following these steps: (1) Transformation of the components of **A** according to Eq. (1-90). (2) Transformation of the coordinates of the location of **A** according to Eq. (1-91).

Although the same **T** is used for the transformations of the coordinates of a point and the components of a vector in the Cartesian-Cartesian transformation as shown in Eqs. (1-90) and (1-91), this is not true in the general case, which will be discussed in the next section.

Example 1-11

With reference to Fig. 1.28, find the coordinate transformation of a vector field $\mathbf{A} = 2x\,\mathbf{a}_x$ from the unprimed to the primed system.

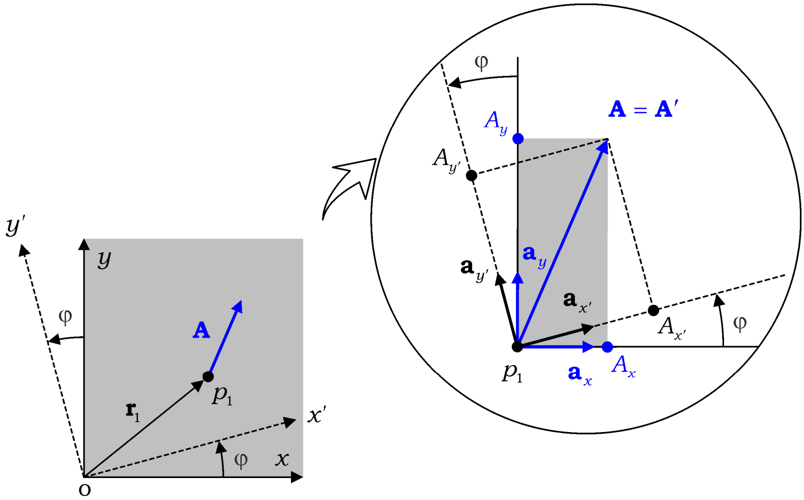


Fig. 1.28 Two Cartesian coordinate systems rotated to each other about the z-axis.

Solution

Coordinate transform of the components of **A** is, from Eq. (1-90),

$$\begin{bmatrix} A_{x'} \\ A_{y'} \\ A_{z'} \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2x \\ 0 \\ 0 \end{bmatrix} \tag{1-92}$$

Rewriting Eq. (1-92) we have

$$\begin{aligned} \mathbf{A}' &= A_{x'}\,\mathbf{a}_{x'} + A_{y'}\,\mathbf{a}_{y'} + A_z\,\mathbf{a}_z \\ &= 2x\cos\varphi\,\mathbf{a}_{x'} - 2x\sin\varphi\,\mathbf{a}_{y'} \end{aligned} \tag{1-93}$$

From Eq. (1-91) we obtain

$$x = x'\cos\varphi - y'\sin\varphi \tag{1-94}$$

By inserting Eq. (1-94) into Eq. (1-93) we obtain

$$\mathbf{A} = 2(x'\cos\varphi - y'\sin\varphi)\cos\varphi\,\mathbf{a}_{x'} - 2(x'\cos\varphi - y'\sin\varphi)\sin\varphi\,\mathbf{a}_{y'}$$

Exercise 1.23

Given $\mathbf{A} = 4\,\mathbf{a}_x + 2\,\mathbf{a}_y + 3\,\mathbf{a}_z$ in unprimed coordinates. Express the vector at point p:(2, 1, 3) in terms of the primed coordinates that is rotated by $\varphi = 30^o$.

Ans. $\mathbf{A} = (2\sqrt{3} + 1)\,\mathbf{a}_{x'} + (-2 + \sqrt{3})\,\mathbf{a}_{y'} + 3\,\mathbf{a}_{z'}$.

1.4.2 Cylindrical-Cartesian Transformation

Let us consider Fig. 1.21, in which vector **A** is given at a point $p_1{:}(\rho_1,\ \phi_1,\ z_1)$ in cylindrical coordinates. At point p_1, the base vectors $(\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z)$ are rotated about $\mathbf{a}_z$ by an angle of $\varphi = -\phi_1$, relative to the base vectors $(\mathbf{a}_\rho, \mathbf{a}_\phi, \mathbf{a}_z)$. Comparison of Fig. 1.21 and Fig. 1.28 reveals that we can conduct the transformation of the components of **A** from cylindrical to Cartesian coordinates simply by substituting $\varphi = -\phi_1$ into the transformation matrix in Eq. (1-90). Transformation of the vector components from cylindrical to Cartesian coordinates is done as follows:

$$\boxed{\begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} = \begin{bmatrix} \cos(-\phi_1) & \sin(-\phi_1) & 0 \\ -\sin(-\phi_1) & \cos(-\phi_1) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_\rho \\ A_\phi \\ A_z \end{bmatrix}} \tag{1-95}$$

It is important to note that ϕ_1 in Eq. (1-95) is the ϕ-coordinate of the location of the vector **A**. The direction of the transformation may be reversed simply by replacing $-\phi_1$ with ϕ_1 in the transformation matrix. Transformation of the vector components from Cartesian to cylindrical coordinates is conducted as follows:

$$\boxed{\begin{bmatrix} A_\rho \\ A_\phi \\ A_z \end{bmatrix} = \begin{bmatrix} \cos(\phi_1) & \sin(\phi_1) & 0 \\ -\sin(\phi_1) & \cos(\phi_1) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix}} \tag{1-96}$$

Again, the angle ϕ_1 is the ϕ-coordinate of the location of the vector **A**.

We may use Eq. (1-95) for transforming the coordinates of a point from cylindrical to Cartesian system. However, Eq. (1-96) cannot be used for transforming the coordinates of the point. To avoid confusion, we do not use the transformation matrix for transforming the coordinates of a point. Alternatively, we obtain the relation between the cylindrical and Cartesian coordinates of point p_1 by applying Pythagorean theorem and trigonometry to Fig. 1.21. The relations between the cylindrical and Cartesian coordinates of a point are given as

$$\boxed{\begin{aligned} x &= \rho\cos\phi \\ y &= \rho\sin\phi \\ z &= z \end{aligned}} \tag{1-97}$$

and

$$\boxed{\begin{aligned} \rho &= \sqrt{x^2 + y^2} \\ \phi &= \tan^{-1}(y / x) \\ z &= z \end{aligned}} \tag{1-98}$$

Subscript 1 is omitted in the above equations for generalization.

Example 1-12

Express the following vectors in Cartesian coordinates:

(a) $\mathbf{A} = 2\,\mathbf{a}_\rho + 3\,\mathbf{a}_\phi + \sqrt{3}\,\mathbf{a}_z$ at $(\rho, \phi, z) = (4, 60^0, 5)$,

(b) $\mathbf{B} = 2\,\mathbf{a}_\rho + 3\,\mathbf{a}_\phi + \sqrt{3}\,\mathbf{a}_z$ at $(x, y, z) = (\sqrt{2}, 2, 1)$, and

(c) Vector $\mathbf{C}$ drawn from the origin to point $(\rho, \phi, z) = (3, 45^0, 4)$.

Solution

(a) From Eq. (1-95) we obtain

$$\begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} = \begin{bmatrix} \cos(-60^o) & \sin(-60^o) & 0 \\ -\sin(-60^o) & \cos(-60^o) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ \sqrt{3} \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ \sqrt{3} \end{bmatrix}$$

Thus,

$$A_x = \frac{2 - 3\sqrt{3}}{2}, \quad A_y = \frac{2\sqrt{3} + 3}{2}, \text{ and } A_z = \sqrt{3}$$

The answer is

$$\mathbf{A} = \left(1 - \frac{3}{2}\sqrt{3}\right)\mathbf{a}_x + \left(\sqrt{3} + \frac{3}{2}\right)\mathbf{a}_y + \sqrt{3}\mathbf{a}_z\,.$$

(b) Substituting $(x, y, z) = (\sqrt{2}, 2, 1)$ into Eq. (1-98) we obtain

$$\phi_1 = \tan^{-1}\left(2/\sqrt{2}\right) = 54.8^o$$

Thus we have

$$\sin\phi_1 = 2/\sqrt{6} \text{ and } \cos\phi_1 = 1/\sqrt{3} \qquad (1\text{-}99)$$

Inserting $\mathbf{B}$ and Eq. (1-99) into Eq. (1-95) we obtain

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \begin{bmatrix} \cos(-\phi_1) & \sin(-\phi_1) & 0 \\ -\sin(-\phi_1) & \cos(-\phi_1) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ \sqrt{3} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 2/\sqrt{6} & 1/\sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ \sqrt{3} \end{bmatrix}$$

Thus,

$$B_x = \frac{2}{\sqrt{3}} - \sqrt{6}, \quad B_y = \frac{4}{\sqrt{6}} + \sqrt{3}, \text{ and } B_z = \sqrt{3}$$

The answer is

$$\mathbf{B} = \left(\frac{2}{\sqrt{3}} - \sqrt{6}\right)\mathbf{a}_x + \left(\frac{4}{\sqrt{6}} + \sqrt{3}\right)\mathbf{a}_y + \sqrt{3}\mathbf{a}_z\,.$$

(c) **C** is a position vector.

Substituting $(\rho, \phi, z) = (3, 45^0, 4)$ into Eq. (1-97), we obtain

$$x = 3\cos 45^o = 3/\sqrt{2}$$
$$y = 3\sin 45^o = 3/\sqrt{2}$$
$$z = 4$$

The answer is

$$\mathbf{C} = \frac{3}{\sqrt{2}}\mathbf{a}_x + \frac{3}{\sqrt{2}}\mathbf{a}_y + 4\,\mathbf{a}_z\,.$$

Exercise 1.24
Show that the transformation matrices in Eqs. (1-95) and (1-96) are the inverse matrix of the other.

Exercise 1.25
Starting from $d\mathbf{l} = dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z$ and applying Eqs. (1-96) and (1-97), find an expression for $d\mathbf{l}$ in cylindrical coordinates.
Ans. $d\mathbf{l} = d\rho\,\mathbf{a}_\rho + \rho d\phi\,\mathbf{a}_\phi + dz\,\mathbf{a}_z$.

1.4.3 Spherical-Cartesian Transformation

When vector **A** is specified at a point $p_1{:}(R_1, \theta_1, \phi_1)$ in spherical coordinates as shown in Fig. 1.26, it can be expanded in component form either in Cartesian coordinates or in spherical coordinates as

$$\mathbf{A} = A_x\mathbf{a}_x + A_y\mathbf{a}_y + A_z\mathbf{a}_z \tag{1-100a}$$

$$\mathbf{A} = A_R\,\mathbf{a}_R + A_\theta\,\mathbf{a}_\theta + A_\phi\,\mathbf{a}_\phi \equiv \mathbf{A}' \tag{1-100b}$$

Transformation of the components of $\mathbf{A}'$ into Cartesian coordinates can be done by projecting $\mathbf{A}'$ in the directions of $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$. The dot product of $\mathbf{A}'$ and $\mathbf{a}_x$ gives A_x as follows:

$$\begin{aligned}\mathbf{A}'\bullet\mathbf{a}_x &= A_x \\ &= \left(A_R\mathbf{a}_R + A_\theta\mathbf{a}_\theta + A_\phi\mathbf{a}_\phi\right)\bullet\mathbf{a}_x \\ &= A_R\sin\theta_1\cos\phi_1 + A_\theta\cos\theta_1\cos\phi_1 - A_\phi\sin\phi_1\end{aligned}$$

where we used $\mathbf{a}_R\bullet\mathbf{a}_x = \sin\theta_1\cos\phi_1$, $\mathbf{a}_\theta\bullet\mathbf{a}_x = \cos\theta_1\cos\phi_1$, and $\mathbf{a}_\phi\bullet\mathbf{a}_x = -\sin\phi_1$ as given in Eqs. (1-82), (1-85) and (1-87). Following the same procedure as for A_x, we can obtain A_y and A_z from $\mathbf{A}'\bullet\mathbf{a}_y$ and $\mathbf{A}'\bullet\mathbf{a}_z$.

Transformation of the components of a vector **A** from spherical to Cartesian coordinates is expressed in matrix form as

$$\begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} = \begin{bmatrix} \sin\theta_1 \cos\phi_1 & \cos\theta_1 \cos\phi_1 & -\sin\phi_1 \\ \sin\theta_1 \sin\phi_1 & \cos\theta_1 \sin\phi_1 & \cos\phi_1 \\ \cos\theta_1 & -\sin\theta_1 & 0 \end{bmatrix} \begin{bmatrix} A_R \\ A_\theta \\ A_\phi \end{bmatrix} \tag{1-101}$$

It should be noted that the transformation matrix depends on the angles θ_1 and ϕ_1, which are the spherical coordinates of the location of **A**.

The same procedure used for Eq. (1-101) can be followed to obtain the transformation of the components of **A** from Cartesian to spherical coordinates given as follows:

$$\begin{bmatrix} A_R \\ A_\theta \\ A_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta_1 \cos\phi_1 & \sin\theta_1 \sin\phi_1 & \cos\theta_1 \\ \cos\theta_1 \cos\phi_1 & \cos\theta_1 \sin\phi_1 & -\sin\theta_1 \\ -\sin\phi_1 & \cos\phi_1 & 0 \end{bmatrix} \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} \tag{1-102}$$

Equation (1-102) is based on the projections of **A** onto $\mathbf{a}_R$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ at point p_1, and thus the transformation matrix obviously depends on the θ- and ϕ-coordinates of p_1. It can be shown that the transformation matrix in Eq. (1-102) is the inverse matrix of that in Eq. (1-101), and vice versa.

To avoid confusion, we do not use the transformation matrices given in Eq. (1-101) and Eq. (1-102) for transforming the coordinates of a point. Alternatively, we obtain the relation between the spherical and Cartesian coordinates of point p_1 by applying Pythagorean theorem and trigonometry to Fig. 1.26. The relations between the spherical and Cartesian coordinates of a point is given as

$$\begin{aligned} x &= R\sin\theta\cos\phi \\ y &= R\sin\theta\sin\phi \\ z &= R\cos\theta \end{aligned} \tag{1-103}$$

and

$$\begin{aligned} R &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \tan^{-1}\left(\sqrt{x^2 + y^2} / z\right) \\ \phi &= \tan^{-1}(y / x) \end{aligned} \tag{1-104}$$

Subscript 1 is omitted in the above equations for generalization.

Example 1-13

Verify, by coordinate transformation,

(a) $\dfrac{\partial \mathbf{a}_R}{\partial \theta} = \mathbf{a}_\theta$

(b) $\dfrac{\partial \mathbf{a}_R}{\partial \phi} = \sin\theta\, \mathbf{a}_\phi$

Solution

Transformation of $\mathbf{a}_R$ into Cartesian coordinates is, from Eq. (1-101),

$$\begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \cos\theta\cos\phi & -\sin\phi \\ \sin\theta\sin\phi & \cos\theta\sin\phi & \cos\phi \\ \cos\theta & -\sin\theta & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Rewriting the above equation we get

$$\begin{aligned} \mathbf{a}_R &= A_x\,\mathbf{a}_x + A_y\,\mathbf{a}_y + A_z\,\mathbf{a}_z \\ &= \sin\theta\cos\phi\,\mathbf{a}_x + \sin\theta\sin\phi\,\mathbf{a}_y + \cos\theta\,\mathbf{a}_z \end{aligned} \tag{1-105}$$

Taking the derivatives of both sides of Eq. (1-105) with respect to θ, we get

$$\frac{\partial \mathbf{a}_R}{\partial \theta} = \cos\theta\cos\phi\,\mathbf{a}_x + \cos\theta\sin\phi\,\mathbf{a}_y - \sin\theta\,\mathbf{a}_z \tag{1-106}$$

Taking the derivatives of both sides of Eq. (1-105) with respect to ϕ, we get

$$\frac{\partial \mathbf{a}_R}{\partial \phi} = -\sin\theta\sin\phi\,\mathbf{a}_x + \sin\theta\cos\phi\,\mathbf{a}_y\,. \tag{1-107}$$

(a) Transforming Eq. (1-106) back into spherical coordinates, by using Eq. (1-102), we have

$$\begin{bmatrix} A_R \\ A_\theta \\ A_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix} \begin{bmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{bmatrix}$$

Rearranging the above equation we get

$A_R = 0$, $A_\theta = 1$, and $A_\phi = 0$

Thus,

$$\begin{aligned} \frac{\partial \mathbf{a}_R}{\partial \theta} &= A_R\,\mathbf{a}_R + A_\theta\,\mathbf{a}_\theta + A_\phi\,\mathbf{a}_\phi \\ &= \mathbf{a}_\theta . \end{aligned}$$

(b) Transforming Eq. (1-107) back into spherical coordinates, by using Eq. (1-102), we have

$$\begin{bmatrix} A_R \\ A_\theta \\ A_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix} \begin{bmatrix} -\sin\theta\sin\phi \\ \sin\theta\cos\phi \\ 0 \end{bmatrix}$$

Rearranging the above equation we get

$A_R = 0$, $A_\theta = 0$, and $A_\phi = \sin\theta\sin^2\phi + \sin\theta\cos^2\phi = \sin\theta$

Thus,

$$\frac{\partial \mathbf{a}_R}{\partial\phi} = A_R\,\mathbf{a}_R + A_\theta\,\mathbf{a}_\theta + A_\phi\,\mathbf{a}_\phi$$
$$= \sin\theta\,\mathbf{a}_\phi .$$

Exercise 1.26
Transform a vector field $\mathbf{A(r)} = \mathbf{a}_R$ into Cartesian coordinates.
Ans. $\mathbf{A(r)} = (x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z)/\sqrt{x^2+y^2+z^2}$.

Exercise 1.27
Show that the transformation matrix is a unitary matrix whose determinant has absolute value one.
Ans. $|\det\mathbf{T}| = 1$.

Review Questions with Hints

RQ 1.17 Does **T** for Cartesian-Cartesian transformation depend on the location of a given vector? [Eq.(1-90)Fig.1.28]
RQ 1.18 Write the relations between cylindrical and Cartesian coordinates of a point. [Eqs.(1-97)(1-98)]
RQ 1.19 State the transformation of the components of a vector between cylindrical and Cartesian coordinates. [Eqs.(1-95)(1-96)]
RQ 1.20 Write the relations between spherical and Cartesian coordinates of a point. [Eqs.(1-103)(1-104)]
RQ 1.21 State the transformation of the components of a vector between spherical and Cartesian coordinates. [Eqs.(1-101)(1-102)]
RQ 1.22 What are the significances of the angles appearing in the transformation matrices in Eqs. (1-90), (1-95) and (1-101)? [Figs.1.28,1.21,1.26]

Problems

1-1 Verify graphically the associative law of vector addition, as given in Eq. (1-5a), for three vectors shown in Fig. 1.29.

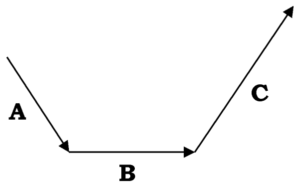


Fig. 1.29 Three vectors (Problem 1-1).

1-2 Two vectors **A** and **B** form a parallelogram in three-dimensional space. Verify that the projection of the parallelogram onto the *xy*-plane is also a parallelogram.

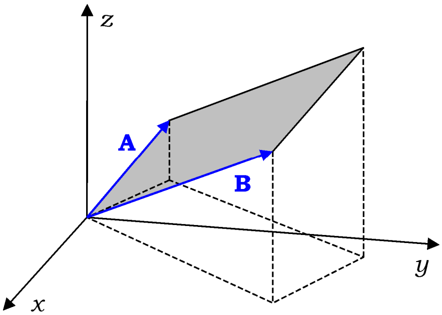


Fig. 1.30 Projection of the parallelogram formed by **A** and **B**(Problem 1-2).

1-3 With reference to **Example 1-2**, in which the distributive law of dot product was verified for three vectors lying in a common plane, verify the distributive law for three arbitrary vectors by using two-step projection.

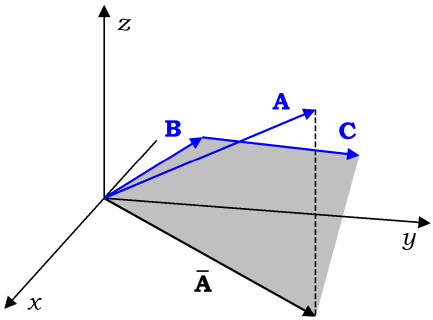


Fig. 1.31 Distributive law of dot product for three arbitrary vectors(Problem 1-3).

1-4 With reference to **Example 1-3**, in which the distributive law of cross product was verified for three vectors(**A** is normal to the plane formed by **B** and **C**), verify the distributive law for three arbitrary vectors. [Hint: $\mathbf{A}\times\mathbf{B} = \mathbf{A}\times\bar{\mathbf{B}}$.]

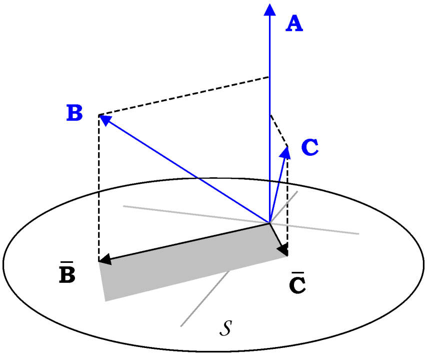


Fig. 1.32 Distributive law of cross product for three arbitrary vectors(Problem 1-4).

1-5 Two vectors **A** and **B** form a parallelogram.
(a) Find expressions for two diagonals in terms of **A** and **B**.
(b) Under what condition are the diagonals perpendicular to each other?
(c) Two diagonal vectors are displaced such that they meet at the tail. Find the area of the new parallelogram.

1-6 Three vectors **A**, **B**, and $\mathbf{A}-\mathbf{B}$ form a triangle. Show the followings:
(a) Parallelograms formed by any two vectors have the same area.
(b) Law of sines, $\dfrac{A}{\sin\theta_A}=\dfrac{B}{\sin\theta_B}=\dfrac{C}{\sin\theta_C}$.
(Sides *A*, *B*, and *C* subtend angles θ_A, θ_B, and θ_C, respectively.)

1-7 Two vectors $\mathbf{A}=2\mathbf{a}_x+\mathbf{a}_y+3\mathbf{a}_z$ and $\mathbf{B}=-\mathbf{a}_x+2\mathbf{a}_y+\mathbf{a}_z$ are at the same point in Cartesian coordinates. Find
(a) $\mathbf{a}_A$
(b) $\mathbf{B}-\mathbf{A}$
(c) $\mathbf{A}\bullet\mathbf{B}$
(d) $\mathbf{A}\times\mathbf{B}$
(e) θ_{AB}
(f) vector component of **B** in the direction of **A**
(g) vector component of **B** in the direction perpendicular to **A**

1-8 Given a vector $\mathbf{A}=\mathbf{a}_x+2\,\mathbf{a}_y+2\,\mathbf{a}_z$ at a point with position vector $\mathbf{r}=2\,\mathbf{a}_x+3\,\mathbf{a}_y+\sqrt{3}\,\mathbf{a}_z$. Find the angles that (a) **A**, and (b) **r** makes with three Cartesian coordinate axes.

1-9 A straight line passes through two points $p_1\!:(1,1,0)$ and $p_2\!:(3,2,2)$ in Cartesian coordinates. Find
(a) unit vector $\mathbf{a}_n$ along the direction from p_1 to p_2, and
(b) position vector of a point on the line, which is at a distance *t* from p_1.

1-10 Two straight lines intersect at right angles at point p_1 in Cartesian coordinates. The first line passes through a point $p_2{:}(2,1,1)$, while the second one passes through two points $p_3{:}(1,-1,3)$ and $p_4{:}(-1,1,2)$ simultaneously. Find the distance between the two points

(a) p_1 and p_2

(b) p_1 and p_3

1-11 For a plane defined by three points $p_1{:}(2,0,0)$, $p_2{:}(0,1,0)$, and $p_3{:}(0,0,1)$ in Cartesian coordinates, find

(a) unit vector normal to the plane, directed toward the origin, and

(b) perpendicular distance from the origin to the plane.

1-12 A plane intersects three Cartesian coordinate axes at $x = a$, $y = b$, and $z = c$, respectively. Find

(a) unit normal to the plane, and

(b) expression for the plane.

1-13 A plane S passes through the point $p_1 : (3,1,2)$ and is perpendicular to the vector $\mathbf{k} = \mathbf{a}_x + 2\mathbf{a}_y + 4\mathbf{a}_z$ in Cartesian coordinates. Find an expression for the plane.

1-14 Two vectors $\mathbf{A}$ and $\mathbf{B}$ define a parallelogram P in three-dimensional space. Projections of $\mathbf{A}$ and $\mathbf{B}$ onto the xy-plane define another parallelogram P'. Show that the area of P' is given by $\left|\mathbf{a}_z \bullet (\mathbf{A} \times \mathbf{B})\right|$.

1-15 A parallelogram P is formed by two vectors $\mathbf{A} = 3\,\mathbf{a}_x - \mathbf{a}_y + 2\,\mathbf{a}_z$ and $\mathbf{B} = \mathbf{a}_x + 2\,\mathbf{a}_y + \mathbf{a}_z$. Projections of P onto the $x = 0$, $y = 0$, and $z = 0$ planes result in three parallelograms P_1, P_2, and P_3 on their respective planes. Find

(a) unit vector $\mathbf{a}_n$ normal to P,

(b) areas of P, P_1, P_2, and P_3, and

(c) relation between the scalar components of $\mathbf{a}_n$ and the areas of P_1, P_2, and P_3.

1-16 Given two vector fields $\mathbf{A} = 3\,\mathbf{a}_\rho + (4\sin\phi)\,\mathbf{a}_\phi$ and $\mathbf{B} = \rho^2\,\mathbf{a}_\rho + (\rho + z + 2)\,\mathbf{a}_z$ in cylindrical coordinates, find at point $p_1{:}(2, 30^o, 1)$

(a) $\mathbf{A} + \mathbf{B}$

(b) $\mathbf{A} \bullet \mathbf{B}$

(c) $\mathbf{A} \times \mathbf{B}$

1-17 A triangle is defined by three points $p_1{:}(2,0,1)$, $p_2{:}(-1,1,3)$, and $p_3{:}(4,2,-1)$ in Cartesian coordinates. Find the area of the triangle.

1-18 A vector field is given as $\mathbf{H}(\mathbf{r}) = (16/\rho^2)\mathbf{a}_\rho$ in cylindrical coordinates. It specifies two vectors $\mathbf{H}(\mathbf{r}_1) = \mathbf{A}$ and $\mathbf{H}(\mathbf{r}_2) = \mathbf{B}$ at two points p_1:$(2, 30^o, 5)$ and p_2:$(2, 90^o, 5)$, respectively. Find
(a) expressions for **A** and **B** in cylindrical coordinates, and
(b) $\mathbf{A} \times \mathbf{a}_B$ at point p_1, where $\mathbf{a}_B$ is the unit vector of **B**

1-19 Given two points p_1:$(2, 90^o, 5)$ and p_2:$(2, 60^o, 5)$ in cylindrical coordinates, find the expression, in cylindrical coordinates, for
(a) position vectors of p_1 and p_2 (or $\mathbf{r}_1$ and $\mathbf{r}_2$), and
(b) distance vector $\mathcal{R}_{1-2} = \mathbf{r}_1 - \mathbf{r}_2$ expanded by the base vectors at p_1.

1-20 A half cylinder is defined by the surfaces $\rho = 4$, $\phi = 30^o$, $\phi = 210^o$, $z = 0$, and $z = 5$ as shown in Fig. 1.33. Find the differential area vector $d\mathbf{s}$ at the following points: (a) p_1, (b) p_2 with $\rho = 2$, (c) p_3, and (d) p_4.

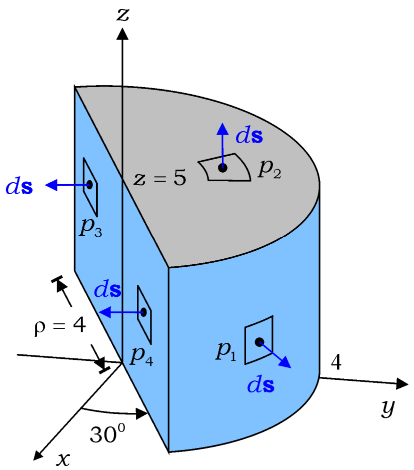


Fig. 1.33 half-cylinder(Problem 1-20).

1-21 With reference to the object shown in Fig. 1.34, which is defined by the surfaces $R = 6$, $\theta = 30^o$, $\phi = 45^o$, and $\phi = 225^o$ in spherical coordinates, find the differential area vectors at the following points:
(a) p_1 with $\theta = 20^o$
(b) p_2 with $R = 4$
(c) p_3 with $R = 4$
(d) p_4 with $R = 5$

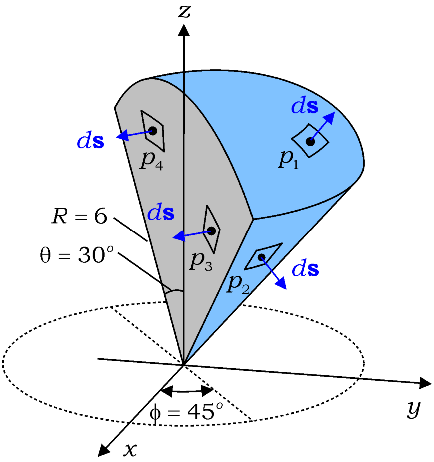


Fig. 1.34 A object in spherical coordinates (Problem 1-21).

1-22 A vector field is defined as $\mathbf{E} = 2R\,\mathbf{a}_R + R^2 \cos\phi\,\mathbf{a}_\theta + 100\sin\theta\cos\phi\,\mathbf{a}_\phi$, in spherical coordinates, inside and outside the object shown in Fig. 1.34. At three points $p_1{:}(6, 20^o, 60^o)$, $p_2{:}(4, 30^o, 50^o)$, and $p_3{:}(4, 20^o, 45^o)$ on the object, find
(a) $\mathbf{E}$, and
(b) vector components of $\mathbf{E}$ parallel and perpendicular to the surface.

1-23 Given a point $p_1{:}(2, \sqrt{3}, 3)$ in Cartesian coordinates, find an expression for p_1 in (a) cylindrical coordinates, and (b) spherical coordinates.

1-24 Find expressions for the unit vector $\mathbf{a}_x$ in terms of spherical coordinates at the following points:
(a) $(x, y, z) = (1, -2, 3)$
(b) $(\rho, \phi, z) = (2, 30^o, \sqrt{2})$
(c) $(R, \theta, \phi) = (2, 45^o, 60^o)$

1-25 Given that $\mathbf{A} = \mathbf{a}_x + \mathbf{a}_y + 0.5\,\mathbf{a}_z$ at point $(2, 2, 1)$ in Cartesian coordinates, transform $\mathbf{A}$ into (a) cylindrical and (b) spherical coordinates.

1-26 Given two points $p_1{:}(1, 1, 3)$ and $p_2{:}(-1, 2, 1)$ in Cartesian coordinates, find a unit vector
(a) at p_1 directed toward p_2 in Cartesian coordinates,
(b) at p_2 directed toward p_1 in Cartesian coordinates,
(c) at p_1 directed toward p_2 in cylindrical coordinates, and
(d) at p_2 directed toward p_1 in cylindrical coordinates.

1-27 In the presence of a vector field $\mathbf{A} = (2z + 10)\,\mathbf{a}_x + x^2\,\mathbf{a}_y - y\,\mathbf{a}_z$, two points are defined as $p_1{:}(-1, \sqrt{3}, 1)$, in Cartesian coordinates, and

p_2:$(4, 60^o, 3)$, in cylindrical coordinates. Find the vector component of $\mathbf{A}$ at p_2 directed toward p_1 in terms of (a) Cartesian and (b) cylindrical coordinates.

1-28 Perform coordinate transformations of the following quantities from Cartesian to spherical coordinates:
(a) unit vector $\mathbf{a}_x$,
(b) point p:(x, y, z),
(c) differential length vector $d\mathbf{l}$, and
(d) position vector $\mathbf{r}$

1-29 A top shown in Fig. 1.35 is defined by a conical surface of half-angle 30^o and a disk of radius 5[cm]. It spins about the z-axis at an angular speed $\omega[\mathrm{rad/s}]$. (a) Find an expression for the linear velocity $\mathbf{v}$ at any point on the surface of the top in terms of spherical coordinates. (b) Find the coordinate transformation of the result in part (a) into Cartesian coordinates.

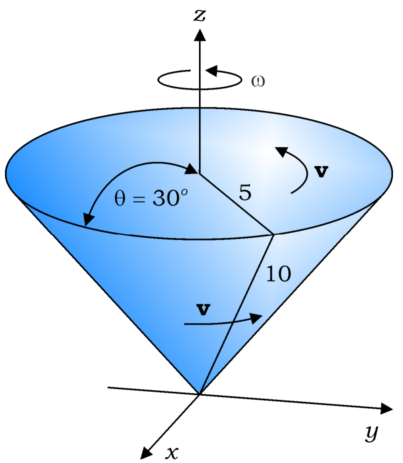


Fig. 1.35 A top spinning at an angular speed ω(Problem 1-29).

1-30 Consider Fig. 1.36, in which $\mathbf{r}$ is the position vector of a point on the conical surface of $\theta = 30^o$ and $\mathbf{r}'$ is the position vector of a point on the z-axis at $z = 4$ from the origin. In the presence of a vector field $\mathbf{A} = \mathbf{r} \times (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'|$, find an expression for $\mathbf{A}$ on the conical surface in terms of spherical coordinates.

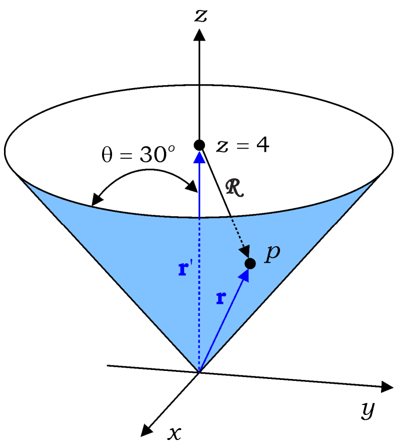


Fig. 1.36 Vector field on a conical surface(Problem 1-30).

Chapter 2
Vector Calculus

In Chapter 1, we discussed the basic concepts of vector algebra developed in the three common coordinate systems. Through the use of the position and base vectors, we could specify points and vectors in three-dimensional space. We reviewed typical symmetries of vector fields that easily reveal themselves in the cylindrical and spherical coordinate systems. We defined differential quantities such as differential length vectors, differential area vectors, and differential volumes in different coordinate systems. We also learned how to transform the coordinates of a point or the components of a vector from one coordinate system to another.

Electromagnetics deals with quantities that are usually distributed in a region of space. If a scalar or a vector quantity is specified at each and every point in a region of space, which is called a scalar or a vector field, and if it represents a real physical quantity, the scalar or vector field should be a smooth function of position so that its derivative is continuous in the given region. This is because the spatial derivative of a field has as much physical significance as the field itself does. For instance, the spatial variation of temperature in a room must be caused by something, and the change of temperature from one point to another in space must be governed by some laws of nature, which, in general, do not allow an abrupt change. Since electromagnetic fields are defined in three-dimensional space, electromagnetic laws are generally expressed in terms of the partial derivatives of the fields.

In this chapter, we discuss the integrals of a vector field along a line and across a surface, in three-dimensional space. We then explore the spatial derivatives of a scalar or a vector field, and define vector operators such as the gradient, divergence, and curl, which are essential for the study of electromagnetics. Spatial derivatives of a field are of course specific to a point in space. Accordingly, electromagnetic laws expressed in differential form are useful for describing local effects, whereas the laws in integral form are useful for describing nonlocal effects, in which the effects are observed at different points than the point of cause. We will learn about divergence and Stokes's theorems, which allow us to convert the differential form of an electromagnetic law into the integral form, and vice versa. We also discuss Helmholtz's theorem.

2.1 Line and Surface Integrals

When a vector field **A**(**r**) is given in a region of space, the line integral of **A** along a path is defined as the integration of the tangential component of **A** along the path. Note that line integral is the standard name whether or not the path of integration is curved. If the initial and terminal points of the path of integration coincide with each other, forming a closed path C, the integration is called the closed line integral of **A** around C, or the circulation of **A** around C. In contrast, the surface integral of **A** across a surface is defined as the integration of the normal component of **A** over the surface. If the surface has no opening, forming a closed surface, the integration is called the closed surface integral of **A**.

2.1.1 Curves

A curve may represent the path of an electric charge moving in space, for instance. In three-dimensional space, we can describe a curve C by the position vector, expressed in terms of a parameter t, as

$$\boxed{\mathbf{r}(t) = x(t)\,\mathbf{a}_x + y(t)\,\mathbf{a}_y + z(t)\,\mathbf{a}_z} \tag{2-1}$$

where t varies from $t = t_i$ to $t = t_t$, which are the initial and terminal points of the curve respectively. Eq. (2-1) is called a parametric representation of C. The sense of increasing t on C defines the *positive direction* of C, or the *direction of travel* on the curve.

If a curve $\mathbf{r}(t)$ is a smooth function of position, there is no abrupt change in the magnitude and direction of the position vector from a point to another in space. In this case the vector function $\mathbf{r}(t)$ is differentiable. The rate of change of $\mathbf{r}(t)$ with respect to t, at $t = t_1$, is expressed as

$$\mathbf{r}'(t)\Big|_{t=t_1} = \frac{d\mathbf{r}}{dt}\bigg|_{t=t_1} = \lim_{\Delta t \to 0} \frac{\mathbf{r}(t_1 + \Delta t) - \mathbf{r}(t_1)}{\Delta t} \tag{2-2}$$

While the denominator is an increment of parameter t, the numerator is a subtraction of the position vectors of the two neighboring points on C. In view of this, we see that the numerator of Eq. (2-2) is a differential length vector, $d\mathbf{r} = d\mathbf{l}$, located at the point with $t = t_1$, directed along the curve C. Therefore, the derivative $\mathbf{r}'(t_1)$ corresponds to a tangent to the curve at the point with $t = t_1$ on C. From Fig. 2.1, in which a straight line L passes through two points p_1 and p_2 with position vectors $\mathbf{r}(t_1)$ and $\mathbf{r}(t_1 + \Delta t)$ respectively, we see that as $\Delta t \to 0$, the line L becomes a tangent to the curve at p_1. Consequently, $\mathbf{r}'(t)$ is recognized as an expression for the tangent to curve C.

Omitting 1 in t_1 for generalization, the derivative $\mathbf{r}'(t)$ is useful for expressing the differential length vector $d\mathbf{l}$:

$$\boxed{d\mathbf{l} = \frac{d\mathbf{r}}{dt}dt = \mathbf{r}'(t)dt} \tag{2-3}$$

Again, $\mathbf{r}$ is the position vector of a point on curve C, t is parameter, and $d\mathbf{l}$ is the differential length vector, tangent to C, directed along the positive direction of C. It should be noted that the magnitude and direction of $d\mathbf{l}$ belong to a point with position vector $\mathbf{r}(t)$ on C, which happens to be the initial point of $d\mathbf{l}$.

A unit vector along the tangent to C is given as

$$\boxed{\mathbf{a}_T = \frac{\mathbf{r}'}{|\mathbf{r}'|} = \frac{d\mathbf{l}}{|d\mathbf{l}|}} \tag{2-4}$$

The unit vector of course points in the positive direction of C.

Alternatively, we may express a curve C by taking the x-coordinate as a parameter such as

$$\boxed{\mathbf{r} = x\,\mathbf{a}_x + g(x)\,\mathbf{a}_y + h(x)\,\mathbf{a}_z} \tag{2-5}$$

As we see in Fig. 2.1, $g(x)$ is the projection of C onto the xy-plane, and $h(x)$ is the projection of C onto the xz-plane.

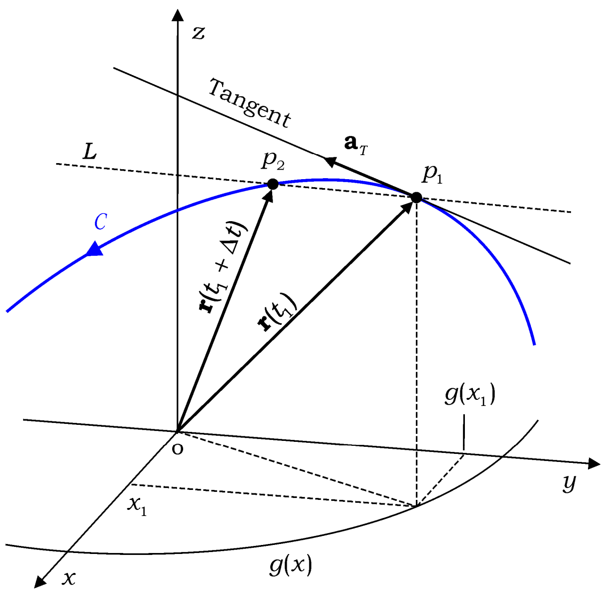


Fig. 2.1 A tangent to curve C.

Example 2-1

Given a point $p_1\!:\!\left(2, 2\sqrt{3}, 0\right)$ on the circle $x^2 + y^2 = 16$, find a unit vector along the tangent to the circle at p_1 by using (a) $\mathbf{r}(\phi)$, (b) $\mathbf{r}(x)$, and (c) $d\mathbf{l}$ in cylindrical coordinates.

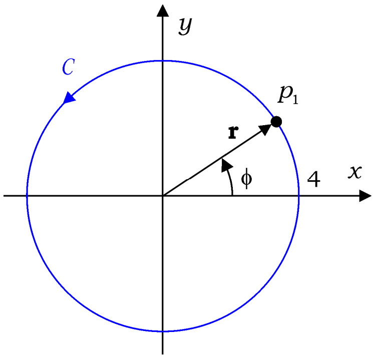


Fig. 2.2 A circular path of radius 4.

Solution

(a) Parametric representation of the circle is

$$\mathbf{r}(\phi) = x(\phi)\,\mathbf{a}_x + y(\phi)\,\mathbf{a}_y = 4\cos\phi\,\mathbf{a}_x + 4\sin\phi\,\mathbf{a}_y \tag{2-6}$$

Unit vector along the tangent to the circle is

$$\mathbf{a}_T = \frac{\mathbf{r}'}{|\mathbf{r}'|} = \frac{-4\sin\phi\,\mathbf{a}_x + 4\cos\phi\,\mathbf{a}_y}{\sqrt{(4\sin\phi)^2 + (4\cos\phi)^2}} = -\sin\phi\,\mathbf{a}_x + \cos\phi\,\mathbf{a}_y$$

Noting that p_1 corresponds to $\phi = 60^o$, we obtain

$$\mathbf{a}_T = -\left(\sqrt{3}/2\right)\mathbf{a}_x + (1/2)\,\mathbf{a}_y\,. \tag{2-7}$$

(b) In the region $y \ge 0$, the circle is expressed as $y = \sqrt{16 - x^2}$. Therefore its parametric representation is

$$\mathbf{r}(x) = x\,\mathbf{a}_x + y\,\mathbf{a}_y = x\,\mathbf{a}_x + \sqrt{16 - x^2}\,\mathbf{a}_y \qquad (y \ge 0) \tag{2-8}$$

Taking the derivatives of both sides of Eq. (2-8) with respective to *x*, we have

$$\mathbf{r}'(x) = \mathbf{a}_x - \frac{x}{\sqrt{16 - x^2}}\,\mathbf{a}_y$$

Inserting the x-coordinate of p_1, or $x = 2$, we have

$$\mathbf{a}_T = \frac{\mathbf{r}'}{|\mathbf{r}'|} = (\sqrt{3}/2)\,\mathbf{a}_x - (1/2)\,\mathbf{a}_y \tag{2-9}$$

As ϕ increases in Eq. (2-6), p_1 moves in the counterclockwise direction on the circle. In contrast, as x increases in Eq. (2-8), p_1 moves in the clockwise direction on the circle. For this reason, the sign in Eq. (2-7) is opposite to that in Eq. (2-9).

(c) The differential length vector in cylindrical coordinates has been previously defined in Chapter 1 as

$$d\mathbf{l} = d\rho\,\mathbf{a}_\rho + \rho d\phi\,\mathbf{a}_\phi + dz\,\mathbf{a}_z \tag{2-10}$$

On the circle we have $d\rho = 0 = dz$, and thus

$$d\mathbf{l} = 4\,d\phi\,\mathbf{a}_\phi \tag{2-11}$$

Inserting Eq. (2-11) into Eq. (2-4), we have

$$\mathbf{a}_T = \mathbf{a}_\phi$$

The cylindrical coordinate system greatly simplifies the solution.

Exercise 2.1

Find a parametric representation of a straight line parallel to the z-axis, passing through a point $(x, y, z) = (2, 1, 3)$.

Ans. $\mathbf{r}(t) = 2\,\mathbf{a}_x + \mathbf{a}_y + t\,\mathbf{a}_z \quad (-\infty < t < \infty)$.

Exercise 2.2

Find $|d\mathbf{r}/dx|$ if $\mathbf{r}(x)$ is a parametric representation of a straight line $y = x$.

Ans. $\sqrt{2}$.

2.1.2 Line Integral

The line integral is an extension of a definite integral into three dimensions. The line integral of a vector field $\mathbf{E}(\mathbf{r})$ along a path C is defined as

$$\boxed{\int_C \mathbf{E}\cdot d\mathbf{l} = \int_C E_t(x, y, z)\,dl} \tag{2-12}$$

where C is called the path of integration, $d\mathbf{l}$ is the differential length vector along C, and E_t is the component of $\mathbf{E}$ tangential to C. ***The line integral of* $\mathbf{E}$ *along a path* C *is an integration of the component of* $\mathbf{E}$ *tangent to* C.**

With reference to Fig. 2.3 in which the path C is subdivided into N line segments, we represent each line segment by an incremental length vector $\Delta\mathbf{l}$. As discussed in Chapter 1, the magnitude and direction of $\Delta\mathbf{l}_j$ belong to the point specified by the position vector $\mathbf{r}_j$. The line integral of $\mathbf{E}$ along C is expressed as

$$\int_C \mathbf{E}\cdot d\mathbf{l} = \lim_{\substack{N\to\infty \\ |\Delta\mathbf{l}|\to 0}} \sum_j^N \mathbf{E}(\mathbf{r}_j)\cdot\Delta\mathbf{l}_j \tag{2-13}$$

In the limit as $N\to\infty$, $\Delta\mathbf{l}$ becomes the differential length vector, that is,

$$\boxed{d\mathbf{l} = d\mathbf{r} = \mathbf{r}'dt} \tag{2-14}$$

which is the same as Eq. (2-3). Inserting Eq. (2-14) into Eq. (2-13), the line integral of $\mathbf{E}$ simply reduces to a definite integral over t, that is,

$$\boxed{\int_C \mathbf{E}\cdot d\mathbf{l} = \int_A^B \mathbf{E}\cdot\mathbf{r}'dt} \tag{2-15}$$

where $\mathbf{r}' = d\mathbf{r}/dt$, A and B represent the initial and terminal points of C, and t is the parameter used for the parametric representation of C.

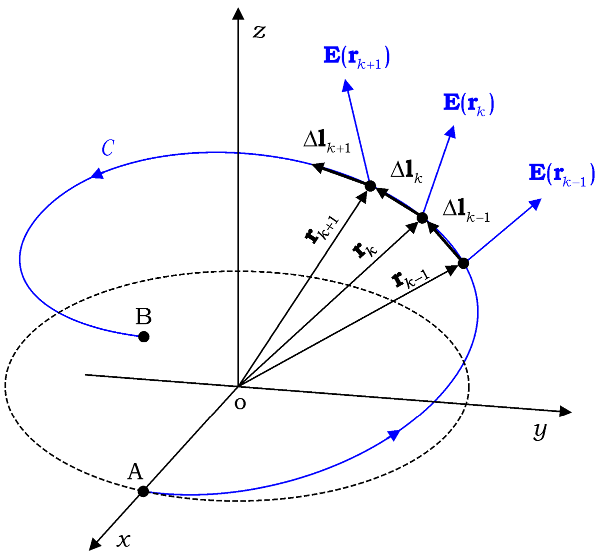


Fig. 2.3 Line integral of $\mathbf{E}$ along path C.

In the three coordinate systems, the line integral of $\mathbf{E}$ along path C also can be expressed as follows:

In Cartesian system

$$\int_C \mathbf{E} \bullet d\mathbf{l} = \int_C \left(E_x \mathbf{a}_x + E_y \mathbf{a}_y + E_z \mathbf{a}_z\right) \bullet \left(dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z\right) = \int_{x_1}^{x_2} E_x dx + \int_{y_1}^{y_2} E_y dy + \int_{z_1}^{z_2} E_z dz \tag{2-16a}$$

In cylindrical system

$$\int_C \mathbf{E} \bullet d\mathbf{l} = \int_C \left(E_\rho \mathbf{a}_\rho + E_\phi \mathbf{a}_\phi + E_z \mathbf{a}_z\right) \bullet \left(d\rho\,\mathbf{a}_\rho + \rho d\phi\,\mathbf{a}_\phi + dz\,\mathbf{a}_z\right) = \int_{\rho_1}^{\rho_2} E_\rho d\rho + \int_{\phi_1}^{\phi_2} E_\phi \rho d\phi + \int_{z_1}^{z_2} E_z dz \tag{2-16b}$$

In spherical system

$$\int_C \mathbf{E} \bullet d\mathbf{l} = \int_C \left(E_R \mathbf{a}_R + E_\theta \mathbf{a}_\theta + E_\phi \mathbf{a}_\phi\right) \bullet \left(dR\,\mathbf{a}_R + Rd\theta\,\mathbf{a}_\theta + R\sin\theta\, d\phi\,\mathbf{a}_\phi\right) = \int_{R_1}^{R_2} E_R dR + \int_{\theta_1}^{\theta_2} E_\theta R d\theta + \int_{\phi_1}^{\phi_2} E_\phi R\sin\theta\, d\phi \tag{2-16c}$$

In the above equations, subscript 1 stands for the initial point of C, and subscript 2 stands for the terminal point of C.

If path C lies in the $z = z_o$ plane in Cartesian coordinates, for instance, which is of the form $y = f(x)$, the differential length vector reduces to $d\mathbf{l} = dx\,\mathbf{a}_x + dy\,\mathbf{a}_y$. In this case the line integral of $\mathbf{E}$ along C becomes

$$\int_C \mathbf{E} \bullet d\mathbf{l} = \int_C \left(E_x \mathbf{a}_x + E_y \mathbf{a}_y + E_z \mathbf{a}_z\right) \bullet \left(dx\,\mathbf{a}_x + dy\,\mathbf{a}_y\right) = \int_{x_1}^{x_2} E_x\left(x, f(x), z_o\right) dx + \int_{y_1}^{y_2} E_y\left(f^{-1}(y), y, z_o\right) dy \tag{2-17}$$

where the initial point is at (x_1, y_1, z_o), and the terminal point is at (x_2, y_2, z_o).

If the path C is parallel to a coordinate axis, the line integral further reduces to an integral similar to a definite integral. For instance, if C is a straight line parallel to the x-axis, the differential length vector reduces to $d\mathbf{l} = dx\,\mathbf{a}_x$, and the line integral of $\mathbf{E}$ becomes

$$\int_C \mathbf{E} \bullet d\mathbf{l} = \int_{x_1}^{x_2} E_x(x, y_o, z_o) dx \tag{2-18}$$

where it is assumed that the initial point is at (x_1, y_o, z_o) and the terminal point is at (x_2, y_o, z_o). Even if the line integral in Eq. (2-18) is to be conducted in the negative x-direction, it is customary to use $d\mathbf{l} = dx\,\mathbf{a}_x$, instead of $d\mathbf{l} = -dx\,\mathbf{a}_x$, but interchange the limits of integration to account for the reversed direction, that, is from x_2 to x_1. The line integral in Eq. (2-18) is different than a definite integral in that a definite integral is defined graphically as the net area under the curve of a function, and thus has the upper limit of integration always larger than the lower limit.

When a line integral of $\mathbf{E}$ is conducted around a closed path C, it is denoted by a small circle on the integral sign such as

$$\boxed{\oint_C \mathbf{E} \bullet d\mathbf{l} = \oint_C E_t\, dl} \qquad (2\text{-}19)$$

The closed line integral is also referred to as the circulation of $\mathbf{E}$ around C.

The direction of travel on a closed path C is called the positive direction of C. The closed path involves a surface S, not necessarily a planar surface, which is bounded by C. The positive direction of C, or the direction of $d\mathbf{l}$ on C, is related to the direction of the differential area vector $d\mathbf{s}$ on S through the right-hand rule, as shown in Fig. 2.4: the right thumb points in the direction of $d\mathbf{s}$ on S when the four fingers advance in the direction of $d\mathbf{l}$ on C. When a closed line integral is conducted in the negative direction of C, it is customary to interchange the limits of integration instead changing the sign of $d\mathbf{l}$; from $\phi_1 = 2\pi$ to $\phi_2 = 0$ in place of from $\phi_1 = 0$ to $\phi_2 = 2\pi$, for instance. In other words, the expression for $d\mathbf{l}$ is always given as one of those in Eqs. (1-39), (1-55), and (1-77), regardless of whether the positive direction of C is clockwise or counterclockwise.

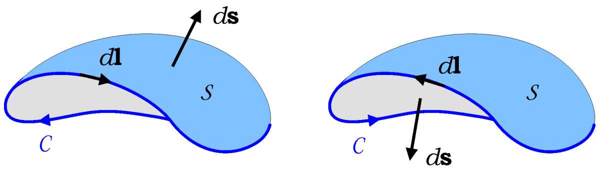


Fig. 2.4 Right-hand rule for the directions of $d\mathbf{l}$ and $d\mathbf{s}$.

Example 2-2

Find the line integral of $\mathbf{E} = y^2\,\mathbf{a}_x + (2xy + 4y)\,\mathbf{a}_y$ from point A:(2, 0, 0) to point B:(0, 2, 0) along

(a) straight line $y = -x + 2$, and

(b) arc of a circle $x^2 + y^2 = 4$ in the counterclockwise direction.

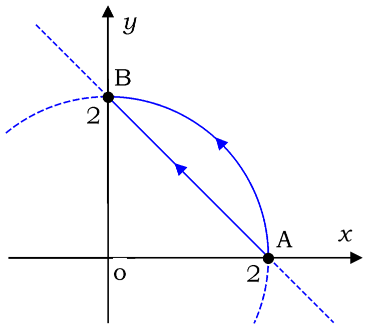


Fig. 2.5 Two paths of integration.

Solution

(a) The line integral is written as

$$\int_C \mathbf{E}\cdot d\mathbf{l} = \int_A^B \left[y^2\,\mathbf{a}_x + (2xy+4y)\,\mathbf{a}_y\right]\cdot\left(dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z\right)$$
$$= \int_2^0 y^2\,dx + \int_0^2 (2xy+4y)\,dy \tag{2-20}$$

Inserting $y = -x + 2$ into Eq. (2-20) leads to

$$\int_C \mathbf{E}\cdot d\mathbf{l} = \int_2^0 (-x+2)^2\,dx + \int_0^2 [2y(2-y)+4y]\,dy = 8\,.$$

(b) Inserting $x^2 + y^2 = 4$ into Eq. (2-20) leads to

$$\int_C \mathbf{E}\cdot d\mathbf{l} = \int_2^0 (4-x^2)\,dx + \int_0^2 \left[2y\sqrt{4-y^2} + 4y\right]dy = 8$$

The results in parts (a) and (b) are equal: the line integral of the given vector field $\mathbf{E}$ is independent of the path of integration. Such a field is called a conservative field.

Example 2-3

Determine the circulation of $\mathbf{E} = \rho^2\,\mathbf{a}_\rho + 3\sin\phi\,\mathbf{a}_\phi + 5z\,\mathbf{a}_z$ around the closed path shown in Fig. 2.6.

Solution

We split the closed line integral into two parts

$$\oint_C \mathbf{E}\cdot d\mathbf{l} = \int_A^B \mathbf{E}\cdot d\mathbf{l} + \int_B^A \mathbf{E}\cdot d\mathbf{l} \tag{2-21}$$

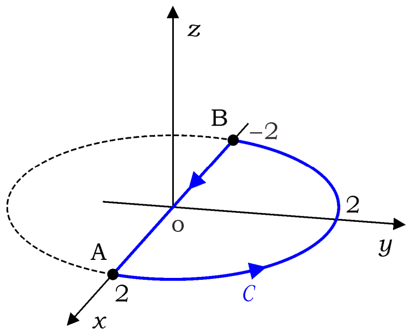


Fig. 2.6 Closed path for line integral.

Using $d\mathbf{l} = 2d\phi\,\mathbf{a}_\phi$ for the semi-circle, we write the first part as

$$\int_A^B \mathbf{E}\cdot d\mathbf{l} = \int_A^B (\rho^2\,\mathbf{a}_\rho + 3\sin\phi\,\mathbf{a}_\phi + 5z\,\mathbf{a}_z)\cdot(2d\phi\,\mathbf{a}_\phi)$$
$$= \int_{\phi=0}^{\phi=\pi} 6(\sin\phi)\,d\phi = 12$$

On the straight line from B to A, we have $d\mathbf{l} = dx\,\mathbf{a}_x$, $\rho^2 = x^2$, and $\mathbf{a}_\phi\cdot\mathbf{a}_x = 0$. Furthermore, we have $\mathbf{a}_\rho\cdot\mathbf{a}_x = -1$ on the line segment $\overline{Bo}$ but $\mathbf{a}_\rho\cdot\mathbf{a}_x = 1$ on $\overline{oA}$. The second part is therefore

$$\int_B^A \mathbf{E}\cdot d\mathbf{l} = \int_B^A \left(\rho^2\,\mathbf{a}_\rho + 3\sin\phi\,\mathbf{a}_\phi + 5z\,\mathbf{a}_z\right)\cdot(dx\,\mathbf{a}_x)$$
$$= \int_{x=-2}^{x=0} -x^2dx + \int_{x=0}^{x=2} x^2dx = 0$$

Thus, the answer is

$$\oint_C \mathbf{E}\cdot d\mathbf{l} = 12.$$

Exercise 2.3
Repeat the closed line integral in **Example 2-3**, by assuming C to be displaced in the $+z$-direction and placed in the $z = 4$ plane.
Ans. 12.

Exercise 2.4
If the vector field **E** has the unit of the volt per meter, what is the unit of the line integral of **E**?
Ans. Volt.

2.1.3 Surface Integral

An open surface is generally expressed as $z = f(x,y)$, or $g(x,y,z) = k$, where k is a constant, in three-dimensional space. In view of this, we write the position vector of a point on the surface as

$$\boxed{\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + f(x,y)\,\mathbf{a}_z} \tag{2-22}$$

which is called a parametric representation of the surface.

As an example, let us consider a planar surface $\mathcal{S}$ as shown in Fig. 2.7. Two straight lines $\mathcal{L}_1$ and $\mathcal{L}_2$ on $\mathcal{S}$ represent the intersections of $\mathcal{S}$ with the $y = y_1$ and $x = x_1$ planes respectively, where x_1 and y_1 are Cartesian coordinates of point p_1 on $\mathcal{S}$. Using the space coordinates x and y as parameters, we obtain parametric representations of $\mathcal{L}_1$ and $\mathcal{L}_2$ as

$$\mathbf{r}(x) = x\,\mathbf{a}_x + y_1\,\mathbf{a}_y + f(x,y_1)\,\mathbf{a}_z \qquad \text{(for } \mathcal{L}_1\text{)} \tag{2-23a}$$

$$\mathbf{r}(y) = x_1\,\mathbf{a}_x + y\,\mathbf{a}_y + f(x_1,y)\,\mathbf{a}_z \qquad \text{(for } \mathcal{L}_2\text{)} \tag{2-23b}$$

Note that y_1 is considered to be constant in Eq. (2-23a), and, similarly, x_1 is considered to be constant in Eq. (2-23b). From Eq. (2-23), we obtain the differential length vectors directed along $\mathcal{L}_1$ and $\mathcal{L}_2$, at point p_1, as follows:

$$d\mathbf{l}^1 = \left.\frac{\partial \mathbf{r}}{\partial x}\right|_{y=y_1} dx \equiv \mathbf{r}'_x dx \tag{2-24a}$$

$$d\mathbf{l}^2 = \left.\frac{\partial \mathbf{r}}{\partial y}\right|_{x=x_1} dy \equiv \mathbf{r}'_y dy \tag{2-24b}$$

where superscripts 1 and 2 denote lines $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively, and $\mathbf{r}'_x$ and $\mathbf{r}'_y$ denote the partial derivatives of $\mathbf{r}$ with respect to x and y respectively. Two differential length vectors $d\mathbf{l}^1$ and $d\mathbf{l}^2$ form a parallelogram of a differential area $\left|d\mathbf{l}^1 \times d\mathbf{l}^2\right|$ on surface $\mathcal{S}$. In view of these discussions, we define a differential area vector on a general surface $\mathcal{S}$ as

$$\boxed{d\mathbf{s} = d\mathbf{l}^1 \times d\mathbf{l}^2 = (\mathbf{r}'_x \times \mathbf{r}'_y)dxdy} \tag{2-25}$$

It should be noted that the differential area vectors previously defined in the three coordinate systems in Chapter 1 are special cases of Eq. (2-25). For instance, for the $z = c$ plane, Eq. (2-25) reduces to $d\mathbf{s} = dxdy\,\mathbf{a}_z$, as was shown in Chapter 1. We also note that Eq. (2-25) is true for any smooth surface, not necessarily a planar surface.

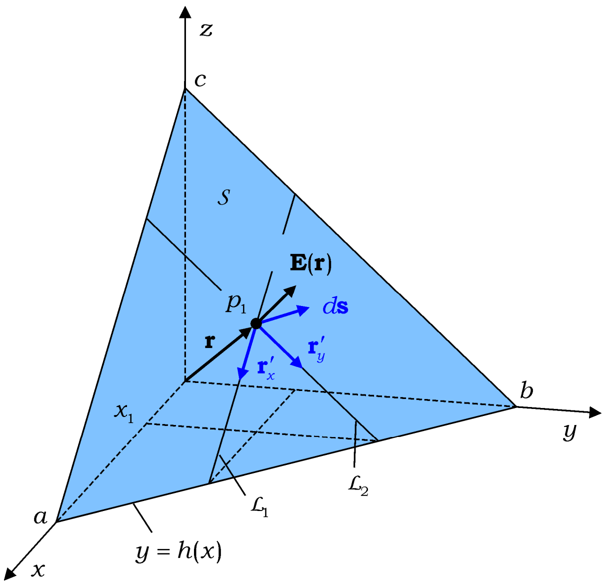


Fig. 2.7 Differential area vector $d\mathbf{s} = (\mathbf{r}'_x \times \mathbf{r}'_y)dxdy$ on a general surface.

The surface integral of a vector field $\mathbf{E}(\mathbf{r})$ over a surface $\mathcal{S}$ is defined as

$$\int_{\mathcal{S}} \mathbf{E}(\mathbf{r}) \cdot d\mathbf{s} = \int_{\mathcal{S}} E_n ds \tag{2-26}$$

where E_n is the component of $\mathbf{E}$ normal to $\mathcal{S}$, and $d\mathbf{s}$ is the differential area vector on $\mathcal{S}$. ***The surface integral of a vector field* $\mathbf{E}$ *over a surface* $\mathcal{S}$ *is the integration of the component of* $\mathbf{E}$ *normal to* $\mathcal{S}$.**

Using Eqs. (2-22) and (2-25) in Eq. (2-26), we write the surface integral of $\mathbf{E}(\mathbf{r})$ over $\mathcal{S}$ as

$$\int_{\mathcal{S}} \mathbf{E}(\mathbf{r}) \cdot d\mathbf{s} = \int_{x=0}^{x=a} \int_{y=0}^{y=h(x)} \mathbf{E}\left(x, y, f(x,y)\right) \cdot \left(\mathbf{r}'_x \times \mathbf{r}'_y\right) dxdy \tag{2-27}$$

We note from Eq. (2-27) that the surface integral is converted to a double integral over the triangular region defined by the x-axis, y-axis, and the straight line $y = h(x)$ in the xy-plane. We also note that the triangular region is the projection of the surface $\mathcal{S}$ onto the xy-plane.

If the surface $\mathcal{S}$ is a rectangular plane residing in the $z = z_o$ plane, with the sides parallel to the x- and y-axes, the surface integral in Eq. (2-27) further reduces to

$$\int_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{s} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} E_z(x, y, z_o) dxdy \tag{2-28}$$

where we used $d\mathbf{s} = dxdy\,\mathbf{a}_z$. We note that the two space coordinates x and y are independent of each other in Eq. (2.28), while they are interrelated by $y = h(x)$ in Eq. (2-27).

Example 2-4
With reference to Fig. 2.7, determine the surface area of the plane $\mathcal{S}$ with $a = b = c = 1$.

Solution

Let us begin with the equation of the plane $\mathcal{S}$, i.e.,

$$x + y + z = 1 \tag{2-29}$$

Parametric representation of $\mathcal{S}$ is

$$\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + (1 - x - y)\,\mathbf{a}_z$$

Partial derivatives of $\mathbf{r}$ with respect to x and y are

$$\mathbf{r}'_x = \mathbf{a}_x - \mathbf{a}_z \tag{2-30a}$$
$$\mathbf{r}'_y = \mathbf{a}_y - \mathbf{a}_z \tag{2-30b}$$

Inserting Eq. (2-30) into Eq. (2-25), the differential area vector on $\mathcal{S}$ is

$$d\mathbf{s} = (\mathbf{a}_x - \mathbf{a}_z) \times (\mathbf{a}_y - \mathbf{a}_z)dxdy = (\mathbf{a}_x + \mathbf{a}_y + \mathbf{a}_z)dxdy \tag{2-31}$$

The projection of surface $\mathcal{S}$ onto the xy-plane forms a triangle with the hypotenuse given by

$$x + y = 1 \tag{2-32}$$

Upon using Eqs. (2-31) and (2-32), the surface area of $\mathcal{S}$ is

$$A = \int_{\mathcal{S}} |d\mathbf{s}| = \int_{x=0}^{x=1}\int_{y=0}^{y=1-x} \sqrt{3}\; dxdy = \frac{\sqrt{3}}{2}.$$

Example 2-5
Determine the surface integral of a vector field $\mathbf{A} = 2\,\mathbf{a}_z$ over a hemispherical surface of radius 5 as shown in Fig. 2.8, by assuming $d\mathbf{s}$ to be directed away from the origin.

Solution

Two differential length vectors $d\mathbf{l}^1$ and $d\mathbf{l}^2$ in spherical coordinates are

$$d\mathbf{l}^1 = R\,d\theta\,\mathbf{a}_\theta = 5\,d\theta\,\mathbf{a}_\theta$$
$$d\mathbf{l}^2 = R\sin\theta\,d\phi\,\mathbf{a}_\phi = 5\sin\theta\,d\phi\,\mathbf{a}_\phi$$

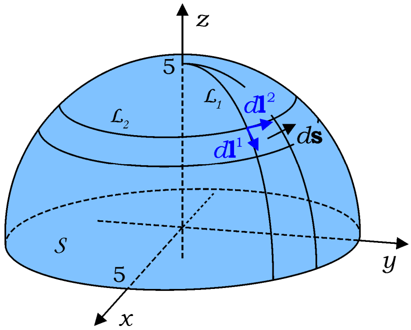


Fig. 2.8 Hemispherical surface of radius 5.

Differential area vector on the hemisphere is

$$d\mathbf{s} = d\mathbf{l}^1 \times d\mathbf{l}^2 = 25 \sin\theta\, d\theta\, d\phi\, \mathbf{a}_R$$

Using $\mathbf{a}_z \cdot \mathbf{a}_R = \cos\theta$, the surface integral is

$$\int_S \mathbf{A} \cdot d\mathbf{s} = \int_S (2\mathbf{a}_z) \cdot (25 \sin\theta\, d\theta\, d\phi\, \mathbf{a}_R) = 50 \int_{\theta=0}^{\pi/2} \int_{\phi=0}^{2\pi} \cos\theta \sin\theta\, d\theta\, d\phi$$
$$= 50\pi.$$

Exercise 2.5
A sphere of radius 5 is centered at the origin. Find the surface area in the region $z \geq 4$.
Ans. 10π.

Exercise 2.6
If the vector field **D** has the unit of the coulomb per square meter, what is the unit of the surface integral of **D**?
Ans. Coulomb.

Review Questions with Hints

RQ 2.1 Which component of **A** is used for the line integral? [Eq. (2-12)]
RQ 2.2 Which component of **A** is used for the surface integral? [Eq. (2-26)]
RQ 2.3 Under what conditions does a line integral reduce to a definite integral? [Eq. (2-18)]
RQ 2.4 Under what conditions does a surface integral reduce to a double integral? [Eq. (2-28)]

2.2 Directional Derivative and Gradient

A scalar function $V(\mathbf{r})$ is usually used to describe a distribution of the scalar quantity V in a region of space. If V represents a physical quantity such as temperature in a room, or the electric potential due to an electric charge, the scalar field must be a smooth function of position so that $V(\mathbf{r})$ may be differentiable at each and every point in the region. Under this condition, at a point in the three-dimensional space, the space rate of change of $V(\mathbf{r})$ must exist for any direction. The space rate of change of V in an $\mathbf{a}_l$-direction at point p is called the directional derivative of V in the $\mathbf{a}_l$-direction at p. The maximum of the directional derivatives at a given point has a special meaning, and is formulated into an operator called the gradient.

The directional derivative of $V(\mathbf{r})$ is denoted as dV/dl, which represents the space rate of change of V in the direction of increase of the differential length dl. At a point with position vector $\mathbf{r}_1$, the directional derivative of V in the $\mathbf{a}_l$-direction is written as

$$\left.\frac{dV}{dl}\right|_{\mathbf{r}_1,\mathbf{a}_l} = \lim_{dl\to 0}\frac{V(\mathbf{r}_1 + dl\,\mathbf{a}_l) - V(\mathbf{r}_1)}{dl} \tag{2-33}$$

where dl is the differential length along the unit vector $\mathbf{a}_l$. It is evident from Eq. (2-33) that the directional derivative depends not only on the position of the point in space but also on the choice of the direction. Accordingly, there are an infinite number of directional derivatives at a given point in space; there must be the maximum of the directional derivatives.

The gradient of V at point p is defined as a vector representing the magnitude and direction of the maximum directional derivative of V at p.

The gradient of V at p is denoted as

$$\boxed{grad\, V = \frac{dV}{dn}\mathbf{a}_n} \tag{2-34}$$

where the unit vector $\mathbf{a}_n$ points in the direction of increase of the differential length dn, or the direction along which the derivative dV/dn is maximum.

When a scalar field $V(\mathbf{r})$ is defined in a region of space, we can always find the spatial points at which V is constant such that $V(\mathbf{r}) = V_1$. These points should form a smooth surface in the three-dimensional space, if V represents a real physical quantity. By the same token, we should be able to find another surface of $V_1 + dV$ in the immediate vicinity of the surface of V_1 as illustrated in Fig. 2.9. Since $V(\mathbf{r})$ is a single-valued function of position, the two surfaces never cross each other. With reference to Fig. 2.9, suppose we intend to move from point p_1 on surface V_1 to a point on surface $V_1 + dV$; i.e., one of p_2, p_3, and p_4. In doing so, we will observe the same change dV, although the travel distance is obviously different for the three cases. If the line segment $\overline{p_1 p_2}$ is perpendicular to the two surfaces, it must be the shortest distance from p_1 to the neighboring surface, and thus the directional derivative of V at p_1 is maximum

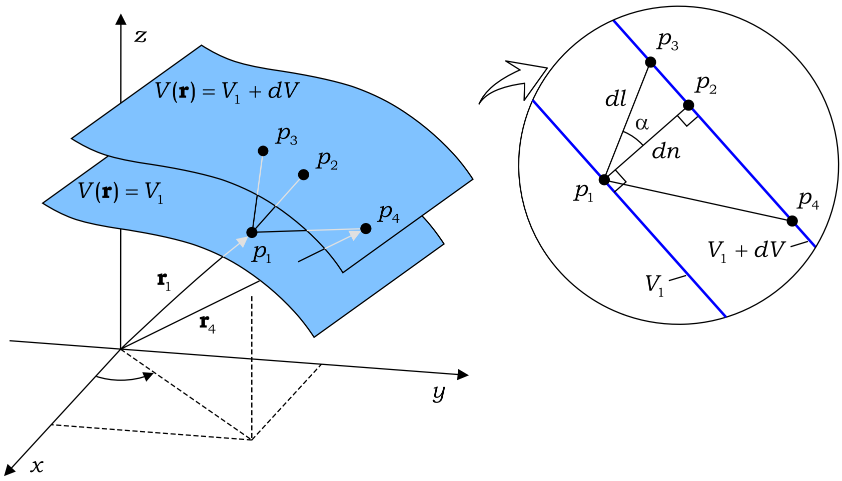


Fig. 2.9 Surfaces of constant *V*.

along the line segment $\overline{p_1 p_2}$. In consequence, ***the gradient of V at a point in space is normal to the surface of constant V passing through that point***.

The gradient of V at a point in space is useful for determining the directional derivative of V in an arbitrary direction at that point. Applying the chain rule from calculus to Eq. (2-34), we have

$$\frac{dV}{dl} = \frac{dV}{dn}\frac{dn}{dl} = \frac{dV}{dn}\cos\alpha = \frac{dV}{dn}\mathbf{a}_n \bullet \mathbf{a}_l \tag{2-35}$$

where we used the relations, $dn/dl = \cos\alpha$ and $\mathbf{a}_n \bullet \mathbf{a}_l = \cos\alpha$, as can be easily obtained from Fig. 2.9. The unit vector $\mathbf{a}_n$ is normal to the surface of constant V, and $\mathbf{a}_l$ is along the direction of increasing dl. Combining Eqs. (2-34) and (2-35) we write the direction derivative of V in the direction of $\mathbf{a}_l$ as

$$\boxed{\frac{dV}{dl} = grad\, V \bullet \mathbf{a}_l \equiv \nabla V \bullet \mathbf{a}_l} \tag{2-36}$$

Note that the gradient of V is denoted either as $grad\, V$ or as ∇V (read "del" V). ***The directional derivative of V in the direction of*** $\mathbf{a}_l$ ***is equal to the dot product of*** ∇V ***and*** $\mathbf{a}_l$.

Next, From Eq. (2-36) we can obtain

$$\boxed{dV = \nabla V \bullet dl\,\mathbf{a}_l = \nabla V \bullet d\mathbf{l}} \tag{2-37}$$

The differential of V in the $\mathbf{a}_l$-direction is equal to the dot product of ∇V and the differential length vector $d\mathbf{l}$ directed along $\mathbf{a}_l$. Note that Eqs. (2-36) and (2-37) are two important applications where the gradient is conveniently used.

From calculus, the total differential of V in Cartesian coordinates is given as

$$\begin{aligned} dV &= \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz \\ &= \left(\frac{\partial V}{\partial x}\mathbf{a}_x + \frac{\partial V}{\partial y}\mathbf{a}_y + \frac{\partial V}{\partial z}\mathbf{a}_z\right)\bullet(dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z) \\ &= \left(\frac{\partial V}{\partial x}\mathbf{a}_x + \frac{\partial V}{\partial y}\mathbf{a}_y + \frac{\partial V}{\partial z}\mathbf{a}_z\right)\bullet d\mathbf{l} \end{aligned} \tag{2-38}$$

In Eq. (2-38), the right-hand side is separated into two parts, and written in vector notation in such a way that the second part is equal to the differential length vector in Cartesian coordinates. Comparison of Eq. (2-37) with Eq. (2-38) reveals that the gradient of V in Cartesian coordinates is

$$\nabla V = \frac{\partial V}{\partial x}\mathbf{a}_x + \frac{\partial V}{\partial y}\mathbf{a}_y + \frac{\partial V}{\partial z}\mathbf{a}_z = \left(\mathbf{a}_x\frac{\partial}{\partial x} + \mathbf{a}_y\frac{\partial}{\partial y} + \mathbf{a}_z\frac{\partial}{\partial z}\right)V \tag{2-39}$$

In view of Eq. (2-39), we define the del operator as

$$\boxed{\nabla \equiv \mathbf{a}_x\frac{\partial}{\partial x} + \mathbf{a}_y\frac{\partial}{\partial y} + \mathbf{a}_z\frac{\partial}{\partial z}} \tag{2-40}$$

It should be noted that the del operator is defined in Cartesian coordinates only.

The expression for the gradient of V in cylindrical coordinates can also be derived from Eq. (2-37). From calculus, in cylindrical coordinates, the total differential of V at point $p_1:(\rho_1, \phi_1, z_1)$ is written as

$$\begin{aligned} dV &= \frac{\partial V}{\partial \rho}d\rho + \frac{\partial V}{\partial \phi}d\phi + \frac{\partial V}{\partial z}dz \\ &= \left(\frac{\partial V}{\partial \rho}\mathbf{a}_\rho + \frac{\partial V}{\partial \phi}\mathbf{a}_\phi + \frac{\partial V}{\partial z}\mathbf{a}_z\right)\bullet(d\rho\,\mathbf{a}_\rho + d\phi\,\mathbf{a}_\phi + dz\,\mathbf{a}_z) \\ &= \left(\frac{\partial V}{\partial \rho}\mathbf{a}_\rho + \frac{1}{\rho_1}\frac{\partial V}{\partial \phi}\mathbf{a}_\phi + \frac{\partial V}{\partial z}\mathbf{a}_z\right)\bullet d\mathbf{l} \end{aligned} \tag{2-41}$$

Following the same procedure, the right-hand side of Eq. (2-41) is separated into two parts, and written in vector notation so that the second part is equal to the differential length vector in cylindrical coordinates. Comparison of Eq. (2-37) with Eq. (2-41) shows that the first part on the right-hand side of Eq. (2-41) is the gradient of V in cylindrical coordinates.

Similarly, in spherical coordinates, the total differential of V is written as

$$dV = \frac{\partial V}{\partial R} dR + \frac{\partial V}{\partial \theta} d\theta + \frac{\partial V}{\partial \phi} d\phi$$
$$= \left(\frac{\partial V}{\partial R} \mathbf{a}_R + \frac{\partial V}{\partial \theta} \mathbf{a}_\theta + \frac{\partial V}{\partial \phi} \mathbf{a}_\phi \right) \bullet (dR\,\mathbf{a}_R + d\theta\,\mathbf{a}_\theta + d\phi\,\mathbf{a}_\phi) \qquad (2\text{-}42)$$
$$= \left(\frac{\partial V}{\partial R} \mathbf{a}_R + \frac{1}{R_1} \frac{\partial V}{\partial \theta} \mathbf{a}_\theta + \frac{1}{R_1 \sin\theta_1} \frac{\partial V}{\partial \phi} \mathbf{a}_\phi \right) \bullet d\mathbf{l}$$

where $d\mathbf{l}$ is the differential length vector in spherical coordinates. We identify the first part on the right-hand side of Eq. (2-42) with the gradient of V in spherical coordinates.

Although the del operator is defined in Cartesian coordinates only, it is customary to use ∇V to denote the gradient of V in other coordinate system. The gradient of V in Cartesian, cylindrical, and spherical coordinates are

$$\boxed{\nabla V = \frac{\partial V}{\partial x} \mathbf{a}_x + \frac{\partial V}{\partial y} \mathbf{a}_y + \frac{\partial V}{\partial z} \mathbf{a}_z} \qquad \text{(Cartesian)} \qquad (2\text{-}43a)$$

$$\boxed{\nabla V = \frac{\partial V}{\partial \rho} \mathbf{a}_\rho + \frac{1}{\rho} \frac{\partial V}{\partial \phi} \mathbf{a}_\phi + \frac{\partial V}{\partial z} \mathbf{a}_z} \qquad \text{(cylindrical)} \qquad (2\text{-}43b)$$

$$\boxed{\nabla V = \frac{\partial V}{\partial R} \mathbf{a}_R + \frac{1}{R} \frac{\partial V}{\partial \theta} \mathbf{a}_\theta + \frac{1}{R \sin\theta} \frac{\partial V}{\partial \phi} \mathbf{a}_\phi} \qquad \text{(spherical)} \qquad (2\text{-}43c)$$

Subscript 1 is omitted in Eq. (2-43) for generalization.

The gradient of V in the general coordinates (u, v, w) is written as

$$\boxed{\nabla V = \left(\mathbf{a}_u \frac{1}{h_1} \frac{\partial}{\partial u} + \mathbf{a}_v \frac{1}{h_2} \frac{\partial}{\partial v} + \mathbf{a}_w \frac{1}{h_3} \frac{\partial}{\partial w} \right) V} \qquad (2\text{-}44)$$

where the metric coefficients are given as follows:

$$h_1 = 1, \quad h_2 = 1, \quad h_3 = 1 \qquad (u, v, w) = (x, y, z) \qquad (2\text{-}45a)$$
$$h_1 = 1, \quad h_2 = \rho, \quad h_3 = 1 \qquad (u, v, w) = (\rho, \phi, z) \qquad (2\text{-}45b)$$
$$h_1 = 1, \quad h_2 = R, \quad h_3 = R \sin\theta \qquad (u, v, w) = (R, \theta, \phi) \qquad (2\text{-}45c)$$

The metric coefficients are necessary to convert differential angles into differential lengths. It should be noted that the total differential of V however involves no metric coefficient, because it is defined in terms of the differential coordinates, whether or not the coordinates represent angles.

Example 2-6

Given a scalar field $V(\mathbf{r}) = x^2 + 4yz$ in Cartesian coordinates, find, at point p_1:$(4,-1,3)$,

(a) gradient of V

(b) directional derivative along the direction of a vector $\mathbf{l} = 3\,\mathbf{a}_x - 2\,\mathbf{a}_y - \sqrt{3}\,\mathbf{a}_z$.

Solution

(a) From Eq. (2-43a), the gradient of V in Cartesian coordinates is

$$\nabla V = \frac{\partial V}{\partial x}\mathbf{a}_x + \frac{\partial V}{\partial y}\mathbf{a}_y + \frac{\partial V}{\partial z}\mathbf{a}_z = 2x\,\mathbf{a}_x + 4z\,\mathbf{a}_y + 4y\,\mathbf{a}_z \tag{2-46}$$

Inserting the coordinates of p_1 into Eq. (2-46), we have

$$\nabla V = 8\,\mathbf{a}_x + 12\,\mathbf{a}_y - 4\,\mathbf{a}_z$$

(b) The unit vector of $\mathbf{l}$ is

$$\mathbf{a}_l = \frac{3}{4}\mathbf{a}_x - \frac{1}{2}\mathbf{a}_y - \frac{\sqrt{3}}{4}\mathbf{a}_z$$

From Eq. (2-36), the directional derivative of V in the $\mathbf{a}_l$-direction is

$$\frac{dV}{dl} = \nabla V \bullet \mathbf{a}_l = (8\,\mathbf{a}_x + 12\,\mathbf{a}_y - 4\,\mathbf{a}_z) \bullet \left(\frac{3}{4}\mathbf{a}_x - \frac{1}{2}\mathbf{a}_y - \frac{\sqrt{3}}{4}\mathbf{a}_z\right) = \sqrt{3}$$

Example 2-7

Find $\nabla \dfrac{1}{\mathcal{R}}$ in Cartesian coordinates, where $\mathcal{R} = |\mathbf{r} - \mathbf{r}'|$.

Solution

Position vectors in the unprimed and primed systems are

$$\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z$$
$$\mathbf{r}' = x'\,\mathbf{a}_x + y'\,\mathbf{a}_y + z'\,\mathbf{a}_z$$

Thus, the distance vector is given as

$$\boldsymbol{\mathcal{R}} = \mathbf{r} - \mathbf{r}' = (x - x')\,\mathbf{a}_x + (y - y')\,\mathbf{a}_y + (z - z')\,\mathbf{a}_z \tag{2-47}$$

The reciprocal of $|\mathbf{r} - \mathbf{r}'|$ is

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \left[(x - x')^2 + (y - y')^2 + (z - z')^2\right]^{-1/2} \tag{2-48}$$

Partial derivatives of Eq. (2-48) are

$$\frac{\partial}{\partial x}\frac{1}{|\mathbf{r}-\mathbf{r}'|} = -(x-x')\left[(x-x')^2+(y-y')^2+(z-z')^2\right]^{-3/2}$$

$$\frac{\partial}{\partial y}\frac{1}{|\mathbf{r}-\mathbf{r}'|} = -(y-y')\left[(x-x')^2+(y-y')^2+(z-z')^2\right]^{-3/2}$$

$$\frac{\partial}{\partial z}\frac{1}{|\mathbf{r}-\mathbf{r}'|} = -(z-z')\left[(x-x')^2+(y-y')^2+(z-z')^2\right]^{-3/2}$$

Combining the above equations we obtain

$$\nabla\frac{1}{|\mathbf{r}-\mathbf{r}'|} = -\frac{(x-x')\,\mathbf{a}_x+(y-y')\,\mathbf{a}_y+(z-z')\,\mathbf{a}_z}{\left[(x-x')^2+(y-y')^2+(z-z')^2\right]^{3/2}}$$

Thus, we have

$$\boxed{\nabla\frac{1}{\mathcal{R}} = -\frac{\boldsymbol{\mathcal{R}}}{\mathcal{R}^3}} \tag{2-49}$$

Note that the del operator is independent of the primed coordinates, and acts on the x-, y-, and z-coordinates only.

Example 2-8

A spherical surface is given by $x^2+y^2+z^2=9$ in Cartesian coordinates. Find the outward unit normal to the surface at point $(2,2,1)$.

Solution

The given spherical surface is a member of the family of spheres defined as $f(x,y,z)=x^2+y^2+z^2$, in which it is specified as $f(x,y,z)=9$. The gradient of f is normal to the surface of constant f.

We obtain the gradient of f as $\nabla f = 2x\,\mathbf{a}_x + 2y\,\mathbf{a}_y + 2z\,\mathbf{a}_z$, and a unit vector in the direction of ∇f as

$$\frac{\nabla f}{|\nabla f|} = \frac{x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z}{\sqrt{x^2+y^2+z^2}} \tag{2-50}$$

Inserting $(x,y,z)=(2,2,1)$ into Eq. (2-50), the unit normal to the surface passing through the given point is

$$\frac{1}{3}\left(2\,\mathbf{a}_x + 2\,\mathbf{a}_y + \mathbf{a}_z\right) \tag{2-51}$$

It points out of the sphere and is thus the answer. Otherwise, we would multiply it by -1.

Alternatively, we may begin with the family of spheres expressed as $g(R,\theta,\phi) = R^2$ in spherical coordinates, in which the given sphere is specified as $g = 9$. The gradient of g in spherical coordinates is

$$\nabla g = \frac{\partial g}{\partial R}\mathbf{a}_R + \frac{1}{R}\frac{\partial g}{\partial \theta}\mathbf{a}_\theta + \frac{1}{R\sin\theta}\frac{\partial g}{\partial \phi}\mathbf{a}_\phi = 2R\,\mathbf{a}_R$$

The unit vector of ∇g is simply

$$\mathbf{a}_R \tag{2-52}$$

It is the outward unit normal to the given sphere in spherical coordinates.

Example 2-9

In the presence of a scalar field $V = (x^2+3)y^2$, a circular path of radius 1 is centered at the origin in the $z = 0$ plane.

(a) Evaluate $\int_C \nabla V \bullet d\mathbf{l}$ along the circle, in the counterclockwise direction, from $p_1 : (1,0,0)$ to $p_2 : (0,1,0)$.

(b) Verify $\int_C \nabla V \bullet d\mathbf{l} = V(p_2) - V(p_1)$.

Solution

(a) Gradient of V in Cartesian coordinates is

$$\nabla V = 2xy^2\,\mathbf{a}_x + 2y(x^2+3)\,\mathbf{a}_y$$

The line integral of ∇V is written from the equation of the circle $x^2 + y^2 = 1$ as

$$\int_{p_1}^{p_2} \nabla V \bullet d\mathbf{l} = \int_{(1,0,0)}^{(0,1,0)} \left[2xy^2\,\mathbf{a}_x + 2y(x^2+3)\,\mathbf{a}_y\right] \bullet (dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z)$$
$$= \int_{x=1}^{x=0} 2x(1-x^2)\,dx + \int_{y=0}^{y=1} 2y(1-y^2+3)\,dy = 3.$$

(b) The line integral of ∇V is written from the definitions of ∇V and $d\mathbf{l}$ as

$$\int_{p_1}^{p_2} \nabla V \bullet d\mathbf{l} = \int_{p_1}^{p_2} \left(\frac{\partial V}{\partial x}\mathbf{a}_x + \frac{\partial V}{\partial y}\mathbf{a}_y + \frac{\partial V}{\partial z}\mathbf{a}_z\right) \bullet (dx\,\mathbf{a}_x + dy\,\mathbf{a}_y + dz\,\mathbf{a}_z)$$
$$= \int_{p_1}^{p_2} \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz = \int_{p_1}^{p_2} dV$$
$$= V(p_2) - V(p_1) = 3$$

The line integral of the gradient of a scalar field is independent of the path of integration; it only depends on the initial and terminal points of the path.

Exercise 2.7
Given a scalar field $V = \cos(2x + 3y + z)$, find, at point $p{:}(2,1,1)$, (a) gradient of *V*, and (b) directional derivative of *V* along $\mathbf{a}_x$.
Ans. (a) $-(2\,\mathbf{a}_x + 3\,\mathbf{a}_y + \mathbf{a}_z)\sin(8)$, (b) $-2\sin(8)$.

Exercise 2.8
If a scalar field *V* has the unit of the volt, what is the unit of ∇V ?
Ans. Volt per meter.

Exercise 2.9
Are the following expressions true? (a) $\nabla(U+V) = \nabla U + \nabla V$,
(b) $\nabla(UV) = V\nabla U + U\nabla V$, (c) $\nabla\left(U^4\right) = 4U^3\left(\nabla U\right)$, and
(d) $\nabla(1/U) = 1/(\nabla U)$.
Ans. (a) Yes, (b) Yes, (c) Yes, (d) No.

Review Questions with Hints

RQ 2.5 What are the significances of the magnitude and direction of the gradient of *V*? [Eq.(2-34)]
RQ 2.6 Does $V = 0$ at a point in space imply $\nabla V = 0$ at the point?[Eq.(2-34)]
RQ 2.7 State the relation between the directional derivative of *V* in the $\mathbf{a}_l$-direction and the gradient of *V*. [Eq.(2-36)]
RQ 2.8 Define the del operator in Cartesian coordinates. [Eq.(2-40]
RQ 2.9 Express ∇V in the three coordinate systems. [Eq.(2-43)]

2.3 Flux and Flux Density

The flux usually means a flow of something, and is defined as a scalar quantity crossing a reference point per unit area per unit time. When the concept of the flux is used in the study of electromagnetics, the flux may represent the number of field lines passing through a given surface. In addition, the flux density is used to represent the number of field lines per unit area. The flux density forms a vector field, which is generally given as a smooth function of position in three-dimensional space.

From geometry, a smooth curve is defined as one on which the tangent line varies only continuously as we move along the curve. Since a physical vector quantity exhibits no abrupt changes in direction as a function of position, we can draw a smooth line connecting spatial points in the vector field in such a way that the vector at each point is tangential to the line. These lines are called the field lines, or flux lines. The tangent to the field line represents the direction of the vector at the point, while the density of the field lines in a region surrounding the point represents the magnitude of the vector. Electromagnetics involves two flux densities: the electric flux density **D** measured in coulombs per square meter and the magnetic flux density **B** measured in webers per square meter.

The total flux passing through a surface is equal to the surface integral of the flux density over the given surface. As an example, let us consider an electron cloud of a number density $n_e[\mathrm{m}^{-3}]$ flowing with a velocity $\mathbf{v}$[m/s]. The flow of electrons can be conveniently described by a flux density, which is defined as the number of electrons crossing per unit area of a cross section per unit time. Here, the cross section is a plane perpendicular to the direction of flow of the electrons. The flux density of flow of the electrons, **N**, is equal to the product of the number density and the velocity. i.e.,

$$\mathbf{N} = n_e \mathbf{v} \qquad [\mathrm{m}^{-2} \bullet \mathrm{s}^{-1}] \qquad (2\text{-}53)$$

With reference to Fig. 2.10, the differential area $|d\mathbf{s}|$ on surface S is equivalent to an area $|d\mathbf{s}|\cos\alpha$ on the cross section. Thus, the differential flux $d\Psi$ passing through the area $|d\mathbf{s}|$, or the number of electrons passing through $|d\mathbf{s}|$ per unit time, is written as

$$d\Psi = |\mathbf{N}||d\mathbf{s}|\cos\alpha = |\mathbf{N}||d\mathbf{s}|\,\mathbf{a}_N \bullet \mathbf{a}_s = \mathbf{N} \bullet d\mathbf{s}$$

where $\cos\alpha$ is conveniently replaced by the dot product of $\mathbf{a}_N$ and $\mathbf{a}_s$, which are unit vectors in the directions of **N** and $d\mathbf{s}$, respectively. The total flux passing through surface S is therefore

$$\boxed{\Psi = \int_S d\Psi = \int_S \mathbf{N} \bullet d\mathbf{s}} \qquad (2\text{-}54)$$

Again, ***the total flux through a surface S is equal to the surface integral of the flux density over S.*** It is important to note that the magnitude of the flux density is the quantity crossing per unit area of a cross section, or a plane perpendicular to the direction of the flux density.

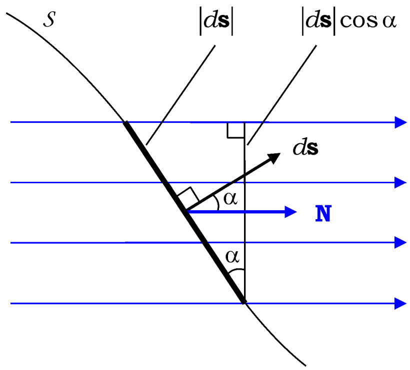


Fig. 2.10 The differential flux crossing a differential area, $d\Psi = \mathbf{N} \bullet d\mathbf{s}$.

Field lines can graphically show the spatial distribution of a given vector field, as illustrated in Fig. 2.11.

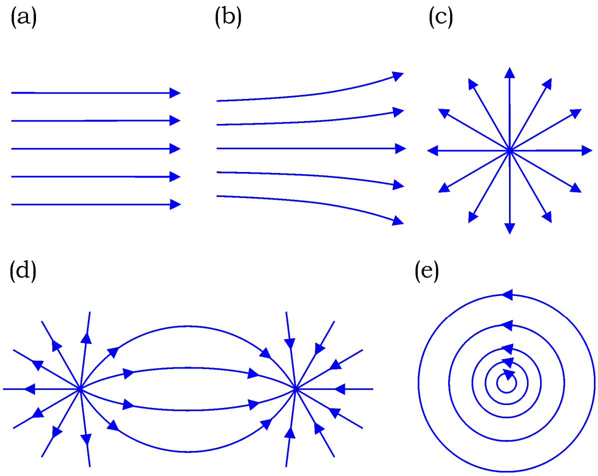


Fig. 2.11 Flux lines (a) uniform flux lines (b) slightly diverging flux lines (c) radial flux lines produced by a point divergence source (d) flux lines starting at a point source and ending at a point sink (e) concentric flux lines produced by a circulation source.

Exercise 2.10
What should be the functional form of the flux density illustrated in Fig. 2.11(c), if its total flux is conserved in three dimensional space?

Ans. $\sim (1/R^2)\,\mathbf{a}_R$

Exercise 2.11
What should be the functional form of the flux density illustrated in Fig. 2.11(e), if its circulation is conserved in three dimensional space?

Ans. $\sim (1/\rho)\,\mathbf{a}_\phi$

Review Questions with Hints

RQ 2.10 State the relation between the flux and flux density. [Eq.(2-54)]
RQ 2.11 When a flux density is defined as the quantity per unit area, identify the surface of which the unit area is a part. [Fig.2.10]
RQ 2.12 What is the significance of the dot product in Eq. (2-54)? [Fig.2.10]

2.4 Divergence and Divergence Theorem

The divergence is a vector operator acting on a flux density, or a vector field with the dimension of a scalar quantity per unit area. The divergence of a flux density is defined in such a way that the net outward flux through a closed surface is directly linked to the source enclosed by the surface. The divergence is based on a premise that the total flux coming out of a source is conserved in three dimensional space. The net outward flux through a close surface remains the same only if the surface encloses the same source. By the same token, if no source is enclosed by the surface, the amount of the outward flux is exactly the same as that of the inward flux, no net outward flux. From the definition of the divergence, we can derive the divergence theorem, which is very useful for converting a closed surface integral of a flux density into a volume integral, and vice versa.

2.4.1 Divergence of a Flux Density

The divergence of a flux density **D**, which is denoted as $div\,\mathbf{D}$, results in a scalar quantity at a point in space. The divergence of **D** involves an imaginary surface of an infinitesimal extent surrounding the given point such that

the divergence of* D *at point* p_1 *is the ratio of the net outward flux, passing through a closed surface centered at* p_1*, to the enclosed volume as the volume shrinks to a point at* p_1*.

The divergence of **D** at point p_1 is written as

$$div\,\mathbf{D} \equiv \lim_{\Delta v \to 0} \frac{\oint_S \mathbf{D} \cdot d\mathbf{s}}{\Delta v} \tag{2-55}$$

The small circle on the integration sign signifies that surface S is a closed one centered at p_1. In the above equation, Δv is the volume bounded by the closed surface S, and $d\mathbf{s}$ is the differential area vector on S, pointing out of the volume enclosed by S.

With reference to Fig. 2.12, we apply the definition of the divergence in Eq. (2-55) to a rectangular parallelepiped of an incremental volume $\Delta v = \Delta x \Delta y \Delta z$, which is centered at point p_1 with the edges parallel to the Cartesian axes. In the presence of a flux density $\mathbf{D} = D_x \mathbf{a}_x + D_y \mathbf{a}_y + D_z \mathbf{a}_z$, we need to know the values of **D** on the six faces of the parallelepiped to compute the closed surface integral of **D**. Since the parallelepiped will eventually shrink to a point at p_1, it is more convenient to expand **D** on the face in Taylor series centered at p_1. Separating the closed surface integral in Eq. (2-55) into six parts, which are the surface integrals over the six faces of the parallelepiped, we have

$$\begin{aligned} &\lim_{\Delta v \to 0} \oint_S \mathbf{D} \cdot d\mathbf{s} \\ &= \lim_{\Delta v \to 0} \left[\int_{S^1} \mathbf{D} \cdot d\mathbf{s} + \int_{S^2} \mathbf{D} \cdot d\mathbf{s} + \int_{S^3} \mathbf{D} \cdot d\mathbf{s} + \int_{S^4} \mathbf{D} \cdot d\mathbf{s} + \int_{S^5} \mathbf{D} \cdot d\mathbf{s} + \int_{S^6} \mathbf{D} \cdot d\mathbf{s} \right] \\ &= \lim_{\Delta v \to 0} \left[\mathbf{D}^1 \cdot d\mathbf{s}^1 + \mathbf{D}^2 \cdot d\mathbf{s}^2 + \mathbf{D}^3 \cdot d\mathbf{s}^3 + \mathbf{D}^4 \cdot d\mathbf{s}^4 + \mathbf{D}^5 \cdot d\mathbf{s}^5 + \mathbf{D}^6 \cdot d\mathbf{s}^6 \right] \end{aligned} \tag{2-56}$$

where superscript 1, for instance, denotes that $\mathbf{D}^1$ is the value of $\mathbf{D}$ at the center of face 1. Upon expanding $\mathbf{D}$ on each face in Taylor series centered at p_1, the outward fluxes through the six faces of the parallelepiped are

$$\mathbf{D}^1 \bullet d\mathbf{s}^1 = \left[\mathbf{D}_o + \frac{\Delta x}{2} \frac{\partial \mathbf{D}}{\partial x}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{s}^1, \quad d\mathbf{s}^1 = \Delta y \Delta z \; \mathbf{a}_x \tag{2-57a}$$

$$\mathbf{D}^2 \bullet d\mathbf{s}^2 = \left[\mathbf{D}_o + \frac{\Delta y}{2} \frac{\partial \mathbf{D}}{\partial y}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{s}^2, \quad d\mathbf{s}^2 = \Delta x \Delta z \; \mathbf{a}_y \tag{2-57b}$$

$$\mathbf{D}^3 \bullet d\mathbf{s}^3 = \left[\mathbf{D}_o - \frac{\Delta x}{2} \frac{\partial \mathbf{D}}{\partial x}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{s}^3, \quad d\mathbf{s}^3 = -\Delta y \Delta z \; \mathbf{a}_x \tag{2-57c}$$

$$\mathbf{D}^4 \bullet d\mathbf{s}^4 = \left[\mathbf{D}_o - \frac{\Delta y}{2} \frac{\partial \mathbf{D}}{\partial y}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{s}^4, \quad d\mathbf{s}^4 = -\Delta x \Delta z \; \mathbf{a}_y \tag{2-57d}$$

$$\mathbf{D}^5 \bullet d\mathbf{s}^5 = \left[\mathbf{D}_o + \frac{\Delta z}{2} \frac{\partial \mathbf{D}}{\partial z}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{s}^5, \quad d\mathbf{s}^5 = \Delta x \Delta y \; \mathbf{a}_z \tag{2-57e}$$

$$\mathbf{D}^6 \bullet d\mathbf{s}^6 = \left[\mathbf{D}_o - \frac{\Delta z}{2} \frac{\partial \mathbf{D}}{\partial z}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{s}^6, \quad d\mathbf{s}^6 = -\Delta x \Delta y \; \mathbf{a}_z \tag{2-57f}$$

where the higher-order terms in the Taylor series are ignored. The minus sign inside the bracket signifies that the corresponding face is receded from point p_1 in the negative direction along the coordinate axis, whereas the minus sign in the expression for $d\mathbf{s}$ is to make $d\mathbf{s}$ point away from the enclosed volume. It should be noted, for instance, that $\mathbf{D}_o$ and the partial derivative of $\mathbf{D}$ in Eq. (2-57a) are those evaluated at point p_1, although they, when combined as in the bracket, represent $\mathbf{D}^1$ at the center of face 1. Combining the six expressions given in Eq. (2-57) we have

$$\begin{aligned} &\mathbf{D}^1 \bullet d\mathbf{s}^1 + \mathbf{D}^2 \bullet d\mathbf{s}^2 + \mathbf{D}^3 \bullet d\mathbf{s}^3 + \mathbf{D}^4 \bullet d\mathbf{s}^4 + \mathbf{D}^5 \bullet d\mathbf{s}^5 + \mathbf{D}^6 \bullet d\mathbf{s}^6 \\ &= \Delta x \Delta y \Delta z \left[\frac{\partial \mathbf{D}}{\partial x}\bigg|_{\text{at } p_1} \bullet \mathbf{a}_x + \frac{\partial \mathbf{D}}{\partial y}\bigg|_{\text{at } p_1} \bullet \mathbf{a}_y + \frac{\partial \mathbf{D}}{\partial z}\bigg|_{\text{at } p_1} \bullet \mathbf{a}_z \right] \\ &= \Delta x \Delta y \Delta z \left[\frac{\partial D_x}{\partial x}\bigg|_{\text{at } p_1} + \frac{\partial D_y}{\partial y}\bigg|_{\text{at } p_1} + \frac{\partial D_z}{\partial z}\bigg|_{\text{at } p_1} \right] \end{aligned} \tag{2-58}$$

where we have used $\partial \mathbf{D} / \partial x = (\partial D_x / \partial x)\mathbf{a}_x + (\partial D_y / \partial x)\mathbf{a}_y + (\partial D_z / \partial x)\mathbf{a}_z$, and the like.

Substituting Eq. (2-56), Eq. (2-58) and the differential volume $\Delta v = \Delta x \Delta y \Delta z$ into Eq. (2-55), the divergence of $\mathbf{D}$ in Cartesian coordinates is

$$\boxed{div\,\mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}} \tag{2-59}$$

The divergence of **D** is a scalar quantity specific to the point in space where the partial derivatives of **D** are evaluated. The del operator enables us to express the divergence of **D** as

$$\boxed{div\,\mathbf{D} = \nabla \bullet \mathbf{D}} \tag{2-60}$$

Following the same procedure, we can obtain the expression for $div\,\mathbf{D}$ in other coordinates. The divergence of **D** in the general coordinates (u, v, w) is

$$\boxed{\nabla \bullet \mathbf{D} = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial u}(h_2 h_3 D_u) + \frac{\partial}{\partial v}(h_1 h_3 D_v) + \frac{\partial}{\partial w}(h_1 h_2 D_w)\right]} \tag{2-61}$$

where h_1, h_2, and h_3 are the metric coefficients as in Eq. (2-45).

In the three coordinate systems, the divergence of **D** is written as follows:

$$\boxed{\nabla \bullet \mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}} \quad \text{(Cartesian)} \tag{2-62a}$$

$$\boxed{\nabla \bullet \mathbf{D} = \frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho D_\rho) + \frac{1}{\rho}\frac{\partial D_\phi}{\partial \phi} + \frac{\partial D_z}{\partial z}} \quad \text{(cylindrical)} \tag{2-62b}$$

$$\boxed{\nabla \bullet \mathbf{D} = \frac{1}{R^2}\frac{\partial}{\partial R}(R^2 D_R) + \frac{1}{R\sin\theta}\frac{\partial}{\partial \theta}(D_\theta \sin\theta) + \frac{1}{R\sin\theta}\frac{\partial D_\phi}{\partial \phi}} \tag{2-62c}$$

(spherical)

Although the del operator is defined in Cartesian coordinates only, the symbol $\nabla \bullet \mathbf{D}$ is used to represent $div\,\mathbf{D}$ in cylindrical and spherical coordinates as well, without suggesting a dot product between ∇ and **D**.

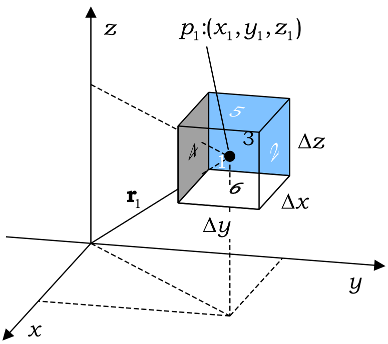


Fig. 2.12 A rectangular parallelepiped with a differential volume.

Example 2-10
In the region $R > 0$ in spherical coordinates, find the divergence of the following vector fields:

(a) $\mathbf{D} = \frac{1}{R^2}\mathbf{a}_R$, and

(b) $\mathbf{A} = \frac{1}{R}\mathbf{a}_R$.

Solution
(a) From (2-62c)

$$\nabla \bullet \mathbf{D} = \frac{1}{R^2}\frac{\partial}{\partial R}(R^2 D_R) + \frac{1}{R\sin\theta}\frac{\partial}{\partial\theta}(D_\theta \sin\theta) + \frac{1}{R\sin\theta}\frac{\partial D_\phi}{\partial\phi}$$
$$= \frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2 \frac{1}{R^2}\right) = 0.$$

(b) From (2-62c)

$$\nabla \bullet \mathbf{A} = \frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2 \frac{1}{R}\right) = \frac{1}{R^2}$$

A vector field with a zero divergence is said to be a solenoidal field. The field lines of **D** and **A** will look the same as in Fig. 2.11(c), if drawn on a paper, even though they yield different divergences $\nabla \bullet \mathbf{D} = 0$ and $\nabla \bullet \mathbf{A} \neq 0$. From $\nabla \bullet \mathbf{D} = 0$ we see that there is no source or sink in the region $R > 0$. The field **D** must be produced by a point source located at the origin, because the integral of **D** over a surface enclosing the origin is nonzero. The vector field **A** is a non-solenoidal field. The field **A** must be produced by the source distributed in the region $R > 0$, as is evident from $\nabla \bullet \mathbf{A} \neq 0$.

Example 2-11
Verify the vector identity $\nabla \bullet (V\mathbf{A}) = (\nabla V) \bullet \mathbf{A} + V(\nabla \bullet \mathbf{A})$.

Solution
In Cartesian coordinates, we write

$$\nabla \bullet (V\mathbf{A}) = \left[\frac{\partial}{\partial x}(VA_x) + \frac{\partial}{\partial y}(VA_y) + \frac{\partial}{\partial z}(VA_z)\right]$$
$$= \left[\frac{\partial V}{\partial x}A_x + \frac{\partial V}{\partial y}A_y + \frac{\partial V}{\partial z}A_z\right] + \left[V\frac{\partial A_x}{\partial x} + V\frac{\partial A_y}{\partial y} + V\frac{\partial A_z}{\partial z}\right]$$
$$= \left(\frac{\partial V}{\partial x}\mathbf{a}_x + \frac{\partial V}{\partial y}\mathbf{a}_y + \frac{\partial V}{\partial z}\mathbf{a}_z\right) \bullet \left(A_x\,\mathbf{a}_x + A_y\,\mathbf{a}_y + A_z\,\mathbf{a}_z\right)$$
$$+ V\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right)$$

Thus,

$$\boxed{\nabla \cdot (V\mathbf{A}) = (\nabla V) \cdot \mathbf{A} + V(\nabla \cdot \mathbf{A})} \tag{2-63}$$

Exercise 2.12

Find the divergence of the following vector fields at point $(x, y, z) = (1, 1, \sqrt{6})$

(a) $\mathbf{A} = x^2\,\mathbf{a}_x + 3z\,\mathbf{a}_y + yz\,\mathbf{a}_z$,

(b) $\mathbf{A} = \left(3\cos\phi / R^2\right)\mathbf{a}_R + (4\sin\theta / R)\,\mathbf{a}_\theta - 2\,\mathbf{a}_\phi$.

Ans. (a) 3, (b) $\sqrt{3}/2$.

Exercise 2.13

Given a vector field $\mathbf{D}(\mathbf{r}) = \mathbf{A}\cos(\mathbf{k} \cdot \mathbf{r})$, where $\mathbf{r}$ is position vector, and $\mathbf{A}$ and $\mathbf{k}$ are constants in Cartesian coordinates, under what condition does $\mathbf{D}$ become solenoidal.

Ans. $\mathbf{A} \perp \mathbf{k}$.

2.4.2 Divergence Theorem

Although the divergence of a vector field is given by the partial derivatives of the field, it inherently involves a closed surface, enclosed volume, and the surface integral of the field. The divergence theorem follows from the definition of the divergence, stating that ***the volume integral of*** $\nabla \cdot \mathbf{D}$ ***over a volume is equal to the closed surface integral of*** $\mathbf{D}$ ***over the surface bounding the volume.*** The divergence theorem is expressed as

$$\boxed{\int_{\mathcal{V}} \nabla \cdot \mathbf{D}\, dv = \oint_{\mathcal{S}} \mathbf{D} \cdot d\mathbf{s}} \tag{2-64}$$

where $\mathcal{S}$ is the bounding surface of the volume $\mathcal{V}$, and $d\mathbf{s}$ is the differential area vector on $\mathcal{S}$, which is always directed out of the enclosed volume.

To verify the divergence theorem, let us consider a finite volume $\mathcal{V}$ as shown in Fig. 2.13, which is bounded by a surface $\mathcal{S}$. For the sake of argument, we subdivide $\mathcal{V}$ into a large number of infinitesimal elements of volume. Figure 2.13 shows three elements, Δv_{k-1} , Δv_k , and Δv_{k+1} , positioned next to each other with their top faces coincident with $\mathcal{S}$. We see that the element Δv_k shares its faces with the adjoining elements except for the top face. Under these conditions, if the face on the right-hand side of Δv_k is considered to belong to Δv_k , the differential area vector $d\mathbf{s}$ on the face is directed toward the right, but should be directed toward the left if the face is considered to belong to Δv_{k+1} . Here, let us consider the divergence of $\mathbf{D}$ at the center of the volume element Δv_k , which is expressed as

$$(\nabla \cdot \mathbf{D})_{\Delta v_k} = \lim_{\Delta v_k \to 0} \frac{\oint_{\Delta s_k} \mathbf{D} \cdot d\mathbf{s}}{\Delta v_k} \tag{2-65}$$

where Δs_k is the bounding surface of Δv_k. Rewriting Eq. (2-65) leads to

$$\lim_{\Delta v_k \to 0} \left[(\nabla \cdot \mathbf{D})_{\Delta v_k} \Delta v_k \right] = \lim_{\Delta s_k \to 0} \left[\oint_{\Delta s_k} \mathbf{D} \cdot d\mathbf{s} \right] \tag{2-66}$$

The term in bracket on the right-hand side of Eq. (2-66) represents the net flux coming out of Δv_k. In view of the principle of conservation of the flux, this flux will eventually cross the outer surface $\mathcal{S}$ with no change in the amount. Therefore, the total flux passing through the boundary surface $\mathcal{S}$ is equal to the sum of the fluxes coming out of all the individual elements of volume. The net outward flux through $\mathcal{S}$ is written from Eq. (2-66) as

$$\lim_{\substack{N \to \infty \\ \Delta v_k \to 0}} \left[\sum_{k=1}^{k=N} (\nabla \cdot \mathbf{D})_{\Delta v_k} \Delta v_k \right] = \lim_{\substack{N \to \infty \\ \Delta s_k \to 0}} \left[\sum_{k=1}^{k=N} \oint_{\Delta s_k} \mathbf{D} \cdot d\mathbf{s} \right] \tag{2-67}$$

The left-hand side of Eq. (2-67) just equals the volume integral of $\nabla \cdot \mathbf{D}$ over $\mathcal{V}$, whereas the right-hand side equals the surface integral of $\mathbf{D}$ over all the faces of all the elements of volume. If a face is shared by two adjoining elements of volume, it contributes nothing to the surface integral because of the opposite directions of $d\mathbf{s}$ on the face. As a consequence, the right-hand side of Eq. (2-67) reduces to the surface integral of $\mathbf{D}$ over the boundary surface $\mathcal{S}$. The divergence theorem is therefore verified.

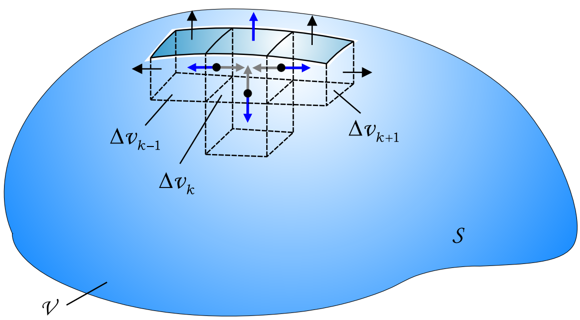


Fig. 2.13 A finite volume $\mathcal{V}$ with a boundary surface $\mathcal{S}$ is subdivided into many infinitesimal elements of volume.

Example 2-12

For a flux density $\mathbf{A(r)} = x^2 y\, \mathbf{a}_x - x^2 y\, \mathbf{a}_y + z\, \mathbf{a}_z$, verify the divergence theorem over a cube two units on a side centered at the origin as shown in Fig. 2.14.

Solution

The surface integral of $\mathbf{A}$ over each face of the cube is as follows:

$$\int_{S_1} \mathbf{A} \cdot d\mathbf{s} = \int_{y=-1}^{y=1}\int_{z=-1}^{z=1} \left(x^2 y\,\mathbf{a}_x - x^2 y\,\mathbf{a}_y + z\,\mathbf{a}_z\right) \cdot (dydz\,\mathbf{a}_x) \qquad \text{(face at } x = 1\text{)}$$
$$= (1)^2 \int_{y=-1}^{y=1} ydy \int_{z=-1}^{z=1} dz = 0$$

$$\int_{S_2} \mathbf{A} \cdot d\mathbf{s} = \int_{y=-1}^{y=1}\int_{z=-1}^{z=1} \mathbf{A} \cdot (-dydz\,\mathbf{a}_x) = -(-1)^2 \int_{y=-1}^{y=1} ydy \int_{z=-1}^{z=1} dz = 0 \quad (x = -1)$$

$$\int_{S_3} \mathbf{A} \cdot d\mathbf{s} = \int_{x=-1}^{x=1}\int_{z=-1}^{z=1} \mathbf{A} \cdot (dxdz\,\mathbf{a}_y) = -(1)\int_{x=-1}^{x=1} x^2 dx \int_{z=-1}^{z=1} dz = -\frac{4}{3} \quad (y = 1)$$

$$\int_{S_4} \mathbf{A} \cdot d\mathbf{s} = \int_{x=-1}^{x=1}\int_{z=-1}^{z=1} \mathbf{A} \cdot (-dxdz\,\mathbf{a}_y) = (-1)\int_{x=-1}^{x=1} x^2 dx \int_{z=-1}^{z=1} dz = -\frac{4}{3} \quad (y = -1)$$

$$\int_{S_5} \mathbf{A} \cdot d\mathbf{s} = \int_{x=-1}^{x=1}\int_{y=-1}^{y=1} \mathbf{A} \cdot (dxdy\,\mathbf{a}_z) = (1)\int_{x=-1}^{x=1} dx \int_{y=-1}^{y=1} dy = 4 \quad (z = 1)$$

$$\int_{S_6} \mathbf{A} \cdot d\mathbf{s} = \int_{x=-1}^{x=1}\int_{y=-1}^{y=1} \mathbf{A} \cdot (-dxdy\,\mathbf{a}_z) = -(-1)\int_{x=-1}^{x=1} dx \int_{y=-1}^{y=1} dy = 4 \quad (z = -1)$$

Combining the above results, the closed surface integral of $\mathbf{A}$ is

$$\oint_S \mathbf{A} \cdot d\mathbf{s} = \frac{16}{3} \tag{2-68}$$

Next, the divergence of $\mathbf{A}$ is computed as

$$\nabla \cdot \mathbf{A} = \frac{\partial}{\partial x}(x^2 y) + \frac{\partial}{\partial y}(-x^2 y) + \frac{\partial}{\partial z}(z) = 2xy - x^2 + 1$$

The volume integral of $\nabla \cdot \mathbf{A}$ over the cube is

$$\int_V \nabla \cdot \mathbf{A}\, dv = \int_{x=-1}^{x=1}\int_{y=-1}^{y=1}\int_{z=-1}^{z=1} (2xy - x^2 + 1)\, dxdydz$$
$$= 2\int_{x=-1}^{x=1} xdx \int_{y=-1}^{y=1} ydy \int_{z=-1}^{z=1} dz - \int_{x=-1}^{x=1} x^2 dx \int_{y=-1}^{y=1} dy \int_{z=-1}^{z=1} dz \tag{2-69}$$
$$+ \int_{x=-1}^{x=1} dx \int_{y=-1}^{y=1} dy \int_{z=-1}^{z=1} dz = \frac{16}{3}$$

The two results in Eqs. (2-68) and (2-69) are equal, and the divergence theorem is therefore verified.

Example 2-13

For a flux density $\mathbf{B}(\mathbf{r}) = \sin\phi\,\mathbf{a}_\phi$ given in cylindrical coordinates, verify the divergence theorem over a cylinder of radius 2 and height 2, centered at the origin, as shown in Fig. 2.14.

Solution

Surface integrals of **B** over the faces of the cylinder are

$$\int_{cylin} \mathbf{B}\cdot d\mathbf{s} = \int_{\phi=0}^{\phi=2\pi}\int_{z=-1}^{z=1} \sin\phi\, \mathbf{a}_\phi \cdot (2d\phi dz\, \mathbf{a}_\rho) = 0 \qquad \text{(cylindrical surface)}$$

$$\int_{top} \mathbf{B}\cdot d\mathbf{s} = \int_{\phi=0}^{\phi=2\pi}\int_{\rho=0}^{\rho=2} \sin\phi\, \mathbf{a}_\phi \cdot (\rho d\phi d\rho\, \mathbf{a}_z) = 0 \qquad \text{(top plate)}$$

$$\int_{bottom} \mathbf{B}\cdot d\mathbf{s} = \int_{\phi=0}^{\phi=2\pi}\int_{\rho=0}^{\rho=2} \sin\phi\, \mathbf{a}_\phi \cdot (-\rho d\phi d\rho\, \mathbf{a}_z) = 0 \qquad \text{(bottom surface)}$$

Closed surface integral of **B** is therefore

$$\oint_S \mathbf{B}\cdot d\mathbf{s} = \int_{cylind} \mathbf{B}\cdot d\mathbf{s} + \int_{top} \mathbf{B}\cdot d\mathbf{s} + \int_{bottom} \mathbf{B}\cdot d\mathbf{s} = 0 \qquad (2\text{-}70)$$

Next, the divergence of **B** is

$$\nabla\cdot\mathbf{B} = \frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho B_\rho) + \frac{1}{\rho}\frac{\partial B_\phi}{\partial\phi} + \frac{\partial B_z}{\partial z} = \frac{1}{\rho}\cos\phi \qquad (2\text{-}71)$$

Volume integral of $\nabla\cdot\mathbf{B}$ over the cylinder is

$$\int_V \nabla\cdot\mathbf{B}\, dv = \int_{\rho=0}^{\rho=2}\int_{\phi=0}^{\phi=2\pi}\int_{z=-1}^{z=1} \frac{1}{\rho}\cos\phi\, \rho\, d\rho d\phi dz = 2\sin\phi\Big|_{\phi=0}^{\phi=2\pi} = 0 \qquad (2\text{-}72)$$

Two results in Eqs. (2-70) and (2-72) are equal, and the divergence theorem is therefore verified.

Although the integral of $\nabla\cdot\mathbf{B}$ over the cylinder is zero in Eq. (2-72), $\nabla\cdot\mathbf{B}$ may not be necessarily zero inside the cylinder as can be seen from Eq. (2-71). In view of these we conclude that the flux arises from the source distributed in the region $-\pi/2<\phi<\pi/2$ inside the cylinder, and completely terminates at the sink distributed in the region $\pi/2<\phi<3\pi/2$. The divergence provides a useful way of checking the presence of the divergence source or sink distributed in a region of space.

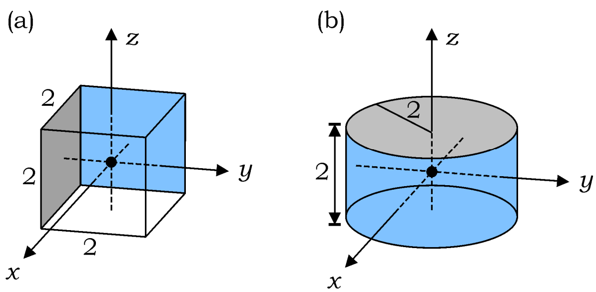


Fig. 2.14 A cube and a cylinder.

Example 2-14

Given a vector field $\mathbf{A(r)} = (1/R^2)\mathbf{a}_R$ in spherical coordinates, verify the divergence theorem over the spherical shell of an inner radius R_1 and an outer radius R_2 as shown in Fig. 2.15.

Solution

The differential area vector $d\mathbf{s}$ is always directed outward from the enclosed volume. Thus, $d\mathbf{s}$ is along the direction of $\mathbf{a}_R$ on the outer sphere, but along the direction of $-\mathbf{a}_R$ on the inner sphere.

Surface integrals of $\mathbf{A}$ over two spheres are

$$\int_{outer} \mathbf{A} \bullet d\mathbf{s} = \int_{\phi=0}^{\phi=2\pi} \int_{\theta=0}^{\theta=\pi} \frac{1}{R_2^2} \mathbf{a}_R \bullet (R_2^2 \sin\theta \, d\phi d\theta \, \mathbf{a}_R) = 4\pi \tag{2-73a}$$

$$\int_{inner} \mathbf{A} \bullet d\mathbf{s} = \int_{\phi=0}^{\phi=2\pi} \int_{\theta=0}^{\theta=\pi} \frac{1}{R_1^2} \mathbf{a}_R \bullet (-R_1^2 \sin\theta \, d\phi d\theta \, \mathbf{a}_R) = -4\pi \tag{2-73b}$$

Closed surface integral of $\mathbf{A}$ is therefore

$$\oint_S \mathbf{A} \bullet d\mathbf{s} = 4\pi - 4\pi = 0$$

Next, the divergence of $\mathbf{A}$ in spherical coordinates is

$$\nabla \bullet \mathbf{A} = \frac{1}{R^2} \frac{\partial}{\partial R} \left(R^2 \frac{1}{R^2} \right) = 0$$

The volume integral of $\nabla \bullet \mathbf{A}$ and the closed surface integral of $\mathbf{A}$ are both zero. The divergence theorem is therefore verified.

We note from Eq. (2-73) that a total flux of 4π enters the volume through the inner sphere, while a total flux of 4π leaves it through the outer sphere. Furthermore, the divergence of $\mathbf{A}$ is zero at every point in the given volume, implying that no flux is generated from inside the volume.

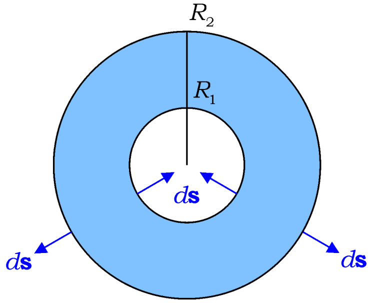


Fig. 2.15 A spherical shell bounded by two spheres.

Exercise 2.14

Under what conditions do you have $\oint_{\mathcal{S}} \mathbf{A} \cdot d\mathbf{s} = 0$, where the surface $\mathcal{S}$ is bounding a volume $\mathcal{V}$?

Ans. (1) $\mathbf{A} = 0$ in $\mathcal{V}$, (2) no source and no sink in $\mathcal{V}$, (3) all the fluxes from the source in $\mathcal{V}$ terminate at the sink in $\mathcal{V}$.

Review Questions with Hints

RQ 2.13 State the divergence of a flux density in words. [Eq.(2-55)]

RQ 2.14 Does $\nabla \cdot \mathbf{A} = 0$ at a point in space directly mean $\mathbf{A} = 0$ at the point? [Eq. (2-55)]

RQ 2.15 What is the unit of $\nabla \cdot \mathbf{D}$, if the unit of $\mathbf{D}$ is the coulomb per square meter? [Eq. (2-55)]

RQ 2.16 State the divergence theorem in words. [Eq. (2-64)]

RQ 2.17 What should be the unit of a vector field if the divergence and the divergence theorem are to be meaningful? [Eq. (2-64)]

RQ 2.18 Is $\nabla \cdot \mathbf{D}$ useful for checking the presence a discrete source other than-the distributed source or sink? If not, how can you detect a point source? [Eq.(2-64)]

2.5 Curl and Stokes's Theorem

A simple circulation source generates field lines in the form of concentric circles. For instance, an infinitely long straight wire carrying a dc current produces magnetic field lines around the wire in the form of concentric circles. As was stated earlier, the closed line integral of a vector field **H** around a closed path $\mathcal{C}$, centered at point p_1, is referred to as the circulation of **H** around $\mathcal{C}$ at p_1. As the total flux from a point divergence source is conserved in space, the circulation of a vector field due to a simple circulation source is conserved in space; that is, the circulation is the same if the path of integration encloses the same source, and zero if the path encloses no source, regardless of the shape and orientation of the path. The curl is a vector operator acting on a vector field of the dimension of a scalar quantity per unit length. The curl of a vector field is defined in such a way that it is directly linked to the circulation source enclosed by the path of integration. From the definition of the curl, we can derive the Stokes's theorem, which is very useful for converting the closed line integral of a vector field into a surface integral, and vice versa.

2.5.1 Curl of a Vector Field

The curl of a vector field **H**, which is denoted as $curl\,\mathbf{H}$, results in a vector quantity at a point in space. The curl of **H** involves an imaginary loop of an infinitesimal extent around the point in space such that

the component of $curl\,\mathbf{H}$ ***in the*** $\mathbf{a}_k$***-direction at point*** p_1 ***is the ratio of the circulation of*** $\mathbf{H}$ ***around a loop, centered at*** p_1 ***with the loop surface perpendicular to*** $\mathbf{a}_k$***, to the loop area as the loop shrinks to a point at*** p_1***.***

The k-component of $curl\,\mathbf{H}$ at point p_1 is written as

$$\boxed{curl\,\mathbf{H}\bullet\mathbf{a}_k = \lim_{\Delta s\to 0}\frac{\oint_C \mathbf{H}\bullet d\mathbf{l}}{\Delta s}} \tag{2-74}$$

where C is the closed loop centered at p_1 having $\mathbf{a}_k$ as the unit normal to the loop surface, and Δs is the area of the loop surface. The direction of $\mathbf{a}_k$ on Δs and the direction of $d\mathbf{l}$ on C follow the right-hand rule: the right thumb points in the direction of $\mathbf{a}_k$ when the fingers advance in the direction of $d\mathbf{l}$.

The curl of $\mathbf{H}$ ***at point*** p_1 ***is a vector whose magnitude is the maximum circulation of*** $\mathbf{H}$ ***per unit area of the loop surface, as the loop shrinks to a point at*** p_1***, and whose unit vector is normal to the loop surface oriented for such maximum.***

The curl of $\mathbf{H}$ at point p_1 is expressed as

$$\boxed{curl\,\mathbf{H} = \lim_{\Delta s\to 0}\frac{\oint_C \mathbf{H}\bullet d\mathbf{l}}{\Delta s}\mathbf{a}_n} \tag{2-75}$$

where C is the closed loop centered at p_1 oriented for the maximum circulation per unit area, Δs is the loop area, and $\mathbf{a}_n$ is the unit normal to the loop surface. Again, the direction of $\mathbf{a}_n$ on Δs and the direction of $d\mathbf{l}$ on C follow the right-hand rule.

The $curl\,\mathbf{H}$ is a vector, and is thus expanded in component form as

$$curl\,\mathbf{H} = (curl\,\mathbf{H})_x\mathbf{a}_x + (curl\,\mathbf{H})_y\mathbf{a}_y + (curl\,\mathbf{H})_z\mathbf{a}_z$$

The x-component is equal to the projection of the $curl\,\mathbf{H}$ onto the x-axis, that is, $(curl\,\mathbf{H})_x = (curl\,\mathbf{H})\bullet\mathbf{a}_x$, and of course the circulation of $\mathbf{H}$ per unit area of the yz-plane as the loop area shrinks to zero.

With reference to a rectangular loop of width Δw and height Δh centered at point p_1 as shown in Fig. 2.16, we obtain the component of $curl\,\mathbf{H}$ in the direction of $\mathbf{a}_k = \sin\alpha\,\mathbf{a}_x + \cos\alpha\,\mathbf{a}_z$. In the limit as the loop tends to zero, we write the circulation of $\mathbf{H}$ around the rectangular loop C as

$$\lim_{\Delta s\to 0}\left[\oint_C \mathbf{H}\bullet d\mathbf{l}\right] = \lim_{\Delta s\to 0}\left[\mathbf{H}^1\bullet d\mathbf{l}^1 + \mathbf{H}^2\bullet d\mathbf{l}^2 + \mathbf{H}^3\bullet d\mathbf{l}^3 + \mathbf{H}^4\bullet d\mathbf{l}^4\right] \tag{2-76}$$

where superscript 1 stands for side ① of loop C on which the vector field is $\mathbf{H} = \mathbf{H}^1$. Since the loop will eventually shrink to a point at p_1, it is more convenient to expand $\mathbf{H}$ on the side of the loop in Taylor series centered at p_1. With the help of the Taylor series expansion of $\mathbf{H}$, the terms on the right-hand side of Eq. (2-76) are written as

$$\mathbf{H}^1 \bullet d\mathbf{l}^1 = \left[\mathbf{H}_o + \frac{\Delta w}{2} \frac{\partial \mathbf{H}}{\partial y}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{l}^1, \qquad d\mathbf{l}^1 = -\Delta x\, \mathbf{a}_x + \Delta z\, \mathbf{a}_z \quad \text{(2-77a)}$$

$$\mathbf{H}^2 \bullet d\mathbf{l}^2 = \left[\mathbf{H}_o - \frac{\Delta x}{2} \frac{\partial \mathbf{H}}{\partial x}\bigg|_{\text{at } p_1} + \frac{\Delta z}{2} \frac{\partial \mathbf{H}}{\partial z}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{l}^2, \; d\mathbf{l}^2 = -\Delta w\, \mathbf{a}_y \quad \text{(2-77b)}$$

$$\mathbf{H}^3 \bullet d\mathbf{l}^3 = \left[\mathbf{H}_o - \frac{\Delta w}{2} \frac{\partial \mathbf{H}}{\partial y}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{l}^3, \qquad d\mathbf{l}^3 = +\Delta x\, \mathbf{a}_x - \Delta z\, \mathbf{a}_z \quad \text{(2-77c)}$$

$$\mathbf{H}^4 \bullet d\mathbf{l}^4 = \left[\mathbf{H}_o + \frac{\Delta x}{2} \frac{\partial \mathbf{H}}{\partial x}\bigg|_{\text{at } p_1} - \frac{\Delta z}{2} \frac{\partial \mathbf{H}}{\partial z}\bigg|_{\text{at } p_1} \right] \bullet d\mathbf{l}^4, \; d\mathbf{l}^4 = \Delta w\, \mathbf{a}_y \quad \text{(2-77d)}$$

It should be noted, for instance, that $\mathbf{H}_o$ and the partial derivatives of $\mathbf{H}$ in Eq. (2-77d) are those evaluated at point p_1, although they, when combined as in the bracket, represent $\mathbf{H}^4$ on side ④ of the loop. Combining the expressions in Eq. (2-77), we obtain

$$\begin{aligned}
&(\mathbf{H}^1 \bullet d\mathbf{l}^1 + \mathbf{H}^3 \bullet d\mathbf{l}^3) + (\mathbf{H}^2 \bullet d\mathbf{l}^2 + \mathbf{H}^4 \bullet d\mathbf{l}^4) \\
&= -\Delta w \Delta x \frac{\partial \mathbf{H}}{\partial y}\bigg|_{p_1} \bullet \mathbf{a}_x + \Delta w\, \Delta z \frac{\partial \mathbf{H}}{\partial y}\bigg|_{p_1} \bullet \mathbf{a}_z + \Delta w \Delta x \frac{\partial \mathbf{H}}{\partial x}\bigg|_{p_1} \bullet \mathbf{a}_y - \Delta w \Delta z \frac{\partial \mathbf{H}}{\partial z}\bigg|_{p_1} \bullet \mathbf{a}_y \\
&= \Delta w \Delta z \left[\frac{\partial H_z}{\partial y}\bigg|_{p_1} - \frac{\partial H_y}{\partial z}\bigg|_{p_1} \right] + \Delta w \Delta x \left[\frac{\partial H_y}{\partial x}\bigg|_{p_1} - \frac{\partial H_x}{\partial y}\bigg|_{p_1} \right]
\end{aligned} \quad \text{(2-78)}$$

where we have used $\partial \mathbf{H} / \partial x = (\partial H_x / \partial x)\,\mathbf{a}_x + (\partial H_y / \partial x)\,\mathbf{a}_y + (\partial H_z / \partial x)\,\mathbf{a}_z$, and the like. Substituting Eqs. (2-76) and (2-78) into Eq. (2-74), the *circulation* of $\mathbf{H}$ per unit area of the plane perpendicular to the direction of $\mathbf{a}_k$, which we denote as $(CIR\,\mathbf{H})_k$, is

$$\begin{aligned}
(CIR\,\mathbf{H})_k &= \underset{\Delta s \to 0}{\text{limit}} \left\{ \frac{\Delta z}{\Delta h} \left(\frac{\partial H_z}{\partial y}\bigg|_{p_1} - \frac{\partial H_y}{\partial z}\bigg|_{p_1} \right) + \frac{\Delta x}{\Delta h} \left(\frac{\partial H_y}{\partial x}\bigg|_{p_1} - \frac{\partial H_x}{\partial y}\bigg|_{p_1} \right) \right\} \\
&= \sin\alpha \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \right) + \cos\alpha \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \right)
\end{aligned} \quad \text{(2-79)}$$

where we used $\Delta s = \Delta w \Delta h$ for the surface area of the loop, and the relations $\Delta z / \Delta h = \sin\alpha$ and $\Delta x / \Delta h = \cos\alpha$ obtained from Fig. 2.16. For the moment, we use the notation $(CIR\,\mathbf{H})_k$ to represent the circulation of **H**, per unit area, around the loop with the loop surface perpendicular to $\mathbf{a}_k$, and the notation $(curl\,\mathbf{H})_k$ to represent the k-component of $curl\,\mathbf{H}$ that is obtained from the projection of $curl\,\mathbf{H}$ in the direction of $\mathbf{a}_k$.

If $\alpha = 90^o$, the unit normal to the loop surface is along the x-axis; that is, $\mathbf{a}_k = \mathbf{a}_x$. Under this condition Eq. (2-79) becomes

$$\boxed{(CIR\,\mathbf{H})_x = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}} \tag{2-80a}$$

This is the circulation of **H** per unit area of the $x = x_1$ plane at p_1. Similarly, if $\alpha = 0^o$, the unit surface normal is along the z-axis; that is, $\mathbf{a}_k = \mathbf{a}_z$. Under this condition Eq. (2-79) becomes

$$\boxed{(CIR\,\mathbf{H})_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}} \tag{2-80b}$$

This is the circulation of **H** per unit area of the $z = z_1$ plane at p_1. The same procedure can be followed to obtain

$$\boxed{(CIR\,\mathbf{H})_y = \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}} \tag{2-80c}$$

Next, in view of Eq. (2-80), we rewrite Eq. (2-79) as

$$(CIR\,\mathbf{H})_k = (\mathbf{a}_x \cdot \mathbf{a}_k)(CIR\,\mathbf{H})_x + (\mathbf{a}_z \cdot \mathbf{a}_k)(CIR\,\mathbf{H})_z$$

where we used $\mathbf{a}_x \cdot \mathbf{a}_k = \sin\alpha$ and $\mathbf{a}_z \cdot \mathbf{a}_k = \cos\alpha$, which can be obtained from Fig. 2.16. Extending the above expression to an arbitrary unit vector $\mathbf{a}_k$, we express $(CIR\,\mathbf{H})_k$ as

$$\begin{aligned}(CIR\,\mathbf{H})_k &= (CIR\,\mathbf{H})_x \mathbf{a}_x \cdot \mathbf{a}_k + (CIR\,\mathbf{H})_y \mathbf{a}_y \cdot \mathbf{a}_k + (CIR\,\mathbf{H})_z \mathbf{a}_z \cdot \mathbf{a}_k \\ &= \left[(CIR\,\mathbf{H})_x \mathbf{a}_x + (CIR\,\mathbf{H})_y \mathbf{a}_y + (CIR\,\mathbf{H})_z \mathbf{a}_z\right] \cdot \mathbf{a}_k\end{aligned} \tag{2-81}$$

Again, the left-hand side of Eq. (2-81) is the circulation of **H** per unit area of the plane perpendicular to an arbitrary unit vector $\mathbf{a}_k$. Suppose the bracket in Eq. (2-81) represents a vector in the direction of a unit vector $\mathbf{a}_n$. In view of the fact that the right-hand side of Eq. (2-81) is maximum for $\mathbf{a}_k = \mathbf{a}_n$, we identify the bracket in Eq. (2-81) with ***the maximum circulation of* H *per unit area***, occurring in the direction of $\mathbf{a}_n$, which we call $curl\,\mathbf{H}$. We also note that the circulation of **H** per unit area of the plane perpendicular to $\mathbf{a}_x$, or $(CIR\ \mathbf{H})_x$, is the x-component of $curl\,\mathbf{H}$, and so on.

From Eq. (2-80), the component form of $curl\,\mathbf{H}$ in Cartesian coordinates is

$$\boxed{curl\ \mathbf{H} = \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right)\mathbf{a}_x + \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}\right)\mathbf{a}_y + \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right)\mathbf{a}_z} \tag{2-82}$$

The del operator allows us to write $curl\,\mathbf{H}$ in determinant form as

$$curl\ \mathbf{H} = \nabla \times \mathbf{H} = \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix}$$

In the general coordinates, (u, v, w), the curl of $\mathbf{H}$ is expressed as

$$\boxed{\nabla \times \mathbf{H} = \frac{1}{h_1 h_2 h_3} \begin{vmatrix} h_1 \mathbf{a}_u & h_2 \mathbf{a}_v & h_3 \mathbf{a}_w \\ \dfrac{\partial}{\partial u} & \dfrac{\partial}{\partial v} & \dfrac{\partial}{\partial w} \\ h_1 H_u & h_2 H_v & h_3 H_w \end{vmatrix}} \tag{2-83}$$

where h_1, h_2, and h_3 are the metric coefficients as given in Eq. (2-45). In the three coordinate systems, the curl of $\mathbf{H}$ is expressed as follows:

$$\boxed{\nabla \times \mathbf{H} = \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix}} \quad \text{(Cartesian)} \tag{2-84a}$$

$$\boxed{\nabla \times \mathbf{H} = \frac{1}{\rho} \begin{vmatrix} \mathbf{a}_\rho & \rho\,\mathbf{a}_\phi & \mathbf{a}_z \\ \dfrac{\partial}{\partial \rho} & \dfrac{\partial}{\partial \phi} & \dfrac{\partial}{\partial z} \\ H_\rho & \rho H_\phi & H_z \end{vmatrix}} \quad \text{(cylindrical)} \tag{2-84b}$$

$$\boxed{\nabla \times \mathbf{H} = \frac{1}{R^2 \sin\theta} \begin{vmatrix} \mathbf{a}_R & R\,\mathbf{a}_\theta & R\sin\theta\,\mathbf{a}_\phi \\ \dfrac{\partial}{\partial R} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial \phi} \\ H_R & R H_\theta & R\sin\theta\, H_\phi \end{vmatrix}} \quad \text{(spherical)} \tag{2-84c}$$

In cylindrical and spherical coordinates, the curl of $\mathbf{H}$ is also denoted as $\nabla \times \mathbf{H}$, without suggesting a cross product between ∇ and $\mathbf{H}$.

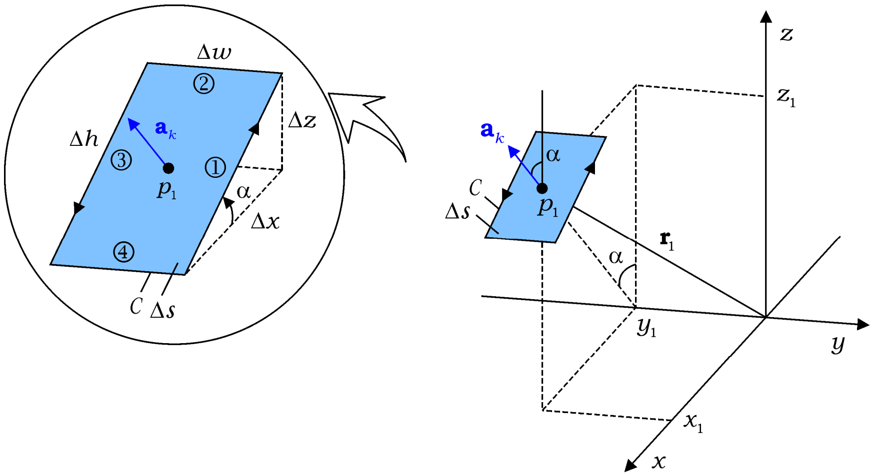


Fig. 2.16 A rectangular loop for the calculation of $(curl\,\mathbf{H})\bullet\mathbf{a}_k$.

Example 2-15
In the region $\rho > 0$ in cylindrical coordinates, find the curl of the following vector fields: (a) $\mathbf{F}_1 = \dfrac{1}{\rho^2}\mathbf{a}_\phi$, (b) $\mathbf{F}_2 = \dfrac{1}{\rho}\mathbf{a}_\phi$, (c) $\mathbf{F}_3 = \mathbf{a}_\phi$, and (d) $\mathbf{F}_4 = \rho^2\,\mathbf{a}_\phi$.

Solution
From Eq. (2-84b) we obtain the curl of the given vector fields as follows:

(a) $$\nabla\times\mathbf{F}_1 = \frac{1}{\rho}\begin{vmatrix} \mathbf{a}_\rho & \rho\,\mathbf{a}_\phi & \mathbf{a}_z \\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\phi} & \dfrac{\partial}{\partial z} \\ 0 & 1/\rho & 0 \end{vmatrix} = -\frac{1}{\rho^3}\mathbf{a}_z\,.$$

(b) $$\nabla\times\mathbf{F}_2 = \frac{1}{\rho}\begin{vmatrix} \mathbf{a}_\rho & \rho\,\mathbf{a}_\phi & \mathbf{a}_z \\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\phi} & \dfrac{\partial}{\partial z} \\ 0 & 1 & 0 \end{vmatrix} = 0\,.$$

(c) $$\nabla\times\mathbf{F}_3 = \frac{1}{\rho}\begin{vmatrix} \mathbf{a}_\rho & \rho\,\mathbf{a}_\phi & \mathbf{a}_z \\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\phi} & \dfrac{\partial}{\partial z} \\ 0 & \rho & 0 \end{vmatrix} = \frac{1}{\rho}\mathbf{a}_z\,.$$

$$\text{(d)}\quad \nabla\times\mathbf{F}_4 = \frac{1}{\rho}\begin{vmatrix} \mathbf{a}_\rho & \rho\,\mathbf{a}_\phi & \mathbf{a}_z \\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\phi} & \dfrac{\partial}{\partial z} \\ 0 & \rho^3 & 0 \end{vmatrix} = 3\rho\,\mathbf{a}_z\,.$$

The field lines of these vector fields are schematically drawn in Fig. 2.17. The curl of $\mathbf{F}_2$ is zero at every point in the region $\rho > 0$. A vector field with zero curl is called an irrotational field, or a conservative field. The field $\mathbf{F}_2$ must be produced by a circulation source located at $\rho = 0$, because $\nabla\times\mathbf{F}_2 = 0$ (no circulation source in the region $\rho > 0$), and the circulation of $\mathbf{F}_2$ around any loop enclosing the z-axis is nonzero. The other fields must be produced by the circulation source distributed in the region $\rho > 0$, because they have a nonzero curl in the region $\rho > 0$. There may or may not be a circulation source along the z-axis for these fields. For instance, from $\lim_{\delta\to 0}\oint_C \mathbf{F}_3 \bullet d\mathbf{l} = \lim_{\delta\to 0}(\mathbf{a}_\phi \bullet \mathbf{a}_\phi)2\pi\delta = 0$, where δ is the radius of C, we see that there is no circulation source for $\mathbf{F}_3$ lying along the z-axis.

A paddle wheel may be used as a detector for the curl source in a region of space: a nonzero circulation will make the wheel rotate. The direction of the curl and the direction of rotation of the paddle wheel follow the right-hand rule: the right thumb points in the direction of the curl when the fingers follow the rotation of the wheel.

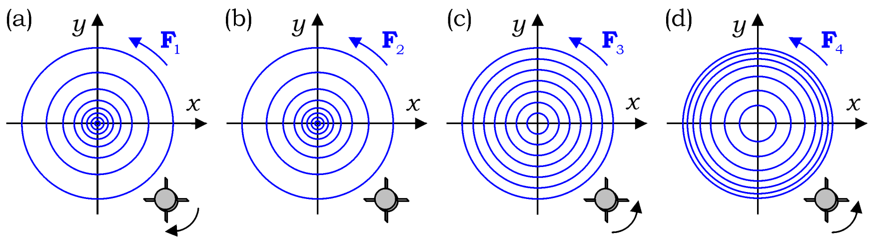


Fig. 2.17 Field lines of the four vector fields, directed along the ϕ-direction. A paddle wheel is used to detect the curl of the vector field.

Example 2-16

Verify the identity $\nabla\times(V\mathbf{A}) = (\nabla V)\times\mathbf{A} + V(\nabla\times\mathbf{A})$ in Cartesian coordinates.

Solution

The left-hand side of the identity is expanded in component form as

$$\nabla \times (V\mathbf{A}) = \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ VA_x & VA_y & VA_z \end{vmatrix}$$

$$= \mathbf{a}_x \left(\frac{\partial}{\partial y} VA_z - \frac{\partial}{\partial z} VA_y \right) + \mathbf{a}_y \left(\frac{\partial}{\partial z} VA_x - \frac{\partial}{\partial x} VA_z \right) + \mathbf{a}_z \left(\frac{\partial}{\partial x} VA_y - \frac{\partial}{\partial y} VA_x \right)$$

$$= \mathbf{a}_x \left(\frac{\partial V}{\partial y} A_z - \frac{\partial V}{\partial z} A_y \right) + \mathbf{a}_y \left(\frac{\partial V}{\partial z} A_x - \frac{\partial V}{\partial x} A_z \right) + \mathbf{a}_z \left(\frac{\partial V}{\partial x} A_y - \frac{\partial V}{\partial y} A_x \right)$$

$$+ \mathbf{a}_x V \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) + \mathbf{a}_y V \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) + \mathbf{a}_z V \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right)$$

Rewriting the above equation, we have

$$\nabla \times (V\mathbf{A}) = \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \dfrac{\partial V}{\partial x} & \dfrac{\partial V}{\partial y} & \dfrac{\partial V}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} + V \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} = (\nabla V) \times \mathbf{A} + V(\nabla \times \mathbf{A})$$

The vector identity is therefore verified:

$$\boxed{\nabla \times (V\mathbf{A}) = (\nabla V) \times \mathbf{A} + V(\nabla \times \mathbf{A}).} \tag{2-85}$$

Exercise 2.15

Find the curl of a vector field $\mathbf{H}(\mathbf{r}) = \mathbf{A}\sin(\mathbf{k}\bullet\mathbf{r})$, where $\mathbf{r}$ is position vector, and $\mathbf{A}$ and $\mathbf{k}$ are constants in Cartesian coordinates.

Ans. $\nabla \times \mathbf{H} = (\mathbf{k} \times \mathbf{A})\cos(\mathbf{k}\bullet\mathbf{r})$.

Exercise 2.16

With reference to $\mathbf{F}_2$ in **Example 2-15**, assuming a unit circulation source produces a unit circulation of $\mathbf{F}_2$, find the magnitude of the source at $\rho = 0$.

Ans. $\lim_{\delta \to 0} \oint_C \mathbf{F}_2 \bullet d\mathbf{l} = \lim_{\delta \to 0} \frac{1}{\delta} \mathbf{a}_\phi \bullet \mathbf{a}_\phi 2\pi\delta = 2\pi$.

2.5.2 Stokes's Theorem

Although the curl of a vector field is given by the partial derivatives of the field, it inherently involves a closed loop, loop area, and the closed line integral of the field. The Stokes's theorem follows from the definition of the curl, stating that ***the surface integral of*** $\nabla\times\mathbf{H}$ ***over an open surface is equal to the closed line integral of*** $\mathbf{H}$ ***around the loop bounding the surface.*** The Stokes's theorem is expressed as

$$\boxed{\int_{S}(\nabla\times\mathbf{H})\cdot d\mathbf{s}=\oint_{C}\mathbf{H}\cdot d\mathbf{l}} \tag{2-86}$$

where the surface S is bounded by the closed loop C. The direction of $d\mathbf{l}$ on C and the direction of $d\mathbf{s}$ on S are governed by the right-hand rule: the right thumb points in the direction of $d\mathbf{s}$ when the four fingers advance in the direction of $d\mathbf{l}$.

To verify the Stokes's theorem, let us consider an open surface S as shown in Fig. 2.18, which is bounded by a contour C. For the sake of argument, we subdivide S into a large number of infinitesimal elements of surface. Figure 2.18 shows three elements, Δs_{k-1}, Δs_k, and Δs_{k+1}, which are positioned next to each other with their top sides coincident with C. When the positive direction of C and the normal direction to S, or $\mathbf{a}_s$, satisfy the right-hand rule as shown in Fig. 2.18, the unit normal to Δs_k, or $\mathbf{a}_k$, should follow the direction of $\mathbf{a}_s$. Then, according to the right-hand rule, the direction of travel on Δc_k, or the contour bounding Δs_k, should be counterclockwise. As we see in the figure, the right side of Δs_k is shared with the adjoining element Δs_{k+1}. The direction of travel on this line segment is upward if the line segment is considered to belong to Δs_k, but is downward if it is considered to belong to Δs_{k+1}. From the definition of the curl, at the center of a surface element Δs_k, we can write

$$\lim_{\Delta s_k\to 0}\left[(\nabla\times\mathbf{H})\cdot\mathbf{a}_k\Delta s_k\right]=\lim_{\Delta c_k\to 0}\left[\oint_{\Delta c_k}\mathbf{H}\cdot d\mathbf{l}\right] \tag{2-87}$$

Adding circulations of $\mathbf{H}$ on all individual surface elements, we have

$$\lim_{\substack{N\to\infty\\ \Delta s_k\to 0}}\left[\sum_{k=1}^{N}(\nabla\times\mathbf{H})\cdot(\mathbf{a}_k\Delta s_k)\right]=\lim_{\substack{N\to\infty\\ \Delta c_k\to 0}}\left[\sum_{k=1}^{N}\oint_{\Delta c_k}\mathbf{H}\cdot d\mathbf{l}\right] \tag{2-88}$$

From calculus we recognize the left-hand side of Eq. (2-88) as the surface integral of $\nabla\times\mathbf{H}$ over S. Meanwhile, the right-hand side of Eq. (2-88) is the sum of the line integrals of $\mathbf{H}$ along all the contours of all the surface elements. If a part of the contour is shared by two adjoining surface elements, it contributes nothing to the net line integral, because of the opposite directions of travel on the path. As a result, the right-hand side of Eq. (2-88) reduces to the closed line integral of $\mathbf{H}$ around the contour C bounding the given surface S. The Stokes's theorem is therefore verified.

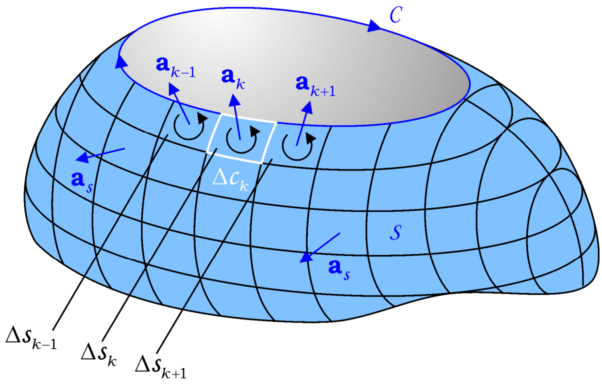


Fig. 2.18 Surface elements of surface S bounded by a contour C.

Example 2-17

For the vector field $\mathbf{A} = \rho \cos\phi\, \mathbf{a}_\phi$ given in cylindrical coordinates, verify Stokes's theorem over a quadrant of a disk of radius a as shown in Fig. 2.19.

Solution

Curl of $\mathbf{A}$ in cylindrical coordinates is

$$\nabla \times \mathbf{A} = \frac{1}{\rho}\begin{vmatrix} \mathbf{a}_\rho & \rho\,\mathbf{a}_\phi & \mathbf{a}_z \\ \dfrac{\partial}{\partial \rho} & \dfrac{\partial}{\partial \phi} & \dfrac{\partial}{\partial z} \\ 0 & \rho^2 \cos\phi & 0 \end{vmatrix} = 2\cos\phi\, \mathbf{a}_z$$

Integrating $\nabla \times \mathbf{A}$ over surface S, we have

$$\int_S (\nabla \times \mathbf{A}) \cdot d\mathbf{s} = \int_{\rho=0}^{\rho=a} \int_{\phi=0}^{\phi=\pi/2} (2\cos\phi\, \mathbf{a}_z) \cdot (\rho\, d\rho\, d\phi\, \mathbf{a}_z) = a^2 \qquad (2\text{-}89)$$

Next, separating the circulation of $\mathbf{A}$ around C into three parts, we have

$$\oint_C \mathbf{A} \cdot d\mathbf{l} = \int_{①} \mathbf{A} \cdot d\mathbf{l} + \int_{②} \mathbf{A} \cdot d\mathbf{l} + \int_{③} \mathbf{A} \cdot d\mathbf{l}$$

where

$$\int_{①} \mathbf{A} \cdot d\mathbf{l} = \int_{\rho=0}^{\rho=a} (\rho \cos 0^o\, \mathbf{a}_\phi) \cdot (d\rho\, \mathbf{a}_\rho) = 0$$

$$\int_{②} \mathbf{A} \cdot d\mathbf{l} = \int_{\phi=0}^{\phi=\pi/2} (a \cos\phi\, \mathbf{a}_\phi) \cdot (a\, d\phi\, \mathbf{a}_\phi) = a^2$$

$$\int_{③} \mathbf{A} \cdot d\mathbf{l} = \int_{\rho=a}^{\rho=0} (\rho \cos 90^o\, \mathbf{a}_\phi) \cdot (d\rho\, \mathbf{a}_\rho) = 0$$

The closed line integral is therefore

$$\oint_C \mathbf{A} \cdot d\mathbf{l} = a^2 \tag{2-90}$$

The two results in Eqs. (2-89) and (2-90) are equal, and the Stokes's theorem is thus verified.

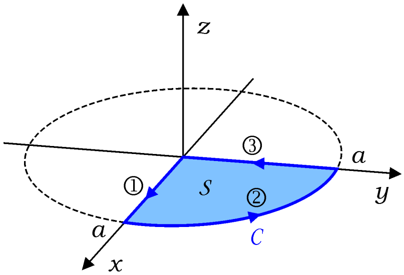


Fig. 2.19 A quadrant of a disk.

Exercise 2.17
Verify the distributive law, $\nabla \times (\mathbf{A} + \mathbf{B}) = \nabla \times \mathbf{A} + \nabla \times \mathbf{B}$, by Stokes's theorem.

Ans. $\int_S [\nabla \times (\mathbf{A} + \mathbf{B})] \cdot d\mathbf{s} = \oint_C \mathbf{A} \cdot d\mathbf{l} + \oint_C \mathbf{B} \cdot d\mathbf{l} = \int_S [(\nabla \times \mathbf{A}) + (\nabla \times \mathbf{B})] \cdot d\mathbf{s}$.

Review Questions with Hints

RQ 2.19 State the curl of a vector field at a point in space in words. [Eq.(2-75)]
RQ 2.20 If $\nabla \times \mathbf{A} = 0$ at a point, does it means $\mathbf{A} = 0$ at the point? [Eq.(2-75)]
RQ 2.21 If the dimension of $\mathbf{A}$ is a scalar per unit length, what is the dimension of $curl\mathbf{A}$? [Eq.(2-75)]
RQ 2.22 How can you identify a conservative field? [Fig.2.17]
RQ 2.23 State Stokes's theorem in words. [Eq.(2-86)]
RQ 2.24 Can Stokes's theorem be applied to any vector field, regardless of the dimension of the vector field? Eq.(2-86)]
RQ 2.25 Is $\nabla \times \mathbf{H}$ useful for checking the presence of a discrete source other than the distributed circulation source? If not, how can you detect it? [Eq.(2-86)]

2.6 Dual Operations of ∇

To summarize, the vector operators such as gradient, divergence, and curl embody special arrangements of the partial derivatives such that each has certain physical significance. The gradient represents the maximum directional derivative of the scalar field at a given point in space. The existence of a distribution of divergence sources gives rise to a nonzero divergence of the vector field in the given region. Similarly, the existence of a distribution of vortex sources gives rise to a nonzero curl of the vector field in the given region. These vector operations can be easily represented by the ∇ operator acting on a scalar or a vector field.

A second del operator may act on the gradient, divergence, or curl. Although the del operator is defined in the Cartesian coordinates only, the second operation is simply interpreted as an additional gradient, divergence, or curl in the other coordinates. A dual operation of ∇ on a scalar or a vector field involves the second partial derivatives. Even if dual operations may not have physical significance, some are very useful for mathematically expressing electromagnetic processes. The dual operations of ∇, which are frequently encountered in electromagnetics, are as follows:

(a) The divergence of the gradient of a scalar field V results in a scalar field. In Cartesian coordinates, it is express as

$$\nabla \bullet \nabla V = \nabla \bullet \left(\frac{\partial V}{\partial x}\mathbf{a}_x + \frac{\partial V}{\partial y}\mathbf{a}_y + \frac{\partial V}{\partial z}\mathbf{a}_z \right) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} \tag{2-91}$$

This is called the Laplacian of V. In Cartesian coordinates, the Laplacian operator is defined as

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

The Laplacian operator in the general coordinates (u, v, w) is

$$\begin{aligned}\nabla^2 = \nabla \bullet \nabla = \frac{1}{h_1 h_2 h_3}\left[\mathbf{a}_u \frac{\partial}{\partial u}(h_2 h_3) + \mathbf{a}_v \frac{\partial}{\partial v}(h_1 h_3) + \mathbf{a}_w \frac{\partial}{\partial w}(h_1 h_2) \right] \\ \bullet \left(\mathbf{a}_u \frac{1}{h_1}\frac{\partial}{\partial u} + \mathbf{a}_v \frac{1}{h_2}\frac{\partial}{\partial v} + \mathbf{a}_w \frac{1}{h_3}\frac{\partial}{\partial w} \right)\end{aligned} \tag{2-92}$$

where the metric coefficients are as follows:

$h_1 = 1,\ \ h_2 = 1,\ \ h_3 = 1$ $\qquad (u, v, w) = (x, y, z)$ (Cartesian)

$h_1 = 1,\ \ h_2 = \rho,\ \ h_3 = 1$ $\qquad (u, v, w) = (\rho, \phi, z)$ (cylindrical)

$h_1 = 1,\ \ h_2 = R,\ \ h_3 = R\sin\theta$ $\qquad (u, v, w) = (R, \theta, \phi)$ (spherical)

The dot product precedes any differentiation in Laplacian operator. For example,

$$\mathbf{a}_u \frac{\partial}{\partial u} \bullet \mathbf{a}_v \frac{\partial}{\partial v} = \mathbf{a}_u \bullet \mathbf{a}_v \frac{\partial}{\partial u}\frac{\partial}{\partial v} \neq \mathbf{a}_u \bullet \frac{\partial}{\partial u}\left(\mathbf{a}_v \frac{\partial}{\partial v} \right)$$

In the three coordinate systems, the Laplacian of V is written as follows:

$$\boxed{\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}} \qquad \text{(Cartesian)} \tag{2-93a}$$

$$\nabla^2 V = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial V}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 V}{\partial \phi^2} + \frac{\partial^2 V}{\partial z^2} \quad \text{(cylindrical)} \tag{2-93b}$$

$$\nabla^2 V = \frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2 \frac{\partial V}{\partial R}\right) + \frac{1}{R^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial V}{\partial \theta}\right) + \frac{1}{R^2 \sin^2\theta}\frac{\partial^2 V}{\partial \phi^2} \quad \text{(spherical)} \tag{2-93c}$$

(b) The curl of the gradient of a scalar field is always zero

$$\boxed{\nabla \times (\nabla V) = 0} \tag{2-94}$$

Equation (2-94) can be readily verified by directly evaluating the gradient and the curl in Cartesian coordinates.

Alternatively, applying Stokes's theorem to the left-hand side of Eq. (2-94), we obtain

$$\int_{S} [\nabla \times (\nabla V)] \bullet d\mathbf{s} = \oint_{C} \nabla V \bullet d\mathbf{l} = \oint_{C} dV \tag{2-95}$$

We used Eq. (2-37) in Eq. (2-95). The closed line integral on the right-hand side is identically zero because the initial and terminal points are the same on the closed loop C. The integrand on the left-hand side of Eq. (2-95) should be zero at every point on surface S, because S may be arbitrary. The vector identity is thus verified.

(c) The divergence of the curl of a vector field is always zero

$$\boxed{\nabla \bullet (\nabla \times \mathbf{A}) = 0} \tag{2-96}$$

Equation (2-96) can be readily verified by direct substitution.

Alternatively, by successively applying the divergence and Stokes's theorems to the left-hand side of Eq. (2-96), we have

$$\int_{V} \nabla \bullet (\nabla \times \mathbf{A})\, dv = \oint_{S} (\nabla \times \mathbf{A}) \bullet d\mathbf{s} = \oint_{C} \mathbf{A} \bullet d\mathbf{l} \tag{2-97}$$

Since the surface S, bounding an arbitrary volume V, has no contour, $C = 0$, the closed line integral on the right-hand side of Eq. (2-97) is always zero. Since V is arbitrary, the integrand on the left-hand side of Eq. (2-97) should be zero at every point in V to satisfy the equality. The vector identity is therefore verified.

(d) The Laplacian of a vector field results in a vector field

$$\nabla^2 \mathbf{E} = \frac{\partial^2 \mathbf{E}}{\partial x^2} + \frac{\partial^2 \mathbf{E}}{\partial y^2} + \frac{\partial^2 \mathbf{E}}{\partial z^2} = (\nabla^2 E_x)\,\mathbf{a}_x + (\nabla^2 E_y)\,\mathbf{a}_y + (\nabla^2 E_z)\,\mathbf{a}_z \tag{2-98}$$

(e) The gradient of the divergence of a vector field results in a vector field

$$\nabla\nabla\bullet\mathbf{E} = \left(\mathbf{a}_x\frac{\partial}{\partial x} + \mathbf{a}_y\frac{\partial}{\partial y} + \mathbf{a}_z\frac{\partial}{\partial z}\right)\left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right)$$
$$= \left(\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_y}{\partial x\partial y} + \frac{\partial^2 E_z}{\partial x\partial z}\right)\mathbf{a}_x + \left(\frac{\partial^2 E_x}{\partial y\partial x} + \frac{\partial^2 E_y}{\partial y^2} + \frac{\partial^2 E_z}{\partial y\partial z}\right)\mathbf{a}_y \qquad (2\text{-}99)$$
$$+ \left(\frac{\partial^2 E_x}{\partial z\partial x} + \frac{\partial^2 E_y}{\partial z\partial y} + \frac{\partial^2 E_z}{\partial z^2}\right)\mathbf{a}_z$$

In many cases, it is easier to take the gradient of $\nabla\bullet\mathbf{E}$ rather than directly evaluating Eq. (2-99).

(f) The curl of the curl of a vector field results in a vector field

$$\nabla\times\nabla\times\mathbf{E} = \nabla\nabla\bullet\mathbf{E} - \nabla^2\,\mathbf{E} \qquad (2\text{-}100)$$

Equation Eq. (2-100) can be readily verified by direct substitution.

Exercise 2.18
Verify the identities in Eqs. (2-94) and (2-96) by direct substitution.

2.7 Helmholtz's Theorem

A vector field is uniquely defined in a region of space, if its divergence and its curl are both specified within the region, and the normal component of the field is specified on the boundary.

If the boundary is at infinity where the vector field diminishes to zero, the vector field is uniquely determined by its divergence and its curl only.

Vector fields can be classified into four categories according to the divergence and the curl as follows:

(1) $\nabla\bullet\mathbf{A} = 0$ and $\nabla\times\mathbf{A} = 0$. (2-101a)
(2) $\nabla\bullet\mathbf{A} = 0$ and $\nabla\times\mathbf{A} \neq 0$. (2-101b)
(3) $\nabla\bullet\mathbf{A} \neq 0$ and $\nabla\times\mathbf{A} = 0$. (2-101c)
(4) $\nabla\bullet\mathbf{A} \neq 0$ and $\nabla\times\mathbf{A} \neq 0$. (2-101d)

A vector field is said to be a solenoidal field if its divergence is zero, and said to be an irrotational field if its curl is zero.

Example 2-18
Show that a vector field $\mathbf{A}$ is a many-valued function of position, if it is defined by $\nabla\times\mathbf{A} = \mathbf{B}$ only, with no information about its divergence.

Solution

With the help of (2-94), we rewrite the equation as

$$\nabla\times\mathbf{A}+\nabla\times(\nabla V)=\mathbf{B} \qquad \rightarrow \qquad \nabla\times(\mathbf{A}+\nabla V)=\mathbf{B}$$

Since V is arbitrary, the solution $\mathbf{A}+\nabla V$ is a multi-valued function of position.

Example 2-19

Show that a vector field $\mathbf{B}$ is a many-valued function of position, if it is defined by $\nabla\bullet\mathbf{B}=0$ only, with no information about its curl.

Solution

With the help of Eq. (2-96), we rewrite the equation as

$$\nabla\bullet\mathbf{B}+\nabla\bullet(\nabla\times\mathbf{K})=0 \qquad \rightarrow \qquad \nabla\bullet(\mathbf{B}+\nabla\times\mathbf{K})=0$$

Since $\mathbf{K}$ is arbitrary, the solution $\mathbf{B}+\nabla\times\mathbf{K}$ is a multi-valued function of position.

Exercise 2.19

Determine if the following vector fields are solenoidal, irrotational, or both:
(a) $-2xy\,\mathbf{a}_x+y^2\,\mathbf{a}_y+x^2\,\mathbf{a}_z$, (b) $\cos\phi\,\mathbf{a}_\rho-\sin\phi\,\mathbf{a}_\phi$, and (c) $\mathbf{a}_R\sin(3R)/R^2$.

Ans. (a) solenoidal, (b) both, (c) irrotational.

Review Questions with Hints

RQ 2.26 State Helmholtz's theorem in words.
RQ 2.27 What is a solenoidal field? [Eq.(2-101)]
RQ 2.28 What is an irrotational field? [Eq.(2-101)]

Problems

2-1 For the ellipse $(x^2/16)+(y^2/4)=1$, determine the unit tangent vector at a point p_1:$(2, \sqrt{3}, 0)$ on the ellipse by using a parametric representation of the ellipse with a parameter

(a) t, as in $\cos t$ and $\sin t$, and
(b) x, the space coordinate.
(c) Is t in part (a) the polar angle ϕ shown in Fig. 2.20?

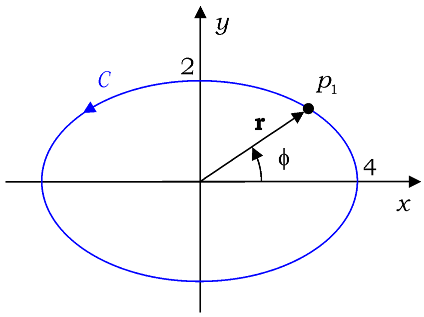


Fig. 2.20 An ellipse(Problem 2-1).

2-2 A closed loop C forms a circle of radius 0.5, centered at a point ($x = 2$, $y = 1$) in the xy-plane as shown in Fig. 2.21. Find
(a) parametric representation of C in terms of φ, and
(b) expression for $d\mathbf{l}$ on C.

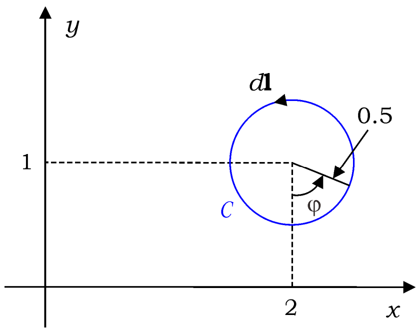


Fig. 2.21 A circle centered at point $(x = 2, y = 1)$ (Problem 2-2).

2-3 Given a vector field $\mathbf{E} = (x + y)^2(\mathbf{a}_x + \mathbf{a}_y)$, find the line integral of $\mathbf{E}$ from point A:$(-1, 0, 0)$ to point B:$(1, 2, 0)$ along a path (a) C_1, and (b) C_2 as shown in Fig. 2.22.

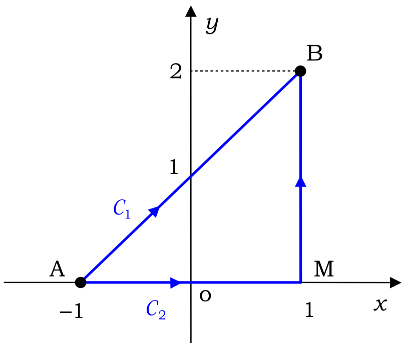


Fig. 2.22 Two paths of integration(Problem 2-3).

2-4 Find the line integral of $\mathbf{E} = y\,\mathbf{a}_x + z\,\mathbf{a}_y + x\,\mathbf{a}_z$ from point $A{:}\left(x = 1, y = 0, z = 0\right)$ to point $B{:}\left(x = 0, y = 1, z = \pi/2\right)$ along a spiral defined by $\mathbf{r}\left(t\right) = \cos t\,\mathbf{a}_x + \sin t\,\mathbf{a}_y + t\,\mathbf{a}_z$.

2-5 Given a vector field $\mathbf{A} = R\sin\theta\cos\phi\,\mathbf{a}_R + \cos\theta\cos\phi\,\mathbf{a}_\theta - \sin\phi\,\mathbf{a}_\phi$ in spherical coordinates, determine the close line integral of $\mathbf{A}$ around path C shown in Fig. 2.23.

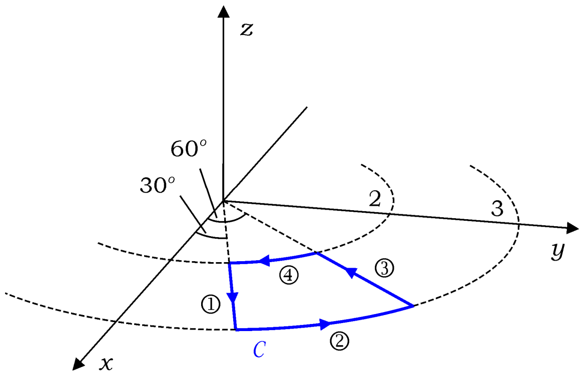


Fig. 2.23 A closed path(Problem 2-5).

2-6 Determine the closed line integral of a vector field $\mathbf{A} = \mathbf{a}_x$ around the circle that is defined by $R = 4$, $\theta = 30^o$, and $0 \le \phi \le \pi/2$ in spherical coordinates, in the direction of increasing ϕ.

2-7 Find the area of a circular strip shown in Fig. 2.24.

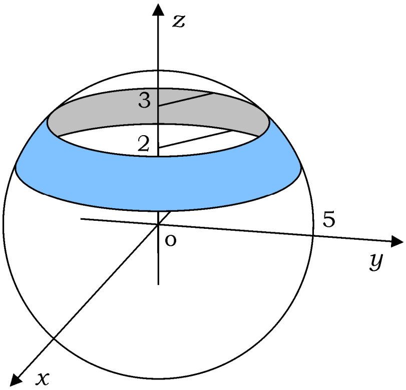


Fig. 2.24 A circular strip (Problem 2-7).

2-8 Given that $\mathbf{A} = \rho\,\mathbf{a}_\rho - 3\rho\cos\phi\,\mathbf{a}_\phi - 2z\,\mathbf{a}_z$ in cylindrical coordinates, determine the closed surface integral of **A** over the bounding surface of a half cylinder defined by $\rho \le 4$, $30^o \le \phi \le 210^o$, and $0 \le z \le 5$ as show in Fig. 1.33.

2-9 For the vector field $\mathbf{D} = \left(3/R^2\right)\mathbf{a}_R + \cos^2\phi\,\mathbf{a}_\phi$ given in spherical coordinates, determine the closed surface integral of **D** over the bounding surface of a volume defined by $0 \le R \le 4$ and $30^o \le \phi \le 60^o$ as shown in Fig, 2.25.

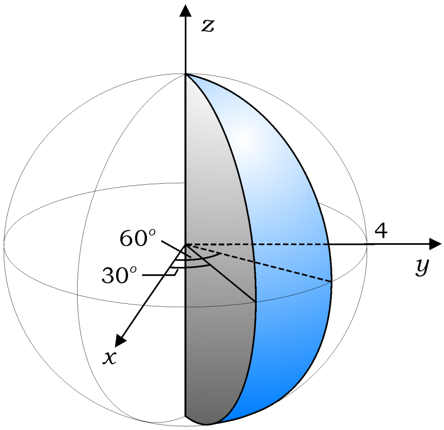


Fig. 2.25 A segment of a sphere (Problem 2-9).

2-10 Determine the gradient of the following scalar fields:

(a) $U = 3ze^{2x+y} + 10$,

(b) $V = 5\rho\sin\phi - \ln\left(z^2 + 1\right)$,

(c) $W = \sin\theta\cos\phi / R^2$.

2-11 Take the gradient of a scalar function $V(\mathbf{r}) = z$ in two different ways and compare the results following these steps:
(a) Find ∇V in Cartesian coordinates.
(b) Transform V into spherical coordinates and then take the gradient.
(c) Transform the result in (b) back into Cartesian coordinates.

2-12 Two families of curves are given by $f(x,y) = x^2 + y^2 = c_1$ and $g(x,y) = (x-6)^2 + y^2 = c_2$ in the $z = 0$ plane, where c_1 and c_2 are constants. Determine the smaller angle between the two curves at point $p{:}(3,4,0)$.

2-13 An ellipsoid is defined by $x^2 + y^2 + z^2/4 = 5$. Find, at point $p_1{:}(2, \sqrt{3}/2, 1)$ on the ellipsoid,
(a) outward unit normal vector, and
(b) expression for the tangent plane.

2-14 A family of surfaces is defined as $f(x,y,z) = x^2 + y^2 + 4z^2 = c$ in Cartesian coordinates. A curve crosses the family of surfaces at right angles, and passes through point $p{:}(2,6,32)$. Find the parametric representation of the curve by using x as parameter.

2-15 An ellipsoidal surface is defined by $x^2 + y^2 + 2z^2 = 8$ as shown in Fig. 2.26. Find
(a) parametric representation of the surface by using ρ and ϕ as parameters,
(b) differential area vector at point $p_1{:}(\sqrt{3}, 1, \sqrt{2})$ on the ellipsoid, and
(c) outward surface normal at p_1 by using the gradient, and compare its direction with that of (b).

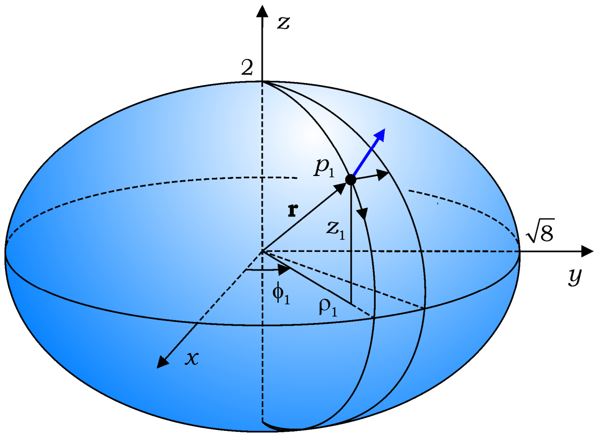


Fig. 2.26 An ellipsoid (Problem 2-15).

2-16 Consider a scalar field $f(\mathbf{r}) = \mathbf{k} \cdot \mathbf{r}$ given in Cartesian coordinates, where $\mathbf{r}$ is position vector, and $\mathbf{k} = k_x \mathbf{a}_x + k_y \mathbf{a}_y + k_z \mathbf{a}_z$ (a constant vector).

(a) Find ∇f.

(b) What is the geometric shape that is defined by $f(\mathbf{r}) = c_1$ (a constant)?

(c) What is the significance of c_1 in part (b)?

2-17 Under the condition $x \approx x_o$, a smooth function $f(x)$ can be expanded by Taylor series as $f(x) = f(x_o) + f'(x_o)(x - x_o)$, where f' is the first derivative of *f*. Verify the Taylor series by use of the gradient of *f*.

2-18 For the scalar fields *U*, *V*, and *W* given in **Problem 2-10**, determine the line integral of the gradient of each field from point $p_1:(2,2,0)$ to point $p_2:(0,0,2)$ in Cartesian coordinates along the intersection between a surface $x^2 + y^2 + 2z^2 = 8$ and the $\phi = 45^0$ plane:

(a) $\int_{p_1}^{p_2} \nabla U \cdot d\mathbf{l}$ in Cartesian coordinates,

(b) $\int_{p_1}^{p_2} \nabla V \cdot d\mathbf{l}$ in cylindrical coordinates,

(c) $\int_{p_1}^{p_2} \nabla W \cdot d\mathbf{l}$ in spherical coordinates.

2-19 Determine the divergence of the following vector fields:

(a) $\mathbf{A} = \left[x^2 + y^2 + z^2\right]^{-1/2}\left(\mathbf{a}_x + \mathbf{a}_y + \mathbf{a}_z\right)$, $\quad (x^2 + y^2 + z^2 > 0)$,

(b) $\mathbf{B} = \ln\rho\, \mathbf{a}_\rho + \rho\cos\phi\, \mathbf{a}_\phi + z^3\, \mathbf{a}_z$,

(c) $\mathbf{C} = R^{-3}\, \mathbf{a}_R + Re^{-R}\sin\theta\, \mathbf{a}_\theta + (R\sin\theta\cos\phi)^2\, \mathbf{a}_\phi$.

2-20 Take the divergence of a vector field $\mathbf{D} = x\,\mathbf{a}_x$ in two different ways and compare the results following these steps:

(a) Find $\nabla \cdot \mathbf{D}$ in Cartesian coordinates.

(b) Transform **D** into spherical coordinates and then take the divergence.

2-21 A vector field is defined as $\mathbf{A} = \rho\sin\phi\left(\sin\phi\,\mathbf{a}_\rho + \cos\phi\,\mathbf{a}_\phi\right)$ in cylindrical coordinates. Verify divergence theorem over the volume enclosed by the four surfaces $\rho = 2$, $\rho = 4$, $z = 0$, and $z = 3$, by computing

(a) closed surface integral $\oint_S \mathbf{A} \cdot d\mathbf{s}$, and

(b) volume integral $\int_V \nabla \cdot \mathbf{A}\, dv$ in cylindrical coordinates.

2-22 Given a vector field $\mathbf{D} = (\sin\phi / R)\,\mathbf{a}_R + \cos\phi\,\mathbf{a}_\phi$ in spherical coordinates, verify divergence theorem over the volume defined by $1 \le R \le 2$ and $\pi \le \phi \le 2\pi$, as shown in Fig. 2.27, by computing

(a) closed surface integral $\oint_S \mathbf{D} \cdot d\mathbf{s}$, and

(b) volume integral $\int_V \nabla \cdot \mathbf{D}\, dv$ in spherical coordinates.

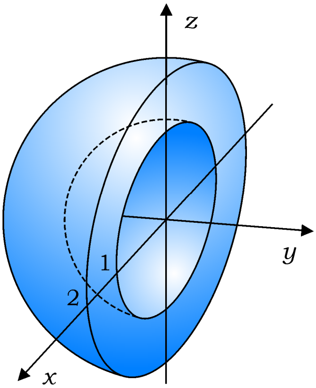


Fig. 2.27 A hemispherical shell(Problem 2-22).

2-23 A vector field is defined as $\mathbf{D} = 3z^2\,\mathbf{a}_z$ in Cartesian coordinates. Verify divergence theorem over the volume bounded by a cone of half angle $\theta = 30^o$ and the $z = 2$ plane, by computing

(a) closed surface integral $\oint_S \mathbf{D} \cdot d\mathbf{s}$, and

(b) volume integral $\int_v \nabla \cdot \mathbf{D} dv$.

2-24 Take the curl of a vector field $\mathbf{A} = x^2\,\mathbf{a}_y - z\,\mathbf{a}_z$ in two different ways and compare the results following these steps:

(a) Find $\nabla \times \mathbf{A}$ in Cartesian coordinates.

(b) Transform **A** into cylindrical coordinates and then take the curl.

2-25 Show that the line integral of a vector field **H** is independent of the path of integration, if **H** is an irrotational field.

2-26 For the vector field $\mathbf{H} = \left(x^2 + (y-2)^2\right)^{-1}\left[-(y-2)\mathbf{a}_x + x\mathbf{a}_y\right]$, find

(a) $\nabla \times \mathbf{H}$ everywhere except for the point($x = 0$, $y = 2$),

(b) $\oint_{C_1} \mathbf{H} \cdot d\mathbf{l}$, and

(c) $\oint_{C_2} \mathbf{H} \cdot d\mathbf{l}$.

Here C_1 and C_2 are concentric circles of radius 1 and 3, respectively, centered at the origin in the xy-plane.

2-27 For the vector field $\mathbf{A} = \left(x - 2y^2\right)\mathbf{a}_x + \left(2xy + y^2\right)\mathbf{a}_y$ given in Cartesian coordinates, verify Stokes's theorem over a triangle residing in the xy-plane as shown in Fig. 2.28.

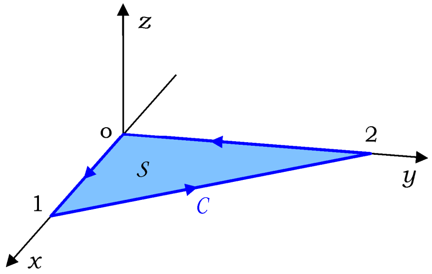


Fig. 2.28 A triangular region in the xy-plane (Problem 2-27).

2-28 Given that $\mathbf{B} = (3\rho)^{-1}\mathbf{a}_\rho + \rho\cos^2\phi\,\mathbf{a}_\phi + \rho(1+\sin\phi)\mathbf{a}_z$ in cylindrical coordinates, verify Stokes's theorem over the outer surface of a hollow cylinder, which is closed except for a circular opening on top as shown in Fig. 2.29.

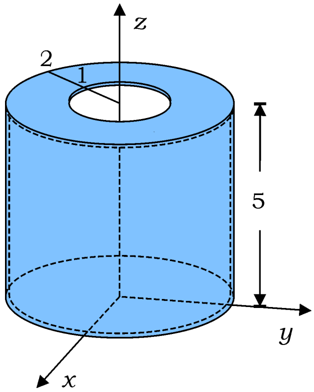


Fig. 2.29 A hollow cylinder with an opening on top (Problem 2-28).

2-29 A vector field $\mathbf{H} = \mathbf{a}_R R + \mathbf{a}_\phi 3R\sin\theta$ is given in spherical coordinates. Over the conical surface shown in Fig. 2.29, verify Stokes's theorem by computing

(a) closed line integral $\oint_C \mathbf{H} \cdot d\mathbf{l}$, and

(b) surface integral $\int_S (\nabla \times \mathbf{H}) \cdot d\mathbf{s}$.

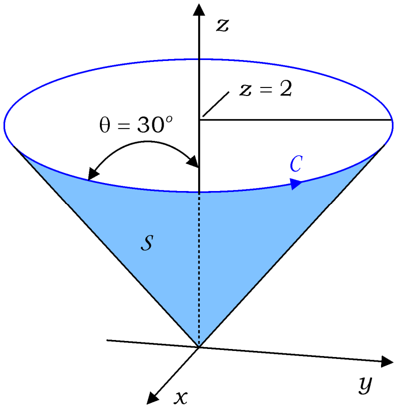


Fig. 2.30 A conical surface (Problem 2-29).

2-30 Consider the vector field $\mathbf{E}(x, y, z) = \mathbf{A}\cos(\mathbf{k}\cdot\mathbf{r})$, where $\mathbf{A}$ and $\mathbf{k}$ are constant vectors, and $\mathbf{r}$ is position vector in Cartesian coordinates. Under what condition is $\mathbf{E}$

(a) solenoidal?

(b) irrotational?

Chapter 3
Electrostatics

In the first two chapters, we discussed in detail mathematical topics, such as vector algebra, coordinate systems, and vector calculus, which provide essential tools for the study of electromagnetics. We are now ready to study the basic concepts of electromagnetics and learn about their applications. This chapter focuses on electrostatics, which deals with static electric fields induced by static electric charges. The static electric fields and charges are constant in time, although they may vary in space. In the present chapter, we will see that all discussions of static electric fields can be boiled down to the divergence and the curl of the electric field. They are two fundamental relations in the sense that they allow us to uniquely determine a static electric field in a given region of space according to Helmholtz's theorem.

After completing the discussion of steady electric currents in Chapter 4, we study magnetostatics in Chapter 5, which deals with static magnetic fields induced by steady electric currents. We will also see that all discussions of the static magnetic fields can be boiled down to the divergence and the curl of the magnetic field. Chapter 6 discusses the modification of the above two curl equations under time-varying conditions. The two modified curl equations and the two divergence equations comprise Maxwell's equations, which form the foundation of electromagnetic theory. Chapters 7 and 8 explore the electromagnetic waves formed by mutually coupled time-varying electric and magnetic fields.

The fundamental laws of electromagnetics were established through experimental observations and a generalization process. They are usually expressed in the form of mathematical formulas. This does not necessarily mean that an electromagnetic law can be derived from another through mathematical manipulations, although some auxiliary relations may be derived from the fundamental laws. Whereas the steps taken for solving an electromagnetic problem may have their origins in either physical arguments or mathematical theorems, it is important to distinguish between those origins to avoid confusion and best comprehend the underlying concepts.

Although one may consider electrostatics as a simpler case of electromagnetics, a complete mastery of electrostatics is essential not only for better understanding of general electromagnetic theory, but also for solving practical problems, such as those encountered with laser printers, Liquid Crystal Displays (LCDs), and electrostatic actuators in Microelectromechanical Systems (MEMS).

In the present chapter, our discussion starts with Coulomb's law, followed by the concept of electric fields. We then introduce Gauss's law and define electric potential, and show how they can be used for determining the electric field in a given problem. Extending our discussion of static electric fields into material media, we define electric flux density. We compute the energy stored in electric fields and examine how a capacitor can store electric energy. We will see how the divergence and the curl of a static electric field can be combined into Poisson's and Laplace's equations, and discuss applications of these equations to boundary value problems for electric potential, and thus, electric fields.

3.1 Coulomb's Law

Static electricity was known to ancient Greeks who observed that rubbing a piece of amber on fur or silk attracted straw, lint or feather. Electron is the Greek word for amber. It took many centuries before the magic revealed its connection to static electric charges. In a simple atomic model, an atom consists of a positively charged nucleus and negatively charged electrons orbiting around the nucleus. Rubbing amber against fur removes the bound electrons of the fur and imparts them to the amber, causing the fur to be positively charged and the amber to be negatively charged.

Coulomb's law is a physical law; it was established through elaborate experiments on two static electric charges separated in free space. Coulomb's law states that ***the electric force exerted on a charge is proportional to the product of two charges and inversely proportional to the square of the distance between the charges***. The mathematical expression for Coulomb's law is

$$F = \frac{1}{4\pi\varepsilon_o}\frac{q_1 q_2}{\mathcal{R}^2} \tag{3-1}$$

where q_1 and q_2 are the charges measured in coulombs [C], and $\mathcal{R}$ is the distance between the charges measured in meters [m]. The universal constant ε_o is called the permittivity of free space (or vacuum), measured in farads per meter [F/m], having the value of

$$\varepsilon_o = 8.854\times10^{-12} \qquad \text{[F/m]}$$

It is acceptable to use $\varepsilon_o \cong (1/36\pi)\times10^{-9}$ [F/m] ignoring a small error. In a material medium, ε_o is replaced by a material constant ε called a permittivity, as will be discussed in Section 3-5.

Two point charges of the same polarity repel each other along the line joining two charges, whereas two point charges of the opposite polarities attract each other along the line. Coulomb force is another name for the electric force exerted on a point charge due to the other. Using vector notation, we express the Coulomb force exerted on q_1 due to q_2 as

$$\boxed{\mathbf{F}_1 = \frac{q_1 q_2}{4\pi\varepsilon_o \mathcal{R}_{1-2}^2}\mathbf{a}_{1-2}} \quad \text{[N]} \tag{3-2}$$

In the above equation, $\mathcal{R}_{1-2}$ and $\mathbf{a}_{1-2}$ are the magnitude and the unit vector of the distance vector defined by

$$\boldsymbol{\mathcal{R}}_{1-2} = \mathbf{r}_1 - \mathbf{r}_2 \tag{3-3}$$

This is a vector drawn from the point of q_2 to the point of q_1 in space. The subscript 1-2 on $\boldsymbol{\mathcal{R}}$ suggests that $\boldsymbol{\mathcal{R}}_{1-2}$ is a vector "from 2 to 1", while mimicking the subtraction on the right-hand side of the equal sign. Throughout the text, subscript a-b denotes something from b to a, whereas subscript ab may stand for something from a to b. The unit vector of $\boldsymbol{\mathcal{R}}_{1-2}$ can be expressed in terms of two position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, that is,

$$\mathbf{a}_{1-2} = \frac{\boldsymbol{\mathcal{R}}_{1-2}}{\mathcal{R}_{1-2}} = \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{3-4}$$

Inserting Eq. (3-4) into Eq. (3-2), the Coulomb force on point charge q_1 due to point charge q_2 is expressed as follows:

$$\boxed{\mathbf{F}_1 = \frac{q_1 q_2}{4\pi\varepsilon_o}\frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3}} \quad \text{[N]} \tag{3-5}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the position vectors of q_1 and q_2, respectively.

The relation between the vectors in Eq. (3-5) is shown in Fig. 3.1, in which charge q_1 is located at point p_1 with position vector $\mathbf{r}_1$, while charge q_2 is located at point p_2 with position vector $\mathbf{r}_2$. In view of Eq. (3-5), point p_1 represents a point in space at which the electric force $\mathbf{F}_1$ is observed; it is called a field point. Meanwhile, point p_2 represents a point at which the source charge is located; it is called a source point. From now on, the distance vector refers to a vector from a source point to a field point. In this case, $\mathbf{r}_1$ or $\mathbf{r}$ stands for the position vector of the field point, whereas $\mathbf{r}_2$ or $\mathbf{r}'$ stands for the position vector of the source point. Since the Coulomb force is observed at the field point, we express those vector quantities such as $\mathbf{F}_1$, $\mathbf{r}_1$, $\boldsymbol{\mathcal{R}}_{1-2}$, and $\mathbf{a}_{1-2}$ in terms of the base vectors $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$ at the field point p_1. It should be noted that their scalar components may depend on the coordinates of both source and field points.

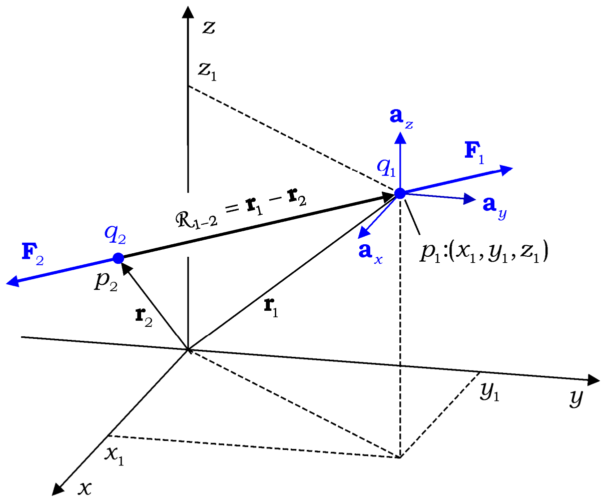


Fig. 3.1 Two point charges are at position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$.

The Coulomb force is a mutual force such that the force acting on q_2 due to q_1 has the same magnitude as $\mathbf{F}_1$ but is directed along the opposite direction, $\mathbf{F}_2 = -\mathbf{F}_1$. Expressed mathematically,

$$\mathbf{F}_2 = \frac{q_2 q_1}{4\pi\varepsilon_o} \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3} = -\mathbf{F}_1 \tag{3-6}$$

It is important to note that since $\mathbf{F}_2$ is observed at point p_2, this force should be described by the base vectors at point p_2, which is now a field point.

Example 3-1
Determine the Coulomb force acting on a point charge q_1 located at $(2,3,5)$ in Cartesian coordinates due to a point charge q_2 located at $(1,1,3)$.

Solution

Position vectors of q_1 and q_2 are, respectively,

$$\mathbf{r}_1 = 2\,\mathbf{a}_x + 3\,\mathbf{a}_y + 5\,\mathbf{a}_z$$

$$\mathbf{r}_2 = \mathbf{a}_x + \mathbf{a}_y + 3\,\mathbf{a}_z$$

Distance vector is

$$\mathcal{R}_{1-2} = \mathbf{r}_1 - \mathbf{r}_2 = \mathbf{a}_x + 2\mathbf{a}_y + 2\mathbf{a}_z$$

The magnitude and unit vector of $\mathcal{R}_{1-2}$ are

$$\mathcal{R}_{1-2} = |\mathbf{r}_1 - \mathbf{r}_2| = \sqrt{1 + 2^2 + 2^2} = 3 \tag{3-7a}$$

$$\mathbf{a}_{1-2} = \frac{1}{3}\mathbf{a}_x + \frac{2}{3}\mathbf{a}_y + \frac{2}{3}\mathbf{a}_z \tag{3-7b}$$

Substituting Eq. (3-7) into Eq. (3-2), we obtain

$$\mathbf{F}_1 = \frac{q_1 q_2}{4\pi\varepsilon_o 3^3}(\mathbf{a}_x + 2\,\mathbf{a}_y + 2\,\mathbf{a}_z)\,.$$

Example 3-2

Find an expression for the electric force acting on q_1 at $(x, y, z) = (2, 3, \sqrt{3})$ due to q_2 located at the origin in (a) Cartesian coordinates, and (b) spherical coordinates

Solution

(a) Position vectors of the charges are, respectively,

$$\mathbf{r}_1 = 2\mathbf{a}_x + 3\mathbf{a}_y + \sqrt{3}\mathbf{a}_z$$

$$\mathbf{r}_2 = 0$$

Distance vector and its magnitude are

$$\mathcal{R}_{1-2} = \mathbf{r}_1 - \mathbf{r}_2 = 2\mathbf{a}_x + 3\mathbf{a}_y + \sqrt{3}\mathbf{a}_z$$

$$|\mathcal{R}_{1-2}| = \sqrt{2^2 + 3^2 + (\sqrt{3})^2} = 4$$

From Eq. (3-5), the electric force on q_1 is

$$\mathbf{F}_1 = \frac{q_1 q_2}{4\pi\varepsilon_o \mathcal{R}_{1-2}^3}\mathcal{R}_{1-2} = \frac{q_1 q_2}{\pi\varepsilon_o 4^4}(2\,\mathbf{a}_x + 3\,\mathbf{a}_y + \sqrt{3}\,\mathbf{a}_z)\,. \tag{3-8}$$

(b) Radial distance from the origin to the point of q_1 is

$$R = \sqrt{2^2 + 3^2 + (\sqrt{3})^2} = 4$$

Since the source point is at the origin, the distance vector is simply given by

$$\mathcal{R}_{1-2} = 4\mathbf{a}_R$$

The electric force on q_1 due to q_2 is therefore

$$\mathbf{F}_1 = \frac{q_1 q_2}{4\pi\varepsilon_o \mathcal{R}_{1-2}^2}\mathbf{a}_{1-2} = \frac{q_1 q_2}{\pi\varepsilon_o 4^3}\mathbf{a}_R \tag{3-9}$$

Coordinate transformation of Eq. (3-9) into Cartesian coordinates can show that the results in Eq. (3-9) and Eq. (3-8) are the same.

Exercise 3.1
A charge q_2 experiences an electric force of $1[\text{N}]$ at a distance $1[\text{m}]$ from another charge. Find the distance at which the force on q_2 is reduced to $0.5[\text{N}]$.
Ans. $1.414[\text{m}]$.

3.2 Electric Field Intensity

The electric field intensity is defined as ***the electric force exerted on a unit test charge***. If a point charge q is located at a point with position vector $\mathbf{r}'$ in free space, the electric field intensity at a point with position vector $\mathbf{r}$ is expressed as

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\varepsilon_o}\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}} \qquad [\text{V/m}] \qquad (3\text{-}10)$$

which is measured in volts per meter. $\mathbf{E}(\mathbf{r})$ in Eq. (3-10) is a vector function. For the charge q located at the position vector $\mathbf{r}'$, $\mathbf{E}(\mathbf{r})$ specifies the electric force vector acting on a unit charge located at the position vector $\mathbf{r}$. The vector function $\mathbf{E}(\mathbf{r})$ defines a vector field, called an electric field, in the region surrounding the source charge. As an example, the electric field due to a point charge q is depicted in Fig. 3.2, in which shades are used for visual effects.

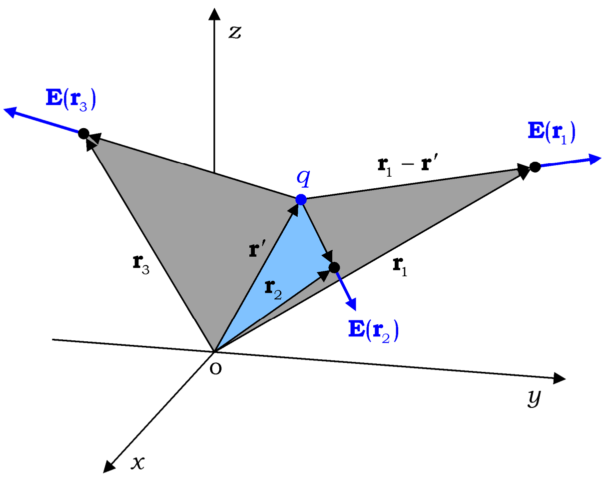


Fig. 3.2 The electric field due to point charge q at position vector $\mathbf{r}'$.

Example 3-3
Find an expression for the electric field of a point charge q located at the origin in
(a) Cartesian coordinate system, and
(b) spherical coordinate system.

Solution

(a) Position vectors of the field and source points are

$$\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z$$

$$\mathbf{r}' = 0$$

Magnitude of the distance vector is

$$|\mathcal{R}| = |\mathbf{r} - \mathbf{r}'| = \sqrt{x^2 + y^2 + z^2}$$

From Eq. (3-10), the electric field of a point charge q located at the origin in Cartesian system is

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\varepsilon_o}\frac{x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z}{[x^2 + y^2 + z^2]^{3/2}}.} \tag{3-11}$$

(b) For a point charge located at the origin, the distance vector in spherical coordinates is

$$\mathcal{R} = R\,\mathbf{a}_R$$

The electric field of the charge q located at the origin in spherical system is

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\varepsilon_o R^2}\mathbf{a}_R} \tag{3-12}$$

The point charge placed at the origin has spherical symmetry: it appears the same as we move around it varying θ and ϕ while keeping R constant. As was discussed in Chapter 1, the resulting vector field should be independent of θ and ϕ, and have neither the θ- nor the ϕ-component.

In Eqs. (3-11) and (3-12), the origin is excluded, because it is physically occupied by the point charge, and therefore the electric field is not defined.

Exercise 3.2

Show that $\mathbf{E}$ in Eq. (3-12) is irrotational, forming a conservative field, and also solenoidal in the region $R > 0$.

Ans. $\nabla \times \mathbf{E} = 0$ and $\nabla \bullet \mathbf{E} = 0$.

3.2.1 Electric Field due to Discrete Charges

The electric field intensity obeys the principle of superposition such that ***the total*** $\mathbf{E}$ ***due to multiple charges is equal to the vector sum of*** $\mathbf{E}$***'s due to all the individual charges***. The superposition of the electric field is based on the fact that the electric field of a charge is unaffected by the existence of the other charges or other electric fields. According to the principle of superposition, the total $\mathbf{E}$ at position vector $\mathbf{r}$ due to N point charges is expressed as

$$\mathbf{E}(\mathbf{r}) = \sum_{j=1}^{N}\frac{q_j}{4\pi\varepsilon_o}\frac{\mathbf{r} - \mathbf{r}_j}{|\mathbf{r} - \mathbf{r}_j|^3} \tag{3-13}$$

where q_j is a charge at position vector $\mathbf{r}_j$. As an example, the electric field intensity due to two point charges q_1 and q_2 is depicted in Fig. 3.3, in which shades are used for visual effects.

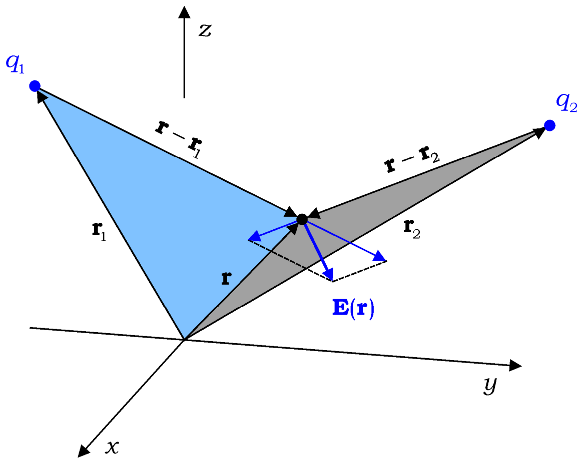


Fig. 3.3 Principle of superposition of the electric field intensity.

Example 3-4
Determine $\mathbf{E}$ due to two point charges $+q$ and $-q$ separated by a small distance d, which is called an electric dipole, as shown in Fig. 3.4.

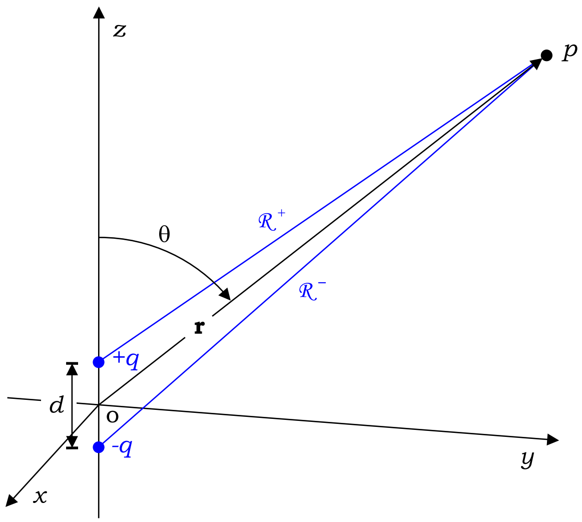


Fig. 3.4 An electric dipole consists of two identical charges of the opposite polarities separated by a small distance.

Solution

Position vector of a field point in Cartesian coordinates is

$$\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z$$

Position vectors of two source points are

$$\mathbf{r}^+ = (d/2)\,\mathbf{a}_z$$

$$\mathbf{r}^- = (-d/2)\,\mathbf{a}_z$$

Two distance vectors are given as follows:

$$\mathcal{R}^+ \equiv \mathbf{r} - \mathbf{r}^+ = x\,\mathbf{a}_x + y\,\mathbf{a}_y + (z - d/2)\,\mathbf{a}_z \tag{3-14a}$$

$$\mathcal{R}^- \equiv \mathbf{r} - \mathbf{r}^- = x\,\mathbf{a}_x + y\,\mathbf{a}_y + (z + d/2)\,\mathbf{a}_z \tag{3-14b}$$

Applying the law of cosines to two triangles, $(+q)op$ and $(-q)op$, we have

$$\mathcal{R}^+ = \sqrt{r^2 + (\tfrac{1}{2}d)^2 - 2r(\tfrac{1}{2}d)\cos\theta} \tag{3-15a}$$

$$\mathcal{R}^- = \sqrt{r^2 + (\tfrac{1}{2}d)^2 - 2r(\tfrac{1}{2}d)\cos(\pi - \theta)} \tag{3-15b}$$

From Eq. (3-13), the electric field of a dipole is in general written as

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\varepsilon_o}\left[\frac{\mathcal{R}^+}{(\mathcal{R}^+)^3} - \frac{\mathcal{R}^-}{(\mathcal{R}^-)^3}\right] \tag{3-16}$$

Under the condition $|\mathbf{r}| >> d$, the term $(\tfrac{1}{2}d)^2$ in the radicand in Eq. (3-15) is ignored compared with the other two. With the aid of the binomial expansion, $(1 \pm a)^n \approx 1 \pm na$ (for $a << 1$ and for any real n), we write

$$\frac{1}{(\mathcal{R}^+)^3} \cong \frac{1}{[r^2 - rd\cos\theta]^{3/2}} \cong \frac{1}{r^3}\left[1 + \frac{3}{2}\frac{d}{r}\cos\theta\right] \tag{3-17a}$$

$$\frac{1}{(\mathcal{R}^-)^3} \cong \frac{1}{[r^2 + rd\cos\theta]^{3/2}} \cong \frac{1}{r^3}\left[1 - \frac{3}{2}\frac{d}{r}\cos\theta\right] \tag{3-17b}$$

Inserting Eqs. (3-14) and (3-17), into Eq. (3-16), and substituting $|\mathbf{r}| = R$, we obtain

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\varepsilon_o R^3}\left[-d\,\mathbf{a}_z + \left(x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z\right)3\frac{d}{R}\cos\theta\right] \tag{3-18}$$

Transforming the coordinates, we have

$$x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z = R\,\mathbf{a}_R \tag{3-19a}$$

$$-d\,\mathbf{a}_z = -d\cos\theta\,\mathbf{a}_R + d\sin\theta\,\mathbf{a}_\theta \tag{3-19b}$$

Substituting Eq. (3-19) into Eq. (3-18), the electric field intensity of an electric dipole is

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{qd}{4\pi\varepsilon_o R^3}[2\cos\theta\,\mathbf{a}_R + \sin\theta\,\mathbf{a}_\theta]} \tag{3-20}$$

The electric dipole is the simplest model of an atom for electricity. It is useful for examining the interaction between a material and an external electric field for the electrical property of the material.

Exercise 3.3
Two identical point charges are at $x=-1$ and $x=3$ on the x-axis, respectively. Locate the point where **E** is zero.

Ans. A point at $x=1$ on the x-axis.

Exercise 3.4
Is it always true that **E** due to multiple charges is irrotational only because **E** due to a single point charge is irrotational?

Ans. Yes, because $\nabla\times\mathbf{E}=0$ is a linear equation.

3.2.2 Electric Field due to a Continuous Charge Distribution

When a large number of discrete charges are concentrated in a small region of space, and the spacing between adjacent charges is much shorter than the distance from the charge to the field point, we may disregard the discrete nature of the charges and treat the charges as a continuous quantity in the given region. Under these conditions, the given region of space is subdivided into many infinitesimal elements of volume, each of which is large enough to contain a large number of charges yet small enough to be considered as a point when seen from the field point. Then the total electric field intensity at the field point is obtained by summing the contributions from all the individual volume elements.

To describe the charges distributed in a volume, we define a volume charge density ρ_v as charges per unit volume, which is measured in coulombs per cubic meter $[\mathrm{C/m^3}]$. The volume charge density allows us to use a volume integral, instead of the summation, in computing the total electric field intensity. Similarly, for the charges confined to a surface, we define a surface charge density ρ_s as charges per unit area, which is measured in coulombs per square meter $[\mathrm{C/m^2}]$. If the charges are confined to a line, we define a line charge density ρ_ℓ as charges per unit length, which is measured in coulombs per meter $[\mathrm{C/m}]$. For the total electric field intensity due to ρ_s, or ρ_ℓ, a surface integral is taken over the surface of ρ_s, or a line integral is taken along the line of ρ_ℓ.

Let Δv be an incremental volume centered at position vector $\mathbf{r}'$, containing an incremental charge Δq. The volume charge density at the point is then defined as

$$\rho_v(\mathbf{r}') = \lim_{\Delta v \to 0} \frac{\Delta q}{\Delta v} \qquad [\mathrm{C/m^3}] \qquad (3\text{-}21)$$

Here, Δv is large enough to contain a large number of discrete charges but small enough to be regarded as a point at position vector $\mathbf{r}'$. The volume charge density $\rho_v(\mathbf{r}')$ is generally a smooth function of position in three-dimensional space, and thus has continuous partial derivatives.

When a volume charge of a density $\rho_v[\mathrm{C/m^3}]$ is present in a region of space, the region is first subdivided into N incremental elements of volume before we can compute the electric field of the charge. An element of volume centered at position vector $\mathbf{r}'_j$, which is denoted as $\Delta v(\mathbf{r}'_j)$, will contain an incremental charge $q_j = \rho_v(\mathbf{r}'_j)\Delta v(\mathbf{r}'_j)$. Next, we apply Coulomb's law to these incremental charges, and add the electric field intensities due to all the individual elements. Taking the limit as $N \to \infty$ and $\Delta v \to 0$, we obtain the electric field intensity at position vector $\mathbf{r}$ as

$$\mathbf{E}(\mathbf{r}) = \lim_{\substack{N\to\infty \\ \Delta v \to 0}} \sum_{j=1}^{N} \frac{\rho_v(\mathbf{r}'_j)\Delta v(\mathbf{r}'_j)}{4\pi\varepsilon_o} \frac{\mathbf{a}_{\mathcal{R}}}{\left|\mathbf{r}-\mathbf{r}'_j\right|^2} \qquad (3\text{-}22)$$

where $\mathbf{a}_{\mathcal{R}}$ is a unit vector in the direction of the distance vector $\mathcal{R} = \mathbf{r} - \mathbf{r}'_j$. We note that the charge contained in an element of volume $\Delta v(\mathbf{r}'_j)$ is regarded as a point charge in Eq. (3-22). Nevertheless, Eq. (3-22) is distinguished from Eq. (3-13) in that the volume charge density ρ_v is a continuous function of position. We know from calculus that the right-hand side of Eq. (3-22) is a volume integral of the summand over a total volume $N\Delta v$. The electric field intensity due to a volume charge density ρ_v is therefore

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o}\int_{V'} \frac{\rho_{v'}}{\mathcal{R}^2}\mathbf{a}_{\mathcal{R}}\, dv'} \qquad [\mathrm{V/m}] \qquad (3\text{-}23)$$

Here, $\mathcal{R}$ and $\mathbf{a}_{\mathcal{R}}$ are the magnitude and direction of the distance vector $\mathcal{R} = \mathbf{r} - \mathbf{r}'$, and $\mathbf{r}$ and $\mathbf{r}'$ are the position vectors of the field and source points, respectively. It should be noted that $\mathbf{E}$ in Eq. (3-23) is expressed in terms of mixed coordinates: the primed coordinates are used to express the quantities at the source point, whereas the unprimed coordinates are used to express the quantities at the field point. We also note that the coordinate axes of the primed system exactly coincide with those of the unprimed system; they are only called differently. The volume integral in Eq. (3-23) is conducted in the primed coordinates, while the unprimed coordinates are regarded as constants. The unit vector $\mathbf{a}_{\mathcal{R}}$

cannot be taken outside the integral in Eq. (3-23), because it varies with the primed coordinates, as is evident from the distance vector $\mathbf{r} - \mathbf{r}' = (x - x')\mathbf{a}_x + (y - y')\mathbf{a}_y + (z - z')\mathbf{a}_z$.

Following the same procedure as was used for the volume charge, the surface with a surface charge density ρ_s is subdivided into many incremental elements of surface. A surface element of an area ds' carries an incremental charge $\rho_{s'} ds'$, which is regarded as a point charge at position vector $\mathbf{r}'$. Applying Coulomb's law to these incremental charges, we obtain the electric field intensity due to a surface charge density ρ_s as

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o} \int_{s'} \frac{\rho_{s'}}{\mathcal{R}^2} \mathbf{a}_{\mathcal{R}} ds'} \qquad \text{[V/m]} \qquad (3\text{-}24)$$

where $\boldsymbol{\mathcal{R}} = \mathcal{R}\,\mathbf{a}_{\mathcal{R}} = |\mathbf{r} - \mathbf{r}'|\,\mathbf{a}_{\mathcal{R}} = [(x - x')^2 + (y - y')^2 + (z - z')^2]^{1/2}\,\mathbf{a}_{\mathcal{R}}$ in Cartesian coordinates. Again, the surface integral in Eq. (3-24) is conducted in the primed coordinates, while the unprimed coordinates are regarded as constants.

The same procedure is followed to obtain the electric field intensity of a line charge density ρ_ℓ. A line segment of a differential length $d\ell'$ contains an incremental charge $\rho_{\ell'} d\ell'$, which is regarded as a point charge at position vector $\mathbf{r}'$. Applying Coulomb's law to these incremental charges, we obtain the electric field intensity due to a line charge density ρ_ℓ as

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o} \int_{\mathcal{L}'} \frac{\rho_{\ell'}}{\mathcal{R}^2} \mathbf{a}_{\mathcal{R}} d\ell'} \qquad \text{[V/m]} \qquad (3\text{-}25)$$

where $\boldsymbol{\mathcal{R}} = \mathcal{R}\,\mathbf{a}_{\mathcal{R}} = |\mathbf{r} - \mathbf{r}'|\,\mathbf{a}_{\mathcal{R}}$.

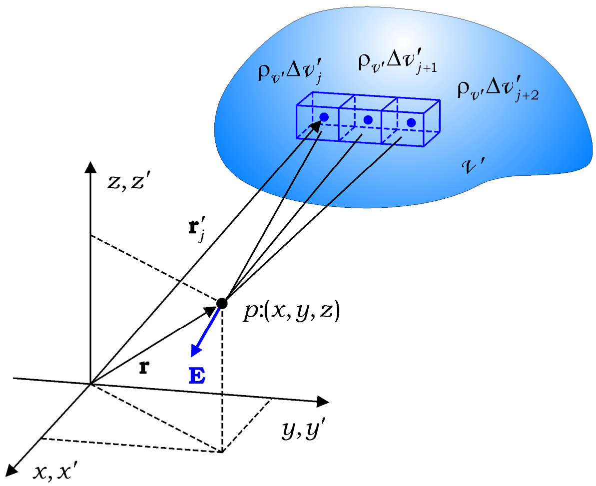


Fig. 3.5 Electric field intensity due to a volume charge density.

Example 3-5
Determine the electric field intensity of an infinitely long, straight, line-charge of a uniform density $\rho_{\ell o}$ that is along the z-axis, as shown in Fig. 3.6.

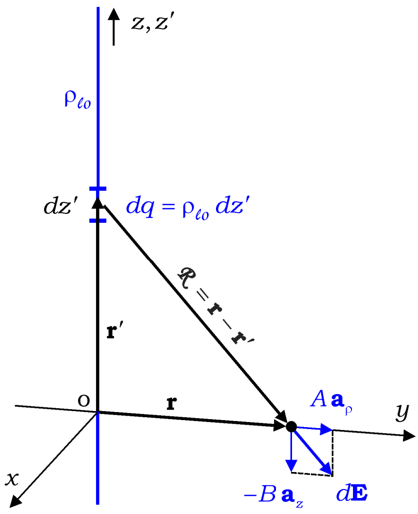


Fig. 3.6 An infinitely long, straight, line-charge.

Solution

In view of the cylindrical symmetry of the line charge, the field point is assumed to be on the y-axis without loss of generality. The position vectors of the field and source points are

$$\mathbf{r} = \rho\,\mathbf{a}_\rho$$

$$\mathbf{r}' = z'\,\mathbf{a}_{z'}$$

The distance vector is expressed in terms of the base vectors $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$ at the field point. Noting that $\mathbf{a}_{z'} = \mathbf{a}_z$, we have

$$\mathcal{R} = \mathbf{r} - \mathbf{r}' = \rho\,\mathbf{a}_\rho - z'\,\mathbf{a}_z \tag{3-26}$$

Differential charge at the source point is

$$dq = \rho_{\ell o}\,dz'$$

From Coulomb's law, the differential field $d\mathbf{E}$ at the field point due to dq at the source point is

$$d\mathbf{E} = \frac{\rho_{\ell o}\, dz'}{4\pi\varepsilon_o} \frac{\mathcal{R}}{\mathcal{R}^3} = \frac{\rho_{\ell o}\, dz'}{4\pi\varepsilon_o} \frac{\rho\, \mathbf{a}_\rho - z'\mathbf{a}_z}{(\rho^2 + z'^2)^{3/2}} \tag{3-27}$$

We see from Eq. (3-27) that two differential charges of an equal amount $\rho_{\ell o} dz'$, located at $z' = a$ and $z' = -a$ will result in $d\mathbf{E} \bullet \mathbf{a}_z = 0$.

Adding only the ρ-component of $d\mathbf{E}$, we obtain

$$\mathbf{E} = \int d\mathbf{E} = \frac{\rho_{\ell o}\, \rho\, \mathbf{a}_\rho}{4\pi\varepsilon_o} \int_{-\infty}^{\infty} \frac{dz'}{(\rho^2 + z'^2)^{3/2}} = \frac{\rho_{\ell o}\, \rho\, \mathbf{a}_\rho}{4\pi\varepsilon_o} \left[\left. \frac{z'/\rho^2}{\sqrt{\rho^2 + z'^2}} \right|_{z'=-\infty}^{z'=\infty} \right] \tag{3-28}$$

The electric field intensity at a distance ρ from an infinitely long, straight, line-charge of a density $\rho_{\ell o}$, which lies along the z-axis, is

$$\boxed{\mathbf{E} = \frac{\rho_{\ell o}}{2\pi\varepsilon_o \rho} \mathbf{a}_\rho} \qquad [\text{V/m}] \tag{3-29}$$

Here, $\rho_{\ell o}$ is the line charge density, while ρ is the radial distance in cylindrical coordinates.

For future reference

$$\boxed{\int \frac{dx}{(x^2 + a^2)^{3/2}} = \frac{x / a^2}{\sqrt{x^2 + a^2}}} \tag{3-30}$$

Let us now examine the line charge for symmetry. The infinitely long line charge lying along the z-axis has cylindrical symmetry, translational symmetry in the z-direction, and twofold rotational symmetry about the x-axis. The line charge appears the same even if it is rotated about the z-axis, or displaced in the z-direction, or rotated about the x-axis by 180^o. Under these conditions, the resultant vector field is independent of ϕ and z, and has no ϕ- and no z-component such that it is of the form $\mathbf{E}(\mathbf{r}) = E_\rho(\rho)\, \mathbf{a}_\rho$, as we see from Eq. (3-29).

Example 3-6

A uniform line charge density $\rho_{\ell o}$ forms a circle of radius a in the $z = 0$ plane, with the center at the origin, as shown in Fig. 3.7.

(a) Find $\mathbf{E}$ at a point p_1:$(0, 0, b)$ on the z-axis.

(b) Show that the line charge can be regarded as a point charge positioned at the origin as $b \to \infty$.

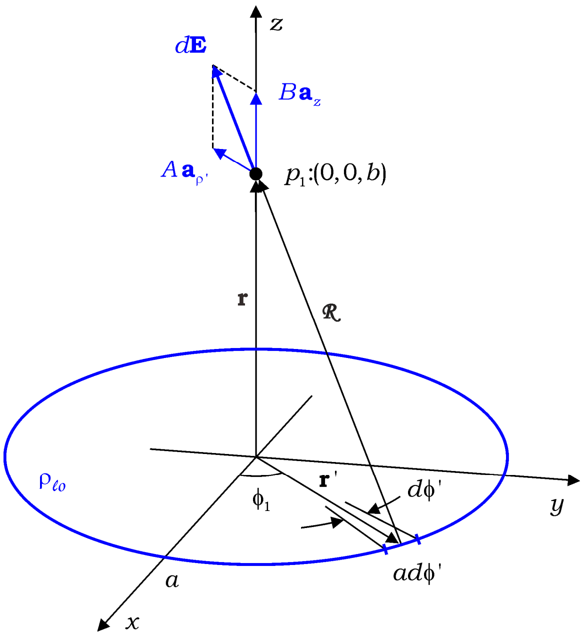


Fig. 3.7 A line charge density on a circle.

Solution

(a) Position vectors of the field and source points are

$$\mathbf{r} = b\,\mathbf{a}_z$$

$$\mathbf{r}' = a\,\mathbf{a}_{\rho'} = a(\cos\phi_1\,\mathbf{a}_x + \sin\phi_1\,\mathbf{a}_y)$$

In a mixed coordinate system, the distance vector is expressed as

$$\mathcal{R} = \mathbf{r} - \mathbf{r}' = b\,\mathbf{a}_z - a\,\mathbf{a}_{\rho'}$$

The unit vector $\mathbf{a}_{\rho'}$ will be transformed to the unprimed coordinates if it does not vanish in the final expression for $\mathbf{E}$.

Differential charge at the source point with position vector $\mathbf{r}'$ is

$$dq = \rho_{\ell o} a\, d\phi'$$

Differential field at the field point with position vector $\mathbf{r}$ is

$$d\mathbf{E} = \frac{\rho_{\ell o} a\, d\phi'}{4\pi\varepsilon_o}\frac{\mathcal{R}}{\mathcal{R}^3} = \frac{\rho_{\ell o} a\, d\phi'}{4\pi\varepsilon_o}\frac{(-a\,\mathbf{a}_{\rho'} + b\,\mathbf{a}_z)}{(a^2+b^2)^{3/2}} \tag{3-31}$$

Although the field $d\mathbf{E}$ due to charge dq at $\phi' = \phi_1$ has exactly the same form as the field $d\mathbf{E}$ due to charge dq at $\phi' = \phi_1 + \pi$, the direction of $\mathbf{a}_{\rho'}$ at $\phi' = \phi_1$ is opposite to the direction of $\mathbf{a}_{\rho'}$ at $\phi' = \phi_1 + \pi$. Thus, the terms with $\mathbf{a}_{\rho'}$ do not contribute to $d\mathbf{E}$. By adding the z-components of $d\mathbf{E}$, we obtain the total electric field intensity at the field point as

$$\mathbf{E} = \int d\mathbf{E} = \frac{\rho_{\ell o}}{4\pi\varepsilon_o} \frac{ab\,\mathbf{a}_z}{(a^2+b^2)^{3/2}} \int_{\phi'=0}^{\phi'=2\pi} d\phi' = \frac{\rho_{\ell o}}{2\varepsilon_o} \frac{ab\,\mathbf{a}_z}{(a^2+b^2)^{3/2}}. \qquad (3\text{-}32)$$

(b) The total charge on the circle is

$$Q = 2\pi a \rho_{\ell o} \qquad (3\text{-}33)$$

Substituting Eq. (3-33) for $\rho_{\ell o}$ in Eq. (3-32), and using the approximation $(a^2+b^2)^{3/2} \approx b^3$ as $b \to \infty$, we obtain

$$\mathbf{E} = \frac{Q}{4\pi\varepsilon_o b^2}\mathbf{a}_z \qquad (3\text{-}34)$$

Eq. (3-34) can be viewed as the electric field intensity at a distance b from a point charge Q positioned at the origin.

Exercise 3.5

A uniform line charge density $\rho_{\ell o}$ extends from $z=-10$ to $z=10$ along the z-axis. Determine **E** at a point $(3,4,0)$ in Cartesian coordinates. [Hint: Eq. (3-28]]

Ans. $\mathbf{E} = \dfrac{\rho_{\ell o}}{\pi\varepsilon_o 5\sqrt{5}}\left(\dfrac{3}{5}\mathbf{a}_x + \dfrac{4}{5}\mathbf{a}_y\right)$.

Exercise 3.6

An infinitely long line charge of a density $\rho_\ell = 3[\text{nC/m}]$ is parallel to the z-axis, passing through a point (0,2,0) in Cartesian coordinates. Determine **E** at a point (4,5,-2).

Ans. $\mathbf{E} = 8.6\,\mathbf{a}_x + 6.5\,\mathbf{a}_y[\text{V/m}]$.

Review Questions with Hints

RQ 3.1 State Coulomb's law in words. [Eq.(3-1)]

RQ 3.2 What are the source and field points? [Fig.3.2]

RQ 3.3 Define electric field intensity. [Eq.(3-10)]

RQ 3.4 State the principle of superposition of **E**. [Eq.(3-13)]

RQ 3.5 What are the units of volume, surface, and line charge density? [Eqs.(3-23)(3-24)(3-25)]

RQ 3.6 Describe how **E** varies with distance for a point charge, an electric dipole, and an infinitely long line charge, respectively. [Eqs.(3-12)(3-20)(3-29)]

RQ 3.7 Describe how **E** due to an electric dipole varies with the separation between two charges. [Eq.(3-20)]

3.3 Electric Flux Density and Gauss's Law

In section 3-2, the electric field of a point charge was obtained as in Eqs. (3-11) and Eqs. (3-12). We also saw that a point charge is the simplest source of the electric field. Figure 3-8(a) illustrates the electric field intensities observed at different points in space, which are produced by a point charge. In Fig. 3-8(b), electric field lines are used to depict the electric field, which are also called electric flux lines. Electric field lines originate from positive charges and terminate on negative charges. The electric field vector at a point is tangential to the field line, and the magnitude of the field is described by the density of the field lines in the neighborhood of the point. Extending the electric field to a distribution of charges, the total electric field intensity at a point is the sum of those due to the individual charges, in accordance with the principle of superposition. Both a single and multiple charges produce the electric field showing no abrupt change in the magnitude and direction in free space. Therefore, the electric field lines of a charge distribution are given by smooth curves in a homogenous medium.

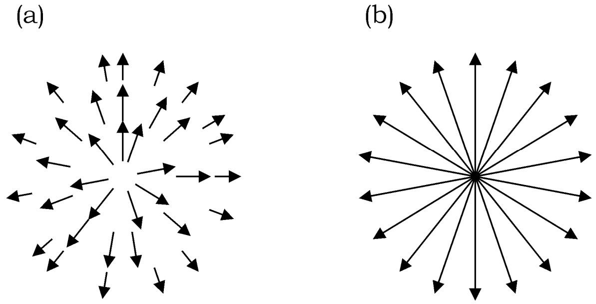


Fig. 3.8 A point charge (a) electric field vectors (b) electric field lines.

3.3.1 Electric Flux Density

We now examine the experiment on electrostatic induction as depicted in Fig. 3.9. The space between two concentric and perfectly conducting spheres is filled with a dielectric, or an insulating material, such that no free charge may be in motion in the region. The net charges of $+Q$[C] are imparted to the inner sphere where they are free to move. The charges should distribute themselves uniformly on the surface of the conductor in order not to produce any electric field inside the conductor. Otherwise, the internal electric field would accelerate the charges instantaneously, generating an infinite conduction current, and thus causing redistribution of the charges until there is no electric field inside the conductor. Just after the outer conductor is connected momentarily to ground for discharge, the induced charges of $-Q$[C] are observed on the outer conductor, independent of the dielectric filling in the gap. This electrostatic induction was interpreted by early pioneers as the result of some sort of displacement, or the electric flux, which starts at the inner conductor and ends at the outer conductor.

The total electric flux Ψ coming out of the inner conductor equals the net charge on the conductor. That is,

$$\Psi = +Q \tag{3-35}$$

which is measured in coulombs[C]. Here, we define the electric flux density **D** as the electric flux per unit area, which is measured in coulombs per square meter $[\mathrm{C/m^2}]$. The electric flux density is a vector field, generally given as a smooth function of position. It is important to note that the magnitude of **D** represents *the net flux crossing a unit area of the cross section*, or the plane perpendicular to the direction of **D**.

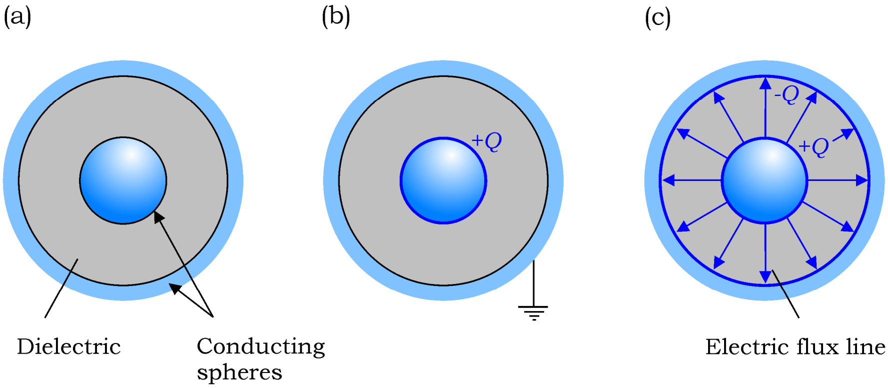


Fig. 3.9 Electrostatic induction. The total electric flux equals the net charge $+Q$[C].

Let us digress briefly and explore the uniform surface charge on a perfectly conducting sphere for symmetry. The uniform surface charge has spherical symmetry: it appears the same as we move around it varying θ and ϕ while keeping R constant in spherical coordinates. As was mentioned in Chapter 1, the spherical symmetry assures that the resultant vector field is independent of θ and ϕ, and has neither the θ- nor the ϕ-component. Therefore the electric flux density induced by the uniform surface charge on the sphere should be of the form $\mathbf{D} = G(R)\,\mathbf{a}_R$ in spherical coordinates.

We see from Fig. 3.9 that the geometry of the two-sphere configuration has spherical symmetry, and thus the uniform distribution of charges on the conductors also has the spherical symmetry. As a result, the electric flux density in the gap should be of the form $\mathbf{D} = D_R(R)\,\mathbf{a}_R$. To compute the total electric flux, we construct a hypothetical sphere of radius R_o in the space between the conductors, and take the surface integral of **D** over the sphere as follows:

$$\Psi = \oint_S \mathbf{D}\cdot d\mathbf{s} = \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} D_R R_o^{\,2}\sin\theta\, d\theta\, d\phi = 4\pi R_o^{\,2} D_R \tag{3-36}$$

Noting that the total electric flux $\Psi = +Q$, the electric flux density on the hypothetical sphere of radius R_o is obtained from Eq. (3-36) as

$$\mathbf{D} = D_R\,\mathbf{a}_R = \frac{+Q}{4\pi R_o^{\,2}}\,\mathbf{a}_R \tag{3-37}$$

The electric flux density given in Eq. (3-37) would remain the same, even if the gap were filled with a different dielectric, or even if the inner sphere were shrunken to a point at the origin and the outer sphere were expanded to infinity, keeping the total charges on the spheres constant. In view of these discussions, the electric flux density due to a point charge q located at the origin is given, in spherical coordinates, as

$$\boxed{\mathbf{D} = \frac{q}{4\pi R^2}\,\mathbf{a}_R} \qquad [\mathrm{C/m^2}] \tag{3-38}$$

It is should be noted that Eq. (3-38) is true regardless of the dielectric surrounding the point charge.

Borrowing from Section 3-2, the electric field intensity due to a point charge q located at the origin in free space is

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_o R^2}\,\mathbf{a}_R \tag{3-12}$$

Comparison of Eq. (3-38) with Eq. (3-12) leads to the relation between **D** and **E** in free space, that is,

$$\boxed{\mathbf{D} = \varepsilon_o\mathbf{E}} \qquad [\mathrm{C/m^2}] \tag{3-39}$$

where ε_o is the permittivity of free space. This is the constitutive relation for **D** and **E** in free space.

Following the same procedure as was used for the electric field, the electric flux density due to a continuous distribution of charges is expressed as follows:

$$\boxed{\mathbf{D(r)} = \frac{1}{4\pi}\int_{V'}\frac{\rho_{v'}}{\mathcal{R}^2}\,\mathbf{a}_{\mathcal{R}}\,dv'} \qquad \text{(volume charge)} \tag{3-40a}$$

$$\boxed{\mathbf{D(r)} = \frac{1}{4\pi}\int_{S'}\frac{\rho_{s'}}{\mathcal{R}^2}\,\mathbf{a}_{\mathcal{R}}\,ds'} \qquad \text{(surface charge)} \tag{3-40b}$$

$$\boxed{\mathbf{D(r)} = \frac{1}{4\pi}\int_{L'}\frac{\rho_{\ell'}}{\mathcal{R}^2}\,\mathbf{a}_{\mathcal{R}}\,d\ell'} \qquad \text{(line charge)} \tag{3-40c}$$

where $\mathcal{R} = |\mathbf{r} - \mathbf{r}'| = [(x-x')^2 + (y-y')^2 + (z-z')^2]^{1/2}$ in Cartesian coordinates, and $\mathbf{a}_{\mathcal{R}}$ is a unit vector in the direction of $\mathcal{R}$. We note that Eq. (3-40) is written in terms of mixed coordinates: the primed coordinates are used to express the quantities at the source point, while the unprimed coordinates are used to express the quantities at the field point.

In a material medium, the expression for **D** is the same as Eq. (3-40). However, the constitutive relation in Eq. (3-39) needs to be modified. The behavior of a dielectric in an external electric field will be discussed in Section 3-5.

Exercise 3.7
A net charge of $+Q$[C] is imparted to a conducting cube. Find the total electric flux leaving the cube.

Ans. $+Q$[C]

Exercise 3.8
One sphere in Fig. 3.9 is brought closer to the other(not concentric any more). (a) Are $+Q$ and $-Q$ still uniform on the spheres? If not, (b) locate the point where the charges are concentrated the most.
[Hint: $\mathbf{E} = 0$ inside the conductor due to electric dipoles straddling the gap.]

Ans. (a) No, (b) The point where two spheres are closest to each other.

Exercise 3.9
Determine the divergence of the electric flux density that is produced by a point charge in free space.

Ans. $\nabla \cdot \mathbf{D} = 0$.

3.3.2 Gauss's Law

Earlier, the electric flux density caused by a point charge located at the origin was obtained as in Eq. (3-38). By taking the divergence of both sides of Eq. (3-38), we obtain $\nabla \cdot \mathbf{D} = 0$ in the region $R > 0$. This implies that the integral of **D** over any closed surface, not enclosing the origin, is identically zero. Next, from Eq. (3-38) we can easily show that the closed surface integral of **D** over a sphere centered at the origin is equal to q. Does this imply that the integral of **D** over any closed surface S, enclosing the origin, is always equal to q? To answer this question, we separate the volume enclosed by S into two parts: a small sphere centered at the origin and the remainder between the sphere and S. Since the closed surface integral of **D** over the bounding surface of the remainder is always zero, the integral of **D** over S shrinks to the closed surface integral of **D** over the small sphere center at the origin. Thus the integral of **D** over any closed surface, enclosing the origin, is always equal to q. The principle of superposition makes it possible to extend the electric flux of a point charge to an arbitrary distribution of charges.

Generalizing the electric flux of a point charge to an arbitrary distribution of charges leads to the Gauss's law, which states that

the net outward electric flux passing through any closed surface is equal to the total charge enclosed by that surface.

Gauss's law is expressed mathematically as

$$\boxed{\oint_{\mathcal{S}} \mathbf{D} \bullet d\mathbf{s} = Q} \qquad \text{[C]} \tag{3-41}$$

The small circle on the integral sign signifies that surface $\mathcal{S}$ is closed. The dot product in the integrand is required to convert the differential area $|d\mathbf{s}|$ on surface $\mathcal{S}$ to an equivalent area on the cross section of **D**. It should be noted that the differential area vector $d\mathbf{s}$, on a closed surface, always points outward from the enclosed volume. For this reason, a positive value of the integral on the left-hand side of Eq. (3-41) means that the electric flux of Q coulombs is coming out of the volume enclosed by surface $\mathcal{S}$.

Gauss's law can also be written in terms of a volume charge density ρ_v as

$$\boxed{\oint_{\mathcal{S}} \mathbf{D} \bullet d\mathbf{s} = \int_{\mathcal{V}} \rho_v dv} \tag{3-42}$$

where $\mathcal{V}$ is the volume bounded by the surface $\mathcal{S}$.

Coulomb's law is the fundamental law of electrostatics, which we can apply any time to determine **E** of a charge distribution. Meanwhile, Gauss's law is particularly useful for determining **D** when the charge distribution has certain symmetry. The application of Gauss's law, however, is subject to the proper choice of a closed surface $\mathcal{S}$, called a Gaussian surface. It should be noted that the Gaussian surface is a hypothetical surface defined mathematically. Through the Gaussian surface, the closed surface integral of **D** can be greatly simplified to a simple algebraic equation. The Gaussian surface $\mathcal{S}$ should be chosen such that

D on the Gaussian surface is always ***constant and normal*** to the surface($\mathbf{D} \bullet d\mathbf{s} = Dds = \text{constant}$). Otherwise, **D** on a part of the Gaussian surface is ***tangential*** to the surface($\mathbf{D} \bullet d\mathbf{s} = 0$).

To obtain the point form of Gauss's law, which is also called the differential form, both sides of Eq. (3-42) are divided by an incremental volume Δv bounded by a small closed surface $\mathcal{S}$. Taking the limit as $\Delta v \to 0$, we have

$$\lim_{\Delta v \to 0} \frac{\oint_{\mathcal{S}} \mathbf{D} \bullet d\mathbf{s}}{\Delta v} = \lim_{\Delta v \to 0} \frac{Q}{\Delta v} \tag{3-43}$$

We immediately recognize that the left-hand side of Eq. (3-43) is the divergence of **D**, and the right-hand side is the volume charge density at a point in space, or the center of Δv. The point form of Gauss's law is therefore,

$$\boxed{\nabla \bullet \mathbf{D} = \rho_v} \qquad [\text{C/m}^3] \tag{3-44}$$

Alternatively, we can derive the point form by applying divergence theorem to the left-hand side of Eq. (3-42) such as

$$\int_{\mathcal{V}} \nabla \bullet \mathbf{D} dv = \int_{\mathcal{V}} \rho_v dv \tag{3-45}$$

Since the volume $\mathcal{V}$ may be arbitrary, only if it encloses all the charges under consideration, the two integrands in Eq. (3-45) should be the same at every point in $\mathcal{V}$ to satisfy the equality, namely, $\nabla \bullet \mathbf{D} = \rho_v$.

Example 3-7

Determine the electric field intensity due to a uniform volume charge density ρ_{vo} forming a spherical shell of an inner radius a and an outer radius b, as shown in Fig. 3.10.

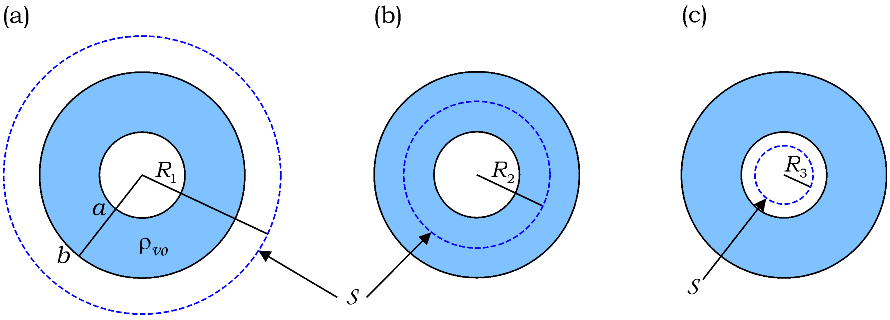


Fig. 3.10 A volume charge forming a spherical shell.

Solution

The volume charge has spherical symmetry, and thus the resulting **D** is expected to be of the form $\mathbf{D} = D_R(R)\,\mathbf{a}_R$. In view of the functional form of **D**, we construct a Gaussian surface in the form of a spherical surface concentric with the charge.

(a) In the region $R \ge b$, the net outward electric flux through the Gaussian surface of radius $R_1 \ge b$ is

$$\oint_S \mathbf{D} \bullet d\mathbf{s} = \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} D_R R_1^2 \sin\theta \, d\theta d\phi = D_R 4\pi R_1^2 \tag{3-46}$$

where the differential area vector $d\mathbf{s} = R_1^2 \sin\theta \, d\theta d\phi \, \mathbf{a}_R$

Next, the total charge enclosed by the Gaussian surface is

$$Q = \int_{\mathcal{V}} \rho_v dv = \int_{R=a}^{b}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \rho_{vo} R^2 \sin\theta \, dR \, d\theta \, d\phi = \rho_{vo}\frac{4\pi}{3}(b^3 - a^3) \tag{3-47}$$

where the differential volume $dv = R^2 \sin\theta \, dR \, d\theta \, d\phi$

Combining Eqs. (3-46) and (3-47) leads to

$$D_R = \frac{\rho_{vo}}{3R_1^2}(b^3 - a^3)$$

Thus, in the region $R \ge b$, omitting subscript 1 for generalization,

$$\mathbf{D} = D_R\mathbf{a}_R = \frac{\rho_{vo}}{3R^2}(b^3 - a^3)\,\mathbf{a}_R$$

$$\mathbf{E} = \frac{\mathbf{D}}{\varepsilon_o} = \frac{\rho_{vo}}{3\varepsilon_o R^2}(b^3 - a^3)\,\mathbf{a}_R\,. \tag{3-48}$$

(b) In the region $a \le R \le b$, the net outward electric flux through the Gaussian surface of radius $a \le R_2 \le b$ is

$$\oint_S \mathbf{D}\cdot d\mathbf{s} = \oint_S D_R\,\mathbf{a}_R \cdot d\mathbf{s} = D_R 4\pi R_2^2$$

The total charge enclosed by the Gaussian surface is

$$Q = \int_V \rho_v dv = \int_{R=a}^{R_2}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \rho_{vo} R^2 \sin\theta\, dR\, d\theta\, d\phi = \rho_{vo}\frac{4\pi}{3}(R_2^3 - a^3)$$

Combining the two equations gives

$$D_R = \frac{\rho_{vo}}{3R_2^2}(R_2^3 - a^3)$$

Thus, in the region $a \le R \le b$, omitting subscript 2 for generalization,

$$\mathbf{D} = \frac{\rho_{vo}}{3R^2}(R^3 - a^3)\,\mathbf{a}_R$$

$$\mathbf{E} = \frac{\mathbf{D}}{\varepsilon_o} = \frac{\rho_{vo}}{3\varepsilon_o R^2}(R^3 - a^3)\,\mathbf{a}_R\,. \tag{3-49}$$

(c) In the region $R \le a$, no charge is enclosed by the Gaussian surface of radius $R_3 \le a$, and thus $D_R = 0$. In the region $R \le a$, we have

$$\mathbf{D} = 0 = \mathbf{E}\,. \tag{3-50}$$

Even if the closed surface integral of **D** is always zero regardless of the closed surface, in a region of space, it does not necessarily means that $\mathbf{D} = 0$ at every point in the region. In this particular example, however, **D** is known to be of the form $\mathbf{D} = D_R(R)\,\mathbf{a}_R$. Thus, if the closed surface integral of **D** over the Gaussian surface is zero, we have $D_R = 0$ at every point on the Gaussian surface, and thus $\mathbf{D} = 0$ in the given region.

The magnitude of the electric field intensity given in Eqs. (3-48)-(3-50) is normalized by $\rho_{vo} / 3\varepsilon_o$, and plotted as a function of R for $a = 2$ and $b = 4$ in Fig. 3.11. We note that $|\mathbf{E}|$ is smooth in each region, and continuous everywhere.

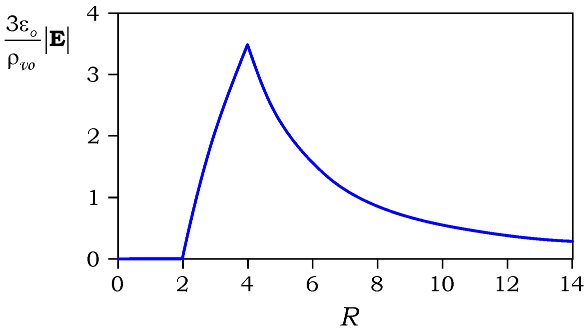


Fig. 3.11 Normalized $|\mathbf{E}|$ versus R

Let us rewrite **E** expressed by Eq. (3-48) as

$$\mathbf{E} = \frac{1}{4\pi\varepsilon_o R^2}\left[\rho_{vo}\frac{4\pi}{3}(b^3 - a^3)\right]\mathbf{a}_R = \frac{Q}{4\pi\varepsilon_o R^2}\mathbf{a}_R \tag{3-51}$$

The term in bracket represents the total charge Q contained in the spherical shell. In view of Eq. (3-51), we note that, in the region $R \geq b$, the electric field intensity is the same as if all charges were concentrated on a point at the center. This is because the spherical shell and the point have the same spherical symmetry.

Example 3-8

Determine **E** due to an infinitely long and straight line charge of a uniform density $\rho_{\ell o}$, which is along the z-axis in free space as shown in Fig. 3.12.

Solution

As was discussed in **Example 3-5**, the infinitely long line charge has cylindrical, translational, and twofold rotational symmetries. Thus the resulting **D** is expected to be of the form $\mathbf{D} = D_\rho(\rho)\,\mathbf{a}_\rho$. In view of the functional form of **D**, we construct a Gaussian surface in the form of a cylinder of radius ρ_1 and length $\mathcal{L}$.

On the side surface, the differential area vector $d\mathbf{s} = \rho_1\, d\phi dz\, \mathbf{a}_\rho$, and thus

$$\mathbf{D}\bullet d\mathbf{s} = D_\rho \rho_1\, d\phi dz$$

On the top and bottom plates of the cylinder, **D** is normal to $d\mathbf{s}$. Thus,

$$\mathbf{D}\bullet d\mathbf{s} = 0$$

The net outward electric flux through the Gaussian surface is therefore

$$\begin{aligned}\oint_S \mathbf{D}\bullet d\mathbf{s} &= \int_{side}\mathbf{D}\bullet d\mathbf{s} + \int_{top}\mathbf{D}\bullet d\mathbf{s} + \int_{bottom}\mathbf{D}\bullet d\mathbf{s} \\ &= \int_{\phi=0}^{2\pi}\int_{z=-\mathcal{L}/2}^{\mathcal{L}/2} D_\rho \rho_1\, d\phi dz = 2\pi D_\rho \rho_1 \mathcal{L}\end{aligned} \tag{3-52}$$

Next, the total charge enclosed by the Gaussian surface is

$$Q = \int_{z=-\mathcal{L}/2}^{z=\mathcal{L}/2} \rho_\ell \, dz = \rho_{\ell o}\mathcal{L} \tag{3-53}$$

From Eqs. (3-52) and (3-53), we obtain

$$D_\rho = \frac{\rho_{\ell o}}{2\pi\rho_1}$$

Omitting subscript 1 in ρ_1 for generalization, **D** and **E** due to the line charge density are

$$\mathbf{D} = D_\rho \mathbf{a}_\rho = \frac{\rho_{\ell o}}{2\pi\rho}\mathbf{a}_\rho$$

$$\mathbf{E} = \frac{\mathbf{D}}{\varepsilon_o} = \frac{\rho_{\ell o}}{2\pi\varepsilon_o\rho}\mathbf{a}_\rho \tag{3-54}$$

The line charge density $\rho_{\ell o}$ should not be confused with ρ that is the radial distance in cylindrical coordinates. Eq. (3-54) is the same as Eq. (3-29) that was obtained by Coulomb's law.

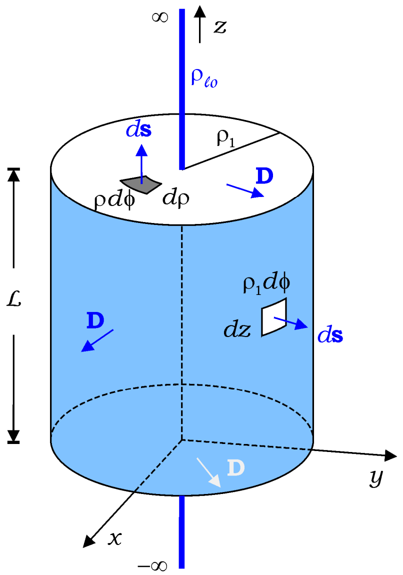


Fig. 3.12 Gaussian surface for an infinitely long line charge

Example 3-9
Determine **E** due to an infinite sheet of a uniform surface charge density ρ_{so}, which coincides with the $z = 0$ plane as shown in Fig. 3.13.

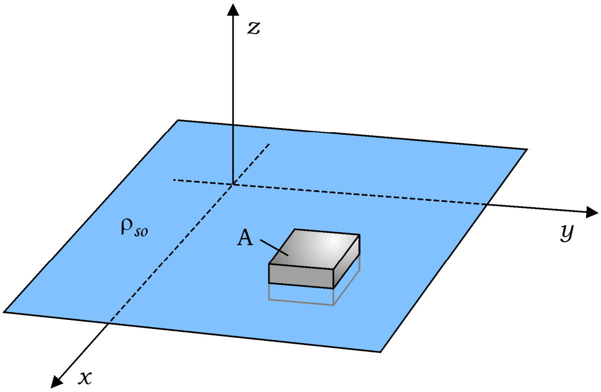


Fig. 3.13 An infinite sheet of a uniform surface charge density.

Solution

We explore the charge distribution for symmetry. The infinite sheet of surface charges has the following symmetries:

(1) Translational symmetries in the x- and y-directions such that **D** is independent of x and y.

(2) Rotational symmetry about the z-axis such that **D** has neither the x- nor the y-component.

(3) Twofold rotational symmetry about the x-axis such that

$\mathbf{D} = D_z(z)\,\mathbf{a}_z$ for $z > 0$, and

$\mathbf{D} = -D_z(z)\,\mathbf{a}_z$ for $z < 0$.

In view of the functional form of **D**, we construct a Gaussian surface in the form of a rectangular parallelepiped, which extends across the sheet of the charge, as shown in Fig. 3.13.

The electric flux through each face of the parallelepiped is

$$\int_{top} \mathbf{D} \cdot d\mathbf{s} = \int_{top} D_z \mathbf{a}_z \cdot dxdy\, \mathbf{a}_z = AD_z$$

$$\int_{bottom} \mathbf{D} \cdot d\mathbf{s} = \int_{bottom} -D_z \mathbf{a}_z \cdot (-dxdy)\, \mathbf{a}_z = AD_z$$

$$\int_{sides} \mathbf{D} \cdot d\mathbf{s} = 0$$

The net outward electric flux through the Gaussian surface is

$$\oint_S \mathbf{D} \cdot d\mathbf{s} = 2AD_z \tag{3-55}$$

Next, the total charge enclosed by the Gaussian surface is

$$Q = \int_A \rho_{so} ds = \rho_{so} A \tag{3-56}$$

Equating Eq. (3-55) with Eq. (3-56) we obtain

$$D_z = \frac{\rho_{so}}{2}$$

Therefore, in the region $z > 0$,

$$\mathbf{D} = \frac{\rho_{so}}{2}\mathbf{a}_z \qquad [\mathrm{C/m^2}]$$

$$\boxed{\mathbf{E} = \frac{\rho_{so}}{2\varepsilon_o}\mathbf{a}_z} \qquad [\mathrm{V/m}] \qquad (3\text{-}57)$$

Similarly, in the region $z < 0$,

$$\mathbf{D} = -\frac{\rho_{so}}{2}\mathbf{a}_z \qquad [\mathrm{C/m^2}]$$

$$\boxed{\mathbf{E} = -\frac{\rho_{so}}{2\varepsilon_o}\mathbf{a}_z} \qquad [\mathrm{V/m}] \qquad (3\text{-}58)$$

If the sheet of surface charges is finite in extent, it has no such symmetries as (1) and (2). In that case, it is impossible to construct a Gaussian surface, and thus we use Coulomb's law for finding **E**.

Exercise 3.10

Two infinite sheets of uniform surface charges ρ_{so} and $-\rho_{so}$ are coincident with the $z = 3$ and $z = 0$ planes, respectively, in free space. Find **E** everywhere.

Ans. $\mathbf{E} = -(\rho_{so}/\varepsilon_o)\mathbf{a}_z$ for $0 < z < 3$, and $\mathbf{E} = 0$ for $z > 3$ or $z < 0$.

Exercise 3.11

A uniform volume charge density ρ_{vo} forms a sphere of radius a in free space. Find $\nabla \bullet \mathbf{D}$ in the regions (a) $R < a$, and (b) $R > a$.

Ans. (a) $\nabla \bullet \mathbf{D} = \rho_{vo}$, (b) $\nabla \bullet \mathbf{D} = 0$.

Exercise 3.12

A hollow sphere with a negligible thickness carries a uniform surface charge density in free space. Show $\mathbf{E} = 0$ inside the sphere.

Ans. From spherical symmetry $\mathbf{D} = D_R(R)\mathbf{a}_R$. In the interior, $4\pi R^2 D_R = 0$.

Review Questions with Hints

RQ 3.8 State Gauss's law in words. [Eq.(3-41)]

RQ 3.9 Under what conditions can Gauss's law be used for **E** ? [Fig.3.12]

RQ 3.10 Does $\nabla \bullet \mathbf{D} = 0$ at a point p_1 mean $\mathbf{E} = 0$ at the point? [Eqs.(3-12)(3-38)]

RQ 3.11 Are electric flux lines of an arbitrary distribution of charges always smooth in free space? [Eq.(3-13),Fig.3.8]

3.4 Electric Potential

In the previous sections we saw that Coulomb's law and Gauss's law provide the direct method of determining the electric field of a given charge distribution, although Coulomb's law involves summing vectors, which may not be straightforward in many cases, and the application of Gauss's law is limited to a charge distribution having a certain symmetry. In search of an indirect method of obtaining the electric field, we define an electric potential from the relation between an electric force and the work done. The electric potential is a scalar quantity, which can be obtained from a charge distribution or another electric potential, and is thus much easier to deal with. Moreover, the electric potential carries physical significance; namely, the work done in moving a unit charge from a point in space to another in an electric field. It also plays an important role in boundary value problems, in which we determine the electric field in a region of space from the charges or electric potentials specified at the boundaries.

3.4.1 Work Done in Moving a Charge

If an electric field exists in a region of space, we can move a charge from a point in space to another only after the electric force acting on the charge is offset by an external force, a mechanical force, for instance. Let us consider a situation in which we move a charge q in the direction of $\mathbf{a}_l$ in an electric field $\mathbf{E}$. To start with, we compute the electric force acting on q:

$$\mathbf{F}_e = q\,\mathbf{E} \qquad \text{[N]}$$

To cancel out $\mathbf{F}_e$, we need to apply an external force $\mathbf{F}_{ext}$ to the charge in such a way that we get $\mathbf{F}_e + \mathbf{F}_{ext} = 0$. Therefore,

$$\mathbf{F}_{ext} = -q\,\mathbf{E} \qquad \text{[N]}$$

When we move a charge q by a small distance dl in the $\mathbf{a}_l$-direction, the work done is

$$dW = \mathbf{F}_{ext} \bullet \mathbf{a}_l\, dl = -q\left(\mathbf{E} \bullet \mathbf{a}_l\right) dl \tag{3-59}$$

Upon substituting $d\mathbf{l} = dl\,\mathbf{a}_l$ into Eq. (3-59), the work done, or energy expended, in moving a charge q by a vector differential length $d\mathbf{l}$ in an electric field $\mathbf{E}$ is

$$\boxed{dW = -q\,\mathbf{E} \bullet d\mathbf{l}} \qquad \text{[J]} \tag{3-60}$$

It is important to note that the minus sign on the right-hand side of Eq. (3-60) is to make a positive value of dW on the left-hand side the work done by the external force, or by us.

The total work done in carrying a charge q from point B to point A in an electric field $\mathbf{E}$ is therefore written as follows:

$$\boxed{W = \int_B^A dW = -q\int_B^A \mathbf{E}\bullet d\mathbf{l}} \qquad \text{[J]} \qquad (3\text{-}61)$$

Although the work done is a scalar, it involves a sense of direction: a positive W represents the work done by us, while a negative W represents the work done by the electric field.

Exercise 3.13

In an electric field $\mathbf{E} = y\,\mathbf{a}_x + 2x\,\mathbf{a}_y$, a charge located at point $(2, 3, 0)$ is moved a unit distance in an $\mathbf{a}_l$-direction. Find $\mathbf{a}_l$ for the maximum work done.

Ans. $\mathbf{a}_l = -(3/5)\,\mathbf{a}_x - (4/5)\,\mathbf{a}_y$.

3.4.2 Electric Potential due to a Charge Distribution

The potential difference V_{A-B} is defined as the work done in carrying a positive unit charge from initial point B to terminal point A. In an electric field $\mathbf{E}$, the potential difference between two points A and B is written as

$$\boxed{V_{A-B} = -\int_B^A \mathbf{E}\bullet d\mathbf{l}} \qquad \text{[V]} \qquad (3\text{-}62)$$

which has the unit of the volt[V]. As was mentioned earlier, in our notations, subscript A-B stands for something “from B to A”.

The absolute electric potential, or electric potential, is the potential difference between a given point in space and the zero reference point, which is usually chosen to be at infinity. Otherwise, the zero reference point is specified in the given system as a point, a line, or a surface in space. The electric potential at point A is defined as

$$\boxed{V_A = -\int_\infty^A \mathbf{E}\bullet d\mathbf{l}} \qquad \text{[V]} \qquad (3\text{-}63)$$

The potential difference is expressed in terms of the electric potentials as

$$\boxed{V_{A-B} = V_A - V_B = -\int_\infty^A \mathbf{E}\bullet d\mathbf{l} + \int_\infty^B \mathbf{E}\bullet d\mathbf{l}} \qquad \text{[V]} \qquad (3\text{-}64)$$

The electric potential is generally given as a function of position; it forms a scalar field in a region of space.

Let us now consider the electric potential due to a point charge q located at the origin. Substituting the electric field expressed by Eq. (3-12) into Eq. (3-63), the electric potential at point A:(R_1, θ_1, ϕ_1) in spherical coordinates is

$$V = -\int_{\infty}^{A} \mathbf{E} \bullet d\mathbf{l} = -\int_{\infty}^{R_1} \frac{q}{4\pi\varepsilon_o R^2} dR = \frac{q}{4\pi\varepsilon_o R_1} \tag{3-65}$$

where we used $d\mathbf{l} = dR\,\mathbf{a}_R + R\,d\theta\,\mathbf{a}_\theta + (R\sin\theta)\,d\phi\,\mathbf{a}_\phi$. Omitting 1 in R_1 for generalization, the electric potential of a point charge located at the origin is

$$\boxed{V = \frac{q}{4\pi\varepsilon_o R}} \qquad \text{[V]} \tag{3-66}$$

We see from Eq. (3-66) that V depends on R only, implying that V is independent of the path of integration.

Wirth reference to Fig. 3.14, we can show that the path independence of the potential difference is rooted in the law of conservation of energy. Consider two separate paths between two points A and B in an electric field as shown in Fig. 3.14. If the potential difference V_{A-B} depended on the path, we would carry a charge from B to A along a path that demands less work done, and return to B along the other path gaining a net energy. To satisfy the law of conservation of energy, ***the potential difference should be independent of the path of integration***.

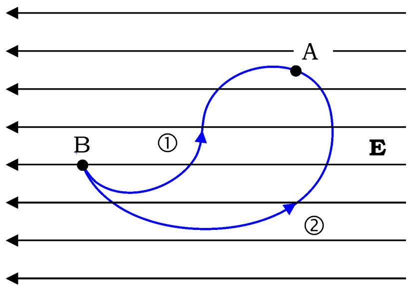


Fig. 3.14 V_{A-B} is independent of the path of integration.

The electric potential obeys the principle of superposition as the electric field. Thus the electric potential due to N point charges q_1, q_2,…, q_N located at points with positions vector $\mathbf{r}_1$, $\mathbf{r}_2$,…, $\mathbf{r}_N$ is expressed as

$$\boxed{V = \frac{1}{4\pi\varepsilon_o} \sum_{j=1}^{N} \frac{q_j}{\left|\mathbf{r} - \mathbf{r}_j\right|}} \tag{3-67}$$

The electric potential due to a continuous distribution of charges can be obtained from Eq. (3-67) by replacing q_j with $\rho_{v'}dv'$, $\rho_{s'}ds'$, or $\rho_{\ell'}d\ell'$, substituting $\mathbf{r} - \mathbf{r}_j$ with $\mathcal{R} = \mathbf{r} - \mathbf{r}'$, and taking the limits as $N \to \infty$. Following this procedure we express the electric potential as follows:

$$V = \frac{1}{4\pi\varepsilon_o}\int_{v'} \frac{\rho_{v'}}{\mathcal{R}} dv' \qquad \text{(volume charge)} \qquad (3\text{-}68a)$$

$$V = \frac{1}{4\pi\varepsilon_o}\int_{s'} \frac{\rho_{s'}}{\mathcal{R}} ds' \qquad \text{(surface charge)} \qquad (3\text{-}68b)$$

$$V = \frac{1}{4\pi\varepsilon_o}\int_{L'} \frac{\rho_{\ell'}}{\mathcal{R}} d\ell' \qquad \text{(line charge)} \qquad (3\text{-}68c)$$

Note that the zero reference point is taken at infinity in the above equations.

The electric potential of a point charge varies as $1/R$, whereas the electric field varies as $1/R^2$.

Example 3-10

An infinitely long straight line is along the z-axis, carrying a uniform line charge density $\rho_{\ell o}$ in free space, as shown in Fig. 3.15. Find an expression for V_{A-B} between two points A:$(\rho_A, \phi_A, 0)$ and B:$(\rho_B, \phi_B, 0)$ in cylindrical coordinates.

Solution

We try two different methods for the solution: the first method uses Eq. (3-62), and the second one uses Eq. (3-68c) and (3-64).

(a) V_{A-B} from Eq. (3-62)

Inserting $\mathbf{E}$ of a line charge $\rho_{\ell o}$ expressed by Eq. (3-54) into Eq. (3-62), noting that $d\mathbf{l} = d\rho\,\mathbf{a}_\rho + \rho d\phi\,\mathbf{a}_\phi + dz\,\mathbf{a}_z$ in cylindrical coordinates, we get

$$V_{A-B} = -\int_B^A \mathbf{E}\cdot d\mathbf{l} = -\int_B^A \frac{\rho_{\ell o}}{2\pi\varepsilon_o\rho}\mathbf{a}_\rho \cdot \left(d\rho\,\mathbf{a}_\rho + \rho d\phi\,\mathbf{a}_\phi + dz\,\mathbf{a}_z\right)$$

$$= -\int_{\rho=\rho_B}^{\rho=\rho_A} \frac{\rho_{\ell o}}{2\pi\varepsilon_o\rho} d\rho$$

Here, ρ_A, ρ_B, and $\rho_{\ell o}$ are constants, while ρ is a variable.

The potential difference between two points A and B, due to a line charge density $\rho_{\ell o}$, is

$$V_{A-B} = \frac{\rho_{\ell o}}{2\pi\varepsilon_o}\ln\left(\frac{\rho_B}{\rho_A}\right) \qquad (3\text{-}69)$$

Here, ρ_A and ρ_B are radial distances in cylindrical coordinates, while $\rho_{\ell o}$ is the line charge density.

(b) V_{A-B} from Eqs. (3-68c) and (3-64)

By using the position vectors of the field and source points, $\mathbf{r} = \rho_A \mathbf{a}_\rho$ and $\mathbf{r}' = z'\mathbf{a}_{z'} = z'\mathbf{a}_z$, the distance vector and its magnitude are written as

$$\mathcal{R} = \mathbf{r} - \mathbf{r}' = \rho_A\, \mathbf{a}_\rho - z'\,\mathbf{a}_z$$

$$\mathcal{R} = \sqrt{\rho_A^2 + z'^{\,2}}$$

Inserting $d\ell' = dz'$ into Eq. (3-68c), the electric potential at point A is

$$V_A = \frac{1}{4\pi\varepsilon_o}\int_{z'=-\infty}^{z'=\infty} \frac{\rho_{\ell o}dz'}{\sqrt{\rho_A^2 + z'^{\,2}}} = \frac{\rho_{\ell o}}{4\pi\varepsilon_o}\ln\left[z' + \sqrt{\rho_A^2 + z'^{\,2}}\right]\Bigg|_{z'=-\infty}^{z'=\infty} = \infty$$

Similarly, the electric potential at point B is obtained as

$$V_B = \frac{1}{4\pi\varepsilon_o}\int_{z'=-\infty}^{z'=\infty} \frac{\rho_{\ell o}dz'}{\sqrt{\rho_B^2 + z'^2}} = \infty$$

Note that the potential difference cannot be expressed in terms of the electric potentials in this case.

For future reference

$$\boxed{\int \frac{dx}{\sqrt{x^2 + a^2}} = \ln\left(x + \sqrt{x^2 + a^2}\right).} \qquad (3\text{-}70)$$

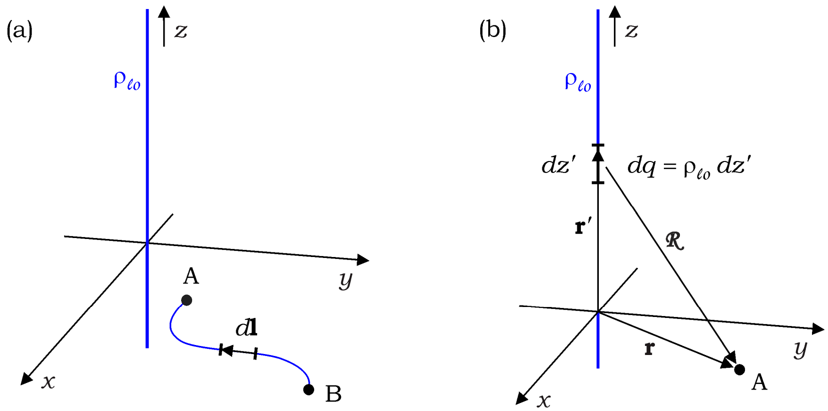


Fig. 3.15 An infinitely long line charge density.

Example 3-11

A uniform line charge density $\rho_{\ell o}$ forms a quadrant of a radius a in the $z = 0$ plane as shown in Fig. 3.16. Determine V at point $(0, 0, 1)$ in cylindrical coordinates.

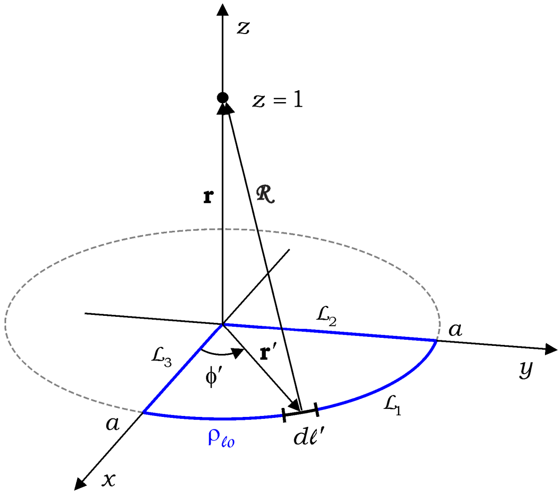


Fig. 3.16 Line charge density forming a quadrant.

Solution

Position vector of the field point on the z-axis is

$$\mathbf{r} = \mathbf{a}_z$$

In mixed coordinates, position vectors of the source points are expressed as

$$\mathbf{r}' = a\,\mathbf{a}_{\rho'} \qquad \text{on } \mathcal{L}_1$$
$$\mathbf{r}' = y'\,\mathbf{a}_{y'} \qquad \text{on } \mathcal{L}_2$$
$$\mathbf{r}' = x'\,\mathbf{a}_{x'} \qquad \text{on } \mathcal{L}_3$$

Magnitudes of the distance vectors are

$$\mathcal{R} = |\mathbf{r}-\mathbf{r}'| = |\mathbf{a}_z - a\,\mathbf{a}_{\rho'}| = \sqrt{1+a^2} \qquad \text{on } \mathcal{L}_1 \qquad (3\text{-}71\text{a})$$

$$\mathcal{R} = |\mathbf{r}-\mathbf{r}'| = |\mathbf{a}_z - y'\,\mathbf{a}_{y'}| = \sqrt{1+y'^2} \qquad \text{on } \mathcal{L}_2 \qquad (3\text{-}71\text{b})$$

$$\mathcal{R} = |\mathbf{r}-\mathbf{r}'| = |\mathbf{a}_z - x'\,\mathbf{a}_{x'}| = \sqrt{1+x'^2} \qquad \text{on } \mathcal{L}_3 \qquad (3\text{-}71\text{c})$$

Differential lengths are

$$d\ell' = a\,d\phi' \qquad \text{on } \mathcal{L}_1 \qquad (3\text{-}72\text{a})$$

$$d\ell' = dy' \qquad \text{on } \mathcal{L}_2 \qquad (3\text{-}72\text{b})$$

$$d\ell' = dx' \qquad \text{on } \mathcal{L}_3 \qquad (3\text{-}72\text{c})$$

Separating Eq. (3-68c) into three parts, we have

$$V = \frac{1}{4\pi\varepsilon_o}\int_{\mathcal{L}'}\frac{\rho_{\ell o}}{\mathcal{R}}d\ell' = \frac{\rho_{\ell o}}{4\pi\varepsilon_o}\left[\int_{\mathcal{L}_1}\frac{1}{\mathcal{R}}d\ell' + \int_{\mathcal{L}_2}\frac{1}{\mathcal{R}}d\ell' + \int_{\mathcal{L}_3}\frac{1}{\mathcal{R}}d\ell'\right] \tag{3-73}$$

Inserting Eqs. (3-71) and (3-72) into Eq. (3-73), we have

$$V = \frac{\rho_{\ell o}}{4\pi\varepsilon_o}\left[\int_{\phi'=0}^{\pi/2}\frac{1}{\sqrt{1+a^2}}a\,d\phi' + \int_{y'=0}^{a}\frac{1}{\sqrt{1+y'^2}}dy' + \int_{x'=0}^{a}\frac{1}{\sqrt{1+x'^2}}dx'\right]$$

$$= \frac{\rho_{\ell o}}{4\pi\varepsilon_o}\left[\frac{a\pi}{2\sqrt{1+a^2}} + \ln\left(y' + \sqrt{1+y'^2}\right)\Big|_{y'=0}^{a} + \ln\left(x' + \sqrt{1+x'^2}\right)\Big|_{x'=0}^{a}\right]$$

Thus,

$$V = \frac{\rho_{\ell o}}{4\pi\varepsilon_o}\left[\frac{a\pi}{2\sqrt{1+a^2}} + 2\ln\left(a + \sqrt{1+a^2}\right)\right]$$

It is important to note that the three integrals on the right-hand side of Eq. (3-73) should be evaluated as three definite integrals, not as three parts comprising a closed line integral. This is obvious from the fact that *a positive charge should result in a positive electric potential.* Thus, each integral in Eq. (3-73) is conducted along the line charge in the direction of increasing coordinate.

It should be noted that the electric potentials expressed by Eq. (3-63) and Eq. (3-68c) involve different kinds of integrals: V in Eq. (3-63) involves ***a line integral of* E,** while V in Eq. (3-68c) involves ***a definite integral of the line charge density.***

Exercise 3.14
With reference to ρ_{vo} in Fig. 3.10, find the electric potential in the region $R > b$.

Ans. $V = \dfrac{1}{4\pi\varepsilon_o R}\left[\rho_{vo}\dfrac{4\pi}{3}(b^3 - a^3)\right]$.

Exercise 3.15
Explain why **E** and V, due to ρ_{vo} in Fig. 3.10, can be obtained as if the charges were concentrated on the origin.

Ans. The electric flux density is always of the form $\mathbf{D} = D_R(R)\,\mathbf{a}_R$.

3.4.3 Conservative Field

To recapitulate, the potential difference V_{A-B} is the work done in carrying a unit charge from point B to point A, and is given by the negative line integral of **E** from point B to point A as shown in Eq. (3-62). If the direction of travel is reversed, such as from point A to point B, the negative line integral now represents the potential difference V_{B-A}, which is just equal to $-V_{A-B}$ simply because the limits of integration is interchanged. With reference to Fig. 3.14, consider the the situation in which we carry a unit charge from point B to point A along path ①, and return to point B along path ②. In this case, the work done is expressed as

$$V_{A-B}^{①} + V_{B-A}^{②} = -\int_{B}^{①A} \mathbf{E} \cdot d\mathbf{l} - \int_{A}^{②B} \mathbf{E} \cdot d\mathbf{l} \tag{3-74}$$

On the left-hand side of Eq. (3-74), we have $V_{A-B}^{①} = V_{A-B}^{②}$, because the potential difference is independent of the path, and $V_{B-A}^{②} = -V_{A-B}^{②}$, because the limits of integration are interchanged. Consequently, the left-hand side of Eq. (3-74) is zero. Meanwhile, the right-hand side of Eq. (3-37) is obviously the negative of the closed line integral of $\mathbf{E}$ around the closed path formed by ① and ②. In view of these, we conclude that the closed line integral of $\mathbf{E}$ is always zero, namely,

$$\boxed{\oint_C \mathbf{E} \cdot d\mathbf{l} = 0} \tag{3-75}$$

Since Eq. (3-75) is based on the law of conservation of energy, it is one of the fundamental relations for the static electric field.

Next, applying Stokes's theorem to Eq. (3-75), we have

$$\int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{s} = 0$$

Since the surface S may be arbitrary, the equality is satisfied only if the integrand is zero at every point on S. Thus, we have

$$\boxed{\nabla \times \mathbf{E} = 0} \tag{3-76}$$

The static electric field is an irrotational field, and called a conservative field.

It is important to note that Eqs. (3-75) and (3-76) are true only for the static electric field. Under time-varying conditions, neither the curl nor the closed line integral of the time-varying electric field is zero.

Exercise 3.16

If the term with $\mathbf{a}_\theta$ is missing in Eq. (3-20), may the remaining $\mathbf{E}(\mathbf{r})$ be a static electric field, which may be produced by some other charge distribution?

Ans. No, because $\nabla \times \mathbf{E} \neq 0$.

Exercise 3.17

For the three fields, $\mathbf{E}$ due to a point charge at the origin as Eq. (3-12), $\mathbf{E}'$ due to a point charge not at the origin, and $\mathbf{E}''$ due to arbitrary multiple charges, show $\nabla \times \mathbf{E} = 0$ from Eq. (3-12). What are the laws underlying $\nabla \times \mathbf{E}' = 0 = \nabla \times \mathbf{E}''$?

Ans. $\nabla \times \mathbf{E}$ is independent of coordinate system, and $\mathbf{E}$ satisfies superposition principle.

*3.4.4 **E** as the Negative Gradient of **V***

The differential of electric potential, dV, represents the work done in moving a unit charge along a differential length vector $d\mathbf{l}$, that is

$$dV = -\mathbf{E} \cdot d\mathbf{l} \tag{3-77}$$

Borrowing from section 2-2, the differential of a smooth function V, in general, is

$$dV = (\nabla V) \cdot d\mathbf{l} \tag{3-78}$$

This is the dot product between the gradient of V and the differentia length vector $d\mathbf{l}$(see Eq. (2-37)). Comparison of Eq. (3-77) with Eq. (3-78) leads to

$$\boxed{\mathbf{E} = -\nabla V} \qquad [\text{V/m}] \tag{3-79}$$

The static electric field equals the negative gradient of the electric potential. Equation (3-79) enables us to obtain $\mathbf{E}$ first by calculating V from a charge distribution, and then taking the negative gradient of V. We note that Eq. (3-79) conforms with the conservative nature of $\mathbf{E}$; that is, $\nabla \times \mathbf{E} = 0$.

The electric potential has the physical significance of the work done, and is thus a smooth function of position. Otherwise, its derivative, or the electric force, might show an abrupt change as a function of position. If an electric potential field V is present in a region of space, the spatial points of a constant V form a smooth surface, called an equipotential surface. This surface is always normal to the electric field lines for the obvious reason that no work should be done in moving a charge around on an equipotential surface.

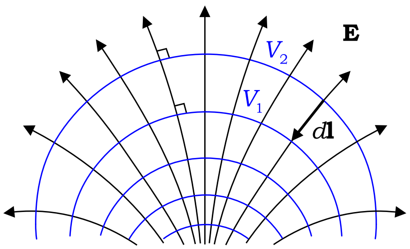


Fig. 3.17 Equipotential surfaces (blue lines) are perpendicular to electric field lines (black lines). $V_1 - V_2 = -\mathbf{E} \cdot d\mathbf{l}$.

Example 3-12

Referring to Fig. 3.4, find V and then $\mathbf{E}$ of an electric dipole located at the origin in free space.

Solution

We replace R in Eq. (3-66) with $\mathcal{R}^+$ or $\mathcal{R}^-$, which is the distance between the field point and a charge of the electric dipole. Applying the principle of superposition, the electric potential V due to the electric dipole is written as

$$V = \frac{q}{4\pi\varepsilon_o}\left(\frac{1}{\mathcal{R}^+} - \frac{1}{\mathcal{R}^-}\right) \tag{3-80}$$

Applying the law of cosines to triangles $(+q)op$ and $(-q)op$, and applying the binomial expansion to Eq. (3-15), under the condition $r >> d$, we obtain

$$\mathcal{R}^+ \cong r - \tfrac{1}{2}d\cos\theta \tag{3-81a}$$

$$\mathcal{R}^- \cong r + \tfrac{1}{2}d\cos\theta \tag{3-81b}$$

Inserting Eq. (3-81) into Eq. (3-80), we have

$$V \cong \frac{q}{4\pi\varepsilon_o}\frac{d\cos\theta}{(r - \frac{1}{2}d\cos\theta)(r + \frac{1}{2}d\cos\theta)} \cong \frac{1}{4\pi\varepsilon_o}\frac{qd\cos\theta}{r^2} \tag{3-82}$$

We define the electric dipole moment as $\mathbf{p} = q\,\mathbf{d}$, where $\mathbf{d}$ is a vector drawn from the point of $-q$ to the point of $+q$. With the aid of the electric dipole moment $\mathbf{p} = q\,\mathbf{d}$ and the position vector $\mathbf{r} = R\,\mathbf{a}_R$, we rewrite Eq. (3-82) as

$$\boxed{V = \frac{1}{4\pi\varepsilon_o}\frac{\mathbf{p}\bullet\mathbf{a}_R}{R^2}} \qquad [\mathrm{V}] \tag{3-83}$$

Taking the negative gradient of V in spherical coordinates, we have

$$\mathbf{E} = -\nabla V = -\left(\mathbf{a}_R\frac{\partial}{\partial R} + \mathbf{a}_\theta\frac{1}{R}\frac{\partial}{\partial\theta} + \mathbf{a}_\phi\frac{1}{R\sin\theta}\frac{\partial}{\partial\phi}\right)\left(\frac{qd\cos\theta}{4\pi\varepsilon_o R^2}\right)$$

The electric field due to an electric dipole located at the origin is

$$\boxed{\mathbf{E} = \frac{p}{4\pi\varepsilon_o R^3}(2\cos\theta\,\mathbf{a}_R + \sin\theta\,\mathbf{a}_\theta)} \qquad [\mathrm{V/m}] \tag{3-84}$$

Here, p is the magnitude of the dipole moment $\mathbf{p}$. Eq. (3-84) is of course the same as Eq. (3-20). Note that $\mathbf{E}$ of an electric dipole is inversely proportional to the cube of distance, whereas $\mathbf{E}$ of a point charge is inversely proportional to the square of distance.

By using a numerical method, we compute, from Eq. (3-84), the electric field lines and equipotential surfaces in the yz-plane, and plot them to scale in Fig. 3.18. The electric field lines intersect the equipotential surfaces at right angles. We note that the $z = 0$ plane is an equipotential surface of $V = 0$, as is evident from Eq. (3-83)($\mathbf{p}\bullet\mathbf{a}_R = 0$).

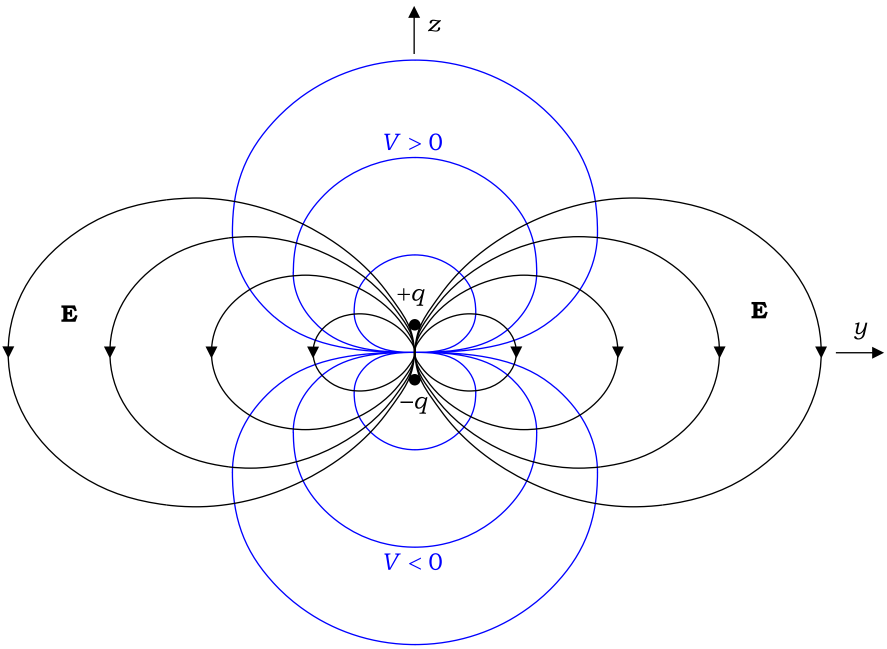


Fig. 3.18 Electric field lines (black lines) and equipotential surfaces (blue lines) of an electric dipole.

Example 3-13

A uniform surface charge density ρ_{so} forms a disc of radius a centered at the origin in the $z = 0$ plane, as shown in Fig. 3.19. Find V and $\mathbf{E}$ at point $(0, 0, b)$.

Solution

Using the position vectors of the field and source points, $\mathbf{r} = b\,\mathbf{a}_z$ and $\mathbf{r}' = \rho'\,\mathbf{a}_{\rho'}$, the distance between two points is written as

$$\mathcal{R} = |\mathbf{r} - \mathbf{r}'| = |b\,\mathbf{a}_z - \rho'\,\mathbf{a}_{\rho'}| = \sqrt{b^2 + \rho'^2}$$

Differential surface area on the disk is

$$ds' = \rho'\, d\rho' d\phi'$$

From Eq. (3-68b), we obtain

$$V = \frac{1}{4\pi\varepsilon_o}\int_{s'} \frac{\rho_{so}}{\mathcal{R}} ds' = \frac{\rho_{so}}{4\pi\varepsilon_o}\int_{\rho'=0}^{a}\int_{\phi'=0}^{2\pi} \frac{\rho'}{\sqrt{b^2 + \rho'^2}} d\rho' d\phi'$$
$$= \frac{\rho_{so}}{2\varepsilon_o}\left[\sqrt{b^2 + a^2} - b\right] \tag{3-85}$$

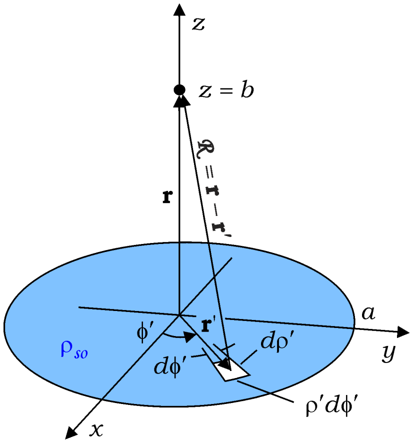


Fig. 3.19 A disk of a uniform surface charge density ρ_{so} .

Replacing b in Eq. (3-85) with z for generalization, the electric potential on the z-axis is

$$V = \frac{\rho_{so}}{2\varepsilon_o}\left[\sqrt{z^2+a^2} - z\right] \tag{3-86}$$

Since Eq. (3-86) is obtained by assuming $x = y = 0$, it has no directional derivative in the $\mathbf{a}_x$- or $\mathbf{a}_y$-direction. Nevertheless, we can obtain $\mathbf{E}$ from Eq. (3-86) by taking its negative gradient, because $\mathbf{E}$ of the surface charge only has the z–component on the z-axis. That is,

$$E_z = -\frac{dV}{dz} = \frac{\rho_{so}}{2\varepsilon_o}\left[1 - \frac{z}{\sqrt{z^2+a^2}}\right]$$

Thus, at point $(0, 0, b)$, we have

$$\mathbf{E} = \frac{\rho_{so}}{2\varepsilon_o}\left[1 - \frac{b}{\sqrt{b^2+a^2}}\right]\mathbf{a}_z\ . \tag{3-87}$$

Exercise 3.18

Describe the equipotential surfaces of (a) a point charge placed at the origin, and (b) an infinitely long line charge lying along the z-axis.

Ans. (a) Concentric spheres, (b) Coaxial circular cylinders.

Exercise 3.19

Justify $V_1 > V_2$ in Fig. 3.17.

Ans. $\mathbf{E} \bullet d\mathbf{l} < 0$.

Exercise 3.20

Justify that the $z = 0$ plane in Fig. 3.18 is an equipotential surface of $V = 0$.

Ans. For the $z = 0$ plane, $\mathbf{p} \bullet \mathbf{a}_R = 0$ in Eq. (3-83).

Review Questions with Hints

RQ 3.12 What is meant by a negative work done. [Eq.(3-60)]

RQ 3.13 Does $\mathbf{E} = 0$ at a point p_1 in space mean $V = 0$ at p_1? [Eq.(3-63)]

RQ 3.14 Does $V = 0$ at a point p_1 in space mean $\mathbf{E} = 0$ at p_1? [Eq.(3-63)]

RQ 3.15 Explain why $\nabla \times \mathbf{E} = 0$ is considered as a fundamental relation for electrostatics. [Fig.3.14]

RQ 3.16 What is the significance of the minus sign in $\mathbf{E} = -\nabla V$. [Fig.3.17]

RQ 3.17 Distinguish between two different kinds of integrals that may be involved in an expression for *V.* [Eqs.(3-63)(3-68)]

3.5 Dielectric in a Static Electric Field

Up to this point our discussion was limited to static electric fields in free space. We now turn our attention to material media placed in an externally applied electric field. In the atomic model, a material is viewed as an aggregate of atoms arranged in a three-dimensional array in free space. In the shell model of an atom, electrons occupy the shells around a nucleus in an ordered manner. The electrons at the outermost shell, called the valence electrons, are responsible for the electrical property of a material. According to their electrical properties, material media are classified into three categories such as conductor, semiconductor, and insulator. In the conductor, the binding force on the valence electrons is so weak that the electrons easily detach themselves from the atoms and migrate from atom to atom. These electrons are called the free electrons or the conduction electrons. In the presence of an externally applied electric field, the free electrons are accelerated in a short period of time before they collide with impurities or imperfections of the lattice, and scatter in random directions. As the result of these repeated acceleration and random scattering, the free electrons can move at a constant speed, constituting a steady conduction current in the conductor. In the insulator, or the dielectric, the valence electrons are tightly bound by the atomic force so that they may not be freed by an external electric field of a moderate strength. Nevertheless, the external field causes the valence electrons to be displaced with respect to the much heavier, positively charged nucleus, resulting in a separation of charges. These charges are called the bound charges. Although the bound charges cannot contribute to a conduction current, they induce electric dipoles in the material, each of which comprises of two identical charges of the opposite polarities separated by a small distance. The semiconductor contains a relatively small number of free electrons, and is thus positioned between the conductor and the insulator from the standpoint of conductivity.

Although the atoms of a dielectric may be electrically neutral under normal conditions, an externally applied electric field always induces electric dipoles in the material, whether or not the dipole moment is strong enough to be detected. At the macroscopic scale in which the discrete nature of the electric dipoles is ignored, the sum of the electric fields of the dipoles forms a polarization field, which is generally given as a function of position in the material. Since the polarization field is always opposite in direction to the external electric field, under static conditions, the sum of the two fields, called the internal electric field, is always smaller than the external field. The ratio between the external and internal fields is defined as the relative permittivity of the material, which is the characteristic parameter of a dielectric determining the electrical property of the material.

Polar molecules of some dielectrics exhibit permanent dipole moments even in the absence of an external electric field. The permanent dipole moment originates from unequal sharing of the valence electrons within a bond. For example, the water molecule is a polar molecule in which each hydrogen-oxygen bond is polar covalent in a bent structure. When there is no external electric field, polar molecules are randomly oriented, and thus yield no net dipole moment in the material. In the presence of an externally applied electric field, however, the molecular dipoles align with the electric field, resulting in a net dipole moment in the material. Whereas nonpolar molecules return to their original neutral state when the external electric field is removed, polar molecules may or may not return to the initial random state, depending on the material.

3.5.1 Electric Polarization

For an electric dipole, consisting of two charges of $+q$ and $-q$ separated by a small distance $|\mathbf{d}|$, the dipole moment is a vector defined as

$$\mathbf{p} = q\,\mathbf{d} \qquad [\text{C}\bullet\text{m}] \qquad (3\text{-}88)$$

where $\mathbf{d}$ is a vector drawn from the point of $-q$ to the point of $+q$. At the macroscopic scale in which the discrete nature of the electric dipole moments can be ignored, it is convenient to define the electric polarization $\mathbf{P}$, or the dipole moment density, as follows:

$$\mathbf{P}(\mathbf{r}) = \lim_{\Delta v \to 0} \frac{1}{\Delta v} \sum_{j=1}^{n\Delta v} \mathbf{p}_j \qquad [\text{C/m}^2] \qquad (3\text{-}89)$$

where n is the number density of dipoles(the number of dipoles per unit volume), and Δv is the incremental volume centered at position vector $\mathbf{r}$. Although the dipole moments, $\mathbf{p}_j$, are discrete vectors enclosed in Δv, the polarization vector $\mathbf{P}(\mathbf{r})$ is a smooth function of position in the material.

The electric polarization can induce electric charges on the surface of a dielectric. Consider Fig. 3.21, in which electric dipoles are induced in a dielectric by an external electric field. The induced dipole moments are directed along the direction of a unit vector $\mathbf{a}_p$, which is at an angle θ with respective to an outward unit normal to the surface, $\mathbf{a}_n$. In the figure, the black dot represents the center of an electric dipole whose charges are denoted by + and –, separated by a distance d. For the sake of argument, the dielectric is sliced into hypothetical layers of a thickness $\frac{1}{2}d\cos\theta$. As can be seen in Fig. 3.20(a), layer 2 contains an equal amount of the positive and negative charges that stem from the dipoles with their centers in layer 3 or 1. In contrast, as shown in Fig. 3.20(b), there are net positive charges in layer 1 and free space that stem from the dipoles with their centers in layer 2 or 1. The net positive charges in layer 1 and free space constitute a surface charge on the surface of the dielectric.

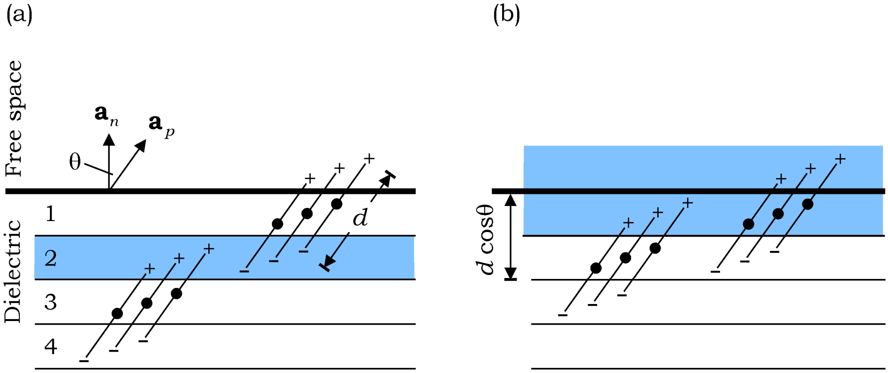


Fig. 3.20 The polarization surface charges induced by an electric polarization (a) No net charge in layer 2, (b) Net positive charges in layer 1 and free space.

For an incremental area Δs on the surface of the dielectric, the net surface charge contained in Δs is computed as follows:

$$\begin{aligned}\Delta Q &= qn[d\Delta s\cos\theta] \\ &= (qnd)(\Delta s)\,\mathbf{a}_p\bullet\mathbf{a}_n = \mathbf{P}\bullet\mathbf{a}_n(\Delta s)\end{aligned} \qquad (3\text{-}90)$$

where the term in bracket represents the volume of a parallelepiped, $\Delta s\times d\cos\theta[\mathrm{m}^3]$, which extends from the surface to the bottom of layer 2. In Eq. (3-90), q is the bound charge of an electric dipole, n is the volume number density of the dipoles, and the electric polarization $\mathbf{P}=(qnd)\mathbf{a}_p$. Dividing both sides of Eq. (3-90) by Δs, and taking the limit as $\Delta s\to 0$, we define the polarization surface charge density ρ_{PS}, i.e.,

$$\boxed{\rho_{PS}\equiv\mathbf{P}\bullet\mathbf{a}_n} \qquad [\mathrm{C/m^2}] \qquad (3\text{-}91)$$

Here, $\mathbf{P}$ is the electric polarization on the surface of the dielectric, and $\mathbf{a}_n$ is an outward unit normal to the surface. The polarization surface charge Q_{PS} is the total surface charge induced on the dielectric. i.e.,

$$
\begin{aligned}
\mathcal{Q}_{PS} &= \oint_{\mathcal{S}} \rho_{PS} ds \\
&= \oint_{\mathcal{S}} (\mathbf{P} \bullet \mathbf{a}_n) ds = \oint_{\mathcal{S}} \mathbf{P} \bullet d\mathbf{s}
\end{aligned} \tag{3-92}
$$

where $\mathcal{S}$ is the closed surface bounding the dielectric.

If the dielectric is electrically neutral, having no net charge, the polarization surface charge $\mathcal{Q}_{PS}$ should be counterbalanced by the volume charge inside the material, called the polarization volume charge $\mathcal{Q}_{PV}$. Namely,

$$
\begin{aligned}
\mathcal{Q}_{PV} = -\mathcal{Q}_{PS} &= -\oint_{\mathcal{S}} \mathbf{P} \bullet d\mathbf{s} \\
&= -\int_{\mathcal{V}} \nabla \bullet \mathbf{P} \, dv
\end{aligned} \tag{3-93}
$$

where the divergence theorem is used. In view of the volume integral in Eq. (3-93), we define the polarization volume charge density ρ_{PV} as

$$
\boxed{\rho_{PV} \equiv -\nabla \bullet \mathbf{P}} \qquad [\mathrm{C/m^3}] \tag{3-94}
$$

If the electric polarization $\mathbf{P}$ is constant in the dielectric such that $\nabla \bullet \mathbf{P} = 0$, then we have $\rho_{PV} = 0$ in the interior, and thus $\mathcal{Q}_{PV} = 0 = \mathcal{Q}_{PS}$. Although the net polarization volume charge and the net polarization surface charge may be zero, the polarization surface charge density ρ_{PS} may be nonzero on the surface of the dielectric.

The polarization charge densities ρ_{PS} and ρ_{PV} produce a polarization electric field in the dielectric such as

$$
\mathbf{E}_P = \frac{1}{4\pi\varepsilon_o} \oint_{\mathcal{S}'} \frac{\rho_{PS'}}{\mathcal{R}^2} \mathbf{a}_{\mathcal{R}} ds' + \frac{1}{4\pi\varepsilon_o} \int_{\mathcal{V}'} \frac{\rho_{PV'}}{\mathcal{R}^2} \mathbf{a}_{\mathcal{R}} dv' \tag{3-95}
$$

where $\mathcal{R}$ and $\mathbf{a}_{\mathcal{R}}$ are the magnitude and unit vector of $\mathcal{R} = \mathbf{r} - \mathbf{r}'$. In view of ε_o in Eq. (3-95), we see that the polarization charges ρ_{PS} and ρ_{PV} are assumed to be in free space as far as the polarization field $\mathbf{E}_p$ is concerned.

Example 3-14

An external electric field induces a constant electric polarization, $\mathbf{P} = P_o \mathbf{a}_z$, in a dielectric cylinder of radius *a* and height *b*, lying along the *z*-axis. Find

(a) polarization charge densities ρ_{PS} and ρ_{PV}, and

(b) polarization charges $\mathcal{Q}_{PS}$ and $\mathcal{Q}_{PV}$.

Solution

(a) From Eq. (3-91),

$$\rho_{PS} = \mathbf{P} \bullet \mathbf{a}_n = (P_o \mathbf{a}_z) \bullet \mathbf{a}_\rho = 0 \qquad \text{(side surface)}$$

$$\rho_{PS} = (P_o \mathbf{a}_z) \bullet (\pm \mathbf{a}_z) = \pm P_o \qquad (\pm\text{, top and bottom surfaces})$$

From Eq. (3-94),

$$\rho_{PV} = -\nabla \bullet \mathbf{P} = -\nabla (P_o\, \mathbf{a}_z) = 0\ .$$

(b) Total polarization surface charge is

$$Q_{PS} = \oint_S \rho_{PS} ds = \int_{side} \rho_{PS} ds + \int_{top} \rho_{PS} ds + \int_{bottom} \rho_{PS} ds$$
$$= 0 + P_o \pi a^2 - P_o \pi a^2 = 0$$

Total polarization volume charge is

$$Q_{PV} = \int_V \rho_{PV}\, dv = 0$$

The constant **P** induces ρ_{PS} only on the top and bottom surfaces.

Exercise 3.21
A constant electric polarization, $\mathbf{P} = P_o\, \mathbf{a}_x$, is induced in an isotropic medium by an external electric field **E**. Determine the direction of **E**.

Ans. $+\mathbf{a}_x$. If **P** is parallel to **E**, it is called an isotropic material.

Exercise 3.22
Show that ρ_{PV} is always zero in an isotropic material having no net charge

Ans. $\nabla \bullet \mathbf{D} = 0$ in the material, and thus $\nabla \bullet \mathbf{P} = 0$.

3.5.2 Dielectric Constant

When a net charge of $\rho_v\,[\mathrm{C/m^3}]$ is injected into a dielectric, its electric field behaves as an external field that induces the electric polarization **P** and the polarization volume charge density ρ_{PV} in the material. The sum of the external electric field due to ρ_v and the polarization electric field due to ρ_{PV} forms the internal electric field. From Eq. (3-44), the relation between the internal electric field and the charges in the material is written as

$$\begin{aligned} \nabla \bullet (\varepsilon_o \mathbf{E}) &= \rho_v + \rho_{PV} \\ &= \rho_v - \nabla \bullet \mathbf{P} \end{aligned} \qquad (3\text{-}96)$$

where we used Eq. (3-94) for ρ_{PV}. It should be noted that ρ_v and ρ_{PV} are assumed to reside in free space as far as the internal electric field is concerned. Rewriting Eq. (3-96), we have

$$\nabla \bullet (\varepsilon_o \mathbf{E} + \mathbf{P}) = \rho_v \qquad (3\text{-}97)$$

At this point, we redefine the electric flux density **D** in a dielectric such that

$$\boxed{\mathbf{D} = \varepsilon_o \mathbf{E} + \mathbf{P}} \qquad [\mathrm{C/m^2}] \qquad (3\text{-}98)$$

Another name for **D** is the displacement density. The newly defined electric flux density, if it is used in Eq. (3-97), enables us to express Gauss's law in the dielectric as

$$\boxed{\nabla \cdot \mathbf{D} = \rho_v} \qquad [\mathrm{C/m^3}] \qquad (3\text{-}99)$$

It is important to note that the electric flux density in the material is related, through Eq. (3-99), only to the net charge ρ_v that is imparted to the material, apart from the induced polarization charges. Therefore, ***Gauss's law is independent of the material***. Upon applying divergence theorem to Eq. (3-99), the integral form of Gauss's law is

$$\boxed{\oint_S \mathbf{D} \cdot d\mathbf{s} = Q} \qquad (3\text{-}100)$$

Gauss's law states that the net outward electric flux through any closed surface is equal to the net charge enclosed by that surface irrespective of the material surrounding the charge.

In a homogenous, linear, and isotropic dielectric, also called a simple medium, the electric polarization is directly proportional to the electric field in the material such that

$$\boxed{\mathbf{P} = \varepsilon_o \chi_e \mathbf{E}} \qquad [\mathrm{C/m^2}] \qquad (3\text{-}101)$$

The constant χ_e is called the electric susceptibility. In general, χ_e may vary with position(inhomogeneous), and depend on the magnitude(nonlinear medium) and direction(anisotropic) of **E**. In the scope of this book, we only deal with simple media for which χ_e is constant, and independent of the magnitude and direction of **E**. Inserting Eq. (3-101) into Eq. (3-98) leads to

$$\mathbf{D} = \varepsilon_o (1 + \chi_e)\mathbf{E} \equiv \varepsilon_o \varepsilon_r \mathbf{E} \equiv \varepsilon \mathbf{E} \qquad [\mathrm{C/m^2}] \qquad (3\text{-}102)$$

The constitutive relation of a material is obtained as

$$\boxed{\mathbf{D} = \varepsilon \mathbf{E}} \qquad (3\text{-}103)$$

Here , ε is the permittivity of the material, ε_r is the relative permittivity or the dielectric constant, and ε_o is the permittivity of free space. The dielectric constant is a measure of the electrical property of a material medium. If ε_r of a material is given, the electric susceptibility can be obtained as $\chi_e = \varepsilon_r - 1$ from Eq. (3-102).

Consider Fig. 3.21, in which a dielectric is placed is an external electric field $\mathbf{E}_o$. Although the induced polarization vector **P** is parallel to the electric field in the material, the polarization field due to **P** is opposite in direction to the electric field such that the static internal field is always weaker than the static external field. The dielectric constant is a proportionality factor between the external and internal electric fields in the material. A larger dielectric constant means that the atoms of the material are more susceptible to an electric field, producing a stronger polarization field and thus a weaker internal field in the material.

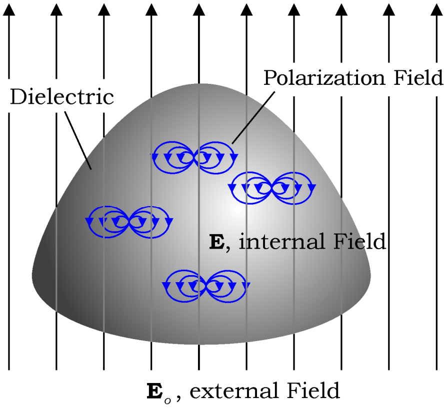


Fig. 3.21 The internal field **E** is the sum of the external and polarization fields.

Example 3-15

A net charge of $+Q$[C] is imparted to a conducting sphere of radius a. The sphere is then enclosed by a spherical dielectric shell of ε_r, with inner radius b and outer radius c, as shown in Fig. 3.22(a). Find the polarization surface charge densities on two surfaces of the dielectric shell.

Solution

From the spherical symmetry of the system, **D** is expected to be $\mathbf{D} = D_R(R)\,\mathbf{a}_R$ everywhere, if $D_R \neq 0$.

In the region $b \le R \le c$, from Gauss's law we obtain

$$4\pi R^2 D_R = 4\pi R^2(\varepsilon_o \varepsilon_r E_R) = +Q$$

Thus,

$$\mathbf{E} = E_R\,\mathbf{a}_R = \frac{Q}{4\pi\varepsilon_o\varepsilon_r R^2}\mathbf{a}_R \tag{3-104}$$

Inserting Eq. (3-104) into Eq. (3-101), with the aid of $\chi_e = \varepsilon_r - 1$, we get

$$\mathbf{P} = \varepsilon_o \chi_e \mathbf{E} = \frac{\varepsilon_r - 1}{\varepsilon_r}\frac{Q}{4\pi R^2}\mathbf{a}_R \tag{3-105}$$

Inserting Eq. (3-105) into Eq. (3-91), noting that the outward unit normal $\mathbf{a}_n = -\mathbf{a}_R$ on the inner surface at $R = b$, we get

$$\rho_{PS1} = \mathbf{P} \cdot \mathbf{a}_n = -\frac{\varepsilon_r - 1}{\varepsilon_r} \frac{Q}{4\pi b^2} \qquad \text{(at } R = b\text{)} \qquad (3\text{-}106)$$

Inserting Eq. (3-105) into Eq. (3-91), noting that the outward unit normal $\mathbf{a}_n = \mathbf{a}_R$ on the outer surface at $R = c$, we get

$$\rho_{PS2} = \mathbf{P} \cdot \mathbf{a}_n = \frac{\varepsilon_r - 1}{\varepsilon_r} \frac{Q}{4\pi c^2} \qquad \text{(at } R = c\text{)} \qquad (3\text{-}107)$$

From Eq. (3-105), in the dielectric, $\nabla \cdot \mathbf{P} = 0$ and therefore $\rho_{PV} = 0$.

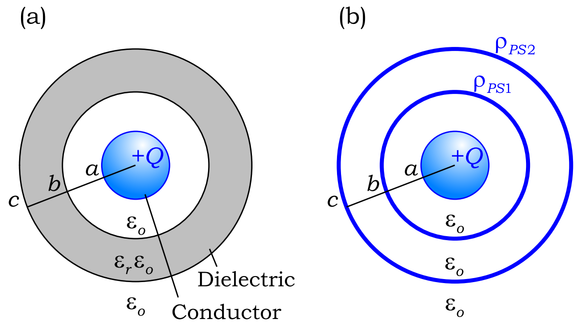


Fig. 3.22 A net charge is enclosed by a dielectric shell.

As far as the internal electric field is concerned, the dielectric may be replaced with the induced polarization charges residing in free space, as shown in Fig. 3.22(b). Applying Gauss's law in the region $b < R < c$, we have

$$4\pi R^2(\varepsilon_o E_R) = Q + 4\pi b^2 \rho_{PS1} \qquad (3\text{-}108)$$

Inserting Eq. (3-106) into Eq. (3-108), we obtain

$$\mathbf{E} = \frac{Q}{4\pi\varepsilon_o\varepsilon_r R^2}\mathbf{a}_R \qquad (b < R < c) \qquad (3\text{-}109)$$

Equation (3-109) is the same as Eq. (3-104). In the region $b < R < c$, $\mathbf{D}$ is obtained by multiplying $\mathbf{E}$ in Eq. (3-109) by the permittivity $\varepsilon = \varepsilon_o\varepsilon_r$. It should be remembered that $\mathbf{D}$ cannot be obtained directly from ρ_{PS1} and ρ_{PS2}.

Exercise 3.23
In view of the permanent dipole moment of water molecules, explain why ε_r of water decreases with temperature?

Ans. Random orientation due to thermal agitation.

Exercise 3.24
Can ε_r be less than one in an isotropic material under static conditions?

Ans. No, **P** is always parallel to **E**.

Exercise 3.25
Explain why ρ_{PS} does not have to be included in Eq. (3-96) or (3-99)?

Ans. $\nabla \cdot (\varepsilon_o \mathbf{E}) = 0$ for **E** of ρ_{PS}.

3.5.3 Boundary Conditions at a Dielectric Interface

By now, we should have a good understanding of the two relations, $\oint_S \mathbf{D} \cdot d\mathbf{s} = Q$ and $\oint_C \mathbf{E} \cdot d\mathbf{l} = 0$, which represent Gauss's law and irrotational nature of **E**. They are two fundamental relations for static electric fields in the sense that their point forms allow us to uniquely determine **E** in a region of space, in accordance with Helmholtz's theorem, regardless of the material. To be specific, Gauss's law is independent of the material because of the new definition of **D** in the material, and the irrotational nature of **E** is rooted in the law of conservation of energy, which is independent of the material.

In a homogeneous material, the static electric field is given as a smooth function of position, meaning that the direction of **E** does not change abruptly from a point in space to another. This is evident from the fact that **E** of a point charge is a smooth function of position, and **E** obeys the principle of superposition. However, this is not the case for **E** at an interface between two materials of different permittivities. The conditions for **E** and **D** at an interface are called the boundary conditions.

To obtain the boundary conditions for **E** and **D**, let us consider an interface formed by two adjoining dielectrics of permittivities ε_1 and ε_2 as shown in Fig. 3.23. We first compute the circulation of **E** around a rectangular loop $abcda$ by assuming that $\mathbf{E}_1$ and $\mathbf{E}_2$ are the electric field intensities on the upper and lower sides of the loop, respectively. As the height Δh tends to zero, $\mathbf{E}_1$ and $\mathbf{E}_2$ represent the electric field intensities on the opposite sides of the interface. According to the irrotational nature of **E**, the circulation of **E** around the loop should be zero, namely,

$$\begin{aligned}\oint_{abcd} \mathbf{E} \cdot d\mathbf{l} &= \int_a^b \mathbf{E} \cdot d\mathbf{l} + \int_b^c \mathbf{E} \cdot d\mathbf{l} + \int_c^d \mathbf{E} \cdot d\mathbf{l} + \int_d^a \mathbf{E} \cdot d\mathbf{l} \\ &= (E_{1t})\Delta w + \int_b^c \mathbf{E} \cdot d\mathbf{l} + (-E_{2t})\Delta w + \int_d^a \mathbf{E} \cdot d\mathbf{l} \\ &= 0\end{aligned} \tag{3-110}$$

where subscript t stands for the tangential component. As the height Δh goes to zero, the line integrals along the left and right sides of the loop, bc and da, vanish in Eq. (3-110). Under these conditions, Eq. (3-110) becomes

$$(E_{1t})\Delta w + (-E_{2t})\Delta w = 0$$

The boundary condition for the tangential component of **E** is therefore

$$\boxed{E_{1t} = E_{2t}} \qquad [\text{V/m}] \qquad (3\text{-}111)$$

The tangential component of* E *is continuous across the interface between two different dielectrics. Applying Eq. (3-103) to Eq. (3-111), the boundary condition for the tangential component of **D** is

$$\frac{D_{1t}}{\varepsilon_1} = \frac{D_{2t}}{\varepsilon_2} \qquad (3\text{-}112)$$

The tangential component of **D** is discontinuous across an interface formed by two different dielectrics.

We next apply Gauss's law to the circular cylinder of a cross section Δs and a height Δh, as shown in Fig. 3.23, by assuming that $\mathbf{D}_1$ and $\mathbf{D}_2$ are the electric flux densities on the top and bottom surfaces of the cylinder. As the height Δh tends to zero, $\mathbf{D}_1$ and $\mathbf{D}_2$ represent the electric flux densities on the opposite sides of the interface. The integral of **D** over the bounding surface of the cylinder is equal to the surface charge enclosed in the cylinder, namely,

$$\begin{aligned} \oint_S \mathbf{D} \cdot d\mathbf{s} &= \int_{top} \mathbf{D} \cdot d\mathbf{s} + \int_{bottom} \mathbf{D} \cdot d\mathbf{s} + \int_{side} \mathbf{D} \cdot d\mathbf{s} \\ &= D_{1n}\Delta s - D_{2n}\Delta s + \int_{side} \mathbf{D} \cdot d\mathbf{s} \\ &= \rho_S \Delta s \end{aligned} \qquad (3\text{-}113)$$

where subscript n stands for the normal component. In the above equation, ρ_S is the surface charge density on the interface, and thus $\rho_S\Delta s$ is the net surface charge enclosed in the cylinder as the height Δh goes to zero. The integral of **D** over the side surface of the cylinder vanishes as Δh tends to zero. Under these conditions, Eq. (3-113) becomes

$$D_{1n}\Delta s - D_{2n}\Delta s = \rho_S \Delta s$$

The boundary condition for the normal component of **D** is therefore

$$\boxed{D_{1n} - D_{2n} = \rho_S} \qquad [\text{C/m}^2] \qquad (3\text{-}114)$$

If there is no surface charge on the interface, $\rho_s = 0$, we have

$$\boxed{D_{1n} = D_{2n}} \tag{3-115}$$

The normal component of* D *is continuous across the interface between two different dielectrics, if there is no surface charge at the interface. Applying Eq. (3-103) to Eq. (3-115), the boundary condition for the normal component of **E** is

$$\varepsilon_1 E_{1n} = \varepsilon_2 E_{2n} \tag{3-116}$$

The normal component of **E** is discontinuous across the interface between two different materials.

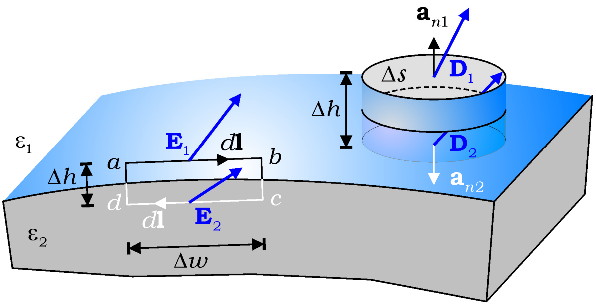


Fig. 3.23 An interface between two different dielectrics of permittivites ε_1 and ε_2.

Example 3-16

Two vectors $\mathbf{E}_1$ and $\mathbf{E}_2$ represent the electric field intensities in two adjoining dielectrics with permittivities ε_1 and ε_2, respectively, as shown in Fig. 3.24. Find E_2 and θ_2 in terms of E_1, ε_1, ε_2, and θ_1, assuming no surface charge at the interface.

Solution

Two components of $\mathbf{E}_1$ that are normal and tangential to the interface are

$$E_{1n} = E_1 \cos\theta_1$$

$$E_{1t} = E_1 \sin\theta_1$$

The tangential component of **E** is continuous cross the interface such that

$$E_{2t} = E_{1t} = E_1 \sin\theta_1 \tag{3-117}$$

The normal component of **D** is continuous cross the interface such that

$$\begin{aligned} D_{2n} &= D_{1n} \\ &= \varepsilon_1 E_{1n} = \varepsilon_1 E_1 \cos\theta_1 \end{aligned} \tag{3-118}$$

Rewriting Eq. (3-118), with the aid of $D_{2n} = \varepsilon_2 E_{2n}$, we have

$$E_{2n} = \frac{\varepsilon_1}{\varepsilon_2} E_1 \cos\theta_1 \tag{3-119}$$

Combining Eqs. (3-117) and (3-119), we have

$$E_2 = \sqrt{(E_{2t})^2 + (E_{2n})^2} = E_1\sqrt{\sin^2\theta_1 + (\varepsilon_1/\varepsilon_2)^2\cos^2\theta_1}$$

The rotation angle of $\mathbf{E}_2$, with respect to the surface normal, is obtained from Eqs. (3-117) and (3-119) as

$$\theta_2 = \tan^{-1}\frac{E_{2t}}{E_{2n}} = \tan^{-1}\left[\frac{\varepsilon_2}{\varepsilon_1}\tan\theta_1\right] \tag{3-120}$$

In view of $\theta_1 > \theta_2$ in Fig. 3.24, we note that $\varepsilon_1 > \varepsilon_2$ in Eq. (3-120).

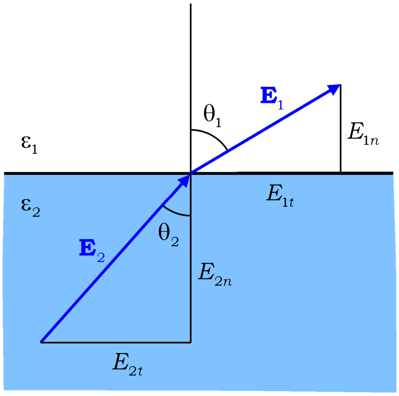


Fig. 3.24 An interface between two adjoining dielectrics.

Exercise 3.26
What is the boundary condition for the normal component of **D** at an interface between free space and a material with a volume charge density ρ_v.

Ans. $D_{1n} = D_{2n}$.

Review Questions with Hints

RQ 3.18 Sate the relation between the electric polarization and induced polarization charge in a material. [Eqs.(3-91)(3-94)]

RQ 3.19 What is Gauss's law in material media? [Eqs.(3-99)(3-100)]

RQ 3.20 What material parameter determines the polarization charge induced in a dielectric? [Eq.(3-101)]

RQ 3.21 State the relation between **D** and **E** in a dielectric. [Eq.(3-103)]

RQ 3.22 State the boundary conditions for E_t and D_n at an interface between two different dielectrics with no surface charge. [Eqs.(3-111)(3-115)]

RQ 3.23 Do static electric field lines reflect from an interface between two different dielectrics? [Fig.3.24]

RQ 3.24 Is the total electric flux conserved at the opposite sides of an interface between unlike dielectrics? [Eq.(3-115)]

RQ 3.25 Can you derive the boundary conditions for **D** and **E** from the point form of the two fundamental relations? [Eqs.(3-99)(3-76)]

3.6 Perfect Conductor in a Static Electric Field

Although the conductor contains a very large number of free electrons, it is electrically neutral under normal conditions. The negative charges of the free electrons are counterbalanced by the positive charges of the ionized lattice atoms. Thus there is no net charge in the conductor. In the absence of an externally applied electric field, the free electrons collide with the impurities and imperfections of the lattice, and scatter in random directions. Consequently, there is no net motion of the free electrons in the conductor. In contrast, in the presence of an external electric field, the free electrons gain a net speed during the mean time between collisions and give rise to a conduction current in the material. The conductivity is a measure of how easily the free electrons can produce the conduction current under the influence of an electric field. A perfect conductor is one with an infinite conductivity, whereas a perfect dielectric is one with zero conductivity.

If a net charge is injected into a perfect conductor, the charge should be distributed on the surface of the conductor in order not to create any electric field inside the conductor. Otherwise, an internal electric field would generate an infinite conduction current, which in turn redistributes the charge instantaneously until there is no electric field inside the conductor. In a perfect conductor,

$$\boxed{\mathbf{E} = 0} \tag{3-121}$$

$$\boxed{\rho_v = 0} \tag{3-122}$$

where ρ_v is the net volume charge density in the conductor.

Consider a perfect conductor placed in free space as shown in Fig. 3.25. The circulation of **E** around a rectangular loop $abcda$ is zero, irrespective of the material, namely,

$$\begin{aligned}\oint_{abcd} \mathbf{E}\cdot d\mathbf{l} &= \int_a^b \mathbf{E}\cdot d\mathbf{l} + \int_b^c \mathbf{E}\cdot d\mathbf{l} + \int_c^d \mathbf{E}\cdot d\mathbf{l} + \int_d^a \mathbf{E}\cdot d\mathbf{l} \\ &= (E_{1t})\Delta w + \int_b^c \mathbf{E}\cdot d\mathbf{l} + 0 + \int_d^a \mathbf{E}\cdot d\mathbf{l} \\ &= 0\end{aligned} \tag{3-123}$$

where t stands for the tangential component. The line integral along the bottom side of the loop, cd, is zero because $\mathbf{E} = 0$ in the perfect conductor. The line integrals along the left and right sides of the loop, da and bc, vanish as Δh tends to zero. Under these conditions, Eq. (3-123) becomes

$$(E_{1t})\Delta w = 0$$

Thus, on the surface of a perfect conductor, we have

$$\boxed{E_t = 0} \tag{3-124}$$

The tangential component of* E *is always zero on the surface of a perfect conductor. This implies that no work is done in moving a charge around on the surface of a perfect conductor; ***the surface of a perfect conductor is an equipotential surface***.

Next, we apply Gauss's law to a circular cylinder half embedded in the conductor as shown in Fig. 3.25. The integral of **D** over the bounding surface of the cylinder is equal to the surface charge enclosed in the cylinder, namely,

$$\begin{aligned}\oint_s \mathbf{D}\cdot d\mathbf{s} &= \int_{top} \mathbf{D}\cdot d\mathbf{s} + \int_{bottom} \mathbf{D}\cdot d\mathbf{s} + \int_{side} \mathbf{D}\cdot d\mathbf{s} \\ &= D_{1n}\Delta s + 0 + \int_{side} \mathbf{D}\cdot d\mathbf{s} \\ &= \rho_s \Delta s\end{aligned} \tag{3-125}$$

where n stands for the normal component. Here, ρ_s is the surface charge density on the perfect conductor, and $\rho_s\Delta s$ is thus the total surface charge enclosed in the cylinder as the height Δh goes to zero. The surface integral over the bottom plate is zero because $\mathbf{E} = 0 = \mathbf{D}$ in a perfect conductor. The surface integral over the side surface vanishes as $\Delta h \to 0$. Under these conditions, Eq. (3-125) becomes

$$D_{1n}\Delta s = \rho_s \Delta s$$

Thus, on the surface of a perfect conductor, we have

$$\boxed{D_n = \rho_s} \tag{3-126}$$

The normal component of* D *on the surface of a perfect conductor is equal to the net surface charge density on the conductor. It is important to note that ρ_s in Eq. (3-126) is the net surface charge on the conductor, except the polarization charges.

The boundary conditions expressed by Eqs. (3-124) and (3-126) can also be applied to an interface between a perfect conductor and a perfect dielectric.

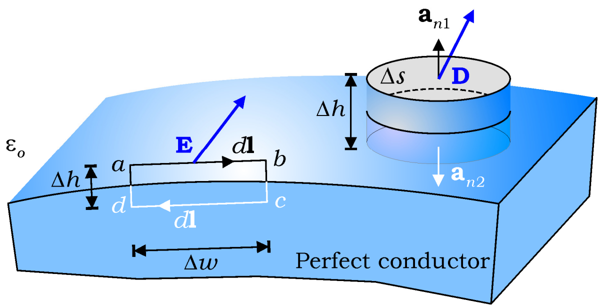


Fig. 3.25 A perfect conductor in free space.

Example 3-17

A net charge of $+Q$[C] is given to a perfectly conducting sphere of radius a, which is then enclosed by a perfectly conducting spherical shell of inner radius b and outer radius c, as shown in Fig. 3.26. Determine the surface charge densities induced on the inner surface at $R = b$ and the outer surface at $R = c$, and identify the source of the surface charges for the two cases: (a) the outer conductor is isolated (b) the outer conductor is grounded.

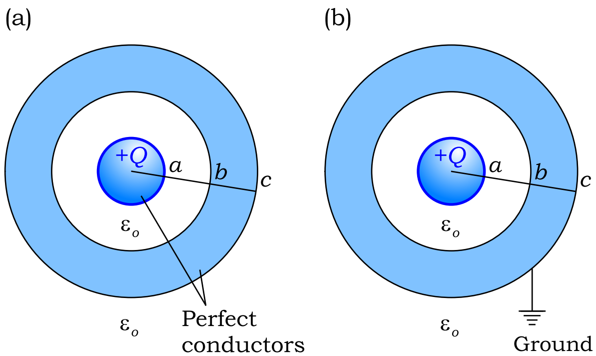


Fig. 3.26 A conducting sphere with a net charge $+Q$[C] is enclosed by a conducting shell.

Solution

From the spherical symmetry of the system, $\mathbf{D}$ is expected to be $\mathbf{D} = D_R(R)\,\mathbf{a}_R$ everywhere, whether or not it is grounded.

(a) In the region $a < R \le b$, applying Gauss's law we obtain

$$4\pi R^2 D_R = +Q$$

Electric flux density at $R = b$ is

$$\mathbf{D} = D_R\,\mathbf{a}_R = \frac{Q}{4\pi b^2}\mathbf{a}_R \tag{3-127a}$$

On the surface at $R = b$, using Eqs. (3-127a) and (3-126), and noting that the outward unit normal to the surface is $\mathbf{a}_n = -\mathbf{a}_R$, we obtain the surface charge density as

$$\rho_s^i = -D_R = -\frac{Q}{4\pi b^2} \qquad \text{(at } R = b\text{)} \qquad \text{(3-127b)}$$

In the region $R \geq c$, the net charge enclosed by the Gaussian surface is $+Q$ because the outer shell is electrically neutral. From Gauss's law we obtain

$$4\pi R^2 D_R = +Q$$

The electric flux density at $R = c$ is

$$\mathbf{D} = D_R\,\mathbf{a}_R = \frac{Q}{4\pi c^2}\mathbf{a}_R \qquad \text{(3-128a)}$$

On the surface at $R = c$, using Eqs. (3-128a) and (3-126), and noting that the outward unit normal to the surface is $\mathbf{a}_n = \mathbf{a}_R$, we obtain the surface charge density as

$$\rho_s^o = D_R = \frac{Q}{4\pi c^2} \qquad \text{(at } R = c\text{)} \qquad \text{(3-128b)}$$

The negative surface charge ρ_s^i at $R = b$ stems from the free electrons of the outer conductor, whereas the positive surface charge ρ_s^o at $R = c$ stems from the ionized lattice atoms of the same conductor. Note that $+Q$ and ρ_s^i jointly result in $\mathbf{E} = 0$ in the interior of the outer conductor.

(b) The outer conductor is at zero potential, because of its connection to ground. No electric field line starts at the conductor and ends at ground, or vice versa. Thus,

$$D_n = 0 = \rho_s^o \qquad (R = c)$$

In the region $a < R \leq b$, $\mathbf{D}$ is the same as that in part (a). Thus,

$$\rho_s^i = -\frac{Q}{4\pi b^2} \qquad (R = b)$$

Note that two charges $+Q$ and ρ_s^i jointly result in $\mathbf{E} = 0$ in the region $b < R < \infty$. The surface charge ρ_s^i is the net charge came from ground.

Exercise 3.27

A point charge $+Q$ is at a distance 1[m] from an infinite conducting surface at $z = 0$ connected to ground. Find the total charge induced on the conductor.

Ans. $-Q$.

Exercise 3.28
Describe the shape of the electric field lines between $+q$ and the conducting surface in **Exercise 3.27**.

Ans. Field lines start at $+q$ and end on the conducting surface at right angles.

Review Questions with Hints

RQ 3.26 Why is it that $\mathbf{E} = 0 = \rho_v$ in a perfect conductor. [Eqs.(3-121)(3-122)]

RQ 3.27 What are the boundary conditions for **E** and **D** at an interface between a perfect conductor and a perfect dielectric? [Eqs.(3-124)(3-126)]

RQ 3.28 Is the surface of a perfect conductor an equipotential surface, even if it has net charges? [Eqs.(3-121)(3-124)]

3.7 Electrostatic Potential Energy

As was stated earlier, the electric potential at a point in space is the work done in bringing a unit charge from infinity to the point against the electric field. In other words, the product of the electric potential and the charge located at the point is the potential energy of the charge. To hold the charge in place, however, we need to counterbalance the Coulomb force exerted on the charge by applying an external force on the charge, for instance a mechanical force. If we remove our hold on the charge, the potential energy is transformed into the kinetic energy of the charge, which will accelerate the charge and send it back to infinity. Extending this concept to a system of charges, the work done in assembling the charges is stored as the potential energy of the system of charges.

To compute the potential energy of a system of charges of the same polarity, we add the energies expended in bringing the individual charges from infinity to the predetermined positions. When we bring the first charge q_1 from infinity to a point in free space, no energy is expended. That is,

$$W_1 = 0 \tag{3-129}$$

When we bring a second charge q_2 to a prearranged point near charge q_1, we have to move against the electric field produced by q_1, and thus expend the energy given by

$$W_2 = q_2 \left[\frac{q_1}{4\pi\varepsilon_o \mathcal{R}_{2-1}} \right] \tag{3-130}$$

This is simply the product of q_2 and the electric potential due to q_1. In our notations, the subscript 2-1 in $\mathcal{R}_{2-1} = |\mathbf{r}_2 - \mathbf{r}_1|$ is to denote the distance from charge "1" to charge "2". With the help of the obvious relation $\mathcal{R}_{2-1} = \mathcal{R}_{1-2}$, Eq. (3-130) is rewritten as

$$W_2 = \frac{1}{2}\frac{q_2 q_1}{4\pi\varepsilon_o \mathcal{R}_{2-1}} + \frac{1}{2}\frac{q_1 q_2}{4\pi\varepsilon_o \mathcal{R}_{1-2}} \tag{3-131}$$

Following the same procedure used for W_2, we express the energy expended in bringing a third charge q_3 to a prearranged point near the charges q_1 and q_2 as

$$\begin{aligned} W_3 &= q_3\left[\frac{q_1}{4\pi\varepsilon_o \mathcal{R}_{3-1}}\right] + q_3\left[\frac{q_2}{4\pi\varepsilon_o \mathcal{R}_{3-2}}\right] \\ &= \frac{1}{2}\left[\frac{q_3 q_1}{4\pi\varepsilon_o \mathcal{R}_{3-1}} + \frac{q_3 q_2}{4\pi\varepsilon_o \mathcal{R}_{3-2}}\right] + \frac{1}{2}\left[\frac{q_1 q_3}{4\pi\varepsilon_o \mathcal{R}_{1-3}} + \frac{q_2 q_3}{4\pi\varepsilon_o \mathcal{R}_{2-3}}\right] \end{aligned} \tag{3-132}$$

Again, W_3 is the product of q_3 and the electric potential due to the previous charges q_1 and q_2. Note that Eq. (3-132) has been rewritten by making use of the relation $\mathcal{R}_{a-b} = \mathcal{R}_{b-a}$. Next, by summing the three terms W_1, W_2, and W_3, the total energy expended in assembling the first three charges is

$$\begin{aligned} &W_1 + W_2 + W_3 \\ &= \frac{q_1}{2}\left[\frac{q_2}{4\pi\varepsilon_o \mathcal{R}_{1-2}} + \frac{q_3}{4\pi\varepsilon_o \mathcal{R}_{1-3}}\right] + \frac{q_2}{2}\left[\frac{q_1}{4\pi\varepsilon_o \mathcal{R}_{2-1}} + \frac{q_3}{4\pi\varepsilon_o \mathcal{R}_{2-3}}\right] \\ &\quad + \frac{q_3}{2}\left[\frac{q_1}{4\pi\varepsilon_o \mathcal{R}_{3-1}} + \frac{q_2}{4\pi\varepsilon_o \mathcal{R}_{3-2}}\right] = \frac{1}{2}\left(q_1 V_1 + q_2 V_2 + q_3 V_3\right) \end{aligned} \tag{3-133}$$

We see that the three brackets in Eq. (3-133) represent the electric potentials V_1, V_2, and V_3 at the points of q_1, q_2, and q_3, respectively. By following the same procedure, we obtain the total energy stored in a system of N point charges as

$$\boxed{W_E = \frac{1}{2}\sum_{j=1}^{N} q_j V_j} \qquad \text{[J]} \tag{3-134}$$

Here, V_j is the electric potential at the location of the charge q_j, which is caused by all other charges except q_j.

We can obtain the energy stored in a continuous distribution of charges of a volume charge density ρ_v in the same manner as for discrete charges. By replacing q_j in Eq. (3-134) with the incremental charge $\rho_v \Delta v_j$, and taking the limit as $N \to \infty$ and $\Delta v \to 0$, we write the energy of the volume charge density ρ_v as

$$W_E = \lim_{\substack{N\to\infty \\ \Delta v \to 0}} \frac{1}{2}\sum_{j=1}^{N} \rho_v \Delta v_j V_j \tag{3-135}$$

From calculus we recognize the right-hand side of Eq. (3-135) as a volume integral of the function $\rho_v V$. The electrostatic potential energy of a volume charge density ρ_v is therefore,

$$\boxed{W_E = \frac{1}{2}\int_{\mathcal{V}} \rho_v V\, dv} \qquad \text{[J]} \qquad (3\text{-}136)$$

where $\mathcal{V}$ is the volume occupied by the volume charge density ρ_v, and V is the electric potential at a point in volume $\mathcal{V}$.

It is often more convenient to express the potential energy in terms of $\mathbf{E}$ and $\mathbf{D}$ than the charge density ρ_v. Upon applying Gauss's law, Eq. (3-136) becomes

$$W_E = \frac{1}{2}\int_{\mathcal{V}} (\nabla\cdot\mathbf{D})V\, dv \qquad (3\text{-}137)$$

Rewriting Eq. (3-137), with the help of the vector identity $\nabla\cdot(V\,\mathbf{D}) = V\,(\nabla\cdot\mathbf{D}) + \mathbf{D}\cdot(\nabla V)$, we have

$$\begin{aligned} W_E &= \frac{1}{2}\int_{\mathcal{V}} \nabla\cdot(V\,\mathbf{D})\, dv - \frac{1}{2}\int_{\mathcal{V}} \mathbf{D}\cdot(\nabla V)\, dv \\ &= \frac{1}{2}\oint_{\mathcal{S}} V\,\mathbf{D}\cdot d\mathbf{s} + \frac{1}{2}\int_{\mathcal{V}} \mathbf{D}\cdot\mathbf{E}\, dv \end{aligned} \qquad (3\text{-}138)$$

where we have used divergence theorem and the relation $\mathbf{E} = -\nabla V$. In the above equation, $\mathcal{S}$ is the bounding surface of the volume $\mathcal{V}$, which may be arbitrary only if it is large enough to enclose all the charges. Let us examine the closed surface integral on the right-hand side of Eq. (3-138). Let the surface $\mathcal{S}$ be a spherical surface of radius R, which tends to infinity. Then, as $R \to \infty$, the terms V, $|\mathbf{D}|$, and $|d\mathbf{s}|$ ($= R^2 \sin\theta d\theta d\phi$), in the closed surface integral, vary as $1/R$, $1/R^2$, and R^2, respectively. In view of these, we see that the closed surface integral varies as $1/R$, and becomes zero as $R \to \infty$. The electrostatic potential energy is therefore expressed as

$$\boxed{W_E = \frac{1}{2}\int_{\mathcal{V}} \mathbf{D}\cdot\mathbf{E}\, dv} \qquad \text{[J]} \qquad (3\text{-}139)$$

Upon using the constitutive relation $\mathbf{D} = \varepsilon\mathbf{E}$, Eq. (3-139) becomes

$$\boxed{W_E = \frac{1}{2}\int_{\mathcal{V}} \varepsilon E^2 dv} \qquad (3\text{-}140)$$

In view of Eq. (3-139), we can define the electrostatic energy density w_e as

$$\boxed{w_e = \frac{1}{2}\mathbf{D}\cdot\mathbf{E}} \qquad \text{[J/m}^3\text{]} \qquad (3\text{-}141)$$

which has the unit of the joule per cubic meter.

Example 3-18

Determine the potential energy of a volume charge of a uniform density ρ_v that is assembled in a spherical shell of inner radius a and outer radius b, in free space, as shown in Fig. 3.27.

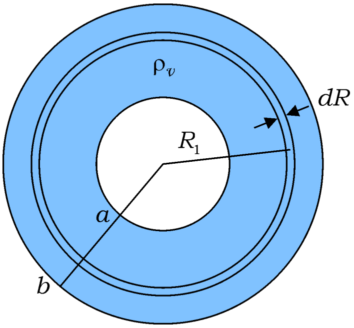


Fig. 3.27 A uniform volume charge in a spherical shell.

Solution

We compute the work done in bringing and stacking thin spherical layers of charges one by one. To start with, the total charge enclosed in a sphere of radius R_1 is

$$Q = \rho_v \frac{4\pi}{3}\left(R_1^3 - a^3\right)$$

The electric potential at $R = R_1$ is the same as if Q were concentrated on a point at the origin, i.e.,

$$V = \frac{Q}{4\pi\varepsilon_o R_1} = \frac{1}{4\pi\varepsilon_o R_1}\left[\rho_v \frac{4\pi}{3}\left(R_1^3 - a^3\right)\right] \tag{3-142}$$

A thin spherical layer of radius R_1 and thickness dR contains a net charge of

$$dq = \rho_v 4\pi R_1^2 dR \tag{3-143}$$

The energy expended in bringing dq from infinity to the spherical surface at $R = R_1$ is obtained from Eqs. (3-142) and (3-143) as

$$dW = Vdq = \left[\rho_v^2 \frac{4\pi}{3\varepsilon_o} R_1 \left(R_1^3 - a^3\right)\right] dR$$

The total energy expended in assembling the volume charge is

$$\begin{aligned} W_E = \int dW &= \rho_v^2 \frac{4\pi}{3\varepsilon_o} \int_{R=a}^{R=b} R\left(R^3 - a^3\right) dR \\ &= \rho_v^2 \frac{4\pi}{3\varepsilon_o}\left[\frac{1}{5}(b^5 - a^5) - a^3 \frac{1}{2}(b^2 - a^2)\right] \end{aligned} \tag{3-144}$$

We see from Eq. (3-144) that if $a = 0$, the volume charge forms a sphere of radius b, and the potential energy is simply given as kb^5, where k is a constant. We note that W_E in Eq. (3-144) cannot be expressed in the form of $k(b^5 - a^5)$, which is simply a subtraction of V due to the charges in a sphere of radius a from that due to the charges in a sphere of radius b. This leads us to conclude that the potential energy does not follow the principle of superposition.

Exercise 3.29
Find the potential energy of the three identical point charges of $1[\mu C]$ being $1[\mathrm{m}]$ apart, in free space, along the x-axis.
Ans. $0.022[\mathrm{J}]$.

Exercise 3.30
Two parallel, infinite sheets with surface charge densities ρ_s and $-\rho_s$ are separated by a distance a in free space. Find the energy density in between.
Ans. $w_e = \rho_s^2 / (2\varepsilon_o)$.

Review Questions with Hints

RQ 3.29 Express the electrostatic potential energy of a volume charge density. [Eq.(3-136)]
RQ 3.30 Express the electrostatic potential energy in terms of field quantities. [Eq.(3-140)]
RQ 3.31 Define the electrostatic energy density. [Eq.(3-141)]

3.8 Electrostatic Boundary Value Problems

Thus far we learned about the methods of obtaining the electric field from a given charge distribution. In most practical problems, however, the charge distribution may not be known in the given region. We now introduce Poisson's equation, Laplace's equation, and the method of images, which allow us to determine V and $\mathbf{E}$ in a region of space from the charges and electric potentials specified at the boundaries. Electrostatic problems involving boundary values are called boundary value problems.

3.8.1 Poisson's and Laplace's Equations

We start with two fundamental relations governing the static electric field, that is,

$$\nabla \cdot \mathbf{D} = \rho_v \tag{3-145}$$

$$\nabla \times \mathbf{E} = 0 \tag{3-146}$$

Limiting our discussion to a homogeneous, linear, and isotropic medium, the constitutive relation between **E** and **D** is

$$\mathbf{D} = \varepsilon\mathbf{E} \tag{3-147}$$

where the permittivity ε is constant, independent of the magnitude and direction of **E**. Let us recall the relation between **E** and V, i.e.,

$$\mathbf{E} = -\nabla V \tag{3-148}$$

Inserting Eqs. (3-147) and (3-148) into Eq. (3-145), we have

$$-\nabla \bullet \varepsilon(\nabla V) = \rho_v \tag{3-149}$$

Since ε is constant in the simple medium, it can be taken outside the divergence operator in Eq. (3-149). The Poisson's equation is therefore,

$$\boxed{\nabla^2 V = -\frac{\rho_v}{\varepsilon}} \tag{3-150}$$

The operator ∇^2, read "del squared", is the Laplacian operator, which represents the divergence of the gradient of a scalar field. If there is no net volume charge in the given region, the Poisson's equation reduces to the Laplace's equation, that is,

$$\boxed{\nabla^2 V = 0} \tag{3-151}$$

The Poisson's and Laplace's equations are second-order differential equations, each of which requires two independent boundary values for the determination of two constants of integration. These constants then uniquely specify a particular solution in the given region. When Laplace's equation is solved in a region where there is no volume charge, other charges, such as point, line, and surface charges, may be used as boundary values.

The Laplace's equation in Cartesian coordinates is

$$\boxed{\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0} \qquad \text{(Cartesian)} \tag{3-152a}$$

The Laplacian of V in cylindrical and spherical coordinates can be obtained by taking the divergence of the gradient of V in their respective coordinates. The Laplace's equations in cylindrical and spherical coordinates are as follows:

$$\boxed{\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial V}{\partial \rho}\right) + \frac{1}{\rho^2}\left(\frac{\partial^2 V}{\partial \phi^2}\right) + \frac{\partial^2 V}{\partial z^2} = 0} \qquad \text{(cylindrical)} \tag{3-152b}$$

$$\boxed{\frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2\frac{\partial V}{\partial R}\right) + \frac{1}{R^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial V}{\partial \theta}\right) + \frac{1}{R^2 \sin^2\theta}\frac{\partial^2 V}{\partial \phi^2} = 0} \qquad \text{(spherical)} \tag{3-152c}$$

Exercise 3.31
Show that the following electric potentials satisfy Laplace's equation:
(a) $V = e^{-2x}\cos(2y)$, and (b) $V = \ln[\tan(\theta/2)]$.

3.8.2 Uniqueness Theorem

From calculus we know that a second-order differential equation such as Poisson's equation should have two independent homogeneous solutions, which are obtained as if there were no source in the given region of space. If they are linearly combined in such a way as to satisfy two independent boundary conditions, the linear combination must be a unique solution, or the only solution, in the given region. This is called the uniqueness theorem, which asserts that ***the solution of Poisson's equation satisfying the given boundary conditions is a unique solution in the given region.*** The uniqueness theorem allows us to solve Poisson's equation in an intuitive manner, in which we only need to guess the solution satisfying the given boundary conditions.

To verify the uniqueness theorem, we proceed as follows. Suppose V_1 and V_2 are two solutions of Poisson's equation in a region $\mathscr{R}$, which has a finite volume $\mathcal{V}$ and a bounding surface $\mathcal{S}$, namely,

$$\nabla^2 V_1(\mathbf{r}) = -\frac{\rho_v}{\varepsilon} \tag{3-153a}$$

$$\nabla^2 V_2(\mathbf{r}) = -\frac{\rho_v}{\varepsilon} \tag{3-153b}$$

Let us assume that V_1 and V_2 satisfy the same boundary condition such that

$$V_1(\mathbf{r}_s) = V_2(\mathbf{r}_s) = V_s \tag{3-154}$$

where $\mathbf{r}_s$ is the position vector of a point on the boundary surface $\mathcal{S}$.

In the region $\mathscr{R}$, we define a scalar field as

$$\tilde{V}(\mathbf{r}) \equiv V_1(\mathbf{r}) - V_2(\mathbf{r}) \tag{3-155}$$

With the help of Eqs. (3-153) and (3-154), we obtain the Laplacian of $\tilde{V}(\mathbf{r})$ and the value of $\tilde{V}(\mathbf{r})$ at the boundary as follows:

$$\nabla^2 \tilde{V}(\mathbf{r}) = \nabla^2 V_1(\mathbf{r}) - \nabla^2 V_2(\mathbf{r}) = 0 \tag{3-156a}$$

$$\tilde{V}(\mathbf{r}_s) = V_1(\mathbf{r}_s) - V_2(\mathbf{r}_s) = 0 \tag{3-156b}$$

The differential equation and the boundary value given in Eq. (3-156a) and Eq. (3-156b) can be viewed as another boundary value problem in the region $\mathscr{R}$.

Upon substituting $\tilde{V}$ and $\nabla\tilde{V}$ for A and $\mathbf{V}$ in the vector identity $\nabla\bullet(A\mathbf{V}) = A(\nabla\bullet\mathbf{V}) + \mathbf{V}\bullet(\nabla A)$, we have

$$\nabla\bullet\left(\tilde{V}\,\nabla\tilde{V}\right) = \tilde{V}\left(\nabla\bullet\nabla\tilde{V}\right) + \nabla\tilde{V}\bullet\left(\nabla\tilde{V}\right)$$

Integrating both sides of the above equation over a volume $\mathcal{V}$, we have

$$\int_{\mathcal{V}} \nabla\bullet\left(\tilde{V}\,\nabla\tilde{V}\right)dv = \int_{\mathcal{V}} \tilde{V}\left(\nabla\bullet\nabla\tilde{V}\right)dv + \int_{\mathcal{V}} \nabla\tilde{V}\bullet\left(\nabla\tilde{V}\right)dv \tag{3-157}$$

The first term on the right-hand side of Eq. (3-157) should vanish, because of Eq. (3-156a). Rewriting the left-hand side of Eq. (3-157) by use of divergence theorem, and substituting the boundary condition given in Eq. (3-156b), we have

$$\int_{\mathcal{V}} \nabla\bullet\left(\tilde{V}\,\nabla\tilde{V}\right)dv = \oint_{\mathcal{S}} \left(\tilde{V}\,\nabla\tilde{V}\right)\bullet d\mathbf{s} = 0$$

Thus Eq. (3-157) becomes

$$\int_{\mathcal{V}} \nabla\tilde{V}\bullet\left(\nabla\tilde{V}\right)dv = 0 \tag{3-158}$$

The integrand in the volume integral in Eq. (3-158) is always positive. Thus we should have $\nabla\tilde{V} = 0$ at every point in $\mathcal{V}$ to satisfy the equality. In the region $\mathscr{R}$, the function $\tilde{V}(\mathbf{r})$ is obtained as

$$\tilde{V}(\mathbf{r}) = V_1(\mathbf{r}) - V_2(\mathbf{r}) = C \tag{3-159}$$

where C is a constant, and $\mathbf{r}$ is the position vector of a point in $\mathscr{R}$. Equation (3-159) should be satisfied at every point in $\mathscr{R}$, including the points on the boundary surface $\mathcal{S}$. Applying the boundary condition expressed by Eq. (3-156b) to Eq. (3-159) gives $C = 0$. Thus, at every point in the region $\mathscr{R}$, we have

$$V_1(\mathbf{r}) = V_2(\mathbf{r}) \tag{3-160}$$

which verifies the uniqueness theorem.

Since the above proof holds true irrespective of ρ_v, the uniqueness of the solution of Laplace's equation is also verified. Even if we find a solution of Laplace's equation by trial and error such that it satisfies the given boundary conditions, the solution is the only solution in the given region.

Exercise 3.32

A long conducting trough is along the z-axis, maintained at zero potential. Its cross section is a square of side b as ⊏ (bottom side, x-axis; left side, y-axis). (a) Does $V = \sinh(n\pi x/b)\sin(n\pi y/b)$, $n = 1, 2, ..$, satisfy Laplace's equation and the boundary condition? (b) Why no unique solution?

Ans. (a) Yes, (b) Laplace's equation requires two boundary conditions.

3.8.3 *Examples of Boundary Values Problems*

Example 3-19
Two parallel conducting plates are separated by a distance d, and maintained at $V=0$ and $V=V_o$, as shown in Fig. 3.28. The gap is filled with a volume charge density $\rho_v = \rho_o \sin(\pi z/d)$. Ignoring the fringing effect of **E** at the edges of the conductors, find
(a) electric potential in the gap, and
(b) surface charge densities induced on the conductors.

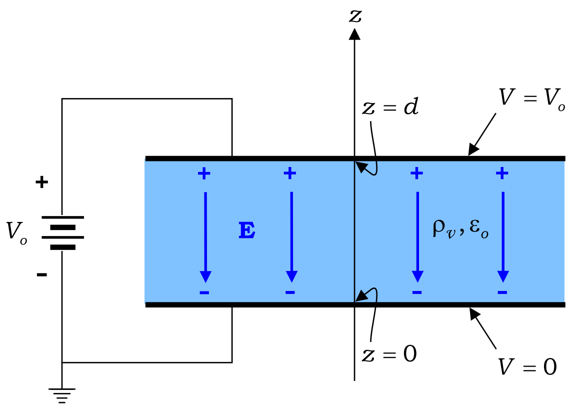


Fig. 3.28 Two parallel plates filled with volume charges.

Solution
(a) If the edge effects are ignored, **E** and V are obtained as if the plates and the charge distribution were infinite in extent in the xy-plane. From translational symmetries in the x- and y-directions, Poisson's equation is written as

$$\frac{d^2V}{dz^2} = -\frac{\rho_o}{\varepsilon_o}\sin\left(\frac{\pi}{d}z\right)$$

Upon integrating both sides with respective to z twice, we obtain

$$V = \frac{\rho_o}{\varepsilon_o}\left(\frac{d}{\pi}\right)^2 \sin\left(\frac{\pi}{d}z\right) + c_1 z + c_2 \tag{3-161}$$

where c_1 and c_2 are constants of integration

Applying the boundary conditions to Eq. (3-161), we have

$$V(z=0) = 0 = c_2 \tag{3-162a}$$

$$V(z=d) = V_o = c_1 d \tag{3-162b}$$

Combining Eqs. (3-162) and (3-161), V in the gap is

$$V = \frac{\rho_o}{\varepsilon_o}\left(\frac{d}{\pi}\right)^2 \sin\left(\frac{\pi}{d}z\right) + \frac{V_o}{d}z\,. \tag{3-163}$$

(b) From Eq. (3-163), the electric field in the gap is

$$\mathbf{E} = -\nabla V = -\left[\frac{\rho_o d}{\varepsilon_o \pi}\cos\left(\frac{\pi}{d}z\right) + \frac{V_o}{d}\right]\mathbf{a}_z$$

On the conducting surfaces at $z = 0$ and $z = d$, we have

$$\mathbf{D}(0) = \varepsilon_o \mathbf{E}(0) = \varepsilon_o\left[-\frac{\rho_o d}{\varepsilon_o \pi} - \frac{V_o}{d}\right]\mathbf{a}_z$$

$$\mathbf{D}(d) = \varepsilon_o \mathbf{E}(d) = \varepsilon_o\left[\frac{\rho_o d}{\varepsilon_o \pi} - \frac{V_o}{d}\right]\mathbf{a}_z$$

A unit normal to the surface at $z = 0$ is $\mathbf{a}_n = \mathbf{a}_z$, and a unit normal to the surface at $z = d$ is $\mathbf{a}_n = -\mathbf{a}_z$. From Eq. (3-126), the induced surface charge densities are

$$\rho_s = D_z(0) = -\frac{\rho_o d}{\pi} - \frac{\varepsilon_o V_o}{d} \qquad \text{at } z = 0 \tag{3-164a}$$

$$\rho_s = -D_z(d) = -\frac{\rho_o d}{\pi} + \frac{\varepsilon_o V_o}{d} \qquad \text{at } z = d \tag{3-164b}$$

The surface charge densities in Eq. (3-164) are not equal in magnitude. The charge densities $\pm\varepsilon_o V_o / d$ stem from the potential difference V_o, whereas the charge density $-\rho_o d/\pi$ originates from the net charge $2\rho_o d/\pi$ contained in a volume $1\times 1\times d$ [m^3] in the gap.

Example 3-20

Two semi-infinite conducting plates take a wedge form as shown in Fig. 3.29. They are at angles ϕ_1 and ϕ_2 with respect to the x-axis, and maintained at potentials V_1 and V_2, respectively. Find V in the regions:

(a) $\phi_1 \le \phi \le \phi_2$, and

(b) $\phi_2 \le \phi \le (\phi_1 + 2\pi)$.

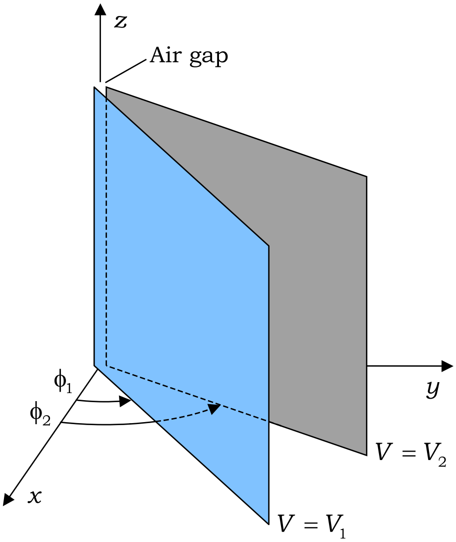


Fig. 3.29 Two semi-infinite conducting plates in a wedge form.

Solution

Keeping the uniqueness theorem in mind, and noting that $\phi = \phi_1$ and $\phi = \phi_2$ planes are equipotential surfaces, we assume that any surface of constant ϕ is an equipotential surface. Under these conditions, V is independent of ρ and z, and Laplace's equation is written in cylindrical coordinates as

$$\nabla^2 V = \frac{1}{\rho^2}\frac{\partial^2 V}{\partial \phi^2} = 0 \qquad (3\text{-}165)$$

Upon integrating both sides of the equation twice with respect to ϕ, we obtain a general solution, with the constants of integration c_1 and c_2, as

$$V = c_1\phi + c_2 . \qquad (3\text{-}166)$$

(a) In the region $\phi_1 \le \phi \le \phi_2$

Applying the boundary conditions to Eq. (3-166), we obtain

$$V(\phi_1) = V_1 = c_1\phi_1 + c_2 \qquad (3\text{-}167a)$$

$$V(\phi_2) = V_2 = c_1\phi_2 + c_2 \qquad (3\text{-}167b)$$

Solving the equations for c_1 and c_2, we obtain

$$c_1 = \frac{V_1 - V_2}{\phi_1 - \phi_2}$$

$$c_2 = V_1 - \frac{V_1 - V_2}{\phi_1 - \phi_2}\phi_1$$

Thus,

$$V = \frac{V_1 - V_2}{\phi_1 - \phi_2}(\phi - \phi_1) + V_1 . \qquad (\phi_1 \le \phi \le \phi_2) \tag{3-168}$$

(b) In the region $\phi_2 \le \phi \le (\phi_1 + 2\pi)$

Applying the boundary conditions to the general solution in Eq. (3-166), we obtain

$$V(\phi_2) = V_2 = c_1\phi_2 + c_2$$

$$V(\phi_1 + 2\pi) = V_1 = c_1(\phi_1 + 2\pi) + c_2$$

Solving the equations for c_1 and c_2, we obtain

$$c_1 = \frac{V_2 - V_1}{\phi_2 - \phi_1 - 2\pi}, \text{ and } c_2 = V_2 - \frac{V_2 - V_1}{\phi_2 - \phi_1 - 2\pi}\phi_2$$

Thus,

$$V = \frac{V_2 - V_1}{\phi_2 - \phi_1 - 2\pi}(\phi - \phi_2) + V_2 \qquad \left(\phi_2 \le \phi \le (\phi_1 + 2\pi)\right) \tag{3-169}$$

The electric potentials given in Eqs. (3-168) and (3-169) are unique solutions in their respective regions because they satisfy Laplace's equation and the given boundary conditions. They conform with the initial assumption that any surface of constant ϕ is an equipotential surface.

Example 3-21

The space between two concentric, perfectly conducting, spherical shells of negligible thickness is filled with a dielectric of permittivity ε as shown in Fig. 3.30. Two spheres have radii a and b, and maintained at potentials V_1 and V_2, respectively. Find V in the regions: (a) $R > b$, (b) $a \le R \le b$, and (c) $0 < R \le a$.

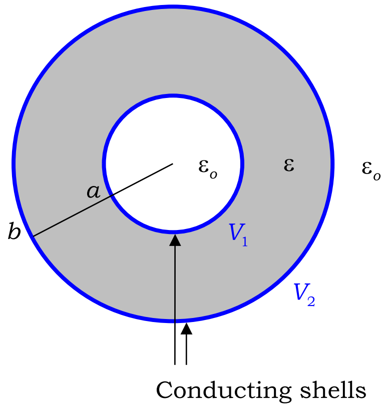


Fig. 3.30 Two conducting shells separated by a dielectric.

Solution

From spherical symmetry, Laplace's equation is written in spherical coordinates as

$$\nabla^2 V = \frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2\frac{\partial V}{\partial R}\right) = 0$$

Upon integrating both sides of the equation twice with respect to *R*, we obtain a general solution, with the constants of integration c_1 and c_2, as

$$V = \frac{c_1}{R} + c_2 \,. \tag{3-170}$$

(a) In the region $R > b$

Applying the boundary conditions to Eq. (3-170), we get

$$V(R = b) = V_2 = \frac{c_1}{b} + c_2 \tag{3-171a}$$

$$V(R = \infty) = 0 = c_2 \tag{3-171b}$$

Solving Eq. (3-171) for c_1 and c_2, we get

$$c_1 = bV_2$$

$$c_2 = 0$$

Thus,

$$V = \frac{bV_2}{R} \,. \qquad (R > b) \tag{3-172a}$$

(b) In the region $a \le R \le b$

Applying the boundary conditions to Eq. (3-170), we get

$$V(R = a) = V_1 = \frac{c_1}{a} + c_2$$

$$V(R = b) = V_2 = \frac{c_1}{b} + c_2$$

Solving the equations for c_1 and c_2, we get

$$c_1 = \frac{ab}{b-a}(V_1 - V_2)$$

$$c_2 = V_1 - \frac{b}{b-a}(V_1 - V_2)$$

Thus,

$$V = V_1 + \left[\frac{a}{R} - 1\right]\frac{b}{b-a}(V_1 - V_2) \,. \qquad (a \le R \le b) \tag{3-172b}$$

(c) There is no electric field in the region $0 < R \le a$. Thus,

$$V = V_1 \,. \tag{3-172c}$$

3.8.4 Method of Images

The electric field lines of an electric dipole are symmetrical about the plane placed in the middle as shown in Fig. 3.31(a). Furthermore, the electric field lines cross the middle plane at right angles, implying that the plane is an equipotential surface. Upon substituting $\mathbf{p} \cdot \mathbf{a}_R = 0$, which is an expression for the middle plane of the dipole, into Eq. (3-83), we see that the plane is at a potential of $V = 0$. Let us digress briefly and consider a boundary value problem in which a point charge $+q$ is at a distance $d/2$ from an infinitely large, perfectly conducting surface that is maintained at a zero potential, as illustrated in Fig. 3.31(b). From potential theory we know that the electric field lines originate on the positive charge and terminate on the conducting surface at right angles. It is apparent that the electric field lines shown in Fig. 3.31(b) are the same as those in the upper half of Fig. 3.31(a). This is justified by the fact that the two field lines satisfy the same boundary condition. Note that the solution satisfying Laplace's equation and given boundary conditions is a unique solution.

The method of images provides another way of determining the electric field in a given problem that involves perfectly conducting planar surfaces of zero potential. In the method of images, the charge distributions above the conducting surface are replaced with the combination of the given charges and their image charges, with the conducting surface removed, as illustrated in Fig. 3.32(a) and Fig. 3.32(b). The electric field of the given problem is the one obtained in the region above the middle plane. The principle of superposition allows us to extend the method of images to continuous distributions of charges and even equipotential surfaces.

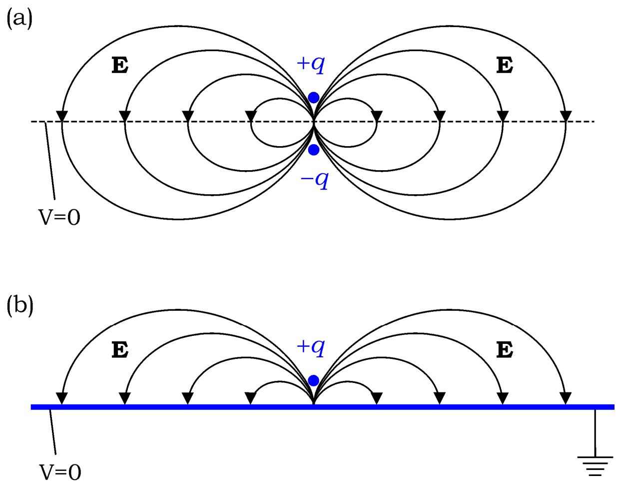


Fig. 3.31 (a) Dipole electric field (b) A point charge above a conducting surface.

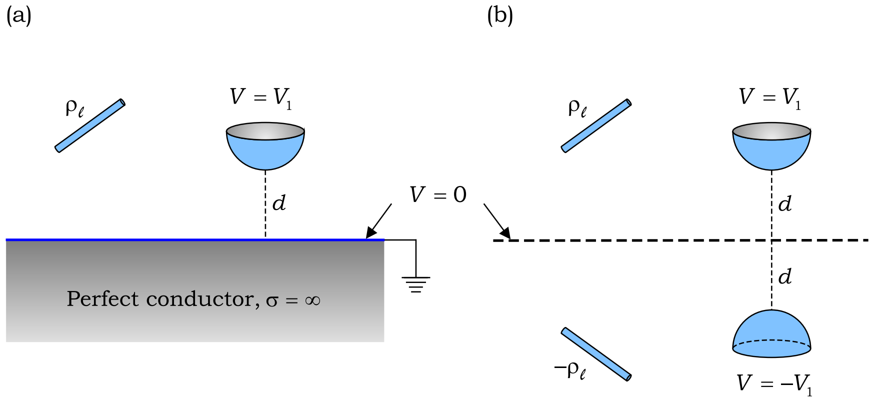


Fig. 3.32 (a) Boundary value problem (b) Method of images.

Example 3-22
Two charges $+Q$ and $-Q$ are images of the other as shown in Fig. 3.33. From symmetry considerations we see that the electric field lines should be symmetrical about the middle plane.

Should the electric field lines be, in addition, normal to the middle plane?

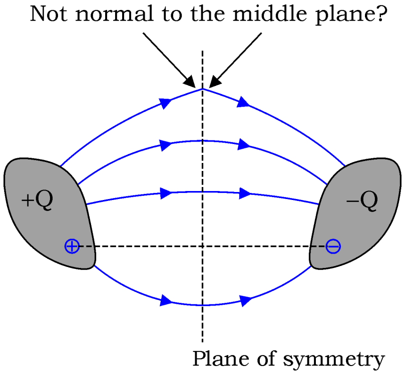


Fig. 3.33 Symmetrical field lines due to symmetrical charges.

Solution
Two symmetrical charges $+Q$ and $-Q$ can be regarded as the sum of many electric dipoles straddling the middle plane. The electric field lines of an electric dipole are smooth everywhere, except for the source points occupied by the two charges of the dipole. According to the principle of superposition, the field lines of $+Q$ and $-Q$ are smooth in space, and thus normal to the middle plane.

Example 3-23
Find an expression for the equipotential surface of

(a) two infinitely long, straight, and thin lines are parallel to each other, separated by a distance $2d$. They carry uniform line charge densities ρ_ℓ and $-\rho_\ell$, respectively.
(b) two infinitely long conducting wires of a finite radius β_1 are parallel to each other, separated by a center-to-center distance of $2\alpha_1$. They are maintained at potentials V_1 and $-V_1$, respectively.

Solution

(a) Referring to Fig. 3.34(a), we first obtain the electric potential V at a point p from Eq. (3-69) as

$$V = \frac{\rho_\ell}{2\pi\varepsilon_o}\left[\ln\frac{\mathcal{R}_{1-0}}{\mathcal{R}_1} - \ln\frac{\mathcal{R}_{2-0}}{\mathcal{R}_2}\right] \tag{3-173}$$

where $\mathcal{R}_1$ (or $\mathcal{R}_2$) is the perpendicular distance from the point p to the line charge ρ_ℓ (or $-\rho_\ell$), and $\mathcal{R}_{1-0}$ (or $\mathcal{R}_{2-0}$) is the perpendicular distance from a zero reference point to the line charge ρ_ℓ (or $-\rho_\ell$).

We set $\mathcal{R}_{1-0} = \mathcal{R}_{2-0}$ by assuming the $y = 0$ plane to be the zero reference point, which will be validated later. Under this condition, Eq. (3-173) becomes

$$V = \frac{\rho_\ell}{2\pi\varepsilon_o}\ln\frac{\mathcal{R}_2}{\mathcal{R}_1} \tag{3-174}$$

Inserting expressions for $\mathcal{R}_1$ and $\mathcal{R}_2$ in Cartesian coordinates into Eq. (3-174), we have

$$\boxed{V = \frac{\rho_\ell}{4\pi\varepsilon_o}\ln\left[\frac{(y+d)^2 + z^2}{(y-d)^2 + z^2}\right]} \tag{3-175}$$

We rewrite Eq. (3-175) and define a parameter K as follows:

$$\boxed{\exp\left[\frac{4\pi\varepsilon_o V}{\rho_\ell}\right] = \frac{(y+d)^2 + z^2}{(y-d)^2 + z^2} \equiv K} \tag{3-176}$$

The second equality in Eq. (3-176) leads to

$$(y+d)^2 + z^2 = K\left[(y-d)^2 + z^2\right] \tag{3-177}$$

Rearranging terms in Eq. (3-177), we have

$$\boxed{(y-\alpha)^2 + z^2 = \beta^2} \tag{3-178}$$

where

$$\boxed{\alpha = d\frac{K+1}{K-1}} \tag{3-179a}$$

$$\boxed{\beta = \frac{2d\sqrt{K}}{K-1}} \tag{3-179b}$$

To obtain the equipotential surface of a potential V_1, we proceed as follows:

(1) Insert $V = V_1$ into Eq. (3-176) to obtain $K = K_1$.

(2) Insert $K = K_1$ into Eq. (3-179) to obtain $\alpha = \alpha_1$ and $\beta = \beta_1$.

(3) Inserting α_1 and β_1 into Eq. (3-178), we obtain the equation of the equipotential surface in the yz-plane.

We see that the equipotential surface of $V = V_1$ is a cylinder of radius β_1, centered at $y = \alpha_1$, as shown in Fig. 3.34(b). We see from Eq. (3-176) that $V > 0$ for $K > 1$ (the cylinder is in the region $y > 0$), and $V < 0$ for $K < 1$ (the cylinder is in the region $y < 0$).

Let us now check the zero reference point. For $V = 0$, we can obtain $K = 1$ and $\alpha = \beta = \infty$ from Eqs. (3-176) and (3-179), which correspond to the $y = 0$ plane as can be seen from Eq. (3-178). The initial assumption for the zero reference point is thus validated.

(b) The cross sections of the two wires are shown in Fig. 3.34(c), which are images of the other with respect to the middle plane, or the $y = 0$ plane. According to the method of images, we replace the left conductor with an infinite conducting plane of $V = 0$ placed in the middle, the $y = 0$ plane, as shown in Fig. 3.34(b). We then solve this boundary value problem by making use of the equivalent parallel line charges as shown in Fig. 3.34(a).

For the given values of α_1 and β_1 ($\alpha_1 > \beta_1$), we obtain K and d from Eq. (3-179) as

$$\sqrt{K_1} = \frac{\alpha_1 + \sqrt{\alpha_1^2 - \beta_1^2}}{\beta_1} \tag{3-180a}$$

$$d = \alpha_1 \frac{K_1 - 1}{K_1 + 1} = \frac{\beta_1}{2} \frac{K_1 - 1}{\sqrt{K_1}} \tag{3-180b}$$

Inserting K_1 and V_1 into Eq. (3-176), we obtain the equivalent line charge density as

$$\rho_\ell = \frac{4\pi\varepsilon_o V_1}{\ln K_1} \tag{3-181}$$

Next, inserting ρ_ℓ and d into Eq. (3-175), we obtain the expression for the equipotential surface of an electric potential V as

$$V = \frac{\rho_\ell}{4\pi\varepsilon_o} \ln\left[\frac{(y+d)^2 + z^2}{(y-d)^2 + z^2}\right] \tag{3-182}$$

Following the same procedure, we can also solve a boundary value problem involving an infinitely long line charge and an infinitely long conducting cylinder, which are parallel to each other.

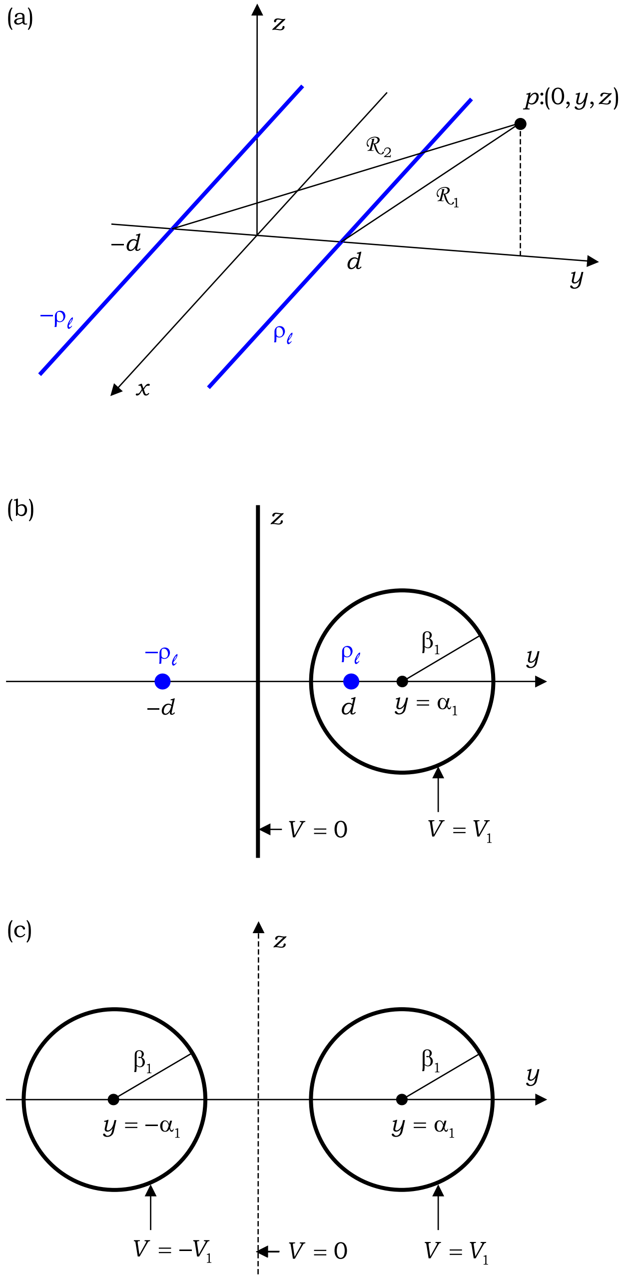


Fig. 3.34 (a) Two parallel line charges (b) Two equipotential surfaces ($V = 0$ and $V = V_1$) (c) Two equipotential surfaces ($V = V_1$ and $V = -V_1$).

When two infinitely long line charges of uniform densities $\pm 1[\mathrm{nC/m}]$ are separated by $2[\mathrm{m}]$ in free space, the equipotential surfaces are drawn to scale in the yz-plane in Fig. 3.35.

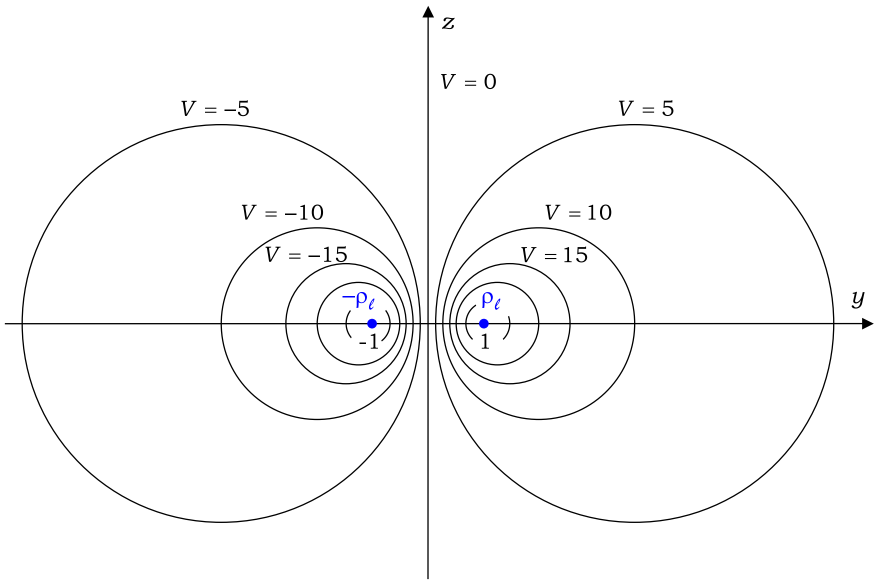


Fig. 3.35 Equipotential surfaces of two parallel line charges of densities $\pm 1[\mathrm{nC/m}]$, separated by $2[\mathrm{m}]$ in free space.

Exercise 3.33
The conducting surface in Fig. 3.31(b) is not at zero potential. (a) Are the field lines still normal to the conductor? (b) Are they still equal to the upper half of Fig. 3.31(a)? (c) If no, what is the source of the distortion?

Ans. (a) Yes, (b) No, (c) Net surface charges on the conductor.

Exercise 3.34
For a boundary value problem involving ρ_ℓ and the $y = 0$ plane maintained at $V = 0$ as in Fig. 3.34(b), find the surface charge density induced on the $y = 0$ plane per unit length of the x-axis.

Ans. $-\dfrac{\rho_\ell d}{\pi(d^2 + z^2)}[\mathrm{C/m}]$.

Review Questions with Hints

RQ 3.32 Write Poisson's and Laplace's equations. [Eqs.(3-150)(3-151)]

RQ 3.33 Can Poisson's and Laplace's equations be applied to an inhomogeneous, nonlinear, and anisotropic material? [Eq.(3-149)]

RQ 3.34 State the uniqueness theorem for electric potential. [Eq.(3-160)]

RQ 3.35 How many boundary conditions are required for a unique determination of *V*? [Eq.(3-161)]

RQ 3.36 Can an arbitrary scalar function represent an electric potential in free space even if it does not satisfy Laplace's equation? [Eq.(3-149)]

RQ 3.37 Under what conditions, is a trial solution a unique solution to a boundary value problem involving no volume charge? [Eqs.(3-165)(3-167)]

RQ 3.38 What are the principles underlying the method of images? [Figs.3.31,3.32]

3.9 Capacitance and Capacitors

Any two conducting objects can form a capacitor, regardless of their shapes and sizes, when they are separated by a dielectric. A capacitor can store energy in the electric field that is induced in the dielectric by the charges accumulated on the conductors. With reference to Fig. 3.36, in which two electrically neutral conductors are embedded in a dielectric of a permittivity ε, let us consider the case in which some free electrons are taken out of conductor 1 and imparted to conductor 2. By doing so, conductor 2 is negatively charged due to the injected electrons, while conductor 1 is positively charged due to the ionized host atoms. Two conductors have a net charge of an equal amount but opposite polarity. In practical situations, the separation of charges may be accomplished by a dc voltage source connected to the conductors.

As was stated earlier, the excess charges should be distributed on the surface of the conductor so that there is no electric field inside the conductor; the conductor is at an equipotential. The potential difference between the conductors, *V*, is of course generated by the electric field **E** that is induced in the dielectric by the net charges on the conductors. At this point, we see that a further separation of charges will only increase the surface charge density ρ_s on the conductor, with no change in the charge distribution; otherwise, the charge distribution would induce an electric field inside the conductor. Thus, if the amount of the separated charges is increased by a factor of *k*, the magnitudes of ρ_s, **E**, and *V* are all increased by the same factor. This is evident from the relations $\rho_s = D_n = \varepsilon E_n$ and $\mathbf{E} = -\nabla V$. Because of the linear relationship between ρ_s and *V*, the ratio between the net charge on the conductor and the potential difference between the conductors should be constant for a given capacitor.

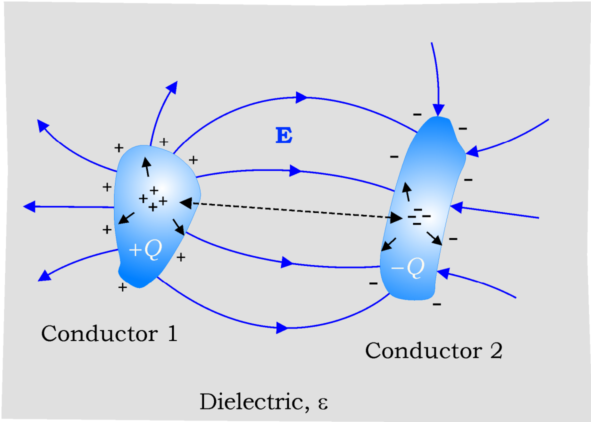


Fig. 3.36 Two conductors separated by a dielectric.

Energy is expended in separating the charges between the two conductors because the charge separation is done against the electric field already established in the dielectric. The work done is stored as the potential energy of the capacitor. For a fixed amount of the net charge on the conductor, the induced **E** strongly depends on the dielectric, and the geometry of the conductors such as shape, surface area, and distance between the conductors.

The capacitance of a capacitor is defined as the amount of the separated charges on the conductor required for building up a potential difference of 1 *V* between the conductors, i.e.,

$$\boxed{C = \frac{Q}{V}} \qquad [\text{F}] \qquad (3\text{-}183)$$

The capacitance is measured in farads[F], which is equivalent to coulombs per volt. The capacitance is independent of the total charge *Q* and the potential difference *V*, because of the linear relationship between *Q* and *V*.

The capacitance between two conductors can be obtained from Eq. (3-183) by following these steps:

1. Assume charges $+Q$ and $-Q$ on two conductors.
2. Choose a coordinate system considering symmetry.
3. Find **E** due to *Q* by Coulomb's law, Gauss's law, or other methods.
4. Find *V* by the negative line integral of **E**.
5. Calculate *C* from Q/V.

Alternatively,

1. Assume potentials V_1 and V_2 on two conductors.
2. Solve Laplace's equation to find *V* between two conductors.
3. Find **E** by the negative gradient of *V*.
4. Find the surface charge density from $\rho_s = \varepsilon E_n$, and the total surface charge *Q*.
5. Calculate *C* from $Q/(V_1 - V_2)$.

3.9.1 Parallel-Plate Capacitor

A parallel-plate capacitor consists of two parallel conducting plates of a surface area $\mathcal{S}$, separated by a dielectric of thickness d and permittivity ε, as shown in Fig. 3.37. To determine the capacitance, we assume charges $+Q$ and $-Q$ on the upper and lower conductors, respectively. By assuming the gap size d to be much smaller than the lateral dimension of the plate, we ignore the fringing effects of **E** at the edges of the plates, and obtain **E** in the gap as if the plates were infinite in extent. Under these conditions, the charges are uniform on the conducting plates with uniform surface densities given as

$$\rho_s = \pm \frac{Q}{\mathcal{S}} \tag{3-184}$$

From Eqs. (3-57) and (3-58), with the aid of the principle of superposition, we obtain $\mathbf{E} = 0$ outside the capacitor, and, in the gap,

$$\mathbf{E} = \frac{-\rho_s}{\varepsilon} \mathbf{a}_z \tag{3-185}$$

The potential difference between the conducting plates is

$$V_{1-2} = -\int_2^1 \mathbf{E} \bullet d\mathbf{l} = \frac{\rho_s}{\varepsilon} d \tag{3-186}$$

The capacitance of a parallel-plate capacitor is therefore

$$\boxed{C = \frac{Q}{V_{1-2}} = \frac{\varepsilon \mathcal{S}}{d}} \qquad \text{[F]} \tag{3-187}$$

The capacitance is directly proportional to the surface area and the permittivity, but inversely proportional to the separation between the two conductors.

Upon inserting Eq. (3-185) into Eq. (3-139), we obtain the electric energy stored in the capacitor as follows:

$$\begin{aligned} W_E &= \frac{1}{2} \int_{\mathcal{V}} \mathbf{D} \bullet \mathbf{E} dv = \frac{\varepsilon}{2} \int_{\mathcal{V}} E^2 dv \\ &= \frac{\rho_s^2}{2\varepsilon} \mathcal{S} d = \frac{1}{2} \left(\frac{\varepsilon \mathcal{S}}{d} \right) \left[\frac{\rho_s d}{\varepsilon} \right]^2 \end{aligned} \tag{3-188}$$

We see from Eq. (3-188) that the term in parenthesis represents the capacitance and the term in bracket represents the potential difference across the capacitor. In view of these, +the electrostatic energy stored in a parallel-plate capacitor is expressed as

$$\boxed{W_E = \frac{1}{2} C V^2} \qquad \text{[J]} \tag{3-189}$$

The stored energy is proportional to the capacitance and the square of the potential difference.

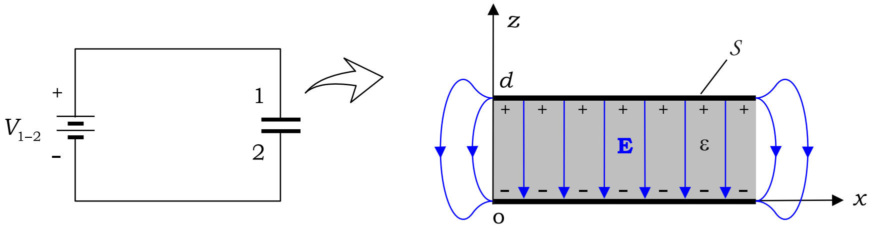


Fig. 3.37 A parallel-plate capacitor.

Let us imagine that we enlarge the surface area of the conducting plates, while maintaining the total charge on the conductor constant. Under these conditions, both ρ_s on the conductor and $\mathbf{E}$ in the gap are reduced, meaning that a further separation of charges is required for a potential difference of $1V$ between the conductors; we have a higher capacitance. Similarly, if the dielectric is replaced with one having a larger ε, the electric field in the dielectric is reduced, meaning that an additional separation of charges is required for a buildup of $1V$; we also have a higher capacitance. Even if the separation between the two conductors is reduced, there is no change in the electric field in the gap. In this case, however, the path of integration for V is reduced, and more separation of charges is required for the $1V$; we have a higher capacitance.

3.9.2 Examples of Capacitors

Example 3-24

A coaxial capacitor of a length $\mathcal{L}$ consists of two concentric cylindrical conductors of radius a and b as shown in Fig. 3.38. The space between two conductors is filled with a dielectric of permittivity ε. Ignoring the fringing effects of $\mathbf{E}$ at the edges of the conductors, determine the capacitance by assuming

(a) charges $\pm Q$ on the conductors

(b) potentials $\pm V_1$ on the conductors.

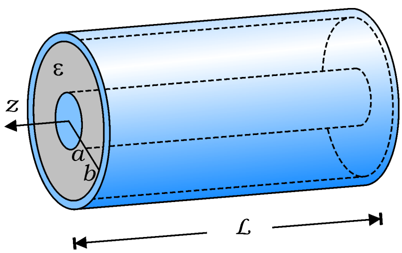


Fig. 3.38 A coaxial capacitor.

Solution

(a) We assume charges $+Q$ and $-Q$ on the inner and outer conductors, respectively. If the fringing effects are ignored, the charges are uniform on the conductors, and $\mathbf{E}$ in the gap is the same as if the capacitor were infinitely long. The charge distribution has cylindrical, translational(in the z-direction), and twofold rotational(about the x-axis) symmetries. From the symmetries, the resultant $\mathbf{E}$ is expected to be of the form $\mathbf{E} = E_\rho(\rho)\,\mathbf{a}_\rho$ everywhere.

In the region $a < \rho < b$, from Gauss's law we obtain

$$\varepsilon E_\rho 2\pi\rho \mathcal{L} = Q$$

Electric field intensity in the gap is

$$\mathbf{E} = E_\rho \mathbf{a}_\rho = \frac{Q}{2\pi\varepsilon\rho\mathcal{L}}\mathbf{a}_\rho$$

Potential difference between two conductors is calculated as follows:

$$V_{a-b} = -\int_{\rho=b}^{\rho=a} \mathbf{E}\bullet d\mathbf{l} = -\int_{\rho=b}^{\rho=a} \frac{Q}{2\pi\varepsilon\rho\mathcal{L}}\mathbf{a}_\rho \bullet \left(\mathbf{a}_\rho d\rho\right) = \frac{Q}{2\pi\varepsilon\mathcal{L}}\ln\left(\frac{b}{a}\right)$$

The capacitance of the coaxial capacitor is therefore

$$C = \frac{Q}{V_{a-b}} = \frac{2\pi\varepsilon\mathcal{L}}{\ln(b/a)}. \qquad \text{[F]} \qquad (3\text{-}190)$$

(b) We assume potentials $+V_1$ and $-V_1$ on the inner and outer conductors, respectively. Keeping the uniqueness theorem in mind, and noting the given boundary conditions, we assume any cylindrical surface centered on the z-axis to be an equipotential surface. Under these conditions, V is independent of ϕ and z, and Laplace's equation in cylindrical coordinates is thus reduced to

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial V}{\partial\rho}\right) = 0$$

A general solution is written, with the constants of integration c_1 and c_2, as

$$\mathrm{V} = c_1 \ln\rho + c_2 \qquad (3\text{-}191)$$

Applying the boundary conditions to Eq. (3-191), we get

$$\mathrm{V}_1 = c_1 \ln a + c_2 \qquad (3\text{-}192a)$$

$$-\mathrm{V}_1 = c_1 \ln b + c_2 \qquad (3\text{-}192b)$$

Inserting c_1 and c_2, obtained from Eq. (3-192), into Eq. (3-191), we get

$$V = V_1 - 2V_1 \frac{\ln(\rho / a)}{\ln(b / a)}$$

Taking the negative gradient of V, we have

$$\mathbf{E} = \frac{2V_1}{\ln(b / a)} \frac{1}{\rho} \mathbf{a}_\rho$$

$$\mathbf{D} = \varepsilon \mathbf{E} = \frac{2\varepsilon V_1}{\ln(b / a)} \frac{1}{\rho} \mathbf{a}_\rho$$

On the cylindrical surface at $\rho = a$, the surface charge density ρ_s and the total surface charge Q are calculated as follows:

$$\rho_s = D_\rho = \frac{2\varepsilon V_1}{\ln(b / a)} \frac{1}{a}$$

$$Q = 2\pi a \mathcal{L} \rho_s = \frac{4\pi\varepsilon V_1 \mathcal{L}}{\ln(b / a)}$$

The capacitance is

$$C = \frac{Q}{2V_1} = \frac{2\pi\varepsilon\mathcal{L}}{\ln(b / a)} . \qquad \text{[F]} \qquad (3\text{-}193)$$

Example 3-25
Determine the capacitance of an isolated conducting sphere of radius a, residing in free space.

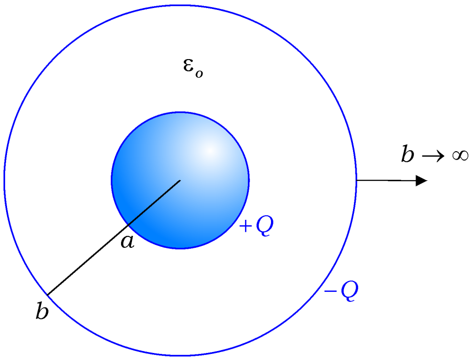


Fig. 3.39 An isolated sphere.

Solution

Frist, we compute the capacitance of a spherical capacitor consisting of two concentric conducting spheres of radius a and b as shown in Fig. 3.39, and then make the outer sphere expand to infinity. If we assume charges $+Q$ and $-Q$ on the spheres, we see that the charge distribution has spherical symmetry. From the spherical symmetry, $\mathbf{E}$ is expected to be of the form $\mathbf{E} = E_R(R)\,\mathbf{a}_R$.

In the region $a < R < b$, from Gauss's law we obtain $D_R 4\pi R^2 = Q$. Thus, the electric field intensity is

$$\mathbf{E} = E_R \mathbf{a}_R = \frac{Q}{4\pi\varepsilon_o R^2}\mathbf{a}_R$$

Potential difference between two spheres is

$$V_{a-b} = -\int_{R=b}^{R=a} \mathbf{E}\cdot d\mathbf{l} = -\int_{R=b}^{R=a} \frac{Q}{4\pi\varepsilon_o R^2}\mathbf{a}_R\cdot(\mathbf{a}_R dR) = \frac{Q}{4\pi\varepsilon_o}\left(\frac{1}{a}-\frac{1}{b}\right)$$

The capacitance of the spherical capacitor is

$$C = \frac{Q}{V_{a-b}} = 4\pi\varepsilon_o\left(\frac{1}{a}-\frac{1}{b}\right)^{-1} \qquad \text{[F]} \qquad (3\text{-}194)$$

Taking the limit as $b \to \infty$, the capacitance of an isolated sphere is therefore

$$C = 4\pi\varepsilon_o a. \qquad \text{[F]} \qquad (3\text{-}195)$$

Example 3-26

Two parallel conducting wires of radius β_1 and length $\mathcal{L}$ are separated by a center-to-center distance of $2\alpha_1$ in free space as shown in Fig. 3.40. Ignoring the fringing effects at the edges, determine the capacitance between two wires.

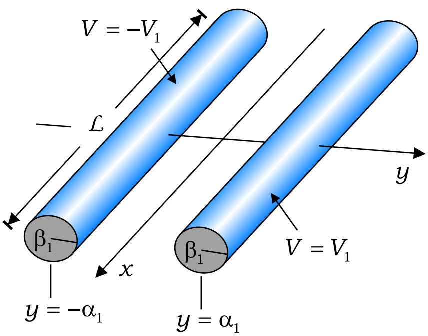


Fig. 3.40 Two parallel conducting wires.

Solution

Assuming potentials V_1 and $-V_1$ on two wires leads to a boundary value problem as shown in Fig. 3.34(c), which we can solve by the equivalent line charges as shown in Fig. 3.34(a). We first obtain K_1 from Eq. (3-180a), and then, by inserting it into Eq. (3-181), the equivalent line charge density as

$$\rho_\ell = \frac{4\pi\varepsilon_o V_1}{\ln K_1}$$

Total charge on the wire of length $\mathcal{L}$ is

$$Q = \rho_\ell \mathcal{L} = \frac{4\pi\varepsilon_o V_1}{\ln K_1}\mathcal{L}$$

The capacitance is therefore

$$C = \frac{Q}{2V_1} = \frac{\pi\varepsilon_o \mathcal{L}}{\ln\left[\alpha_1 + \sqrt{\alpha_1^2 - \beta_1^2}\right] - \ln\beta_1}$$

With an aid of the identity $\ln(a + \sqrt{a^2 - 1}) = \cosh^{-1} a$ ($a > 1$), the capacitance can be written as

$$C = \frac{\pi\varepsilon_o \mathcal{L}}{\cosh^{-1}(\alpha_1 / \beta_1)} \qquad \text{[F]} \qquad (3\text{-}196)$$

By assuming $\alpha_1 >> \beta_1$, the capacitance can be approximated as

$$C = \frac{\pi\varepsilon_o \mathcal{L}}{\ln(2\alpha_1 / \beta_1)}. \qquad \text{[F]} \qquad (3\text{-}197)$$

Exercise 3.35

With reference to Fig. 3.39, find the capacitance of a spherical capacitor by solving Laplace's equation.

Ans. $C = 4\pi\varepsilon_o(a^{-1} - b^{-1})^{-1}$ [F].

Exercise 3.36

What physical quantities are the same for two capacitors connected (a) in parallel, and (b) in series?

Ans. (a) Potential difference, (b) Total charge.

Exercise 3.37

Show that the capacitance of a coaxial cable in Eq. (3-190) reduces to that of a parallel-pate capacitor in Eq. (3-187) under the condition of $b / a \approx 1$.

[Hint: $\ln(b / a) \approx (b / a) - 1$].

Review Questions with Hints

RQ 3.39 Define capacitance. [Eq.(3-183)]

RQ 3.40 Write the capacitances of a parallel-plate, a coaxial, and a spherical capacitor. [Eqs.(3-187)(3-190)(3-194)]

RQ 3.41 Express the energy stored in a capacitor. [Eq.(3-189)]

Problems

3-1 Three point charges are arranged on the x-axis in free space such that $q_1 = +Q$[C] at $x = \ell$, $q_2 = -Q$[C] at $x = 2\ell$, and an unknown charge q_3 is at $x = 3\ell$. Determine q_3 so that the net force on q_1 is zero.

3-2 A point charge, $q_1 = +Q$[C], is at $z = a$ on the z-axis while a second point charge, $q_2 = +Q$[C], is at a point on the y-axis. Locate q_2 on the y-axis so that the force on q_2 in the direction of $\mathbf{a}_y$ is maximum.

3-3 Identical point charges alternate in sign at eight corners of a cube of side a as shown in Fig. 3.41. Find the net electric force acting on the positive charge located at the origin.

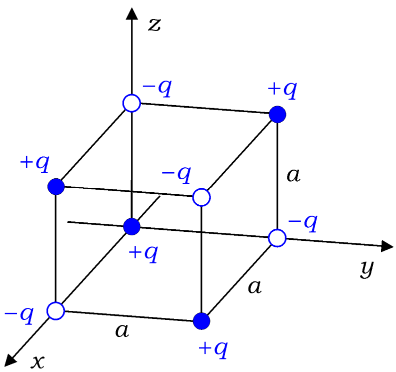


Fig. 3.41 Point charges at eight corners of a cube (Problem 3-3).

3-4 Three identical point charges of q[C] are equally spaced on the perimeter of a circle of radius a. Find the net force on a charge due to the others.

3-5 To examine the relation between the electric and gravitational forces of an electron, consider Fig. 3.42 in which three electrons are thrown into an insulating hemispherical bowl of radius $a = 0.1$[m]. Assuming no friction between the electron and the bowl, find the height of the electrons, h. [Hint: For an electron $q = -1.6\times10^{-19}$[C], $m_e = 9.1\times10^{-31}$[Kg].]

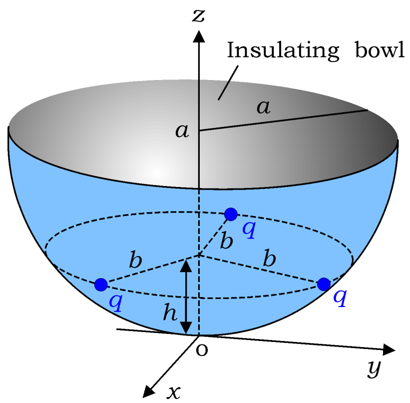


Fig. 3.42 Three electrons in an insulating hemispherical bowl (Problem 3-5).

3-6 A uniform line charge density ρ_ℓ is along an arc that is defined by $\rho = a$, and $0 \le \phi \le \phi_1$ in the $z = 0$ plane, as shown in Fig. 3.43. Find
(a) **E** at the origin, and
(b) ϕ_1 for the maximum $|\mathbf{E}|$ at the origin.

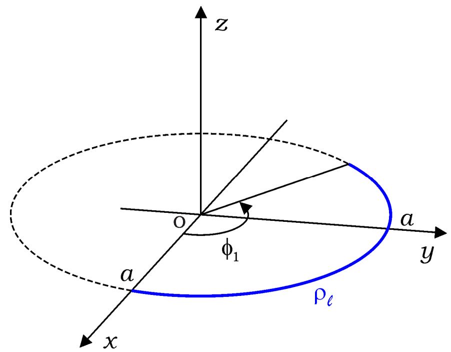


Fig. 3.43 A line charge density along an arc (Problems 3-6, 3-7, 3-29).

3-7 With reference to ρ_ℓ shown in Fig. 3.43 in which the line charge subtends an angle $\phi_1 = \pi$, determine **E** at a point on the z-axis, which is at a distance b from the origin.

3-8 An infinitely long, straight, uniform line charge density ρ_ℓ is parallel to the z-axis and passes through a point $(x, y, z) = (2, 1, 3)$ in free space. Find **E** everywhere.

3-9 Two infinitely long line charges of a uniform density ρ_ℓ are both parallel to the z-axis and pass through points $(a, 0, 0)$ and $(-a, 0, 0)$, respectively, in Cartesian coordinates. Determine **E** at point $(0, b, 0)$ in free space.

3-10 A uniform surface charge density ρ_s is confined in an infinitely long and narrow region defined by $-a \le x \le a$ and $-\infty \le z \le \infty$ in the $y = 0$ plane. Find $\mathbf{E}$ at point $(x, y, z) = (0, b, 0)$ in free space.

3-11 The region occupied by the surface charge density ρ_s in **Problem 3-10** can be subdivided into many infinitely long and narrow strips of a width dx, each of which can be considered as an infinitely long line charge. Find
(a) equivalent line charge density of a strip, and
(b) $\mathbf{E}$ at a point $(x, y, z) = (0, b, 0)$ by adding $\mathbf{E}$'s of the line charges.

3-12 A net charge of Q[C] is uniformly distributed on a disc of radius a, which is placed in the $z = 0$ plane with the center at the origin. A point p is on the +z-axis, at a distance b from the origin. Find
(a) surface charge density ρ_s on the disk, and
(b) electric field intensity $\mathbf{E}_1$ at p.
Let $\mathbf{E}_2$ be the electric field intensity at p obtained as if ρ_s were infinite in extent in the xy-plane, with no change in its value, and $\mathbf{E}_3$ be that at p obtained as if ρ_s were a point charge of Q[C] at the origin. Find the ranges of b in which $\mathbf{E}_1$ can approximate
(c) $\mathbf{E}_2$ with an error less than 1%, and
(d) $\mathbf{E}_3$ with an error less than 1%.

3-13 A uniform volume charge density ρ_v forms a hemispherical shell of inner radius 1 and outer radius 2, as shown in Fig. 2.27. Determine $\mathbf{E}$ at the origin.

3-14 The electric field intensity of an electric dipole is given as $\mathbf{E}(\mathbf{r}) = \left(qd / 4\pi\varepsilon_o R^3\right)\left[2\cos\theta\,\mathbf{a}_R + \sin\theta\,\mathbf{a}_\theta\right]$. Find an expression for the electric field lines in a plane of constant ϕ. [Hint: $dR / (Rd\theta) = E_R / E_\theta$.]

3-15 Three point charges, $3[\mu\text{C}]$, $-5[\mu\text{C}]$, and $4[\mu\text{C}]$, are located at three points, $(x, y, z) = (1[\text{cm}], 0, 0)$, $(0, 1[\text{cm}], 0)$, and $(0, 0, 3[\text{cm}])$, respectively. Find the net outward flux through a cube 4[cm] on a side, which is centered at the origin with the faces normal to the coordinate axes.

3-16 An infinitely long line charge of a density 20[nC/m] is oriented along the x-axis. Find the net electric flux through a sphere of radius 10[cm] centered at the origin.

3-17 In the experiment of electrostatic induction, the outer spherical shell is assumed to be perfectly conducting. Show by Gauss's law that all the induced charges of $-Q$[C] should be distributed on the inner surface of the conductor.

3-18 Two uniform surface charge densities ρ_{s1} and ρ_{s2} form two concentric spheres of radius a and b ($0 < a < b$) with the center at the origin in free space. Determine $\mathbf{D}$ everywhere.

3-19 A uniform surface charge of a density $5[\mu C/m^2]$ forms a sphere of radius $10[cm]$, with the center at the origin, in free space. Meanwhile, an infinitely long line charge of a density $0.1[\mu C/m]$ is parallel to the y-axis and passes through a point $(0,0,20[cm])$ in Cartesian coordinates. Find $\mathbf{D}$ at the point $(0,1[m],0)$.

3-20 An electric flux density is given as $\mathbf{D} = \mathbf{a}_R / R$ in spherical coordinates. Find the total charge contain in a spherical shell of an inner radius a and an outer radius $b(0 < a < b)$, with the center at the origin.

3-21 Given a volume charge density $\rho_v = e^{-|z|}$ $[C/m^3]$ in Cartesian coordinates, find the electric field intensity everywhere.

3-22 Given a volume charge density $\rho_v = 1/\cosh^2(z)$ $[C/m^3]$ in Cartesian coordinates, find $\mathbf{D}$ by using the point form of Gauss's law and considering the symmetry of the charge. [Hint: $d\tanh(x)/dx = 1/\cosh^2(x)$.]

3-23 An infinite plane with a uniform surface charge density ρ_s is defined by $6x + 3y + 2z = 12$. Determine $\mathbf{E}$ at the point p_o:(0.5, 1, 1) in Cartesian coordinates.

3-24 Two infinite, parallel, perfectly conducting plates are separated by an air gap of length d as shown in Fig. 3.44. Two plates carry uniform charges of the opposite polarities. Show that the charges should be distributed on the inner surfaces(bottom of the upper plate and top of the lower plate).
[Hint: Principle of superposition and Gauss's law.]

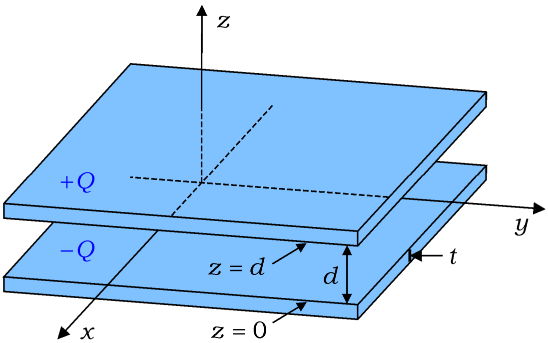


Fig. 3.44 Two parallel conducting plates (Problem 3-24).

3-25 In the presence of an electric field $\mathbf{E} = y\,\mathbf{a}_x + x\,\mathbf{a}_y$ in Cartesian coordinates, find the work done in carrying a charge $3[\mu C]$ from a point p_1:$(3,4,0)$ to a point p_2:$(0,1,0)$ along

(a) parabola defined by $y = (x-1)^2$, and

(b) straight line defined by $y = x + 1$.

(c) Is $\mathbf{E}$ a conservative field?

(d) Check the answer in part (c) with $\nabla \times \mathbf{E}$.

3-26 Consider two vector fields, $\mathbf{E}_1 = e^{-|z|}\mathbf{a}_x$ and $\mathbf{E}_2 = \mathbf{a}_\theta / R^2$. Do they represent static electric fields?

3-27 Two infinitely long, parallel, uniform line charges of densities ρ_ℓ and $-\rho_\ell$ are parallel to the z-axis and pass through two points $(a, 0, 0)$ and $(-a, 0, 0)$, respectively. Find V by taking the origin as the zero reference point.

3-28 For the same line charges as given in **Problem 3-27**, find V everywhere by assuming the zero reference point to be at the point $(b, 0, 0)$, where $|b| < |a|$.

3-29 With reference to the line charge density ρ_ℓ shown in Fig. 3.43, (a) find V at point p:$(0, 0, b)$ if $\phi_1 \neq 2\pi$, (b) is it possible to obtain **E**, at point p, by taking the negative gradient of the result in part (a)? If not, why not?

3-30 An electric potential is given in Cartesian coordinates as $V = 8\ln\left(\sqrt{(x+2)^2 + (y+3)^2} \,/\, \sqrt{(x-2)^2 + (y-3)^2}\right)$, find the location of the zero reference point.

3-31 An electric dipole moment, $\mathbf{p} = \mathbf{a}_z$, is located at the origin. Find an expression for the equipotential surfaces in the $x = 0$ plane.

3-32 Two infinite, parallel, perfectly conducting plates of a negligible thickness intersect the z-axis at right angles at $z = 0.2\,[\text{m}]$ and $z = 0$, respectively carrying uniform surface charge densities $\rho_s = 45\,[\mu\text{C/m}^2]$ and $\rho_s = -45\,[\mu\text{C/m}^2]$, respectively. The lower half of the space in between is filled with a dielectric of $\varepsilon_r = 1.5$. When the lower conducting plate is maintained at a potential $V = 0$, find
(a) **D**, **E**, **P**, and V in the region $0 < z < 0.2\,[\text{m}]$, and
(b) total surface charge density on each plate, the net and induced charges.

3-33 A point charge of q[C] is embedded at the center of a dielectric sphere($\varepsilon_r = 2.5$) of radius a. Find
(a) **D**, **E**, and **P** everywhere, and
(b) ρ_{PV} and ρ_{PS} induced in the dielectric.

3-34 A very large dielectric slab of thickness d is placed on the $z = 0$ plane in an electric field, $\mathbf{E}_1 = 5\,\mathbf{a}_y + 5\sqrt{3}\,\mathbf{a}_z$, as shown in Fig. 3.45. Inside the dielectric, the electric field vector $\mathbf{E}_2$ is observed to be at an angle 45^o to the z -axis. Find
(a) E_2 and ε_r in the dielectric, and
(b) ρ_{PS} induced on the surfaces at $z = 0$ and $z = d$.

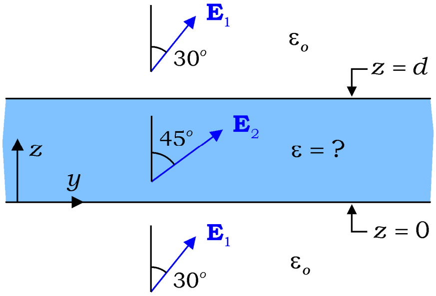


Fig. 3.45 A dielectric slab in an electric field (Problem 3-34).

3-35 With reference to Fig. 3.45, we can replace the dielectric slab with the polarization surface charges ρ_{PS} and $-\rho_{PS}$ residing in an external electric field, $\mathbf{E}_1 = 5\,\mathbf{a}_y + 5\sqrt{3}\,\mathbf{a}_z$, as far as the internal electric field is concerned, as shown in Fig. 3.46. Find the internal electric field $\mathbf{E}_2$ in the region $0 < z < d$, and compare it with the result in **Problem 3-34**.

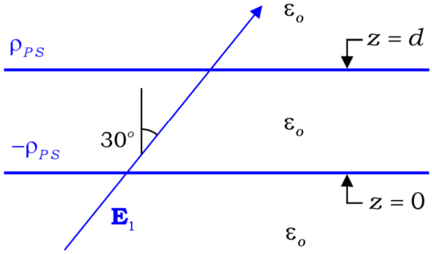


Fig. 3.46 The dielectric slab in Fig. 3.45 is replaced with the polarization surface charges(Problem 3-35).

3-36 A perfectly conducting sphere of radius a is enclosed by a perfectly conducting spherical shell of radii b and c, as shown in Fig. 3.47. The lower half of the space in between is filled with a dielectric of ε_r . If the net charges $+Q$ and $-Q$ are given to the inner and outer conductors, respectively, find

(a) **E** in the gap(air and dielectric),
(b) distributions of $+Q$ and $-Q$ on the conductors, and
(c) polarization surface charge at the dielectric-conductor interface.
(d) Is **E** zero inside the conductors?

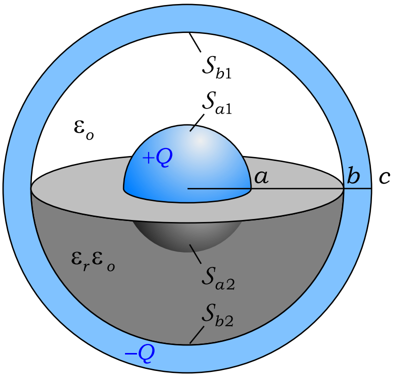


Fig. 3.47 Concentric spheres half filled with a dielectric (Problem 3-36).

3-37 An infinitely long line charge, lying along the z-axis, is parallel to an infinitely long dielectric bar with a concave surface, as shown in Fig. 3.48. We desire to find an expression for the concave surface $\mathcal{S}$, which will cause the electric field $\mathbf{E}_2$ to be parallel to the x-axis inside the dielectric. The surface $\mathcal{S}$ can be represented by the position vector $\mathbf{r} = x(t)\,\mathbf{a}_x + y(t)\,\mathbf{a}_y$, where t is a parameter. In this case, a unit vector tangent to $\mathcal{S}$ is given by $\mathbf{a}_T = A(x'\mathbf{a}_x + y'\mathbf{a}_y)$, and a unit vector normal to $\mathcal{S}$ is given by $\mathbf{a}_N = A(-y'\mathbf{a}_x + x'\mathbf{a}_y)$, where $x' = dx/dt$, $y' = dy/dt$, and A is a constant. Find, in terms of x, y, x', y', and E_1,

(a) tangential and normal components, $|\mathbf{E}_{1T}|$, $|\mathbf{E}_{1N}|$, $|\mathbf{E}_{2T}|$, and $|\mathbf{E}_{2N}|$, on $\mathcal{S}$, and

(b) differential equation for $\mathcal{S}$.

(c) Show that $\mathcal{S}$ is expressed as $(\varepsilon_r - 1)x^2 + \varepsilon_r y^2 = (\varepsilon_r - 1)c^2$ for $h \ll c$ and $(dx/dy) \ll 1$.

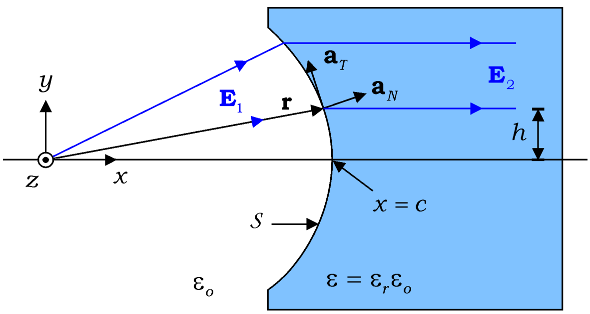


Fig. 3.48 Shaping of electric field lines by a concave surface (Problem 3-37).

3-38 Three identical point charges of 5[nC] are located at the three vertices of a right triangle, in free space, which correspond to points (1,0,0), (0,0,0), and (0,1,0) in Cartesian coordinates(in units of meters). Find the potential energy of the three charges.

3-39 A uniform surface charge density $2[\mu C/m^2]$ forms a sphere of radius 10[cm], with the center at the origin. Find the potential energy.

3-40 With reference to the volume charge density ρ_v as shown in Fig. 3.27, find the potential energies by using
(a) electric field as in Eq. (3-140), and
(b) electric potential as in Eq. (3-136).

3-41 A perfectly conducting spherical surface is defined by $x^2 + y^2 + (z-2)^2 = 4$, which is maintained at a potential 10[V] in free space. Let us try three functions for V outside the sphere, which are chosen so as to satisfy the boundary condition, $V = 10[V]$ on the sphere. Which of the following functions represents V outside the sphere?
(a) $V = \left[x^2 + y^2 + (z-2)^2\right] + 6$.
(b) $V = \sqrt{x^2 + y^2 + (z-2)^2} + 8$.
(c) $V = 20 / \sqrt{x^2 + y^2 + (z-2)^2}$.

3-42 An electric potential is given as $V = V_o e^{-2R} / R$ in spherical coordinates in free space. Find
(a) volume charge density ρ_v everywhere,
(b) electric field intensity everywhere, and
(c) total charge in the whole space.

3-43 A perfectly conducting, elongated ellipsoidal surface is defined as $4x^2 + 4y^2 + z^2 = 4$. It is filled with a volume charge density $\rho_o[C/m^3]$ and connected to ground. By assuming $\varepsilon = \varepsilon_o$ everywhere, find V and $\mathbf{E}$ everywhere.

3-44 Two infinite, parallel, perfectly conducting plates of a negligible thickness intersect the z-axis at right angles at $z = 20[cm]$ and $z = 0$. The upper plate is at a potential 10[V] while the lower one is at zero potential. The lower half of the space in between is filled with an imperfect dielectric containing a volume charge density $4[nC/m^3]$ ($\varepsilon_r = 2.0$). Find V and $\mathbf{E}$ in the gap, air and dielectric.

3-45 In free space, two parallel conducting planes are defined by $2x + y + 3z = 6$ and $2x + y + 3z = 12$, respectively. The former is at a potential 10[V], while the latter is at zero potential. Find V everywhere.

3-46 A conical surface with half angle θ_o is perfectly conducting and is at a potential V_o. It is insulated from another perfectly conducting surface(the

$z = 0$ plane) of zero potential by a tiny air gap, as shown in Fig. 3.49. Find

(a) V and $\mathbf{E}$ in the region $\theta_o < \theta < \pi/2$, and

(b) surface charge densities on two conducting surfaces.

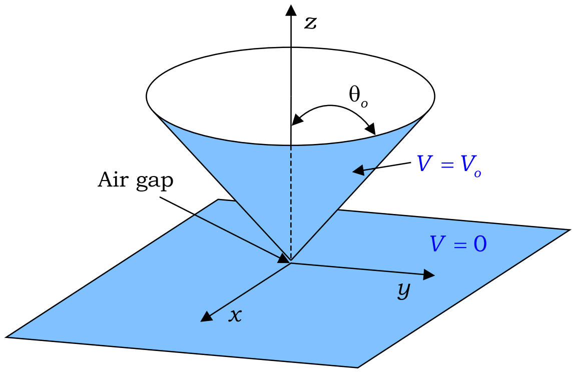


Fig. 3.49 A boundary value problem (Problem 3-46).

3-47 A point charge q is at a distance d from an infinitely large, bent into an L shape, grounded, perfectly conducting plate, as shown in Fig. 3.50. Find

(a) V and $\mathbf{E}$ in the first quadrant of the xy-plane, and

(b) ρ_s induced on the conductor.

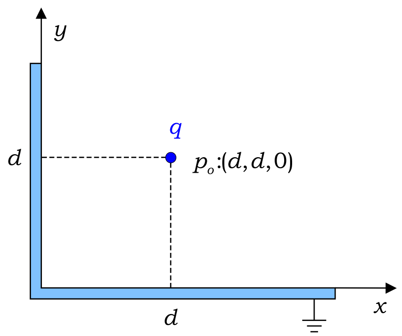


Fig. 3.50 An infinite conducting sheet bent into an L shape (Problem 3-47).

3-48 A very long, straight wire with a uniform line charge density $5[\text{nC/m}]$ is parallel to the x-axis at a distance $2[\text{m}]$ above an infinite, grounded, perfectly conducting surface, coincident with the xy-plane. Find an expression for $\mathbf{E}$ in the region $z > 0$.

3-49 Two very long, parallel, perfectly conducting cylinders of a radius $0.1[\text{m}]$ are separated by a center-to-center distance of $1[\text{m}]$ as shown in Fig. 3.51. Two conductors are maintained at potentials $-100[\text{V}]$ and $100[\text{V}]$, respectively. Find the location and density of the equivalent line charges.

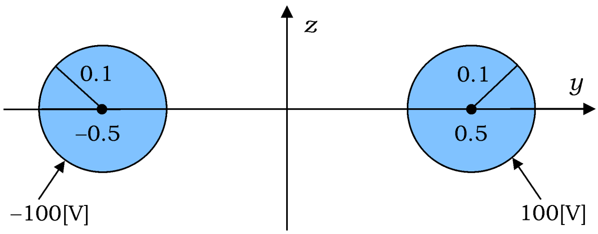


Fig. 3.51 Two parallel conducting cylinders (Problem 3-49).

3-50 A very long straight wire with a line charge density $\rho_\ell = -2[\text{nC/m}]$ is parallel to a very long perfectly conducting cylinder of a radius 0.2[m], carrying a net charge of 2[nC] per unit length, as shown in Fig. 3.52. They are separated by a center-to-center distance of 0.85[m]. Find
(a) V outside the cylinder, and
(b) surface charge densities at points p_1, p_2, and p_3 on the cylinder.
Is the charge uniform on the perfectly conducting cylinder?

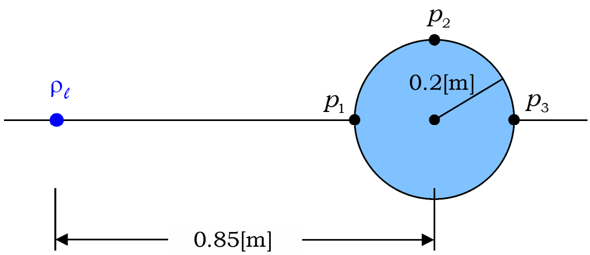


Fig. 3.52 A very long line charge and a very long conducting cylinder (Problem 3-50).

3-51 A parallel-plate capacitor consists of two conducting plates of an area S, separated by a distance d. The lower half of the space in between is filled with a dielectric of a permittivity ε_1 as shown in Fig. 3.53. Ignoring the fringing effects at the edges, show that the capacitance is $C^{-1} = C_1^{-1} + C_2^{-1}$, where C_1 and C_2 are the capacitances of the individual layers.

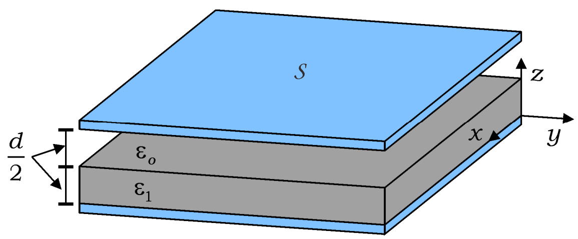


Fig. 3.53 A parallel-plate capacitor (Problem 3-51).

3-52 Consider a parallel-plate capacitor as shown in Fig. 3.54. The left half of the gap is filled with a dielectric of permittivity ε_1, and the plates are maintained at potentials V_o and 0, respectively. Ignoring the fringing effects at the edges, show that the capacitance is $C = C_1 + C_2$, where C_1 and C_2 are the capacitances of the individual sections.

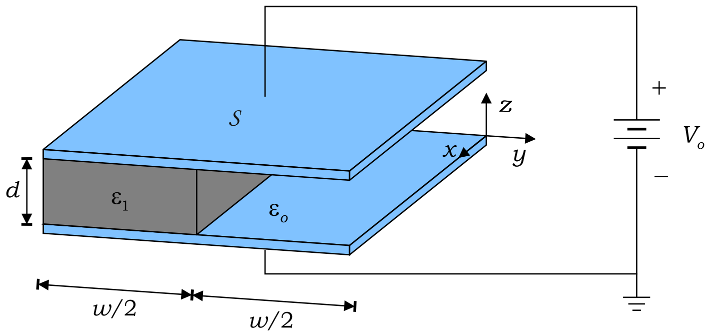


Fig. 3.54 A parallel-plate capacitor (Problem 3-52).

3-53 A parallel-plate capacitor consists of two conducting plates of an area S, which coincide with the $z = d$ and $z = 0$ planes, respectively. The capacitor is filled with an inhomogeneous dielectric of $\varepsilon_r = \sqrt{z+1}$. Ignoring the fringing effects at the edges, find the capacitance.

3-54 A parallel-plate capacitor with a dielectric($\varepsilon = \varepsilon_1$) is charged to a voltage V_o for $-\infty < t < t_o$ (see Fig. 3.55). At time $t = t_o$, the switch is opened, and a new dielectric slab($\varepsilon = \varepsilon_2$) is slowly pushed into the capacitor while the old one is pushed out. Find

(a) force **F** required for inserting the dielectric, and

(b) potential difference between two plates as a function of y.

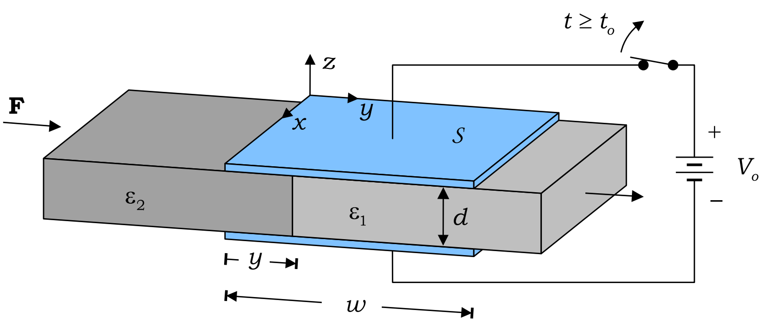


Fig. 3.55 A parallel-pate capacitor (Problem 3-54).

Chapter 4
Steady Electric Current

In the previous chapter, we focused our attention on static electric charges that are fixed in space and constant in time. Otherwise, we assumed that the charges relax to a steady distribution in an instant. Electric charges, however, can move under the influence of an electric field. The charges moving in a conductor constitute a conduction current, while those moving in a vacuum constitute a convection current. From basic circuit theory, the readers should be familiar with the conduction current flowing in a simple electric circuit, which is governed by Ohm's law, stating that the voltage across a resistor is equal to the product of the resistance and the current passing through it. According to the principle of conservation of charge, electric charges cannot be created or destroyed. This principle manifests itself as the equation of continuity in electromagnetics, and Kirchhoff's current law in circuit theory, stating that the sum of all currents entering a junction in an electric circuit is equal to zero. On a macroscopic scale, when we are concerned with the currents flowing in conducting wires, the current is defined as charges passing through a reference point per unit time. At the microscopic scale, when the magnitude and direction of the current are assumed to vary as functions of position in a region of space, we define the current density as charges passing through a reference point per unit area per unit time.

In electromagnetics, we frequently encounter three types of currents: conduction current, convection current, and displacement current density. In the conductor, the loosely bound valence electrons easily detach themselves from the host atoms and make up a sea of electrons, called free electrons or conduction electrons. In the presence of an externally applied electric field, the free electrons gain an average velocity called a drift velocity, and form a conduction current. In contrast, charged particles moving in a vacuum or in a rarefied gas constitute a convection current. The electron beam in a cathode-ray tube, the accelerated electrons in a photomultiplier, and the electrical discharge in a bolt of lightning are a few examples of convection currents. While the conduction and convection currents are directly related to the motion of electric charges, the displacement current density is an equivalent current that involves no electric charges, but behaves as a conduction or convection current as far as the time-varying magnetic field is concerned.

4.1 Convection Current

The convection current is formed by the electric charges moving in a vacuum. In order to describe the spatial variation of the current, we define the current density as charges passing through a unit area of the cross section per unit time. In view of the definition, the current density can be regarded as a kind of flux density. The current density is measured in amperes per square meter $[\mathrm{A/m^2}]$, or coulombs per square meter per second $[\mathrm{C/m^2 \cdot sec}]$. While the current is a scalar quantity, the current density is a vector quantity, which may vary from point to point in space, forming a vector field in a region of space. It is important to note that the current density is defined as the current through a cross section, or a plane perpendicular to the direction of the current.

Let us consider Fig. 4.1, in which a volume charge of an uniform density $\rho_v\,[\mathrm{C/m^3}]$ moves with a constant velocity $\mathbf{v}$, passing through a surface S. The total charge crossing an incremental area Δs in a short period of time Δt is

$$\Delta Q = \rho_v\,|\mathbf{v}|\,\Delta t\,\Delta s\cos\theta \tag{4-1}$$

In the above equation the term $\Delta s\cos\theta$ represents the equivalent area of Δs, which is given in the cross section, or the projection of Δs onto the plane perpendicular to the direction of $\mathbf{v}$. Upon using the relation $\cos\theta = \mathbf{a_v} \bullet \mathbf{a}_s$, where $\mathbf{a_v}$ is a unit vector in the direction of $\mathbf{v}$ and $\mathbf{a}_s$ is a unit normal to Δs, Eq. (4-1) becomes

$$\Delta Q = \rho_v\,|\mathbf{v}|\,\Delta t\,\Delta s(\mathbf{a_v} \bullet \mathbf{a}_s) = \rho_v \mathbf{v} \bullet \Delta\mathbf{s}\,\Delta t \tag{4-2}$$

where we used $\mathbf{v} = |\mathbf{v}|\,\mathbf{a_v}$ and $\Delta\mathbf{s} = \Delta s\,\mathbf{a}_s$. The incremental current through the incremental area Δs is therefore

$$\Delta I = \frac{\Delta Q}{\Delta t} = \rho_v \mathbf{v} \bullet \Delta\mathbf{s} \equiv \mathbf{J} \bullet \Delta\mathbf{s} \tag{4-3}$$

The current density $\mathbf{J}$ is defined, from Eq. (4-3), as

$$\boxed{\mathbf{J} \equiv \rho_v \mathbf{v}} \qquad [\mathrm{A/m^2}] \tag{4-4}$$

where ρ_v is the volume charge density, and $\mathbf{v}$ is the velocity of flow of the charges. The current density has the unit of the ampere per square meter $[\mathrm{A/m^2}]$. The current density is a vector whose unit vector points in the direction of flow of the current, and whose magnitude equals the charges crossing a unit area of the cross section per unit time. The total current through a surface $\mathcal{S}$ is therefore given by the surface integral of $\mathbf{J}$ over $\mathcal{S}$, that is,

$$\boxed{I = \int_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{s}} \qquad \text{[A]} \tag{4-5}$$

The current is measured in amperes [A].

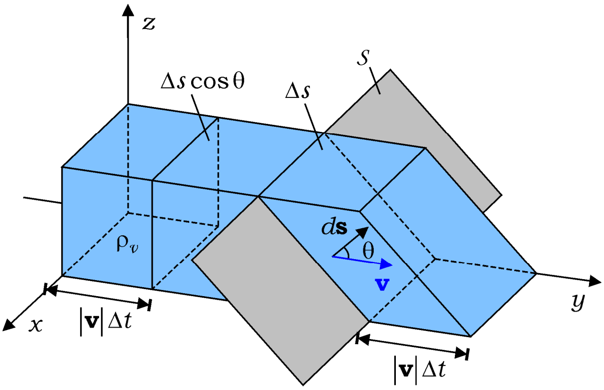


Fig. 4.1 Volume charge density ρ_v moving with a velocity $\mathbf{v}$.

Example 4-1
An electron cloud is accelerated between two electrodes in a photomultiplier as depicted in Fig. 4.2. In the $y = 0$ plane, the volume charge density is assumed to be uniform as $\rho_o[\text{C/m}^3]$, which is negative because of the negative electron charge. The electron cloud crosses the electrodes at $y = 0$ and $y = d$ at the speeds $v_o[\text{m/s}]$ and $v_1[\text{m/s}]$, respectively. Find the current densities in (a) $y = 0$ plane, and (b) $y = d$ plane.

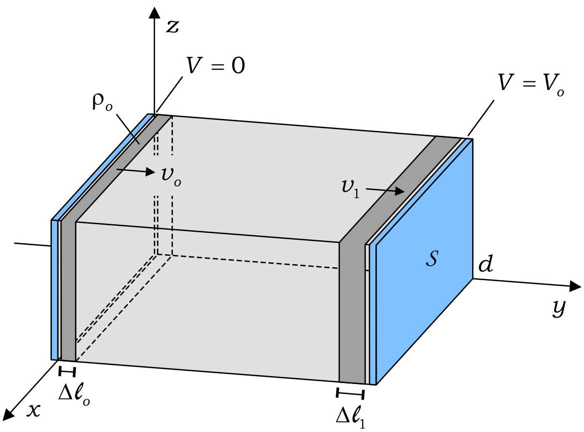


Fig. 4.2 An electron cloud accelerated between two electrodes at $y = 0$ and $y = d$.

Solution

(a) Current density at a point on the $y = 0$ plane is

$$\mathbf{J} = \rho_o \upsilon_o \mathbf{a}_y \qquad (4\text{-}6)$$

(b) The incremental distances traveled by the electrons in a time Δt are

$$\Delta \ell_o = \upsilon_o \Delta t \qquad \text{at } y = 0$$

$$\Delta \ell_1 = \upsilon_1 \Delta t \qquad \text{at } y = d$$

Combining the two equations gives

$$\upsilon_1 = \upsilon_o \frac{\Delta \ell_1}{\Delta \ell_o} \qquad (4\text{-}7)$$

In a time Δt, the net charge crossing the $y = 0$ plane is $\rho_o \Delta \ell_o S$, while the net charge crossing the $y = d$ plane is $\rho_1 \Delta \ell_1 S$, where S is the cross sectional area of the electron beam. From the principle of conservation of charge, the amount of the two charges should be the same, i.e.,

$$\rho_1 = \rho_o \frac{\Delta \ell_o}{\Delta \ell_1} \qquad (4\text{-}8)$$

Inserting Eqs. (4-7) and (4-8) into Eq. (4-4), the current density at a point on the $y = d$ plane is

$$\mathbf{J} = \rho_1 \upsilon_1 \, \mathbf{a}_y = \left(\rho_o \frac{\Delta \ell_o}{\Delta \ell_1} \right) \left(\upsilon_o \frac{\Delta \ell_1}{\Delta \ell_o} \right) \mathbf{a}_y = \rho_o \upsilon_o \, \mathbf{a}_y \qquad (4\text{-}9)$$

We see from Eq. (4-6) and Eq. (4-9) that the current densities are the same even if the charges are accelerated.

Exercise 4.1

The cathode of a CRT provides a current density $0.5[\text{A/cm}^2]$ at 800^oC for an electron beam of a radius $0.24[\text{mm}]$. Find the current carried by the beam.

Ans. $0.9[\text{mA}]$.

Review Questions with Hints

RQ 4.1 What is the relation between current density, volume charge density, and velocity of the charge? [Eq.(4-4)]

RQ 4.2 What is the relation between the current density and total current passing through a surface? [Eq.(4-5)]

RQ 4.3 May accelerated charges constitute a steady current? [Fig.4.2]

4.2 Conduction Current and Ohm's Law

The solid copper is a typical example of a good conductor having a very high conductivity. Copper atoms, having a single valence electron, arrange themselves at regular sites in the face-centered cubic crystal structure. The loosely bound valence electrons easily detach themselves from the lattice atoms and make up an electron cloud in the solid. The electrostatic force between the negatively charged electron cloud and the positively charged copper ions is the origin of the metallic bonding of the solid copper. Initiated by the thermal energy of the solid, the free electrons migrate from atoms to atoms, and, in the course of the migration, collide with the crystal defects, impurities, and mostly with vibrating lattice atoms. In the absence of an externally applied electric field, the free electrons move in random directions exhibiting no net displacement. However, an external electric field can cause the free electrons to accelerate in one direction before they collide with the lattice. In endless cycles of the acceleration and the randomizing collision, the free electrons gain a constant velocity called a drift velocity, and form a steady current called a conduction current in the material.

Under the influence of an externally applied electric field **E**, an electron is accelerated to a velocity **v** in the conductor before it collides with the lattice. Expressed mathematically,

$$\mathbf{v} = \mathbf{v}_o + \frac{e\mathbf{E}}{m_e} t_o \qquad \text{[m/s]}$$

where $\mathbf{v}_o$ is the velocity of the electron just after the collision, t_o is the period of time before the electron collides with the lattice again, and e and m_e are the electron charge and electron mass, respectively. The average of the velocities of all individual electrons is expressed as

$$\bar{\mathbf{v}} = \frac{1}{N}\sum_{j=1}^{N}\left[\mathbf{v}_{oj} + \frac{e\mathbf{E}}{m_e} t_{oj}\right] = \frac{e\mathbf{E}}{m_e}\bar{t}_o \qquad \text{[m/s]} \qquad (4\text{-}10)$$

Since the velocity of an electron just after a collision, $\mathbf{v}_{oj}$, is random in the magnitude and direction, the average of the initial velocities of the electrons is zero in Eq. (4-10). In the above equation, $\bar{t}_o$ is the average time called the mean time between collisions and $\bar{\mathbf{v}}$ is the average velocity called the drift velocity of the free electrons

As can be seen from Eq. (4-10), the drift velocity of the free electrons is directly proportional to the electric field in the metallic conductor, that is,

$$\boxed{\mathbf{v}_e = -\mu_e \mathbf{E}} \qquad \text{[m/s]} \qquad (4\text{-}11)$$

where μ_e is the electron mobility measured in square meters per volt per second $[\mathrm{m^2/V \cdot sec}]$. For examples, $\mu_e = 0.0032[\mathrm{m^2/V \cdot sec}]$ for copper, and $\mu_e = 0.0056[\mathrm{m^2/V \cdot sec}]$ for silver.

Substitution of Eq. (4-11) into Eq. (4-4) leads to the conduction current density, i.e.,

$$\mathbf{J} = -\rho_e \mu_e \mathbf{E} \qquad [\mathrm{A/m^2}] \qquad (4\text{-}12)$$

Here, ρ_e is the volume charge density of the free electrons given by

$$\rho_e = n_e e \qquad (4\text{-}13)$$

where n_e is the number density of the free electrons(number of electrons per unit volume), and e is the electron charge, $e = -1.602 \times 10^{-19}[\mathrm{C}]$. In a metallic conductor, **J** is always parallel to **E**; the minus sign in Eq. (4-12) is canceled by the negative value of ρ_e.

Upon inserting Eq. (4-13) into Eq. (4-12), we obtain the point form of Ohm's law as

$$\boxed{\mathbf{J} = \sigma \mathbf{E}} \qquad [\mathrm{A/m^2}] \qquad (4\text{-}14)$$

The conductivity σ of a metallic conductor is defined as

$$\boxed{\sigma = -n_e e \mu_e} \qquad [\mathrm{S/m}] \qquad (4\text{-}15)$$

which is measured in siemens per meter[S/m], or amperes per volt per mete $[\mathrm{A/V \cdot m}]$. Note that the conductivity σ is always positive. For examples, $\sigma = 5.80 \times 10^7[\mathrm{S/m}]$ for copper, and $\sigma = 6.17 \times 10^7[\mathrm{S/m}]$ for silver.

Semiconductors contain two kinds of charge carriers, electrons and holes. They both contribute to the conductivity of the semiconductor such that

$$\sigma = n_e |e| \mu_e + n_h |e| \mu_h \qquad [\mathrm{S/m}] \qquad (4\text{-}16)$$

where $|e|$ is the absolute value of the electron charge, n_e and n_h are the number densities, and μ_e and μ_h are the mobilities of the electrons and holes, respectively. The conductivity of the intrinsic semiconductor increases with temperature, whereas that of the metal decreases with temperature. This is because the total number of charge carriers increases with temperature in the intrinsic semiconductor, while the electron mobility decreases with temperature in the conductor. In semiconductors, the conductivity is in the range of 10 to $10^{-10}[\mathrm{S/m}]$. For intrinsic silicon, $\mu_e = 0.14[\mathrm{m^2/V \cdot s}]$, $\mu_h = 0.045[\mathrm{m^2/V \cdot s}]$, and

$n_e = n_h = 1.0 \times 10^{16}[\mathrm{m}^{-3}]$ at $300\mathrm{K}$. For germanium, $\mu_e = 0.39[\mathrm{m}^2/\mathrm{V}\cdot\mathrm{s}]$, $\mu_h = 0.19[\mathrm{m}^2/\mathrm{V}\cdot\mathrm{s}]$, and $n_e = n_h = 2.3\times 10^{19}[\mathrm{m}^{-3}]$ at $300\mathrm{K}$.

Example 4-2

Given a copper wire of a diameter $2[\mathrm{mm}]$ for which the conductivity $\sigma = 5.8\times 10^7[\mathrm{S/m}]$ and the electron mobility $\mu_e = 0.0032[\mathrm{m}^2/\mathrm{V}\cdot\mathrm{sec}]$, find

(a) volume density of free electrons, and

(b) drift velocity for a current $25[\mathrm{A}]$ in the wire.

Solution

(a) From Eq. (4-15), we obtain

$$n_e = \frac{\sigma}{|e|\mu_e} = \frac{5.8\times 10^7}{(1.6\times 10^{-19})(0.0032)}$$
$$= 1.13\times 10^{29}[\mathrm{m}^{-3}]$$

(b) Current density in the wire is

$$\mathrm{J} = \frac{current}{area} = \frac{25}{\pi(10^{-3})^2} = 7.96\times 10^6[\mathrm{A/m}^2]$$

The drift velocity is, from Eq. (4-4),

$$\upsilon = \frac{\mathrm{J}}{\rho_\upsilon} = \frac{\mathrm{J}}{n_e|e|} = \frac{7.96\times 10^6}{1.13\times 10^{29}\bullet 1.6\times 10^{-19}}$$
$$= 4.4\times 10^{-4}[\mathrm{m/s}]$$

The free electrons in a copper wire carrying a current $25[\mathrm{A}]$ have a drift velocity of a mere $0.44[\mathrm{mm/s}]$, which is amazingly slow compared with a current pulse propagating at a speed close to light. This discrepancy will be explained in terms of the wave theory in Chapter 7.

Exercise 4.2

Find the ratio between the free electron densities of solid copper and silver by using the known values of σ and μ.

Ans. $n_e^{Cu} / n_e^{Ag} = 1.65$.

Exercise 4.3

In the intrinsic silicon at $300\mathrm{K}$, find the ratio between the conductivities due to the electrons and the holes.

Ans. $\sigma_e / \sigma_h = 3.1$.

Review Questions with Hints

RQ 4.4 Why is it that the drift velocity is constant in a metallic conductor, even though free electrons are accelerated between collisions? [Eq.(4-10)]

RQ 4.5 What causes the conductivity to be different in conductors? [Eq.(4-15)]

RQ 4.6 What makes Ag have a larger σ than Cu(n_e , μ_e , or both)? [Eq.(4-15)]

4.3 Resistance

A conducting body of a finite conductivity is called a resistor. By applying the point form of Ohm's law to a homogeneous conducting material as shown in Fig. 4.3, we can derive the voltage-current relationship for conducting bodies in terms of the conductor length $\mathcal{L}$, cross-sectional area $\mathcal{S}$, and conductivity σ. When a voltage V_{a-b} is applied across terminals a and b, an electric field is established in the conductor in such a way that **E** is directed from the terminal a of a higher potential to the terminal b of a lower potential. If the electric field **E** can be assumed to be uniform over the cross section $\mathcal{S}$, the induced current density **J** is also uniform over $\mathcal{S}$, and such a material is called a homogeneous conducting material. In this case, the total current I flowing through the cross section $\mathcal{S}$ is expressed as

$$I = \int_{\mathcal{S}} \mathbf{J} \bullet d\mathbf{s} = J\mathcal{S} = \sigma E\mathcal{S} \tag{4-17}$$

where we used Eq. (4-14). From Eq. (3-62) we obtain the relationship between the voltage across the terminals and the electric field intensity in the conductor as

$$V_{a-b} = -\int_b^a \mathbf{E} \bullet d\mathbf{l} = E\mathcal{L} \tag{4-18}$$

Combination of Eq. (4-17) with Eq. (4-18) leads to the voltage-current relationship, that is,

$$\boxed{\frac{V_{a-b}}{I} = \frac{\mathcal{L}}{\sigma\mathcal{S}} \equiv R} \qquad [\Omega] \tag{4-19}$$

where $\mathcal{L}$ is the conductor length, $\mathcal{S}$ is the cross-sectional area, and σ is the conductivity. The ratio between the voltage and current is referred to as the resistance, which is measured in ohms[Ω]. The resistance also can be expressed in terms of the resistivity ρ, namely,

$$\boxed{R = \frac{\rho\mathcal{L}}{\mathcal{S}}} \qquad [\Omega] \tag{4-20}$$

The resistivity is the reciprocal of the conductivity, $\rho = 1/\sigma$, and is measured in units of ohm meters $[\Omega \cdot \mathrm{m}]$. The resistivity ρ should not be confused with a radial distance ρ in cylindrical coordinates or a line charge density ρ_ℓ .

From Eq. (4-19) we derive Ohm's law as

$$\boxed{V = RI} \qquad \text{[V]} \qquad (4\text{-}21)$$

This states that the voltage across a resistor is equal to the product of the resistance and the current flowing through the resistor. Notice that subscript a-b is omitted in Eq. (4-21) for simplicity.

The conductance G is the reciprocal of the resistance R such that

$$G = \frac{1}{R} = \frac{\sigma S}{\mathcal{L}} \qquad \text{[S]} \qquad (4\text{-}22)$$

The conductance is measured in units of siemens[S], while the conductivity is measured in units of siemens per meter[S/m].

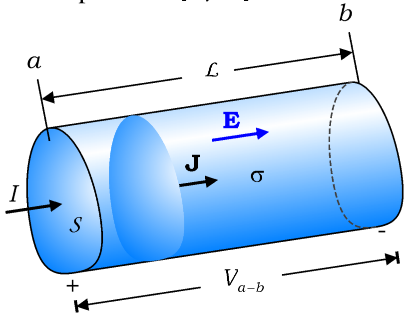


Fig. 4.3 A homogeneous conducting body of a finite conductivity.

The resistance of a homogenous conducting body can be obtained by solving a boundary value problem. The procedure for finding the resistance is as follows:

1. Assume potential difference V_o between conductor terminals.
2. Choose a coordinate system.
3. Find V by solving Laplace's equation.
4. Determine electric field from $\mathbf{E} = -\nabla V$.
5. Obtain current density and total current from $\mathbf{J} = \sigma\mathbf{E}$ and $I = \int_S \mathbf{J} \cdot d\mathbf{s}$.
6. Find R from V_o / I .

Example 4-3

A conducting material of a conductivity σ forms a half-ring with a rectangular cross section, as shown in Fig. 4.4(a). The same material is used for a straight bar having the same cross section and volume as the half-ring, as shown in Fig. 4.4(b). Ignoring the fringing effects at the edges, find the resistances of the two conducting bodies.

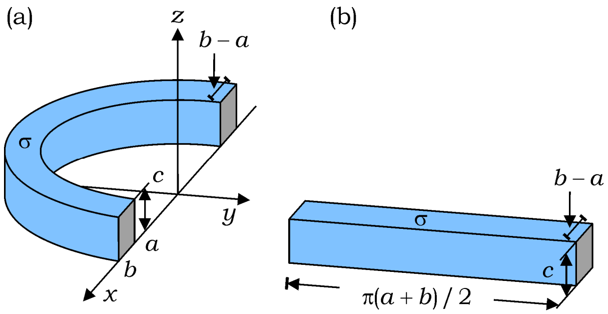


Fig. 4.4 A half-ring and a straight bar with the same cross section and volume.

Solution

(a) Ignoring the edge effects, we can obtain V in the half-ring as if the half-ring were infinite in extent in the ρ- and z-directions in cylindrical coordinates. For the boundary conditions, we assume $V = V_o$ at $\phi = \pi$, and $V = 0$ at $\phi = 2\pi$. With the uniqueness theorem in mind, we assume the planes of constant ϕ to be equipotential surfaces. Under these conditions, V is independent of ρ and z everywhere, and Laplace's equation reduces to

$$\frac{d^2V}{d\phi^2} = 0$$

A general solution is, with the constants of integration c_1 and c_2, written as

$$V = c_1\phi + c_2 \tag{4-23}$$

Applying the boundary conditions to Eq. (4-23), we have

$$V(\phi = \pi) = V_o = \pi c_1 + c_2$$

$$V(\phi = 2\pi) = 0 = 2\pi c_1 + c_2$$

Solving the above equations for c_1 and c_2, and inserting the result into Eq. (4-23), we have

$$V = -\frac{V_o}{\pi}\phi + 2V_o$$

Current density in the conductor is therefore

$$\mathbf{J} = \sigma\mathbf{E} = -\sigma\nabla V$$
$$= \frac{\sigma V_o}{\rho\pi}\mathbf{a}_\phi$$

Total current through the rectangular cross section is given as

$$I = \int_s \mathbf{J} \cdot d\mathbf{s} = \int_{\rho=a}^{\rho=b} \int_{z=0}^{z=c} \frac{\sigma V_o}{\rho \pi} \mathbf{a}_\phi \cdot (d\rho dz\, \mathbf{a}_\phi)$$
$$= \frac{\sigma V_o}{\pi} c \ln(b/a)$$

The resistance of the half-ring is therefore

$$R_1 = \frac{V_o}{I} = \frac{\pi}{\sigma c \ln(b/a)}. \quad (4\text{-}24)$$

(b) The resistance of the bar is obtained simply from Eq. (4-19) as

$$R_2 = \frac{\mathcal{L}}{\sigma \mathcal{S}} = \frac{\pi(a+b)}{2\sigma c(b-a)} \quad (4\text{-}25)$$

It can be shown by a numerical method that the resistance of the straight bar is larger than that of the half-ring, $R_2 > R_1$.

Exercise 4.4
A copper wire of No. 10 AWG is listed as one with a diameter $2.588[\text{mm}]$ and a resistance $3.277[\Omega/\text{Km}]$. Verify the listed resistance.

Exercise 4.5
Find the relationship between the resistance and the wire length, when a conducting wire is being lengthened, while keeping the volume constant.

Ans. $R \sim \mathcal{L}^2$.

Review Questions with Hints

RQ 4.7	State Ohm's law in words.	[Eq.(4-21)]
RQ 4.8	Express the resistance in terms of conductivity.	[Eq.(4-19)]
RQ 4.9	Express the resistance in terms of resistivity.	[Eq.(4-20)]

4.4 Equation of Continuity

According to the principle of conservation of charge, electric charges cannot be created or destroyed. Charges are generated only in pairs of a positive and a negative charge. Therefore, in an electrically neutral conductor, the negative charges of the free electrons are counterbalanced by the positive charges of the ionized lattice atoms; the net charge is zero in a discharged conductor. Certainly, the recombination of an electron with an ionized atom exactly consumes one negative and one positive charge, without changing the net charge in the conductor. The principle of conservation of charge is a law of nature, which is known as the equation of continuity in electromagnetics.

Let us consider the net current flowing out of a closed surface $\mathcal{S}$. The net current can be obtained by the integral of the current density **J** over the closed surface $\mathcal{S}$, that is,

$$I = \oint_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{s} \tag{4-26}$$

For simplicity, let us assume that the current is generated by the positive charges in motion. If the net current I flowing out of the closed surface $\mathcal{S}$ is nonzero, the total charge Q enclosed by the surface should decrease in accordance with the principle of conservation of charge, i.e.,

$$I = -\frac{dQ}{dt} = -\frac{d}{dt}\int_{\mathcal{V}} \rho_v dv \tag{4-27}$$

where ρ_v is the volume charge density, and $\mathcal{V}$ is the volume enclosed by $\mathcal{S}$. With the help of divergence theorem, combination of Eq. (4-26) with Eq. (4-27) leads to

$$\begin{aligned}\int_{\mathcal{V}} \nabla \cdot \mathbf{J}\, dv &= -\frac{d}{dt}\int_{\mathcal{V}} \rho_v dv \\ &= \int_{\mathcal{V}} (-\frac{d\rho_v}{dt}) dv\end{aligned} \tag{4-28}$$

In the above equation, the time derivative is taken inside the volume integral, because $\mathcal{V}$ is independent of time. Next, noting that $\mathcal{V}$ may be arbitrary only if it encloses all charges, the integrands in Eq. (4-28) should be the same at every point in $\mathcal{V}$ for the equality. Namely,

$$\boxed{\nabla \cdot \mathbf{J} = -\frac{\partial \rho_v}{\partial t}} \tag{4-29}$$

This equation is referred to as the equation of continuity, stating that

the net current flowing out of a closed surface is equal to the time rate of decrease of the charge enclosed by the surface.

The current density **J** in Eq. (4-29) may be the conduction or the convection current density, or both.

Under static conditions, the charge density ρ_v is independent of time, and thus Eq. (4-29) becomes

$$\boxed{\nabla \cdot \mathbf{J} = 0} \tag{4-30}$$

Upon taking the integral of Eq. (4-30) over a volume $\mathcal{V}$, and applying divergence theorem, we have

$$\oint_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{s} = 0$$
$$= \int_{\mathcal{S}_1} \mathbf{J} \cdot d\mathbf{s} + \int_{\mathcal{S}_2} \mathbf{J} \cdot d\mathbf{s} + \int_{\mathcal{S}_3} \mathbf{J} \cdot d\mathbf{s} + \ldots \tag{4-31}$$

Here, the closed surface integral of $\mathbf{J}$ over $\mathcal{S}$ is broken up into the integrals of $\mathbf{J}$ over the parts comprising $\mathcal{S}$. Each term on the right-hand side of Eq. (4-31) represents the current through a part of $\mathcal{S}$, as shown in Fig. 4.5. Rewriting Eq. (4-31) in terms of the currents, we obtain Kirchhoff's current law, i.e.,

$$\sum_j I_j = 0 \tag{4-32}$$

It states that the sum of all the currents flowing out of a junction is zero.

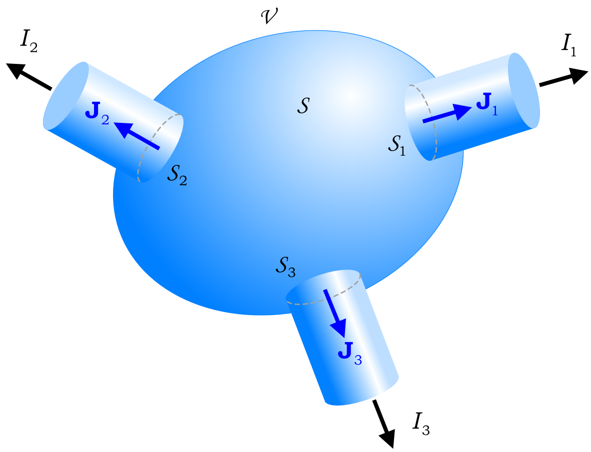


Fig. 4.5 Kirchhoff's current law.

4.4.1 Relaxation Time Constant

As was stated earlier, the net charge imparted to a conductor should be distributed on the surface of the conductor in such a way that $\rho_v = 0$ and $\mathbf{E} = 0$ inside the conductor. By applying the equation of continuity we can compute the time taken by the excess charges to settle in a static distribution. Let us begin with a volume charge density ρ_v introduced in the interior of a good conductor. The charge will induce an electric field in the conductor, according to Gauss's law, as

$$\nabla \cdot \mathbf{E} = \frac{\rho_v}{\varepsilon}$$

By using the point form of Ohm's law, $\mathbf{J} = \sigma\mathbf{E}$, we rewrite the above equation as

$$\nabla \cdot \mathbf{J} = \frac{\sigma}{\varepsilon}\rho_v \tag{4-33}$$

Here, the conductivity σ is constant, independent of **E**, by assuming a homogenous, linear, and isotropic material. Inserting Eq. (4-33) into Eq. (4-29), we obtain

$$\frac{\partial \rho_v}{\partial t} + \frac{\sigma}{\varepsilon}\rho_v = 0 \tag{4-34}$$

Upon solving Eq. (4-34) for ρ_v, applying the initial condition $\rho_v(t = 0) = \rho_o$, we have

$$\boxed{\rho_v(t) = \rho_o e^{-(\sigma/\varepsilon)t}} \tag{4-35}$$

As the excess charges spread out to the surface of the conductor, the initial volume charge density ρ_o exponentially decreases with time in the conductor. The volume charge density reduces to $1/e$, or 36.8%, of its initial value in a relaxation time constant, which is given by

$$\boxed{\tau = \frac{\varepsilon}{\sigma}} \qquad \text{[s]} \tag{4-36}$$

For instance, inserting $\sigma = 5.80 \times 10^7 [\text{S/m}]$ and $\varepsilon \cong \varepsilon_o = 8.854 \times 10^{-12} [\text{F/m}]$ into Eq. (4-36), we obtain the relaxation time constant in copper as $\tau = 1.53 \times 10^{-19} [\text{s}]$, which is the shortest time that we may encounter in electromagnetics.

Exercise 4.6
Determine the relaxation time constant in amber for which dielectric constant is 2.7 and resistivity is $5 \times 10^{14} [\Omega \cdot \text{m}]$.

Ans. 3.3 hours.

Exercise 4.7
Find the ratio between the relaxation time constants in copper and silver, by assuming $\varepsilon \cong \varepsilon_o$ for two conductors.

Ans. $\tau_{Cu} / \tau_{Ag} = 1.06$.

Review Questions with Hints

RQ 4.10 State the equation of continuity. [Eq.(4-29)]
RQ 4.11 State the equation of continuity under static conditions. [Eq.(4-30)]
RQ 4.12 State Kirchhoff's current law. [Eq.(4-32)]

RQ 4.13 Express the relaxation time constant. [Eq.(4-36)]
RQ 4.14 Can you explain the linear relationship between τ and ε in terms of the electric polarization and the internal electric field. [Fig.3.21]

4.5 Power Dissipation and Joules's Law

To recapitulate briefly, an externally applied electric field accelerates the free electrons in a conductor during the mean time between collisions. The free electrons soon collide with the lattice atoms and scatter in random directions. In the course of the acceleration and collision, the free electrons gain kinetic energy from the electric field and dissipate it as thermal energy of the lattice atoms. In other words, some of the potential energy of the electric field is converted to the kinetic energy of the free electrons, and is completely dissipated as heat.

From circuit theory we are already familiar with the electrical power expressed by the product of the voltage and current. Nevertheless, in the present section, we derive the electrical power from the general relation between energy and force. Suppose that an electric field **E** exerts a Coulomb force on a free electron, causing it to move by a distance $|d\mathbf{l}|$ in the direction of $d\mathbf{l}$. If the displacement of the electron is done in a short period of time dt, the power delivered by the electric field to the electron is

$$p = \frac{e\mathbf{E}\cdot d\mathbf{l}}{dt} = e\mathbf{E}\cdot\mathbf{v} \qquad \text{[W]} \qquad (4\text{-}37)$$

where e is the electron charge, and $\mathbf{v} = d\mathbf{l}/dt$ represents the drift velocity of the electron in the material. In view of Eq. (4-37), the total power delivered to the free electrons contained in a differential volume dv is expressed as

$$\begin{aligned} dP &= (n_e dv)\,p = (n_e dv)\,e\mathbf{E}\cdot\mathbf{v} \\ &= \mathbf{E}\cdot\mathbf{J}\,dv \end{aligned} \qquad \text{[W]} \qquad (4\text{-}38)$$

where n_e and **v** are the number density and drift velocity of the free electrons, respectively, and **J** is the current density expressed by Eq. (4-14).

The total power dissipated in a volume $\mathcal{V}$ is equal to the integral of dP over the given volume, i.e.,

$$\boxed{P = \int_{\mathcal{V}} \mathbf{E}\cdot\mathbf{J}\,dv} \qquad \text{[W]} \qquad (4\text{-}39)$$

This is known as Joule's law. Note that the power is measured in watts[W], or joules per second.

From Eq. (4-39) we define the volume power density as

$$\frac{dP}{dv} = \mathbf{E}\cdot\mathbf{J} \qquad [\text{W/m}^3] \qquad (4\text{-}40)$$

which is called the point form of Joule's law.

Joule's law in Eq. (4-39) can be transformed into a more common form, that is, $P = VI$. Consider a straight conducting wire in which a differential volume is defined as $dv = d\ell ds$, where $d\ell$ is a differential length along the wire, and ds is a differential area in the cross section. Assuming that the electric field intensity $\mathbf{E}$ is directed along the wire, and is uniform throughout the cross section, we rewrite Eq. (4-39) as

$$P = \int_{\mathcal{V}} \mathbf{E} \bullet \mathbf{J}\, dv = \left(\int_{\mathcal{L}} E\, d\ell\right)\left(\int_{\mathcal{S}} J\, ds\right) = VI \qquad \text{[W]} \tag{4-41}$$

Substitution of Ohm's law expressed by Eq. (4-21) into Eq. (4-41) leads to the familiar expression for the ohmic power loss:

$$\boxed{P = I^2 R} \qquad \text{[W]} \tag{4-42}$$

which is the power dissipated in a resistance *R*.

Example 4-4
A silver fuse of slow-blow type is designed to disconnect the circuit in 10 seconds at a current twice as large as the current rating of the fuse. For a metallic silver, specific heat is $0.233[\mathrm{J/g \cdot K}]$, density is $10.49[\mathrm{g/cm^3}]$, resistivity is $1.59 \times 10^{-8}[\Omega \cdot \mathrm{m}]$, and melting point is $1235[\mathrm{K}]$. Assuming these coefficients are independent of temperature, determine the diameter of a silver fuse rated at 20[A] of a dc-current.

Solution

Let the silver fuse have cross section $\mathcal{S}[\mathrm{m^2}]$ and length $\mathcal{L}[\mathrm{m}]$. The thermal energy required to raise the fuse temperature from 293[K] (room temperature) to 1235[K] (melting point) is

$$W_T = 0.233\left[\frac{\mathrm{J}}{\mathrm{g \cdot K}}\right] \times 10.49\left[\frac{\mathrm{g}}{\mathrm{cm^3}}\right] \times \left(\mathcal{S} \times \mathcal{L} \times 10^6\right)[\mathrm{cm^3}] \times (1235 - 293)[\mathrm{K}]$$
$$= (\mathcal{S}\mathcal{L}) \times 2.30 \times 10^9 [\mathrm{J}]$$

Total electrical energy expended in the resistance *R*, in 10[s], by a dc-current of 40[A] is

$$W_E = I^2 R \times 10[\mathrm{s}] = 40^2 \times \frac{1.59 \times 10^{-8}\mathcal{L}}{\mathcal{S}} \times 10 = 2.54 \times 10^{-4}\frac{\mathcal{L}}{\mathcal{S}}[\mathrm{J}]$$

Equating W_T with W_E, we obtain the diameter D as follows:

$$\mathcal{S}^2 \equiv (\pi D^2/4)^2 = \frac{2.54 \times 10^{-4}}{2.30 \times 10^9} = 1.10 \times 10^{-13}$$

Thus,

$$D = 0.65[\mathrm{mm}]$$

Exercise 4.8
A copper wire of No. 10 AWG, as previously mentioned in **Exercise 4.4**, is rated at 30[A]. Find the maximum power dissipated in the wire per kilometer.

Ans. 2.95[KW].

Exercise 4.9
Consider two voltage sources of 110[V] and 220[V], which are to deliver the same power to identical loads through identical pairs of wires. Find the ratio between (a) currents, and (b) power losses in two pairs of wires.

Ans. (a) $I_{110} / I_{220} = 2$, (b) $P_{110} / P_{220} = 4$.

Exercise 4.10
Is it true that the total power dissipated in the resistors connected (a) in parallel, and (b) in series is the sum of the powers dissipated in the individual resistors?

Ans. (a) Yes, (b) Yes.

Review Questions with Hints

RQ 4.15 Express Joule's law in terms of current density. [Eq.(4-39)]
RQ 4.16 Express ohmic power loss. [Eq.(4-42)]
RQ 4.17 How can you reduce the ohmic power loss in a power transmission line? [Eq.(4-42)]

4.6 Steady Currents at an Interface

The steady current density **J** can be uniquely determined in a region of space if its divergence and its curl are specified in the given region, according to Helmholtz's theorem. The equation of continuity for the steady current given in Eq. (4-30) defines the divergence of **J** as

$$\boxed{\nabla \cdot \mathbf{J} = 0} \tag{4-43}$$

It is important to note that Eq. (4-43) is independent of the material occupying the region under consideration. Next, upon combining one of the fundamental laws of electrostatics, $\nabla \times \mathbf{E} = 0$, and the point form of Ohm's law, $\mathbf{J} = \sigma\mathbf{E}$, we obtain the curl of **J** as follows:

$$\boxed{\nabla \times (\mathbf{J} / \sigma) = 0} \tag{4-44}$$

In general, the conductivity σ may not be taken outside the curl operator, because it may vary as a function of position in an inhomogeneous material, and change across an interface between two differential materials. We note that Eqs. (4-43) and (4-44)constitute two fundamental relations for the current density under static conditions.

The integral forms for Eqs. (4-43) and (4-44) can be obtained by taking a volume integral of Eq. (4-43) and a surface integral of Eq. (4-44), and then applying divergence and Stokes's theorems, respectively. These procedures lead to the integral form of the governing equations for the steady current density:

$$\oint_S \mathbf{J} \cdot d\mathbf{s} = 0 \tag{4-45}$$

$$\oint_C \frac{1}{\sigma} \mathbf{J} \cdot d\mathbf{l} = 0 \tag{4-46}$$

Here, S is a closed surface and C is a closed loop; they are independent of each other. The conductivity σ in Eq. (4-46) may not be taken outside the integral, because the closed loop C may traverse materials of different conductivities.

The same procedure as was used for the boundary conditions for **E** and **D** in Chapter 3 can be followed to obtain the boundary conditions for **J** at an interface between two different conductors of conductivities σ_1 and σ_2. That is,

$$\boxed{J_{1n} = J_{2n}} \tag{4-47}$$

$$\boxed{\frac{J_{1t}}{\sigma_1} = \frac{J_{2t}}{\sigma_2}} \tag{4-48}$$

where n and t stand for the normal and tangential components, respectively. The normal component of **J** is continuous across the interface, whereas the tangential component of **J** is discontinuous across the interface.

Example 4-5

The $z = 0$ plane is an interface between two lossy dielectrics with permittivities ε_1 and ε_2, and conductivities σ_1 and σ_2, respectively, as shown in Fig. 4.6. In region 2($z < 0$), a steady current is given as $\mathbf{J}_2 = b\,\mathbf{a}_y + c\,\mathbf{a}_z$ [A/m^2]. Find

(a) $\mathbf{J}_1$ in region 1($z > 0$),

(b) $\mathbf{E}_1$ and $\mathbf{E}_2$, and

(c) surface charge density induced on the interface.

Solution

(a) Let $\mathbf{J}_1 = J_{1y}\,\mathbf{a}_y + J_{1z}\,\mathbf{a}_z$ in region 1. Then, from the boundary conditions,

$$J_{1y} = \frac{\sigma_1}{\sigma_2} b \qquad \text{(tangential component)}$$

$$J_{1z} = c \qquad \text{(normal component)}$$

Combining two components we get

$$\mathbf{J}_1 = \frac{\sigma_1}{\sigma_2} b\,\mathbf{a}_y + c\,\mathbf{a}_z .$$

(b) By the point form of Ohm's law we obtain

$$\mathbf{E}_1 = \frac{\mathbf{J}_1}{\sigma_1} = \frac{b}{\sigma_2}\mathbf{a}_y + \frac{c}{\sigma_1}\mathbf{a}_z$$

$$\mathbf{E}_2 = \frac{\mathbf{J}_2}{\sigma_2} = \frac{b}{\sigma_2}\mathbf{a}_y + \frac{c}{\sigma_2}\mathbf{a}_z \,.$$

(c) Normal components of **D** at the interface are

$$D_{1z} = \varepsilon_1 E_{1z} = c\frac{\varepsilon_1}{\sigma_1}$$

$$D_{2z} = \varepsilon_2 E_{2z} = c\frac{\varepsilon_2}{\sigma_2}$$

From the boundary condition for the normal component of **D** at the interface, we obtain

$$\rho_s = D_{1z} - D_{2z} = c\left(\frac{\varepsilon_1}{\sigma_1} - \frac{\varepsilon_2}{\sigma_2}\right) \qquad [\text{A/m}^2] \qquad (4\text{-}49)$$

We note that the induced ρ_s is proportional to the difference between the relaxation time constants in two adjoining conductors.

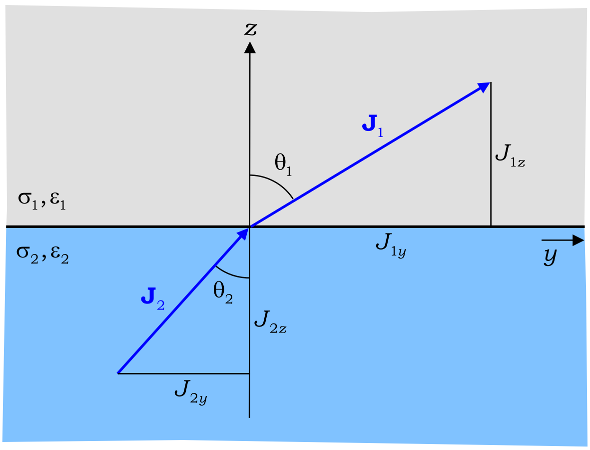


Fig. 4.6 Steady currents in two adjoining conductors.

Exercise 4.11
With reference to Fig. 4.6, express the ratio between two conductivities in terms of the tilt angles of **J**.

Ans. $\tan\theta_1 / \tan\theta_2 = \sigma_1 / \sigma_2$.

Review Questions with Hints

RQ 4.18 State two governing equations for **J**. [Eqs.(4-43)(4-44)]

RQ 4.19 State boundary conditions for **J**. [Eqs.(4-47)(4-48)]

RQ 4.20 Which component of **J** is responsible for the surface charge induced on a conductor-conductor interface? [Eq.(4-49)]

4.7 Analogy between D and J

The permittivity ε is a characteristic of the dielectric, whereas the conductivity σ is a characteristic of the conductor. When a material of a permittivity ε and a conductivity σ is in an electric field **E**, the flux density **D** is related to **E** by ε, and the current density **J** is related to **E** by σ, i.e.,

$$\mathbf{D} = \varepsilon\mathbf{E} \tag{4-50a}$$

$$\mathbf{J} = \sigma\mathbf{E} \tag{4-50b}$$

In the same way as the closed surface integral of **D** is equal to the enclosed net charge Q, the closed surface integral of **J** is equal to the outward total current I, i.e.,

$$\oint_S \mathbf{D} \cdot d\mathbf{s} = Q \tag{4-51a}$$

$$\oint_S \mathbf{J} \cdot d\mathbf{s} = I \tag{4-51b}$$

In the above equations, S may refer to the same closed surface. In addition, we see from the boundary conditions that both **D** and **J** terminate on the surface of the perfect conductor at right angles. In view of these considerations, we note that **D**, ε, and Q are analogous to **J**, σ, and I, respectively.

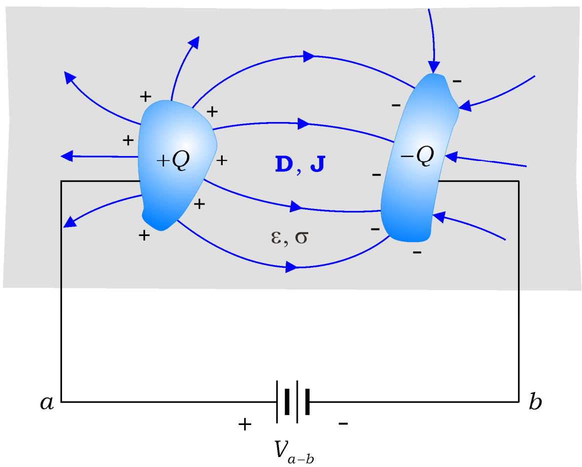


Fig. 4.7 Analogy between **D** and **J**.

Let us now consider two perfect conductors embedded in a lossy dielectric, which are connected to a dc-voltage source as illustrated in Fig. 4.7. From the analogy between **D** and **J**, we can derive a simple relation between the capacitance and the leakage resistance between two conductors, which may be useful for determining the capacitance from the known resistance, and vice versa. With reference to Fig. 4.7, the potential difference between the two conductors is

$$V_{a-b} = -\int_b^a \mathbf{E} \bullet d\mathbf{l} \tag{4-52}$$

From Eq. (4-51b) and Eq. (4-52) we can express the leakage resistance between two conductors as

$$R = \frac{V_{a-b}}{I} = \frac{-\int_b^a \mathbf{E} \bullet d\mathbf{l}}{\oint_S \mathbf{J} \bullet d\mathbf{s}} \tag{4-53}$$

Next, from Eq. (4-51a) and Eq. (4-52) we can express the capacitance between two conductors as

$$C = \frac{Q}{V_{a-b}} = \frac{\oint_S \mathbf{D} \bullet d\mathbf{s}}{-\int_b^a \mathbf{E} \bullet d\mathbf{l}} \tag{4-54}$$

The line integral of **E** in Eq. (4-53) is exactly the same as that in Eq. (4-54). In a simple medium in which ε and σ are constant, the product of Eq. (4-53) and Eq. (4-54), with the aid of Eq. (4-50), leads to

$$\boxed{RC = \frac{\varepsilon}{\sigma}} \tag{4-55}$$

The product of R and C of a two-conductor configuration is equal to the relaxation time constant in the material in between. By using Eq. (4-55) we can determine the leakage resistance from the known capacitance, and vice versa. We can rewrite Eq. (4-55), with the aid of $R = 1/G$, as

$$\boxed{\frac{C}{G} = \frac{\varepsilon}{\sigma}} \tag{4-56}$$

It is important to note that Eqs. (4-55) and (4-56) are true only in a simple medium in which both ε and σ are constant.

Example 4-6

Find the leakage resistance per unit length of a coaxial capacitor, which is filled with a lossy dielectric of permittivity ε and conductivity σ. The radius of the inner conductor is a, and the inner radius of the outer conductor is b.

Solution

From Eq. (3-190), the capacitance per unit length of the coaxial capacitor is

$$C = \frac{2\pi\varepsilon}{\ln(b/a)} \qquad [\text{F/m}]$$

The leakage resistance is, from Eq. (4-55),

$$R = \frac{\varepsilon}{\sigma}\frac{1}{C} = \frac{\ln(b/a)}{2\pi\sigma} \qquad [\Omega] \tag{4-57}$$

We of course obtain the same resistance as Eq. (4-57) if we follow the procedure for R given in Section 4-3.

Exercise 4.12

A parallel-plate capacitor of capacitance $0.01[\mu\text{F}]$ has a lossy dielectric of $\varepsilon_r = 2.5$ and $\sigma = 10^{-3}[\text{S/m}]$ between two plates. Find its leakage resistance.

Ans. $R = 2.2[\Omega]$.

Exercise 4.13

With reference to the coaxial capacitor given in **Example 4-6**, find the leakage resistance following the procedure for R given in Section 4.3.

Review Questions with Hints

RQ 4.21 What makes the field lines of **D** and **J** appear the same? [Eq. (4-50)]

RQ 4.22 Write the relations that exhibit analogy between **D** and **J**. [Eqs. (4-50)(4-51)]

RQ 4.23 Write the relation between the capacitance and the leakage resistance. [Eq. (4-55)]

Problems

4-1 Given that a current density $\mathbf{J} = 0.3\,\mathbf{a}_y + 0.4\,\mathbf{a}_z[\text{A/cm}^2]$, find the total current crossing the cylindrical surface defined by $\rho = a = 2[\text{cm}]$, $0 \le \phi \le \pi$, and $0 \le z \le b = 2[\text{cm}]$, as shown in Fig. 4.8.

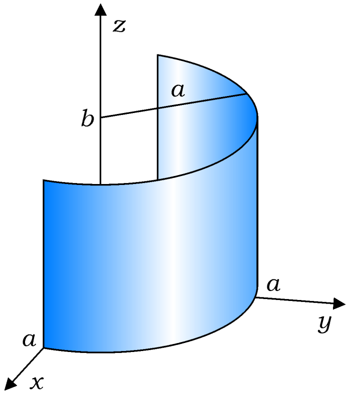


Fig. 4.8 A cylindrical surface(Problem 4-1).

4-2 A current density is given as $\mathbf{J} = 20e^{-10\rho^2}\,\mathbf{a}_z$ $[\mathrm{A/m^2}]$ in cylindrical coordinates(ρ in units of meters). Find the total current passing through the spherical surface defined by $R = 0.4\,[\mathrm{m}]$ and $0 \le \theta \le 50^o$ as shown in Fig. 4.9.

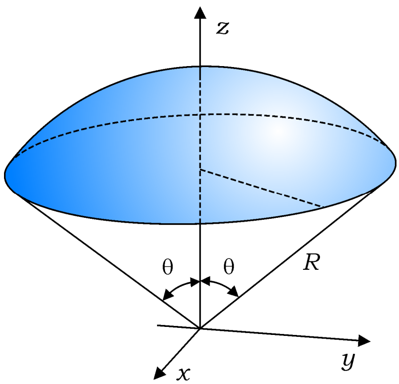


Fig. 4.9 A spherical surface (Problem 4-2).

4-3 With reference to the spherical surface($R = 0.4\,[\mathrm{m}]$ and $\theta = 50^o$) as shown in Fig. 4.9, if the current density is given as $\mathbf{J} = 6\cos(\rho^2)\,\mathbf{a}_z$ $[\mathrm{A/m^2}]$ in cylindrical coordinates(ρ in units of meters), find the total current through the surface.
[Hint: Equation of continuity.]

4-4 The heated cathode residing in the $z = 0$ plane induces thermionic emission of electrons of a density N $[\mathrm{electrons/m^2 \cdot sec}]$. The anode residing

in the $z = d$ plane is at a higher potential than the cathode by V_o [V], accelerating the electrons in the z-direction. Find, as a function of position,
(a) velocity of the electrons,
(b) volume charge density, and
(c) current density between the electrodes.
(Electron mass is m, and electron charge is e.)

4-5 A straight wire of radius a is made of a conducting material($\sigma = \sigma_1$) and coated with another conducting material($\sigma = \sigma_2$). Determine the outer radius of the coated wire that carries the same amount current in the core and the coating, if the electric field is uniform in the cross section.

4-6 A silicon bar with a rectangular cross section $4[\text{mm}] \times 2[\text{mm}]$ is $10[\text{mm}]$ long. For intrinsic silicon, $\mu_e = 0.14[\text{m}^2/\text{V}\cdot\text{s}]$, $\mu_h = 0.045[\text{m}^2/\text{V}\cdot\text{s}]$, and $n_e = n_h = 1.0 \times 10^{16}[\text{m}^{-3}]$ at 300K. If the voltage drop along the length of the bar is 5[V], find
(a) conductivity,
(b) total current,
(c) resistance, and
(d) power dissipated in the silicon bar.

4-7 When an intrinsic silicon is doped with a small amount of donors, the electron density is increased by 10% at 300K. How much is the conductivity increased due to the doping?
[Hint: Use typical values of the electron and hole mobilities.]

4-8 Two concentric, perfectly conducting cylindrical surfaces of radii $a = 5[\text{cm}]$ and $b = 6[\text{cm}]$ are $10[\text{cm}]$ long. The space in between is filled with graphite of a conductivity $\sigma = 7 \times 10^4[\text{S/m}]$. Ignoring the edge effects, find the resistance between two conductors.

4-9 An ohmic medium of conductivity σ fills the gap between two concentric hemispheres of radii a and b as shown in Fig. 4.10. Ignoring the edge effects, find the resistance.

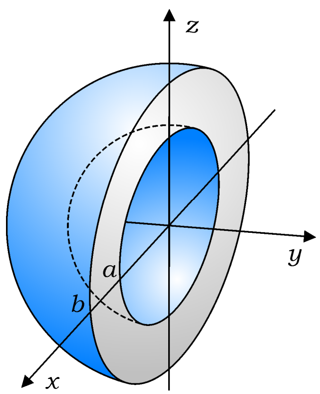


Fig. 4.10 Concentric hemispheres filled with an ohmic medium(Problem 4-9).

4-10 The conductivity of copper depends of temperature as $\sigma = \sigma_o\left[1+\alpha\left(T-T_o\right)\right]^{-1}$, where T is temperature in degrees Celsius, $\sigma_o = 5.8\times10^7$[S/m] at $T_o = 20^o\mathrm{C}$, and the temperature coefficient $\alpha = 0.0039$ [$^o\mathrm{C}^{-1}$]. Determine the resistance of 10[m]-long copper wire of diameter 0.2[mm] at temperature (a) $T = -30^o\mathrm{C}$, and (b) $T = 40^o\mathrm{C}$.

4-11 Consider two concentric, perfectly conducting spherical shells of radii a and b as shown in Fig. 4.11. The space in between is filled with a lossy dielectric of conductivity σ and permittivity ε. This lossy capacitor has been charged to $\pm Q_o$[C] and, at time $t = 0$, disconnected from the source. Find the current density in the gap for $t > 0$, ignoring the electromagnetic induction that may be caused by the time-varying current.

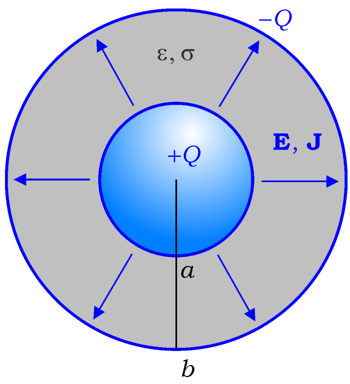


Fig. 4.11 Two concentric spherical shells filled with a lossy dielectric(Problem 4-11).

4-12 A cylindrical capacitor of length L consists of two concentric, perfectly conducting cylindrical surfaces of radii a and c as shown in Fig. 4.12. The gap is filled with two different lossy dielectrics: (σ_1 and ε_1) for $a < \rho < b$ and (σ_2 and ε_2) for $b < \rho < c$. It is connected to a voltage source of V_o[V]. Find (a) current density, and (b) resistance between two conductors.

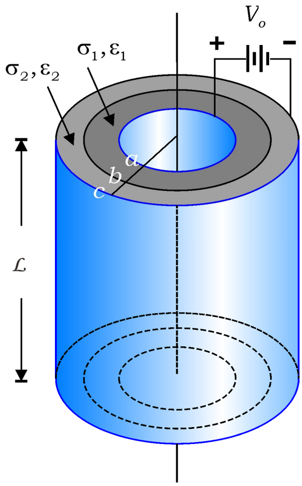


Fig. 4.12 Cylindrical capacitor filled with two different materials(Problems 4-12, 4-13).

4-13 When the cylindrical capacitor shown in Fig. 4.12 was disconnected from the source at time $t = t_o$, a net charge of Q_o [C] was observed on the inner conductor. Find the current density between two conductors for $t > t_o$.

4-14 Two parallel conducting plates of an area S coincide with the $z = d$ and $z = 0$ planes, which are at potentials $V = V_o$ and $V = 0$, respectively. The gap in between is filled with an inhomogeneous material of conductivity $\sigma = \sigma_o \sqrt{1 + (z/d)}$. Ignoring the edge effects, find

(a) current density,
(b) electric field, and
(c) resistance between two conductors.

4-15 The $z = 0$ plane is an interface between two lossy dielectrics: (σ_1 and ε_1) for $z < 0$ and (σ_2 and ε_2) for $z > 0$. A uniform current density $\mathbf{J}_1 = 2\mathbf{a}_y + 3\mathbf{a}_z$ [A/m^2] exists in the region $z < 0$. Find

(a) $\mathbf{J}_2$ in the region $z > 0$,
(b) electric fields in both regions, and
(c) surface charge density on the interface.

4-16 A parallel-plate capacitor has two conducting plates of an area S, which are separated by a lossy dielectric of thickness d, conductivity σ, and permittivity ε. The capacitor has been fully charged to a voltage $V = V_o$ by a voltage source, and then disconnected from the source at time $t = 0$. Find, for $t > 0$,

(a) charge density on the conductor,
(b) power dissipated in the dielectric, and
(c) Show that the total energy expended in the dielectric equals the total energy stored in the capacitor at $t = 0$.

Chapter 5
Magnetostatics

In Chapter 3, we focused our discussions on electrostatics, which is primarily concerned with the electric field **E** and electric flux density **D**. Initially, we introduced the electric field to describe the interaction between static electric charges separated by free space. We defined the electric flux density to account for the interaction between an external electric field and a material medium. We could define the permittivity ε of a material medium from the relation between **D** and **E**. We saw that the static electric field, in a region of space, is governed by the two fundamental relations, $\nabla \cdot \mathbf{D} = \rho_v$ and $\nabla \times \mathbf{E} = 0$, which are specified in the region.

In this chapter, we turn our attention to magnetostatics, which is primarily concerned with magnetic field **H** and magnetic flux density **B**. As **E** and **D** are related through the constitutive relation $\mathbf{D} = \varepsilon \mathbf{E}$ in a simple medium, **H** and **B** are related through $\mathbf{B} = \mu \mathbf{H}$ in a simple magnetic medium. As the static electric field is governed by the two fundamental relations, the static magnetic field, in a region of space, is governed by the fundamental relations, $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{H} = \mathbf{J}$, which are specified in the region. In view of these relations, an analogy between **E** and **H** and between **D** and **B** is apparent. Although the analogy can help us comprehend the concept of magnetostatics to a certain degree, we should not rely entirely on this analogy in studying magnetostatics. For instance, **B** is responsible for the magnetic force on an electric charge in motion, as **E** is responsible for the electric force exerted on an electric charge. In the same way as **D** is defined in a dielectric, **H** is newly defined in a magnetic material.

Magnetism was first discovered by ancient people in pieces of lodestone, which attracted iron pieces. Lodestone is a naturally magnetized piece of magnetite found near the ancient city called Magnesia. The word magnet is derived from the Greek word for Magnesian stone. Magnetism was thought to be independent of electricity until 1819, when Hans Christian Oersted observed that the needle of a compass was deflected when placed near a current-carrying wire. Oersted's observation established a link between magnetic and electric fields, in that when an electric field produces a current in a wire, the current generates a magnetic field. The link is not complete, however, because a static magnetic field cannot generate a static electric field. We will learn in Chapter 6 that a time-varying **B** induces a time-varying **E**, completing the link between magnetic and electric fields under time-varying conditions.

A steady current, or a direct current, generates a static magnetic field. Our discussion of static magnetic fields starts with the Biot-Savart law, which relates a static magnetic field to a steady current. We next explore Ampere's circuital law, which is a special case of the Biot-Savart law, just as Gauss's law is a special case of Coulomb's law. Ampere's circuital law is particularly useful when the geometry under consideration has certain symmetry. We introduce the concept of magnetization vector to account for the interaction between an external magnetic field and a magnetic material. The magnetic field intensity **H** is redefined in the magnetic material to account for the effect of the magnetization on the total magnetic flux in the material. We will learn that the magnetization leads to the constitutive relation of the magnetic material, $\mathbf{B} = \mu\mathbf{H}$, where the permeability μ is characteristic of the material. After we obtain two fundamental relations for static magnetic fields, $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{E} = \mathbf{J}$, we discuss the boundary conditions for **B** and **H** at an interface between two different magnetic materials. We continue with our discussion to cover inductance, magnetic energy, and torque.

5.1 Lorentz Force Equation

We recall from Chapter 3 that an electric charge q, either at rest or in motion, experiences an electric force $\mathbf{F}_e$ in the presence of an electric field **E**, that is,

$$\mathbf{F}_e = q\mathbf{E} \qquad \text{[N]} \tag{5-1}$$

As the electric field intensity **E** is defined as the electric force on a unit charge, the magnetic flux density **B** can be defined as the magnetic force exerted on a unit charge in motion. It was found from experiments that an electric charge q moving with a velocity **v** in a magnetic flux density **B** experiences a magnetic force $\mathbf{F}_e$. The magnetic force is expressed mathematically as

$$\mathbf{F}_m = q\mathbf{v} \times \mathbf{B} \qquad \text{[N]} \tag{5-2}$$

The magnetic flux density **B** has the unit of the tesla[T] or the weber per square meter $[\mathrm{Wb/m^2}]$.

If a charge q moves with a velocity **v** in the presence of both an electric field **E** and a magnetic flux density **B**, the total force exerted on the charge is therefore

$$\boxed{\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right)} \qquad \text{[N]} \tag{5-3}$$

which is called the Lorentz force equation.

It is important to note that $\mathbf{F}_e$ is exerted on an electric charge, whether or not it is in motion, but $\mathbf{F}_m$ is exerted on an electric charge in motion only. Because $\mathbf{F}_m$ is always perpendicular to the direction of motion of the charge, or the direction of **v**, there is no expenditure of the magnetic energy, even though the charge is displaced in the force field that is produced by the magnetic field. In other words, there is no transfer of the magnetic energy stored in **B** to the kinetic energy of the charge. For this reason, the magnetic field cannot change the speed of the charge but changes the direction of motion of the charge.

Example 5-1

In a uniform magnetic flux density, $\mathbf{B} = B_o \mathbf{a}_z$, an electron crosses the x-axis at right angles, with an initial velocity $\mathbf{v} = \upsilon_o\, \mathbf{a}_y$, at time $t = 0$.

(a) Show that a circle is traced by the electron in the xy-plane for time $t \ge 0$.

(b) Find the radius r_o of the trace in terms of the electron charge e, electron mass m, and initial speed υ_o.

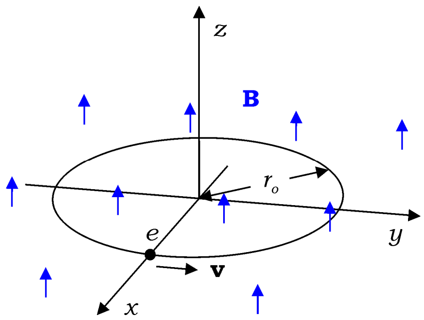


Fig. 5.1 An electron moving in a magnetic flux density.

Solution

(a) Magnetic force on the electron is

$$\mathbf{F}_m = e\,\mathbf{v} \times \mathbf{B} \tag{5-4}$$

B does not change $|\mathbf{v}|$. Therefore, $|\mathbf{F}_m|$ is constant at time $t \ge 0$ and $\mathbf{F}_m$ is always perpendicular to the direction of move of the electron. $\mathbf{F}_m$ behaves as a centripetal force, and thus the electron moves along the circumference of a circle.

(b) If the electron moves on a circle with a velocity $\mathbf{v} = \upsilon_o\, \mathbf{a}_\phi$, from Eq. (5-4) we write

$$\mathbf{F}_m = -|e|\upsilon_o B_o(\mathbf{a}_\phi \times \mathbf{a}_z) = -|e|\upsilon_o B_o\, \mathbf{a}_\rho \qquad (e < 0) \tag{5-5}$$

At steady state, $\mathbf{F}_m$ is balanced by ***mass*** x ***acceleration*** of the electron, which is expressed as

$$\begin{aligned} m\frac{d\mathbf{v}}{dt} = m\frac{d\mathbf{v}}{d\phi}\frac{d\phi}{dt} &= m\upsilon_o \frac{d\mathbf{a}_\phi}{d\phi}\frac{d\phi}{dt} \\ &= m\upsilon_o(-\mathbf{a}_\rho)\omega \end{aligned} \tag{5-6}$$

where we used $d\mathbf{a}_\phi / d\phi = -\mathbf{a}_\rho$ as in Eq. (1-68), and $\phi = \omega t$ (angular velocity multiplied by time).

Equating Eq. (5-5) with Eq. (5-6) we have

$$\omega = \frac{|e| B_o}{m} \qquad \text{[rad/s]} \qquad (5\text{-}7)$$

Inserting $v_o = r_o \omega$ into Eq. (5-7), the radius of the circular path is

$$r_o = \frac{m v_o}{|e| B_o} \qquad \text{[m]}.$$

Exercise 5.1
With reference to Fig. 5.1, find the electric field intensity **E** which, together with **B**, will cause the electron to move along a straight line for time $t \geq 0$.

Ans. $\mathbf{E} = -v_o B_o \, \mathbf{a}_x$.

5.2 The Biot-Savart Law

Earlier, we defined the magnetic flux density **B** as a sort of a force field. We now turn our attention to the magnetic field **H**, which is defined as a vector field induced by a steady current.

The Biot-Savart law is an experimental law, which states that the differential magnetic field intensity $d\mathbf{H}$ due to a steady current *I* flowing through a vector differential length $d\mathbf{l}'$ is

$$\boxed{d\mathbf{H} = \frac{I d\mathbf{l}' \times \mathbf{a}_{\mathcal{R}}}{4\pi \mathcal{R}^2}} \qquad \text{[A/m]} \qquad (5\text{-}8)$$

The differential field intensity $d\mathbf{H}$ is viewed as a vector located at the field point with position vector **r**, whereas the differential current element $I d\mathbf{l}'$ as a vector located at the source point with position vector $\mathbf{r}'$. The current element $I d\mathbf{l}'$ corresponds to a small segment of the conducting wire of negligible thickness carrying the steady current *I*. A conducting wire of negligible thickness is referred to as a filament throughout the text. In Eq. (5-8), $\mathcal{R}$ and $\mathbf{a}_{\mathcal{R}}$ are the magnitude and unit vector of the distance vector defined by $\mathcal{R} = \mathbf{r} - \mathbf{r}'$. The filamentary current *I* has the unit of the ampere[A], and the magnetic field intensity **H** has the unit of the ampere per meter[A/m].

In view of Eq. (5-8), we see that the cross product is conveniently used to specify the direction of $d\mathbf{H}$, which is normal to the plane formed by $I d\mathbf{l}'$ and $\mathbf{a}_{\mathcal{R}}$, and the sine of the angle between $d\mathbf{l}'$ and $\mathbf{a}_{\mathcal{R}}$, simultaneously. It is important to remember that the cross product $d\mathbf{l}' \times \mathbf{a}_{\mathcal{R}}$ should be treated as a vector located at the field point in unprimed coordinates(see Fig. 5.2), even though it may contain scalar components expressed in primed coordinates. We see from Eq. (5-8) that the magnetic field intensity is inversely proportional to the square of the distance between the current element and the field point, as **E** is inversely proportional to the square of the distance between the charge and the field point. However, in contrast to **E**, which is directed along the distance vector $\mathcal{R}$, the magnetic field vector $d\mathbf{H}$ is always normal to $\mathcal{R}$.

The total magnetic field intensity **H**, due to a steady current I flowing through a filament, is equal to the sum of the contributions from all the individual current elements comprising the filamentary current, i.e.,

$$\boxed{\mathbf{H} = \int_{C'} \frac{I d\mathbf{l}' \times \mathbf{a}_{\mathcal{R}}}{4\pi \mathcal{R}^2}} \qquad \text{[A/m]} \tag{5-9}$$

where the path of integration C' corresponds to the filament carrying the current I.

As, was discussed previously in Chapter 4, if a current is distributed over a volume $\mathcal{V}$, it can be described by the current density $\mathbf{J} = \rho_v \mathbf{v}$, which is the product of the volume charge density ρ_v and the velocity of flow of the charges, $\mathbf{v}$. Similarly, if the current is confined to a surface, it can be described by the surface current density $\mathbf{J}_s = \rho_s \mathbf{v}$, where ρ_s is the surface charge density. For these volume and surface currents, the equivalent current elements are defined as

$$\boxed{I d\mathbf{l}' = \mathbf{J}' dv' = \mathbf{J}_s' ds'} \tag{5-10}$$

The differential volume dv' or the differential area ds' is assumed to be centered at the source point with position vector $\mathbf{r}'$. The symbol "′" in Eq. (5-10) is to denote that the quantity is closely related to the current located at the source point. Inserting Eq. (5-10) into Eq. (5-9) allows us to write the Biot-Savart law as

$$\boxed{\mathbf{H} = \int_{\mathcal{V}'} \frac{\mathbf{J}' \times \mathbf{a}_{\mathcal{R}} dv'}{4\pi \mathcal{R}^2}} \qquad \text{[A/m]} \tag{5-11}$$

$$\boxed{\mathbf{H} = \int_{S'} \frac{\mathbf{J}_s' \times \mathbf{a}_{\mathcal{R}} ds'}{4\pi \mathcal{R}^2}} \qquad \text{[A/m]} \tag{5-12}$$

Again, $\mathcal{R}$ and $\mathbf{a}_{\mathcal{R}}$ are the magnitude and direction of the distance vector defined by $\mathcal{R} = \mathbf{r} - \mathbf{r}'$. It is important to note that even though $\mathbf{J}'$ and $\mathbf{a}_{\mathcal{R}}$ may be expressed in primed coordinates, the cross product $\mathbf{J}' \times \mathbf{a}_{\mathcal{R}}$ should be expanded by the base vectors at the field point.

Although the expression for the Biot-Savart law, as given in Eqs. (5-9), (5-11), and (5-12), employs a mixed coordinate system, the coordinate axes of the primed system coincide with those of the unprimed system; they are only referred to by different names. The purpose of the mixed system is to distinguish variables, or the coordinates of the source point, from constants, or the coordinates of the field point, in the integral for **H**.

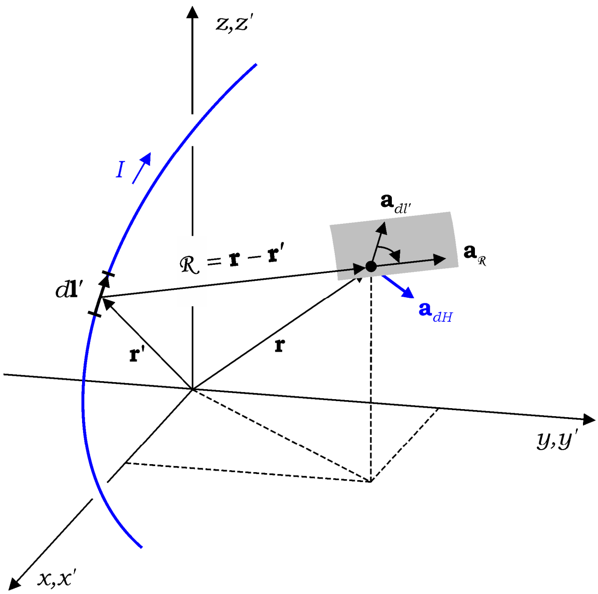


Fig. 5.2 A magnetic field $d\mathbf{H}$ at position vector $\mathbf{r}$ due to a current element $Id\mathbf{l}'$ at position vector $\mathbf{r}'$.

Example 5-2
Determine $\mathbf{H}$ due to an infinitely long straight filament lying along the z-axis and carrying a steady current I.

Solution

From symmetry considerations, we employ cylindrical coordinate system, and assume the field point to be on the $z = 0$ plane for simplicity in calculation, as shown in Fig. 5.3.

Using the relation $\mathbf{a}_z = \mathbf{a}_{z'}$, the distance vector is expanded by the base vectors at the field point, i.e.,

$$\begin{aligned}\mathcal{R} = \mathbf{r} - \mathbf{r}' &= \rho\,\mathbf{a}_\rho - z'\,\mathbf{a}_{z'} \\ &= \rho\,\mathbf{a}_\rho - z'\,\mathbf{a}_z\end{aligned}$$

Differential length vector along the filament is also expressed in terms of the base vectors at the field point as

$$d\mathbf{l}' = dz'\,\mathbf{a}_z$$

From Eq. (5-9) we obtain

$$\begin{aligned}\mathbf{H} &= \int_{z'=-\infty}^{z'=+\infty} \frac{I dz'\,\mathbf{a}_z \times (\rho\,\mathbf{a}_\rho - z'\,\mathbf{a}_z)}{4\pi\mathcal{R}^3} = \frac{I\,\mathbf{a}_\phi}{4\pi} \int_{z'=-\infty}^{z'=+\infty} \frac{\rho dz'}{(\rho^2 + z'^2)^{3/2}} \\ &= \mathbf{a}_\phi \frac{I\rho}{4\pi} \frac{z'/\rho^2}{\sqrt{\rho^2 + z'^2}}\Bigg|_{z'=-\infty}^{z'=+\infty}\end{aligned} \tag{5-13}$$

Here, $\mathbf{a}_\phi$ is taken outside the integral, because it is independent of z'. We also used the relations, $\mathbf{a}_z \times \mathbf{a}_\rho = \mathbf{a}_\phi$, $\mathbf{a}_z \times \mathbf{a}_z = 0$, and the integral $\int (x^2 + a^2)^{-3/2} dx = (x/a^2)(x^2 + a^2)^{-1/2} + C$.

The magnetic field intensity due to an infinitely long filament carrying a steady current I is therefore

$$\boxed{\mathbf{H} = \frac{I}{2\pi\rho}\mathbf{a}_\phi} \qquad \text{[A/m]} \qquad (5\text{-}14)$$

The direction of* H *and the direction of flow of* I *are related through the right-hand rule: the right thumb points in the direction of I when the fingers follow the direction of **H**.

Let us now examine the symmetry to see what it can tell us about **H**. The infinitely long filamentary current has cylindrical symmetry(no change after rotated about the z-axis) and translational symmetry in the z-direction(no change after displaced along the z-axis). Consequently, the resulting **H** should be independent of ϕ(cylindrical symmetry) and z(translational symmetry). We also note that the vector $d\mathbf{l}' \times \mathbf{a}_{\mathcal{R}}$ is always in the direction of $\mathbf{a}_\phi$ for the infinitely long filamentary current I. Therefore, the resultant **H** only has the ϕ-component that only depends on ρ, as is evident from Eq. (5-14).

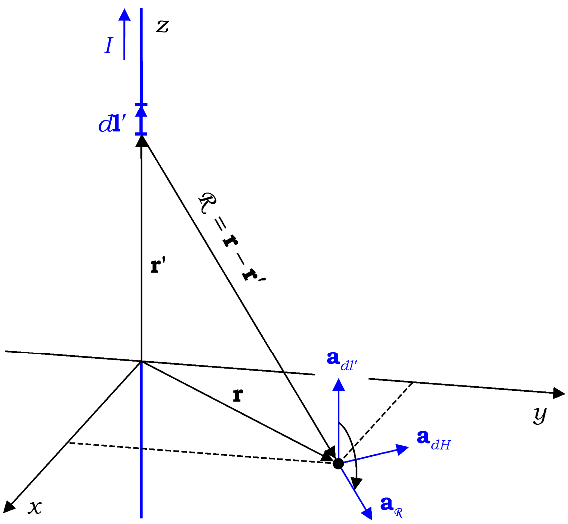


Fig. 5.3 An infinitely long straight filament carrying a steady current I.

Let us now digress briefly and derive a formula useful for finding **H** of a finite filamentary current extending between $z = z_1$ and $z = z_2$ $(z_1 < z_2)$, as shown in Fig. 5.4. From Eq. (5-13), at a point on the xy-plane, **H** is obtained as

$$\mathbf{H} = \mathbf{a}_\phi \frac{I}{4\pi\rho}\left(\frac{z'}{\sqrt{\rho^2 + z'^2}} \Bigg|_{z'=z_1}^{z'=z_2} \right) = \mathbf{a}_\phi \frac{I}{4\pi\rho}\left(\frac{z_2}{\sqrt{\rho^2 + z_2^2}} - \frac{z_1}{\sqrt{\rho^2 + z_1^2}} \right)$$

For the angles θ_1 and θ_2 as defined in Fig. 5.4, we express the magnetic field intensity at a distance ρ from a finite filamentary current *I* as

$$\boxed{\mathbf{H} = \frac{I}{4\pi\rho}\left[\cos\theta_1 + \cos\theta_2\right]\mathbf{a}_\phi} \qquad \text{[A/m]} \qquad (5\text{-}15)$$

Equation (5-15) always holds true for any z_1 and z_2. The direction of **H** is determined by the right-hand rule: When the right thumb points in the direction of *I*, the fingers rotate in the direction of **H**.

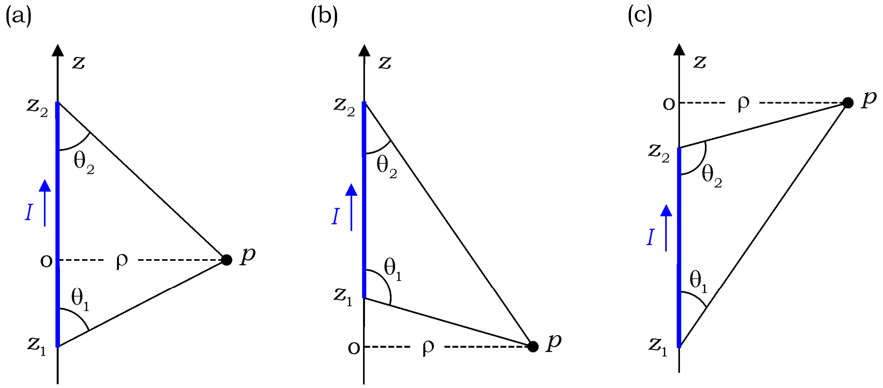


Fig. 5.4 **H** due to a filamentary current of a finite extent.

Example 5-3
Determine **H** at a field point $p{:}(a, b, 0)$ in Cartesian coordinates($a > 0,\ b > 0$), which is produced by the steady current *I* flowing in the filament bent into an L shape as shown in Fig. 5.5.

Solution
For the segment of the filament existing in the region $y < 0$, substituting $\theta_1 = 0$, $\theta_2 = \alpha$, and $\rho = a$ into Eq. (5-15) we obtain its magnetic field intensity as

$$\mathbf{H}_1 = \frac{I}{4\pi a}\left[1 + \cos\alpha\right](-\mathbf{a}_z)$$

For the segment of the filament existing in the region $x < 0$, substituting $\theta_1 = \beta$, $\theta_2 = 0$, and $\rho = b$ into Eq. (5-15) we obtain its magnetic field intensity as

$$\mathbf{H}_2 = \frac{I}{4\pi b}[\cos\beta + 1](-\mathbf{a}_z)$$

In the above equations, the direction, $-\mathbf{a}_z$, has been determined by the right-hand rule.

From Fig. 5.5 we obtain

$$\cos\alpha = -\sin(\alpha - 90^o) = -\frac{b}{\sqrt{a^2 + b^2}}$$

$$\cos\beta = -\sin(\beta - 90^o) = -\frac{a}{\sqrt{a^2 + b^2}}.$$

The magnetic field intensity at the field point p is therefore

$$\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2 = \frac{I}{4\pi a}\left[1 - \frac{b}{\sqrt{a^2+b^2}}\right](-\mathbf{a}_z) + \frac{I}{4\pi b}\left[-\frac{a}{\sqrt{a^2+b^2}} + 1\right](-\mathbf{a}_z)$$

$$= \mathbf{a}_z \frac{I}{4\pi ab}\left[\sqrt{a^2+b^2} - (a+b)\right]$$

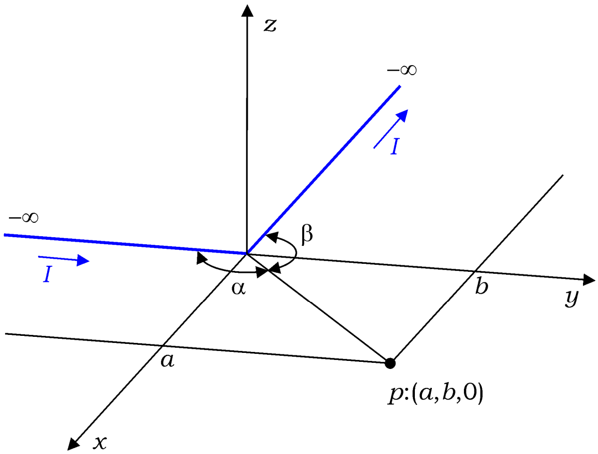


Fig. 5.5 An L shaped filament carrying a current I.

Example 5-4

A filament is formed into a circle of radius a in the xy-plane as shown in Fig. 5.6. Determine **H** at a field point p:(0, 0, b) due to a steady current I flowing in the filament in the counterclockwise direction.

Solution

In cylindrical coordinates, the differential length and the distance vectors are

$$d\mathbf{l}' = a\,d\phi'\,\mathbf{a}_{\phi'}$$

$$\mathcal{R} = \mathbf{r} - \mathbf{r}' = b\,\mathbf{a}_z - a\,\mathbf{a}_{\rho'}$$

Although two vectors $d\mathbf{l}'$ and $\mathcal{R}$ are expressed in terms of $\mathbf{a}_{\rho'}$ and $\mathbf{a}_{\phi'}$, we just insert them into Eq. (5-9), understanding that their cross product $d\mathbf{l}' \times \mathbf{a}_{\mathcal{R}}$ can be expressed later in terms of $\mathbf{a}_\rho$, $\mathbf{a}_\phi$, and $\mathbf{a}_z$, and obtain

$$\mathbf{H} = \int_{C'} \frac{I d\mathbf{l}' \times \mathbf{a}_{\mathcal{R}}}{4\pi \mathcal{R}^2} = \int_{\phi'=0}^{\phi'=2\pi} \frac{Ia\, d\phi' \mathbf{a}_{\phi'} \times (b\, \mathbf{a}_z - a\, \mathbf{a}_{\rho'})}{4\pi \mathcal{R}^3}$$

Separating the integral into two parts, we have

$$\mathbf{H} = \frac{Iab}{4\pi(a^2+b^2)^{3/2}} \int_{\phi'=0}^{\phi'=2\pi} \mathbf{a}_{\rho'} d\phi' + \frac{Ia^2 \mathbf{a}_z}{4\pi(a^2+b^2)^{3/2}} \int_{\phi'=0}^{\phi'=2\pi} d\phi'$$

The unit vector $\mathbf{a}_{\rho'}$ depends on ϕ', and the integral of $\mathbf{a}_{\rho'}$ with respect to ϕ', from 0 to 2π, is zero.

Thus, at a point on the z-axis, the magnetic field intensity is

$$\boxed{\mathbf{H} = \frac{Ia^2}{2(a^2+b^2)^{3/2}} \mathbf{a}_z} \qquad \text{[A/m]} \qquad (5\text{-}16)$$

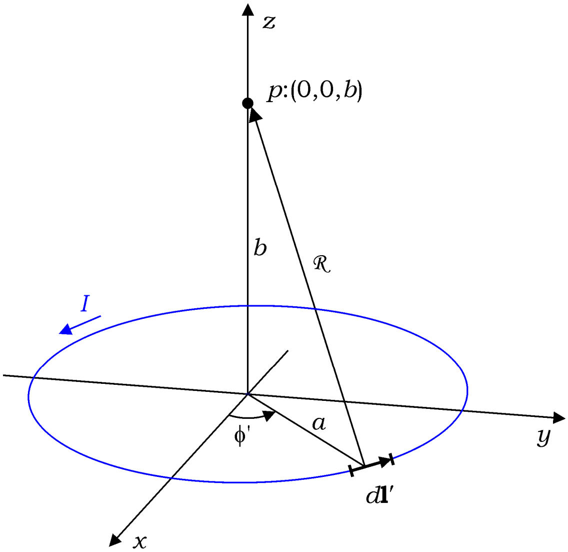


Fig. 5.6 A circular filament carrying a current I.

Exercise 5.2
A filament forms a square of side a centered at the origin in the xy-plane, with the sides perpendicular to the coordinate axes. Find $\mathbf{H}$ at the origin due to I flowing counterclockwise in the filament.

Ans. $\mathbf{H} = \mathbf{a}_z I 2\sqrt{2} / (\pi a)$.

Exercise 5.3

For a finite filament with current *I*, extending from $z = z_1$ to $z = z_2$, show that **H** is the sum of the contributions from the segments $\overline{z_1 z_o}$ and $\overline{z_o z_2}$ ($z_1 < z_o < z_2$).

Ans. $\cos\theta + \cos(\pi - \theta) = 0$.

Review Questions with Hints

RQ 5.1 State Lorentz force equation. [Eq.(5-3)]

RQ 5.2 Explain the relative orientation of the coordinate axes of the primed and unprimed systems used for the Biot-Savart law. [Fig.5.2]

RQ 5.3 What is the significance of the cross product used in the expression for the Biot-Savart law? [Fig.5.2]

RQ 5.4 In what way does **H** due to *I* flowing in an infinite filament vary with distance? [Eq.(5-14)]

RQ 5.5 Write the formula for **H** due to *I* flowing in a finite filament. [Eq.(5-15)]

5.3 Ampere's Circuital Law

Ampere's circuital law is analogous to Gauss's law for electricity in the sense that it allows us to solve the problem with certain symmetry by solving a simple algebraic equation. Ampere's circuital law is a special case of the Biot-Savart law; the former can be derived from the latter(see Section 5-5). As the point form of Gauss's law relates the divergence of **D** to a charge distribution, the point form of Ampere's circuital law relates the curl of **H** to a current distribution.

Ampere's circuital law states that ***the line integral of* H *around a closed path is equal to the current enclosed by that path.*** Ampere's circuital law is expressed as

$$\boxed{\oint_C \mathbf{H} \cdot d\mathbf{l} = I} \tag{5-17}$$

The direction of travel on C and the direction of flow of I are governed by the right-hand rule: the right thumb points in the direction of I when the fingers follow the positive direction on C.

With reference to Fig. 5.7, we can show that the *enclosed current I* in Eq. (5-17) is equal to the integral of the current density **J** over any surface bounded by the closed path C. From the equation of continuity in a static condition, that is, $\nabla \cdot \mathbf{J} = 0$, the integral of **J** over any closed surface S_o is zero. Expressed mathematically,

$$\oint_{S_o} \mathbf{J} \cdot d\mathbf{s} = 0 \tag{5-18}$$

Suppose the closed surface S_o forms a closed hemispherical shell as shown in Fig. 5.7(a). In view of Eq. (5-18), we see that the total current entering the hemisphere through the base plate is equal to the total current leaving the hemisphere through the spherical surface. If we remove the base plate from the hemisphere as shown in Fig. 5.7(b), the circular rim of the hemisphere forms a closed path C. As is evident in Fig. 5.7(b), the current enclosed by C is equal to the current flowing through the surface S bounded by C. Since S_o may be arbitrary, the surface S can also be arbitrary only if it is bounded by C.

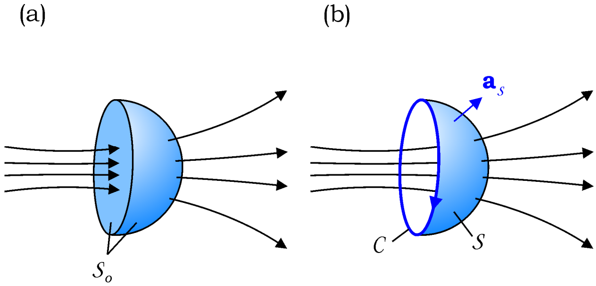


Fig. 5.7 The current enclosed by C equals the current flowing through S.

From the above discussion, we can express Ampere's circuital law in terms of the current density $\mathbf{J}$ as

$$\boxed{\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot d\mathbf{s}} \tag{5-19}$$

where S is the surface bounded by the closed path C. The direction of $d\mathbf{l}$ on C and the direction of $d\mathbf{s}$ on S are related by the right-hand rule: the right thumb points in the direction of $d\mathbf{s}$ when the fingers advance in the direction of $d\mathbf{l}$. Ampere's circuital law states that ***the line integral of* H *around a closed path is equal to the total current flowing through the surface bounded by that path.***

Applying Stokes's theorem to the left-hand side of Eq. (5-19) we have

$$\int_S \nabla \times \mathbf{H} \cdot d\mathbf{s} = \int_S \mathbf{J} \cdot d\mathbf{s} \tag{5-20}$$

Since S may be arbitrary in Eq. (5-20), the two integrands should be the same at every point on S in order to satisfy the equality. The point form of Ampere's circuital law is therefore

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J}} \tag{5-21}$$

Equation (5-21) is one of the two fundamental relations governing static magnetic fields. It should be noted that Ampere's circuital law, as given in Eqs. (5-17), (5-19), and (5-21), holds true only for static magnetic fields. Under time-varying conditions, Ampere's circuital law should be modified to account for the electromagnetic induction observed by Faraday, which we will discuss in Chapter 6.

Ampere's circuital law expressed by Eq. (5-19) is very useful for determining the magnetic field intensity **H**, only if ***the closed path** C **can be chosen in such a way that H at every point on** C **is constant and tangential to** C.* Such a path is called an Amperian path.

Example 5-5
An infinitely long filament is oriented along the z-axis, with a steady current I flowing in the $+z$-direction. Determine **H** by using Ampere's circuital law.

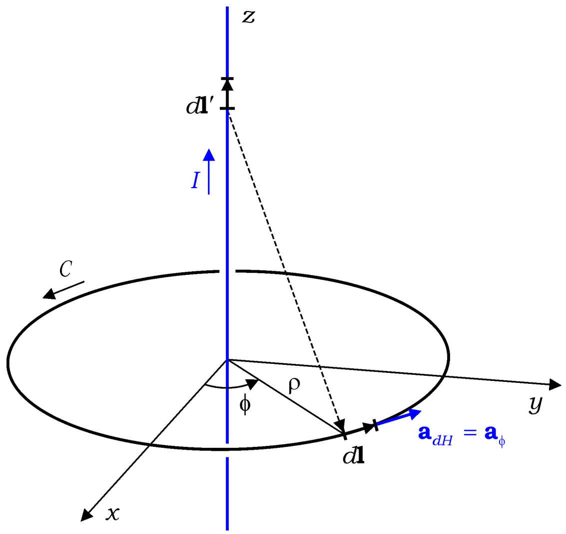


Fig. 5.8 An infinitely long filamentary current and an Amperian path.

Solution

As was stated previously in Example 5-2, an infinitely long filamentary current has cylindrical and translational symmetries. In addition, the Biot-Savart law, or the term $I d\mathbf{l}' \times \mathbf{a}_{\mathcal{R}}$, tells us that $d\mathbf{H}$ is always in the direction of $\mathbf{a}_\phi$. Consequently, the resulting **H** should be of the form $\mathbf{H} = H_\phi(\rho)\,\mathbf{a}_\phi$: **H** has the ϕ-component only, which depends on ρ only. From these discussions, we take a circle centered at the current as an Amperian path.

Applying Ampere's circuital law to the Amperian circle of radius ρ gives

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_{\phi=0}^{2\pi} H_\phi(\rho)\,\mathbf{a}_\phi \cdot \left(\rho\, d\phi\, \mathbf{a}_\phi\right) = H_\phi(\rho)\,2\pi\rho$$
$$= I$$

Thus, we have

$$\mathbf{H} = H_\phi(\rho)\,\mathbf{a}_\phi = \frac{I}{2\pi\rho}\,\mathbf{a}_\phi \tag{5-22}$$

The result is of course the same as Eq. (5-14) that was obtained by the Biot-Savart law.

Example 5-6
An infinitely long coaxial cable consists of an inner conductor of radius a, and an outer conductor of radii b and c, as shown in Fig. 5.9. It carries a steady current I, flowing in the opposite directions, uniformly distributed in the conductors. Determine **H** everywhere by using Ampere's circuital law.

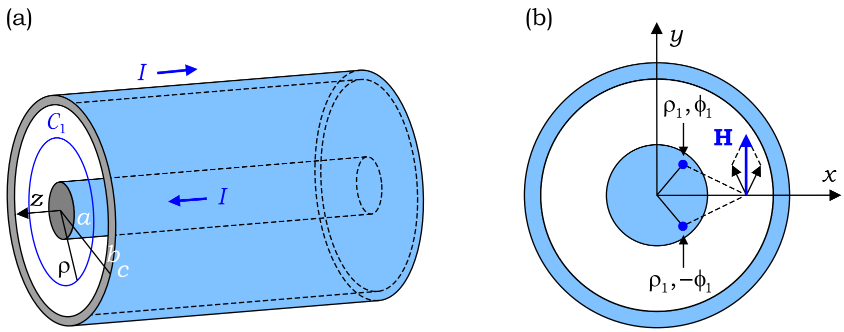


Fig. 5.9 A coaxial cable carrying a steady current I.

Solution
A close examination of the geometry reveals that the coaxial cable has cylindrical symmetry, and translational symmetry in the z-direction. Therefore, the resultant **H** is independent of ϕ and z. Next, suppose that the current in the inner conductor is composed of a large number of filamentary currents. Since **H** of a filamentary current is circumferential, a pair of filamentary currents located at (ρ_1, ϕ_1) and at $(\rho_1, -\phi_1)$, as shown in Fig. 5.9(b), jointly produce a magnetic field intensity directed only in the direction of $\mathbf{a}_\phi$. Thus, the resultant **H** should be of the form $\mathbf{H} = H_\phi(\rho)\,\mathbf{a}_\phi$ everywhere, even in the interior of the conductor.

In the region $0 < \rho \le a$, the closed line integral of **H** around a circle of radius ρ is

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_{\phi=0}^{2\pi} H_\phi(\rho)\,\mathbf{a}_\phi \cdot \left(\rho\, d\phi\, \mathbf{a}_\phi\right) = 2\pi\rho H_\phi(\rho) \tag{5-23}$$

The current enclosed by the circular path is

$$\frac{\pi\rho^2}{\pi a^2} I$$

Combining the two equations, we obtain

$$\mathbf{H} = H_\phi\,\mathbf{a}_\phi = \frac{I\rho}{2\pi a^2}\mathbf{a}_\phi \qquad (0 < \rho \le a) \tag{5-24a}$$

In the other regions, the closed line integral of $\mathbf{H}$ is the same as Eq. (5-23), except that the enclosed current is different in different regions such as

$$I \qquad (a \le \rho \le b)$$

$$I\frac{c^2-\rho^2}{c^2-b^2} \qquad (b \le \rho \le c)$$

$$0 \qquad (\rho \ge c)$$

Thus, the magnetic field intensity in the coaxial cable is

$$\mathbf{H} = \frac{I}{2\pi\rho}\mathbf{a}_\phi \qquad (a \le \rho \le b) \qquad \text{(5-24b)}$$

$$\mathbf{H} = \frac{I}{2\pi\rho}\frac{c^2-\rho^2}{c^2-b^2}\mathbf{a}_\phi \qquad (b \le \rho \le c) \qquad \text{(5-24c)}$$

$$\mathbf{H} = 0 \qquad (\rho \ge c) \qquad \text{(5-24d)}$$

A circular Amperian path around the coaxial cable encloses no net current, and thus $\mathbf{H}$ is zero outside the coaxial cable as given in Eq. (5-24d). In general, a zero circulation of $\mathbf{H}$ around a closed path does not necessarily mean $\mathbf{H}=0$ at every point on the path; $\oint_C \mathbf{H}\cdot d\mathbf{l} = 0$ in a uniform field $\mathbf{H}=\mathbf{a}_z$. In this problem, however, we are aware of the functional form of $\mathbf{H}$; namely, $\mathbf{H} = H_\phi(\rho)\,\mathbf{a}_\phi$. Thus, a zero circulation of $\mathbf{H}$ directly leads to $\mathbf{H}=0$ at every point on the path, outside the coaxial cable.

Example 5-7

A toroidal coil has N turns closely wound on an air core and carries a steady current I as shown in Fig. 5.10. The toroid has an inner radius a and an outer radius b. Find $\mathbf{H}$ everywhere.

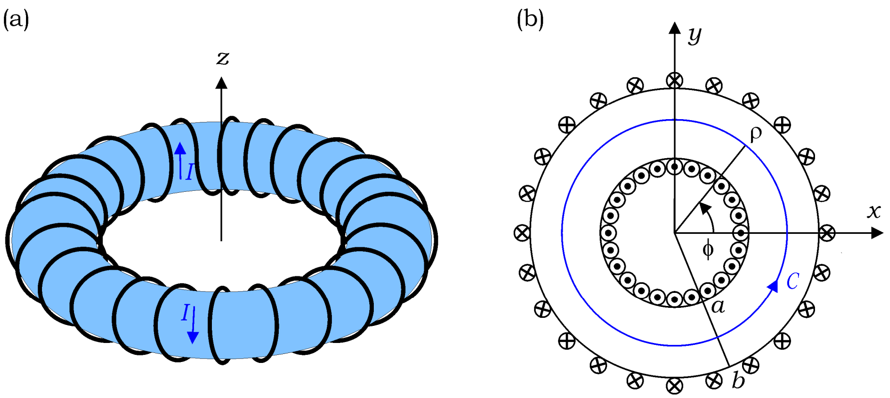


Fig. 5.10 A toroidal coil.

Solution

The closely wound toroidal coil has cylindrical symmetry(rotational symmetry about the z-axis), and thus the resulting **H** is independent of ϕ. If the toroidal coil is rotated about the x-axis by 180^o, it appears the same except for the reversed direction of the current in the coil. Note that if we reverse the current in the original coil, the direction of the resultant **H** should also be reversed. In view of this, we see that **H** cannot have the ρ-component, because the vector component $H_\rho \mathbf{a}_\rho$ appears the same even if it is rotated about the x-axis by 180^o. From these symmetry considerations so far, we arrive at a conclusion that **H** is of the form $\mathbf{H} = H_\phi(\rho, z)\,\mathbf{a}_\phi + H_z(\rho, z)\,\mathbf{a}_z$. Assuming an ideal toroidal coil in which no current penetrates the $\phi = \phi_1$ plane(ϕ_1 is constant), we see that the line integral of **H** around any closed path C_1, which lies flat on the $\phi = \phi_1$ plane, is zero, i.e.,

$$\oint_{C_1} \mathbf{H}\cdot d\mathbf{l} = \oint_{C_1} (H_\phi\,\mathbf{a}_\phi + H_z\,\mathbf{a}_z)\cdot(d\rho\,\mathbf{a}_\rho + dz\,\mathbf{a}_z) = \oint_{C_1} H_z(\rho, z)dz \\ = 0 \tag{5-25}$$

Here, we use $d\mathbf{l} = d\rho\,\mathbf{a}_\rho + dz\,\mathbf{a}_z$ in the $\phi = \phi_1$ plane. Suppose C_1 is a rectangle with the left side at $\rho = \rho_1$ and the right side at $\rho = \rho_2$, being parallel to the z-axis in the $\phi = \phi_1$ plane. Since Eq. (5-25) should be satisfied regardless of the height of C_1, we obtain the relation $H_z(\rho_1, z) = H_z(\rho_2, z)$ from Eq. (5-25), implying that H_z is constant everywhere. The finite extent of the toroidal coil, however, assures that $\mathbf{H} = 0$ at infinity, meaning that $H_z = 0$ everywhere. From the discussions so far, we conclude that **H** of the toroidal coil is of the form $\mathbf{H} = H_\phi(\rho, z)\,\mathbf{a}_\phi$.

Inside the toroidal coil($a < \rho < b$), taking a circle of radius ρ as an Amperian path C, from Ampere's circuital law we obtain

$$\oint_C \mathbf{H}\cdot d\mathbf{l} = \oint_C H_\phi \mathbf{a}_\phi \cdot \rho d\phi\,\mathbf{a}_\phi = \rho H_\phi \int_0^{2\pi} d\phi = 2\pi\rho H_\phi$$

Since C encloses the total current of NI, Ampere's circuital law gives

$$H_\phi = \frac{NI}{2\pi\rho}$$

Inside the toroidal coil, we have

$$\mathbf{H} = H_\phi \mathbf{a}_\phi = \frac{NI}{2\pi\rho}\mathbf{a}_\phi \qquad \text{(inside)} \tag{5-26a}$$

Outside the toroidal coil, no net current is enclosed by $\mathcal{C}$. Thus,

$$\mathbf{H} = 0 \qquad \text{(outside)} \qquad (5\text{-}26b)$$

We notice that the final magnetic field intensity given in Eq. (5-26a) is of the form $\mathbf{H} = H_\phi(\rho)\mathbf{a}_\phi$; it is not against our expectation of $\mathbf{H} = H_\phi(\rho, z)\mathbf{a}_\phi$.

Exercise 5.4
A thin conducting coating on a long dielectric cylinder of radius a carries a uniform surface current density $\mathbf{J}_s = J_o\mathbf{a}_z$ [A/m] . Find **H** everywhere.

Ans. $\mathbf{H} = 0$ for $\rho < a$, and $\mathbf{H} = \mathbf{a}_\phi J_o a / \rho$ for $\rho > a$.

Exercise 5.5
What would be the functional form of **H** inside the toroidal coil with a rectangular cross section?

Ans. $\mathbf{H} = H_\phi(\rho)\mathbf{a}_\phi$.

Review Questions with Hints

RQ 5.6 State Ampere's circuital law in words. [Eq.(5-17)]
RQ 5.7 What is meant by the current enclosed by a closed loop? [Fig.5.7]
RQ 5.8 What determines the positive direction of $\mathcal{C}$ in Eqs. (5-17) and (5-19)? [Fig.5.7]
RQ 5.9 How do you choose an Amperian path? [Fig.5.8]
RQ 5.10 Is it possible to apply Ampere's circuital law to a current element of a finite extent? [Eq.(5-15)]
RQ 5.11 Does zero circulation of **H** mean $\mathbf{H} = 0$ at every point on the loop? [Eq.(5-22), Fig.5.8]
RQ 5.12 Write the point form of Ampere's circuital law. [Eq.(5-21)]

5.4 Magnetic Flux Density

Earlier, the magnetic flux density **B** was defined in terms of the magnetic force exerted on a charge moving in **B**. As **D** is related to **E** by the constitutive relation $\mathbf{D} = \varepsilon_o\mathbf{E}$, in free space, **B** is related to **H** by the constitutive relation given, in free space, as

$$\boxed{\mathbf{B} = \mu_o\mathbf{H}} \qquad [\text{T}] \text{ or } [\text{Wb/m}^2] \qquad (5\text{-}27)$$

The magnetic flux density is measured in tesla[T] or webers per square meter $[\text{Wb/m}^2]$. The proportionality constant μ_o is called the permeability of free space, and has the value of

$$\mu_o = 4\pi \times 10^{-7} \qquad [\text{H/m}] \qquad (5\text{-}28)$$

which is measured in henries per meter.

The total magnetic flux passing through a surface S is equal to the integral of **B** over the surface, that is,

$$\Phi = \int_S \mathbf{B} \cdot d\mathbf{s} \qquad \text{[Wb]} \tag{5-29}$$

which is measured in webers. The dot product in Eq. (5-29) is necessary to convert the differential area $|d\mathbf{s}|$ in S into an equivalent area in the cross section of **B**. This is because $|\mathbf{B}|$ is originally defined as the magnetic flux passing through a unit area of the plane perpendicular to **B**.

We recall from electrostatics that the electric flux lines always begin at a positive charge and end at a negative one. Thus the net outward electric flux through a closed surface is not zero if the surface encloses a net charge. In contrast, the magnetic flux lines always close upon themselves because there are no isolated magnetic charges or no isolated magnetic poles. For instance, even if a permanent magnet is broken up into pieces on an atomic scale for the purpose of separating the north and south poles, each piece always has a north and a south pole.

From the Biot-Savart law given in Eq. (5-8), we acknowledge that a differential current element $Id\mathbf{l}'$ produces the magnetic field lines of the form of concentric circles. Applying the principle of superposition of the magnetic field leads us to conclude that the magnetic field lines should be of the form of closed lines regardless of the current distribution.

Gauss's law for magnetism states that ***the net outward magnetic flux through a closed surface is always zero***. Expressed mathematically,

$$\boxed{\oint_S \mathbf{B} \cdot d\mathbf{s} = 0} \tag{5-30}$$

Equation (5-30) can also be viewed as an expression for nonexistence of isolated magnetic charges, and an expression for the law of conservation of magnetic flux. Applying divergence theorem to Eq. (5-30), we obtain the point form of Gauss's law for magnetism as

$$\boxed{\nabla \cdot \mathbf{B} = 0} \tag{5-31}$$

It should be noted that Eq. (5-21) and Eq. (5-31) constitute two fundamental relations for static magnetic fields.

So far, we obtained the curl of **H** and the divergence of **B**, which are essential for a unique determination of the static magnetic field in a region of space according to the Helmholtz's theorem. To repeat, in Chapter 3, we obtained the curl of **E** and the divergence of **D**, which are essential for a unique determination of the static electric field in a region of space. For future reference, collecting the fundamental relations for static electric and static magnetic fields, we have

$$\boxed{\begin{aligned} &\nabla \times \mathbf{E} = 0 \\ &\nabla \cdot \mathbf{D} = \rho_v \\ &\nabla \times \mathbf{H} = \mathbf{J} \\ &\nabla \cdot \mathbf{B} = 0 \end{aligned}} \tag{5-32}$$

Applying the divergence and Stokes's theorems, the integral form of the fundamental relations is obtained as follows:

$$\boxed{\begin{aligned} &\oint_C \mathbf{E}\cdot d\mathbf{l} = 0 \\ &\oint_S \mathbf{D}\cdot d\mathbf{s} = \int_V \rho_v dv \\ &\oint_C \mathbf{H}\cdot d\mathbf{l} = \int_S \mathbf{J}\cdot d\mathbf{s} \\ &\oint_S \mathbf{B}\cdot d\mathbf{s} = 0 \end{aligned}} \tag{5-33}$$

Although there is no coupling between $\mathbf{E}$ and $\mathbf{H}$, time-varying electric and magnetic fields become coupled to each other such that a time-varying electric field generates a time-varying magnetic field, and vice versa. Under time-varying conditions, the two divergence equations in Eq. (5-32) remain valid, but the two curl equations need to be modified to account for Faraday's electromagnetic induction and the concept of displacement current density.

Example 5-8

A thin conducting shell of a radius a is connected to two very long filaments as shown in Fig. 5.11. The filamentary current I becomes a surface current, $\mathbf{J}_s = J_s\mathbf{a}_\theta$, which is uniform in the $\mathbf{a}_\phi$-direction on the sphere. Show

(a) $H_z = 0$ on the $z = 0$ plane,

(b) $\mathbf{H} = -\mathbf{a}_\phi \dfrac{I}{2\pi\rho}$ outside the sphere.

[Hint: Helmholtz's theorem.]

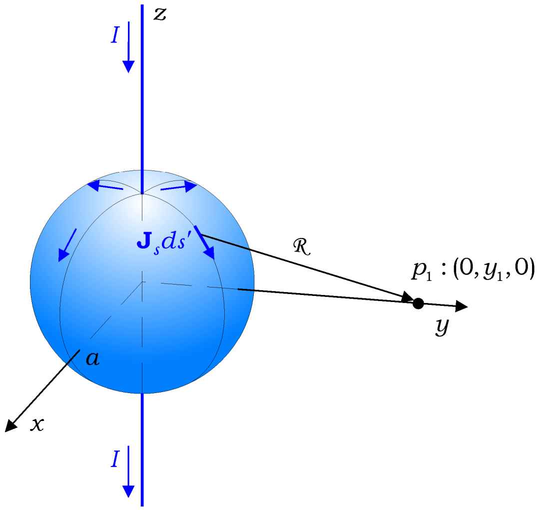


Fig. 5.11 A conducting spherical shell connected to two very long filaments.

Solution

(a) For the differential surface current $J_s ds' \mathbf{a}_{\theta'}$ given at the source point (a, θ', ϕ'), the differential field $d\mathbf{H}$ at the field point $p_1:(0, y_1, 0)$ is obtained from Eq. (5-12) as

$$d\mathbf{H} = \frac{J_s ds' \mathbf{a}_{\theta'} \times \mathbf{a}_{\mathcal{R}}}{4\pi \mathcal{R}^2} \tag{5-34}$$

Position vectors of the source and field points are

$$\mathbf{r}' = a \sin\theta' \cos\phi' \mathbf{a}_{x'} + a \sin\theta' \sin\phi' \mathbf{a}_{y'} + a \cos\theta' \mathbf{a}_{z'}$$

$$\mathbf{r} = y_1 \mathbf{a}_y$$

Distance vector and its magnitude are

$$\mathcal{R} = \mathbf{r} - \mathbf{r}' = \mathcal{R}\,\mathbf{a}_{\mathcal{R}} = -a \sin\theta' \cos\phi' \mathbf{a}_x + (y_1 - a \sin\theta' \sin\phi')\,\mathbf{a}_y - a \cos\theta' \mathbf{a}_z$$

$$\mathcal{R} = \left[a^2 \sin^2\theta' \cos^2\phi' + (y_1 - a\sin\theta' \sin\phi')^2 + a^2 \cos^2\theta'\right]^{1/2}$$

In the above equation $\mathcal{R}$ is expanded by $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$, because of $\mathbf{a}_{x'} = \mathbf{a}_x$, $\mathbf{a}_{y'} = \mathbf{a}_y$, and $\mathbf{a}_{z'} = \mathbf{a}_z$.

Applying the coordinate transformation $\mathbf{a}_{\theta'} = \cos\theta' \cos\phi' \mathbf{a}_x + \cos\theta' \sin\phi' \mathbf{a}_y - \sin\theta' \mathbf{a}_z$ to Eq. (5-34), we obtain

$$d\mathbf{H} = \frac{J_s ds'}{4\pi \mathcal{R}^3} \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \cos\theta' \cos\phi' & \cos\theta' \sin\phi' & -\sin\theta' \\ -a \sin\theta' \cos\phi' & (y_1 - a \sin\theta' \sin\phi') & -a\cos\theta' \end{vmatrix}$$

The z-component of $d\mathbf{H}$ is therefore

$$d\mathbf{H} \bullet \mathbf{a}_z = \frac{J_s ds'}{4\pi \mathcal{R}^3} y_1 \cos\theta' \cos\phi' \mathbf{a}_z$$

The two differential current elements at points (a, θ', ϕ') and $(a, \pi - \theta', \phi')$ jointly produce $d\mathbf{H} \bullet \mathbf{a}_z = 0$ at p_1 on the $z = 0$ plane. In addition, the filamentary current produces $\mathbf{H}$ in the direction of $\mathbf{a}_\phi$ only, as is evident from the Biot-Savart law. From these discussions and the cylindrical symmetry we conclude that $\mathbf{H} \bullet \mathbf{a}_z = 0$ at every point on the $z = 0$ plane outside the sphere. We may consider the $z = 0$ plane as a boundary on which the normal component of $\mathbf{H}$ is specified to be zero, $\mathbf{H} \bullet \mathbf{a}_z = 0$. This boundary condition is useful for part (b), which will be solved by applying Helmholtz's theorem that requires the normal component of the field to be specified at the boundary.

(b) In the region $z \geq 0$ outside the sphere, a trial solution $\mathbf{H} = -\mathbf{a}_\phi I/(2\pi\rho)$ satisfies not only Ampere's circuital law, but also the fundamental relations($\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{H} = 0$ in the source free region), and the boundary condition($H_z = 0$). Thus, it must be a unique solution in the given region according to Helmholtz's theorem. By the same token, the trial solution is also a unique solution in the region $z \leq 0$ outside the sphere.

Review Questions with Hints

RQ 5.13 State Gauss's law for magnetism in words. [Eq.(5-30)]
RQ 5.14 Write a mathematical expression for the statement that the magnetic flux line is always closed. [Eq.(5-30)]
RQ 5.15 Express the law of conservation of magnetic flux. [Eq.(5-30)]
RQ 5.16 What are two fundamental relations for static magnetic fields? [Eqs.(5-32)(5-33)]

5.5 Vector Magnetic Potential

In Chapter 3, we saw that the irrotational nature of **E**, $\nabla \times \mathbf{E} = 0$, leads to the electric potential *V* defined through the relation $\mathbf{E} = -\nabla V$. We also witnessed that the electric potential is much easier to handle than **E** because of its scalar nature. For instance, the electric potential is conveniently used for determining **E** by first calculating *V* from a given charge distribution and then taking the negative gradient of *V*. In the same way as the electric potential, we can define the scalar magnetic potential V_m through the relation $\mathbf{H} = -\nabla V_m$. However, the potential V_m is a many-valued function because of the non-conservative nature of **H**. As an example, the scalar magnetic potential is given as $V_m = -I\phi/(2\pi)$ in the presence of **H** that is produced by an infinitely long filamentary current *I*. In this case, the potential V_m has many values at a point in space as ϕ increases past 2π. The relation, $\nabla \times \mathbf{H} = \nabla \times (-\nabla V_m) = \mathbf{J}$, states that V_m is useful only in the region with $\mathbf{J} = 0$. In this section, we focus our attention on the vector magnetic potential **A**, which is useful for determining **B**, whether or not the given region contains a current distribution.

The solenoidal nature of **B** is mathematically expressed as $\nabla \cdot \mathbf{B} = 0$. By making use of the vector identity $\nabla \cdot (\nabla \times \mathbf{U}) = 0$, we can define the vector magnetic potential **A** as

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}} \tag{5-35}$$

The unit of **A** is the weber per meter[Wb/m]. If we find the vector potential **A** from a given current distribution, we can obtain **B** by taking the curl of **A**.

We can derive the relation between the vector magnetic potential **A** and the current density **J** from the Biot-Savart law. Rewriting the Biot-Savart law for **B**, we have

$$\mathbf{B} = \frac{\mu_o}{4\pi}\int_{C'} \frac{I\,d\mathbf{l}'\times\boldsymbol{\mathcal{R}}}{\mathcal{R}^3} \tag{5-36}$$

Borrowing from Section 2-2, we have

$$\nabla\frac{1}{\mathcal{R}} = -\frac{\boldsymbol{\mathcal{R}}}{\mathcal{R}^3} \tag{2-49}$$

where the distance vector $\boldsymbol{\mathcal{R}} = \mathbf{r} - \mathbf{r}'$. Inserting Eq. (2-49) into Eq. (5-36) we get

$$\mathbf{B} = -\frac{\mu_o I}{4\pi}\int_{C'} d\mathbf{l}'\times\left(\nabla\frac{1}{\mathcal{R}}\right)$$

Substituting $U = 1/\mathcal{R}$ and $\mathbf{L} = d\mathbf{l}'$ into the vector identity $\nabla\times(U\mathbf{L}) = (\nabla U)\times\mathbf{L} + U(\nabla\times\mathbf{L})$, and inserting the identity into the right-hand side of the above equation, we get

$$\mathbf{B} = \frac{\mu_o I}{4\pi}\int_{C'}\left[\nabla\times\left(\frac{d\mathbf{l}'}{\mathcal{R}}\right) - \frac{1}{\mathcal{R}}(\nabla\times d\mathbf{l}')\right] = \frac{\mu_o I}{4\pi}\int_{C'}\nabla\times\left(\frac{d\mathbf{l}'}{\mathcal{R}}\right) \tag{5-37}$$

In Eq. (5-37) we used $\nabla\times d\mathbf{l}' = 0$, which is true because ∇ acts on the unprimed coordinates only. On the right-hand side of Eq. (5-37), the curl operator can be taken outside the line integral because it is independent of the path C'. Thus Eq. (5-37) reduces to

$$\mathbf{B} = \nabla\times\int_{C'}\frac{\mu_o I\,d\mathbf{l}'}{4\pi\mathcal{R}} \tag{5-38}$$

Comparison of Eq. (5-38) with Eq. (5-35) gives an expression for **A**, i.e.,

$$\boxed{\mathbf{A} = \int_{C'}\frac{\mu_o I\,d\mathbf{l}'}{4\pi\mathcal{R}}} \qquad \text{[Wb/m]} \qquad \text{(line current)} \tag{5-39a}$$

where the line integral is performed along the filamentary current *I*. If the current is distributed over a volume $\mathcal{V}$ or across a surface $\mathcal{S}$, we can obtain the expression for **A** due to **J** or $\mathbf{J}_s$ simply by substituting the equivalent current element given in Eq. (5-10) into Eq. (5-39a):

$$\boxed{\mathbf{A} = \int_{\mathcal{V}'}\frac{\mu_o\mathbf{J}'}{4\pi\mathcal{R}}dv'} \qquad \text{[Wb/m]} \qquad \text{(volume current)} \tag{5-39b}$$

$$\boxed{\mathbf{A} = \int_{\mathcal{S}'}\frac{\mu_o\mathbf{J}'_s}{4\pi\mathcal{R}}ds'} \qquad \text{[Wb/m]} \qquad \text{(surface current)} \tag{5-39c}$$

where $\mathcal{R}$ is the magnitude of the distance vector $\boldsymbol{\mathcal{R}} = \mathbf{r} - \mathbf{r}'$.

We see that Eq. (5-39) is much simpler than the Biot-Savart law, since it does not involve the cross product. The vector potential **A** may be a many-valued function in space, because $\nabla \cdot \mathbf{A}$ is not yet specified. For instance, there is no change in **B** even if $\mathbf{A} + \nabla\varphi$ is used in place of **A** in the equation $\mathbf{B} = \nabla \times \mathbf{A}$. We find it convenient to set $\varphi = 0$, because of $\nabla \cdot \mathbf{A} = 0$, under static conditions, as can be shown in the next section.

The vector magnetic potential **A** does have physical significance: the closed line integral of **A** equals the magnetic flux passing through the surface bounded by the path of integration. The magnetic flux through a surface $\mathcal{S}$ is

$$\Phi = \int_{\mathcal{S}} \mathbf{B} \cdot d\mathbf{s}$$

Using the relation $\mathbf{B} = \nabla \times \mathbf{A}$ and applying Stokes's theorem, we have

$$\Phi = \int_{\mathcal{S}} (\nabla \times \mathbf{A}) \cdot d\mathbf{s} = \oint_{C} \mathbf{A} \cdot d\mathbf{l} \tag{5-40}$$

***The circulation of* A *around a closed loop* C *equals the magnetic flux enclosed by* C.**

5.5.1 Ampere's Circuital Law from the Biot-Savart Law

The point form of Ampere's circuital law can be derived from Biot-Savart law with the help of the vector magnetic potential. To start with, we rewrite the curl of **H**, in free space, by use of $\mathbf{B} = \mu_o \mathbf{H}$ and $\mathbf{B} = \nabla \times \mathbf{A}$, as

$$\nabla \times \mathbf{H} = \frac{1}{\mu_o} \nabla \times \mathbf{B} = \frac{1}{\mu_o} \nabla \times \nabla \times \mathbf{A} \tag{5-41}$$

Applying the vector identity $\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$ to Eq. (5-41), we have

$$\nabla \times \mathbf{H} = \frac{1}{\mu_o} \left[\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \right] \tag{5-42}$$

At this point we digress briefly and, from Eq. (5-39b), show that the divergence of **A** is always zero. By taking the divergence of both sides of Eq. (5-39c), we have

$$\nabla \cdot \mathbf{A} = \nabla \cdot \int_{\mathcal{V}'} \frac{\mu_o \mathbf{J}'}{4\pi \mathcal{R}} dv' = \frac{\mu_o}{4\pi} \int_{\mathcal{V}'} \nabla \cdot \left(\frac{\mathbf{J}'}{\mathcal{R}} \right) dv' \tag{5-43}$$

In the above equation the divergence operator is taken inside the volume integral because it is independent of $\mathcal{V}'$. The volume integral in Eq. (5-43) is zero, as can be shown by making use of the vector identities below:

$$\nabla \bullet \left(\frac{\mathbf{J}'}{\mathcal{R}} \right) = \left(\nabla \frac{1}{\mathcal{R}} \right) \bullet \mathbf{J}' + \frac{1}{\mathcal{R}} \nabla \bullet \mathbf{J}' = \left(\nabla \frac{1}{\mathcal{R}} \right) \bullet \mathbf{J}' \tag{5-44a}$$

$$\nabla' \bullet \left(\frac{\mathbf{J}'}{\mathcal{R}} \right) = \left(\nabla' \frac{1}{\mathcal{R}} \right) \bullet \mathbf{J}' + \frac{1}{\mathcal{R}} \nabla' \bullet \mathbf{J}' = \left(\nabla' \frac{1}{\mathcal{R}} \right) \bullet \mathbf{J}' \tag{5-44b}$$

In Eq. (5-44), we used $\nabla \bullet \mathbf{J}' = 0$, which is due to the fact that ∇ is independent of the primed coordinates, and $\nabla' \bullet \mathbf{J}' = 0$, which is the equation of continuity that steady currents should satisfy at all times. Using the identity $\nabla(1/\mathcal{R}) = -\nabla'(1/\mathcal{R})$, where $\mathcal{R} = \left[(x-x')^2 + (y-y')^2 + (z-z')^2\right]^{1/2}$, we can combine Eq. (5-44a) and Eq. (5-44b) such that

$$\boxed{\nabla \bullet \left(\frac{\mathbf{J}'}{\mathcal{R}} \right) = -\nabla' \bullet \left(\frac{\mathbf{J}'}{\mathcal{R}} \right)} \tag{5-45}$$

Inserting Eq. (5-45) into Eq. (5-43) and applying the divergence theorem, we have

$$\nabla \bullet \mathbf{A} = -\frac{\mu_o}{4\pi} \int_{\mathcal{V}'} \nabla' \bullet \left(\frac{\mathbf{J}'}{\mathcal{R}} \right) dv' = -\frac{\mu_o}{4\pi} \oint_{\mathcal{S}'} \frac{\mathbf{J}' \bullet d\mathbf{s}'}{\mathcal{R}} \tag{5-46}$$

The volume $\mathcal{V}'$ in Eq. (5-46) may be chosen in such a way that it includes all the current $\mathbf{J}'$ and an empty space surrounding $\mathbf{J}'$. In that case, no current passes through the surface $\mathcal{S}'$ bounding the volume $\mathcal{V}'$, and thus the closed surface integral on the right-hand side of Eq. (5-46) becomes zero:

$$\boxed{\nabla \bullet \mathbf{A} = 0} \tag{5-47}$$

We note that the vector magnetic potential **A** is solenoidal.

Next, inserting Eq. (5-47) into Eq. (5-42), we obtain

$$\boxed{\nabla \times \mathbf{H} = -\frac{1}{\mu_o} \nabla^2 \mathbf{A}} \tag{5-48}$$

The curl of **H** is linearly proportional to the Laplacian of **A**.

We now obtain the Laplacian of **A** in term of the source current **J**. To start with, note that the Laplacian of **A** is a vector with three vector components:

$$\nabla^2 \mathbf{A} = (\nabla^2 A_x)\,\mathbf{a}_x + (\nabla^2 A_y)\,\mathbf{a}_y + (\nabla^2 A_z)\,\mathbf{a}_z \tag{5-49}$$

Inserting **A** expressed by Eq. (5-39b) into Eq. (5-49), and collecting the terms with the unit vector $\mathbf{a}_x$, we have

$$\nabla^2 A_x = \nabla^2 \int_{\mathcal{V}'} \frac{\mu_o}{4\pi} \left(\frac{J'_x}{\mathcal{R}} \right) dv' = \frac{\mu_o}{4\pi} \int_{\mathcal{V}'} J'_x \left(\nabla^2 \frac{1}{\mathcal{R}} \right) dv' \tag{5-50}$$

where $\mathcal{R}$ is the magnitude of the distance vector $\boldsymbol{\mathcal{R}} = \mathbf{r} - \mathbf{r}'$. In Eq. (5-50), the Laplacian operator ∇^2 is taken inside the integral sign, because it is independent of the primed coordinates.

Let us focus our attention on a small region surrounding a field point, and image that the source point with position vector $\mathbf{r}'$ gradually moves toward the field point with position vector $\mathbf{r}$. Under these conditions, the term $\nabla^2(1/\mathcal{R})$ in Eq. (5-50) increases very rapidly compared with the source current J_x', and becomes infinite when $\mathbf{r}' = \mathbf{r}$. In view of these considerations, the volume integral in Eq. (5-50) can be performed over an infinitesimal volume Δ' surrounding the field point, with an infinitely small error. Since, however, $\nabla^2(1/\mathcal{R}) = \infty$ at $\mathbf{r}' = \mathbf{r}$, we may exclude the center point, a single point, from Δ' without affecting the volume integral. The new infinitesimal volume surrounding the field point is denoted by $\overline{\Delta}'$ as shown in Fig. 5.12. For the moment, $\overline{\Delta}'$ is a hollow sphere with an outer radius α and an inner radius δ, although δ shrinks to zero in the end. The outer radius α is such that it is short enough so that J_x' is assumed to be constant in the interior of $\overline{\Delta}'$, and yet long enough so that $\nabla^2(1/\mathcal{R}) \approx 0$ on the outer surface of $\overline{\Delta}'$ (see Fig. 5.12). Thus, we have $J_x'(\mathbf{r}') = J_x(\mathbf{r})$ at any point in the volume $\overline{\Delta}'$, and the source current J_x' can be taken outside the volume integral in Eq. (5-50) such that

$$\nabla^2 A_x = \frac{\mu_o J_x}{4\pi} \int_{\overline{\Delta}'} \left(\nabla^2 \frac{1}{\mathcal{R}} \right) dv' \tag{5-51a}$$

By making use of the identity $\nabla^2 U = \nabla \bullet \nabla U$ and the divergence theorem we write Eq. (5-51a) as

$$\nabla^2 A_x = \frac{\mu_o J_x}{4\pi} \int_{\overline{\Delta}'} \nabla \bullet \nabla \frac{1}{\mathcal{R}} dv' = \frac{\mu_o J_x}{4\pi} \lim_{\delta \to 0} \oint_{\mathcal{S}_1 + \mathcal{S}_2} \left(\nabla \frac{1}{\mathcal{R}} \right) \bullet d\mathbf{s}' \tag{5-51b}$$

where $\mathcal{S}_1 + \mathcal{S}_2$ is the bounding surface of $\overline{\Delta}'$. In Eq. (5-51b) we let the inner radius shrink to zero($\delta \to 0$), while the center point is excluded from the volume integral, which gives $\nabla^2(1/\mathcal{R}) = \infty$. Since we have $\nabla^2(1/\mathcal{R}) \approx 0$ on the outer surface $\mathcal{S}_2$, the integral on the right-hand side of Eq. (5-51b) reduces to a closed surface integral over the inner surface $\mathcal{S}_1$. Using the relation $\nabla(1/\mathcal{R}) = -(\mathbf{a}_{\mathcal{R}} / \mathcal{R}^2)$, and noting that the unit normal to surface $\mathcal{S}_1$ is along the direction of the distance vector $\mathbf{a}_{s'} = \mathbf{a}_{\mathcal{R}}$, we write Eq. (5-51b) as

$$\begin{aligned} \nabla^2 A_x &= -\frac{\mu_o J_x}{4\pi} \lim_{\delta \to 0} \oint_{S_1} \frac{\mathbf{a}_{\mathcal{R}}}{\mathcal{R}^2} \bullet d\mathbf{s}' = -\frac{\mu_o J_x}{4\pi} \lim_{\delta \to 0} \left[\frac{4\pi\delta^2}{\delta^2} \right] \\ &= -\mu_o J_x \end{aligned} \tag{5-52a}$$

where $4\pi\delta^2$ is the area of the surface $\mathcal{S}_1$ at $\mathcal{R} = \delta$. Following the same procedure as was used for A_x, we can obtain the Laplacian of A_y and A_z:

$$\nabla^2 A_y = -\mu_o \mathrm{J}_y \tag{5-52b}$$

$$\nabla^2 A_z = -\mu_o \mathrm{J}_z \tag{5-52c}$$

Combining the results in Eq. (5-52) leads to the vector Poisson's equation:

$$\boxed{\nabla^2 \mathbf{A} = -\mu_o \mathbf{J}} \tag{5-53}$$

Next, inserting Eq. (5-53) into Eq. (5-48), we obtain the point form of Ampere's circuital law, that is,

$$\nabla \times \mathbf{H} = \mathbf{J}$$

As was mentioned earlier, while the Biot-Savart law relates the current element $Id\mathbf{l}'$ at a source point to the differential field $d\mathbf{H}$ at a field point, the point form of Ampere's circuital law relates the curl of $\mathbf{H}$ at a point in space to the current density $\mathbf{J}$ at that point. The point form of Ampere's circuital law is therefore advantageous in describing the local effect in magnetostatics.

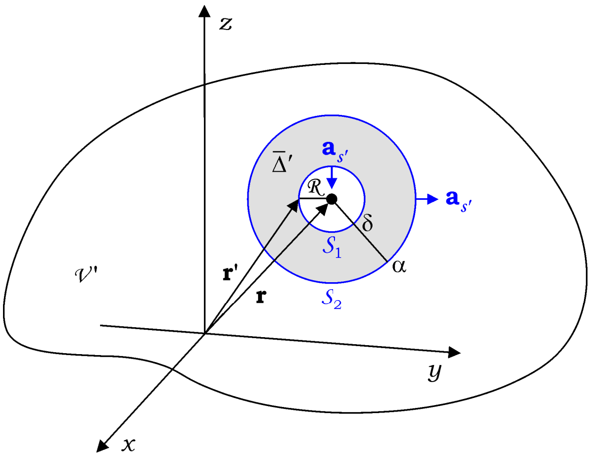


Fig. 5.12 An infinitesimal volume $\bar{\Delta}'$, bounded by surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$, excludes the field point at the center.

Example 5-9
An infinite sheet of current, with a uniform current density $\mathbf{J}_s = \mathrm{J}_o\, \mathbf{a}_y$, coincides with the $z = 0$ plane. Determine $\mathbf{H}$ on the z-axis by using vector magnetic potential.

Solution

From Eq. (5-39c), the vector potential $d\mathbf{A}$ at the position vector $\mathbf{r}$, due to the current element $\mathbf{J}_s ds'$ at the position vector $\mathbf{r}'$, is

$$d\mathbf{A} = \frac{\mu_o J_o \,\mathbf{a}_y}{4\pi\mathcal{R}} ds'$$

where $\mathcal{R} = |\mathbf{r} - \mathbf{r}'| = \left[(x - x')^2 + (y - y')^2 + z^2\right]^{1/2}$

Vector magnetic potential at a field point (x, y, z), in general, is

$$\mathbf{A} = \int_{x'=-\infty}^{\infty}\int_{y'=-\infty}^{\infty} \frac{\mu_o J_o \,\mathbf{a}_y}{4\pi\mathcal{R}} dx'dy'$$

Magnetic flux density is

$$\mathbf{B} = \nabla \times \mathbf{A} = \frac{\mu_o J_o}{4\pi}\int_{x'=-\infty}^{\infty}\int_{y'=-\infty}^{\infty} \nabla \times \frac{\mathbf{a}_y}{\mathcal{R}} dx'dy' \qquad (5\text{-}54)$$

The integrand in Eq. (5-54) is computed in Cartesian coordinates by use of $\mathcal{R} = |\mathbf{r} - \mathbf{r}'| = \left[(x - x')^2 + (y - y')^2 + z^2\right]^{1/2}$:

$$\nabla \times \frac{\mathbf{a}_y}{\mathcal{R}} = \begin{vmatrix} \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ 0 & \left[(x - x')^2 + (y - y')^2 + z^2\right]^{-1/2} & 0 \end{vmatrix}$$
$$= \left[z\,\mathbf{a}_x - (x - x')\,\mathbf{a}_z\right]\left[(x - x')^2 + (y - y')^2 + z^2\right]^{-3/2} \qquad (5\text{-}55)$$

Inserting Eq. (5-55) into Eq. (5-54), we have

$$\mathbf{B} = \frac{\mu_o J_o}{4\pi}\int_{x'=-\infty}^{\infty}\int_{y'=-\infty}^{\infty} \frac{\mathbf{a}_x z\, dx'dy'}{\left[(x - x')^2 + (y - y')^2 + z^2\right]^{3/2}}$$
$$- \frac{\mu_o J_o}{4\pi}\int_{x'=-\infty}^{\infty}\int_{y'=-\infty}^{\infty} \frac{\mathbf{a}_z (x - x')\, dx'dy'}{\left[(x - x')^2 + (y - y')^2 + z^2\right]^{3/2}} \qquad (5\text{-}56)$$

On the right-hand side of Eq. (5-56), the second integral vanishes because the integrand is an odd function of x'. In view of the limits of integration, $-\infty$ and ∞, we rewrite the first integral as

$$\mathbf{B} = \frac{\mu_o J_o}{4\pi}\int_{x'=-\infty}^{\infty}\int_{y'=-\infty}^{\infty} \frac{\mathbf{a}_x z\, dx'dy'}{\left[(x')^2 + (y')^2 + z^2\right]^{3/2}} \qquad (5\text{-}57)$$

Substituting $(x')^2 + (y')^2 = (\rho')^2$ and $dx'dy' = \rho' d\rho' d\phi'$ into Eq. (5-57), we evaluate the integral in cylindrical coordinates as

$$\mathbf{B} = \frac{\mu_o J_o\, z\, \mathbf{a}_x}{4\pi} \int_{\rho'=0}^{\infty} \int_{\phi'=0}^{2\pi} \frac{\rho' d\rho' d\phi'}{\left[(\rho')^2 + z^2\right]^{3/2}} = \frac{\mu_o J_o\, z\, \mathbf{a}_x}{4\pi} \left. \frac{-2\pi}{\left[(\rho')^2 + z^2\right]^{1/2}} \right|_{\rho'=0}^{\rho'=\infty}$$

$$= \frac{\mu_o J_o\, z\, \mathbf{a}_x}{2|z|}$$

Thus,

$$\mathbf{B} = \frac{\mu_o J_o\, \mathbf{a}_x}{2} \qquad (z > 0)$$

$$\mathbf{B} = -\frac{\mu_o J_o\, \mathbf{a}_x}{2} \qquad (z < 0)$$

The magnetic field intensity is

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_o} = \frac{J_o}{2}\mathbf{a}_x \qquad (z > 0) \tag{5-58a}$$

$$\mathbf{H} = -\frac{J_o}{2}\mathbf{a}_x \qquad (z < 0) \tag{5-58b}$$

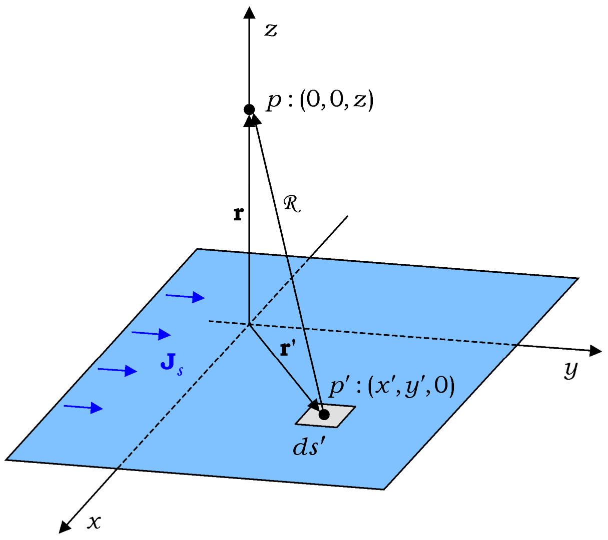


Fig. 5.13 An infinite sheet of a uniform surface current.

Exercise 5.6
A circular filament of radius a is centered at the origin in the $z = 0$ plane, carrying a steady current *I*. Find **A** at point $(0, 0, b)$.

Ans. 0.

Exercise 5.7
The surface current $\mathbf{J}_s$ in Fig. 5.13 has translational symmetries in the *x*- and *y*-directions. By using Eqs. (5-39c) and (5-53), find the functional form of **A**.

Ans. $\mathbf{A} = A_y(z)\,\mathbf{a}_y$ from Eq. (5-39c), $\mathbf{A} = (C_1 z + C_2)\,\mathbf{a}_y$ from Eq. (5-53).

Review Questions with Hints

RQ 5.17 Define vector magnetic potential. [Eq.(5-35)]
RQ 5.18 Express the relation between **A** and the source current. [Eq.(5-39)]
RQ 5.19 Express vector Poisson's equation. [Eq.(5-53)]
RQ 5.20 Which of the thee expressions in Eq. (5-39) is the solution of the vector Poisson's equation. [Eq.(5-39b)]
RQ 5.21 What are two fundamental relations for **A**. [Eqs.(5-36)(5-47)]

5.6 The Magnetic Dipole

A magnetic dipole is a small loop carrying a current; the diameter of the loop is much smaller than the distance between the loop and the field point. The magnetic dipole is the simplest model of an atom for magnetism. It provides a simple means of studying the magnetic behavior of material media. To explore the magnetic dipole, let us consider a circular filament of radius a centered at the origin in the $z = 0$ plane, carrying a steady current *I*, as shown in Fig. 5.14. Without loss generality, we assume a field point p to be on the yz-plane. Point p' is the source point at which a differential current element $I d\mathbf{l}'$ is located. From Eq. (5-39a), the vector magnetic potential at point p is

$$\mathbf{A} = \frac{\mu_o I}{4\pi} \oint_{C'} \frac{d\mathbf{l}'}{\mathcal{R}} \tag{5-59}$$

Employing a mixed coordinate system, we expand the vectors $\mathbf{r}$, $\mathbf{r}'$, and $\mathcal{R}$ by the base vectors in Cartesian coordinates, $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$, while expressing their scalar components in terms of spherical coordinates. Namely,

$$\mathbf{r} = R\sin\theta\,\mathbf{a}_y + R\cos\theta\,\mathbf{a}_z$$

$$\mathbf{r}' = a\cos\phi'\,\mathbf{a}_x + a\sin\phi'\,\mathbf{a}_y$$

$$\mathcal{R}^{-1} = |\mathbf{r} - \mathbf{r}'|^{-1} = \left[R^2 + a^2 - 2aR\sin\theta\sin\phi'\right]^{-1/2}$$

At far distances($R >> a$), by using the binomial expansion, the inverse distance $\mathcal{R}^{-1}$ can be approximated as

$$\begin{aligned}\mathcal{R}^{-1} &= \frac{1}{R}\left[1+\frac{a^2}{R^2}-\frac{2a}{R}\sin\theta\sin\phi'\right]^{-1/2} \\ &\cong \frac{1}{R}\left(1+\frac{a}{R}\sin\theta\sin\phi'\right)\end{aligned} \tag{5-60}$$

The higher order terms such as a^2 / R^2 are omitted in Eq. (5-60).

The differential length vector $d\mathbf{l}' = a\,d\phi'\,\mathbf{a}_{\phi'}$, located at the source point p' on the circular loop, is expressed in terms of the base vectors at the field point p as

$$d\mathbf{l}' = a\,d\phi'(\cos\phi'\sin\theta\,\mathbf{a}_R + \cos\phi'\cos\theta\,\mathbf{a}_\theta + \sin\phi'\,\mathbf{a}_\phi) \tag{5-61}$$

Substituting Eqs. (5-60) and (5-61) into Eq. (5-59), we obtain

$$\begin{aligned}\mathbf{A} &= \frac{\mu_o I}{4\pi}\int_{\phi'=0}^{2\pi}\frac{1}{R}\left(1+\frac{a}{R}\sin\theta\sin\phi'\right)a\,d\phi'\big(\cos\phi'\sin\theta\,\mathbf{a}_R \\ &\qquad\qquad + \cos\phi'\cos\theta\,\mathbf{a}_\theta + \sin\phi'\,\mathbf{a}_\phi\big) \\ &= \frac{\mu_o I}{4\pi}\frac{a}{R}\mathbf{a}_\phi\int_{\phi'=0}^{2\pi}\left(\frac{a}{R}\sin\theta\sin\phi'\right)\sin\phi'\,d\phi'\end{aligned}$$

Thus, we have

$$\boxed{\mathbf{A} = \frac{\mu_o I a^2}{4R^2}\sin\theta\,\mathbf{a}_\phi} \qquad [\mathrm{Wb/m}] \tag{5-62}$$

In vector notation, the vector magnetic potential due to the magnetic dipole with a magnetic dipole moment $\mathbf{m}$ is

$$\boxed{\mathbf{A} = \frac{\mu_o \mathbf{m}\times\mathbf{a}_R}{4\pi R^2}} \qquad [\mathrm{Wb/m}] \tag{5-63}$$

where the magnetic dipole moment is defined as

$$\boxed{\mathbf{m} = I\pi a^2\mathbf{a}_z = IS\,\mathbf{a}_z} \qquad [\mathrm{A\cdot m^2}] \tag{5-64}$$

The magnetic dipole moment is a vector whose magnitude is the product of the loop area and the current flowing in the loop, and whose unit vector is normal to the loop surface, obeying the right-hand rule: the right thumb points in the direction of $\mathbf{m}$ when the fingers follow the direction of I in the loop.

By taking the curl of **A** in spherical coordinates, we obtain the magnetic flux density **B** as

$$\boxed{\mathbf{B} = \frac{\mu_o m}{4\pi R^3}\left(2\cos\theta\,\mathbf{a}_R + \sin\theta\,\mathbf{a}_\theta\right)} \qquad [\mathrm{T}] \qquad (5\text{-}65)$$

We see from (5-65) that **B** of a magnetic dipole is analogous to **E** of an electric dipole. The expression for **B** in Eq. (5-65) will be identical to that of **E** in Eq. (3-84), if we replace m and μ_o in Eq. (5-65) with p and $1/\varepsilon_o$, respectively. Both Eqs. (3-84) and (5-65) represent the far-field patterns under the conditions of $R >> d$ and $R >> a$.

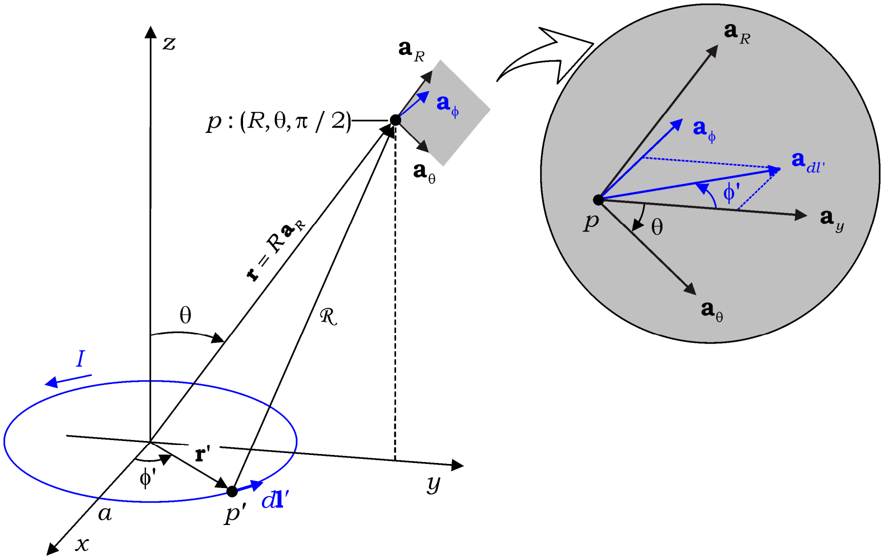


Fig. 5.14 A magnetic dipole.

The electric field lines of an electric dipole and the magnetic field lines of a magnetic dipole are obtained by numerical methods, without relying on the binomial expansion, and drawn to scale in Fig. 5.15.

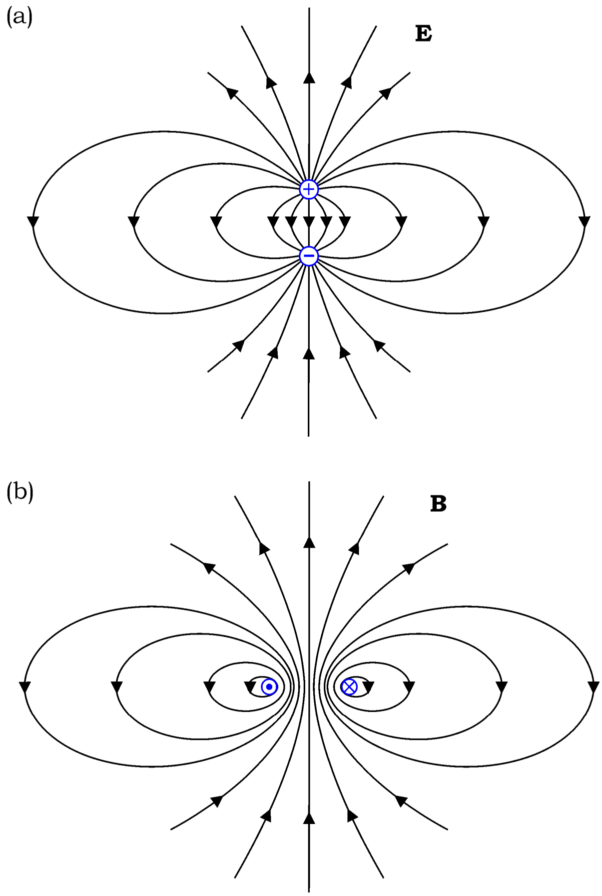


Fig. 5.15 The near field patterns of (a) an electric dipole (b) a magnetic dipole.

Exercise 5.8
Find the differential equation for the magnetic flux lines, in the yz-plane, at far distances from the origin where the dipole is located[Hint: Eq. (5-65)].

Ans. $\dfrac{dz}{dy} = \dfrac{B_z}{B_y} = \dfrac{2\cos^2\theta - \sin^2\theta}{3\sin\theta\cos\theta}$, where $\theta = \tan^{-1}\left(\dfrac{y}{z}\right)$.

Exercise 5.9
For a magnetic dipole moment $\mathbf{m} = m\,\mathbf{a}_z$ located at the origin, find the point on a sphere, centered at the origin, at which $|\mathbf{B}|$ is maximum.

Ans. The intersection of the sphere and the z-axis.

Review Questions with Hints

RQ 5.22 Define magnetic dipole and magnetic dipole moment. [Fig.5.14, Eq.(5-64)]

RQ 5.23 Can you distinguish between far field patterns of an electric and a magnetic dipole? [Eqs.(3-84)(5-65)]

RQ 5.24 Sketch near field patterns of electric and magnetic dipoles. [Fig.5.15]

5.7 Magnetic Materials

The simplest atomic model comprises a positively charged nucleus and negatively charged electrons orbiting around the nucleus. The orbiting electrons induce a magnetic dipole moment of the atom. The electron spin and nuclear spin are other sources of the magnetic dipole moment, which are rooted in the quantum theory. The effect of the nuclear spin is however three orders of magnitude smaller than those of the orbiting electrons and the electron spin, and is therefore neglected in most cases.

The magnetic property of a material is determined by the interaction between the lattice atoms and an externally applied magnetic field. In the absence of an external magnetic field in a diamagnetic material, the magnetic dipole moment due to the orbiting electrons cancels that due to the electron spin, and thus the lattice atom exhibits no net dipole moment. An externally applied magnetic field, however, exerts either a centrifugal or a centripetal force on the orbiting electrons. Since the electron orbits are quantized and thus cannot be changed, the electron velocity is reduced to counterbalance the magnetic force exerted on the orbiting electrons. The external field causes an imbalance between the dipole moments due to the orbiting electrons and the electron spin, and results in a net dipole moment in the diamagnetic material, which is in the opposite direction to the external magnetic field. Diamagnetic materials, such as gold, silver, copper, lead, and silicon, have a negative magnetic susceptibility of the order of -10^{-5}, and thus a relative permeability slightly less than one. Superconductors are perfect diamagnetic materials with a susceptibility $\chi_m = -1$ and a relative permeability $\mu_r = 0$. No magnetic flux density is allowed inside a superconductor($\mathbf{B} = 0$). This implies that a superconductor completely expels the magnetic field, and is thus levitated in the magnetic field. All materials exhibit diamagnetism, although the diamagnetic effect may be overwhelmed by other stronger magnetic effects in some materials.

When the magnetic dipole moments are caused by the orbiting electrons and the electron spin, they do not cancel completely in a paramagnetic material. Whereas the dipole moment is induced only by an externally applied magnetic field in a diamagnetic material, the atoms in a paramagnetic material have a permanent dipole moment even in the absence of an external magnetic field. Under normal conditions, the permanent dipole moments are randomly oriented in the paramagnetic material, yielding no net dipole moment. In the presence of an externally applied magnetic field, the permanent dipole moments align with the external field and result in a net dipole moment in the material, which is turn increase the total magnetic field in the interior of the material. Paramagnetic materials, such as air, aluminum, platinum, titanium, and tungsten, have a magnetic susceptibility of the order of 10^{-5} and a relative permeability slightly larger than one. For most practical purposes, we may assume $\mu = \mu_o$ for diamagnetic and paramagnetic materials, and regard them as nonmagnetic materials.

The atoms in a ferromagnetic material have a large permanent dipole moment because of the dominant electron spin. In the absence of an external magnetic field in a ferromagnetic material, the electron spins are coupled together by strong interatomic forces, and aligned in the same direction in microscopic regions of the material, called the magnetic domains. In view of a complete alignment of the dipole moments in a domain, the magnetic domain is said to be fully magnetized. Under normal conditions, the domains of different spin orientations are randomly distributed in the material yielding no net dipole moment. However, in the presence of an external magnetic field, the domains with a spin orientation parallel to the external field grow at the expense of the other domains, and result in a large net dipole moment in the material. Ferromagnetic materials, such as iron, cobalt, and nickel, have a very large μ_r ranging from 250 to 5000, and their alloys, such as permalloy and mumetal, have a relative permeability as large as 10^5. Unlike the diamagnetic and paramagnetic materials, the ferromagnetic material is a nonlinear material in that μ_r depends not only on the magnetic field intensity, but also on the course of the previous magnetic states of the material.

5.7.1 Magnetization and Equivalent Current Densities

According to the atomic model of a material, a magnetized material can be regarded as an aggregate of discrete magnetic dipole moments positioned at the equivalent lattice points in free space. For the macroscopic magnetic property of a material, we define the magnetization **M** as the magnetic dipole moments per unit volume, i.e.,

$$\mathbf{M} = \lim_{\Delta v \to 0} \frac{1}{\Delta v} \sum_{j=1}^{n\Delta v} \mathbf{m}_j \qquad \text{[A/m]} \tag{5-66}$$

where n is the number density of the lattice atoms, or the magnetic dipoles, and $\mathbf{m}_j$ represents the dipole moment vector within the incremental volume Δv. We note that **M** is a continuous function of position, whereas $\mathbf{m}_j$ is a discrete vector located at a point in space. The magnetization **M** is measured in amperes per meter. A magnetic material with a nonzero **M** is said to be magnetized.

The magnetization **M** induces surface currents on the material. To investigate the magnetization surface current, we consider an interface between a magnetized material, in the region $z \le 0$, and free space(or air with little error), in the region $z > 0$, as shown in Fig. 5.16. For simplicity, the magnetic dipole is represented by a square loop of side a, carrying a current I. Note that the center of the loop is marked with a dot in the figure. The current loop is oriented in such a way that the direction of its dipole moment $\mathbf{m}$, or the unit vector $\mathbf{a}_m$, is at an angle θ to the outward unit normal to the interface, $\mathbf{a}_n$, which is along the direction of $\mathbf{a}_z$. It is assumed that the material is fully magnetized to a magnetization $\mathbf{M}$ by an external magnetic field. Accordingly, the directions of the external field, $\mathbf{M}$, and $\mathbf{m}$ are all parallel to the unit vector $\mathbf{a}_m$. For the sake of argument, we slice the material into hypothetical layers of thickness $(a/2)\sin\theta$. A side view of the material is given in Fig. 5.16(a), in which a small circle with a dot(or a cross) indicates the direction of the current flowing in the upper(or lower) side of the square loop. We see in Fig. 5.16(a) that the dipoles with the center within layer 1 contribute to the net current flowing in the air, flowing out of the paper. Similarly, the dipoles with the center within layer 2 contribute to the net current flowing in layer 1, also flowing out of the paper. In contrast, the dipoles with the center within layer 1 or layer 3 jointly produce zero net current in layer 2. The net currents flowing in the air and in layer 1 constitute the surface current induced by the magnetization.

We now compute the surface current density induced on the surface of the material, referring to Fig. 5.16(b), in which the front view of the material is given. In the figure, the tilt of the magnetic dipole moment $\mathbf{m}$, with respect to the surface normal $\mathbf{a}_n$, causes the square loop to appear as a rectangle of $a \times a\sin\theta$. As was stated earlier, only the dipoles with the center within layer 1 or layer 2 can contribute to the surface current. To be specific, a single magnetic dipole contributes to a filamentary current of a length a on the surface, which is done by the upper side of the square loop. Thus only the dipoles with the center within the shaded region shown in Fig. 5.16(b) can contribute to the surface current flowing through the $x = 0$ plane, which is taken as a reference plane. In view of these considerations, the surface current density is defined as the current flowing through the reference plane per unit length of the y-axis. The magnitude of the surface current density is therefore

$$J_{ms} = In(a^2 \sin\theta) = n|\mathbf{m}|\sin\theta$$

where n is the number density of the dipoles, and the term in parenthesis is the area of the shaded region shown in Fig. 5.16(b). Upon using the relation $\mathbf{M} = n|\mathbf{m}|$, and noting that the surface current flows in the x-direction, the magnetization surface current density $\mathbf{J}_{ms}$ is defined as

$$\boxed{\mathbf{J}_{ms} = \mathbf{M} \times \mathbf{a}_n} \qquad \text{[A/m]} \qquad (5\text{-}67)$$

where $\mathbf{a}_n$ is the unit normal to the surface pointing out of the material. The unit of $\mathbf{J}_{ms}$ is the Ampere per meter.

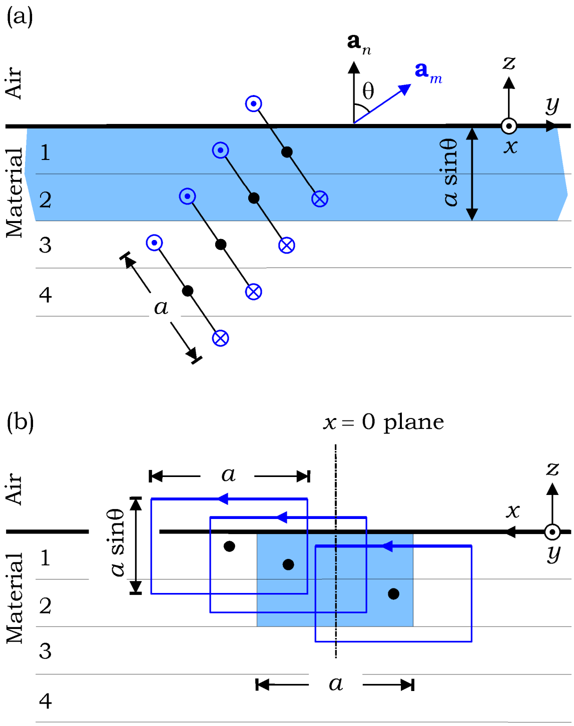


Fig. 5.16 Magnetization surface current. A magnetic dipole is represented by a square current loop. (a) Side view (b) Front view.

The magnetization $\mathbf{M}$ may also induce the magnetization current density $\mathbf{J}_m$ inside the material, which is measured in units of Amperes per square meter. Consider Fig. 5.17, in which the magnetization surface current density $\mathbf{J}_{ms}$ is induced on the surface of a material with the magnetization $\mathbf{M}$, filling the region $z \leq 0$. In accordance with the equation of continuity, if the net surface current crossing the circumference of a closed loop C, which lies flat on the surface of the material, is nonzero, there must be a net current flowing from the interior of the material to the surface. This current is referred to as the magnetization current. The magnetization current and the magnetization surface current are both rooted in the current of the magnetic dipole. The current in a square loop shown in Fig. 5.17 behaves as the magnetization current inside the material, and, at the same time, as the magnetization surface current on the surface of the material.

With reference to Fig. 5.17, the differential surface current dI_{ms} passing through the differential length $|d\mathbf{l}|$ on the circular loop C is written in terms of the magnetization surface current density $\mathbf{J}_{ms}$ as follows:

$$dI_{ms} = \mathbf{J}_{ms} \cdot (|d\mathbf{l}|\,\mathbf{a}_\rho) = (\mathbf{M} \times \mathbf{a}_n) \cdot \mathbf{a}_\rho\, |d\mathbf{l}|$$

where $\mathbf{a}_\rho$ represents a unit vector normal to $d\mathbf{l}$, or the loop, and we use Eq. (5-67) for $\mathbf{J}_{ms}$. Upon applying the vector identity $(\mathbf{A}\times\mathbf{B})\bullet\mathbf{C} = \mathbf{A}\bullet(\mathbf{B}\times\mathbf{C})$, the differential surface current crossing $d\mathbf{l}$ becomes

$$dI_{ms} = \mathbf{M}\bullet(\mathbf{a}_n \times \mathbf{a}_\rho)|d\mathbf{l}| = \mathbf{M}\bullet d\mathbf{l}$$

In the above equation, the relation, $\mathbf{a}_n \times \mathbf{a}_\rho = \mathbf{a}_\phi$, is used to represent the direction of $d\mathbf{l}$. The net surface current flowing out of the loop C, crossing the circumference of C on the surface of the material, is therefore

$$I_{ms} = \oint_C \mathbf{M}\bullet d\mathbf{l} = \int_S (\nabla\times\mathbf{M})\bullet d\mathbf{s} \qquad \text{[A]} \qquad (5\text{-}68)$$

where S is the surface of the material that is enclosed by C. Notice that Stokes's theorem is used in Eq. (5-68). Next, from the equation of continuity, the net surface current I_{ms} in Eq. (5-68) should be equal to the net current passing through the loop surface S, which is given by the surface integral of the current density $\mathbf{J}_m$, called the magnetization current density. That is,

$$I_{ms} = \int_S \mathbf{J}_m\bullet d\mathbf{s} \qquad \text{[A]} \qquad (5\text{-}69)$$

Since, in Eqs. (5-68) and (5-69), the closed loop C and the enclosed surface S may be arbitrary, the two integrands of the two surface integrals should be the same at every point on the surface of the material. Thus,

$$\boxed{\mathbf{J}_m = \nabla\times\mathbf{M}} \qquad [\text{A/m}^2] \qquad (5\text{-}70)$$

As we can see from Eqs. (5-67) and (5-70), the magnetization $\mathbf{M}$ of a magnetized material induces the magnetization current density $\mathbf{J}_m$ in the material, and the magnetization surface current density $\mathbf{J}_{ms}$ on the surface of the material. Although the current $\mathbf{J}_m$ may be zero in the material, because of a constant $\mathbf{M}$, the surface current $\mathbf{J}_{ms}$ is always induced on the surface of the material.

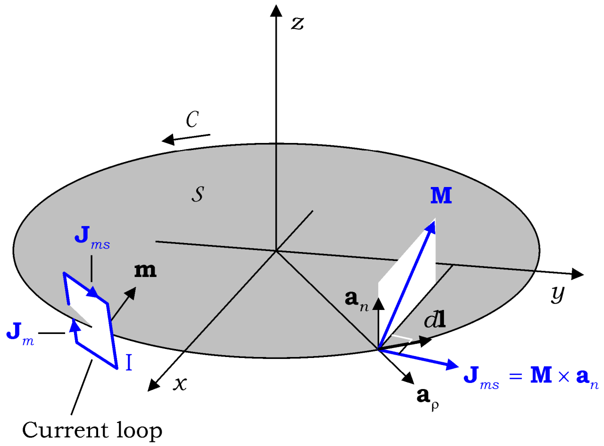


Fig. 5.17 Magnetization current and magnetization surface current.

Example 5-10
A permanent magnet forms a circular cylinder of radius a and height b, having a uniform magnetization $\mathbf{M} = M_o \mathbf{a}_z$. Find $\mathbf{J}_m$ and $\mathbf{J}_{ms}$.

Solution

From Eq. (5-70), we have

$$\mathbf{J}_m = \nabla \times (M_o \mathbf{a}_z) = 0$$

From Eq. (5-67), we have

$$\begin{aligned}\mathbf{J}_{ms} &= \mathbf{M} \times \mathbf{a}_n \\ &= M_o \mathbf{a}_z \times \mathbf{a}_z = 0\end{aligned} \qquad \text{(top surface)}$$

$$\mathbf{J}_{ms} = M_o \mathbf{a}_z \times (-\mathbf{a}_z) = 0 \qquad \text{(bottom surface)}$$

$$\mathbf{J}_{ms} = M_o \mathbf{a}_z \times \mathbf{a}_\rho = M_o \mathbf{a}_\phi \qquad \text{(side surface)}$$

From the above results, we note that the permanent magnet is equivalent to a cylindrical conducting sheet of radius a and height b carrying a surface current density $\mathbf{J}_s = M_o \mathbf{a}_\phi$.

Exercise 5.10
For a sphere of radius a formed by the magnetized material with a magnetization $\mathbf{M} = M_o \mathbf{a}_z$, find $\mathbf{J}_{ms}$ on the surface of the sphere.

Ans. $M_o \sin\theta \, \mathbf{a}_\phi$ [A/m].

Review Questions with Hints

RQ 5.25 Define magnetization **M**. [Eq. (5-66)]
RQ 5.26 Relate $\mathbf{J}_{ms}$ and $\mathbf{J}_m$ to the magnetization **M**. [Eqs.(5-67)(5-70)]
RQ 5.27 What are the units of $\mathbf{J}_{ms}$ and $\mathbf{J}_m$? [Eqs.(5-67)(5-70)]
RQ 5.28 Does $\mathbf{J}_m = 0$ at a point in a magnetized material mean $\mathbf{M} = 0$ at that point? [Eq.(5-70)]

5.7.2 Permeability

In the study of electrostatics in Chapter 3 we discussed the interaction between a dielectric and an externally applied electric field. When an electric polarization is induced in the material by the external field, it gives rise to an electric field called a polarization field. We saw that the sum of the external and the polarization fields constitutes the internal electric field in the material. We redefined the electric flux density **D** inside the material such that $\nabla \bullet \mathbf{D}$ is related to the net charge only, which is responsible for the external electric field. We also learned that the relation between **D** and **E** determines the permittivity of the material.

The magnetic flux density **B** in a magnetic material is analogous to **E** in a dielectric. When an externally applied magnetic field induces the magnetization **M**, and thus the magnetization current density $\mathbf{J}_m$ in the material, the sum of the external magnetic flux and that due to $\mathbf{J}_m$ constitutes the internal magnetic flux density **B** in the material. We redefine **H** in the material such that it is related to the free current **J**, excluding $\mathbf{J}_m$ induced in the material, which is responsible for the external field. In a magnetic material, the internal **B** is the sum of the magnetic fluxes produced by the free current and the magnetization current, i.e.,

$$\begin{aligned}\nabla \times \frac{\mathbf{B}}{\mu_o} &= \mathbf{J} + \mathbf{J}_m \\ &= \mathbf{J} + \nabla \times \mathbf{M}\end{aligned} \tag{5-71}$$

Here, **J** is the free current(conduction and/or convection current), and $\mathbf{J}_m$ is the magnetization current induced in the material. In view of μ_o in Eq. (5-71), we note that Eq. (5-71) is based on the atomic model of a material, in which the magnetic material is considered as an aggregate of the magnetic dipole moments positioned at the equivalent lattice points in free space. Rewriting Eq. (5-71), we have

$$\nabla \times \left(\frac{\mathbf{B}}{\mu_o} - \mathbf{M} \right) = \mathbf{J} \tag{5-72}$$

We now redefine **H** in the magnetic material as

$$\mathbf{H} \equiv \frac{\mathbf{B}}{\mu_o} - \mathbf{M} \tag{5-73}$$

or

$$\boxed{\mathbf{B} = \mu_o (\mathbf{H} + \mathbf{M})} \qquad \text{[T]} \tag{5-74}$$

The newly defined **H** makes the point form of Ampere's circuital always hold true irrespective of the material such that

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J}} \qquad [\mathrm{A/m^2}] \tag{5-75}$$

It is important to note that **J** in Eq. (5-75) is the free current density, excluding the magnetization current induced in the material.

To find the integral form of Eq. (5-75) we take the surface integral of both sides of Eq. (5-75) over a surface $\mathcal{S}$, i.e.,

$$\int_{\mathcal{S}} (\nabla \times \mathbf{H}) \cdot d\mathbf{s} = \int_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{s}$$

Application of Stokes's theorem to the left-hand side of the above equation, noting that the right-hand side is the total current *I*, leads to Ampere's circuital law, which always holds regardless of the material. Namely,

$$\boxed{\oint_C \mathbf{H} \cdot d\mathbf{l} = I} \qquad \text{[A]} \qquad (5\text{-}76)$$

Again, *I* is the total current enclosed by the loop *C*, or passing through the surface *S* bounded by *C*.

The magnetization depends on the magnetic field intensity such that

$$\boxed{\mathbf{M} = \chi_m \mathbf{H}} \qquad \text{[A/m]} \qquad (5\text{-}77)$$

where χ_m is the magnetic susceptibility, which is a constant, independent of the magnitude and direction of **H**, in a homogeneous, linear, and isotropic material.

Inserting Eq. (5-77) into Eq. (5-74) leads to the constitutive relation between **B** and **H**, which may be given in different forms as follows:

$$\boxed{\mathbf{B} = \mu_o(1+\chi_m)\mathbf{H} = \mu_o\mu_r\mathbf{H} = \mu\mathbf{H}} \qquad \text{[T]} \qquad (5\text{-}78)$$

where μ is the permeability measured in henrys per meter[H/m], μ_o is the permeability of free space, and μ_r is the relative permeability, which is a dimensionless constant in a simple medium.

From Eq. (5-78) we obtain the relation between χ_m and μ, which may be useful for solving practical problems, i.e.,

$$\chi_m = \frac{\mu}{\mu_o} - 1 \qquad (5\text{-}79)$$

This is a dimensionless constant, independent of **H**, in a simple medium. In general, χ_m may be a function of position(inhomogeneous medium), depending on the magnitude of **H**(nonlinear medium), or the direction of **H**(anisotropic medium).

Magnetic materials can be classified into three groups according to μ_r :

$$\mu_r \lesssim 1 \qquad \text{(diamagnetic)} \qquad (5\text{-}80a)$$

$$\mu_r \gtrsim 1 \qquad \text{(paramagnetic)} \qquad (5\text{-}80b)$$

$$\mu_r >> 1 \qquad \text{(ferromagnetic)} \qquad (5\text{-}80c)$$

For diamagnetic materials, the typical value of χ_m is of the order of $-10^{-6} \sim -10^{-4}$. For paramagnetic materials, χ_m is of the order of $10^{-5} \sim 10^{-3}$. For ferromagnetic materials, χ_m is of the order of $10^{1} \sim 10^{5}$.

To recapitulate, Ampere's circuital law and Gauss's law for magnetism constitute two fundamental relations for the static magnetic field in that they allow us to uniquely determine the magnetic field in a region of space, regardless of the material medium filling the region, in accordance with Helmholtz's theorem. These relations may be expressed either in point form or in integral form as

$$\boxed{\oint_C \mathbf{H} \bullet d\mathbf{l} = I} \tag{5-81a}$$

$$\boxed{\oint_S \mathbf{B} \bullet d\mathbf{s} = 0} \tag{5-81b}$$

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J}} \tag{5-81c}$$

$$\boxed{\nabla \bullet \mathbf{B} = 0} \tag{5-81d}$$

Two vector fields **B** and **H** are related by the constitutive relation:

$$\mathbf{B} = \mu \mathbf{H} \tag{5-82}$$

where the permeability μ is a constant, independent of **H**, in a simple medium.

Example 5-11

A long filament with a current I is oriented along the z-axis in free space. It is then enclosed by a long hollow cylinder of permeability μ_1, with an inner radius a and an outer radius b, as shown in Fig. 5.18. Find

(a) **H** and **B** everywhere by Ampere's circuital law and
(b) magnetization current density in the hollow cylinder,
(c) magnetization surface current densities on the inner and outer surfaces.

Solution

As far as **H** is concerned, the hollow cylinder can be completely ignored, because it carries no free current. From symmetry considerations, **H** is of the form $\mathbf{H} = H_\phi(\rho)\,\mathbf{a}_\phi$, and thus we choose a circle centered at the filament as the Amperian path.

(a) In the region $0 < \rho < a$, Ampere's circuital law gives

$$\mathbf{H} = H_\phi\,\mathbf{a}_\phi = \frac{I}{2\pi\rho}\mathbf{a}_\phi, \text{ and } \mathbf{B} = \mu_o \mathbf{H} = \frac{I\mu_o}{2\pi\rho}\mathbf{a}_\phi \tag{5-83a}$$

In the region $a \le \rho \le b$, we have

$$\mathbf{H} = \frac{I}{2\pi\rho}\mathbf{a}_\phi, \text{ and } \mathbf{B} = \mu_1 \mathbf{H} = \frac{I\mu_1}{2\pi\rho}\mathbf{a}_\phi \tag{5-83b}$$

In the region $\rho > b$, we have

$$\mathbf{H} = \frac{I}{2\pi\rho}\mathbf{a}_\phi, \text{ and } \mathbf{B} = \mu_o \mathbf{H} = \frac{I\mu_o}{2\pi\rho}\mathbf{a}_\phi \tag{5-83c}$$

Note that the expressions for **H** in Eq. (5-83) are the same, independent of the hollow cylinder.

(b) In the region $a \le \rho \le b$, inserting Eq. (5-83b) into Eq. (5-77) we obtain

$$\mathbf{M} = \chi_m \mathbf{H} = \left(\frac{\mu_1}{\mu_o} - 1 \right) \frac{I}{2\pi\rho} \mathbf{a}_\phi \tag{5-84}$$

Taking the curl of Eq. (5-84) in cylindrical coordinates, we have

$\nabla \times \mathbf{M} = 0$

Then, from Eq. (5-70) we obtain

$\mathbf{J}_m = 0$.

(c) On the cylindrical surface at $\rho = a$, from Eq. (5-84) we have

$$\mathbf{M} = \left(\frac{\mu_1}{\mu_o} - 1 \right) \frac{I}{2\pi a} \mathbf{a}_\phi \tag{5-85}$$

The unit normal to the surface at $\rho = a$ is $\mathbf{a}_n = -\mathbf{a}_\rho$. From Eqs. (5-67) and (5-85) we obtain

$$\mathbf{J}_{ms1} = \mathbf{M} \times \mathbf{a}_n = \left(\frac{\mu_1}{\mu_o} - 1 \right) \frac{I}{2\pi a} \mathbf{a}_z \qquad (\rho = a) \tag{5-86}$$

Similarly, on the cylindrical surface at $\rho = b$, from Eq. (5-84) we obtain

$$\mathbf{M} = \left(\frac{\mu_1}{\mu_o} - 1 \right) \frac{I}{2\pi b} \mathbf{a}_\phi \tag{5-87}$$

The unit normal to the surface at $\rho = b$ is $\mathbf{a}_n = \mathbf{a}_\rho$, From Eqs. (5-67) and (5-87) we obtain

$$\mathbf{J}_{ms2} = \mathbf{M} \times \mathbf{a}_n = -\left(\frac{\mu_1}{\mu_o} - 1 \right) \frac{I}{2\pi b} \mathbf{a}_z \qquad (\rho = b) \tag{5-88}$$

From Eqs. (5-86) and (5-88), we note that two surfaces of the hollow cylinder carry the magnetization surface currents of an equal amount, but flowing in the opposite directions.

If there is no free current in a magnetized material, Ampere's circuital law tells us that $\nabla \times \mathbf{H} = 0$ in the material. In this case, according to Eq. (5-77) and Eq. (5-70), we have $\nabla \times \mathbf{M} = 0$ and $\mathbf{J}_m = 0$ in the material. Under these conditions, the total magnetic flux can be obtained by adding the external magnetic flux and the flux produced by $\mathbf{J}_{ms}$ (see **Example 5-12**).

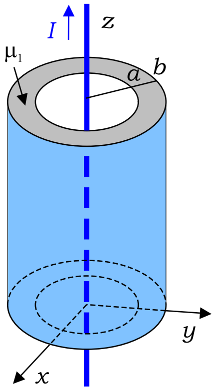


Fig. 5.18 A long filamentary current enclosed by a long hollow cylinder.

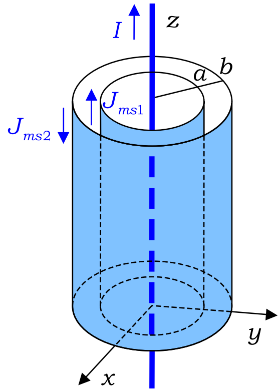


Fig. 5.19 Cylindrical sheets carrying the magnetization surface currents.

Example 5-12

Since there is no free current in the hollow cylinder in **Example 5-11**, the total magnetic flux is given by the sum of the external magnetic flux and the flux produced by the magnetization surface currents $\mathbf{J}_{ms1}$ and $\mathbf{J}_{ms2}$ residing in free space as shown in Fig. 5.19. Find $\mathbf{B}$ everywhere and compare it with the previous result in **Example 5-11**.

Solution

From symmetry considerations, $\mathbf{B}$ is of the form $\mathbf{B} = B_\phi(\rho)\,\mathbf{a}_\phi$ everywhere. In the region $0 < \rho < a$, Ampere's circuital law gives

$$\oint_C (\mathbf{B}/\mu_o)\cdot d\mathbf{l} = I\text{, or } (B_\phi/\mu_o)2\pi\rho = I$$

The magnetic flux density is therefore

$$\mathbf{B} = B_\phi\,\mathbf{a}_\phi = \frac{I\mu_o}{2\pi\rho}\mathbf{a}_\phi \qquad (0 < \rho < a) \qquad \text{(5-89a)}$$

In the region $a < \rho < b$, Ampere's circuital law gives

$$\oint_C (\mathbf{B}/\mu_o)\cdot d\mathbf{l} = I + 2\pi a J_{ms1}$$

Inserting J_{ms1} expressed by Eq. (5-86) into the above equation, we have

$$\left(\frac{B_\phi}{\mu_o}\right)2\pi\rho = I + 2\pi a\left(\frac{\mu_1}{\mu_o} - 1\right)\frac{I}{2\pi a}$$

The magnetic flux density is therefore

$$\mathbf{B} = B_\phi\,\mathbf{a}_\phi = \frac{I\mu_1}{2\pi\rho}\mathbf{a}_\phi \qquad (a < \rho < b) \qquad \text{(5-89b)}$$

In the region $\rho > b$, Ampere's circuital law gives

$$\oint_C (\mathbf{B}/\mu_o)\cdot d\mathbf{l} = I + 2\pi a J_{ms1} - 2\pi b J_{ms2}$$

Inserting J_{ms1} and J_{ms2} given in Eqs. (5-86) and (5-88) into the above equation we get

$$\left(\frac{B_\phi}{\mu_o}\right)2\pi\rho = I + 2\pi a\left(\frac{\mu_1}{\mu_o}-1\right)\frac{I}{2\pi a} - 2\pi b\left(\frac{\mu_1}{\mu_o}-1\right)\frac{I}{2\pi b}$$

The magnetic flux density is therefore

$$\mathbf{B} = B_\phi \mathbf{a}_\phi = \frac{I\mu_o}{2\pi\rho}\mathbf{a}_\phi \qquad (\rho > b) \qquad \text{(5-89c)}$$

The three results in Eq. (5-89) are the same as those in Eq. (5-83). In each region, **H** is obtained from **B** by the relation $\mathbf{H} = \mathbf{B}/\mu$. An important point to remember is that **H** cannot be obtained directly from the magnetization currents.

If the magnetization **M** is induced in a magnetic material of a high permeability, by an externally applied magnetic field, it will in turn distort the magnetic field lines both inside and outside the material. Consider Fig. 5.20(a), which depicts the case when a small cylinder of a high permeability μ is brought into the region of a uniform **B**. For the sake of argument, we replace the magnetization surface current induced on the surface of the cylinder with a simple circular loop carrying a steady current. Then the magnetic flux lines are obtained, by using numerical methods, by computing and summing the external magnetic flux and the flux due to the current loop, and plotted to scale in Fig. 5.20(b). We see that the magnetic flux lines are concentrated in the interior of the material of a high μ, while the total magnetic flux is conserved inside and outside the material. This example shows that a structure made of a magnetic material of a high μ may be used to confine and guide magnetic flux lines.

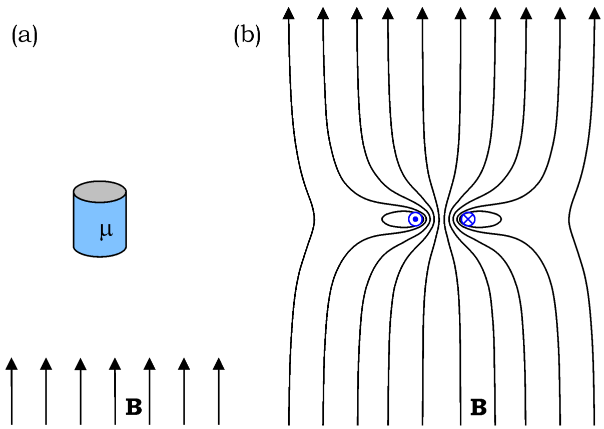


Fig. 5.20 A magnetic material in a uniform **B** (a) a cylinder of a high μ (b) the induced **M** distorts **B** inside and outside the cylinder.

Exercise 5.11

Determine χ_m and μ_r of a material in which $|\mathbf{M}|$ is 10 times as large as $|\mathbf{H}|$.

Ans. $\chi_m = 10$, and $\mu_r = 11$.

Exercise 5.12

For $|\mathbf{B}| = 0.02\,[\mathrm{T}]$ in a solid nickel of $\mu_r = 250$, find χ_m, H, and M inside the material.

Ans. $\chi_m = 249$, $H = 63.9\,[\mathrm{A/m}]$, and $M = 15.9\times 10^3\,[\mathrm{A/m}]$.

Review Questions with Hints

RQ 5.29 What are two sources of $\mathbf{B}$ in a magnetic material? [Eq.(5-74)]
RQ 5.30 What is the source of $\mathbf{H}$ in a magnetic material? [Eq.(5-75)]
RQ 5.31 Define χ_m, μ_r, and μ. [Eq.(5-78)]
RQ 5.32 Categorize magnetic materials in terms of μ_r. [Eq.(5-80)]
RQ 5.33 Can you make a finite cylinder of a high μ uniformly magnetized as $\mathbf{M} = M_o\,\mathbf{a}_z$ by applying an external magnetic field? [Fig.5.20]

5.7.3 Hysteresis of a Ferromagnetic Material

An externally applied magnetic field can make a ferromagnetic material strongly magnetized. Ferromagnetic materials exhibit unique magnetic properties, which are mainly due to the magnetized domains formed inside the material. The relation between $\mathbf{B}$ and $\mathbf{H}$ is highly nonlinear in the ferromagnetic material in that the material retains a fair amount of $\mathbf{M}$ even after the applied field is removed, and the relative permeability depends on the magnetic field intensity. However, the ferromagnetic material completely loses its magnetization and behaves like a paramagnetic material, when it is heated above the Curie temperature.

The atoms of the ferromagnetic material have a permanent magnetic dipole moment originating from the electron spin moment. The magnetic dipole moments tend to align with each other over small regions called magnetic domains owing to the strong interatomic forces. The magnetic domains have various shapes and sizes ranging from micrometers to millimeters in linear dimension. Every magnetic domain is fully magnetized, having a uniform magnetization, meaning that all constituent magnetic dipole moments point in the same direction within a domain. In the absence of an external field, the magnetic domains are randomly oriented yielding no net dipole moment as a whole. Under the influence of an externally applied magnetic field, however, the domains with the magnetization vector parallel to the applied field grow at the expense of the others, and result in a net dipole moment very large in magnitude. When the external field is removed, it is not possible to have the domains broken again into pieces with the original orientations, but a residual magnetization remains in the material. This phenomenon is referred to as the magnetic hysteresis; the magnetic state at the present time depends on the past magnetic history of the material. Above the Curie temperature, the ferromagnetic material behaves like a parametric material in the sense that the permanent dipole moments are randomly oriented, not forming magnetic domains. As an example, the Curie temperature for iron is $770^{o}C$.

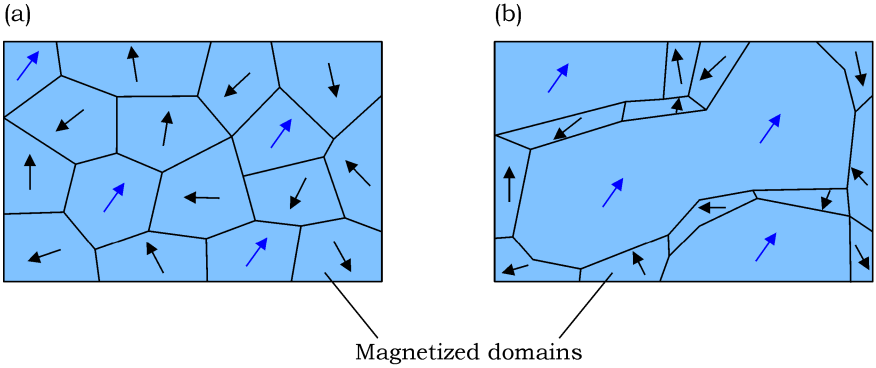


Fig. 5.21 A ferromagnetic material with magnetic domains (a) No external **B** (b) An external **B** of a moderate strength is applied along the blue arrow.

The magnetic behavior of the ferromagnetic material can be well described by the *B*-*H* magnetization curve, which is a plot of the magnetic flux density *B* measured in the material versus the applied magnetic field intensity *H*, as shown in Fig. 5.22. Let us begin with a ferromagnetic material that is completely demagnetized. Suppose we apply a magnetic field to the material increasing its magnitude from zero to H_1, which may be accomplished by inserting the material in a solenoidal coil and increasing the current in the coil. Then, the magnetic flux density grows from zero to B_1 in the material along the dotted line called the initial magnetization curve, as shown in Fig. 5.22(a). Even if we decrease *H* from H_1 to zero, *B* does not go to zero in the material, but it follows the upper path of the *B*-*H* curve and remains at a nonzero value B_r called a residual flux density. The phenomenon that the magnetization lags behind the field is called the hysteresis, meaning "to lag" in Greek. To make the residual flux density vanish from the material, we need to apply the magnetic field H_c called the coercive field intensity in the opposite direction. In the ferromagnetic material, both B_r and H_c depend on the maximum field intensity H_1.

As the applied field *H* varies periodically between two maximum values $\pm H_1$, the trace of *B* forms a closed loop in the *B*-*H* plot, per cycle of the periodic magnetization, which we call the hysteresis loop. The shape and size of the hysteresis loop depend on the ferromagnetic material and the maximum applied field. The hysteresis loop is a direct consequence of the nonlinear relation between *B* and *H*. In the ferromagnetic material, the permeability μ is a nonlinear function of *H*, and is equal to the ratio B/H in the *B*-*H* curve, which is not necessarily a tangent to the curve. The area of the hysteresis loop corresponds to the energy dissipated in the form of heat, per unit volume of the material per cycle of the periodic magnetization. Hard ferromagnetic materials have a wide hysteresis loop as shown in Fig. 5.22(a), whereas soft ferromagnetic materials have a narrow hysteresis loop as shown in Fig. 5.22(b). The permanent magnet is possible because of the existence of the residual flux density in the ferromagnetic material. The hard ferromagnetic material, having a larger coercive field intensity, makes a better permanent magnet because it is hardly demagnetized by an external magnetic field of a moderate strength.

If the externally applied field *H* is strong enough to cause a total alignment of the magnetic dipole moments with the field, the ferromagnetic material is said to be saturated. A further increase in *H* will result in no change in the magnetization. However, *B* increases in the material beyond the saturation, because of the added magnetic flux by the external field.

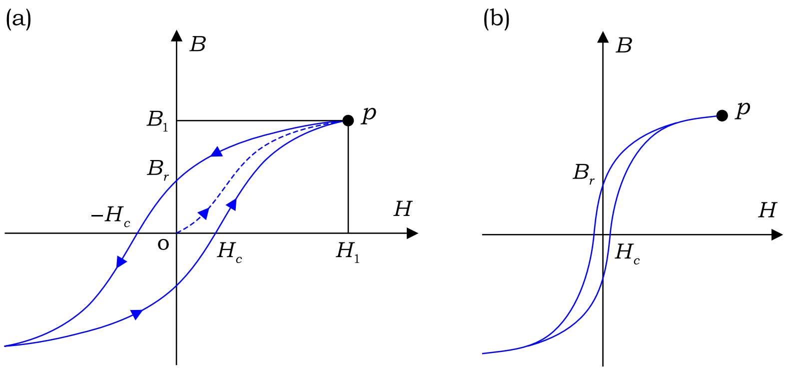


Fig. 5.22 Hysteresis loop for (a) hard ferromagnetic material (b) soft ferromagnetic material.

Exercise 5.13

For an iron metal, saturation magnetization is 7.9×10^5 [A/m], and number density of atoms is 8.5×10^{28} [m^{-3}]. Find magnetic dipole moment of an atom.

Ans. 9.3×10^{-24} [A·m^2].

Exercise 5.14

At saturation point *p* in Fig. 5.22(b), the tangent to the curve is at an angle 7^o to the *H*-axis, while the line from the origin to *p* is at an angle 45^o. Find μ_r at *p*.

Ans. $\mu_r = \tan 45^o / \tan 7^o = 8.1$ (the first angle is for μ and the second for μ_o).

Review Questions with Hints

RQ 5.34 What is hysteresis loop? [Fig.5.22]
RQ 5.35 Explain residual flux density and coercive field intensity. [Fig.5.22]
RQ 5.36 Explain magnetic domains and Curie temperature. [Fig.5.21]

5.7.4 Magnetic Boundary Conditions

Gauss's law for magnetism is also called the law of conservation of magnetic flux. It is so named because the net magnetic flux through any closed surface is zero, regardless of the material medium. Moreover, Ampere's circuital law should also hold regardless of the material medium. In view of these considerations, the two fundamental relations for static magnetic fields must be satisfied in any region of space, even at an interface between two adjoining materials. To obtain the boundary conditions for **B** and **H** at an interface between two different magnetic materials, we follow the same procedure used for **E** and **D**.

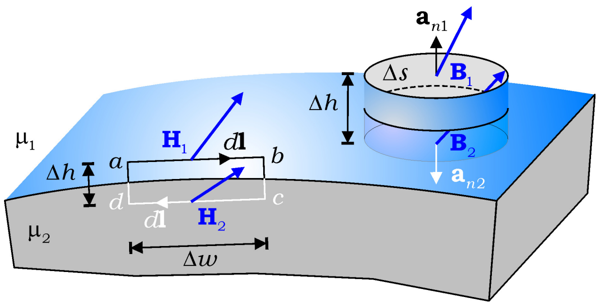


Fig. 5.23 An interface of two adjoining magnetic materials of μ_1 and μ_2.

With reference to Fig. 5.23, applying Ampere's circuital law to the rectangular loop *abcda*, we have

$$\begin{aligned}\oint_C \mathbf{H}\cdot d\mathbf{l} &= H_{1t}\Delta w - H_{2t}\Delta w + \int_b^c \mathbf{H}\cdot d\mathbf{l} + \int_d^a \mathbf{H}\cdot d\mathbf{l} \\ &= \mathrm{J}_{sn}\Delta w\end{aligned} \tag{5-90}$$

where t denotes the tangential component, and J_{sn} denotes the surface current density flowing on the interface in the direction perpendicular to the loop surface. Even though the line integral of **H** along the left or the right side of the loop vanishes as the loop height Δh tends to zero, the surface current $\mathrm{J}_{sn}\Delta w$ remains enclosed by the loop. The boundary condition for the tangential component of **H** is therefore

$$\boxed{H_{1t} - H_{2t} = \mathrm{J}_{sn}} \tag{5-91}$$

In the presence of the surface current density J_{sn} on the interface, the tangential component of **H** is discontinuous across the interface. A more general expression for the boundary condition for **H** is possible if a unit normal to the interface $\mathbf{a}_{1-2}$ is employed, which is directed from medium 2 to medium 1, i.e.,

$$\boxed{\mathbf{a}_{1-2} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_s} \tag{5-92}$$

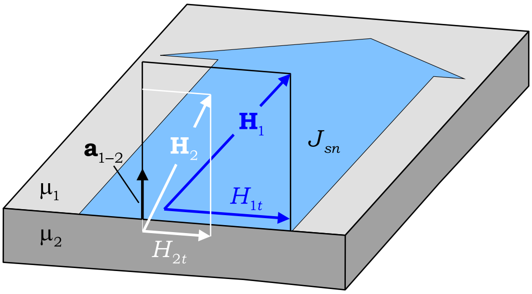


Fig. 5.24 Boundary condition for H_t at an interface with a surface current.

If an interface is formed by two adjoining materials of finite conductivities, it cannot support a surface current; a surface current can exist only on a perfectly conducting surface. In most practical cases, the surface current is zero at the interface, $\mathbf{J}_s = 0$, and the boundary condition for the tangential component of **H** is

$$\boxed{H_{1t} = H_{2t}} \tag{5-93}$$

***The tangential component of* H *is continuous across the interface*.** Applying the constitutive relation, $\mathbf{B} = \mu\mathbf{H}$, to Eq. (5-93), we obtain

$$\frac{B_{1t}}{\mu_1} = \frac{B_{2t}}{\mu_2} \tag{5-94}$$

The tangential component of **B** is discontinuous across the interface.

We now apply Gauss's law for magnetism to a cylinder extending across the interface as shown in Fig. 5.23. Following the same procedure used for **D**, we obtain the boundary condition for the normal component of **B** as

$$\boxed{B_{1n} = B_{2n}} \tag{5-95}$$

***The normal component of* B *is continuous across the interface*.** Applying the constitutive relation, $\mathbf{B} = \mu\mathbf{H}$, to Eq. (5-95), we obtain

$$\mu_1 H_{1n} = \mu_2 H_{2n} \tag{5-96}$$

The normal component of **H** is discontinuous across the interface.

Example 5-13

Consider an interface between free space and an iron metal with $\mu = 1000\mu_o$. In free space, $\mathbf{B}_1$ makes an angle of $\theta_1 = 1.0^o$ with the normal to the interface. Find (a) direction of $\mathbf{B}_2$ in the iron metal, and (b) ratio B_2 / B_1.

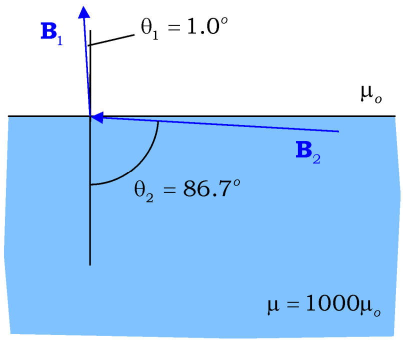


Fig. 5.25 An interface between free space and an iron metal with $\mu = 1000\mu_o$.

Solution

(a) From the boundary conditions for **H** and **B**, we get

$$\frac{1}{\mu}B_{2t} = \frac{1}{\mu_o}B_1\sin(1^o) \qquad \text{(tangential component of } \mathbf{H}) \qquad (5\text{-}97a)$$

$$B_{2n} = B_1\cos(1^o) \qquad \text{(normal component of } \mathbf{B}) \qquad (5\text{-}97b)$$

In the iron metal, we have

$$\tan\theta_2 = \frac{B_{2t}}{B_{2n}} = \frac{(\mu/\mu_o)B_1\sin(1^o)}{B_1\cos(1^o)} = 17.46$$

Thus, $\theta_2 = 86.7^o$.

(b) Combining Eqs. (5-97a) and (5-97b), we obtain

$$B_2 = \sqrt{(B_{2t})^2 + (B_{2n})^2} = B_1\sqrt{(1000\mu_o/\mu_o)^2\sin^2(1^o) + \cos^2(1^o)}$$

Thus, $B_2/B_1 = 17.5$.

Example 5-14

A long filament with a steady current I_o penetrates a dielectric sphere of radius a, for which $\mu = \mu_o\mu_r$ and $\sigma = 0$, as shown in Fig 5.26. Verify by use of Helmholtz's theorem that the magnetic field $\mathbf{H} = (I_o/2\pi\rho)\mathbf{a}_\phi$ is a unique solution in the region $\rho > 0$ in cylindrical coordinates, irrespective of the sphere.

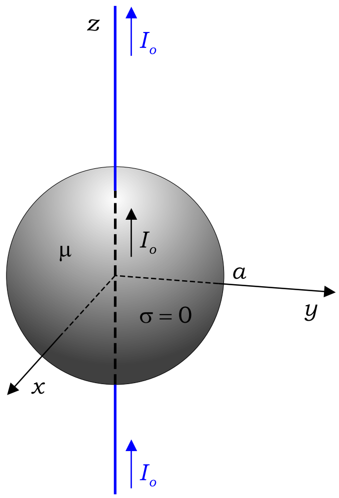


Fig. 5.26 A long filament with a current I_o penetrating a dielectric sphere.

Solution

The given **H** satisfies Ampere's circuital law inside and outside the sphere. It is tangential to the surface of the sphere, and thus continuous across the interface(the spherical surface), satisfying the boundary condition. The given **H** satisfies both $\nabla \times \mathbf{H} = 0$ and $\nabla \bullet \mathbf{B} = 0$ in the region $\rho > 0$ and vanishes at infinity. Thus, the given **H** is a unique solution in accordance with the Helmholtz's theorem.

If the dielectric formed a cube, the given **H** would not be a solution, because it does not satisfy the boundary condition at the surface of the cube.

Exercise 5.15

At an interface between two materials of μ_1 and μ_2, find the relation between the tilt angles of **B** in the two regions(see Fig. 5.25) in terms of μ_1 and μ_2.

Ans. $\tan\theta_1 / \tan\theta_2 = \mu_1 / \mu_2$.

Exercise 5.16

Which boundary condition represents the conservation of magnetic flux?

Ans. $B_{1n} = B_{2n}$.

Review Questions with Hints

RQ 5.37 What are the boundary conditions for **B** and **H** at an interface with no surface current? [Eqs.(5-93)-(5-96)]

RQ 5.38 Is the magnetic flux conserved across an interface carrying a surface current? [Eq.(5-95)]

RQ 5.39 Does an interface with no surface current always cause a bending of the magnetic flux line? [Eqs.(5-94)(5-95)]

5.8 Inductance and Inductors

An inductor in magnetostatics is analogous to a capacitor in electrostatics. As a capacitor can store energy in its electric field, an inductor can store energy in its magnetic field. A single turn of wire may be the simplest inductor, as a typical inductor consists of many turns of wire. When magnetic flux lines are produced by a single current-carrying loop, they always pass through the loop surface and form closed lines around the loop. As the capacitance is defined as the ratio between the charge accumulated on the conducting plate and the potential difference across the plates, the inductance is defined as the ratio between the magnetic flux linking the loop and the current flowing in the loop.

Let us consider a closed loop $\mathcal{C}_1$ carrying a steady current I_1, and another closed loop $\mathcal{C}_2$ situated in the neighborhood of the first one, as shown in Fig. 5.27. When the current I_1 in $\mathcal{C}_1$ produces the magnetic flux with a density $\mathbf{B}_1$, some flux will pass through the surface $\mathcal{S}_2$ bounded by $\mathcal{C}_2$. The mutual flux Φ_{12} is defined as the magnetic flux that is produced by I_1 and links with $\mathcal{C}_2$, or passing through the surface $\mathcal{S}_2$, i.e.,

$$\Phi_{12} = \int_{\mathcal{S}_2} \mathbf{B}_1 \bullet d\mathbf{s} \tag{5-98}$$

If $\mathcal{C}_2$ has N_2 turns, the magnetic flux linkage Λ_{12} is given by

$$\Lambda_{12} = N_2 \Phi_{12} \tag{5-99}$$

where Φ_{12} is the magnetic flux linking with a single turn in $\mathcal{C}_2$. In our notations, subscript "12" denotes something "from 1 to 2", whereas subscript "1-2" denotes something "from 2 to 1".

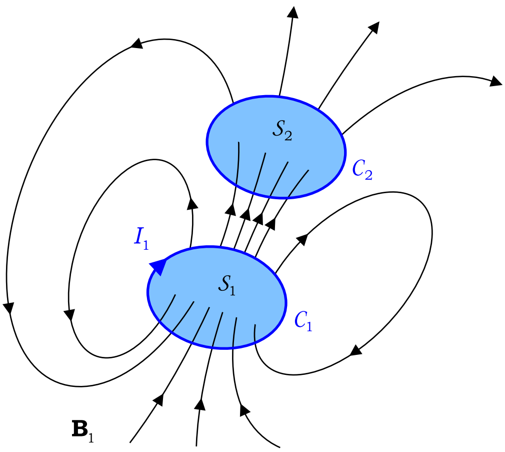


Fig. 5.27 Magnetic coupling between two conducting loops.

With reference to Fig. 5.28, we can show that the magnetic flux linkage with a loop $\mathcal{C}_2$ is equal to the total magnetic flux passing through the total surface enclosed by $\mathcal{C}_2$. Suppose that $\mathcal{C}_2$ has two turns, and that Σ_1 and Σ_2 are the surfaces bounded by the individual turns as shown in Fig. 5.28. If the loop $\mathcal{C}_2$ is uncoiled to a big loop of a single turn, we will immediately recognize that the total surface enclosed by $\mathcal{C}_2$ is given by $\mathcal{S} = \Sigma_1 + \Sigma_2$, and that the same magnetic flux Φ_{12} passes through $\mathcal{S}$ twice in the same direction. Thus, the magnetic flux linking with $\mathcal{C}_2$ must be $2\Phi_{12}$. In view of these considerations, we conclude that the magnetic flux linkage is equal to the product of the magnetic flux linking with each turn and the number of turns.

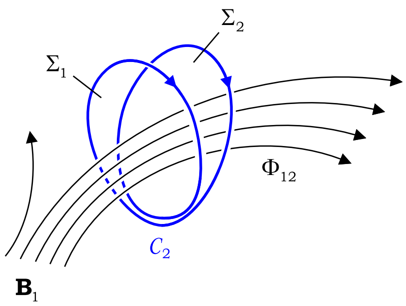


Fig. 5.28 Flux linkage with a loop of 2 turns.

Two neighboring loops C_1 and C_2 as shown in Fig. 5.27 are magnetically coupled through the mutual flux. To describe the magnetic coupling, we define the mutual inductance between two loops C_1 and C_2 as

$$\boxed{M_{12} = \frac{\Lambda_{12}}{I_1} = \frac{N_2}{I_1}\int_{\mathcal{S}_2} \mathbf{B}_1 \cdot d\mathbf{s}} \qquad \text{[H]} \qquad (5\text{-}100)$$

where N_2 is the number of turns in $\mathcal{C}_2$, $\mathcal{S}_2$ is the surface enclosed by a single turn in $\mathcal{C}_2$, and $\mathbf{B}_1$ is the magnetic flux density produced by I_1 flowing in $\mathcal{C}_1$. The mutual inductance is measured in henrys[H], and should not be confused with the magnetization $\mathbf{M}$. In a simple medium in which the permeability μ is a constant, independent of the magnitude and direction of **B**, Ampere's circuital law $\nabla \times (\mathbf{B}/\mu) = \mathbf{J}$ shows that $\mathbf{B}_1$ is linearly proportional to I_1 flowing in $\mathcal{C}_1$. Thus the mutual inductance defined by Eq. (5-100) should be independent of the current I_1.

In the same manner, the mutual inductance M_{21} is defined as the ratio between the magnetic flux linkage with $\mathcal{C}_1$ and the current I_2 flowing in $\mathcal{C}_2$. In a linear medium in which the permeability μ is independent of the magnitude of **B**, we have

$$\boxed{M_{12} = M_{21}} \qquad \text{[H]} \tag{5-101}$$

This can be readily verified by use of the concept of mutual energy in Section 5-5.

When the magnetic flux is produced by the current I_1 flowing in C_1, it can link with the loop itself, which we call the magnetic flux linkage Λ_{11}. The self-inductance, or inductance, of a loop C_1 is defined as the ratio between the magnetic flux linkage with C_1 and the current I_1 flowing in C_1. That is,

$$\boxed{L_1 = \frac{\Lambda_{11}}{I_1} = \frac{N_1}{I_1}\int_{S_1} \mathbf{B}_1 \cdot d\mathbf{s}} \qquad \text{[H]} \tag{5-102}$$

where N_1 is the number of turns in C_1, and S_1 is the surface enclosed by a single turn in C_1. The inductance is measured in henrys[H]. Although the inductance may be independent of the current, it is highly dependent on the geometry of the conducting structure and the permeability of the surrounding medium.

The inductance of an inductor can be determined by taking the following steps:

1. Assume current I in the conductor.
2. Choose a coordinate system considering symmetry.
3. Find **B** from I by the Biot-Savart law or Ampere's law.
4. Find flux linkage $\Lambda = N\Phi = N\int \mathbf{B}\cdot d\mathbf{s}$.
5. Find L from $L = \Lambda / I$.

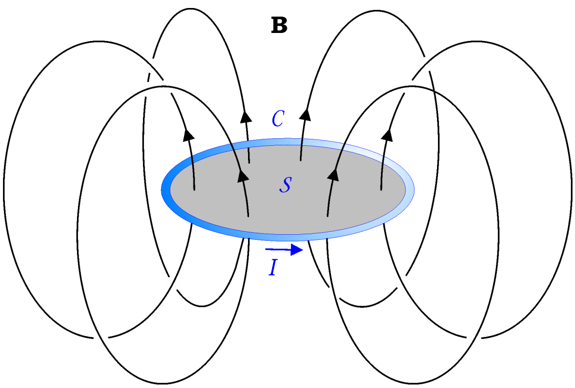


Fig. 5.29 Inductance of a loop C is the ratio between the magnetic flux linking with C and the current I.

Example 5-15

Find the mutual inductance between a very long straight wire and a rectangular loop residing in free space as shown in 5-30.

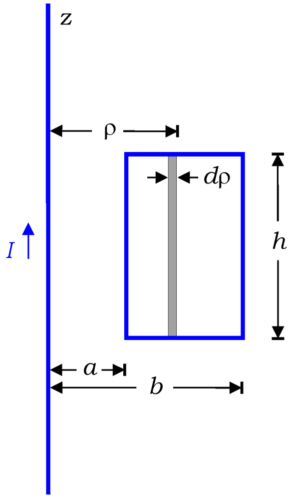


Fig. 5.30 A very long straight wire and a rectangular loop.

Solution

When a steady current I is assumed in the straight wire, the induced magnetic flux density is

$$\mathbf{B}_1 = \frac{\mu_o I}{2\pi\rho}\mathbf{a}_\phi$$

Magnetic flux linking the rectangular loop is

$$\Lambda_{12} = \Phi_{12} = \int_{S_2} \mathbf{B}_1 \bullet d\mathbf{s} = \int_{\rho=a}^{\rho=b} \frac{\mu_o I}{2\pi\rho}\mathbf{a}_\phi \bullet (h d\rho\, \mathbf{a}_\phi)$$
$$= \frac{\mu_o I h}{2\pi}\ln\frac{b}{a}$$

In the above equation, the differential area vector $d\mathbf{s}$ is perpendicular to the narrow strip of width $d\rho$, and is directed into the paper, parallel to $\mathbf{B}_2$.

Thus the mutual inductance is

$$M_{12} = \frac{\Lambda_{12}}{I} = \frac{\mu_o h}{2\pi}\ln\frac{b}{a} \qquad \text{[H]} \qquad (5\text{-}103)$$

It would be demanding to find the mutual inductance by assuming a current I in the rectangular loop and computing the flux linkage with the straight wire. The mutual inductance, obtained in this way, must be the same as Eq. (5-103), since free space is a linear medium.

Example 5-16

Find the self-inductance per unit length of a very long solenoid, which consists of N turns per unit length tightly wound on an air core of radius a.

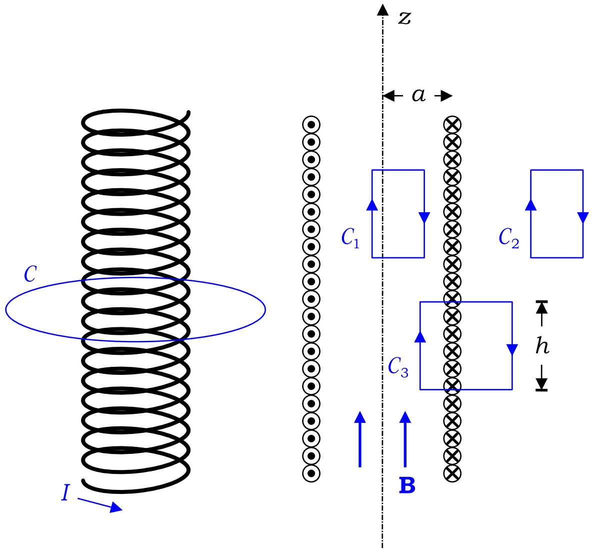


Fig. 5.31 A very long solenoid.

Solution

Let us explore the solenoid for symmetry:

(1) The solenoid has cylindrical symmetry(no change after rotated about the z-axis), and translational symmetry in the z-direction(no change after displaced in the z-direction). Thus the resulting **B** should be of the form $\mathbf{B} = B_\rho(\rho)\,\mathbf{a}_\rho + B_\phi(\rho)\,\mathbf{a}_\phi + B_z(\rho)\,\mathbf{a}_z$ at the most.

(2) The solenoid appears the same even if it is vertically flipped, or rotated about the x-axis by 180^o, except for the reversed direction of the current. If the current is reversed in the original solenoid, the direction of **B** should be reversed at every point in space. Since the direction of $B_\rho \mathbf{a}_\rho$ remains the same when it is vertically flipped, **B** of the solenoid should have no ρ-component such that $\mathbf{B} = B_\phi(\rho)\,\mathbf{a}_\phi + B_z(\rho)\,\mathbf{a}_z$.

(3) The current flows only in the $\mathbf{a}_\phi$-direction in an ideal solenoid, and thus no current passes through the surface of the loop C shown in Fig. 5.31. Applying Ampere's circuital law to C, we have $\oint_C (\mathbf{B}/\mu_o)\cdot d\mathbf{l} = \int_{\phi=0}^{2\pi} (B_\phi(\rho)/\mu_o)\,\mathbf{a}_\phi \cdot (\rho d\phi\,\mathbf{a}_\phi) = 0$, which results in $B_\phi = 0$. Thus, the functional form of **B** should be reduced to $\mathbf{B} = B_z(\rho)\,\mathbf{a}_z$.

Applying Ampere's circuital law to loops C_1 and C_2 as shown in Fig. 5.31, by assuming $\mathbf{B} = B_z(\rho)\,\mathbf{a}_z$, leads to a conclusion that **B** is constant in the interior and exterior of the solenoid. Thus, **B** should be so far of the form $\boxed{\mathbf{B} = B_z\,\mathbf{a}_z}$ everywhere.

Next, we can show that $\mathbf{B} = 0$ at infinity, and thus $\mathbf{B} = 0$ outside the solenoid. To start with, the solenoid is considered to be a stack of a large number of identical current loops. The magnetic field intensity of a single current loop varies with distance as $1/R^3$, or $1/(z^2+\rho^2)^{3/2}$ in cylindrical coordinates, as can be seen from Eq. (5-65). Thus, the total magnetic field intensity at far distances can be approximated as $B \sim \int (z^2+\rho^2)^{-3/2} dz \sim 1/\rho^2$, which becomes zero at infinity. Thus, $\mathbf{B} = 0$ outside the solenoid.

Applying Ampere's circuital law to loop C_3 shown in Fig. 5.31, we obtain

$$hH_z = NIh$$

In the interior of the solenoid, we have

$$\mathbf{H} = NI\,\mathbf{a}_z \tag{5-104a}$$

$$\mathbf{B} = \mu_o NI\,\mathbf{a}_z \tag{5-104b}$$

Total magnetic flux through the cross-section of the solenoid is

$$\Phi = B\pi a^2 = \mu_o NI\pi a^2$$

Magnetic flux linkage per unit length of the z-axis is

$$\Lambda = N\Phi = \mu_o N^2 I\pi a^2$$

Self-inductance per unit length of the solenoid is therefore

$$L = \frac{\Lambda}{I} = \mu_o N^2 \pi a^2 \qquad [\text{H/m}] \tag{5-105}$$

Example 5-17

A very long coaxial cable consists of a solid inner conductor, having a radius a, and a cylindrical shell of negligible thickness, having a radius b. Assuming that the current I flows uniformly in the inner conductor, and returns through the outer conductor as a uniform surface current, find the inductance per unit length of the cable.

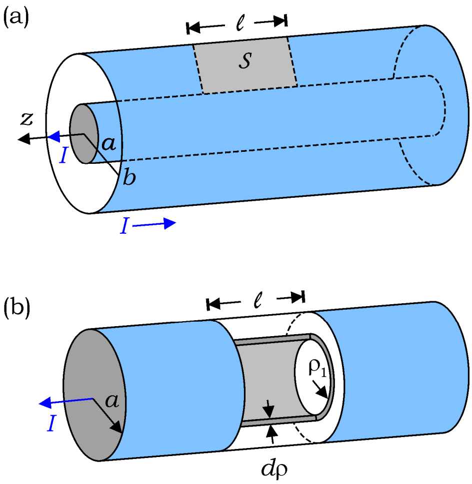


Fig. 5.32 A coaxial cable consisting of a solid conductor and a thin cylindrical shell.

Solution

From symmetry considerations, we anticipate $\mathbf{B} = B_\phi(\rho)\,\mathbf{a}_\phi$ everywhere.

By using Ampere's circuital law we obtain

$$\mathbf{B} = 0 \qquad (\rho > b) \qquad \text{(5-106a)}$$

$$\mathbf{B} = \frac{\mu_o I}{2\pi\rho}\mathbf{a}_\phi \qquad (a < \rho < b) \qquad \text{(5-106b)}$$

$$\mathbf{B} = \frac{\mu_o \rho I}{2\pi a^2}\mathbf{a}_\phi \qquad (0 \le \rho \le a) \qquad \text{(5-106c)}$$

(1) In the region $a < \rho < b$, the magnetic flux passing through surface S shown in Fig. 5.32(a) is

$$\Phi = \int_S \mathbf{B} \cdot d\mathbf{s} = \int_{\rho=a}^{\rho=b} \frac{\mu_o I}{2\pi\rho}\ell d\rho = \frac{\mu_o I\ell}{2\pi}\ln\frac{b}{a}$$

Magnetic flux linkage with the inner conductor per unit length of the cable is

$$\Lambda_1 = \frac{\Phi}{\ell} = \frac{\mu_o I}{2\pi}\ln\frac{b}{a} \qquad \text{(5-107)}$$

(2) In the region $0 \le \rho \le a$, let us consider a cylindrical shell of radius ρ_1, thickness $d\rho$, and length ℓ as shown in Fig. 5.32(b). The magnetic flux confined to this shell is obtained from (5-106c) as

$$d\Phi = B_\phi \ell d\rho = \frac{\mu_o \rho_1 I}{2\pi a^2}\ell d\rho \qquad \text{(5-108)}$$

It should be noted that $d\Phi$ in Eq. (5-108) is due to the partial current $I_1 = \rho_1^2 I / a^2$, which is enclosed by the shell, *not due to the total current I.* The differential magnetic flux linkage with the cable, per unit length, is therefore given by the product of $d\Phi / \ell$ and I_1 / I :

$$d\Lambda = \frac{d\Phi}{\ell}\frac{\rho_1^2}{a^2} = \frac{\mu_o \rho_1^3 I}{2\pi a^4} d\rho$$

The total magnetic flux linking the whole inner conductor, per unit length, is

$$\Lambda_2 = \int d\Lambda = \int_{\rho=0}^{\rho=a} \frac{\mu_o \rho^3 I}{2\pi a^4} d\rho \;= \frac{\mu_o I}{8\pi} \qquad (5\text{-}109)$$

In the above equation, subscript 1 is dropped from ρ for generalization.

The total magnetic flux linkage with the coaxial cable, per unit length, equals the sum of Λ_1 and Λ_2. The inductance per unit length of the coaxial cable is therefore

$$L = \frac{\Lambda_1 + \Lambda_2}{I} = \frac{\mu_o}{2\pi}\ln\frac{b}{a} + \frac{\mu_o}{8\pi} \qquad \text{[H/m]} \qquad (5\text{-}110)$$
$$\equiv L_{ex} + L_{in}$$

In Eq. (5-110), L_{ex} is called the external inductance and L_{in} is called the internal inductance.

Example 5-18

Two very long conducting wires of an equal radius a are parallel to each other in free space, with the axis-to-axis distance of b(b>>a). They carry uniform currents flowing in the opposite directions as shown in Fig. 5.33. Compute the external and internal inductances per unit length of the pair of wires.

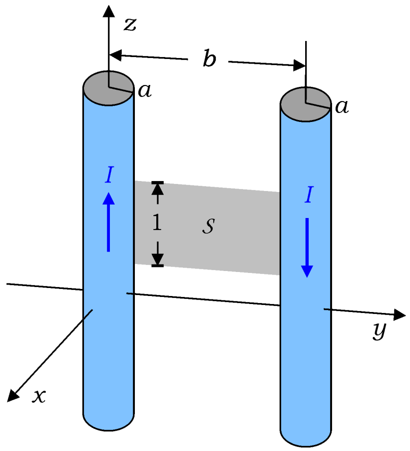


Fig. 5.33 A pair of wires carrying uniform currents in the opposite directions.

Solution

In the region $a < y < (b-a)$ in the yz-plane, **B** due to the left conductor is obtained by Ampere's circuital law as

$$\mathbf{B}_1 = -\frac{\mu_o I}{2\pi y}\mathbf{a}_x$$

In the same region, **B** due to the right conductor is

$$\mathbf{B}_2 = -\frac{\mu_o I}{2\pi(b-y)}\mathbf{a}_x$$

Under the condition $b >> a$, the magnetic flux linkage with the pair of wires, per unit length, is given by the surface integral of $\mathbf{B}_1 + \mathbf{B}_2$ over the surface S shown in Fig. 5.33, that is,

$$\Lambda_{ex} = \int_S (\mathbf{B}_1 + \mathbf{B}_2) \bullet d\mathbf{s} = \frac{\mu_o I}{2\pi}\int_{y=a}^{y=b-a}\left[\frac{1}{y} + \frac{1}{b-y}\right]dy$$
$$= \frac{\mu_o I}{\pi}\ln\left(\frac{b-a}{a}\right) \cong \frac{\mu_o I}{\pi}\ln\frac{b}{a}$$

The external inductance per unit length of the wires is therefore

$$L_{ex} = \frac{\Lambda_{ex}}{I} = \frac{\mu_o}{\pi}\ln\frac{b}{a} \quad (5\text{-}111)$$

In the interior of the left conductor, for instance, there exist two **B**'s originating from two different sources: one is due to the internal current I and the other is due to the current flowing in the other conductor. Under the condition $b >> a$, **B** due to I flowing in the other conductor is ignored compared with that due to the internal current. Thus the internal inductance of a wire, per unit length, is equal to $\mu_o / 8\pi$ as given by Eq. (5-110). The internal inductance of the pair of wires is therefore

$$L_{in} = \frac{\mu_o}{4\pi} \qquad [\text{H/m}] \quad (5\text{-}112)$$

The total inductance per unit length of a pair of wires is therefore

$$L = L_{ex} + L_{in} = \frac{\mu_o}{\pi}\left[\ln\frac{b}{a} + \frac{1}{4}\right] \qquad [\text{H/m}] \quad (5\text{-}113)$$

Exercise 5.17

The coaxial cable shown in Fig. 5.32 is made of conductors of μ_1 and the gap material of μ_2. Find the external and internal inductances per unit length.

Ans. $L_{ex} = \frac{\mu_2}{2\pi}\ln\frac{b}{a}$ [H/m] and $L_{in} = \frac{\mu_1}{8\pi}$.

Exercise 5.18
The toroid shown in Fig. 5.10 has a mean radius $\rho_o = (a+b)/2 \gg (b-a)$ such that **B** in the interior can be assumed to be constant. Find the inductance.

Ans. $L = \dfrac{\mu_o N^2 (b-a)^2}{4(a+b)}$ [H].

Exercise 5.19
One of two turns in a loop C_1 is on top of the other, while the two turns are at right angles in another loop C_2. Compare the inductances of C_1 and C_2, having the same radius.
Ans. $L_1 > L_2$.

Exercise 5.20
Find the mutual inductance between a straight wire along the z-axis and a circular loop residing in the xy-plane.
Ans. $M_{12} = 0$.

Review Questions with Hints

RQ 5.40 What is meant by magnetic flux linkage? [Eq.(5-99)]
RQ 5.41 Define mutual and self inductances. [Eqs.(5-100)(5-102)]
RQ 5.42 Distinguish between the external and internal inductances of a coaxial cable. [Eq.(5-110)]

5.9 Magnetic Energy

As the electric energy density is defined as $w_e = \frac{1}{2}\varepsilon E^2$, the magnetic energy density is defined as $w_m = \frac{1}{2}\mu H^2$, which is the subject of the present section. In Chapter 3, to derive the expression for w_e, we started with a distribution of charges, and obtained the potential energy by computing the work done in moving the individual charges from infinity to the predetermined points in space. However, the magnetic energy of a conducting system cannot be obtained by following the procedure used for the electric energy. This is because there is no magnetic equivalence to the discrete electric charge. Seeing that a steady current flowing in a wire generates a static magnetic field, one may be tempted to compute the work done in bringing current elements from infinity to the prearranged positions. To show that this does not work, let us consider the case in which two parallel conducting wires carry a steady current in the opposite directions. As the wires are brought closer together, work must be done against the magnetic force exerted on the wire. Meanwhile, the magnetic field between two wires increases, which in turn induces a voltage on the wire to reduce the current and thus oppose the change in the magnetic field, in accordance with Faraday's electromagnetic induction. Accordingly, additional work must be done by the current source to maintain the current in the wire at a constant level. As a consequence, the energy stored in the magnetic field is the sum of the work done by us and done by the current source.

5.9.1 Magnetic Energy in an Inductor

The inductor is a conducting device, which has a self-inductance and stores energy in the magnetic field. We know from circuit theory that when an *ac*-current $\mathscr{i}$ flows in an inductor, the voltage across it is given by $\mathscr{v} = L d\mathscr{i} / dt$. From now on, time-varying quantities are denoted by letters in a script font. Whereas the $\mathscr{v} - \mathscr{i}$ relation for the inductor can be derived from Faraday's law later in Chapter 6, we just use the relation to obtain the expression for the magnetic energy in the present section.

Let us start by considering a conducting loop with an inductance L. As we increase the current $\mathscr{i}$ in the loop from zero to a constant value I, the current induces not only the magnetic flux linking the loop, but also an opposing voltage $\mathscr{v}$ in the loop according to Faraday's law of induction. The product of the voltage and current, $\mathscr{v}\mathscr{i}$, represents the power delivered by the current source, and is equal to the energy stored in the magnetic field per unit time. The total magnetic energy is thus obtained as

$$W_m = \int \mathscr{v}\mathscr{i}\, dt = \int \mathscr{i}\, L \frac{d\mathscr{i}}{dt} dt = L \int_{\mathscr{i}=0}^{\mathscr{i}=I} \mathscr{i}\, d\mathscr{i} \tag{5-114}$$

From Eq. (5-114), the magnetic energy stored in an inductor is

$$\boxed{W_m = \frac{1}{2} L I^2} \qquad \text{[J]} \tag{5-115}$$

Upon substituting the inductance $L = \Lambda / I = N\Phi / I$ into Eq. (5-115), the magnetic energy of an inductor is

$$\boxed{W_m = \frac{1}{2} \Lambda I = \frac{1}{2} N I \Phi} \qquad \text{[J]} \tag{5-116}$$

where I is the steady current in the loop, N is the number of turns in the loop, Φ is the magnetic flux linking a single turn, and Λ is the magnetic flux linkage with the loop.

Next, we consider two neighboring loops C_1 and C_2 carrying steady currents I_1 and I_2, respectively, as illustrated in Fig. 5.34. The magnetic flux linking with each loop is as follows:

$$\Lambda_1 = N_1 \left(\Phi_{11} + \Phi_{21} \right) \tag{5-117a}$$

$$\Lambda_2 = N_2 \left(\Phi_{12} + \Phi_{22} \right) \tag{5-117b}$$

Here, N_1 and N_2 are the numbers of turns in C_1 and C_2, respectively. In the above equations, Φ_{12} is the magnetic flux produced by the current I_1 flowing in the loop C_1, which links with the loop C_2, whereas Φ_{21} is the magnetic flux produced by I_2 in C_2, which links with C_1. Similarly, Φ_{11} (or Φ_{22}) is the magnetic flux due to the current I_1 (or I_2) flowing in C_1 (or C_2), which links with C_1 (or C_2). Inserting Eq. (5-117) into Eq. (5-116), the magnetic energy stored in the two-loop system is

$$W_m = \frac{1}{2}\left(\Lambda_1 I_1 + \Lambda_2 I_2\right) = \frac{1}{2}N_1 I_1\left(\Phi_{11} + \Phi_{21}\right) + \frac{1}{2}N_2 I_2\left(\Phi_{12} + \Phi_{22}\right)$$

With the help of the self-inductances $L_1 = N_1\Phi_{11}/I_1$ and $L_2 = N_2\Phi_{22}/I_2$, the mutual inductances $M_{12} = N_2\Phi_{12}/I_1$ and $M_{21} = N_1\Phi_{21}/I_2$, and the relation $M_{12} = M_{21}$, the magnetic energy stored in the two-loop system is

$$\boxed{W_m = \frac{1}{2}L_1 I_1^2 + \frac{1}{2}L_2 I_2^2 \pm M_{12} I_1 I_2} \qquad \text{[J]} \tag{5-118}$$

The plus sign is for the case when I_1 and I_2 flow in the same direction, whereas the minus sign is for the case when the two currents flow in the opposite directions.

To examine the physical meaning underlying the terms on the right-hand side of Eq. (5-118), we consider two loops with no current, $i_1 = i_2 = 0$, at time $t = 0$. While maintaining i_1 at zero in the loop C_1, the current i_2 is increased from zero to I_2 in the loop C_2. In this case, the work done by the current source, connected to C_2, is given by the second term on the right-hand side of Eq. (5-118); no work is done in the loop C_1 because $i_1 = 0$. Next, while maintaining the current in C_2 at I_2, the current i_1 is increased from zero to I_1 in the loop C_1. The work done by the current source connected to C_1 is given by the first term on the right-hand side of Eq. (5-118). In this case, however, as i_1 is increased from zero to I_1, the mutual flux Φ_{12} varies with time and induces a voltage v_{12} in the loop C_2 in such a way as to reduce the current in the loop C_2, opposing the change in the magnetic flux linking with C_2. Additional work must be done by the current source connected to the loop C_2 to maintain the current at I_2. The additional work done is computed by use of the relation $v_{12} = M_{12} di_1/dt$ as

$$W_{12} = \int v_{12} I_2 dt = M_{12} I_2 \int_{i_1=0}^{i_1=I_1} di_1 = M_{12} I_1 I_2 \tag{5-119}$$

We now see that the third term on the right-hand side of Eq. (5-118) just represents the additional work done as given in Eq. (5-119).

If I_1 and I_2 flow in the directions opposite to each other, the mutual flux Φ_{12} tends to reduce the flux linkage with C_2. In this case, v_{12} is induced in C_2 in such a way as to increase the current in C_2, and thus the additional work done is negative($W_{12} < 0$).

Even if we started our calculation of W_m by increasing i_2 from zero to I_2, while maintaining $i_1 = 0$ in C_1, we would have the same result as Eq. (5-118), except that M_{12} is replaced with M_{21}. The magnetic energy of the two-loop system is the same regardless of whether we start with the current in C_1 or that in C_2. The relation $M_{12} = M_{21}$ given in Eq. (5-101) is therefore verified.

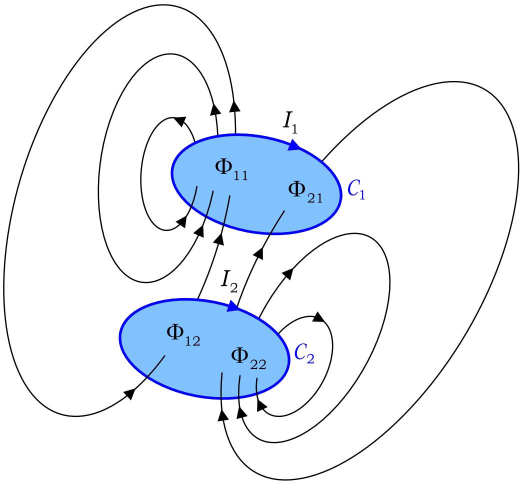


Fig. 5.34 Two neighboring loops with currents flowing in the same direction.

5.9.2 Magnetic Energy in Terms of Magnetic Field

It is more convenient to express the magnetic energy in terms of the field quantities, **B** and **H**, as the electrostatic energy is expressed in terms of **D** and **E** in Chapter 3. Let us consider a rectangular loop carrying a steady current I in the counterclockwise direction in the $z = 0$ plane, as shown in Fig. 5.35, which is designated as C'. It is obvious that the magnetic field lines generated by I always pass through the surface S' bounded by C', and close upon themselves. We assume a simple case in which the current-carrying loop is surrounded by a homogeneous, linear, and isotropic medium. Another closed loop C shown in Fig. 5.35 is a hypothetical loop, coincident with one of the magnetic field lines, which will be used as an Amperian path. Applying Ampere's circuital law to the loop C we have

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = NI \tag{5-120}$$

where N is the number of turns in C', and $d\mathbf{l}$ is the differential length vector on the loop C, which is always parallel to the direction of $\mathbf{H}$ and $\mathbf{B}$ because C represents one of the magnetic field lines.

The total magnetic flux due to I in C' is given by the surface integral the magnetic flux density $\mathbf{B}'$ across the surface S':

$$\Phi = \int_{S'} \mathbf{B}' \cdot d\mathbf{s}'$$

Inserting the above equation into Eq. (5-116), the magnetic energy due to the current I flowing in the loop C' is expressed as

$$W_m = \frac{1}{2} NI \int_{S'} \mathbf{B}' \cdot d\mathbf{s}' \tag{5-121}$$

where S' is the surface bounded by C' in the $z = 0$ plane, and $\mathbf{B}'$ is the magnetic flux density on S'. Upon substituting Eq. (5-120) into Eq. (5-121), we obtain

$$\begin{aligned} W_m &= \frac{1}{2}\left[\oint_C \mathbf{H} \cdot d\mathbf{l}\right]\left[\int_{S'} \mathbf{B}' \cdot d\mathbf{s}'\right] = \int_{S'} \frac{1}{2}\left[\oint_C (\mathbf{B}' \cdot d\mathbf{s}')\, \mathbf{H} \cdot d\mathbf{l}\right] \\ &\equiv \int_{S'} dW \end{aligned} \tag{5-122}$$

From the fact that S' and C, and thus $d\mathbf{s}'$ and $d\mathbf{l}$, are totally independent of each other, the closed line integral is taken inside the surface integral in Eq. (5-122). Notice that dW is to denote the inside of the surface integral in Eq. (5-122). Let us now consider a ring as shown in Fig. 5.35, which is formed by a bundle of the magnetic field lines that pass through the differential area ds' in S'. Thus the edges and sides of the ring correspond to the magnetic field line. From the law of conservation of the magnetic flux we immediately recognize the following:

$$d\Phi = \mathbf{B}' \cdot d\mathbf{s}' = \mathbf{B} \cdot d\mathbf{s} \tag{5-123}$$

Although $d\mathbf{s}'$ is always oriented along the z-axis on S', the magnitude and direction of $d\mathbf{s}$ may vary with position in the ring. The magnitude $|d\mathbf{s}|$ (or ds) represents the cross sectional area of the ring, whereas $|d\mathbf{s}'|$ (or ds') is the cross sectional area of the ring measured on S' (or the $z = 0$ plane). Inserting Eq. (5-123) into Eq. (5-122) we have

$$dW = \frac{1}{2}\left[\oint_C (\mathbf{B} \cdot d\mathbf{s})\, \mathbf{H} \cdot d\mathbf{l}\right]$$

Noting that the vectors $\mathbf{H}$, $d\mathbf{l}$, and $d\mathbf{s}$ are all tangential to C, we rewrite the above equation as

$$dW = \frac{1}{2}\left[\oint_C (\mathbf{B} \cdot \mathbf{H})\, d\mathbf{s} \cdot d\mathbf{l}\right] = \oint_C \frac{1}{2}(\mathbf{B} \cdot \mathbf{H})\, ds\, dl \tag{5-124}$$

Identifying $dsdl$ in Eq. (5-124) with the differential volume in the ring, we note that the right-hand side of Eq. (5-124) is the integral of $\frac{1}{2}\mathbf{B}\bullet\mathbf{H}$ over the volume of the ring. Combination of Eq. (5-122) with Eq. (5-124) leads to the conclusion that the total magnetic energy W_m is equal to the integral of $\frac{1}{2}\mathbf{B}\bullet\mathbf{H}$ throughout the volume occupied by the magnetic field lines. Namely,

$$\boxed{W_m = \frac{1}{2}\int_{\mathcal{V}} \mathbf{B}\bullet\mathbf{H}\,dv} \qquad \text{[J]} \qquad (5\text{-}125)$$

In a simple medium in which μ is a constant, independent of **B**, Eq. (5-125) becomes

$$\boxed{W_m = \frac{\mu}{2}\int_{\mathcal{V}} H^2 dv = \frac{1}{2\mu}\int_{\mathcal{V}} B^2 dv} \qquad \text{[J]} \qquad (5\text{-}126)$$

In view of Eq. (5-126), the magnetic energy density is defined as $\frac{1}{2}\mathbf{B}\bullet\mathbf{H}$, which has the unit of the joule per cubic meter.

If the magnetic energy of a conducting device is calculated from Eq. (5-126), we can obtain the inductance from Eq. (5-115) as

$$\boxed{L = \frac{2W_m}{I^2}} \qquad \text{[J]} \qquad (5\text{-}127)$$

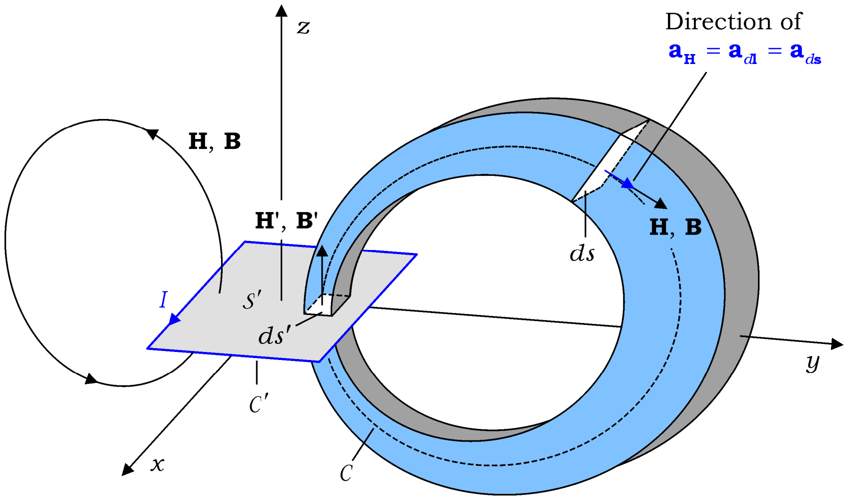


Fig. 5.35 A bundle of magnetic field lines forming a ring.

Example 5-19

Referring to the coaxial cable as shown in Fig. 5.32, determine the inductance per unit length of the cable from the magnetic energy stored in the cable.

Solution

In the region $a < \rho < b$, inserting Eq. (5-106b) into Eq. (5-126), the magnetic energy is

$$W_{m1} = \frac{1}{2\mu_o}\int_{\rho=a}^{\rho=b}\int_{\phi=0}^{\phi=2\pi}\left[\frac{\mu_o I}{2\pi\rho}\right]^2 \rho d\rho d\phi = \frac{\mu_o I^2}{4\pi}\ln\frac{b}{a} \tag{5-128a}$$

In the region $0 \le \rho \le a$, inserting Eq. (5-106c) into Eq. (5-126), the magnetic energy is

$$W_{m2} = \frac{1}{2\mu_o}\int_{\rho=0}^{\rho=a}\int_{\phi=0}^{\phi=2\pi}\left[\frac{\mu_o \rho I}{2\pi a^2}\right]^2 \rho d\rho d\phi = \frac{\mu_o I^2}{16\pi} \tag{5-128b}$$

The inductance per unit length of the cable is obtained by inserting Eq. (5-128) into Eq. (5-127):

$$L = \frac{2}{I^2}\left(W_{m1} + W_{m2}\right) = \frac{\mu_o}{2\pi}\ln\frac{b}{a} + \frac{\mu_o}{8\pi} \qquad [\text{H/m}] \tag{5-129}$$

We have the same result as Eq. (5-110).

Example 5-20

The coaxial cable shown in Fig. 5.32 carries a current I [A], which is non-uniform in the inner conductor such that $\mathbf{J} = 2I\rho^2/(\pi a^4)\,\mathbf{a}_z$ [A/m^2] in cylindrical coordinates. Determine the internal inductance per unit length from (a) magnetic flux linkage, and (b) magnetic energy.

Solution

In the region $0 \le \rho \le a$, applying Ampere's circuital law to an Amperian path of a radius ρ_1 ($0 \le \rho_1 \le a$), we obtain

$$2\pi\rho_1 H_\phi = \int_{\rho=0}^{\rho=\rho_1}\int_{\phi=0}^{\phi=2\pi}\frac{2I\rho^2}{\pi a^4}\rho d\rho d\phi = I\frac{\rho_1^4}{a^4}$$

Omitting 1 in ρ for generalization, we have

$$H_\phi = \frac{\rho^3 I}{2\pi a^4}\,. \tag{5-130}$$

(a) Magnetic flux confined to the cylindrical shell of a radius ρ_1, thickness $d\rho$, and length ℓ as shown in Fig. 5.32(b) is

$$d\Phi = \mu_o H_\phi \ell d\rho = \frac{\mu_o \rho_1^3 I}{2\pi a^4}\ell d\rho$$

The differential flux $d\Phi$ is produced by a partial current I_1 flowing in a cylinder of radius ρ_1. To find I_1, we integrate $\mathbf{J}$ over the disc of radius ρ_1 in the cross section of the inner conductor:

$$I_1 = I\frac{\rho_1^4}{a^4}$$

The magnetic flux linking with the cylindrical shell, per unit length, is given by the product of $d\Phi/\ell$ and I_1/I, i.e.,

$$d\Lambda = \frac{d\Phi}{\ell}\frac{I_1}{I} = \frac{\mu_o\rho_1^7 I}{2\pi a^8}d\rho$$

Total flux linkage with the whole inner conductor, per unit length, is

$$\Lambda = \int d\Lambda = \int_{\rho=0}^{\rho=a}\frac{\mu_o\rho^7 I}{2\pi a^8}d\rho = \frac{\mu_o I}{16\pi}$$

The internal inductance per unit length is therefore

$$L = \frac{\Lambda}{I} = \frac{\mu_o}{16\pi} \qquad \text{[H/m]}. \qquad (5\text{-}131)$$

(b) The magnetic energy stored per unit length of the inner conductor is obtained by inserting Eq. (5-130) into Eq. (5-126):

$$W_m = \frac{\mu_o}{2}\int_{\rho=0}^{\rho=a}\int_{\phi=0}^{\phi=2\pi}H^2\rho d\rho d\phi = \frac{\mu_o}{2}2\pi\left(\frac{I}{2\pi a^4}\right)^2\int_{\rho=0}^{\rho=a}\rho^7 d\rho$$
$$= \frac{\mu_o I^2}{32\pi}\text{[J/m]}$$

The internal inductance per unit length is therefore,

$$L = \frac{2W_m}{I^2} = \frac{\mu_o}{16\pi} \qquad \text{[H/m]} \qquad (5\text{-}132)$$

The result is the same as that in part (a).

Exercise 5.21
Two solenoids have the same number of turns on air cores of the same radius but different lengths, ℓ and 2ℓ. Find the ratio between the stored magnetic energies.

Ans. $W_m^{\ell}/W_m^{2\ell} = 2$.

Review Questions with Hints

RQ 5.43 Express magnetic energy in terms of inductance. [Eq.(5-115)]
RQ 5.44 Express magnetic energy in terms of flux linkage. [Eq.(5-116)]
RQ 5.45 Express magnetic energy in terms of field quantities. [Eq.(5-125)]
RQ 5.46 Write an expression for the magnetic energy stored in a two-loop system. [Eq.(5-118)]

5.10 Magnetic Force and Torque

Lorentz force equation states that a charge q moving with a velocity $\mathbf{v}$ in a magnetic flux density $\mathbf{B}$ experiences a magnetic force given by $\mathbf{F}_m = q\mathbf{v}\times\mathbf{B}$. Since a current-carrying wire unavoidably involves motion of free electrons, the wire will experience a magnetic force even if it is fixed in position in a magnetic flux density.

5.10.1 Magnetic Force on a Current-Carrying Conductor

Consider a conducting wire of a uniform cross section $\mathcal{S}\,[\mathrm{m}^2]$, which contains conduction electrons of a uniform number density $n\,[\mathrm{m}^{-3}]$, moving with a velocity $\mathbf{v}$[m/s]. A segment of the wire of a differential length $|d\mathbf{l}|$ contains the charge of $en\mathcal{S}|d\mathbf{l}|$, where the electron charge $e = -1.6\times 10^{-19}\,[\mathrm{C}]$. In the presence of a magnetic flux density $\mathbf{B}$, the magnetic force exerted on the segment of the wire is

$$d\mathbf{F}_m = en\mathcal{S}|d\mathbf{l}|\,\mathbf{v}\times\mathbf{B} \tag{5-133a}$$

$$= -en\mathcal{S}|\mathbf{v}|\,d\mathbf{l}\times\mathbf{B} \tag{5-133b}$$

It should be noted that $d\mathbf{l}$ is along the direction of flow of the current in the wire, which is opposite to the direction of flow of the electrons, or $\mathbf{v}$. With the aid of the current density $\mathbf{J} = en\mathbf{v}$ and the total current $I = |\mathbf{J}|\mathcal{S}$, Eqs. (5-133a) and (5-133b) are rewritten as

$$d\mathbf{F}_m = \mathbf{J}\times\mathbf{B}\,dv \tag{5-134a}$$

$$d\mathbf{F}_m = I\,d\mathbf{l}\times\mathbf{B} \tag{5-134b}$$

Here, the differential volume $dv = \mathcal{S}|d\mathbf{l}|$. In view of Eq. (5-134), the total magnetic force exerted on a current-carrying wire can be expressed in terms of either the current density $\mathbf{J}$ or the total current I as

$$\boxed{\mathbf{F}_m = \int_{\mathcal{V}} \mathbf{J}\times\mathbf{B}\,dv} \quad [\mathrm{N}] \tag{5-135a}$$

$$\boxed{\mathbf{F}_m = I\oint_{\mathcal{C}} d\mathbf{l}\times\mathbf{B}} \quad [\mathrm{N}] \tag{5-135b}$$

In Eq. (5-135b), the differential length vector $d\mathbf{l}$ is along the direction of flow of the current I, which is the direction of travel on the loop $\mathcal{C}$. Since current-carrying wires usually form closed lines, the line integral for the magnetic force is conducted around a closed path, in general. It is important to remember that the conducting wire carrying the current I is assumed to be infinitely thin in Eq. (5-135b).

We now extend our discussion of the magnetic force to two parallel conducting wires carrying currents I_1 and I_2. The first wire experiences a magnetic force in the magnetic field that is due to the current flowing in the second wire, and vice versa. The magnetic flux density, at a point on the first wire C_1, which is produced by I_2 flowing in the second wire C_2, is denoted by $\mathbf{B}_{1-2}$. From the Biot-Savart law given in Eq. (5-9) we write

$$\mathbf{B}_{1-2} = \frac{\mu_o I_2}{4\pi} \oint_{C_2} \frac{d\mathbf{l}_2 \times \mathbf{a}_{\mathcal{R}}}{\mathcal{R}^2} \tag{5-136}$$

where $\mathbf{a}_{\mathcal{R}}$ is the unit vector of the distance vector $\mathcal{R} = \mathbf{r}_1 - \mathbf{r}_2$. Next, from Eq. (5-135b), the magnetic force on the wire C_1, carrying the current I_1 and residing in the magnetic flux density $\mathbf{B}_{1-2}$, is written as

$$\mathbf{F}_1 = I_1 \oint_{C_1} d\mathbf{l}_1 \times \mathbf{B}_{1-2} \tag{5-137}$$

Inserting Eq. (5-136) into Eq. (5-137), we have

$$\boxed{\mathbf{F}_1 = \frac{\mu_o}{4\pi} I_1 I_2 \oint_{C_1} \oint_{C_2} \frac{d\mathbf{l}_1 \times (d\mathbf{l}_2 \times \mathbf{a}_{\mathcal{R}})}{\mathcal{R}^2}} \quad \text{[N]} \tag{5-138}$$

It is not easy to directly evaluate the double line integral in Eq. (5-138). In many cases, it is more convenient to break the integral into two parts, as Eq. (5-136) and Eq. (5-137), and evaluate them separately.

Example 5-21

A very long transmission line consists of two parallel wires separated by a distance d[m] in free space. A steady current of I [A] flows in the opposite directions as shown in Fig. 5.36. Find the magnetic force per unit length of the line.

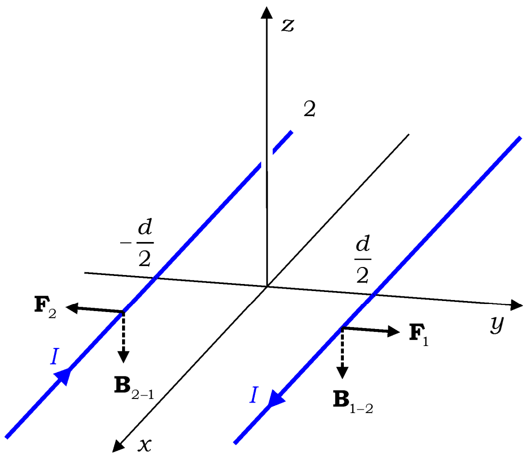


Fig. 5.36 Two parallel wires carrying a current in the opposite directions.

Solution

From Ampere's circuital law, **B** on wire 1 due to current I in wire 2 is

$$\mathbf{B}_{1-2} = -\frac{\mu_o I}{2\pi d}\mathbf{a}_z \qquad (5\text{-}139)$$

Differential length vector on wire 1 is

$$d\mathbf{l}_1 = dx\,\mathbf{a}_x$$

From Eq. (5-137), the magnetic force per unit length of wire 1 is

$$\mathbf{F}_1 = I\oint_{C_1} d\mathbf{l}_1 \times \mathbf{B}_{1-2} = I\int_{x=0}^{x=1} dx\,\mathbf{a}_x \times \left(-\frac{\mu_o I}{2\pi d}\mathbf{a}_z\right)$$
$$= \frac{\mu_o I^2}{2\pi d}\mathbf{a}_y$$

In view of the opposite currents in the wires, the magnetic forces acting on the wires are repulsive. Thus the magnetic force on wire 2 is

$$\mathbf{F}_2 = -\mathbf{F}_1 = -\frac{\mu_o I^2}{2\pi d}\mathbf{a}_y\,.$$

Exercise 5.22

A circular loop, centered at the origin in the xy-plane, carries a steady current in the presence of a uniform field $\mathbf{B} = B_o\,\mathbf{a}_z$. Find the net force on the loop.

Ans. Zero.

5.10.2 Magnetic Force Involved in a Virtual Work

In the previous section, Lorentz force equation was used to determine the magnetic force exerted on a charge in motion and a current-carrying conductor placed in a magnetic flux density. We now introduce an alternative method of finding the magnetic force, which is called the method of virtual displacement. This method is based on the fundamental relation between energy and force; that is, the energy is equal to the line integral of the force. Consider an electromagnet as shown in Fig. 5.37 in which a current flowing in the coil induces a magnetic flux density **B** in the core, gap, and armature. If the fringing effects at the edges are ignored, the magnetic flux density in the gap is the same as that in the core. This is justified by the boundary condition asserting that the normal component of **B** is continuous across the interface. Suppose we apply an external force **F** to the armature, and move it downward by a small distance $d\ell$. The work done in moving the armature is stored as the magnetic energy in the increase of the gap volume. In view of these discussions, we can determine the attractive force on the armature, $\mathbf{F}_m$, by assuming a virtual displacement of the armature in the upward direction, and by calculating the magnetic energy stored in the decrease of the gap volume.

If we allow a virtual displacement of the armature in the upward direction by a small distance $|d\mathbf{l}|$, the displacement is such as if the magnetic force $\mathbf{F}_m$ moved the armature in the upward direction by a vector differential length $d\mathbf{l}$. Since the work is done at the expense of the magnetic energy W_m, we write

$$\mathbf{F}_m \cdot d\mathbf{l} = -dW_m \tag{5-140}$$

The right-hand side of Eq. (5-140) represents the decrement in the magnetic energy stored in the gap. Expressing the differential of a scalar quantity W_m in terms of the gradient of W_m, Eq. (5-140) becomes

$$\mathbf{F}_m \cdot d\mathbf{l} = -(\nabla W_m) \cdot d\mathbf{l} \tag{5-141}$$

From Eq. (5-140) we obtain

$$\boxed{\mathbf{F}_m = -\nabla W_m} \qquad \text{[N].} \tag{5-142}$$

The attractive force on the armature is equal to the negative gradient of the magnetic energy stored in the gap.

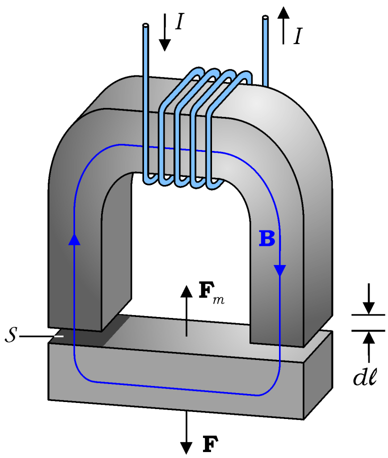


Fig. 5.37 An electromagnet and an armature.

Example 5-22
Referring to the electromagnet shown in Fig. 5.37, in which the current I in the coil of N turns produces a uniform magnetic flux density $\mathbf{B}$ in the core, armature, and the gap with a cross section S, express $\mathbf{F}_m$ on the armature in terms of $\mathbf{B}$ and S.

Solution

Allowing a virtual displacement $d\ell$ of the armature in the upward direction, the magnetic energy in the gap is reduced by

$$dW_m = -2\left[\frac{B^2}{2\mu_o} \mathcal{S}\, d\ell\right] \tag{5-143}$$

Inserting Eq. (5-143) into Eq. (5-140), and noting that $d\mathbf{l} = d\ell\, \mathbf{a}_z$, where $\mathbf{a}_z$ is along the upward direction, we obtain

$$\mathbf{F}_m = \frac{B^2}{\mu_o} \mathcal{S}\, \mathbf{a}_z \qquad \text{[N]} \tag{5-144}$$

Exercise 5.23

With reference to the electromagnet in Fig. 5.37, show that the tractive pressure on the surface $\mathcal{S}$ is equal to the magnetic energy density in the gap.

Ans. $P = |\mathbf{F}_m| / 2\mathcal{S} = B^2 / 2\mu_o$ [N/m^2].

5.10.3 Magnetic Torque

We can make a rigid body rotate about its pivot axis by applying a force on the body. Referring to Fig. 5.38, the moment arm is defined as a distance vector $\mathbf{r}$ that is drawn from the pivot axis to the point of application of the force. The angular acceleration of the body depends on the length of the moment arm and the force normal to the moment arm. The torque is defined as the cross product of the moment arm and the applied force such that

$$\boxed{\mathbf{T} = \mathbf{r} \times \mathbf{F}} \qquad [\text{N} \cdot \text{m}] \tag{5-145}$$

The torque has the unit of the newton-meter.

Consider a simple case shown in Fig. 5.38, in which a solid bar extends from the origin to point p in the xy-plane. The bar will pivot about the z-axis when a force $\mathbf{F}$, lying in the xy-plane, is applied to point p. From Eq. (5-145), the magnitude of the torque is obtained as

$$T = rF_\phi = dF \tag{5-146}$$

where r is the length of the moment arm, F_ϕ is the component of $\mathbf{F}$ normal to $\mathbf{r}$, and d is the *effective moment arm*, representing the perpendicular distance from the origin to the vector $\mathbf{F}$. The direction of the torque obeys the right-hand rule: the right thumb points in the direction of the torque when the fingers follow the direction of rotation of the body.

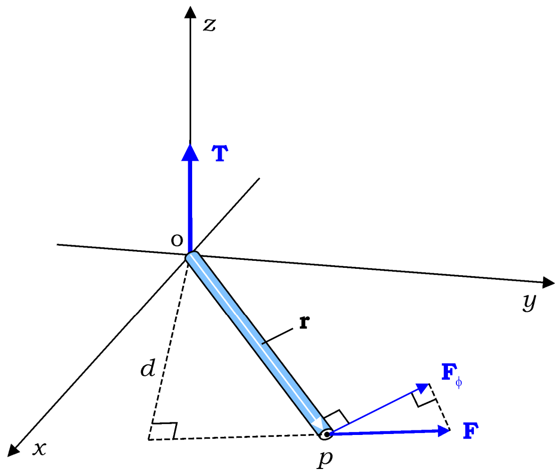


Fig. 5.38 A torque **T** on a rigid bar.

An externally applied magnetic field may exert a torque on a current-carrying loop, and cause it to rotate about its pivot axis until the magnetic dipole moment of the loop is aligned with the external field. Let us consider a small rectangular loop carrying a steady current I as shown in Fig. 5.39. The loop is constrained to rotate about the x-axis in the presence of a uniform magnetic flux density **B** applied along the y-axis. At time $t = 0$, the magnetic dipole moment of the loop, **m**, is at an angle α to the y-axis. The field **B** exerts magnetic forces on the four sides of the loop with a current I, according to the Lorentz force equation. The magnetic forces acting on the top and bottom sides, $\mathbf{F}_1$ and $\mathbf{F}_2$, exert a torque on the loop, i.e.,

$$\mathbf{T} = -d\left(F_1 + F_2\right)\mathbf{a}_x \qquad (5\text{-}147)$$

where d is the effective moment arm for the top and bottom sides of the loop. In contrast, the magnetic forces on the left and right sides, $\mathbf{F}_3$ and $\mathbf{F}_4$, are either parallel or anti-parallel to the pivot axis, and thus contribute nothing to the torque. From Eq. (5-135b), the magnitudes of $\mathbf{F}_1$ and $\mathbf{F}_2$ are

$$F_1 = F_2 = IwB \qquad (5\text{-}148)$$

where w is the width of the rectangular loop. Inserting Eq. (5-148) and the relation $d = (h/2)\sin\alpha$ into Eq. (5-147), the torque exerted on the loop is

$$\mathbf{T} = -hwIB\sin\alpha\,\mathbf{a}_x = -mB\sin\alpha\,\mathbf{a}_x \qquad (5\text{-}149)$$

where h is the height of the loop. In Eq. (5-149), m is the magnitude of the magnetic dipole moment defined by $\mathbf{m} = hwI\,\mathbf{a}_n$: the unit vector $\mathbf{a}_n$ is normal to the loop surface, directed along the right thumb when the fingers follow I in the loop. In the presence of **B**, in vector notation, the torque exerted on a magnetic dipole moment **m** is

$$\boxed{\mathbf{T} = \mathbf{m} \times \mathbf{B}} \qquad [\mathrm{N \cdot m}] \qquad (5\text{-}150)$$

The torque tends to rotate the current loop in such a way that its magnetic dipole moment **m** is aligned with the applied **B**.

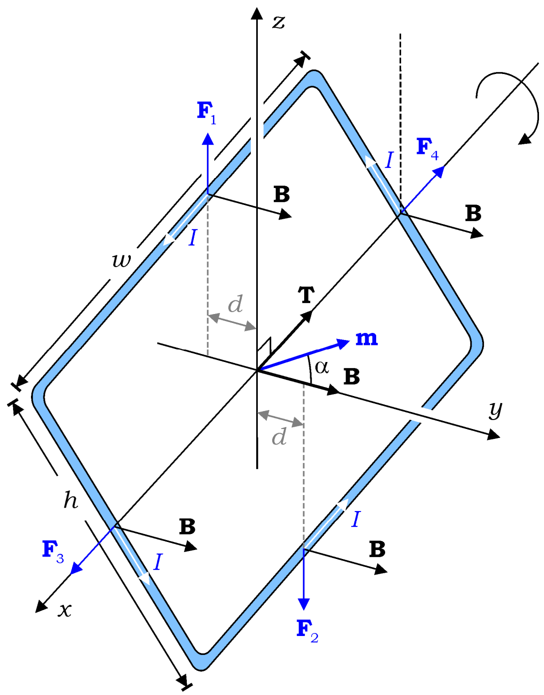


Fig. 5.39 A rectangular loop carrying a current in a uniform **B**.

Example 5-23
A small circular loop of radius a is centered at the origin in the xy-plane, carrying a current I in the counterclockwise direction in a uniform magnetic flux density $\mathbf{B} = B_y\mathbf{a}_y + B_z\mathbf{a}_z$, as shown in Fig. 5.40. Find the torque on the loop about the origin.

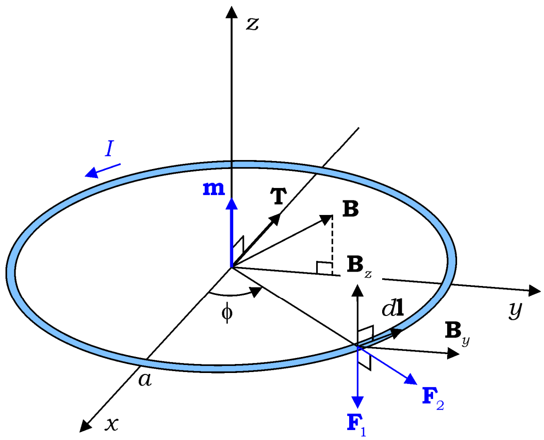


Fig. 5.40 A circular loop carrying a current in a uniform **B**.

Solution

From Eq. (5-135b), the differential magnetic force on a differential current element $I d\mathbf{l} = I a d\phi\, \mathbf{a}_\phi$ is

$$\begin{aligned} d\mathbf{F} &= Iad\phi\, \mathbf{a}_\phi \times (B_y \mathbf{a}_y + B_z \mathbf{a}_z) \\ &= -aIB_y \sin\phi\, d\phi\, \mathbf{a}_z + aIB_z d\phi\, \mathbf{a}_\rho \\ &\equiv \mathbf{F}_1 + \mathbf{F}_2 \end{aligned}$$

The differentia torque about the origin due to $Id\mathbf{l}$ is

$$\begin{aligned} d\mathbf{T} = \mathbf{r} \times (\mathbf{F}_1 + \mathbf{F}_2) &= a\, \mathbf{a}_\rho \times (-aIB_y \sin\phi\, d\phi\, \mathbf{a}_z + aIB_z d\phi\, \mathbf{a}_\rho) \\ &= a^2 IB_y \sin\phi\, d\phi\, \mathbf{a}_\phi \end{aligned}$$

The net torque on the loop is

$$\mathbf{T} = \int d\mathbf{T} = a^2 IB_y \int_{\phi=0}^{\phi=2\pi} \mathbf{a}_\phi \sin\phi\, d\phi \tag{5-151}$$

The unit vector $\mathbf{a}_\phi$ cannot be taken outside the integral sign, because it is a function of ϕ. Using the relation $\mathbf{a}_\phi = -\sin\phi\, \mathbf{a}_x + \cos\phi\, \mathbf{a}_y$, Eq. (5-151) is rewritten as

$$\begin{aligned} \mathbf{T} &= a^2 IB_y \int_{\phi=0}^{\phi=2\pi} (-\sin^2\phi\, \mathbf{a}_x + \sin\phi \cos\phi\, \mathbf{a}_y) d\phi \\ &= -\pi a^2 IB_y \mathbf{a}_x \end{aligned}$$

With the aid of the magnetic dipole moment $\mathbf{m} = \pi a^2 I\, \mathbf{a}_z$, the torque on the loop is, in vector notation,

$$\mathbf{T} = \mathbf{m} \times \mathbf{B} \tag{5-152}$$

We see from Eqs. (5-150) and (5-152) that the torque on a current-carrying loop has the same form, in vector notation, regardless of the shape of the loop.

Exercise 5.24

Referring to the rectangular loop in Fig. 5.39, find the torque on the loop if it is constrained to pivot about an axis coincident with the bottom side of the loop.

Ans. $\mathbf{T} = \mathbf{m} \times \mathbf{B}$

Exercise 5.25

For a square and a circular loop with the same perimeter and current, placed separately in a uniform **B**, find the ratio between the torques on the two loops.

Ans. $T_{square} / T_{circle} = \pi / 4$

Review Questions with Hints

RQ 5.47 Express the magnetic force on a current-carrying wire placed in a magnetic flux density. [Eq.(5-135b)]

RQ 5.48 Express the relation between the magnetic force and energy from the point of view of virtual displacement. [Eq.(5-142)]

RQ 5.49 Express the torque exerted on a magnetic dipole placed in a uniform magnetic flux density. [Eq.(5-150)]

RQ 5.50 Does $\mathbf{T} = 0$ on a current-carrying loop mean $\mathbf{F}_m = 0$ at every point on the loop? [Fig.5.40]

RQ 5.51 Is the relation $\mathbf{T} = \mathbf{m} \times \mathbf{B}$ true even if **B** is not uniform? [Eq.(5-151)]

Problems

5-1 In a uniform magnetic flux density $\mathbf{B} = B_o \, \mathbf{a}_z$ in free space, a charge q is observed to move with a velocity $\mathbf{v} = a\,\mathbf{a}_x + b\,\mathbf{a}_y$ at time $t = 0$. Find an electric field **E** required to make the charge move along a straight line for $t \geq 0$.

5-2 A conductor with a rectangular cross section $h \times w$ [m^2] extends along the y-axis from $-\infty$ to ∞ as shown in Fig. 5.41. It carries a steady current I [A], uniformly distributed in the cross section, in the presence of a uniform magnetic flux density $\mathbf{B} = B_o \, \mathbf{a}_z$ [T].

(a) Find the drift velocity of free electrons of a uniform density n[m^{-3}].

(b) The magnetic force on the electrons establishes an electric field inside the conductor, which in turn offsets the magnetic force. Show that Hall voltage $V_H = IB_o \left(hn|e|\right)^{-1}$ [V], where e is the electron charge.

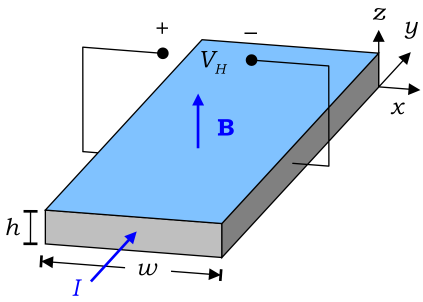


Fig. 5.41 Hall voltage (Problem 5-2).

5-3 A very long, filamentary, steady current I flows in the $+z$-direction, passing through a point $(0,1,0)$ in Cartesian coordinates as shown in Fig. 5.42. Find $\mathbf{H}$ everywhere.

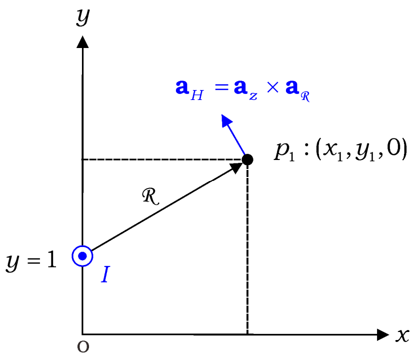


Fig. 5.42 A very long filamentary current (Problem 5-3).

5-4 Two very long filaments are both parallel to the z-axis, passing through points $p_1{:}(0,-1,0)$ and $p_2{:}(0,1,0)$ in Cartesian coordinates and carrying a current I in the $-\mathbf{a}_z$ and $+\mathbf{a}_z$ directions, respectively. Find $\mathbf{H}$ at points (a) $(0,2,0)$, and (b) $(2,0,0)$.

5-5 A conducting wire forms an equilateral triangle of side a in the xy-plane, with the center at the origin, as shown in Fig. 5.43. When a steady current I flows in the counterclockwise direction, find $\mathbf{H}$ at the origin.

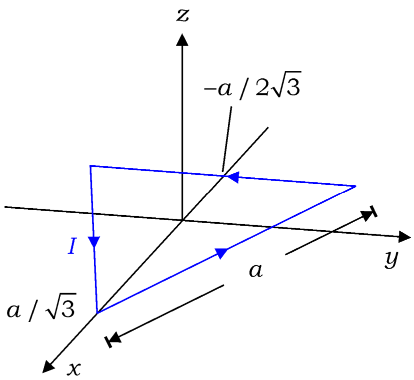


Fig. 5.43 A triangular loop with a current I (Problem 5-5).

5-6 A surface current of a density $\mathbf{J}_s = J_o\,\mathbf{a}_y$ is confined to a region defined by $z = 0$ and $-a \le x \le a$ as shown in Fig. 5.44. Find **H** at a point $p{:}(0, 0, b)$ by using

(a) the Biot-Savart law, and

(b) the equivalent line currents, flowing in infinitely long and narrow strips of a width dx, comprising $\mathbf{J}_s$.

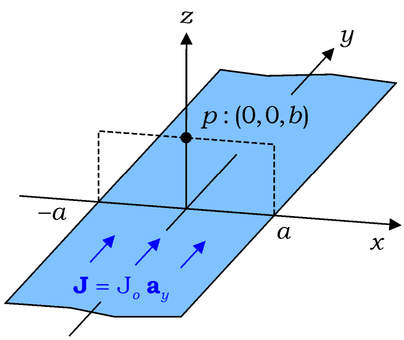


Fig. 5.44 A surface current density(Problem 5-6).

5-7 A surface current of a density $\mathbf{J}_s = (\pi/\rho)\,\mathbf{a}_\phi$ is confined to a region defined by $z = 0$ and $a \le \rho \le b$ as shown in Fig. 5.45. Find **H** at a point $(0,0,d)$ on the z-axis.

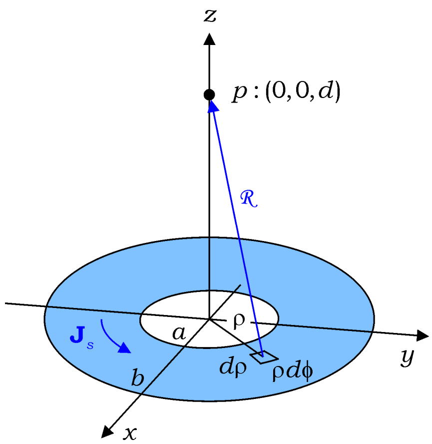


Fig. 5.45 A surface current density(Problem 5-7).

5-8 A surface current of a density $\mathbf{J}_s = 3\,\mathbf{a}_\phi$ [A/m^2] is confined to a region defined by $\rho = 0.4$[m] and $0 \le z \le 0.2$[m] as shown in Fig. 5.46. Find $\mathbf{H}$ at point (0,0,0.5[m]) on the z-axis.

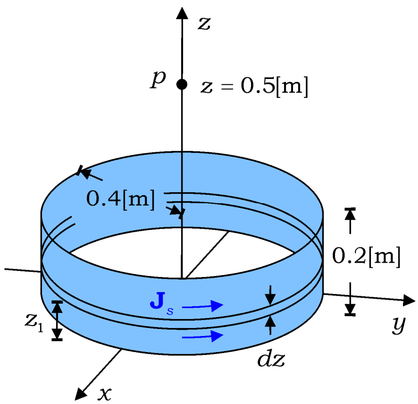


Fig. 5.46 A surface current (Problem 5-8).

5-9 A very long conducting wire of a radius a is oriented along the z-axis, and carries a steady current I [A] in the +z-direction, which is assumed to be uniform in the cross section.
(a) Find the current density $\mathbf{J}$.
(b) Find $\mathbf{H}$ inside the conductor.
(c) Show that $\nabla \times \mathbf{H} = \mathbf{J}$ in the conductor.

5-10 A volume current density, $\mathbf{J} = J_o\,\mathbf{a}_y$[A/m^2], forms an infinite slab of current in the region $-d \le z \le d$ and $-\infty \le (x, y) \le \infty$. From symmetry considerations, (a) show that $\mathbf{H}$ is independent of x and y. In view of the fact that the magnetic field intensity of an infinitely long line current is circumferential, (b) show that **H** only has the x-component: $\mathbf{H} = H_x(z)\mathbf{a}_x$. (c) Also show that if $\mathbf{H} = H_x(z)\mathbf{a}_x$ for $z > 0$, then $\mathbf{H} = -H_x(|z|)\mathbf{a}_x$ for $z < 0$. (d) Find **H** everywhere by using Ampere's circuital law.

5-11 A uniform surface current of a density $\mathbf{J}_s = J_o\,\mathbf{a}_y$ is infinite in extent in the xy-plane. Following the same procedure as was used for **H** in **Problem 5-10**, (a) show that $\mathbf{H} = H_o\,\mathbf{a}_x$ for $z > 0$ and $\mathbf{H} = -H_o\,\mathbf{a}_x$ for $z < 0$, where H_o is constant. (b) Find **H** everywhere.

5-12 A semi-infinite filament is normal to an infinitely large conducting sheet lying in the xy-plane as shown in Fig. 5.47. A steady current I flows in the $-z$-direction in the filament, and becomes a surface current density $\mathbf{J}_s$ flowing in the $\mathbf{a}_\rho$-direction on the conducting sheet, which is uniform in the $\mathbf{a}_\phi$-direction.

(a) Using the Biot-Savart law and symmetry, show that $\mathbf{J}_s$ produces **H** of the form $\mathbf{H} = H_o(\rho, z)\,\mathbf{a}_\phi$, at the most, in the region $z \geq 0$.

(b) Using Helmholtz's theorem, show that a trial solution $\mathbf{H} = -I\,\mathbf{a}_\phi/(2\pi\rho)$ is a unique solution in the region $z \geq 0$.

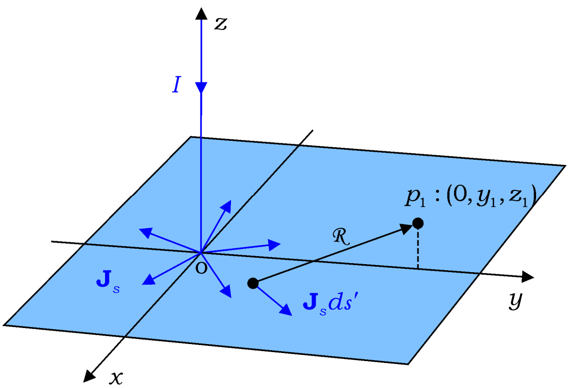


Fig. 5.47 A very long filament is normal to an infinite current sheet(Problem 5-12).

5-13 A perfectly conducting, hollow cylinder of radius a and height b is connected to two very long wires carrying a steady current I flowing in the $-z$-direction, as shown in Fig. 5.48. The surface current densities $\mathbf{J}_{s1}$, $\mathbf{J}_{s2}$, and $\mathbf{J}_{s3}$ are uniform in the ϕ-direction on the top, side, and bottom surfaces of the cylinder, respectively.

(a) Using the Biot-Savart law and symmetry, show that **H** is of the form $\mathbf{H} = H_o(\rho)\,\mathbf{a}_\phi$ in the xy-plane outside the cylinder.

(b) Using Helmholtz's theorem, show that a trial solution $\mathbf{H} = -I\,\mathbf{a}_\phi/(2\pi\rho)$ is a unique solution outside the cylinder.

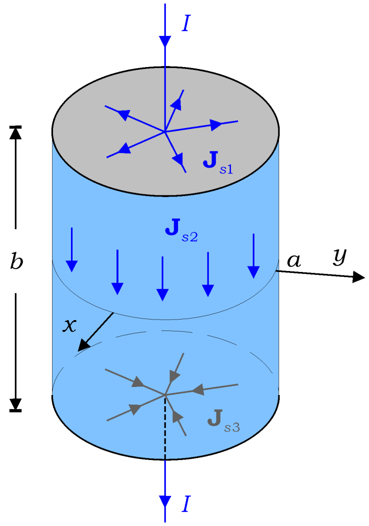


Fig. 5.48 A hollow cylinder connected to two very long wires(Problem 5-13).

5-14 A finite filament extends from $z=-a$ to $z=a$, and carries a steady current I in the $+z$-direction.
(a) Find $\mathbf{A}$ at point $p:(\rho,\phi,z)$ in cylindrical coordinates.
(b) Determine $\mathbf{B}$ from $\mathbf{A}$.

5-15 A very long solenoid is centered on the z-axis, having N turns per unit length closely wound on an air core of radius a, and carries a current I in the ϕ-direction.
(a) Using Eq. (5-39) and symmetry, show that $\mathbf{A}$ is of the form $\mathbf{A}=A_\phi(\rho)\,\mathbf{a}_\phi$.
(b) Determine $A_\phi(\rho)$ by making use of the fact that $\mathbf{B}$ is constant in the interior and zero in the exterior of the solenoid.

5-16 An infinitely long cylindrical sheet of a radius a is centered on the z-axis, carrying a surface current of a uniform density $\mathbf{J}_s = J_o\,\mathbf{a}_\phi$, as shown in Fig. 5.49. Prove that $\mathbf{H}=0$ at a point $(b,0,0)$ outside the solenoid by decomposing the current sheet into many magnetic dipoles.
[Hint: $\int \left(x^2+a^2\right)^{-3/2} dx = \left(x/a^2\right)\left(x^2+a^2\right)^{-1/2} + C$ and $\int_0^\infty \left(x^2+a^2\right)^{-5/2} dx = 2/3a^4 + C$.]

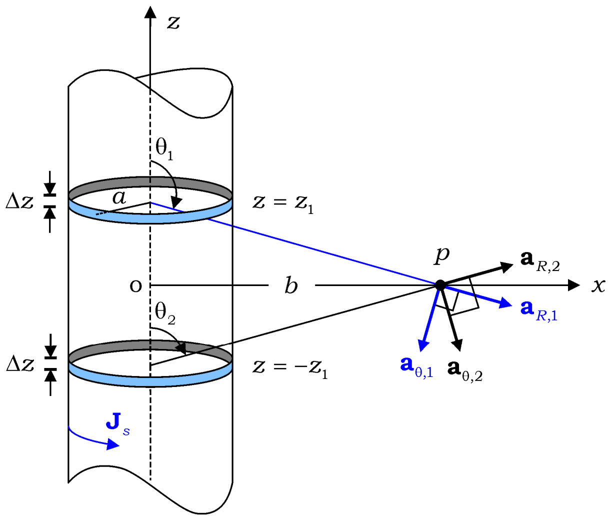


Fig. 5.49 A cylindrical current sheet (Problem 5-16).

5-17 For a nickel metal, the relative permeability $\mu_r = 50$, atomic weight is 58.69 [g/mol], and density is 8.91[g/cm^3]. The material is fully magnetized by a magnetic flux density $B = 0.4$[T]. Find
(a) magnitude of the magnetization **M**, and
(b) magnetic dipole moment **m** per atom.

5-18 A magnetic slab of a thickness t is infinite in extent, and is placed on the xy-plane. The magnetization is uniform with $\mathbf{M} = 10^5\,\mathbf{a}_z$ [A/m] inside the slab($0 \le z \le t$), when a uniform field, $\mathbf{H} = 1.02 \times 10^5\,\mathbf{a}_z$ [A/m], is present below the slab($z < 0$). Find μ of the material.

5-19 The space between two coaxial cylindrical sheets of radii a and b is filled with two dissimilar magnetic materials(μ_1 for $0 < \phi < \pi$ and μ_2 for $\pi < \phi < 2\pi$) as shown in Fig. 5.50. Two cylinders extend between $z = -\infty$ and $z = \infty$, and carry a current I in the opposite directions. Find
(a) **B** and **H** in the gap,
(b) distribution of I on the surface at $\rho = a$, and
(c) magnetization surface current density on the surface at $\rho = a$.

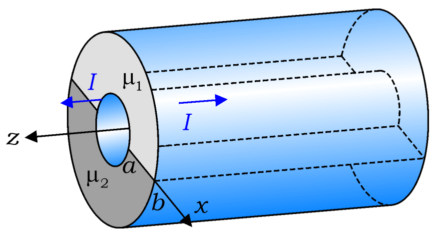


Fig. 5.50 A coaxial cable filled with two dissimilar materials(Problems 5-19 and 5-25).

5-20 A very long solenoid of a radius b has N turns per unit length wound on a core consisting of two dissimilar materials(μ_1 for $\rho < a$ and μ_2 for $a < \rho \le b$. The solenoid is centered on the z-axis, carrying a current I in the $\mathbf{a}_\phi$-direction. Find **B** and **H** inside the solenoid.

5-21 The $z = 0$ plane is an interface between two dissimilar magnetic materials of $\mu_1 = 10\mu_o$ ($z > 0$) and $\mu_2 = 20\mu_o$ ($z < 0$). The interface carries a uniform surface current of $\mathbf{J}_s = 5\mathbf{a}_y$ [A/m]. The magnetic field intensity is $\mathbf{H} = 12\mathbf{a}_x + 20\mathbf{a}_y + 23\mathbf{a}_z$ [A/m] in the region $z > 0$. Find **H** in the region $z < 0$.

5-22 An infinite magnetic slab of permeability μ has a thickness d, and lies on the $z = 0$ plane, as shown in Fig. 5.51. Below the slab($z < 0$), the magnetic flux density $\mathbf{B}_1$ makes an angle θ_1 with respect to the z-axis. In terms of the given values B_1, θ_1, and μ, determine

(a) B_2, B_3, θ_2, and θ_3,

(b) magnetization surface current densities on the $z = 0$ and $z = d$ planes.

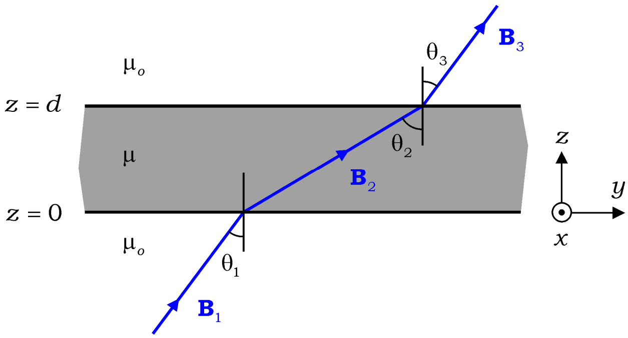


Fig. 5.51 An infinite magnetic slab of thickness d(Problem 5-22).

5-23 The magnetic slab shown in Fig. 5.51 can be replaced with the magnetization surface current densities, $\mathbf{J}_{ms1}$ at $z = 0$ and $\mathbf{J}_{ms2}$ at $z = d$, residing in free space as shown in Fig. 5.52. Find **B** by summing the external field $\mathbf{B}_1$ and the **B**'s due to $\mathbf{J}_{ms1}$ and $\mathbf{J}_{ms2}$, and compare it with the result in **Problem 5-22**.

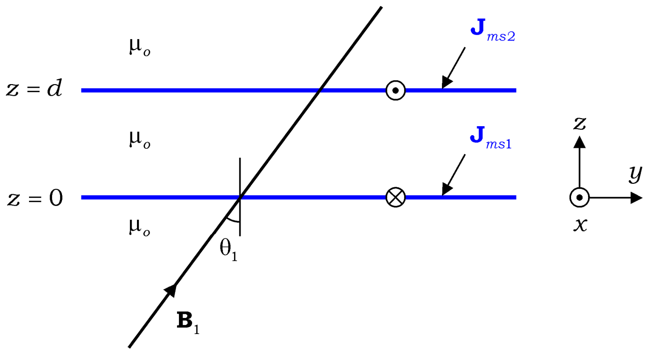


Fig. 5.52 The surface currents in free space equivalent to the magnetization surface currents(Problem 5-23).

5-24 A permanent cylindrical magnet is along the z-axis, having a uniform magnetization $\mathbf{M} = M_o\,\mathbf{a}_z$. The magnetic field is assumed to be uniform inside the magnet by ignoring the fringing effects at the edges.
(a) Find **B** on the top and in the interior of the magnet.
(b) If $\mathbf{B} = B_1\,\mathbf{a}_z$ on the top of a second, identical magnet, find **H** in the interior.

5-25 With reference to the coaxial cable shown in Fig. 5.50, which is filled with two dissimilar materials of μ_1 and μ_2, carrying a steady current I in the opposite directions in the conductors, find, per unit length of the cable,
(a) stored energy, and
(b) inductance.

5-26 A totroidal coil has N turns closely wound around a toroid with a permeability μ and a rectangular cross section, $(b-a)\times h$, as shown in Fig. 5.53. Find the inductance.

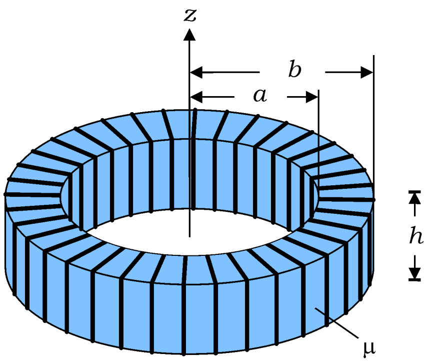


Fig. 5.53 A toroidal coil.

5-27 A toroidal coil has N turns closely wound around two adjoining toroids with different permeabilities but the same rectangular cross section as shown in Fig. 5.54. Find the inductance.

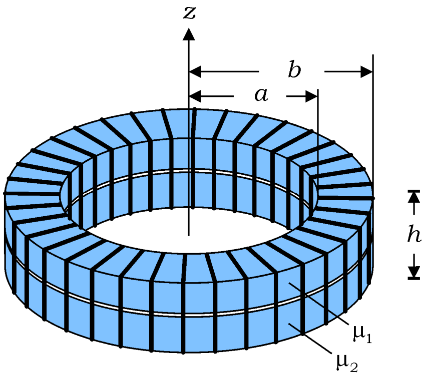


Fig. 5.54 A toroidal coil around two adjoining toroids(Problem 5-27).

5-28 Two identical, circular, conducting loops are arranged along the z-axis as shown in Fig. 5.55. Assuming $a \ll b$ and using the vector magnetic potential of a magnetic dipole moment, find the mutual inductance.

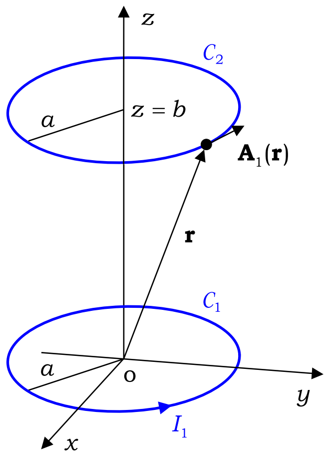


Fig. 5.55 Two identical loops(Problem 5-28).

5-29 A very long coaxial cable consists of a solid inner conductor of radius $a = 2[\text{cm}]$, and a hollow cylinder of inner radius $b = 4[\text{cm}]$ and outer radius $c = 5[\text{cm}]$. Assuming $\mu = \mu_o$ in the conductors and the current I flows in the opposite directions, uniformly distributed in their respective conductors, find the internal inductance of the outer conductor, per unit length of the cable, and compare it with that of the inner conductor.

5-30 Consider the same coaxial cable as in **Problem 5-29**, which involves two conductors with radii $a = 2[\text{cm}]$, $b = 4[\text{cm}]$, and $c = 5[\text{cm}]$. The inner conductor carries a current of $I = 10[A]$ in the $+z$-direction, while the outer conductor carries the same current in the $-z$-direction, which are uniform in the cross sections. Find the accumulated magnetic force, per unit area, on the outermost surface at $\rho = 5[\text{cm}]$.

5-31 A very long solenoid of radius a has N turns per unit length, closely wound around two iron rods $(\mu >> \mu_o)$ separated by a small air gap, as shown in Fig. 5.56. For a current I in the coil, ignoring the fringing effects at the edges, find the force between two rods.

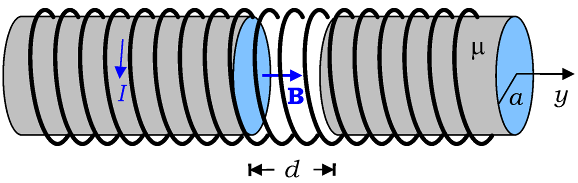


Fig. 5.56 Two iron rods in a solenoid.

5-32 Two identical magnetic dipoles of a dipole moment, $\mathbf{m}_1 = \mathbf{m}_2 = m_o\,\mathbf{a}_z$, are located at different points in free space: $\mathbf{m}_1$ is at the origin, whereas $\mathbf{m}_2$ is at a point $p{:}(1, \sqrt{2}, 1)$ in Cartesian coordinates. Find the torques exerted on the two dipoles.

5-33 A printed circuit board is constrained to pivot about the x-axis which coincides with one of the edges, as shown in Fig. 5.57. The board has a trace of an equilateral triangle of sides a, which carries a current I in the counterclockwise direction in a uniform field $\mathbf{B} = B_o\,\mathbf{a}_y$.

(a) Find the torque about the x-axis if the board is in the xy-plane.
(b) Show that the torque can be expressed as $\mathbf{T} = \mathbf{m} \times \mathbf{B}$.

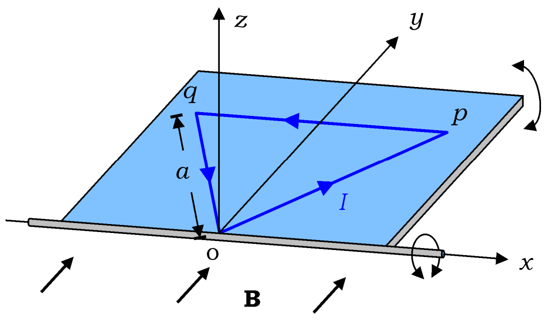


Fig. 5.57 A trace on a circuit board pivoting about the x-axis in **B**.

Chapter 6
Time-Varying Fields and Maxwell's Equations

Until now, we have devoted ourselves to static electric and static magnetic fields that are constant in time. To summarize the discussions up to this point, the electric field and electric flux density due to a distribution of static electric charges are related by the constitutive relation $\mathbf{D} = \varepsilon\mathbf{E}$, from which we define the permittivity of the material. The irrotational nature of **E**, expressed by $\nabla \times \mathbf{E} = 0$, originates from the principle of conservation of energy, while Gauss's law, expressed by $\nabla \bullet \mathbf{D} = \rho_v$, is based on the discrete nature of the electric charges. They constitute two fundamental relations for static electric fields in the sense that they allow us to uniquely determine **E** and **D** in a given region of space. Similarly, the magnetic field and magnetic flux density due to a distribution of steady electric currents are related by the constitutive relation $\mathbf{B} = \mu\mathbf{H}$, from which we define the permeability of the material. The solenoidal nature of **B**, expressed by $\nabla \bullet \mathbf{B} = 0$, follows directly from the closed nature of the magnetic flux lines, while Ampere's circuital law, expressed by $\nabla \times \mathbf{H} = \mathbf{J}$, is based on the fact that the steady current is a vortex source causing a circulation of **H** around it. They constitute two fundamental relations for static magnetic fields in the sense that they allow us to uniquely determine **B** and **H** in a given region of space.

There is no mutual relationship between the static electric and static magnetic fields, although two fields may coexist in a conductor in the sense that a static electric field generates a steady current in the conductor, which in turn induces a static magnetic field around it. However, a static magnetic field cannot produce a static electric field. Under time-varying conditions, the electric and magnetic fields due to a time-varying current will be coupled to each other, and form an electromagnetic wave propagating through free space and in material media.

In the present chapter, we focus our attention on time-varying electric and magnetic fields. We then introduce the concept of wave motion in chapter 7, and discuss electromagnetic waves propagating in different media in chapter 8. From now on, to avoid confusion, we use bold-script letters such as $\boldsymbol{\mathscr{E}}$ and $\boldsymbol{\mathscr{H}}$ to denote time-varying electric and time-varying magnetic fields, while using script letters such as $\mathscr{D}_x$ and $\mathscr{B}_x$ to denote scalar components of $\boldsymbol{\mathscr{D}}$ and $\boldsymbol{\mathscr{B}}$. For other time-varying vector and time-varying scalar quantities, such as time-varying current density and time-varying volume charge density, we use bold-italic letters and italic letters, respectively, such as $\boldsymbol{J}$ and ρ_v.

Time-varying electromagnetic fields differ from static electric and magnetic fields in a number of important respects, apart from the fact that they vary in time. First, the time-varying electromagnetic fields are generated by accelerated charges or electric currents varying in time. Second, the time-varying electric and magnetic fields are coupled to each other in that the time-varying electric field induces the time-varying magnetic field, and vice versa. In the same way that a static field is governed by its divergence and curl, the time-varying electric and magnetic fields are governed by their divergences and curls. Gauss's laws for electricity and magnetism remain valid under time-varying conditions: $\nabla \bullet \mathscr{D} = \rho_v$ and $\nabla \bullet \mathscr{B} = 0$. However, the two curl equations of $\mathscr{E}$ and $\mathscr{H}$ need to be modified to conform with Faraday's electromagnetic induction and the concept of displacement current density introduced by Maxwell, respectively. The four equations, which are the curl of $\mathscr{E}$, curl of $\mathscr{H}$, divergence of $\mathscr{D}$, and divergence of $\mathscr{B}$, are called Maxwell's equations that the time-varying electromagnetic field must satisfy in a given region of space at all times.

6.1 Faraday's Law

Michael Faraday observed that an electromotive force (*emf*) was induced in a wire loop when a permanent magnet moved near the loop or the loop moved near the magnet. The electromotive force is nothing but a voltage induced in the loop. Faraday's law of electromagnetic induction states that ***the induced emf in a wire loop is equal to the negative time rate of change of the magnetic flux linkage with the loop***.

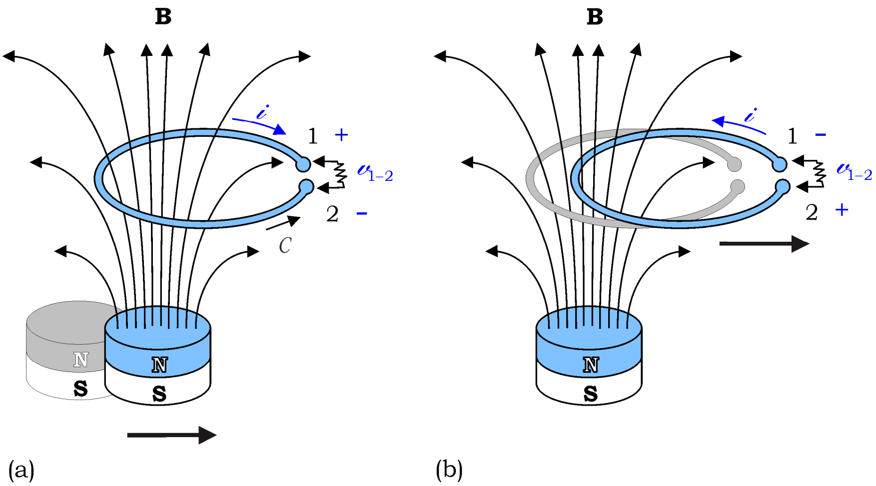


Fig. 6.1 Electromagnetic induction. The induced *emf* gives rise to the current i and the terminal voltage v_{1-2}: (a) $v_{1-2} > 0$ and $emf < 0$, and (b) $v_{1-2} < 0$ and $emf > 0$, (the counterclockwise direction is the positive direction of C).

According to Faraday's law of electromagnetic induction, the induced *emf* in a loop of N turns is expressed as

$$\boxed{emf = -\frac{d\Lambda}{dt} = -N\frac{d\Phi}{dt}} \qquad \text{[V]} \qquad (6\text{-}1)$$

where Λ is the magnetic flux linkage with the loop and Φ is that with a single turn of the loop. The *emf* is equal to the closed line integral of the induced electric field in the loop such that

$$\boxed{emf = \oint_C \mathscr{E} \cdot d\mathbf{l}} \qquad \text{[V]} \qquad (6\text{-}2)$$

It is important to note that the sign of *emf* depends not only on the direction of $\mathscr{E}$, but also on the positive direction of C. Thus, a positive *emf* signifies that the induced $\mathscr{E}$ is directed in the positive direction of C. Although the *emf* or $\mathscr{E}$, in Eq. (6-2), is induced by the time-varying magnetic flux as given in Eq. (6-1), they may or may not vary with time. Unlike the static electric field that is an irrotational field, or a conservative field, the closed line integral of $\mathscr{E}$ is nonzero, and is a non-conservative field. As a result, $\mathscr{E}$ is not given by the negative gradient of the electric potential. Note that $\mathscr{E}$ generates a current i in the loop, which flows in the direction of $\mathscr{E}$.

Seeing Eqs. (6-1) and (6-2), we can express Faraday's law in terms of the integral of $\mathscr{E}$, conducted around the loop C, and the integral of $\mathscr{B}$, conducted over the surface S bounded by C, that is,

$$\boxed{emf = \oint_C \mathscr{E} \cdot d\mathbf{l} = -\frac{d}{dt}\int_S \mathscr{B} \cdot d\mathbf{s}} \qquad (6\text{-}3)$$

If the loop has many turns, the closed line integral is conducted around each and every turn of the loop, and the surface integral is performed over all the surfaces enclosed by all the turns of the loop. The direction of $d\mathbf{l}$ on C and the direction of $d\mathbf{s}$ on S are governed by the right-hand rule: the right thumb points in the direction of $d\mathbf{s}$, when the four fingers rotate in the direction of $d\mathbf{l}$, or the positive direction of C.

The negative sign in Eq. (6-1) and Eq. (6-3) is compatible with Lenz's law, which states that ***the emf is such that it induces a current in the loop and thus a magnetic flux so as to oppose the change in the total magnetic flux linking with the loop***.

Let us use the simply electromagnetic induction depicted in Fig. 6.1 to clarify the relation between the *emf*, terminal voltage v_{1-2}, induced current i, and electric field $\mathscr{E}$ in the loop, which is very confusing in the discussion of electromagnetic induction. Figure 6.1(a) shows that the magnetic flux linkage with the loop increases with time, as the permanent magnet moves to the right. According to Faraday's and Lentz's laws, the electric field $\mathscr{E}$ and thus the current i are induced in the loop both in the clockwise direction. If the terminal 1-2 is open, the current

$\mathscr{i}$ tends to accumulate positive charges at terminal 1 and negative charges at terminal 2, resulting in a positive terminal voltage, $\mathscr{v}_{1-2} > 0$: terminal 1 is at a higher potential than terminal 2. The terminal voltage counterbalances the *emf*, and there is no current flowing in the loop. If a resistance R is connected cross the terminals, an ohmic current will of course flow from terminal 1 to terminal 2 through the resistor, which is the current $\mathscr{i}$ induced in the loop. The *emf* is simply equal to the closed line integral of $\mathscr{E}$ around the loop; its sign signifies the direction of $\mathscr{E}$ relative to the direction of travel on the loop. It is important to note that the terminal voltage $\mathscr{v}_{1-2}$ is an induced voltage, not an externally applied voltage. If $\mathscr{v}_{1-2}$ were a voltage externally applied across the terminals, the electric field and the current in the loop would be directed opposite to those shown in the figure. Similarly, as we see in Fig. 6.1(b), the magnetic flux linkage with the loop decreases with time as the loop moves to the right. Therefore, the induced $\mathscr{E}$ and $\mathscr{i}$ are directed in the counterclockwise direction in the loop, and we have a negative terminal voltage, $\mathscr{v}_{1-2} < 0$, and a positive *emf*. This is because the counterclockwise direction is taken as the positive direction of C, as shown in the figure.

Exercise 6.1
If the clockwise direction is taken as the positive direction of C in Fig. 6.1, what is the sign of the *emf* induced in the loop in Figs. 6.1(a) and (b)?

Ans. (a) Positive, and (b) Negative

6.1.1 Transformer emf

The surface integral on the right-hand side of Eq. (6-3) can be viewed as the sum of incremental magnetic fluxes passing through different portions of the given surface S. In this case, the incremental flux is given by the dot product of the magnetic flux density $\mathscr{B}$ and an incremental area vector $\Delta\mathbf{s}$; that is, $\Delta\Phi = \mathscr{B} \cdot \Delta\mathbf{s}$. Noting that $\Delta\mathbf{s}$ is on the surface S that is bounded by the loop C, the direction and location of $\Delta\mathbf{s}$ in space are closely related to the motion of C as a function of time. As is evident from Eq. (6-3), the *emf* induced in C depends on the time derivative of $\Delta\Phi$. There are three cases in which the time derivative of $\Delta\Phi$ is nonzero:

(1) Loop is stationary in a time-varying field: $d(\Delta\Phi)/dt = (d\mathscr{B}/dt) \cdot \Delta\mathbf{s}$.

(2) Loop moves in a static magnetic field: $d(\Delta\Phi)/dt = \mathbf{B} \cdot d(\Delta\mathbf{s})/dt$.

(3) Loop moves in a time-varying field: $d(\Delta\Phi)/dt = d(\mathscr{B} \cdot \Delta\mathbf{s})/dt$.

The transformer *emf* is one that is induced in a stationary loop placed in a time-varying magnetic field. In this case, $d\mathbf{s}$ has nothing to do with $d(\Delta\Phi)/dt$, and thus the time derivative can be taken inside the integral sign in Eq. (6-3). The transformer *emf* is therefore

$$\boxed{emf_t = \oint_C \mathscr{E} \cdot d\mathbf{l} = -\int_S \frac{\partial \mathscr{B}}{\partial t} \cdot d\mathbf{s}} \qquad [\text{V}] \qquad (6\text{-}4)$$

where t stands for transformer *emf*. Again, the time derivative in Eq. (6-3) is moved inside the integral sign and changed to the partial derivative in Eq. (6-4). Next, applying Stokes's theorem to Eq. (6-4) we obtain

$$\int_S (\nabla \times \mathscr{E}) \bullet d\mathbf{s} = -\int_S \frac{\partial \mathscr{B}}{\partial t} \bullet d\mathbf{s} \tag{6-5}$$

The surface S may be arbitrary only if it is bounded by the loop C. Thus, to satisfy the equality in Eq. (6-5), the two integrands should be the same at every point on S. The point form of Faraday's law is therefore

$$\boxed{\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t}} \tag{6-6}$$

This is a member of Maxwell's equations.

Example 6-1

A circular loop of radius a is stationary in the xy-plane in the presence of a magnetic flux density $\mathscr{B}(t) = \mathbf{a}_z B_o \cos(\omega t)$. Find *emf* induced in the loop.

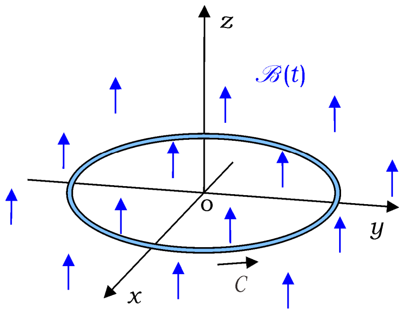


Fig. 6.2 A stationary circular loop in a time-varying magnetic field.

Solution

The counterclockwise direction is taken as the positive direction of the loop, as shown in Fig. 6.2, and thus the differential area vector $d\mathbf{s}$ is directed along the positive z-axis, according to the right-hand rule. The magnetic flux linking the loop is

$$\Phi = \int_S \mathscr{B} \bullet d\mathbf{s} = \int_{\rho=0}^{\rho=a} \int_{\phi=0}^{\phi=2\pi} \mathbf{a}_z B_o \cos(\omega t) \bullet (\mathbf{a}_z \rho\, d\rho\, d\phi)$$
$$= \pi B_o a^2 \cos(\omega t)$$

The transformer *emf* induced in the loop is

$$emf_t = -\frac{d\Phi}{dt} = \pi B_o a^2 \omega \sin(\omega t)$$

When the *emf* is positive, it means that the induced electric field is directed in the positive direction of C, the counterclockwise direction in this case.

The normalized flux linkage $\overline{\Phi}$ and the normalized electromotive force $\overline{emf}$ are plotted as functions of time in Fig. 6.3. The shade area is the time interval in which $\overline{\Phi}$ is increasing, and thus $\overline{emf}$ is negative($\mathscr{E}$ is directed in the clockwise direction in the loop).

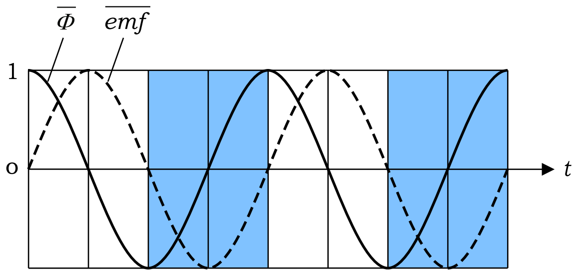


Fig. 6.3 The normalized flux linkage and *emf* as functions of time.

Exercise 6.2
For a circular loop of radius a and a square loop of side b residing separately in $\mathscr{B}(t) = \mathbf{a}_z B_o \cos\omega t$, find the relation between a and b for the same (a) *emf* (b) $\mathscr{E}$ in the two loops.

Ans. (a) $\pi a^2 = b^2$, (b) $2a = b$

6.1.1.1 Ideal Transformer
The transformer is an alternating current device, which operates on Faraday's law of electromagnetic induction. Fig. 6.4 shows a transformer consisting of two coils wound around a common core. An ideal magnetic core is made of a lossless material having infinite permeability, $\mu = \infty$, so that the magnetic flux may be limited to the interior of the core, and thus the two coils may be coupled magnetically with no flux leakage.

Let us consider the case when the primary coil with N_1 turns is connected to an *ac*-voltage source v_1, and the secondary coil with N_2 turns is connected to a load resistance R_L. The source voltage v_1 and the current i_1 in the primary circuit are responsible for the magnetic flux Φ that is established in the magnetic core. Note that the same magnetic flux Φ links with both the primary and secondary coils. Under these conditions, according to Faraday's law, the *emf*'s across the two coils are given as follows:

$$emf_1 = v_1 = -N_1 \frac{d\Phi}{dt} \tag{6-7a}$$

$$emf_2 = v_2 = -N_2 \frac{d\Phi}{dt} \tag{6-7b}$$

We see from Eq. (6-7a) that the emf_1 counterbalances the source voltage v_1. Otherwise, an infinite current would flow in the primary circuit. On the other hand, the emf_2 in the secondary coil provides a voltage v_2 across the load resistance R_L, and produces a current i_2 in the secondary circuit according to Ohm's law (see Fig. 6.4). The power dissipated in the primary circuit, $P_1 = v_1 i_1$, should be the same as that in the secondary circuit, $P_2 = v_2 i_2$, to satisfy the principle of conservation of energy:

$$\begin{aligned} P_1 &= v_1 i_1 \\ &= P_2 = v_2 i_2 \end{aligned} \tag{6-8}$$

Combination of Eqs. (6-7) and (6-8) gives

$$\frac{v_1}{v_2} = \frac{i_2}{i_1} = \frac{N_1}{N_2} \equiv \alpha \tag{6-9}$$

where α is called the turns ratio. The voltage ratio v_1 / v_2 is proportional to the turns ratio, whereas the current ratio i_1 / i_2 is inversely proportional to the ratio.

The two transformers shown in Figs. 6.4(a) and 6.4(b) are identical except that their second coils are wound in the opposite directions. For this reason, the polarity of v_2 and the direction of i_2 in Fig. 6.4(b) are the reverse of those in Fig. 6.4(a).

As we see from Eq. (6-9), the transformer can transform voltages and currents. Furthermore, the transformer can change the impedance of a circuit. In the primary circuit, the transformer can be regarded as an equivalent load with an impedance Z_1, that is,

$$Z_1 = \frac{v_1}{i_1} = \frac{(N_1 / N_2) v_2}{(N_2 / N_1) i_2} \tag{6-10}$$

If the source voltage v_1 varies sinusoidally with time, and the load impedance is given by Z_2, the effective impedance seen by the source, in the primary circuit, is

$$Z_1 = \left(\frac{N_1}{N_2} \right)^2 Z_2 \tag{6-11}$$

The impedance transformation involves the square of the turns ratio.

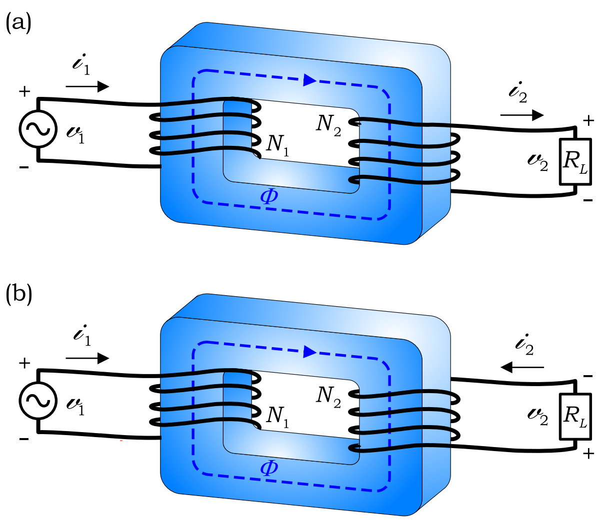


Fig. 6.4 Two identical transformers with the secondary coils wound in the opposite directions.

Exercise 6.3
What is the turns ratio in a step-up transformer changing 110V to 220V?

Ans. 0.5.

6.1.2 Motional emf

The motional *emf* is one that is induced in a loop moving in a static magnetic field. Expressed mathematically,

$$\boxed{emf_m = \oint_C \mathscr{E} \bullet d\mathbf{l} = -\frac{d}{dt}\int_S \mathbf{B} \bullet d\mathbf{s}} \qquad \text{[V]} \qquad (6\text{-}12)$$

For instance, an *emf* may be induced in a conducting loop, if the loop rotates in a uniform static field **B**, or moves rectilinearly in a non-uniform static field **B**.

As an example, consider a rectangular loop of sides x_o and y_o, which enters the region of a uniform field $\mathbf{B} = B_o\,\mathbf{a}_z$ ($y \geq y_o$) at time $t = 0$, and moves with a constant velocity $\boldsymbol{v} = v_o\,\mathbf{a}_y$, as depicted in Fig. 6.5. During the time $0 \leq t \leq y_o / v_o$, the motional *emf* induced in the loop is computed from Eq. (6-12) as follows:

$$\begin{aligned} emf_m &= -\frac{d}{dt}\int_S \mathbf{B} \bullet d\mathbf{s} \\ &= -\frac{d}{dt}(B_o x_o v_o t) = -B_o x_o v_o \end{aligned} \qquad (6\text{-}13)$$

In the above equation, the counterclockwise direction is taken as the positive direction of the loop, and $d\mathbf{s}$ is thus directed in the $+z$-direction according to the right-hand rule. The negative *emf* in Eq. (6-13) signifies that the electric field and the current are induced in the clockwise direction in the loop. The induced current i generates a magnetic flux in such a way as to oppose the increase in the flux linkage with the loop in accordance with Lenz's law.

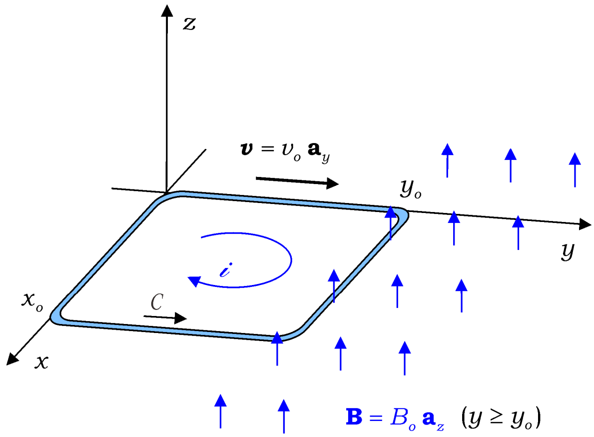


Fig. 6.5 A motional *emf* induced in a loop moving in a static magnetic field.

The motional *emf* is closely related to the magnetic force exerted on the conduction electrons, move together with the loop in the magnetic field. By use of the Lorentz force equation, we can express the magnetic force on a free electron in the conductor moving with a velocity $\boldsymbol{v}$ in a static field $\mathbf{B}$ as

$$\boldsymbol{F}_m = e\boldsymbol{v} \times \mathbf{B} \tag{6-14}$$

where e is the electron charge of -1.6×10^{-19}[C]. The magnetic force exerted on a unit charge in the conducting loop is referred to as the motional electric field intensity, which is expressed as

$$\boxed{\mathscr{E}_m = \frac{\boldsymbol{F}_m}{e} = \boldsymbol{v} \times \mathbf{B}} \tag{6-15}$$

By integrating $\mathscr{E}_m$ around the wire loop we obtain the motional *emf*, that is,

$$\boxed{emf_m = \oint_C \mathscr{E}_m \cdot d\mathbf{l} = \oint_C (\boldsymbol{v} \times \mathbf{B}) \cdot d\mathbf{l}} \qquad \text{[V]} \tag{6-16}$$

where $\boldsymbol{v}$ is the velocity of the conducting loop, and $\mathbf{B}$ is the static magnetic flux density.

To see how Eq. (6-16) can be used for obtaining the motional *emf*, we again consider the rectangular loop shown in Fig. 6.5. By applying Eq. (6-15) to the loop, we see that only the right side of the loop contributes to emf_m. There is no contribution from the top or the bottom side of the loop, because the induced $\mathscr{E}_m$ is perpendicular to the wire. There is no contribution from the left side, because it travels in the region of $\mathbf{B} = 0$ for the time $0 < t < y_o / v_o$. Thus, the motional *emf* is obtained from Eq. (6-16) as

$$\begin{aligned} emf_m &= \int_{x=x_o}^{x=0} (\boldsymbol{v} \times \mathbf{B}) \cdot d\mathbf{l} \\ &= \int_{x=x_o}^{x=0} v_o B_o \, \mathbf{a}_x \cdot \mathbf{a}_x dx = -B_o x_o v_o \end{aligned} \tag{6-17}$$

This is the same as Eq. (6-13).

We next consider a classical example of motional *emf* as illustrated in Fig. 6.6, in which a conducting bar slides on a pair of conducting rails with a constant velocity $\boldsymbol{v} = v_o \, \mathbf{a}_y$ in a uniform magnetic flux density $\mathbf{B} = B_o \, \mathbf{a}_z$. When switch S_1 is opened, the induced motional *emf* appears as a terminal voltage between the terminals of S_1. From Eq. (6-16), the motional *emf* induced in the sliding bar is

$$\begin{aligned} emf_m &= \int_u^l (\boldsymbol{v} \times \mathbf{B}) \cdot d\mathbf{l} \\ &= \int_{x=0}^{x=x_o} v_o B_o \, \mathbf{a}_x \cdot \mathbf{a}_x dx = v_o B_o x_o \end{aligned} \tag{6-18}$$

where u and l stand for the upper and lower points on the sliding bar at which the bar touches the rails. The interval of integration in Eq. (6-18), from u to l, signifies that the clockwise direction has been taken as the positive direction of the closed path, which is formed by the sliding bar, rails, and terminals 1 and 2. The positive *emf* in Eq. (6-18) implies that the induced $\mathscr{E}_m$ is directed in the $+x$-direction in the bar. In view of Eq. (6-18), from Fig. 6.6, the terminal voltage is given as

$$\mathscr{V}_{1-2} = v_o B_o x_o \tag{6-19}$$

The positive value of $\mathscr{V}_{1-2}$ means that terminal 1 is at a higher potential than terminal 2.

Referring to Fig. 6.6, let us now examine the relation between $\mathscr{E}_m$ and $\mathscr{V}_{1-2}$. As we can see from Eq. (6-15), the motional electric field $\mathscr{E}_m$ represents the magnetic force in the sliding bar, which causes free electrons to move towards point u, while accumulating positive charges at point l. The internal electric field due to the separated charges then counteracts the motional electric field $\mathscr{E}_m$, opposing a further separation of charges. Consequently, when the motional electric field $\mathscr{E}_m$ results in the separation of charges in the sliding bar, the charges give rise to a potential difference between the two end points of the bar, which is equal to the terminal voltage $\mathscr{V}_{1-2}$.

The motional *emf* can be obtained directly from Eq. (6-12) as

$$\begin{aligned} emf_m &= -\frac{d}{dt}\int_s \mathbf{B}\bullet d\mathbf{s} \\ &= -\frac{d}{dt}\int_s B_o\,\mathbf{a}_z\bullet(-dxdy\,\mathbf{a}_z) = \frac{d}{dt}\left[B_o x_o\left(y_o + v_o t\right)\right] \end{aligned} \tag{6-20}$$

In the above equation, $d\mathbf{s}$ is directed in the $-z$-direction in accordance with the right-hand rule. Note that the result in Eq. (6-20) is the same as in Eq. (6-18).

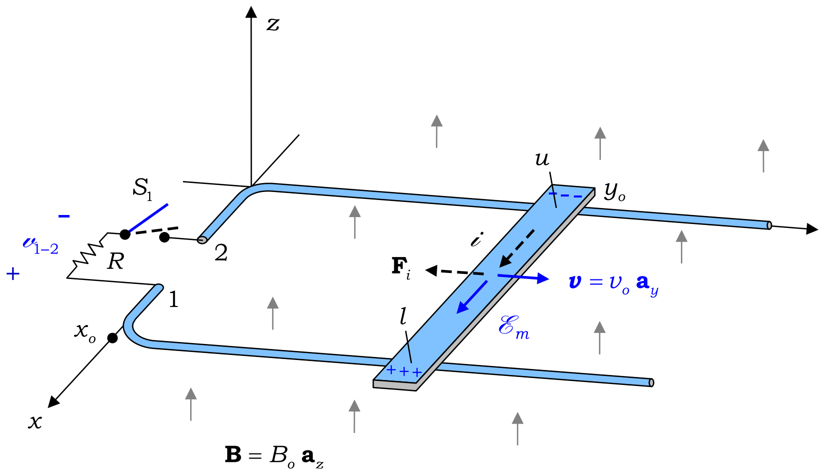


Fig. 6.6 A sliding bar moving on a pair of conducting rails.

If the switch S_1 is closed, the motional *emf* induces a current i in the closed path that comprises the sliding bar, rails, and the resistor *R*. When the sliding bar, carrying the current, moves with a velocity $\boldsymbol{v}$ in the magnetic field, it experiences a magnetic force directed in the negative direction of $\boldsymbol{v}$. Accordingly, energy is expended in moving the bar in the magnetic field. The magnetic force acting on the bar with the current i is obtained from Eq. (5-135b) as follows:

$$\begin{aligned} \mathbf{F}_i &= i\int_C d\mathbf{l}\times\mathbf{B} \\ &= i\int_{x=0}^{x=x_o} B_o(-\mathbf{a}_y)dx = -iB_o x_o\,\mathbf{a}_y \end{aligned} \qquad \text{[N]} \tag{6-21}$$

where we use $d\mathbf{l} = dx\,\mathbf{a}_x$ and $\mathbf{B} = B_o\,\mathbf{a}_z$. Note that the integral in Eq. (6-21) is conducted along the direction of the current in the sliding bar. Because the magnetic force $\mathbf{F}_i$ opposes the bar to move with the velocity $\boldsymbol{v} = v_o\,\mathbf{a}_y$, work must be done in moving the bar. The power dissipated in moving the bar with a velocity $\boldsymbol{v} = v_o \, \mathbf{a}_y$ is therefore

$$P_{me} = -\mathbf{F}_i \bullet \boldsymbol{v} = \mathscr{i} B_o x_o v_o \qquad \text{[W]} \qquad (6\text{-}22)$$

where $-\mathbf{F}_i$ is the mechanical force applied to the bar in order to cancel the magnetic force. According to the principle of conservation of energy, the mechanical power should be equal to the electrical power dissipated in the resistance *R*. With the help of $\mathscr{v}_{1-2}$ expressed by Eq. (6-19), the electrical power is written as

$$P_{el} = \mathscr{i}\mathscr{v}_{1-2} = \mathscr{i} B_o x_o v_o \qquad \text{[W]} \qquad (6\text{-}23)$$

We note that the two results in Eqs. (6-22) and (6-23) are the same: $P_{me} = P_{el}$.

Exercise 6.4
Find the *emf* induced in the loop in Fig. 6.5, if the magnetic flux density, $\mathbf{B} = B_o \, \mathbf{a}_z$, is confined to the region $0 \le y \le y_o$.

Ans. $B_o x_o v_o$ [V] for $0 \le t \le (y_o / v_o)$.

Exercise 6.5
What will become of the work done in moving the bar in Fig 6.6, if switch S_1 is opened?

Ans. $\mathscr{v}_{1-2}$.

6.1.3 A Loop Moving in a Time-Varying Magnetic Field

We now consider the more general case of *emf* where a closed wire loop moves in a time-varying magnetic field. In this case, the total *emf* is given by the sum of the transformer *emf* and the motional *emf*, that is,

$$\boxed{\begin{aligned} emf &= \oint_C \mathscr{E} \bullet d\mathbf{l} \\ &= -\int_S \frac{\partial \mathscr{B}}{\partial t} \bullet d\mathbf{s} + \oint_C (\boldsymbol{v} \times \mathscr{B}) \bullet d\mathbf{l} \end{aligned}} \qquad \text{[V]} \qquad (6\text{-}24)$$

where C represents the closed conducting loop moving with a velocity $\boldsymbol{v}$ in a time-varying magnetic flux density $\mathscr{B}$, and S represents the surface bounced by C. The directions of $d\mathbf{s}$ and $d\mathbf{l}$ follow the right-hand rule. The surface integral on the right-hand side of Eq. (6-24) represents the transformer *emf*, which is evaluated by assuming the surface S to be fixed in space. On the other hand, the closed line integral represents the motional *emf*, which is evaluated by assuming $\mathscr{B}$ to be constant in time. We note that the total *emf* in Eq. (6-24) can be simply written as Eq. (6-3), namely

$$emf = \oint_C \mathscr{E} \bullet d\mathbf{l} = -\frac{d}{dt}\int_S \mathscr{B} \bullet d\mathbf{s} \qquad \text{[V]} \qquad (6\text{-}3)(6\text{-}25)$$

Upon applying Stokes's theorem to the closed line integrals in Eq. (6-24), noting that the surface S may be arbitrary, we obtain

$$\boxed{\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t} + \nabla \times (\boldsymbol{v} \times \mathscr{B})} \qquad (6\text{-}26)$$

The first term on the right-hand side of Eq. (6-26) originates from the transformer *emf*, while the second term comes from the motional *emf*.

Example 6-2

A rectangular wire loop of sides a and b rotates about the x-axis with an angular speed ω_o in a time-varying magnetic flux density $\mathscr{B} = B\sin(\omega t)\,\mathbf{a}_y$, as shown in Fig. 6.7. The unit normal to the loop surface, $\mathbf{a}_S$, makes an angle $\varphi = \omega_o t + \alpha$ with respect to the $+y$-axis, where α is constant. Find at time $t = t_1$ (a) transformer *emf*, (b) motional *emf*, and (c) total *emf* from Eq. (6-25).

Solution

(a) Total magnetic flux linking the loop at $t = t_1$ is

$$\begin{aligned}\Phi &= \int \mathscr{B} \bullet d\mathbf{s} = B\sin(\omega t_1)\,\mathbf{a}_y \bullet ab\,\mathbf{a}_s \\ &= abB\sin(\omega t)\big|_{t=t_1} \cos(\varphi(t_1))\end{aligned} \qquad (6\text{-}27)$$

Transformer *emf* is obtained from Eq. (6-27) by fixing φ at a constant and taking the time derivative of the flux linkage,

$$emf_t = -\left.\frac{d\Phi}{dt}\right|_{\substack{t=t_1,\\ \varphi=\text{constant}}} = -\omega abB\cos(\omega t_1)\cos(\varphi(t_1)) \qquad (6\text{-}28)$$

In view of the unit surface normal $\mathbf{a}_S$ we see that the loop should be traversed, passing corners 1, 2, 3, and 4 in accordance with the right-hand rule, for the calculation of the *emf*. Thus a positive *emf* in Eq. (6-28) means that terminal *I* is at a higher potential than terminal *II* in Fig. 6.7.

(b) Motional *emf* is obtained from Eq. (6-16) by fixing $\mathscr{B}$ at a value given at time $t = t_1$:

$$\begin{aligned}emf_m &= \oint_C (\boldsymbol{v} \times \mathscr{B}) \bullet d\mathbf{l} \\ &= \int_1^2 \left[\omega_o \left(\tfrac{1}{2}b\right)\mathbf{a}_S \times \mathbf{a}_y B\sin(\omega t_1)\right] \bullet (dx\,\mathbf{a}_x) \\ &\quad + \int_3^4 \left[\omega_o \left(\tfrac{1}{2}b\right)(-\mathbf{a}_S) \times \mathbf{a}_y B\sin(\omega t_1)\right] \bullet (dx\,\mathbf{a}_x) \\ &= \omega_o abB\sin(\omega t_1)\sin\varphi(t)\big|_{t=t_1}\end{aligned} \qquad (6\text{-}29)$$

In Eq. (6-29), we used $\boldsymbol{v} = \omega_o\left(\frac{1}{2}b\right)(\pm\mathbf{a}_S)$ for the lower and upper sides of the loop, respectively. There is no contribution from the left or the right side of the loop because $\boldsymbol{v} \times \mathscr{B}$ is perpendicular to the wire.

(c) Upon inserting $\varphi = \omega_o t + \alpha$ in Eq. (6-27), the instantaneous magnetic flux linkage with the loop is

$$\Phi = abB\sin(\omega t)\cos(\omega_o t + \alpha)$$

Total *emf* at $t = t_1$ is therefore

$$\begin{aligned} emf &= -\left.\frac{d\Phi}{dt}\right|_{t=t_1} \\ &= -\omega abB\cos(\omega t_1)\cos(\omega_o t_1 + \alpha) + \omega_o abB\sin(\omega t_1)\sin(\omega_o t_1 + \alpha) \end{aligned} \tag{6-30}$$

In Eq. (6-30), ω should not be confused with ω_o : ω is the angular frequency of $\mathscr{B}$, whereas ω_o is the angular speed of the loop. We note that the result in Eq. (6-30) is equal to the sum of those in Eqs. (6-28) and (6-29).

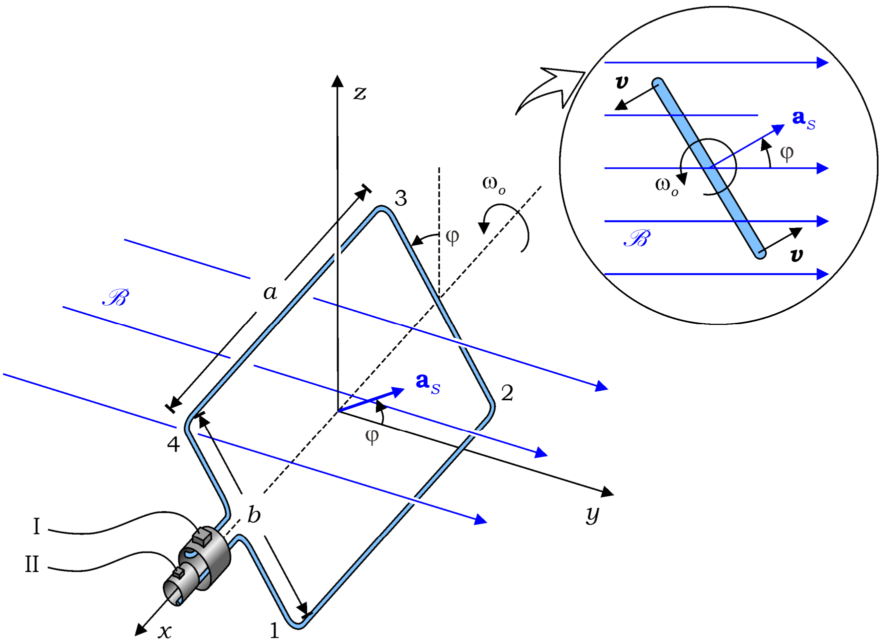


Fig. 6.7 A rectangular loop rotating in a time-varying magnetic field.

Exercise 6.6

In Fig. 6.6, the bar moves with a velocity $\boldsymbol{v} = v_o \mathbf{a}_y$ in a time-varying field $\mathscr{B} = \mathbf{a}_z B_o \cos(\omega_o t)$, crossing the *x*-axis at $t = 0$. Find v_{1-2} across the open terminal for $t \geq 0$.

Ans. $v_{1-2} = x_o v_o B_o \left[\cos(\omega_o t) - \omega_o t \sin(\omega_o t)\right]$

Review Questions with Hints

RQ 6.1	State Faraday's law of electromagnetic induction in words.	[Eq.(6-1)]
RQ 6.2	State Lenz's law in words.	[Eq.(6-1)]
RQ 6.3	What is the significance of a negative *emf*?	[Eq.(6-2), Fig.6.3]
RQ 6.4	Explain the case in which the *emf* is the sum of the transformer and motional *emf*s?	[Eq.(6-24)]

6.2 Displacement Current Density

As we saw in section 6-1, the time-varying electric field is not a conservative field. Under time-varying conditions, the expression for the irrotational nature of the static electric field, $\nabla \times \mathbf{E} = 0$, is modified to $\nabla \times \mathscr{E} = -\partial \mathscr{B} / \partial t$ in order to conform with Faraday's electromagnetic induction. Similarly, under time-varying conditions, the expression for Ampere's circuital law, $\nabla \times \mathbf{H} = \mathbf{J}$, should be modified in order to incorporate Maxwell's hypothesis of displacement current density.

Ampere's circuital law is not consistent with the equation of continuity under time-varying conditions. To start with, the point form of Ampere's circuital law is

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{6-31}$$

Taking the divergence of both sides of Eq. (6-31), we have

$$\nabla \cdot \nabla \times \mathbf{H} = \nabla \cdot \mathbf{J}$$

The left-hand side of the equation is identically zero, because of the vector identity $\nabla \cdot \nabla \times \mathbf{U} = 0$, and thus the equation simply reduces to $\nabla \cdot \mathbf{J} = 0$. Invoking the equation of continuity, $\nabla \cdot \boldsymbol{J} = -\partial \rho_v / \partial t$, we see that Ampere's circuital law holds true under static conditions only.

Ampere's circuital law leads to a contradiction, if it is applied to a parallel-plate capacitor being charged by an increasing current i as illustrated in Fig. 6.8. Let us consider a closed loop C around the capacitor lead, and suppose S_1 and S_2 are two separate surfaces bounded by the same loop C: the bounded surface S_1 is flat in Fig. 6.8(a), while S_2 shown in Fig. 6.8(b) resembles the surface of an untied balloon, covering one of the two conducting plates of the capacitor. Applying Ampere's circuital law to the loop C with the surface S_1 as shown in Fig. 6.8(a), we see that the closed line integral of $\mathbf{H}$ around C is nonzero, because the conduction current i

passes through $\mathcal{S}_1$. However, if we apply Ampere's circuital law to the same loop $\mathcal{C}$ but with the surface $\mathcal{S}_2$ as shown in Fig. 6.8(b), no conduction current actually passes through $\mathcal{S}_2$, and thus we arrive at a contradictory conclusion that the closed line integral of **H** around $\mathcal{C}$ is zero.

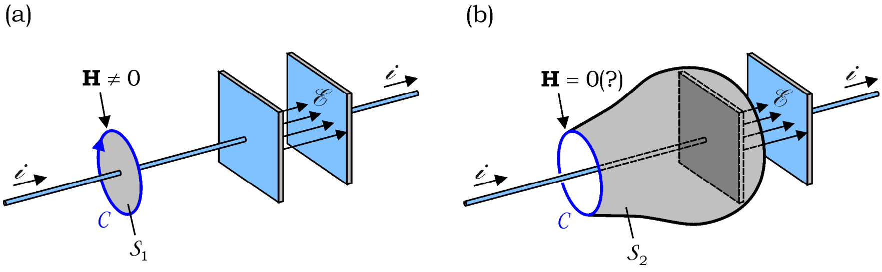


Fig. 6.8 A capacitor is being charged by an increasing current $\imath$. A loop $\mathcal{C}$ encloses a surface either $\mathcal{S}_1$ or $\mathcal{S}_2$.

The conduction current is not the only source of a time-varying magnetic field. If the conduction current $\imath$ is increasing in the parallel-plate capacitor, a time-varying magnetic field is also induced in the space between the conducting plates where there is no conduction current. When the capacitor is charged to $\pm\mathcal{Q}$ [C], the electric field intensity in the air gap is written as

$$\mathcal{E} = \frac{\mathcal{Q}}{\varepsilon_o A}\mathbf{a}_z \tag{6-32}$$

where A is the surface area of the conducting plates assumed to be perpendicular to the z-axis. Note that the fringing effects at the edges are ignored in Eq. (6-32). Next, by taking the time derivatives of both sides of Eq. (6-32) and invoking the relation $\imath = d\mathcal{Q}/dt$, we have

$$\frac{\partial(\varepsilon_o\mathcal{E})}{\partial t} = \frac{\imath}{A}\mathbf{a}_z \tag{6-33}$$

The right-hand side of Eq. (6-33) has the same unit as a current density, and the term in parenthesis on the left-hand side is the electric flux density. In view of Eq. (6-33), the displacement current density $\boldsymbol{J}_D$ is defined as the time derivative of the electric flux density $\mathcal{D}$, that is,

$$\boxed{\boldsymbol{J}_D \equiv \frac{\partial\mathcal{D}}{\partial t}} \qquad [\text{A/m}^2] \tag{6-34}$$

Although the displacement current density does not involve any motion of electric charges, it behaves like a conduction or convection current density as far as the time-varying magnetic field is concerned. In the parallel-plate capacitor, the conduction current is responsible for the magnetic field induced around the conducting lead, whereas the displacement current density is responsible for the magnetic field induced in the space between the conducting plates. The conduction and displacement currents may coexist in the lossy dielectric of a finite conductivity inserted between the conducting plates, both contributing to the magnetic field in the material. In this case, the sum of the conduction and displacement currents in the dielectric is equal to the conduction current in the capacitor lead, where the displacement current is ignored owing to the high conductivity.

Under time-varying conditions, the point form of Ampere's circuital law is modified in order to incorporate the displacement current density, that is,

$$\boxed{\nabla \times \mathscr{H} = \boldsymbol{J} + \frac{\partial \mathscr{D}}{\partial t}} \tag{6-35}$$

This equation is referred to as the generalized Ampere's law or Ampere's law. The current density $\boldsymbol{J}$ in Eq. (6-35) represents the conduction or the convection current varying with time. Applying Stokes's theorem to Eq. (6-35) we obtain the integral form of Ampere's law:

$$\boxed{\oint_C \mathscr{H} \bullet d\mathbf{l} = \int_s \left(\boldsymbol{J} + \frac{\partial \mathscr{D}}{\partial t} \right) \bullet d\mathbf{s}} \tag{6-36}$$

Let us now check if Eq. (6-35) satisfies the equation of continuity. By taking the divergence of both sides of Eq. (6-35), and applying the divergence theorem to the term on the left-hand side, we have

$$\nabla \bullet \nabla \times \mathscr{H} = \nabla \bullet \boldsymbol{J} + \frac{\partial}{\partial t}(\nabla \bullet \mathscr{D}) \tag{6-37}$$

For the right-most term in Eq. (6-37), the divergence and the time derivative were interchanged. By using the vector identity $\nabla \bullet \nabla \times \mathbf{U} = 0$ and Gauss's law $\nabla \bullet \mathscr{D} = \rho_v$, Eq. (6-37) reduces to

$$0 = \nabla \bullet \boldsymbol{J} + \frac{\partial \rho_v}{\partial t} \tag{6-38}$$

This equation always holds true because it is nothing but the equation of continuity.

Example 6-3

An ac voltage $v_{1-2} = V_o \sin \omega t$ is applied to a parallel-plate capacitor shown in Fig. 3.37. Ignoring the fringing effects at the edges, show that the displacement current in the lossless dielectric between the plates is equal to the conduction current in the capacitor lead.

Solution

Assuming a uniform $\mathscr{E}$ in the region between the plates, $\mathscr{D}$ in the region is obtained by use of the relations $\mathscr{V}_{1-2} = \mathscr{E}d$ and $\mathscr{D} = \varepsilon\mathscr{E}$:

$$\mathscr{D} = -\mathbf{a}_z \varepsilon \frac{V_o}{d} \sin \omega t$$

The displacement current density is therefore

$$\boldsymbol{J}_D = \frac{\partial \mathscr{D}}{\partial t} = -\mathbf{a}_z \omega\varepsilon \frac{V_o}{d} \cos \omega t$$

Total displacement current flowing in the -z-direction is

$$\begin{aligned} I_D = \int_{\mathcal{S}} J_D ds &= \mathcal{S}\omega\varepsilon \frac{V_o}{d} \cos \omega t \\ &= CV_o \omega \cos \omega t \end{aligned} \qquad (6\text{-}39)$$

In Eq. (6-39), $\mathcal{S}$ is the surface area of the conducting plates, and C is the capacitance given by $C = \varepsilon\mathcal{S}/d$

From circuit theory, the conduction current in the lead is

$$I_C = C\frac{dV_{1-2}}{dt} = CV_o \omega \cos \omega t \qquad (6\text{-}40)$$

Two results in Eqs. (6-39) and (6-40) are equal.

Exercise 6.7

For the time-varying electric field $\mathscr{E} = \mathbf{a}_z E_o \cos(\omega t - k\rho)/\rho\,[\mathrm{V/m}]$ given in free space, find the displacement current density in cylindrical coordinates.

Ans. $\boldsymbol{J}_D = -\mathbf{a}_z \omega\varepsilon_o E_o \sin(\omega t - k\rho)/\rho\,[\mathrm{A/m^2}]$.

Review Questions with Hints

RQ 6.5	What is the significance of displacement current density?	[Eq.(6-35)]
RQ 6.6	What is the unit of $\partial\mathscr{D}/\partial t$?	[Eq.(6-34)]
RQ 6.7	What physical law justifies the displacement current?	[Eq.(6-38)]

6.3 Maxwell's Equations

In 1873, James Clerk Maxwell published the unified theory of electricity and magnetism by formulating previously known experimental results of Coulomb, Gauss, Ampere, Faraday, and others, and by incorporating the concept of displacement current. The theory comprises four fundamental relations called Maxwell's equations that any electromagnetic field should satisfy under time-varying conditions. Maxwell's equations always hold regardless of the material medium. Maxwell's equations are available either in integral form or in differential form.

The integral form is advantageous in describing the underlying physical concepts, whereas the differential form or point form is advantageous in specifying the electromagnetic field intensities at each and every point in a given region of space.

Maxwell's equations in point form are

$$\boxed{\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t}} \tag{6-41a}$$

$$\boxed{\nabla \times \mathscr{H} = \boldsymbol{J} + \frac{\partial \mathscr{D}}{\partial t}} \tag{6-41b}$$

$$\boxed{\nabla \bullet \mathscr{D} = \rho_v} \tag{6-41c}$$

$$\boxed{\nabla \bullet \mathscr{B} = 0} \tag{6-41d}$$

Theses equations are individually referred to as Faraday's law, Ampere's law, Gauss's law, and Gauss's law for magnetism.

There are auxiliary equations that are essential for solving electromagnetic problems. The relation between $\mathscr{D}$ and $\mathscr{E}$, and that between $\mathscr{B}$ and $\mathscr{H}$ are called the constitutive relations of the electromagnetic medium:

$$\mathscr{D} = \varepsilon \mathscr{E} \tag{6-42a}$$

$$\mathscr{B} = \mu \mathscr{H} \tag{6-42b}$$

where ε and μ are the permittivity and permeability, respectively.

The conduction and convection current densities are, respectively, defined as

$$\boldsymbol{J} = \sigma \mathscr{E} \tag{6-43a}$$

$$\boldsymbol{J} = \rho_v \boldsymbol{v} \tag{6-43b}$$

where σ is the conductivity, ρ_v is the volume charge density, and $\boldsymbol{v}$ is the velocity of the charge density.

The total force exerted on a charge q moving in the presence of $\mathscr{E}$ and $\mathscr{B}$ is specified by Lorentz force equation given by

$$\boldsymbol{F} = q\left(\mathscr{E} + \boldsymbol{v} \times \mathscr{B}\right) \tag{6-44}$$

where $\boldsymbol{v}$ is the velocity of the charge.

Under time-varying conditions, the equation of continuity is

$$\nabla \bullet \boldsymbol{J} = -\frac{\partial \rho_v}{\partial t} \tag{6-45}$$

which is based on the principle of conservation of electric charges.

Example 6-4

For given $\mathscr{E} = E_o \mathbf{a}_x \cos(\omega t - kz)$ and $\mathscr{H} = E_o \mathbf{a}_y \sqrt{\varepsilon_o / \mu_o} \cos(\omega t - kz)$ in free space, with the propagation constant $k = \omega\sqrt{\mu_o \varepsilon_o}$, show that $\mathscr{E}$ and $\mathscr{H}$ satisfy the Maxwell's equations and thus constitute an electromagnetic field.

Solution

Substituting $\mathscr{E}$ and $\mathscr{H}$ into Eq. (6-41a), we obtain

$$kE_o\,\mathbf{a}_y \sin(\omega t - kz) = \omega E_o\,\mathbf{a}_y\sqrt{\mu_o\varepsilon_o}\sin(\omega t - kz) \tag{6-46a}$$

Substituting the fields into Eq. (6-41b), by assuming $\boldsymbol{J} = 0$ (free space), we obtain

$$-kE_o\,\mathbf{a}_x\sqrt{\varepsilon_o/\mu_o}\sin(\omega t - kz) = -\omega\varepsilon_o E_o\,\mathbf{a}_x \sin(\omega t - kz) \tag{6-46b}$$

Eqs. (6-46a) and (6-46b) always hold through the relation $k = \omega\sqrt{\mu_o\varepsilon_o}$.

Substituting the field into Eq. (6-41c), by assuming $\rho_v = 0$ (free space), we obtain

$$\nabla \bullet \mathscr{D} = \nabla \bullet [\varepsilon_o E_o\,\mathbf{a}_x \cos(\omega t - kz)] = 0 \tag{6-46c}$$

Substituting the field into Eq. (6-41d), we obtain

$$\nabla \bullet \left[\mu_o E_o \mathbf{a}_y\sqrt{\varepsilon_o/\mu_o}\cos(\omega t - kz)\right] = 0 \tag{6-46d}$$

Eq. (6-46c) holds because $\mathscr{D}$ is independent of x, and Eq. (6-46d) holds because $\mathscr{B}$ is independent of y. Thus, the given $\mathscr{E}$ and $\mathscr{H}$ satisfy Maxwell's equations, forming an electromagnetic field in free space.

Similarly, we can show that the fields $\mathscr{E} = E_o\,\mathbf{a}_x \cos(\omega t + kz)$ and $\mathscr{H} = -E_o\,\mathbf{a}_y\sqrt{\varepsilon_o/\mu_o}\cos(\omega t + kz)$ satisfy Maxwell's equations, and thus form an electromagnetic field in free space.

Exercise 6.8

Do the following vector fields satisfy Maxwell's equations in free space? (a) $\mathscr{E} = \mathbf{a}_x E_o e^{(\omega t - kz)}$, (b) $\mathscr{E} = \mathbf{a}_z E_o \cos(k\rho)\cos(\omega t)$, (c) $\mathscr{E} = \mathbf{a}_z E_o \cos(ky)\cos(\omega t)$, and (d) $\mathscr{E} = \mathbf{a}_z E_o \sin(\omega t - kz)$, where E_o , ω, and $k = \omega\sqrt{\varepsilon_o\mu_o}$ are constants, and ρ is the radial distance in cylindrical coordinates.
[Hint: Find $\mathscr{H}$ from $\mathscr{E}$ and see if it in turn induces $\mathscr{E}$]

Ans. (a) yes, (b) no, (c) yes, (d) no.

6.3.1 Maxwell's Equations in Integral Form

Maxwell's equations in point form can be converted to the integral form by using divergence and Stokes's theorems. Since the integral form involves geometric figures such as line, surface, and volume, it is useful for obtaining the boundary conditions for the electromagnetic field at an interface between two dissimilar media.

The integral form of Maxwell's equations is

$$\boxed{\oint_C \mathscr{E} \cdot d\mathbf{l} = -\int_S \frac{\partial \mathscr{B}}{\partial t} \cdot d\mathbf{s}} \tag{6-47a}$$

$$\boxed{\oint_C \mathscr{H} \cdot d\mathbf{l} = \int_S \left[\boldsymbol{J} + \frac{\partial \mathscr{D}}{\partial t} \right] \cdot d\mathbf{s}} \tag{6-47b}$$

$$\boxed{\oint_S \mathscr{D} \cdot d\mathbf{s} = \int_V \rho_v \, dv} \tag{6-47c}$$

$$\boxed{\oint_S \mathscr{B} \cdot d\mathbf{s} = 0} \tag{6-47d}$$

The directions of $d\mathbf{l}$ and $d\mathbf{s}$ are governed by the right-hand rule: $d\mathbf{s}$ on the surface S points in the direction of the right thumb when the fingers follow $d\mathbf{l}$ on the loop C.

6.3.2 Electromagnetic Boundary Conditions

In the previous Chapters 3 and 5 we derived the boundary conditions for **E**, **D**, **H**, and **B** at an interface between two different media by applying the fundamental relations for the static electric and magnetic fields. Following the same procedure used for the static fields, we can obtain the boundary conditions for $\mathscr{E}$, $\mathscr{D}$, $\mathscr{H}$, and $\mathscr{B}$ at an interface between two different materials. The boundary conditions for the tangential components of $\mathscr{E}$ and $\mathscr{H}$ are obtained by applying Faraday's law and Ampere's law to a rectangular loop straddling the interface, whereas the boundary conditions for the normal components of $\mathscr{D}$ and $\mathscr{B}$ are obtained by applying Gauss's law and Gauss's law for magnetism to a circular cylinder extending from one material to the other. Although Faraday's law includes the extra term $\partial \mathscr{B} / \partial t$ and Ampere's law includes the displacement current density $\partial \mathscr{D} / \partial t$ under time-varying conditions, these additional terms have no effect on the boundary conditions, because the integral of these terms over the surface enclosed by the loop is zero as the height of the loop tends to zero. Consequently, the previous boundary conditions for the static fields remain valid for the time-varying fields.

To summarize, the boundary conditions for the tangential components of $\mathscr{E}$ and $\mathscr{H}$ are

$$\boxed{\mathscr{E}_{1t} = \mathscr{E}_{2t}} \tag{6-48a}$$

$$\boxed{\mathscr{H}_{1t} - \mathscr{H}_{2t} = J_s} \tag{6-48b}$$

where t stands for the tangential component, and J_s represents the surface current flowing on the interface in the direction normal to the tangential components of $\mathscr{H}$.

The boundary conditions for the normal components of $\mathscr{D}$ and $\mathscr{B}$ are

$$\boxed{\mathscr{D}_{1n} - \mathscr{D}_{2n} = \rho_s} \tag{6-48c}$$

$$\boxed{\mathscr{B}_{1n} = \mathscr{B}_{2n}} \tag{6-48d}$$

where n stands for the normal component, and ρ_s is the surface charge density existing on the interface.

Let us now examine two most important interfaces frequently encountered in the study of electromagnetics: (1) the interface between two lossless dielectrics (2) the interface between a lossless dielectric and a perfect conductor.

Because lossless dielectrics have a zero conductivity ($\sigma = 0$), there is no surface charge and no surface current at the interface between two lossless dielectrics, that is,

$$\rho_s = J_s = 0 \tag{6-49}$$

The boundary conditions for the time-varying electric and magnetic fields at the interface between two lossless dielectrics are therefore

$$\boxed{\mathscr{E}_{1t} = \mathscr{E}_{2t}} \tag{6-50a}$$

$$\boxed{\mathscr{H}_{1t} = \mathscr{H}_{2t}} \tag{6-50b}$$

$$\boxed{\mathscr{D}_{1n} = \mathscr{D}_{2n}} \tag{6-50c}$$

$$\boxed{\mathscr{B}_{1n} = \mathscr{B}_{2n}} \tag{6-50d}$$

The tangential components of $\mathscr{E}$ and $\mathscr{H}$ and the normal components of $\mathscr{D}$ and $\mathscr{B}$ are all continuous across the interface.

Second, we examine perfect conductors having an infinite conductivity ($\sigma = \infty$). In most practical cases, good conductors such as silver, copper, gold, and aluminum with a high conductivity of the order of $\sigma \sim 10^7$ [S/m] can be regarded as perfect conductors as far as the boundary condition is concerned. The infinite conductivity is connected to a unique characteristic of the perfect conductor, that is, a zero electric field inside the perfect conductor. Thus, any net charges the perfect conductor will have should reside on the surface only, and any currents the perfect conductor will carry should flow on the surface only. The interrelationship between $\mathscr{E}$ and $\mathscr{H}$ assures that the time-varying magnetic field is also zero in the perfect conductor. Thus, in region 2 occupied by a perfect conductor, we have

$$\mathscr{E}_2 = 0 = \mathscr{D}_2 \tag{6-51a}$$

$$\mathscr{H}_2 = 0 = \mathscr{B}_2 \tag{6-51b}$$

The boundary conditions at an interface between a lossless dielectric(region 1) and a perfect conductor(region 2) are therefore

$$\boxed{\mathscr{E}_{1t} = 0} \tag{6-52a}$$

$$\boxed{\mathscr{H}_{1t} = J_s} \tag{6-52b}$$

$$\boxed{\mathscr{D}_{1n} = \rho_s} \tag{6-52c}$$

$$\boxed{\mathscr{B}_{1n} = 0} \tag{6-52d}$$

where t and n denote the tangential and normal components, respectively. It is important to remember that the unit surface normal is in the direction away from the conductor. For instance, $\mathscr{D}_{1n}$ is the normal component of $\mathscr{D}$, at the interface, pointing out of the conductor.

The electric field intensity must be zero in the perfect conductor owing to the infinite conductivity of the material, whether or not it is time-varying. Although $\mathscr{H}$ is also zero in the perfect conductor, owing to the interrelationship between $\mathscr{E}$ and $\mathscr{H}$, the static magnetic field **H** may not necessarily be zero in the perfect conductor because it is not coupled to **E**. If there is a static surface charge ρ_s on the interface between a lossless dielectric and a perfect conductor, the surface charge may be interpreted either as the net charge on the conductor, which will generate an electric field in the dielectric, or as the surface charge induced on the conductor by the electric field that is generated elsewhere and terminated on the conductor. Although the steady surface current $\mathbf{J}_s$ on the perfect conductor may be responsible for **H** existing in the dielectric and the conductor, it may not be the one that is induced on the conductor by an external static magnetic field. This is because the tangential component of an external field **H** may be continuous across the interface, or the surface of the perfect conductor, without inducing a surface current. However, under time-varying conditions, we see from the boundary conditions expressed by Eqs. (6-51b) and (6-52b) that the time-varying $\boldsymbol{J}_s$ flowing on the perfect conductor may be interpreted either as the source current responsible for $\mathscr{E}$ and $\mathscr{H}$ existing outside the conductor or as the surface current induced on the conductor by an external electromagnetic field impinging on the conductor. Similarly, we see from Eqs. (6-51a) and (6-52c) that the time-varying ρ_s may be interpreted either as the source charge producing an electromagnetic wave outside the conductor or as the surface charge induced on the conductor by an external electromagnetic wave impinging on the conductor.

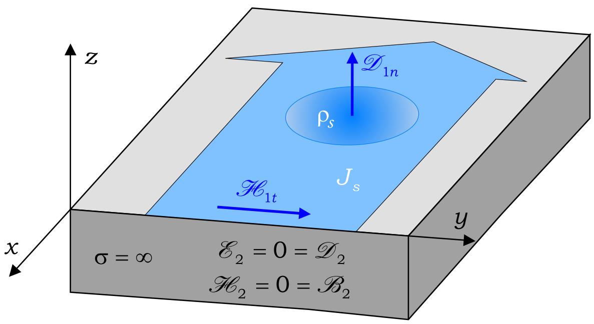


Fig. 6.9 An interface between a perfect conductor and a lossless dielectric.

Exercise 6.9

Two fields $\mathscr{H}_1 = \mathbf{a}_x H_o \cos(\omega t - kz)$ and $\mathscr{H}_2 = \mathbf{a}_x H_o \cos(\omega t + kz)$ coexist in free space $(z > 0)$. Find $\boldsymbol{J}_s$ induced on the perfect conductor($z \le 0$).

Ans. $\boldsymbol{J}_s = \mathbf{a}_y 2H_o \cos(\omega t)$ [A/m] .

Review Questions with Hints

RQ 6.8 Write Maxwell's equations in point form. [Eq.(6-41)]

RQ 6.9 Write Maxwell's equations in integral form. [Eq.(6-47)]

RQ 6.10 What are the experimental laws that Maxwell's equations are rooted in? [Eq.(6-41)]

RQ 6.11 What causes $\mathscr{E}$ and $\mathscr{H}$ to be coupled to each other? [Eq.(6-41b)]

RQ 6.12 State the general boundary conditions for $\mathscr{E}$ and $\mathscr{H}$ [Eq.(6-48)]

RQ 6.13 Why is it that $\mathscr{E} = 0 = \mathscr{H}$ in a perfect conductor?[Eqs.(6-43a)(6-41a)]

RQ 6.14 Why is it that $\mathscr{E}_t = 0$ and $\mathscr{H}_n = 0$ on the surface of a perfect conductor? [Eq.(6-51)]

6.4 Retarded Potentials

In Chapters 3 and 5 we saw that the electric potential *V*, due to a static charge ρ_v distributed in a volume $\mathcal{V}'$, is expressed by Eq. (3-68a), and the vector magnetic potential **A**, due to a steady current **J** distributed over $\mathcal{V}'$, is expressed by Eq. (5-39b). Rewriting them,

$$V(\mathbf{r}) = \int_{\mathcal{V}'} \frac{\rho_v(\mathbf{r}')dv'}{4\pi\varepsilon\mathcal{R}} \qquad \text{[V]} \qquad \text{(3-68a)(6-53a)}$$

$$\mathbf{A}(\mathbf{r}) = \int_{\mathcal{V}'} \frac{\mu\mathbf{J}(\mathbf{r}')dv'}{4\pi\mathcal{R}} \qquad \text{[Wb/m]} \qquad \text{(5-39b)(6-53b)}$$

where $\mathcal{R} = |\mathbf{r} - \mathbf{r}'|$ is the distance between the field point at position vector **r** and the source point at position vector $\mathbf{r}'$. When the source charge and current vary in time, one might be tempted to rewrite the above equations by replacing $\rho_v(\mathbf{r}')$ with $\rho_v(\mathbf{r}', t)$, and $\mathbf{J}(\mathbf{r}')$ with $\boldsymbol{J}(\mathbf{r}', t)$, for the time-varying potentials. Even if the negative gradient of $V(\mathbf{r}, t)$ and the curl of $\boldsymbol{A}(\mathbf{r}, t)$ may be computed mathematically, they do not represent the time-varying electric field and time-varying magnetic flux density, because they do not satisfy Maxwell's equations.

If the source charge or current varies with time, the potential at far distances cannot respond instantaneously. For instance, let V_1 be the static electric potential observed at **r**, which is due to a static charge ρ_1 located at $\mathbf{r}'$. If the charge is suddenly changed from ρ_1 to ρ_2 at $\mathbf{r}'$, it takes a finite period of time for the potential observed at **r** to change from V_1 to V_2. In other words, the response to the change occurred at $\mathbf{r}'$ is observed at **r** at a later time, delayed by $\mathcal{R}/v$, where v is the speed of propagation from the source to the observation point, and $\mathcal{R}$ is the distance between the two points in space. Thus the potentials $V(\mathbf{r}, t_1)$ and $\boldsymbol{A}(\mathbf{r}, t_1)$ at time $t = t_1$ correspond to the sources $\rho_v(\mathbf{r}', t_1 - \mathcal{R}/v)$ and $\boldsymbol{J}(\mathbf{r}', t_1 - \mathcal{R}/v)$ at an earlier time $t = t_1 - \mathcal{R}/v$. In view of this reaction time, the time-varying electric potential and vector magnetic potential at position vector **r** are written as

$$\boxed{V(\mathbf{r},t) = \int_{\mathcal{V}'} \frac{\rho_v(t-\mathcal{R}/v)dv'}{4\pi\varepsilon\mathcal{R}}} \qquad [\text{V}] \qquad (6\text{-}54a)$$

$$\boxed{\boldsymbol{A}(\mathbf{r},t) = \int_{\mathcal{V}'} \frac{\mu \boldsymbol{J}(t-\mathcal{R}/v)}{4\pi\mathcal{R}}dv'} \qquad [\text{Wb/m}] \qquad (6\text{-}54b)$$

which are called the retarded scalar potential and retarded vector potential, respectively. In free space, the speed of light is $v = c = 1/\sqrt{\varepsilon_o\mu_o}$, where ε_o is the permittivity and μ_o is the permeability of free space.

The retarded potentials are very useful for solving radiation problems. In the following we will begin with the definition of the vector magnetic potential ***A***, and derive useful formulas related to the retarded potential. Following the same way in which the vector magnetic potential **A** in magnetostatics is defined from the solenoidal nature of **B** and the vector identity $\nabla \bullet (\nabla \times \mathbf{U}) = 0$, the time-varying vector magnetic potential ***A*** is defined by

$$\boxed{\mathscr{B} = \nabla \times \boldsymbol{A}} \qquad (6\text{-}55)$$

Inserting Eq. (6-55) into Faraday's law, we have

$$\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t} = -\frac{\partial}{\partial t}(\nabla \times \boldsymbol{A}) \qquad (6\text{-}56)$$

Interchanging the time derivative and the curl on the right-hand side of Eq. (6-56), we get

$$\nabla \times \left(\mathscr{E} + \frac{\partial \boldsymbol{A}}{\partial t} \right) = 0 \qquad (6\text{-}57)$$

Comparison of Eq. (6-57) with the vector identity $\nabla \times (\nabla W) = 0$ leads to the definition of a time-varying scalar field V, that is,

$$\mathscr{E} + \frac{\partial \boldsymbol{A}}{\partial t} = -\nabla V \qquad (6\text{-}58)$$

or

$$\boxed{\mathscr{E} = -\nabla V - \frac{\partial \boldsymbol{A}}{\partial t}} \qquad (6\text{-}59)$$

We note that Eq. (6-59) reduces to the famous relation, $\mathbf{E} = -\nabla V$, under static conditions in which $\partial \boldsymbol{A}/\partial t = 0$ is assumed.

Next, inserting Eq. (6-55) into Ampere's law expressed by Eq. (6-41b) we obtain

$$\nabla \times \nabla \times \boldsymbol{A} = \mu \boldsymbol{J} + \mu \frac{\partial \mathscr{D}}{\partial t} \qquad (6\text{-}60)$$

Inserting Eq. (6-59) into Eq. (6-60) and applying the vector identity $\nabla\times\nabla\times\mathbf{U} = \nabla(\nabla\bullet\mathbf{U}) - \nabla^2\mathbf{U}$ we obtain

$$\nabla^2\boldsymbol{A} - \mu\varepsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2} = -\mu\boldsymbol{J} + \nabla\left(\nabla\bullet\boldsymbol{A} + \mu\varepsilon\frac{\partial V}{\partial t}\right) \tag{6-61}$$

where we used the constitutive relation $\mathscr{D} = \varepsilon\mathscr{E}$.

If the electric scalar potential V and the vector magnetic potential $\boldsymbol{A}$ were obtained from Eq. (6-61), we could determine $\mathscr{E}$ and $\mathscr{B}$ from Eq. (6-59) and Eq. (6-55), respectively. We know from linear algebra that two equations are required for a unique determination of two unknowns. Moreover, Helmholtz's theorem requires the specification of the divergence of $\boldsymbol{A}$, in addition to the curl of $\boldsymbol{A}$ given in Eq. (6-55), for a unique determination of $\boldsymbol{A}$. In view of these considerations, if we let

$$\boxed{\nabla\bullet\boldsymbol{A} = -\mu\varepsilon\frac{\partial V}{\partial t}} \tag{6-62}$$

which is called the Lorentz condition for potentials, then Eq. (6-61) reduces to the inhomogeneous wave equation for vector potential, that is,

$$\boxed{\nabla^2\boldsymbol{A} - \mu\varepsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2} = -\mu\boldsymbol{J}} \tag{6-63}$$

Under static conditions, Eq. (6-62) reduces to $\nabla\bullet\mathbf{A} = 0$ as given in Eq. (5-47), while Eq. (6-63) reduces to the vector Poisson's equation as given in Eq. (5-53).

Substitution of Eq. (6-59) into Gauss's law leads to

$$\begin{aligned}\frac{\rho_v}{\varepsilon} &= \nabla\bullet\left(-\nabla V - \frac{\partial\boldsymbol{A}}{\partial t}\right)\\ &= -\nabla^2 V - \frac{\partial}{\partial t}(\nabla\bullet\boldsymbol{A})\end{aligned} \tag{6-64}$$

Upon inserting the Lorentz condition expressed by Eq. (6-62) into Eq. (6-64), we obtain the inhomogeneous wave equation for scalar potential, that is,

$$\boxed{\nabla^2 V - \mu\varepsilon\frac{\partial^2 V}{\partial t^2} = -\frac{\rho_v}{\varepsilon}} \tag{6-65}$$

Under static conditions, Eq. (6-65) reduces to Poisson's equation. We see from Eqs. (6-63) and (6-65) that the wave equations for $\boldsymbol{A}$ and V are decoupled from each other through the Lorentz condition, and there exists a remarkable symmetry between the two equations.

If we set $\rho_v = 0 = \boldsymbol{J}$ in Eq. (6-63) and Eq. (6-65), the equations reduce to the homogeneous differential wave equations for $\boldsymbol{A}$ and V, respectively, having general solutions of the form $\boldsymbol{U}(t \pm \mathcal{R}/\upsilon)$ and $W(t \pm \mathcal{R}/\upsilon)$. These solutions represent the waves propagating through the source free region with a velocity $\upsilon = 1/\sqrt{\mu\varepsilon}$.

Example 6-5
Find the retarded vector magnetic potential at point p in free space, which is caused by an incremental current element of a linear dimension h[m] located at the origin, carrying an ac current $I(t) = I_o \cos(\omega t)$ [A].

Solution

Rewriting Eq. (6-54b) for a line current, with the speed of light in free space $c = 1/\sqrt{\varepsilon_o \mu_o}$, we have

$$\boldsymbol{A}(\mathbf{r}, t) = \int_{L'} \frac{\mu_o I(t - \mathcal{R}/c)}{4\pi\mathcal{R}} \mathbf{a}_z dl' \tag{6-66}$$

Assuming h to be vanishingly small, we use an approximate expression, $\mathcal{R} = |\mathbf{r} - \mathbf{r}'| \cong R$, where $\mathbf{r}'$ is the position vector of a point on the line current and R is the radial distance in spherical coordinates. Evaluating the line integral in Eq. (6-66), we get

$$\boldsymbol{A}(\mathbf{r}, t) = \frac{\mu_o h}{4\pi R} I_o \cos[\omega(t - R/c)] \mathbf{a}_z \tag{6-67}$$

where c is the speed of light in free space. The vector magnetic potential at a distance R from the origin is retarded by R/c in time with respect to the current at the origin. The retardation time is the travel time of the electromagnetic wave from the origin to the field point p.

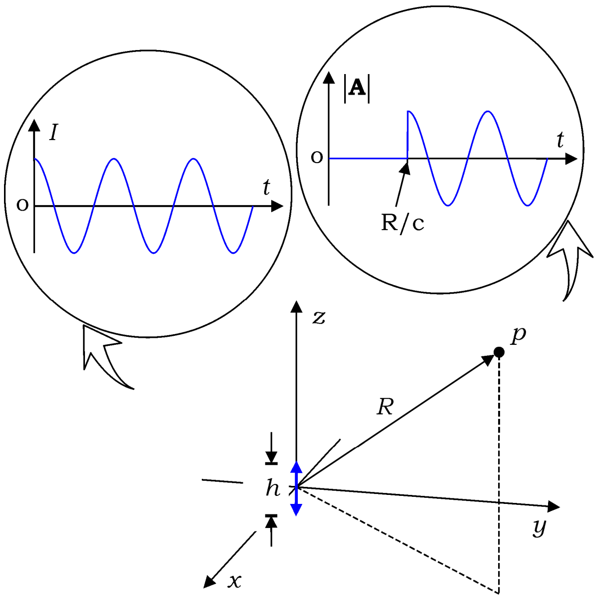


Fig. 6.10 Retarded potential due to a current element at the origin.

Exercise 6.10
Derive the point form of Ampere's circuital law by assuming static field conditions in Eqs. (6-55), (6-62) and (6-63).

Ans. $\nabla \times \mathbf{H} = \mathbf{J}$.

Exercise 6.11
Derive Faraday's law from Eq. (6-59).

Ans. $\nabla \times \mathscr{E} = -\dfrac{\partial \mathscr{B}}{\partial t}$.

Review Questions with Hints

RQ 6.15 Define scalar electric potential V and vector magnetic potential $\boldsymbol{A}$ under time-varying conditions. [Eqs.(6-59)(6-55)]
RQ 6.16 Explain how $\mathscr{E}$ and $\mathscr{B}$ are obtained from V and $\boldsymbol{A}$. [Eqs.(6-59)(6-55)]
RQ 6.17 Express inhomogeneous wave equations for V and $\boldsymbol{A}$. [Eqs.(6-65)(6-63)]
RQ 6.18 Do V and $\boldsymbol{A}$ always form waves propagating through space? [Eq.(6-54)]

Problems

6-1 Find the electromotive force in a stationary circular loop of radius a placed in the xy-plane with the center at the origin, which is caused by the magnetic flux density $\mathscr{B}(t) = \mathbf{a}_z B_o \rho^{-1} \cos(\omega t - k\rho)$, where ρ is the radial distance in cylindrical coordinates, t is time, and B_o, ω, and k are constants.

6-2 A rectangular wire loop is connected to a resistor R and placed in the yz-plane in the magnetic flux density $\mathscr{B} = B_o\, \mathbf{a}_x \sin(ky)\sin(\omega t)$ as shown in Fig. 6.11. Find
(a) voltage $\mathscr{V}_{1-2}$ induced across terminals 1 and 2,
(b) current flowing in the loop having a self-inductance L.

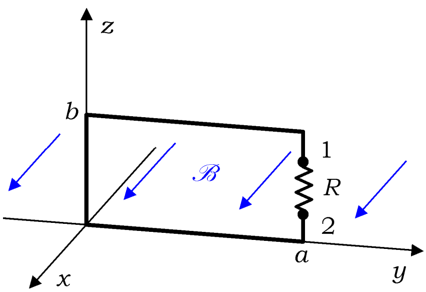


Fig. 6.11 A rectangular wire loop (Problem 6-2).

6-3 A rectangular wire loop rotates about the z-axis with an angular velocity ω_o in the static magnetic flux density $\mathbf{B} = B_o \mathbf{a}_x$ as shown in Fig. 6.12. Find the induced voltage in the loop by calculating
(a) motional emf,
(b) magnetic flux linkage with the loop.

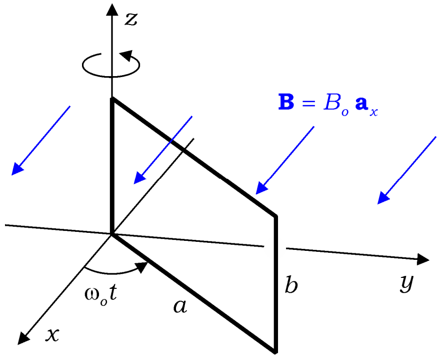


Fig. 6.12 A wire loop rotates in a static **B** (Problem 6-3).

6-4 With reference to Fig. 5.30 shown in **Example 5-15**, in which the rectangular loop is now moving in the $\mathbf{a}_\rho$-direction with a constant velocity $\mathbf{v} = v_o \mathbf{a}_\rho$ in the presence of **B**, which is produced by the current I flowing in the straight wire oriented along the z-axis, find emf induced in the loop.

6-5 A conducting disk of radius a rotates with an angular velocity ω_o in a uniform static field $\mathbf{B} = B_o \mathbf{a}_z$, as shown in Fig. 6.13. Find the induced emf between terminals 1 and 2, assuming the shaft to be negligibly thin.

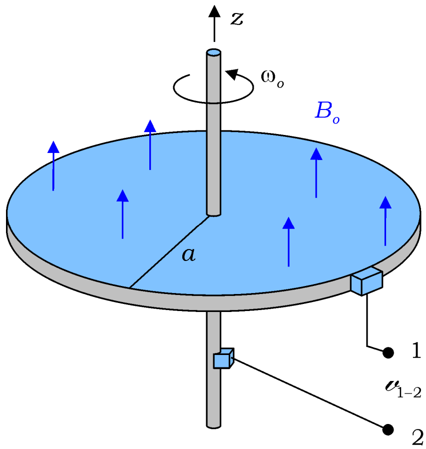


Fig. 6.13 A conducting disk rotates in a static field **B** (Problem 6-5).

6-6 A conducting disk of a radius a and a small height h is made of a material of a finite conductivity σ and a permeability μ_o. It is placed on the *xy*-plane in the presence of a uniform, time-varying, magnetic flux density $\mathscr{B} = \mathbf{a}_z B_o \cos \omega t$ as shown in Fig. 6.14. Ignoring time-delay of *emf* at different points on the disk, and neglecting the magnetic field induced by the current in the disk, compute
(a) induced *emf* along a circle of radius ρ,
(b) total time-average power dissipated in the disk.

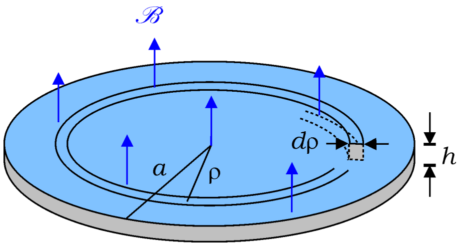


Fig. 6.14 A conducting disk in a time varying field $\mathscr{B}$ (Problem 6-6).

6-7 An electric field is given by $\mathscr{E} = E_o \mathbf{a}_x \cos(\omega t - kz)$ in a lossy dielectric of a finite conductivity σ and permittivity ε. Find
(a) conduction current density $\boldsymbol{J}_C$,
(b) displacement current density $\boldsymbol{J}_D$,
(c) ratio between the amplitudes of $\boldsymbol{J}_C$ and $\boldsymbol{J}_D$.

6-8 A coaxial capacitor of a finite length $\mathcal{L}$ consists of two cylindrical shells of radius a and radius b ($a < b$). It is filled with a dielectric(ε and μ_o), and connected to an ac-voltage source, $\mathscr{v} = V_o \cos \omega t$. Ignoring the fringing effects at the edges and the retardation of $\mathscr{v}$ in the gap, find
(a) conduction current,
(b) displacement current in the gap.

6-9 A lossless dielectric with ε and μ_o is infinite in extent.
(a) Write Maxwell's equations in point form,
(b) For the electric field $\mathscr{E} = (E_1 \mathbf{a}_x + E_2 \mathbf{a}_y)\cos(\omega t - kz)$, find $\mathscr{H}$, and
(c) Express k in terms of ω, ε, and μ_o.

6-10 Consider a time-varying but spatially uniform field $\mathscr{E}(t) = E_o \mathbf{a}_x \cos \omega t$. To show that the given field cannot be an electromagnetic field in free space, follow the steps below:
(a) Obtain $\mathscr{H}$ by inserting the given $\mathscr{E}$ into Maxwell's equations.
(b) Obtain $\mathscr{E}$ by inserting $\mathscr{H}$ given in part (a) into Maxwell's equations.
(c) Compare $\mathscr{E}$ obtained in part (b) with the given $\mathscr{E}$.

6-11 For the electric field $\mathscr{E} = \mathbf{a}_x f(t - z\sqrt{\mu_o\varepsilon_o})$ given in free space,

(a) obtain $\mathscr{H}$ by substituting $\mathscr{E}$ into Maxwell's equations,

(b) show that both $\mathscr{E}$ and $\mathscr{H}$ satisfy Maxwell's four equations,

(c) under what condition for f do both $\mathscr{E}$ and $\mathscr{H}$ satisfy Maxwell's equations?

6-12 The $z = 0$ plane is an interface between free space $(z < 0)$ and a lossless dielectric $(z \geq 0)$ with $\varepsilon = n^2\varepsilon_o$ and μ_o. The electric field is given as $\mathscr{E}_o = E_i\, \mathbf{a}_x \cos(\omega t - z\omega\sqrt{\mu_o\varepsilon_o}) + E_r\, \mathbf{a}_x \cos(\omega t + z\omega\sqrt{\mu_o\varepsilon_o})$ in free space, whereas the electric field is given as $\mathscr{E}_1 = E_t\, \mathbf{a}_x \cos(\omega t - zn\omega\sqrt{\mu_o\varepsilon_o})$ in the dielectric(n, ω, E_i, E_r, and E_t are constants).

(a) Show that the magnetic fields are given, in their respective regions, as

$$\mathscr{H}_o = \sqrt{\varepsilon_o / \mu_o}\left[E_i\, \mathbf{a}_y \cos(\omega t - z\omega\sqrt{\mu_o\varepsilon_o}) - E_r\, \mathbf{a}_y \cos(\omega t + z\omega\sqrt{\mu_o\varepsilon_o})\right]$$

$$\mathscr{H}_1 = n\sqrt{\varepsilon_o / \mu_o}\, E_t\, \mathbf{a}_y \cos(\omega t - zn\omega\sqrt{\mu_o\varepsilon_o}).$$

(b) Express E_t and E_r in terms of E_i and n[Hint: boundary conditions].

6-13 The $z = 0$ plane is an interface between a perfect conductor $(z \geq 0)$ and free space $(z < 0)$. Two electric fields coexist in free space such that $\mathscr{E}_1 = E_1\, \mathbf{a}_x \cos(\omega t - k_y y - k_z z)$ and $\mathscr{E}_2 = E_2\, \mathbf{a}_x \cos(\omega t - k_y y + k_z z)$, where E_1, E_2, ω, k_y, and k_z are constants. Find

(a) ratio E_1 / E_2,

(b) magnetic field in free space,

(c) surface current density induced on the interface at $z = 0$.

6-14 Given the vector magnetic potential $\boldsymbol{A} = A_o\mathbf{a}_x \cos\left[\omega(t - z / c)\right]$ in free space, find $\mathscr{E}$ by making use of

(a) $\mathscr{B} = \nabla \times \boldsymbol{A}$ and $\nabla \times \mathscr{H} = \dfrac{\partial \mathscr{D}}{\partial t}$.

(b) $\nabla \bullet \boldsymbol{A} = -\mu\varepsilon \dfrac{\partial V}{\partial t}$ and $\mathscr{E} = -\nabla V - \dfrac{\partial \boldsymbol{A}}{\partial t}$.

6-15 A vector magnetic potential is given as $\boldsymbol{A} = A_o\mathbf{a}_z(1 / R)\cos\left[\omega(t - R / c)\right]$ in free space, where R is the radial distance in spherical coordinates, and the speed of light $c = 1 / \sqrt{\varepsilon_o\mu_o}$. Show that the electric and magnetic fields are given as $\mathscr{E} = -A_o\omega\mathbf{a}_\theta(1 / R)\sin\left[\omega(t - R / c)\right]\sin\theta$ and $\mathscr{H} = -A_o\omega\sqrt{\varepsilon_o / \mu_o}\,\mathbf{a}_\phi(1 / R)\sin\left[\omega(t - R / c)\right]\sin\theta$, respectively, at far distances($R \to \infty$).

Chapter 7
Wave Motion

7.1 One-Dimensional Waves

A wave is the disturbance of a medium that travels through space with no change in its shape. A mechanical wave in a stretched spring, a water wave on the surface of a lake, and a sound wave in the air are good examples of traveling waves. There are two types of traveling waves. For instance, the wave in a spring as shown in Fig. 7.1(a) is a longitudinal wave for which the medium is displaced in the same direction as the direction of propagation of the wave. The other wave as shown in Fig. 7.1(b) is a transverse wave for which the medium is displaced in the direction perpendicular to the direction of propagation of the wave. Although a wave can transport energy from one point to another in space, the material medium itself does not advance in space. This is the reason that a wave can propagate at a great speed.

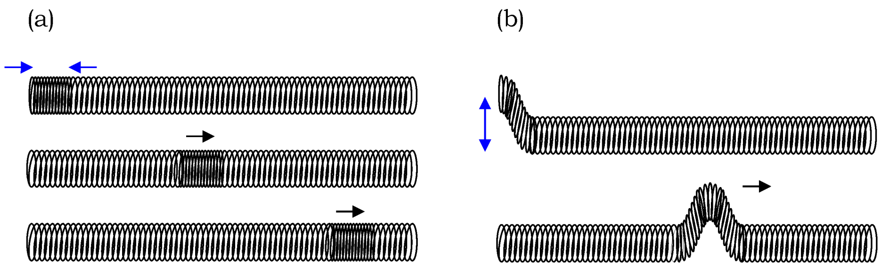


Fig. 7.1 Waves in a spring: (a) longitudinal wave, and (b) transverse wave.

The wave shown in Fig. 7.1(b) propagates along the horizontal axis, which is designated $+x$-axis. The disturbance of the medium, or the spatial variation of the spring, varies along the x-axis only, and thus the wave is called a one-dimensional wave. The expression for the disturbance is called the wavefunction, which is generally given as functions of space and time such as

$$\psi = f(x, t) \tag{7-1}$$

To be a wavefunction, it is necessary that the function f is smooth in space and time so that it is differentiable with respect to x and t. The wavefunction contains all the information about the wave such as wave shape, propagation direction, and wave velocity. Although a traveling wave may be well described mathematically by its wavefunction, unfortunately, it cannot be drawn on a paper because its disturbance varies simultaneously with position and time. To avoid this difficulty, we take a detour. First, by holding the variable t constant at a particular time, we draw ψ as a function of position, which is called the profile of the wave. The profile corresponds to taking a picture of the wave at that time. Second, by fixing the variable x at a point in space, we draw ψ as a function of time.

Consider Fig. 7.2, which simultaneously shows the two profiles of a wave taken at $t = 0$ and $t = t_1$, respectively. The first one is described by a function $f(x)$, and the second one by a function $\bar{f}(x)$, namely

$$\psi(x, 0) = f(x) \tag{7-2a}$$

$$\psi(x, t_1) = \bar{f}(x) \tag{7-2b}$$

where ψ denotes the wavefunction, in general, while f and $\bar{f}$ represent wavefunctions, in particular, at times $t = 0$ and t_1. As we see in Fig. 7.2, the second profile is displaced to the right by a distance vt_1, with respect to the first one, where v is the velocity of the wave. In view of this, we write the second profile as

$$\psi(x, t_1) = f(x - vt_1) \tag{7-3}$$

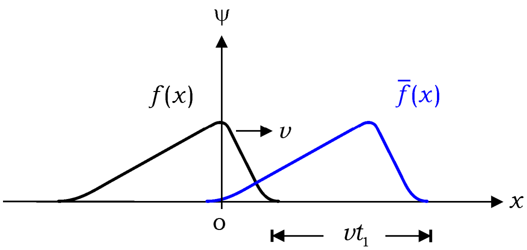


Fig. 7.2 Two profiles of a wave taken at $t = 0$ and $t = t_1$: $\psi(x, 0) = f(x)$ and $\psi(x, t_1) = \bar{f}(x)$.

Allowing t_1 in Eq. (7-3) to vary between zero and infinity, we can obtain a general expression for the wavefunction as

$$\boxed{\psi(x, t) = f(x - vt)} \tag{7-4}$$

If the wave is traveling in the $-x$-direction, its wavefunction is, of course, expressed as

$$\boxed{\psi(x, t) = f(x + vt)} \tag{7-5}$$

The two waves expressed by Eqs. (7-4) and (7-5) have the same profile at time $t = 0$. At time $t > 0$, however, two waves move farther away from each other as illustrated in Fig. 7.3.

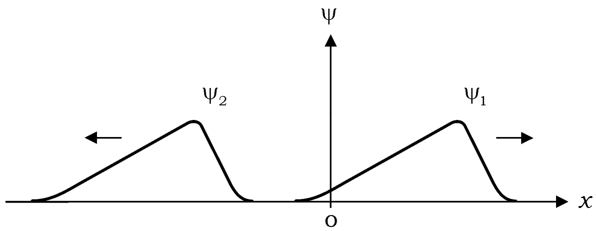


Fig. 7.3 Two waves of $\psi_1 = f(x - \upsilon t)$ and $\psi_2 = f(x + \upsilon t)$.

The differential wave equation had been well known long before Maxwell and used for describing wave motions in different types of media. In a lossless medium, it is a homogeneous, second order, linear, partial differential equation. In one-dimensional space, the differential wave equation becomes

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{\upsilon^2} \frac{\partial^2 \psi}{\partial t^2} \tag{7-6}$$

where the wave velocity υ is constant in a homogeneous and linear medium. As we will see in Section 7-3, Maxwell's equations can be combined into a three-dimensional differential wave equation. However, in this section, we focus our discussion on the wave propagating in one dimensional space. Through direct substitution, we can show that the second-order differential wave equation expressed by Eq. (7-6) has two independent solutions of the form

$$\psi = f(x \pm \upsilon t) \tag{7-7}$$

By direct substitution we can show that the following formulas are also solutions to Eq. (7-6):

$$\psi = f(\upsilon t \pm x) \tag{7-8}$$

It is important to note that the function *f* in Eqs. (7-7) and (7-8) may be arbitrary only if it is twice differentiable. The plus sign in the argument signifies that the wave propagates in the $-x$-direction, whereas the minus sign signifies that the wave propagates in the $+x$-direction.

Example 7-1
If a function $f(x)$ is given as in Fig. 7.4, sketch the two traveling waves expressed by $\psi_1 = f(x - \upsilon t)$ and $\psi_2 = f(\upsilon t - x)$ as functions of position at time $t = t_1$.

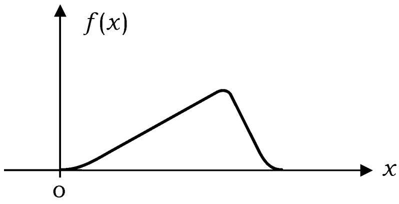


Fig. 7.4 A smooth function.

Solution

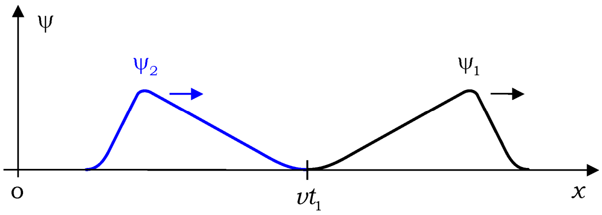


Fig. 7.5 Two waves, $\psi_1 = f(x - vt_1)$ and $\psi_2 = f(vt_1 - x)$.

Example 7-2

Do the following functions represent traveling waves in a homogenous and linear medium? (a) $\psi = e^{-(x-3t)^2}$, and (b) $\psi = \sin(2xt)$.

Solution

(a) $$\frac{\partial^2\psi}{\partial x^2} = -2e^{-(x-3t)^2} + 4(x-3t)^2 e^{-(x-3t)^2}$$

$$\frac{\partial^2\psi}{\partial t^2} = -18e^{-(x-3t)^2} + 36(x-3t)^2 e^{-(x-3t)^2}$$

Taking the ratio between the two terms, we have

$$v^2 = \frac{\partial^2\psi}{\partial t^2} \Big/ \frac{\partial^2\psi}{\partial x^2} = 9$$

It is a wave traveling with a constant speed $v = 3$.

(b) $$\frac{\partial^2\psi}{\partial x^2} = -(2t)^2 \sin(2xt)$$

$$\frac{\partial^2\psi}{\partial t^2} = -(2x)^2 \sin(2xt)$$

$$\frac{\partial^2\psi}{\partial t^2} \Big/ \frac{\partial^2\psi}{\partial x^2} = \left(\frac{x}{t}\right)^2 \neq \text{constant}$$

It is not a traveling wave.

Exercise 7.1
Determine the propagation direction of the following waves: (a) $\psi = f(-x-t)$, (b) $\psi = 10\sin^2(-z+5t)$, and (c) $\psi = \cos(4z-3t+10)$.

Ans. (a) -x-direction, (b) +z-direction, (c) +z-direction.

Exercise 7.2
Show by direct substitution that the following functions satisfy one-dimensional differential wave equation: (a) $\psi = f(-x-\upsilon t)$, (b) $\psi = f(t+x/\upsilon)$, and (c) $\psi = f(t-x/\upsilon)$.

Exercise 7.3
Is ψ a longitudinal or transverse wave ?
(a) $\psi = 5\,\mathbf{a}_x\cos(2x-3t)$, and (b) $\psi = 4\,\mathbf{a}_y\cos(2x-3t)$

Ans. (a) Longitudinal wave, (b) Transverse wave.

Review Questions with Hints

RQ 7.1 Why is it necessary for the wavefunction to be smooth in space and time in a homogeneous and linear medium? [Eq.(7-6)]
RQ 7.2 What determines the direction of propagation of the wave in the wave-function? [Eq.(7-7)]

7.1.1 Harmonic Waves

If the profile of a wave is given by a sine or cosine curve, the wave is referred to as a harmonic wave. It is customary to use cosine function for the harmonic wave. The wavefunction of a harmonic wave is written in general as

$$\boxed{\psi = A\cos\left[k(x-\upsilon t)+\varphi_o\right] = A\cos\left[kx-\omega t+\varphi_o\right]} \tag{7-9}$$

where A is the amplitude, k is the propagation constant, ω is the angular frequency, which is the frequency multiplied by 2π: $\omega = 2\pi f$. The entire argument of the cosine function in Eq. (7-9) is called the phase. The phase of the wave, φ, is functions of space and time, that is,

$$\varphi = kx - \omega t + \varphi_o \tag{7-10}$$

where ϕ_o is a constant called an initial phase.

The phase velocity of a harmonic wave is the velocity with which the crest of the wave travels along the +x-axis. For instance, if a harmonic wave travels with a phase velocity υ_p, the crest of the wave is displaced along the x-axis by a distance $\upsilon_p\Delta t$ in a time Δt. In this case, if the displacement $\upsilon_p\Delta t$ corresponds to a distance Δx along the +x-axis, or $\Delta x = \upsilon_p\Delta t$, it means that the crest at a point $x = x_1$ at time

$t = t_1$ will appear at another point $x = x_1 + \Delta x$ at a later time $t = t_1 + \Delta t$. These statements can be expressed in terms of the phase of the wave as follows:

$$\varphi_1 = kx_1 - \omega t_1 + \varphi_o = k(x_1 + \Delta x) - \omega(t_1 + \Delta t) + \varphi_o$$

This relation is based on the fact that the crests at different points in space are at the same phase; two crests with a phase difference of an integer multiple of 2π are indistinguishable in a harmonic wave. Rewriting the above equation we obtain

$$k\Delta x - \omega \Delta t = 0$$

As was discussed earlier, the phase velocity of the wave is defined by $\upsilon_p = \Delta x / \Delta t$. By inserting this relation into the above equation, we obtain the phase velocity as

$$\boxed{\upsilon_p = \frac{\omega}{k}} \qquad \text{[m/s]} \qquad (7\text{-}11)$$

The phase velocity of a harmonic wave is the ratio between the angular frequency and the propagation constant.

The harmonic wave is periodic both in space and time, and thus has two periods: spatial period, also called wavelength λ, and temporal period τ. If a harmonic wave is the end result of an electromagnetic process, the given system is said to be at sinusoidal steady-state. Under this condition, it is of little consequence to distinguish between the crests of a harmonic wave. If two crests are observed at two different points in space, at the same time, it simply means that the two spatial points are at the same time-phase. At two points $x = x_1$ and $x = x_1 + \lambda$, separated by the wavelength, the harmonic wave exhibits the same time-behavior, which is expressed as

$$A\cos\left[k(x_1 + \lambda) - \omega t\right] = A\cos\left(kx_1 - \omega t\right) \qquad (7\text{-}12)$$

Note that this expression leads to $k\lambda = 2\pi$. By the same token, at two different times $t = t_1$ and $t = t_1 + \tau$, separated by the temporal period, the spatial field patterns of the wave should be the same such that

$$A\cos\left[kx - \omega(t_1 + \tau)\right] = A\cos\left(kx - \omega t_1\right) \qquad (7\text{-}13)$$

which leads to the relation $\omega\tau = 2\pi$. Rewriting these relations, the propagation constant and the angular frequency of a harmonic wave are expressed as

$$\boxed{k = \frac{2\pi}{\lambda}} \qquad \text{[rad/m]} \qquad (7\text{-}14)$$

$$\boxed{\omega = \frac{2\pi}{\tau} = 2\pi f} \qquad \text{[rad/s]} \qquad (7\text{-}15)$$

where λ is the wavelength, τ is the temporal period, and f is the frequency. Inserting Eqs. (7-14) and (7-15) into Eq. (7-11) we obtain a useful relation, that is,

$$\boxed{\upsilon_p = f\lambda} \qquad [\mathrm{m/s}] \qquad (7\text{-}16)$$

The phase velocity of a harmonic wave is equal to the product of the frequency and the wavelength.

To see the relation between the time-behavior and the spatial field pattern of a harmonic wave, let us consider Fig. 7.6, in which a harmonic wave is generated by a source located at $x = 0$, and propagates along the x-axis at time $0 \le t \le t_1$ with a velocity υ. The spatial field pattern at time $t = t_1$ is drawn in the figure, while the time-behaviors at points $x = x_1$ and $x = x_2$ are shown in the insets. We note that the curve for the wavefunction in time domain is reversed with respect to that in space domain. This is due to the negative sign in the argument of the wavefunction given in Eq. (7-9).

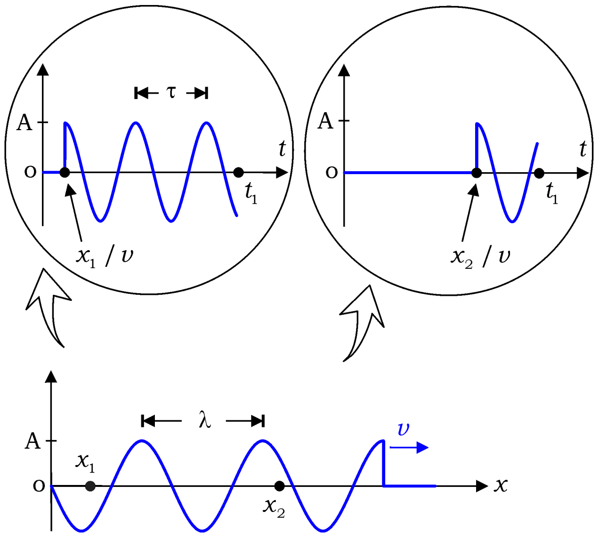


Fig. 7.6 Spatial field pattern of a harmonic wave. Time-behaviors at points $x = x_1$ and $x = x_2$ are shown in the insets.

Example 7-3

Given a harmonic wave $\psi = -4\cos 2\pi(0.2x - 3t)$, find (a) amplitude, (b) direction of propagation, (c) wavelength, (d) temporal period, (e) frequency, and (f) phase velocity.

Solution

(a) 4,

(b) $+x$ -direction,

(c) $k = \frac{2\pi}{\lambda} = 2\pi \times 0.2 \quad \rightarrow \quad \lambda = 5[\mathrm{m}]$,

(d) $\omega = \frac{2\pi}{\tau} = 2\pi \times 3 \quad \rightarrow \quad \tau = \frac{1}{3}[\mathrm{sec}]$,

(e) $\omega = 2\pi f = 2\pi \times 3 \quad \rightarrow \quad f = 3[1/\mathrm{sec}]$,

(f) $\upsilon_p = \frac{\omega}{k} = f\lambda = 15[\mathrm{m/s}]$.

Exercise 7.4

Express the wavefunction in cosine for a wave, propagating along the $+x$-direction with $\upsilon_p = 10[\mathrm{m/s}]$, whose time-behavior at $x = 0$ is shown in Fig. 7.7.

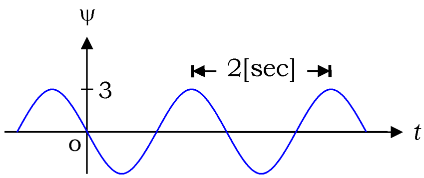


Fig. 7.7 Time-behavior of a harmonic wave.

Ans. $\psi = 3\cos(0.1\pi x - \pi t - \pi/2)$.

7.1.2 Complex Form of a Harmonic Wave

Harmonic waves are mathematically demanding because they are expressed in terms of trigonometric functions such as sine and cosine. For instance, summing two simple harmonic waves expressed by cosine functions may require tedious manipulations of the trigonometric rules. We can circumvent this difficulty by using complex exponentials in place of the cosine functions. This method has a couple of advantages in dealing with harmonic waves. First, the complex exponential provides an easy way of describing the superposition of harmonic waves. Second, it provides a simply way of handling the phase and impedance.

As we already know, a complex number $\hat{z}$ is comprised of a real part and an imaginary part such as

$$\hat{z} = x + iy \tag{7-17}$$

where $i = \sqrt{-1}$, and both x and y are real numbers. The real exponential function $\exp(x)$ can be extended to a complex exponential $\exp(\hat{z})$ by making use of the Euler formula given as

$$\boxed{e^{\pm i\theta} \equiv \cos\theta \pm i\sin\theta} \tag{7-18}$$

Algebraic manipulations of Eq. (7-18) give

$$\cos\theta = \mathrm{Re}\left[e^{i\theta}\right] = \frac{1}{2}\left[e^{i\theta} + e^{-i\theta}\right] \tag{7-19a}$$

$$\sin\theta = \mathrm{Im}\left[e^{i\theta}\right] = \frac{1}{2i}\left[e^{i\theta} - e^{-i\theta}\right] \tag{7-19b}$$

where $\mathrm{Re}[.]$ and $\mathrm{Im}[.]$ denote the real and imaginary parts, respectively.

It is customary to use the real part of a complex exponential for a harmonic wave such as

$$\psi(x,t) = A\cos\left(kx - \omega t + \varphi_o\right) = \mathrm{Re}\left[Ae^{i(kx-\omega t+\varphi_o)}\right] \tag{7-20}$$

which we call the real instantaneous form of the harmonic wave. From this point, with the understanding that the actual wave is its real part, we express the wavefunction of a harmonic wave simply as

$$\psi = Ae^{i(kx-\omega t+\varphi_o)} \tag{7-21}$$

This is called the complex form of the harmonic wave. In solving a problem under sinusoidal steady-state conditions, we use the complex form until the final result is reached, and then take the real part to express the solution in real instantaneous form.

Alternatively, the complex form of the harmonic wave is expressed as

$$\psi = \left[(Ae^{i\varphi_o})e^{ikx}\right]e^{-i\omega t} \equiv \left[\hat{A}e^{ikx}\right]e^{-i\omega t} \tag{7-22}$$

In the above equation $\hat{A}$ is called the complex amplitude, which is defined as

$$\hat{A} = Ae^{j\varphi_o} \tag{7-23}$$

where A is the amplitude and φ_o is the phase angle. Throughout the text, a complex number is denoted by ^ on the top (called “caret” or “hat”).

The frequency of a harmonic wave does not change, as the wave propagates through linear media. In view of the constant frequency, we can drop the harmonic time-dependence term $e^{-i\omega t}$ from Eq. (7-22), and write the wavefunction of the harmonic wave simply as

$$\boxed{\psi = \hat{A}e^{ikx}} \tag{7-24}$$

where $\hat{A}$ is the complex amplitude. If a harmonic wave is expressed as the above time-independent complex form, we can obtain the real instantaneous wavefunction by multiplying it by $e^{-i\omega t}$ and taking the real part.

Even if we use $-i$ instead of i in the complex form in Eq. (7-21), there will be no change in the real instantaneous form in Eq. (7-20). In solving practical electromagnetic problems, the use of $-i$ is preferred to that of i, even though the latter may offer a definite advantage in certain cases. To avoid confusion, it is customary to express $-i$ as $-j$, where $j = \sqrt{-1}$. Thus, the complex form of the harmonic wave is expressed alternatively as

$$\psi(x,t) = \hat{A}e^{j(\omega t - kx)} \tag{7-25}$$

where $j = \sqrt{-1}$. In the same way as was done for $e^{-i\omega t}$, the time-dependence term $e^{j\omega t}$ is dropped from Eq. (7-25), with the understanding that there is no change in the frequency of a harmonic wave propagating through linear media. Thus the harmonic wave is simply expressed as

$$\boxed{\psi = \hat{A}e^{-jkx}} \tag{7-26}$$

This is called the phasor form of the harmonic wave. The real instantaneous value is obtained by multiplying the phasor by $e^{j\omega t}$ and taking the real part, that is,

$$\psi = \mathrm{Re}\left[\hat{A}e^{-jkx}e^{j\omega t}\right] = A\cos\left(\omega t - kx + \varphi_o\right) \tag{7-27}$$

where the complex amplitude $\hat{A} = Ae^{j\varphi_o}$.

Two complex forms shown in Eqs. (7-24) and (7-26) have distinct advantages and disadvantages of their own. The complex form shown in Eq. (7-24) enables us to easily visualize the spatial field pattern of the harmonic wave at a particular time. For example, the spatial field pattern of a given wave at a later time Δt can be obtained simply by making the former field pattern displaced by a distance $\upsilon\Delta t$ along the direction of propagation of the wave. On the other hand, the phasor form shown in Eq. (7-26) is useful for specifying time delays of the harmonic wave at different points in space; a positive phase simply means a lead in time phase. It also allows us to handle the impedance in a straightforward manner. For instance, a positive phase angle of the intrinsic impedance simply means that the electric field leads the magnetic field in time phase. It is important to remember that only one complex form should be used throughout a given problem. The phasor form will be extensively used in the following chapters.

If the time-dependence of the harmonic wave is specified by the term $e^{-i\omega t}$, the exponent of the complex exponential in Eq. (7-24) represents the space-phase. In contrast, if the time-dependence is specified by the term $e^{j\omega t}$, the exponent of the complex exponential in Eq. (7-26) represents the time-phase.

Example 7-4

Consider two harmonic waves of the same wavefunction $\psi_o = \psi_1 = Ae^{i\pi x - i(3\pi\times10^8)t}$ given in free space, as shown in Fig. 7.8. Wave ψ_1 passes through a lossless medium ($2[\text{m}] \le x \le 3.5[\text{m}]$) with a propagation constant $k = 2\pi[\text{rad/m}]$. For the sake of simplicity, it is assumed that there is no reflection at the interface, and the wavefunction is always continuous across the interface. Find the real instantaneous values of (a) ψ_2, (b) ψ_3, and (c) phase difference $\Delta\varphi$.

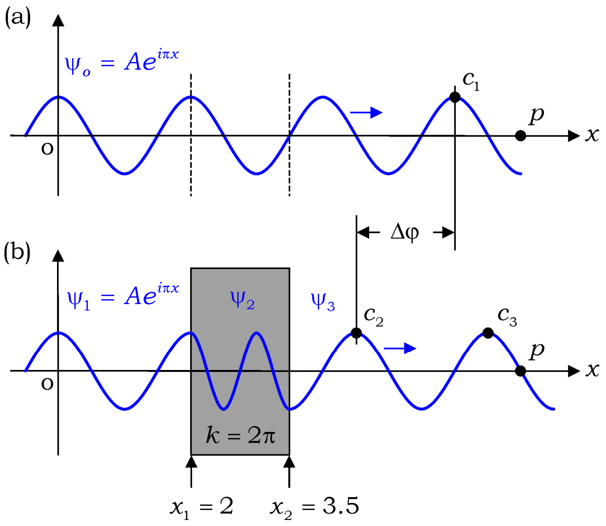


Fig. 7.8 Two harmonic waves.

Solution

(a) Just before crossing the interface at $x = 2[\text{m}]$, the time-independent complex form of ψ_1 is

$$\psi_1(x=2) = Ae^{i\pi 2} = A \tag{7-28}$$

In the material of $k = 2\pi$, the time-independent complex form of the wave is, in general,

$$\psi_2(x) = Ce^{ikx} = Ce^{i2\pi x} \qquad C\text{, constant} \tag{7-29}$$

Two disturbances at $x = 2[\text{m}]$, as given in Eqs. (7-28) and (7-29), should be equal according to the boundary condition. Thus,

$$A = Ce^{j4\pi}$$

Inserting the above equation into Eq. (7-29), we have

$$\psi_2(x) = Ae^{i2\pi x} \tag{7-30}$$

In the region $2[\text{m}] \le x \le 3.5[\text{m}]$, the real instantaneous value is

$$\psi_2 = \text{Re}\left[Ae^{i2\pi x}e^{-i\omega t}\right] = A\cos\left(2\pi x - 3\pi \times 10^8 t\right).$$

(b) Just before crossing the interface at $x = 3.5[\text{m}]$, from Eq. (7-30) we get

$$\psi_2(x = 3.5) = Ae^{i2\pi\times 3.5} = Ae^{i\pi} \tag{7-31}$$

The general wavefunction in the region $x \ge 3.5[\text{m}]$ is the same as that in the region $x < 2[\text{m}]$

$$\psi_3(x) = C'e^{i\pi x} \qquad C', \text{constant} \tag{7-32}$$

Two disturbances at $x = 3.5[\text{m}]$, as given in Eqs. (7-31) and (7-32), should be equal according to the boundary condition. Thus,

$$\psi_3(x = 3.5) = C'e^{i\pi 3.5} = Ae^{i\pi}$$

or

$$C' = Ae^{-i0.5\pi}$$

Inserting the above equation into Eq. (7-32), we have

$$\psi_3(x) = Ae^{-i0.5\pi}e^{i\pi x}$$

In the region $x \ge 3.5[\text{m}]$, the real instantaneous value is

$$\psi_3 = \text{Re}\left[Ae^{-i0.5\pi}e^{i\pi x}e^{-i\omega t}\right] = A\cos\left[\pi x - 3\pi \times 10^8 t - 0.5\pi\right]. \tag{7-33}$$

(c) At point p with $x = x_p$, the two waves are given as

$$\psi_o = A\cos(\pi x_p - 3\pi \times 10^8 t)$$
$$\psi_3 = A\cos(\pi x_p - 3\pi \times 10^8 t - 0.5\pi)$$

In view of the crests c_1 and c_3 shown in Fig. 7.8, ψ_3 appears to reach point p earlier than ψ_o, and thus ψ_3 leads ψ_o by $0.5\pi[\text{rad}]$ in time-phase. On the other hand, in view c_1 and c_2, ψ_3 lags behind ψ_o by $1.5\pi[\text{rad}]$ in time-phase.

Example 7-5

A harmonic wave, $\psi = Ae^{-j0.5\pi x + j\omega t}$ ($e^{j\omega t}$ is used instead of $e^{-i\omega t}$), propagates in the $+x$-direction in free space as shown in Fig. 7.9.

(a) Plot ψ versus ωt at $x = x_1$ and $x = x_2$.

(b) How much does $\psi(x_2, t)$ lag behind $\psi(x_1, t)$ in time-phase?

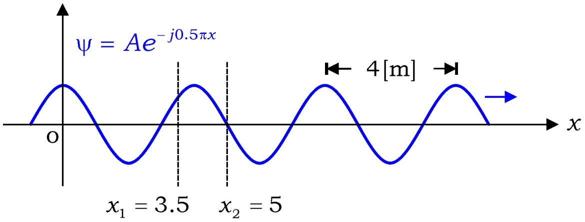


Fig. 7.9 A harmonic wave.

Solution

(a) Disturbances at $x = x_1$ and $x = x_2$ are

$$\psi(x_1) = Ae^{-j0.5\pi\times3.5} = Ae^{-j1.75\pi}$$
$$\psi(x_2) = Ae^{-j0.5\pi\times5} = Ae^{-j2.5\pi}$$

In real instantaneous form

$$\psi(x_1, t) = A\cos(\omega t - 1.75\pi)$$
$$\psi(x_2, t) = A\cos(\omega t - 2.5\pi)$$

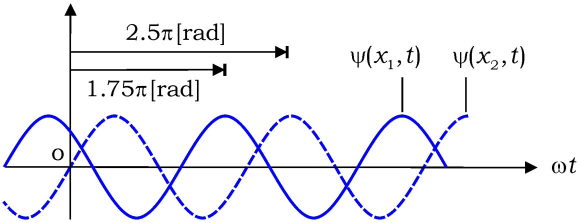


Fig. 7.10 Time-behaviors of the harmonic wave at $x = x_1$ and $x = x_2$.

(b) The phase difference is

$$\begin{aligned}\Delta\varphi = \varphi_2 - \varphi_1 &= (\omega t - 2.5\pi) - (\omega t - 1.75\pi)\\ &= -0.75\pi\end{aligned}$$

The minus sign signifies that $\psi(x_2, t)$ lags behind $\psi(x_1, t)$ by $0.75\pi\,[\text{rad}]$ in time-phase.

Exercise 7.5

Given a phasor $\psi = 2e^{j4.3}$, find the real instantaneous value oscillating at an angular frequency ω.

Ans. $\psi = 2\cos(\omega t + 4.3)$

Exercise 7.6

For the following pairs of phasors, determine the time-phase by which ψ_2 lags behind ψ_1. (a) $\psi_1 = 2e^{j4.3}$ and $\psi_2 = 3e^{j2.1}$, (b) $\psi_1 = 4e^{j\pi}$ and $\psi_2 = -4e^{j1.2\pi}$, (c) $\psi_1 = 6e^{j0.7\pi}$ and $\psi_2 = 5e^{j1.8\pi}$, and (d) $\psi_1 = 2e^{-j0.6}$ and $\psi_2 = 2e^{-j1.8}$.

Ans. (a) $2.2[\text{rad}]$, (b) $0.8\pi[\text{rad}]$, (c) $0.9\pi[\text{rad}]$, (d) $1.2[\text{rad}]$.

Review Questions with Hints

RQ 7.3	Define the phase of one-dimensional harmonic wave?	[Eq.(7-10)]
RQ 7.4	Define phase velocity?	[Eqs.(7-11)(7-16)]
RQ 7.5	Define propagation constant and specify its unit?	[Eq.(7-14)]
RQ 7.6	Write Euler formula?	[Eq.(7-18)]
RQ 7.7	Can the frequency of a wave be determined from the phasor?	[Eqs.(7-11)(7-26)]

7.2 Plane Waves in Three-Dimensional Space

An extension of the differential wave equation expressed by Eq. (7-6) to three-dimensions leads to the three-dimensional differential wave equation, that is,

$$\nabla^2\psi = \frac{1}{\upsilon^2}\frac{\partial^2\psi}{\partial t^2} \tag{7-34}$$

where ∇^2 is the Laplacian operator, and υ is the phase velocity of the wave in a given medium. The propagation of a wave in three-dimensional space is totally governed by Eq. (7-34). In this section, we will show that the simplest solution to Eq. (7-34) is a uniform plane wave, in which the spatial points of a constant phase form a plane, and the disturbance is constant and uniform on that plane. Since the energy of a three-dimensional wave gradually spreads out over the transverse plane as the wave propagates along a given direction, the profile of the wave generally varies in space. Nevertheless, the profile of the wave as a whole does not change with time. A uniform plane wave is the only wave that has the same profile at every point in space. From now on, we define a three-dimensional wave as a time-varying field satisfying the differential wave equation given in Eq. (7-34).

The three-dimensional differential wave equation has two independent general solutions of the form

$$\boxed{\psi = A_o \cos\left[(k_x x + k_y y + k_z z) \pm \omega t + \varphi_o\right]} \tag{7-35}$$

where φ_o is a constant phase. The general solution given in Eq. (7-35) can be readily verified by directly substituting it into Eq. (7-34). If the two general solutions are linearly combined in such a way as to satisfy given boundary conditions, such a particular solution must be a unique solution in the given region, according to the uniqueness theorem given from calculus. We rewrite Eq. (7-35) in complex form as

$$\boxed{\psi = \hat{A}e^{i\mathbf{k}\cdot\mathbf{r}\pm i\omega t}} \tag{7-36}$$

where the complex amplitude $\hat{A} = A_o \exp(i\varphi_o)$, and **k** is the wavevector defined in Cartesian coordinates as

$$\mathbf{k} = k_x\mathbf{a}_x + k_y\mathbf{a}_y + k_z\mathbf{a}_z \tag{7-37a}$$

$$|\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{2\pi}{\lambda} \tag{7-37b}$$

In the above equations, λ is the wavelength, and **r** is the position vector expressed in Cartesian coordinates as

$$\mathbf{r} = x\,\mathbf{a}_x + y\,\mathbf{a}_y + z\,\mathbf{a}_z \tag{7-38}$$

We note that the three-dimensional wave shown in Eq. (7-35) reduces to a one-dimensional wave as given in Eq. (7-9), if it is constrained to propagate along the x-axis only.

If a three-dimensional wave exists in a region of space, its phase forms a scalar field in the region, which is a smooth function of position at a given time. Accordingly, the spatial points of an equal phase define a smooth surface in the given region, which we call a phase front or wavefront. We now explore the wave expressed by Eq. (7-36) for the distribution of the phase in space. We see from Eq. (7-36) that the phase is solely determined by the term $\mathbf{k}\cdot\mathbf{r}$, which is the dot product of the wavevector **k**, which is a constant, and the position vector **r**, which varies from point to point in space. To specify the spatial points at which the phase is given by a constant a, we desire to find the position vectors that satisfy the following relation:

$$\mathbf{k}\cdot\mathbf{r} = a \tag{7-39}$$

Again, **r** is the position vector, and **k** is the wavevector. If $\mathbf{r}_o$ is one of the position vectors satisfying Eq. (7-39), such that $\mathbf{k}\cdot\mathbf{r}_o = a$, we can rewrite Eq. (7-39) as

$$\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}_o) = 0 \tag{7-40}$$

where **k** and $\mathbf{r}_o$ are constant, while **r** defines those points whose position vectors satisfy Eq. (7-40). We see from the dot product in Eq. (7-40), that the vector $\mathbf{r} - \mathbf{r}_o$ should be in a plane perpendicular to **k** as shown in Fig. 7.11. In other words, the position vector $\mathbf{r}$ in Eq. (7-40) defines a plane that is perpendicular to the wavevector $\mathbf{k}$, and passes through the point specified by $\mathbf{r}_o$. Consequently, the wave expressed by Eq. (7-35) or (7-36) has planar wavefronts, infinite in extent in three-dimensional space, on which the disturbance is constant and uniform; the wave is called a uniform plane wave.

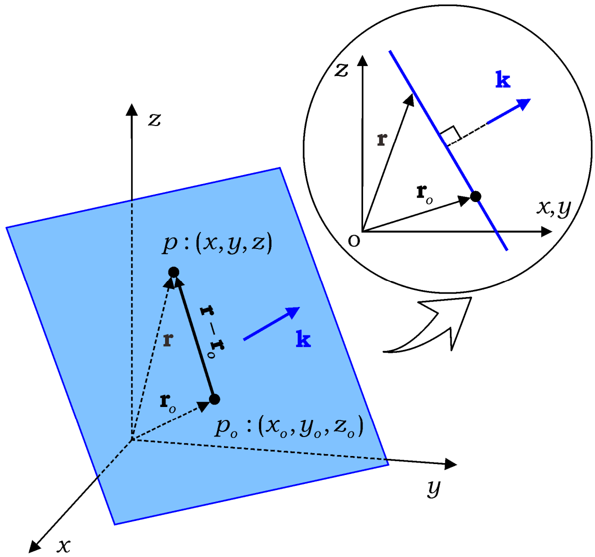


Fig. 7.11 The wavefront is perpendicular to the wavevector $\mathbf{k}$.

The wavefront of a uniform plane wave is periodic along the direction of its wavevector, repeating itself every λ, or the wavelength. Consider two points separated by a distance λ along the $\mathbf{k}$-direction. In vector notation, the two points are conveniently specified by position vectors $\mathbf{r}$ and $\mathbf{r}+\lambda\mathbf{a}_k$, respectively. The periodic nature of the wave assures that the disturbances at the two points are the same, that is,

$$\hat{A}e^{i\mathbf{k}\cdot\mathbf{r}} = \hat{A}e^{i\mathbf{k}\cdot(\mathbf{r}+\lambda\mathbf{a}_k)} \tag{7-41}$$

The equality in Eq. (7-41) leads to $\lambda\mathbf{k}\cdot\mathbf{a}_k = 2\pi$. Thus, we have

$$\boxed{k = \frac{2\pi}{\lambda}} \qquad [\text{rad/m}] \tag{7-42}$$

This is also called the wavenumber.

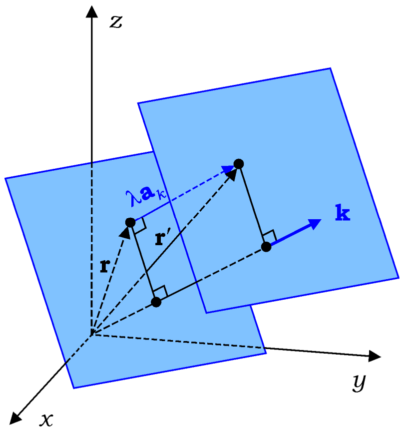


Fig. 7.12 Two planar wavefronts with the same phase.

Figure 7.13 can help us visualize the propagation of the uniform plane wave expressed by Eq. (7-36). The parallel planes represent the wavefronts of the wave, which are perpendicular to the direction of **k**, and move in unison in the direction of **k** as time increases. A cosine curve can be conveniently used to represent these planar wavefronts, which are infinite in extent and thus difficult to draw on a paper. Each point on the cosine curve corresponds to a planar wavefront that passes through the point and intersects the axis of the curve at right angles. The abscissa represents the phase that is constant on a given wavefront, and the ordinate represents the magnitude of the disturbance that is uniform on the wavefront.

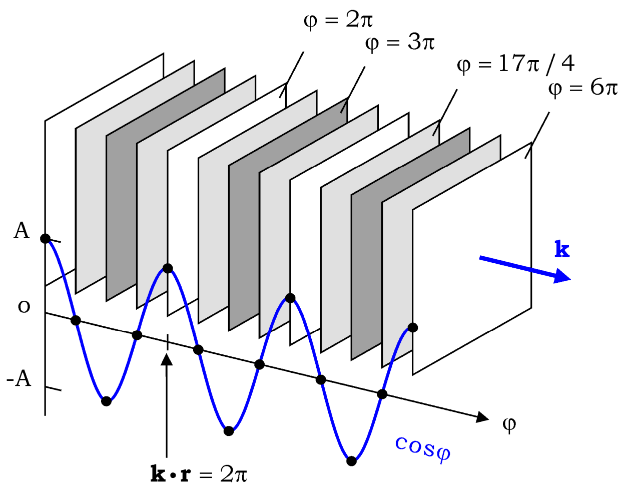


Fig. 7.13 Infinite planar wavefronts moving in unison along the **k**-direction.

Example 7-6

A uniform plane wave of an amplitude A propagates with a wavevector $\mathbf{k}$, which makes an angle θ with the y-axis in the yz-plane as shown in Fig. 7.14.

(a) Write the wavefunction in complex form.

(b) Find the spatial periods λ_y and λ_z, which are measured along the y- and z-axes, respectively.

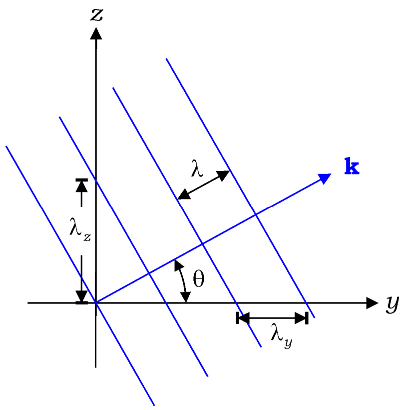


Fig. 7.14 Planar wavefronts of a uniform plane wave.

Solution

(a) Wavevector in component form

$$\mathbf{k} = \frac{2\pi}{\lambda}(\cos\theta\,\mathbf{a}_y + \sin\theta\,\mathbf{a}_z)$$

Wavefunction in complex form is therefore

$$\psi = Ae^{i\frac{2\pi}{\lambda}(y\cos\theta + z\sin\theta) - i\omega t}. \tag{7-43}$$

(b) From Eq. (7-43), the phases at $y = y_1$ and $y = y_1 + \lambda_y$ are

$$\varphi_1 = \frac{2\pi}{\lambda}y_1\cos\theta - \omega t \qquad \text{at } y = y_1 \tag{7-44a}$$

$$\varphi_2 = \frac{2\pi}{\lambda}\left(y_1 + \lambda_y\right)\cos\theta - \omega t \qquad \text{at } y = y_1 + \lambda_y \tag{7-44b}$$

Since the phase difference between the two points is 2π, or $\varphi_2 - \varphi_1 = 2\pi$, from Eq. (7-44), we obtain

$$\lambda_y = \frac{\lambda}{\cos\theta}$$

Following the same procedure as for λ_y, noting that the phase difference between two points at $z = z_1$ and $z = z_1 + \lambda_z$ is 2π, we obtain

$$\lambda_z = \frac{\lambda}{\sin\theta}.$$

Exercise 7.7
Referring to Fig. 7.11**,** find the distance between the wavefront with a phase a and the origin with a zero phase.

Ans. a / k.

Exercise 7.8
Referring to Fig. 7.14, determine the phase velocity of the wave measured along the x-axis, in terms of ω and λ_x.

Ans. $\omega\lambda_x / 2\pi$.

Review Questions with Hints

RQ 7.8 What information about the wave is contained in the wavevector? [Eq.(7-37)]

RQ 7.9 Explain the relation between a point on a cosine wavefunction and the wavefront of a given wave? [Fig.7.13]

RQ 7.10 Is the phase velocity of a uniform plane wave independent of the direction of measurement? [Fig.7.14]

7.3 Electromagnetic Plane Waves

In free space, there are no free charges and no free currents. Maxwell's equations in free space are therefore

$$\nabla \times \mathscr{E} = -\mu_o \frac{\partial \mathscr{H}}{\partial t} \tag{7-45a}$$

$$\nabla \times \mathscr{H} = \varepsilon_o \frac{\partial \mathscr{E}}{\partial t} \tag{7-45b}$$

$$\nabla \bullet \mathscr{E} = 0 \tag{7-45c}$$

$$\nabla \bullet \mathscr{H} = 0 \tag{7-45d}$$

where ε_o and μ_o are the permittivity and permeability of free space, respectively. The instantaneous electric and magnetic fields are vectors with three scalar components that are functions of space and time. The general expressions for $\mathscr{E}$ and $\mathscr{H}$ are

$$\mathscr{E}(\mathbf{r},t) = \mathscr{E}_x(\mathbf{r},t)\,\mathbf{a}_x + \mathscr{E}_y(\mathbf{r},t)\,\mathbf{a}_y + \mathscr{E}_z(\mathbf{r},t)\,\mathbf{a}_z \tag{7-46a}$$

$$\mathscr{H}(\mathbf{r},t) = \mathscr{H}_x(\mathbf{r},t)\,\mathbf{a}_x + \mathscr{H}_y(\mathbf{r},t)\,\mathbf{a}_y + \mathscr{H}_z(\mathbf{r},t)\,\mathbf{a}_z \tag{7-46b}$$

Taking the curl of both sides of Eq. (7-45a) and applying Eq. (7-45b), we have

$$\nabla\times\nabla\times\mathscr{E} = -\varepsilon_o\mu_o\frac{\partial^2\mathscr{E}}{\partial t^2}$$

Upon using the vector identity $\nabla\times\nabla\times\mathbf{U} = \nabla\nabla\bullet\mathbf{U} - \nabla^2\mathbf{U}$ and applying Eq. (7-45c) to the above equation, we obtain the three-dimensional differential wave equation, that is,

$$\boxed{\nabla^2\mathscr{E} = \varepsilon_o\mu_o\frac{\partial^2\mathscr{E}}{\partial t^2}} \tag{7-47}$$

Since the three vector components of $\mathscr{E}$ are mutually exclusive, Eq. (7-47) can be decomposed into three scalar differential wave equations. In Cartesian coordinates,

$$\nabla^2\mathscr{E}_x = \varepsilon_o\mu_o\frac{\partial^2\mathscr{E}_x}{\partial t^2} \tag{7-48a}$$

$$\nabla^2\mathscr{E}_y = \varepsilon_o\mu_o\frac{\partial^2\mathscr{E}_y}{\partial t^2} \tag{7-48b}$$

$$\nabla^2\mathscr{E}_z = \varepsilon_o\mu_o\frac{\partial^2\mathscr{E}_z}{\partial t^2} \tag{7-48c}$$

Three uniform plane wave are the solutions to these three differential wave equations, namely

$$\mathscr{E}_x(\mathbf{r},t) = E_1\cos\left(\mathbf{k}\bullet\mathbf{r} - \omega t + \varphi_x\right) \tag{7-49a}$$

$$\mathscr{E}_y(\mathbf{r},t) = E_2\cos\left(\mathbf{k}\bullet\mathbf{r} - \omega t + \varphi_y\right) \tag{7-49b}$$

$$\mathscr{E}_z(\mathbf{r},t) = E_3\cos\left(\mathbf{k}\bullet\mathbf{r} - \omega t + \varphi_z\right) \tag{7-49c}$$

Where **k** is wavevector, **r** is position vector, ω is angular frequency, and φ_x, φ_y, and φ_z are constant phases. From Eq. (7-48), we see that the magnitude of **k** is always given by $k = \omega\sqrt{\varepsilon_o\mu_o}$, regardless of the direction of **k**, in Eq. (7-49). For the reason that the three components comprise a single electromagnetic wave propagating in free space, the direction of **k** should be the same for the three components. Otherwise, the three would represent three independent waves in free space. It should be noted that the uniform plane waves are independent of each other, if they have different angular frequencies or different wavevectors that are different either in magnitude or in direction. Upon combining the three components given in Eq. (7-49), we obtain a uniform plane wave propagating in three-dimensional space, that is,

$$\boxed{\mathscr{E}(\mathbf{r},t) = \mathrm{Re}\left[\mathbf{E}_o e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\right]} \tag{7-50}$$

where $\mathbf{E}_o$ is the vector complex amplitude defined as

$$\mathbf{E}_o = E_1 e^{i\varphi_x}\,\mathbf{a}_x + E_2 e^{i\varphi_y}\,\mathbf{a}_y + E_3 e^{i\varphi_z}\,\mathbf{a}_z \tag{7-51a}$$

In the common case, in which the three component waves are in phase such that $\varphi_x = \varphi_y = \varphi_z = \varphi_o$, the vector complex amplitude is simply written as

$$\mathbf{E}_o = E_o e^{j\varphi_o}\,\mathbf{a}_E \tag{7-51b}$$

where E_o is the amplitude, φ_o is the phase angle, and $\mathbf{a}_E$ is a unit vector in the direction of the electric field intensity. Note that $\mathbf{E}_o$ is a vector whose scalar components are complex numbers, in general. In space, the electromagnetic wave expressed by Eq. (7-50) behaves like the scalar wave as discussed in section 7-2, except for the vector complex amplitude $\mathbf{E}_o$. In other words, Eq. (7-50) describes the spatiotemporal variation of a vector quantity, instead of a scalar quantity as in (7-36), in three-dimensional space. The electromagnetic plane wave expressed by Eq. (7-50) propagates in the direction of $\mathbf{k}$ with a phase velocity,

$$\upsilon_p = \frac{\omega}{k} = \frac{1}{\sqrt{\varepsilon_o\mu_o}} \cong 3\times 10^8 \qquad [\mathrm{m/s}] \tag{7-52}$$

which is true in free space. Since the phase velocity is equal to the speed of light in free space, *c*, we express the wavenumber as

$$\boxed{k = \frac{\omega}{c}} \tag{7-53}$$

This is true in free space only. In a material medium, c in the above equation is replaced with a phase velocity υ_p that is a distinct characteristic of the medium.

Following the same procedure as was used for $\mathscr{E}$, we can derive the differential wave equation for the magnetic field as

$$\boxed{\nabla^2\mathscr{H} = \varepsilon_o\mu_o\frac{\partial^2\mathscr{H}}{\partial t^2}} \tag{7-54}$$

The simplest solution to Eq. (7-54) is a uniform plane wave expressed as

$$\boxed{\mathscr{H}(\mathbf{r},t) = \mathrm{Re}\left[\mathbf{H}_o e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\right]} \tag{7-55}$$

where $\mathbf{H}_o$ is the vector complex amplitude. Since $\mathscr{E}$ and $\mathscr{H}$ are interrelated, we only need to solve either one of Eqs. (7-47) and (7-54), and then obtain the other field from one of Maxwell's equations, either Eq. (7-45a) or Eq. (7-45b), for an electromagnetic wave propagating in free space.

Exercise 7.9
Show by direct substitution that Eq. (7-50) is a solution of Eq. (7-47).

Exercise 7.10
Show that Eq. (7-47) has another solution $\mathscr{E}(\mathbf{r},t) = \mathrm{Re}\left[\mathbf{E}_o e^{i\mathbf{k}\cdot\mathbf{r}+i\omega t}\right]$.

Exercise 7.11
Find the ratio $|\mathbf{E}_o / \mathbf{B}_o|$ in free space, by inserting Eqs. (7-50) and (7-55) into Eq. (7-45).
Ans. c.

7.3.1 Transverse Electromagnetic Waves

Maxwell's equations tell us that a time-varying electric field $\mathscr{E}$ induces a time-varying magnetic field $\mathscr{H}$ in the direction perpendicular to $\mathscr{E}$. The time-varying $\mathscr{H}$ in turn generates a time-varying $\mathscr{E}$ in the direction perpendicular to $\mathscr{H}$. The time-varying electric and magnetic fields regenerate each other in the cycle, and form an electromagnetic wave propagating at the speed of light in space. The interrelationship, perpendicularity, and symmetry between $\mathscr{E}$ and $\mathscr{H}$ lead us to a conclusion that the direction of propagation of an electromagnetic wave is perpendicular to both $\mathscr{E}$ and $\mathscr{H}$.

If a uniform plane wave propagates along the z-direction, it means that the planar wavefronts are perpendicular to the z-axis. Since the electric field intensity is constant and uniform on the wavefront, $\mathscr{E}$ is independent of the x- and y-coordinates. That is,

$$\frac{\partial \mathscr{E}(\mathbf{r},t)}{\partial x} = \frac{\partial \mathscr{E}(\mathbf{r},t)}{\partial y} = 0 \tag{7-56}$$

In view of Eq. (7-56), the general expression for $\mathscr{E}$ is reduced to

$$\mathscr{E}(\mathbf{r},t) = \mathscr{E}_x(z,t)\,\mathbf{a}_x + \mathscr{E}_y(z,t)\,\mathbf{a}_y + \mathscr{E}_z(z,t)\,\mathbf{a}_z \tag{7-57}$$

Inserting Eq. (7-57) into Eq. (7-45c), we have

$$\nabla \bullet \mathscr{E} = \frac{\partial \mathscr{E}_z}{\partial z} = 0 \tag{7-58}$$

This relation will be satisfied if $\mathscr{E}_z$ is constant along the z-axis, which is the direction of propagation of the wave. In the context of electromagnetics, in which the electric and magnetic fields vary in space and time, a spatially uniform electric field is not supported. Thus, we should have

$$\mathscr{E}_z = 0 \tag{7-59}$$

Upon inserting Eq. (7-59) into Eq. (7-57), the electric field intensity of a uniform plane wave propagating in the z-direction is expressed as

$$\mathscr{E} = \mathscr{E}_x(z,t)\,\mathbf{a}_x + \mathscr{E}_y(z,t)\,\mathbf{a}_y \tag{7-60}$$

The direction of $\mathscr{E}$ must be perpendicular to the direction of propagation of the wave.

Next, inserting Eq. (7-60) into Eq. (7-45a), we have

$$-\frac{\partial \mathscr{E}_y}{\partial z}\mathbf{a}_x + \frac{\partial \mathscr{E}_x}{\partial z}\mathbf{a}_y = -\mu_o \frac{\partial \mathscr{H}_x}{\partial t}\mathbf{a}_x - \mu_o \frac{\partial \mathscr{H}_y}{\partial t}\mathbf{a}_y - \mu_o \frac{\partial \mathscr{H}_z}{\partial t}\mathbf{a}_z \tag{7-61}$$

The equality in Eq. (7-61) is satisfied mathematically if $\mathscr{H}_z$ is constant in time. In the context of electromagnetics, we set

$$\mathscr{H}_z = 0$$

The direction of $\mathscr{H}$ is also perpendicular to the direction of propagation of the wave. In view of these discussions, ***the electromagnetic wave is a transverse wave.***

The time-varying electric and magnetic fields, $\mathscr{E}$ and $\mathscr{H}$, are mutually orthogonal, and both perpendicular to the wavevector $\mathbf{k}$. The directions of the three vectors obey the right-hand rule: the right thumb points in the direction of $\mathbf{k}$ when the fingers rotate from $\mathscr{E}$ to $\mathscr{H}$, that is,

$$\boxed{\mathbf{a}_E \times \mathbf{a}_H = \mathbf{a}_k} \tag{7-62}$$

where $\mathbf{a}_E$, $\mathbf{a}_H$, and $\mathbf{a}_k$ are the unit vectors of $\mathscr{E}$, $\mathscr{H}$, and $\mathbf{k}$, respectively. Eq. (7-62) is true at all times at every point in space.

Example 7-7

The electric field $\mathscr{E} = E_o\,\mathbf{a}_x \cos(kz - \omega t)$ represents a uniform plane wave propagating in free space. Find the magnetic field.

Solution

Inserting $\mathscr{E}$ into Eq. (7-45a),

$$\mathbf{a}_y E_o \frac{\partial}{\partial z}\cos(kz-\omega t) = -\mu_o \frac{\partial}{\partial t}\left(\mathscr{H}_x\,\mathbf{a}_x + \mathscr{H}_y\,\mathbf{a}_y + \mathscr{H}_z\,\mathbf{a}_z\right)$$

From the above equation, we obtain

$\mathscr{H}_x = \mathscr{H}_z = 0$, and thus

$$kE_o \sin(kz-\omega t) = \mu_o \frac{\partial \mathscr{H}_y}{\partial t} \tag{7-63}$$

Integrating both sides of Eq. (7-63) with respect to t, we obtain

$$\mathscr{H}_y = \sqrt{\frac{\varepsilon_o}{\mu_o}} E_o \cos(kz-\omega t)$$

The magnetic field of the wave is therefore

$$\mathscr{H} = \sqrt{\frac{\varepsilon_o}{\mu_o}} E_o\,\mathbf{a}_y \cos(kz-\omega t) \tag{7-64}$$

The expressions for $\mathscr{E}$ and $\mathscr{H}$ are exactly the same except for the amplitudes, which differ by a factor of $\sqrt{\varepsilon_o/\mu_o}$. In this case, the electric and magnetic fields are said to be in phase. The electric and magnetic field vectors are perpendicular to each other in such a way that $\mathscr{E}\times\mathscr{H}$ points in the direction of propagation of the wave, which is the direction of the wavevector $\mathbf{k}$.

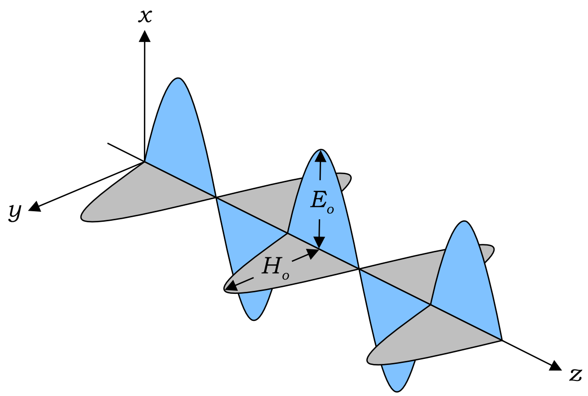


Fig. 7.15 An electromagnetic plane wave propagating in the z-direction.

Exercise 7.12

A uniform plane wave has a wavevector directed in the direction of $\mathbf{a}_x + 2\,\mathbf{a}_y$ and $\mathscr{E}$ directed in the direction of $\mathbf{a}_z$. Find the direction of $\mathscr{H}$.

Ans. It is along the direction of $2\,\mathbf{a}_x - \mathbf{a}_y$.

Review Questions with Hints

RQ 7.11 Write Maxwell's equations in free space. [Eq.(7-45)]

RQ 7.12 Write the differential wave equation for $\mathscr{E}$ and express its plane wave solution. [Eqs.(7-47)(7-50)]

RQ 7.13 Define the wavenumber. [Eqs.(7-42)(7-53)]

RQ 7.14 What is the relation among the unit vectors of $\mathscr{E}$, $\mathscr{H}$, and **k**? [Eq.(7-62)]

RQ 7.15 What is the ratio between the scalar amplitudes of $\mathscr{E}$ and $\mathscr{H}$ of an electromagnetic plane wave? [Eq.(7-64)]

Problems

7-1 Is ψ a traveling wave in free space ?

(a) $\psi = (2x - 3t)^2$

(b) $\psi = \sin(2x^2 - 3t^2)$

(c) $\psi = \cos(x^2 + t^2 + 2xt)$

(d) $\psi = e^{-3x} e^{j2t - jx}$

7-2 Is ψ a traveling wave in free space, whether or not it is an electromagnetic wave?

(a) $\psi = (5\,\mathbf{a}_x + 4\,\mathbf{a}_z)\cos(2z - 3t)$

(b) $\psi = 5\,\mathbf{a}_x \cos(2z - 3t) + 4\,\mathbf{a}_y \sin(2z - 3t)$

7-3 Given a traveling wave $\psi = e^{2+j1.2} e^{j100\pi t - j0.02\pi z}$, find

(a) amplitude,
(b) wavelength,
(c) frequency,
(d) phase velocity,
(e) direction of propagation.

7-4 At time $t = 0.5\,[\mathrm{s}]$, the profile is given as Fig. 7.16 for the harmonic wave traveling in the +x-direction, with a phase velocity $\upsilon_p = 10\,[\mathrm{m/s}]$. Express the wavefunction as a cosine.

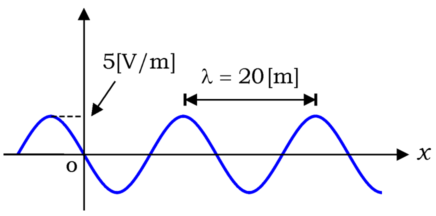


Fig. 7.16 Profile of a harmonic wave(Problem 7-14).

7-5 A harmonic wave is given as $\psi = A_o e^{ikx - i\omega t}$ in free space $(x < 0)$, where k is the propagation constant. The wave passes through a lossless dielectric, occupying the region $0 \le x \le d$, with a propagation constant $k' = nk$ as shown in Fig. 7.17. In the region $x > d$, the wave lags behind the wave, which has propagated as if there were no dielectric block, by 90^o. For the sake of simplicity, we assume that there is no reflection at the interface and the wavefunction is continuous across the interface. Find
(a) ψ_1 in the dielectric,
(b) ψ_2 after the dielectric,
(c) n in terms of d and λ.

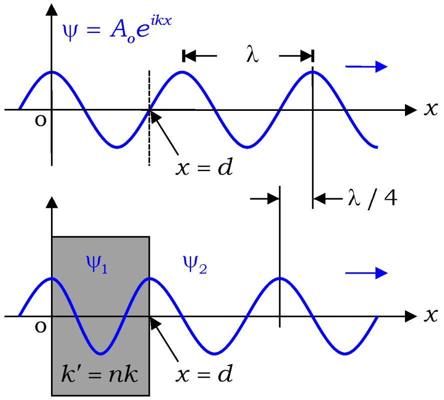


Fig. 7.17 A phase delay due to a dielectric(Problem 7-5).

7-6 The wavefront of a uniform plane wave($\lambda = 0.25\pi$[m]) intersects Cartesian coordinate axes at $x = 1/2$, $y = 1/\sqrt{3}$, and $z = 1/3$, respectively, at time $t = t_o$. The wave is known to propagate away from the origin. If the phase at the origin is measured to be zero at time $t = t_o$. find
(a) wavevector,
(b) phase on this wavefront at $t = t_o$.

7-7 For the uniform plane wave with a wavevector $\mathbf{k} = 5\,\mathbf{a}_x + 3\,\mathbf{a}_y + \sqrt{2}\,\mathbf{a}_z$ and a frequency 60[Hz], find
(a) phase velocity,
(b) phase velocity measured along the x-axis.

7-8 Given the uniform plane wave of a wavevector $\mathbf{k} = 10(\mathbf{a}_x + 2\,\mathbf{a}_y + 2\,\mathbf{a}_z)$, find
(a) wavelength,
(b) spatial period measured along the x-axis.

7-9 A uniform plane wave, propagating with a phase velocity $v_p = 50[\mathrm{m/s}]$, is expressed as $\psi = 5e^{i40\pi(x+2y+2z)-i\omega t}$. At time $t = 0$, the wavefront intersects the x-axis at $x = 0.5[\mathrm{m}]$. Find the x-intercept of this wavefront at time $t = 0.02[\mathrm{sec}]$.

7-10 A planar wavefront intersects Cartesian coordinate axes at $x = 3$, $y = 2\sqrt{3}$, and $z = 2$, respectively, at time $t = 0$. The wavefront passes the origin at time $t = 0.2[\mathrm{sec}]$. Find

(a) unit wavevector,

(b) phase velocity.

7-11 Is ψ a transverse electromagnetic wave propagating in free space?

(a) $\psi = (\mathbf{a}_x - i2\mathbf{a}_y)e^{i10z}$

(b) $\psi = (1 + i0.5)\mathbf{a}_x e^{-i12.5z+i0.3}$

(c) $\psi = 3e^{i1.4}\,\mathbf{a}_x e^{i(2y-5z)}$

(d) $\psi = 2\mathbf{a}_y e^{i(3y+8z)}$

(e) $\psi = (15\mathbf{a}_x + 12\mathbf{a}_y)e^{i(4x-5y+3z)}$

(f) $\psi = (15\mathbf{a}_x + i12\mathbf{a}_y)e^{i(4x-5y+3z)}$

(g) $\psi = 12\mathbf{a}_x e^{i(-20x+15y+34z)}$

7-12 An electromagnetic plane wave has the electric field expressed as $\mathbf{E} = (3\,\mathbf{a}_x + 4\,\mathbf{a}_y)e^{i10^6 z}$, in free space, and the harmonic time dependence of the form $e^{-i\omega t}$. Find the real instantaneous expressions for

(a) **E** and (b) **H**.

7-13 In free space, the electric field of an electromagnetic plane wave is expressed as $\mathbf{E} = \left[(1 + i2)\,\mathbf{a}_x + \mathbf{a}_y\right]e^{i10^4 z-i0.8}$ with the harmonic time dependence of the form $e^{-i\omega t}$. Find the real instantaneous expressions for

(a) **E** and (b) **H**.

7-14 The electric field, $\mathscr{E} = \mathbf{E}_o \cos(3x + 4y - \omega t)[\mathrm{V/m}]$, represents an electromagnetic plane wave propagating in free space. The amplitudes of the x- and z-components of $\mathscr{E}$ are given by $E_{ox} = 12[\mathrm{V/m}]$ and $E_{oz} = 10[\mathrm{V/m}]$, respectively. Find (a) wavevector **k**, (b) angular frequency ω, and (c) an expression for $\mathbf{E}_o$.

7-15 At the origin in free space, $\mathscr{E}$ and $\mathscr{H}$ oscillate with an angular frequency, $\omega = 9\times10^9[\mathrm{rad/s}]$. The amplitude and direction of $\mathscr{E}$ are given by $20\,\mathbf{a}_z[\mathrm{V/m}]$, while those of $\mathscr{H}$ are given by $(1/90\pi)(9\,\mathbf{a}_x - 12\,\mathbf{a}_y)[\mathrm{A/m}]$, at the origin. Find the real instantaneous values of $\mathscr{E}$ and $\mathscr{H}$ everywhere.

Chapter 8
Time-Harmonic Electromagnetic Waves

We saw in Chapter 7 that Maxwell's equations can be combined into the differential wave equations for the electric field $\mathscr{E}$ and the magnetic field $\mathscr{H}$. Since $\mathscr{E}$ and $\mathscr{H}$ are interrelated under time-varying conditions, the differential wave equation for $\mathscr{E}$ alone is enough to describe the general behavior of an electromagnetic wave propagating in a region of space. The differential wave equation is a homogeneous, second-order, partial differential equation. In one-dimensional space, a general solution is given by a linear combination of two waves propagating in the opposite directions. In three-dimensional space, on the other hand, the general solution is given by a linear combination of uniform plane waves with wavevectors of the same magnitude but different unit vectors. Applying two boundary conditions to the general solution, we obtain a particular solution, or a unique solution, in the given region.

When an electromagnetic wave propagates through linear media, there is no change in the frequency of the wave, although the wavevector and wave velocity may be altered in different media. If this is the case, it will be more convenient to separate the space coordinates x, y, and z from the time variable t in the wavefunction of a time-harmonic electromagnetic wave, and represent the wave by the phasor, or the part independent of t. From now on, we will use the term $e^{j\omega t}$, where $j = \sqrt{-1}$, for describing the temporal behavior of time-harmonic quantities. Accordingly, a uniform plane wave with a unit amplitude and a wavevector $\mathbf{k}$ is simply expressed as $e^{-j\mathbf{k}\cdot\mathbf{r}}$.

In the present chapter, we focus our attention on time-harmonic electromagnetic waves propagating in different media, including free space, good conductors, lossless media, and lossy dielectrics. We also discuss different polarizations of an electromagnetic wave, and the power delivered by the electromagnetic wave. We will study the reflection and refraction of an electromagnetic wave that is incident on an interface between different media.

8.1 Phasors

A perfect dielectric has no free charges and no free currents so that $\rho_v = 0$ and $\boldsymbol{J} = 0$ in the inside. If the material is homogeneous, linear, and isotropic, the permittivity ε and permeability μ are constant, independent of the magnitudes and directions of the electric and magnetic fields. Under these conditions, Maxwell's equations are written, in the dielectric, as follows:

$$\nabla \times \mathscr{E} = -\mu \frac{\partial \mathscr{H}}{\partial t} \tag{8-1a}$$

$$\nabla \times \mathscr{H} = \varepsilon \frac{\partial \mathscr{E}}{\partial t} \tag{8-1b}$$

$$\nabla \bullet \mathscr{E} = 0 \tag{8-1c}$$

$$\nabla \bullet \mathscr{H} = 0 \tag{8-1d}$$

where $\mathscr{E}$ and $\mathscr{H}$ are the instantaneous electric and magnetic fields. As shown previously in Chapter 7, Maxwell's equations can be combined into a differential wave equation such as

$$\nabla^2 \mathscr{E} = \mu\varepsilon \frac{\partial^2 \mathscr{E}}{\partial t^2} \tag{8-2}$$

In a medium of infinite extent, which is said to be unbounded, a general solution of Eq. (8-2) is given by a uniform plane wave simply expressed as

$$\mathscr{E}(\mathbf{r}, t) = \mathrm{Re}\left[\mathbf{E}_o e^{-j\mathbf{k}\bullet\mathbf{r}+j\omega t}\right] \tag{8-3}$$

where $j = \sqrt{-1}$. As was discussed in Chapter 7, the complex amplitude $\mathbf{E}_o$ in Eq. (8-3), which is yet to be determined, is a vector with three, complex, scalar components. From the phase term $-\mathbf{k}\bullet\mathbf{r}$ in Eq. (8-3), we see that the wavefront is a plane which is infinite in extent in three-dimensional space and perpendicular to the wavevector $\mathbf{k}$. If $\mathbf{E}_o$ is constant in space, the electric field intensity is uniform on the wavefront, and the wave is called a uniform plane wave. The time-dependence term $e^{j\omega t}$ in Eq. (8-3) shows that $\mathscr{E}$ varies sinusoidally with time, and the wave is called a time-harmonic wave. We saw in Chapter 7 that the differential wave equation as in Eq. (8-2) specifies the wavevector to be of the form

$$\mathbf{k} = \omega\sqrt{\mu\varepsilon}\,\mathbf{a}_k \tag{8-4}$$

where ω is the angular frequency and $\mathbf{a}_k$ is the unit vector. It should be noted that although Eq. (8-2) provides the information about the magnitude of $\mathbf{k}$ through the material parameters ε and μ, it imposes no limit on the direction of $\mathbf{k}$. As we will see later, the direction of $\mathbf{k}$ is determined by the position and orientation of the source, if the medium is unbounded, and, additionally, by the boundary conditions, if the medium is finite in extent.

As an example, consider the case in which a uniform plane wave impinges on an interface between two adjoining media of different kinds. In this case, a part of the incident wave is reflected back into the first medium, and thus the total electric field in the medium is equal to the sum of the electric fields of the incident and the reflected waves. Expressed mathematically, in the first medium, the total electric field intensity is

$$\mathscr{E}(\mathbf{r},t) = \text{Re}\left[\left(\mathbf{E}_o^i e^{-j\mathbf{k}_i\cdot\mathbf{r}} + \mathbf{E}_o^r e^{-j\mathbf{k}_r\cdot\mathbf{r}}\right)e^{j\omega t}\right] \tag{8-5}$$

where i and r denote the incident and reflected waves, respectively. It can be readily shown that the general solution given in Eq. (8-5) satisfies the differential wave equation only if $|\mathbf{k}_i| = |\mathbf{k}_r| = \omega\sqrt{\mu\varepsilon}$. While the direction of $\mathbf{k}_i$ (the wavevector of the incident wave) is usually given in the problem, the direction of $\mathbf{k}_r$ (the wavevector of the reflected wave) is yet to be determined from the boundary conditions at the interface. Rewriting Eq. (8-5) in a more general form, we have

$$\boxed{\mathscr{E}(\mathbf{r},t) = \text{Re}\left[\mathbf{E}(x,y,z)e^{j\omega t}\right]} \tag{8-6}$$

In Eq. (8-6), the space variables are completely separated from the time variable. The vector function $\mathbf{E}(x,y,z)$ in Eq. (8-6) is called the phasor, which is generally a complex number, varying from point to point in space, independent of time. The phasor should not be confused with the complex amplitudes $\mathbf{E}_o^i$ and $\mathbf{E}_o^r$, which are usually constant in space. In the sinusoidal steady state, a phasor generally represents a linear combination of uniform plane waves with the wavevectors of the same magnitude but different unit vectors. The real instantaneous value of $\mathbf{E}$ can be obtained, at any time, by multiplying the phasor by the harmonic time-dependence term $e^{j\omega t}$ and taking the real part.

Following the same procedure as for $\mathscr{E}$, Maxwell's equations can be combined into the differential wave equation for the magnetic field $\mathscr{H}$, which has exactly the same form as Eq. (8-2). Therefore, the general solution for $\mathscr{H}$ should have the same form as Eq. (8-6), that is,

$$\boxed{\mathscr{H}(\mathbf{r},t) = \text{Re}\left[\mathbf{H}(x,y,z)e^{j\omega t}\right]} \tag{8-7}$$

where $\mathbf{H}(x,y,z)$ is the phasor of the magnetic field intensity.

Example 8-1

Resolve the phasor $\mathbf{E} = \mathbf{a}_y\, E_o \cos(k_1 x)e^{-jk_3 z}$ into two uniform plane waves and sketch the wavefronts.

Solution

Using Euler formula we rewrite the phasor as

$$\mathbf{E} = \mathbf{a}_y E_o \frac{1}{2}\left[e^{jk_1x} + e^{-jk_1x}\right]e^{-jk_3z} = \mathbf{a}_y \frac{E_o}{2} e^{-j(-k_1x+k_3z)} + \mathbf{a}_y \frac{E_o}{2} e^{-j(k_1x+k_3z)}$$
$$\equiv \mathbf{E}' + \mathbf{E}''$$

In the above equation, $\mathbf{E}'$ and $\mathbf{E}''$ represent two uniform plane waves of wavevectors $\mathbf{k}' = -k_1\,\mathbf{a}_x + k_3\,\mathbf{a}_z$ and $\mathbf{k}'' = k_1\,\mathbf{a}_x + k_3\,\mathbf{a}_z$, respectively. Since the two waves propagate in the same medium, we should have $|\mathbf{k}'| = |\mathbf{k}''| = \sqrt{k_1^2 + k_3^2}$. As shown in Fig. 8.1, wavefronts are infinite, parallel planes perpendicular to the wavevector. From Fig. 8.1 we can determine the wavelength of the wave as $\lambda = 2\pi / \sqrt{k_1^2 + k_3^2}$, and the tilt angle of the wavefront as $\theta = \tan^{-1}(k_1 / k_3)$.

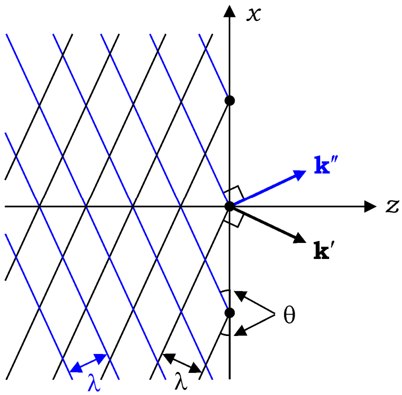


Fig. 8.1 Superposition of two uniform plane waves.

Exercise 8.1
Find the real instantaneous value of the phasor $\mathbf{A} = A_o(\mathbf{a}_x + j\mathbf{a}_y)$.

Ans. $\mathscr{A} = A_o\mathbf{a}_x \cos(\omega t) - A_o\mathbf{a}_y \sin(\omega t)$.

Exercise 8.2
Is it possible to express $\mathscr{B} = B_o\mathbf{a}_x \cos(10t)\cos(15t - 30z)$ in phasor form? If not, why?

Ans. No. It has two frequencies, $\omega = 15$ and $\omega = 25$.

8.1.1 Maxwell's Equations in Phasor Form

In the sinusoidal steady state, Maxwell's equations can be expressed in terms of phasors. First, upon substituting the phasor form of $\mathscr{E}$ and $\mathscr{H}$, as given in Eq. (8-6) and Eq. (8-7), into Eq. (8-1a), we have

$$\nabla \times \mathrm{Re}\left[\mathbf{E}(x,y,z)e^{j\omega t}\right] = -\mu \frac{\partial}{\partial t}\mathrm{Re}\left[\mathbf{H}(x,y,z)e^{j\omega t}\right] \tag{8-8}$$

Since the three operations, such as curl, time derivative, and taking the real part, are mutually exclusive in Eq. (8-8), we can write the equation as

$$\mathrm{Re}\left[e^{j\omega t}\,\nabla\times\mathbf{E}(x,y,z)\right]=-\mu\,\mathrm{Re}\left[\mathbf{H}(x,y,z)j\omega e^{j\omega t}\right]$$

In dropping the operator Re[..] and the time-harmonic term $e^{j\omega t}$ from both sides of the above equation, with the understating that the real instantaneous value is the real part of the product of the phasor and $e^{j\omega t}$, we obtain the phasor form of Faraday's law:

$$\nabla\times\mathbf{E}=-j\omega\mu\mathbf{H}$$

Following the same procedure as for Faraday's law, we can obtain the phasor form of the other equations. In a homogeneous, linear, and isotropic perfect dielectric, the phasor form of Maxwell's equations is written as follows:

$$\boxed{\nabla\times\mathbf{E}=-j\omega\mu\mathbf{H}} \qquad (8\text{-}9a)$$

$$\boxed{\nabla\times\mathbf{H}=j\omega\varepsilon\mathbf{E}} \qquad (8\text{-}9b)$$

$$\boxed{\nabla\bullet\mathbf{E}=0} \qquad (8\text{-}9c)$$

$$\boxed{\nabla\bullet\mathbf{H}=0} \qquad (8\text{-}9d)$$

Again, the electric and magnetic field phasors, **E** and **H**, depend on space coordinates only, independent of time. They are vectors with three scalar components, which are complex numbers in general.

Combination of Eq. (8-9a) and Eq. (8-9b) leads to a time-harmonic wave equation expressed in terms of phasors. Upon taking the curl of both sides of Eq. (8-9a), followed by a substitution of Eq. (8-9b), we have

$$\begin{aligned}\nabla\times\nabla\times\mathbf{E}&=-j\omega\mu\nabla\times\mathbf{H}\\&=\omega^2\mu\varepsilon\mathbf{E}\end{aligned} \qquad (8\text{-}10)$$

Applying the vector identity $\nabla\times\nabla\times\mathbf{E}=\nabla\nabla\bullet\mathbf{E}-\nabla^2\mathbf{E}$ to the left-hand side of Eq. (8-10), followed by a substitution of Eq. (8-9c), we obtain the vector Helmholtz's equation:

$$\boxed{\nabla^2\mathbf{E}+k^2\mathbf{E}=0} \qquad (8\text{-}11)$$

Here, the wavenumber k is defined as

$$\boxed{k=\omega\sqrt{\mu\varepsilon}=\frac{\omega}{\upsilon}=\frac{2\pi}{\lambda}} \qquad \text{[rad/m]} \qquad (8\text{-}12)$$

where ω is angular frequency, μ is permeability, ε is permittivity, and λ is wavelength. In a material, the speed of wave is expressed as $\upsilon=c/\sqrt{\mu_r\varepsilon_r}$, where

μ_r and ε_r are the relative permeability and relative permittivity, respectively, and $c = 1/\sqrt{\mu_o \varepsilon_o}$ is the speed of light in free space.

Exercise 8.3

For an electric field phasor $\mathbf{E} = A\,\mathbf{a}_x e^{jkz + j\varphi}$ in free space, where A and φ are constants, find an expression for k.

Ans. $k = \omega\sqrt{\mu_o \varepsilon_o}$.

Review Questions with Hints

RQ 8.1 Write the harmonic electric field intensity, in general, in phasor form. [Eq.(8-6)]

RQ 8.2 Write Maxwell's equations in phasor form. [Eq.(8-9)]

RQ 8.3 Write vector Helmholtz's equation. [Eq.(8-11)]

8.2 Waves in Homogenous Media

A homogeneous medium consists of one kind of material, in which the permittivity and permeability are constant. There is no boundary or interface in the medium, and thus a single uniform plane wave can be supported uninterrupted by a reflection or refraction.

8.2.1 Uniform Plane Wave in a Lossless Dielectric

A lossless dielectric is a perfect dielectric, where an electromagnetic wave can propagate without loss of energy. The propagation behavior of an electromagnetic wave in a lossless dielectric depends on the constitutive material parameters ε and μ as well as the frequency of operation. These three factors comprise the wavenumber of the wave; that is, $k = \omega\sqrt{\varepsilon\mu}$.

Let us begin with the vector Helmholtz's equation, expressed in Cartesian coordinates as

$$\nabla^2\left(E_x\,\mathbf{a}_x + E_y\,\mathbf{a}_y + E_z\,\mathbf{a}_z\right) + k^2\left(E_x\,\mathbf{a}_x + E_y\,\mathbf{a}_y + E_z\,\mathbf{a}_z\right) = 0 \qquad (8\text{-}13)$$

where E_x, E_y, and E_z are three components of the electric field phasor $\mathbf{E}$. Since the unit vectors in Cartesian coordinates are mutually exclusive, Eq. (8-13) can be resolved into three scalar Helmholtz's equations:

$$\nabla^2 E_x + k^2 E_x = 0 \qquad (8\text{-}14a)$$

$$\nabla^2 E_y + k^2 E_y = 0 \qquad (8\text{-}14b)$$

$$\nabla^2 E_z + k^2 E_z = 0 \qquad (8\text{-}14c)$$

For example, rewriting Eq. (8-14a), the three-dimensional differential wave equation for E_x is obtained as

$$\frac{\partial^2 E_x}{\partial x^2}+\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial^2 E_x}{\partial z^2}+k^2 E_x=0 \tag{8-15}$$

Direct substitution can show that a uniform plane wave is a general solution of Eq. (8-15), that is,

$$E_x=\hat{E}_{ox}e^{-j\left(k_x x+k_y y+k_z z\right)}=\hat{E}_{ox}e^{-j\mathbf{k}\cdot\mathbf{r}} \tag{8-16}$$

where $\hat{E}_{ox}$ is the complex amplitude. A similar procedure can be followed to obtain the y- and z-components of $\mathbf{E}$:

$$E_y=\hat{E}_{oy}e^{-j\mathbf{k}\cdot\mathbf{r}} \tag{8-17}$$

$$E_z=\hat{E}_{oz}e^{-j\mathbf{k}\cdot\mathbf{r}} \tag{8-18}$$

The three components given in Eqs. (8-16)-(8-18) can be combined into a single uniform plane wave only if they have the same wavevector. Consequently, in a lossless dielectric of infinite extent, a general solution of the vector Helmholtz's equation is a uniform plane wave propagating along the direction of the wavevector $\mathbf{k}$, in three-dimensional space. The electric field phasor of the wave takes on different forms:

$$\mathbf{E}=E_x\,\mathbf{a}_x+E_y\,\mathbf{a}_y+E_z\,\mathbf{a}_z=\left(\hat{E}_{ox}\,\mathbf{a}_x+\hat{E}_{oy}\,\mathbf{a}_y+\hat{E}_{oz}\,\mathbf{a}_z\right)e^{-j\mathbf{k}\cdot\mathbf{r}}$$

or

$$\boxed{\mathbf{E}=\mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}}} \tag{8-19}$$

Here, the complex amplitude vector $\mathbf{E}_o$ represents the amplitude and direction of the electric field intensity, which is generally expressed as

$$\boxed{\mathbf{E}_o=\hat{E}_o\,\mathbf{a}_E=E_o e^{j\varphi_o}\,\mathbf{a}_E} \tag{8-20}$$

where $\hat{E}_o$ is the complex amplitude, E_o is the amplitude, φ_o is the phase angle, and $\mathbf{a}_E$ is a unit vector in the direction of $\mathbf{E}$. For a wave propagating in a medium of infinite extent, the constant phase φ_o has no significance, and is therefore set to zero for simplicity.

The uniform plane wave expressed by Eq. (8-19) involves three kinds of vectors, which should be distinguished clearly from each other. The wavevector $\mathbf{k}$ tells us the wavelength, and shows the direction of propagation of the wave. The position vector $\mathbf{r}$ in the exponent $-\mathbf{k}\cdot\mathbf{r}$ is to specify that the time-phase of the electric field at a point with position vector $\mathbf{r}$ is equal to $-\mathbf{k}\cdot\mathbf{r}$. Finally, the complex amplitude vector $\mathbf{E}_o$ specifies the amplitude and direction of oscillation of the electric field intensity, in time.

The uniform plane wave expressed by Eq. (8-19) is a solution of the vector Helmholtz's equation that is derived from the two curl equations of Maxwell's equations. If the wave is an electromagnetic wave, it should also satisfy Gauss's law expressed by Eq. (8-9c) such that

$$\nabla \cdot \mathbf{E} = -j(\mathbf{k} \cdot \mathbf{E}_o) e^{-j\mathbf{k}\cdot\mathbf{r}} = 0$$

The equality in the above equation can be satisfied only if we have

$$\boxed{\mathbf{k} \cdot \mathbf{E}_o = 0} \tag{8-21}$$

Equation (8-21) shows that the complex amplitude vector $\mathbf{E}_o$ is always perpendicular to the wavevector $\mathbf{k}$. In other words, the electric field vector $\mathbf{E}$ should lie in the wavefront; the electromagnetic wave is a transverse wave.

Upon substituting the electric field expressed by Eq. (8-19) into Eq. (8-9a), we have

$$\begin{aligned} \nabla \times \mathbf{E} &= -j\mathbf{k} \times \mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}} \\ &= -j\omega\mu\mathbf{H} \end{aligned} \tag{8-22}$$

Inserting the expression for k as given in Eq. (8-12) into Eq. (8-22), the relation between the electric and magnetic field phasors is

$$\boxed{\mathbf{H} = \frac{1}{\eta}(\mathbf{a}_k \times \mathbf{E})} \tag{8-23}$$

Here, $\mathbf{a}_k$ is a unit vector in the direction of the wavevector $\mathbf{k}$. For a medium with a permeability μ and a permittivity ε, the intrinsic impedance η is defined as

$$\boxed{\eta = \sqrt{\frac{\mu}{\varepsilon}}} \qquad [\Omega] \tag{8-24}$$

It is a characteristic parameter of the medium. Next, substitution of Eq. (8-19) into Eq. (8-23) reveals that the magnetic field is also a uniform plane wave propagating with the same wavevector as the electric field, that is,

$$\begin{aligned} \mathbf{H} &= \frac{1}{\eta}(\mathbf{a}_k \times \mathbf{E}_o) e^{-j\mathbf{k}\cdot\mathbf{r}} \\ &\equiv \mathbf{H}_o e^{-j\mathbf{k}\cdot\mathbf{r}} \end{aligned} \tag{8-25}$$

Although $\mathbf{E}$ and $\mathbf{H}$ have exactly the same form as in Eqs. (8-19) and (8-25), the complex amplitude vectors $\mathbf{E}_o$ and $\mathbf{H}_o$ are mutually orthogonal in three-dimensional space. From Eq. (8-25), with the help of the expression $\mathbf{E}_o = \hat{E}_o \mathbf{a}_E$, the complex amplitude vector $\mathbf{H}_o$ is obtained as

$$\mathbf{H}_o = \frac{\hat{E}_o}{\eta}(\mathbf{a}_k \times \mathbf{a}_E) \equiv \hat{H}_o \mathbf{a}_H \tag{8-26}$$

It is evident from Eq. (8-26) that ***the intrinsic impedance*** η ***is the ratio between the complex amplitudes*** $\hat{E}_o$ ***and*** $\hat{H}_o$. The intrinsic impedance is measured in units of ohm $[\Omega]$, and, in free space, $\eta_o = \sqrt{\mu_o / \varepsilon_o} = 120\pi \cong 377[\Omega]$. In a lossless dielectric, η is real, meaning that the magnetic field is always in phase with the electric field. In contrast, in a lossy dielectric, η becomes a complex number, meaning that the magnetic field is generally out of phase with the electric field. Nevertheless, the directions of the three vectors, $\mathbf{E}_o$, $\mathbf{H}_o$, and $\mathbf{k}$, are perpendicular to each other, in three-dimensional space, such that $\mathbf{a}_k = \mathbf{a}_E \times \mathbf{a}_H$. The three unit vectors follow the right-hand rule: the right thumb point in the direction of the wavevector, when the fingers rotate from the electric field vector to the magnetic field vector. In view of these discussions, we note that the electric and magnetic field vectors lie in the wavefront, perpendicular to $\mathbf{k}$. This means that ***the electromagnetic wave is a transverse wave***.

The instantaneous electric and magnetic fields are obtained from their phasors $\mathbf{E}$ and $\mathbf{H}$, expressed by Eq. (8-19) and Eq. (8-23), as follows:

$$\mathscr{E}(\mathbf{r}, t) = \mathbf{a}_E E_o \cos(\omega t - \mathbf{k} \bullet \mathbf{r}) \tag{8-27a}$$

$$\mathscr{H}(\mathbf{r}, t) = \mathbf{a}_H \frac{E_o}{\eta} \cos(\omega t - \mathbf{k} \bullet \mathbf{r}) \tag{8-27b}$$

From Eq. (8-26), we obtain the orthogonality relation between the unit vectors of the wavevector, electric field vector, and magnetic field vector:

$$\boxed{\mathbf{a}_k = \mathbf{a}_E \times \mathbf{a}_H} \tag{8-28}$$

In a lossless dielectric, the electric and magnetic fields are always in phase, and propagates in the direction of $\mathbf{k}$, as illustrated in Fig. 8.2.

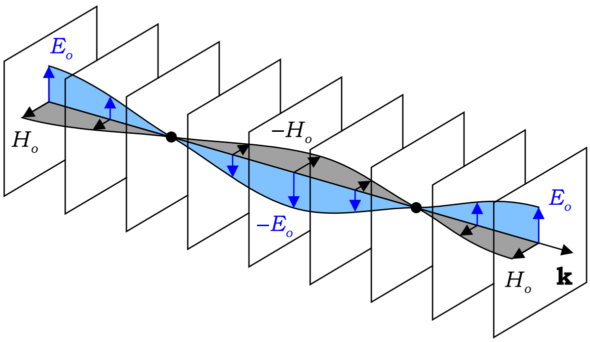


Fig. 8.2 A uniform plane wave in a lossless medium.

Example 8-2

A uniform plane wave $\mathbf{E} = \mathbf{a}_x E_o(1+j)e^{-jkz}$, operating at a frequency 2[GHz], propagates in a dielectric of $\varepsilon_r = 4$ and $\mu_r = 1$. Find

(a) $\mathscr{E}$ and $\mathscr{H}$, and

(b) wavelength.

Solution

(a) Rewriting the phasor, we have

$$\mathbf{E} = \mathbf{a}_x E_o \sqrt{2} e^{j\pi/4} e^{-jkz} \tag{8-29}$$

The real instantaneous value is

$$\mathscr{E} = \text{Re}\left[\mathbf{a}_x E_o \sqrt{2} e^{j\pi/4} e^{-jkz} e^{j\omega t}\right] = \mathbf{a}_x E_o \sqrt{2}\cos(\omega t - kz + \pi/4)$$

Inserting Eq. (8-29) into Eq. (8-23), noting that $\mathbf{a}_k = \mathbf{a}_z$ in Eq. (8-29), we have

$$\mathbf{H} = \frac{1}{\eta}(\mathbf{a}_k \times \mathbf{E}) = \frac{1}{\eta}(\mathbf{a}_z \times \mathbf{a}_x) E_o \sqrt{2} e^{j\pi/4} e^{-jkz}$$
$$= \frac{1}{\eta}\mathbf{a}_y E_o \sqrt{2} e^{j\pi/4} e^{-jkz}$$

The intrinsic impedance is

$$\eta = \sqrt{\frac{\mu_o}{\varepsilon_o \varepsilon_r}} = 188.5[\Omega]$$

The real instantaneous value is

$$\mathscr{H} = \text{Re}\left[\frac{1}{\eta}\mathbf{a}_y E_o \sqrt{2} e^{j\pi/4} e^{-jkz} e^{j\omega t}\right] = \mathbf{a}_y \frac{E_o}{\eta}\sqrt{2}\cos\left(\omega t - kz + \frac{\pi}{4}\right).$$

(b) From Eq. (8-12)

$$k = \omega\sqrt{\varepsilon\mu} = \frac{\omega\sqrt{\varepsilon_r}}{1/\sqrt{\mu_o \varepsilon_o}} = \frac{2\pi \times 2 \times 10^9 \times 2}{3 \times 10^8}$$

Thus, $\lambda = \dfrac{2\pi}{k} = 7.5[\text{cm}]$.

Exercise 8.4

Given the electric field phasor $\mathbf{E} = 4(2\,\mathbf{a}_x + j\,\mathbf{a}_y)e^{-j600z}$ of a uniform plane wave, propagating in a medium of $\eta = 200$, find the magnetic field phasor.

Ans. $\mathbf{H} = 0.02(-j\,\mathbf{a}_x + 2\,\mathbf{a}_y)e^{-j600z}$.

Review Questions with Hints

RQ 8.4	Express the uniform plane wave in phasor form.	[Eq.(8-19)]
RQ 8.5	Define intrinsic impedance of a lossless medium.	[Eq.(8-24)]
RQ 8.6	Define complex amplitude vector of **E**.	[Eq.(8-20)]
RQ 8.7	Write the relation between two phasors **H** and **E** in a lossless medium.	[Eq.(8-23)]

8.2.2 Poynting Vector and Power Flow

In the previous chapters we learned that energy can be stored in the static electric and magnetic fields, and that the static electric and magnetic energy densities are defined as $(1/2)\mathbf{D}\bullet\mathbf{E}$ and $(1/2)\mathbf{B}\bullet\mathbf{H}$, respectively. Energy can also be stored in the time-varying electric and magnetic fields. In this case, the stored energy propagates with the same velocity and direction as the given electromagnetic wave. In view of these, we note that an electromagnetic wave can transfer energy from one point to another in space and deliver power to the load. In the present section, we are mainly concerned with the power density of an electromagnetic wave and its flow in space.

Let us start with two curl equations of Maxwell's equations:

$$\nabla\times\mathscr{E} = -\frac{\partial\mathscr{B}}{\partial t} \tag{8-30a}$$

$$\nabla\times\mathscr{H} = \boldsymbol{J} + \frac{\partial\mathscr{D}}{\partial t} \tag{8-30b}$$

We note that the following vector identity always holds true for two smooth vector fields:

$$\nabla\bullet(\mathscr{E}\times\mathscr{H}) = \mathscr{H}\bullet(\nabla\times\mathscr{E}) - \mathscr{E}\bullet(\nabla\times\mathscr{H}) \tag{8-31}$$

Substituting Eq. (8-30) into Eq. (8-31), we have

$$-\nabla\bullet(\mathscr{E}\times\mathscr{H}) = \mathscr{E}\bullet\boldsymbol{J} + \mathscr{E}\bullet\frac{\partial\mathscr{D}}{\partial t} + \mathscr{H}\bullet\frac{\partial\mathscr{B}}{\partial t} \tag{8-32}$$

In a simple medium, in which ε , μ, and σ are constant, we can rewrite the terms on the right-hand side of Eq. (8-32) as follows:

$$\mathscr{E}\bullet\frac{\partial\mathscr{D}}{\partial t} = \frac{\varepsilon}{2}\frac{\partial}{\partial t}(\mathscr{E}\bullet\mathscr{E}) = \frac{1}{2}\frac{\partial}{\partial t}(\mathscr{D}\bullet\mathscr{E}) \tag{8-33a}$$

$$\mathscr{H}\bullet\frac{\partial\mathscr{B}}{\partial t} = \frac{\mu}{2}\frac{\partial}{\partial t}(\mathscr{H}\bullet\mathscr{H}) = \frac{1}{2}\frac{\partial}{\partial t}(\mathscr{B}\bullet\mathscr{H}) \tag{8-33b}$$

Inserting Eq. (8-33) back into Eq. (8-32), we have

$$\boxed{-\nabla\bullet(\mathscr{E}\times\mathscr{H}) = \mathscr{E}\bullet\boldsymbol{J} + \frac{1}{2}\frac{\partial}{\partial t}(\mathscr{D}\bullet\mathscr{E}) + \frac{1}{2}\frac{\partial}{\partial t}(\mathscr{B}\bullet\mathscr{H})} \tag{8-34}$$

Next, upon integrating both sides of Eq. (8-34) over a volume $\mathcal{V}$ and applying divergence theorem, we obtain

$$-\oint_S (\mathscr{E} \times \mathscr{H}) \cdot d\mathbf{s} = \int_{\mathcal{V}} \mathscr{E} \cdot \boldsymbol{J}\, dv + \frac{\partial}{\partial t}\left[\int_{\mathcal{V}} \frac{1}{2}(\mathscr{D} \cdot \mathscr{E})\, dv\right] + \frac{\partial}{\partial t}\left[\int_{\mathcal{V}} \frac{1}{2}(\mathscr{B} \cdot \mathscr{H})\, dv\right] \tag{8-35}$$

The first term on the right-hand side of Eq. (8-35) is the ohmic power-loss suffered in a given volume $\mathcal{V}$, whereas the two terms in bracket on the right-hand side represent the energies stored in $\mathscr{E}$ and $\mathscr{H}$, respectively, within $\mathcal{V}$. Accordingly, the sum of the second and third terms, on the right-hand side of Eq. (8-35), represents the time rate of increase of the total electromagnetic energy stored in $\mathcal{V}$. In view of these considerations, the left-hand side of Eq. (8-35) is interpreted as the total instantaneous power flowing into the volume $\mathcal{V}$, a part of which is dissipated in $\mathcal{V}$, and the rest is stored as the electromagnetic energy in the volume. Equation (8-35) is referred to as the Poynting's theorem, which states that ***the net power flowing into a given volume is equal to the sum of the ohmic power loss in the volume and the time rate of increase of the energy stored in the volume.***

In view of the surface integral on the left-hand side of Eq. (8-35), the integrand $\mathscr{E} \times \mathscr{H}$ is interpreted as the instantaneous power density measured in watts per square meter. For an electromagnetic wave with the instantaneous electric field $\mathscr{E}$ and the instantaneous magnetic field $\mathscr{H}$, the Poynting vector $\mathbf{S}$ is defined as

$$\boxed{\mathbf{S} = \mathscr{E} \times \mathscr{H}} \qquad [\mathrm{W/m^2}] \tag{8-36}$$

The unit of $\mathbf{S}$ is the watt per square meter. The Poynting vector is normal to both the electric and the magnetic field vectors, and parallel to the wavevector. It also points in the direction of flow of the power.

In most practical problems, the time-average power density is more useful than its instantaneous counterpart. It can be shown that the time-average power density is expressed in terms of the electric and the magnetic field phasors. In deriving the expression for the time-average power density, it is more convenient to work with the electric and magnetic fields expressed in terms of complex exponentials than cosines. To show how the amplitude E_o, complex amplitude $\hat{E}_o$, complex amplitude vector $\mathbf{E}_o$, and phasor $\mathbf{E}$ are used for the expressions for $\mathscr{E}$, we write

$$\begin{aligned}
\mathscr{E} = \mathbf{a}_E E_o \cos(\omega t - \mathbf{k}\cdot\mathbf{r} + \varphi_e) &= \frac{1}{2}\left[\mathbf{a}_E\left(E_o e^{j\varphi_e}\right)e^{j\omega t - j\mathbf{k}\cdot\mathbf{r}} + \mathbf{a}_E\left(E_o e^{-j\varphi_e}\right)e^{-j\omega t + j\mathbf{k}\cdot\mathbf{r}}\right] \\
&\equiv \frac{1}{2}\left[\left(\mathbf{a}_E \hat{E}_o\right)e^{-j\mathbf{k}\cdot\mathbf{r}}e^{j\omega t} + \left(\mathbf{a}_E \hat{E}_o^*\right)e^{+j\mathbf{k}\cdot\mathbf{r}}e^{-j\omega t}\right] \\
&\equiv \frac{1}{2}\left[\mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}}e^{j\omega t} + \mathbf{E}_o^* e^{+j\mathbf{k}\cdot\mathbf{r}}e^{-j\omega t}\right] \\
&\equiv \frac{1}{2}\left[\mathbf{E}e^{j\omega t} + \mathbf{E}^* e^{-j\omega t}\right]
\end{aligned} \tag{8-37a}$$

where * denotes the complex conjugate. Following the same procedure, the instantaneous magnetic field is expressed as

$$\mathscr{H} = \mathbf{a}_H H_o \cos(\omega t - \mathbf{k}\bullet\mathbf{r} + \varphi_h) = \frac{1}{2}\left[\mathbf{H}e^{j\omega t} + \mathbf{H}^* e^{-j\omega t}\right] \tag{8-37b}$$

Inserting Eqs. (8-37a) and (8-37b) into Eq. (8-36), we have

$$\begin{aligned}\mathbf{S} &= \frac{1}{2}\left[\mathbf{E}e^{j\omega t} + \mathbf{E}^* e^{-j\omega t}\right] \times \frac{1}{2}\left[\mathbf{H}e^{j\omega t} + \mathbf{H}^* e^{-j\omega t}\right] \\ &= \frac{1}{4}\left[\mathbf{E}\times\mathbf{H}e^{j2\omega t} + \mathbf{E}^*\times\mathbf{H}^* e^{-j2\omega t} + \mathbf{E}\times\mathbf{H}^* + \mathbf{E}^*\times\mathbf{H}\right]\end{aligned} \tag{8-38}$$

In general, the time average of an instantaneous value $\mathbf{G}$ is defined as

$$\langle\mathbf{G}\rangle = \frac{1}{T}\int_{-T/2}^{T/2}\mathbf{G}dt$$

where T is the temporal period of $\mathbf{G}$. Upon taking the time average of Eq. (8-38), the time-average power density, or time-average Poynting vector is

$$\boxed{\langle\mathbf{S}\rangle = \frac{1}{2}\mathrm{Re}\left[\mathbf{E}\times\mathbf{H}^*\right]} \qquad [\mathrm{W/m^2}] \tag{8-39}$$

where $\mathbf{E}$ and $\mathbf{H}$ are the phasors of the electric and magnetic fields, and * denotes complex conjugate.

Since the electric and magnetic fields in Eq. (8-34) are all time-harmonic quantities, the time averages of the second and third terms on the right-hand side of Eq. (8-34) vanish. Thus, by taking the time average of both sides of Eq. (8-34), the Poynting's theorem becomes

$$-\nabla\bullet\langle\mathscr{E}\times\mathscr{H}\rangle = \langle\mathscr{E}\bullet\mathbf{J}\rangle \tag{8-40a}$$

Alternatively, using the phasors of the electric field intensity and current density, we write

$$-\nabla\bullet\langle\mathbf{S}\rangle = \frac{1}{2}\mathrm{Re}\left[\mathbf{E}\bullet\mathbf{J}^*\right] \tag{8-40b}$$

Equation Eq. (8-40b) shows that the time-average ohmic power loss per unit volume is equal to the negative divergence of the time-average Poynting vector.

The time-average Poynting vector expressed by Eq. (8-39) can also be used for computing the total time-average power density carried by the two waves co-propagating in a region of space. The total power density is of course obtained from the total electric and magnetic field intensities. Through the time-average Poynting vector, it can be readily shown that the total power density is equal to the sum of the time-average power densities of the individual waves, if the waves are independent of each other. By using the time-average Poynting vector, for instance, we can verify that the two waves co-propagating with $\mathbf{E}_1 = 2\mathbf{a}_x$ and

$\mathbf{E}_2 = 3\,\mathbf{a}_y$ are independent of each other, and show that the two waves co-propagating with $\mathbf{E}_3 = 2\,\mathbf{a}_x$ and $\mathbf{E}_4 = 3\,\mathbf{a}_x$ should be treated as a single wave having an electric field phasor $\mathbf{E} = \mathbf{E}_3 + \mathbf{E}_4 = 5\,\mathbf{a}_x$.

Example 8-3

Given an electromagnetic wave with $\mathscr{E}(\mathbf{r},t) = \mathbf{a}_x E_o \cos(\omega t - kz)$ and $\mathscr{H}(\mathbf{r},t) = \mathbf{a}_y E_o \sqrt{\varepsilon/\mu}\cos(\omega t - kz)$, determine $\langle \mathbf{S} \rangle$ by using

(a) $\mathscr{E}$ and $\mathscr{H}$

(b) $\mathbf{E}$ and $\mathbf{H}$

Solution

(a) From Eq. (8-36), we have

$$\mathbf{S} = \mathscr{E} \times \mathscr{H} = \mathbf{a}_z E_o^2 \sqrt{\varepsilon/\mu}\cos^2(\omega t - kz)$$

With the help of trigonometry, we rewrite the above equation as

$$\mathbf{S} = \mathbf{a}_z E_o^2 \sqrt{\varepsilon/\mu}\,\tfrac{1}{2}\left[1 + \cos(2\omega t - 2kz)\right]$$

Taking the time-average of the above equation, we obtain

$$\langle \mathbf{S} \rangle = \mathbf{a}_z \frac{E_o^2}{2}\sqrt{\frac{\varepsilon}{\mu}}. \tag{8-41}$$

(b) Phasors of the electric and magnetic fields are

$$\mathbf{E} = \mathbf{a}_x E_o e^{-jkz}$$

$$\mathbf{H} = \mathbf{a}_y E_o \sqrt{\varepsilon/\mu}\, e^{-jkz}$$

From Eq. (8-39), we obtain

$$\begin{aligned}\langle \mathbf{S} \rangle = \tfrac{1}{2}\mathrm{Re}\left[\mathbf{E}\times\mathbf{H}^*\right] &= \tfrac{1}{2}\mathrm{Re}\left[\left(\mathbf{a}_x E_o e^{-jkz}\right)\times\left(\mathbf{a}_y E_o \sqrt{\varepsilon/\mu}\, e^{-jkz}\right)^*\right] \\ &= \tfrac{1}{2} E_o^2 \sqrt{\varepsilon/\mu}\,\mathrm{Re}\left[\left(\mathbf{a}_x e^{-jkz}\right)\times\left(\mathbf{a}_y e^{jkz}\right)\right]\end{aligned}$$

Thus,

$$\langle \mathbf{S} \rangle = \mathbf{a}_z \frac{E_o^2}{2}\sqrt{\frac{\varepsilon}{\mu}}$$

This is the same as Eq. (8-41).

Exercise 8.5

Given a wave $\mathbf{E} = 10\mathbf{a}_x e^{-j600(3y+4z)}$ in free space, find the time-average power flowing through a unit area of the $z = 0$ plane.

Ans. $1/(3\pi)$ [W].

Review Questions with Hints

RQ 8.8 State Poynting's theorem in words. [Eq.(8-34)]
RQ 8.9 Express Poynting vector in terms of $\mathscr{E}$ and $\mathscr{H}$. [Eq.(8-36)]
RQ 8.10 Express the time-average power density in terms of the electric and magnetic field phasors. [Eq.(8-39)]

8.2.3 Polarization of a Uniform Plane Wave

The electromagnetic uniform plane wave is a transverse wave with the electric field vector lying in the wavefront, or a plane perpendicular to the direction of propagation of the wave. The polarization of a wave describes the way in which the electric field vector changes as a function of time, at a point in space. If the tip of the electric field vector moves back and forth along a straight line in time, at a fixed point in space, the wave is called linearly polarized. If the locus traced by the tip of the electric field vector forms a circle, the wave is called circularly polarized. In general, the locus forms an ellipse in the transverse plane. For a uniform plane wave, we only need to trace the electric field vector, because the magnetic field vector is related to the electric field vector by Eq. (8-25).

8.2.3.1 Linearly Polarized Wave

For a uniform plane wave propagating in the $+z$-direction, the electric field vector should be in the xy-plane, having a x-component and a y-component, in general. The electric field phasor of the wave is expressed, in general, as

$$\boxed{\mathbf{E} = \mathbf{E}_o e^{-jkz} = \left(\hat{E}_{ox}\mathbf{a}_x + \hat{E}_{oy}\mathbf{a}_y\right)e^{-jkz}} \tag{8-42}$$

where $\hat{E}_{ox}$ and $\hat{E}_{oy}$ are the complex amplitudes of the x- and y-components, respectively.

If the amplitudes are real such as $\hat{E}_{ox} = E_{ox}$ and $\hat{E}_{oy} = E_{oy}$, the instantaneous electric field intensity is written as

$$\begin{aligned}\mathscr{E} &= \mathrm{Re}\left[\left(E_{ox}\mathbf{a}_x + E_{oy}\mathbf{a}_y\right)e^{-jkz}e^{j\omega t}\right] \\ &= \left(E_{ox}\mathbf{a}_x + E_{oy}\mathbf{a}_y\right)\cos(\omega t - kz)\end{aligned} \tag{8-43}$$

At a point on a plane of constant z, the electric field intensity shown in Eq. (8-43) is vibrating along a straight line parallel to the direction of $E_{ox}\mathbf{a}_x + E_{oy}\mathbf{a}_y$. This wave is said to be linearly polarized in the direction of $E_{ox}\mathbf{a}_x + E_{oy}\mathbf{a}_y$. In view of Eq. (8-43), we may consider the wave as the sum of two uniform plane waves: one is linearly polarized in the x-direction and the other is linearly polarized in the y-direction (see Fig. 8.3). Since the amplitudes E_{ox} and E_{oy} are both real, the two component waves are in phase, and the two cosine curves cross the z-axis at the same point as can be seen in Fig. 8.3.

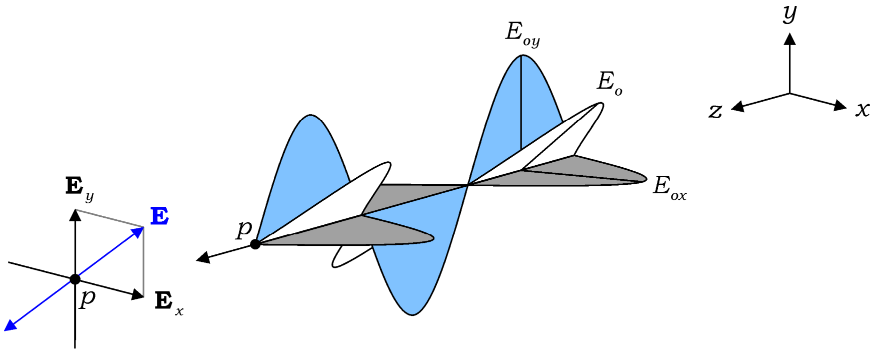


Fig. 8.3 Linearly polarized wave.

Example 8-4

Given a linearly polarized wave, $\mathbf{E} = E_o\left(\mathbf{a}_x + \sqrt{3}\mathbf{a}_y\right)e^{-jkz}$, propagating in free space, find
(a) angle between $\mathbf{E}$ and the x-axis, and
(b) time-average power density.

Solution

(a) The rotation angle of $\mathbf{E}$ is

$$\theta = \tan^{-1}\left(\sqrt{3}\right) = 60^o.$$

(b) From Eq. (8-23) we obtain the magnetic field phasor as

$$\mathbf{H} = \sqrt{\varepsilon_o/\mu_o}\,\mathbf{a}_z \times E_o\left(\mathbf{a}_x + \sqrt{3}\mathbf{a}_y\right)e^{-jkz} = E_o\sqrt{\varepsilon_o/\mu_o}\left(\mathbf{a}_y - \sqrt{3}\mathbf{a}_x\right)e^{-jkz}$$

From Eq. (8-39), we obtain

$$\begin{aligned}\langle\mathbf{S}\rangle &= \tfrac{1}{2}\mathrm{Re}\left[E_o\left(\mathbf{a}_x + \sqrt{3}\mathbf{a}_y\right)e^{-jkz} \times \left\{E_o\sqrt{\varepsilon_o/\mu_o}\left(\mathbf{a}_y - \sqrt{3}\mathbf{a}_x\right)e^{-jkz}\right\}^*\right] \\ &= \tfrac{1}{2}E_o^2\sqrt{\varepsilon_o/\mu_o}\left(\mathbf{a}_x + \sqrt{3}\mathbf{a}_y\right)\times\left(\mathbf{a}_y - \sqrt{3}\mathbf{a}_x\right) \\ &= \tfrac{1}{2}E_o^2\sqrt{\varepsilon_o/\mu_o}\,(1+3)\,\mathbf{a}_z\end{aligned}$$

The time-average power density is equal to the sum of those of two component waves.

Exercise 8.6

Are the following waves linearly polarized? (a) $\mathbf{E} = 4\left(\mathbf{a}_x e^{j0.2} + \mathbf{a}_y e^{j0.2}\right)e^{-j300z}$, (b) $\mathbf{E} = \left(3\mathbf{a}_x - \mathbf{a}_y\right)e^{+j400z}$, (c) $\mathbf{E} = 2\left(\mathbf{a}_x e^{-j0.3} + \mathbf{a}_y e^{j0.3}\right)e^{-j500z}$, and (d) $\mathbf{E} = \left[\mathbf{a}_x \sin(30^o) + \mathbf{a}_y \cos(30^o)\right]e^{-j600z}$

Ans. (a) yes, (b) yes, (c) no, (d) yes.

8.2.3.2 Circularly Polarized Wave

Let us now consider the case, in which the complex amplitudes $\hat{E}_{ox}$ and $\hat{E}_{oy}$ of $\mathbf{E}$ have the same magnitude, $|\hat{E}_{ox}| = |\hat{E}_{oy}| = E_o$, but the y-component lags behind the x-component by 90^o in time dimension. This wave is called a right-hand circularly polarized wave for the reason that will become evident shortly. For a right-hand circularly polarized wave, the electric field phasor is written as

$$\boxed{\mathbf{E} = \left(E_o\mathbf{a}_x - jE_o\mathbf{a}_y\right)e^{-jkz} = \left(E_o\mathbf{a}_x + E_oe^{-j\pi/2}\mathbf{a}_y\right)e^{-jkz}} \tag{8-44}$$

In real instantaneous form, the electric field intensity is expressed as

$$\mathscr{E}(z,t) = \mathbf{a}_xE_o\cos\left(\omega t - kz\right) + \mathbf{a}_yE_o\cos\left(\omega t - kz - \pi/2\right) \tag{8-45}$$

Here, let us examine the electric field vector at a point on a plane of constant z, or simply the $z = 0$ plane without loss of generality. In this case, Eq. (8-45) reduces to

$$\mathscr{E}\left(0,t\right) = \mathbf{a}_xE_o\cos\left(\omega t\right) + \mathbf{a}_yE_o\sin\left(\omega t\right) \tag{8-46}$$

At time $t = 0$, we have $\mathscr{E}(0,0) = E_o\mathbf{a}_x$, which is parallel to the $+x$-axis. At a later time $t = \pi/(2\omega)$, we have $\mathscr{E}(0,\pi/2\omega) = E_o\mathbf{a}_y$, which is parallel to the $+y$-axis. In view of these, we note that as time increases, the electric field vector rotates in the counterclockwise direction in the xy-plane, with no change in the magnitude (see Fig. 8.4). The electric field vector completes a turn as ωt changes by 2π, while the tip of the electric field vector follows a circular locus. This type of wave is called a right-hand circularly polarized wave. The polarization handedness is determined by the rotation direction of the electric field vector relative to the propagation direction of the wave; the right thumb points in the direction of propagation of the wave when the fingers follow the rotation of the electric field vector.

A right-hand circularly polarized wave is illustrated in Fig. 8.4. The y-component lags behind the x-component by 90^o in time phase. It means that the y-component arrives at a reference point p at a later time than the x-component by a quarter of the temporal period. In space dimension, the y-component curve appears to be pushed back in the $-z$-direction by a quarter wavelength relative to the x-component curve. Since both component waves propagate in the same direction, with the same velocity, the phase relationship between them is maintained at all times, at any point in space. It should be noted that the phase difference is a relative quantity so that the x-component may be said to lead the y-component by 270^o in time phase.

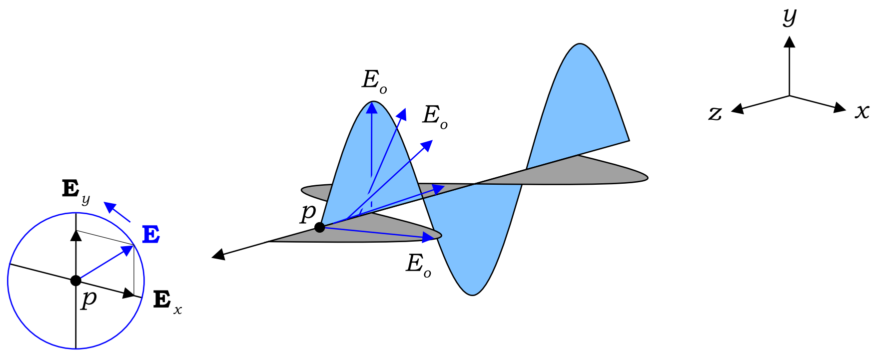


Fig. 8.4 A right-hand circularly polarized wave.

Next, we consider the case in which the y-component leads the x-component by 90^o in time phase, while the amplitudes of the two components are the same. This wave is called a left-hand circularly polarized wave; the left thumb points in the direction of propagation of the wave when the fingers follow the rotation of the electric field vector. For a left-hand circularly polarized wave, the electric field phasor is expressed as

$$\boxed{\mathbf{E} = \left(E_o\mathbf{a}_x + jE_o\mathbf{a}_y\right)e^{-jkz} = \left(E_o\mathbf{a}_x + E_o e^{j\pi/2}\mathbf{a}_y\right)e^{-jkz}} \tag{8-47}$$

In real instantaneous form, the electric field intensity is written as

$$\mathscr{E}(z,t) = \mathbf{a}_x E_o \cos\left(\omega t - kz\right) + \mathbf{a}_y E_o \cos\left(\omega t - kz + \pi/2\right) \tag{8-48}$$

Following the same procedure, we consider a point on the $z = 0$ plane, where the electric field intensity is given as a function of time only as

$$\mathscr{E}(0,t) = \mathbf{a}_x E_o \cos\left(\omega t\right) - \mathbf{a}_y E_o \sin\left(\omega t\right) \tag{8-49}$$

At time $t = 0$, we have $\mathscr{E}(0,0) = E_o\mathbf{a}_x$, which is parallel to the +x-axis. At a later time $t = \pi/(2\omega)$, we have $\mathscr{E}(0,\pi/2\omega) = -E_o\mathbf{a}_y$, which is parallel to the −y-axis. In view of these, we note that as time increases, the electric field vector rotates in the clockwise direction in the xy-plane, with no change in the magnitude. The electric field vector completes a turn as ωt changes by 2π, while the tip of the electric field vector follows a circular locus.

A left-hand circularly polarized wave is illustrated in Fig. 8.5. The y-component leads the x-component by 90^o in time phase. Thus, the y-component arrives at a reference point p at an earlier time than the x-component by a quarter of the temporal period. In space dimension, the y-component curve appears to be pulled forward in the +z-direction by a quarter wavelength relative to the x-component curve.

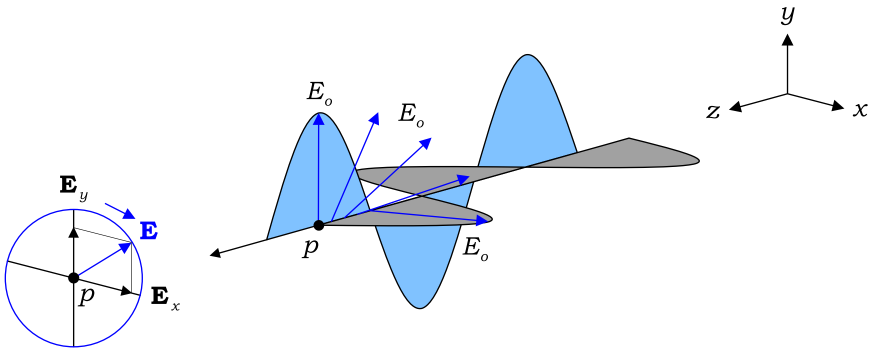


Fig. 8.5 A left-hand circularly polarized wave.

Example 8-5

Resolve a linearly polarized wave $\mathbf{E}_1 = E_o\,\mathbf{a}_x e^{-jkz}$ into two circularly polarized waves.

Solution

By adding and subtracting the term $(1/2)jE_o\,\mathbf{a}_y$ on the right-hand side of the expression for $\mathbf{E}_1$, we have

$$\mathbf{E}_1 = \frac{1}{2}\left[\left(E_o\,\mathbf{a}_x - jE_o\,\mathbf{a}_y\right) + \left(E_o\,\mathbf{a}_x + jE_o\,\mathbf{a}_y\right)\right]e^{-jkz} \tag{8-50}$$

The first term on the right-hand side of Eq. (8-50) represents a right-hand circularly polarized wave, whereas the second term a left-hand circularly polarized wave.

From the opposite point of view, if two electric field vectors of the same magnitude rotate in the opposite directions in the xy-plane, and cross the $+x$-axis at the same time, the sum of these waves is one that is linearly polarized in the x-direction.

Exercise 8.7

Determine the polarization handedness of the following circularly polarized waves: (a) $\mathbf{E} = 10\left(j\mathbf{a}_x + \mathbf{a}_y\right)e^{-j300z}$, (b) $\mathbf{E} = 20\left(j\mathbf{a}_x - \mathbf{a}_y\right)e^{-j400z}$, and (c) $\mathbf{E} = 30\left(\mathbf{a}_x - j\mathbf{a}_y\right)e^{+j500z}$.

Ans. (a) right-hand, (b) left-hand, (c) left-hand.

8.2.3.3 Elliptically Polarized Wave

A uniform plane wave is said to be elliptically polarized if it is not linearly or circularly polarized. The tip of the electric field vector follows an elliptical locus in the plane perpendicular to the direction of propagation of the wave. The linearly and circularly polarized waves are special cases of the elliptically polarized wave. The electric field phasor of an elliptically polarized wave is generally expressed as

$$\boxed{\mathbf{E} = \left(E_{ox}\mathbf{a}_x + E_{oy}e^{j\varphi}\mathbf{a}_y\right)e^{-jkz}} \tag{8-51}$$

Here, E_{ox} and E_{oy} are positive real. The phase angle φ determines the phase difference between the two components. The instantaneous form of the elliptically polarized wave is

$$\mathscr{E} = E_{ox}\mathbf{a}_x \cos(\omega t - kz) + E_{oy}\mathbf{a}_y \cos(\omega t - kz + \varphi) \tag{8-52}$$

The electric field vector changes in both the magnitude and the direction with time.

We now examine the electric field vector of an elliptically polarized wave at a point on the $z = 0$ plane. At the point, from Eq. (8-52), the two components are given as functions of time only as

$$\mathscr{E}_x = E_{ox}\cos(\omega t) \tag{8-53a}$$

$$\mathscr{E}_y = E_{oy}\cos(\omega t + \varphi) \tag{8-53b}$$

For $0 < \varphi < \pi$, at time $t = 0$, we have $\mathscr{E}_x = E_{ox}$ and $\mathscr{E}_y = E_{oy}\cos\varphi$, meaning that the electric field vector is either in the first or fourth quadrant in the xy-plane with a magnitude $\sqrt{E_{ox}^2 + E_{oy}^2\cos^2\varphi}$, depending φ. At a later time $t = \pi/(2\omega)$, we have $\mathscr{E}_x = 0$ and $\mathscr{E}_y = -E_{oy}\sin\varphi$, meaning that the electric field vector is parallel to the $-y$-axis with a magnitude $|E_{oy}\sin\varphi|$. We see that as time increases, the electric field vector rotates in the clockwise direction in the xy-plane, although the magnitude may change with time. This wave is said to be left-hand elliptically polarized.

If the y-component lags behind the x-component in time phase such that $-\pi < \varphi < 0$ in Eq. (8-51), the wave is called a right-hand elliptically polarized wave.

The shape of the polarization ellipse depends on the ratio E_{oy}/E_{ox} and the phase angle φ. Referring to Fig.8.6, in which an ellipse has the major axis $2E'_{ox}$ along the x'-axis, and the minor axis $2E'_{oy}$ along the y'-axis, we can obtain an expression for the rotation angle θ as follows. Consider a left-hand elliptically polarized wave, having the electric field vector whose tip follows the same elliptical locus as shown in Fig. 8.6. In this case, the electric field intensity of the wave is expressed in the primed coordinates as

$$\mathscr{E}'_x = E'_{ox}\cos(\omega t + \varphi') \tag{8-54a}$$

$$\mathscr{E}'_y = -E'_{oy}\sin(\omega t + \varphi') \tag{8-54b}$$

The two amplitudes E'_{ox} and E'_{oy}, and the phase angle φ' are to be determined from E_{ox}, E_{oy}, and φ given in the general expression in Eq. (8-53). Taking the coordinate transformation of the electric field intensities shown in Eq. (8-53) into the primed coordinates, we have

$$\mathscr{E}'_x = E_{ox}\cos(\omega t)\cos\theta + E_{oy}\cos(\omega t+\varphi)\sin\theta \qquad (8\text{-}55a)$$

$$\mathscr{E}'_y = -E_{ox}\cos(\omega t)\sin\theta + E_{oy}\cos(\omega t+\varphi)\cos\theta \qquad (8\text{-}55b)$$

By equating Eq. (8-54) with Eq. (8-55) we obtain

$$\tan(2\theta) = \frac{2E_{ox}E_{oy}\cos\varphi}{\left(E_{ox}\right)^2 - \left(E_{oy}\right)^2} \qquad (8\text{-}56)$$

If $\varphi = 0$, Eq. (8-56) reduces to $\tan\theta = E_{oy}/E_{ox}$, representing a linearly polarized wave.

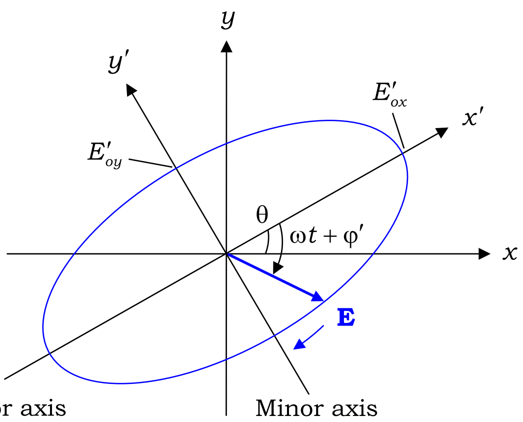


Fig. 8.6 An ellipse with a rotation angle θ.

Example 8-6
Find the time-average power density of a uniform plane wave, propagating in free space, with the electric field phasor $\mathbf{E} = \left(E_{ox}\mathbf{a}_x + E_{oy}e^{j\varphi}\mathbf{a}_y\right)e^{-jkz}$.

Solution
From Eq. (8-23) we obtain the magnetic field phasor as

$$\mathbf{H} = \frac{1}{\eta_o}(\mathbf{a}_k \times \mathbf{E}) = \frac{1}{\eta_o}\mathbf{a}_z \times \left(E_{ox}\mathbf{a}_x + E_{oy}e^{j\varphi}\mathbf{a}_y\right)e^{-jkz}$$
$$= \frac{1}{\eta_o}\left(E_{ox}\mathbf{a}_y - E_{oy}e^{j\varphi}\mathbf{a}_x\right)e^{-jkz}$$

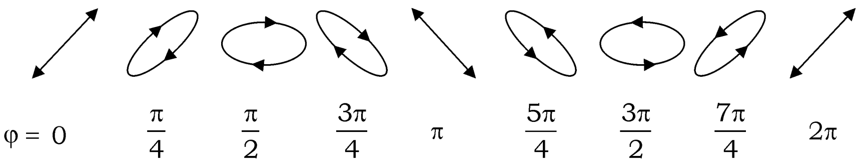
The time-average Poynting vector is

$$\begin{aligned}\langle \mathbf{S} \rangle &= \frac{1}{2}\mathrm{Re}\left[\mathbf{E}\times\mathbf{H}^*\right] \\ &= \frac{1}{2\eta_o}\mathrm{Re}\left[\left(E_{ox}\mathbf{a}_x + E_{oy}e^{j\varphi}\mathbf{a}_y\right)e^{-jkz}\times\left(E_{ox}\mathbf{a}_y - E_{oy}e^{-j\varphi}\mathbf{a}_x\right)e^{jkz}\right] \\ &= \frac{1}{2\eta_o}\left(E_{ox}^2 + E_{oy}^2\right)\mathbf{a}_z\end{aligned}$$

The time-average power density is equal to the sum of those of two linearly-polarized component waves.

Exercise 8.8
Sketch the loci of **E** expressed by Eq. (8-51) for $0 \le \varphi \le 2\pi$ and $E_{ox} = 2E_{oy}$.

Ans.



Review Questions with Hints

RQ 8.11 What is the polarization of a uniform plane wave? [Figs.8.3,8.4,8.5]
RQ 8.12 Under what conditions is a uniform plane wave linearly polarized, and circularly polarized? [Eqs.(8-42)(8-44)(8-47)]
RQ 8.13 Are the right-hand and left-hand circularly polarized waves independent of each other? [Eqs.(8-44)(8-47)]

8.2.4 Uniform Plane Wave in a Lossy Medium

We now turn our attention to the electromagnetic wave propagating in a lossy medium. Dielectrics with free electrons and conductors of a finite conductivity are good examples of lossy media. A plane wave can be supported in a lossy medium of infinite extent, although the wave suffers power loss as it propagates through the medium. There are two mechanisms of power loss in a lossy dielectric. First, when the electric field of the wave induces electric dipoles in the material, the bound charges of the dipole vibrate to and fro along a straight line at the same frequency as the electric field. However, the strong interaction between the neighboring atoms tends to hinder the electric dipoles from oscillating in unison with the electric field. This phase delay causes a damping of the electric dipole, and thus a dissipation of the incident power in the form of heat. Second, the free electrons in a lossy dielectric constitute a conduction current flowing in the electric field of the wave, and thus cause ohmic power loss of the wave in the material.

The power loss in a lossy dielectric can be easily incorporated into the differential wave equation by means of a complex permittivity, $\hat{\varepsilon} = \varepsilon' - j\varepsilon''$. While the real part determines the phase constant of the wave propagating in the medium, the imaginary part describes the attenuation of the wave in the medium.

The magnetic field of the incident wave may interact with the lattice atoms through the magnetic dipole moment, and thus the wave may suffer an additional power loss because of the damping of the magnetic dipole. Following the same procedure, we may introduce a complex permeability to account for the power dissipation in a lossy magnetic material. In most material media of our interest, however, the magnetic response of the medium is so weak in comparison with the electric counterpart that we may ignore the magnetic interaction and assume $\mu = \mu_o$.

8.2.4.1 Lossy Dielectric with a Damping Force

In a homogeneous(ε and μ being independent of position), linear(ε and μ being independent of the magnitude of $\mathscr{E}$ and $\mathscr{H}$), isotropic(ε and μ being independent of the direction of $\mathscr{E}$ and $\mathscr{H}$), and lossy(ε being a complex number) dielectric, we begin with Maxwell's equations given in phasor form as

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{8-57a}$$

$$\nabla \times \mathbf{H} = j\omega\hat{\varepsilon}\mathbf{E} = j\omega(\varepsilon' - j\varepsilon'')\mathbf{E} = j\omega\varepsilon_o(\varepsilon_r' - j\varepsilon_r'')\mathbf{E} \tag{8-57b}$$

$$\nabla \bullet \mathbf{E} = 0 \tag{8-57c}$$

$$\nabla \bullet \mathbf{H} = 0 \tag{8-57d}$$

Here, the material is assumed to have no free charges as in Eq. (8-57c), and no conduction currents as in Eq. (8-57b). We assume that the complex permittivity in Eq. (8-57b) comes from the damping of the electric dipole in the material. From Eq. (8-57), we write Helmholtz's equation as

$$\nabla^2\mathbf{E} + \hat{k}^2\mathbf{E} = 0 \tag{8-58}$$

where the complex wavenumber is defined as

$$\boxed{\hat{k} = \omega\sqrt{\mu\hat{\varepsilon}} = \omega\sqrt{\mu(\varepsilon' - j\varepsilon'')}} \tag{8-59}$$

Noting that Eq. (8-58) only differs from Eq. (8-11) in the constant $\hat{k}$, the general solution for Eq. (8-58) is a uniform plane wave with the complex wavenumber, that is

$$\boxed{\mathbf{E} = \mathbf{E}_o e^{-j\hat{k}\mathbf{a}_k \bullet \mathbf{r}}} \tag{8-60}$$

where $\mathbf{a}_k$ is a unit vector in the direction of propagation of the wave. Without loss of generality, in a medium of infinite extent, we set $\mathbf{a}_k = \mathbf{a}_z$. Under this condition, the electric field phasor is simply given as

$$\boxed{\begin{aligned}\mathbf{E} &= \mathbf{E}_o e^{-j\hat{k}z} \\ &\equiv \mathbf{E}_o e^{-\gamma z}\end{aligned}} \tag{8-61}$$

where the propagation constant γ is defined as

$$\boxed{\begin{aligned}\gamma &= j\hat{k} = j\omega\sqrt{\mu(\varepsilon' - j\varepsilon'')} \\ &\equiv \alpha + j\beta\end{aligned}} \qquad [\mathrm{m}^{-1}] \tag{8-62}$$

Here, α is the attenuation constant measured in nepers per meter $[\mathrm{Np/m}]$, and β is the phase constant measured in radians per meter $[\mathrm{rad/m}]$. From Eq. (8-62) we obtain expressions for α and β as follows:

$$\boxed{\alpha = \omega\sqrt{\frac{\mu\varepsilon'}{2}}\left[\sqrt{1+\left(\frac{\varepsilon''}{\varepsilon'}\right)^2} - 1\right]^{1/2}} \qquad [\mathrm{Np/m}] \tag{8-63a}$$

$$\boxed{\beta = \omega\sqrt{\frac{\mu\varepsilon'}{2}}\left[\sqrt{1+\left(\frac{\varepsilon''}{\varepsilon'}\right)^2} + 1\right]^{1/2}} \qquad [\mathrm{rad/m}] \tag{8-63b}$$

In a lossless dielectric for which $\varepsilon'' = 0$, the attenuation constant reduces to $\alpha = 0$, as expected. The physical significance of α and β will become evident shortly. The ratio $\varepsilon''/\varepsilon'$ is called the loss tangent, which is denoted as

$$\boxed{\tan\xi = \frac{\varepsilon''}{\varepsilon'}} \tag{8-64}$$

This can be used as a measure of power loss in the medium. The angle ξ is called the loss angle.

Let us now examine the electromagnetic wave propagating in a lossy dielectric characterized by the parameters expressed by Eqs. (8-62)-(8-64). Inserting Eq. (8-62) into Eq. (8-61), we write the electric field of the wave, in phasor and instantaneous forms, as

$$\mathbf{E} = (E_o\,\mathbf{a}_E)\,e^{-\alpha z}e^{-j\beta z} \tag{8-65a}$$

$$\mathscr{E} = (E_o\,\mathbf{a}_E)\,e^{-\alpha z}\cos(\omega t - \beta z) \tag{8-65b}$$

where $\mathbf{a}_E$ is the unit vector of the electric field vector. From Eq. (8-65) we see that the attenuation constant α shows how fast the amplitude decreases in the $+z$-direction, while the phase constant β describes how the phase changes in the $+z$-

direction at a fixed time. The attenuation constant α must be positive in a passive medium in which no amplification of the electric field can occur.

If two points are separated by a wavelength λ on the z-axis, it means that the phase difference between the electric field intensities observed at the points is 2π ; that is, $\beta\lambda = 2\pi$. Thus, the relation between the phase constant and wavelength is obtained as

$$\boxed{\beta = \frac{2\pi}{\lambda}} \qquad \text{[rad/m]} \qquad (8\text{-}66)$$

The wavelength λ depends on ε'' as well as ε' in a lossy dielectric, as is evident from Eq. (8-63b).

Whether the material medium is lossless or not, the phase velocity of the wave is always given by

$$\boxed{\upsilon_p = \frac{\omega}{\beta}} \qquad \text{[m/s]} \qquad (8\text{-}67)$$

which also depends on both the real and imaginary parts of $\hat{\varepsilon}$.

We note that the loss tangent $\varepsilon''/\varepsilon'$ determines the deviations of β expressed by Eq. (8-63b), λ expressed by Eq. (8-66), and υ_p expressed by Eq. (8-67) from their counterparts in an identical medium with no loss.

Substitution of the electric field phasor expressed by Eq. (8-65a) into Eq. (8-57a) leads to the magnetic field phasor, that is,

$$\begin{aligned}\nabla \times \mathbf{E} &= \nabla \times \left(E_o\, \mathbf{a}_E e^{-\gamma z}\right) = -\gamma \mathbf{a}_z \times \left(E_o\, \mathbf{a}_E e^{-\gamma z}\right) \\ &= -j\omega\mu \mathbf{H}\end{aligned}$$

$$\mathbf{H} = \sqrt{\frac{\hat{\varepsilon}}{\mu}} E_o \left(\mathbf{a}_z \times \mathbf{a}_E \right) e^{-\gamma z} = \frac{E_o}{\hat{\eta}} \left(\mathbf{a}_z \times \mathbf{a}_E \right) e^{-\alpha z} e^{-j\beta z} \qquad (8\text{-}68)$$

Here, the complex intrinsic impedance $\hat{\eta}$ is defined as

$$\boxed{\hat{\eta} = \sqrt{\frac{\mu}{\hat{\varepsilon}}} = \sqrt{\frac{\mu}{\varepsilon' - j\varepsilon''}}} \qquad [\Omega] \qquad (8\text{-}69)$$

This is the ratio between the complex amplitudes of the electric and magnetic fields; that is, $\hat{\eta} = E_o / H_o$. In view of Eq. (8-69), we note that **E** and **H** are always out of phase in a lossy dielectric.

Example 8-7

Given that $\mathbf{E} = \mathbf{a}_x E_o e^{-0.3z} e^{-j0.5z}$ for a wave propagating in a lossy dielectric of $\varepsilon' = 2\varepsilon_o$ and $\mu = \mu_o$, determine the intrinsic impedance of the medium.

Solution

From the given electric field phasor, we obtain $\alpha = 0.3$ and $\beta = 0.5$

From Eq. (8-63), we obtain

$$\left(\frac{\alpha}{\beta}\right)^2 = \frac{\sqrt{1+(\varepsilon''/\varepsilon')^2}-1}{\sqrt{1+(\varepsilon''/\varepsilon')^2}+1} \tag{8-70}$$

Inserting the values of α and β into Eq. (8-70), the loss tangent is

$$\frac{\varepsilon''}{\varepsilon'} = 1.88 \tag{8-71}$$

Substituting $\varepsilon' = 2\varepsilon_o$ and the loss tangent into Eq. (8-69), we obtain the complex intrinsic impedance as

$$\hat{\eta} = \sqrt{\frac{\mu_o}{2\varepsilon_o}}\frac{1}{\sqrt{1-j(\varepsilon''/\varepsilon')}} = \frac{120\pi}{\sqrt{2}}\frac{1}{\sqrt{1-j1.88}} = 183e^{j0.54} \tag{8-72}$$

We note that **H** lags behind **E** by 0.54 radian in time phase.

Exercise 8.9

For a wave $\mathscr{E} = -\mathbf{a}_y 0.5e^{-0.3z}\sin\left(10^8 t + 2z\right)$, find (a) α, (b) β, (c) υ_p, (d) direction of propagation, and (e) $\mathbf{E}_o$.

Ans. (a) $\alpha = 0.3[\text{Np/m}]$, (b) $\beta = 2[\text{rad/m}]$, (c) $\upsilon_p = 5\times10^7[\text{m/s}]$, (d) −*z*-direction, (e) $\mathbf{E}_o = 0.5e^{j\pi/2}[\text{V/m}]$.

8.2.4.2 Lossy Dielectric of a Low Conductivity

The conduction current is the dominant source of power loss in lossy media. In a dielectric with a finite conductivity and a negligible damping, Maxwell's equations are written as

$$\nabla\times\mathbf{E} = -j\omega\mu\mathbf{H} \tag{8-73a}$$
$$\nabla\times\mathbf{H} = \mathbf{J} + j\omega\varepsilon\mathbf{E} = (\sigma + j\omega\varepsilon)\mathbf{E} \tag{8-73b}$$
$$\nabla\bullet\mathbf{E} = 0 \tag{8-73c}$$
$$\nabla\bullet\mathbf{H} = 0 \tag{8-73d}$$

In a material of a negligibly small damping, ε and μ are real. From the above equations we note that although the given material contains free electrons, $\sigma \neq 0$, it has no net charges, $\rho_v = 0$. The current density phasor **J** shown in Eq. (8-73b) represents the conduction current generated by the electric field of the incident wave. Rewriting Eq. (8-73b), we have

$$\begin{aligned}\nabla\times\mathbf{H} &= j\omega\left(\varepsilon - j\frac{\sigma}{\omega}\right)\mathbf{E} \\ &\equiv j\omega(\varepsilon' - j\varepsilon'')\mathbf{E} \equiv j\omega\hat{\varepsilon}\mathbf{E}\end{aligned} \tag{8-74}$$

Here, the complex permittivity $\hat{\varepsilon}$ is newly defined as

$$\boxed{\hat{\varepsilon} \equiv \varepsilon' - j\varepsilon'' = \varepsilon - j\frac{\sigma}{\omega}} \qquad \text{[F/m]} \tag{8-75}$$

A lossy dielectric with a finite conductivity, infinite in extent, can support a plane wave of the form

$$\boxed{\mathbf{E} = \mathbf{E}_o e^{-\gamma z}} \tag{8-76}$$

Substitution of $\varepsilon' = \varepsilon$ and $\varepsilon'' = \sigma/\omega$ into Eqs. (8-62) and (8-63) leads to the propagation, attenuation, and phase constants in a lossy dielectric of a finite conductivity expressed as

$$\boxed{\gamma \equiv \alpha + j\beta = j\omega\sqrt{\mu\varepsilon}\sqrt{\left(1 - j\frac{\sigma}{\omega\varepsilon}\right)}} \qquad [\text{m}^{-1}] \tag{8-77}$$

$$\boxed{\alpha = \omega\sqrt{\frac{\mu\varepsilon}{2}}\left[\sqrt{1 + \left(\frac{\sigma}{\omega\varepsilon}\right)^2} - 1\right]^{1/2}} \qquad [\text{Np/m}] \tag{8-78a}$$

$$\boxed{\beta = \omega\sqrt{\frac{\mu\varepsilon}{2}}\left[\sqrt{1 + \left(\frac{\sigma}{\omega\varepsilon}\right)^2} + 1\right]^{1/2}} \qquad [\text{rad/m}] \tag{8-78b}$$

Following a similar procedure, we obtain the complex intrinsic impedance from Eq. (8-69) as

$$\boxed{\hat{\eta} = \sqrt{\frac{\mu}{\varepsilon}}\sqrt{\frac{1}{1 - j\sigma/(\omega\varepsilon)}}} \qquad [\Omega] \tag{8-79}$$

The loss tangent is now given by

$$\boxed{\tan\xi = \frac{\varepsilon''}{\varepsilon'} = \frac{\sigma}{\omega\varepsilon}} \tag{8-80}$$

For instance, in a low-loss dielectric with a loss tangent $\tan\xi = \sigma/\omega\varepsilon < 0.1$, we obtain $\alpha/\beta < 0.05$ and a phase angle of $\hat{\eta}$ being less than 0.05[rad].

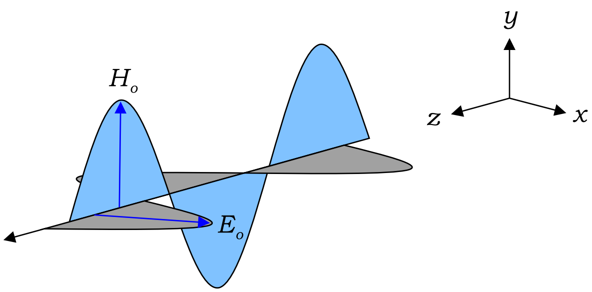


Fig. 8.7 In a low-loss dielectric, **H** always lags behind **E** in time phase.

Example 8-8
Find the attenuation and phase constants in a nonmagnetic material($\mu = \mu_o$) in which the complex intrinsic impedance is $\hat{\eta} = 200e^{j0.5}$ at a frequency $100[\text{MHz}]$.

Solution
The term $1 - j\sigma/(\omega\varepsilon)$ in Eq. (8-79) is written in polar form as

$$1 - j\sigma/(\omega\varepsilon) = \sqrt{1 + (\sigma/\omega\varepsilon)^2}\, e^{-j\xi}$$

Inserting it into Eq. (8-79), the intrinsic impedance is expressed as

$$\hat{\eta} = \sqrt{\frac{\mu_o}{\varepsilon_o\varepsilon_r}} \frac{e^{j\xi/2}}{\left[1 + (\sigma/\omega\varepsilon)^2\right]^{1/4}} = 200e^{j0.5} \tag{8-81}$$

Equating the two exponentials in Eq. (8-81), we obtain $\xi = 1$, and thus $\tan\xi = \tan(1) = 1.56$.

Inserting the loss tangent into Eq. (8-81), we have

$$\frac{377}{\sqrt{\varepsilon_r}} \frac{1}{\left[1 + (1.56)^2\right]^{1/4}} = 200 \quad \rightarrow \quad \varepsilon_r = 1.92$$

From Eq. (8-78a), we obtain

$$\alpha = \omega\sqrt{\mu_o\varepsilon_o}\sqrt{\frac{\varepsilon_r}{2}}\left[\sqrt{1 + \left(\frac{\sigma}{\omega\varepsilon}\right)^2} - 1\right]^{1/2} = \frac{2\pi\times10^8}{3\times10^8}\sqrt{\frac{1.92}{2}}\left[\sqrt{1 + (1.56)^2} - 1\right]^{1/2}$$
$$= 1.90[\text{Np/m}] \tag{8-82}$$

From (8-78b), we obtain

$$\beta = \omega\sqrt{\mu_o\varepsilon_o}\sqrt{\frac{\varepsilon_r}{2}}\left[\sqrt{1+\left(\frac{\sigma}{\omega\varepsilon}\right)^2}+1\right]^{1/2} = \frac{2\pi\times10^8}{3\times10^8}\sqrt{\frac{1.92}{2}}\left[\sqrt{1+(1.56)^2}+1\right]^{1/2}$$
$$= 3.47[\text{rad/m}] \tag{8-83}$$

In view of the ratio $\alpha/\beta = 0.55 >> 0.05$ and the loss tangent $\tan\xi = 1.56$, this is a high-loss material.

Example 8-9

For an electromagnetic wave with $\mathbf{E} = \mathbf{a}_x 300 e^{-0.21z - j(2.2z)}$ [V/m], propagating in a medium of $\hat{\eta} = 220 + j21$ [Ω], find $\langle \mathbf{S} \rangle$ at the $z = 0.5[\text{m}]$ plane.

Solution

Magnetic field phasor is

$$\mathbf{H} = \frac{1}{\hat{\eta}}(\mathbf{a}_z \times \mathbf{E}) = \mathbf{a}_y \frac{300}{220 + j21} e^{-0.21z - j(2.2z)}$$

Time-average Poynting vector is

$$\langle \mathbf{S} \rangle = \frac{1}{2}\text{Re}\left[\mathbf{E}\times\mathbf{H}^*\right] = \mathbf{a}_z \frac{1}{2}\text{Re}\left[\frac{300^2 e^{-0.42z}}{220 - j21}\right] \tag{8-84}$$
$$= \mathbf{a}_z (202.7) e^{-0.42z}$$

At the $z = 0.5[\text{m}]$ plane

$$\langle \mathbf{S} \rangle = \mathbf{a}_z (202.7) e^{-0.42\times0.5} = 164.3\,\mathbf{a}_z[\text{W/m}^2].$$

Exercise 8.10

Write an expression for the ratio α/β in terms of the loss tangent in a low-loss dielectric.

Ans. $\alpha/\beta \cong (1/2)\tan\xi$.

Exercise 8.11

Express, in terms of the loss angle, how much **H** lags behind **E** in time phase.

Ans. Half the loss angle, or $\xi/2$.

Review Questions with Hints

RQ 8.14 Define attenuation and phase constants of the wave in a lossy dielectric. [Eqs.(8-63)(8-78)]
RQ 8.15 Define loss tangent of a lossy medium. [Eqs.(8-64)(8-80)]
RQ 8.16 Under what conditions does the permittivity become complex? [Eqs.(8-57b)(8-73b)]
RQ 8.17 Express intrinsic impedance of a lossy dielectric. [Eqs.(8-69)(8-79)]

8.2.4.3 Good Conductors

Although the conductivity may not be infinite in a good conductor, it is very large such that we may assume $\sigma/\omega\varepsilon >> 1$ in Eqs. (8-77)-(8-80). Under this condition, it is evident from Eq. (8-73b) that the displacement current can be neglected in the conductor in comparison with the conduction current. In a good conductor, the propagation constant γ in Eq. (8-77) simply reduces to

$$\gamma \cong j\omega\sqrt{\mu\varepsilon}\sqrt{-j\frac{\sigma}{\omega\varepsilon}} = \frac{1+j}{\sqrt{2}}\sqrt{\omega\mu\sigma} \equiv \alpha + j\beta \tag{8-85}$$

In the above equation, we used the following relation:

$$\sqrt{-j} = \left[e^{-j\pi/2}\right]^{1/2} = e^{-j\pi/4} = \frac{1-j}{\sqrt{2}} \tag{8-86}$$

As can be seen from Eq. (8-85), the attenuation and phase constants are equal in magnitude in a good conductor:

$$\boxed{\alpha = \beta = \sqrt{\pi f \mu\sigma}} \tag{8-87}$$

where $\omega = 2\pi f$. In a good conductor, both α and β increase with $\sqrt{f}$ and $\sqrt{\sigma}$.

In a good conductor, the phasor and instantaneous forms of the electric field are given as

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_o e^{-\alpha z} e^{-j\beta z} \\ &= (E_o\,\mathbf{a}_E)e^{-z\sqrt{\pi f\mu\sigma}}e^{-jz\sqrt{\pi f\mu\sigma}}\end{aligned} \tag{8-88}$$

$$\mathscr{E} = (E_o\,\mathbf{a}_E)e^{-z\sqrt{\pi f\mu\sigma}}\cos\left(\omega t - z\sqrt{\pi f\mu\sigma}\right) \tag{8-89}$$

The amplitude of the electric field intensity varies as $e^{-\alpha z}$ and, therefore, it will be attenuated by a factor of $e^{-1} = 0.368$ at a distance $1/\sqrt{\pi f\mu\sigma}$ along the z-axis. This distance is designated by δ and called the depth of penetration or skin depth of the conductor. Namely

$$\boxed{\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} = \frac{1}{\alpha} = \frac{1}{\beta}} \qquad \text{[m]} \tag{8-90}$$

The skin depth is measured in meters[m]. It tells us how deep an electromagnetic wave can penetrate into the conductor. At microwave frequencies, the skin depth is so short that the electric field and the induced current may be considered to be confined within a very thin layer of a thickness δ on the surface of the conductor. For example, silver-plated brass waveguides can be used in place of those made of solid silver with little degradation in performance, but with much reduced material cost.

Upon using the relation $\beta = 2\pi/\lambda$, the skin depth is simply written as

$$\delta = \frac{\lambda}{2\pi} \tag{8-91}$$

where λ is the wavelength measured in the conductor. We see from Eq. (8-91) that only a fraction, less than one sixth, of a cycle of the wave is packed within a skin depth. To show an electromagnetic wave in a good conductor occupying the region $z \geq 0$, the electric field intensity $|\mathscr{E}| = e^{-z/\delta}\cos(\omega t - z/\delta)$ is plotted as a function of z at times $t = 0$ and $t = \pi/(2\omega)$ in Fig. 8.8.

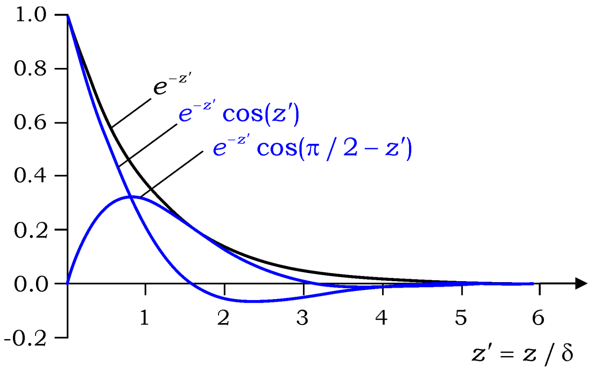


Fig. 8.8 $|\mathscr{E}| = e^{-z/\delta}\cos(\omega t - z/\delta)$ at times $t = 0$ and $t = \pi/(2\omega)$ in a good conductor.

The phase velocity of a wave propagating in a good conductor is, from Eq. (8-90),

$$\upsilon_p = \frac{\omega}{\beta} = 2\sqrt{\frac{\pi f}{\mu\sigma}} \qquad \text{[m/s]} \tag{8-92}$$

The phase velocity depends on frequency in a good conductor, and is generally very small compared with that in free space. For instance, the phase velocity in copper is $\upsilon_p = 1.3 \times 10^4$[m/s] at 1[GHz].

Example 8-10

Express the skin depth of copper, having a conductivity $\sigma = 5.8\times10^{7}[\text{S/m}]$, as a function of frequency.

Solution

From Eq. (8-90), with $\mu \approx \mu_o$, we write

$$\delta = \frac{1}{\sqrt{\pi f \mu_o \sigma}} = \frac{1}{\sqrt{\pi f \times 4\pi\times10^{-7}\times5.8\times10^{7}}}$$

Skin depth of copper is therefore

$$\delta = \frac{0.066}{\sqrt{f}} \quad [\text{m}]$$

For instance, at a frequency $f = 1[\text{GHz}]$, the skin depth of copper is only $2.1[\mu\text{m}]$.

A good conductor is characterized by a property $\sigma/\omega\varepsilon >> 1$. The intrinsic impedance of a good conductor is therefore, from Eq. (8-79),

$$\hat{\eta} \cong \sqrt{\frac{j\omega\mu}{\sigma}} = (1+j)\sqrt{\frac{\pi f\mu}{\sigma}}$$

The intrinsic impedance of a good conductor is expressed in terms of the conductivity and skin depth as

$$\boxed{\hat{\eta} = \frac{(1+j)}{\sigma\delta}} \qquad [\Omega] \qquad (8\text{-}93)$$

Equation (8-93) implies that the magnetic field always lags behind the electric field by 45^{o} in time phase in a good conductor.

Let us now turn our attention to the power density carried by an electromagnetic wave propagating in a good conductor. If an electromagnetic plane wave is polarized in the x-direction, and propagates in the $+z$-direction in a good conductor, the electric and magnetic field phasors are written as

$$\boxed{\mathbf{E} = \mathbf{a}_x E_o e^{-(1+j)z/\delta}} \qquad (8\text{-}94a)$$

$$\boxed{\mathbf{H} = \frac{1}{\hat{\eta}}(\mathbf{a}_z\times\mathbf{E}) = \mathbf{a}_y \frac{\sigma\delta}{1+j} E_o e^{-(1+j)z/\delta}} \qquad (8\text{-}94b)$$

The time-average Poynting vector is

$$\langle\mathbf{S}\rangle = \frac{1}{2}\text{Re}\left[\mathbf{E}\times\mathbf{H}^*\right] = \frac{1}{2}\text{Re}\left[\mathbf{a}_z E_o^2 \frac{\sigma\delta(1+j)}{2} e^{-2z/\delta}\right]$$

Thus, in a good conductor in the region $z \geq 0$, we have

$$\boxed{\langle \mathbf{S} \rangle = \mathbf{a}_z \frac{\sigma\delta}{4} E_o^2 e^{-2z/\delta}} \qquad [\mathrm{W/m^2}] \qquad (8\text{-}95)$$

Here, E_o is the amplitude of the electric field at the surface of the conductor defined at $z = 0$, σ is the conductivity, and δ is the skin depth of the conductor. The power density decreases rapidly as the wave propagates into the conductor. At a distance of δ from the surface, $\langle \mathbf{S} \rangle$ is only a fraction, $e^{-2} = 0.135$, of its initial value specified at the surface.

The skin depth enables us to express the power loss in a good conductor in terms of an ac resistance. Consider the case as shown in Fig. 8.9, in which an electromagnetic plane wave, polarized in the *x*-direction, propagates in the +*z*-direction in a good conductor of a cross section $w \times \ell$ [m^2]. The electric field of the wave induces a time-varying current density in the conductor, which is expressed in phasor form as

$$\mathbf{J} = \mathbf{a}_x \sigma E_o e^{-(1+j)z/\delta} = \mathbf{a}_x \mathrm{J}_o e^{-(1+j)z/\delta} \qquad (8\text{-}96)$$

The total time-average power dissipated in the conductor of a conductivity σ and a volume $w \times \ell \times d$ [m^3] is computed as follows:

$$\begin{aligned}\langle P \rangle = \int_V \frac{1}{2} \mathrm{Re}\left[\mathbf{J} \bullet \mathbf{E}^*\right] dv &= \frac{1}{2\sigma} w\ell \,\mathrm{Re}\left[\int_{z=0}^{z=d} \mathbf{J} \bullet \mathbf{J}^* dz\right] \\ &= \frac{1}{2\sigma} w\ell \mathrm{J}_o^2 \int_{z=0}^{z=d} e^{-2z/\delta} dz\end{aligned}$$

Under the condition of a small skin depth, $\delta << d$, the total time-average power loss in the conductor is expressed as

$$\boxed{\langle P \rangle = \frac{1}{4\sigma} w\ell\, \delta \mathrm{J}_o^2} \qquad [\mathrm{W}] \qquad (8\text{-}97)$$

where $w\ell$ represents the cross sectional area of the conductor, and δ is the skin depth. It should be noted that J_o is the amplitude of the current density on the surface of the conductor, not a surface current density. Let us now calculate, from Eq. (8-96), the total *ac* current flowing in the conductor. The phasor of the total *ac* current, I_T, is

$$\begin{aligned}\mathrm{I}_T = \int_S \mathbf{J} \bullet d\mathbf{s} &= w\mathrm{J}_o \int_{z=0}^{z=d} e^{-(1+j)z/\delta} dz \\ &= \frac{w\delta \mathrm{J}_o}{1+j}\end{aligned} \qquad [\mathrm{A}] \qquad (8\text{-}98)$$

Now let us assume that the total current I_T flows uniformly in a thin slab of a cross section $w \times \delta [\mathrm{m}^2]$ and a length $\ell [\mathrm{m}]$ for the reason that will become evident shortly. We next express the total time-average power dissipated in the conductor, Eq. (8-97), in terms of the total current I_T as

$$\boxed{\langle P \rangle = \frac{1}{2} R_{ac} \operatorname{Re}\left[I_T I_T^* \right]} \qquad [\mathrm{W}] \qquad (8\text{-}99)$$

Comparing Eq. (8-97) and Eq. (8-99), with the aid of Eq. (8-98), we define the ac resistance R_{ac} as follows:

$$\boxed{R_{ac} = \frac{\ell}{\sigma \delta w}} \qquad [\Omega] \qquad (8\text{-}100)$$

The total power dissipated in a conductor with a dc-resistance $R_{dc} = \ell / \sigma w d \ [\Omega]$ is equal to the total power dissipated in a thin slab of a cross section $w \times \delta [\mathrm{m}^2]$ and a length $\ell [\mathrm{m}]$, having an *ac*-resistance $R_{ac} = \ell / \sigma \delta w \ [\Omega]$ that is experienced by the total current uniformly distributed in the slab. Again, the total time-average power dissipated in a conductor can be expressed in terms of the current density J_o specified on the surface of the conductor, as shown in Eq. (8-97), or in terms of the total current I_T flowing in the conductor, as shown in Eq. (8-99).

The total current in the conductor in real instantaneous form is, from Eq. (8-98),

$$i_T = \operatorname{Re}\left[I_T e^{j\omega t} \right] = w\delta \frac{J_o}{\sqrt{2}} \cos\left(\omega t - \frac{\pi}{4} \right) \qquad (8\text{-}101)$$

Substituting $J_o = \sigma E_o$ into Eq. (8-101), we obtain the relation between the total instantaneous current i_T and the electric field intensity on the surface of the conductor, E_o.

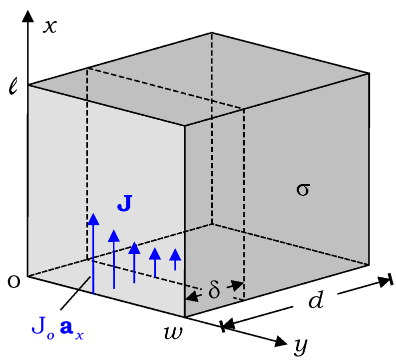


Fig. 8.9 Penetration of an electromagnetic wave into a good conductor.

Example 8-11

Referring to Fig. 8.9, find the total power transmitted into the conductor through the entrance surface of the cross section $w \times \ell\,[\text{m}^2]$.

Solution

Time-average power density at the $z = 0$ plane is, from Eq. (8-95),

$$\langle \mathbf{S} \rangle = \mathbf{a}_z \frac{\sigma\delta}{4} E_o^2$$

Total power through the cross section is

$$\langle P \rangle = \int_{x=0}^{\ell} \int_{y=0}^{w} \langle \mathbf{S} \rangle \cdot d\mathbf{s} = \frac{1}{4}\sigma\delta E_o^2 w\ell = \frac{1}{4\sigma} w\delta\ell J_o^2 \tag{8-102}$$

The direction of $d\mathbf{s}$ is along the direction of $\langle \mathbf{S} \rangle$, not outward away from the volume. The transmitted power given by Eq. (8-102) is equal to the total power dissipated in the conductor given by Eq. (8-97), conforming with the law of conservation of energy.

Exercise 8.12

What is the skin depth of sea water having $\sigma = 4\,[\text{S/m}]$ at a frequency 1[KHz] ?

Ans. $8.0\,[\text{m}]$.

Exercise 8.13

At what distance into a good conductor is the time-average power density of the incident wave reduced to 1% of its initial value at the surface?

Ans. $2.3 \times \delta$.

Exercise 8.14

What is the *ac*-resistance of a copper wire of radius 1[mm] and length 1[Km] at a frequency 1[MHz]?

Ans. $R_{ac} = 41.5\,[\Omega]$.

Review Questions with Hints

RQ 8.18 Define the attenuation and phase constants of a plane wave propagating in a good conductor. [Eq.(8-87)]

RQ 8.19 What is the skin depth of a conductor? [Eq.(8-90)]

RQ 8.20 Express the intrinsic impedance of a good conductor. [Eq.(8-93)]

RQ 8.21 Express the time-average power density of an electromagnetic wave in a good conductor. [Eq.(8-95)]

8.3 Plane Waves at an Interface

When an electromagnetic wave impinges on an interface between two different media, a portion of the wave is reflected from the interface, while the remainder is transmitted through the interface. The electric and magnetic fields of these three waves should satisfy the boundary conditions at the interface, which are imposed by Maxwell's equations. Since the phase of the incident wave generally varies from point to point in the interface (see the $z = 0$ plane in Fig. 8.1, for instance), the spatial variations of the phases of the reflected and transmitted waves should be the same as that of the incident wave, in the interface, in order to satisfy the boundary conditions; otherwise, the three waves cannot be related to each other at the interface. As we will see later, this condition of an identical phase variation leads to the law of reflection and the law of refraction at the interface.

In this section, we first examine the uniform plane wave normally incident on a planar interface between two different media, and learn about the standing wave. Next, we extend our discussion to the uniform plane wave obliquely incident on the interface, and study the laws of reflection and refraction, followed by Fresnel's equations of reflection and transmission coefficients. We compute the reflected and transmitted powers, and discuss the reflectance and transmittance at the interface.

8.3.1 Normal Incidence of a Plane Wave

In this subsection, we limit our discussion to a uniform plane wave normally incident on an infinitely large planar interface between two lossless media, as depicted in Fig. 8.10. We assume that the wave is linearly polarized in the x-direction, and propagates in the $+z$-direction in medium 1 in the region $z \le 0$, before it impinges on the interface at $z = 0$. To start with, we write the electric and magnetic field phasors of the incident wave as

$$\mathbf{E}^i = \mathbf{a}_x E_o^i e^{-j\beta_1 z} \tag{8-103a}$$

$$\mathbf{H}^i = \mathbf{a}_y \frac{E_o^i}{\eta_1} e^{-j\beta_1 z} \tag{8-103b}$$

where i stands for the incident wave, and 1 stands for medium 1. The phase constant β_1 and the intrinsic impedance η_1 are both real in the lossless medium 1. We assume the amplitude E_o^i to be real without loss of generality. In this case, the magnetic field vector is directed along the $+y$-axis in accordance with Eq. (8-28).

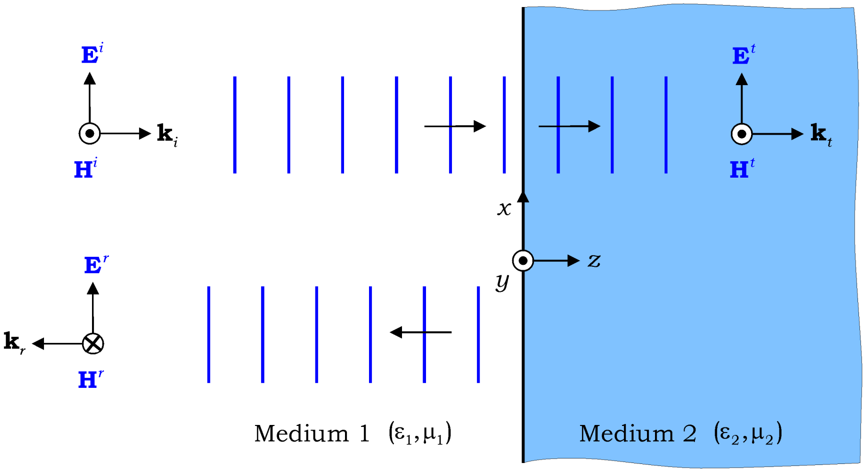


Fig. 8.10 Normal incidence of a plane wave on a planar interface.

The electric field intensity $\mathbf{E}^i$ of the incident wave is constant in the interface at $z = 0$, as can be seen from Eq. (8-103). Under this condition, the electric field intensities of the reflected and transmitted waves should also be constant in the interface. These are prerequisites for the boundary condition, which requires that the tangential component of $\mathbf{E}$ is continuous across the interface. The preconditions can be met only if the reflected and transmitted waves are uniform plane waves propagating in either the $+z$ or the $-z$ direction, and their wavefronts are parallel to the $z = 0$ plane.

The reflected wave propagates in the $-z$-direction, and its electric and magnetic field phasors are expressed as

$$\mathbf{E}^r = \mathbf{a}_x E_o^r e^{j\beta_1 z} \tag{8-104a}$$

$$\mathbf{H}^r = -\mathbf{a}_y \frac{E_o^r}{\eta_1} e^{j\beta_1 z} \tag{8-104b}$$

where r stands for the reflected wave. It should be noted that $\mathbf{E}^r$ is assumed to be directed in the $+x$-direction, instead of the $-x$-direction, for simplicity. Under this condition, $\mathbf{H}^r$ is directed in the $-y$-direction in accordance with Eq. (8-28). Note that the phase constant and the intrinsic impedance shown in Eq. (8-104) are the same as those in Eq. (8-103) because the two waves are in the same medium.

The transmitted wave propagates in the $+z$-direction, and its electric and magnetic field phasors are expressed as

$$\mathbf{E}^t = \mathbf{a}_x E_o^t e^{-j\beta_2 z} \tag{8-105a}$$

$$\mathbf{H}^t = \mathbf{a}_y \frac{E_o^t}{\eta_2} e^{-j\beta_2 z} \tag{8-105b}$$

where t stands for the transmitted wave. In medium 2, the phase constant β_2 and the intrinsic impedance η_2 are, in general, not the same as those in medium 1.

We are now ready to apply the boundary conditions to the electric and magnetic field intensities expressed by Eqs. (8-103)-(8-105), and determine the amplitudes E_o^r and E_o^t in terms of E_o^i. The tangential components of **E** and **H** should be continuous across the interface at $z = 0$ such that

$$E_o^i + E_o^r = E_o^t \tag{8-106a}$$

$$\frac{E_o^i}{\eta_1} - \frac{E_o^r}{\eta_1} = \frac{E_o^t}{\eta_2} \tag{8-106b}$$

The left-hand side of Eq. (8-106a) is the sum of the two electric field intensities in medium 1, observed on the interface, while the right-hand side is the electric field intensity in medium 2, observed on the interface. Eq. (8-106b) obviously comes from the boundary condition for the magnetic field intensity at the interface.

The reflection coefficient Γ and the transmission coefficient τ are readily obtained from Eqs. (8-106a) and (8-106b) as

$$\boxed{\Gamma \equiv \frac{E_o^r}{E_o^i} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1}} \tag{8-107a}$$

$$\boxed{\tau \equiv \frac{E_o^t}{E_o^i} = \frac{2\eta_2}{\eta_2 + \eta_1}} \tag{8-107b}$$

It should be noted that Eq. (8-107) is true only for the normal incidence of a uniform plane wave on an infinitely large planar interface. The reflection coefficient Γ becomes negative for $\eta_1 > \eta_2$, implying that the direction of the electric field intensity is reversed after a reflection from the interface. In contrast, the transmission coefficient τ is always positive. It can be shown that Eq. (8-107) is also valid for an interface between two lossy media with complex intrinsic impedances; the reflection and transmission coefficients are generally complex numbers.

Equation (8-107) leads to

$$\boxed{1 + \Gamma = \tau} \tag{8-108}$$

It is important to remember that this is true only for the normal incidence.

Example 8-12

For a uniform plane wave normally incident on an interface between two media of η_1 and η_2, derive $|\Gamma|^2 + (\eta_1 / \eta_2)|\tau|^2 = 1$ from the law of conservation of energy.

Solution

Time-average power densities of the incident, reflected, and transmitted waves are, from Eqs. (8-103)-(8-105),

$$\left\langle \mathbf{S}^i \right\rangle = \frac{1}{2}\mathrm{Re}\left[\mathbf{E}^i \times \mathbf{H}^{i*}\right] = \mathbf{a}_z \frac{1}{2\eta_1}\left|E_o^i\right|^2 \tag{8-109a}$$

$$\left\langle \mathbf{S}^r \right\rangle = -\mathbf{a}_z \frac{1}{2\eta_1}\left|E_o^r\right|^2 = -\mathbf{a}_z \frac{1}{2\eta_1}\left|\Gamma\right|^2\left|E_o^i\right|^2 \tag{8-109b}$$

$$\left\langle \mathbf{S}^t \right\rangle = \mathbf{a}_z \frac{1}{2\eta_2}\left|E_o^t\right|^2 = \mathbf{a}_z \frac{1}{2\eta_2}\left|\tau\right|^2\left|E_o^i\right|^2 \tag{8-109c}$$

From the law of conservation of energy, we write

$$\left|\left\langle \mathbf{S}^i \right\rangle\right| = \left|\left\langle \mathbf{S}^r \right\rangle\right| + \left|\left\langle \mathbf{S}^t \right\rangle\right| \tag{8-110}$$

Substituting Eq. (8-109) into Eq. (8-110), we have

$$\boxed{\left|\Gamma\right|^2 + \frac{\eta_1}{\eta_2}\left|\tau\right|^2 = 1.} \tag{8-111}$$

Exercise 8.15

For a wave $\mathbf{E}^i = \mathbf{a}_x 10 e^{-j\beta_1 z}$ with $f = 2\,[\mathrm{GHz}]$, propagating in free space $(z < 0)$, impinging on a dielectric $(z \geq 0)$ with $\varepsilon_2 = 2\varepsilon_o$ and $\mu_2 = \mu_o$, find $\mathbf{E}^r$ and $\mathbf{E}^t$.

Ans. $\mathbf{E}^r = -\mathbf{a}_x 1.72 e^{-j41.9z}$, and $\mathbf{E}^t = \mathbf{a}_x 8.28 e^{-j59.2z}$.

Exercise 8.16

A light of $\lambda = 500\,[\mathrm{nm}]$ is normally incident on a glass plate of $n = 1.5$. What portion of the incident power is reflected from the first surface of the glass plate?

Ans. 4%.

8.3.1.1 Standing Wave Ratio

The total electric field intensity in medium 1 is equal to the sum of those of the incident and reflected waves, which propagate in the opposite directions. That is,

$$\begin{aligned}\mathbf{E}_1 = \mathbf{E}^i + \mathbf{E}^r &= \mathbf{a}_x E_o^i e^{-j\beta_1 z}\left[1 + \Gamma e^{j2\beta_1 z}\right] \\ &\equiv \mathbf{a}_x \hat{E}_1\end{aligned} \tag{8-112}$$

where Γ is the reflection coefficient as given in Eq. (8-107a). As mentioned earlier, if one of the two adjoining media is a lossy material, the reflection coefficient is given by a complex number such as

$$\Gamma = \left|\Gamma\right| e^{j\phi} \tag{8-113}$$

where ϕ is the phase angle. Inserting Eq. (8-113) into Eq. (8-112), the magnitude of the electric field vector phasor $\mathbf{E}_1$ in medium 1 is expressed as

$$\hat{E}_1 = E_o^i e^{-j\beta_1 z}\left[1 + |\Gamma| e^{j(2\beta_1 z + \phi)}\right] \tag{8-114}$$

The complex amplitude $\hat{E}_1$ is a complex number, whose magnitude represents the amplitude of the total electric field intensity, oscillating with an angular frequency ω in medium 1, and whose phase angle represents the phase delay of the time-harmonic electric field.

The maximum amplitude is observed at a point on the z-axis where the two terms in bracket on the right-hand side of Eq. (8-114) are in phase. Namely,

$$\boxed{\left|\hat{E}_1\right|_{\max} = E_o^i\left(1 + |\Gamma|\right)} \tag{8-115a}$$

$$\boxed{z_{\max} = -\frac{1}{2\beta_1}(\phi + 2\pi n)} \qquad (n = 0, \pm 1, \pm 2, \ldots) \tag{8-115b}$$

The maximum is closely related to the reflection coefficient in such a way that the maximum amplitude $\left|\hat{E}_1\right|_{\max}$ depends on the magnitude of Γ, whereas the location of the maximum, $z_{\max}$, depends on the phase angle of Γ. On the other hand, the minimum amplitude is observed at a point on the z-axis where the two terms in bracket are out of phase by 180^o. Namely,

$$\boxed{\left|\hat{E}_1\right|_{\min} = E_o^i\left(1 - |\Gamma|\right)} \tag{8-116a}$$

$$\boxed{z_{\min} = -\frac{1}{2\beta_1}(\phi + 2\pi n + \pi)} \qquad (n = 0, \pm 1, \pm 2, \ldots) \tag{8-116b}$$

We see from Eqs. (8-115b) and (8-116b) that two adjacent maxima or two adjacent minima are separated by a distance π/β_1 along the z-axis, which is a half wavelength of the incident or the reflected wave in medium 1.

Let us rewrite Eq. (8-114) as

$$\begin{aligned}\hat{E}_1 &= E_o^i e^{-j\beta_1 z}\left[1 - |\Gamma| + |\Gamma| + |\Gamma| e^{j(2\beta_1 z + \phi)}\right] \\ &= E_o^i e^{-j\beta_1 z}\left(1 - |\Gamma|\right) + E_o^i |\Gamma| e^{j\phi/2}\left[2\cos\left(\beta_1 z + \phi/2\right)\right]\end{aligned} \tag{8-117}$$

The instantaneous form of the total electric field intensity in medium 1 is, from (8-117),

$$\mathscr{E}_1 = E_o^i\left(1 - |\Gamma|\right)\cos\left(\omega t - \beta_1 z\right) + 2E_o^i |\Gamma| \cos\left(\beta_1 z + \phi/2\right)\cos\left(\omega t + \phi/2\right) \tag{8-118}$$

The first term on the right-hand side of Eq. (8-118) is a traveling wave, having an amplitude $E_o^i(1-|\Gamma|)$, whereas the second term is a standing wave, oscillating with an amplitude $2E_o^i|\Gamma|\cos(\beta_1 z+\phi/2)$ at an angular frequency ω. Although the standing wave comprises of two plane waves of equal amplitudes, it does not travel through the medium, because of the opposite directions of travel of the two component waves. From Eq. (8-118), we see that when a wave of an amplitude E_o^i impinges normally on the interface, a portion of the wave, $|\Gamma|E_o^i$, is reflected from the interface, and propagates in the direction opposite to that of the incident wave. The reflected wave interferes with the incident wave and produces a standing wave in medium 1. The other portion of the incident wave, $(1-|\Gamma|)E_o^i$, just travels in the $+z$-direction, without being interrupted. We note that the amplitude of the traveling wave is constant as a function of z in medium 1, whereas the amplitude of the standing wave varies sinusoidally with z, because of the term $\cos(\beta_1 z+\phi/2)$ in Eq. (8-118).

For the case of $|\Gamma|=0.5$ and $\phi=0$, the electric field intensities of the traveling and the standing waves given in Eq. (8-118) are both plotted in Fig. 8.11 as a function of $\beta_1 z$, at times $\omega t=0,\ \pi/4$, and π, Blue lines represent the traveling wave, and black lines represent the standing wave.

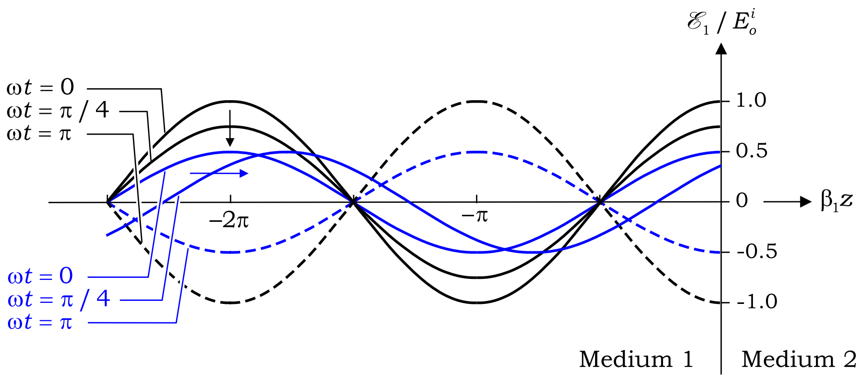


Fig. 8.11 Traveling wave(blue line) and standing wave(black line) in medium 1.

The total electric field intensity given in Eq. (8-118) is plotted in Fig. 8.12 as a function of $\beta_1 z$ at times $\omega t=0, \pi/4, \pi/2, 3\pi/4,$ and π, by assuming $|\Gamma|=0.5$ and $\phi=0$. Fig. 8.12 shows that the electric field intensity curve moves to the right with a decreasing amplitude for $0<\omega t<\pi/2$, and with an increasing amplitude for $\pi/2<\omega t<\pi$, and so on. The two blue lines represent the envelope of the electric field intensities such that the electric field intensity oscillates between the two blue lines, with an angular frequency ω, at a point on the z-axis. The location of the peak depends on the phase angle of the reflection coefficient as shown in Eq. (8-115b).

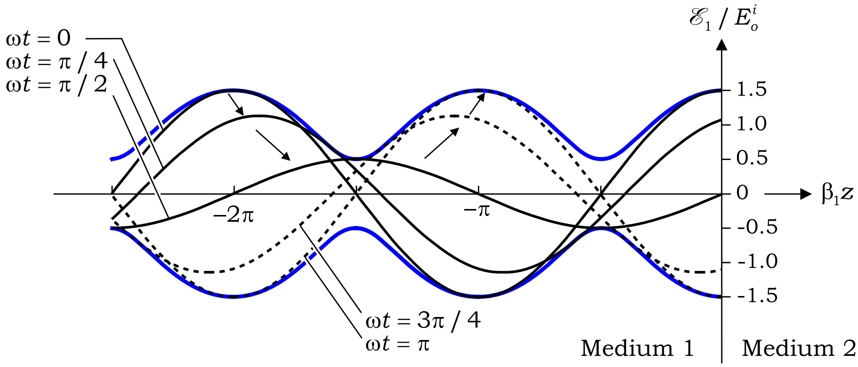


Fig. 8.12 Total electric field in medium 1.

We define the standing wave ratio S as the ratio of the maximum to the minimum amplitude of the total electric field intensity in medium 1. That is,

$$\boxed{S = \frac{\left|\hat{E}_1\right|_{\max}}{\left|\hat{E}_1\right|_{\min}} = \frac{1+|\Gamma|}{1-|\Gamma|}} \tag{8-119}$$

This is a dimensionless quantity. While $|\Gamma|$ ranges from 0 to 1, the standing wave ratio S ranges from 1 to ∞. It is customary to express the standing wave ratio on a logarithmic scale, in decibels; that is, $20\log_{10} S$ [dB]. For example, $S = 4$ corresponds to a standing wave ratio of $20\log_{10} 4 = 12.04[\text{dB}]$. The standing wave ratio S should not be confused with the Poynting vector **S**, or the unit [S] representing the siemens.

Example 8-13

A standing wave of $S = 4$ is formed in free space $(z < 0)$. The first maximum is observed at a distance $0.2[\text{m}]$ from the interface at $z = 0$, and two adjacent maxima are found to be separated by $0.5[\text{m}]$. Determine η_2 of the material in the region $z \geq 0$.

Solution

The distance between two adjacent maxima is equal to a half wavelength:

$$\lambda = 2 \times 0.5[\text{m}], \text{ and thus } \beta_1 = \frac{2\pi}{\lambda} = 2\pi. \tag{8-120}$$

The first maximum corresponds to $n = 0$ in Eq. (8-115b):

$$z_{\max} = -\frac{1}{2\beta_1}\phi = -0.2[m] \tag{8-121}$$

Inserting Eq. (8-120) into Eq. (8-121), we get

$$\phi = 0.8\pi \tag{8-122}$$

From Eq. (8-119), we get

$$|\Gamma| = \frac{S-1}{S+1} = \frac{4-1}{4+1} = 0.6 \tag{8-123}$$

Combining Eq. (8-122) and Eq. (8-123), the reflection coefficient is

$$\Gamma = 0.6e^{j0.8\pi} \tag{8-124}$$

Rewriting Eq. (8-107a), we have

$$\frac{\eta_2}{\eta_1} = \frac{1+\Gamma}{1-\Gamma} \tag{8-125}$$

Substituting Eq. (8-124) and $\eta_1 = \eta_o = 377[\Omega]$ into Eq. (8-125), we get the intrinsic impedance of medium 2 as

$$\eta_2 = 377\frac{1+0.6e^{j0.8\pi}}{1-0.6e^{j0.8\pi}} = 154e^{j0.83}[\Omega].$$

Example 8-14

A uniform plane wave, $\mathscr{E}^i = \mathbf{a}_x 20\cos(3\times10^9 t - 17.3z)$, travels in a lossless dielectric $(z<0)$, and impinges normally on the surface of a lossy dielectric $(z \ge 0)$, for which $\varepsilon_r = 4$ and $\sigma = 0.5[\text{S/m}]$. Find Γ, S, and the distance of the first maximum from the interface at $z = 0$.

Solution

For the incident wave in the region $z<0$, we have

$$\omega = 3\times10^9\,[\text{rad/s}]$$

$$\beta_1 = \omega\sqrt{\mu_o\varepsilon_o\varepsilon_r} = 17.3, \text{ and thus } \sqrt{\varepsilon_r} = 1.73$$

$$\eta_1 = \sqrt{\frac{\mu_o}{\varepsilon_o\varepsilon_r}} = \frac{377}{1.73} = 217.92$$

In the region $z \ge 0$, the intrinsic impedance is, from Eq. (8-79),

$$\eta_2 = \sqrt{\frac{\mu_o}{\varepsilon}}\sqrt{\frac{1}{1-j\sigma/(\omega\varepsilon)}} = \frac{377}{2}\sqrt{\frac{1}{1-j0.5/\left(3\times10^9\times4\times8.854\times10^{-12}\right)}}$$
$$= 188.50\sqrt{\frac{1}{1-j4.71}} = 85.90e^{j0.68}$$

The reflection coefficient is, from Eq. (8-107a),

$$\Gamma = \frac{\eta_2-\eta_1}{\eta_2+\eta_1} = \frac{85.90\cos(0.68)-217.92+j85.90\sin(0.68)}{85.90\cos(0.68)+217.92+j85.90\sin(0.68)}$$
$$= \frac{-151.13+j54.01}{284.71+j54.01} = \frac{160.49e^{j2.80}}{289.79e^{j0.19}} = 0.55e^{j2.61}$$

The standing wave ratio is, from Eq. (8-119),

$$S = \frac{1+|\Gamma|}{1-|\Gamma|} = \frac{1+0.55}{1-0.55} = 3.44$$

Substituting $\phi = 2.61$, $\beta_1 = 17.3$, and $n = 0$ into Eq. (8-115b), we get

$$z_{\max} = -\frac{\phi}{2\beta_1} = -\frac{2.61}{2\times 17.3} = -7.5[\text{cm}].$$

Exercise 8.17

If the incident wave $\mathbf{E}^i$ is delayed by $0.5[\text{rad}]$ such as $\mathbf{E}^i e^{-j0.5}$, are there any changes in Eqs. (8-115b), (8-116b), and (8-119)?

Ans. No.

8.3.1.2 Interface Involving a Perfect Conductor

The perfect conductor is one with an infinite conductivity. The intrinsic impedance of a perfect conductor is zero, as is evident from Eq. (8-93). Let us consider an interface formed by a lossless dielectric in the region $z < 0$ and a perfect conductor in the region $z \geq 0$. We see that a substitution of $\eta_2 = 0$ into Eq. (8-107) results in $\Gamma = -1$, which means that the incident wave is totally reflected by the perfect conductor, and the reflected wave is 180^o out of phase with respect to the incident wave.

If the incident wave is a uniform plane wave linearly polarized in the x-direction, propagating in the $+z$-direction in medium 1, the expressions for the incident and reflected waves are the same as Eq. (8-103) and Eq. (8-104), respectively. With the aid of $\Gamma = E_o^r / E_o^i = -1$ given on the perfect conductor, the total electric and magnetic field intensities are expressed in the dielectric(medium 1) as

$$\begin{aligned}\mathbf{E}_1 = \mathbf{E}^i + \mathbf{E}^r &= \mathbf{a}_x E_o^i \left[e^{-j\beta_1 z} - e^{j\beta_1 z}\right] \\ &= -\mathbf{a}_x j2E_o^i \sin(\beta_1 z)\end{aligned} \tag{8-126a}$$

$$\begin{aligned}\mathbf{H}_1 = \mathbf{H}^i + \mathbf{H}^r &= \mathbf{a}_y \frac{E_o^i}{\eta_1}\left[e^{-j\beta_1 z} + e^{j\beta_1 z}\right] \\ &= \mathbf{a}_y \frac{2E_o^i}{\eta_1}\cos(\beta_1 z)\end{aligned} \tag{8-126b}$$

The instantaneous form of the total electric and magnetic field intensities in medium 1 is

$$\begin{aligned}\mathscr{E} &= \mathbf{a}_x 2E_o^i \sin(\beta_1 z)\cos(\omega t - \pi/2) \\ &= \mathbf{a}_x 2E_o^i \sin(\beta_1 z)\sin(\omega t)\end{aligned} \tag{8-127a}$$

$$\mathscr{H} = \mathbf{a}_y \frac{2E_o^i}{\eta_1}\cos(\beta_1 z)\cos(\omega t) \tag{8-127b}$$

We see from Eq. (8-127) that the incident and reflected waves form a complete standing wave, containing no traveling wave component. The space coordinate z is completely separated from the time variable t in Eq. (8-127). The standing wave is not a traveling wave and thus cannot deliver power to the load. No transfer of power can be verified by inserting Eq. (8-126) into Eq. (8-39), in which $\mathbf{E}_1 \times \mathbf{H}_1^*$ becomes imaginary, resulting in $\langle \mathbf{S} \rangle = 0$, or simply by inserting $\eta_2 = 0$ into Eq. (8-39) to obtain $\tau = 0$.

The total $\mathscr{E}$ and $\mathscr{H}$ given in Eq. (8-127) are plotted in Fig. 8.13 as a function of $\beta_1 z$ at five times. It is apparent from Fig. 8.13 that the total $\mathscr{E}$ is zero on the surface of the perfect conductor at all times to satisfy the boundary condition, which requires that the tangential component of $\mathscr{E}$ should vanish on the surface of the perfect conductor. The point of zero amplitude, called a node, occurs periodically along the z-axis with a period of a half wavelength of the incident wave. The node of $\mathscr{H}$ occurs with the same spatial period as that of $\mathscr{E}$, but is sifted by a distance $\lambda / 4$ with respect to that of $\mathscr{E}$. As we can see in Fig. 8.13, the first node of $\mathscr{H}$ occurs at a distance $\lambda / 4$ from the surface of the conductor.

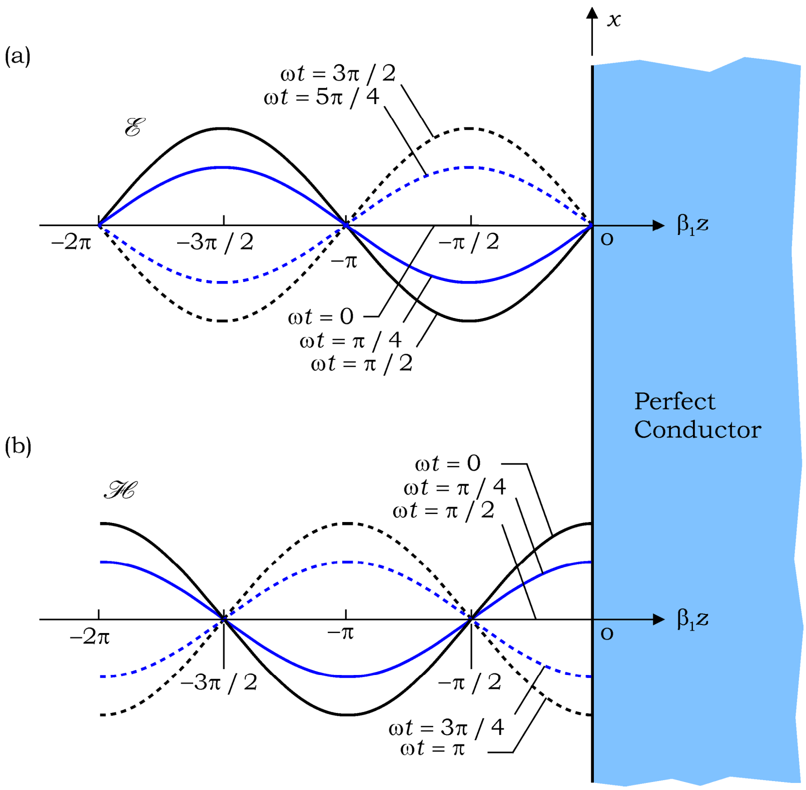


Fig. 8.13 Electric and magnetic field intensities of a complete standing wave caused by a perfectly conducting surface.

Example 8-15

A right-hand circularly polarized wave, $\mathbf{E}^i = \left(E_o \mathbf{a}_x - jE_o \mathbf{a}_y\right)e^{-j\beta z}$, propagates in free space $(z < 0)$ and impinges normally on the surface of a perfect conductor at $z = 0$. Determine

(a) $\mathscr{E}^r$ and its polarization state,

(b) surface current density induced on the surface of the perfect conductor, and

(c) instantaneous form of the total electric field intensity in free space.

Solution

(a) Using $\Gamma = -1$, the electric field phasor of the reflected wave is written as

$$\mathbf{E}^r = \Gamma\left(E_o \mathbf{a}_x - jE_o \mathbf{a}_y\right)e^{j\beta z} = \left(-E_o \mathbf{a}_x + jE_o \mathbf{a}_y\right)e^{j\beta z} \tag{8-128}$$

The instantaneous form of the reflected wave is

$$\begin{aligned}\mathscr{E}^r &= \mathrm{Re}\left[\left(-E_o \mathbf{a}_x + e^{j\pi/2} E_o \mathbf{a}_y\right)e^{j\beta z} e^{j\omega t}\right] \\ &= -\mathbf{a}_x E_o \cos(\omega t + \beta z) - \mathbf{a}_y E_o \sin(\omega t + \beta z)\end{aligned} \tag{8-129}$$

At the $z = 0$ plane, we have

$$\mathscr{E}^r = -\mathbf{a}_x E_o \qquad \text{at } \omega t = 0 \tag{8-130a}$$

$$\mathscr{E}^r = -\mathbf{a}_y E_o \qquad \text{at } \omega t = \pi/2 \tag{8-130b}$$

If the left four fingers follow the rotation of $\mathscr{E}^r$ in Eq. (8-130), the left thumb points in the direction of propagation of the reflected wave, as shown in Fig. 8.14. Thus, the reflected wave is left-hand circularly polarized. Note that the polarization state of a circularly polarized wave is reversed when the wave is reflected by a perfect conductor.

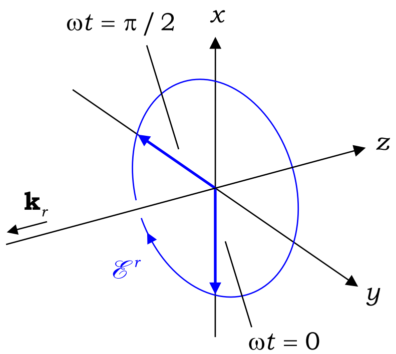


Fig. 8.14 Reflection of a circularly polarized wave by a perfectly conducting surface at $z = 0$.

(b) Magnetic field intensities of the incident and reflected waves are expressed in free space as

$$\mathbf{H}^i = \frac{1}{\eta_o}\left(\mathbf{a}_k \times \mathbf{E}^i\right) = \frac{1}{\eta_o}\mathbf{a}_z \times \left(E_o\mathbf{a}_x - jE_o\mathbf{a}_y\right)e^{-j\beta z} = \frac{E_o}{\eta_o}\left(\mathbf{a}_y + j\mathbf{a}_x\right)e^{-j\beta z} \tag{8-131}$$

$$\mathbf{H}^r = \frac{1}{\eta_o}\left(\mathbf{a}_k \times \mathbf{E}^r\right) = \frac{1}{\eta_o}(-\mathbf{a}_z) \times \left(-E_o\mathbf{a}_x + jE_o\mathbf{a}_y\right)e^{j\beta z} = \frac{E_o}{\eta_o}\left(\mathbf{a}_y + j\mathbf{a}_x\right)e^{j\beta z} \tag{8-132}$$

The total magnetic field intensity at the $z = 0$ plane is obtained from Eqs. (8-131) and (8-132) as

$$\mathbf{H}(0) = \mathbf{H}^i(0) + \mathbf{H}^r(0) = \frac{2E_o}{\eta_o}\left(\mathbf{a}_y + j\mathbf{a}_x\right) \tag{8-133}$$

Using the outward unit normal to the surface, $\mathbf{a}_n = -\mathbf{a}_z$, the surface current density induced on the conductor is

$$\mathbf{J}_s = \mathbf{a}_n \times \mathbf{H}(0) = (-\mathbf{a}_z) \times \frac{2E_o}{\eta_o}\left(\mathbf{a}_y + j\mathbf{a}_x\right) = \frac{2E_o}{\eta_o}\left(\mathbf{a}_x - j\mathbf{a}_y\right) \tag{8-134}$$

(c) Total electric field intensity in free space is written as

$$\mathbf{E} = \mathbf{E}^i + \mathbf{E}^r = \left(E_o\mathbf{a}_x - jE_o\mathbf{a}_y\right)e^{-j\beta z} + \left(-E_o\mathbf{a}_x + jE_o\mathbf{a}_y\right)e^{j\beta z}$$
$$= -\mathbf{a}_x 2jE_o\sin(\beta z) - \mathbf{a}_y 2E_o\sin(\beta z)$$

The instantaneous form of the total electric field intensity in free space is therefore

$$\mathscr{E} = \mathrm{Re}\left[\mathbf{E}e^{j\omega t}\right] = \mathrm{Re}\left[\mathbf{a}_x 2E_o\sin(\beta z)e^{j(\omega t - \pi/2)} - \mathbf{a}_y 2E_o\sin(\beta z)e^{j\omega t}\right]$$
$$= 2E_o\sin(\beta z)\left[\mathbf{a}_x\sin(\omega t) - \mathbf{a}_y\cos(\omega t)\right].$$

Exercise 8.18

A standing wave with $\omega = 5\times10^8$ [rad/s] exists in free space $(z < 0)$ due to a perfectly conducting surface at $z = 0$. Find the location of the first maximum.

Ans. $z = -0.94$ [m].

Exercise 8.19
With reference to Fig. 8.13, find (a) standing wave ratio in medium 1, and (b) phase angle of the reflection coefficient.

Ans. (a) $S = \infty$, (b) $\phi = \pi[\text{rad}]$.

Review Questions with Hints

RQ 8.22 Write the reflection and transmission coefficients for the wave normally incident on a dielectric-dielectric interface. [Eq.(8-107)]
RQ 8.23 Express the law of conservation of energy in terms of Γ and τ of the wave normally incident on an interface. [Eq.(8-111)]
RQ 8.24 What are the standing wave and standing wave ratio? [Eq.(8-119)]
RQ 8.25 Express the separation between two adjacent electric field intensity maxima in a standing wave pattern? [Eq.(8-115)]
RQ 8.26 Write the reflection coefficient of the wave normally incident on the surface of a perfect conductor. [Eq.(8-126)]

8.3.2 Oblique Incidence of a Plane Wave

In the general case, a uniform plane wave may be obliquely incident on an interface between two different dielectrics. The incident wave is partly reflected from the interface, while the remainder is transmitted into the second medium, as in the case of the normal incidence. Under certain conditions, the incident wave may undergo a transmission only or a reflection only at the interface.

The behavior of the wave at the interface strongly depends on the direction of propagation and the polarization state of the incident wave. The plane of incidence is a reference plane that is defined by the wavevector of the incident wave and a unit normal to the interface. It can be shown that all three wavevectors of the incident, reflected, and transmitted waves are confined in the plane of incidence. Thus, it would be most convenient to specify the directions of propagation of the waves by the rotation angles of the wavevectors with respect to the unit normal to the interface. Figure 8.15 shows the definitions of the angle of incidence, angle of reflection, and angle of transmission, which are denoted by θ_i, θ_r, and θ_t, respectively. These angles are always positive, ranging from zero to $\pi/2$. Next, to describe the polarization state of the wave at the interface, we define the perpendicular polarization and the parallel polarization, which are so named because the electric field vectors are perpendicular and parallel to the plane of incidence, respectively. Note that the two polarization states of a wave are mutually orthogonal at the interface.

With reference to the plane of incidence shown in Fig. 8.15, we generally express the electric field phasors of the incident, reflected, and transmitted waves as follows:

$$\mathbf{E}^i = \mathbf{E}_o^i e^{-j\mathbf{k}_i \cdot \mathbf{r}} = \mathbf{E}_o^i e^{-j\left[(k_i \sin\theta_i)x + (k_i \cos\theta_i)z\right]} \tag{8-135a}$$

$$\mathbf{E}^r = \mathbf{E}_o^r e^{-j\mathbf{k}_r \cdot \mathbf{r}} = \mathbf{E}_o^r e^{-j\left[(k_r \sin\theta_r)x - (k_r \cos\theta_r)z\right]} \tag{8-135b}$$

$$\mathbf{E}^t = \mathbf{E}_o^t e^{-j\mathbf{k}_t \cdot \mathbf{r}} = \mathbf{E}_o^t e^{-j\left[(k_t \sin\theta_t)x + (k_t \cos\theta_t)z\right]} \tag{8-135c}$$

where i, r, and t stand for the incident, reflected, and transmitted waves, respectively, and k is the magnitude of the wavevector $\mathbf{k}$. For known values of $\mathbf{E}_o^i$, k_i, and θ_i, we can determine the values of $\mathbf{E}_o^r$, $\mathbf{E}_o^t$, k_r, k_t, θ_r, and θ_t by applying the boundary conditions to the three waves at the interface.

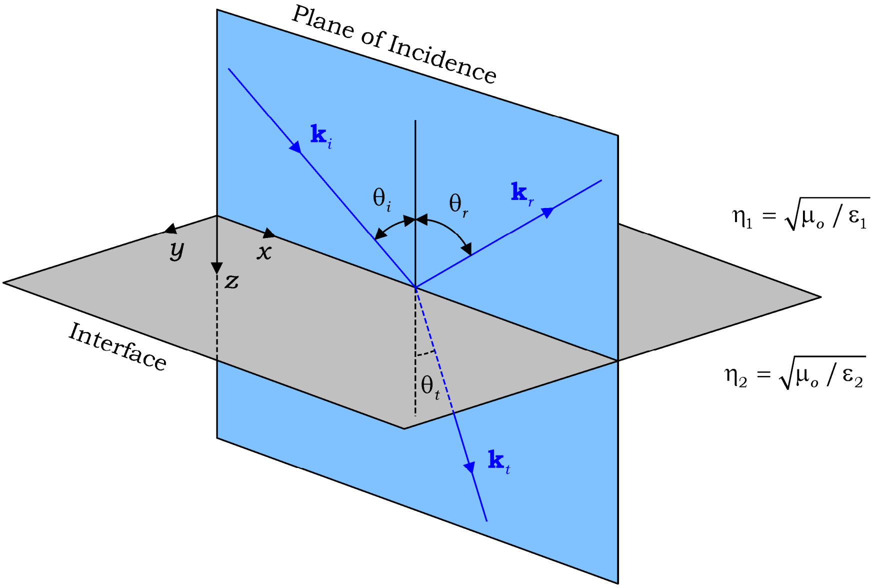


Fig. 8.15 Plane of incidence at an interface between two dielectrics.

We consider an interface between two dielectrics of different permittivities, which are lossless and nonmagnetic($\alpha = 0$ and $\mu = \mu_o$). We first apply the boundary condition for $\mathbf{E}$ to the electric field intensities given in Eq. (8-135), and make the tangential component of the electric field intensity be continuous across the interface such as $\left(\mathbf{E}^i\right)_{\tan} + \left(\mathbf{E}^r\right)_{\tan} = \left(\mathbf{E}^t\right)_{\tan}$, where *tan* stands for the tangential component. Substituting $z = 0$ into Eq. (8-135), and applying the boundary condition, we have

$$\left(\mathbf{E}_o^i\right)_{\tan} e^{-j(k_i \sin\theta_i)x} + \left(\mathbf{E}_o^r\right)_{\tan} e^{-j(k_r \sin\theta_r)x} = \left(\mathbf{E}_o^t\right)_{\tan} e^{-j(k_t \sin\theta_t)x} \tag{8-136}$$

In the above equation, we notice that the sum of two spatial functions on the left-hand side is equated with another spatial function on the right-hand side. The equality in Eq. (8-136) can be satisfied at every point on the interface only if the three terms have the same functional form. Therefore, the three exponents should be identical, that is,

$$\boxed{k_i \sin\theta_i = k_r \sin\theta_r = k_t \sin\theta_t} \tag{8-137}$$

This implies that the tangential components of $\mathbf{k}_r$ and $\mathbf{k}_t$ should be the same as that of $\mathbf{k}_i$.

Here, let us digress briefly and introduce the refractive index of a lossless and nonmagnetic material, which is defined as

$$n = \sqrt{\varepsilon / \varepsilon_o} = \sqrt{\varepsilon_r} \tag{8-138}$$

where ε_r is the relative permittivity, or the dielectric constant of the material. Thus, the wavenumber, or the magnitude of the wavevector, is expressed as

$$k = \omega\sqrt{\mu_o \varepsilon} = n\frac{\omega}{c} \tag{8-139}$$

It is important to note that the refractive index is related to the intrinsic impedance of the material by

$$\eta = \eta_o / n \tag{8-140}$$

where the intrinsic impedance of free space $\eta_o = \sqrt{\mu_o / \varepsilon_o}$.

Noting that the incident and reflected waves are in the same medium, we obtain the relation $k_i = k_r$. Therefore, the first equality in Eq. (8-137) leads to the law of reflection:

$$\boxed{\theta_i = \theta_r} \tag{8-141}$$

Next, with the help of Eq. (8-139), the second equality in Eq. (8-137) leads to Snell's law of refraction:

$$\boxed{n_1 \sin\theta_i = n_2 \sin\theta_t} \tag{8-142}$$

Again, θ_i and θ_t are the angles of incidence and transmission, respectively, while n_1 and n_2 are the refractive indices of medium 1 and 2, respectively. Because the laws of reflection and refraction are derived from the relation between the wavenumbers, or Eq. (8-137), they hold true regardless of the polarization state of the wave.

Example 8-16

A uniform plane wave of a wavevector $\mathbf{k}_i = \mathbf{a}_x + \sqrt{3}\,\mathbf{a}_y + 2\sqrt{3}\,\mathbf{a}_z$ propagates in free space in the region $z < 0$, and impinges on the surface of a dielectric of

$\varepsilon_r = 1.69$ and $\mu_r = 1$, occupying the region defined by $z \geq 0$. Find (a) $\mathbf{k}_r$, and (b) $\mathbf{k}_t$.

Solution

(a) The law of reflection comprises of two parts:

(I) $\theta_r = \theta_i$.

(II) $\mathbf{k}_i$ and $\mathbf{k}_r$ are in the plane of incidence.

Thus

$$\mathbf{k}_r = \mathbf{a}_x + \sqrt{3}\,\mathbf{a}_y - 2\sqrt{3}\,\mathbf{a}_z .$$

(b) The law of refraction comprises of two parts:

(I) $n_i \sin\theta_i = n_t \sin\theta_t$

(II) $\mathbf{k}_i$ and $\mathbf{k}_t$ are in the plane of incidence

From Snell's law, the tangential component of $\mathbf{k}_t$ is the same as that of $\mathbf{k}_i$, that is,

$$\mathbf{k}_t = \mathbf{a}_x + \sqrt{3}\,\mathbf{a}_y + k_{tz}\,\mathbf{a}_z$$

where k_{tz} is an unknown.

Using $k_t = n_t k_i = 1.3 \times 4$ in the above equation, we write

$$(5.2)^2 = 1 + (\sqrt{3})^2 + (k_{tz})^2$$

Thus,

$$\mathbf{k}_t = \mathbf{a}_x + \sqrt{3}\,\mathbf{a}_y + 4.8\,\mathbf{a}_z .$$

Exercise 8.20

What is the maximum angle of transmission of sunlight into the water of a placid lake with a refractive index 1.33.

Ans. 48.8^o.

Exercise 8.21

An incident wave of $\mathbf{k}_i = 3\mathbf{a}_x + 4\mathbf{a}_y + 5\mathbf{a}_z$ in medium 1($z < 0$) impinges on an interface at $z = 0$. For $n_1 = 1.2$ and $n_2 = 2.4$, find $\mathbf{k}_r$ and $\mathbf{k}_t$.

Ans. $\mathbf{k}_r = 3\mathbf{a}_x + 4\mathbf{a}_y - 5\mathbf{a}_z$, and $\mathbf{k}_t = 3\mathbf{a}_x + 4\mathbf{a}_y + \sqrt{175}\,\mathbf{a}_z$.

8.3.2.1 Perpendicular Polarization

A uniform plane wave with the perpendicular polarization, or *s*-polarization, has the electric field vector always perpendicular to the plane of incidence. Let us consider the case as illustrated in Fig. 8.16, in which the wave is obliquely incident on an interface. Although the directions of propagation are yet to be determined, the reflected and transmitted waves should also be uniform plane waves of perpendicular polarization. Otherwise, the boundary conditions would not be satisfied at the interface. Assuming the three waves to be perpendicularly polarized, we write the electric field phasors of the waves as follows:

$$\mathbf{E}^i = \mathbf{a}_y E_o^i e^{-j\mathbf{k}_i \cdot \mathbf{r}} = \mathbf{a}_y E_o^i e^{-j\left[(k_i \sin\theta_i)x + (k_i \cos\theta_i)z\right]} \tag{8-143a}$$

$$\mathbf{E}^r = \mathbf{a}_y E_o^r e^{-j\mathbf{k}_r \cdot \mathbf{r}} = \mathbf{a}_y E_o^r e^{-j\left[(k_r \sin\theta_r)x - (k_r \cos\theta_r)z\right]} \tag{8-143b}$$

$$\mathbf{E}^t = \mathbf{a}_y E_o^t e^{-j\mathbf{k}_t \cdot \mathbf{r}} = \mathbf{a}_y E_o^t e^{-j\left[(k_t \sin\theta_t)x + (k_t \cos\theta_t)z\right]} \tag{8-143c}$$

In the above equations we assumed that all three electric field vectors are directed in the $+y$-direction for simplicity. It is implicit that the angles θ_r and θ_t are related to θ_i through the laws of reflection and refraction, respectively. Considering Eq. (8-26), we write the magnetic field phasors of the three waves as

$$\mathbf{H}^i = \left(\mathbf{a}_i \times \mathbf{a}_y\right)\frac{E_o^i}{\eta_1} e^{-j\mathbf{k}_i \cdot \mathbf{r}} = \left(-\mathbf{a}_x \cos\theta_i + \mathbf{a}_z \sin\theta_i\right)\frac{E_o^i}{\eta_1} e^{-j\mathbf{k}_i \cdot \mathbf{r}} \tag{8-144a}$$

$$\mathbf{H}^r = \left(\mathbf{a}_r \times \mathbf{a}_y\right)\frac{E_o^r}{\eta_1} e^{-j\mathbf{k}_r \cdot \mathbf{r}} = \left(\mathbf{a}_x \cos\theta_r + \mathbf{a}_z \sin\theta_r\right)\frac{E_o^r}{\eta_1} e^{-j\mathbf{k}_r \cdot \mathbf{r}} \tag{8-144b}$$

$$\mathbf{H}^t = \left(\mathbf{a}_t \times \mathbf{a}_y\right)\frac{E_o^t}{\eta_2} e^{-j\mathbf{k}_t \cdot \mathbf{r}} = \left(-\mathbf{a}_x \cos\theta_t + \mathbf{a}_z \sin\theta_t\right)\frac{E_o^t}{\eta_2} e^{-j\mathbf{k}_t \cdot \mathbf{r}} \tag{8-144c}$$

where we have used $\mathbf{a}_i = \mathbf{a}_x \sin\theta_i + \mathbf{a}_z \cos\theta_i$, $\mathbf{a}_r = \mathbf{a}_x \sin\theta_r - \mathbf{a}_z \cos\theta_r$, and $\mathbf{a}_t = \mathbf{a}_x \sin\theta_t + \mathbf{a}_z \cos\theta_t$, which are the unit vectors of $\mathbf{k}_i$, $\mathbf{k}_r$, and $\mathbf{k}_t$, respectively. In Eq. (8-114), the position vector is given as $\mathbf{r} = x\mathbf{a}_x + y\mathbf{a}_y + z\mathbf{a}_z$ in Cartesian coordinates. It should be noted that the magnetic field vectors, represented by the phasors $\mathbf{H}^i$, $\mathbf{H}^r$, and $\mathbf{H}^t$, all lie in the plane of incidence.

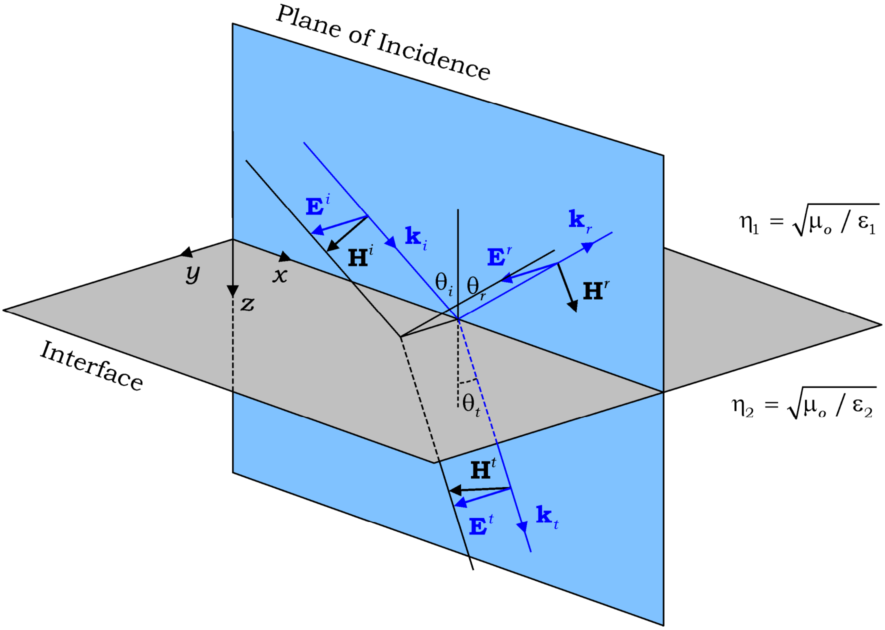


Fig. 8.16 Uniform plane waves of perpendicular polarization at an interface.

Upon applying the boundary conditions for **E** and **H** to the interface coincident with the $z = 0$ plane, we have

$$\mathbf{E}^i\Big|_{z=0,\tan} + \mathbf{E}^r\Big|_{z=0,\tan} = \mathbf{E}^t\Big|_{z=0,\tan}$$
$$\mathbf{H}^i\Big|_{z=0,\tan} + \mathbf{H}^r\Big|_{z=0,\tan} = \mathbf{H}^t\Big|_{z=0,\tan}$$

where tan stands for the tangential component. Substitution of Eq. (8-143) and Eq. (8-144) into the above equations leads to

$$E_o^i + E_o^r = E_o^t \tag{8-145a}$$

$$-\cos\theta_i \frac{E_o^i}{\eta_1} + \cos\theta_r \frac{E_o^r}{\eta_1} = -\cos\theta_t \frac{E_o^t}{\eta_2} \tag{8-145b}$$

In view of Eq. (8-137), the common exponential factor has been dropped from both sides of the above equation.

The reflection coefficient $\Gamma_\perp$ and the transmission coefficient $\tau_\perp$ for perpendicular polarization are obtained from Eq. (8-145):

$$\boxed{\Gamma_\perp = \frac{E_o^r}{E_o^i} = \frac{\eta_2\cos\theta_i - \eta_1\cos\theta_t}{\eta_2\cos\theta_i + \eta_1\cos\theta_t}} \tag{8-146a}$$

$$\boxed{\tau_\perp = \frac{E_o^t}{E_o^i} = \frac{2\eta_2\cos\theta_i}{\eta_2\cos\theta_i + \eta_1\cos\theta_t}} \tag{8-146b}$$

These are also called the Fresnel's equations for perpendicular polarization. It should be noted that positive values of $\Gamma_\perp$ and $\tau_\perp$ imply that $\mathbf{E}^r$ and $\mathbf{E}^t$ are directed along the direction of $\mathbf{E}^i$.

From Eq. (8-146) we obtain

$$\boxed{1+\Gamma_\perp = \tau_\perp} \tag{8-147}$$

which is true for any angle of incidence.

At normal incidence($\theta_i = 0 = \theta_t$), Eq. (8-146) reduces to Eq. (8-107) as expected. If medium 2 is a perfect conductor($\eta_2 = 0$), we have $\Gamma_\perp = -1$ and $\tau_\perp = 0$, irrespective of the angle of incidence. In this case, from Eq. (8-143) we see that $\mathbf{E}^i$ and $\mathbf{E}^r$ completely cancel each other on the surface of the perfect conductor.

In Eqs. (8-143) and (8-144), the wavevectors $\mathbf{k}_r$ and $\mathbf{k}_t$ were assumed to be confined in the plane of incidence as $\mathbf{k}_i$. Otherwise, $\mathbf{k}_r$ and $\mathbf{k}_t$ would have a component in the y-direction, causing $\mathbf{E}^r$ and $\mathbf{E}^t$ to vary with y, and the boundary condition could not be satisfied at the interface, because $\mathbf{E}^i$ is independent of y.

Example 8-17

A uniform plane wave, $\mathscr{E}^i = 10\mathbf{a}_y \cos\left(\omega t - \sqrt{3}\,x - z\right)$, propagates in free space $(z < 0)$ and impinges on a lossless dielectric $(z \geq 0)$ of $\varepsilon_r = 4$. Find

(a) $\mathbf{k}_i$ and θ_i,

(b) $\mathscr{E}^r$, and

(c) total electric field intensity in the region $z < 0$.

Solution

(a) Phasor form of $\mathscr{E}^i$ is

$$\mathbf{E}^i = 10\mathbf{a}_y e^{-j\left(\sqrt{3}\,x+z\right)}$$

The wavevector of $\mathbf{E}^i$ is

$$\mathbf{k}_i = \sqrt{3}\,\mathbf{a}_x + \mathbf{a}_z .$$

The angle of incidence is the smaller angle between $\mathbf{k}_i$ and $\mathbf{a}_z$

$$\theta_i = \tan^{-1}\frac{\sqrt{3}}{1} = 60^0 .$$

(b) Inserting $\theta_i = 60^0$, $n_1 = 1$, and $n_2 = 2$ into Snell's law of refraction($n_1 \sin\theta_i = n_2 \sin\theta_t$), we have

$$\cos\theta_t = \sqrt{1 - \left(\frac{n_1}{n_2}\sin\theta_i\right)^2} = 0.901$$

Rewriting Eq. (8-146a) by use of $\eta = \eta_o / n$, we have

$$\Gamma_\perp = \frac{n_1\cos\theta_i - n_2\cos\theta_t}{n_1\cos\theta_i + n_2\cos\theta_t} = \frac{\cos 60^o - 2\times 0.901}{\cos 60^o + 2\times 0.901} = -0.57$$

Substituting $k_r = k_i = 2$ and $\theta_r = \theta_i = 60^o$ into Eq. (8-143b), we have

$$\mathbf{E}^r = \mathbf{a}_y(-0.57)10e^{-j\left[\left(2\sin 60^o\right)x - \left(2\cos 60^o\right)z\right]} = \mathbf{a}_y(-5.7)e^{-j\left(\sqrt{3}x - z\right)}$$

$$\mathscr{E}^r = \text{Re}\left[\mathbf{E}^r e^{j\omega t}\right] = \mathbf{a}_z(-5.7)\cos\left(\omega t - \sqrt{3}x + z\right).$$

(c) In the region $z < 0$, the total electric field intensity is expressed in phasor form as

$$\begin{aligned}\mathbf{E} = \mathbf{E}^i + \mathbf{E}^r &= \mathbf{a}_y e^{-j\sqrt{3}x}\left[10e^{-jz} - 5.7e^{+jz}\right] \\ &= \mathbf{a}_y e^{-j\sqrt{3}x}\left[(15.7 - 5.7)e^{-jz} - 5.7e^{+jz}\right] \\ &= \mathbf{a}_y e^{-j\sqrt{3}x}\left[15.7e^{-jz} - 11.4\cos z\right]\end{aligned}$$

The total electric field intensity is written in instantaneous form as

$$\begin{aligned}\mathscr{E} &= \text{Re}\left[\mathbf{E}e^{\omega t}\right] \\ &= \mathbf{a}_y\left[15.7\cos\left(\omega t - \sqrt{3}x - z\right) - 11.4\cos(z)\cos\left(\omega t - \sqrt{3}x\right)\right].\end{aligned}$$

Exercise 8.22

Under what condition is $\Gamma_\perp$ (a) positive, and (b) negative at an interface between two lossless dielectrics.

Ans. (a) $\Gamma_\perp > 0$ for $\eta_1 < \eta_2$, (b) $\Gamma_\perp < 0$ for $\eta_1 > \eta_2$.

8.3.2.2 Parallel Polarization

A uniform plane wave with the parallel polarization, or *p*-polarization, has the electric field vector parallel to the plane of incidence. The same procedure as was used for the perpendicular polarization can be followed to obtain the reflection and transmission coefficients for the parallel polarization. The sign convention for the positive direction of an electric field vector is taken so that the projection of **E** onto the interface is directed along the projection of the wavevector **k** onto the same interface, as shown in Fig. 8.17. Then the direction of **H** is determined by the relation $\mathbf{a}_k = \mathbf{a}_E \times \mathbf{a}_H$. For the uniform plane wave obliquely incident on the interface as shown in Fig. 8.17, the reflected and transmitted waves should also be uniform plane waves of a parallel polarization. Assuming the three waves to be parallel polarized, we write the electric field phasors of the waves as follows:

$$\mathbf{E}^i = \left(\mathbf{a}_x \cos\theta_i - \mathbf{a}_z \sin\theta_i\right) E_o^i e^{-j\left[(k_i \sin\theta_i)x + (k_i \cos\theta_i)z\right]} \tag{8-148a}$$

$$\mathbf{E}^r = \left(\mathbf{a}_x \cos\theta_r + \mathbf{a}_z \sin\theta_r\right) E_o^r e^{-j\left[(k_r \sin\theta_r)x - (k_r \cos\theta_r)z\right]} \tag{8-148b}$$

$$\mathbf{E}^t = \left(\mathbf{a}_x \cos\theta_t - \mathbf{a}_z \sin\theta_t\right) E_o^t e^{-j\left[(k_t \sin\theta_t)x + (k_t \cos\theta_t)z\right]} \tag{8-148c}$$

The electric field vector is said to be in the positive direction if the projections of $\mathbf{E}$ and $\mathbf{k}$ onto the interface are parallel to each other; the x-component of $\mathbf{E}$ is positive as shown in Fig. 8.17, in this case. It is implicit in Eq. (8-148) that the angles θ_r and θ_t are related to θ_i through the laws of reflection and refraction, respectively. In view of Eq. (8-26), we write the magnetic field phasors of the three waves as

$$\mathbf{H}^i = \mathbf{a}_y \frac{E_o^i}{\eta_1} e^{-j\left[(k_i \sin\theta_i)x + (k_i \cos\theta_i)z\right]} \tag{8-149a}$$

$$\mathbf{H}^r = -\mathbf{a}_y \frac{E_o^r}{\eta_1} e^{-j\left[(k_r \sin\theta_r)x - (k_r \cos\theta_r)z\right]} \tag{8-149b}$$

$$\mathbf{H}^t = \mathbf{a}_y \frac{E_o^t}{\eta_2} e^{-j\left[(k_t \sin\theta_t)x + (k_t \cos\theta_t)z\right]} \tag{8-149c}$$

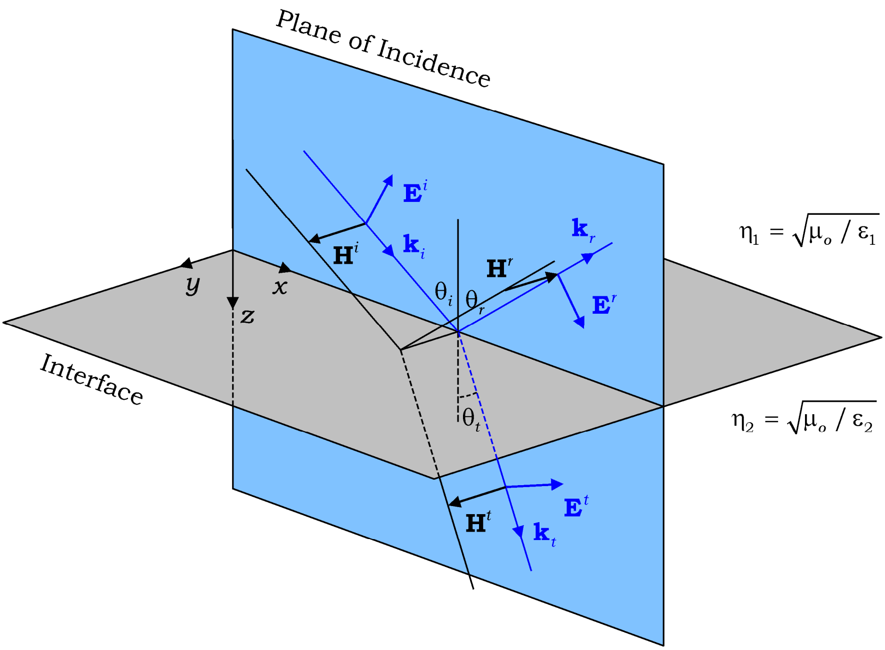


Fig. 8.17 Uniform plane waves of parallel polarization at an interface.

Upon applying the boundary conditions for **E** and **H** to the interface at the $z = 0$ plane, we have

$$\mathbf{E}^i\Big|_{z=0,\tan} + \mathbf{E}^r\Big|_{z=0,\tan} = \mathbf{E}^t\Big|_{z=0,\tan}$$

$$\mathbf{H}^i\Big|_{z=0,\tan} + \mathbf{H}^r\Big|_{z=0,\tan} = \mathbf{H}^t\Big|_{z=0,\tan}$$

Substitution of Eqs. (8-148) and (8-149) into the above equations gives

$$\cos\theta_i E_o^i + \cos\theta_r E_o^r = \cos\theta_t E_o^t \tag{8-150a}$$

$$\frac{E_o^i}{\eta_1} - \frac{E_o^r}{\eta_1} = \frac{E_o^t}{\eta_2} \tag{8-150b}$$

In view of Eq. (8-137), the common exponential factor has been dropped from both sides of the above equation.

The reflection coefficient $\Gamma_{\|}$ and the transmission coefficient $\tau_{\|}$ for parallel polarization are obtained from Eq. (8-150):

$$\boxed{\Gamma_{\|} = \frac{E_o^r}{E_o^i} = \frac{\eta_2\cos\theta_t - \eta_1\cos\theta_i}{\eta_2\cos\theta_t + \eta_1\cos\theta_i}} \tag{8-151a}$$

$$\boxed{\tau_{\|} = \frac{E_o^t}{E_o^i} = \frac{2\eta_2\cos\theta_i}{\eta_2\cos\theta_t + \eta_1\cos\theta_i}} \tag{8-151b}$$

which are called the Fresnel's equations for parallel polarization. It should be noted that positive values of $\Gamma_{\|}$ and $\tau_{\|}$ imply that the *x*-components of $\mathbf{E}^r$ and $\mathbf{E}^t$ are parallel to that of $\mathbf{E}^i$.

From Eq. (8-151) we obtain

$$\boxed{1 + \Gamma_{\|} = \tau_{\|}\left(\frac{\cos\theta_t}{\cos\theta_i}\right)} \tag{8-152}$$

It is important to remember that the right-hand side of the above equation differs from that of Eq. (8-147).

At normal incidence($\theta_i = 0 = \theta_t$), Eq. (8-151) and Eq. (8-152) reduce to Eq. (8-107) and Eq. (8-108), respectively, as expected. If medium 2 is a perfect conductor($\eta_2 = 0$), we have $\Gamma_{\|} = -1$ and $\tau_{\|} = 0$, irrespective of the angle of incidence. In this case, we see from Eq. (8-148) that the tangential components of $\mathbf{E}^i$ and $\mathbf{E}^r$ cancel each other on the surface of the perfect conductor. However, the sum of the normal components of $\mathbf{E}^i$ and $\mathbf{E}^r$ is not necessarily zero on the perfect conductor.

If a wave, traveling in the first medium, reflects off the second medium of a larger refractive index, $n_1 < n_2$, this is called the external reflection. Otherwise, it is called the internal reflection. The reflection and transmission coefficient are plotted as a function of θ_i in Fig. 8.18(a) for the case of external reflection($n_1 = 1$ and $n_2 = 1.5$), and in Fig. 8.18(b) for the case of internal reflection($n_1 = 1.5$ and $n_2 = 1$). Both plots show that $|\Gamma_\perp|$ monotonically increases from an initial value to unity as θ_i increases from 0^o to 90^o. In contrast, $|\Gamma_\parallel|$ first decreases from the initial value to zero and then increases to unity as θ_i increases from 0^o to 90^o. In view of these, the wave with perpendicular polarization reflects better than the wave with parallel polarization.

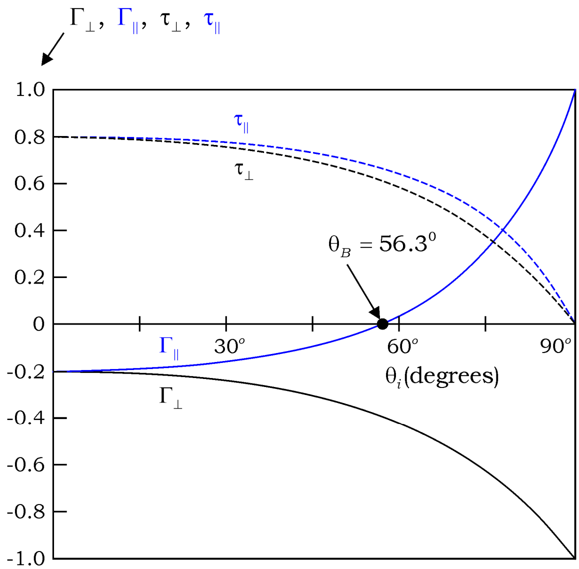


Fig. 8.18(a) External reflection($n_1 = 1$ and $n_2 = 1.5$).

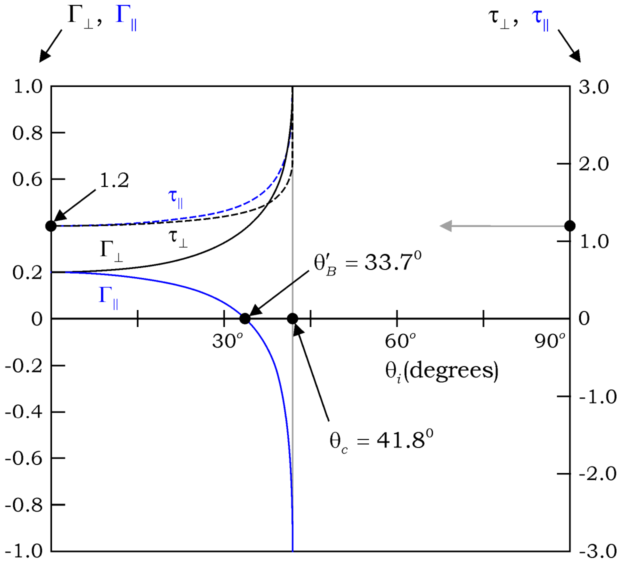


Fig. 8.18(b) Internal reflection($n_1 = 1.5$ and $n_2 = 1$).

Example 8-18

The uniform plane wave with parallel polarization, $\mathbf{E}^i = 4\mathbf{a}_E e^{-j(2x+3z)}$, propagates in free space($z < 0$), and impinges obliquely on a lossless dielectric of $\varepsilon_r = 2.25$ occupying the region $z \geq 0$. Find

(a) expression for $\mathbf{a}_E$ in Cartesian coordinates,

(b) $\mathbf{E}^r$, and

(c) $\mathbf{E}^t$.

Solution

(a) From the wavevector $\mathbf{k}_i = 2\mathbf{a}_x + 3\mathbf{a}_z$, the angle of incidence is

$$\theta_i = \tan^{-1}\frac{2}{3} \cong 33.7^o$$

The direction of the parallel polarization is determined from the two conditions; that is, $\mathbf{a}_E \bullet \mathbf{k}_i = 0$ and $(\mathbf{a}_E)_x \parallel (\mathbf{k}_i)_x$.

From the first condition : $\mathbf{a}_E \parallel \pm(3\mathbf{a}_x - 2\mathbf{a}_z)$

From the second condition : $\mathbf{a}_E \parallel (3\mathbf{a}_x - 2\mathbf{a}_z)$

Thus,

$$\mathbf{a}_E = \frac{1}{\sqrt{13}}(3\mathbf{a}_x - 2\mathbf{a}_z)$$

The electric field phasor of the incident wave is

$$\mathbf{E}^i = \frac{4}{\sqrt{13}}(3\mathbf{a}_x - 2\mathbf{a}_z)e^{-j(2x+3z)}. \tag{8-153a}$$

(b) Substituting $n_1 = 1$ and $n_2 = \sqrt{2.25} = 1.5$ into Snell's law of refraction, $n_1 \sin\theta_i = n_2 \sin\theta_t$, we obtain

$$\theta_t = \sin^{-1}\left(\frac{1}{1.5}\sin 33.7^o\right) \cong 21.7^o$$

From Eq. (8-151a), by use of $\eta = \eta_o / n$, we obtain

$$\Gamma_{||} = \frac{n_1\cos\theta_t - n_2\cos\theta_i}{n_1\cos\theta_t + n_2\cos\theta_i} = \frac{\cos 21.7^o - 1.5\times\cos 33.7^o}{\cos 21.7^o + 1.5\times\cos 33.7^o} = -0.15$$

Substituting $E_o^i = 4$, $\Gamma_{||} = -0.15$, $k_r = k_i = \sqrt{13}$, and $\theta_r = \theta_i = 33.7^o$ into Eq. (8-148b), we obtain

$$\begin{aligned}\mathbf{E}^r &= (\mathbf{a}_x\cos\theta_r + \mathbf{a}_z\sin\theta_r)\Gamma_{||}E_o^i e^{-j[(k_r\sin\theta_r)x-(k_r\cos\theta_r)z]}\\ &= (0.83\mathbf{a}_x + 0.55\mathbf{a}_z)(-0.15\times 4)e^{-j[2x-3z]}\\ &= -(0.50\mathbf{a}_x + 0.33\mathbf{a}_z)e^{-j[2x-3z]}.\end{aligned} \tag{8-153b}$$

(c) From Eq. (8-151b) we obtain

$$\tau_{||} = \frac{2n_1\cos\theta_i}{n_1\cos\theta_t + n_2\cos\theta_i} = \frac{2\cos 33.7^o}{0.929 + 1.5\cos 33.7^o} = 0.76$$

Substituting $k_t = k_i\sqrt{\varepsilon_r} = 1.5\sqrt{13}$ and $\theta_t = 21.7^o$ into Eq. (8-148c),

$$\begin{aligned}\mathbf{E}^t &= (\mathbf{a}_x\cos\theta_t - \mathbf{a}_z\sin\theta_t)\tau_{||}E_o^i e^{-j[(k_t\sin\theta_t)x+(k_t\cos\theta_t)z]}\\ &= (0.93\mathbf{a}_x - 0.37\mathbf{a}_z)0.76\times 4e^{-j[2.00x+5.03z]}\\ &= (2.84\mathbf{a}_x - 1.13\mathbf{a}_z)e^{-j[2.00x+5.03z]}.\end{aligned} \tag{8-153c}$$

Exercise 8.23

Given $\mathbf{E} = \mathbf{a}_E 10e^{-j[3x+2y+4z]}$, traveling in free space $(z < 0)$ and impinging on an interface at $z = 0$, find the expression for $\mathbf{a}_E$ for the parallel polarization.

Ans. $\mathbf{a}_E = (12\mathbf{a}_x + 8\mathbf{a}_y - 13\mathbf{a}_z)/\sqrt{377}$.

8.3.2.3 Brewster Angle

A uniform plane wave with parallel polarization undergoes no reflection if it impinges on the interface at an angle called the Brewster angle. The wave is totally transmitted into the second medium. There is, however, no Brewster angle for perpendicular polarization. The Brewster angle for external reflection, θ_B, is complementary to the Brewster angle for internal reflection, θ'_B, that is, $\theta_B + \theta'_B = 90^o$ (see Fig. 8.18). The Brewster angle is also called the polarization angle. When an unpolarized wave is incident upon a surface at the Brewster angle, only the component with perpendicular polarization undergoes a reflection, and thus the surface behaves like a polarizer.

If a material medium is lossless and nonmagnetic, its intrinsic impedance is expressed in terms of the refractive index as $\eta = \eta_o / n$. At an interface between two such materials, the reflection coefficient expressed by Eq. (8-151a) is rewritten, with the aid of Snell's law of refraction, as

$$\boxed{\Gamma_{\parallel} = \frac{\tan(\theta_t - \theta_i)}{\tan(\theta_t + \theta_i)}} \tag{8-154}$$

In the case of $\theta_t + \theta_i = 90^o$, Eq. (8-154) gives $\Gamma_{\parallel} = 0$. Therefore we obtain the relation $\theta_i = 90^o - \theta_t = \theta_B$. At the Brewster angle, Snell's law of refraction is rewritten as

$$\begin{aligned} n_1 \sin\theta_B &= n_2 \sin\theta_t \\ &= n_2 \sin\left(90^o - \theta_B\right) \end{aligned}$$

Rearranging the terms in the above equation, we can express the Brewster angle in terms of the refractive indices, namely

$$\boxed{\tan\theta_B = \frac{n_2}{n_1}} \tag{8-155}$$

This is the Brewster angle for the wave traveling in medium 1, impinging on medium 2. By the same token, the Brewster angle for the wave traveling in medium 2, impinging on medium 1, is

$$\tan\theta'_B = \frac{n_1}{n_2} \tag{8-156}$$

Combining Eq. (8-155) and Eq. (8-156) we obtain

$$\theta_B + \theta'_B = 90^o \tag{8-157}$$

The two Brewster angles are complementary to each other.

Similarly, at an interface between two lossless, nonmagnetic materials, Eq. (8-146a) is rewritten as

$$\boxed{\Gamma_{\perp} = \frac{\sin(\theta_t - \theta_i)}{\sin(\theta_t + \theta_i)}} \tag{8-158}$$

The reflection coefficient $\Gamma_\perp$ is always positive regardless of the angle of incidence; there is no Brewster angle for perpendicular polarization(see also Fig. 8.18).

Example 8-19
On the front surface of a dielectric slab as shown in Fig. 8.19, a uniform plane wave with parallel polarization is incident at Brewster angle θ_B. Show that the wave also experiences no reflection at the rear surface which is parallel to the front one.

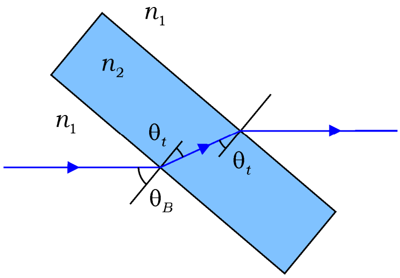


Fig. 8.19 A wave incident on a dielectric slab at Brewster angle.

Solution

Substituting $\theta_B = 90^o - \theta_t$, which is obtained from Eq. (8-154), into Eq. (8-155), we write

$$\frac{n_2}{n_1} = \frac{\sin\theta_B}{\cos\theta_B} = \frac{\cos\theta_t}{\sin\theta_t} \tag{8-159}$$

Comparing Eq. (8-159) with Eq. (8-156), we obtain $\theta_t = \theta'_B$, which implies that the internal wave is incident on the rear surface at Brewster angle. Thus, there is no reflection at the rear surface.

Exercise 8.24
Find θ_B and θ'_B for a diamond of $n = 2.42$ placed in free space.

Ans. $\theta_B = 67.5^o$, and $\theta'_B = 22.5^o$.

Review Questions with Hints

RQ 8.27 What is the plane of incidence? [Fig.8.15]
RQ 8.28 Define the perpendicular and parallel polarizations. [Figs.8.16,8.17]
RQ 8.29 State Snell's law of refraction. [Eq. (8-142)]
RQ 8.30 Write Fresnel's equations. [Eqs.(8-146)(8-151)]
RQ 8.31 Define Brewster angle. [Eq.(8-155)]

8.3.3 Total Internal Reflection

We now turn our attention to the internal reflection of a uniform plane wave at a dielectric-dielectric interface, and discuss the total internal reflection(TIR). The internal reflection takes place for a wave incident from medium 1 onto medium 2 of a lower refractive index($n_1 > n_2$). In this case, according to Snell's law of refraction, the angle of transmission θ_t in medium 2 is always larger than the angle of incidence θ_i in medium 1. As we gradually increase the angle θ_i from 0^o to a higher value, the angle θ_t increases from 0^o to 90^o, before θ_i reaches 90^o. The particular angle of incidence corresponding to $\theta_t = 90^o$ is called the critical angle θ_c. If we further increase θ_i past the critical angle, there is no transmission of the wave into medium 2, and the incident wave is totally reflected by the interface. This is known as the total internal reflection. Substitution of $\theta_t = 90^o$ into Snell's law of refraction leads to the critical angle defined as

$$\boxed{\sin\theta_c = \frac{n_2}{n_1}} \tag{8-160}$$

where $n_1 > n_2$ is assumed.

Although there is no transmission of electromagnetic energy into medium 2 under the conditions of total internal reflection, a residual electromagnetic field should exist in medium 2. Otherwise, the boundary conditions would not be satisfied at the interface. For $\theta_i > \theta_c$, Snell's law of refraction does not allow a real solution for θ_t such as

$$\cos\theta_t = \pm\sqrt{1-\sin^2\theta_t} = \pm j\sqrt{(n_1/n_2)^2\sin^2\theta_i - 1} \tag{8-161}$$

where $n_1 > n_2$. To obtain an expression for the electric field in medium 2, we insert Eq. (8-161) into either Eq. (8-143b) or Eq. (8-148b):

$$\boxed{\begin{aligned}\mathbf{E}^t &= \mathbf{a}_E E_o^t e^{-j[(k_t\sin\theta_t)x+(k_t\cos\theta_t)z]} \\ &= \mathbf{a}_E E_o^t e^{-\alpha_e z} e^{-j\beta_e x}\end{aligned}} \tag{8-162}$$

where

$\mathbf{a}_E = \mathbf{a}_y$ (perpendicular polarization)

$\mathbf{a}_E = \mathbf{a}_x\cos\theta_t - \mathbf{a}_z\sin\theta_t$ (parallel polarization)

By use of Snell's law of refraction, the positive constants α_e and β_e are defined as

$$\begin{aligned} k_t \sin\theta_t &= k_t (n_1/n_2)\sin\theta_i \\ &\equiv \beta_e \end{aligned} \tag{8-163a}$$

$$\begin{aligned} k_t \cos\theta_t &= \pm j k_t \sqrt{(n_1/n_2)^2 \sin^2\theta_i - 1} \\ &\equiv -j\alpha_e \end{aligned} \tag{8-163b}$$

In Eq. (8-163b), the minus sign is taken for the reason that will become evident shortly. The wave given in Eq. (8-162) is called the evanescent wave. The amplitude decays exponentially in the z-direction, while the phase changes sinusoidally in the x-direction. The evanescent wave propagates along the interface, with the amplitude being attenuated in the direction normal to the interface. This wave is bound to the surface, forming a surface wave.

Substitution of Eq. (8-161) into Eqs. (8-146a) and (8-151a) gives the reflection coefficients for the case of total internal reflection:

$$\Gamma_\perp = \frac{\eta_2\cos\theta_i - j\eta_1\sqrt{(n_1/n_2)^2\sin^2\theta_i - 1}}{\eta_2\cos\theta_i + j\eta_1\sqrt{(n_1/n_2)^2\sin^2\theta_i - 1}} \tag{8-164a}$$

$$\Gamma_\parallel = \frac{j\eta_2\sqrt{(n_1/n_2)^2\sin^2\theta_i - 1} - \eta_1\cos\theta_i}{j\eta_2\sqrt{(n_1/n_2)^2\sin^2\theta_i - 1} + \eta_1\cos\theta_i} \tag{8-164b}$$

In the above equations, $\Gamma_\perp$ is of the form $(a - jb)/(a + jb)$, whereas $\Gamma_\parallel$ is of the form $(-c + jd)/(c + jd)$. Therefore, we always have

$$|\Gamma_\perp| = |\Gamma_\parallel| = 1 \tag{8-165}$$

All the incident energy is reflected by the interface, irrespective of the polarization state of the incident wave.

Example 8-20

The dielectric slab waveguide shown in Fig. 8.20 is infinite in extent in the x- and y-directions, having refractive indices $n_1 > n_2$. Find the maximum angle of incidence θ_m for total internal reflection at the interface between the core and cladding.

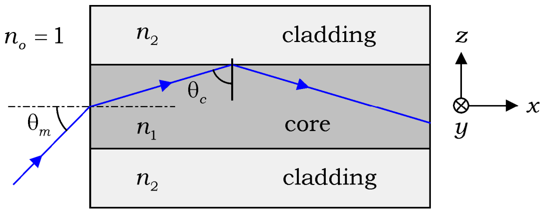


Fig. 8.20 A dielectric slab waveguide.

Solution

Applying Snell's law to the entrance surface, we have

$$n_o \sin\theta_m = n_1 \sin\left(90^o - \theta_c\right)$$
$$= n_1 \cos\theta_c$$

The critical angle at the interface between the core and cladding is

$$\sin\theta_c = n_2 / n_1$$

Combining the two equations, we have

$$n_o \sin\theta_m = n_1\sqrt{1-\left(n_2/n_1\right)^2}$$

The maximum angle of incidence is therefore

$$\theta_m = \sin^{-1}\left[\frac{1}{n_o}\sqrt{n_1^2 - n_2^2}\right]$$

The wave incident at an angle less than θ_m is guided through the dielectric slab by successive total internal reflections at the interface between the core and cladding.

Exercise 8.25

Find θ_c for a diamond of $n = 2.42$ placed in free space.

Ans. $\theta_c = 24.4^o$.

Exercise 8.26

Distinguish the evanescent wave in the second medium of a dielectric-dielectric interface from the evanescent wave penetrating into a good conductor.

Review Questions with Hints

RQ 8.32	What is the total internal reflection?	[Fig.8.20]
RQ 8.33	Define the critical angle.	[Eq.(8-160)]
RQ 8.34	State the evanescent wave in the second medium.	[Eq.(8-162)]

8.3.4 Reflectance and Transmittance

As was discussed in Section 8-2.2, the time-average power per unit area is defined by the Poynting vector $\langle \mathbf{S} \rangle$. When we are concerned with electromagnetic power reaching a given surface, the time-average power per unit area is also called the radiant flux density, or irradiance *I*, and expressed as

$$I \equiv \langle \mathbf{S} \rangle = \frac{1}{2}\mathrm{Re}\left[\mathbf{E}\times\mathbf{H}^*\right] = \mathrm{Re}\left[\frac{E_o^2}{2\eta^*}\right] \qquad [\mathrm{W/m^2}] \qquad (8\text{-}166)$$

where E_o is the amplitude of the electric field intensity, η is the intrinsic impedance of the medium, and $*$ stands for complex conjugate. In a lossless and nonmagnetic medium, the irradiance is simply written as

$$\boxed{I = \frac{1}{2}\sqrt{\frac{\varepsilon}{\mu_o}}E_o^2 = \frac{n}{2\eta_o}E_o^2} \qquad [\mathrm{W/m^2}] \qquad (8\text{-}167)$$

where n is the refractive index of the medium, and η_o is the intrinsic impedance of free space. The irradiance is measured in units of watts per square meter $[\mathrm{W/m^2}]$.

Consider Fig. 8.21, in which the area *A* represents the intersection of the incident, reflected, and transmitted waves of a finite cross section with a given interface. In view of the area *A* in the interface, the cross sections of the three waves are given by $A\cos\theta_i$, $A\cos\theta_r$, and $A\cos\theta_t$, respectively. If the radiant flux densities of the three waves are I_i, I_r, and I_t, respectively, the total powers carried by the incident, reflected, and transmitted waves are $I_i A\cos\theta_i$, $I_r A\cos\theta_r$, and $I_t A\cos\theta_t$, respectively.

The reflectance *R* is defined as the ratio of the reflected power to the incident power, that is,

$$\boxed{R = \frac{I_r A\cos\theta_r}{I_i A\cos\theta_i} = \left(\frac{E_o^r}{E_o^i}\right)^2 = \Gamma^2} \qquad (8\text{-}168)$$

where Γ is the reflection coefficient, and $\theta_r = \theta_i$, from the law of reflection. We note that the reflectance is simply equal to the square of the reflection coefficient.

Similarly, the transmittance *T* is defined as the ratio of the transmitted power to the incident power, that is,

$$\boxed{T = \frac{I_t A\cos\theta_t}{I_i A\cos\theta_i} = \left(\frac{n_2\cos\theta_t}{n_1\cos\theta_i}\right)\tau^2} \qquad (8\text{-}169)$$

where n_1 and n_2 are the refractive indices of the incident and transmitting media, respectively.

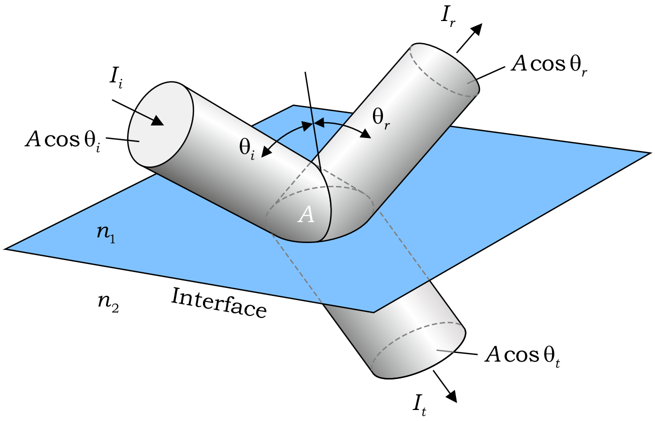


Fig. 8.21 Three waves of a finite cross section at an interface.

According to the law of conservation of energy, the sum of the powers of the reflected and transmitted waves is equal to that of the incident wave, that is,

$$I_i A\cos\theta_i = I_r A\cos\theta_r + I_t A\cos\theta_t \tag{8-170}$$

Dividing both sides of Eq. (8-170) by $I_i A\cos\theta_i$, we have

$$1 = \frac{I_r}{I_i} + \frac{I_t\cos\theta_t}{I_i\cos\theta_i} \tag{8-171}$$

Substitution of Eqs. (8-168) and (8-169) into Eq. (8-171) leads to

$$\boxed{1 = R + T} \tag{8-172}$$

The sum of the reflectance and transmittance is always one, irrespective of the polarization state of the incident wave and the losses in the material media. Since the reflectance is always given by $R = |\Gamma|^2$ at the interface, regardless of the polarization state and material media, the transmittance can be obtained from the relation $T = 1 - |\Gamma|^2$.

Example 8-21
With reference to the three waves expressed by Eqs. (8-153a), (8-153b), and (8-153c) in **Example 8-18**,
(a) find the power of each wave per unit area of the interface, and
(b) verify the law of conservation of energy at the interface.

Solution

(a) Incident power per unit area of the interface is

$$I_i \cos\theta_i = \frac{1}{2}\sqrt{\frac{\varepsilon_o}{\mu_o}}\left(E_o^i\right)^2 \cos\theta_i = \frac{377}{2}\left[\left(\frac{12}{\sqrt{13}}\right)^2 + \left(\frac{8}{\sqrt{13}}\right)^2\right]\cos 33.7^o$$
$$= 2,509\,[\mathrm{W}]$$

Reflected power per unit area of the interface is

$$I_r \cos\theta_r = \frac{1}{2}\sqrt{\frac{\varepsilon_o}{\mu_o}}\left(E_o^r\right)^2 \cos\theta_r = \frac{377}{2}\left[(0.50)^2 + (0.33)^2\right]\cos 33.7^o$$
$$= 56\,[\mathrm{W}]$$

Transmitted power per unit area of the interface is

$$I_t \cos\theta_t = \frac{1}{2}\sqrt{2.25}\sqrt{\frac{\varepsilon_o}{\mu_o}}\left(E_o^t\right)^2 \cos\theta_t = 1.5\frac{377}{2}\left[(2.84)^2 + (1.13)^2\right]\cos 21.7^o$$
$$= 2,454\,[\mathrm{W}].$$

(b) The incident power is $2,509\,[\mathrm{W}]$, whereas the sum of the reflected and transmitted powers is calculated as $2,510\,[\mathrm{W}]$. Ignoring floating point errors, the law of conservation of energy is satisfied.

Exercise 8.27

If a uniform plane wave with parallel polarization is incident on an interface at Brewster angle, find $\tau_{||}$ and $T_{||}$ in terms of the refractive indices of the two media.

Ans. $\tau_{||} = n_1 / n_2$, and $T_{||} = 1$.

Review Questions with Hints

RQ 8.35 Define the irradiance. [Eq.(8-166)]

RQ 8.36 What are the reflectance and transmittance. [Eq.(8-168)(8-169)]

RQ 8.37 Express the law of conservation of energy in terms of the reflectance and transmittance at the interface. [Eq.(8-172)]

8.4 Waves in Dispersive Media

According to the atomic model of a material, a dielectric may be regarded as an assemblage of discrete atoms placed at the equivalent lattice points in free space. When an electromagnetic wave of a frequency ω is introduced into the dielectric, the electric field induces electric dipole moments, vibrating at the same frequency ω, in the material. The electric dipoles generate secondary wavelets, which oscillate with the same frequency and propagate with the same velocity as the incident wave. However, the damping of the electric dipole gives rise to a phase lag or lead of the secondary wave with respect to the primary wave; the sum of the primary and secondary waves constitutes the internal wave in the material. As the wave propagates through the material, the phase delay accumulates so that the phase velocity in the material is different from that in free space. The damping of the electric dipole depends on the frequency. Thus, the phase velocity in the material depends on the frequency, which is called the dispersion.

Until now, we limited our discussion to a uniform plane wave of a single frequency, traveling with a single phase velocity. If a signal wave is to transport meaningful information, it should be finite in time dimension such as a short digital pulse. A wavepacket is a wave of a finite extent in both space and time dimensions, which can be shown to consist of a band of frequencies, by making use of Fourier analysis. In a dispersive medium, the different frequency components travel with different phase velocities, and a broadening of the wavepacket will result, in general. The wavepacket propagates with a group velocity, which is the velocity of propagation of energy.

We now consider a wavepacket, resulting from interference between two harmonic waves of equal amplitudes but slightly different angular frequencies $\omega^- = \omega_o - \Delta\omega$ and $\omega^+ = \omega_o + \Delta\omega$, propagating in free space. The two waves will have slightly different phase constants $\beta^- = \beta_o - \Delta\beta$ and $\beta^+ = \beta_o + \Delta\beta$, respectively. Assuming that the waves are polarized in the x-direction, and propagate in the z-direction, the total electric field intensity is written as

$$\begin{aligned}\mathscr{E}(z,t) &= \mathbf{a}_x E_o \cos\left(\omega^- t - \beta^- z\right) + \mathbf{a}_x E_o \cos\left(\omega^+ t - \beta^+ z\right) \\ &= \mathbf{a}_x 2E_o \cos\left(\Delta\omega t - \Delta\beta z\right)\cos\left(\omega_o t - \beta_o z\right)\end{aligned} \tag{8-173}$$

where we use

$$\Delta\omega = \frac{1}{2}\left(\omega^+ - \omega^-\right) \tag{8-174a}$$

$$\Delta\beta = \frac{1}{2}\left(\beta^+ - \beta^-\right) \tag{8-174b}$$

$$\omega_o = \frac{1}{2}\left(\omega^+ + \omega^-\right) \tag{8-174c}$$

$$\beta_o = \frac{1}{2}\left(\beta^+ + \beta^-\right) \tag{8-174d}$$

Equation (8-173) shows that the electric field intensity of the wavepacket rapidly oscillates with an angular frequency ω_o, while its envelope slowly varies with an angular frequency $\Delta\omega$ (see Fig. 8.22(b)). As in the case of a uniform plane wave, the phase velocity of the wavepacket is the velocity of a point of constant phase, or a crest of the electric field intensity curve, for instance. We can obtain the phase velocity from the ratio between the traveled distance Δz and the elapsed time

Δt, which are required of an observer moving with the crest. The phase velocity is therefore

$$\boxed{\upsilon_p = \frac{\omega_o}{\beta_o}} \qquad [\text{m/s}] \qquad (8\text{-}175)$$

In contrast to the phase velocity, which is the velocity of a wave crest, the group velocity is the velocity of the envelope of the wavepacket. By following the same procedure as for the phase velocity, we obtain the group velocity from the ratio between the traveled distance Δz and the elapsed time Δt, which are required of an observer moving with the peak of the envelope. From Eq. (8-173) we see that the peak of the envelope is defined by $(\Delta\omega t - \Delta\beta z) = 0$. By taking the limit as $\Delta\omega \to 0$ and $\Delta\beta \to 0$, the group velocity is expressed as

$$\boxed{\upsilon_g = \frac{d\omega}{d\beta}} \qquad [\text{m/s}] \qquad (8\text{-}176)$$

The group velocity is the velocity of the modulated amplitude of the wavepacket, and the velocity of propagation of the time-average energy of the wave. On the other hand, the phase velocity is the velocity of the wavefront, which has an important role in describing interference of waves(see Fig. 8.22). The phase and group velocities are equal for a uniform plane wave of a single frequency.

Although the phase and group velocities both depend on frequency in a dispersive medium, they behave differently in different frequency ranges. In the region of the normal dispersion, in which the refractive index of the medium, n, increases with frequency ω, the group velocity υ_g is smaller than the phase velocity υ_p. On the other hand, in the region of the anomalous dispersion, in which the refractive index decreases with frequency, υ_g is always larger than υ_p.

The relationship between υ_g and υ_p can be explained qualitatively by use of Fig. 8.22, in which three points *A*, *B*, and *C* represent three crests of the wave of a higher frequency ω^+. Let us assume that we move with these points at the velocity υ_p^+. In the region of normal dispersion in which $n^+ > n^-$, the wave of a lower frequency ω^- moves faster than the wave of ω^+. Thus, as time increases, the peak of the envelope, or the point of overlap of the two crests, appears to move from point *B* to point *A*, to an observer moving with the wave of ω^+ (see Fig. 8.22(a)). Accordingly, υ_g is smaller than υ_p in the region of normal dispersion. On the other hand, in the region of anomalous dispersion in which $n^+ < n^-$, the wave of ω^+ moves faster than the wave of ω^-. As time increases, the peak of the envelope appears to move from point *B* to point *C*, to the observer moving with the wave of ω^+. Accordingly, υ_g is larger than υ_p in this case.

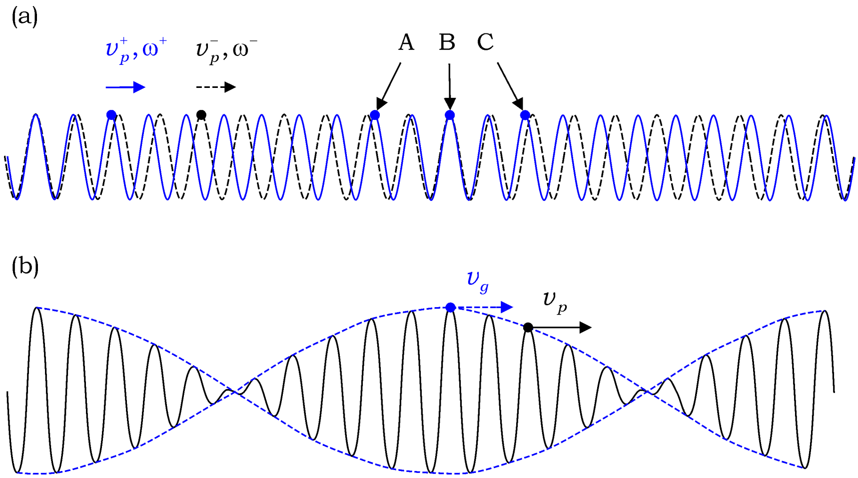


Fig. 8.22 Interference of two waves of slightly different frequencies.

In Fig. 8.23, the $\omega-\beta$ diagram is plotted in a dispersive medium for the case of (a) normal dispersion, and (b) anomalous dispersion. The slope of the line drawn from the origin to a point on the curve represents the phase velocity of the wave at a given frequency ω_o. On the other hand, the tangent to the curve represents the group velocity of the wavepacket at a mean frequency ω_o.

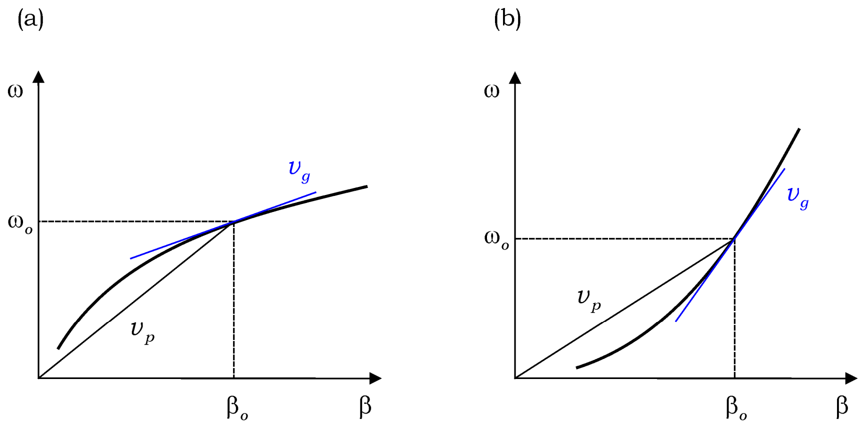


Fig. 8.23 $\omega-\beta$ diagram of a dispersive medium (a) normal dispersion (b) anomalous dispersion.

Example 8-22
With reference to the wavepacket expressed by Eq. (8-173), propagating in free space, find
(a) $\mathscr{H}(z,t)$, and
(b) $\langle \mathbf{S} \rangle$.

Solution

(a) With the help of the relations $|\mathscr{H}| = |\mathscr{E}|\sqrt{\varepsilon_o / \mu_o}$ and $\mathbf{a}_k = \mathbf{a}_E \times \mathbf{a}_H$, we obtain

$$\mathscr{H}(z,t) = \mathbf{a}_y 2E_o\sqrt{\varepsilon_o / \mu_o}\cos(\Delta\omega t - \Delta\beta z)\cos(\omega_o t - \beta_o z) \tag{8-177}$$

The temporal period of the phase is given by $\tau_1 = 2\pi / \omega_o$, while that of the envelope is given by $\tau_2 = 2\pi / \Delta\omega$, where $\tau_1 << \tau_2$.

(b) Poynting vector is

$$\begin{aligned}\mathbf{S} &= \mathscr{E} \times \mathscr{H} \\ &= \mathbf{a}_z 4E_o^2\sqrt{\varepsilon_o / \mu_o}\cos^2(\Delta\omega t - \Delta\beta z)\cos^2(\omega_o t - \beta_o z)\end{aligned} \tag{8-178}$$

Time-average power density is

$$\langle \mathbf{S} \rangle = \frac{1}{T}\int_{-T/2}^{T/2} \mathbf{S}\, dt \qquad : \ \tau_1 << T << \tau_2 \tag{8-179}$$

Inserting Eq. (8-178) into Eq. (8-179), we obtain

$$\langle \mathbf{S} \rangle = \mathbf{a}_z 2E_o^2\sqrt{\varepsilon_o / \mu_o}\cos^2(\Delta\omega t - \Delta\beta z) \tag{8-180}$$

It is evident from Eq. (8-180) that the time-average power density of the wavepacket propagates with the group velocity $\Delta\omega / \Delta\beta$.

Exercise 8.28

The refractive index of a dispersive material is given by $n = n_o(1 + \omega / \omega_o)$. Find the phase and group velocities at $\omega = \omega_o$.

Ans. $\upsilon_p = c / 2n_o$ and $\upsilon_g = c / 3n_o$ (c, speed of light in vacuum).

Review Questions with Hints

RQ 8.38	State dispersion in words.	[Fig.8.23]
RQ 8.39	Define group velocity.	[Eq.(8-176)]
RQ 8.40	Define the phase and group velocities in $\omega - \beta$ diagram.	[Fig.8.23]

Problems

8-1 Given the instantaneous electric field $\mathscr{E}$ in the material of the intrinsic impedance η, find the electric and magnetic field phasors **E** and **H**:

(a) $\mathscr{E} = \mathbf{a}_x E_o e^{-\alpha z}\sin(\omega t - \beta z + \pi / 4)$ in a material with $\eta = \eta_o e^{j\pi/4}$.

(b) $\mathscr{E} = -\mathbf{a}_x E_o e^{-\alpha z}\cos(\omega t + \beta z)$ in a material with $\eta = \eta_o e^{-j\pi/4}$.

(c) $\mathscr{E} = \mathbf{a}_y E_o \cos(kz)\cos(\omega t)$ in free space.

8-2 For the electric field $\mathscr{E}$ as given in **Problem 8-1**, find the time-average power density $\langle \mathbf{S} \rangle$.

8-3 In free space, the electric field phasor is given by $\mathbf{E} = 6\mathbf{a}_x e^{-j3y-j4z}$ [V/m]. Find
(a) direction of propagation, wavelength, and frequency,
(b) magnetic field phasor **H**, and
(c) time-average power through the area defined by $0 \le x \le 1$[m], $0 \le y \le 2$[m], and $z = 0$.

8-4 A uniform plane wave of a frequency 200[MHz] propagates in a lossless medium($\varepsilon_r = 12$ and $\mu_r = 5$). Find (a) β, (b) λ, (c) v_p, and (d) η.

8-5 Given the magnetic field phasor $\mathbf{H} = (10 - j2)(3\mathbf{a}_y + j5\mathbf{a}_z)e^{-j20x}$ in a lossless medium with $\varepsilon_r = 2.5$ and $\mu_r = 4$, find the electric field phasor **E**.

8-6 At a point a in free space, the electric field intensity $\mathscr{E}$ varies sinusoidally in time, with an amplitude 5.3[V/m], along the direction of a vector $\mathbf{A} = (2\mathbf{a}_x - \sqrt{3}\,\mathbf{a}_y)$. Point b is a nearest point at which $\mathscr{E}$ oscillates with the same phase as point a. The distance vector from a to b is given by $\mathscr{R} = \sqrt{3}\,\mathbf{a}_x + 2\mathbf{a}_y + 3\mathbf{a}_z$ [m]. Find $\mathscr{E}$ everywhere.

8-7 An electromagnetic plane wave is in general represented by the electric field phasor $\mathbf{E} = \mathbf{a}_E E_o e^{-j\mathbf{k}\cdot\mathbf{r}}$, where $\mathbf{a}_E$ is a unit vector, **k** is the wavevector, and **r** is the position vector. From the phasor form of Maxwell's equations, derive the following relations:
(a) $\mathbf{k} \times \mathbf{E} = \omega\mu\mathbf{H}$
(b) $\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) = -\omega^2\mu\varepsilon\mathbf{E}$
(c) $\mathbf{k} \cdot \mathbf{E} = 0$
(d) $\mathbf{k} \cdot (\mathbf{k} \times \mathbf{E}) = 0$

8-8 Given the electric field phasor $\mathbf{E} = E_o \rho^{-1} e^{-jkz}\,\mathbf{a}_\rho$ in cylindrical coordinates, where $k = \omega\sqrt{\mu_o\varepsilon_o}$, show that it represents an electromagnetic wave propagating in free space by using (a) Maxwell's equations, and (b) Helmholtz equation. (c) Find the magnetic field phasor **H**.

8-9 Given the electric field phasor $\mathbf{E} = \left(E_o\mathbf{a}_x + jE_o\mathbf{a}_y\right)e^{+jkz}$ in free space, determine (a) propagation direction, (b) **H**, (c) $\mathscr{E}$, and (d) polarization state.

8-10 Resolve the linearly polarized wave $\mathbf{E} = \left(E_{ox}\,\mathbf{a}_x + E_{oy}\,\mathbf{a}_y\right)e^{-jkz}$ into two circularly polarized waves.

8-11 When two circularly polarized waves with $\mathbf{E}_R = (E_o\,\mathbf{a}_x - jE_o\,\mathbf{a}_y)e^{-jkz}$ and $\mathbf{E}_L = (E_o\,\mathbf{a}_x + jE_o\,\mathbf{a}_y)e^{-jkz}$ pass through a thin wave plate at the same time, the phase of $\mathbf{E}_R$ is increased by θ[rad], while there no change in $\mathbf{E}_L$. Determine the polarization state of the combined wave, $\mathbf{E}_R + \mathbf{E}_L$, (a) before, and (b) after the wave plate.

8-12 A linearly polarized wave, $\mathbf{E} = E_o\,\mathbf{a}_y e^{-jkz}$, is incident on a quarter-wave plate whose slow axis is rotated by 30^o with respect to the x-axis as shown in Fig. 8.24. The electric field component parallel to the slow axis is delayed by 90^o compared with that parallel to the fast axis, which is normal to the slow axis, after the plate. Determine the polarization state of the output wave.

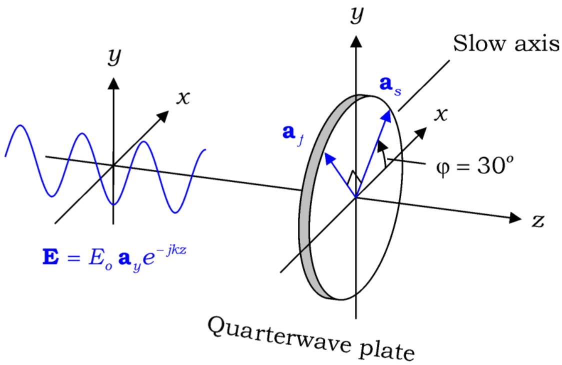


Fig. 8.24 A quarter-wave plate (Problem 8-12).

8-13 A 100[MHz] electromagnetic plane wave is represented by the electric field phasor $\mathbf{E} = 2.4\,\mathbf{a}_x e^{-0.1z} e^{-j6z}$ in a nonmagnetic dielectric. Find ε_r', ε_r'', and **H**.

8-14 A uniform plane wave propagates in the +z-direction in a lossy dielectric for which $\varepsilon'_r = 2.53$, $\mu = \mu_o$, and the loss tangent is 0.03. At a point on the $z = 0.5$[m] plane, the electric field intensity varies sinusoidally in time as $\mathscr{E}(z = 0.5, t) = 400\,\mathbf{a}_x \sin(2\pi \times 10^8 t - \pi/8)$. Determine **E** in the material.

8-15 A uniform plane wave of an angular frequency, $\omega = 3\times 10^9$[rad/s], propagates in the +z-direction in an infinite, nonmagnetic, lossy dielectric. The power is reduced by 10[dB] over a distance of 32.3[cm], and **H** always lags behind **E** by 0.22[rad]. Determine $\hat{\eta}$ of the medium.

8-16 Two nonmagnetic dielectrics($\mu_r = 1$) have the same dielectric constant($\varepsilon_r = 2.25$), but different conductivities($\sigma = 0$ and $\sigma = 50$ [S/m]). If a wave of a frequency 100[MHz] propagates in each dielectric, compare the values of (a) β, (b) λ, and (c) v_p of the two waves.

8-17 An electromagnetic plane wave propagating in a nonmagnetic dielectric has the instantaneous electric and magnetic field intensities given as

$\mathscr{E}(z,t) = \mathbf{a}_x 967.1 e^{-\alpha z} \cos\left(1.5\times10^9 t - \beta z\right)$ and

$\mathscr{H}(z,t) = \mathbf{a}_y 10 e^{-\alpha z} \cos\left(1.5\times10^9 t - \beta z - 0.7194\right)$.

Determine (a) η, (b) ε_r, (c) σ, (d) α, and(e) β.

8-18 A nonmagnetic dielectric supports an electromagnetic plane wave of $\alpha = 20.32[\text{Np/m}]$ and $\beta = 38.85[\text{rad/m}]$ at a frequency $f = 1$[GHz]. Find (a) loss tangent, (b) dielectric constant, and (c) intrinsic impedance.

8-19 A lossy dielectric of $\varepsilon_r = 1.5$ and $\mu_r = 1$ support an electromagnetic plane wave of $\beta = 120$ [rad/m] at a frequency 50[MHz]. Find (a) loss tangent, and (b) σ.

8-20 An electromagnetic wave with $\mathbf{E} = \mathbf{a}_x 300 e^{-0.21z - j(2.2z)}$[V/m] propagates in a medium of $\hat{\eta} = 220 + j21[\Omega]$. Find (a) expression for $\langle \mathbf{S} \rangle$. Next, obtain the time-average ohmic power-loss per unit volume at $z = 0.5$[m] by use of (b) $-\nabla \bullet \langle \mathbf{S} \rangle$, and (c) $(1/2)\text{Re}\left[\mathbf{E} \bullet \mathbf{J}^*\right]$.

8-21 A typical laser pointer has an intensity of $1[\text{mW/mm}^2]$. By assuming the laser light to be a monochromatic plane wave, find the electric field intensity.

8-22 The power density of a 30[MHz]-plane wave is reduced by 10[dB] over a distance of 0.15[mm] in a good conductor. Assuming a nonmagnetic material, find (a) σ, and (b) $\hat{\eta}$.

8-23 For a 100[MHz]-plane wave, propagating in the +*z*-direction in a good conductor of $\sigma = 10^7$[S/m] and $\mu = \mu_o$, find (a) δ, (b) $\hat{\eta}$, and (c) v_p.

8-24 An air-gap coaxial cable consists of an inner conductor of radius a and an outer conductor of inner radius b. By assuming the skin depth δ to be much smaller than the radius a and the thickness of the outer conductor, find the ac-resistance per unit length of the coaxial cable.

8-25 A uniform plane wave is normally incident on an interface between two lossless media. The power densities of the reflected and transmitted waves are the same. Find (a) standing wave ratio, and (b) ratio η_2 / η_1.

8-26 A uniform plane wave is normally incident on the interface at $z = 0$, and produces a standing wave with $S = 4.0$ in air $(z < 0)$. The power density transmitted into the nonmagnetic, lossless dielectric in the region $z \geq 0$ is $5[\text{W/m}^2]$. Find the amplitude of the incident wave.

8-27 A red light of wavelength $630[\text{nm}]$ forms a uniform plane wave in free space, and is normally incident on the planar surface of a good conductor with $\sigma = 5\times10^7[\text{S/m}]$ and $\mu = \mu_o$.

(a) Find $|\Gamma|$.

(b) Find the percentage change in $|\Gamma|$ when a violet light of $350[\text{nm}]$ is incident on the same conductor.

8-28 The $z=0$ plane coincides with the front surface of an aluminum foil ($\sigma = 3.82\times10^7[\text{S/m}]$ and $\mu = \mu_o$), which is one skin depth thick ($t=\delta$). In free space ($z<0$), a blue light of wavelength $0.5[\mu\text{m}]$ forms a uniform plane wave whose electric field intensity is given by $\mathscr{E} = 100\,\mathbf{a}_x \cos(2\pi ft)[\text{V/m}]$ at $z=0$. Inside the foil, find (a) frequency *f*, (b) skin depth δ, and (c) complex intrinsic impedance $\hat{\eta}$.

8-29 With reference to the blue light normally incident on the aluminum foil in **Problem 8-28**, by taking account of the reflections at the front and rear surfaces, but ignoring multiple internal reflections inside the foil, find the electric field phasor **E** in the region $z>t$ (after the foil).

8-30 A uniform plane wave with $\mathbf{E}^i = -300\mathbf{a}_y e^{-j(10x+15z)}$ propagates in free space $(z<0)$, and impinges on the surface of a dielectric($n=1.30$) at $z=0$. Find

(a) angle of incidence,

(b) $\mathbf{E}^r$ and $\mathbf{E}^t$, and

(c) R and T.

8-31 Given a uniform plane wave with $\mathbf{E}^i = 100(-4\mathbf{a}_x + 3\mathbf{a}_z)e^{-j(9x+12z)}$ in free space $(z<0)$, find $\mathbf{E}^r$ and $\mathbf{E}^t$ due to the dielectric($n=1.30$) occupying the region $z\ge 0$.

8-32 Consider a wave with $\mathbf{E}^i = (2\mathbf{a}_x - \mathbf{a}_z)30\sqrt{5}\,e^{-j(10x+20z)}$, which propagates in free space $(z<0)$, and is reflected by a perfect conductor occupying the region $z\ge 0$. Find

(a) $\mathbf{H}^i$, $\mathbf{E}^r$, and $\mathbf{H}^r$,

(b) total **E** and total **H** on the $z=0$ plane, and

(c) $\mathbf{J}_s$ induced on the $z=0$ plane.

8-33 Given that $\mathbf{E}^i = (2\mathbf{a}_x + 2\mathbf{a}_y - 4\mathbf{a}_z)e^{-j(4x+8y+6z)}$ in free space $(z<0)$, resolve the wave into two components, having the electric field vectors perpendicular and parallel to the $z=0$ plane, respectively.

8-34 With reference to $\mathbf{E}^i$ in **Problem 8-33**, find the electric field phasor of the reflected wave if the $z=0$ plane is the surface of a perfect conductor occupying the region $z>0$.

8-35 A uniform plane wave with $\mathbf{E}^i = 4\mathbf{a}_E e^{-j\left(x+\sqrt{3}z\right)}$ propagates in free space in the region $z<0$, and impinges on a lossless dielectric of $\varepsilon_r = 2.25$ and

$\mu_r = 1$, occupying the region $z \geq 0$. The incident wave is known to have a perpendicular-polarization component of an amplitude 3[V/m].

(a) Resolve $\mathbf{E}^i$ into perpendicular- and parallel-polarization components.

(b) Find expression for $\mathbf{E}^r$.

8-36 The $z = 0$ plane serves as an interface between free space $(z < 0)$ and a fused quartz $(z \geq 0)$ of $n_2 = 1.46$. Consider a uniform plane wave with a wavelength $\lambda_o = 0.633$[nm] and a power density $\langle \mathbf{S} \rangle = 10$[mW/mm^2] in free space. If the wave is incident on the interface at Brewster angle and is totally transmitted into the crystal, find

(a) wavevectors $\mathbf{k}_i$ and $\mathbf{k}_t$,

(b) unit polarization vectors of the incident and transmitted waves, and

(c) expressions for $\mathscr{E}^i$ and $\mathscr{E}^t$.

8-37 A circularly polarized wave with $\mathbf{E}^i = (3\mathbf{a}_x + j5\mathbf{a}_y - 4\mathbf{a}_z)e^{-j10^6(4.8x+3.6z)}$ propagates in free space($z < 0$) and is reflected from the surface of a dielectric($n = 1.52$) at $z = 0$. Determine

(a) polarization state of $\mathbf{E}^i$,

(b) expression for $\mathbf{E}^r$, and

(c) polarization state of $\mathbf{E}^r$.

8-38 A right-angle prism is made of a crown glass of $n = 1.52$. Determine the range of the angles of incidence over which the internal wave will undergo a total internal reflection at the hypotenuse.

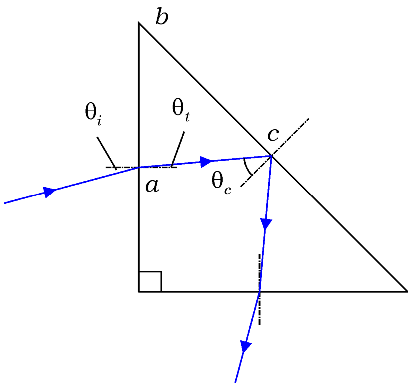


Fig. 8.25 A right-angle prism(Problem 8-38).

8-39 The evanescent wave as given in Eq. (8-162) is perpendicularly polarized such that $\mathbf{a}_E = \mathbf{a}_y$.

(a) Find $\mathbf{H}^t$.

(b) Show that no power is transmitted through the interface.

8-40 A 500[MHz]-plane wave is incident from air upon the surface of a lossy medium with $\sigma = 10[\mathrm{S/m}]$, $\varepsilon = \varepsilon_o$, and $\mu = \mu_o$. Find the ratio between the transmitted and incident powers.

8-41 The normal incidence of a uniform plane wave of $E_o = 30[\mathrm{V/m}]$ onto an interface at $z = 0$ generates a standing wave of $S = 4$ in air($z < 0$). Find the power dissipated in the lossy medium extending from $z = 0$ to $z = \infty$, having a cross section $1[\mathrm{m}] \times 1[\mathrm{m}]$.

8-42 Assuming the conductivity is independent of frequency in a nonmagnetic, good conductor, find (a) phase velocity, and (b) group velocity.

Chapter 9
Transmission Lines

So far most discussion of electromagnetic waves was in a medium of infinite extent, or two semi-infinite media adjoined to each other. By solving the differential wave equation we saw that the electromagnetic waves take the form of a uniform plane wave in such media. The uniform plane wave is an unbounded and unguided wave in the sense that the electric field exists in all space and the electromagnetic energy spreads over the whole space. While an unbounded and unguided wave may be useful for broadcasting radio and TV signals, it is inefficient when used for point-to-point transmission of electromagnetic power or information.

A transmission line may be a much more efficient way of transmitting the electromagnetic power and information from the source to the load, which is composed of two parallel conductors in general. The transmission line can support a transverse electromagnetic (TEM) wave, which is an electromagnetic wave whose electric and magnetic field vectors are both normal to the direction of propagation of the wave. The TEM wave traveling on a transmission line has the same propagation characteristics as those of a uniform plane wave traveling in an infinite medium.

The most common transmission lines are two-wire lines, coaxial cables, and striplines. Examples of two-wire lines are overhead power lines, and twin-lead cables connecting TV sets to the antenna. Coaxial cables find their uses in transmitting TV signals, interconnecting high-frequency precision electronic equipment, and Ethernet. Striplines can be easily fabricated on dielectric substrates by printed-circuit technologies, and used for connecting electronic components in integrated circuits.

To elucidate the difference between a transmission line and an electric circuit, consider an AM radio frequency of $1000[\text{KHz}]$, which corresponds to a wavelength of $300[\text{m}]$ in free space. If a resistor is connected to an AC voltage source of $1000[\text{KHz}]$ through a pair of $10[\text{cm}]$-long copper wires, every point on the wire and the resistor itself can be considered to be in phase in the time domain, because of the long wavelength of the voltage wave. In contrast, if a microwave of $2.45[\text{GHz}]$, having a wavelength $12.2[\text{cm}]$, is guided by a transmission line, the electric field intensity of the wave changes by 180^o in time phase every $6.1[\text{cm}]$ along the transmission line.

We can analyze the characteristics of the wave traveling on a transmission line by use of either electromagnetic theory or circuit theory. Circuit theory deals with voltage and current waves that are simple one-dimensional scalar waves, whereas electromagnetic theory deals with electric and magnetic fields that are vector fields propagating through three-dimensional space, in general. In the present chapter, we use circuit theory for describing the operation of the transmission line.

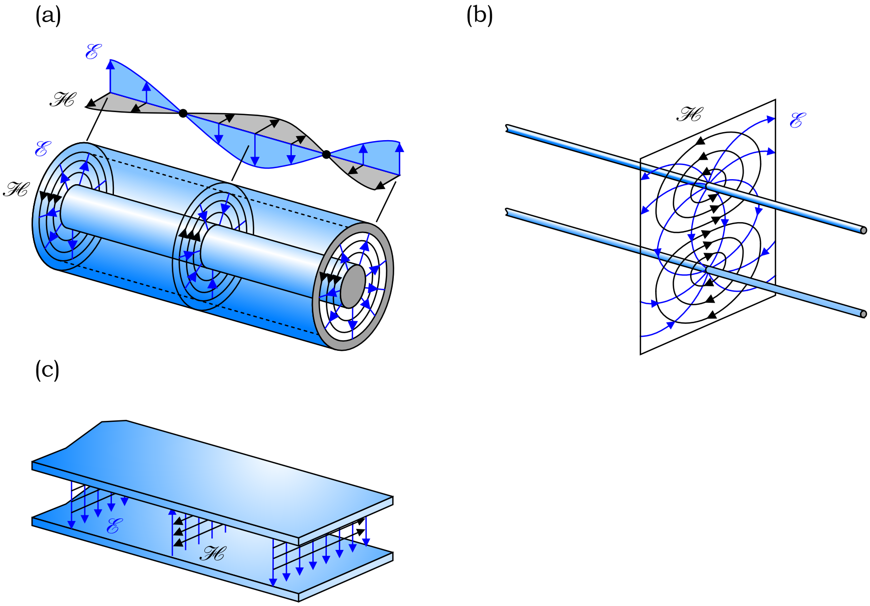


Fig. 9.1 Three common transmission lines: (a) coaxial cable, (b) two-wire line, and (c) parallel-plate line.

Figure 9.1 illustrates the three most common transmission lines, and their electric and magnetic field patterns in the cross sections. We notice that the direction of the electric and magnetic fields reverses at every half wavelength along the line. Let us digress briefly and examine the electromagnetic field pattern between the two conductors of a coaxial transmission line, as shown in Fig. 9.1(a). The exact field pattern can be obtained by solving the differential wave equation and applying boundary conditions. At this point, we will not take the exact steps for the field pattern. Alternatively, we use a simpler approach, and extend the static electric and magnetic fields previously obtained in a coaxial cable to the time-varying case, with the help of the basic wave theory and the symmetry of the given configuration. If the coaxial cable is oriented along the z-axis, the field distribution in the cross section is independent of z under static field conditions, because of the translational symmetry in the z-direction. Under time-varying conditions, although the field vectors vary in time, the field distribution in the cross section is assumed to be the same as that of a static field, which is called the quasi-static field conditions. This holds true if the source current or voltage can be assumed to vary slowly compared with the time-varying electromagnetic field. In this case, the field pattern in the cross section behaves as a profile of the wave in the transmission line. From the Poynting vector $\mathbf{S} = \mathscr{E} \times \mathscr{H}$, we see that the electromagnetic fields propagate in the z-direction, and are guided by the transmission line without loss of energy. The variation of the electric and magnetic fields along the transmission line can be described by a simple cosine curve, in view of the fact that the field distribution in the cross section is independent of z.

9.1 Transmission Line Equations

The transmission line equations are a pair of first-order partial differential equations governing the relation between the voltage and current in a transmission line. For a TEM wave guided by a transmission line, the transverse electric and magnetic fields are related to the time-varying voltage $\mathscr{v}$ and time-varying current $\mathscr{i}$ by

$$\mathscr{v} = -\int \mathscr{E} \cdot d\mathbf{l} \tag{9-1a}$$

$$\mathscr{i} = \oint \mathscr{H} \cdot d\mathbf{l} \tag{9-1b}$$

The voltage and current vary with position on the transmission line as well as with time, and are well explained by waves traveling along the transmission line. Eq. (9-1) shows that a transmission line problem can be solved by means of either electromagnetic waves or voltage and current waves. The voltage and current waves are one-dimensional scalar waves, whereas the electromagnetic waves are three-dimensional vector waves, in general. The former is easier to deal with mathematically. Furthermore, it is preferred when we are concerned with the impedance in a low loss transmission line. However, if a complete analysis of the transmission line is required, the electromagnetic field theory should be employed.

A transmission line can be viewed as an electric circuit that is composed of four circuit elements such as series resistance **R**, inductance **L**, capacitance **C**, and shunt conductance **G**, as shown in Fig. 9.2(a). The series resistance **R** arises from a finite conductivity of the conductors, while the shunt conductance **G** arises from a nonzero conductivity of the insulating material between the conductors. Since these circuit parameters linearly depend on the length of the transmission line, the equivalent electric circuit should be a distributed-parameter network, instead of a lumped-element circuit, as depicted in Fig. 9.2(b). The circuit parameters R, L, C, and G in Fig. 9.2(b) are measures in units of ohms per meter $[\Omega/\mathrm{m}]$, henrys per meter$[\mathrm{H/m}]$, farads per meter$[\mathrm{F/m}]$, and siemens per meter$[\mathrm{S/m}]$, respectively.

As mentioned earlier, if a TEM wave is guided in a transmission line, the distribution of $\mathscr{E}$ and $\mathscr{H}$ in the cross section is the same as that of the static electric and magnetic fields. In view of these, we can obtain the circuit parameters by assuming static electric and magnetic fields in the transmission line.

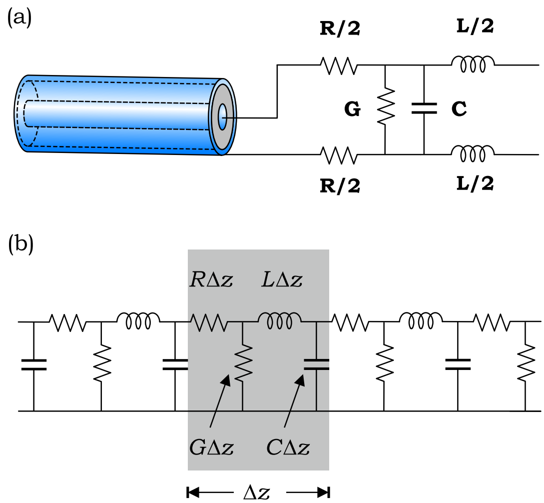


Fig. 9.2 Transmission line and equivalent distributed-parameter network.

To derive the relationship between the voltage and current in the transmission line, let us consider a small segment of length Δz of the transmission line. The segment can be represented by an equivalent circuit as shown in Fig. 9.3, where $R\Delta z$ is the resistance, $L\Delta z$ is the inductance, $C\Delta z$ is the capacitance, and $G\Delta z$ is the conductance. By applying Kirchhoff's voltage law to the circuit, we obtain

$$\mathscr{v}(z,t) = R\Delta z\,\mathscr{i}(z,t) + L\Delta z\frac{\partial \mathscr{i}(z,t)}{\partial t} + \mathscr{v}(z+\Delta z,t)$$

We rewrite the equation as

$$-\frac{\mathscr{v}(z+\Delta z,t) - \mathscr{v}(z,t)}{\Delta z} = R\mathscr{i}(z,t) + L\frac{\partial \mathscr{i}(z,t)}{\partial t} \tag{9-2}$$

Taking the limit of Eq. (9-2) as Δz approaches zero we have

$$\boxed{-\frac{\partial \mathscr{v}(z,t)}{\partial z} = R\mathscr{i}(z,t) + L\frac{\partial \mathscr{i}(z,t)}{\partial t}} \tag{9-3}$$

Next, by applying Kirchhoff's current law to the upper node of the circuit we obtain

$$\mathscr{i}(z,t) = \mathscr{i}(z+\Delta z,t) + G\Delta z\,\mathscr{v}(z+\Delta z,t) + C\Delta z\frac{\partial \mathscr{v}(z+\Delta z,t)}{\partial t}$$

We rewrite the equations as

$$-\frac{i(z+\Delta z,t)-i(z,t)}{\Delta z}=Gv(z+\Delta z,t)+C\frac{\partial v(z+\Delta z,t)}{\partial t} \tag{9-4}$$

Taking the limit of Eq. (9-4) as $\Delta z \to 0$ we have

$$\boxed{-\frac{\partial i(z,t)}{\partial z}=Gv(z,t)+C\frac{\partial v(z,t)}{\partial t}} \tag{9-5}$$

Equations (9-3) and (9-5) constitute the transmission-line equations.

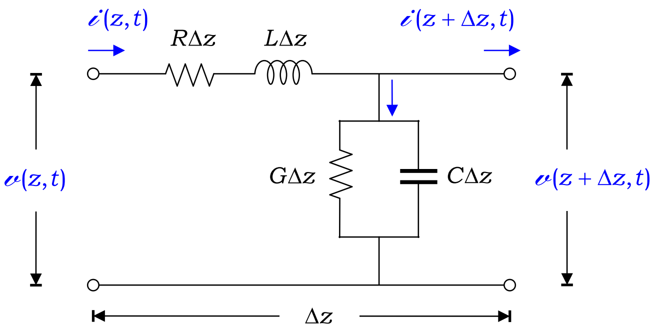


Fig. 9.3 Equivalent circuit for a segment of length Δz of the transmission line.

9.1.1 Phasor Form of Transmission Line Equations

Under sinusoidal steady-state conditions, the voltage and current are given by time-harmonic functions, and thus expressed in terms of phasors as

$$v(z,t)=\text{Re}\left[V(z)e^{j\omega t}\right] \tag{9-6a}$$

$$i(z,t)=\text{Re}\left[I(z)e^{j\omega t}\right] \tag{9-6b}$$

where $V(z)$ and $I(z)$ are voltage and current phasors, respectively. The phasors are complex numbers in general, and depend on the space coordinate z only. Substitution of Eq. (9-6) into Eqs. (9-3) and (9-5) leads to the phasor form of transmission–line equations, that is,

$$\boxed{-\frac{\partial V}{\partial z}=(R+j\omega L)I} \tag{9-7a}$$

$$\boxed{-\frac{\partial I}{\partial z}=(G+j\omega C)V} \tag{9-7b}$$

Next, we take the second derivatives of V and I in Eq. (9-7a) and (9-7b), and replace $\partial V/\partial z$ and $\partial I/\partial z$ on the right-hand side with Eqs. (9-7a) and (9-7b), respectively, as

$$\frac{\partial^2 V}{\partial z^2} = (R + j\omega L)(G + j\omega C)V \tag{9-8a}$$

$$\frac{\partial^2 I}{\partial z^2} = (G + j\omega C)(R + j\omega L)I \tag{9-8b}$$

From Eq. (9-8) we obtain Helmholtz's equations for V and I, that is,

$$\boxed{\frac{\partial^2 V}{\partial z^2} - \gamma^2 V = 0} \tag{9-9a}$$

$$\boxed{\frac{\partial^2 I}{\partial z^2} - \gamma^2 I = 0} \tag{9-9b}$$

Here, the propagation constant γ is defined as

$$\boxed{\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}} \tag{9-10}$$

The voltage and current satisfying Eq. (9-9) form waves traveling on a transmission line, whose propagation property is determined by the propagation constant γ.

9.1.2 Relationship between Parameters

By solving the differential wave equations given in Eq. (9-9) we obtain the voltage and current waves propagating on a transmission line of an infinite extent. The propagation properties of the waves are the same as those of a uniform plane wave in a medium of an infinite extent. If the transmission line is made of a good conductor with a high conductivity, the series resistance R can be neglected in Eq. (9-10). Under this condition, the propagation constant of the transmission line is

$$\gamma = \sqrt{j\omega L(G + j\omega C)} = j\omega\sqrt{LC}\sqrt{1 - j\frac{G}{\omega C}} \qquad \text{(transmission line)} \tag{9-11}$$

If the series resistance is zero, it implies that the electric field should be normal to the surface of the conductor; the tangential would produce an infinite conduction current. Thus the electric field vectors should lie in the transverse plane, meaning that the wave in the transmission line is a TEM wave.

We recall from Eq. (8-77) that the propagation constant of a uniform plane wave propagating in a lossy dielectric of an infinite extent is

$$\gamma = j\omega\sqrt{\mu\varepsilon}\sqrt{\left(1 - j\frac{\sigma}{\omega\varepsilon}\right)} \qquad \text{(dielectric)} \qquad \text{(8-77)(9-12)}$$

It should be noted that σ in Eq. (9-12) is the conductivity of the dielectric.

We recall from Eq. (4-55) that the product of R and C of a two-conductor configuration is equal to the relaxation time constant of the dielectric between the conductors. Alternatively, as given in Eq. (4-56), the ratio between the capacitance and the shunt conductance is equal to the relaxation time constant of the dielectric, that is,

$$\boxed{\frac{C}{G} = \frac{\varepsilon}{\sigma}} \qquad (9\text{-}13)$$

where σ is the conductivity of the dielectric between two conductors. Comparison of Eq. (9-11) and Eq. (9-12), with the aid of Eq. (9-13), leads to

$$\boxed{LC = \mu\varepsilon} \qquad (9\text{-}14)$$

From Eqs. (9-13) and (9-14) we see that the transmission line parameters, such as L, G, and C, are related to the parameters of the dielectric surrounding the conductors, such as μ, ε, and σ. For a known capacitance C of a transmission line, we can obtain the shunt conductance G from Eq. (9-13), and the inductance L from Eq. (9-14), for instance.

Exercise 9.1
Identify the loss tangent in Eq. (9-11).

Ans. $\tan\xi = \dfrac{\varepsilon''}{\varepsilon'} = \dfrac{G}{\omega C}$.

Review Questions with Hints

RQ 9.1 Distinguish between transmission line and electric circuit. [Fig.9.1]
RQ 9.2 Write transmission line equations in instantaneous and phasor forms. [Eqs.(9-3)(9-5)(9-7)]
RQ 9.3 Write Helmholtz's equations for V and I on a transmission line. [Eq.(9-9)]
RQ 9.4 Define propagation constant of a transmission line. [Eq.(9-10)]
RQ 9.5 Under what conditions can a transmission line support a TEM wave? [Eq.(9-11)]
RQ 9.6 State the relationships between parameters of a transmission line and those of the insulating material. [Eqs.(9-13)(9-14)]

9.2 Transmission Line Parameters

The relation between the voltage and current on a transmission line depends on the four distributed line parameters, which are the series resistance R, inductance L, shunt conductance G, and capacitance C. If the series resistance is negligibly small, the electric field lines are confined in the transverse plane of the transmission line, and the guided wave is approximated as a TEM wave. Under these conditions, the electric and magnetic field patterns in the transverse plane are the same as those of static fields. Thus, we can obtain the line parameters by assuming the transmission line to be under static field conditions. We have already obtained some parameters in the previous chapters, and therefore we can determine the rest from the relations given in Eqs. (9-13) and (9-14). In this section, we list the line parameters for the three most common transmission lines.

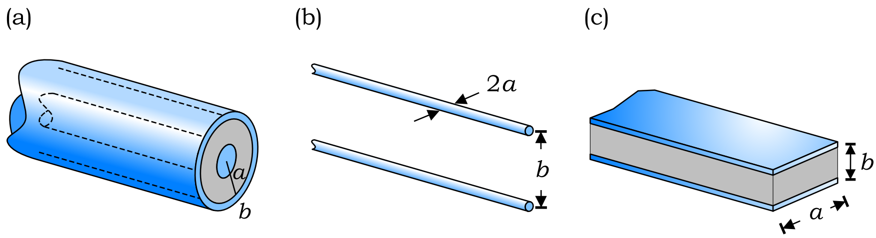


Fig. 9.4 Three most common transmission lines: (a) coaxial line, (b) two-wire line, and (c) parallel-plate line.

9.2.1 Coaxial Transmission Lines

In a coaxial transmission line as shown in Fig. 9.4(a), the inner conductor of radius a is separated from the outer conductor of inner radius b by a dielectric with ε and μ. From Eq. (3-193), the capacitance per unit length of the coaxial transmission line is

$$C = \frac{2\pi\varepsilon}{\ln(b/a)} \qquad \text{[F/m]} \qquad \text{(9-15a)}$$

The inductance per unit length of the line can be either borrowed from Eq. (5-110) or obtained from Eq. (9-14) as

$$L = \frac{\mu}{2\pi}\ln\frac{b}{a} \qquad \text{[H/m]} \qquad \text{(9-15b)}$$

Note that this is the external inductance. The internal inductance can be ignored at high frequencies, because the magnetic field hardly penetrates the conductor due to an extremely short skin depth. By inserting Eq. (9-15a) into Eq. (9-13), the shunt conductance per unit length of the coaxial transmission line is

$$G = \frac{2\pi\sigma}{\ln(b/a)} \qquad \text{[S/m]} \qquad \text{(9-15c)}$$

where σ is the conductivity of the dielectric between two conductors. The series resistance is obtained from the ac-resistance defined by Eq. (8-100) by assuming the total current to be uniform within a skin depth. The cross section of the current path is $\mathcal{S}_i = 2\pi a\delta$ in the inner conductor, whereas it is $\mathcal{S}_o = 2\pi b\delta$ in the outer conductor. The series resistance per unit length of the coaxial transmission line is therefore

$$R = \frac{1}{\sigma_c \mathcal{S}_i} + \frac{1}{\sigma_c \mathcal{S}_o} = \frac{1}{2\pi}\left(\frac{1}{a} + \frac{1}{b}\right)\sqrt{\frac{\pi f \mu_c}{\sigma_c}} \qquad [\Omega / \mathrm{m}] \tag{9-15d}$$

where σ_c and μ_c are the conductivity and permeability of the conductor, not those of the dielectric between the conductors. In Eq. (9-15d), it is assumed that the skin depth is much smaller than the radius of the inner conductor and the thickness of the outer conductor.

9.2.2 Two-Wire Transmission Lines

In a two-wire transmission line as shown in Fig. 9.4(b), two parallel conducting wires of the same radius a are separated by a center-to-center distance of b in a dielectric of ε and μ. From Eq. (3-196), the capacitance per unit length of the two-wire transmission line is

$$C = \frac{\pi\varepsilon}{\cosh^{-1}(b/2a)} \qquad [\mathrm{F/m}] \tag{9-16a}$$

Inserting Eq. (9-16a) into Eq. (9-14) gives the inductance per unit length of the line as

$$L = \frac{\mu}{\pi}\cosh^{-1}(b/2a) \qquad [\mathrm{H/m}] \tag{9-16b}$$

Under the condition $b >> a$, if we use the approximation $\cosh^{-1}(b/2a) \approx \ln(b/a)$, Eq. (9-16b) reduces to the external inductance as given in Eq. (5-110). Next, upon inserting Eq. (9-16a) into Eq. (9-13), the shunt conductance per unit length of the line is

$$G = \frac{\pi\sigma}{\cosh^{-1}(b/2a)} \qquad [\mathrm{S/m}] \tag{9-16c}$$

If we assume a uniform distribution of the total current within a skin depth of the conductor, the cross section of the current path is $\mathcal{S} = 2\pi a\delta$ in the conductor. The series resistance per unit length of the two-wire transmission line is, from Eq. (8-100),

$$R = \frac{2}{2\pi a\delta\sigma_c} = \frac{1}{\pi a}\sqrt{\frac{\pi f \mu_c}{\sigma_c}} \qquad [\Omega / \mathrm{m}] \tag{9-16d}$$

where σ_c and μ_c are those of the conductors. In the above equation, the skin depth is assumed to be much smaller than the conductor's radius a. Notice that the numerator in Eq. (9-16d) contains 2 to account for the effects of the two conductors.

9.2.3 Parallel-Plate Transmission Lines

In a parallel-plate transmission line as shown in Fig. 9.4(c), the two parallel conducting plates of the same width a are separated by a dielectric of thickness b. The capacitance per unit length of the parallel-plate transmission line is, from Eq. (3-187),

$$C = \frac{\varepsilon a}{b} \qquad \text{[F/m]} \qquad (9\text{-}17a)$$

where the fringing effects at the edges are ignored by assuming $b << a$. Upon inserting Eq. (9-17a) into Eq. (9-14), the inductance per unit length of the line is

$$L = \frac{\mu b}{a} \qquad \text{[H/m]} \qquad (9\text{-}17b)$$

Next, Substituting Eq. (9-17b) into Eq. (9-13), the shunt conductance per unit length of the line is

$$G = \frac{\sigma a}{b} \qquad \text{[S/m]} \qquad (9\text{-}17c)$$

If the total current is assumed to be uniform within a skin depth of the conductor, the cross section of the current path in the conductor is given by $S = a\delta$. The series resistance per unit length of the parallel-plate transmission line is, from Eq. (8-100),

$$R = \frac{2}{a\delta\sigma_c} = \frac{2}{a}\sqrt{\frac{\pi f \mu_c}{\sigma_c}} \qquad [\Omega/\mathrm{m}] \qquad (9\text{-}17d)$$

where σ_c and μ_c are those of the conductors. Here, the skin depth is assumed to be much smaller than the thickness of the conducting plate. The numerator in Eq. (9-17d) contains 2 to account for the effect of two conductors.

It should be noted that the inductances given by Eqs. (9-15b), (9-16b), and (9-17b) are the external inductances. At high frequencies, the internal inductances are ignored, because of the short skin depth of the conductor.

Exercise 9.2
Do R, L, G, and C depend on frequency even if ε, μ, and σ of the dielectric are independent of frequency?

Ans. $R \sim \sqrt{f}$.

9.3 Infinite Transmission Lines

The differential wave equation given in Eq. (9-9) determines the behavior of V and I on a transmission line. A general solution of Eq. (9-9) is given by a linear combination of two waves propagating in the opposite directions on the transmission line, namely

$$V(z) = V_o^+ e^{-\gamma z} + V_o^- e^{+\gamma z} \equiv V^+(z) + V^-(z) \tag{9-18a}$$

$$I(z) = I_o^+ e^{-\gamma z} + I_o^- e^{+\gamma z} \equiv I^+(z) + I^-(z) \tag{9-18b}$$

where V_o^+, V_o^-, I_o^+, and I_o^- are the complex amplitudes, which will be determined by the boundary conditions that are specified by the generator, connected at one end, and the load, connected at the other end. In the above equations, $V^+(z)$ and $V^-(z)$ are the phasors of time-harmonic voltage waves propagating in the $+z$- and $-z$-directions, respectively, whereas $I^+(z)$ and $I^-(z)$ are those of current waves propagating in the $+z$- and $-z$-directions, respectively. In an infinitely long transmission line, the + or – wave can propagates without being interrupted by other waves, and without reflection, in the wave path. Thus the two voltage waves in Eq. (9-18a) are independent of each other. The same is true for the two current waves in Eq. (9-18b).

The propagation constant γ is, from Eq. (9-10),

$$\boxed{\begin{aligned}\gamma &= \sqrt{(R + j\omega L)(G + j\omega C)} \\ &\equiv \alpha + j\beta\end{aligned}} \tag{9-19}$$

Here, α is called the attenuation constant measured in nepers per meter[Np/m], and β is called the phase constant measured in radians per meter[rad/m].

The sign convention for V and I on a transmission line is shown in Fig. 9.5. A positive voltage, either $V^+(z_1) > 0$ or $V^-(z_1) > 0$, signifies that the upper conductor is at a higher potential than the lower one at $z = z_1$, regardless of the direction of propagation of the wave. In contrast, the polarity of the current depends on the direction of propagation of the wave. A positive current, $I^+(z_1) > 0$, signifies that the current flows in the $+z$-direction in the upper conductor at $z = z_1$, and a positive current, $I^-(z_1) > 0$, means that the current flows in the $-z$-direction in the upper conductor at $z = z_1$.

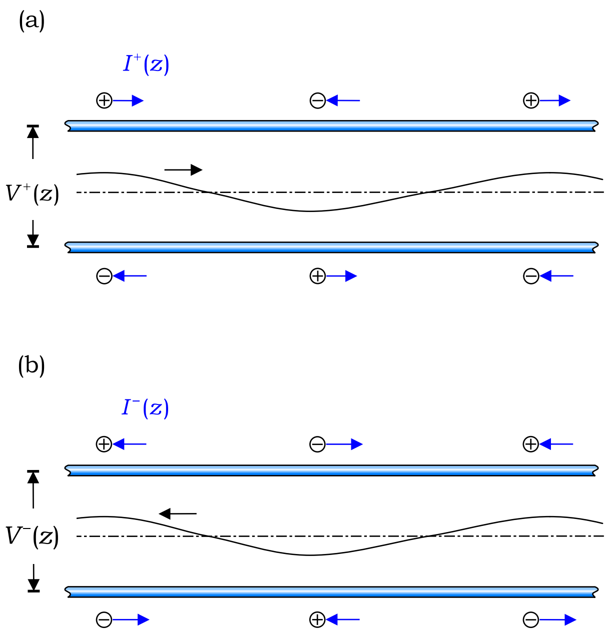


Fig. 9.5 Sign convention for V and I on a transmission line. The directions of propagation of the waves are along (a) $+z$-axis, and (b) $-z$-axis. Blue arrow denotes the direction of current, and a pair of circles along a vertical line shows the polarity of voltage.

Although the V^+ and V^- waves are independent of each other in an infinitely lone transmission line, they are related to the I^+ and I^- waves, respectively. We next find the relations among the complex amplitudes in the general solutions given in Eq. (9-18). Substituting the general solutions expressed by Eq. (9-18) into the transmission-line equation given in Eq. (9-7a), we have

$$\gamma V_o^+ e^{-\gamma z} - \gamma V_o^- e^{+\gamma z} = (R + j\omega L)\left[I_o^+ e^{-\gamma z} + I_o^- e^{+\gamma z}\right] \tag{9-20}$$

By equating the coefficients of the $e^{-\gamma z}$ terms (or the $e^{\gamma z}$ terms) on both sides of Eq. (9-20), we define the characteristic impedance of the transmission line as

$$\boxed{\begin{aligned} Z_o &\equiv \frac{V_o^+}{I_o^+} = -\frac{V_o^-}{I_o^-} \\ &= \sqrt{\frac{R + j\omega L}{G + j\omega C}} \equiv R_o + jX_o \end{aligned}} \qquad [\Omega] \tag{9-21}$$

The general name for the real part of the characteristic impedance is resistance, and that for the imaginary part is reactance. The resistance R_o is measured in ohms. It should not be confused with the series resistance R measured in ohms per meter. The characteristic impedance Z_o of a transmission line is analogous to the intrinsic impedance η of a medium supporting a uniform plane wave. Both Z_o and η represent the ratio between two complex amplitudes: Z_o is the ratio between the complex amplitudes of the voltage and current, whereas η is the ratio between those of the electric and magnetic fields.

The propagation constant γ and the characteristic impedance Z_o are the two distinctive constants of a transmission line; they determine the propagation property of the voltage and current waves on the line. The two constants depend on frequency as well as the line parameters *R*, *L*, *G*, and *C*.

The reciprocal of Z_o is called the characteristic admittance; that is, $Y_o = 1/Z_o$.

Exercise 9.3
Find the total current on a transmission line with $Z_o = 100\,[\Omega]$, when the total voltage is given by $\mathcal{v}(t) = 20\cos(10^7 t - 0.13z) + 10\cos(10^7 t + 0.13z)[\mathrm{V}]$.

Ans. $\mathcal{i}(t) = 0.2\cos(10^7 t - 0.13z) - 0.1\cos(10^7 t + 0.13z)[\mathrm{A}]$.

Exercise 9.4
For the transmission line made of a perfect conductor, (a) is R_o always zero? (b) may R_o be negative?

Ans. (a) No, (b) No.

Exercise 9.5
In a transmission line, may we assume (a) $R = 0 = G$? (b) $L = 0 = C$?
Ans. (a) Yes, (b) No.

9.3.1 Lossless Transmission Lines

A lossless transmission line is made of a perfect conductor $(\sigma_c = \infty)$ and a perfect dielectric $(\sigma = 0)$. A wave can propagate on a lossless transmission line without loss of energy. For a lossless transmission line,

$$R = 0 = G \tag{9-22}$$

$$\gamma = j\beta = j\omega\sqrt{LC} \quad \text{(propagation constant)} \tag{9-23a}$$

$$Z_o = R_o = \sqrt{L/C} \quad \text{(characteristic resistance)} \tag{9-23b}$$

$$\upsilon_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \quad \text{(phase velocity)} \tag{9-23c}$$

The propagation constant γ is purely imaginary, and thus the attenuation constant α is zero, meaning that there is no attenuation of the wave traveling on the lossless line. The characteristic impedance is real, implying that the voltage and current are in phase in time dimension at every point on the lossless line. Substitution of Eq. (9-14) into Eq. (9-23c) shows that the phase velocity of the voltage and current waves on a lossless transmission line is the same as that of a uniform plane wave propagating in the dielectric between the two conductors as if it were infinite in extent.

Exercise 9.6
A lossless transmission line is made of a perfect dielectric of ε and μ. Find (a) phase velocity, and (b) phase constant.

Ans. (a) $\upsilon_p = 1/\sqrt{\mu\varepsilon}$, (b) $\beta = \omega\sqrt{\mu\varepsilon}$.

Exercise 9.7
A lossless line is made of two parallel conducting plates of width a, separated by a dielectric of $\varepsilon = \varepsilon_o\varepsilon_r$ with a thickness b. Find the characteristic impedance.

Ans. $Z_o = 120\pi b/(a\sqrt{\varepsilon_r})$.

9.3.2 Distortionless Transmission Lines

In a lossy transmission line, the attenuation constant is nonzero, and varies with frequency, as we can see from Eq. (9-19). Furthermore, the phase constant is not a linear function of frequency, and therefore the phase velocity also varies with frequency. Such a transmission line is said to be dispersive. If an information-bearing signal is transmitted through a dispersive transmission line, it suffers distortion. This is because a signal usually comprises of a band of frequencies, and the component waves of different frequencies travel with different phase velocities. Moreover, the component waves have different attenuation constants, and thus the signal wave suffers additional distortion. A distortionless transmission line is one that can transmit a signal wave with no change in the shape even if the signal intensity may be reduced. For a distortionless transmission line, the attenuation constant α is independent of frequency, while the phase constant β linearly varies with frequency. We obtain a distortionless transmission line if the line parameters in Eq. (9-19) satisfy

$$\boxed{\frac{R}{L} = \frac{G}{C}} \tag{9-24}$$

Under this condition, the propagation constant in a distortionless line becomes

$$\gamma = \sqrt{LC\left(\frac{R}{L} + j\omega\right)\left(\frac{G}{C} + j\omega\right)} = \sqrt{LC}\left(\frac{G}{C} + j\omega\right) \tag{9-25a}$$
$$\equiv \alpha + j\beta$$

where

$$\alpha = G\sqrt{L/C} = R\sqrt{C/L} \tag{9-25b}$$

$$\beta = \omega\sqrt{LC} \tag{9-25c}$$

Other parameters of the distortionless line are

$$Z_o = \sqrt{\frac{L(R/L + j\omega)}{C(G/C + j\omega)}} = \sqrt{\frac{L}{C}} \quad \text{(characteristic impedance)} \tag{9-25d}$$

$$\upsilon_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \quad \text{(phase velocity)} \tag{9-25e}$$

The distortionless-line parameters, such as β, Z_o, and υ_p, are the same as those of a lossless transmission line. The only difference is that the attenuation constant α is nonzero in the distortionless line.

Example 9-1
A lossless transmission line has a characteristic impedance $100[\Omega]$. When it operates at a frequency 200[MHz], the phase constant is 5[rad/s] on the line. Determine the capacitance and inductance per unit length of the line.

Solution
The phase constant is, from Eq. (9-23a),

$$\beta = \omega\sqrt{LC} = 2\pi \times 2 \times 10^8 \sqrt{LC} = 5$$

The characteristic impedance is, from Eq. (9-23b),

$$Z_o = \sqrt{L/C} = 100$$

The product of β and Z_o gives

$$2\pi \times 2 \times 10^8 L = 500$$

Thus

$$L = 0.40[\mu\text{H/m}]$$
$$C = 40[\text{pF/m}].$$

Example 9-2
A distortionless line has a characteristic impedance $Z_o = 50[\Omega]$ and a propagation constant $\gamma = 0.01 + j4.0[\text{m}^{-1}]$ at frequency $f = 100[\text{MHz}]$. Find R, L, G, C, and υ_p.

Solution
From Eqs. (9-25c) and (9-25d), we have

$$\beta = \omega\sqrt{LC} = 2\pi \times 10^8 \sqrt{LC} = 4.0$$

$$Z_o = \sqrt{L/C} = 50$$

From the above equations we obtain

$$L = 318\,[\mathrm{nH/m}]$$
$$C = 127\,[\mathrm{pF/m}]$$

From Eq. (9-25b) we obtain

$\alpha = G\sqrt{L/C} = GZ_o = 0.01$, and thus
$$G = 200\,[\mu\mathrm{S/m}]$$

From the relation $RC = GL$ we obtain

$$R = \frac{GL}{C} = \frac{200\times10^{-6}\times318\times10^{-9}}{127\times10^{-12}} = 0.50\,[\Omega/\mathrm{m}]$$

Phase velocity is therefore

$$\upsilon_p = \frac{\omega}{\beta} = \frac{2\pi\times10^8}{4} = 1.57\times10^8\,[\mathrm{m/s}].$$

Exercise 9.8
A transmission line is distortionless at $f = 100\,[\mathrm{MHz}]$. Is it still distortionless at $f = 200\,[\mathrm{MHz}]$ if ε, μ, and σ are independent of frequency?

Ans. No.

9.3.3 Power Transmission and Power Loss

The voltage and current waves can jointly transmit power along a transmission line. If the line is oriented along the z-axis, the instantaneous power transmitted in the $+z$-direction is equal to the product of the instantaneous voltage and current of the waves propagating in the $+z$-direction, namely

$$\begin{aligned}\mathscr{P}(z,t) &= \mathscr{v}(z,t)\mathscr{i}(z,t)\\ &= \frac{1}{2}\left[V^+e^{j\omega t} + V^{+*}e^{-j\omega t}\right]\frac{1}{2}\left[I^+e^{j\omega t} + I^{+*}e^{-j\omega t}\right]\\ &= \frac{1}{4}\left[V^+I^{+*} + V^{+*}I^+ + V^+I^+e^{j2\omega t} + V^{+*}I^{+*}e^{-j2\omega t}\right]\\ &= \frac{1}{2}\mathrm{Re}\left[V^+I^{+*} + V^+I^+e^{j2\omega t}\right]\end{aligned} \tag{9-26}$$

where V^+ and I^+ are the phasors of the voltage and current waves, propagating in the $+z$-direction. In the above equation, asterisk $*$ stands for complex conjugate. It should be noted that $\mathscr{P}$ cannot be expressed as $\mathscr{P} = \mathrm{Re}[V^+e^{j\omega t}]\,\mathrm{Re}[I^+e^{j\omega t}]$.

The time-average power transmitted along the transmission line is expressed as

$$\langle \mathscr{P} \rangle = \frac{1}{T}\int_0^T \mathscr{P}(z,t)dt \tag{9-27}$$

where T is the temporal period of the time harmonic function $\mathscr{P}(z,t)$. Inserting Eq. (9-26) into Eq. (9-27), the time-average power is expressed in terms of the voltage and current phasors as

$$\boxed{\langle \mathscr{P} \rangle = \frac{1}{2}\mathrm{Re}\left[V^+ I^{+*}\right]} \qquad \text{[W]} \tag{9-28}$$

Substitution of V^+ and I^+ expressed by Eq. (9-18) into Eq. (9-28) leads to

$$\begin{aligned}\langle \mathscr{P} \rangle &= \frac{1}{2}\mathrm{Re}\left[V_o^+ e^{-(\alpha+j\beta)z} I_o^{+*} e^{-(\alpha-j\beta)z}\right] = \frac{1}{2}\mathrm{Re}\left[Z_o I_o^+ I_o^{+*} e^{-2\alpha z}\right] \\ &= \frac{1}{2}\left|I_o^+\right|^2 R_o e^{-2\alpha z}\end{aligned} \tag{9-29}$$

where we used $V_o^+ / I_o^+ = Z_o = R_o + jX_o$. Rewriting Eq. (9-29), the time-average power at a point on the transmission line is

$$\boxed{\langle \mathscr{P}(z) \rangle = \langle \mathscr{P}(0) \rangle e^{-2\alpha z}} \qquad \text{[W]} \tag{9-30}$$

The time-average power exponentially decreases with distance on the transmission line.

It is customary to express the power loss in decibel such as

$$\boxed{\text{dB power loss} \equiv -10\log_{10}\left[\frac{\mathscr{P}(z)}{\mathscr{P}(0)}\right] = -20\log_{10}\left[\frac{\left|V^+(z)\right|}{\left|V^+(0)\right|}\right]} \tag{9-31}$$

By using $V^+(z) = V_o^+ e^{-(\alpha+j\beta)z}$, Eq. (9-31) can be reduced to a compact form such as

$$\text{dB power loss} = -20\log_{10}\left[e^{-\alpha z}\right] = 8.69\alpha z \tag{9-32}$$

This enables us to convert an attenuation constant α into a dB power loss easily.

Example 9-3

The power reduces by 1.5dB every 100[m] on a transmission line. Find

(a) attenuation constant, and

(b) fraction of power at a distance 1[Km] from the input end.

Solution

(a) From Eq. (9-32) we obtain

$$\alpha = \frac{\text{dB power loss}}{8.69z} = \frac{1.5}{8.69\times 100} = 0.0017\,[\text{Np/m}].$$

(b) From Eq. (9-30) we obtain

$$\frac{\langle \mathscr{P}(1000) \rangle}{\langle \mathscr{P}(0) \rangle} = e^{-2\times 0.0017\times 1000} = 0.033 .$$

Exercise 9.9
Find the three numbers after the decimal point on the right side of Eq. (9-32).

Ans. 686.

Exercise 9.10
If two lines with 2.3dB and 1.8dB power losses are joined together, the joint additionally causes 3.5dB power loss. Find the total dB-power-loss.

Ans. 7.6dB.

Exercise 9.11
What is the dB-power-loss corresponding to 0.1[Np]?

Ans. 0.869dB.

Review Questions with Hints

RQ 9.7 Express general solutions for the time-harmonic V and I on a transmission line. [Eq.(9-18)]
RQ 9.8 Define propagation constant of a transmission line. [Eq.(9-19]
RQ 9.9 Define characteristic impedance of a transmission line. [Eq.(9-21)]
RQ 9.10 Under what conditions does a transmission line become lossless? Write the characteristic parameters of a lossless line. [Eqs.(9-22)(9-23)]
RQ 9.11 Under what conditions does a line become distortionless? Write the characteristic parameters of a distortionless line. [Eqs.(9-24)(9-25)]
RQ 9.12 Express time-average power in terms of phasors V and I. [Eq.(9-28)]
RQ 9.13 Write the relation between the attenuation constant and the dB power loss on a transmission line. [Eq.(9-32)]

9.4 Finite Transmission Lines

A finite transmission line is one that usually connects the source to a load on the other end. The load may be a short circuit, an open circuit, or another transmission line. If the source generates a sinusoidal voltage or current, it induces time-harmonic voltage and current waves on the transmission line. As was discussed earlier, the propagation of the waves is governed by the transmission line equations or the Helmholtz's equations given in Eq. (9-9). The voltage and current on the line can be simply described by the general solutions given in Eq. (9-18) combined with the boundary conditions given at the two ends. While the two waves V^+ and V^- are independent of each other on an infinitely long transmission line, they are related to each other on a finite transmission line, because of the load impedance, so that V^+ represents an incident wave and V^- represents the wave reflected from the load.

Let us consider a finite transmission line of a length ℓ as shown in Fig. 9.6. In the discussion of a finite transmission line, we always assume that the generator is at $z=0$ and the load is at $z=\ell$. The internal impedance of the generator is denoted by Z_g while the load impedance is denoted by Z_L. On the finite transmission line of a characteristic impedance Z_o, general solutions for the voltage and current waves are written as

$$\boxed{V(z) \equiv V^+(z) + V^-(z) = V_o^+ e^{-\gamma z} + V_o^- e^{\gamma z}} \tag{9-33a}$$

$$\boxed{I(z) \equiv I^+(z) + I^-(z) = \frac{V_o^+}{Z_o} e^{-\gamma z} - \frac{V_o^-}{Z_o} e^{\gamma z}} \tag{9-33b}$$

where V_o^+ and V_o^- are the complex amplitudes of the forward and backward voltage waves, respectively, and γ is the propagation constant as given in Eq. (9-19). If the forward wave, with the voltage-current relation given by $V^+(z)/I^+(z) = Z_o$, reaches the load of impedance Z_L, it alone cannot satisfy the voltage-current relation $V_L/I_L = Z_L$ given at the load. The impedance mismatch inevitably leads to a reflection of the incident wave, inducing a backward wave, in such a way that the relation between the total voltage and current at the load are consistent with the load impedance.

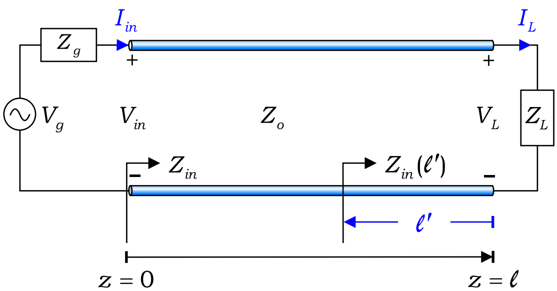


Fig. 9.6 A finite transmission line between a generator and a load.

9.4.1 Input Impedance

The wave launched at $z=0$ may take endless round-trips between Z_g and Z_L because of the impedance mismatches at $z=0$ and $z=\ell$. In order to avoid dealing with all individual reflections we add all the forward reflections into the forward wave of a complex amplitude V_o^+, and all the backward reflections into the backward wave of a complex amplitude V_o^-. For this reason, the forward wave represents the net wave incident on the load, and the backward wave represents the net wave reflected from the load. The total voltage and current at the load are called the load voltage V_L and the load current I_L, respectively. The load voltage and current at $z = \ell$ are written form the general solutions given in Eq. (9-33) as

$$V_L = V_o^+ e^{-\gamma\ell} + V_o^- e^{\gamma\ell} \tag{9-34a}$$

$$I_L = \frac{V_o^+}{Z_o} e^{-\gamma\ell} - \frac{V_o^-}{Z_o} e^{\gamma\ell} \tag{9-34b}$$

Again, the load voltage and current are related to each other by $V_L = I_L Z_L$, where Z_L is the load impedance. Combination of the above equations, with the aid of $V_L = I_L Z_L$, leads to

$$V_o^+ = \frac{I_L}{2}(Z_L + Z_o)e^{\gamma\ell} \tag{9-35a}$$

$$V_o^- = \frac{I_L}{2}(Z_L - Z_o)e^{-\gamma\ell} \tag{9-35b}$$

Upon substituting Eq. (9-35) into Eq. (9-33), we obtain the phasors of the time-harmonic voltage and current at a point on the transmission line as

$$V(z) = \frac{I_L}{2}\left[(Z_L + Z_o)e^{\gamma(\ell-z)} + (Z_L - Z_o)e^{-\gamma(\ell-z)}\right] \qquad \text{[V]} \tag{9-36a}$$

$$I(z) = \frac{I_L}{2Z_o}\left[(Z_L + Z_o)e^{\gamma(\ell-z)} - (Z_L - Z_o)e^{-\gamma(\ell-z)}\right] \qquad \text{[A]} \tag{9-36b}$$

The total voltage and current at the input end $(z = 0)$ are called the input voltage V_{in} and the input current I_{in} respectively. The input impedance of the line is the ratio between the input voltage and the input current, that is,

$$Z_{in} = \frac{V_{in}}{I_{in}} = \frac{V(0)}{I(0)} = Z_o \frac{Z_L\left(e^{\gamma\ell} + e^{-\gamma\ell}\right) + Z_o\left(e^{\gamma\ell} - e^{-\gamma\ell}\right)}{Z_o\left(e^{\gamma\ell} + e^{-\gamma\ell}\right) + Z_L\left(e^{\gamma\ell} - e^{-\gamma\ell}\right)} \tag{9-37}$$

With the help of the hyperbolic functions, $\cosh\theta = (e^{\theta} + e^{-\theta})/2$, $\sinh\theta = (e^{\theta} - e^{-\theta})/2$, and $\tanh\theta = \sinh\theta/\cosh\theta$, Eq. (9-37) can be rewritten in a compact form. The input impedance of a finite line of a length ℓ, terminated by a load impedance Z_L, is therefore

$$\boxed{Z_{in} = Z_o \frac{Z_L + Z_o \tanh\gamma\ell}{Z_o + Z_L \tanh\gamma\ell}} \qquad [\Omega] \qquad \text{(lossy line)} \tag{9-38}$$

where Z_o is the characteristic impedance and γ is the propagation constant on the transmission line, which is complex in general.

If we are only concerned with the power fed into a transmission line, by the generator, the input impedance Z_{in} may be used in place of the terminated transmission line as shown in Fig. 9.7. The equivalent circuit provides a simple way of finding the input voltage V_{in} and the input current I_{in}.

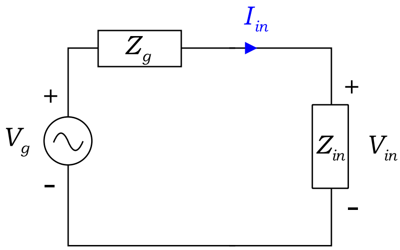


Fig. 9.7 Equivalent circuit for the source connected to a terminated transmission line.

For a lossless transmission line, the propagation constant is purely imaginary ($\gamma = j\beta$) and the characteristic impedance is purely real ($Z_o = R_o$) as given in Eq. (9-23). Under these conditions, the term $\tanh(j\beta\ell)$ in Eq. (9-38) becomes $j\tan(\beta\ell)$. Thus, the input impedance of a lossless transmission line of length ℓ, terminated by load impedance Z_L, is

$$\boxed{Z_{in} = R_o \frac{Z_L + jR_o \tan\beta\ell}{R_o + jZ_L \tan\beta\ell}} \qquad [\Omega] \qquad \text{(lossless line)} \qquad (9\text{-}39)$$

where R_o is the characteristic impedance that is purely real, and β is the phase constant. We note that the input impedance in Eq. (9-39) varies periodically with the line length ℓ.

Example 9-4

A 2[m]-transmission line is terminated in a load impedance $Z_L = 10 + j40[\Omega]$, and connected to a generator with $v_g(t) = 20\cos(10^7 t)[\mathrm{V}]$ and $Z_g = 50[\Omega]$. On the line $\gamma = 0.15 + j3.5[\mathrm{m}^{-1}]$ and $Z_o = 110 + j80[\Omega]$. Find

(a) input impedance,
(b) input current, and
(c) input voltage.

Solution

(a) From the values given in the problem, we write

$$\gamma\ell = 0.3 + j7.0$$

$$\tanh(\gamma\ell) = \frac{e^{\gamma\ell} - e^{-\gamma\ell}}{e^{\gamma\ell} + e^{-\gamma\ell}} = \frac{1 - e^{-0.6-j14}}{1 + e^{-0.6-j14}} = 0.482 + j0.749$$

From Eq. (9-38), we obtain

$$\begin{aligned} Z_{in} &= Z_o \frac{Z_L + Z_o \tanh \gamma \ell}{Z_o + Z_L \tanh \gamma \ell} \\ &= (110 + j80)\frac{(10 + j40) + (110 + j80)(0.482 + j0.749)}{(110 + j80) + (10 + j40)(0.482 + j0.749)} \\ &= 45.86 + j153.9 \end{aligned} \tag{9-40a}$$

(b) From the equivalent circuit, the input current is obtained as

$$\begin{aligned} I_{in} &= \frac{V_g}{Z_g + Z_{in}} = \frac{20}{50 + 45.86 + j153.9} \\ &= 0.0583 - j0.0936[\mathrm{A}] \end{aligned} \tag{9-40b}$$

(c) Input voltage is thus

$$\begin{aligned} V_{in} &= Z_{in} I_{in} = (45.86 + j153.9)(0.0583 - j0.0936) \\ &= 17.08 + j4.68[\mathrm{V}] \end{aligned} \tag{9-40c}$$

Exercise 9.12

From Eq. (9-33), show $V_o^+ = (1/2)(V_{in} + Z_o I_{in})$ and $V_o^- = (1/2)(V_{in} - Z_o I_{in})$.

Exercise 9.13

For a lossless line of $Z_o = R_o$, terminated in Z_L, find Z_{in} if the line is (a) a quarter wavelength long, and (b) a half wavelength long.

Ans. (a) $Z_{in} = R_o^2 / Z_L$, (b) $Z_{in} = Z_L$.

Exercise 9.14

For complex values of Z_g and Z_{in}, find the condition under which the maximum power is fed into the transmission line, by the generator.

Ans. $Z_{in} = Z_g^*$.

9.4.2 Reflection Coefficient and Standing Wave Ratio

The forward wave on a transmission line of characteristic impedance Z_o undergoes a reflection at the load, if the load impedance Z_L is not equal to Z_o. The ratio between the complex amplitudes of the reflected and the incident voltage waves, measured at the load, is called the voltage reflection coefficient Γ. At this point, we introduce a new space coordinate ℓ' that is defined as $\ell' = \ell - z$. It represents the distance from the load toward the generator, or simply the distance from the load. It should be noted that ℓ' is a variable, while the line length ℓ is a constant. Then the time-harmonic voltage on the transmission line given in Eq. (9-36a) can be rewritten in terms of ℓ' as

$$V(\ell') = \frac{I_L}{2}\left[(Z_L + Z_o)e^{\gamma\ell'} + (Z_L - Z_o)e^{-\gamma\ell'}\right] \equiv V_o'^{+}e^{\gamma\ell'} + V_o'^{-}e^{-\gamma\ell'} \tag{9-41}$$

It is important to remember that the term with $e^{\gamma\ell'}$ represents the forward wave of complex amplitude $V_o'^{+}$ and the term with $e^{-\gamma\ell'}$ represents the backward wave of complex amplitude $V_o'^{-}$. From Eq. (9-41), the voltage reflection coefficient of the load impedance Z_L is simply defined as $\Gamma = V_o'^{-}/V_o'^{+}$. Thus,

$$\boxed{\Gamma = \frac{Z_L - Z_o}{Z_L + Z_o} \equiv |\Gamma|e^{j\phi}} \tag{9-42}$$

which is a complex number, in general, with a magnitude not larger than one; that is, $|\Gamma| \le 1$. If Z_o and Z_L in Eq. (9-42) are replaced with η_1 and η_2, respectively, which are the intrinsic impedances of two adjoining media, the reflection coefficient in Eq. (9-42) is the same as the reflection coefficient for a uniform plane wave normally incident on the interface, as given in Eq. (8-107a). If $Z_L = Z_o$, we obtain $\Gamma = 0$ from Eq. (9-42). In this case, the transmission line is said to be ***matched to the load***. Similarly, the current reflection coefficient is defined as the ratio between the complex amplitudes of the reflected and the incident current waves at the load. It can be shown that the current reflection coefficient is equal to the negative of the voltage reflection coefficient.

We now express the voltage (or current) on the transmission line in terms of the complex amplitude of the forward voltage wave (or current wave), by substituting Eq. (9-35) and Eq. (9-42) into Eq. (9-36). Thus, on the transmission line,

$$\boxed{V(z) = V_o^{+}e^{-\gamma z}\left[1 + \Gamma e^{-2\gamma\ell'}\right]} \quad \text{[V]} \tag{9-43a}$$

$$\boxed{I(z) = I_o^{+}e^{-\gamma z}\left[1 - \Gamma e^{-2\gamma\ell'}\right]} \quad \text{[A]} \tag{9-43b}$$

Here, the length of the transmission line, ℓ, is a constant while ℓ' is a variable defined by $\ell' = \ell - z$. Comparison of Eq. (9-33a) and Eq. (9-43a) reveals that the first term inside the bracket in Eq. (9-43a) represents the forward wave, observed at a distance ℓ' from the load, whereas the second term represents the backward wave, reflected from the load impedance and observed at the distance ℓ' from the load. Manipulating the terms in Eq. (9-43a) we write the total voltage as

$$\begin{aligned} V(z) &= V_o^{+}e^{-\gamma z}\left(1 - |\Gamma| + |\Gamma| + \Gamma e^{-2\gamma\ell'}\right) \\ &= V_o^{+}e^{-\gamma z}\left(1 - |\Gamma|\right) + V_o^{+}e^{-\gamma z}|\Gamma|\left(1 + e^{j\phi}e^{-2\gamma\ell'}\right) \\ &= V_o^{+}e^{-\gamma z}\left(1 - |\Gamma|\right) + V_o^{+}|\Gamma|e^{j\phi/2-\gamma\ell}\left(e^{-j\phi/2}e^{\gamma\ell'} + e^{j\phi/2}e^{-\gamma\ell'}\right) \end{aligned} \tag{9-44}$$

where we used $\Gamma = |\Gamma| e^{j\phi}$. By assuming a lossless line with $\gamma = j\beta$, the instantaneous voltage on the line is obtained from Eq. (9-44) as

$$\begin{aligned} v(z,t) &= V_o^+ \left(1 - |\Gamma|\right) \cos\left(\omega t - \beta z\right) \\ &\quad + 2V_o^+ |\Gamma| \cos\left(\beta \ell' - \phi/2\right) \cos\left(\omega t + \phi/2 - \beta \ell\right) \end{aligned} \tag{9-45}$$

The first term on the right-hand side of Eq. (9-45) represents a traveling wave with amplitude $V_o^+ \left(1 - |\Gamma|\right)$, whereas the second term represents a standing wave with amplitude $2V_o^+ |\Gamma|$. In other words, a portion of the forward wave is reflected from the load, and interferes with the incoming wave to form a standing wave, while the rest propagating toward the load.

The standing wave causes the amplitude of the time-harmonic voltage to vary with position on the transmission line. The voltage on the lossless transmission line is therefore, from Eq. (9-43a),

$$V(z) = V_o^+ e^{-j\beta z} \left[1 + |\Gamma| e^{j(\phi - 2\beta \ell')}\right] \tag{9-46}$$

The spatial variation of the amplitude of V is solely determined by the terms in bracket in Eq. (9-46), because V_o^+ is constant and $\left|e^{-j\beta z}\right| = 1$. The voltage maximum occurs at a distance $\ell'_{\max}$ from the load. That is,

$$\boxed{\ell'_{\max} = \frac{1}{2\beta}\left(\phi + 2m\pi\right)} \qquad (m = 0, 1, 2..) \tag{9-47a}$$

where β is the phase constant of the wave and ϕ is the phase angle of the reflection coefficient. The voltage maximum occurs periodically, with half-wavelength spacing, along the line. By inserting Eq. (9-47a) into Eq. (9-46), the voltage maximum is

$$V_{\max} = \left|V_o^+\right| \left(1 + |\Gamma|\right) \tag{9-47b}$$

A similar procedure can be followed to obtain the voltage minimum on the transmission line. The voltage minimum occurs at a distance $\ell'_{\min}$ from the load, that is,

$$\boxed{\ell'_{\min} = \frac{1}{2\beta}\left[\phi + (2m+1)\pi\right]} \qquad (m = 0, 1, 2..) \tag{9-48a}$$

The voltage minimum is, from Eq. (9-46),

$$V_{\min} = \left|V_o^+\right| \left(1 - |\Gamma|\right) \tag{9-48b}$$

It is evident from Eq. (9-48a) that two adjacent minima are separated by a half-wavelength of the wave.

The voltage standing wave ratio (SWR) is defined as the ratio between the voltage maximum and the voltage minimum observed on the transmission line, that is,

$$\boxed{S = \frac{V_{\max}}{V_{\min}} = \frac{1+|\Gamma|}{1-|\Gamma|}} \tag{9-49}$$

It should be noted that $V_{\max}$ and $V_{\min}$ in Eq. (9-49) are measured at different points on the transmission line. As $|\Gamma|$ increases from 0 to 1, S increases from 1 to ∞. A higher standing wave ratio is less desirable because it means a large reflection at the load and a poor power delivery to the load. The inverse of Eq. (9-49) is

$$\boxed{|\Gamma| = \frac{S-1}{S+1}} \tag{9-50}$$

The unknown impedance of a load can be easily determined by measuring the voltage standing wave ratio on a slotted-line that is terminated by the load. The slotted-line consists of a lossless coaxial line with a long narrow slit in the outer conductor and a small probe movable along the slit. The probe can sample the electric field and measure the voltage maximum and minimum together with their separations. The ratio between the voltage maximum and minimum gives the standing wave ratio, which in turn tells us the magnitude of the reflection coefficient through Eq. (9-50). The separation between adjacent voltage maxima or voltage minima contains the information about the phase angle of the reflection coefficient, as can be seen from Eq. (9-47a) and Eq. (4-48a). By inserting the characteristic impedance of the slotted-line and the measured reflection coefficient into Eq. (9-42) we can determine the load impedance Z_L.

We now compute the power transmitted from the generator to the load by means of a lossless transmission line. The time-average power at a point on the transmission line is expressed by use of Eq. (9-43) as

$$\begin{aligned}\langle \mathscr{P} \rangle &= \frac{1}{2}\mathrm{Re}\left[V(z)I^*(z)\right] \\ &= \frac{1}{2}\mathrm{Re}\left[V_o^+ e^{-j\beta z}\left(1+\Gamma e^{-j2\beta\ell'}\right)\left(\frac{V_o^+}{R_o}e^{-j\beta z}\left(1-\Gamma e^{-j2\beta\ell'}\right)\right)^*\right] \\ &= \frac{1}{2}\frac{|V_o^+|^2}{R_o}\mathrm{Re}\left[1-|\Gamma|^2+\left\{\Gamma e^{-j2\beta\ell'}-\Gamma^* e^{j2\beta\ell'}\right\}\right]\end{aligned} \tag{9-51}$$

The brace { } on the right-hand side of Eq. (9-51) is given by a number that is purely imaginary, and thus contributes nothing to the time-average power. The time-average power delivered to the load is therefore

$$\langle \mathscr{P} \rangle = \frac{1}{2}\frac{|V_o^+|^2}{R_o}\left(1-|\Gamma|^2\right) \tag{9-52}$$

The first term inside the parenthesis in Eq. (9-52) represents the power of the incident wave, whereas the second term represents that of the reflected wave.

Example 9-5
With reference to the transmission line given in **Example 9-4**, find
(a) reflection coefficient,
(b) expression for the total voltage on the line,
(c) power dissipated at the load, and
(d) total power dissipated in the line.

Solution
(a) The reflection coefficient is, from Eq. (9-42),

$$\Gamma = \frac{Z_L - Z_o}{Z_L + Z_o} = \frac{(10+j40)-(110+j80)}{(10+j40)+(110+j80)} = 0.635e^{j2.74}.$$

(b) The input voltage is obtained from Eq. (9-43a) by substituting $z = 0$ and $\ell' = 2[\text{m}]$:

$$V_{in}(z=0, \ell'=2) = V_o^+\left[1+\Gamma e^{-4\gamma}\right] = V_o^+\left[1+0.635e^{j2.74}e^{-0.6-j14}\right]$$
$$= V_o^+(1.0911 + j0.336)[\text{V}]$$

Equating this with Eq. (9-40c), the amplitude of the forward wave is

$$V_o^+ = 15.51e^{-j0.031}[\text{V}]$$

Total voltage on the line is therefore

$$\begin{aligned} V(z) &= V_o^+ e^{-\gamma z}\left[1+\Gamma e^{-2\gamma\ell'}\right] \\ &= 15.51e^{-j0.031}e^{-(0.15+j3.5)z}\left[1+0.635e^{j2.74}e^{-(0.3+j7.0)\ell'}\right][\text{V}] \end{aligned} \tag{9-53}$$

(c) By substituting $z = 2[\text{m}]$ and $\ell' = 0$ into Eq. (9-53), the load voltage is

$$V_L = 15.51e^{-j0.031}e^{-(0.3+j7.0)}\left[1+0.635e^{j2.74}\right] = 5.56e^{-j0.210}[\text{V}]$$

Time-average power dissipated at the load is therefore

$$\langle \mathscr{P}_L \rangle = \frac{1}{2}\text{Re}\left[V_L\left(\frac{V_L}{Z_L}\right)^*\right] = \frac{1}{2}\text{Re}\left[\frac{(5.56)^2}{10-j40}\right] = 90.9[\text{mW}].$$

(d) By using the input voltage and current given in Eqs. (9-40b) and (9-40c), we compute the total power dissipated in the line and the load:

$$\langle \mathscr{P}_{in} \rangle = \frac{1}{2}\text{Re}\left[V_{in}I_{in}^*\right] = \frac{1}{2}\text{Re}[(17.08+j4.68)(0.0583+j0.0936)]$$
$$= 279[\text{mW}]$$

Thus the power dissipated in the line is

$279-90.9=188.1[\text{mW}]$.

Example 9-6

A lossless line of characteristic resistance R_o is terminated in a load impedance Z_L. The waves on the line have a phase constant β. At a distance ℓ' from the load toward the generator, the input impedance looking toward the load is defined as $Z_{in}(\ell') \equiv V(\ell')/I(\ell')$, and the reflection coefficient is defined as $\Gamma(\ell') = V^-(\ell')/V^+(\ell')$. Find

(a) expression for $\Gamma(\ell')$,

(b) $\Gamma(\ell')$ at the locations of $V_{\max}$ and $V_{\min}$, and

(c) $Z_{in}(\ell')$ at the locations of $V_{\max}$ and $V_{\min}$.

(d) Show that $S = Z_{in}(\ell')/R_o$ at the location of $V_{\max}$.

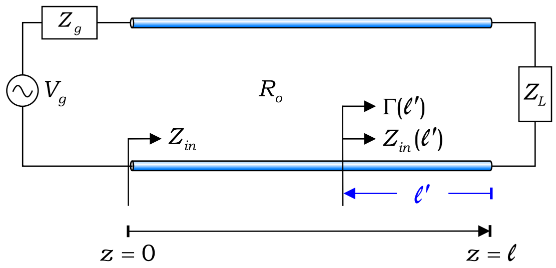


Fig. 9.8 Input impedance and reflection coefficient at a distance ℓ' from the load.

Solution

(a) At a point with $z = z_1$ on the line, or at a distance $\ell' = \ell - z_1$ from the load, the forward voltage wave is given by the first term on the right-hand side of Eq. (9-46):

$$V^+(z_1) = V_o^+ e^{-j\beta z_1}$$

At the same point on the line, the backward voltage wave is given by the second term on the right-hand side of Eq. (9-46):

$$V^-(z_1) = V_o^+ e^{-j\beta z_1} |\Gamma| e^{j(\phi - 2\beta\ell')}$$

The reflection coefficient at a distance ℓ' from the load is therefore

$$\Gamma(\ell') = \frac{V^-(z_1)}{V^+(z_1)} = |\Gamma| e^{j(\phi-2\beta\ell')}. \tag{9-54}$$

(b) From Eq. (9-46), we have a voltage maximum if the relation $e^{j(\phi-2\beta\ell')}=1$ is satisfied, and a voltage minimum if $e^{j(\phi-2\beta\ell')}=-1$. Therefore, from Eq. (9-54), we obtain

$$\Gamma(\ell') = |\Gamma| \qquad \text{: at the location of } V_{max} \tag{9-55a}$$

$$\Gamma(\ell') = -|\Gamma| \qquad \text{: at the location of } V_{min} \tag{9-55b}$$

We note that $\Gamma(\ell')$ is purely real at the locations of V_{max} and V_{min}.

(c) From Eq. (9-43), the input impedance at a distance ℓ' from the load is written as

$$\boxed{Z_{in}(\ell') = \frac{V_o^+ e^{-j\beta z}\left[1+\Gamma e^{-j2\beta\ell'}\right]}{I_o^+ e^{-j\beta z}\left[1-\Gamma e^{-j2\beta\ell'}\right]} = R_o \frac{\left[1+|\Gamma| e^{j\phi} e^{-j2\beta\ell'}\right]}{\left[1-|\Gamma| e^{j\phi} e^{-j2\beta\ell'}\right]}} \tag{9-56}$$

were we used $\gamma = j\beta$ and $Z_o = R_o$.

Noting that $e^{j(\phi-2\beta\ell')}=1$ for the voltage maximum and $e^{j(\phi-2\beta\ell')}=-1$ for the voltage minimum, from Eq. (9-56) we obtain

$$Z_{in}(\ell') = R_o \frac{1+|\Gamma|}{1-|\Gamma|} \qquad \text{: at the location of } V_{max} \tag{9-57a}$$

$$Z_{in}(\ell') = R_o \frac{1-|\Gamma|}{1+|\Gamma|} \qquad \text{: at the location of } V_{min} \tag{9-57b}$$

We note that $Z_{in}(\ell')$ is purely real at the locations of V_{max} and V_{min}.

(d) At the location of V_{max}, dividing both sides of Eq. (9-57a) by R_o we obtain

$$\frac{Z_{in}(\ell')}{R_o} = \frac{1+|\Gamma|}{1-|\Gamma|} = S\,. \tag{9-58}$$

Exercise 9.15
What is the shortest length of a transmission line on which we can observe a standing wave with one voltage maximum and one voltage minimum?

Ans. Quarter wavelength.

9.4.3 Short-Circuited and Open-Circuited Lines

The transmission line can be used for other purposes than guiding waves. A short-circuited or open-circuited line may be used as a circuit element with an inductive or capacitive reactance.

9.4.3.1 Short-Circuited Line ($Z_L = 0$)

A short circuit has a zero load impedance. If a lossless line of length ℓ is terminated in a short-circuit, the input impedance is obtained from Eq. (9-39) as

$$\boxed{Z_{in}^{s} = jR_o \tan\beta\ell} \tag{9-59}$$

The input impedance of a shorted line becomes purely inductive for $0 < \beta\ell < \pi/2$: it is a positive, imaginary number. The input impedance becomes purely capacitive for $\pi/2 < \beta\ell < \pi$: it is a negative, imaginary number. It is interesting to note that the input impedance becomes infinite for $\beta\ell = \pi/2$, or $\ell = \lambda/4$. The short-circuited line behaves like an open terminal, in this case.

9.4.3.2 Open-Circuited Line ($Z_L = \infty$)

An open terminal has infinite load impedance. If a lossless line of length ℓ is open circuited, the input impedance is obtained from Eq. (9-39) as

$$\boxed{Z_{in}^{o} = -jR_o \cot\beta\ell} \tag{9-60}$$

The input impedance becomes purely capacitive for $0 < \beta\ell < \pi/2$, but purely inductive for $\pi/2 < \beta\ell < \pi$. The input impedance becomes zero if the line is a quarter-wavelength long, or $\ell = \lambda/4$. In this case, the open-circuited line behaves like a shorted terminal.

The two input impedances Z_{in}^{s} and Z_{in}^{o} enable us to determine the characteristic impedance and the propagation constant of a given transmission line as follows:

$$R_o = \sqrt{Z_{in}^{s} Z_{in}^{o}} \qquad [\Omega] \tag{9-61a}$$

$$\beta\ell = \tan^{-1}\sqrt{\left|\frac{Z_{in}^{s}}{Z_{in}^{o}}\right|} + n\pi \qquad (n = 0, 1, 2..) \tag{9-61b}$$

where n is a positive integer, including zero.

Example 9-7
A lossless transmission line is known to be shorter than $10[\text{m}]$ in length. If it is open-circuited, the input impedance is given by $Z_{in}^{o} = -j81[\Omega]$. If short-circuited, the input impedance is given by $Z_{in}^{s} = j144[\Omega]$, and the first voltage maximum is observed at a distance $2.5[\text{m}]$ from the shorted end. Find

(a) R_o, (b) β, and (c) ℓ.

Solution

(a) The characteristic impedance is, from Eq. (9-61a),

$$R_o = \sqrt{Z_{in}^s Z_{in}^o} = \sqrt{(j144)(-j81)} = 108[\Omega] .$$

(b) From Eq. (9-42), the reflection coefficient of a short circuit ($Z_L = 0$) is obtained as $\Gamma = -1 = e^{j\pi}$. Thus we obtain $\phi = \pi$.

Substituting $\phi = \pi$, $m = 0$, and $\ell'_{\max} = 2.5[\mathrm{m}]$ into Eq. (9-47a), we get

$$\beta = 0.2\pi . \qquad (9\text{-}62)$$

(c) From Eq. (9-61b) we get

$$\beta\ell = \tan^{-1}\sqrt{144/81} + n\pi = 0.93 + n\pi . \qquad (9\text{-}63)$$

Inserting Eq. (9-62) into Eq. (9-63), the line length is

$$\ell = (0.93 + n\pi)/(0.2\pi)$$
$$= 1.48[\mathrm{m}],\ 6.48[\mathrm{m}],\ 11.48[\mathrm{m}],\ldots$$

The length ℓ should be longer than $2.5[\mathrm{m}]$, but shorter than $10[\mathrm{m}]$.

Therefore, we have $\ell = 6.48[\mathrm{m}]$.

Example 9-8

A $75[\Omega]$ -lossless line is terminated in an unknown impedance Z_L . Measurements show that $S = 3$ on the line, and that the first and second voltage maxima occur at distances $4[\mathrm{cm}]$ and $14[\mathrm{cm}]$ from the load, respectively. Find (a) Γ , and (b) Z_L .

Solution

(a) The distance between two adjacent maxima is a half-wavelength. Thus, the wavelength and the phase constant are

$$\lambda = 2\times(14-4) = 20[\mathrm{cm}]$$
$$\beta = 2\pi/\lambda = 10\pi[\mathrm{rad/m}]$$

Substituting $\ell'_{\max} = 0.04[\mathrm{m}]$, $m = 0$, and $\beta = 10\pi$ into Eq. (9-47a) we get

$$\phi = 0.8\pi[\mathrm{rad}] \qquad (9\text{-}64a)$$

Substituting $S = 3$ into Eq. (9-50) we get

$$|\Gamma| = \frac{3-1}{3+1} = 0.5 \qquad (9\text{-}64b)$$

From Eqs. (9-64a) and (9-64b) we get

$$\Gamma = 0.5e^{j0.8\pi}. \tag{9-65}$$

(b) Inserting Eq. (9-65) into Eq. (9-42) we get

$$\frac{Z_L - 75}{Z_L + 75} = 0.5e^{j0.8\pi}$$

Thus

$$Z_L = 46 + j88[\Omega].$$

Exercise 9.16
What are the reflection coefficients of (a) shorted terminal, and (b) open terminal?
Ans. (a) $\Gamma = -1$, (b) $\Gamma = 1$.

Review Questions with Hints

RQ 9.14 State the relation between the general solutions for V and I. [Eq.(9-33)]
RQ 9.15 Distinguish between characteristic impedance and input impedance. [Eq.(9-38)]
RQ 9.16 Write the reflection coefficient of a load impedance Z_L. [(9-42)]
RQ 9.17 Write the general solutions for V and I in terms of Γ. [Eq.(9-43)]
RQ 9.18 Express $V_{\max}$ and $V_{\min}$ in terms of $|\Gamma|$, and find their locations on a transmission line. [Eqs.(9-47)(9-48)]
RQ 9.19 Define standing wave ratio. [Eq.(9-49)]
RQ 9.20 Write the time-average power delivered to the load impedance in terms of Γ. [Eq.(9-52)]
RQ 9.21 Write the input impedances of short- and open-circuited lines. [Eqs.(9-59)(9-60)]

9.5 The Smith Chart

The Smith chart provides with a graphical means of finding solutions to various transmission line problems, avoiding tedious manipulations of complex numbers. It is a chart, in which the reflection coefficient is plotted as a function of the real (or imaginary) part of the normalized load impedance, while the other part is held fixed, in a complex plane whose abscissa and ordinate represent the real and imaginary parts of the voltage reflection coefficient, respectively. As we can see from Eq. (9-43), the voltage and current on a transmission line are uniquely determined by the reflection coefficient, meaning that all wave behaviors on the transmission line can be described in terms of the reflection coefficient. The Smith chart can be conveniently used for determining the standing wave ratio, input impedance, and load admittance, and solving impedance matching problems.

9.5.1 Relationship between Γ *and* z_L

To start with, we consider a lossless transmission line of characteristic impedance $Z_o = R_o$, which is terminated in a load impedance Z_L. From Eq. (9-42), the reflection coefficient at the load is written as

$$\begin{aligned}\Gamma &= \frac{Z_L - R_o}{Z_L + R_o} \\ &\equiv \Gamma_R + j\Gamma_I \equiv |\Gamma| e^{j\phi}\end{aligned} \tag{9-66}$$

where Γ_R and Γ_I are the real and imaginary parts of the voltage reflection coefficient Γ, respectively, and ϕ is the phase angle of Γ. Normalizing the load impedance Z_L to R_o we write

$$\begin{aligned}z_L &= \frac{Z_L}{R_o} = \frac{R_L + jX_L}{R_o} \\ &\equiv r + jx\end{aligned} \tag{9-67}$$

Here, R_L and X_L are the resistance and the reactance of Z_L, respectively, while r and x are the normalized load resistance and the normalized load reactance, respectively. We can express the relation between Γ and z_L in different forms as

$$\Gamma = \frac{z_L - 1}{z_L + 1} \tag{9-68a}$$

$$\boxed{z_L = \frac{1+\Gamma}{1-\Gamma} = \frac{1+|\Gamma| e^{j\phi}}{1-|\Gamma| e^{j\phi}}} \tag{9-68b}$$

$$r + jx = \frac{1 + \Gamma_R + j\Gamma_I}{1 - \Gamma_R - j\Gamma_I} \tag{9-68c}$$

Separating the real and imaginary parts of Eq. (9-68c), we have

$$r = \frac{1 - \Gamma_R^2 - \Gamma_I^2}{(1-\Gamma_R)^2 + \Gamma_I^2} \tag{9-69a}$$

$$x = \frac{2\Gamma_I}{(1-\Gamma_R)^2 + \Gamma_I^2} \tag{9-69b}$$

Rewriting Eq. (9-69), we have

$$\boxed{\left(\Gamma_R - \frac{r}{1+r}\right)^2 + \Gamma_I^2 = \left(\frac{1}{1+r}\right)^2} \tag{9-70a}$$

$$\boxed{(\Gamma_R - 1)^2 + \left(\Gamma_I - \frac{1}{x}\right)^2 = \left(\frac{1}{x}\right)^2} \tag{9-70b}$$

For a fixed value of r, Eq. (9-70a) represents a circle of radius $1/(1+r)$ with the center at a point $(r/(1+r),\ 0)$ in the (Γ_R, Γ_I)-plane. Similarly, for a fixed value of x, Eq. (9-70b) represents a circle of radius $1/|x|$ with the center at a point $(1,\ 1/x)$ in the (Γ_R, Γ_I)-plane. From the above equations, it is implicit that Γ varies with x in Eq. (9-70a), and with r in Eq. (9-70b).

Let us suppose that we are given a normalized load impedance with $r = r'$ and $x = x'$. Then we can solve Eqs. (9-70a) and (9-70b) for the reflection coefficient, which is given either by $(\Gamma_R, \Gamma_I) = (\Gamma'_R, \Gamma'_I)$ or by $(\Gamma_R, \Gamma_I) = (1, 0)$. The second solution is trivial because it satisfies Eq. (9-70a) and Eq. (9-70b) regardless of r and x. The first solution represents the intersection between two circles of Eq. (9-70a) and Eq. (9-70b), which are drawn by assuming $r = r'$ and $x = x'$, respectively, in the (Γ_R, Γ_I)-plane, as shown in Fig. 9.9. It can be shown that the two circles expressed by Eq. (9-70a) and Eq. (9-70b) always intersect at right angles, and all the circles pass through the point $(\Gamma_R, \Gamma_I) = (1, 0)$. Consequently, the reflection coefficient of the normalized load impedance, $z_L = r' + jx'$, is obtained as $\Gamma = \Gamma'_R + j\Gamma'_I$.

The Smith chart contains two families of circles: one comprises of the circles of constant r, while the other comprises of the circles of constant x. The circles of a positive x reside in the region $\Gamma_I > 0$, whereas those of a negative x reside in the region $\Gamma_I < 0$. If the intersection between two circles of $r = r'$ and $x = x'$ is identified on the Smith chart, we can find the magnitude of Γ by measuring the relative length of the line segment $\overline{OA'}$ with respect to the axial distance of a point with $|\Gamma| = 1$, on the chart, and the phase angle of Γ by measuring the rotation angle of $\overline{OA'}$ with respect to the positive Γ_R-axis, as shown in Fig. 9.9.

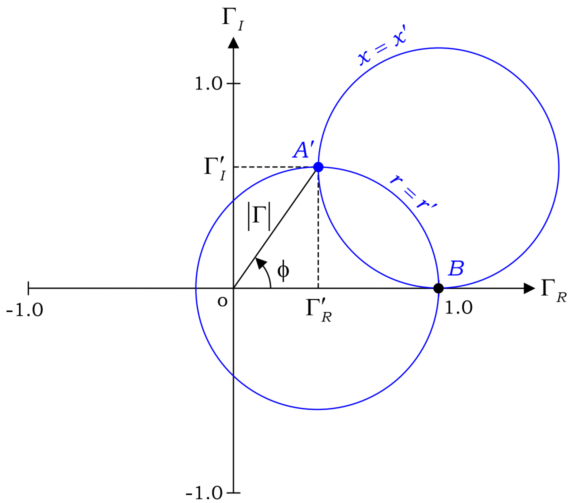


Fig. 9.9 Reflection coefficient is $\Gamma'_R + j\Gamma'_I$ for a normalized load impedance $r' + jx'$.

Exercise 9.17
Use the Smith chart to find Γ of the following normalized load impedances:
(a) $z_L = 3$, (b) $z_L = 1 + j1$, (c) $z_L = -j10$, (d) $z_L = 0$, and (e) $z_L = \infty$.

Ans. (a) $\Gamma = 0.5$, (b) $\Gamma = 0.45\angle 63.4^o$, (c) $\Gamma = 1\angle -11.4^o$, (d) $\Gamma = -1$, (e) $\Gamma = 1$.

Exercise 9.18
Use the Smith chart to find z_L responsible for the following reflection coefficients:
(a) $\Gamma = 1\angle 45^o$, (b) $\Gamma = -0.5$, and(c) $\Gamma = j$.

Ans. (a) $z_L = j2.41$, (b) $z_L = 0.33$, (c) $z_L = j$.

9.5.2 Relationship between Γ *and* Z_{in}

Let us consider a lossless transmission line($\gamma = j\beta$) of a length ℓ. The voltage and current at a point on the line are written from Eq. (9-43) as

$$V(z) = V_o^+ e^{-j\beta z}\left[1 + \Gamma e^{-j2\beta\ell'}\right] \tag{9-71a}$$

$$I(z) = I_o^+ e^{-j\beta z}\left[1 - \Gamma e^{-j2\beta\ell'}\right] \tag{9-71b}$$

Here, $\ell' = \ell - z$, representing the distance from the load toward the generator. The input impedance of the finite line is defined as the ratio between V and I measured at $z = 0$, or at $\ell' = \ell$. Noting the relation $V_o^+ = R_o I_o^+$ on the lossless line of characteristic resistance R_o, the input impedance of the line of the length ℓ is

$$Z_{in} = \frac{V(\ell' = \ell)}{I(\ell' = \ell)} = R_o \frac{\left[1 + \Gamma e^{-j2\beta\ell}\right]}{\left[1 - \Gamma e^{-j2\beta\ell}\right]} = R_o \frac{\left[1 + |\Gamma| e^{j(\phi - 2\beta\ell)}\right]}{\left[1 - |\Gamma| e^{j(\phi - 2\beta\ell)}\right]} \tag{9-72}$$

where ϕ is the phase angle of Γ.

Upon normalizing Z_{in} in Eq. (9-72) to R_o, we have

$$z_{in} \equiv \frac{Z_{in}}{R_o} = \frac{\left[1 + |\Gamma| e^{j(\phi - 2\beta\ell)}\right]}{\left[1 - |\Gamma| e^{j(\phi - 2\beta\ell)}\right]} \tag{9-73}$$

We recognize that z_{in} in Eq. (9-73) is the same as z_L in Eq. (9-68b), except for the phase angle $\phi - \beta\ell$. This means that the term $|\Gamma| e^{j(\phi - 2\beta\ell)}$ may behave as $|\Gamma| e^{j\phi}$ on the Smith chart, if the two families of circles on the Smith chart are assumed to be the circles of constant $\mathrm{Re}[z_{in}]$ and the circles of constant $\mathrm{Im}[z_{in}]$, respectively. Consider Fig. 9.10, in which the point of $|\Gamma| e^{j\phi}$ is marked as A' on the Smith chart, which is the intersection of the $r = r'$ circle and the $x = x'$ circle as was shown in Fig. 9.9. If we can locate the point of $|\Gamma|e^{j(\phi-2\beta\ell)}$ on the Smith chart, the two circles passing through that point will correspond to the real and imaginary parts of z_{in}, according to Eq. (9-73). It is easy to locate the point of $|\Gamma|e^{j(\phi-2\beta\ell)}$; we only need to reduce the phase angle of Γ by $2\beta\ell$ [rad]. In Fig. 9.10, the straight-line segment $\overline{OA'}$ is rotated about the origin by $2\beta\ell$ [rad] in the clockwise direction so that the tip of $\overline{OA'}$ moves from point A' to point A'' along "circle of constant $|\Gamma|$". Then we notice that the $r = r''$ circle and the $x = x''$ circle intersect at point A''. Thus, the normalized input impedance is obtained as $z_{in} = r'' + jx''$ from the Smith chart. The input impedance of the lossless line of length ℓ, having characteristic resistance R_o, terminated in load impedance Z_L, is therefore

$$Z_{in} = R_o z_{in} = R_o r'' + jR_o x''[\Omega]. \tag{9-74}$$

These are summarized as follows:

(1) Normalize Z_L to R_o and obtain $z_L = r' + jx'$.
(2) Find $r = r'$ and $x = x'$ circles on the Smith chart, and locate the intersection.
(3) Draw a circle centered at the origin, passing through the intersection, called "circle of constant $|\Gamma|$".
(4) Move the point on "circle of constant $|\Gamma|$" in the clockwise direction, reducing polar angle by $2\beta\ell$ [rad].
(5) Find two circles passing through the new point.
(6) Read $r = r''$ and $x = x''$.
(7) The input impedance is given by $Z_{in} = R_o r'' + jR_o x''[\Omega]$.

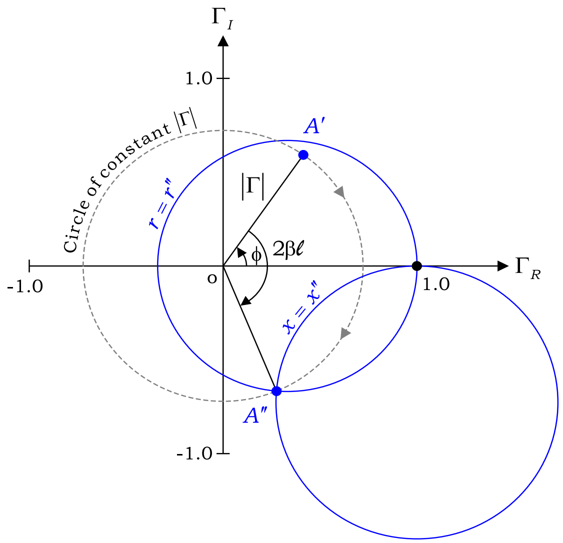


Fig. 9.10 Relation between z_{in}, at point A'', and Γ, at point A'.

We recall from Fig. 9.8 that $Z_{in}(\ell')$ is the input impedance looking toward the load at a distance ℓ' from the load, which is simply the ratio between the voltage and current at a distance ℓ' from the load. That is,

$$Z_{in}(\ell') = \frac{V(\ell')}{I(\ell')} = R_o \frac{\left[1+|\Gamma| e^{j(\phi-2\beta\ell')}\right]}{\left[1-|\Gamma| e^{j(\phi-2\beta\ell')}\right]} \tag{9-75}$$

We recognize that $Z_{in}(\ell')$ is the input impedance as if the line were ℓ'[m] long. Thus, the relation between the normalized impedance, $z_{in}(\ell') = Z_{in}(\ell')/R_o$, and the reflection coefficient is the same as that between z_{in} and Γ given in Eq. (9-73). In view of these, we can obtain $z_{in}(\ell')$ on the Smith chart by reducing the polar angle of the point of Γ by $2\beta\ell'$ [rad].

A complete turn on "circle of constant $|\Gamma|$" corresponds to a half-wavelength on the transmission line; that is, $2\beta\ell' = 2\pi\,[\text{rad}]$. To denote the movement on the transmission line by a distance ℓ', the scale called "wavelengths toward generator"(*wtg*) is constructed around the perimeter of the circle of constant $|\Gamma|$, with radius 1, or the perimeter of the Smith chart. The *wtg* is measured in units of the wavelength. It is customary to set the zero point of the *wtg* scale on the negative Γ_R-axis; *wtg* increases in the clockwise direction.

As an example, consider a lossless line of length 0.169λ, which is terminated in a load impedance that is responsible for the reflection coefficient $\Gamma = 0.73\angle 55^o$. The normalized input impedance of this line is obtained on the Smith chart as illustrated in Fig. 9.11. We first convert the phase angle $\phi = 55^o$ into a point on the *wtg* scale; that is, $wtg = 0.174\lambda = (\lambda/2)(180^o - 55^o)/360^o$. Here, the subtraction of 55^o from 180^o is to reflect the fact that ϕ has zero point on the positive Γ_R-axis, increasing in the counterclockwise direction, whereas *wtg* has zero point on the negative Γ_R-axis, increasing in the clockwise direction. The factor $\lambda/2$ is to reflect the fact that a complete turn on the circle of constant $|\Gamma|$ corresponds to a half-wavelength. The point of Γ is found, at a distance 0.73 from the origin, on a straight line drawn from the origin to the point 0.174λ on the *wtg* scale, which is assumed to be of length 1. Next, the point of Γ is moved on the circle of constant $|\Gamma|$, from point *A* to point *B*, so that the *wtg* increases from 0.174λ to $0.343\lambda(= 0.174\lambda + 0.169\lambda)$, as shown in Fig. 9.11. We read $r = 0.49$ and $x = -1.40$ at point *B*. Thus, the normalized input impedance is obtained as $z_{in} = 0.49 - j1.40$.

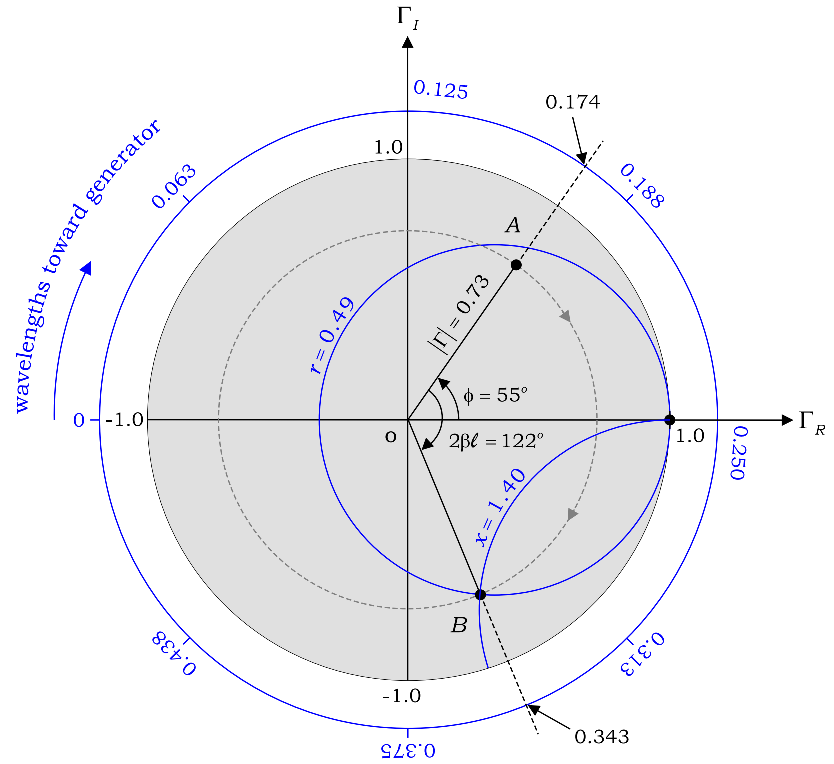


Fig. 9.11 Reflection coefficient Γ at point A and normalized input impedance z_{in} at point B.

Exercise 9.19
Use the Smith chart to show that the input impedance of a half-wavelength line is equal to the load impedance.

9.5.3 Relationship between Γ and Standing Wave Ratio

We begin with a lossless line of characteristic resistance R_o terminated in a pure resistance R_L ($R_L > R_o$). The reflection coefficient at the load is, from Eq. (9-42),

$$\Gamma = \frac{R_L - R_o}{R_L + R_o} \tag{9-76}$$

The reflection coefficient of a purely resistive load R_L, connect to a lossless line of R_o $(R_L > R_o)$, is a positive real number. Rewriting Eq. (9-76), by use of $\Gamma = |\Gamma|$, we have

$$\frac{R_L}{R_o} = \frac{1+|\Gamma|}{1-|\Gamma|} \equiv S \tag{9-77}$$

From Eq. (9-77), we notice that the normalized load impedance on the left-hand side of the equation is equal to the standing wave ratio on the line. This holds true only if the load impedance is larger than the characteristic resistance ($R_L > R_o$).

We next consider a lossless line of R_o terminated in an arbitrary load impedance Z_L as shown in Fig. 9.12(a). As the distance from the load, ℓ', is increased on the transmission line, the point of $|\Gamma|e^{j(\phi-2\beta\ell')}$ rotates in the clockwise direction on the circle of constant $|\Gamma|$ on the Smith chart. When the point arrives at the positive Γ_R-axis, which correspond to $\phi - 2\beta\ell' = 0$, the input impedance becomes purely resistive, $Z_{in}(\ell') \equiv R'_{in}$, as can be seen from Eq. (9-75). If we cut off the transmission line at the distance ℓ' from the load and terminate it by a load of a resistance R'_{in} as depicted in Fig. 9.12(b), there is no change in the waves traveling on the shortened transmission line. In view of Eq. (9-77), the standing wave ratio on the shortened line is simply given by $S = R'_{in} / R_o$. In other words, the standing wave ratio on a lossless line is equal to the normalized input impedance, measured at a distance ℓ' from the load looking toward the load, which is purely real and larger than 1. It should be noted that the standing wave ratio is measured on the positive Γ_R-axis of the Smith chart.

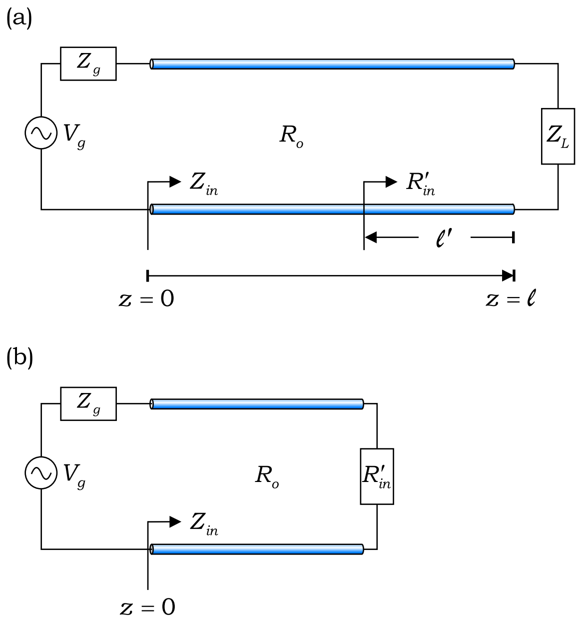


Fig. 9.12 (a) Purely resistive input impedance at a distance ℓ' from the load. (b) Equivalent line, shortened and terminated in the resistance.

As an example, for the reflection coefficient, $\Gamma = 0.73\angle 55^o$, of a given load, we can obtain the standing wave ratio on the lossless transmission line from the Smith chart by taking the following steps. First, we locate the point of Γ on the Smith chart, which is marked as *A* in Fig. 9.13. Next, we move on the circle of constant $|\Gamma|$ from point *A* to point *B*, which is positioned on the positive Γ_R-axis. We read $r = 6.4$ and $x = 0$ at point *B*. The normalized input impedance, $z_{in} = 6.4 + j0$, is purely resistive and larger than 1, and therefore the standing wave ratio on the transmission line is $S = 6.4$.

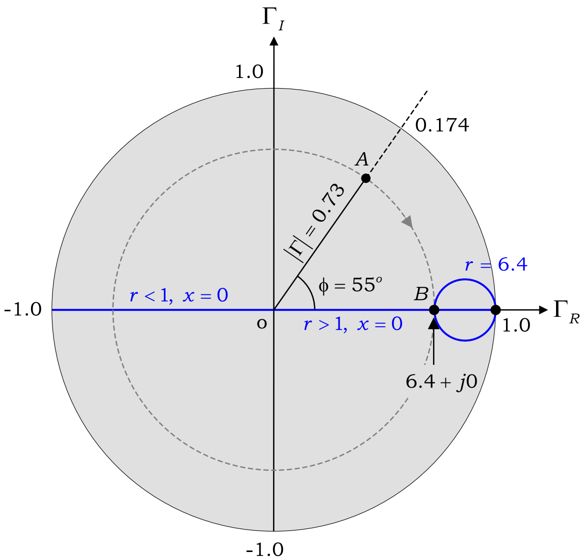


Fig. 9.13 Standing wave ratio is measured at point *B*.

Exercise 9.20
Find the locus of Γ on the Smith chart, if S is fixed in Eq. (9-49).
Ans. A circle centered at the origin.

9.5.4 Admittances on the Smith Chart

A lossless line of characteristic resistance R_o is a quarter-wavelength long, and is terminated with a load impedance Z_L. Upon substituting $\beta\ell = \pi/2$ into Eq. (9-39), we obtain the input impedance, that is,

$$Z_{in} = R_o \frac{Z_L + jR_o \tan(\pi/2)}{R_o + jZ_L \tan(\pi/2)} = \frac{R_o^2}{Z_L} \qquad \text{(9-78)}$$

By using the admittances $Y_L = 1/Z_L$ and $Y_o = 1/R_o$, we rewrite Eq. (9-78) as

$$\frac{Z_{in}}{R_o} = \frac{Y_L}{Y_o} \tag{9-79}$$

We recognize the left-hand side of Eq. (9-79) as the normalized input impedance, z_{in}, and the right-hand side as the normalized load admittance, y_L. That is,

$$z_{in} = y_L \equiv g + jb \tag{9-80}$$

Here, g is the normalized load conductance and b is the normalized load susceptance.

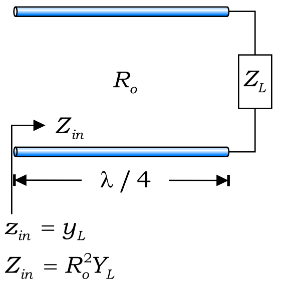


Fig. 9.14 Input impedance of a quarter-wave line.

The Smith chart may also be used as an admittance chart. In that case, the circles of constant r are regarded as the circles of constant g, and the circles of constant x as the circles of constant b. Eq. (9-80) allows us to obtain the load admittance, $Y_L = 1/Z_L$, by use of the Smith chart (see Fig. 9.15). If the load impedance is given as Z_L, we first locate the point of the normalized load impedance, $z_L = Z_L / R_o = r' + jx'$, on the Smith chart, which is the intersection between the $r = r'$ circle and the $x = x'$ circle. The point is marked as A in Fig. 9.15, where we read $\Gamma = |\Gamma| \angle \phi^o$ as the reflection coefficient. Next, to find the input impedance of the quarter-wave line, we move along the circle of constant $|\Gamma|$ from the point of Γ, or point A, to point B by reducing the phase angle ϕ by 180^o or by increasing the wtg by 0.25λ. We read $r = r''$ and $x = x''$ at point B as the normalized input impedance. Thus, according to Eq. (9-80), the normalized load admittance is obtained as $y_L = r'' + jx''$. The admittance of the load impedance Z_L is therefore

$$\begin{aligned} Y_L &= \frac{y_L}{R_o} = \frac{z_{in}}{R_o} \\ &= (r'' + jx'') / R_o \end{aligned} \tag{9-81}$$

The admittance is measured in units of siemens[S]. As can be seen from Fig. 9.15, ***the point of y_L on the Smith chart is the inversion of the point of z_L through the origin, and vice versa.***

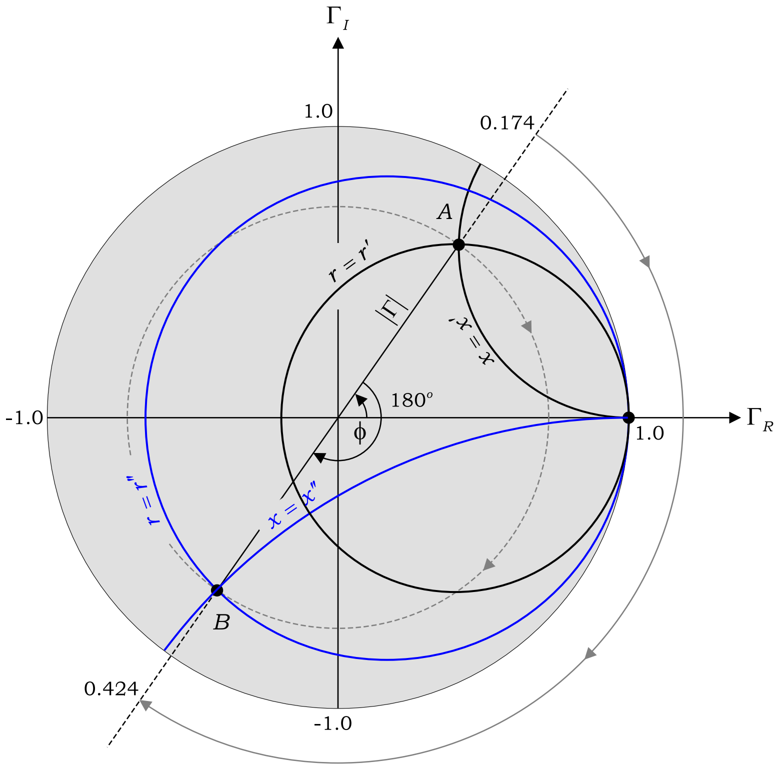


Fig. 9.15 Normalized load impedance at A and normalized load admittance at B.

Example 9-9

A lossless transmission line of $R_o = 75[\Omega]$ has a length of $\ell = \lambda/8$. Locate the point of z_L on the Smith chart, and find Γ and z_{in}, if the line is terminated in

(a) a short circuit, and

(b) an open circuit.

Solution

(a) Shorted terminal has $Z_L = 0$: $z_L = r + jx = 0 + j0$.
The $r = 0$ circle and the $x = 0$ circle intersect at point V (see Fig. 9.16)
Point V gives $|\Gamma| = 1$ and $\phi = \pi$. : $\Gamma = -1$.
The line length $\ell = \lambda/8$ corresponds to a *wtg* of 0.125λ.
Move from point V to point W by increasing the *wtg* by 0.125λ.
Point W gives $r = 0$ and $x = 1$: $Z_{in} = j75[\Omega]$.

(b) Open terminal has $Z_L = \infty$: $z_L = r + jx = \infty + j\infty$.
The $r = \infty$ circle and the $x = \infty$ circle intersect at point X (see Fig. 9.16).
Point X gives $|\Gamma| = 1$ and $\phi = 0$: $\Gamma = 1$.
Move from point X to point Y by increasing the *wtg* by 0.125λ.
Point Y gives $r = 0$ and $x = -1$: $Z_{in} = -j75[\Omega]$.

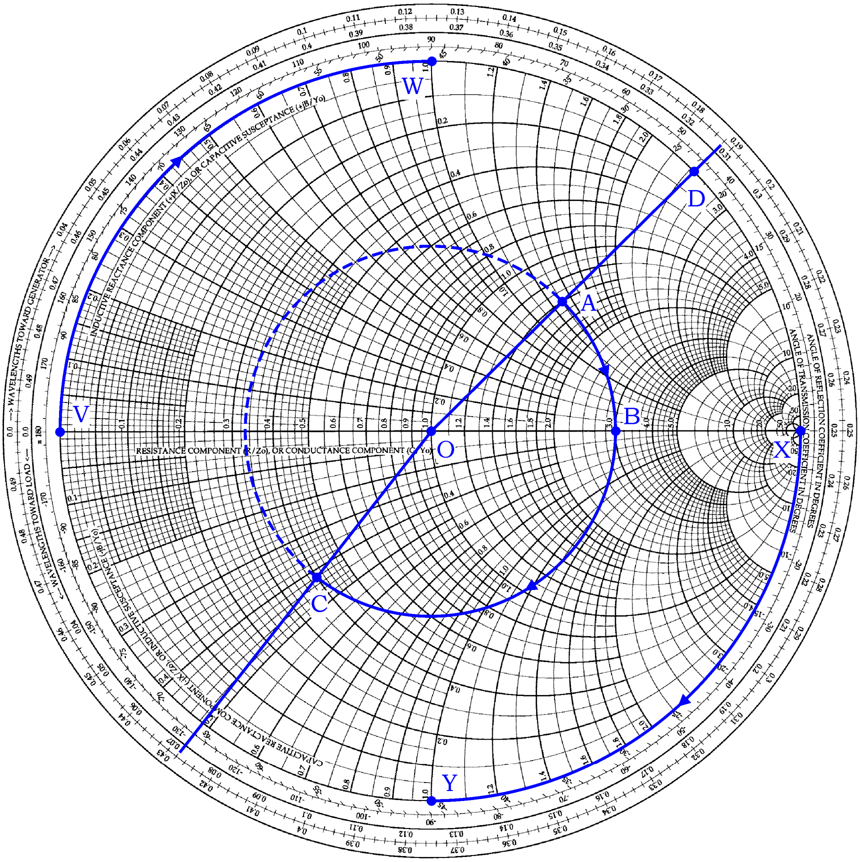


Fig. 9.16 Smith chart for Example 9-9 and Example 9-10.

Example 9-10

A $100[\Omega]$-lossless line is 0.24λ long, and is terminated in a load impedance $Z_L = 140 + j130[\Omega]$. By using the Smith chart, find

(a) voltage reflection coefficient,
(b) input impedance,
(c) standing wave ratio, and
(d) location of the first voltage maximum from the load.

Solution

(a) Normalized load impedance is

$$z_L = (140 + j130)/100 = 1.4 + j1.3 .$$

The intersection of the $r = 1.4$ circle and the $x = 1.3$ circle is marked as *A* on the Smith chart in Fig. 9.16.

From the Smith chart we find that

(1) $\frac{\overline{OA}}{\overline{OD}} = 0.5 = |\Gamma|$, and

(2) point A is located at 0.188λ on the *wtg* scale.

Noting that the *wtg* of 1λ corresponds to $4\pi[\text{rad}]$, the phase angle of point A is calculated as follows:

$\phi = (0.25 - 0.188) \times 4\pi = 0.78[\text{rad}]$.

Thus, $\Gamma = 0.5e^{j0.78}$.

(b) Move from point A to point C along the circle of $|\Gamma| = 0.5$, in the clockwise direction, by increasing the *wtg* by 0.24λ.

At point C, we read $r = 0.40$ and $x = -0.42$.

The input impedance is therefore

$Z_{in} = 100(0.40 - j0.42) = 40 - j42[\Omega]$.

(c) The circle of $|\Gamma| = 0.5$ crosses the positive Γ_R-axis at point B, where we read $r = 3.0$ and $x = 0$.

Thus, the standing wave ratio is $S = 3.0$.

(d) A voltage maximum occurs at a distance ℓ' from the load, at which $z_{in}(\ell') = r' + j0$ ($r' > 1$). This point is marked as B. The difference in *wtg* between points B and A is equal to ℓ', that is,

$\ell' = (0.25 - 0.188)\lambda = 0.062\lambda$.

Exercise 9.21

A $50[\Omega]$-lossless line of length 0.5λ is terminated in $Z_L = j200[\Omega]$. By using the Smith chart, find (a) *wtg* for Γ, (b) Γ, (c) Z_{in}, (d) Y_L, (e) location of V_{max} from the load, (f) location of V_{min} from the load, and (g) S.

Ans. (a) 0.211λ, (b) $|\Gamma| = 1$, $\phi = (0.25 - 0.211) \times 4\pi = 0.49[\text{rad}]$, (c) $Z_{in} = 50(j4) = j200[\Omega]$, (d) $Y_L = (-j0.25)/50 = -j5[\text{mS}]$, (e) 0.039λ, (f) 0.289λ, (g) $S = \infty$.

Exercise 9.22

Solve $\coth(a + jb) = 0.5e^{j0.78}$ for the constants a and b by using Smith chart. [Hint: Eq. (9-68b).]

Ans. $a = 0.324$ and $b = 0.374$.

9.5.5 Impedance Matching with a Single-Stub

The transmission line can transmit the maximum power from the generator to the load, if the load impedance is equal to the characteristic impedance of the line so that there is no standing wave on the line. It is not possible, in general, to match a load to the transmission line because the load may be designed to perform other tasks independently. Alternatively, a matching network can be connected between the line and the load; stub matching networks and transformers are two commonly used impedance matching networks.

In the stub matching, the impedance on the transmission line is altered by connecting short- or open-circuited stubs in parallel or in series with the main line. It is rather difficult to work with a series stub, and an open stub radiates undesirable electromagnetic energy at high frequencies. A short-circuited stub, having the same characteristic impedance as the main line, can be connected in parallel with the line so that the load impedance is matched to the line. This method is called the single-stub matching. Since the admittances simply add in parallel connections, it is more convenient to use the Smith chart as an admittance chart in solving impedance matching problems involving the parallel connection.

Consider Fig. 9.17, which illustrates a single-stub matching. We change the length and location of the stub in such a way that the normalized input admittance looking toward the load at the junction is unity, that is, $y_{in} = 1$. Since the normalized input admittance of a shorted stub is purely imaginary such as $y_d = -jb_d$, the normalized input admittance of the main line looking toward the load should be $y_\ell = 1 + jb_d$ so that we have $y_{in} = y_\ell + y_d = 1$ at the junction.

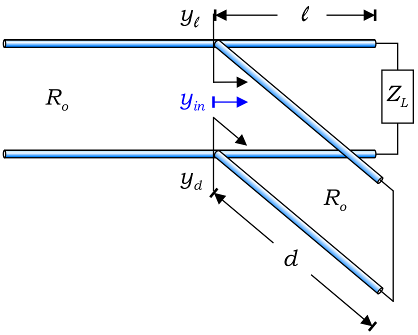


Fig. 9.17 Single-stub matching.

Example 9-11

A $50[\Omega]$-lossless line is terminated in a load impedance $Z_L = 17.5 - j55[\Omega]$. Find the location and length of the short-circuited stub to be connected in parallel for the impedance matching.

Solution

Normalized load impedance is

$$z_L = (17.5 - j55)/50 = 0.35 - j1.1.$$

The intersection between the $r = 0.35$ circle and the $x = -1.1$ circle is marked as A on the Smith chart in Fig. 9.18. It is located at 0.363λ on the wtg scale.

To obtain the normalized load admittance, at the inversion of point A through the origin, or at point B, we read $r = 0.26$ and $x = 0.83$. Point B is located at 0.113λ on the *wtg* scale.

Thus, $y_L = 0.26 + j0.83$.

We next move along the circle of constant $|\Gamma|$ from point B to point C, which is a point on the $r = 1$ circle. At point C, we read $x = 2.2$ and $wtg = 0.191\lambda$. The difference in *wtg* between points C and B is $(0.191 - 0.113)\lambda = 0.078\lambda$.

Thus, we have $y_\ell = 1 + j2.2$ at a distance $\ell = 0.078\lambda$ from the load.

A short circuit has $Z_L = 0$ and $Y_L = \infty$, which is marked as D on the admittance chart in Fig. 9.18. We move from point D to point E, which is the intersection between the circle of $|\Gamma| = 1$ and the circle of $b = -2.2$. At point E, we read 0.318λ on the *wtg* scale.

The difference in *wtg* between points E and D is the length of the stub, that is, $d = (0.318 - 0.25)\lambda = 0.068\lambda$.

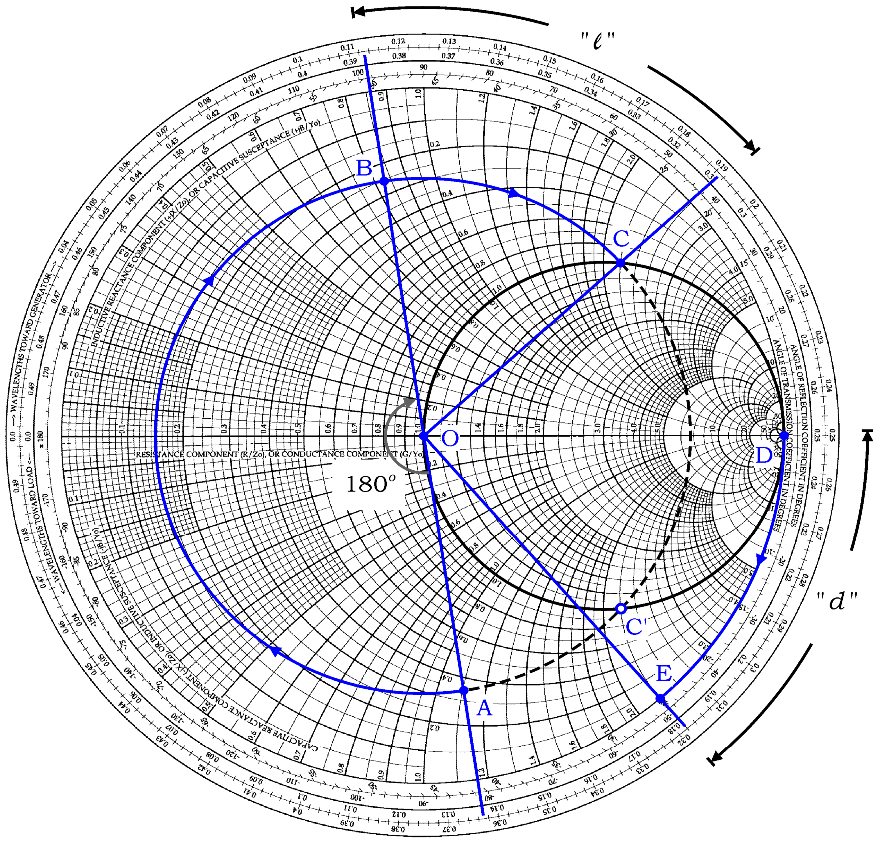


Fig. 9.18 Impedance matching with a single-stub.

Exercise 9.23
Locate the point of the normalized load impedance, on the Smith chart, matched to a given line.

Ans. Center.

Exercise 9.24
On the $100[\Omega]$-lossless line terminated in $Z_L = 160 + j220[\Omega]$, the separation between adjacent voltage maxima is 5[cm]. For impedance matching, (a) locate a shorted stub closest to the load, and (b) find the shortest length.

Ans. (a) $2.31[\text{cm}]$, (b) $0.81[\text{cm}]$.

Review Questions with Hints

RQ 9.22 What are the circles in the Smith chart? [Eq.(9-70)]
RQ 9.23 What are the horizontal and vertical axes in the Smith chart? [Eq.(9-70)]
RQ 9.24 What do the polar coordinates of a point on the Smith chart represent? [Fig.9.9]
RQ 9.25 What are the values of r and x for the point on the horizontal axis in the Smith chart? [Fig.9.13]
RQ 9.26 What is the value of $|\Gamma|$ on the perimeter of the Smith chart? [Fig.9.13]
RQ 9.27 What is meant by a movement on the circle of constant $|\Gamma|$? [Fig.9.10]
RQ 9.28 What is meant by a complete turn on the circle of constant $|\Gamma|$? [Eq.(9-75)]
RQ 9.29 Locate, on the circle of constant $|\Gamma|$, the points corresponding to $V_{\max}$ and $V_{\min}$ on the transmission line. [Eqs.(9-47)(9-48)(9-75)]
RQ 9.30 Explain how to use the Smith chart for Γ and S. [Fig.9.13]
RQ 9.31 Describe the geometrical relation between the normalized load impedance and admittance on the Smith chart. [Fig.9.15]
RQ 9.32 Explain why it is more convenient to use the admittance chart for a single-sub matching. [Fig.9.17]

Problems

9-1 A coaxial transmission line consists of an inner conductor of radius $a = 0.51[\text{mm}]$ and an outer conductor of inner radius $b = 2.4[\text{mm}]$, which are separated by polyethylene of $\varepsilon_r = 2.26$ and $\varepsilon''/\varepsilon' = 0.0002$. At a frequency $f = 1[\text{GHz}]$, find

(a) R, L, G, and C
(b) α, β, and Z_o
(c) dB power loss at $30[\text{m}]$

9-2 A lossless transmission line of characteristic resistance $75[\Omega]$ operates at a frequency 80[MHz], and the phase constant is measured as $8.42[\text{rad/m}]$. Find L and C of the line.

9-3 With reference to the coaxial line in **Problem 9-1**, by assuming that the permittivity, permeability, and conductivity are independent of frequency, find the frequency at which the line becomes distortionless.

9-4 Verify that the following statements always hold true in an air-spaced lossy transmission line regardless of the frequency:
(a) Phase angle of Z_o is in the range $-\pi/4 \le \theta_{Z_o} \le 0$.
(b) Phase angle of γ is in the range $\pi/4 \le \theta_\gamma \le \pi/2$.
(c) $\theta_\gamma - \theta_{Z_o} = \pi/2$.

9-5 A generator with $V_g = 10\angle 0^o[\text{V}]$ and $Z_g = 20 + j15[\Omega]$ is connected to a half-wave lossless line that is terminated in a load impedance Z_L. Find
(a) Z_L for the maximum power transmission from the generator to the load, and
(b) maximum power delivered to the load.

9-6 A voltage source with $Z_g = 75 + j75[\Omega]$ is directly connected to a load that consists of a $75[\Omega]$ resistor connected in series with a short-circuited stub of characteristic impedance $75[\Omega]$. Determine the shortest length of the stub, in units of λ, for the maximum power delivered to the resistor.

9-7 A $100[\Omega]$-coaxial line is made of a perfect conductor and a perfect dielectric($\varepsilon = 2.56\varepsilon_o$ and $\mu = \mu_o$). The cable is 1.5[m] long and is terminated in a load impedance $Z_L = 60 + j50$. For the frequency of operation of 200[MHz], find the input impedance.

9-8 If a $100[\Omega]$, air-spaced, 0.6[m] long, lossless line is terminated in an unknown load impedance Z_L, the input impedance is measured as $Z_{in} = 100 + j130[\Omega]$ at $f = 50[\text{MHz}]$. Find the load impedance Z_L.

9-9 If a lossless line of $200[\Omega]$ is terminated in an unknown load impedance, the input impedance is measured as $Z_{in} = 169 + j210[\Omega]$, and a quarter of the power fed into the line is reflected back. Find the voltage reflection coefficient at the input end.

9-10 The standing wave ratio is measured to be 3.58 on a lossless line that is terminated in a load impedance $Z_L = 280 + j140[\Omega]$. Find the characteristic impedance, which is known to be less than $150[\Omega]$.

9-11 The input impedance of a 2[m]-line is measured as $Z_{in}^o = 56.81 - j213.5[\Omega]$ when open circuited, and as $Z_{in}^s = 182.9 + j16.37[\Omega]$ when short circuited. Find Z_o and γ.

9-12 A lossless line with $Z_o = 50[\Omega]$ and $\beta = 1.2\pi[\text{rad/m}]$ is terminated in a load impedance $Z_L = 80 + j120[\Omega]$. The amplitude of the forward wave is measured to be $V_o^+ = 10[\text{V}]$. Find

(a) reflection coefficient at the load,

(b) voltage on the line as a function of ℓ'(distance from the load), and

(c) distance from the load to the first voltage maximum.

9-13 The input impedance looking toward the load at a distance ℓ' from the load is denoted as $Z_{in}(\ell')$. Show that

(a) the maximum input impedance $|Z_{in}(\ell')|_{\max}$ occurs at the same location as the voltage maximum $V_{\max}$, and

(b) the maximum normalized input impedance $z_{in}(\ell')|_{\max}$ is equal to the standing wave ratio.

9-14 A $100[\Omega]$-lossless line is connected to another $50[\Omega]$-lossless line of 0.4λ in length, which is then terminated with a load resistance $100[\Omega]$. Find the standing wave ratio on each line.

9-15 A $100[\Omega]$-lossless line of length 0.2λ is inserted between a $50[\Omega]$-lossless line and a load resistance $75[\Omega]$. The time-average power of 4[W] is fed into the combined line. Calculate the time-average power dissipated at the load by using the reflection coefficient at the joint of the two lines.

9-16 Repeat **Problem 9-15** by using the voltage across the load.

9-17 A lossless line of characteristic impedance Z_o may be matched to a load impedance Z_L via a quarter-wave section, which is also lossless, as shown in Fig. 9.19.

(a) Find Z_1 if Z_L is purely resistive.

(b) Can the line also be matched to a load that is not purely resistive?

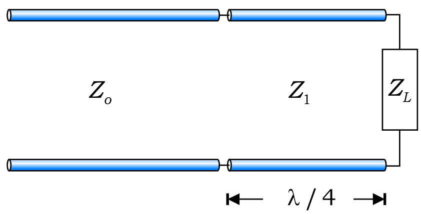


Fig. 9.19 In-series quarter-wave transformer(Problem 9-17).

9-18 A lossless line of characteristic impedance $Z_{o1} = 100[\Omega]$ is of length $\ell_1 = 2[\text{m}]$, and is connected end-to-end to another lossless line of characteristic impedance $Z_{o2} = 50[\Omega]$. The second line is of length $\ell_2 = 1.25[\text{m}]$, and is terminated by a load impedance $Z_L = 75[\Omega]$. Both lines are air-spaced lines. If the input voltage at the input end of the first line is $v_{in} = 10\cos(3\pi\times 10^8 t)$, determine

(a) standing wave ratio on each line,
(b) expression for voltage on each line,
(c) time-average power dissipated at the load, and
(d) time-average input power.

9-19 A lossless line of characteristic resistance R_1 and length ℓ_1 is inserted between the lossless line of $Z_o = 75[\Omega]$ and the load impedance $Z_L = 100 - j50[\Omega]$ to match the load to the line of Z_o. Find R_1 and ℓ_1 in terms of λ.

9-20 Solve $\tanh(a + jb) = -1.01 - j0.73$ for a and b by using the Smith chart.

9-21 A load impedance $Z_L = 30 + j44[\Omega]$ is connected to a lossless line of characteristic impedance $100[\Omega]$, which is 0.336λ long. Determine, using the Smith chart,
(a) Γ, (b) S, (c) location of $|V|_{\max}$, (d) Z_{in}, and (e) Y_{in}.

9-22 Given the intersection of the $r = r'$ circle and the $x = x'$ circle on the Smith chart, find, in terms of r' and x',
(a) polar angle of the intersection, and
(b) wtg of the intersection.

9-23 The standing wave ratio is measured to be 3.0 on a lossless line of $75[\Omega]$. The distance between a voltage maximum and a nearby voltage minimum is 5[cm], and the first voltage minimum is at 2.5[cm] from the load. Determine, using the Smith, (a) load impedance, and (b) reflection coefficient.

9-24 A $120[\Omega]$-lossless line is terminated in an unknown load impedance. From measurements, the standing wave ratio is 3.1, the first voltage minimum is at 2.52[cm] from the load, and the first voltage maximum is at 6.27[cm] from the load. Using the Smith chart, determine
(a) load impedance Z_L, and
(b) position and length of a short-circuited stub to match the line to the load.

9-25 A lossless line of $50[\Omega]$, 29.6[cm] long, is made of a lossless dielectric($\varepsilon = 2.25\varepsilon_o$ and μ_o). If the line is terminated by an unknown load impedance Z_L, the input impedance is measured as $Z_{in} = 140 + j100[\Omega]$ at frequency $f = 100[\text{MHz}]$. Find, using the Smith chart,
(a) load impedance Z_L, and
(b) voltage reflection coefficient at the load.

9-26 When a $200[\Omega]$-lossless line is terminated in an unknown load impedance, the input impedance is measured as $Z_{in} = 56 - j68[\Omega]$. The input impedance, however, is measured as $Z_{in}^s = j250[\Omega]$, if it is terminated in a short circuit. Determine the load impedance by using the Smith chart.

Chapter 10
Waveguides

In Chapter 9, we saw that a transmission line is used in transmitting electromagnetic signals from a point to anther by means of a transverse electromagnetic (TEM) wave. At an operating frequency in the SHF (3-30[GHz]) or EHF (30-300[GHz]) bands, the two-conductor transmission line becomes highly lossy owing to the skin effect in the conductors with a finite conductivity. Alternatively, waveguides can provide a much more efficient way of transmitting electromagnetic signals with the frequency in those bands. A waveguide is usually a hollow metal pipe with a uniform rectangular or circular cross section. It has a lower attenuation constant because of the large surface area of the conductors. The waveguide differs from the transmission line in various important aspects. The waveguide can support transverse electric (TE) waves with no longitudinal electric field, and transverse magnetic (TM) waves with no longitudinal magnetic field. In contrast, the transmission line supports TEM waves only. The waveguide operates only above a particular frequency known as a cutoff frequency, while the transmission line has no cutoff frequency, and can thus be used in transmitting even DC voltages and currents. A rigorous electromagnetic approach is required for the analysis of the waveguide, whereas the simple circuit theory dealing with voltages and currents can be used for the analysis of the transmission line. When, however, the operating frequency approaches the optical frequency of visible or infrared light, the loss due to the skin effect becomes excessively high in the hollow metallic waveguide. To circumvent these difficulties, the air-conductor interfaces are replaced with dielectric-dielectric interfaces, which allow for low-loss transmission of the optical signals, as in optical fibers and dielectric waveguides. Fig. 10.1 illustrates different hollow metallic waveguides, and Fig. 10.2 shows an optical fiber and a dielectric slab waveguide.

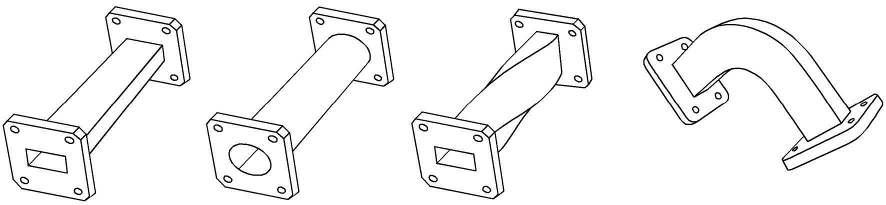


Fig. 10.1 Metallic waveguides.

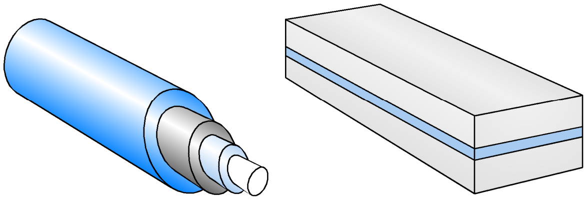


Fig. 10.2 An optical fiber and a dielectric slab waveguide.

10.1 Parallel-Plate Waveguides

All metallic waveguides operate on the same basic principles. To see how electromagnetic fields propagate in the waveguide, we may use a parallel-plate waveguide consisting of two parallel conducting plates as shown in Fig. 10.3. The parallel-plate waveguide is the simplest form of a metallic waveguide in the sense that the electromagnetic fields are uniform in the transverse plane or otherwise vary only in one spatial dimension in the transverse plane. The electromagnetic wave propagating in a waveguide has particular field patterns that are determined by Maxwell's equations. To be specific, the time-varying electric and magnetic fields should satisfy differential wave equations in the interior of the waveguide, and obey boundary conditions at the conductor-dielectric interfaces.

10.1.1 Transverse Electromagnetic(TEM) Waves

As was discussed in Chapter 7, the uniform plane wave is a solution to the three-dimensional differential wave equation. In time domain, it is also referred to as a time-harmonic wave. Since the electromagnetic wave obeys the principle of superposition, the sum of the uniform plane waves with the same wavenumber but with different directions of travel constitutes a unique solution in a given region of space. For this reason, the behavior of the electromagnetic wave in a waveguide can be conveniently analyzed by means of uniform plane waves.

Let us consider a parallel-plate waveguide as shown in Fig. 10.3, in which two parallel plates are assumed to be perfectly conducting, and the fringing effects at the edges are ignored. A single uniform plane wave can exist in the space between the plates if the wavevector is parallel to the surface of the plates, and the electric fields are perpendicular to the conducting surfaces, satisfying the boundary condition which requires no tangential components of the electric field on the perfectly conducting surfaces. Under this condition, both the electric and magnetic field vectors lie in the transverse plane, or the plane perpendicular to the direction of propagation of the wave, as illustrated in Fig. 10.4. A wave with these field patterns is called a transverse electromagnetic(TEM) mode; the specific field patterns of the wave in a waveguide is called the waveguide mode, which can propagate in the waveguide with no change in shape. Only if the aforementioned conditions are satisfied, two uniform plane waves may be combined in such a way that they satisfy the boundary condition at the conducting surfaces, and constitute a waveguide mode propagating in the parallel-plate waveguide.

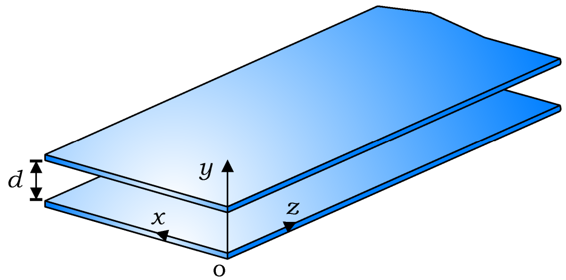


Fig. 10.3 A parallel-plate waveguide.

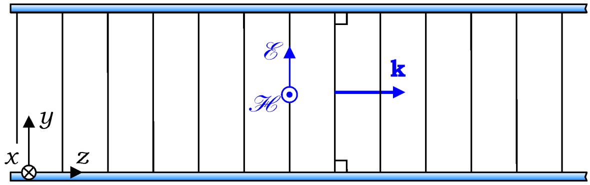


Fig. 10.4 TEM mode in a parallel-plate waveguide.

10.1.2 Transverse Electric(TE) Waves

The aforementioned TEM mode is not the only mode of propagation in the parallel-plate waveguide. Consider a uniform plane wave that is incident on the upper conducting plate at an angle θ with respect to the z-axis as shown in Fig. 10.5. This wave will be reflected from the upper plate, and then from the lower plate, and so on. From the point of view of the zigzagging, the wave propagates along the waveguide bouncing up and down endlessly between the two conducting plates. From another point of view, in which the wave propagates in free space between the two plates, we can assume that two uniform plane waves propagate in the space in between, one in the upward direction and the other in the downward direction, satisfying the differential wave equation. Note that the two viewpoints come from the boundary condition at the conducting surfaces and the differential wave equation in the homogeneous medium, for the electromagnetic wave, respectively.

In Fig. 10.5, the wavefronts of the upward wave are denoted by solid lines, whereas the wavefronts of the downward wave are denoted by dotted lines. In the figure the blue line, either a solid or a dotted line, corresponds to the wavevector. Although the two waves may have the electric field components parallel to the conducting surfaces, they may be linearly combined so as to have no tangential components on the surface of the plates, and thus comprise a mode of propagation in the waveguide. Depending on the spatial distribution of the electric and magnetic fields, the waveguide mode may be called the transverse electric (TE) mode, having no longitudinal component of the electric field, or the transverse magnetic(TM) mode, having no longitudinal component of the magnetic field.

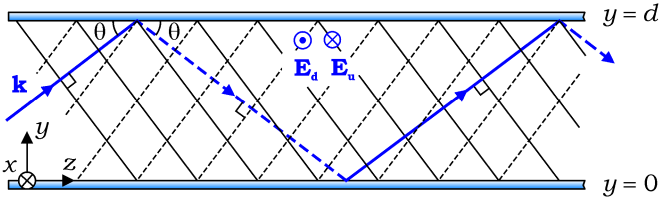


Fig. 10.5 TE mode in a parallel-plate waveguide.

We now examine the TE mode of a parallel-plate waveguide, and obtain the phase relationship between the upward and downward plane waves, as shown in Fig. 10.5. We begin with a uniform plane wave with the electric field vector perpendicular to the plane of incident, or the yz-plane in Fig. 10.5. The wave is incident on the upper plate at an angle θ with respect to the z-axis, which we call the upward wave. The electric field of the upward wave is written in phase form as

$$\mathbf{E}_u = \mathbf{a}_x E_o e^{-j(k_y y + k_z z)} \tag{10-1}$$

where E_o is the amplitude, and $\mathbf{a}_x$ is a unit vector in the direction of the electric field vector. In the above equation, k_y and k_z are the y- and z-components of the wavevector given by $\mathbf{k} = k_y\,\mathbf{a}_y + k_z\,\mathbf{a}_z = k\sin\theta\,\mathbf{a}_y + k\cos\theta\,\mathbf{a}_z$. The wavenumber k is therefore

$$k^2 = k_x^2 + k_y^2 \tag{10-2}$$

According to the law of reflection, the reflected wave, or the downward wave, should have the wavevector of the form $\mathbf{k}' = -k\sin\theta\,\mathbf{a}_y + k\cos\theta\,\mathbf{a}_z$, which is the same as that of the incident wave except for the negative sign in the y-component. The magnitudes of $\mathbf{k}$ and $\mathbf{k}'$ are the same because the upward and downward waves propagate in the same medium, or air. The electric field of the downward wave is therefore written in phasor form as

$$\mathbf{E}_d = -\mathbf{a}_x E_o e^{-j(-k_y y + k_z z)} \tag{10-3}$$

The minus sign on the right-hand side of Eq. (10-3) stems from the Fresnel's equation for the perpendicular polarization; the reflection coefficient for the perpendicular polarization at the perfectly conducting surface is given by $\Gamma_\perp = -1$. It shows that the electric field $\mathbf{E}_d$ is 180° out of phase with $\mathbf{E}_u$. The phase difference of 180° is denoted by the dotted line in Fig. 10.5. The total electric field in the waveguide is equal to the sum of $\mathbf{E}_u$ and $\mathbf{E}_d$. That is,

$$\mathbf{E} = \mathbf{E}_u + \mathbf{E}_d = E_o\,\mathbf{a}_x\left[e^{-jk_y y} - e^{jk_y y}\right]e^{-jk_z z} \tag{10-4a}$$

Rewriting Eq. (10-4a) we have

$$\boxed{\mathbf{E} = -j2E_o\,\mathbf{a}_x \sin(k_y y)\,e^{-jk_z z}} \qquad \text{(10-4b)}$$

We now apply the boundary condition for electric field to the conducting surfaces at $y = 0$ and $y = d$, which requires that the tangential component of $\mathbf{E}$ is zero: $\mathbf{E}(y = 0) = 0$ and $\mathbf{E}(y = d) = 0$. Application of the boundary condition to Eq. (10-4b) leads to

$$\boxed{k_y = \frac{m\pi}{d}} \qquad (m = 1, 2, 3, \ldots) \qquad \text{(10-5)}$$

With the aid of Eq. (10-5), Eq. (10-4b) constitutes a solution to the differential wave equation, satisfying the boundary conditions. In other words, the electric field in Eq. (10-4b) is a mode of propagation in the parallel-plate waveguide. Examination of Eq. (10-4b) reveals that the wave propagates in the z-direction with the phase constant k_z, and the electric field of the wave is directed along the x-axis, varying with y in the transverse plane, or the xy-plane. The waveguide mode expressed by Eq. (10-4b) has no longitudinal component of the electric field, and is called the transverse electric(TE) mode. The integer m in Eq. (10-5) is called the mode number.

We can obtain the eigenvalues of TE modes given in Eq. (10-5) from the transverse resonance condition for the upward and downward waves. The sine term in Eq. (10-4b) signifies that a standing wave is formed along the y-direction in the transverse plane, because of the interference of two plane waves $e^{-jk_y y - jk_z z}$ and $e^{jk_y y - jk_z z}$ (see Eq. (10-4a)). Although the two plane waves propagate with different wavevectors, they behave in unison in the waveguide so that they may be viewed as a single plane wave traveling along a zigzag path. During the round trip from the upper plate to the lower plate and then back to the upper plate, the wave experiences a total phase shift of ψ, that is,

$$\begin{aligned}\psi &= k_y d + \phi + k_y d + \phi \\ &= 2\pi m\end{aligned} \qquad (m = 1, 2, 3, \ldots) \qquad \text{(10-6)}$$

where $k_y d$ is the phase shift due to the travel in the upward or downward direction by a distance d, and ϕ is the phase shift due to the reflection at the perfectly conducting surface. Note that Fresnel's equations give $\Gamma_\perp = -1$ for the perpendicular polarization, and therefore $\phi = \pi$ for the TE mode. The transverse resonance condition requires that the phase shift ψ should be an integer multiple of 2π. Otherwise, the two waves would interfere destructively, and no electric field could exist inside the waveguide. We notice that the eigenvalue k_y given in Eq. (10-6) is the same as that given in Eq. (10-4).

It is more convenient to identify the TE mode by giving k_z instead of k_y. Inserting Eq. (10-5) into Eq. (10-2), we obtain

$$\begin{aligned}\beta_m \equiv k_z &= \pm\sqrt{k^2 - k_y^2} \\ &= \pm\sqrt{\left(\frac{n\omega}{c}\right)^2 - \left(\frac{m\pi}{d}\right)^2} \qquad (m = 1, 2, 3, \dots)\end{aligned} \tag{10-7}$$

where we used the relation $k = n\omega / c$ by assuming the dielectric between the plates, in general, to be nonmagnetic and lossless; that is, $\mu = \mu_o$ and $\varepsilon = \varepsilon_r \varepsilon_o = n^2 \varepsilon_o$. In the above equation, n is the refractive index, m is the mode number, c is the speed of light in free space, and d is the separation between the plates.

At an operating frequency f, the mode of operation propagates along the waveguide in the $+z$-direction with the phase constant β_m as given in Eq. (10-7). We rewrite Eq. (10-7), by taking the positive sign, as

$$\boxed{\beta_m = k\sqrt{1 - \left(\frac{f_{c(m)}}{f}\right)^2}} \qquad \text{[rad/m]} \tag{10-8}$$

Here, the cutoff frequency $f_{c(m)}$ is defined as

$$\boxed{f_{c(m)} = \frac{mc}{2nd}} \qquad \text{[rad/s]} \tag{10-9}$$

It is important to note that k is the phase constant of the wave assumed to propagate in an unbounded dielectric, whereas β_m is the phase constant of the mode m propagating actually along the waveguide in the $+z$-direction.

If an operating frequency f is lower than the cutoff frequency $f_{c(m)}$, the radicands in Eqs. (10-7) and (10-8) become negative. Under this condition, we can rewrite Eq. (10-7) as $k_z \equiv -j\left|\alpha_m\right|$ by taking the minus sign for the reason that will become evident shortly. Substitution of Eq. (10-7) into Eq. (10-4) shows that the electric field of a given mode decays exponentially as $e^{-|\alpha_m|z}$ along the waveguide; we note that the positive sign would have meant a wave with an amplitude increasing indefinitely. ***Below the cutoff frequency, the mode will not propagate in the waveguide*.**

The ray angle θ is measured between the wavevector $\mathbf{k}$ and the z-axis as shown in Fig. 10.5. We now obtain the relation between the angle θ_m for mode m and the cutoff frequency $f_{c(m)}$. Suppose we gradually decrease the frequency of operation toward the cutoff frequency. Since the frequency is still larger than the cutoff frequency, the relation $f_{c(m)} / f = k_y / k$ is still valid, which is evident from Eqs. (10-7) and (10-8). Combining this with the relation $\sin\theta_m = k_y / k$, which is apparent from Fig. 10.5, the ray angle θ_m for mode m, above cutoff, is given by

$$\sin\theta_m = \frac{k_y}{k} = \frac{f_{c(m)}}{f} \tag{10-10}$$

where the wavenumber $k = n\omega / c$, which is obtained by assuming a nonmagnetic, lossless material between the parallel plates. We see from Eq. (10-10) that the angle θ_m becomes $\pi/2$ as f approaches $f_{c(m)}$. In this case, the wave moves in the transverse direction, and no propagation of the wave takes place in the waveguide.

Equation (10-4) shows that the TE_m mode is comprised of two plane-waves, which propagate with the same phase velocity $\upsilon_p = \omega / k = 1/\sqrt{\mu\varepsilon}$, but in different directions with θ_m and $-\theta_m$. Note that υ_p is the phase velocity in an unbounded dielectric. In conjunction with the relations $\beta_m = k_z = k\cos\theta_m$ and $k = n\omega / c$, the phase velocity of mode m, which is measured along the waveguide, is expressed as

$$\boxed{\upsilon_{p(m)} = \frac{\omega}{\beta_m} = \frac{c}{n\cos\theta_m}} \qquad [\mathrm{m/s}] \tag{10-11}$$

Although $\upsilon_{p(m)}$ is larger than the speed of light in an unbounded dielectric(c/n), it is not against the special relativity because $\upsilon_{p(m)}$ is a mere phase velocity along the waveguide.

The electromagnetic energy propagates with a group velocity in the waveguide. From Eq. (10-8), the group velocity of mode m is

$$\boxed{\upsilon_{g(m)} = \frac{d\omega}{d\beta_m} = \frac{1}{d\beta_m / d\omega} = \frac{c}{n}\sqrt{1-\left(\frac{f_{c(m)}}{f}\right)^2}} \qquad [\mathrm{m/s}] \tag{10-12}$$

Rewriting Eq. (10-12) with the help of Eq. (10-10), the group velocity of mode m is simply expressed as

$$\upsilon_{g(m)} = \frac{c}{n}\cos\theta_m \qquad [\mathrm{m/s}] \tag{10-13}$$

From Eq. (10-13) we see that the group velocity of mode m is equal to the projection of the phase velocity of the component plane-wave in the direction of propagation of the mode, or the z-direction. Therefore the group velocity is always smaller than the speed of light in an unbounded dielectric.

The real instantaneous electric field of mode m is obtained by multiplying Eq. (10-4) by the harmonic time-dependence term $e^{j\omega t}$ and by taking the real part. Above cutoff, the electric field of the TE_m mode is therefore

$$\mathscr{E}_{TE} = 2E_o\mathbf{a}_x \sin\left(\frac{m\pi}{d}y\right)\sin(\omega t - \beta_m z) \qquad (\omega > \omega_{c(m)}) \qquad (10\text{-}14)$$

Below the cutoff frequency, $f < f_{c(m)}$, the phase constant β_m expressed by Eq. (10-7) becomes imaginary such that $\beta_m = -j\alpha_m$. Here, α_m is called the attenuation constant. Therefore, below cutoff, the electric field of the TE_m mode is expressed as

$$\mathscr{E}_{TE} = 2E_o\mathbf{a}_x \sin\left(\frac{m\pi}{d}y\right)e^{-\alpha_m z}\sin(\omega t) \qquad (\omega < \omega_{c(m)}) \qquad (10\text{-}15)$$

The attenuation constant is defined as

$$\boxed{\alpha_m = \frac{n\omega_{c(m)}}{c}\sqrt{1-\left(\frac{f}{f_{c(m)}}\right)^2}} \qquad \text{[Np/m]} \qquad (10\text{-}16)$$

where $\omega_{c(m)} = 2\pi f_{c(m)}$. Below the cutoff frequency, the mode is nonpropagating in the waveguide, and is called the evanescent mode.

Exercise 10.1
For the parallel-plate waveguide with an air gap of $d = 5$ [mm], find the cutoff frequency for the TE mode with $m = 1$.

Ans. 30[GHz].

Exercise 10.2
What is the mode number of the TE mode illustrated in Fig. 10.5?

Ans. $m = 4$.

10.1.3 Transverse Magnetic(TM) Waves

Figure 10.6 illustrates a uniform plane wave with parallel polarization launched into a parallel-plate waveguide. In our notation, for the parallel polarization, the positive direction of the electric field vector **E** is such that the projection of **E** onto the interface is parallel to the projection of the wavevector **k** onto the same plane (see Fig. 8.17). From Fresnel's formula in Eq. (8-151a), the reflection coefficient for the parallel polarization is $\Gamma_{\parallel} = -1$ at the perfectively conducting surface. In view of these considerations, the electric field vectors of the upward and downward waves, $\mathbf{E}_u$ and $\mathbf{E}_d$, should lie in the plane of incident, or the plane of the paper, as shown in Fig. 10.6. The total electric field in the waveguide is thus written as

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_u + \mathbf{E}_d \\ &= \left(-\mathbf{a}_y \cos\theta + \mathbf{a}_z \sin\theta\right) E_o e^{-j(k_y y + k_z z)} - \left(\mathbf{a}_y \cos\theta + \mathbf{a}_z \sin\theta\right) E_o e^{j(k_y y - k_z z)}\end{aligned}$$

where the wavevector $\mathbf{k} = k_y\,\mathbf{a}_y + k_z\,\mathbf{a}_z = k\sin\theta\,\mathbf{a}_y + k\cos\theta\,\mathbf{a}_z$, as is apparent from Fig. 10.6. Rewriting the above equation we have

$$\boxed{\mathbf{E} = -2E_o\left[\mathbf{a}_y \cos\theta\cos(k_y y) + \mathbf{a}_z j\sin\theta\sin(k_y y)\right]e^{-jk_z z}} \tag{10-17}$$

Inserting Eq. (10-17) into Eq. (8-9a), the magnetic field in the waveguide is

$$\begin{aligned}\mathbf{H} &= -\frac{1}{j\omega\mu}\nabla\times\mathbf{E} \\ &= \mathbf{a}_x \frac{2E_o}{\eta}\cos(k_y y)e^{-jk_z z}\end{aligned} \tag{10-18}$$

where $\eta = \sqrt{\mu/\varepsilon}$ is the intrinsic impedance of the dielectric inside the waveguide. The electric and magnetic fields expressed by Eq. (10-17) and Eq. (10-18) form a waveguide mode, which propagates along the waveguide in the +z-direction with the phase constant k_z. The waveguide mode has no longitudinal component of the magnetic field, and is called the transverse magnetic(TM) mode.

The boundary condition for $\mathbf{E}$ requires that the tangential component of $\mathbf{E}$ should be zero on the conducting surfaces at $y = 0$ and $y = d$. To satisfy the boundary condition, the sine term in Eq. (10-17) should vanish at $y = 0$ and $y = d$. It can be shown that the boundary condition leads to the same eigenvalues as given in Eq. (10-5). Accordingly, ***the relations given in Eqs. (10-5), (10-8)-(10-13), and (10-16) remain valid for the*** TM ***modes in the parallel-plate waveguide***.

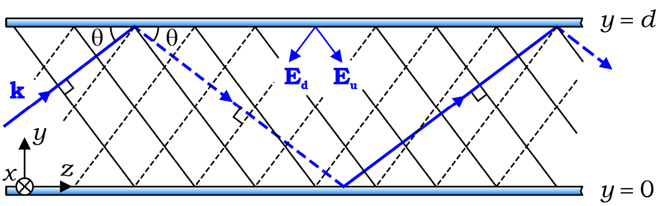


Fig. 10.6 TM mode in a parallel-plate waveguide.

Example 10-1
A parallel-plate waveguide as shown in Fig. 10.3 has an air gap of d[m].
(a) Show that the TE modes expressed by Eq. (10-4b) are mutually orthogonal such that $\int_{y=0}^{y=d} \mathbf{E}_m \cdot \mathbf{E}_n^* dy = 0$ for $m \neq n$.

(b) Find the orthonormal set in TE modes satisfying

$$\int_{y=0}^{y=d} \mathbf{E}_m \bullet \mathbf{E}_m^* dy = 1 .$$

Solution

(a) Upon using Eq. (10-4b), we write

$$\begin{aligned}\int_{y=0}^{y=d} \mathbf{E}_m \bullet \mathbf{E}_n^* dy &= 4E_o^2 \int_{y=0}^{y=d} \sin\left(\frac{m\pi}{d}y\right)\sin\left(\frac{n\pi}{d}y\right)dy \\ &= 2E_o^2 \int_{y=0}^{y=d}\left[\cos\left\{(m-n)\frac{\pi y}{d}\right\} - \cos\left\{(m+n)\frac{\pi y}{d}\right\}\right]dy \\ &= 0.\end{aligned}$$

(b) Assuming $m = n$ in part (a) we have

$$\int_{y=0}^{y=d} \mathbf{E}_m \bullet \mathbf{E}_m^* dy = 2dE_o^2$$

Dividing Eq. (10-4b) by $\sqrt{2dE_o^2}$, the orthonormal set in TE modes is

$$\boxed{\mathbf{E}_m = -j\sqrt{\frac{2}{d}}\,\mathbf{a}_x \sin\left(\frac{m\pi}{d}y\right)e^{-jk_z z}} \qquad (m = 1, 2, 3, \ldots) \qquad (10\text{-}19)$$

If an electromagnetic field is propagating in a parallel-plate waveguide, which is known to have no longitudinal component of **E**, then its electric field can be expanded in terms of the orthonormal set given in Eq. (10-19).

Example 10-2

A parallel-plate waveguide with an air gap of 1[cm] operates at a frequency 35[GHz]. The electric field is given as $\mathbf{E} = 10\,\mathbf{a}_x \sin^2(200\pi y)$ in the cross section at $z = 0$ in the waveguide. For TE_1, TE_2, and TE_3 modes, find

(a) cutoff frequencies,

(b) electric fields in the $z = 0$ plane, and

(c) electric fields in the $z = 2$[cm] plane.

[Hint: $\sin^2 a \sin b = \frac{1}{2}\sin b - \frac{1}{4}\sin(2a+b) + \frac{1}{4}\sin(2a-b)$.]

Solution

(a) From Eq. (10-9), the cutoff frequency for TE_m mode is

$$f_{c(m)} = \frac{mc}{2nd} = \frac{m \times 3\times10^8}{2\times10^{-2}} = m \times 15[\text{GHz}] .$$

Cutoff frequencies for TE_1, TE_2, and TE_3 modes are

$f_{c(1)} = 15[\text{GHz}]$, $f_{c(2)} = 30[\text{GHz}]$, and $f_{c(3)} = 45[\text{GHz}]$.

We note that $f_{c(1)}$ and $f_{c(2)}$ are above cutoff, while $f_{c(3)}$ is below cutoff.

(b) From Eq. (10-19) we can obtain the amplitude of the TE_1 mode, which is one of the component waves comprising **E** in the $z = 0$ plane, that is,

$$\int_{y=0}^{y=d} \mathbf{E} \bullet \mathbf{E}_1^* dy = j100\sqrt{2}\int_{y=0}^{y=0.01} \sin^2(200\pi y)\sin(100\pi y)\,dy \\ = j0.48. \tag{10-20a}$$

The amplitude of the TE_2 mode is

$$\int_{y=0}^{y=d} \mathbf{E} \bullet \mathbf{E}_2^* dy = j100\sqrt{2}\int_{y=0}^{y=0.01} \sin^2(200\pi y)\sin(200\pi y)\,dy \\ = 0. \tag{10-20b}$$

The amplitude of the TE_3 mode is

$$\int_{y=0}^{y=d} \mathbf{E} \bullet \mathbf{E}_3^* dy = j100\sqrt{2}\int_{y=0}^{y=0.01} \sin^2(200\pi y)\sin(300\pi y)\,dy \\ = j0.34. \tag{10-20c}$$

By use of Eqs. (10-19) and (10-20), the electric fields of the three component waves are obtained, in the $z = 0$ plane, as

$$\mathbf{E}_1 = 4.8\sqrt{2}\,\mathbf{a}_x \sin(100\pi y)\,. \tag{10-21a}$$

$$\mathbf{E}_2 = 0\,. \tag{10-21b}$$

$$\mathbf{E}_3 = 3.4\sqrt{2}\,\mathbf{a}_x \sin(300\pi y)\,. \tag{10-21c}$$

(c) The phase constant for the TE_1 mode is obtained from Eq. (10-7):

$$\beta_1 = \sqrt{\left(\frac{2\pi\times 35\times 10^9}{3\times 10^8}\right)^2 - \left(\frac{\pi}{0.01}\right)^2} = 662.3[\text{rad/m}]\,. \tag{10-22a}$$

The attenuation constant for the TE_3 mode is obtained from Eq. (10-16):

$$\alpha_3 = \frac{2\pi\times 45\times 10^9}{3\times 10^3}\sqrt{1-\left(\frac{35}{45}\right)^2} = 592[\text{Np/m}]\,. \tag{10-22b}$$

Inserting Eq. (10-21) and Eq. (10-22) into Eq. (10-19) we obtain, in the $z = 2[\text{cm}]$ plane,

$$\mathbf{E}_1 = 4.8\sqrt{2}\,\mathbf{a}_x \sin(100\pi y)e^{-j662.3\times 0.02} = 6.79\,\mathbf{a}_x \sin(100\pi y)e^{-j2.11}\,.$$

$$\mathbf{E}_2 = 0\,.$$

$$\mathbf{E}_3 = 3.4\sqrt{2}\,\mathbf{a}_x \sin(300\pi y)e^{-592\times 0.02} = 3.5\times 10^{-5}\,\mathbf{a}_x \sin(300\pi y)\,.$$

Exercise 10.3

What is the maximum frequency at which the parallel-plate waveguide with an air gap of 1.5[cm] is excited with a TEM mode only.

Ans. 10[GHz].

Exercise 10.4

A parallel-plate waveguide has a gap of $d = 1$ [cm], which is filled with a dielectric of $n = 2.5$, and operates at $f = 15$ [GHz]. Find the time delay between TM_1 and TM_2, if they have traveled 1[cm] along the waveguide.

Ans. 49[ps].

Review Questions with Hints

RQ 10.1 What are the three types of waveguide modes. [Figs.10.4,10.5,10.6]
RQ 10.2 Explain the transverse resonance condition. [Eq.(10-6)]
RQ 10.3 Define the phase constant of a mode of operation. [Eq.(10-8)]
RQ 10.4 Define the cutoff frequency of a waveguide. [Eq.(10-9)]
RQ 10.5 Distinguish between the phase and group velocities of a mode of operation in a waveguide. [Eqs.(10-11)(10-12)]

10.2 Rectangular Waveguides

In the discussion of the parallel-plate waveguide in Section 10-1, it was implied that the parallel conducting plates were infinite in extent. Accordingly, the electric field was uniform, or otherwise varied in one dimension, in the transverse plane of the waveguide. In that case, the propagation of an electromagnetic wave along the waveguide was easy to visualize. However, in most practical cases, the parallel conducting plates of a finite width give rise to the fringing effects of the fields at the edges, and cause losses of the electromagnetic energy through the sides of the waveguide. To avoid the energy loss, practical waveguides are made in the form of a hollow metal pipe with a uniform cross section. One of the most common waveguides is the rectangular waveguide, with a rectangular cross section, enclosed by four conducting walls. As in the case of the parallel-plate waveguide, the mode of propagation in the rectangular waveguide can be decomposed into uniform plane-wave components propagating, with the same phase constant, along the waveguide. Although it may not be easy, in general, to visualize the propagation of the electromagnetic wave in a rectangular waveguide, both rectangular and parallel-plate waveguides operate on the same basic principles.

For a rectangular waveguide as shown in Fig. 10.6, the field pattern of the mode of propagation can be obtained in two different ways. Firstly, we may begin with the three-dimensional differential wave equation or the vector Helmholtz equation to obtain general solutions for the electric and magnetic fields in the waveguide. By applying boundary conditions for $\mathscr{E}$ and $\mathscr{H}$ at the four conducting walls of the waveguide, we can obtain particular solutions for $\mathscr{E}$ and $\mathscr{H}$, or the mode of propagation in the waveguide. Secondly, we may begin with uniform plane waves, which are the simplest general solutions obtained from the differential wave equation or the vector Helmholtz equation in the medium, inside the waveguide, as if the material were infinite in extent. By linearly combining these plane waves in such a way as to satisfy the boundary condition at the conducting walls, we can obtain the expression for the mode of propagation in the waveguide. This approach is justified by the principle of superposition of the electromagnetic wave, and the uniqueness theorem, which states that a trial solution is a unique solution only if it satisfies both the differential wave equation and the boundary condition.

Following the second approach, we assume the mode of operation in a rectangular waveguide to be a superposition of the uniform plane waves with the same wavenumber k and the same phase constant in the z-direction, or k_z. In the general case, the rectangular waveguide is filled with a dielectric of permeability μ and permittivity ε, and operates at an angular frequency ω.

We write a general expression for the mode of propagation in a rectangular waveguide as

$$\begin{aligned}\mathscr{E}(\mathbf{r},t) &= \mathrm{Re}\left[\mathbf{E}(\mathbf{r})e^{j\omega t}\right] \\ &= \mathrm{Re}\left[\left\{E_x(x,y)\,\mathbf{a}_x + E_y(x,y)\,\mathbf{a}_y + E_z(x,y)\,\mathbf{a}_z\right\}e^{-\gamma z}e^{j\omega t}\right]\end{aligned} \tag{10-23}$$

From the above equation, we notice that the space coordinates x and y are separated from the space coordinate z, and that the three terms, $E_x(x,y)$, $E_y(x,y)$, and $E_z(x,y)$, describe the distributions of three electric field components in the transverse plane, whereas the term $e^{-\gamma z}$ describes the variation of the electric field in the longitudinal direction along the waveguide. The underlying assumption is that the electric field pattern moves in the z-direction, with the propagation constant γ, with no change in the electric field in the xy-plane. In phasor notation, the electric field in the rectangular waveguide is

$$\boxed{\mathbf{E}(\mathbf{r}) \equiv \left\{E_x(x,y)\,\mathbf{a}_x + E_y(x,y)\,\mathbf{a}_y + E_z(x,y)\,\mathbf{a}_z\right\}e^{-\gamma z}} \tag{10-24}$$

If a time-harmonic field $\mathscr{E}(\mathbf{r},t)$ represents the electric field of a mode of propagation in the waveguide, it should satisfy the differential wave equation, and its phasor $\mathbf{E}(\mathbf{r})$ should satisfy the vector Helmholtz equation, whether or not the medium is limited to the interior of the waveguide, namely,

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = 0 \tag{10-25}$$

The wavenumber $k = \omega\sqrt{\mu\varepsilon}$ in the dielectric of permeability μ and permittivity ε, and the operating angular frequency $\omega = 2\pi f$. Inserting Eq. (10-24) into Eq. (10-25) we have

$$\nabla_{xy}^2\left(E_x\mathbf{a}_x + E_y\mathbf{a}_y + E_z\mathbf{a}_z\right) + \left(\gamma^2 + k^2\right)\left(E_x\mathbf{a}_x + E_y\mathbf{a}_y + E_z\mathbf{a}_z\right) = 0 \tag{10-26}$$

The two-dimensional Laplacian operator is defined as

$$\nabla_{xy}^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{10-27}$$

Since the three unit vectors in Eq. (10-26) are mutually exclusive, Eq. (10-26) can be separated into three partial differential equations, having the same form, as

$$\boxed{\nabla_{xy}^2 E_x + \left(\gamma^2 + k^2\right) E_x = 0} \tag{10-28a}$$

$$\boxed{\nabla_{xy}^2 E_y + \left(\gamma^2 + k^2\right) E_y = 0} \tag{10-28b}$$

$$\boxed{\nabla_{xy}^2 E_z + \left(\gamma^2 + k^2\right) E_z = 0} \tag{10-28c}$$

where γ is the propagation constant measured along the waveguide, and k is the wavenumber of the component plane-waves, that is, $k = \omega\sqrt{\mu\varepsilon}$. Upon solving Eq. (10-28) for general solutions and applying the boundary condition for **E** at the conducting walls of the waveguide, we can uniquely determine the three electric field components, $E_x(x,y)$, $E_y(x,y)$, and $E_z(x,y)$, and the propagation constant γ.

By following the same procedure, we write a general expression for the magnetic field in the rectangular waveguide as

$$\begin{aligned} \mathscr{H}(\mathbf{r},t) &= \mathrm{Re}\left[\mathbf{H}(\mathbf{r})e^{j\omega t}\right] \\ &= \mathrm{Re}\left[\left\{H_x(x,y)\,\mathbf{a}_x + H_y(x,y)\,\mathbf{a}_y + H_z(x,y)\,\mathbf{a}_z\right\}e^{-\gamma z}e^{j\omega t}\right] \end{aligned} \tag{10-29}$$

The underlying assumption is that the magnetic field pattern moves with the same propagation constant as **E**, along the waveguide, with no change in the field distribution in the transverse plane. Similarly, the magnetic field phasor **H(r)** should satisfy the Helmholtz equation, namely

$$\nabla_{xy}^2\left(H_x\mathbf{a}_x + H_y\mathbf{a}_y + H_z\mathbf{a}_z\right) + \left(\gamma^2 + k^2\right)\left(H_x\mathbf{a}_x + H_y\mathbf{a}_y + H_z\mathbf{a}_z\right) = 0 \tag{10-30}$$

By solving Eq. (10-30) for a general solution and applying the boundary conditions for **H** at the conducting walls, we can uniquely determine the magnetic field in the waveguide and the propagation constant.

A rectangular waveguide can be excited with the TE mode containing no E_z-component, or with the TM mode containing no H_z-component. Can the rectangular waveguide also support the TEM mode? To answer this question, let us substitute **E(r)** and **H(r)** expressed by Eqs. (10-24) and (10-29) into the phasor form of Maxwell's equations, that is, $\nabla\times\mathbf{E} = -j\omega\mu\mathbf{H}$ and $\nabla\times\mathbf{H} = j\omega\varepsilon\mathbf{E}$, by assuming $E_z = 0 = H_z$. We then acknowledge that both **E** and **H** should be independent of x and y in the rectangular waveguide such that

$$\mathbf{E}(\mathbf{r}) = \left\{E_{xo}\mathbf{a}_x + E_{yo}\mathbf{a}_y\right\}e^{-\gamma z} \tag{10-31a}$$

$$\mathbf{H}(\mathbf{r}) = \left\{H_{xo}\mathbf{a}_x + H_{yo}\mathbf{a}_y\right\}e^{-\gamma z} \tag{10-31b}$$

where E_{xo}, E_{yo}, H_{xo}, and H_{yo} are constants. Application of the boundary conditions for **E** to the conducting walls of the waveguide leads us to conclude that both E_{xo} and E_{yo} should be zero in the interior and on the walls of the waveguide. It can also be shown that the magnetic field is zero everywhere. In consequence, ***the*** TEM ***mode is not supported in a rectangular waveguide***!

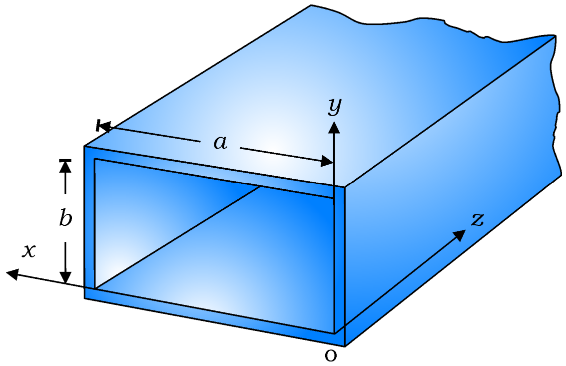


Fig. 10.7 A rectangular waveguide.

Exercise 10.5
Derive Eq. (10-31) by assuming $E_z = 0 = H_z$, and show that the fields are constant in the transverse plane.

Review Questions with Hints

RQ 10.6 What makes the TEM mode a nonpropagating wave in a rectangular waveguide? [Eq.(10-31)]
RQ 10.7 Distinguish between the propagation constant γ and the wavenumber k in a rectangular waveguide. [Eq.(10-28)]
RQ 10.8 What is the underlying assumption about the general expression for the waveguide mode given in Eq. (10-23)? [Eq.(10-4)]

10.2.1 Transverse Magnetic(TM) Modes

Although the TM mode has no longitudinal component of the magnetic field($H_z = 0$), it should have the longitudinal component of the electric field($E_z \neq 0$) in the rectangular waveguide. Otherwise, it would be the TEM mode that is not supported in the waveguide. If the electric field E_z is known in a rectangular waveguide, we can obtain the transverse components of the TM mode, such as E_x, E_y, H_x, and H_y, from E_z.

10.2.1.1 Longitudinal Field Component of a TM Mode

In a rectangular waveguide, the longitudinal component of the TM mode, or the electric field E_z, is governed by the differential equation given in Eq. (10-28c), that is,

$$\boxed{\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + h^2 E_z = 0} \tag{10-32}$$

where h is defined by

$$h^2 \equiv \gamma^2 + k^2 \tag{10-33}$$

In a lossless waveguide, it can be shown that the propagation constant γ is given by $\gamma = jk_z$, where k_z is the phase constant of the waveguide, or the z-component of the wavevector $\mathbf{k}$. In view of this, we note that the eigenvalue h is the projection of $\mathbf{k}$ onto the transverse plane, or onto the xy-plane, and is thus expressed as $h^2 = k_x^2 + k_y^2$. Again, γ is the propagation constant for the wave propagating along the waveguide, whereas k is the propagation constant for the plane wave propagating in the dielectric as if the dielectric were infinite in extent.

As a first step, let us assume that the complex amplitude $E_z(x, y)$ given in Eq. (10-24) is separable such that

$$E_z(x, y) = X(x)Y(y) \tag{10-34}$$

If we can show that Eq. (10-34) satisfies the boundary conditions at the conducting walls of the waveguide, the uniqueness theorem will assure that the trial solution is a unique solution in the waveguide. Inserting Eq. (10-34) into Eq. (10-32), and rearranging the terms, we have

$$-\frac{1}{X}\frac{d^2 X}{dx^2} = \frac{1}{Y}\frac{d^2 Y}{dy^2} + h^2 \tag{10-35}$$

From Eq. (10-35), we notice that the two space variables x and y are completely separated from each other. Since the two variables x and y are independent of each other, the equality in Eq. (10-35) may be satisfied only if both sides are equal to a constant, called a separation constant. By using k_x^2 as a separation constant, Eq. (10-35) can be separated into two parts as

$$\frac{d^2 X}{dx^2} + k_x^2 X = 0 \tag{10-36a}$$

$$\frac{d^2 Y}{dy^2} + \left(h^2 - k_x^2\right) Y = 0 \tag{10-36b}$$

General solutions of Eqs. (10-36a) and (10-36b) are

$$X = A\sin(k_x x) + B\cos(k_x x) \tag{10-37a}$$

$$Y = C\sin(k_y y) + D\cos(k_y y) \tag{10-37b}$$

where use was made of the definition:

$$k_y^2 \equiv h^2 - k_x^2 \tag{10-38}$$

The longitudinal electric field E_z as given in Eq. (10-34) may become the tangential components at the four conducting walls of the rectangular waveguide. Therefore, to satisfy the boundary conditions, the electric field E_z should vanish at the four conducting walls of the waveguide. Namely

$$E_z(x = 0, y) = 0 \tag{10-39a}$$

$$E_z(x = a, y) = 0 \tag{10-39b}$$

$$E_z(x, y = 0) = 0 \tag{10-39c}$$

$$E_z(x, y = b) = 0 \tag{10-39d}$$

Applying these boundary conditions to Eq. (10-37), we get

$$B = 0, \text{ and } k_x = \frac{m\pi}{a} \qquad (m = 1, 2, 3, \dots) \tag{10-40a}$$

$$D = 0, \text{ and } k_y = \frac{n\pi}{b} \qquad (n = 1, 2, 3, \dots) \tag{10-40b}$$

Upon substituting Eqs. (10-37) and (10-40) into Eq. (10-34), the longitudinal component of the TM mode is given as follows:

$$\boxed{E_z(x, y) = E_o \sin\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \qquad (m, n = 1, 2, 3, \dots) \tag{10-41}$$

$$\boxed{\mathscr{E}_z(x, y, z, t) = \mathrm{Re}\left[E_o \sin\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)e^{-\gamma z}e^{j\omega t}\right]} \qquad \text{[V/m]} \tag{10-42}$$

where

$$\boxed{\begin{aligned}\gamma = \pm\sqrt{h^2 - k^2} &= \pm j\sqrt{\omega^2\mu\varepsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \\ &\equiv \alpha_{mn} + j\beta_{mn}\end{aligned}} \qquad [\text{m}^{-1}] \tag{10-43}$$

$$\boxed{h^2 = k_x^2 + k_y^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \tag{10-44}$$

$$\boxed{k = \omega\sqrt{\mu\varepsilon}} \qquad \text{[rad/m]} \tag{10-45}$$

In the above equations, m and n are integers denoting the mode numbers. The mode number n should not be confused with the refractive index n. If the material parameters ε and μ are real, and the operating frequency ω is large enough to make the radicand positive in Eq. (10-43), then the attenuation constant α_{mn} is zero, and the propagation constant is given simply by $\gamma = -j\beta_{mn}$. In this case, the longitudinal electric field $\mathscr{E}_z$ given in Eq. (10-42) represents an electromagnetic wave propagating with the phase constant β_{mn} along the waveguide in the z-direction, while forming standing waves in the x- and y-directions. The sign conventions for α_{mn} and β_{mn} are taken so that both α_{mn} and β_{mn} are positive for a wave propagating in the $+z$-direction. The specific field pattern with integers m and n is designated as the TM_{mn} mode. The first subscript represents the number of maxima of $\mathscr{E}_z$ on the x-axis, whereas the second represents that of $\mathscr{E}_z$ on the y-axis, as are evident from Eq. (10-42).

The propagation constant for the TM_{mn} mode becomes zero at the operating frequency called the cutoff frequency, $f_{c(mn)}$. From Eq. (10-43), the cutoff frequency for the TM_{mn} mode is obtained as

$$\boxed{f_{c(mn)} = \frac{1}{2\sqrt{\mu\varepsilon}}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}} \qquad \text{[Hz]} \qquad (10\text{-}46)$$

Below the cutoff frequency, the radicand in Eq. (10-43) becomes negative. In this case, the positive sign in front of the square-root sign in Eq. (10-43) is taken so that the propagation constant is given by a positive real number, that is, $\gamma = \alpha_{mn} > 0$. Under this condition, the TM_{mn} mode becomes an evanescent mode that cannot propagate in the waveguide.

By making use of the relation $v_p = 1/\sqrt{\mu\varepsilon} = f\lambda$, the cutoff frequency expressed by Eq. (10-46) can be converted into the cutoff wavelength for the TM_{mn} mode, that is,

$$\boxed{\lambda_{c(mn)} = \frac{v_p}{f_{c(mn)}} = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}}} \qquad \text{[m]} \qquad (10\text{-}47)$$

An electromagnetic wave of a wavelength $\lambda < \lambda_{c(mn)}$ may propagate in the rectangular waveguide as the TM_{mn} mode.

The smallest possible value of the mode number m or n for TM_{mn} modes is one. Accordingly, the TM_{11} mode has the lowest cutoff frequency. In a rectangular waveguide with dimensions $a > b$, the TM_{21} mode has the second lowest cutoff frequency. Thus, for instance, the electromagnetic wave with the operating frequency in the range $f_{c(11)} < f < f_{c(21)}$ can propagate in the waveguide only as the TM_{11} mode.

An electromagnetic wave can propagate in the waveguide as the TM_{mn} mode only if the operating frequency is higher than the cutoff frequency, that is, $f > f_{c(mn)}$. Above cutoff, the phase constant for the TM_{mn} mode is

$$\boxed{\beta_{mn} = k\sqrt{1-\left(\frac{f_{c(mn)}}{f}\right)^2}} \qquad \text{[rad/m]} \qquad (10\text{-}48)$$

This has been obtained by inserting Eq. (10-46) into Eq. (10-43). The phase constant β_{mn} is always smaller than the wavenumber $k = \omega\sqrt{\mu\varepsilon}$, which is the phase constant of an unbounded dielectric. As the wavelength λ is related to the wavenumber k by $\lambda = 2\pi / k$, the waveguide wavelength λ_{mn} is related to the phase constant β_{mn} by

$$\boxed{\lambda_{mn} = \frac{2\pi}{\beta_{mn}} = \frac{\lambda}{\sqrt{1-\left(f_{c(mn)}/f\right)^2}}} \qquad \text{[m]} \qquad (10\text{-}49)$$

In the above equation, λ_{mn} is the spatial period of the wave measured along the waveguide, whereas λ is the wavelength measured in an unbounded dielectric. In other words, the field pattern of the TM_{mn} mode repeats itself every distance λ_{mn} along the waveguide.

The waveguide wavelength λ_{mn} expressed by Eq. (10-49) should not be confused with the cutoff wavelength $\lambda_{c(mn)}$ expressed by Eq. (10-47), which is measured in an unbounded dielectric. With the help of the phase velocity $\upsilon_p = f\lambda = f_{c(mn)}\lambda_{c(mn)}$, we rewrite Eq. (10-49) as

$$\boxed{\frac{1}{\lambda^2} = \frac{1}{\lambda_{mn}^2} + \frac{1}{\lambda_{c(mn)}^2}} \qquad (10\text{-}50)$$

where λ is the wavelength, λ_{mn} is the waveguide wavelength, and $\lambda_{c(mn)}$ is the cutoff wavelength.

To summarize, the phase constant β_{mn} and the waveguide wavelength λ_{mn} for the TM_{mn} mode are related by $\beta_{mn} = 2\pi / \lambda_{mn}$; they are both measured along the waveguide. The wavenumber k is defined by $k = \omega\sqrt{\mu\varepsilon} = \omega / \upsilon_p$ in the unbounded dielectric. If the lossless dielectric is nonmagnetic($\mu = \mu_o$), the wavenumber is expressed as $k = n\omega / c$, and the phase velocity as $\upsilon_p = c / n$, where the refractive index $n = \sqrt{\varepsilon / \varepsilon_o}$.

The phase velocity in an unbounded dielectric can be expressed as

$$\upsilon_p = \frac{\omega}{k} = f\lambda = f_{c(mn)}\lambda_{c(mn)} \tag{10-51}$$

where $k = \omega\sqrt{\mu\varepsilon}$. In contrast, the phase velocity of the TM_{mn} mode is expressed as

$$\upsilon_{p(mn)} = \frac{\omega}{\beta_{mn}} = f\lambda_{mn} \tag{10-52a}$$

which is measured along the waveguide. Combining Eq. (10-51) with Eq. (10-52a), we obtain a useful relation, that is,

$$\boxed{\upsilon_{p(mn)} = \frac{\upsilon_p}{\sqrt{1-\left(f_{c(mn)}/f\right)^2}}} \tag{10-52b}$$

where $\upsilon_p = 1/\sqrt{\mu\varepsilon}$.

10.2.1.2 Transverse Field Components of a TM Mode

Earlier, we obtained the expression for the longitudinal component of the TM mode, E_z, which is useful for describing the propagation behavior of the mode in the rectangular waveguide. We now turn our attention to the transverse components of the TM mode, such as E_x, E_y, H_x, and H_y, which are obtained from a given E_z.

As was stated earlier, the TM wave may be considered as a linear combination of the uniform plane waves propagating with the same phase constant, having the same wavenumber, in the waveguide. In view of this, we write general expressions for the electric and magnetic fields of the TM wave in the rectangular waveguide as

$$\begin{aligned}\mathscr{E}(x,y,z,t) &= \text{Re}\left[\mathbf{E}e^{j\omega t}\right] \\ &= \text{Re}\left[\left\{E_x(x,y)\,\mathbf{a}_x + E_y(x,y)\,\mathbf{a}_y + E_z(x,y)\,\mathbf{a}_z\right\}e^{-\gamma z}e^{j\omega t}\right]\end{aligned} \tag{10-53a}$$

$$\begin{aligned}\mathscr{H}(x,y,z,t) &= \text{Re}\left[\mathbf{H}e^{j\omega t}\right] \\ &= \text{Re}\left[\left\{H_x(x,y)\,\mathbf{a}_x + H_y(x,y)\,\mathbf{a}_y\right\}e^{-\gamma z}e^{j\omega t}\right]\end{aligned} \tag{10-53b}$$

where the longitudinal component $E_z(x,y)$ is given by Eq. (10-41).

Upon substituting Eq. (10-53) into Maxwell's equations in phasor form, we have

(1) From $\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}$,

$$\frac{\partial E_z}{\partial y} + \gamma E_y = -j\omega\mu H_x \tag{10-54a}$$

$$-\gamma E_x - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \tag{10-54b}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \tag{10-54c}$$

(2) From $\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E}$,

$$\gamma H_y = j\omega\varepsilon E_x \tag{10-55a}$$

$$-\gamma H_x = j\omega\varepsilon E_y \tag{10-55b}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \tag{10-55c}$$

Combining the expressions given in Eqs. (10-54) and (10-55), we can obtain E_x, E_y, H_x, and H_y in terms of E_z. For instance, combination of Eq. (10-54b) and Eq. (10-55a) gives E_x expressed in terms of E_z. Here, for future reference, we repeat the longitudinal electric field of the TM_{mn} mode, or E_z given in Eq. (10-41), as

$$\boxed{E_z(x,y) = E_o \sin\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \qquad (m,n = 1,2,3,\dots) \tag{10-41}$$

Inserting Eq. (10-41) into Eqs (10-54) and (10-55), and manipulating the derivatives, the transverse fields of the TM_{mn} mode are obtained as

$$\boxed{E_x = -\frac{\gamma}{h^2}\left(\frac{m\pi}{a}\right)E_o \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \qquad (m,n = 1,2,3,\dots) \tag{10-56a}$$

$$\boxed{E_y = -\frac{\gamma}{h^2}\left(\frac{n\pi}{b}\right)E_o \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)} \tag{10-56b}$$

$$\boxed{H_x = \frac{j\omega\varepsilon}{h^2}\left(\frac{n\pi}{b}\right)E_o \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)} \tag{10-56c}$$

$$\boxed{H_y = -\frac{j\omega\varepsilon}{h^2}\left(\frac{m\pi}{a}\right)E_o \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \tag{10-56d}$$

Note that the propagation constant is given by $\gamma = j\beta_{mn}$ in the lossless waveguide. The eigenvalue h is just equal to the transverse component of the wavevector $\mathbf{k}$, that is,

$$h^2 = \gamma^2 + k^2 \\ = k_x^2 + k_y^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 \qquad (10\text{-}57)$$

In the above equations, the mode numbers are positive integers: $m = 1, 2, 3\ldots$ and $n = 1, 2, 3\ldots$. Finally, the electric fields expressed by Eq. (10-41) and Eq. (10-56) comprise the TM_{mn} mode propagating in the rectangular waveguide with dimensions a and b.

The wave impedance of the TM_{mn} mode, propagating in the lossless rectangular waveguide($\alpha_{mn} = 0$), is defined as

$$\boxed{\eta_{TM} = \frac{E_x}{H_y} = -\frac{E_y}{H_x}} \quad [\Omega] \qquad (10\text{-}58a)$$

Substitution of Eq. (10-55a) or Eq. (10-55b) into Eq. (10-58a) leads to the wave impedance of the TM_{mn} mode, that is,

$$\eta_{TM} = \frac{\beta_{mn}}{\omega\varepsilon} \qquad (10\text{-}58b)$$

where use was made of the relation $\gamma = j\beta_{mn}$ by assuming a lossless waveguide with $\alpha_{mn} = 0$. Upon substituting Eq. (10-48) into Eq. (10-58b), the wave impedance of the TM_{mn} mode is expressed as

$$\boxed{\eta_{TM} = \sqrt{\frac{\mu}{\varepsilon}}\sqrt{1 - \left(\frac{f_{c(mn)}}{f}\right)^2}} \quad [\Omega] \qquad (10\text{-}58c)$$

The wave impedance η_{TM} depends on frequency, and is always less than the intrinsic impedance of the dielectric($\eta = \sqrt{\mu/\varepsilon}$) above cutoff($f > f_{c(mn)}$).

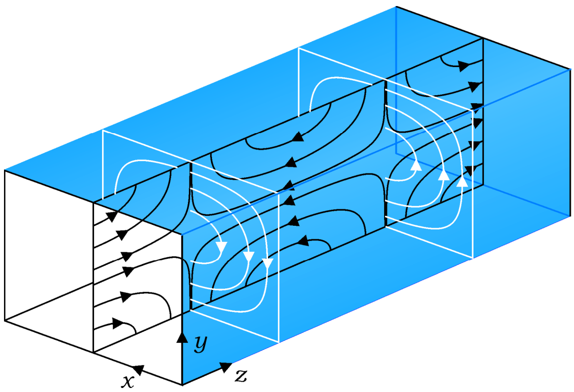


Fig. 10.8 TM_{11} mode in a rectangular waveguide. Black lines, **E**-field lines; White lines, **H**-field lines.

10.2.1.3 Orthonormal Set in TM Modes

The transverse **H**-fields of different TM modes are mutually orthogonal in the transverse plane of the rectangular waveguide such that

$$\int_{x=0}^{x=a}\int_{y=0}^{y=b} \mathbf{H}_{mn} \bullet \mathbf{H}_{m'n'}^{*} dxdy = 0 \qquad (m \neq m' \text{ or } n \neq n') \tag{10-59}$$

This relation can be readily verified by direct substitution of Eq. (10-56). We first normalize the transverse **H**-field of the TM mode, propagating in a lossless rectangular waveguide, so that

$$\int_{x=0}^{x=a}\int_{y=0}^{y=b} \mathbf{H}_{mn} \bullet \mathbf{H}_{mn}^{*} dxdy = 1 \tag{10-60}$$

Upon inserting Eq. (10-56) into Eq. (10-60), assuming a lossless waveguide with $\alpha_{mn} = 0$ and $\gamma = j\beta_{mn}$, we obtain

$$\begin{aligned}
&\int_{x=0}^{x=a}\int_{y=0}^{y=b} \mathbf{H}_{mn} \bullet \mathbf{H}_{mn}^{*} dxdy \\
&= \int_0^a\int_0^b \left[\left(H_x\mathbf{a}_x + H_y\mathbf{a}_y\right)e^{-j\beta_{mn}z}\right]\bullet\left[\left(H_x\mathbf{a}_x + H_y\mathbf{a}_y\right)e^{-j\beta_{mn}z}\right]^{*} dxdy \\
&= \int_0^a\int_0^b \left(\left|H_x\right|^2 + \left|H_y\right|^2\right) dxdy \\
&= \omega^2\varepsilon^2 \frac{E_o^2}{h^2}\frac{ab}{4}
\end{aligned} \tag{10-61}$$

where ω is the operating angular frequency, ε is the permittivity of the dielectric in the waveguide, h is the eigenvalue given in Eq. (10-57), a and b are the dimensions of the waveguide. We take the square root of Eq. (10-61), and use it as a normalization factor for Eqs. (10-56c) and (10-56d). Thus, the normalized **H**-field of the TM_{mn} mode is generally written as

$$\boxed{\bar{\mathbf{H}}_{mn} = \left(\bar{H}_x\,\mathbf{a}_x + \bar{H}_y\,\mathbf{a}_y\right)e^{-j\beta_{mn}z}} \tag{10-62a}$$

where the bar on the top is to denote the normalization. The normalized x- and y-components of the **H**-field of the TM_{mn} mode are

$$\boxed{\bar{H}_x = \frac{2j}{h\sqrt{ab}}\left(\frac{n\pi}{b}\right)\sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)} \qquad (m, n = 1, 2, 3, \ldots) \tag{10-62b}$$

$$\boxed{\bar{H}_y = -\frac{2j}{h\sqrt{ab}}\left(\frac{m\pi}{a}\right)\cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \qquad (m, n = 1, 2, 3, \ldots) \tag{10-62c}$$

where $j = \sqrt{-1}$, $h^2 = (m\pi / a)^2 + (n\pi / b)^2$, and a and b are the dimensions of the rectangular waveguide. The **H**-fields expressed by Eq. (10-62) constitute an orthonormal set in the TM modes in the rectangular waveguide: they are normalized to one, and, at the same time, are mutually orthogonal.

Example 10-3
A rectangular waveguide has interior dimensions $a = 1.5[\text{cm}]$ and $b = 1[\text{cm}]$. It is filled with a nonmagnetic, lossless dielectric($\mu = \mu_o$, $\varepsilon_r = 1.96$), and excited with an electromagnetic field($f = 15[\text{GHz}]$), which propagates as the TM_{11} mode in the waveguide . Find (a) λ, (b) $f_{c(11)}$, and (c) λ_{11}.

Solution
(a) Wavenumber in an unbounded dielectric

$$k = \frac{2\pi}{\lambda} = \omega\sqrt{\mu_o \varepsilon} = \frac{2\pi \times 15 \times 10^9 \times 1.4}{3 \times 10^8}$$

Thus, the wavelength is

$\lambda = 14.3[\text{mm}]$.

(b) From Eq. (10-46), the cutoff frequency is

$$f_{c(11)} = \frac{1}{2 \times 1.4 \times \sqrt{\mu_o \varepsilon_o}} \sqrt{\left(\frac{1}{1.5 \times 10^{-2}}\right)^2 + \left(\frac{1}{10^{-2}}\right)^2} = 12.9[\text{GHz}].$$

(c) From Eq. (10-49), the waveguide wavelength is

$$\lambda_{11} = \frac{\lambda}{\sqrt{1 - \left(f_{c(11)}/f\right)^2}} = \frac{14.3 \times 10^{-3}}{\sqrt{1 - (12.9/15)^2}} = 28.0[\text{mm}].$$

Example 10-4
How many component plane-waves does the TM_{21} mode have in a rectangular waveguide with dimensions a and b?

Solution
From Eq. (10-41) we write

$$\begin{aligned} E_z &= E_o \sin\left(\frac{2\pi}{a}x\right)\sin\left(\frac{\pi}{b}y\right) \\ &= E_0 \frac{1}{2j}\left(e^{j2\pi x/a} - e^{-j2\pi x/a}\right)\frac{1}{2j}\left(e^{j\pi y/b} - e^{-j\pi y/b}\right) \\ &= -\frac{E_0}{4}\left[e^{j2\pi x/a + j\pi y/b} - e^{j2\pi x/a - j\pi y/b} - e^{-j2\pi x/a + j\pi y/b} + e^{-j2\pi x/a - j\pi y/b}\right] \end{aligned}$$

From Eq. (10-56a) write

$$E_x = -\frac{\gamma}{h^2}\left(\frac{2\pi}{a}\right)E_o \cos\left(\frac{2\pi}{a}x\right)\sin\left(\frac{\pi}{b}y\right)$$
$$= -\frac{\gamma}{h^2}\left(\frac{2\pi}{a}\right)E_o \frac{1}{2}\left(e^{j2\pi x/a} + e^{-j2\pi x/a}\right)\frac{1}{2j}\left(e^{j\pi y/b} - e^{-j\pi y/b}\right)$$
$$= -\frac{\gamma}{h^2}\left(\frac{2\pi}{a}\right)\frac{E_o}{4j}\left[e^{j2\pi x/a+j\pi y/b} - e^{j2\pi x/a-j\pi y/b} + e^{-j2\pi x/a+j\pi y/b} - e^{-j2\pi x/a-j\pi y/b}\right]$$

From Eq. (10-56b) write

$$E_y = -\frac{\gamma}{h^2}\left(\frac{\pi}{b}\right)E_o \sin\left(\frac{2\pi}{a}x\right)\cos\left(\frac{\pi}{b}y\right)$$
$$= -\frac{\gamma}{h^2}\left(\frac{\pi}{b}\right)E_o \frac{1}{2j}\left(e^{j2\pi x/a} - e^{-j2\pi x/a}\right)\frac{1}{2}\left(e^{j\pi y/b} + e^{-j\pi y/b}\right)$$
$$= -\frac{\gamma}{h^2}\left(\frac{\pi}{b}\right)\frac{E_o}{4j}\left[e^{j2\pi x/a+j\pi y/b} + e^{j2\pi x/a-j\pi y/b} - e^{-j2\pi x/a+j\pi y/b} - e^{-j2\pi x/a-j\pi y/b}\right]$$

The three components of the electric field can be combined into the electric field phasor, in a lossless rectangular waveguide($\gamma = j\beta_{21}$), as

$$\mathbf{E} = \left\{E_x(x,y)\mathbf{a}_x + E_y(x,y)\mathbf{a}_y + E_z(x,y)\mathbf{a}_z\right\}e^{-j\beta_{21}z}$$

In view of these, we note that the TM_{21} mode comprises four plane waves, $e^{j2\pi x/a+j\pi y/b-j\beta_{21}z}$, $e^{j2\pi x/a-j\pi y/b-j\beta_{21}z}$, $e^{-j2\pi x/a+j\pi y/b-j\beta_{21}z}$, and $e^{-j2\pi x/a-j\pi y/b-j\beta_{21}z}$

It is important to note that the four component plane-waves have the same wavenumber k as well as the same phase constant β_{21}.

Example 10-5

In a hollow rectangular waveguide with interior dimensions $a = 5[\mathrm{cm}]$ and $b = 2.5[\mathrm{cm}]$, an electromagnetic wave of frequency $8[\mathrm{GHz}]$ propagates in the $+z$-direction. In the transverse plane at $z = 0$, the magnetic field is given by $\mathbf{H} = 4\mathbf{a}_x \sin(20\pi x)\sin(60\pi y) + 3\mathbf{a}_y \cos(20\pi x)\cos(60\pi y)[\mathrm{A/m}]$ with no z-component. Find

(a) TM modes of propagation,

(b) α_{mn} for the evanescent mode with the lowest cutoff frequency, and

(c) magnetic field in the transverse plane at $z = 1[\mathrm{m}]$.

Solution

(a) From Eq. (10-46), the cutoff frequencies are

$$f_{c(11)} = \frac{1}{2\sqrt{\mu_o\varepsilon_o}}\sqrt{\left(\frac{1}{0.05}\right)^2 + \left(\frac{1}{0.025}\right)^2} = 6.7[\mathrm{GHz}]$$

$$f_{c(21)} = \frac{1}{2\sqrt{\mu_o \varepsilon_o}} \sqrt{\left(\frac{2}{0.05}\right)^2 + \left(\frac{1}{0.025}\right)^2} = 8.5[\text{GHz}]$$

$$f_{c(12)} = \frac{1}{2\sqrt{\mu_o \varepsilon_o}} \sqrt{\left(\frac{1}{0.05}\right)^2 + \left(\frac{2}{0.025}\right)^2} = 12.4[\text{GHz}]$$

The TM_{21} and TM_{12} modes are the evanescent modes in the waveguide operating at $8[\text{GHz}]$.

(b) From Eq. (10-43), the attenuation constant for TM_{21} mode is

$$\alpha_{21} = \pi \sqrt{\left(\frac{2}{0.05}\right)^2 + \left(\frac{1}{0.025}\right)^2 - \left(\frac{2\times 8\times 10^9}{3\times 10^8}\right)^2} = 59[\text{Np/m}]$$

In view of the large value of α_{21}, the evanescent wave can be neglected at the output end at $z = 1[\text{m}]$ of the waveguide.

(c) We obtain $\bar{\mathbf{H}}_{11}$ from Eq. (10-62) as

$$\bar{\mathbf{H}}_{11} = \frac{2j}{h\sqrt{ab}}\left(\frac{\pi}{0.025}\right)\mathbf{a}_x \sin(20\pi x)\cos(40\pi y) - \frac{2j}{h\sqrt{ab}}\left(\frac{\pi}{0.05}\right)\mathbf{a}_y \cos(20\pi x)\sin(40\pi y)$$

The magnitude of the TM_{11} mode contained in **H** is obtained as follows:

$$\begin{aligned}
&\int_{x=0}^{x=0.05}\int_{y=0}^{y=0.25} \mathbf{H}\bullet\bar{\mathbf{H}}_{11}^{*}\,dxdy \\
&= -\frac{8j}{h\sqrt{ab}}\left(\frac{\pi}{0.025}\right)\int_{x=0}^{x=0.05}\sin^2(20\pi x)dx\int_{y=0}^{y=0.025}\sin(60\pi y)\cos(40\pi y)dy \\
&\quad + \frac{6j}{h\sqrt{ab}}\left(\frac{\pi}{0.05}\right)\int_{x=0}^{x=0.05}\cos^2(20\pi x)dx\int_{y=0}^{y=0.025}\cos(60\pi y)\sin(40\pi y)dy \\
&= -\frac{0.3j}{h\sqrt{ab}}
\end{aligned} \tag{10-63}$$

Multiplying Eq. (10-62) by the factor given by Eq. (10-63) we obtain

$$H_x = 3.06\sin\left(\frac{\pi}{0.05}x\right)\cos\left(\frac{\pi}{0.025}y\right) \quad [\text{A/m}] \tag{10-64a}$$

$$H_y = -1.53\cos\left(\frac{\pi}{0.05}x\right)\sin\left(\frac{\pi}{0.025}y\right) \quad [\text{A/m}] \tag{10-64b}$$

From Eq. (10-48) we obtain the phase constant for the TM_{11} mode as

$$\beta_{mn} = k\sqrt{1-\left(\frac{f_{c(mn)}}{f}\right)^2} = \frac{2\pi\times 8\times 10^9}{3\times 10^8}\sqrt{1-\left(\frac{6.7}{8}\right)^2} = 91.6[\mathrm{rad/m}]$$

(10-65)

In the plane at $z = 1[\mathrm{m}]$, Eqs.(10-64) and (10-65) are combined together to write the magnetic field phasor, that is,

$$\mathbf{H} = \left[H_x(x,y)\,\mathbf{a}_x + H_y(x,y)\,\mathbf{a}_y\right]e^{-j91.6}\,.$$

Exercise 10.6

Two rectangular waveguides are identical in dimensions. The first one is filled with a nonmagnetic, lossless dielectric($\varepsilon_r = 2.25$), while the second one is filled with air. Find the ratio between the cutoff frequencies for TM_{mn} mode.

Ans. $f_{c(mn)}^{①} / f_{c(mn)}^{②} = 1/1.5$.

Exercise 10.7

Find the group velocity of the TM_{mn} mode in a rectangular waveguide.

Ans. $$v_{g(mn)} = \left(d\beta_{mn}/d\omega\right)^{-1} = \left(c/n\right)\sqrt{1-\left(f_{c(mn)}/f\right)^2}\,. \qquad (10\text{-}66)$$

Review Questions with Hints

RQ 10.9 Define h, γ, and k in a rectangular waveguide. [Eq.(10-33)]

RQ 10.10 State the boundary condition for E_z of the TM mode propagating in a rectangular waveguide. [Eq.(10-39)]

RQ 10.11 Which TM mode has the lowest cutoff frequency in a rectangular waveguide? [Eq.(10-46)]

10.2.2 Transverse Electric(TE) Modes

The transverse electric(TE) mode has no longitudinal component of $\mathbf{E}$. Therefore, a waveguide mode with a nonzero H_z must be a TE mode, because it is neither a TEM mode nor a TM mode. The TE mode with the propagation constant γ is generally written as

$$\boxed{\mathscr{H}_z(x,y,z,t) = \mathrm{Re}\left[H_z(x,y)e^{-\gamma z}e^{j\omega t}\right]} \qquad [\mathrm{A/m}] \qquad (10\text{-}67)$$

Recalling that we solved Eq. (10-28c) for E_z of the TM mode, we now solve the same Helmholtz's equation for H_z of the TE mode, namely,

$$\boxed{\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + h^2 H_z = 0} \tag{10-68}$$

where the constant h is the transverse component of the wavevector $\mathbf{k}$(see Eq. (10-33)). Following the same logic that led, from Eq. (10-34), to Eq. (10-37), we write a general solution for H_z, that is,

$$H_z(x, y) = \left[A\sin(k_x x) + B\cos(k_x x)\right]\left[C\sin(k_y y) + D\cos(k_y y)\right] \tag{10-69}$$

Although the longitudinal magnetic field expressed by Eq. (10-69) has unknowns, *A*, *B*, *C*, and *D*, we may proceed to find the relations between H_z and the transverse electric and magnetic fields of the TE mode.

(1) From Faraday's law, $\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}$, we write

$$\gamma E_y = -j\omega\mu H_x \tag{10-70a}$$

$$-\gamma E_x = -j\omega\mu H_y \tag{10-70b}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \tag{10-70c}$$

(2) From Ampere's law, $\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E}$, by assuming a perfect dielectric($\mathbf{J} = 0$) in the waveguide, we write

$$\frac{\partial H_z}{\partial y} + \gamma H_y = j\omega\varepsilon E_x \tag{10-71a}$$

$$-\gamma H_x - \frac{\partial H_z}{\partial x} = j\omega\varepsilon E_y \tag{10-71b}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = 0, \tag{10-71c}$$

Upon combining Eqs. (10-70b) and (10-71a), and combining Eqs. (10-70a) and (10-71b), we have

$$E_x = -\frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y} \tag{10-72a}$$

$$E_y = \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x} \tag{10-72b}$$

$$H_x = -\frac{\gamma}{h^2}\frac{\partial H_z}{\partial x} \tag{10-72c}$$

$$H_y = -\frac{\gamma}{h^2}\frac{\partial H_z}{\partial y} \tag{10-72d}$$

According to the boundary condition for the electric field, the transverse components E_x and E_y should vanish on the conducting wall of the waveguide, if they are tangential to the wall. Namely

$$E_x(x, y=0) = -\frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y}\bigg|_{y=0} = 0 \tag{10-73a}$$

$$E_x(x, y=b) = -\frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y}\bigg|_{y=b} = 0 \tag{10-73b}$$

$$E_y(x=0, y) = \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x}\bigg|_{x=0} = 0 \tag{10-73c}$$

$$E_y(x=a, y) = \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x}\bigg|_{x=a} = 0 \tag{10-73d}$$

Substitution of Eq. (10-69) into Eq. (10-73) simply gives $A = C = 0$. The longitudinal magnetic field of the TE mode is thus given as

$$\boxed{H_z(x,y) = H_o \cos\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)} \qquad (m, n = 0, 1, 2, \ldots) \tag{10-74}$$

Upon inserting Eq. (10-74) into Eq. (10-72), the transverse field components of the TE_{mn} mode are obtained as

$$\boxed{E_x = \frac{j\omega\mu}{h^2}\left(\frac{n\pi}{b}\right)H_o \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \quad (m, n = 0, 1, 2, \ldots) \tag{10-75a}$$

$$\boxed{E_y = -\frac{j\omega\mu}{h^2}\left(\frac{m\pi}{a}\right)H_o \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)} \tag{10-75b}$$

$$\boxed{H_x = \frac{\gamma}{h^2}\left(\frac{m\pi}{a}\right)H_o \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)} \tag{10-75c}$$

$$\boxed{H_y = \frac{\gamma}{h^2}\left(\frac{n\pi}{b}\right)H_o \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \tag{10-75d}$$

Even though the mode number m or n may be zero, they should not be both zero; TE_{00} mode is not defined in the rectangular waveguide.

Comparison of the field patterns of the TE and TM modes shows that the longitudinal field H_z of the TE mode contains two cosine functions, whereas the longitudinal component E_z of the TM mode contains two sine functions. Meanwhile, all the transverse components of the TE and TM modes contain a sine and a cosine function.

The definitions of h, γ, and k for the TM mode given in Eqs. (10-43)-(10-45) are also valid for the TE mode. Thus the expressions for the quantities, such as $f_{c(mn)}$, $\lambda_{c(mn)}$, β_{mn}, λ_{mn}, v_p, and $v_{p(mn)}$ given in Eqs. (10-46)-(10-52), remain valid for the TE_{mn} mode.

The waveguide mode with the lowest possible cutoff frequency is called the dominant mode. In the rectangular waveguide with $a > b$, the TE_{10} mode has the lowest cutoff frequency, and is thus the dominant mode of the waveguide. In this case, the cutoff frequency for the dominant mode is

$$\boxed{f_{c(10)} = \frac{1}{2a\sqrt{\mu\varepsilon}} = \frac{\upsilon_p}{2a}} \quad \text{[Hz]} \tag{10-76}$$

We also note that the TE_{10} mode has the lowest attenuation constant of all modes in the rectangular waveguide; the TE_{10} mode has the electric field directed along the y-axis only.

The wave impedance for the TE_{mn} mode is defined from Eq. (10-70a) and Eq. (10-70b) as

$$\eta_{TE} = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{\omega\mu}{\beta_{mn}} \tag{10-77a}$$

Upon substituting Eq. (10-48) into Eq. (10-77a), the wave impedance for the TE_{mn} mode is expressed as

$$\boxed{\eta_{TE} = \sqrt{\frac{\mu}{\varepsilon}} \frac{1}{\sqrt{1-\left(f_{c(mn)}/f\right)^2}}} \quad [\Omega] \tag{10-77b}$$

The wave impedance, η_{TM} or η_{TE}, is purely resistance for the wave propagating in the rectangular waveguide filled with a lossless dielectric, because of real ε and μ. Although η_{TM} and η_{TE} both depend on the operating frequency, the product of the two wave impedances is given by a constant value; that is, $\eta_{TM}\eta_{TE} = \eta^2$, where the intrinsic impedance of the dielectric $\eta = \sqrt{\mu/\varepsilon}$ [Ω].

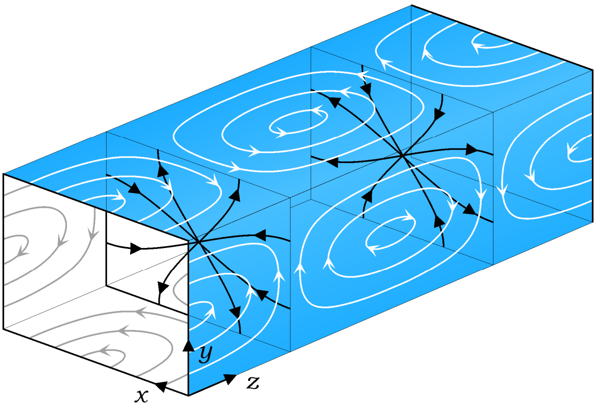


Fig. 10.9 Field patterns of TE_{11} mode. Black lines, **E**-field lines; White lines, **H**-field lines.

Note from Eq. (10-43) that if the TE_{mn} and TM_{mn} waves have the same mode numbers, they propagate with the same phase constant β_{mn} in the rectangular waveguide. From Eq. (10-48) we see that the phase constant depends on the frequency of operation. The relation between β and ω is called the dispersion relation. By making use of the relations $k = \omega / \upsilon_p$ and $\omega_{c(mn)} = 2\pi f_{c(mn)}$ we can rewrite Eq. (10-48) as follows:

$$\omega = \frac{\beta_{mn}\upsilon_p}{\sqrt{1-\left(\omega_{c(mn)}/\omega\right)^2}} \tag{10-78}$$

The ω–β diagram shown in Fig. 10.10 can be conveniently used to explain the dispersion relation for a mode of operation. For the mode with the cutoff frequency $\omega_{c(mn)}$, the slope of the tangent to the ω–β curve, at point *p*, represents the group velocity of the mode of propagation with the phase constant β_{mn}, whereas the slope of the line joining the origin and point *p* represents the phase velocity of the mode. As are apparent from Fig. 10.10, the TM and TE modes have the same phase and group velocities, and the phase velocity $\upsilon_{p(mn)}$ and the group velocity $\upsilon_{g(mn)}$ satisfy the relation $\upsilon_{g(mn)} < \upsilon_p < \upsilon_{p(mn)}$, where υ_p is the phase velocity in an unbounded dielectric. As the operating frequency ω increases much higher than the cutoff frequency, both $\upsilon_{p(mn)}$ and $\upsilon_{g(mn)}$ approach an asymptotic value of υ_p.

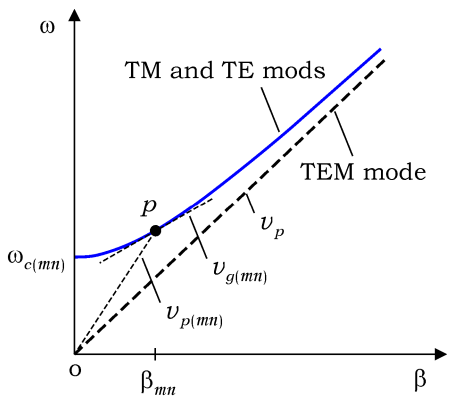


Fig. 10.10 ω-β diagram for a mode with cutoff frequency ω_c.

10.2.2.1 Orthonormal Set in TE Modes

Following the same procedure as was used for the TM mode, we can show that the transverse **E**-fields of different TE modes in a rectangular waveguide are mutually orthogonal such that

$$\int_{x=0}^{x=a}\int_{y=0}^{y=b}\mathbf{E}_{mn}\bullet\mathbf{E}^*_{m'n'}dxdy=0 \qquad (m\neq m' \text{ or } n\neq n') \tag{10-79}$$

This can be readily verified by direct substitution of Eq. (10-75). The transverse electric field of the TE mode can be normalized in the transverse plane of the lossless rectangular waveguide so that

$$\int_{x=0}^{x=a}\int_{y=0}^{y=b}\mathbf{E}_{mn}\bullet\mathbf{E}^*_{mn}dxdy=1 \tag{10-80}$$

Upon substituting Eq. (10-75) into Eq. (10-80), and assuming a lossless waveguide with $\alpha_{mn}=0$ and $\gamma=j\beta_{mn}$, we have

$$\begin{aligned}&\int_{x=0}^{x=a}\int_{y=0}^{y=b}\mathbf{E}_{mn}\bullet\mathbf{E}^*_{mn}dxdy\\&=\int_0^a\int_0^b\left[\left(E_x\mathbf{a}_x+E_y\mathbf{a}_y\right)e^{-j\beta_{mn}z}\right]\bullet\left[\left(E_x\mathbf{a}_x+E_y\mathbf{a}_y\right)e^{-j\beta_{mn}z}\right]^*dxdy\\&=\int_0^a\int_0^b\left(\left|E_x\right|^2+\left|E_y\right|^2\right)dxdy\\&=\omega^2\mu^2\frac{H_o^2}{h^2}\frac{ab}{4}\end{aligned} \tag{10-81}$$

where ω is the operating angular frequency, μ is the permeability of the dielectric in the waveguide, h is the eigenvalue given in Eq. (10-57), and a and b are dimensions of the waveguide. We note that Eq. (10-81) is the same as Eq. (10-61) if μ and H_o are replaced with ε and E_o, respectively. By taking the square root of Eq. (10-81), and using it as a normalization factor for Eqs. (10-75a) and (10-75b), the normalized **E**-field of the TE_{mn} mode is generally written as

$$\boxed{\bar{\mathbf{E}}_{mn}=\left(\bar{E}_x\,\mathbf{a}_x+\bar{E}_y\,\mathbf{a}_y\right)e^{-j\beta_{mn}z}} \tag{10-82}$$

where the bar on the top denotes the normalization. The normalized x- and y-components of the electric field of the TE_{mn} mode are

$$\boxed{\bar{E}_x=\frac{2j}{h\sqrt{ab}}\left(\frac{n\pi}{b}\right)\cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)} \qquad (m,n=0,1,2,\ldots) \tag{10-83a}$$

$$\boxed{\bar{E}_y=-\frac{2j}{h\sqrt{ab}}\left(\frac{m\pi}{a}\right)\sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)} \qquad (m,n=0,1,2,\ldots) \tag{10-83b}$$

where $j=\sqrt{-1}$, $h^2=\left(m\pi/a\right)^2+\left(n\pi/b\right)^2$, and a and b are the dimensions of the rectangular waveguide. The **E**-fields expressed by Eq. (10-83) constitute an orthonormal set in the TE modes in the rectangular waveguide: they are normalized to one, and shown to be mutually orthogonal.

Example 10-6

The TE_{10} mode of frequency 3[GHz] propagates in a rectangular waveguide with dimensions $a = 2b = 4[\text{cm}]$, which is filled with a nonmagnetic, lossless dielectric($n = 1.4$). Find (a) cutoff frequency, (b) phase constant, (c) phase velocity, and (d) wave impedance.

Solution

(a) From Eq. (10-46), the cutoff frequency is

$$f_{c(10)} = \frac{1}{2\sqrt{\mu\varepsilon}}\sqrt{\left(\frac{1}{a}\right)^2 + \left(\frac{0}{b}\right)^2} = \frac{3\times 10^8}{2\times 1.4}\left(\frac{1}{0.04}\right) = 2.68[\text{GHz}].$$

(b) From Eq. (10-48), the phase constant is

$$\beta_{mn} = \frac{n\omega}{c}\sqrt{1-\left(\frac{f_{c(mn)}}{f}\right)^2} = \frac{1.4\times 2\pi\times 3\times 10^9}{3\times 10^8}\sqrt{1-\left(\frac{2.68}{3}\right)^2}$$
$$= 12.58\pi[\text{rad/m}].$$

(c) From Eq. (10-52a), the phase velocity of the TE_{10} mode is

$$v_{p(10)} = \frac{\omega}{\beta_{10}} = \frac{2\pi\times 3\times 10^9}{12.58\pi} = 4.77\times 10^8[\text{m/s}].$$

(d) From Eq. (10-77b), the wave impedance is

$$\eta_{TE} = \frac{377}{1.4}\times\frac{1}{\sqrt{1-(2.68/3)^2}} 599.2[\Omega].$$

Example 10-7

In a hollow rectangular waveguide with an aspect ratio $a/b = 2$ ($a < 10[\text{cm}]$), the TE mode of propagation gives $\mathscr{E} = 4(\mathbf{a}_x - \mathbf{a}_y)\sin\left(20\pi z - 1.8\pi\times 10^{10} t\right)[\text{V/m}]$ at a point ($x = a/8$ and $y = b/4$) on the cross section of the waveguide. Find

(a) phase velocity,
(b) cutoff frequency, and
(c) complete expression for $\mathscr{E}$ in the waveguide.

Solution

(a) From the given $\mathscr{E}$, we obtain $\beta_{mn} = 20\pi$, and $\omega = 1.8\pi\times 10^{10}$.

The phase velocity is, from Eq. (10-52a),

$$v_{p(mn)} = \frac{\omega}{\beta_{mn}} = \frac{1.8\pi\times 10^{10}}{20\pi} = 9\times 10^8[\text{m/s}].$$

(b) Rewriting Eq. (10-48) we have

$$f_{c(mn)} = f\sqrt{1-\left(\frac{\beta_{mn}c}{\omega}\right)^2} = 9\times10^9\sqrt{1-\left(\frac{20\pi\times3\times10^8}{1.8\pi\times10^{10}}\right)^2} = 8.485[\text{GHz}]\,.$$

(c) Inserting $x = a/8$ and $y = b/4$ into the general expression for the TE_{mn} mode given in Eqs. (10-75a) and (10-75b), we have

$$E_x = \frac{j\omega\mu}{h^2}\left(\frac{n\pi}{b}\right)H_o\cos\left(\frac{m\pi}{8}\right)\sin\left(\frac{n\pi}{4}\right) \tag{10-84a}$$

$$E_y = -\frac{j\omega\mu}{h^2}\left(\frac{m\pi}{a}\right)H_o\sin\left(\frac{m\pi}{8}\right)\cos\left(\frac{n\pi}{4}\right) \tag{10-84b}$$

If the electric fields in Eq. (10-84) are combined, it should be of the form $4(\mathbf{a}_x - \mathbf{a}_y)$ as given in the problem. Thus, using $a = 2b$, from Eq. (10-84) we obtain

$$n\pi\cos\left(\frac{m\pi}{8}\right)\sin\left(\frac{n\pi}{4}\right) = (m\pi/2)\sin\left(\frac{m\pi}{8}\right)\cos\left(\frac{n\pi}{4}\right)$$

Rewriting it we obtain

$$\left(\frac{n\pi}{4}\right)\tan\left(\frac{n\pi}{4}\right) = \left(\frac{m\pi}{8}\right)\tan\left(\frac{m\pi}{8}\right)$$

This is of the form $X\tan X = Y\tan Y$, and thus the solution must be $X = Y$, or

$$m = 2n \tag{10-85}$$

Next, from Eq. (10-46), the cutoff frequency is

$$f_{c(mn)} = \frac{1}{2\sqrt{\mu_o\varepsilon_o}}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} = 8.485\times10^9 \tag{10-86}$$

Solving Eqs. (10-85) and (10-86) for the mode numbers, using $a = 2b$, we get

$$n/b = 40 \tag{10-87a}$$

$$m/a = 40 \tag{10-87b}$$

Inserting Eq. (10-87) into the general expressions for **E** given in Eq. (10-75), we get

$$\begin{aligned}\mathscr{E} &= A\left[\,\mathbf{a}_x\cos(40\pi x)\sin(40\pi y) - \mathbf{a}_y\sin(40\pi x)\cos(40\pi y)\right]\\ &\quad\times\sin\left(20\pi z - 1.8\pi\times10^{10}t\right)[\text{V/m}]\end{aligned} \tag{10-88}$$

where A is a constant.

Inserting $x = a/8$ and $y = b/4$ into Eq. (10-88), and comparing Eq. (10-88) with the given $\mathscr{E}$ in the problem, we obtain

$a = 1/20$, $b = 1/40$, and $A = 8$

Finally, the real instantaneous electric field of the mode of operation is

$$\mathscr{E} = 8\left[\mathbf{a}_x \cos(40\pi x)\sin(40\pi y) - \mathbf{a}_y \sin(40\pi x)\cos(40\pi y)\right]$$
$$\times \sin\left(20\pi z - 1.8\pi \times 10^{10} t\right)[\mathrm{V/m}].$$

Example 10-8

How many component plane-waves does the TE_{10} mode contain in a rectangular waveguide with dimensions a and b?

Solution

For the TE_{10} mode, let us rewrite Eq. (10-74) as

$$H_z = H_o \cos\left(\frac{\pi}{a}x\right) = H_0 \frac{1}{2}\left(e^{j\pi x/a} + e^{-j\pi x/a}\right) \tag{10-89}$$

As can be seen from Eq. (10-75), the TE_{10} mode has no E_x- and no H_y-component. Rewriting Eq. (10-75c) for the TE_{10} mode, we have

$$H_x = \frac{\gamma}{h^2}\left(\frac{\pi}{a}\right)H_o \sin\left(\frac{\pi}{a}x\right) = \frac{\gamma}{h^2}\left(\frac{\pi}{a}\right)H_o \frac{1}{2j}\left(e^{j\pi x/a} - e^{-j\pi x/a}\right) \tag{10-90}$$

Combining Eq. (10-89) and Eq. (10-90), assuming a lossless waveguide($\gamma = j\beta_{10}$), the magnetic field phasor is

$$\mathbf{H} = \left\{H_x(x,y)\mathbf{a}_x + H_y(x,y)\mathbf{a}_y + H_z(x,y)\mathbf{a}_z\right\}e^{-j\beta_{10}z}$$

From the above equation, we see that the TE_{10} mode comprises two plane waves of the form $e^{j\pi x/a - j\beta_{10}z}$ and $e^{-j\pi x/a - j\beta_{10}z}$.

Exercise 10.8

Show that the phase velocity $v_{p(mn)}$ and the group velocity $v_{g(mn)}$ of a mode of propagation in a rectangular waveguide are related to the phase velocity in an unbounded dielectric v_p by $v_{p(mn)}v_{g(mn)} = v_p^2$.

Exercise 10.9

Find the frequency range over which an air-filled rectangular wave-guide($a = 1.5b = 3[\mathrm{cm}]$) supports only the dominant mode.

Ans. $5[\mathrm{GHz}] < f < 7.5[\mathrm{GHz}]$.

Exercise 10.10

The electric field of a TE mode is zero at the center of the cross section of a rectangular waveguide. Identify the mode.

Ans. TE_{11} mode.

Exercise 10.11

The TE_{mn} mode propagates along a rectangular waveguide in the $-z$-direction. Compare the component fields of the wave with those of the wave propagating in the $+z$-direction?

Ans. The directions of H_x and H_y are reversed.

Review Questions with Hints

RQ 10.12 Is the phase velocity of a mode of propagation always larger than c near cutoff($f \approx f_{c(mn)}$) in a rectangular waveguide? [Eq.(10-52b)]

RQ 10.13 Which of the following quantities, $f_{c(mn)}$, β_{mn}, $v_{p(mn)}$, and wave impedance, is the same for the TM_{mn} and TE_{mn} modes ($m \neq 0$ and $n \neq 0$) in a rectangular waveguide. [Eqs.(10-46)-(10-52)]

RQ 10.14 Explain the evanescent mode in a rectangular waveguide. [Eq.(10-46)]

RQ 10.15 State the boundary condition for H_z of TE waves. [Eq.(10-73)]

RQ 10.16 Explain the dominant mode of a rectangular waveguide. [Eq.(10-76)]

10.2.3 Power Attenuation

The power of the electromagnetic wave propagating in a waveguide may be attenuated owing to the lossy dielectric and/or imperfect conducting walls. If the attenuation is small as in most practical waveguides, the power loss may be most conveniently described in terms of the attenuation constant, by assuming no change in the field pattern in the transverse plane. Let us first consider the time-average power density of the wave propagating in a waveguide, or the time-average Poynting vector,

$$\langle \mathbf{S} \rangle = \frac{1}{2}\text{Re}\left[\mathbf{E} \times \mathbf{H}^*\right] \tag{10-91}$$

where **E** and **H** are the electric and magnetic field phasors of the wave. The time-average total power through the cross section of the waveguide is therefore

$$\langle P \rangle = \int_s \langle \mathbf{S} \rangle \cdot d\mathbf{s} = \int_s \frac{|E_x(x,y)|^2 + |E_y(x,y)|^2}{2\eta_w} e^{-2\alpha z} dxdy = P_o e^{-2\alpha z} \tag{10-92}$$

where the wave impedance $\eta_w = \eta_{TM}$ for TM modes and $\eta_w = \eta_{TE}$ for TE modes. The attenuation constant α can be separated into two parts,

$$\alpha = \alpha_d + \alpha_c \tag{10-93}$$

The first part, α_d, is due to the lossy dielectric $(\sigma \neq 0)$, whereas the second part, α_c, is due to the imperfect conducting walls $(\sigma_c \neq 0)$.

Even if the rectangular waveguide is filled with a low-loss dielectric, all the discussions and expressions developed previously for the lossless waveguide are valid for the low-loss waveguide, except that the permittivity ε is now given by a complex number:

$$\hat{\varepsilon} \equiv \varepsilon' - j\varepsilon'' = \varepsilon - j\frac{\sigma}{\omega} \tag{10-94}$$

where ε and σ are the permittivity and conductivity of the dielectric, respectively. The caret on the top of ε stands for complex number. By assuming a low-loss in the dielectric, the propagation constant for the waveguide mode is written as

$$\gamma \equiv \alpha_{mn} + j\beta_{mn} = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \omega^2\mu\left(\varepsilon - j\frac{\sigma}{\omega}\right)} \tag{10-95}$$

By separating Eq. (10-95) into the real and imaginary parts, we have

$$\alpha_{mn}^2 - \beta_{mn}^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \omega\mu\varepsilon \tag{10-96a}$$

$$2\alpha_{mn}\beta_{mn} = \omega\mu\sigma \tag{10-96b}$$

Under the condition of a small attenuation, or $\alpha_{mn} << \beta_{mn}$, we can assume $\alpha_{mn}^2 - \beta_{mn}^2 \approx -\beta_{mn}^2$ on the left-hand side of Eq. (10-96a), and write the phase constant as

$$\beta_{mn} = \sqrt{\omega\mu\varepsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} = \omega\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{f_{c(mn)}}{f}\right)^2} \tag{10-97}$$

We note that Eq. (10-97) is the same as that in the lossless waveguide. Inserting Eq. (10-97) into Eq. (10-96), the attenuation constant for the mode of operation, due to the lossy dielectric, is

$$\boxed{\alpha_{mn} = \alpha_d = \frac{\sigma\sqrt{\mu/\varepsilon}}{2\sqrt{1 - \left(f_{c(mn)}/f\right)^2}}} \qquad [\mathrm{m}^{-1}] \tag{10-98}$$

where subscript d stands for dielectric.

Whereas it is tedious to determine analytically the attenuation constant α_c that is caused by the finite conductivity of the conducting walls, we only outline the general procedure for obtaining α_c here. When the finite conductivity causes ohmic power losses in the waveguide, the total power dissipated in the waveguide is of course the sum of the powers dissipated in the four conducting walls, namely,

$$\langle P_c \rangle = \langle P(y=0) \rangle + \langle P(y=b) \rangle + \langle P(x=0) \rangle + \langle P(x=a) \rangle \qquad [W] \qquad (10\text{-}99)$$

where < > denotes time average. Borrowing from section 8-2.4.3, the time-average power dissipated in a conducting plate of an area $w \times \ell$ [m^2], with a skin depth δ[m], is

$$\langle P_c \rangle = \frac{1}{4\sigma} w\ell\, \delta J_o^2 \qquad [W] \qquad (8\text{-}97)$$

In the above equation, J_o is the amplitude of the current density on the surface, not a surface current density. It is equal to the electric field intensity, tangential to the surface, multiplied by the conductivity; that is, $J_o = \sigma E_o$. In a rectangular waveguide with dimensions a and b, the dissipated power per unit length along the z-direction is therefore

$$\langle \bar{P}_c \rangle = \frac{\sigma\delta}{4} e^{-2\alpha z} \int_{x=0}^{x=a} \left[|E_x(y=0)|^2 + |E_z(y=0)|^2 + |E_x(y=b)|^2 + |E_z(y=b)|^2 \right] dx$$
$$+ \frac{\sigma\delta}{4} e^{-2\alpha z} \int_{y=0}^{y=b} \left[|E_y(x=0)|^2 + |E_z(x=0)|^2 + |E_y(x=a)|^2 + |E_z(x=a)|^2 \right] dy$$

(10-100)

which is measured in units of watts per meter [W/m]. The bar on the top is to denote the power dissipation per unit length along the waveguide. Upon assuming a perfect dielectric in the waveguide, the ohmic power loss on the conducting walls expressed by Eq. (10-100) is responsible for the attenuation of the wave propagating in the waveguide. In this case, $\alpha = \alpha_c$ in Eq. (10-93). In view of the principle of conservation of energy, we write the time-average power loss per unit length along the waveguide as

$$\langle \bar{P}_c \rangle = -\frac{d}{dz}\left(P_o e^{-2\alpha z} \right) = 2\alpha P_o e^{-2\alpha z} \qquad [W/m] \qquad (10\text{-}101)$$

where P_o is the time-average power at $z = 0$, and the term in parenthesis is the time-average power propagated along the waveguide by a distance z. Equating Eq. (10-100) with Eq. (10-101), we can obtain the attenuation constant α for the wave propagating in the low-loss waveguide.

Example 10-9

An air-filled rectangular waveguide is 1[m] long, with dimensions $3a = 4b = 12$[cm]. It operates at 4[GHz], supporting a wave with $\alpha = 2 \times 10^{-3}$[Np/m]. When the output power is 1.5[kW], find

(a) total time-average power dissipated in the waveguide,
(b) maximum electric field intensity in the waveguide, and
(c) location of the maximum electric field intensity.

Solution

(a) The input power is

$$P_{in} = P_{out} e^{\alpha \ell} = 1.5 \times 10^3 e^{0.002 \times 2 \times 1} = 1,506[\text{W}] \tag{10-102}$$

Thus, the power dissipated in the waveguide is 6[W].

(b) Cutoff frequency of the dominant mode is

$$f_{c(10)} = \frac{c}{2a} = \frac{3 \times 10^8}{2 \times 0.04} = 3.75[\text{GHz}] \tag{10-103}$$

The next higher mode is TE_{01}, having a cutoff frequency 4.5[GHz]. Because of the operating frequency 4[GHz], the waveguide supports only TE_{10} mode.

From Eqs. (10-75) and (10-77b), the transverse field components of the TE_{10} mode are written as

$$E_y = -\frac{j\omega\mu}{h^2}\left(\frac{\pi}{a}\right) H_o \sin\left(\frac{\pi}{a}x\right) \equiv E_o' \sin\left(\frac{\pi}{a}x\right) \tag{10-104a}$$

$$H_x = \frac{j\beta_{10}}{h^2}\left(\frac{\pi}{a}\right) H_o \sin\left(\frac{\pi}{a}x\right) = -\frac{E_o'}{\eta_{TE}} \sin\left(\frac{\pi}{a}x\right) \tag{10-104b}$$

At the input end of the waveguide($z = 0$), with the aid of Eq. (10-77b), the total time-average power is

$$P_{in} = \frac{1}{2}\int_{x=0}^{x=a}\int_{y=0}^{y=b} \mathbf{E} \times \mathbf{H}^* \bullet \mathbf{a}_z dxdy = \frac{ab\left(E_o'\right)^2}{4} \frac{\sqrt{1-\left(f_{c(10)}/f\right)^2}}{\eta_o} \tag{10-105}$$

Inserting Eqs. (10-102) and (10-103) into Eq. (10-105) we have

$$1,506 = \frac{0.04 \times 0.03 (E_o')^2}{4} \frac{\sqrt{1-(3.75/4)^2}}{377}$$

Thus,

$$E_o' = 73.7[\text{kV/m}].$$

(c) From Eq. (10-104), we see that the maximum electric field intensity is observed at $x = a/2 = 0.02[\text{m}]$.

Example 10-10

Show that the TM_{mn} and TE_{mn} modes are independent of each other in a rectangular waveguide, even though they have the same phase constant and the same group velocity.

Solution

The electric and magnetic field phasors of the TM_{mn} mode, $\mathbf{E}_{\mathrm{TM}}$ and $\mathbf{H}_{\mathrm{TM}}$, are expressed by Eqs. (10-41) and (10-56), whereas the phasors $\mathbf{E}_{\mathrm{TE}}$ and $\mathbf{H}_{\mathrm{TE}}$ of the TE_{mn} mode are expressed by Eqs. (10-74) and (10-75).

The total electric and magnetic fields in the waveguide are therefore

$$\mathbf{E} = \mathbf{E}_{\mathrm{TM}} + \mathbf{E}_{\mathrm{TE}} \tag{10-106a}$$

$$\mathbf{H} = \mathbf{H}_{\mathrm{TM}} + \mathbf{H}_{\mathrm{TE}} \tag{10-106b}$$

The time-average power density in the waveguide is

$$\begin{aligned}\langle P\rangle &= \int_S \langle \mathbf{S}\rangle \bullet d\mathbf{s} = \frac{1}{2}\mathrm{Re}\int_S (\mathbf{E}\times\mathbf{H}^*)\bullet \mathbf{a}_z dxdy \\ &= \frac{1}{2}\mathrm{Re}\left[\int_S (\mathbf{E}_{\mathrm{TM}}\times\mathbf{H}^*_{\mathrm{TM}} + \mathbf{E}_{\mathrm{TE}}\times\mathbf{H}^*_{\mathrm{TE}})\bullet \mathbf{a}_z dxdy\right] \\ &\quad + \frac{1}{2}\mathrm{Re}\left[\int_S (\mathbf{E}_{\mathrm{TM}}\times\mathbf{H}^*_{\mathrm{TE}} + \mathbf{E}_{\mathrm{TE}}\times\mathbf{H}^*_{\mathrm{TM}})\bullet \mathbf{a}_z dxdy\right]\end{aligned} \tag{10-107}$$

The integrand of the second integral on the right-hand side of Eq. (10-107) is written as

$$\begin{aligned}\mathbf{E}_{\mathrm{TM}}\times\mathbf{H}^*_{\mathrm{TE}} + \mathbf{E}_{\mathrm{TE}}\times\mathbf{H}^*_{\mathrm{TM}} &= \left(E_{x,\mathrm{TM}}H^*_{y,\mathrm{TE}} + E_{y,\mathrm{TM}}H^*_{x,\mathrm{TE}}\right) \\ &\quad + \left(E_{x,\mathrm{TE}}H^*_{y,\mathrm{TM}} + E_{y,\mathrm{TE}}H^*_{x,\mathrm{TM}}\right)\end{aligned} \tag{10-108}$$

Substitution of Eqs. (10-56) and (10-75) into Eq. (10-108), in conjunction with $\gamma = j\beta_{mn}$, shows that each parenthesis on the right-hand side of Eq. (10-108) vanishes during the integration. Thus, Eq. (10-107) becomes

$$\langle P\rangle = \frac{1}{2}\mathrm{Re}\left[\int_S (\mathbf{E}_{\mathrm{TM}}\times\mathbf{H}^*_{\mathrm{TM}})\bullet \mathbf{a}_z dxdy\right] + \frac{1}{2}\mathrm{Re}\left[\int_S (\mathbf{E}_{\mathrm{TE}}\times\mathbf{H}^*_{\mathrm{TE}})\bullet \mathbf{a}_z dxdy\right] \tag{10-109}$$

From (10-109), we see that there is no coupling between the two waves: the first term on the right-hand side of Eq.(10-109) is the total power of the TM_{mn} mode, while the second term is that of the TE_{mn} mode. Thus, the TM_{mn} and TE_{mn} waves are independent of each other in the waveguide.

Exercise 10.12

Show that an attenuation constant of $0.05[\mathrm{dB/m}]$ corresponds to $\alpha = 5.75\times10^{-3}[\mathrm{Np/m}]$.

Ans. $20\log_{10}\left(e^{-0.00575}\right) = -0.05$.

Review Questions with Hints

RQ 10.17 TE_{21} and TE_{02} modes have the same phase constant and phase velocity in a rectangular waveguide with $a = 2b$. Can the energy of one mode be transferred to the other? [Eq.(10-83)]

RQ 10.18 What is the relation between the power loss on the conducting walls of a rectangular waveguide and the attenuation constant? [Eq.(10-101)]

Problems

10-1 A parallel-plate waveguide is designed to support the TEM mode only over the frequency range $0 < f < 10$[GHz]. The space between two plates is filled with a dielectric of $n = 1.5$. Find the maximum separation of the plates.

10-2 Two conducting plates of a parallel-plate waveguide is separated by a dielectric of thickness $d = 4$[cm] and refractive index $n = 1.25$. Electromagnetic waves of a frequency $f = 7$[GHz] propagate in the waveguide, which have no electric field component in the *x*-direction. Which modes are propagating?

10-3 With reference to the electromagnetic waves in **Problem 10-2**, find the maximum difference in the arrival times of the modes of propagation over a distance of 20[cm] along the waveguide.

10-4 A TE wave is launched into a parallel-plate waveguide filled with a dielectric of permittivity ε and permeability μ. The electric field phasor of the TE_m mode is expressed by Eqs. (10-4b) and (10-5). Find

(a) magnetic field phasor **H**, and

(b) real instantaneous values, $\mathscr{E}$ and $\mathscr{H}$.

10-5 A parallel-plate waveguide is partially filled with a lossless dielectric ($\varepsilon = 2.25\varepsilon_o$ and $\mu = \mu_o$) as shown in Fig. 10.11. The TE_1 mode of $f = 12$[GHz] propagates in the +*z*-direction in the region $z < 0$, and impinges on the air-dielectric interface at $z = 0$. The electric phasor of the incident wave is given by $\mathbf{E} = -j2\mathbf{a}_x \sin(50\pi y) e^{-j\beta_1 z}$[V/m] in the region $z < 0$.

(a) Find $f_{c(1)}$, β_1, and θ_1 in the region $z < 0$.

(b) Which modes will propagate in the region $z > 0$?

(c) Find the electric field in the region $z > 0$.

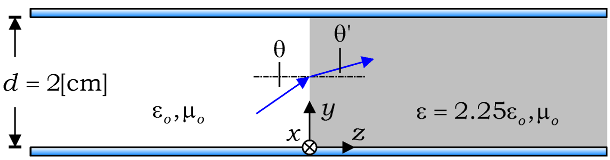


Fig. 10.11 A partially filled parallel-plate waveguide.

10-6 In Fig. 10.11, it is assumed that the left region($z < 0$) is filled with a dielectric($\varepsilon = 2.25\varepsilon_o$), and the right region($z > 0$) with another dielectric($\varepsilon = 1.69\varepsilon_o$). The TE_1 mode propagates in the +z-direction in the left region, and impinges on the dielectric-dielectric interface at $z = 0$.
(a) Find the cutoff frequencies for the TE_1 mode in both regions.
(b) Find the frequency range over which the TE_1 mode can propagates in the left region but is cutoff in the right region.
(c) Show that the cutoff in part (b) is due to the total internal reflection of the component plane-waves of the mode at the interface at $z = 0$.

10-7 A hollow rectangular waveguide has dimensions $a = 2[\text{cm}]$ and $b = 1.5[\text{cm}]$. Find the frequency range over which the waveguide will support (a) a single mode, and (b) two modes.

10-8 The operating frequency is 10% lower than the cutoff frequency for the TE_{mn} in a hollow rectangular waveguide. Express the attenuation constant α_{mn} in terms of $f_{\text{c}(mn)}$.

10-9 In a hollow rectangular waveguide of dimensions $a = 5[\text{cm}]$ and $b = 4[\text{cm}]$, the longitudinal component of the electric field is given as $\mathscr{E}_z = 4\sin(40\pi x)\sin(25\pi y)\sin(16\pi\times 10^9 t - \beta z)$.
(a) Identify the mode of operation.
(b) Find cutoff frequency, phase constant, and wave impedance.

10-10 In an air-filled rectangular waveguide with dimensions $a = 2b = 4[\text{cm}]$, the transverse component of the electric field is given as $\mathscr{E}_t = 1.5\,\mathbf{a}_x \cos(25\pi x)\sin(50\pi y)\cos(20\pi\times 10^9 t)$.
(a) Identify the modes of operation.
(b) Find the cutoff frequency and phase constant of the modes.
(c) Find the instantaneous electric fields of the modes.

10-11 The amplitude of the electric field of the dominant mode is E_o in an air filled $a[\text{m}]\times b[\text{m}]$ rectangular waveguide. Express the total time-average power of the mode as a function of the operating frequency f.

10-12 A rectangular waveguide of dimensions $a = 1.5b = 7.5[\text{cm}]$ is filled with a low-loss dielectric of $\varepsilon_r = 2$ and $\sigma = 3\times 10^{-5}[\text{S/m}]$. If the waveguide operates at frequency 2[GHz], find the attenuation constant of the mode of operation.

10-13 Show that the magnetic field lines of the TM_{11} mode can be obtained from $dy/dx = -\tan(\pi y)/\tan(\pi x)$ in the transverse plane of the square waveguide($a = b = 1$).

10-14 The x-component of the electric field is given by $\mathscr{E}_x = 1.6\cos(50\pi)\sin(40\pi)\cos(18\pi\times 10^9 t)$ in the transverse plane of a

rectangular waveguide filled with a dielectric($n = 1.2$). For the mode of operation in the waveguide, determine
(a) cutoff frequency.
(b) phase velocity.
(c) group velocity.

10-15 Using the expression for the TE_{mn} mode given in Eq. (10-75), (a) decompose the TE_{mn} mode into four plane waves, and (b) show that the power of the mode is equally distributed among the four component plane-waves.

Appendix: Material Constants

A-1 Physical Constants

Constant	Value
Electron Charge	$e = -1.602 \times 10^{-19}$[C]
Electron Mass	$m_e = 9.109 \times 10^{-31}$[Kg]
Radius of Electron	$R_e = 2.81 \times 10^{-15}$[m]

A-2 Constants of Free space

Constant	Value
Permittivity	$\varepsilon_o = 8.854 \times 10^{-12}$[F/m]
	or $\sim \frac{1}{36\pi} \times 10^{-9}$[F/m]
Permeability	$\mu_o = 4\pi \times 10^{-7}$[H/m]
Speed of light	$c = 3.0 \times 10^{8}$[m/s]
Intrinsic Impedance	377 or $\sim 120\pi[\Omega]$

A-3 Relative Permittivities(Dielectric Constants), ε_r

Constant	Value
Vacuum	1
Air(sea level)	1.0006
Amber	2.7
Quartz(fused)	3.8
Glass	4-10
Mica	5.4
Barium titanate	1200
Germanium	16
Silicon	11.7
Paper	3
Rubber	2.5-4
Wood(dry)	1.5-4
Distilled Water	80
Ice	4.2

A-4 Conductivity, σ[S/m]

Constant	Value
Silver	6.17×10^{7}
Copper	5.80×10^{7}
Gold	4.10×10^{7}
Aluminum	3.54×10^{7}
Iron	10^{7}
Glass	10^{-12}
Rubber	10^{-15}
Fuse quartz	10^{-17}
Distilled water	10^{-4}
Sea water	4

A-5 Relative Permeabilities, μ_r

Constant	Value
Diamagnetic	
Silver	$0.99998(\simeq 1)$
Copper	$0.99999(\simeq 1)$
Gold	$0.99996(\simeq 1)$
Water	$0.99999(\simeq 1)$
Paramagnetic	
Air	$1.000004(\simeq 1)$
Aluminum	$1.00002(\simeq 1)$
Titanium	$1.0002(\simeq 1)$
Tungsten	$1.00008(\simeq 1)$
Ferromagnetic	
Cobalt	250
Nickel	600
Iron(pure)	5000
Mumetal	100,000

Subject Index