Makes electromagnetics more interesting and understandable through the use of MATLAB examples, graphics, and exercises.

Building upon the reader's basic understanding of electromagnetics obtained in an introductory course in physics, this book stresses its importance in a modern electrical and computer engineering curriculum. The concepts of static fields, time varying fields, wave propagation, transmission lines, and radiation are described. Most importantly, the text emphasizes the use of MATLAB ${ }^{\text {e }}$, now widely available and frequently used by most students in previous courses, for the calculation and visualization of electromagnetic field characteristics.

## KEY FEATURES

- Electromagnetic theory is introduced using the historical approach in that static fields precede time varying fields.
- Wave propagation and transmission lines are emphasized in this text.
- Provides students with a broader introduction to numerical methods - Method of Moments, FTD, and Finite Element Method - for application in later advanced courses.
- MATLAB supplements analytical mathematical tools in solving boundary value problems for static fields. Traditional vector calculus also incorporates MATLAB techniques.
- Several MATLAB programs are included in this text that can be easily modified by the reader to solve more difficult problems. In addition, the MATLAB programs that are used to generate the figures and solve the examples are available on the Web.
- Answers to problems are available to students and complete solutions to the problems including the MATLAB programs are available to instructors.


## ABOUT THE AUTHORS

Karl E. Lonngren is Professor within the Dept. of Electrical and Computer Engineering at the University of lowa. He received his BS, MS, and Ph.D. at the University of Wisconsin. His research interests are in the area of nonlinear plasma physics. Dr. Lonngren has authored four books and more than two hundred articles. He is a Fellow of the IEEE and of the American Physical Society.
Sava V. Savov is Associate Professor in the Department of Electronic Engineering at the Technical University of Varna, Bulgaria. His research interests are in the area of computational electromagnetics, numerical modeling of antennas, and propagation in wireless communications. He is a Senior Member of IEEE.

www.scitechpub.com

Lonngren Savov

## Fundamentals of Eilchioniciance with MATLAB

Karl E. Lonngren Sava V. Savov



## Contents

Preface ..... ix
Chapter $1 \quad$ MATLAB and Vectors ..... 2
1.1 MATLAB and a Review of Vectors ..... 3
1.2 Coordinate Systems ..... 16
1.3 Integral Relations for Vectors ..... 29
1.4 Differential Relations for Vectors ..... 39
1.5 Phasors ..... 56
1.6 Conclusion ..... 60
1.7 Problems ..... 60
Chapter $2 \quad$ Static Electric and Magnetic Fields ..... 70
2.1 Coulomb's Law ..... 71
2.2 Electric Field ..... 76
2.3 Superposition Principles ..... 78
2.4 Gauss's Law ..... 87
2.5 Potential Energy and Electric Potential ..... 95
2.6 Numerical Integration ..... 111
2.7 Dielectric Materials ..... 121
2.8 Capacitance ..... 126
2.9 Electrical Currents ..... 130
2.10 Fundamentals of Magnetic Fields ..... 135
2.11 Magnetic Vector Potential and the Biot-Savart Law ..... 146
2.12 Magnetic Forces ..... 155
2.13 Magnetic Materials ..... 166
2.14 Magnetic Circuits ..... 172
2.15 Inductance ..... 175
2.16 Boundary Conditions ..... 181
2.17 Conclusion ..... 189
2.18 Problems ..... 190
vi Contents
Chapter 3 Boundary Value Problems Using MATLAB ..... 204
3.1 Poisson's and Laplace's Equations ..... 205
3.2 Analytical Solution in One Dimension-Direct Integration Method ..... 211
3.3 Numerical Solution of a One-Dimensional Equation- Finite Difference Method ..... 222
3.4 Analytical Solution of a Two-Dimensional Equation-Fourier Series Expansion 233
3.5 Finite Difference Method Using MATLAB ..... 243
3.6 Finite Element Method Using MATLAB ..... 249
3.7 Method of Moments Using MATLAB ..... 263
3.8 Conclusion ..... 273
3.9 Problems ..... 274
Chapter 4 Time-Varying Electromagnetic Fields ..... 282
4.1 Faraday's Law of Induction ..... 283
4.2 Equation of Continuity ..... 296
4.3 Displacement Current ..... 301
4.4 Maxwell's Equations ..... 307
4.5 Poynting's Theorem ..... 312
4.6 Time-Harmonic Electromagnetic Fields ..... 318
4.7 Conclusion ..... 322
4.8 Problems ..... 322
Chapter 5 Electromagnetic Wave Propagation ..... 330
5.1 Wave Equation ..... 331
5.2 One-Dimensional Wave Equation ..... 336
5.3 Time-Harmonic Plane Waves ..... 352
5.4 Plane Wave Propagation in a Dielectric Medium ..... 361
5.5 Reflection and Transmission of an Electromagnetic Wave ..... 373
5.6 Waveguide-Propagation with Dispersion ..... 388
5.7 Conclusion ..... 398
5.8 Problems ..... 398
Chapter 6 Transmission Lines ..... 408
6.1 Equivalent Electrical Circuits ..... 409
6.2 Transmission Line Equations ..... 412
6.3 Sinusoidal Waves ..... 418
6.4 Terminations ..... 423
6.5 Impedance on the Transmission Line and Matching ..... 431
6.6 Smith Chart ..... 438
6.7 Transient Effects and the Bounce Diagram ..... 449
6.8 Pulse Propagation ..... 457
6.9 Lossy Transmission Lines ..... 462
6.10 Dispersion and Group Velocity ..... 468
6.11 Conclusion ..... 476
6.12 Problems ..... 477
Chapter $7 \quad$ Radiation of Electromagnetic Waves ..... 484
7.1 Radiation Fundamentals ..... 485
7.2 Short Electric Dipole Antenna ..... 493
7.3 Long Dipole Antenna ..... 501
7.4 Antenna Parameters ..... 506
7.5 Magnetic Dipole Antenna ..... 517
7.6 Aperture Antennas, Diffraction of Waves ..... 522
7.7 Antenna Arrays ..... 535
7.8 Conclusion ..... 546
7.9 Problems ..... 547
Appendix A Mathematical Formulas ..... 552
A. 1 Vector Identities ..... 553
A. 2 Vector Operations in the Three Coordinate Systems ..... 554
A. 3 Summary of the Transformations Between Coordinate Systems ..... 555
A. 4 Integral Relations ..... 557
Appendix B Mathematical Foundation of the Finite Element Method ..... 558
B. 1 Minimum Energy Condition ..... 559
B. 2 Linear Interpolation Coefficients ..... 560
B. 3 S-matrix Elements ..... 560
B. 4 Decoupled and Coupled Node Potentials ..... 561
B. 5 The Matrix Equation for the Unknown Potentials ..... 563
Appendix C Material Parameters ..... 564
Appendix D Transmission Line Parameters of Two Parallel Wires ..... 568
Appendix E Plasma Evolution Adjacent to a Metallic Surface ..... 574
Appendix F References ..... 578
Appendix G Answers ..... 582
Index ..... 635

## Preface

Electromagnetic Field Theory is one of the fundamental courses that an electrical and computer engineering student is required to take in order to gain a physical understanding of the foundations and the heritage of the field that will occupy his or her professional life for the several decades following graduation. The acquiring of an appreciation for the laws of nature that govern and limit the speed of the smallest computer chip continue to be crucial as this speed approaches the ultimate limit. With the many changes that are occurring in undergraduate curriculums due to the rapid development of new technologies and hence additional courses, it is common to find that only one course in electromagnetic theory is now required for students. However, most of the students are "computer savvy" and have been introduced to and have used MATLAB in their previous courses and are motivated by its ability to create pictures on a computer screen that can help illustrate complicated physical phenomena.

## Our Approach

The underlying philosophy of this one semester undergraduate text is to combine the student's computer/MATLAB ability that has been gained in earlier courses with an introduction to electromagnetic theory in a coherent fashion in order to stimulate the physical understanding of this difficult topic. Where two terms of Electromagnetic Theory were once required, the challenge of squeezing study into one term can at least be partially met with the use of MATLAB to diminish the drudgery of numerical computations while enhancing understanding of concepts. Therefore, in this text numerous examples are solved using MATLAB along with the creation of several figures throughout the text, and all of the ".m" files are made available for the reader to examine and to modify. We therefore believe that it is possible to take this seemingly abstract material and make it understandable and interesting to the student. This belief has been confirmed by using the material in classes for six years and continually using student feedback to improve it.

## Preface

## Organization of the Text

We review essential features of MATLAB immediately in Chapter 1 in order to satisfy the novice's initial trepidations and incorporate its MATLAB's capabilities throughout the entire text. After an initial review in Chapter 1 of MATLAB, vector calculus, and phasors, we follow in the footsteps of the giants who have preceded us in and summarize the fundamentals of static electromagnetic fields, including several examples that the reader may have encountered previously. We discuss analytical and MATLAB techniques in order to illustrate the spatial behavior of a static field in a finite boundary in Chapter 3. The majority of the text is directed toward the presentation of time varying electromagnetic fields and Maxwell's equations in Chapter 4. From these equations we derive a wave equation that can be most easily understood using a diverse selection of examples from other disciplines. A study of plane electromagnetic waves directly follows this review of waves in Chapter 5. In Chapter 6 the subject of transmission lines is emphasized, owing to its importance in modern technology. This includes MATLAB programs for the creation of a Smith chart and its application. Finally, in Chapter 7 the subject of radiation of electromagnetic waves is explained, first from a very simple physical interpretation, and then summarizing many of the important parameters associated with antennas.

Anticipating the student's further study of modern topics in electrical engineering, we have tried to present a somewhat broader look in numerical methods than most introductory electromagnetics texts. The Finite Element Method, Method of Moments, and Finite Time Difference are all examples of this effort. With MATLAB, we believe most students can handle this material well and will be better prepared for their application later.

## Aids to Learning

The Appendices and page layout are designed to enhance the reader's understanding and appreciation of electromagnetic theory as it applies to their study.

- Examples have been clearly set of from the text with rule lines.
- Each time MATLAB is employed, whether in examples or chapter problems, the MATLAB icon is used to signify its use.
- The answers to all problems have been included in Appendix G so that students can check all of their work, not just some. Instructors are provided with complete worked-out solutions in hard copy and MATLAB files, to use at their discretion.
- The most important equations occurring in the text have been boxed so that they might be quickly identified as such and committed to memory.
- Important mathematical formulae have been consolidated and placed together in Appendix A.
- Several interesting extensions of text material are offered in Appendix B (Mathematical Foundation of the Finite Element Method) and Appendix E (Plasma Evolution Adjacent to a Metallic Surface).
- Material parameters are listed in Appendix C.
- A fairly extensive and up-to-date list of references for both electromagnetics and MATLAB is provided in Appendix F.


## Aids to Teaching

For instructors, we are working closely with SciTech Publishing to supply our teaching colleagues with ample resources and to add to them continuously, even to invite contributions from them and their students. An initial CDROM offers complete solutions to problems in hard copy (PDF and Word) as well as MATLAB.m files. The code for all MATLAB-generated figures is made available on the CD, as well as on the web for students. All other figures used in the text are provided as EPS graphic files as well as in a PowerPoint file. A modest number of MATLAB animation files have been collected, with the hope that the authors and our adopting instructors will add to them. Check in frequently at our website to see what has been added: www.scitechpub.com/lonngren.htm

## Acknowledgments

The construction of an effective textbook, with its attendant resource materials, is a team effort akin to an engineering marvel. We are fortunate to work with a publisher that believes in our effort and maintains an open and constant dialog. In particular, our editor and SciTech Publishing founder Dudley Kay has extended his considerable years of experience in commercial publishing and at IEEE Press to dispense advice and encouragement, making the finished book even better than we had first envisaged. Robert Kern, Melissa Parker, and the team at TIPS Technical Publishing, Inc. provided an outstanding page design and worked tirelessly on the hundreds of art and equation files to bring text and graphics together into a coherent, attractive whole. Our early figures were brought to professional polish by Michael Georgiev, working under the demanding guidance of Prof. Savov. The striking time-delay

## Preface

photograph of a rocket probe into the aurora borealis was adapted by Brent Beckley into the brilliant cover design.

Of course, despite the most rigorous of efforts and capabilities of our fine team, any errors that occur are ours. We strongly encourage your feedback on any aspect of the book that could be improved and pledge to correct any errors reported to us or the publisher.

The authors have profited from extended discussions with several people who have influenced their thinking concerning the presentation of this material. This includes their former teachers, their past and present electromagnetics colleagues, and the many students who have asked stimulating questions in and outside of class over the last three decades. In particular, Professors Er Wei Bai, Adrian Korpel, and Jon Kuhl provided valuable assistance at crucial times.

Finally, the authors thank their wives Vicki and Rossi for their encouragement and understanding during this endeavor, and this book is dedicated to them.

Karl E. Lonngren
University of Iowa
Lonngren@eng.uiowa.edu
Sava V. Savov
Technical University at Varna
Varna, Bulgaria
svsavov@ms.ieee.bg

## MATLAB and a review of vectors

## 1. MATLAB and vectors

In this chapter, we introduce and summarize several properties of the software program entitled MATLAB. The topics in this summary have been selected based on their later application in our study of electromagnetics. You have probably encountered MATLAB in other courses since the software is widely used in the educational community. In addition, MATLAB is a tool that will permit you to easily obtain pictures of various electromagnetic phenomena that we will encounter in our journey through this book. In addition, vectors which are crucial in describing electromagnetic phenomena can be easily manipulated using MATLAB. Several of the figures in this text have been created using MATLAB. Because of simplicity, we will emphasize Cartesian coordinates in this review. The vector operations in other coordinate systems are included in Appendix 1. Our motivation in employing vectors is that electromagnetic fields are vector quantities and their use will permit us to use a fairly compact notation to represent sets of partial differential equations. This review will include a derivation of the vector differential operations of the gradient, the divergence and the curl. The transformation of a vector from one coordinate system to another will be discussed. Additional symmetry found in a particular problem in one coordinate system over another one may suggest such a transformation. The reader who feels comfortable with vector terminology can easily skip this portion of the chapter and pass on with no loss of continuity. Just remember that in this text, bold-face type will be used to define a vector $\mathbf{A}$ and the symbol $\mathbf{u}_{\mathbf{A}}$ will be

## MATLAB and a review of vectors

used to indicate the unit vector corresponding to this vector. This chapter concludes with a few brief comments on phasors.

## MATLAB and a review of vectors

### 1.1. MATLAB and a review of vectors

MATLAB is a software program that is widely available for digital computers at a large number of universities and on a large variety of machines. As will be noted in this text, we will make extensive use of it. The two and three dimensional plotting capabilities will be exploited throughout this text since a picture or a graph can usually aid in the physical interpretation of an equation. Herein, we will briefly present an introduction of several germane features of this program that will be useful for electromagnetic theory. Various functions such as trig functions appear in a MATLAB library that can be easily called and used. The user can customize and add to this list by writing a program in a ".m" (dot m) file. Several MATLAB figures will be included throughout this text. In addition, the files that have been used to create the figures in the text are available at the following web site: [http://www.scitechpub.com/]. These programs will be characterized with the names: "example_103" and "figure_103" to indicate the third example and the third figure in chapter 1. The example and figure captions are identified in the book with the superscript notation MATLAB. Matrix operations will not be examined since their application will receive minimal attention in this text. We assume that the reader is able to call MATLAB and have the familiar MATLAB prompt ">>" appear on the screen. Typing the words, "help topic" after the prompt brings on-screen help to the user.

For example, we type without the following command after the prompt, press the enter key, and note the following statements that appear on the screen

# MATLAB and a review of vectors 

$$
\begin{aligned}
& \gg x=3 \\
& x=
\end{aligned}
$$

$$
\begin{equation*}
3 \tag{1.1}
\end{equation*}
$$

$$
\gg
$$

The computer has assigned a value for the variable x that it will remember until it is changed or until we exit the program. It is now ready for the next input. Let us choose a value $\mathrm{y}=4$ but desire the computer to not print back this number immediately. This is accomplished with a semicolon ";".

$$
\begin{align*}
& \gg y=4 ; \\
& \gg \tag{1.2}
\end{align*}
$$

This may not seem important to the stage. However, a simple statement in a program could lead to a large waste of "computer screen" or computer paper as the numbers are spewed forth.

Mathematical operations with these two numbers follow and we write a mathematical operation at the prompt. In the table given below, the following three lines will appear after we push the return key.

| Addition | Subtraction | Multiplication | Division |
| :---: | :---: | :---: | :---: |
| $\gg z=x+y$ | $\gg z=x-y$ | > $>\mathrm{z}=\mathrm{x}^{*} \mathrm{y}$ | $\gg z=x / y$ |
| $\mathrm{z}=$ | $\mathrm{z}=$ | Z = | $\mathrm{z}=$ |
| 7 | - 1 | 12 | 0.7500 |
| \gg | \gg | \gg | \gg |

Note the four place accuracy in the last column. The accuracy can be controlled by the user.

## MATLAB and a review of vectors

With the semicolon notation, it is possible to write any of the commands in one line. For example, the addition program can also be written in one line as

$$
\begin{equation*}
\gg x=3 ; y=4 ; z=x+y ; \tag{1.3}
\end{equation*}
$$

In order to obtained the solution using this operation, you just have to type "z" at the MATLAB prompt and computer will respond

$$
\begin{gather*}
\gg z \\
7  \tag{1.4}\\
7 \\
\ggg
\end{gather*}
$$

The semicolon will be very useful in a lengthy calculation if we do not wish to display intermediate results. Another useful tool to remember is the symbol "\%" since anything typed on the line after it will receive no attention by the computer. It is a convenient way to add comments to a program or to an operation.

In electromagnetics, you'll frequently encounter fields that have both a magnitude and a direction associated with them. Examples of physical effects from other disciplines that require a vector notation include force, acceleration, and velocity. A car traveling with a velocity v from a location A to a different location B implies that the car has a certain speed $\mathrm{v}=|\mathbf{v}|$ in a prescribed direction. In this case, the speed is the magnitude of the velocity. The vector should be contrasted with a scalar, a quantity that possesses only a magnitude and no direction. Energy, weight and speed are examples of scalar quantities. Our car can travel with a speed v in any direction but will pass us by with a velocity $\mathbf{v}$ in a definite direction.

## MATLAB and a review of vectors

The convenience of employing vector notation allows us visualize problems with or without the specification of a coordinate system. After choosing the coordinate system that will most accurately describe the field, the field is then specified with the components determined with regard to this coordinate system. Coordinate systems that we will encounter later are Cartesian, cylindrical, and spherical. The derivations of vector operations will be performed in Cartesian coordinates with the equivalent results just stated in the other systems. There are a large number of "orthogonal" coordinate systems and there is a generalized orthogonal coordinate system. The term "orthogonal" implies that every point in a particular coordinate system can be defined as the intersection of three orthogonal surfaces in that coordinate system. This will be further examined later.

A vector can be specified in MATLAB by stating its three components. We will use a capital letter to identify a vector in using MATLAB notation. Lower case letters will be reserved for scalar quantities. This is not required but it does add clarity to the work. The unit vector is defined as a vector whose magnitude is equal to 1 and it is directed in the same direction as the vector. For example in a Cartesian coordinate system, the vector $\mathbf{A}=A_{x} \mathbf{u}_{\mathbf{x}}+\mathrm{A}_{\mathbf{y}} \mathbf{u}_{\mathbf{y}}+\mathrm{A}_{\mathbf{z}} \mathbf{u}_{\mathbf{z}}$ where $\mathrm{A}_{\mathbf{x}}$ is the magnitude of the x component of the vector $\mathbf{A}$ and $\mathbf{u}_{\mathbf{x}}$ is a unit vector directed along the x axis is written as

$$
\begin{align*}
& \gg A=\left[A_{x} A_{y}\right] ; \\
& \gg A=\left[A_{x} A_{y} A_{z}\right] \tag{1.5}
\end{align*}
$$

## MATLAB and a review of vectors

in two and three dimensions respectively. In MATLAB notation, this vector can be displayed by just typing "A" at the prompt

$$
\begin{aligned}
& \gg A \\
& A=
\end{aligned}
$$

$A_{x} A_{y} A_{z}$
\gg
We must insert a space between the components of the vectors that are numbers.

Let us now specify numerical values for the three components $A=\left[\begin{array}{lll}1 & 2 & 3\end{array}\right]$.
A second vector $\mathbf{B}=\mathbf{2} \mathbf{u}_{\mathbf{x}}+3 \mathbf{u}_{\mathbf{y}}+4 \mathbf{u}_{\mathbf{z}}$ is written as

$$
\text { >> B = }\left[\begin{array}{lll}
2 & 3 & 4 \tag{1.7}
\end{array}\right] ;
$$

where we again employed the semicolon in order to save space.
Having stored the two vectors $\mathbf{A}$ and $\mathbf{B}$ in the computer, we can perform the following mathematical operations. The vectors can be added $\mathbf{C}=\mathbf{A}+\mathbf{B}$ by typing

$$
\begin{align*}
& \gg C=A+B \\
& C= \\
&  \tag{1.8}\\
& 357 \\
& \gg
\end{align*}
$$

The vector is interpreted as $\mathbf{C}=3 \mathbf{u}_{\mathbf{x}}+5 \mathbf{u}_{\mathbf{y}}+7 \mathbf{u}_{\mathbf{z}}$.
The two vectors can also be subtracted $\mathbf{D}=\mathbf{A}-\mathbf{B}$ with the command

$$
\begin{aligned}
& \gg D=A-B \\
& D=
\end{aligned}
$$

```
    -1 
> >
```

The vector is interpreted to be $\mathbf{D}=-1 \mathbf{u}_{\mathbf{x}}-1 \mathbf{u}_{\mathbf{y}}-1 \mathbf{u}_{\mathbf{z}}$. We will incorporate other vector operations such as the scalar product, the vector product, and various derivative operations using MATLAB at the location of their introduction.

Finally, the magnitude of a vector can also been computed using the MATLAB command 'norm'. The unit vector can be defined in any direction. It is equal to the vector divided by the magnitude of the vector. This will be illustrated in Example1-1.

Example $1-1^{\text {MATLAB }}$. Using MATLAB, plot the addition of the vectors $\mathbf{A}=3 \mathbf{u}_{\mathbf{x}}$ and $\mathbf{B}=4 \mathbf{u}_{\mathbf{y}}$. The vectors are to start at the point $(1,1)$. Plot and label the unit vectors $\mathbf{u}_{\mathbf{x}}, \mathbf{u}_{\mathbf{y}}$ and $\mathbf{u}_{\mathbf{c}}$ and find the magnitude of the vector $\mathbf{C}$.

Answer. The magnitude of the vector is calculated with the MATLAB command 'norm $(\mathbf{A}+\mathbf{B})$ '. The unit vector is $\mathbf{u}_{\mathbf{c}}=\frac{3 \mathbf{u}_{\mathrm{x}}+4 \mathbf{u}_{\mathrm{y}}}{\sqrt{3^{2}+4^{2}}}$ The default accuracy in MATLAB is 4 decimal places. This value can be controlled by the user. In addition, the 'title' of this figure has been added using another MATLAB command in the program.

At the present time, MATLAB does not have a feature to directly to create a vector with arrows. However, there exists a user contributed file entitled "vector3" at [http://www.mathworks.com/matlabcentral/spotlight/arrows.shtml/]. We acknowledge Jeff Chang and Tom Davis who gave us permission to incorporate this feature into the book.

## MATLAB and a review of vectors



The multiplication of two vectors can be accomplished with two different vector operations that have two different interpretations. The first multiplication is given the name of being a scalar product or a dot product. The definition of a scalar product of two vectors is

$$
\begin{equation*}
\mathbf{A} \bullet \mathbf{B} \equiv A B \cos \theta \tag{1.10}
\end{equation*}
$$

The result of this multiplication is a scaler quantity. The scalar product yields the projection of the vector $\mathbf{A}$ on the vector $\mathbf{B}$ as shown in Figure 1-1. Since $\cos \theta$

## MATLAB and a review of vectors

$=\cos (-\theta)$, this can also be interpreted as the projection of the vector $\mathbf{B}$ on the vector $\mathbf{A}$. If one of the vectors is a unit vector, this is a useful technique to calculate the components of a vector.


Figure $1-1^{\text {MATLAB }}$. Illustration of the scalar product of two vectors $\mathbf{A}=2 \mathbf{u}_{\mathbf{x}}+0$ $\mathbf{u}_{\mathbf{y}}+0 \mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=1 \mathbf{u}_{\mathbf{x}}+2 \mathbf{u}_{\mathbf{y}}+0 \mathbf{u}_{\mathbf{z}}$.

We can also interpret this operation in terms of work which is a scalar quantity. For example, if we are to move a box a distance $\Delta \mathbf{x}$ in a prescribed direction, we must apply a certain force $\mathbf{F}$ in the same direction. The total work $\Delta \mathrm{W}$ is given by the compact expression

## MATLAB and a review of vectors

$$
\begin{equation*}
\Delta \mathrm{W}=\mathrm{F} \cdot \Delta \mathbf{x} \tag{1.11}
\end{equation*}
$$

The MATLAB command that permits you to take a scaler product of the two vectors $\mathbf{A}$ and $\mathbf{B}$ is 'dot $(A, B)$ ' or ' $\operatorname{dot}(B, A)$ '.

The second vector multiplication of two vectors is called the vector product or the cross product and it is defined as

$$
\begin{equation*}
\mathbf{A} \times \mathbf{B} \equiv \mathrm{AB} \sin \theta \mathbf{u}_{\mathbf{A} \times \mathbf{B}} \tag{1.12}
\end{equation*}
$$

This multiplication yields a vector whose direction is determined by the "right hand rule." This implies that if you take the fingers of your right hand (vector $\mathbf{A}$ ) and close them in order to make a fist (vector $\mathbf{B}$ ), the unit vector $\mathbf{u}_{\mathbf{A}} \times \mathbf{B}$ will be in the direction of the thumb. Therefore, we find that $\mathbf{B} \times \mathbf{A}=-\mathbf{A} \mathbf{x} \mathbf{B}$ since you have had to rotate your right arm. The direction of rotation of a boat initially in a pond and carried up in a tornado or down in a whirlpool can be used as an indication of this asymmetry in the direction.

In Cartesian coordinates, we can easily calculate the vector product by remembering the expansion routine of the following determinant

$$
\begin{align*}
& A \times B=\left|\begin{array}{lll}
u_{x} & u_{y} & u_{z} \\
A_{x} & A_{y} & A_{z} \\
B_{x} & B_{y} & B_{z}
\end{array}\right| \\
& =\mathbf{u}_{x}\left(A_{y} B_{z}-A_{z} B_{y}\right)+\mathbf{u}_{y}\left(A_{z} B_{x}-A_{x} B_{z}\right)+\mathbf{u}_{z}\left(A_{x} B_{y}-A_{y} B_{x}\right) \tag{1.13}
\end{align*}
$$

The MATLAB command which effects the vector product of two vectors $\mathbf{A}$ and $\mathbf{B}$ is 'cross (A, B)'. In Figure 1-2, we illustrate the cross product of two vectors.

## MATLAB and a review of vectors



Figure $\mathbf{1 - 2}{ }^{\text {MATLAB }}$. The cross product of the two vectors $\mathbf{A}=2 \mathbf{u}_{\mathbf{X}}+1 \mathbf{u}_{\mathbf{y}}+0 \mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=1 \mathbf{u}_{\mathbf{x}}+2 \mathbf{u}_{\mathbf{y}}+0 \mathbf{u}_{\mathbf{z}}$ are shown The vector product of the two vectors $\mathbf{A}$ and $\mathbf{B}$ is equal to $\mathbf{C}=0 \mathbf{u}_{\mathbf{x}}+0 \mathbf{u}_{\mathbf{y}}+3 \mathbf{u}_{\mathbf{z}}$.

Example 1-2 ${ }^{\text {MATLAB }}$. Show that it is possible to interpret the cross product of two vectors that are in a plane in terms of the area included between the two vectors. Answer. It is possible to give a geometric interpretation for the vector product. We can use the magnitude of the vector product $|\mathbf{A} \mathbf{x} \mathbf{B}|$ to find the area of a parallelogram whose sides are specified by the vectors $\mathbf{A}$ and $\mathbf{B}$ as shown in the figure.


From geometry, the area of a parallelogram is given by

$$
\text { Area }=A B \sin \theta
$$

This is equal to the area of the rectangle whose area is given by the product of the horizontal distance A and the vertical distance $\mathrm{B} \sin \theta$. From the definition of the vector product (1.12), this is equal to $|\mathbf{A} \times \mathbf{B}|$

A convenient method of stating that two non-zero vectors $\mathbf{A}$ and $\mathbf{B}$ are pendicular $\left(\theta=90^{\circ}\right)$ is to use the scalar product. If $\mathbf{A} \cdot \mathbf{B}=0$ and neither vector is zero, then the two vectors are perpendicular since $\cos 90^{\circ}=0$. To state that

## MATLAB and a review of vectors

the two non-zero vectors are parallel $\left(\theta=0^{\circ}\right)$ or antiparallel $\left(\theta=180^{\circ}\right)$, we use the vector product. If $\mathbf{A} \mathbf{x}=\mathbf{B}$, then the two vectors are parallel or antiparallel since $\sin 0^{\circ}=\sin 180^{\circ}=0$.

Two triple products will be encountered in electromagnetic theory and they are included here. The first is called the scalar triple product. It is defined as

$$
\begin{equation*}
A \cdot(B \times C)=B \cdot(C \times A)=C \cdot(A \times B) \tag{1.14}
\end{equation*}
$$

Note the cyclical permutation of the vectors in (1.14). There are several additional possible permutations to this product since the change of the vector product will lead to a minus sign since,

$$
\begin{equation*}
B \times C=-C \times B . \tag{1.15}
\end{equation*}
$$

The second triple product is called the vector triple product.

$$
\begin{equation*}
A \times(B \times C)=B(A \bullet C)-C(A \bullet B) \tag{1.16}
\end{equation*}
$$

The proper inclusion of the parentheses in this triple product is critical. Each of the vectors is mollified with a numerical value that is determined by the scalar product of two vectors. This triple product is sometimes called the "back - cab" rule since this is an easy way to remember the ordering of the vectors. We will use this particular vector identity extensively when the topic of Poynting's vector is introduced.

Example 1-3 MATLAB. Show that the volume $\Delta V$ of a parallelepiped defined by three vectors originating at a point can be defined in terms of the scalar and the vector products of the vectors.

Answer: The volume $\Delta \mathrm{V}$ of the parallelepiped is given by

## MATLAB and a review of vectors

$$
\Delta V=\text { (area of the base of the parallelepiped) } x \text { (height of the }
$$

parallelepiped)

$$
\begin{aligned}
& =(|\mathbf{A} \times \mathbf{B}|)\left(\mathbf{C} \cdot \mathbf{u}_{\mathrm{n}}\right) \\
& =(|\mathbf{A} \times \mathbf{B}|) \cdot\left(\mathbf{C} \cdot\left(\frac{\mathbf{A} \times \mathbf{B}}{|\mathbf{A} \times \mathbf{B}|}\right)\right) \\
& =\mathbf{C} \cdot(\mathbf{A} \times \mathbf{B})
\end{aligned}
$$

This is illustrated with the vectors defined as $A=\left[\begin{array}{lll}3 & 0 & 0\end{array}\right] ; B=\left[\begin{array}{lll}0 & 2 & 0\end{array}\right] ;$ and $C=\left[\begin{array}{ll}0 & 2\end{array}\right.$ 4]. The calculated volume is equal to 24 .


## MATLAB and a review of vectors

Note that the height of the parallelepiped is given by the scalar product of the vector $C$ with the unit to vector $(\mathbf{A} \times \mathbf{B} /|\mathbf{A} \times \mathbf{B}|)$ that is perpendicular to the base.

MATLAB provides extensive two and three dimensional graphical plotting routines. The data to be plotted can be generated internally in a program or it can be imported from an external program. The command 'fplot' specifies and plots a known function that is included in the MATLAB library. Labels and titles using different fonts and font sizes and styles can be placed on the graphs and the plots can be distinguished with different symbols. The MATLAB notation for a superscript and a subscript require the additional statement ' $\wedge$ ' and ',' respectively in the text command. We will present several examples here in order to illustrate the variety of two dimensional plots that are available. Additional graphs can be placed on one plot with the "hold" command. Either axis can have a logarithmic scale. In addition, the command "subplot" permits us to place more than one graph on a page, either vertically or horizontally displaced. The command 'subplot $(1,2,1)$ ' states that there are to be two graphs next to each other and this command will be used to select the left one. The command 'subplot (2, 1, 2)' states that are two graphs, one on top of the other and the command selects the bottom one. Other commands that follow detail the characteristics of that particular graph. This is best illustrated with an example.

Example 1-4 MATLAB. Construct 4 subplots on one figure. (a) A bar graph that contains five numbers $x=2,4,6,8,10$. (b) Plot the numbers $y=5,4,3,2,1$ vs x. (c) Plot two cycles of a sine wave using the 'fplot' command. The introduction
of the symbol $\theta$ is accomplished with the command 'ltheta' in the xlabel or in a text statement. (d) Plot an exponential function in the range $0<x<3$. Calculate this function with the interval $\Delta x=0.01$. Text items such as the ylabel or a statement can include superscripts and subscripts. The superscript is introduced with the command ' $\wedge$ ' and the subscript is introduced with the command '.'.


It is possible to customize a graph by changing the characteristics of the line. This is illustrated by plotting the same function, say $\sin \theta$ vs. the independent variable $\theta$. In addition, the text item "(a)" is sequenced in the

## MATLAB and a review of vectors

program using the command ' $s(2)=s e t s t r(s(2)+1)$ ' after the initial inclusion of the statement $s=$ "(a)" in the program. This is illustrated in Figure 1-3.


Figure 1-3 MATLAB. Illustration of a variety of different styles for the lines in a graph. (a) Solid line. (b) Dashed line. (c) Alternate 'o -'.

MATLAB also permits the graphical representation in a three-dimensional graph. The horizontal space is subdivided into a large number of points $\left(x_{j}, y_{k}\right)$ and the function $z=z\left(x_{j}, y_{k}\right)$ has to be evaluated at each of these points. In order to accomplish this. A "." (period) must follow each of the independent variables in a program. The results of a three-dimensional picture is illustrated in Figure

1-4. There are two distinct plot commands, 'mesh and 'surf'. In addition, there are also commands that allow the user to change the 'viewing angle', both in the rotation and in the elevation.


Figure $1-4$ MATLAB . Three-dimensional plots of a Gaussian function. (a) mesh plot. (b) surf plot.

In addition to plotting the figure in Cartesian coordinates, it is also possible to plot the figure in polar coordinates. This will be useful in examining the radiation pattern of antennas. In addition, one can plot graphs in a semi log format or in a log log format. This graph is useful when one wants to ascertain

## MATLAB and a review of vectors

(a)

(b)


Figure 1-5 ${ }^{\text {MATLAB }}$. Polar plot and a $\log -\log$ graph. (a) $|\sin \theta|$ versus the angle $\theta$ where $0<\theta<2 \pi$. (b) $y=x^{n}$ where $1<x<100$.
the variation of a function $y=x^{n}$, say in the interpretation on the data collected in a laboratory experiment. Examples of these figures are shown in Figure 1-5.

The functions that have been used above are all functions that are included in the MATLAB library of functions. The question then arises, "Can a program written by us which we expect to use again also be included in the library?" The answer is "Yes" and the process is given the name of creating a ".m" (dot m) file with its unique name, say "custom.m". The creation of the "custom.m" file involves a text editor, the form of which depends upon the local computer or work station to which the user has access. Once the file is created, it becomes a part of our personal library. In order to use this file, all we have to do is type the word "custom" after the prompt ">>" and this particular file is activated

## MATLAB and a review of vectors

at that point. These files are frequently shared over the internet and all of the files that have been used to create either the figures or the examples in this book are available at the following web site: [http://www.scitechpub.com/].

In the discussion above, we have focused on single vectors. As we will see later in this text, we will encounter a distribution of vectors that is called a vector field This would be similar to the wind distribution in a region where the wind at any point has a magnitude and a direction associated with it. We typically would ascribe a vector field that had different length vectors as representing a nonuniform distribution of wind. The distribution of temperature throughout the nation would be specified by numbers or scalars and this would be an example of a scalar field. Examples of the two fields are shown in Figure 1-6.


Figure 1-6. Scalar and vector fields. (a) The magnitude of a scalar is specified by the size of the circle. (b) The magnitude and the direction of the vector at any point is indicated with the length and the orientation on the vectors.

## Coordinate systems

### 1.2. Coordinate systems

In this text, we will frequently encounter problems where there is a source of an electromagnetic field. To be able to specify the field at a point in space caused by this source, we have to make reference to a coordinate system. In three dimensions, the coordinate system can be specified by the intersection of three surfaces. Each surface is described with $\xi_{1}=$ constant, $\xi_{2}=$ constant, and $\xi_{3}=$ constant where $\xi_{\mathrm{j}}$ is the $\mathrm{j}^{\text {th }}$ axis of an orthogonal coordinate system. An orthogonal coordinate system is defined when these three surfaces are mutually orthogonal at a point. Thus surfaces may be planar or may be curved. This is illustrated in Figure 1-7.


Figure 1-7. A general orthogonal coordinate system. Three surfaces intersect at a point and the unit vectors are mutually orthogonal at that point.

In Cartesian coordinates, all of the surfaces are planes and they are specified by each of the independent variables $x, y$ and $z$ separately being a constant. In cylindrical coordinates, the surfaces are two planes and a cylinder.

## Coordinate systems

In spherical coordinates, the surfaces are a sphere, a plane, and a cone. We will examine each of these in detail in the following discussion. There are many more coordinate systems that can be employed for particular problems and there even exists a generalized coordinate system that allows one to easily transform the vector operations from one system to another.

The three coordinate systems used in this text are also displayed in Figure 1-8 (a), (b) and (c). The directions along the axes of the coordinate systems are given by the sets of unit vectors $\left(\mathbf{u}_{\mathbf{x}}, \mathbf{u}_{\mathbf{y}}, \mathbf{u}_{\mathbf{z}}\right) ;\left(\mathbf{u}_{\mathbf{r}}, \mathbf{u}_{\phi}, \mathbf{u}_{\mathbf{z}}\right)$; and $\left(\mathbf{u}_{\boldsymbol{\rho}}, \mathbf{u}_{\phi}, \mathbf{u}_{\theta}\right)$ for Cartesian, cylindrical and spherical coordinates respectively. The radial variable in the cylindrical and spherical coordinates have been chosen to have different symbols. In each of the coordinate systems, the unit vectors are orthogonal to each other at every point.


Figure 1-8. The three coordinate systems that will be employed in this text. The unit vectors are indicated. (a) Cartesian coordinates. (b) Cylindrical coordinates. (c) Spherical coordinates.

## Coordinate systems

The cross product of two unit vectors can be used to to define a unit surface whose vector direction is in the direction of the third unit vector. In Cartesian coordinates, the direction of the unit vectors is independent of position. However in cylindrical and spherical coordinate the directions do depend on their positions. For example, in spherical coordinates, The unit vector $\mathbf{u}_{\rho}$ will be directed in the $+z$ axis if $\theta=0$ and it will be directed in the $-z$ axis if $\theta=\pi$. Since we will employ these three coordinate systems extensively in the following chapters, it is useful to summarize germane properties of each one.

### 1.2.1.Cartesian coordinates



Figure 1-9. A point in Cartesian coordinates is defined by the intersection of the three planes: $x=$ constant, $y=$ constant, $z=$ constant. The three unit vectors are normal to each of the three surfaces.

## Coordinate systems

The unit vectors in Cartesian coordinates depicted in Figure 1-8a are normal to the intersection of three planes as shown in Figure 1-9. Each of the surfaces depicted in this figure is a plane that is individually normal to a coordinate axis.

For the unit vectors that are in the directions of the $x, y$, and $z$ axes, we have $\mathbf{u}_{\mathbf{x}} \cdot \mathbf{u}_{\mathbf{x}}=\mathbf{u}_{\mathbf{y}} \cdot \mathbf{u}_{\mathbf{y}}=\mathbf{u}_{\mathbf{z}} \cdot \mathbf{u}_{\mathbf{z}}=1$ and $\mathbf{u}_{\mathbf{x}} \cdot \mathbf{u}_{\mathbf{y}}=\mathbf{u}_{\mathbf{x}} \cdot \mathbf{u}_{\mathbf{z}}=\mathbf{u}_{\mathbf{y}} \cdot \mathbf{u}_{\mathbf{z}}=0$. In Cartesian coordinates, the following rules also apply to the unit vectors since this is a right-handed system.

$$
\left.\begin{array}{l}
\mathbf{u}_{\mathrm{x}} \times \mathbf{u}_{\mathrm{y}}=\mathbf{u}_{\mathrm{z}} \\
\mathbf{u}_{\mathrm{y}} \times \mathbf{u}_{\mathbf{z}}=\mathbf{u}_{\mathrm{x}}  \tag{1.17}\\
\mathbf{u}_{\mathrm{z}} \times \mathbf{u}_{\mathrm{x}}=\mathbf{u}_{\mathrm{y}}
\end{array}\right\}
$$

The vector $\mathbf{A}$ from the origin of the coordinate system to the point determined by the intersection of the three planes in Figure 1-9 is given by

$$
\begin{equation*}
\mathbf{A}=x \mathbf{u}_{\mathbf{x}}+y \mathbf{u}_{\mathbf{y}}+\mathrm{z} \mathbf{u}_{\mathbf{z}} \tag{1.18}
\end{equation*}
$$

Example 1-5 ${ }^{\text {MATLAB }}$. State the MATLAB commands for the three unit vectors in Cartesian coordinates. A general feature of all the orthogonal coordinate systems is that the unit vectors at any point defined by the intersection of the three surfaces is that the unit vectors are mutually orthogonal at that point.

Answer. The MATLAB commands for the unit vectors are written as:

$$
\mathbf{u}_{\mathbf{x}} \rightarrow \text { "ux = }\left[\begin{array}{lll}
1 & 0 & 0
\end{array}\right] ", \mathbf{u}_{\mathbf{y}} \rightarrow \text { "uy = }\left[\begin{array}{lll}
0 & 1 & 0
\end{array}\right] ", \text { and } \mathbf{u}_{\mathbf{z}} \rightarrow \text { "uz = }\left[\begin{array}{lll}
0 & 0 & 1
\end{array}\right] " .
$$

The unit vectors are depicted below.

## Coordinate systems



In Figure 1-10, a differential volume in Cartesian coordinates is illustrated. This small volume has six surfaces. A differential surface area is a vector quantity whose direction is defined to be in the "outward normal" direction. Three of the six differential surface areas are indicated.


Figure 1-10. A differential volume $d v=(d x)(d y)(d z)$ in Cartesian coordinates. Three of the six differential surface areas: $\mathbf{d s}_{\mathbf{x}}=(\mathrm{dy})(\mathrm{dz}) \mathbf{u}_{\mathrm{x}}, \mathbf{d} \mathbf{s}_{\mathbf{y}}=(\mathrm{dx})(\mathrm{dz}) \mathbf{u}_{\mathbf{y}}$, and $\mathbf{d s}_{\mathbf{z}}=(\mathrm{dx})(\mathrm{dy}) \mathbf{u}_{\mathbf{z}}$ are indicated with the vectors. The differential length $\mathbf{d l}$ is defined as $\mathbf{d l}=d x \mathbf{u}_{\mathbf{x}}+d y \mathbf{u}_{\mathbf{y}}+\mathrm{dz} \mathbf{u}_{\mathbf{z}}$.

Example 1-6. Find the vector $\mathbf{G}$ that joins point $A$ to point $B$ on the enclosed graph in Cartesian coordinates. In addition, determine the unit vector.

## Coordinate systems



Answer: The vector is

$$
\mathbf{G}=(6-2) \mathbf{u}_{\mathbf{x}}+(7-2) \mathbf{u}_{\mathbf{y}}+0 \mathbf{u}_{\mathbf{z}}=4 \mathbf{u}_{\mathbf{x}}+5 \mathbf{u}_{\mathbf{y}}+0 \mathbf{u}_{\mathbf{z}}
$$

The unit vector is

$$
\mathbf{u}_{\mathrm{G}}=\frac{4 \mathbf{u}_{\mathrm{x}}+5 \mathbf{u}_{\mathrm{y}}+0 \mathbf{u}_{\mathbf{z}}}{\sqrt{4^{2}+5^{2}+0^{2}}}
$$

Example $\mathbf{1 - 7}^{\text {MATLAB }}$. Given the two vectors $\mathbf{A}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}$ and $\mathbf{B}=12 \mathbf{u}_{\mathbf{x}}+5 \mathbf{u}_{\mathbf{y}}$ in Cartesian coordinates, evaluate the following quantities. In addition, state the MATLAB commands that can be used to check your answers. The vectors are written in MATLAB notation as " $A=\left[\begin{array}{lll}3 & 4 & 0\end{array}\right]$ " and " $B=\left[\begin{array}{lll}12 & 5 & 0\end{array}\right]$ ".
(a) the scalar product $\mathbf{A} \cdot \mathbf{B}$.
(b) the angle between the two vectors.
(c) the scalar product $\mathbf{A} \cdot \mathbf{A}$.

## Coordinate systems

(d) the vector product $\mathbf{A} \times \mathbf{B}$.

## Answer:

(a) The scalar product $\mathbf{A} \cdot \boldsymbol{B}$ is given by

$$
A \bullet B=36 \mathbf{u}_{x} \bullet \mathbf{u}_{x}+15 \mathbf{u}_{x} \bullet \mathbf{u}_{y}+48 \mathbf{u}_{\mathbf{y}} \bullet \mathbf{u}_{x}+20 \mathbf{u}_{\mathbf{y}} \bullet \mathbf{u}_{\mathbf{y}}
$$

Note - the scalar product of two orthogonal unit vectors is equal to 0 and two collinear unit vectors is equal to 1 . This leads to $\mathbf{A} \bullet \mathbf{B}=36+0+0+20=56$. In MATLAB, use 'dot(A, B)'.
(b) The angle between the two vectors is computed from the definition of the scalar product.

$$
\cos \theta=\frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|}=\frac{56}{\sqrt{3^{2}+4^{2}} \sqrt{12^{2}+5^{2}}}=\frac{56}{65} \text { or } \theta=30.5^{\circ}
$$

In MATLAB, 'theta $=\operatorname{acos}\left(\operatorname{dot}(\mathrm{A}, \mathrm{B}) /\left(\operatorname{norm}(\mathrm{A})^{*} \operatorname{norm}(\mathrm{~B})\right)\right)^{*}(180 / \mathrm{pi})$ '
(c) The scalar product $\mathbf{A} \cdot \mathbf{A}$ is given by

$$
A \cdot A=9 u_{\mathbf{x}} \cdot \mathbf{u}_{\mathbf{x}}+16 \mathbf{u}_{\mathbf{y}} \cdot \mathbf{u}_{\mathbf{y}}=25
$$

The scalar product $\mathbf{A} \cdot \mathbf{A}$ is a convenient method to determine the magnitude of the vector $\mathbf{A}$ since $\mathbf{A}=|\mathbf{A}|=\sqrt{\mathbf{A} \bullet \mathbf{A}}$. The MATLAB command is ' $\operatorname{dot}(A, A)$ '.
(d) The vector product $\mathbf{A} \mathbf{x} \mathbf{B}$ is given by

$$
\mathbf{A} \times \mathbf{B}=\left|\begin{array}{ccc}
\mathbf{u}_{\mathrm{x}} & \mathbf{u}_{\mathbf{y}} & \mathbf{u}_{z} \\
3 & 4 & 0 \\
12 & 5 & 0
\end{array}\right|=-33 \mathbf{u}_{\mathbf{z}}
$$

The MATLAB command is ' $\operatorname{cross}(A, B)$ '.

## Coordinate systems

### 1.2.2. Cylindrical coordinates



Figure 1-11. Cylindrical coordinates. The point is defined by the intersection of the cylinder and the two planes. A differential volume is also shown.

The unit vectors in cylindrical coordinates depicted in Figure 1-8b are normal to the intersection of three surfaces as shown in Figure 1-11. Two of the surfaces depicted in this figure are planes and the third surface is a cylinder that is centered on the $z$ axis. A point in cylindrical coordinates is specified by the

## Coordinate systems

intersection of the two planes and the cylinder. The unit vectors $\mathbf{u}_{\mathbf{r}}, \mathbf{u}_{\phi}$, and $\mathbf{u}_{\mathbf{z}}$ are perpendicular to the two planes and to the cylinder.

In cylindrical coordinates, the following rules apply to the unit vectors since this is a "right-handed system."

$$
\left.\begin{array}{l}
\mathbf{u}_{\mathbf{r}} \times \mathbf{u}_{\phi}=\mathbf{u}_{\mathbf{z}}  \tag{1.19}\\
\mathbf{u}_{\phi} \times \mathbf{u}_{\mathbf{z}}=\mathbf{u}_{\mathbf{r}} \\
\mathbf{u}_{\mathbf{z}} \times \mathbf{u}_{\mathbf{r}}=\mathbf{u}_{\phi}
\end{array}\right\}
$$

The vector $\mathbf{A}$ from the origin of the coordinate system to the point determined by the intersection of the two planes and the cylinder is given by

$$
\begin{equation*}
\mathbf{A}=r \mathbf{u}_{\mathbf{r}}+\phi \mathbf{u}_{\phi}+\mathrm{z} \mathbf{u}_{\mathbf{z}} \tag{1.20}
\end{equation*}
$$

In cylindrical coordinates, the differential length dI, the differential surface area ds, and the differential volume dv are respectively written as

$$
\begin{gather*}
\mathbf{d l}=d r \mathbf{u r}_{\mathbf{r}}+\mathrm{rd} \phi \mathbf{u}_{\phi}+\mathrm{dz} \mathbf{u}_{\mathbf{z}}  \tag{1.21}\\
\mathbf{d s}=\mathrm{rd} \phi \mathrm{dz} \mathbf{u}_{\mathbf{r}}+\operatorname{dr} \mathrm{dz} \mathbf{u}_{\phi}+\mathrm{rdr} \mathrm{~d} \phi \mathbf{u}_{\mathbf{z}}  \tag{1.22}\\
\mathrm{dv}=\mathrm{rdrd} d \mathrm{dz} \tag{1.23}
\end{gather*}
$$

A vector in cylindrical coordinates can be transformed to a vector in Cartesian coordinates or vice versa. The vector $\mathbf{A}=A_{\mathbf{x}} \mathbf{u}_{\mathbf{x}}+A_{\mathbf{y}} \mathbf{u}_{\mathbf{y}}+A_{\mathbf{z}} \mathbf{u}_{\mathbf{z}}$ is in Cartesian acoordinates and the same vector $\mathbf{A}=A_{r} \mathbf{u}_{\boldsymbol{r}}+A_{\phi} \mathbf{u}_{\phi}+A_{\mathbf{z}} \mathbf{u}_{\mathbf{z}}$ is in cylindrical coordinates. The unknown quantities in this transformation are the coefficients associated with each of the unit vectors and the relationship between these unit vectors.

## Coordinate systems



Figure $\mathbf{1 - 1 2}{ }^{\text {MATLAB }}$. The transformation of a vector $\mathbf{A}=3 \mathbf{u}_{\mathrm{x}}+2 \mathbf{u}_{\mathbf{y}}+4 \mathbf{u}_{\mathbf{z}}$ in Cartesian coordinates into a vector into a vector in cylindrical coordinates. The unit vectors of the two coordinate systems are indicated.

The transformation between the two coordinate systems is found by taking the scalar product of the unit vector in the cylindrical coordinate system with the vector in the Cartesian coordinate system and correctly interpreting the scalar

## Coordinate systems

products of the unit vectors. We should immediately note that the $A_{z}$ terms are the same in either coordinate system. We write

$$
\begin{aligned}
A_{x}= & A \cdot \mathbf{u}_{\mathbf{x}} \\
& =A_{r} u_{r} \cdot \mathbf{u}_{\mathbf{x}}+A_{\phi} \mathbf{u}_{\phi} \cdot \mathbf{u}_{\mathbf{x}}
\end{aligned}
$$

From Figure 1-12, we note that

$$
\begin{gathered}
\mathbf{u}_{\mathbf{r}} \cdot \mathbf{u}_{\mathbf{x}}=\cos \phi \\
\mathbf{u}_{\varphi} \cdot \mathbf{u}_{\mathbf{x}}=\cos \left(\frac{\pi}{2}+\phi\right)=-\sin \phi
\end{gathered}
$$

Therefore

$$
\begin{equation*}
A_{X}=A r \cos \phi-A_{\phi} \sin \phi \tag{1.24}
\end{equation*}
$$

Similarly

$$
\begin{aligned}
A_{y} & =\mathbf{A} \cdot \mathbf{u}_{\mathbf{y}} \\
& =A_{r} \mathbf{u}_{\mathbf{r}} \cdot \mathbf{u}_{\mathbf{y}}+A_{\phi} \mathbf{u}_{\phi} \cdot \mathbf{u}_{\mathbf{y}}
\end{aligned}
$$

where

$$
\begin{gathered}
\mathbf{u}_{\mathbf{r}} \cdot \mathbf{u}_{\mathbf{y}}=\cos \left(\frac{\pi}{2}-\phi\right)=\sin \phi \\
\mathbf{u}_{\varphi} \cdot \mathbf{u}_{\mathbf{y}}=\cos \phi
\end{gathered}
$$

Hence

$$
\begin{equation*}
A_{y}=A_{r} \sin \phi+A_{\phi} \cos \phi \tag{1.25}
\end{equation*}
$$

In Figure 1-12 the vector $\mathbf{A}$ was assumed to be a constant. In reality, it may also be a function of the independent variables. These variables must also be transformed. From cylindrical to Cartesian coordinates, we write

## Coordinate systems

$$
\left.\begin{array}{c}
x=r \cos \phi \\
y=r \sin \phi  \tag{1.26}\\
z=z
\end{array}\right\}
$$

The inverse transformation from Cartesian to cylindrical coordinates is given by

$$
\left.\begin{array}{c}
r=\sqrt{x^{2}+y^{2}} \\
\phi=\tan ^{-1}\left(\frac{y}{x}\right)  \tag{1.27}\\
z=z
\end{array}\right\}
$$

There are commands in MATLAB that will effect this transformation between cylindrical and Cartesian coordinates. In addition, the command 'cylinder' that includes additional parameters will create a picture of a cylinder.

### 1.2.3. Spherical coordinates

The unit vectors in spherical coordinates depicted in Figure 1-8c are normal to the intersection of three surfaces as shown in Figure 1-13. One of the surfaces depicted in this figure is a plane, another surface is a sphere and the third surface is a cone. The latter two surfaces are centered on the $z$ axis. A point in spherical coordinates is specified by the intersection of the three surfaces. The unit vectors $\mathbf{u}_{\rho}, \mathbf{u}_{\theta}$, and $\mathbf{u}_{\phi}$ are perpendicular to the sphere, the cone, and the plane.

In spherical coordinates, the following rules apply to the unit vectors since this is a right-handed system.

## Coordinate systems

$$
\begin{align*}
& \mathbf{u}_{\rho} \times \mathbf{u}_{\theta}=\mathbf{u}_{\phi} \\
& \mathbf{u}_{\theta} \times \mathbf{u}_{\phi}=\mathbf{u}_{\rho}  \tag{1.28}\\
& \mathbf{u}_{\phi} \times \mathbf{u}_{\rho}=\mathbf{u}_{\theta}
\end{align*}
$$



Figure $1-13^{\text {MATLAB }}$. Spherical coordinates. The point is defined by the intersection of a sphere whose radius is $\rho$, a plane that makes an angle $\phi$ with respect to the $x$ axis, and a cone that makes an angle $\theta$ with respect to the $z$ axis. A differential volume is also shown.

## Coordinate systems

The vector $\mathbf{A}$ from the origin of the coordinate system to the point determined by the intersection of the sphere whose radius is $\rho$, the plane that makes an angle $\phi$ with respect to the x axis, and the cone that makes an angle $\theta$ with respect to the $z$ axis. The radial variable has a different letter in spherical coordinates than it had in cylindrical coordinates in order to avoid confusion.

$$
\begin{equation*}
\mathbf{A}=A_{\rho} \mathbf{u}_{\rho}+A_{\theta} \mathbf{u}_{\theta}+A_{\phi} \mathbf{u}_{\phi} \tag{1.29}
\end{equation*}
$$

The differential length dl, the differential surface area ds, and differential volume dv are respectively written as

$$
\begin{gather*}
d \mathbf{l}=d \rho u_{\rho}+\rho \sin \theta d \theta \mathbf{u}_{\theta}+\rho d \phi \mathbf{u}_{\varphi}  \tag{1.30}\\
d \boldsymbol{s}=\rho^{2} \sin \theta d \theta d \phi \mathbf{u}_{\rho}+\rho \sin \theta d \rho d \phi \mathbf{u}_{\theta}+\rho d \rho d \theta \mathbf{u}_{\phi}  \tag{1.31}\\
d \mathbf{v}=\rho^{2} \sin \theta d \rho d \theta d \phi \tag{1.32}
\end{gather*}
$$

A convenient point to check at this point is the dimensions of these three expressions. In SI units, both the length $\rho$ and the differential length d $\rho$ have units of meters. Hence the dimensions of dl are in meters, ds are in (meters) ${ }^{2}$, and dv are in (meters) ${ }^{3}$.

Example 1-8. Show that a vector given in spherical coordinates can be expressed in Cartesian coordinates.

Answer: The vector $\mathbf{A}=A_{\mathbf{x}} \mathbf{u}_{\mathbf{x}}+\mathrm{A}_{\mathbf{y}} \mathbf{u}_{\mathbf{y}}+\mathrm{A}_{\mathbf{z}} \mathbf{u}_{\mathbf{z}}$ in Cartesian coordinates and $\mathbf{A}=$ $A_{\rho} \mathbf{u}_{\rho}+A_{\theta} \mathbf{u}_{\theta}+A_{\phi} \mathbf{u}_{\phi}$ in spherical coordinates. The transformation between the two coordinate systems is found by taking the scalar product of the unit vector in the new coordinate system with the vector in the other coordinate system and correctly interpreting the scalar products of the unit vectors

## Coordinate systems

$$
\begin{aligned}
A_{\mathbf{X}} & =\mathbf{A} \cdot \mathbf{u}_{\mathbf{X}} \\
& =A_{\rho} \mathbf{u}_{\rho} \cdot \mathbf{u}_{\mathbf{X}}+A_{\theta} \mathbf{u}_{\theta} \cdot \mathbf{u}_{\mathbf{X}}+\mathrm{A}_{\phi} \mathbf{u}_{\phi} \cdot \mathbf{u}_{\mathbf{X}}
\end{aligned}
$$

From the figure, we note that


$$
\begin{gathered}
\mathbf{u}_{\boldsymbol{\rho}} \bullet \mathbf{u}_{\mathbf{x}}=\sin \theta \cos \phi=\frac{\mathrm{x}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}+\mathrm{z}^{2}}} \\
\mathbf{u}_{\theta} \bullet \mathbf{u}_{\mathbf{x}}=\cos \theta \cos \phi=\frac{\mathrm{x} \mathbf{z}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}} \sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}+\mathrm{z}^{2}}} \\
\mathbf{u}_{\varphi} \bullet \mathbf{u}_{\mathbf{x}}=-\sin \phi=-\frac{\mathrm{x}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}}}
\end{gathered}
$$

In a similar fashion, we write

$$
\begin{gathered}
\mathbf{u}_{\boldsymbol{\rho}} \bullet \mathbf{u}_{\mathbf{y}}=\sin \theta \sin \phi=\frac{\mathrm{y}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}+\mathrm{z}^{2}}} \\
\mathbf{u}_{\theta} \bullet \mathbf{u}_{\mathbf{y}}=\cos \theta \sin \phi=\frac{\mathrm{yz}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2} \sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}+\mathrm{z}^{2}}}} \\
\mathbf{u}_{\varphi} \bullet \mathbf{u}_{\mathbf{y}}=\cos \phi=\frac{\mathrm{y}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}}}
\end{gathered}
$$

and

## Coordinate systems

$$
\begin{aligned}
& \mathbf{u}_{\rho} \cdot \mathbf{u}_{\mathbf{z}}= \cos \theta=\frac{\mathrm{z}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}+\mathrm{z}^{2}}} \\
& \mathbf{u}_{\theta} \cdot \mathbf{u}_{\mathbf{z}}=-\sin \theta= \\
&-\frac{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}}}{\sqrt{\mathrm{x}^{2}+\mathrm{y}^{2}+\mathrm{z}^{2}}} \\
& \mathbf{u}_{\varphi} \cdot \mathbf{u}_{\mathbf{z}}=0
\end{aligned}
$$

The transformations of the variables from Cartesian to spherical coordinates yields

$$
\left.\begin{array}{c}
\rho=\sqrt{x^{2}+y^{2}+z^{2}} \\
\theta=\tan ^{-1}\left(\frac{\sqrt{x^{2}+y^{2}}}{z}\right)  \tag{1.33}\\
\phi=\tan ^{-1}\left(\frac{y}{x}\right)
\end{array}\right\}
$$

The transformation on the variables from spherical to Cartesian coordinates is

$$
\left.\begin{array}{rl}
x & =\rho \sin \theta \cos \phi  \tag{1.34}\\
y & =\rho \sin \theta \sin \phi \\
z=\rho \cos \theta
\end{array}\right\}
$$

A summary of the unit vectors, the differential lengths, the differential surfaces, and the differential volumes for the three coordinate systems is given in Table 1-1 (see also Appendix A.1.4).

## Coordinate systems

| Coordinate system | Cartesian $(x, y, z)$ | $\begin{aligned} & \text { Cylindrical } \\ & (r, \phi, z) \end{aligned}$ | Spherical $(\rho, \theta, \phi)$ |
| :---: | :---: | :---: | :---: |
| Unit vectors | $\mathbf{u}_{\mathrm{x}} \mathrm{u}_{\mathrm{y}} \mathrm{u}_{\mathrm{z}}$ | $\mathbf{u}_{\mathrm{r}} \mathbf{u}_{\phi} \mathbf{u}_{\mathbf{z}}$ | $\mathbf{u}_{\mathrm{p}} \mathbf{u}_{\theta} \mathbf{u}_{\phi}$ |
| Differential length dl | $d x \mathbf{u}_{\mathbf{x}}$ <br> dy $\mathbf{u}_{\mathbf{y}}$ <br> $\mathrm{dz} \mathbf{u}_{\mathbf{z}}$ | $\mathrm{dr} \mathbf{u}_{\mathrm{r}}$ <br> $r d \phi \mathbf{u}_{\phi}$ <br> $\mathrm{dz} \mathbf{u}_{\mathbf{z}}$ | $\begin{aligned} & d \rho \mathbf{u}_{\rho} \\ & \rho \sin \theta d \theta \mathbf{u}_{\theta} \\ & \rho d \phi \mathbf{u}_{\phi} \end{aligned}$ |
| Differential surface area ds | $d y d z u_{x}$ <br> $d x d z \mathbf{u}_{\mathbf{y}}$ <br> $d x d y u_{z}$ | $r d \phi d z \mathbf{u}_{r}$ <br> $\mathrm{dr} \mathrm{dz} \mathbf{u}_{\phi}$ <br> $r \operatorname{dr} d \phi \mathbf{u}_{\mathbf{z}}$ | $\begin{aligned} & \rho^{2} \sin \theta d \theta d \phi \mathbf{u}_{\rho} \\ & \rho \sin \theta d \rho d \phi \mathbf{u}_{\theta} \\ & \rho d \rho d \theta \mathbf{u}_{\phi} \end{aligned}$ |
| Differential volume dv | dx dy dz | $\mathrm{r} d \mathrm{dr} \phi \mathrm{dz}$ | $\rho^{2} \sin \theta d \rho d \theta d \phi$ |

Table 1-1. Three orthogonal coordinate systems.

A summary of the transformations of the variables between coordinate systems is given in Table 1-2. In addition, the MATLAB commands that will perform these operations are also presented. Appendix 1 provides a summary of the vector operations which will be defined later in these three coordinate systems.

The simple vector mathematics of addition and subtraction in Cartesian coordinates that was described previously requires a little more care in the other coordinate systems. All of the vectors must be defined with reference to the same point.

## Coordinate systems

Cartesian to cylindrical
$r=\sqrt{x^{2}+y^{2}} \quad \phi=\tan ^{-1}\left(\frac{y}{x}\right) \quad z=z$

$$
\phi=\tan ^{-1}\left(\frac{y}{x}\right)
$$

$$
z=z
$$

MATLAB command
[phi, r, z] = cart2pol(x, y, z)

Cartesian to spherical
$\rho=\sqrt{x^{2}+y^{2}+z^{2}} \quad \phi=\tan ^{-1}\left(\frac{y}{x}\right) \quad \theta=\tan ^{-1}\left(\frac{\sqrt{x^{2}+y^{2}}}{z}\right) \quad \begin{aligned} & {[p h i, p s i, \text { rho] }=} \\ & \operatorname{cart2\operatorname {sph}(x,y,z)}\end{aligned}$ note $\psi=\frac{\pi}{2}-\theta$

## Cylindrical to Cartesian

$$
\begin{array}{lll}
x=r \cos \phi & y=r \sin \phi & z=z \\
& {[x, y, z]=} \\
& \operatorname{pol} 2 \operatorname{cart}(p h i, r, z)
\end{array}
$$

## Spherical to Cartesian

$x=\rho \sin \theta \cos \phi \quad y=\rho \sin \theta \sin \phi \quad z=\rho \cos \theta$
[x, y, z] = sph2cart(phi,psi, rho) note $\psi=\frac{\pi}{2}-\theta$

Table 1-2. Summary of the transformation between coordinate systems.

## Integral relations for vectors

## 1-3. Integral relations for vectors

We will find that certain integrals involving vector quantities will be important in describing the material that is to be presented later in this text. These integrals will be useful initially in deriving vector operations and later in gaining an understanding of electromagnetic fields. For simplicity, the derivations will be presented in Cartesian coordinates. The fact that a field could depend upon its local position should not be too surprising to the reader since you may have noted the effects of a change in the gravitational field while watching the astronauts walking into a satellite when it is on the ground and then floating within the satellite as it soars above the earth.

The integrals on which we will focus are listed in Table 1-3.

| Line integral of a vector field $F$ along a prescribed path from the <br> location a to the location $b$. | $\int_{a}^{b} F \cdot d \mathbf{d l}$ |
| :--- | :--- |
| Surface integral of a vector field $F$ through a surface $\Delta \mathbf{s}$. | $\int_{\Delta s} \cdot d \mathbf{d}$ |
| Volume integral of a density $\rho_{v}$ over the volume $\Delta v$. | $\int_{\Delta v} \rho_{v} d v$ |

Table 1-3. Integrals of vector fields and densities.

## Integral relations for vectors

### 1.3.1. Line integral

The first integral that will be examined is a line integral. One possible application of this integral would be to compute the work $\Delta \mathrm{W}$ that would be required to push the cart with a force $\mathbf{F}$ from point $a$ to the point $b$ along a prescribed path as shown in Figure 1-14. This path could be dictated by metallic rails underneath the cart. The line integral is written as


Figure 1-14. The motion of the cart is constrained to move along the prescribed path from points a to b .

$$
\begin{equation*}
\int_{a}^{b} \mathrm{~F} \cdot \mathrm{dl} \tag{1.35}
\end{equation*}
$$

The differential length element $\mathbf{d l}$ can be written in the three orthogonal coordinate systems and these were included in Table 1-1. The limits $a$ and $b$ will determine the sign of the integration, $i e+$ or - . This integral states that no work will be expended in moving the cart if the direction of the force that is applied to the object is perpendicular to the path of the motion. If we were to push the cart completely around the path so it returned to the original point, we would call this a closed line integral and indicate it with a circle at the center of the integral sign as in (1.36)


Figure 1-15. Two of the many possible paths that the line integral could be evaluated.

To illustrate this point, let us calculate the work required to move the cart along the path 1 indicated in Figure 1-15 against a force field $\mathbf{F}$ where

$$
\begin{equation*}
F=3 x y \mathbf{u}_{\mathbf{x}}+4 x y \mathbf{u}_{\mathbf{y}} \tag{1.37}
\end{equation*}
$$

In this example, we are be able to specify a numerical value for one of the variables along each segment of the total path since each path is chosen to be parallel to an axis of the Cartesian coordinate system. This is not always possible and one of the variables may have to be specified in terms of the other variable or these dependent variables may be a function of another independent parameter, for example time. In this example, the work is found using thise line

## Integral relations for vectors

integral. This integral will consist of two terms since the path of integration is initially parallel to the $x$ axis and then parallel to the $y$ axis. In the first term, the incremental change in y is zero, hence $\mathrm{dy}=0$ and the differential length becomes $\mathbf{d} \mathbf{I}=\mathrm{dx} \mathbf{u}_{\mathbf{x}}$. Similarly, $\mathbf{d} \mathbf{I}=\mathrm{dy} \mathbf{u}_{\mathbf{y}}$ in the second integration since $\mathrm{dx}=0$. Therefore, we write

$$
\begin{gather*}
\Delta W=\int_{(1,1)}^{(4,2)} F \cdot d \mathbf{l l} \\
=\left.\int_{1}^{4}(3 x y+4 x y \quad)\right|_{y=1} \bullet d x \quad+\left.\int_{1}^{4}(3 x y \quad+4 x y \quad)\right|_{x=4} \bullet d y \\
=\left.3 \frac{x^{2}}{2}\right|_{1} ^{4}+\left.16 \frac{y^{2}}{2}\right|_{1} ^{2}=\frac{93}{2} \tag{1.38}
\end{gather*}
$$

We could return from point $b$ back to point a along the same path that we followed earlier or along a different path - say path 2 in Figure 1-15. We calculate the work along this new path. The differential length dl remains the same even though there is a change of direction in the integration. The limits of the integration will specify the final sign that will be encountered from the integration.

$$
\begin{gather*}
\Delta \mathrm{W}=\int_{(4,2)}^{(1,1)} F \cdot d \mathbf{l} \\
=\left.\int_{4}^{1}(3 x y+4 x y \quad)\right|_{y=2} \bullet d x \quad+\left.\int_{2}^{1}(3 x y \quad+4 x y \quad)\right|_{x=1} \bullet d y \\
=\left.6 \frac{x^{2}}{2}\right|_{4} ^{1}+\left.4 \frac{y^{2}}{2}\right|_{2} ^{1}=-\frac{102}{2} \tag{1.39}
\end{gather*}
$$

The total work that is required to move the cart completely around this closed path is not equal to zero! A closed path is defined as any path that returns us to the original point. In Figure 1-15, the cart could have been pushed completely around the loop. There may or may not be something enclosed

## Integral relations for vectors

within the closed path. In order to emphasize this point, think of walking completely around the perimeter of a green on a golf course. This would be an example of a closed path. The entity that would be enclosed within this path and rising above the ground would be the flag. If the closed line integral over all possible paths were equal to zero, then the vector $\mathbf{F}$ would belong to a class of fields that are called conservative fields. ${ }^{1}$ The example that we have just encountered would correspond to the class of nonconservative fields and in our method of exercise on the golf course, the flag would be flying. Both conservative and nonconservative fields will be encountered in electromagnetic fields.

As Electrical and Computer Engineers, you have already encountered this integral in the first course in electrical circuits without knowing it. If we sum up the voltage drops around a closed loop, we find that they are equal to zero. This is, of course, just one of Kirchhoff's laws. For the cases that we have encountered in that early circuit's course, this would be an example of a conservative field.

Example 1-9. Calculate the work $\Delta \mathrm{W}$ required to move the cart along the closed path if the force field is $\mathbf{F}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}$.

Answer. The closed line integral is given by the sum of four integrals.

$$
\begin{gathered}
\Delta W=\oint F \cdot=\int_{(1,1)}^{(4,1)}\left(3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}\right) \cdot d x \mathbf{u}_{\mathbf{x}}+\int_{(4,1)}^{(4,2)}\left(3 \mathbf{u}_{\mathrm{x}}+4 \mathbf{u}_{\mathbf{y}}\right) \bullet d y \mathbf{u}_{\mathbf{y}} \\
\quad+\int_{(4,2)}^{(1,2)}\left(3 \mathbf{u}_{\mathrm{x}}+4 \mathbf{u}_{\mathbf{y}}\right) \bullet d x \mathbf{u}_{\mathrm{x}}+\int_{(1,2)}^{(1,1)}\left(3 \mathbf{u}_{\mathrm{x}}+4 \mathbf{u}_{\mathbf{y}}\right) \bullet \mathrm{dy} \mathbf{u}_{\mathrm{y}} \\
=[3(3)]+[4(1)]+[3(-3)]+[4(-1)]=0
\end{gathered}
$$

[^0]
## Integral relations for vectors

In this case, the force field $\mathbf{F}$ is a conservative field.

Example 1-10. Calculate the work $\Delta \mathrm{W}$ required to move the cart along the circular path from point $A$ to point $B$ if the force field is $F=3 x y u_{\mathbf{x}}+4 x u_{\mathbf{y}}$.


Answer: The integral can be performed in Cartesian coordinates or in cylindrical coordinates. In Cartesian coordinates, we write

$$
\begin{aligned}
\mathbf{F} \cdot \mathbf{d l} & =\left(3 x y \mathbf{u}_{\mathbf{x}}+4 x \mathbf{u}_{\mathbf{y}}\right) \cdot\left(d x \mathbf{u}_{\mathbf{x}}+d y \mathbf{u}_{\mathbf{y}}\right) \\
& =3 x y d x+4 x d y
\end{aligned}
$$

The equation of a circle is $x^{2}+y^{2}=4^{2}$. Hence

$$
\begin{aligned}
& \int_{A}^{B} F \cdot d l=\int_{4}^{0} 3 x \sqrt{16-x^{2}} d x+\int_{0}^{4} 4 \sqrt{16-y^{2}} d y \\
& =-\left.\left(16-x^{2}\right)^{3 / 2}\right|_{4} ^{0}+\left.4\left(\frac{y}{2} \sqrt{16-y^{2}}+8 \sin ^{-1}\left(\frac{y}{4}\right)\right)\right|_{0} ^{4}=-64+16 \pi
\end{aligned}
$$

In cylindrical coordinates, we write

# Integral relations for vectors 

$$
\mathbf{F} \bullet \mathbf{d l}=\left(3 x y \mathbf{u}_{\mathbf{x}}+4 x \mathbf{u}_{\mathbf{y}}\right) \bullet\left(\mathrm{dr} \mathbf{u}_{\mathbf{r}}+\mathrm{rd} \mathrm{~d} \mathbf{u}_{\varphi}+\mathrm{dz} \mathbf{u}_{\mathbf{z}}\right)
$$

Since the integral is to be performed along the indicated path where only the angle $\phi$ is changing, we have $d r=0$ and $d z=0$. Also $r=4$. Therefore

$$
\mathbf{F} \bullet \mathbf{d l}=\left(3 x y \mathbf{u}_{\mathrm{x}}+4 \mathrm{x} \mathbf{u}_{\mathrm{y}}\right) \bullet\left(4 \mathrm{~d} \phi \mathbf{u}_{\phi}\right)
$$

From Table 1-2, we write the scalar products as

$$
\mathbf{u}_{\mathbf{x}} \bullet \mathbf{u}_{\phi}=-\sin \phi \text { and } \mathbf{u}_{\mathbf{y}} \bullet \mathbf{u}_{\phi}=\cos \phi
$$

The integral becomes

$$
\begin{aligned}
& \int_{A}^{B} \mathbf{F} \cdot d \mathrm{~d}=-64 \int_{0}^{\pi / 2}\left[\sin ^{2} \phi \cos \phi-\cos ^{2} \phi\right] \mathrm{d} \phi \\
= & -\left.64\left(3 \frac{\sin ^{3}}{3}-\frac{\phi}{2}-\frac{\sin 2 \phi}{4}\right)\right|_{0} ^{\pi / 2}=-64+16 \pi
\end{aligned}
$$

The results of the two calculations are identical as should be expected.

### 1.3.2. Surface integral

Another integral that will be encountered in the study of electromagnetic fields is the surface integral which is written as

$$
\begin{equation*}
\int_{\Delta s} F \cdot d s \tag{1.40}
\end{equation*}
$$

where $\mathbf{F}$ is the vector field and $\mathbf{d s}$ is the differential surface arrea. The differential surface areas for the three cooordinate systems are given in Table 1-1. This is shown in Figure 1-16 for an arbitrary surface. The vector $\mathbf{F}$ at this stage could represent a fluid flow. The loop, in some sense, monitors the flow of the field.

The differential surface element is a vector since a direction is associated with it. The vector direction of $\mathbf{d s}$ is in the direction normal to the surface and it is

## Integral relations for vectors

directed in the outward direction. For a closed surface, this direction is taken to be in the obvious direction. However, for a non-closed surface such as a plane or our golfing green, this direction has to be specified by the user since there is no obvious outward direction. Using the "right hand rule" convention, it is taken to be in the direction of the thumb if the fingers of the right hand follow the perimeter of the surface in a counterclockwise sense. A person standing above the green would observe a different direction than an individual buried beneath it.


Figure 1-16. A surface integral for an arbitrary surface. At the particular location of the loop, the component of $\mathbf{A}$ that is tangent to the loop does not pass through the loop. The scalar product $\mathbf{A} \cdot \mathbf{d s}$ eliminates this contribution.

The surface integral allows us to ascertain the amount of the vector field $\mathbf{A}$ that is passing through a surface element $\Delta \mathbf{S}$ which has a differential surface element ds. This vector field is frequently called a flux. A vector $\mathbf{A}$ that is directed in a direction such that it is confined to the surface will have the scalar product $\mathbf{A}$ - ds $=0, i$. e. the vector $\mathbf{A}$ does not pass through surface.

## Integral relations for vectors

If we integrated the vector field over the entire closed surface, the notation

$$
\begin{equation*}
\oint A \bullet d s \tag{1.41}
\end{equation*}
$$

would be employed. As we will see later, this closed surface integral can be either: greater than 0 , equal to 0 , or less than 0 depending on what is inside the closed volume. The limits of the integration would be defined by the edges of the surface area.


Figure 1-17. There are six differential surface vectors associated with a cube. The vectors are directed outwards.

For the cubical surface shown in Figure 1-17, there are six vectors ds associated with the six differential surfaces. The vectors

$$
d s=d x d y u_{\mathbf{z}} \text { and } d s=d x d y\left(-u_{\mathbf{z}}\right)
$$

## Integral relations for vectors

for the two surfaces that are perpendicular to the $z$ axis and are opposite from each other have the vector directions also in the opposite directions. The other four surfaces are similarly defined.

Example $1-11^{\text {MAtLAB }}$. Assume that a vector field $\mathbf{A}=\frac{A_{o}}{\rho^{2}}{ }_{\rho}$ exists in a region surrounding the origin of a spherical coordinate system. Find the value of the closed surface integral $\oint \mathbf{A} \bullet$ ds .

Answer: The closed surface integral is given by

$$
\oint \mathbf{A} \bullet \mathbf{d} \mathbf{s}=\int_{\phi=0}^{\phi=2 \pi} \int_{\theta=0}^{\theta=\pi}\left(\frac{\mathbf{A}_{\circ}}{\rho^{2}} \mathbf{u}_{\rho}\right) \bullet\left(\rho^{2} \sin \theta d \theta d \phi \mathbf{u}_{\rho}\right)=4 \pi \mathrm{~A}_{\circ}
$$



In this integral, we have used the differential surface area in spherical

## Integral relations for vectors

coordinates that has a unit vector $\mathbf{u}_{\boldsymbol{\rho}}$. If the vector $\mathbf{A}$ had any additional components directed in the $\mathbf{u}_{\theta}$ or $\mathbf{u}_{\phi}$ directions, their contribution to this surface integral would be zero since the scalar product of these terms will be equal to zero. The MATLAB command 'sphere' is used to produce the plot.

### 1.3.3. Volume integral

Finally, we will encounter various volume integrals of scalar quantities, such as a volume charge density $\rho_{\mathrm{V}}$. A typical integration would involve the computation of the total charge or mass in a volume if the volume charge or mass density were known. It is written as

$$
\begin{equation*}
Q=\int_{\Delta v} \rho_{v} d v \tag{1.42}
\end{equation*}
$$

The differential volumes for the three cooordinate systems are given in Table1-1. This will be demonstrated with an example.

Example 1-12. Find the volume of a cylinder that has a radius a and a length $L$.


Answer: The volume of a cylinder is calculated to be

$$
\Delta v=\int_{\Delta v} d v=\int_{z=0}^{z=L} \int_{\phi=0}^{\phi=2 \pi} \int_{r=0}^{r=a} r d r d \phi d z=\pi a^{2} L .
$$

## Differential relations for vectors

### 1.4. Differential relations for vectors

In addition to the integral relations for vectors, there are also differential operations that will be frequently encountered in our journey through electromagnetic theory. Each of these differential operators can be interpreted in terms of understandable physical phenomena. We will derive these vector operations in Cartesian coordinates. In addition, the operations in cylindrical and spherical coordinates will be included. The three vector operations are given in Table 1-4.

| gradient of a scalar field | $\nabla \mathrm{a}$ |
| :--- | :--- |
| divergence of a vector field | $\nabla \bullet \mathbf{A}$ |
| curl of a vector field | $\nabla \times \mathbf{A}$ |

Table 1-4. The three vector operations.

### 1.4.1. Gradient

It is possible to methodically measure scalar quantities such as a temperature at various locations in space. From this data, it is possible to connect the locations where the temperatures are the same. These equitemperature contours, when placed on a graph in a two dimensional plot are useful in interpreting various effects. This could include the magnitude and the direction where the most rapid changes occur or ascertaining the direction and the amount of heat that will flow. This would also be useful in planning a skiing or sunbathing vacation. The gradient of the scalar quantity which in this case is

## Differential relations for vectors

the temperature allows us compute the magnitude and the required direction that we should follow in order to find the maximum spatial rate of change of the scalar quantity in order to attain the desired conditions.

In Figure 1-18, we sketch two equipotential surfaces in space, the potential of one surface is arbitrarily chosen to have the value V and the potential of the other surface is $V+\Delta \mathrm{V}$. Point 1 is located on the first surface. The unit vector $\mathbf{u}_{\mathbf{n}}$ that is normal to this surface at $\mathrm{P}_{1}$ intersects the second surface at point $\mathrm{P}_{2}$. The magnitude of the distance between these two points is $\Delta \mathrm{n}$. Point $P_{3}$ is another point on the second surface and the vector distance between $P_{1}$ and $P_{3}$ is $\Delta l$. The unit vector from $P_{1}$ to $P_{3}$ is $\mathbf{u}$. The angle between the two vectors is $\zeta$. The distance $\Delta \mathrm{l}$ is greater than $\Delta \mathrm{n}$. Therefore,

$$
\frac{\Delta \mathrm{V}}{\Delta \mathrm{n}} \geq \frac{\Delta \mathrm{V}}{\Delta \mathrm{l}}
$$

This allows us to define two differential operations.
The gradient is defined as the vector that represents both the magnitude and the direction of the maximum spatial rate of increase of a scalar function. It depends upon the position where the gradient is to be evaluated and it may have different magnitudes and directions at different locations in space. In Figure 118, we write the gradient as

$$
\begin{equation*}
\operatorname{grad} \mathrm{V} \equiv \nabla \mathrm{~V}=\frac{\Delta \mathrm{V}}{\Delta \mathrm{n}} \mathbf{u}_{\mathrm{n}} \rightarrow \frac{\mathrm{dV}}{\mathrm{dn}} \mathbf{u}_{\mathrm{n}} \tag{1.43}
\end{equation*}
$$

In writing (1.43), we have used the common notation of replacing grad with $\nabla$.
In addition, we have assumed that the separation distance between the two surfaces is small and let $\Delta \mathrm{n} \rightarrow \mathrm{dn}$ which is indicative of a derivative.

## Differential relations for vectors



Figure 1-18. Equipotential surfaces in space.

The directional derivative is defined as a derivative in a particular direction. From Figure 1-18, this will be in the $\mathbf{u}$ direction and we write

$$
\frac{\Delta \mathrm{V}}{\Delta \mathrm{l}} \mathbf{u}_{1} \rightarrow \frac{\mathrm{dV}}{\mathrm{dl}} \mathbf{u}_{1}
$$

where we have again let $\Delta \mathrm{l} \rightarrow \mathrm{dl}$. Using the chain rule, we find that

$$
\begin{equation*}
\frac{\mathrm{dV}}{\mathrm{dl}}=\frac{\mathrm{dV}}{\mathrm{dn}} \frac{\mathrm{dn}}{\mathrm{dl}}=\frac{\mathrm{dV}}{\mathrm{dn}} \cos \varsigma=\frac{\mathrm{dV}}{\mathrm{dn}} \mathbf{u}_{\mathrm{n}} \cdot \mathbf{u}_{1}=\nabla V \bullet \mathbf{u}_{1} \tag{1.44}
\end{equation*}
$$

We realize that the directional derivative in the $\mathbf{u}$ direction is the projection of the gradient in that particular direction. Equation (1.44) can be written as

$$
\begin{equation*}
\mathrm{dV}=\nabla \mathrm{V} \bullet \mathrm{dl} \mathbf{u}_{\mathbf{l}}=\nabla \mathrm{V} \bullet \mathrm{dl} \tag{1.45}
\end{equation*}
$$

The gradient of the scalar function $\mathrm{a}(\mathrm{x}, \mathrm{y}, \mathrm{z})$ in Cartesian coordinates is

$$
\begin{equation*}
\nabla \mathrm{a}=\frac{\partial \mathrm{a}}{\partial \mathrm{x}} \mathbf{u}_{\mathrm{x}}+\frac{\partial \mathrm{a}}{\partial \mathbf{y}} \mathbf{u}_{\mathrm{y}}+\frac{\partial \mathrm{a}}{\partial \mathbf{z}} \mathbf{u}_{\mathbf{z}} \tag{1.46}
\end{equation*}
$$

The gradient of the scalar function $a(r, \phi, z)$ in cylindrical coordinates is

# Differential relations for vectors 

$$
\begin{equation*}
\nabla \mathbf{a}=\frac{\partial \mathrm{a}}{\partial \mathrm{r}} \mathbf{u}_{\mathrm{r}}+\frac{1}{\mathrm{r}} \frac{\partial \mathrm{a}}{\partial \phi} \mathbf{u}_{\phi}+\frac{\partial \mathrm{a}}{\partial \mathbf{z}} \mathbf{u}_{\mathbf{z}} \tag{1.47}
\end{equation*}
$$

The gradient of the scalar function $\mathrm{a}(\rho, \theta, \phi)$ in spherical coordinates is

$$
\begin{equation*}
\nabla \mathbf{a}=\frac{\partial \mathbf{a}}{\partial \rho} \mathbf{u}_{\rho}+\frac{1}{\rho} \frac{\partial \mathbf{a}}{\partial \theta} \mathbf{u}_{\theta}+\frac{1}{\rho \sin \theta} \frac{\partial \mathbf{a}}{\partial \phi} \mathbf{u}_{\phi} \tag{1.48}
\end{equation*}
$$

MATLAB also provides the capability of performing the gradient operation. In order to use this command, we must first calculate the contours that connect the points that have the same elevation. After this calculation, the gradient operation can then be performed.

Example $1-13^{\text {MATLAB }}$. Assume that there exists a surface that can be modeled with the equation $z=e^{-\left(x^{2}+y^{2}\right)}$ Calculate $\nabla z$ at the point $(x=0, y=0)$. In addition, use MATLAB to illustrate the profile and to calculate and plot this field.

Answer: $\nabla \mathrm{z}=-2 \mathrm{xe}^{-\left(\mathrm{x}^{2}+\mathrm{y}^{2}\right)} \mathbf{u}_{\mathbf{x}}-2 \mathrm{ye}^{-\left(\mathrm{x}^{2}+\mathrm{y}^{2}\right)} \mathbf{u}$. $\quad$ At the point $(\mathrm{x}=0, \mathrm{y}=0), \nabla \mathrm{z}=0$.
Using MATLAB, the function is illustrated in (a). The contours with the same value are connected together and the resulting field is indicated in (b). The length of the vectors and their orientation clearly indicate the distribution of the field in space. The commands 'contour' and 'quiver' have been employed in this calculation in order to create the figure. You should be aware that the accuracy of the calculation and the resulting figures are under the complete control of user.

## Differential relations for vectors


(b)


### 1.4.2. Divergence

The second vector derivative that should be reviewed is the divergence operation. The divergence operator is useful in determining if there is a source or a sink at a certain location in space in a region where a vector field exists. For electromagnetic fields, these sources and sinks will turn out to be positive and negative charges. This region could also be situated in a river where water would

## Differential relations for vectors

be flowing as shown in Figure 1-19. This could be a very porous box that contained either a drain or faucet that was connected with an invisible hose to the shore where the fluid could either be absorbed or from which it could be extracted.

The divergence of a vector that applies at a point is defined from the expression

$$
\begin{equation*}
\operatorname{div} \mathbf{A} \equiv \nabla \cdot \mathbf{A}=\lim _{\Delta \mathrm{v} \rightarrow 0} \frac{\oint \mathbf{A} \cdot \mathbf{d s}}{\Delta \mathrm{v}} \tag{1.49}
\end{equation*}
$$

The symbol $\oint \mathbf{A} \bullet$ ds indicates an integral over the entire closed surface that encloses the volume $\Delta v$. The point where the divergence is evaluated is within the volume $\Delta v$ and the surface for the closed surface integral is the surface that surrounds this volume. As we will see, the application of the " $\nabla \bullet$ " notation where $\nabla$. is the del operator will help us in remembering the terms that will actually appear in the operation.


Figure 1-19. Schematic of a source or a sink in a region where a fluid is flowing.

## Differential relations for vectors

Let us evaluate this surface integral over two of the six surfaces in full detail. If we can do this, then the integrations over the remaining surfaces are straightforward and need not be repeated. To evaluate the surface integrals at $x \pm \frac{\Delta x}{2}$ which have an approximate surface area of $(\Delta y \Delta z)$, we choose a point that is at the center of the volume and expand the $x$ component of the vector $\mathbf{A}=$ $\mathrm{A}_{\mathbf{X}} \mathbf{u}_{\mathbf{X}}$ about this point using a Taylor series. The Taylor series is an infinite series that allows us to approximate a function at an arbitrary point if the function and its derivatives are known at a certain point $\xi_{0}$. It is defined as

$$
\begin{equation*}
f(\xi)=f\left(\xi_{0}\right)+\left.\frac{\partial f(\xi)}{\partial \xi}\right|_{\xi_{0}}\left(\xi-\xi_{0}\right)+\left.\frac{1}{2} \frac{\partial^{2} f(\xi)}{\partial \xi^{2}}\right|_{\xi_{0}}\left(\xi-\xi_{0}\right)^{2}+\bullet \bullet \bullet \tag{1.50}
\end{equation*}
$$

where the function $f(\xi)$ is expanded about the point $\xi=\xi_{0}$. In our application of the Taylor series, we will keep only the first two terms since the terms containing $\left(\xi-\xi_{o}\right)$ to higher powers will be very small and can be neglected. Also, the function $f(\xi)$ will be sufficiently smooth so the higher order derivatives will not blow up at any point of interest to us.


Figure 1-20. Two surfaces that are located at $x$ and at $x+\Delta x$.

## Differential relations for vectors

Hence, we will just evaluate the integrals over these two surfaces. One surface is to the right of the point of interest and the other is to the left. Only the $x$ component of the vector $\mathbf{A}$ will pass-through both surfaces. Since the volume is assumed to be very small, $A_{x}$ can be assumed to be approximately constant when evaluated on either of these two surfaces. The surface integrals will just yield factors of $(\Delta y \Delta z)$. Note that the outward normal vector ds must be employed for both surfaces and this will introduce an additional minus sign in the unit vector directed toward decreasing values of the coordinate $x$. See Figure 1-20.

The surface integral becomes

$$
\begin{align*}
& \oint A_{x} \mathbf{u}_{x} \cdot d \mathbf{d} \approx\left(\left.A_{x} \mathbf{u}_{x}\right|_{x+\Delta x}\right) \cdot\left(\Delta y \Delta z \mathbf{u}_{x}\right)-\left(\left.A_{x} \mathbf{u}_{x}\right|_{x}\right) \bullet\left(-\Delta y \Delta z \mathbf{u}_{x}\right) \\
& \approx\left\{\left[\left(\mathrm{A}_{x}+\Delta x \frac{\partial A_{x}}{\partial x}\right) \mathbf{u}_{x}\right] \bullet\left(\Delta y \Delta z \mathbf{u}_{x}\right)+\left(A_{x} \mathbf{u}_{x}\right) \bullet\left(-\Delta y \Delta z \mathbf{u}_{x}\right)\right\}  \tag{1.51}\\
& =\left(\frac{\partial A_{x}}{\partial x}\right)(\Delta x \Delta y \Delta z)
\end{align*}
$$

If the values of $A_{X}$ that are evaluated at the two surfaces are equal in the limit of $\Delta x \rightarrow 0$, the differential term will be equal to zero since there is no change in the slope of the function $A_{X}$. This states that fluid that flows into one of the surfaces flows out of the other surface and none is lost nor created in the intermediate region.

The divergence of the x component of the vector $\mathbf{A}$ is evaluated from the definition (1.49) to be

## Differential relations for vectors

$$
\begin{equation*}
\nabla \bullet\left(A_{x} u_{x}\right)=\lim _{\Delta v \rightarrow 0} \frac{\left(\frac{\partial A_{x}}{\partial x}\right)(\Delta x \Delta y \Delta z)}{(\Delta x \Delta y \Delta z)} \rightarrow \frac{\partial A_{x}}{\partial x} \tag{1.52}
\end{equation*}
$$

where the volume of the small cube $\Delta V=\Delta x \Delta y \Delta z$. This procedure can be repeated for the other two components by summing up the contributions from the remaining four surfaces of the cube. Adding them up, we obtain

$$
\begin{equation*}
\nabla \bullet \mathbf{A}=\frac{\partial \mathbf{A}_{x}}{\partial \mathbf{x}}+\frac{\partial \mathbf{A}_{\mathrm{y}}}{\partial \mathbf{y}}+\frac{\partial \mathbf{A}_{z}}{\partial \mathbf{z}} \tag{1.53}
\end{equation*}
$$

In writing (1.52), we note that the terms for the divergence of $\mathbf{A}$ in Cartesian coordinates can be remembered by thinking of the $\nabla$ operator as a vector and taking the scalar product of this "vector" with the vector $\mathbf{A}$. The resulting vector operation is analogous to a scalar product in that a scalar quantity ensues. The term " $\nabla \cdot \bullet$ " is, however, an operator and not a vector: $\nabla \bullet \mathbf{A} \neq \mathbf{A} \bullet \nabla$ !

Example 1-14. Find the divergence of the vector $\mathbf{A}$ where the vector is the radius vector $\mathbf{r}$ or

$$
\mathbf{A}=x \mathbf{u}_{\mathrm{x}}+\mathrm{y} \mathbf{u}_{\mathrm{y}}+\mathrm{z} \mathbf{u}_{\mathrm{z}}
$$

Answer: Using (1.53), we write

$$
\nabla \bullet A=1+1+1=3 .
$$

We have found the divergence of a vector and we can suggest a physical interpretation of it. If the divergence of a vector $\mathbf{A}$ is equal to zero, then there are no sources to create the vector $\mathbf{A}$ nor sinks to absorb the vector $\mathbf{A}$ at that location since the Taylor series expanded terms that were evaluated at the opposite surfaces would be the same and would therefore cancel. If the divergence of a

## Differential relations for vectors

vector is equal to zero, then everything that enters the volume will leave the volume unscathed. If there were either a source or a sink at that point, then the divergence would be non-zero and the convention that would be followed is that a source will have a positive divergence and a sink will have a negative divergence. In the context of electromagnetic theory, this will determine whether positive or negative electric charges exist within a volume and the fact that magnetic charges or magnetic monopoles do not exist in nature.

We derived the divergence in Cartesian coordinates. The extension to cylindrical and spherical coordinates follows immediately. In the cylindrical coordinates, we write

$$
\begin{equation*}
\nabla \bullet \mathbf{A}=\frac{1}{r} \frac{\partial\left(r A_{r}\right)}{\partial r}+\frac{1}{r} \frac{\partial A_{\phi}}{\partial \phi}+\frac{\partial A_{z}}{\partial z} \tag{1.54}
\end{equation*}
$$

In spherical coordinates, we write

$$
\begin{equation*}
\nabla \bullet \mathbf{A}=\frac{1}{\rho^{2}} \frac{\partial\left(\rho^{2} A_{\rho}\right)}{\partial \rho}+\frac{1}{\rho \sin \theta} \frac{\partial\left(A_{\theta} \sin \theta\right)}{\partial \theta}+\frac{1}{\rho \sin \theta} \frac{\partial A_{\phi}}{\partial \phi} \tag{1.55}
\end{equation*}
$$

Example 1-15 ${ }^{\text {MATLAB }}$. Find divergence of the 2-D vector field $\mathbf{A}=\mathrm{e}^{-(\mathrm{r} / \alpha)^{2}} \mathbf{r}$, where $\mathbf{r}=x \mathbf{u}_{\mathrm{x}}+\mathrm{y} \mathbf{u}_{\mathrm{y}}$ and $\mathrm{r}^{2}=\mathrm{x}^{2}+\mathrm{y}^{2}$ by application of the MATLAB 'divergence' function ( $\alpha=$ const $)$.

Answer: The plot of the 2-D vector field by the 'quiver' function is presented in figure (a), while the contours of the scalar divergence field $D=\nabla \bullet A$ are presented in figure (b).

# Differential relations for vectors 



From the definition of the divergence (1.49), we can also find a useful relation between a volume integral of the divergence of a vector and the integral of the vector field integrated over the closed surface enclosing the volume $\Delta \mathrm{v}$. This can be obtained from the following "hand-waving" argument. From (1.49), we write that

$$
\begin{equation*}
\oint \mathbf{A} \bullet \mathbf{d} \mathbf{s} \approx(\nabla \bullet \mathbf{A}) \Delta \mathrm{v} \approx \int_{\Delta v}(\nabla \bullet \mathbf{A}) \mathrm{dv} \tag{1.56}
\end{equation*}
$$

In passing from the second term that appears in the definition of the divergence to the integral in the third term, we have let the volume $\Delta v$ be so small that the

## Differential relations for vectors

volume integral of the divergence of the vector is approximately equal to the product of the volume and this divergence operation.

Equating the two terms involving the integrals and replacing the approximately equal symbol with the equal sign, we obtain the divergence (Gauss's) theorem

$$
\begin{equation*}
\oint \mathbf{A} \bullet \mathbf{d s}=\int_{\Delta v}(\nabla \bullet \mathbf{A}) \mathrm{dv} \tag{1.57}
\end{equation*}
$$

This theorem will be very useful in later derivations of electromagnetic fields as it allows us to easily move between a volume integral and a closed surface integral in which the surface encloses the volume $\Delta \mathrm{v}$. It is also known as Gauss's theorem.

Example 1-16. Evaluate the divergence theorem for a vector $\mathbf{A}=x \mathbf{u}_{\mathbf{x}}$ within a unit cube centered about the origin.


Answer: The volume integral is given by

$$
\int_{\Delta v} \nabla \cdot A d v=\int_{x=-1 / 2}^{x=1 / 2} \int_{y=-1 / 2}^{y=1 / 2} \int_{z=-1 / 2}^{z=1 / 2} d x d y d z=1
$$

## Differential relations for vectors

The closed surface integral consists of two terms that are evaluated at $x=-1 / 2$ and at $x=+1 / 2$. We write
$\oint A \cdot d s$

$$
=\int_{y=-1 / 2}^{y=1 / 2} \int_{z=-1 / 2}^{z=1 / 2}\left(\left.x\right|_{x=-1 / 2}\right) \cdot\left(-d z d y \mathbf{u}_{x}\right)+\int_{y=-1 / 2}^{y=1 / 2} \int_{z=-1 / 2}^{z=1 / 2}\left(\left.x\right|_{x=+1 / 2}\right) \bullet\left(d z d y \mathbf{u}_{x}\right)=1
$$

### 1.4.3. Curl

The curl is a vector operation that can be used to state whether there is a rotation associated with a vector field. This is most easily visualized by considering the experiment of inserting a small paddle wheel in a flowing river as shown in Figure 1-21. If the paddle wheel is inserted in the center of the river, it will not rotate since the velocity of the water a small distance on either side of the center will be the same. However, if the paddle wheel were situated near the edge of the river, it would rotate since the velocity just at the edge will be less than in a region further from the edge. Note that the rotation will be in the opposite directions at the two edges of the river. The curl vector operation determines both the sense and the magnitude of the rotation. In German, this vector is given the name "rotation" that is abbreviated as rot (pronounced "rote").

## Differential relations for vectors



Figure 1-21. The paddle wheels inserted in a river will rotate if they are near the edges since the river velocity just at the edge is zero. The wheel at the center of the river will not rotate.

The curl of one component of a vector is defined from

$$
\begin{equation*}
\operatorname{curl} \mathbf{A}=\nabla \times \mathbf{A} \equiv \lim _{\Delta \mathrm{s} \rightarrow 0} \frac{\mathbf{u}_{\mathrm{n}} \oint \mathbf{A} \cdot \mathbf{d l}}{\Delta \mathrm{~s}} \tag{1.58}
\end{equation*}
$$

where we have incorporated the $\nabla$ notation. Note also that we have incorporated the notation " $x$ " in the curl operation. This should remind us that the curl has a sense of rotation associated with it if we think of $\nabla$ as a vector. In order to find the terms that will be used for the curl $\mathbf{A}$, we should examine Figure 1-22.

In order to calculate the $\mathbf{u}_{\mathbf{z}}$ component of curl $\mathbf{A}$, we have to first evaluate the line integral

$$
\begin{equation*}
\oint \mathbf{A} \cdot \mathbf{d} \mathbf{l}=\int_{1}^{2} \mathbf{A} \cdot \mathbf{d} \mathbf{l}+\int_{2}^{3} \mathbf{A} \cdot \mathbf{d} \mathbf{l}+\int_{3}^{4} \mathbf{A} \bullet \mathbf{d} \mathbf{l}+\int_{4}^{1} \mathbf{A} \cdot \mathbf{d} \mathbf{l} \tag{1.59}
\end{equation*}
$$

## Differential relations for vectors



Figure 1-22. Orientation of the loop required to find the $\mathbf{u}_{\mathbf{z}}$ component of curl $\mathbf{A}$ at the point $(x, y)$.

The right-hand convention will be followed for the curl operation in that a counterclockwise rotation determines the $+\mathbf{u}_{\mathbf{z}}$ component. A clockwise path would yield a - $\mathbf{u}_{\mathbf{z}}$ contribution. Let us make the assumption that the square is sufficiently small such that the vector field $\mathbf{A}$ is approximately constant along each segment of the square. With this assumption, we can factor these terms out of the integrals. Hence, the closed line integral becomes approximately

$$
\begin{align*}
\oint \mathbf{A} \bullet d \mathbf{l} \approx & \mathbf{A}(x, y) \cdot \int_{1}^{2} d \mathbf{l}+\mathbf{A}(x+\Delta x, y) \cdot \int_{2}^{3} d \mathbf{l}  \tag{1.60}\\
& +\mathbf{A}(x, y+\Delta y) \bullet \int_{3}^{4} d \mathbf{l}+\mathbf{A}(x, y) \bullet \int_{4}^{1} d \mathbf{l}
\end{align*}
$$

The integrals are equal to the vector distances that specify the square

$$
\begin{equation*}
\Delta \mathbf{x}=\int_{1}^{2} \mathrm{dl} ; \quad \Delta \mathbf{y}=\int_{2}^{3} \mathrm{dl} ; \quad-\Delta \mathbf{x}=\int_{3}^{4} \mathrm{dl} ; \quad \text { and }-\Delta \mathbf{y}=\int_{4}^{1} \mathrm{dl} \tag{1.61}
\end{equation*}
$$

## Differential relations for vectors

If $\Delta x$ and $\Delta y$ are sufficiently small, we can expand a function that is to be evaluated on the lines, at the point $(x, y)$ using a two dimensional generalization of the Taylor series that was defined in (1.50). Keeping only the lowest order terms, we obtain

$$
\begin{equation*}
f(x, y) \approx f\left(x_{0}, y_{0}\right)+\left.\frac{\partial f(x, y)}{\partial x}\right|_{x_{0}, y_{0}} \Delta x+\left.\frac{\partial f(x, y)}{\partial y}\right|_{x_{0}, y_{0}} \Delta y+\bullet \bullet \tag{1.62}
\end{equation*}
$$

Applying this to each term in (1.59), we obtain

$$
\begin{align*}
\oint \mathbf{A} \bullet \mathbf{d} l & \approx\left[A_{x}(x, y)\right] \Delta x+\left[A_{y}(x, y)+\left.\frac{\partial A_{y}(x, y)}{\partial x}\right|_{x, y} \Delta x\right] \Delta y \\
& -\left[A_{x}(x, y)+\left.\frac{\partial A_{x}(x, y)}{\partial y}\right|_{x, y} \Delta y\right] \Delta x-\left[A_{y}(x, y)\right] \Delta y  \tag{1.63}\\
& =\left(\frac{\partial A_{y}(x, y)}{\partial x}-\frac{\partial A_{x}(x, y)}{\partial y}\right) \Delta x \Delta y
\end{align*}
$$

The area of the square is $\Delta x \Delta y$. Using the definition given in (1.58), we have found the component of curl $\mathbf{A}$ that is oriented in the $\mathbf{u}_{\mathbf{z}}$ direction. This can be extended to three dimensions. If we collect all of the terms for all three coordinates in Cartesian coordinates, the general expression for the curl $\mathbf{A}$ results. As with the vector product, we can easily remember the terms from the determinant.

$$
\operatorname{curl} \mathbf{A}=\left|\begin{array}{ccc}
\mathbf{u}_{x} & \mathbf{u}_{\mathbf{y}} & \mathbf{u}_{z} \\
\frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\
\mathrm{~A}_{\mathrm{x}} & \mathrm{~A}_{\mathrm{y}} & \mathrm{~A}_{\mathrm{z}}
\end{array}\right|
$$

We derived the curl in Cartesian coordinates. The extension to cylindrical and spherical coordinates follows. In cylindrical coordinates, we write

## Differential relations for vectors

$$
\operatorname{curl} \mathbf{A}=\frac{1}{r}\left|\begin{array}{ccc}
\mathbf{u}_{r} & r \mathbf{u}_{\phi} & \mathbf{u}_{z}  \tag{1.65}\\
\frac{\partial}{\partial r} & \frac{\partial}{\partial \phi} & \frac{\partial}{\partial z} \\
A_{r} & r A_{\phi} & A_{z}
\end{array}\right|
$$

In spherical coordinates, we write

$$
\operatorname{curl} \mathbf{A}=\frac{1}{\rho^{2} \sin \theta}\left|\begin{array}{ccc}
\mathbf{u}_{\rho} & \rho \mathbf{u}_{\theta} & \rho \sin \theta \mathbf{u}_{\phi}  \tag{1.66}\\
\frac{\partial}{\partial \rho} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \phi} \\
A_{\rho} & \rho A_{\theta} & \rho \sin \theta A_{\phi}
\end{array}\right|
$$

Example 1-18 MATLAB. Find curl of the 2-D vector field $\mathbf{A}=\mathrm{e}^{-(r / \alpha)^{2}} \boldsymbol{\omega} \times \boldsymbol{r}$, where $\mathbf{r}=\mathbf{x} \mathbf{u}_{\mathbf{x}}+\mathrm{y} \mathbf{u}_{\mathbf{y}}, \mathrm{r}^{2}=\mathrm{x}^{2}+\mathrm{y}^{2}$ and $\boldsymbol{\omega}=\omega \mathbf{u}_{\mathbf{z}}$ by application of the 'curl' function.

Answer: The curl operator here produces a vector $\mathbf{C}$ with only z-component $\mathbf{C}=\mathrm{C}_{\mathbf{z}} \mathbf{u}_{\mathbf{z}}$. The plot of the 2-D vector field by the 'quiver' function is presented in figure (a), while the contours of the z-component $C_{z}$ of the vector curl field $\mathbf{C}=\nabla \times \mathbf{A}$ are presented in figure (b).

# Differential relations for vectors 



From the definition of the curl of a vector given in (1.58), we can obtain Stokes' theorem that relates a closed line integral to a surface integral. Following the same "hand waving" procedure that we used to derive the divergence theorem, we write

$$
\begin{equation*}
\oint \mathbf{A} \cdot \mathbf{d l} \approx(\nabla \times \mathbf{A}) \Delta \mathbf{s} \approx \int_{\Delta s} \nabla \times \mathbf{A} \cdot \mathbf{d s} \tag{1.67}
\end{equation*}
$$

This is finally written with the same caveats that we employed previously as

$$
\begin{equation*}
\oint \mathbf{A} \cdot \mathbf{d l}=\int_{\Delta \mathbf{s}} \nabla \times \mathbf{A} \cdot \mathbf{d s} \tag{1.68}
\end{equation*}
$$

This is called Stokes's theorem.

## Differential relations for vectors

In this case, the closed line integral is along the perimeter of the surface over which the surface integral is performed. Recall that in the right-hand convention that we are employing, the fingers of the right hand follow the path of the line integral dl and the thumb points in the direction of the vector surface element ds.

Example 1-19. Given a vector field $\mathbf{A}=x y \mathbf{u}_{\mathbf{x}}-2 x \mathbf{u}_{\mathbf{y}}$, verify Stokes's theorem over one-quarter of a circle whose radius is 3 .


Answer: We must first calculate $\nabla \times \mathbf{A}$ and the surface integral.

$$
\nabla \times \mathbf{A}=\left|\begin{array}{ccc}
\mathbf{u}_{\mathbf{x}} & \mathbf{u}_{\mathbf{y}} & \mathbf{u}_{\mathbf{z}} \\
\frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\
x y & -2 x & 0
\end{array}\right|=-(2+x) \mathbf{u}_{\mathbf{z}}
$$

The surface integral becomes

$$
\begin{aligned}
\int_{\Delta s} \nabla \times \mathbf{A} \bullet \mathbf{d s} & =\int_{y=0}^{y=3} \int_{x=0}^{x=\sqrt{9-y^{2}}}(\nabla \times \mathbf{A}) \bullet d x d y u_{z} \\
& =\int_{y=0}^{y=3} \int_{x=0}^{x=\sqrt{9-y^{2}}}-(2+x) \mathbf{u}_{z} \bullet \mathrm{dxdyu}_{z} \\
& =-\int_{y=0}^{y=3}\left[2 \sqrt{9-y^{2}}+\frac{\left(\sqrt{9-y^{2}}\right)^{2}}{2}\right] d y
\end{aligned}
$$

The integral

## Differential relations for vectors

$$
\int_{y=0}^{y=3} 2 \sqrt{9-y^{2}} d y
$$

requires the substitution $y=3 \sin \varphi$ and the identity $\cos ^{2} \varphi=\frac{1}{2}(1+\cos 2 \varphi)$ to transform it to the integral

$$
9 \int_{0}^{\pi / 2}(1+\cos 2 \varphi) d \varphi
$$

which can be evaluated. Therefore, we obtain

$$
-\int_{y=0}^{y=3}\left[2 \sqrt{9-y^{2}}+\frac{\left(\sqrt{9-y^{2}}\right)^{2}}{2}\right] d y=-9 \frac{\pi}{2}-\frac{27}{2}+\frac{27}{6}=-9\left(1+\frac{\pi}{2}\right)
$$

The closed line integral will involve three terms and using the right-hand convention for the integration sequence, we write

$$
\oint \mathbf{A} \bullet \mathbf{d} \mathbf{l}=\int_{\mathrm{x}=0, \mathrm{y}=0}^{\mathrm{x}=3, \mathrm{y}=0} \mathbf{A} \bullet \mathrm{dx} \mathbf{u}_{\mathrm{x}}+\int_{\mathrm{arc}} \mathbf{A} \bullet \mathbf{d} \mathbf{l}+\int_{\mathrm{x}=0, \mathrm{y}=3}^{\mathrm{x}=0, \mathrm{y}=0} \mathbf{A} \bullet \mathrm{dy} \mathbf{u}_{\mathrm{y}}
$$

The two integrals that are along the two axes will contribute zero to the closed integral since the vector $\mathbf{A}=0$ on the axis. The remaining integral becomes

$$
\begin{aligned}
\oint \mathbf{A} \bullet \mathbf{d} \mathbf{l}=\int_{a \operatorname{arc}} \mathbf{A} \bullet \mathbf{d} \mathbf{l}= & \int_{\operatorname{arc}}\left(x y u_{x}-2 x \mathbf{u}_{y}\right) \bullet\left(d x u_{x}+d y u_{y}+d z u_{z}\right) \\
& =\int(x y d x-2 x d y) \\
& =\int_{3}^{0} x \sqrt{9-x^{2}} d x-2 \int_{0}^{3} \sqrt{9-y^{2}} d y=-9\left(1+\frac{\pi}{2}\right)
\end{aligned}
$$

As we should expect, the two answers are the same.

## Differential relations for vectors

### 1.4.4. Repeated vector operations

Having defined the vector operations of the gradient, the divergence and the curl; we may be curious about a repeated vector operation such as the divergence of the curl of a vector. There are several methods of approaching this topic. A straight forward rigorous approach would be to mechanically perform the vector operations and find the answer. This approach is left for the problems. A second approach that will be followed here is based on intuitive arguments. Hopefully the meaning of the various vector operations will become more clear as the discussion is presented.

The three vector operations that will be examined are:

$$
\begin{gather*}
\nabla \bullet \nabla \times \mathbf{A}=0  \tag{1.69}\\
\nabla \times \nabla \mathrm{a}=0  \tag{1.70}\\
\nabla \bullet \nabla \mathrm{a}=\nabla^{2} \mathrm{a} \tag{1.71}
\end{gather*}
$$

Other vector identities exist and a list of useful vector identities are also included in Appendix 1.

The first equation (1.69) can be interpreted in the following terms. The curl operation gives the magnitude and the sense of the rotation of a vector that remains within a prescribed region. The quantity that this vector represents neither enters nor leaves the region. The divergence operation monitors the entry or departure of a vector field from a region due to a local source or sink within it. Therefore a vector $\mathbf{A}$ that has a nonzero curl just rotates and does not enter nor leave the region. One could think of a boat in a rotating whirlpool that cannot be paddled away from its impending doom as an example of this identity.

## Differential relations for vectors

The second equation (1.70) is understood from the following argument. The gradient of a scalar function expresses the direction and the magnitude that an inertialess ball would take as it rolls down a mountain along the path of least resistance. This path would not be expected to close upon itself. The curl, however, would require that the ball return to the same point on the mountain to indicate rotation. This point could be back at, say back at the top. This would be after it had begun its meander down the mountain under its own volition. Hence we can conclude that $(1.70)$ is correct since it could not return unless there were some new laws of nature such as anti-gravitational forces.

The third equation (1.71) is a definition of the Laplacian operation. It states that there is a vector field $\nabla$ a where a is some scalar quantity. The divergence of this vector field will determine whether a source or a sink exists at that point. In Cartesian coordinates, the Laplacian operator is written as

$$
\begin{equation*}
\nabla^{2} \mathbf{a}=\frac{\partial^{2} a}{\partial \mathbf{x}^{2}}+\frac{\partial^{2} a}{\partial \mathbf{y}^{2}}+\frac{\partial^{2} a}{\partial \mathbf{z}^{2}} \tag{1.72}
\end{equation*}
$$

As will be seen later, this operation will be important for finding the potential distribution caused by a charge distribution.

In cylindrical coordinates, the Laplacian operator is

$$
\begin{equation*}
\nabla^{2} \mathbf{a}=\frac{1}{r} \frac{\partial\left(r \frac{\partial \mathbf{a}}{\partial r}\right)}{\partial \mathbf{r}}+\frac{1}{r^{2}} \frac{\partial^{2} a}{\partial \phi^{2}}+\frac{\partial^{2} a}{\partial z^{2}} \tag{1.73}
\end{equation*}
$$

In spherical coordinates, the Laplacian operator is

$$
\begin{equation*}
\nabla^{2} \mathbf{a}=\frac{1}{\rho^{2}} \frac{\partial\left(\rho^{2} \frac{\partial \mathbf{a}}{\partial \rho}\right)}{\partial \rho}+\frac{1}{\rho^{2} \sin \theta} \frac{\partial\left(\sin \theta \frac{\partial \mathbf{a}}{\partial \theta}\right)}{\partial \theta}+\frac{1}{\rho^{2} \sin ^{2} \theta} \frac{\partial^{2} \mathbf{a}}{\partial \phi^{2}} \tag{1.74}
\end{equation*}
$$

## Phasors

## 1-5. Phasors

We frequently encounter a time-harmonic signal that represents a real physical quantity, say a voltage in a circuit, that varies sinusoidally in time. This can be expressed as

$$
\begin{equation*}
v(t)=V_{0} \cos (\omega t+\phi) \tag{1.75}
\end{equation*}
$$

In (1.75), $\mathrm{V}_{\mathrm{O}}$ is the amplitude of the signal, $\omega$ is the angular frequency $[\omega=2 \pi f$ where f is the frequency in Hertz], and $\phi$ is the phase of the signal. This is shown in Figure 1-23.

An alternative method of expressing (1.75) is to write it as

$$
\begin{equation*}
v(t)=\operatorname{Re}\left[V \mathrm{e}^{\mathrm{j} \omega \mathrm{t}}\right] \tag{1.76}
\end{equation*}
$$

where Re [ ] implies that we are to take the real part of the terms that are confined within the braces [ ] and $\mathrm{j}=\sqrt{-1}$. For (1.76) to represent (1.75), this implies that

$$
\begin{equation*}
V=V_{0} e^{j \phi} \tag{1.77}
\end{equation*}
$$



Figure 1-23. A time-harmonic signal $\mathrm{v}(\mathrm{t})$.

## Phasors

This entity $V$ contains that contains information concerning both the amplitude and the relative phase of the signal and it is given the acronym phasor. The phasor is independent of time although it may depend on a spatial variable $r$. In using phasor notation, we frequently omit the symbol $\operatorname{Re}\left[\mathrm{e}^{\mathrm{j} \omega \mathrm{t}}\right.$ ] for simplicity.

The time derivative and time integral of the signal $v(t)$ are written as

$$
\begin{equation*}
\frac{d v}{d t}=\operatorname{Re}\left[V e^{j \omega t}\right] \text { and } \int v d t^{\prime}=\operatorname{Re}\left[\frac{1}{j \omega} V e^{j \omega t}\right] \tag{1.78}
\end{equation*}
$$

This follows from the operations on the equation (1.75)

$$
\begin{align*}
& \frac{d v}{d t}=-\omega V_{o} \sin (\omega t+\phi)=\omega V_{o} \cos \left(\omega t+\phi+\frac{\pi}{2}\right)=\operatorname{Re}\left[j \omega V e^{j \omega t}\right]  \tag{1.79}\\
& \int v d t^{\prime}=\frac{1}{\omega} V_{o} \sin (\omega t+\phi)=\frac{1}{\omega} V_{o} \cos \left(\omega t+\phi-\frac{\pi}{2}\right)=\operatorname{Re}\left[\frac{1}{j \omega} V e^{j \omega t}\right] \tag{1.80}
\end{align*}
$$

Certain rules apply in the application of phasors. Phasor quantities can be added or subtracted only if they have the same frequency. The use of phasor notation implies signals with the same frequency. The product of two signals, say

$$
[A \cos (\omega t)] \times[B \cos (\omega t+\phi)]
$$

cannot be represented in phasor notation since it is nonlinear. A "trig identity" for this product leads to two signals, one with a frequency $2 \omega$, and one with zero frequency. For the linear world, phasors are a panacea.

Example 1-19. Express $\vee(t)=10 \cos \left(120 \pi t+60^{\circ}\right) V$ in phasor notation.
Answer: This is written as

$$
V=10 \mathrm{e}^{\mathrm{j}}(\pi / 3)=5+\mathrm{j} 8.7 \text { volts }
$$

## Phasors

Example 1-20 ${ }^{\text {MATLAB }}$. Express $v(t)=3 \cos \omega t-4 \sin \omega t$ as $A \cos (\omega t+\phi)$. Use phasor notation. Plot the function.

Answer: Let us use cos $\omega$ t as the reference and add the two phasors.

$$
\begin{gathered}
3 \cos \omega t \Rightarrow 3 \\
-4 \sin \omega t=-4 \cos \left(\omega t-\frac{\pi}{2}\right) \Rightarrow-4 e^{-j \pi / 2}=j 4
\end{gathered}
$$

Therefore, we write

$$
V=3+j 4=5 e^{j \tan ^{-1}(4 / 3)}=5 e^{j 53^{\circ}}
$$

The real part of the product of this phasor and ej $\omega$ t yields

$$
3 \cos \omega t-4 \sin \omega t=\operatorname{Re}\left(5 \mathrm{e}^{\mathrm{j}\left(\omega t+53^{\circ}\right)}\right)=5 \cos \left(\omega t+53^{\circ}\right)
$$

The plot of the time-harmonic function is shown below.

## Phasors



Example 1-21. Express the loop equation for an RLC equation in phasor notation. The applied voltage is $v(t)=V \cos \omega t$ and the loop equation is

$$
\mathrm{L} \frac{\mathrm{di}}{\mathrm{dt}}+\mathrm{Ri}+\frac{1}{\mathrm{C}} \int \mathrm{idt}=\mathrm{v}
$$

## Phasors



Answer: The current $i$ is written as

$$
i(t)=I \cos (\omega t+\phi)
$$

since we have chosen the cosine as the reference. Hence the differential equation is written as

$$
I\left[-\omega L \sin (\omega t+\phi)+R \cos (\omega t+\phi)+\frac{1}{\omega C} \sin (\omega t+\phi)\right]=V \cos \omega t
$$

Mathematical manipulations beyond this point that would be required in order to determine I and $\phi$ are tedious at best and difficult at worst.

In phasor notation, we write

$$
v(t)=V \cos \omega t=\operatorname{Re}\left[V e^{j 0} e^{j \omega t}\right]=\operatorname{Re}\left[V e^{j \omega t}\right]
$$

and

$$
i(t)=\operatorname{Re}\left[I e^{j \phi} e^{j \omega t}\right]=\operatorname{Re}\left[/ e^{j \omega t}\right]
$$

The terms $V$ and $I$ are phasors. They contain both the amplitude and the phase information that has been isolated from the time dependence $t$. The derivative term and the integral term that appear in the loop equation are replaced with the terms from (1.78). The loop equation in phasor notation is

## Phasors

$$
\left[R+j\left(\omega L-\frac{1}{\omega C}\right)\right] I=V
$$

The time factor $\mathrm{e}^{\mathrm{j} \omega \mathrm{t}}$, which is common to all terms in the equation, has been canceled. This algebraic equation can easily be solved for the phasor current $I$ in terms of the phasor voltage $V$. The expression in the brackets is called the "impedance" Z. The current i is obtained by multiplying $/$ by $\mathrm{e}^{\mathrm{j} \omega \mathrm{t}}$ and taking the real part of the product.

## Conclusion

### 1.6. Conclusion

The electromagnetic fields that will be described in the rest of this book will make use of MATLAB, vectors, and the various integral and differential operations that have been given in this chapter. In addition, the two theorems that allowed us to convert a surface integral into a closed line integral (Stokes's theorem) or a volume integral into a closed surface integral (divergence theorem) will be very important in gaining an appreciation of these fields. They will also be employed in later derivations to actually develop the basic laws of electromagnetic theory from the equations that arise from experimental observations. We have initially interpreted several applications of vectors using fluids. This was done since most of us have gone to the beach at one time or have seen mechanical systems. At this stage, electric and magnetic fields may seem rather opaque.

## Problems 1

### 1.7. Problems

1. Find the vector that connects the two opposite corners of a cube whose volume is $a^{3}$. One corner of the cube is located at the center of a Cartesian coordinate system. Write this vector also in terms of the magnitude times a unit vector.

2. Find the vector $\mathbf{B}$ from the origin to the opposite corner that lies in the $x y$ plane.
$3^{\text {MATLAB }}$. Given two vectors $\mathbf{A}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+5 \mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=-5 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}-3 \mathbf{u}_{\mathbf{z}}$, find $\mathbf{C}$ $=\mathbf{A}+\mathbf{B}$ and $\mathbf{D}=\mathbf{A}-\mathbf{B}$. In addition, carefully illustrate these vectors using MATLAB.
$4^{\text {MATLAB }}$. Using the vectors defined in 3 , evaluate $\mathbf{A} \cdot \mathbf{B}$ and $\mathbf{A} \times \mathbf{B}$. Check your answer with MATLAB.
$5^{\mathrm{MATLAB}}$. Given two vectors $\mathbf{A}=\mathbf{u}_{\mathbf{x}}+\mathbf{u}_{\mathbf{y}}+\mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=\mathbf{2} \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+6 \mathbf{u}_{\mathbf{z}}$, find
$\mathbf{C}=\mathbf{A}+\mathbf{B}$ and $\mathbf{D}=\mathbf{A}-\mathbf{B}$. In addition, carefully illustrate these vectors using MATLAB.
$6^{\text {MATLAB }}$. Using the vectors defined in 5, evaluate $\mathbf{A} \cdot \mathbf{B}$ and $\mathbf{B} \times \mathbf{A}$. Check your answer with MATLAB.
$7^{\text {MATLAB }}$. Using MATLAB, write a program to convert degrees $C$ to degrees F. Plot the results.

## Problems 1

$8^{\text {MATLAB }}$. Using MATLAB, write a program to convert a yard stick to a meter stick. Plot the results.
$9^{\text {MATLAB }}$. Using MATLAB, plot $y=e^{-x}$ on a linear and a semilog graph.
$10^{\text {MATLAB }}$. Using MATLAB, plot two cycles of $y=\cos (x)$ on a linear and a polar graph.
$11^{\text {MATLAB }}$. Using MATLAB, carefully plot a vector field defined by $\mathbf{A}=y^{2} \mathbf{u}_{\mathbf{x}}-x \mathbf{u}_{\mathbf{y}}$ in the region $-2<x<+2,-2<y<+2$. The length of the vectors in the field should be proportional to the field at that point. Find the magnitude of this vector at the point $(3,2)$.
$12^{\text {MATLAB }}$. Using MATLAB, carefully plot a vector field defined by in the $\mathbf{A}=\sin \mathrm{x} \mathbf{u}_{\mathbf{x}}-\sin \mathrm{y} \mathbf{u}_{\mathbf{y}}$ in the region $0<\mathrm{x}<\pi, 0<\mathrm{y}<\pi$. The length of the vectors in the field should be proportional to the field at that point. Find the magnitude of this vector at the point $\left(\frac{\pi}{2}, \frac{\pi}{2}\right)$.
$13^{\text {MATLAB }}$. Find the scalar product of the two vectors defined by
$\mathbf{A}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+5 \mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=-5 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}-3 \mathbf{u}_{\mathbf{z}}$. Determine the angle between these two vectors. Check your answer using MATLAB.
$14^{\text {MATLAB }}$. Find the scalar product of the two vectors defined by
$\mathbf{A}=\mathbf{u}_{\mathbf{x}}+\mathbf{u}_{\mathbf{y}}+\mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=\mathbf{2} \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+6 \mathbf{u}_{\mathbf{z}}$. Determine the angle between these two vectors. Check your answer using MATLAB.
$15^{\mathrm{MATLAB}}$. Find the projecthetion of a vector from the origin to a point defined at $(1,2,3)$ on the vector from the origin to a point defined at $(2,1,6)$. Find the angle between these two vectors. Check your answer using MATLAB.

## Problems 1

$16^{\text {MATLAB }}$. Find the vector product of the two vectors defined by
$\mathbf{A}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+5 \mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=-5 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}-3 \mathbf{u}_{\mathbf{z}}$. Check your answer using MATLAB.
$17^{\text {MATLAB }}$. Find the vector product of the two vectors defined by
$\mathbf{A}=\mathbf{u}_{\mathbf{x}}+\mathbf{u}_{\mathbf{y}}+\mathbf{u}_{\mathbf{z}}$ and $\mathbf{B}=\mathbf{2} \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+6 \mathbf{u}_{\mathbf{z}}$. Check your answer using MATLAB.
$18^{\text {MATLAB }}$. Express the vector field $\mathbf{A}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+5 \mathbf{u}_{\mathbf{z}}$ in cylindrical coordinates.
Check your answer using MATLAB.
$19^{\text {MATLAB }}$. Express the vector $\mathbf{B}=3 \mathbf{u}_{\mathbf{r}}+4 \mathbf{u}_{\boldsymbol{\phi}}+5 \mathbf{u}_{\mathbf{z}}$ that is in cylindrical coordinates into Cartesian coordinates. Check your answer using MATLAB.
$20^{\text {MATLAB }}$. Express the vector field $\mathbf{A}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+5 \mathbf{u}_{\mathbf{z}}$ in spherical coordinates. Check your answer using MATLAB.
$21^{\text {MATLAB }}$. Express the vector $\mathbf{B}=3 \mathbf{u}_{\rho}+4 \mathbf{u}_{\theta}+5 \mathbf{u}_{\phi}$ that is in spherical coordinates into Cartesian coordinates. Check your answer using MATLAB.
$22^{\text {MATLAB }}$. For the vectors $\mathbf{A}=\mathbf{u}_{\mathbf{x}}+\mathbf{u}_{\mathbf{y}}+\mathbf{u}_{\mathbf{z}}, \mathbf{B}=2 \mathbf{u}_{\mathbf{x}}+2 \mathbf{u}_{\mathbf{y}}+2 \mathbf{u}_{\mathbf{z}}$, and
$\mathbf{C}=3 \mathbf{u}_{\mathbf{x}}+3 \mathbf{u}_{\mathbf{y}}+3 \mathbf{u}_{\mathbf{z}}$; show that $\mathbf{A} \mathbf{x}(\mathbf{B} \times \mathbf{C})=\mathbf{B}(\mathbf{A} \cdot \mathbf{C})-\mathbf{C}(\mathbf{A} \cdot \mathbf{B})$. Check your answer using MATLAB.
$23^{\text {MATLAB }}$. For the vectors $\mathbf{A}=\mathbf{u}_{\mathbf{x}}+3 \mathbf{u}_{\mathbf{y}}+5 \mathbf{u}_{\mathbf{z}}, \mathbf{B}=2 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+6 \mathbf{u}_{\mathbf{z}}$, and $\mathbf{C}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}+5 \mathbf{u}_{\mathbf{z}}$; show that $\mathbf{A} \mathbf{x}(\mathbf{B} \times \mathbf{C})=\mathbf{B}(\mathbf{A} \cdot \mathbf{C})-\mathbf{C}(\mathbf{A} \cdot \mathbf{B})$. Check your answer using MATLAB.
$24^{\text {MATLAB }}$. Find the area of the parallelogram using vector notation. Compare your result with that found graphically.

$25^{\text {MATLAB }}$. Show that we can use the vector definitions $\mathbf{A} \cdot \mathbf{B}=0$ and $\mathbf{A} \times \mathbf{B}=0$ to express that two vectors are perpendicular and parallel to each other respectively.

$$
26^{\text {MATLAB }} . \text { Let } \mathbf{A}=-2 \mathbf{u}_{\mathbf{x}}+3 \mathbf{u}_{\mathbf{y}}+4 \mathbf{u}_{\mathbf{z}} ; \mathbf{B}=7 \mathbf{u}_{\mathbf{x}}+1 \mathbf{u}_{\mathbf{y}}+2 \mathbf{u}_{\mathbf{z}} ; \text { and }
$$

$$
\mathbf{C}=-1 \mathbf{u}_{\mathbf{x}}+2 \mathbf{u}_{\mathbf{y}}+4 \mathbf{u}_{\mathbf{z}} . \text { Find (a) } \mathbf{A} \times \mathbf{B} .(\mathrm{b})(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} .(\mathrm{c}) \mathbf{A} \cdot(\mathbf{B} \times \mathbf{C})
$$

27. Calculate the work required to move a mass $m$ against a force field $\mathbf{F}=5 \mathbf{u}_{\mathbf{x}}+7 \mathbf{u}_{\mathbf{y}}$ along the indicated direct path from point a to point $b$.

28. Calculate the work required to move a mass m against a force field $\mathbf{F}=\mathbf{y} \mathbf{u}_{\mathbf{x}}+\mathbf{x} \mathbf{u}_{\mathbf{y}}$ along the path abc and along the path adc. Is this field conservative?

29. Calculate the work required to move a mass $m$ against a force field $F=r u_{r}+r \mathbf{u}_{\phi}$ along the path abc.

30. Calculate the work required to move a mass $m$ against a force field $\mathbf{F}=\mathrm{r} \phi \mathbf{u}_{\phi}$ if the radius of the circle is a and $0 \leq \phi \leq 2 \pi$.

31. Calculate the closed surface integral $\oint \mathbf{A} \bullet$ ds if $\mathbf{A}=x \mathbf{u}_{\mathbf{x}}+\mathrm{y} \mathbf{u}_{\mathbf{y}}$ and the surface is the surface of a cube. Then apply divergence theorem to solve the same integral.

32. Evaluate the closed surface integral of the vector
$\mathbf{A}=x y z \mathbf{u}_{\mathbf{x}}+x y z \mathbf{u}_{\mathbf{y}}+x y z \mathbf{u}_{\mathbf{z}}$ over the cubical surface shown in Problem 31.
33 Evaluate the closed surface integral of the vector $\mathbf{A}=3 \mathbf{u}_{\rho}$ over the spherical surface that has a radius a .

## Problems 1

34. Find the surface area of a cylindrical surface by setting up and evaluating the integral $\oint \mathbf{A} \bullet d s$ where
$\mathbf{A}=1 \mathbf{u}_{\mathrm{r}}+2 \mathbf{u}_{\mathbf{z}}$.

35. A hill can be modeled with the equation $H=10-x^{2}-3 y^{2}$ where $H$ is the elevation of the hill. Find the path that a frictionless ball would take in order that it experienced the greatest change of elevation in the shortest change of horizontal position. Assume that the motion of the ball is unconstrained.
36. Find the gradient of the function $H=x^{2} y z$ and also the directional derivative of H specified by the unit vector $\mathbf{u}=\mathrm{a}\left(\mathbf{u}_{\mathbf{x}}+\mathbf{u}_{\mathbf{y}}+\mathbf{u}_{\mathbf{z}}\right)$ where a is a constant at the point (1, 2, 3). State the value for the constant a.
37. By direct differentiation show that $\nabla\left(\frac{1}{\rho}\right)=-\nabla^{\prime}\left(\frac{1}{\rho}\right)$ where

$$
\rho=\sqrt{\left(x-x^{\prime}\right)^{2}+\left(y-y^{\prime}\right)^{2}+\left(z-z^{\prime}\right)^{2}}
$$

and $\nabla^{\prime}$ denotes differentiation with respect to the variables $x^{\prime}, y^{\prime}$, and $z^{\prime}$.
38. Calculate the divergence of the vector

$$
\mathbf{A}=x^{3} y \sin (\pi z) \mathbf{u}_{\mathbf{x}}+x y \sin (\pi z) \mathbf{u}_{\mathbf{y}}+x^{2} y^{2} z^{2} \mathbf{u}_{\mathbf{z}}
$$

at the point $(1,1,1)$.
39. Show that the divergence theorem is valid for a cube located at the center of a Cartesian coordinate system for a vector
$A=x u_{x}+2 u_{y}$.

40. Show that the divergence theorem is valid for a sphere of radius a located at the center of a coordinate system for a vector $\mathbf{A}=\rho \mathbf{u}_{\rho}$.
41. The water that flows in a channel with sides at $x=0$ and $x=a$ has a velocity distribution $\mathbf{v}(x, z)=\left[\left(\frac{a}{2}\right)^{2}-\left(x-\frac{a}{2}\right)^{2}\right] z^{2} \mathbf{u}_{y}$. The bottom of the river is at $z=$ 0. A small paddle wheel with its axis parallel to the $z$ axis is inserted into the channel and is free to rotate. Find the relative rates of rotation at the points

$$
\left(x=\frac{a}{4}, z=1\right),\left(x=\frac{a}{2}, z=1\right), \text { and }\left(x=\frac{3 a}{4}, z=1\right) .
$$

Will the paddle wheel rotate if its axis is parallel to the $x$ axis or the $y$ axis?

## Problems 1


42. Evaluate the line integral of the vector function $\mathbf{A}=x \mathbf{u}_{\mathbf{x}}+\mathrm{x}^{2} \mathbf{y} \mathbf{u}_{\mathbf{y}}+\mathrm{xyz} \mathbf{u}_{\mathbf{z}}$ around the square contour $C$. Integrate $\nabla \mathbf{x} \mathbf{A}$ over the surface bounded by C . Show that this example satisfies Stokes's theorem.

43. Show that $\nabla \times \mathbf{A}=0$ if $\mathbf{A}=\left(\frac{1}{r}\right) \mathbf{u}_{\mathbf{r}}$ in cylindrical coordinates.
44. Show that $\nabla \times \mathbf{A}=0$ if $\mathbf{A}=\rho^{2} \mathbf{u}_{\rho}$ in spherical coordinates.
45. In rectangular coordinates, verify that $\nabla \cdot \nabla \times \mathbf{A}=0$ where

$$
A=x^{2} y^{2} z^{2}\left[u_{x}+u_{y}+u_{z}\right]
$$

by carrying out the detailed differentiations.
46. In rectangular coordinates, verify that $\nabla \mathbf{x} \nabla \mathrm{a}=0$ where

## Problems 1

$$
a=3 x^{2} y+4 z^{2} x
$$

by carrying out the detailed differentiations.
47. In rectangular coordinates, verify that $\nabla \mathbf{x}(\mathrm{aA})=(\nabla \mathrm{a}) \mathbf{x} \mathbf{A}+\mathrm{a} \nabla \mathbf{x} \mathbf{A}$ where $\mathbf{A}=x y z\left[\mathbf{u}_{\mathbf{x}}+\mathbf{u}_{\mathbf{y}}+\mathbf{u}_{\mathbf{z}}\right]$ and $\mathrm{a}=3 \mathrm{xy}+4 \mathrm{zx}$ by carrying out the detailed differentiations.
48. In rectangular coordinates, verify that $\nabla \cdot(\mathrm{aA})=\mathbf{A} \cdot \nabla \mathrm{a}+\mathrm{a} \nabla \cdot \mathbf{A}$ where $A=x y z\left[u_{\mathbf{x}}+\mathbf{u}_{\mathbf{y}}+\mathbf{u}_{\mathbf{z}}\right]$ and $\mathrm{a}=3 x y+4 z x$ by carrying out the detailed differentiations.
49. By direct differentiation, show that $\nabla^{2}\left(\frac{1}{\rho}\right)=0$ at all points where $\rho \neq 0$
where $\rho=\sqrt{\left(x-x^{\prime}\right)^{2}+\left(y-y^{\prime}\right)^{2}+\left(z-z^{\prime}\right)^{2}}$.
50. Express the signal $v(t)=100 \cos \left(120 \pi t-45^{\circ}\right)$ in phasor notation.
51. Given a phasor $V=10+\mathrm{j} 5$. Find the sinusoidal signal this represents if the frequency $=60 \mathrm{~Hz}$.
52. Find the phasor notation of $v(t)=\cos \left[120 \pi t-60^{\circ}\right]-\sin [120 \pi t]$. $53^{\text {MATLAB }}$. Find and plot the current $i(t)$ in the circuit if $v(t)=10 \cos (120 \pi t)$.


## Problems 1

$54^{\mathrm{MATLAB}}$. Repeat problem 53 with $v(\mathrm{t})=10 \cos \left(120 \pi \mathrm{t}+45^{\circ}\right)$.

## Electric field

## 2. Static electric and magnetic fields

The important properties of time independent static electric and magnetic fields will be reviewed in this chapter. This will include a review of the force between two stationary charges, the concept of an electric field, the electric energy, and several procedures that are used to calculate the electric field. It will also include a review of the magnetic effects that will be encountered If the charge is in motion with a constant velocity that can be described as being a current. A magnetic field will follow from this current and there will be a magnetic force between two current elements. Several procedures to calculate the magnetic field will be described. Various electrical circuit elements will be defined at this stage. In addition, the effects that are introduced due to the inclusion of various materials into the space will be summarized. Finally, the boundary conditions that one encounters when different materials are in juxtaposition will be described. This will be important in relating the electric and magnetic fields in one material to those in the other material.

### 2.1. Coulomb's law

The phenomenon that is the basis for the study of static electromagnetic fields has been known since ancient times. As early as 600 BC, Thales of Miletus is given credit for being the first to note that the rubbing of amber against a cloth caused the amber rod to attract light objects to itself. The use of amber by this ancient Greek experimenter has had a dramatic influence on the discipline that we now call Electrical Engineering ${ }^{1}$ and on the subject of electromagnetic fields. Indeed, these ancient observers have given us a word that is still in everyday use - the Greek word for amber is élektron. Materials other than amber also exhibit

[^1]
## Electric field

this process of electrification and we today can observe the same effect when we rub a glass rod on a silk cloth or take off a wool sweater too quickly. Both the rod and the cloth will attract small pieces of "fluff and stuff".

A new entity in nature that we will call a charge has been uncovered in these experiments. It is as fundamental a quantity as those that we have already encountered: mass, length, and time. The charge can be either positive or negative. We will reserve the symbols $Q, M, L$, and $T$ for the quantities charge, mass, length, and time respectively. We will call these "fundamental units" in this book. This nomenclature will be useful in "checking the dimensions" of an equation that we may have derived. We cannot claim that a lengthy derivation is correct if we end up with "apples" on one side of the equality sign and "oranges" on the other side.

As we pass through middle of the eighteenth century, we find the names of many who have contributed to our understanding of this physical phenomenon: Benjamin Franklin, Joseph Priestley, Michael Faraday, Henry Cavendish, and Augustine Coulomb. Who has not heard the story of Franklin flying his kite in a thunderstorm? Through a series of experiments, they uncovered the fact that there would be a force of attraction for unlike charges and a force of repulsion for like charges. This force is somewhat similar to the force of gravity. Both forces have the same geometrical dependence on the separation distance $\rho$ between the two objects. Both forces also depend on the product of the magnitudes of the charges $Q_{j}$ or of the masses $m_{j}$. After much experimentation, they concluded that the magnitude of this electrical force could be written as

$$
\begin{equation*}
F \propto \frac{Q_{1} Q_{2}}{R^{2}} \tag{2.1}
\end{equation*}
$$

## Electric field

An ancient experimental system is depicted in Figure 2-1. A charge is induced on the glass rod by rubbing it on a cloth and touching it simultaneously to the two pith balls. In part (a), the "electrified" glass rod is just used to transfer the electrical charge to two stationary pith balls that are initially touching. The sign of the charge deposited on each ball is the same and it was found that both balls experienced a force that caused the balls to separate. Let us assume that the experimenter performing the experiment shown in Figure 2-1 could accurately measure the following quantities.
a) The magnitude and the sign of the charges.
b) The magnitude and the vector direction of the force
c) The distance between the two pith balls.
d) The masses of the two pith balls.

It was found in this experiment that the magnitude of a charge was an integer multiple of the magnitude of the charge of an electron. If charges with different signs had been individually placed on these two balls, the balls would not separate but would be attracted to each other.

From the experimental results, this force that we will call an "electrostatic force" or a "Coulomb force" can be written in MKS or SI units as

$$
\begin{equation*}
\mathbf{F}=\frac{\mathrm{Q}_{1} \mathrm{Q}_{2}}{4 \pi \varepsilon_{0} \mathrm{R}^{2}} \mathbf{u}_{\mathrm{R}} \quad(\mathrm{~N}) \tag{2.2}
\end{equation*}
$$

The unit of force is measured in terms of Newtons $(\mathrm{N})$, the unit of charge is measured in terms of Coulombs (C) and the distance between the charges is measured in terms of meters (m). The charge of an electron is

$$
\begin{equation*}
Q_{e} \approx-1.602 \times 10-19 C \tag{2.3}
\end{equation*}
$$

and the charge of a singly charged proton is

$$
\begin{equation*}
Q_{p} \approx+1.602 \times 10^{-19} \mathrm{C} \tag{2.4}
\end{equation*}
$$

## Electric field

There are both positive and negative charges that exist in nature. All of the charges have values that are integer multiples of these values.

In a series of reports to the French Academy of Science from 1785 to 1791, Charles-Augustin de Coulomb described the results of a series of experiments involving a carefully constructed torsion balance and which he verified the equation of the electrostatic force (2.2). He also performed a series of experiments using small magnets and verified that the magnetic force between like and unlike magnetic poles would be either repulsive or attractive with the same geometrical dependence.


Figure 2-1. An experiment designed to demonstrate the electrostatic force.
(a) Two uncharged pith balls are hanging from a vertical rod. The length of the string is $L$. The only force is in the vertical direction and it is due to gravity.
(b) The pith balls repel each other due to the Coulomb force in the horizontal direction to a distance b after the same charge is distributed on each ball.

## Electric field

One Coulomb of charge is a very large amount of charge. For example, if we were to collect all the charge that is created by a single lightning stroke, we would only collect a total of approximately 10 to 20 Coulombs. Considering the violent nature of such a stroke, this does not appear to be a very big number. There are, however, a very large number N of charged particles in a lightning stroke. For example, we compute

$$
\begin{equation*}
N=\frac{10}{1.602 \times 10^{-19}} \approx 6 \times 10^{19} \text { particles } \tag{2.5}
\end{equation*}
$$

The constant $\varepsilon_{0}$ in equation (2.2) is called the permittivity of free space. In SI units, it has the numerical value

$$
\begin{equation*}
\varepsilon_{0}=8.854 \times 10^{-12} \approx \frac{1}{36 \pi} \times 10^{-9} \frac{\text { farads }}{\text { meter }} \equiv \frac{\mathrm{F}}{\mathrm{~m}} \tag{2.6}
\end{equation*}
$$

The approximate number $\frac{1}{36 \pi} \times 10^{-9}$ is a convenient number to remember although it is not the exact number. With this approximate number, we frequently will be able to obtain a numerical result without having to resort to a calculator in a computation. As we will see later when we discuss electromagnetic waves, this approximate number is useful and will yield the well known numerical value of $3 \times 10^{8}$ for the velocity of light instead of the more accurate value that is slightly less than this.

We now call equation (2.2): Coulomb's law. At this time, we can still experience the phenomenon of electrification every day when we stroll across a shag rug and receive a shock upon touching someone else or if we comb our hair and later pick up pieces of paper with the comb. But don't attempt to stand outside in a lightning storm! If you did this latter experiment, you might find that your hair would "stand on its end" due to your body conducting like charge from the ground to the tips of the strands of your hair. The charge in one strand would repel the charge in an adjacent strand causing the wild effects suggested in this expression.

## Electric field

Example 2-1. Using Coulomb's law, determine the fundamental units of the permittivity of free space.
Answer: The force equals the mass $x$ acceleration or $M \frac{L}{T^{2}}$. Therefore, we write from Coulomb's law (2.2) that the units of $\varepsilon_{0}$ can be obtained

$$
\mathrm{F}=\frac{\mathrm{Q}_{1} \mathrm{Q}_{2}}{4 \pi \varepsilon_{0} \mathrm{R}^{2}} \mathbf{u}_{\mathrm{R}} \Rightarrow \mathrm{M} \frac{\mathrm{~L}}{\mathrm{~T}^{2}}=\frac{\mathrm{Q}^{2}}{\mathrm{~L}^{2}} \frac{1}{\varepsilon_{0}} \Rightarrow \varepsilon_{o}=\frac{\mathrm{Q}^{2} \mathrm{~T}^{2}}{\mathrm{ML}^{3}}
$$

Remember, 4 and $\pi$ are just numbers that do not have any units associated with them. The unit vector is also dimensionless.

The vector direction of the force acting on charge 1 due to a charge 2 is directed along the line between the two charges. We indicate this direction with the unit vector $\mathbf{u}_{\mathbf{R}}$. The direction of the Coulomb force is either to cause the two charges to attract each other or to repel each other. Attraction or repulsion depends on the relative signs of the two charges. Two charges that have the same sign, either positive or negative, will repel each other. Two charges that have the opposite signs will attract each other. In the experimental system depicted in Figure 2-1, the charge with the same sign that was originally on the rod would be subsequently transferred to both pith balls. Therefore, the two balls will repel each other.

The Coulomb force equation is fundamental in the explanation of electromagnetic fields. It contains the new physical quantity - the charge - that makes electromagnetic theory unique. Charges that are in motion create currents that, in turn, create magnetic fields. The theory of relativity allows us to derive other laws of electromagnetic theory from the Coulomb force equation. Hence, this one simple equation will bear much fruit in our later discussion. Rather than invoke

## Electric field

such an esoteric subject as relativity and considerable mathematical chicanery, we will examine these topics and follow in the footsteps of the giants who have walked ahead of us and who will guide us through the dark forest of seemingly unrelated experimental observations.

Example 2-2. Find the magnitude of the Coulomb force that exists between an electron and a proton in a hydrogen atom. Compare the Coulomb force and the gravitational force between the two particles. The two particles are separated approximately by 1 Ångstrom $=1 \AA=1 \times 10^{-10}$ meters.

Answer: The magnitude of the Coulomb force is computed from (2.2)

$$
\mathrm{F}_{\text {Coulomb }}=\frac{\mathrm{Q}^{2}}{4 \pi \varepsilon_{0} \mathrm{R}^{2}} \approx \frac{\left(1.602 \times 10^{-19}\right)^{2}}{4 \pi\left(\frac{1}{36 \pi} \times 10^{-9}\right)\left(10^{-10}\right)^{2}} \approx 2.3 \times 10^{-8} \mathrm{~N}
$$

The gravitational constant $\kappa=6.67 \times 10^{-11}$

$$
\begin{aligned}
F_{\text {gravitational }}=\kappa \frac{m_{\text {electrons }} M_{\text {proton }}}{R^{2}} & =\left(6.67 \times 10^{-11}\right) \frac{\left(9.11 \times 10^{-31}\right)\left(1836 \times 9.11 \times 10^{-31}\right)}{\left(10^{-10}\right)^{2}} \\
= & 1.02 \times 10^{-47} \mathrm{~N} .
\end{aligned}
$$

The ratio of the two forces is

$$
\frac{F_{\text {Coulomb }}}{F_{\text {gravitational }}}=2.27 \times 10^{39} .
$$

### 2.2. Electric field

When we step on the scale to determine our weight, we do not carry out a detailed calculation involving the mass of the Earth, our mass, and the distance between the center of the Earth and our center of mass. We just assume that there is a gravitational field where we are standing and have the scale calibrated to indicate the multiplication of our mass times the gravitational field. The gravitational field is a vector quantity that is pointing toward the center of the Earth. We also encounter the same phenomenon with electric fields that will be described below.

The electric field $E$ caused by a charge $Q$ is a vector quantity that has the definition

$$
\begin{equation*}
E \equiv \frac{F}{q} \quad\left(\frac{N}{C}\right) \tag{2.7}
\end{equation*}
$$

where $\mathbf{F}$ is the Coulomb force between the two charges $Q$ and $q$. The standard symbol for the electric field is $\mathbf{E}$. Since we're looking at static electric fields that did not depend upon time, this electric field is frequently called an electrostatic field. The electric field in this region due to the charge $Q$ is therefore written as

$$
\begin{equation*}
\mathbf{E}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{R}^{2}} \mathbf{u}_{\mathrm{R}} \tag{2.8}
\end{equation*}
$$

We will see later that the units for the electric field are also (volts / meter)=(V/m).

(a)

(b)

Figure 2-2. Electric fields: (a) emanate from a positive charge. (b) terminate on a negative charge.

## Electric field

Electric fields from a positive and a negative charge are depicted in Figure 2-2. We note that the direction of the electric field depends on the sign of the charge. Gravitational fields only cause two masses to be attracted to each other. In analogy with the relation between the gravitational field and the gravitational force, we can find the force on a charged particle that is brought into a region containing an electric field E by just multiplying the electric field by the charge q , that is $F=q E$. This will be a particularly useful concept when we study the ballistic motion of charged particles in a region containing an electric field, say in a cathode ray tube. Knowing the spatial distribution of the electric field in a particular region will have important practical consequences.

It is worth pointing out a conceptual point at this stage. We might be inclined to compare equations (2.2) and (2.8) and suggest that the electric field could be defined in terms of a derivative. We remember the definition of the derivative to be

$$
\begin{equation*}
E \Rightarrow \frac{d F}{d q} \equiv \lim _{\Delta q \rightarrow 0} \frac{F(q+\Delta q)-F(q)}{\Delta q} \tag{2.9}
\end{equation*}
$$

where the definition of the derivative is also explicitly stated in (2.9). The operation of performing a differentiation certainly appears to give the correct mathematical result. However, this differentiation will not be correct since the smallest charge that has been observed in nature is that of an electron or a singly charged proton whose charge has a magnitude of $1.602 \times 10^{-19} \mathrm{C}$. The limiting procedure that is required in the definition of the derivative cannot be performed since the charge does not continuously and smoothly approach the value of zero. This is because charges have a charge that is an integer multiple of the value of the electron or the proton. Quantum electrodynamics has suggested that entities with a charge magnitude that is equal to $(1 / 3)$ of this value

## Electric field

exist but the mathematical limiting procedure in the derivative still fails. We must use the definition $E \equiv \frac{F}{q}$ for the electric field.

Example 2-3. Calculate the electric field at a distance of $1 \mu \mathrm{~m}\left(1 \mu \mathrm{~m}=10^{-6} \mathrm{~m}\right)$. from a singly ionized proton. Calculate the Coulomb force on a second electron at this location.

Answer: From (2.8), we compute the electric field to be

$$
\mathrm{E}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{R}^{2}} \mathbf{u}_{\mathrm{R}}=\frac{1.602 \times 10^{-19}}{4 \pi\left(\frac{1}{36 \pi} \times 10^{-9}\right)\left(10^{-6}\right)^{2}} \mathbf{u}_{\mathrm{R}}=1440 \mathbf{u}_{\mathrm{R}}
$$

The Coulomb force is

$$
\mathbf{F}=q E=\left(1.602 \times 10^{-19}\right) \times(1440) \mathbf{u}_{\mathrm{R}}=2.3 \times 10^{-16} \mathbf{u}_{\mathrm{k}}
$$

Note that we have employed the approximate value for the permittivity of free space.

## Superposition principles

### 2.3. Superposition principles

If we had more than one charge and each charge were at a different location in a vacuum, the total electric field in the space external to the location of these charges would be the vector summation of the electric field originating from each individual charge. The vacuum is a linear media. In fact, a vacuum has the greatest number of linear properties that can be found in any media. The introduction of an additional material into the vacuum may cause the region to become nonlinear. The principles of superposition apply in a vacuum. The only caveat that we will encounter will be that we must be careful to apply vector superposition principles and just not scalar superposition principles. Both the magnitude and the direction of the individual electric fields from each charge must be included in the addition. For N separate charges in the region of interest, this vector summation can be written as

$$
\begin{equation*}
\mathbf{E}=\mathbf{E}_{1}+\mathbf{E}_{\mathbf{2}}+\mathbf{E}_{3}+\cdots=\sum_{n=1}^{N} \mathbf{E}_{\mathbf{n}} \tag{2.10}
\end{equation*}
$$

The electric field created by each individual charge add them up as vectors and this is illustrated in Figure 2-3 for the particular case of two charges $Q_{1}$ and $Q_{2}$. Remember that we have to include the correct sign of the charge in the vector addition operation. For the case depicted in Figure 2-3, the total electric field $E_{T}$ is calculated from the vector summation of the two individual components. Both charges are assumed to be positive, therefore the electric fields will be directed away from the charges. The total electric field intensity is given by

$$
\begin{equation*}
\mathbf{E}_{\mathrm{T}}=\mathbf{E}_{1}+\mathbf{E}_{\mathbf{2}}=\frac{\mathrm{Q}_{1}}{4 \pi \varepsilon_{0} \mathrm{R}_{1}^{2}} \mathbf{u}_{\mathrm{R}_{1}}+\frac{\mathrm{Q}_{2}}{4 \pi \varepsilon_{0} \mathrm{R}_{2}^{2}} \mathbf{u}_{\mathrm{R}_{2}} \tag{2.11}
\end{equation*}
$$

where $\mathbf{u}_{R_{j}}$ indicates the unit vector associated with each individual charge $Q_{j}$ to the point where the electric field is to be computed. The distance between the

## Superposition principles

charge and this point is given by $\mathrm{R}_{\mathrm{j}}$. This vector addition is best illustrated with two examples.


Figure 2-3. The total electric field is the vector sum of individual components.

Example 2-4. Two charges $Q_{1}=+4 C$ and $Q_{2}=-2 C$ are located at the points indicated on the graph. The units of the graph are in meters. Find the electric field at the origin $(0,0)$ of the coordinate system.


Answer: The electric field $\mathbf{E}$ is computed from (2.11)

$$
\mathbf{E}_{\mathbf{T}}=\mathbf{E}_{1}+\mathbf{E}_{2}=\frac{\mathrm{Q}_{1}}{4 \pi \varepsilon_{0} \mathrm{R}_{1}^{2}} \mathbf{u}_{\mathrm{R}_{1}}+\frac{\mathrm{Q}_{2}}{4 \pi \varepsilon_{0} \mathrm{R}_{2}^{2}} \mathbf{u}_{\mathrm{R}_{2}}
$$

The vector direction of the electric field is directed from the charge \#1 to the charge \#2 at the origin. Using the numerical values specified for the charges and the distances as determined from the graph, we write

## Superposition principles

$\mathbf{E}_{\mathrm{T}}=\mathbf{E}_{1}+\mathbf{E}_{2}=\frac{4}{4 \pi \varepsilon_{0} 3^{2}} \mathbf{u}_{\mathrm{x}}+\frac{2}{4 \pi \varepsilon_{0} 2^{2}} \mathbf{u}_{\mathrm{x}}=\frac{1}{4 \pi \varepsilon_{0}}\left(\frac{4}{9}+\frac{2}{4}\right) \mathbf{u}_{\mathrm{x}}=\frac{1}{4 \pi \varepsilon_{0}}\left(\frac{34}{36}\right) \mathbf{u}_{\times} \quad \mathrm{V} / \mathrm{m}$.
Note that the electric fields from the two charges add up at the origin.

Example 2-5. Three charges $\mathrm{Q}_{1}=+1 \mathrm{C}, \mathrm{Q}_{2}=+2 \mathrm{C}$, and $\mathrm{Q}_{3}=3 \mathrm{C}$ are placed at the indicated points on the graph. Find the electric field at the point $P$.


Answer: The electric field at the point $P$ is computed from a linear superposition on the individual electric field components due to the individual charges. We write

$$
\begin{aligned}
\mathbf{E}(3,4) & =\mathbf{E}_{1}(3,4)+\mathbf{E}_{2}(3,4)+\mathbf{E}_{3}(3,4)=\frac{\mathrm{Q}_{1}}{4 \pi \varepsilon_{0} R_{1}^{2}} \mathbf{u}_{\mathbf{R}_{1}}+\frac{\mathrm{Q}_{2}}{4 \pi \varepsilon_{0} R_{2}^{2}} \mathbf{u}_{\mathbf{R}_{2}}+\frac{\mathbf{Q}_{3}}{4 \pi \varepsilon_{0} R_{3}^{2}} \mathbf{u}_{\mathbf{R}_{3}} \\
& =\frac{1}{4 \pi \varepsilon_{0} 5^{2}} \frac{3 \mathbf{u}_{\mathrm{x}}+4 \mathbf{u}_{\mathrm{y}}}{5}+\frac{2}{4 \pi \varepsilon_{0} 4^{2}} \mathbf{u}_{\mathrm{y}}+\frac{3}{4 \pi \varepsilon_{0} 3^{2}} \mathbf{u}_{x} \\
& =\frac{1}{4 \pi \varepsilon_{0}}\left[\left(\frac{3}{125}+\frac{3}{9}\right) \mathbf{u}_{\mathrm{x}}+\left(\frac{4}{125}+\frac{2}{16}\right) \mathbf{u}_{\mathrm{y}}\right] \mathrm{V} / \mathrm{m}
\end{aligned}
$$

The further evaluation of this electric field is straightforward.

## Superposition principles

Up to this point, our discussion of electrostatic fields has assumed that it was possible to calculate the electric field by merely summing the vector contributions from each individual charge. In theory, this is the correct procedure that should always be followed. However, in many cases that one encounters in practice, we would quickly run out of steam in following such a procedure when describing realistic situations where the number of charged particles in a confined volume may be of the order of a power of ten, say fifteen or twenty. Numerical tools would soon be required to perform this summation.

If we can make certain assumptions concerning the distribution of the charges in a region and realize that an integration of the distributed charges over the region follows directly from a summation if we let a certain parameter become extremely small, then it is possible to obtain analytical solutions for a particular problem. We will include some of these solutions in the following discussion.

The assumption that we will employ is that if a total charge $\Delta \mathrm{Q}$ is distributed within a volume $\Delta \mathrm{v}$ and we take the limit as this volume $\Delta \mathrm{v} \rightarrow 0$, then we can define a volume charge density $\rho_{\mathrm{v}}$ as

$$
\begin{equation*}
\rho_{v}=\frac{\Delta Q}{\Delta v}\left(\frac{C}{m^{3}}\right) \tag{2.12}
\end{equation*}
$$

This charge density may be inhomogeneous such that it depends on the local position $\mathbf{r}$, and we write this charge density as $\rho_{\mathrm{V}}=\rho_{\mathrm{V}}(\mathbf{r})$. There may also be a uniform volume charge distribution.

If the charge is distributed on a surface whose area is $\Delta \mathrm{s}$ and it is independent of the distance normal to the surface, then we can define this as a surface charge density $\rho_{\mathrm{S}}$ as

$$
\begin{equation*}
\rho_{\mathrm{s}}=\frac{\Delta \mathrm{Q}}{\Delta \mathrm{~s}}\left(\frac{\mathrm{C}}{\mathrm{~m}^{2}}\right) \tag{2.13}
\end{equation*}
$$

We take the limit of $\Delta s \rightarrow 0$. This surface charge density could depend on its location on the surface $r$ and we write $\rho_{S}=\rho_{S}(\mathbf{r})$. It could also be uniformly

## Superposition principles

distributed on the surface and the charge distribution would be a constant in this case.

The charge could also be distributed along a line whose length is $\Delta \mathrm{L}$. The charge would have a uniform distribution in the two transverse coordinates of the line. This would yield a linear charge density $\rho \mathrm{L}$ where

$$
\begin{equation*}
\rho_{\mathrm{L}}=\frac{\Delta \mathrm{Q}}{\Delta \mathrm{~L}}\left(\frac{\mathrm{C}}{\mathrm{~m}}\right) \tag{2.14}
\end{equation*}
$$

We again take the limit of $\Delta \mathrm{L} \rightarrow 0$. Once again, the charge could be distributed nonuniformly or uniformly along the line.

We will find it advantageous to use all three definitions in later derivations. We will later encounter "infinite sheets" or "infinite lines" that have charge densities given by (2.13) and (2.14). This merely implies that an infinite amount of charge is distributed over these infinite surfaces or lines but the ratios given in these two equations (2.13) and (2.14) are finite.

(a)

(b)

(c)

Figure 2-4. Distributed charge densities: a) The charge is distributed in a volume $\Delta v$ creating a volume charge density $\rho_{\mathrm{v}}, b$ ) The charge is distributed on a surface $\Delta s$ creating a surface charge density $\rho_{\mathrm{s}}, \mathrm{c}$ ) The charge is distributed along a line $\Delta \mathrm{L}$ creating a linear charge density $\rho_{\mathrm{L}}$.

If we want to calculate the electric field that is created by either of the distributed charge density distributions, we will make use of the principle of superposition that was stated in equation (2.10) and shown in Figure 2-5. A quick

## Superposition principles

glimpse at Figure 2-5 should convince us that numerical techniques may have to be employed for most charge distributions in order to calculate the electric field. Fortunately for us, there are a few examples that can be treated analytically and some of them will appear in this text.


Figure 2-5. In order to calculate the electric field at the point $P$, the differential electric fields $\Delta \mathbf{E}_{\mathbf{j}}$ caused by the charges in the differential volumes $\Delta \mathrm{v}_{\mathrm{j}}$ are added together vectorially.

If we let the differential volumes $\Delta \mathrm{v}_{\mathrm{j}}$ become very small and the number of the small volumes to become very large, then the summation of the distinct electric fields caused by the discrete charges within these volumes will eventually cause the summation to become an integral that must be performed over the entire volume $v$ where the distributed charge density is located. This integration is written as

$$
\begin{equation*}
\mathbf{E}=\frac{1}{4 \pi \varepsilon_{o}} \int_{V} \frac{\rho_{v}}{R^{2}} \mathbf{u}_{R} d v \tag{2.15}
\end{equation*}
$$

Equation (2.15) implies that there exists a differential electric field that is directed radially from each of the differential charges that is enclosed within each of the differential volumes. The total electric field that will emanate from the

## Superposition principles

entire volume v is calculated by integrating the charge density over the entire volume. Each of the incremental electric fields will have their individual unit vectors and the integration must incorporate this fact.

If we are given a particular charged object and wish to analytically calculate the electric field caused by it, the first thing that we must do is to select the proper coordinate system in which the integration must be performed. This choice is usually predicated on any possible symmetry that can be found in the problem. For example, if the charged body were a sphere that was centered on the origin of a coordinate system, we should attempt the solution in spherical coordinates. If the charged body were a long cylindrical rod that was centered at the origin, we should use cylindrical coordinates.

The variables that appear in this integral are defined as follows. The variable $R$ is the distance between the point of observation and the location of a particular charge element $\rho_{\mathrm{v}}\left(\mathbf{r}^{\prime}\right) \mathrm{dv}$ that is within the volume of integration v . In Cartesian coordinates, we write

$$
\begin{equation*}
R=\sqrt{\left(x-x^{\prime}\right)^{2}+\left(y-y^{\prime}\right)^{2}+\left(z-z^{\prime}\right)^{2}} \tag{2.16}
\end{equation*}
$$

where $x^{\prime}, y^{\prime}$, and $z^{\prime}$ specify the location of the differential charge element and $x, y$, and $z$ specify the location where the electric field is to be determined at the point $P$. The unit vector $u_{R}$ is directed from this charge element to the point $P$. The unit vector $\mathbf{u}_{\mathrm{R}}$ will change as the integration is performed. This will be noted when actually performing the integration. In this general equation (2.15), we have to be careful since vectors are present and we would have to perform the integration separately over the three components in the differential volume dv .

In order to illustrate the procedure involved in setting up the integral and identifying each term in the integral, we calculate the electric field from a finite amount of charge that is uniformly distributed on a finite line. The linear charge density on this line will be $\rho_{\mathrm{L}}$. This linear charge density is depicted in Figure 2-6.

## Superposition principles

From this figure, we find that the unit vector from a differential charge that is localized on a section of the line whose length is $d z$ to the point of observation is given by

$$
\begin{equation*}
\mathbf{u}_{\mathrm{R}}=\frac{-\mathrm{z} \mathbf{u}_{\mathrm{z}}+\mathrm{r} \mathbf{u}_{\mathrm{r}}}{\sqrt{\mathrm{z}^{2}+\mathrm{r}^{2}}} \tag{2.17}
\end{equation*}
$$

The variation of the unit vector upon the variable $z$ alluded to earlier is clearly displayed in (2.17). In the calculation, we will assume that there is symmetry in that the point of observation is taken to be at the midpoint of the line. Therefore, for every charge segment at a distance $+z$, there will be an equivalent charge element located at $-z$. This is an example of symmetry and it is shown in Figure 2-6. Because of this symmetry, the components of the electric fields polarized in the $\pm z$ directions will cancel $\left(\Delta \mathrm{E}_{\mathrm{z}^{+}}=\Delta \mathrm{E}_{\mathrm{z}_{-}}\right)$. If the line of charge were infinite in length, the center of symmetry could be placed anywhere along the line.


Figure 2-6. Calculating the electric field from a uniformly distributed finite line of charge. The radial axis is at the center of the charged line. Because of this symmetry, the tangential components $\Delta \mathbf{E}_{\mathbf{z}}$ of the electric field cancel.

## Superposition principles

The term polarization means that the field is directed in that particular direction. Therefore, the radial component of the electric field is given in terms of the differential electric field $\mathbf{d E}$ by

$$
\begin{equation*}
\mathrm{dE}_{\mathrm{r}}=\mathrm{dE} \cos \theta=\mathrm{dE} \frac{\mathrm{r}}{\mathrm{R}}=\mathrm{dE} \frac{\mathrm{r}}{\sqrt{\mathrm{r}^{2}+\mathrm{z}^{2}}} \tag{2.18}
\end{equation*}
$$

where the magnitude of differential electric field $d E$ is calculated from the charge that is contained in the length dz . This charge is equal to $\rho \mathrm{L} \mathrm{dz}$. Therefore, (2.15) becomes

$$
\begin{equation*}
\mathrm{dE}=\frac{\rho_{\mathrm{L}} \mathrm{dz}}{4 \pi \varepsilon_{0}\left(\mathrm{r}^{2}+\mathrm{z}^{2}\right)} \tag{2.19}
\end{equation*}
$$

The total radial electric field is given by the summation of all of the infinitesimal components $d E_{r}$ since this is a linear media and superposition applies. This summation becomes an integration of the linear charge density over the length of the line and it can be performed analytically.

$$
\begin{equation*}
E_{r}=\frac{1}{4 \pi \varepsilon_{0}} \int_{-\mathrm{a}}^{\mathrm{a}} \frac{\rho_{\mathrm{L}} \mathrm{r}}{\left(\mathrm{r}^{2}+\mathrm{z}^{2}\right)^{3 / 2}} d z=\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon_{0} r} \frac{a}{\sqrt{r^{2}+\mathrm{a}^{2}}} \tag{2.20}
\end{equation*}
$$

This integral can be performed with the substitution of $z=r \tan \theta$ or by using an integral table. As the length of the line is made extremely long $(2 a \rightarrow \infty)$, the electric field decreases as this distance increases.

$$
\begin{equation*}
\mathrm{E}_{\mathrm{r}}=\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon_{\mathrm{o}} \mathrm{r}} \tag{2.21}
\end{equation*}
$$

Example 2-6. Calculate the electric field from an infinite charged plane. Assume that the plane consists of an infinite number of parallel charged lines as shown in the figure.

## Superposition principles



Answer: It is possible to consider the infinite plane as a parallel array of juxtaposed infinite charged lines. Hence, we can use (2.21) as our point of embarkation, where the distance $R=\sqrt{x^{2}+y^{2}}$. The linear charge density $\rho_{L}$ of a particular line whose width is $d y$ is just equal to $\rho_{L}=\rho_{S} d x$. Due to symmetry, the components of the electric field that are tangent to the plane will cancel. Therefore, we need only find the component of the electric field that is normal to the plane

$$
\begin{aligned}
E_{y} & =\int_{-\infty}^{\infty} d E \cos \alpha d x=\int_{-\infty}^{\infty} \frac{\rho_{S}}{2 \pi \varepsilon_{0} \sqrt{x^{2}+y^{2}}} \frac{y}{\sqrt{x^{2}+y^{2}}} d x \\
& =\frac{\rho_{S}}{2 \pi \varepsilon_{0}} \int_{-\infty}^{\infty} \frac{y}{x^{2}+y^{2}} d x=\left.\frac{\rho_{S}}{2 \pi \varepsilon_{0}} \tan ^{-1}\left(\frac{x}{y}\right)\right|_{-\infty} ^{\infty}=\frac{\rho_{S}}{2 \varepsilon_{0}}
\end{aligned}
$$

We find that the electric field is independent of the distance that it is above the infinite charged sheet. An alternative integration could be performed by assuming that the differential surface areas are concentric circular washers.

Due to the symmetry found in these two examples, we have been able to obtain analytical solutions for the electric field from two different charge configurations using the integral given in (2.15). We already know the electric field due

## Superposition principles

to a point charge in (2.8). The field varies respectively in distance from the charge region as $R^{-2}, R^{-1}$ and being independent of $R$ for the electric field from a point charge, an infinite line charge and an infinite surface charge. We would expect a difference since the infinite line charge and the infinite surface charge each contain an increasing order of infinity more charge than the point charge.

The assumption of symmetry has made these two examples problems that can easily be solved. There are, however, many more examples in which one cannot invoke these arguments of symmetry. The resulting integration may have to be performed numerically and we'll discuss this topic with reference to MATLAB later after we encounter the subject of the electric potential. This will permit us to neglect any vector notation and this will simplify our discussion of that topic. In the material that immediately follows, we'll continue to make the symmetry assumption.

## Gauss's law

### 2.4. Gauss's law

There are cases where it is possible to find the electric field directly without the integrals that were described previously. The only requirement that we will encounter is that there is sufficient symmetry inherent in the problem. This application will make use of various laws of vector calculus that were described in the first chapter. In order to introduce this procedure, we assume that there is a charge $Q$ that is uniformly distributed within a sphere whose radius is $a$. There will be a uniform volume charge density $\rho_{v}=\frac{Q}{4 \pi a^{3}}$ within the sphere


Figure 2-7. A charge $Q$ is distributed uniformly within a sphere whose radius is a.

From equation (2.8), we write the electric field at the surface of the sphere as

$$
\begin{equation*}
\mathbf{E}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{\mathrm{o}} \mathrm{a}^{2}} \mathbf{u}_{\rho} \tag{2.22}
\end{equation*}
$$

The next step is to integrate both sides of (2.22) over the entire spherical surface. This is a closed surface integral that is

$$
\begin{equation*}
\oint E \bullet d s=\oint \frac{Q}{4 \pi \varepsilon_{0} \mathrm{a}^{2}} \mathbf{u}_{\rho} \bullet \mathbf{d s} \tag{2.23}
\end{equation*}
$$

## Gauss's law

At the surface of the sphere, the electric field is a constant and it is directed in the radial direction. This is in the same direction as the differential surface area ds which implies that the scalar product of the two unit vectors $\mathbf{u}_{\rho} \bullet \mathbf{u}_{\rho}=1$. The closed surface integral yields the spherical surface area $4 \pi a^{2}$. Hence (2.23) becomes

$$
\begin{equation*}
\oint \mathbf{E} \cdot \mathbf{d} \mathbf{s}=\frac{\mathrm{Q}_{\mathrm{enc}}}{\varepsilon_{\mathrm{o}}} \tag{2.24}
\end{equation*}
$$

where we have explicitly stated that $Q=Q_{\text {enc }}$ represents the charge that is enclosed within the closed surface. This is Gauss's law. It is common to refer to the closed surface as a "Gaussian surface." In passing from (2.23) to (2.24), we have invoked symmetry arguments by stating that the electric field had a constant value on the surface.

We can rewrite equation (2.24) using the divergence theorem given in Chapter 1 and express the enclosed charge $Q_{\text {enc }}$ in terms of a charge density $\rho_{v}$ as

$$
\begin{equation*}
\oint E \cdot d s=\int_{\Delta v} \nabla \cdot E d v=\frac{\int_{\Delta v} \rho_{v} d v}{\varepsilon_{0}} \tag{2.25}
\end{equation*}
$$

In order for the two volume integrals in (2.25) to be equal for any arbitrary volume $\Delta \mathrm{v}$, the two integrands must be equal. This implies that

$$
\begin{equation*}
\nabla \bullet E=\frac{\rho_{\mathrm{v}}}{\varepsilon_{\mathrm{o}}} \tag{2.26}
\end{equation*}
$$

Equations (2.24) and (2.26) express one of the fundamental postulates of electrostatics. These equations are the integral form and the differential form of Gauss's law respectively. We will make extensive use of both forms.

As written, (2.24) is Gauss's law that allows us to ascertain the electric field in cases where there is significant symmetry inherent in the problem. This will be demonstrated for a charge $Q$ that is uniformly distributed within a spherical volume and for a charge that is uniformly distributed on a surface.

## Gauss's law

Consider the spherical volume shown in Figure 2-8. A charge $Q$ is uniformly distributed within the spherical volume $\Delta v=\frac{4 \pi \mathrm{a}^{3}}{3}$. The volume charge density $\rho_{v}$ is specified to be

$$
\begin{equation*}
\rho_{v}=\frac{Q}{\Delta v}=\frac{Q}{\left(\frac{4 \pi a^{3}}{3}\right)} \tag{2.27}
\end{equation*}
$$

The total charge $Q_{\text {enc }}$ that is enclosed within the spherical volume is calculated to be

$$
\begin{equation*}
\frac{\int_{\Delta v} \rho_{v} d v}{\varepsilon_{0}}=\frac{Q_{\mathrm{enc}}}{\varepsilon_{0}} \tag{2.28}
\end{equation*}
$$

This integral can be performed in this case and for several other cases. The charge that is enclosed within the volume is given by Qenc and this volume could have a radius $\rho$ that is greater than or less than the radius a of the sphere. This allows us to find the radial electric field both outside of and within the spherical volume.


Figure 2-8. A charge $Q$ is uniformly distributed within a sphere of radius a.

## Gauss's law

Let us first calculate the electric field outside of the spherical volume $\rho>a$. In this case, a spherical surface will entirely enclose the entire charge Q . Therefore Qenc = Q and we have

$$
\begin{equation*}
\frac{\int_{\Delta v} \rho_{\mathrm{v}} \mathrm{dv}}{\varepsilon_{\mathrm{o}}}=\frac{\mathrm{Q}}{\varepsilon_{\mathrm{o}}} \tag{2.29}
\end{equation*}
$$

Since the differential surface area is in the radial direction, we will have only a radial component of the electric field. The closed surface integral can also be performed and this leads to

$$
\begin{equation*}
\oint E \cdot d s=4 \pi \rho^{2} E_{\rho} \tag{2.30}
\end{equation*}
$$

Therefore, we use Gauss's law (2.24) and equate (2.29) and (2.30) to yield

$$
\begin{equation*}
4 \pi \rho^{2} E_{\rho}=\frac{Q}{\varepsilon_{0}} \tag{2.31}
\end{equation*}
$$

The radial electric field $E_{\rho}$ in the region $\rho>a$ is given by

$$
\begin{equation*}
E_{\rho}=\frac{Q}{4 \pi \varepsilon_{0} \rho^{2}} \tag{2.32}
\end{equation*}
$$

This is the same result that was obtained in (2.8) as we should expect since the charge is entirely enclosed within this larger spherical surface.

Within the sphere, $\rho<a$ and the total charge that is enclosed within this volume is given by

$$
\begin{equation*}
\frac{\mathrm{Q}_{\text {enc }}}{\varepsilon_{\mathrm{o}}}=\frac{\int_{\Delta v} \rho_{\mathrm{v}} \mathrm{dv}}{\varepsilon_{\mathrm{o}}}=\frac{1}{\varepsilon_{\mathrm{o}}} \int_{0}^{\pi} \sin \theta \mathrm{d} \theta \int_{0}^{2 \pi} \mathrm{~d} \phi \int_{0}^{\rho} \frac{\mathrm{Q}}{\left(\frac{4 \pi \mathrm{a}^{3}}{3}\right)} \rho^{2} \mathrm{~d} \rho=\frac{\mathrm{Q}}{\varepsilon_{\mathrm{o}}}\left(\frac{\rho}{\mathrm{a}}\right)^{3} \tag{2.33}
\end{equation*}
$$

In (2.33), we have used the definition of the differential volume $d v$ in spherical coordinates. The closed surface integral surrounding this charge is still given by (2.30). Hence the radial electric field within the charged sphere is found from equating these two expressions

$$
\begin{equation*}
4 \pi \rho^{2} E_{\rho}=\frac{Q}{\varepsilon_{0}}\left(\frac{\rho}{a}\right)^{3} \tag{2.34}
\end{equation*}
$$

Solving for the radial electric field with in the sphere, we compute

$$
\begin{equation*}
E_{\rho}=\frac{Q \rho}{4 \pi \varepsilon_{0} \mathrm{a}^{3}} \tag{2.35}
\end{equation*}
$$

We note that the electric field linearly increases with the radius $\rho$. This increase is due to the inclusion of more charge within this expanding spherical surface. A summary of the electric field as a function of radius is shown in Figure 2-9.


Figure 2-9. The variation of the electric field inside and outside of a uniformly charged sphere.

Example 2-7. A linear charge density $+\rho_{\mathrm{L}}$ is distributed on the inner hollow cylinder and a linear charge density $-\rho_{\mathrm{L}}$ is distributed on the outer hollow cylinder.

Find the electric field in all regions of space using Gauss's law.


## Gauss's law

Answer: Due to the cylindrical symmetry that is found in this problem, the Gaussian surface will be a cylinder. In the region $r<a$, the enclosed charge is equal to zero. Hence the electric field within the inner cylinder is equal to zero. In the region $\mathrm{a}<\mathrm{r}<\mathrm{b}$, the enclosed charge in a length $\Delta \mathrm{L}$ is equal to $+\rho_{\mathrm{L}} \Delta \mathrm{L}$. Therefore, from Gauss's law (2.24) in cylindrical coordinates, we write

$$
\oint E \cdot d \mathbf{d s}=\int_{\phi=0}^{\phi=2 \pi} \mathrm{rd} \phi \int_{z=0}^{z=\Delta L} E_{\mathrm{r}} \mathrm{dz}=\frac{\mathrm{Q}_{\text {enc }}}{\varepsilon_{0}}=\frac{+\rho_{\mathrm{L}} \Delta L}{\varepsilon_{0}}
$$

The first integral contributes a factor of $2 \pi$ and the second integral contributes a factor of $\Delta L$. Hence, the radial electric field in the region $a<r<b$ is equal to

$$
\mathrm{E}_{\mathrm{r}}=\frac{+\rho_{\mathrm{L}}}{2 \pi \varepsilon_{0} r}
$$

In the region $r>b$, the enclosed charge is equal to $\left\{+\rho_{L} \Delta L-\rho_{L} \Delta L\right\}=0$ in the length $\Delta \mathrm{L}$. The electric field external to the outer cylinder will be equal to zero.

In order to further emphasize the physical meaning of Gauss's law, let us introduce a slightly different derivation. Its interpretation will allow us to clearly see the meaning of the term "enclosed charge Qenc." This method will apply for problems where there is sufficient symmetry and there is a dependence on only one of the dependent variables such that (2.26) can be written as

$$
\nabla \bullet E=\left\{\begin{array}{ll}
\frac{d E}{d x}=\frac{\rho_{v}(x)}{\varepsilon_{0}} & \text { Cartesian coordinates }  \tag{2.36}\\
\frac{1}{r}(r E) \\
d r & =\frac{\rho_{v}(r)}{\varepsilon_{0}}
\end{array} \quad\right. \text { cylindrical coordinates }
$$

In writing (2.36), the assumption has been made that sufficient symmetry exists such that the electric field depends only on one of the coordinates. Hence we can
use the ordinary derivative rather than the partial derivative. Also, the charge density $\rho_{\mathrm{v}}$ is only a function of this coordinate. For example, the charge could be distributed within a spherical volume has shown in Figure 2-10.


Figure 2-10. Charge is distributed within the spherical volume whose radius is a. The charge Qenc refers to the charge that is enclosed within the sphere whose radius is $\rho$.

The charge Qenc, that is enclosed within the spherical volume depicted in Figure 2-10 is given from (2.33) where the integrations over the transverse coordinates $(\rho$ and $\phi$ ) yield a factor of only $4 \pi$. We are left with the following integral in the radial variable that has yet to be performed.

$$
\begin{equation*}
Q_{\text {enc }}=4 \pi \int_{0}^{\rho} \rho_{v}(\tilde{\rho}) \widetilde{\rho}^{2} d \tilde{\rho} \tag{2.37}
\end{equation*}
$$

where $\tilde{\rho}$ is a dummy variable of integration. Using the chain rule for differentiation, we write the left hand side of (2.36) as

$$
\begin{align*}
& \frac{1}{\rho^{2}} \frac{d\left(\rho^{2} E\right)}{d \rho}=\frac{1}{\rho^{2}} \frac{d\left(\rho^{2} E\right)}{d Q_{\mathrm{enc}}} \frac{d Q_{\mathrm{enc}}}{d \rho}=  \tag{2.38}\\
& =\frac{1}{\rho^{2}} \frac{d\left(\rho^{2} E\right)}{d Q_{\mathrm{enc}}}\left\{4 \pi \rho^{2} \rho_{\mathrm{v}}(\rho)\right\}=\frac{d\left(\rho^{2} E\right)}{d Q_{\mathrm{enc}}}\left\{4 \pi \rho_{\mathrm{v}}(\rho)\right\}
\end{align*}
$$

## Gauss's law

where we have employed the definition for the differentiation of an integral using Leibnitz rule ${ }^{1}$. The term $\left\{4 \pi \rho^{2} \rho_{\mathrm{v}}(\rho)\right\}$ arises from the application of this definition. Since the charge density $\rho_{\mathrm{v}}(\rho)$ is common to both sides of the equation, it cancels and (2.36) becomes

$$
\begin{equation*}
4 \pi \frac{\mathrm{~d}\left(\rho^{2} E\right)}{\mathrm{dQ}} \mathrm{Q}_{\mathrm{enc}}=\frac{1}{\varepsilon_{0}} \tag{2.39}
\end{equation*}
$$

This can be easily integrated to yield

$$
\begin{equation*}
\mathrm{E}=\frac{\mathrm{Q}_{\mathrm{enc}}}{4 \pi \varepsilon_{\mathrm{o}} \rho^{2}} \tag{2.40}
\end{equation*}
$$

Equation (2.40) explicitly states that the electric field external to a surface is determined by the charge Qenc that is enclosed within the surface. This is the physical interpretation of Gauss's law.

We can also use this technique to compute the electric field within a sphere that has a uniform charge density

$$
\begin{equation*}
\rho_{\mathrm{v}}=\frac{Q}{\left(\frac{4 \pi \mathrm{a}^{3}}{3}\right)} \tag{2.41}
\end{equation*}
$$

The charge $Q_{e n c}$ that is enclosed within the spherical volume $\frac{4 \pi \rho^{3}}{3}$ is given by

$$
\begin{equation*}
Q_{\text {enc }}=\int_{\Delta v} \rho_{v} d v=\int_{0}^{\pi} \sin \theta d \theta \int_{0}^{2 \pi} d \phi \int_{0}^{\rho} \frac{Q}{\left(\frac{4 \pi a^{3}}{3}\right)} \rho^{2} d \rho=Q\left(\frac{\rho}{a}\right)^{3} \tag{2.42}
\end{equation*}
$$

Substitute this in (2.40) and obtain

$$
\begin{equation*}
E=\frac{Q_{\mathrm{enc}}}{4 \pi \varepsilon_{\mathrm{o}} \rho^{2}}=\frac{Q\left(\frac{\rho}{\mathrm{a}}\right)^{3}}{4 \pi \varepsilon_{\mathrm{o}} \rho^{2}}=\frac{Q \rho}{4 \pi \varepsilon_{\mathrm{o}} \mathrm{a}^{3}} \tag{2.43}
\end{equation*}
$$

[^2]
## Gauss's law

The result in (2.43) is the same result that was given in (2.35) but from a slightly different point of view. We should not expect to find and do not find a different result. What we have done is employ a modified form of Lagrangian mass variables. This change of variables has been used by our colleagues in fluid mechanics who let the independent variable of space $\rho$ become the total mass $m_{\text {enc }}$ that is enclosed within the volume that is defined by this spatial variable $\rho$. Using this technique, they have been able to advance the solutions for fairly difficult problems. We have merely borrowed and used their technique in order to further interpret that meaning of Gauss's law in electrostatics.

In using Gauss's law, we have made extensive use of various symmetry arguments. Because of this, we have been able to reduce the problem such that it depends on only one spatial variable. If the enclosing sphere is sufficiently larger than the container of the charge, then it may be a good approximation to assume that the enclosed charge is localized at a point that is at the center of the sphere in order to obtain an approximate solution for the electric field. In conclusion, Gauss's law states that there must be a charge that is enclosed within an enclosed surface in order to have an electric field emanate or terminate within the enclosed surface. Otherwise, the electric field will just pass through this region.

## Potential energy and electric potential

### 2.5. Potential energy and electric potential

A charged particle will gain a certain amount of potential energy as the particle is moved in a region against an electric field as shown in Figure 2-11. This is because work has to be done to overcome the force due to the electric field.


Figure 2-11. The transport of a charge $Q$ against an electric field $E$ from point a to point $b$ causes the particle's potential energy to change.

The energy $\Delta \mathrm{W}_{\mathrm{e}}$ in Joules $(\mathrm{J})$ that will be gained by the charged particle is calculated from the line integral

$$
\begin{equation*}
\Delta W_{e}=\int_{a}^{b} F \cdot d I=-Q \int_{a}^{b} E \cdot d l \quad(J) \tag{2.44}
\end{equation*}
$$

Note the appearance of a minus sign in this equation. This indicates that if the charge is positive, work must be done to overcome the electric field. Energy must be conserved in this process. Therefore, the positive charge will gain in energy has calculated from (2.44). The potential energy of the positive charge will be increased. A negatively charged particle will experience a decrease of potential energy if it followed the same path as indicated in Figure 2-11. The fact that the scalar product has been employed in this integral reflects the fact that no work is performed in regions where the force (or electric field) is perpendicular to the direction of the motion. This fact will be important in later work.

## Potential energy and electric potential

It is possible to define the total electrostatic energy stored in a volume using the following gedanken experiment. 1 Let us assume that all charges initially are at $\rho=\infty$ and none exist in the laboratory. When we say that all of the charges are at $\rho=\infty$, we also imply that each of the charges is infinitely far from its neighbor and there are no Coulomb forces between them that will have to be included in our experiment. Any electric field far out at $\rho=\infty$ will have decayed to have a value of zero in the laboratory.


Figure 2-12. Calculation of the work required to bring charges from $x=-\infty$ into the defined space. (a) Moving the first charge $Q_{1}$ requires no work. (b) Moving the second charge $Q_{2}$ requires work since the first charge $Q_{1}$ creates an electric field. (c) Moving the third charge $Q_{3}$ requires work since there are two charges already present.

[^3]
## Potential energy and electric potential

Let us compute the total work required to bring the charges into the shaded region from "where the wild things are." This is illustrated in Figure 2-12. No work is required to bring the first charge $\mathrm{Q}_{1}$ into the shaded region since no force is required to move this charge in our frictionless wagon, hence $W_{1}=0$. However, to bring the second charge $\mathrm{Q}_{2}$ into the region, we will have to do some work since we have to overcome the Coulomb force of repulsion caused by the presence of the first charge in the laboratory. Hence a minus sign will appear in this equation

$$
\begin{equation*}
W_{2}=-\int_{-\infty}^{x_{\mathrm{b}}} \frac{\mathrm{Q}_{1} \mathrm{Q}_{2}}{4 \pi \varepsilon_{0}\left(\mathrm{x}-\mathrm{x}_{\mathrm{a}}\right)^{2}} \mathrm{dx}=\frac{\mathrm{Q}_{1} \mathrm{Q}_{2}}{4 \pi \varepsilon_{0}\left|\mathrm{x}_{\mathrm{b}}-\mathrm{x}_{\mathrm{a}}\right|}=\mathrm{Q}_{2} \mathrm{~V}_{1} \tag{2.45}
\end{equation*}
$$

In (2.45), the energy depends on the magnitude of the distance separating the two charges. We factor the charge $Q_{2}$ from the remaining terms. The remaining terms are due to the presence of the first charge $Q_{1}$. We will call this collection of terms as being the absolute potential $\mathrm{V}_{1}$ that is entirely caused by the charge 1 residing in this region.

$$
\begin{equation*}
\mathrm{V}_{1}=-\int_{-\infty}^{\mathrm{x}_{\mathrm{b}}} \frac{\mathrm{Q}_{1}}{4 \pi \varepsilon_{0}\left(\mathrm{x}-\mathrm{x}_{\mathrm{a}}\right)^{2}} \mathrm{dx}=\frac{\mathrm{Q}_{1}}{4 \pi \varepsilon_{\mathrm{o}}\left|\mathrm{x}_{\mathrm{b}}-\mathrm{x}_{\mathrm{a}}\right|} \tag{2.46}
\end{equation*}
$$

The units of the voltage are volts [V]. We will encounter the term potential again in a few lines where it will be given a physical interpretation. Initially, we will just use this integration as a mathematical entity.

Since the charges that are carried into the room in Figure 2-12 each have a label on them (1 or 2), it behooves us to ask the following question, "Would it have made any difference in calculating the total energy that had been expended if we had brought the charge labeled \#2 in before the charge labeled \#1?" The answer is NO! Equation (2.45) could equivalently be written as $Q_{1} V_{2}$ with no loss of generality. We'll encounter this point in a few minutes when we try to generalize this result.

## Potential energy and electric potential

Passing on to the next iteration of carrying charges into the room as depicted in Figure 2-12, we now bring charge $Q_{3}$ into the shaded region. The work that has to be performed, following the same procedure of calculating the work required to bring in $Q_{2}$, will be against the electric fields due to charges $Q_{1}$ and $Q_{2}$ already being in the laboratory.

$$
\begin{equation*}
W_{3}=-\int_{-\infty}^{x_{c}} \frac{Q_{1} Q_{3}}{4 \pi \varepsilon_{0}\left(x-x_{a}\right)^{2}} d x-\int_{-\infty}^{x_{c}} \frac{Q_{2} Q_{3}}{4 \pi \varepsilon_{0}\left(x-x_{b}\right)^{2}} d x \tag{2.47}
\end{equation*}
$$

leads to the following expression

$$
\begin{equation*}
\mathrm{W}_{3}=\frac{\mathrm{Q}_{1} \mathrm{Q}_{3}}{4 \pi \varepsilon_{0}\left|\mathrm{x}_{\mathrm{c}}-\mathrm{x}_{\mathrm{a}}\right|}+\frac{\mathrm{Q}_{2} \mathrm{Q}_{3}}{4 \pi \varepsilon_{0}\left|\mathrm{x}_{\mathrm{c}}-\mathrm{x}_{\mathrm{b}}\right|}=\mathrm{Q}_{3} \mathrm{~V}_{1}+\mathrm{Q}_{3} \mathrm{~V}_{2} \tag{2.48}
\end{equation*}
$$

The total work that has been expended in bringing the three charges into the shaded region is given by

$$
\begin{equation*}
W_{\text {total }}=W_{1}+W_{2}+W_{3}=0+Q_{2} V_{12}+Q_{3}\left(V_{13}+V_{23}\right) \tag{2.49}
\end{equation*}
$$

The double subscript notation for the potential $V_{i j}$ allows us to explicitly indicate that the potential due to charge i is to be evaluated at the location where charge j is eventually to be located.

The total energy that has been expended in order to bring the charges into the shaded region has to appear somewhere. None was converted into heat and subsequently lost since the charges were transported in frictionless vehicles. This energy is stored in this region as electrostatic stored energy. It can be recovered and used for other purposes at a later time. This energy has the potential to do work at a later time.

We could continue this process to include all N charges that were originally at the location $\xi=-\infty$. The procedure is straight forward and can be generalized to

$$
\begin{equation*}
\mathrm{W}_{\text {total }}=\frac{1}{2} \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\mathrm{j}=1}^{\mathrm{N}(\mathrm{i} i=1)} \frac{\mathrm{Q}_{\mathrm{i}} \mathrm{Q}_{\mathrm{j}}}{4 \pi \varepsilon_{0}\left|\mathrm{x}_{\mathrm{ij}}\right|} \tag{2.50}
\end{equation*}
$$

## Potential energy and electric potential

where $x_{i j}$ is the magnitude of the distance between the charge $Q_{i}$ and the charge $Q_{j}$. The factor of $1 / 2$ arises since the terms are counted twice in using the notation of this double summation. For example, a term with $i=6$ and $j=8$ will have the same value as the term with $\mathrm{i}=8$ and $\mathrm{j}=6$ and thus this value would be counted twice in the summation. The notation $\mathrm{N}(\mathrm{j} \neq \mathrm{i})$ indicates that this particular summation excludes the term $\mathrm{j}=\mathrm{i}$.

Let the potential at the $\mathrm{i}^{\text {th }}$ charge due to all of the other charges be given by $V_{j}$, that is

$$
\begin{equation*}
V_{i}=\sum_{j=1}^{N(j i f)} \frac{Q_{j}}{4 \pi \varepsilon_{0}\left|x_{i j}\right|} \tag{2.51}
\end{equation*}
$$

Hence the total stored electrostatic energy can be explicitly written as

$$
\begin{equation*}
\mathrm{W}_{\mathrm{e}}=\mathrm{W}_{\text {total }}=\frac{1}{2} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{Q}_{\mathrm{i}} \mathrm{~V}_{\mathrm{i}} \tag{2.52}
\end{equation*}
$$

We have used the notation $W_{e}$ to indicate the total stored electrostatic energy. Later, the symbol $W_{m}$ will be used to indicate the total stored magnetic energy when we discuss magnetic fields.

Example 2-8. Demonstrate that the factor of $1 / 2$ must be included in (2.50) for the case of $N=3$ charges. Recall that there is no energy required to bring in the first charge.

Answer: Explicitly expand (2.50) for the case $N=3$. We write

$$
W_{\text {total }}=\frac{1}{2}\left(\frac{Q_{1} Q_{2}}{4 \pi \varepsilon_{0} x_{12}}+\frac{Q_{1} Q_{3}}{4 \pi \varepsilon_{0} x_{13}}+\frac{Q_{2} Q_{1}}{4 \pi \varepsilon_{0} x_{21}}+\frac{Q_{2} Q_{3}}{4 \pi \varepsilon_{0} x_{23}}+\frac{Q_{3} Q_{1}}{4 \pi \varepsilon_{0} x_{31}}+\frac{Q_{3} Q_{2}}{4 \pi \varepsilon_{0} x_{32}}\right)
$$

Since the distances satisfy $x_{j, k}=x_{k, j}$ and the products of the charges satisfy $Q_{j} Q_{k}$ $=Q_{k} Q_{j}$, this can be written as

$$
W_{\text {total }}=\frac{Q_{1} Q_{2}}{4 \pi \varepsilon_{0} x_{12}}+\frac{Q_{1} Q_{3}}{4 \pi \varepsilon_{0} x_{13}}+\frac{Q_{2} Q_{3}}{4 \pi \varepsilon_{0} x_{23}}=Q_{2} V_{12}+Q_{3} V_{13}+Q_{3} V_{23}
$$

This is the same value for the total energy that was obtained in (2.49).

## Potential energy and electric potential

If the charge is distributed within a volume, we can further generalize the expression for the energy that is stored within a volume $\Delta \mathrm{V}$. The charge $\mathrm{Q}_{\mathrm{i}}$ in (2.52) can be replaced with a distribution of charge and the differential volume as $\rho_{\mathrm{v}} \mathrm{dv}$ and the discrete potential $\mathrm{V}_{\mathrm{j}}$ with a continuous term V . The summation is replaced with an integration of the volume charge density over the volume $d v$ where the distributed charge is located. Hence, the stored electrostatic energy is given by

$$
\begin{equation*}
W_{e}=\frac{1}{2} \int_{\Delta v} \rho_{v} V d v \tag{2.53}
\end{equation*}
$$

where we have again defined the total electrostatic stored energy as We . There are alternative methods of writing this expression as will be noted later.

It is now possible to give a physical interpretation to the terms electric potential or the more common expression voltage. The ratio of the work required to move the charge against the electric field from point a to point $b$ divided by the value of that charge is defined as the electric potential difference $\Delta \mathrm{V}_{\mathrm{ab}}$ between the points $a$ and $b$. This can be written with reference to the absolute potentials at the two points $a$ and $b$ as

$$
\mathrm{V}_{\mathrm{a}}-\mathrm{V}_{\mathrm{b}}=\Delta \mathrm{V}_{\mathrm{ab}}=\frac{1}{\mathrm{Q}}\left\{\int_{-\infty}^{\mathrm{b}} \mathbf{F} \cdot \mathbf{d} \mathbf{I}-\int_{-\infty}^{\mathrm{a}} \mathbf{F} \cdot \mathbf{d} \mathbf{I}\right\}=\frac{1}{\mathrm{Q}}\left\{\mathrm{Q} \int_{-\infty}^{\mathrm{b}} \mathbf{E} \cdot \mathbf{d} \mathbf{I}-\mathrm{Q} \int_{-\infty}^{\mathrm{a}} \mathbf{E} \cdot \mathbf{d} \mathbf{I}\right\}
$$

or

$$
\begin{equation*}
\Delta \mathrm{V}_{\mathrm{ab}}=\int_{\mathrm{a}}^{\mathrm{b}} \mathrm{E} \cdot \mathrm{dl} \tag{2.54}
\end{equation*}
$$

The units of the energy that has been expended to effect this action are given in SI units as Joules (J). The units of energy when we consider the individual charge is a very large quantity and we frequently measure the energy in terms of

## Potential energy and electric potential

the energy gained by an electron passing through a potential difference of 1 volt $\equiv \frac{1 \text { Joule }}{1 \text { Coulomb }}$. This energy is given in terms of electron volts $(\mathrm{eV})^{2}$.

To illustrate this method of calculating the potential difference given in (2.47), let us calculate the work (work $=$ charge $\times$ potential difference) required to move a charge q from a radius $=\mathrm{b}$ to a radius $\mathrm{r}=\mathrm{a}$ as shown in Figure 2-13. A charge $Q$ is located at the center of the inner spherical surface. The electric field between the concentric spherical surfaces shown in Figure 2-13 is calculated using (2.8)

$$
\begin{equation*}
\mathbf{E}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{\rho} \rho^{2}} \mathbf{u}_{\rho} \tag{2.55}
\end{equation*}
$$

Hence the potential difference between the two spherical surfaces is computed to be

$$
\begin{equation*}
\Delta \mathrm{V}_{\mathrm{ab}}=\int_{\mathrm{a}}^{\mathrm{b}} \frac{\mathrm{Q}}{4 \pi \varepsilon_{\mathrm{o}} \rho^{2}} \mathbf{u}_{\rho} \bullet \mathrm{d} \rho \mathbf{u}_{\rho}=-\left.\frac{\mathrm{Q}}{4 \pi \varepsilon_{o} \rho}\right|_{\mathrm{a}} ^{\mathrm{b}}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{\mathrm{o}}}\left(\frac{1}{\mathrm{a}}-\frac{1}{\mathrm{~b}}\right) \tag{2.56}
\end{equation*}
$$

If the radius $b$ of the outer sphere increased to a value of $b \rightarrow \infty$, this would then be the potential difference between $\rho=+\infty$ and $\rho=$ a. The potential at $\rho=\infty$ is defined as being equal to zero. It is frequently designated as being the ground potential. In a properly connected three-wire electrical cord, the third wire is connected to this far-off place with a third wire. This far off place in practice may actually be the copper tubing that brings the water into the room. In many student laboratories, copper tubing is frequently located above the pipe that contains the electrical power for the instruments. This tubing is supposed to be connected to ground for safety reasons and to have a well defined ground potential in the laboratory. Imagine the red faces of teachers who might connect a light bulb to copper tubing in different parts of the laboratory or the building and have it shine brightly. The potential at the radius $\rho=$ a may be either positive or

[^4]
## Potential energy and electric potential

negative depending upon the sign of the charge. This electric potential with respect to the ground potential is defined as the absolute potential at that particular point.


Figure 2-13. Two concentric spherical surfaces surrounding a charge Q. The dashed lines indicate a possible path that is to be followed in order to calculate the potential difference between the two spheres.

We should note at this point that there is only a potential difference between the two spherical surfaces in Figure 2-13 (path 1->2 and path 3->4). If we move along a circumferential line on one of the surfaces (path 2->3 which is at a constant radius), no work would be required since $\mathbf{E}$ and $\mathbf{d l}$ are then perpendicular to each other. From (2.54), the work required to effect this move is equal to zero. A surface that has the same potential is called an equipotential surface. This term will be encountered again in this text and we might expect to see it later in several practical situations. For example, a metal container surrounding an electronic device should be an equipotential surface and that surface for safety reasons should be at the ground potential. The third wire in the electric cable connects this outer cover to ground. If we reach for the container and the

## Potential energy and electric potential

connection is faulty, a "zap!" will quickly convince us that it is an unsafe implement to have in the home. The water pipes and the "third wire conductor" in the home are designed to be good connections to the ground potential.

Example 2-9. Calculate the variation of the potential between two concentric cylinders if the potential of the inner cylinder is $\mathrm{V}_{\mathrm{o}}$ and the potential of the outer cylinder is 0 .


Answer: From (2.21), the radial electric field from an infinitely long cylinder that has a uniform linear charge density $\rho_{\mathrm{L}} \mathrm{C} / \mathrm{m}$ on the external surface of the inner conductor is given by $E_{r}=\frac{\rho_{L}}{2 \pi \varepsilon_{0} r}$. This charge density is, however, not known in this example and it must be computed. Only the potential difference between the two cylinders was given. The spatial variation of the potential between the cylinders is computed from the electric field using (2.54) to yield

$$
\mathrm{V}(\mathrm{r})=-\int_{\mathrm{b}}^{r} \mathrm{E}_{\mathrm{r}} \mathrm{dr}=-\int_{\mathrm{b}}^{r} \frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon_{0} r} \mathrm{dr}=-\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon_{0} r} \ln \left(\frac{\mathrm{r}}{\mathrm{~b}}\right)
$$

## Potential energy and electric potential

This satisfies the requirement that the potential at $r=b$ be equal to zero. In this case, the constant of integration is included in the charge density. In order to compute this constant, we evaluate the potential at $\mathrm{r}=\mathrm{a}$ to be $\mathrm{V}_{\mathrm{O}}$. This yields

$$
\mathrm{V}_{\mathrm{o}}=-\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon_{0} \mathrm{r}} \ln \left(\frac{\mathrm{a}}{\mathrm{~b}}\right)
$$

Eliminate the charge density $\rho_{\mathrm{L}}$ between these two expressions to obtain the potential variation between the two cylinders.

$$
V(r)=V_{o} \frac{\ln \left(\frac{r}{b}\right)}{\ln \left(\frac{a}{b}\right)}
$$

This electric potential satisfies the boundary conditions. Recall that when $r=b$, we have $\ln (1)=0$. When $r=a$, the numerator and denominator cancel.

If the separation between two equipotential surfaces is very small and the potential difference is also small, we can approximate the voltage difference between two surfaces using (2.54)

$$
\begin{equation*}
d V \approx-E \cdot d l=-E_{x} d x-E_{y} d y-E_{z} d z \tag{2.57}
\end{equation*}
$$

From the chain rule, we write

$$
\begin{equation*}
\mathrm{dV}=\frac{\partial \mathrm{V}}{\partial \mathrm{x}} \mathrm{dx}+\frac{\partial \mathrm{V}}{\partial \mathrm{y}} \mathrm{dy}+\frac{\partial \mathrm{V}}{\partial \mathrm{z}} \mathrm{dz} \tag{2.58}
\end{equation*}
$$

In comparing (2.57) and (2.58), we are able to relate the various terms of the electric field as $\mathrm{E}_{\mathrm{x}}=-\partial \mathrm{V} / \partial \mathrm{x} ; \mathrm{E}_{\mathrm{y}}=-\partial \mathrm{V} / \partial \mathrm{y}$; and $\mathrm{E}_{\mathrm{z}}=-\partial \mathrm{V} / \partial \mathrm{z}$. Therefore, we can write the electric field in vector notation as

$$
\begin{equation*}
\mathbf{E}=-\frac{\partial V}{\partial \mathbf{x}} \mathbf{u}_{\mathbf{x}}-\frac{\partial V}{\partial \mathbf{y}} \mathbf{u}_{\mathbf{y}}-\frac{\partial \mathrm{V}}{\partial \mathbf{z}} \mathbf{u}_{\mathbf{z}} \tag{2.59}
\end{equation*}
$$

If we knew the location and values of various equipotential surfaces, say from a sequence of measurements, then it would be possible to calculate both the magnitude and the vector direction associated with the electric field. In writing

## Potential energy and electric potential

(2.59), we have made the statement that the electric field is defined to be in the direction of the maximum rate of change of the potential. In addition, we see that the electric field can also have the units of

$$
\mathrm{E}=\left(\frac{\text { volts }}{\text { meter }}\right) \equiv\left(\frac{\mathrm{V}}{\mathrm{~m}}\right)
$$

since the spatial derivative operation will introduce the unit of $1 /$ (length). These units are probably are the most commonly used in practice.

We identify (2.59) as being the three components of the gradient operation of the scalar electric potential in Cartesian coordinates. Hence the electric field can be found analytically by taking the negative gradient of the electrostatic potential. This is a written as

$$
\begin{equation*}
\mathrm{E}=-\nabla \mathrm{V} \tag{2.60}
\end{equation*}
$$

This equation has important ramifications since it is usually easy to measure the electric potential at various points in the space. From these measurements, it is possible to connect all of the points in space that have the same potential. The resulting surfaces or lines are equipotential surfaces or lines. The application of (2.60) will then produce the magnitude and the direction of the resulting electric field in this space.

Example 2-10. The potential is measured at several locations in space. Connecting the points that have the same value of the electric potential with a line produces equipotential contours that can be drawn on a graph. Find the electric field at the point $P$. The graph is 5 meters $\times 5$ meters.

## Potential energy and electric potential



Answer: The electric field is computed from $E=-\nabla V$ in 2 dimensions. The measured equipotential contours are indicated by the solid lines and are separated by a distance of $\sqrt{1^{2}+1^{2}}=\sqrt{2} \mathrm{~m}$. The electric field at the point $P$ is

$$
\mathbf{E}=-\nabla \mathbf{V}=-\frac{\partial V}{\partial \mathbf{x}} \mathbf{u}_{\mathrm{x}}-\frac{\partial V}{\partial \mathbf{y}} \mathbf{u}_{\mathrm{y}}=-2\left(\mathbf{u}_{\mathrm{x}}+\mathbf{u}_{\mathbf{y}}\right) \mathrm{V} / \mathrm{m}
$$

The electric field is a vector that is pointing from the higher potential to the lower potential. There are several cases in practice where this "graphical" procedure can be performed using MATLAB.

We can substitute the electric field that is given in (2.60) into (2.26) in order to obtain the dependence of the electrostatic potential upon the charge density. We write

$$
\nabla \cdot(-\nabla \mathrm{V})=\frac{\rho_{\mathrm{v}}}{\varepsilon_{\mathrm{o}}}
$$

or

$$
\begin{equation*}
\nabla^{2} \mathrm{~V}=-\frac{\rho_{\mathrm{v}}}{\varepsilon_{0}} \tag{2.61}
\end{equation*}
$$

Equation (2.61) is called Poisson's equation. If the charge density $\rho_{v}$ is equal to 0 , this is called Laplace's equation. These two equations are extremely important

## Potential energy and electric potential

in obtaining solutions for the electric potential and the electric field in terms of the charge density. They will be studied in further detail in the next chapter since the methods of solution will involve either analytical or numerical techniques. Rest assured, you'll encounter this equation again.

If the charge is distributed within a volume $\Delta v$, the absolute potential can also be calculated. In this case, the summation of the individual charge contributions that appears in (2.51) is replaced with an integration over the entire volume in which the charge is distributed.


Figure 2-14. The voltage at a location in space that is caused by a volume charge distribution that is located at a different point in space. The distance between the volume charge distribution and the point where the voltage is to be determined is given by the magnitude of the difference of the two vectors $\mathbf{r}$ and $\mathbf{r}^{\prime}$.

From the definition of the absolute potential, we would write the potential caused by a volume distribution of charge that is not centered at the origin of a Cartesian coordinate system as


## Potential energy and electric potential

$$
\begin{equation*}
V(x, y, z)=\frac{1}{4 \pi \varepsilon_{0}} \int_{\Delta v} \frac{\rho_{v}\left(x^{\prime}, y^{\prime}, z^{\prime}\right)}{R} d x^{\prime} d y^{\prime} d z^{\prime} \tag{2.62}
\end{equation*}
$$

where the distance $R$ is given by $R=\sqrt{\left(x-x^{\prime}\right)^{2}+\left(y-y^{\prime}\right)^{2}+\left(z-z^{\prime}\right)^{2}}$. The expressions for the distance in other coordinate systems are given in Appendix 1. This distance is equal to the magnitude of the difference between the two vectors r' and $\mathbf{r}$ shown in Figure 2-14. In MATLAB notation, we define the distances using the command "norm."

The gradient operation is carried forth at the observer's position and we can assume that the two variables defining the coordinates $\mathbf{r}$ and $\mathbf{r}^{\prime}$ are independent from each other. Hence, the gradient operation $\nabla$ that is in the unprimed coordinate system can be freely brought inside the integral since the integration is performed in the primed coordinate system. It is left as a problem to verify that

$$
\begin{equation*}
\nabla\left(\frac{1}{R}\right)=-\frac{\mathbf{u}_{\mathrm{R}}}{\mathrm{R}^{2}} \tag{2.63}
\end{equation*}
$$

Therefore, we can compute the electric field once the potential field is known from (2.60). There may be certain advantages in finding the electrostatic potential first using (2.62) that contains the integration operation. The reason is that there is only an integration that is involved in (2.62) and it is a scalar integral. The direct calculation of electric field from a distributed volume charge distribution given in (2.15) was a vector integral. This implies that the integration must be performed over the three coordinates in order to obtain the three components of the electric field.

In (2.53), we calculated the electrostatic energy stored in the volume $\Delta v$ after bringing in additional charges We are now prepared to obtain other expressions for this energy by replacing the volume charge density $\rho_{v}$ using (2.26)

$$
\begin{equation*}
W_{e}=\frac{1}{2} \int_{\Delta v} \rho_{v} V d v=\frac{1}{2} \int_{\Delta v}\left(\varepsilon_{o} \nabla \cdot E\right) V d v \tag{2.64}
\end{equation*}
$$

## Potential energy and electric potential

This equation contains the product of the divergence of the electric field and the scalar electric potential. Using the vector identity from Appendix 1

$$
\nabla \bullet(a \mathbf{B})=\mathbf{B} \bullet \nabla \mathbf{a}+\mathrm{a} \nabla \bullet \mathbf{B}
$$

we rewrite (2.64)

$$
\begin{equation*}
\mathrm{W}_{\mathrm{e}}=\frac{\varepsilon_{\mathrm{o}}}{2} \int_{\Delta \mathrm{V}}[\nabla \bullet(\mathrm{VE})-\mathrm{E} \cdot \nabla \mathrm{~V}] \mathrm{dv} \tag{2.65}
\end{equation*}
$$

The first term on the right hand side of (2.65) can be converted to a closed surface integral using the divergence theorem

$$
\begin{equation*}
\int_{\Delta v} \nabla \bullet(\mathrm{VE}) \mathrm{dv}=\oint \mathrm{VE} \cdot \mathrm{ds} \tag{2.66}
\end{equation*}
$$

In this case, the surface $\Delta \mathrm{s}$ encloses the volume $\Delta \mathrm{v}$. Let us assume that this volume has a spherical shape and the volume charge density is localized near the center of the sphere. The electric field and the electric potential on the spherical surface will depend upon the radius of the sphere as $\frac{1}{R^{2}}$ and $\frac{1}{R}$ as given in (2.55) and (2.47) respectively. The surface area of the spherical surface will increase with increasing radius as $R^{2}$. Therefore, (2.66) will approach 0 as the radius $R \rightarrow \infty$. The conclusion is that (2.66) is equal to 0 for an infinite volume.

The total electrostatic energy can be written as

$$
\mathrm{W}_{\mathrm{e}}=\frac{\varepsilon_{\mathrm{o}}}{2} \int_{\Delta \mathrm{V}}[-\mathrm{E} \cdot \nabla \mathrm{~V}] \mathrm{dv}=\frac{\varepsilon_{\mathrm{o}}}{2} \int_{\Delta \mathrm{V}}[\mathrm{E} \cdot \mathrm{E}] \mathrm{dv}
$$

or

$$
\begin{equation*}
\mathrm{W}_{\mathrm{e}}=\frac{\varepsilon_{0}}{2} \int_{\Delta v} \mathrm{E}^{2} \mathrm{dv} \tag{2.67}
\end{equation*}
$$

where we have incorporated the relation that the electric field can be derived from the gradient of the scalar electric potential stated in (2.60). Note that the electrostatic energy depends upon the scalar quantity of the magnitude of the square of the electric field. We will encounter (2.67) later.

## Potential energy and electric potential

The following example which makes use of the superposition principles that were described above will introduce the reader to the subject of dielectric materials. A material will consist of a very large number of atoms. The Bohr model of a hydrogen atom assumes that there is a positive charge at the center of the atom and there is a negative charge that is located at a distance of approximately $10^{-10}$ meters from the center. This is a very small separation in distance and our colleagues in physics have defined a new unit called the Ångstrom where $1 \AA=10^{-10}$ meters.

Example 2-11 MATLAB . Find the potential $V$ due to two equal charges that have the opposite sign and are separated by a distance $d$ in a vacuum. If the point where the voltage is to be determined is much greater than the separation distance d , this configuration is known as an electric dipole. Using MATLAB, plot the equipotential contours and the electric field surrounding the charges.


Answer: Superposition will apply and the total electric potential and is computed by adding the individual contributions together. We find

$$
V=\frac{Q}{4 \pi \varepsilon_{0} \rho_{1}}-\frac{Q}{4 \pi \varepsilon_{0} \rho_{2}}
$$

## Potential energy and electric potential

We have chosen that $\rho \gg d$. Therefore, we can assume that the three lines $\rho, \rho_{1}$ and $\rho_{2}$ are almost parallel and the three angles $\theta, \theta_{1}$ and $\theta_{2}$ are approximately equal. With these assumptions, we write that

$$
\rho_{1} \approx \rho-\frac{d}{2} \cos \theta \text { and } \rho_{2} \approx \rho+\frac{d}{2} \cos \theta .
$$

Using these approximations, the voltage is computed to be
$V=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0}\left(\rho-\frac{\mathrm{d}}{2} \cos \theta\right)}-\frac{\mathrm{Q}}{4 \pi \varepsilon_{0}\left(\rho+\frac{\mathrm{d}}{2} \cos \theta\right)}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \rho\left(1-\frac{\mathrm{d}}{2 \rho} \cos \theta\right)}-\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \rho\left(1+\frac{\mathrm{d}}{2 \rho} \cos \theta\right)}$
The ratio of the distances $\frac{d}{2 \rho}$ is a small quantity and we can use it as a small parameter to expand the terms in the denominator and obtain

$$
\mathrm{V} \approx \frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \rho}\left(1+\frac{\mathrm{d}}{2 \rho} \cos \theta-1+\frac{\mathrm{d}}{2 \rho} \cos \theta\right)=\frac{\mathrm{Qd}}{4 \pi \varepsilon_{o} \rho^{2}} \cos \theta
$$

Let us define an electric dipole moment vector $\mathbf{p}=\mathbf{Q} \mathbf{d}$ that is directed from the negative charge to the positive charge. In addition, there is a unit vector $\mathbf{u}_{\rho}$ that is directed from the midpoint between the two charges to the point of observation. The term $\mathrm{Qd} \cos \theta$ can be interpreted as being the scalar product of these two vectors.

## Potential energy and electric potential



## Potential energy and electric potential

(b)

The normalized potential profile resulting from two charges that have the opposite signs at $x=0$ and $y= \pm 0.25$ is shown in (a). Note that it rapidly decays to zero. Equipotential contours and the resulting electric fields are shown in (b). Note that between the two charges, the electric field is directed from the positive charge to the negative charge.

In moving the charge from point $a$ to point $b$ in a region that contained an electric field, we found from equation (2.45) that work was required. If we move it back to point a along a slightly different path as shown in Figure 2-15 in an

## Potential energy and electric potential

electrostatic field, we will find that the expended energy is returned to us. In this case, we write (2.45) as

$$
\begin{equation*}
W_{e} / Q=\oint E \cdot d \mathbf{l}=0 \tag{2.68}
\end{equation*}
$$

where an integral over a closed contour is indicated. Equation (2.68) states that no energy is either expended or created in this process. In this case, the electrostatic field belongs to a class of fields that are called conservative fields. You may recognize that (2.68) is almost identical to the Kirrchhoff's voltage law that states "The sum of the voltage drops around a closed loop is equal to 0 ."

Equation (2.68) is the second postulate of electrostatics. This equation can be converted into a surface integral via Stokes's theorem. We write

$$
\begin{equation*}
0=\oint \mathbf{E} \cdot \mathbf{d l}=\int_{\Delta \mathbf{s}} \nabla \times \mathbf{E} \cdot \mathbf{d s} \tag{2.69}
\end{equation*}
$$

In order for this integral to be zero for any arbitrary surface, the integrand must be equal to zero. This allows us to obtain the second postulate of electrostatic fields in differential form.

$$
\begin{equation*}
\nabla \times \mathbf{E}=0 \tag{2.70}
\end{equation*}
$$

Equation (2.70) states that an electrostatic field is irrotational. We will encounter these postulates of electrostatics later when time-varying fields are described.


## Potential energy and electric potential

Figure 2-15. The closed path of integration from $a$ to $b$ and then back to $a$. There is an electric field in the region.

Example 2-12. Calculate the work required to move a charge $Q=2 \mathrm{C}$ around the closed path if there is an electric field $\mathbf{E}=3 \mathbf{u}_{\mathbf{x}}$ in the region.


Answer: The total work is computed from evaluating the closed line integral
$W_{e}=Q \oint \mathbf{E} \cdot \mathbf{d} \mathbf{I}=\mathbf{Q}\left(\int_{\mathrm{a}}^{\mathrm{b}} \mathbf{E} \cdot \mathbf{d} \mathbf{I}+\int_{\mathrm{b}}^{\mathrm{c}} \mathbf{E} \cdot \mathbf{d} \mathbf{I}+\int_{\mathrm{c}}^{\mathrm{d}} \mathbf{E} \cdot \mathbf{d} \mathbf{I}+\int_{\mathrm{d}}^{\mathrm{a}} \mathbf{E} \cdot \mathbf{d} \mathbf{I}\right)$
$=2\left(\int_{(1,1)}^{(4,1)} 3 \mathbf{u}_{\mathrm{x}} \bullet \mathrm{dx} \mathbf{u}_{\mathrm{x}}+\int_{(4,1)}^{(4,4)} 3 \mathbf{u}_{\mathrm{x}} \bullet \mathrm{dy} \mathbf{u}_{\mathrm{y}}+\int_{(4,4)}^{(1,4)} 3 \mathbf{u}_{\mathrm{x}} \bullet \mathrm{dx} \mathbf{u}_{\mathrm{x}}+\int_{(1,4)}^{(1,1)} 3 \mathbf{u}_{\mathrm{x}} \bullet \mathrm{dy} \mathbf{u}_{\mathrm{y}}\right)$
$=2(9+0-9+0)=0$
The limits that are included in each integral will determine the sign of a particular integration in this closed loop. This electric field is a conservative field.

## Numerical integration

### 2.6. Numerical integration

In the previous section, we have been able to calculate the electric potential that results from a distributed charge density. The examples that were presented all required that there was sufficient symmetry in order to find the electric potential. However, in practice, we usually encounter situations that do not have the required symmetry and we are forced to embark on a slightly different path. One of these paths requires the numerical solution of an electrostatics problem. Remembering that the integration is just a summation in which the number of distinct differential volumes, differential areas, or differential lengths has been allowed to approach 0, we now investigate whether a digital computer can actually perform the summation operation. It turns out that MATLAB provides a solution to this problem that requires minimal effort. In the following, we will initially develop the procedure before making use of the commands that are available in MATLAB.


Figure 2-16. The area under the curve $y=y(x)$ is numerically obtained by subdividing the area into small trapezoidal subareas and adding the areas of the individual trapezoids.

This is first illustrated in Figure 2-16 for a simple one-dimensional integration where the area under the curve is approximated with the summation

## Numerical integration

of a number of trapezoids. MATLAB has a command that creates the trapezoids automatically. This allows us to perform this calculation with just three commands

$$
\begin{equation*}
\mathrm{x}_{\min }: \Delta \mathrm{x}=\mathrm{b}-\mathrm{a}: \mathrm{x}_{\max }, \mathrm{y}(\mathrm{x}) ; \mathrm{z}=\operatorname{trapz}(\mathrm{x}, \mathrm{y}) . \tag{2.71}
\end{equation*}
$$

The choice of the value of $\Delta x$ is somewhat arbitrary and it depends upon the conflict between the desired accuracy and computational time. There are additional numerical integration programs that can be employed and the evaluation of the truncation errors has received mathematical attention.

In addition, this integration can be performed using Simpson's rule with the commands

> quad (func, xmin, xmax)
> dblquad (func, xmin, xmax, ymin, ymax) triplequad (func, xmin, xmax, ymin, ymax, zmin, zmax)
where the function 'func' is defined by the user. The numerical values for the end points of the integration are also stated in these commands. The default tolerance for the integration is $10^{-6}$. For example, this function could be the product of the three variables xyz where we would write

$$
\begin{equation*}
\text { func } \left.=\text { inline ('x. *y. }{ }^{*} z^{\prime}\right) \tag{2.73}
\end{equation*}
$$

In the definition for the function, we must remember to include the "." after each of the first two variables.

Example 2-13 MATLAB. Compare the analytical and the numerical evaluation of the area under the curve defined by the function $y=x^{2}$ in the interval $0 \leq x \leq 4$. Use both the 'trapezoidal' command and the 'quadrature' command.

Answer. The solution that is obtained from an analytical integration is

$$
\text { area }=\int_{0}^{4} x^{2} d x=\left.\frac{x^{3}}{3}\right|_{0} ^{4}=\frac{64}{3}=21.3333
$$

Using the trapezoidal commands, we write

## Numerical integration

$$
\begin{aligned}
& x=0: 0.001: 4 ; y=x . \wedge 2 ; \\
& \text { ztrapezoidal }=\operatorname{trapz}(x, y) ;
\end{aligned}
$$

Typing ztrapezoidal yields the numerical value of 21.3333.
The quadrature command requires the definition of the function

$$
\begin{gathered}
\text { func }=\text { inline ('x. } .^{\wedge} 2^{\prime} \text { ); } \\
\text { zquadrature }=\text { quad (func, } 0,4 \text { ); }
\end{gathered}
$$

Typing zquadrature yields the numerical value of 21.3333 .


Figure 2-17. Electric charge is distributed within a volume whose coordinate system is identified with an '. The voltage is to be determined in the coordinate system that does not have the '. The vector $\mathbf{R}$ is directed from the charge to the point of observation.

We find the electric potential due to an object that has a finite size as shown in Figure 2-17. The potential from an arbitrary body of charge was obtained in (2.62) and for convenience, we rewrite it below

# Numerical integration 

$$
\begin{equation*}
V(x, y, z)=\int \frac{\rho_{v}\left(x^{\prime}, y^{\prime}, z^{\prime}\right)}{4 \pi \varepsilon_{0} R} d x^{\prime} d y^{\prime} d z^{\prime} \tag{2.74}
\end{equation*}
$$

where $R=\sqrt{\left(x-x^{\prime}\right)^{2}+\left(y-y^{\prime}\right)^{2}+\left(z-z^{\prime}\right)^{2}}$. In our later development of a MATLAB program, we'll replace this distance with the command "norm( $\mathbf{r}-\mathbf{r}^{\prime}$ )." The primed variables refer to the location of the charge and the unprimed variables refer to the location at which the potential is to be computed. This permits us to assume that neither the charged object nor the calculated voltage have to be at the origin of the coordinate system.

The procedure to numerically perform the integration will be developed in full detail. After this development, we will use one of the commands that is presented in (2.71). We assume that a finite charge $Q$ is uniformly distributed on a thin finite sheet that is located at $z^{\prime}=0$. This results in a uniform charge distribution $\rho_{\mathrm{s}}$ at every point on the sheet. Let us also assume that the sheet has a rectangular shape that is centered on a Cartesian coordinate system as shown in Figure 2-18. The procedure that we will follow is to subdivide this large sheet into a number of small subareas and assume that the charge in each small subarea is localized at its individual center. Therefore, we have reduced the uniformly distributed charge to a large number of discrete individual charges. This reduction allows us to convert the integration (2.74) into a summation.

Therefore, (2.74) can be written as

$$
\begin{equation*}
V(x, y, z)=\sum_{j=1}^{N-1} \sum_{k=1}^{M-1} \frac{\rho_{s}(j, k) \Delta x^{\prime} \Delta y^{\prime}}{4 \pi \varepsilon_{0} R} \tag{2.75}
\end{equation*}
$$

where the rectangular sheet has been subdivided into $(N-1) \times(M-1)$ subareas. The area of each subarea is given by $\Delta x^{\prime} \Delta y^{\prime}$. The charge that is at the center of an individual subarea is $\rho_{s}(j, k) \Delta x^{\prime} \Delta y^{\prime}$. The charge density is not required to be uniform on the entire sheet. However, it is assumed to be uniform within each subarea.

## Numerical integration



Figure 2-18. The area containing the charge is subdivided into a number of small subareas. Each subarea is replaced with an individual charge whose value is equal to the charge contained in the individual subarea.

It is convenient to assume that the sheet is centered upon a Cartesian coordinate system since we later will be able to invoke certain symmetry arguments in order to simplify the calculation. In addition, we assume that it is located at $z^{\prime}=0$. With these assumptions, (2.74) becomes the following double integral

$$
\begin{equation*}
V(x, y, z)=\int_{-b / 2}^{b / 2} \int_{-a / 2}^{a / 2} \frac{\rho_{s}}{4 \pi \varepsilon_{0} R} d x^{\prime} d y^{\prime} \tag{2.76}
\end{equation*}
$$

where the uniformly distributed charge density is $\rho_{s}$. Using the definition for the distance R, we write.

$$
\begin{equation*}
V(x, y, z)=\int_{-b / 2}^{b / 2} \int_{-a / 2}^{a / 2} \frac{\rho_{s}}{4 \pi \varepsilon_{o} \sqrt{\left(x-x^{\prime}\right)^{2}+\left(y-y^{\prime}\right)^{2}+(z)^{2}}} d x^{\prime} d y^{\prime} \tag{2.77}
\end{equation*}
$$

Since the potential is to be determined along the $z$ axis, this simplifies to

## Numerical integration

$$
\begin{equation*}
\mathrm{V}(0,0, z)=\int_{-b / 2}^{b / 2} \int_{-a / 2}^{\mathrm{a} / 2} \frac{\rho_{\mathrm{s}}}{4 \pi \varepsilon_{o} \sqrt{\left(x^{\prime}\right)^{2}+\left(y^{\prime}\right)^{2}+(z)^{2}}} d x^{\prime} d y^{\prime} \tag{2.78}
\end{equation*}
$$

There are several cases that have a certain degree of symmetry associated with it. Symmetry may reduce the computational time required to perform the resulting calculation. The integration will have to be performed in only one quadrant of the surface, say $0 \leq x^{\prime} \leq \frac{a}{2}, 0 \leq y^{\prime} \leq \frac{b}{2}$. The computed value resulting from the integration will then just have to be multiplied by a factor of 4. Equation (2.78) becomes

$$
\begin{equation*}
V(0,0, z)=4 \int_{0}^{b / 2} \int_{0}^{\mathrm{a} / 2} \frac{\rho_{\mathrm{s}}}{4 \pi \varepsilon_{0} \sqrt{\left(x^{\prime}\right)^{2}+\left(y^{\prime}\right)^{2}+(z)^{2}}} d x^{\prime} d y^{\prime} \tag{2.79}
\end{equation*}
$$

In order to numerically evaluate the integral (2.79), we subdivide the entire plane in the quadrant' into small rectangles We will be required to identify the edges of each of the subareas in a methodical manner. One such procedure assumes that the point at the bottom-left corner is identified as being ( $j=1, k=1$ ) and the point at the upper-right corner is identified as being ( $j=N, k=M$ ). Therefore, there will be $(N-1) \times(M-1)$ small subareas in the subdivision process of the large area. The area of each individual subarea is equal to

$$
\begin{equation*}
\Delta \mathrm{A}=\frac{\mathrm{a}}{(\mathrm{~N}-1)} \frac{\mathrm{b}}{(\mathrm{M}-1)}=\mathrm{h}_{\mathrm{x}} \mathrm{~h}_{\mathrm{y}} \tag{2.80}
\end{equation*}
$$

The total charge $\Delta Q$ within each subarea is

$$
\begin{equation*}
\Delta \mathrm{Q}=\rho_{\mathrm{s}} \Delta \mathrm{~A}=\rho_{\mathrm{s}} \mathrm{~h}_{\mathrm{x}} \mathrm{~h}_{\mathrm{y}} \tag{2.81}
\end{equation*}
$$

and we assume that this charge is localized at the center of the subarea. If the charge has a nonuniform distribution, then the charge distribution $\rho_{\mathrm{s}}$ will have different values at each individual subarea. The incremental potential $\Delta \mathrm{V}_{\mathrm{j}, \mathrm{k}}$ due to the localized charge that is identified with the label " $\mathrm{j}, \mathrm{k}$ " is given by

## Numerical integration

$$
\begin{equation*}
\Delta \mathrm{V}_{\mathrm{j}, \mathrm{k}}=\frac{1}{4 \pi \varepsilon_{\mathrm{o}}}\left(\frac{\Delta \mathrm{Q}}{\sqrt{\left(j h_{x}-\frac{\mathrm{h}_{\mathrm{x}}}{2}\right)^{2}+\left(k h_{\mathrm{y}}-\frac{\mathrm{h}_{\mathrm{y}}}{2}\right)^{2}+\mathrm{z}^{2}}}\right) \tag{2.82}
\end{equation*}
$$

The center of a particular subarea is identified as

$$
\begin{equation*}
\left[\left(j h_{x}-\frac{h_{x}}{2}\right),\left(k h_{y}-\frac{h_{y}}{2}\right)\right] \tag{2.83}
\end{equation*}
$$

All that we need do now is use superposition and sum the incremental potentials due to each incremental charge

$$
\begin{equation*}
\mathrm{V}=4 \sum_{\mathrm{j}=1}^{\mathrm{N}-1} \sum_{\mathrm{k}=1}^{\mathrm{M}-1} \Delta \mathrm{~V}_{\mathrm{j}, \mathrm{k}} \tag{2.84}
\end{equation*}
$$

or

$$
\begin{equation*}
\mathrm{V}=4 \frac{1}{4 \pi \varepsilon_{0}} \sum_{\mathrm{j}=1}^{N-1} \sum_{\mathrm{k}=1}^{\mathrm{M}-1}\left(\frac{\Delta \mathrm{Q}}{\sqrt{\left(j h_{x}-\frac{h_{x}}{2}\right)^{2}+\left(k h_{y}-\frac{h_{y}}{2}\right)^{2}+z^{2}}}\right) \tag{2.85}
\end{equation*}
$$

Hence, the double integral in (2.79) has been converted into a double summation (2.85). The number of small subareas is determined by the compromise that must be made between accuracy and a computational time that is required to effect this calculation.

Example 2-14 ${ }^{\text {MATLAB }}$. Evaluate the potential due to a charge $Q$ being uniformly distributed upon a square surface whose area is equal to $a^{2}$ at the point $z=a$ if the number of subareas is equal to 1 and 4 . The center of the square is the $z$ axis which creates significant symmetry in the problem.

# Numerical integration 



Answer: The first iteration assumes that the charge is localized at the center of the square. Therefore, we compute from (2.46) that

$$
\mathrm{V}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{a}}
$$

The second iteration is evaluated using (2.85) since there are now 4 subareas. We calculate the potential to be
$\mathrm{V}=4 \frac{1}{4 \pi \varepsilon_{0}}\left(\frac{\left(\frac{\mathrm{Q}}{\mathrm{a}^{2}}\right)\left(\frac{\mathrm{a}^{2}}{4}\right)}{\sqrt{\left(\frac{\mathrm{a}}{2}-\frac{\mathrm{a}}{2}\right)^{2}+\left(\frac{\mathrm{a}}{2}-\frac{\mathrm{a}}{2}\right)^{2}+\mathrm{a}^{2}}}\right)=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{a}}\left(\frac{1}{\left.\sqrt{\frac{1}{16}+\frac{1}{16}+1}\right)}=(.9428) \frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{a}}\right.$
We could continue with this analytical procedure. However, we find that this is better left for the computer. Using MATLAB, we obtain the following numerical coefficients for the voltage vs. the number of subareas $n$. The results are presented using the above procedure along with the 'quadrature' command that is included in MATLAB.


The procedure of subdividing an area into a large number of subareas or a volume into a large number of subvolumes with the incremental charges located at the center of the subarea or the subvolume could be continued. As noted In Example 2-13, there are significant inaccuracies in the resulting numerical computation results. Therefore, we will exploit the MATLAB commands given in (2.72) in the following computation.

In particular, we focus on the following question. "If one moves far away from the object, should it not appear that the charges are localized at a point?" Remember, the sun is bigger than the earth although it appears to be a small ball in the sky. We will answer this question by examining the dependence of the

## Numerical integration

voltage as we move far away from the sheet of uniformly distributed charge. This is demonstrated with an example.

Example 2-15 ${ }^{\text {MATLAB }}$. Plot the coefficient for the electric field as a function of distance $z$ from a square that contains a uniform charge distribution of $\rho_{s}$. The $z$ axis is at the center of the square.

Answer: The electric field is computed with the quadrature command that is given in (2.72). In this case, there is an additional variable $z$.


It is convenient to understand the asymptotic values on the spatial dependence of the electric field coefficient by plotting the calculated result using a log-log graph. The slope of the dashed line is equal to $(-2)$. This is the same dependence that was obtained for the electric field from a point charge (2.8).

## Numerical integration

This example clearly illustrates that the potential will decrease at large distances and the unit square of uniform charge density will appear almost as a point charge.

Example 2-16 MATLAB . Plot the potential in the $x-y$ plane in the region $x>0$ due to a uniformly charged line that is 10 units long that is located at $y=0$. Perform the integration with the 'quad' function.

Answer: In MATLAB, a function is defined using the command 'inline'. Since the calculation is to be performed as part of a "for $x=1: 20$ " loop, the numerical value of $x$ must be converted to a string variable that can be incorporated into the inline command. This is accomplished with a command 'num2str' command. The result of the calculation is shown below.


There are other important problems in electromagnetics that will require numerical techniques that are far beyond the brief introduction that is presented here. Some of these will be described in the next chapter.

## Dielectric materials

### 2.7. Dielectric materials

Up to this point, we have examined the behavior of electric fields in a vacuum. The results were correct but we may now be wondering what the effects of applying the electric field in a material would be. The wearing of rubber gloves seems to have some desirable protective features when one is close to touching a high voltage line. Manufacturers of capacitors or integrated circuits usually insert an oxide layer between the two metal surfaces in order to keep the top conductor from falling down and touching the bottom conductor. How do these materials affect the electric field? Some answers will be provided here.


Figure 2-19. A material is placed between two electrodes that are separated by a distance $L$. An electric field is applied between the two electrodes. (a) Random orientation of the atoms before the application of the electric field.
(b) Reorientation of the atoms after the application of the electric field.

As noted earlier, materials consist of atoms and in a simple model, these atoms can be considered to be a large collection of randomly oriented small electric dipoles as shown in Figure 2-19. Certain molecules, called polar molecules normally have a permanent displacement between the positively charged nucleus at the center of the atom and the negatively charged electron at the edge. This distance is of the order of $10^{-10}$ meters. This distance is also

## Dielectric materials

equal to $1 \AA$ in honor of the scientist Anders Jonas Ångstrom. Each pair of charges acts as an electric dipole. If an electric field is externally applied to this material, then the dipoles may reorient themselves. If the field is strong enough, there will actually be an additional displacement of the positive and negative charges. A nonpolar molecule does not have this dipole arrangement of charges unless an external electric field is applied. The positive and negative charges separate by a certain distance after the application of the electric field.

In some materials, the dipoles may reorient themselves such that a large number or even all of the atoms will realign themselves causing the electric field created by the dipoles to add to the applied electric field. In other materials, the reorientation may cause the dipole electric field to subtract from the applied field. This dipole field created by the atoms will be examined here.

After the application of the electric field between the two electrodes in Figure $2-19 b$, the atoms have been reoriented. Since the atomic distances depicted in this figure, it is possible to regroup the electric dipoles and suggest that the positive charge of one atom could unite with the negative charge of the adjacent atom in order to form a new distribution of electric dipoles as depicted in Figure 2-20. This regrouping of the electric dipoles will leave a thin layer of charge of the opposite sign at either edge of the material. This charge which is due to the application of the electric field is called the polarization charge. The polarization charge cannot be found in a vacuum and it does not come out of the battery. It is only due to the fact that the atoms had been reoriented due to the application of the electric field. We will define a polarization charge density using the symbol $\rho_{\mathrm{p}}$ as being the polarization charge per unit volume.

## Dielectric materials



Figure 2-20. The reorientation of the atoms in a material due to the application of an electric field creates polarization charge at the two edges whose density is $\rho_{\mathrm{P} .}$ This polarization charge creates a polarization field $\mathbf{P}$

In the region between the two dashed lines, a positive nucleus of one atom "pairs" with an electron of the adjacent atom. The positive and negative charge centers overlap. However, in the region between the left electrode and the dashed line, there are more positively charged particles. In the region between the second dashed line and the right electrode, there are more negatively charged particles. This effectively states that there is a very narrow region of charge of one sign that has migrated to that edge of the dielectric while there is a narrow region of charge of the opposite sign that has migrated to the other edge of the dielectric. Between these two edges, a charge-neutral region exists. This displaced charge cannot be removed from the material, it is bound to the material. It is given the name of a polarization charge. Herein, we will just describe the polarization charge at the surfaces that is called the surface polarization charge. The density of this polarization charge has the symbol $\rho_{p}$ and it is shown in Figure 2-19b. This bound charge will set up a field that is called the polarization field $\mathbf{P}$ and it is defined as the dipole moment per unit volume. It is written via the relation

## Dielectric materials

$$
\begin{equation*}
\mathbf{P}=\lim _{\Delta \mathrm{v} \rightarrow 0}\left\{\frac{1}{\Delta v} \sum_{j=1}^{N} \mathbf{p}_{\mathrm{j}}\right\} \tag{2.86}
\end{equation*}
$$

where $\mathbf{p}_{\mathbf{j}}=$ Qdud $\mathbf{d}_{\mathbf{d}}$ is the dipole moment of an individual dipole. The units are $(C-m) / m^{3}=\left(C / m^{2}\right)$. Within the volume $\Delta v$, there are $N$ atoms. With the notation given in (2.86), we see that the polarization field depends on position since we have let the differential volume $\Delta v$ shrink to zero. In Figure 2-20, this would imply that the distance separating the two thin layers of polarization charge shrinks to zero. In analogy with Gauss's law, we can relate the polarization charge $\rho_{p}$ to an electric field. This field is called the polarization $\mathbf{P}$ and we write

$$
\begin{equation*}
\rho_{P}=-\nabla \bullet \mathbf{P} \tag{2.87}
\end{equation*}
$$

Let us add the polarization charge density $\rho_{\mathrm{p}}$ to the real charge density $\rho_{\mathrm{v}}$. The real charge density could come from a battery or from the ground. This will dramatically influence the resulting electric field that we calculated from (2.26)

$$
\begin{equation*}
\nabla \cdot \mathbf{E}=\frac{\rho_{v}+\rho_{\mathrm{P}}}{\varepsilon_{o}} \tag{2.88}
\end{equation*}
$$

Replacing the polarization charge density in (2.88) with (2.87), we finally obtain

$$
\begin{equation*}
\nabla \cdot \mathbf{D}=\rho_{v} \tag{2.89}
\end{equation*}
$$

where

$$
\begin{equation*}
\mathbf{D}=\varepsilon_{0} \mathbf{E}+\mathbf{P} \tag{2.90}
\end{equation*}
$$

is called the electric flux density or the displacement flux density. The unit of this quantity is also $\left(\mathrm{C} / \mathrm{m}^{2}\right)$. The total electric flux $\Psi_{\mathrm{e}}$ that passes through a surface equals the surface integral of the electric flux density integrated over the surface $\Delta \mathrm{s}$

$$
\begin{equation*}
\Psi_{\mathrm{e}}=\int_{\Delta \mathrm{s}} \mathrm{D} \bullet \mathrm{ds} \tag{2.91}
\end{equation*}
$$

Note that the displacement flux density has a significant meaning only when materials that can be polarized are discussed. In a vacuum, it is just equal to a constant $\varepsilon_{r}$ times the electric field.

## Dielectric materials

Gauss's law which was used to compute the electric field in a vacuum can be employed to calculate the displacement flux density with the same restrictive limitations of symmetry requirements that were encountered previously. The procedure to develop this equation follows directly from an integration of (2.89) over the same volume. The volume integration of the divergence of the displacement flex density can be converted to a closed surface integral using the divergence theorem. The result of this is

$$
\begin{equation*}
\oint D \cdot d s=Q_{\mathrm{enc}} \tag{2.92}
\end{equation*}
$$

Therefore, the total dielectric flux emanating from or terminating on a closed surface $\Delta \mathrm{s}$ is equal to the total charge that is enclosed within this surface.

A dielectric material is susceptible to being polarized. In many materials, this polarization is linearly proportional to the applied electric field if the electric field remains small. In these cases, we can write that $\mathbf{P}=\varepsilon_{o} X_{e} \mathbf{E}$ where $X_{e}$ is the electric susceptibility. Finally, we obtain

$$
\begin{equation*}
\mathbf{D}=\varepsilon_{\mathrm{O}}\left(1+X_{\mathrm{e}}\right) \mathbf{E}=\varepsilon_{\mathrm{o}} \varepsilon_{\mathrm{r}} \mathbf{E}=\varepsilon \mathbf{E} . \tag{2.93}
\end{equation*}
$$

The term $\varepsilon_{r}$ is the relative dielectric constant for a material. Tabulated values of $\varepsilon_{r}$ for various materials are given in Appendix 3. In a vacuum, $X_{e}=0$ and $\varepsilon_{r}=1$ by definition.

The expression (2.93) applies only for linear and isotropic materials. It is not difficult to create a material that does not satisfy this criterion. For example, the application of an external magnetic field to an ionized gas will make it anisotropic. Large amplitude signals that are applied to a material may cause the material to have a nonlinear response. This case could occur if the relative dielectric constant changed, say due to the dielectric being modified where the modification was proportional to the square of the magnitude of the applied electric field $|E|^{2}$. Such nonlinear materials do exist and are currently under

## Dielectric materials

active investigation. In what follows, we will restrict our discussion to linear materials.

Example 2-17. A dielectric slab is placed between two parallel plates. A battery is connected to one plate and the other plate is grounded. The area of each plate is equal to $A$ and the charge on each plate is $\pm Q$. The separation of the plates is d . Sketch the following quantities between the plates:
a) surface charge density $\rho_{\mathrm{s}}$,
b) displacement flux density D ,
c) electric field $E$,
d) polarization $P$, and
e) the bound surface polarization charge density $\rho_{p s}$.

Answer: a) The real charge $Q$ can come from the battery or from ground. It will be distributed on the surface of the metal plates creating a surface charge density $\rho_{s}=\frac{Q}{A}$.
b) The displacement flux density $D$ will be determined by the real charge from the battery or from the ground. It will not depend on whether a dielectric or a vacuum exists between the plates. It follows from Gauss's law that $D=\rho_{s}$.
c) The electric field $E=D /\left(\varepsilon_{0} \varepsilon_{r}\right)$. Hence the electric field will be decreased within the dielectric below its value in the vacuum since $\varepsilon_{r}>1$.
d) The polarization field $P$ will exist in the dielectric. Its value will be determined from (2.90).
e) The bound surface polarization charge density $\rho_{p s}$ can be evaluated from

$$
\rho_{P S}=\frac{Q_{p}}{A}=-\frac{P \cdot \mathbf{A}}{A}=\frac{\left(\frac{1-\varepsilon_{r}}{\varepsilon_{r}}\right) Q}{A} .
$$

## Dielectric materials



# Capacitance 

### 2.8. Capacitance

The electrical capacitance between two objects in space is defined as the ratio of the charge on one of the objects divided by the potential difference between the two objects. This is expressed as

$$
\begin{equation*}
C=\frac{Q}{V}(F) \tag{2.94}
\end{equation*}
$$

and it is measured in $(F) \equiv$ (farads) $=$ (Coulombs $/$ volts). This is the formal definition for the capacitance and as we will see in the next chapter, it is possible to numerically calculate its value for objects that may have a very complicated shape. This is sometimes called the "self capacitance" since there are only two objects that are being considered. It could also be generalized to incorporate the "mutual capacitance" if there are more than two objects. The procedure will be useful in practical situations such as finding the capacitance of various portions of an integrated circuit or of objects that have an odd shape. This will also be useful in developing various models for a transmission line which will also be described later. In this section, we'll just examine some very simple objects.

Your first encounter with a capacitor was probably in the first course dealing with electrical circuits in which you encountered a very simple expression for the capacitance of the parallel plate capacitor. In this book, we will obtain this expression using the terms that we have already discussed. The area of each plate is equal to $A=w x \Delta z$ and the two plates are separated by a distance $d$ as shown in Figure 2-21. The choice for using these symbols for the dimensions is predicated on our future applications in this book. In addition, let us assume that

## Capacitance

a charge $+Q$ is uniformly distributed on the top plate and a charge $-Q$ is uniformly distributed on the bottom plate. This will result in a uniform charge density of $\pm \rho_{\mathrm{s}}=\frac{\mathrm{Q}}{\mathrm{A}}$ being distributed on the two plates.


Figure 2-21. A parallel plate capacitor is depicted above. The plates whose area $A=w x \Delta z$ are separated by a distance $d$.

We assume that the transverse dimensions are much greater than the distance between the two plates. The electric field surrounding an infinite charged plate was obtained in Example 2-6 or we could obtain it using Gauss's law. We evaluate Gauss's law (2.24) in order to obtain the electric field from one of the plates, say the top plate that has a positive charge density

$$
\begin{equation*}
\oint E \cdot d s=\frac{Q_{\mathrm{enc}}}{\varepsilon_{\mathrm{o}}} \Rightarrow E(2 A)=\frac{\rho_{\mathrm{s}} A}{\varepsilon_{\mathrm{o}}} \Rightarrow E=\frac{\rho_{\mathrm{s}}}{2 \varepsilon_{\mathrm{o}}} \tag{2.95}
\end{equation*}
$$

The electric field between the two plates and in the regions above and below the two plates is evaluated using the principle of superposition. The result is that the electric field is equal to 0 in the regions above and below the two plates since the contributions from the two plates have the same magnitude but are in the opposite directions and will therefore cancel. The field in the external

## Capacitance

region but between the two plates is called a "fringing field" and it will be neglected since it is small in comparison with the field directly between the two plates. Using superposition, the electric field between the two plates becomes

$$
\begin{equation*}
\mathrm{E}=\frac{\rho_{\mathrm{s}}}{\varepsilon_{\mathrm{o}}} \tag{2.96}
\end{equation*}
$$

In addition to neglecting the fringing field, we are also assuming that there is a uniform distribution of the charge on the surface. The actual distribution of the charge on the plate is actually nonuniform and it will be numerically calculated in the next chapter.

The electric potential is calculated using (2.54)

$$
\begin{equation*}
\mathrm{V}_{\mathrm{ab}}=\int_{\mathrm{a}}^{\mathrm{b}} \mathrm{E} \bullet \mathrm{dl} \Rightarrow \mathrm{Ed} \tag{2.97}
\end{equation*}
$$

Therefore, the capacitance of the parallel plate capacitor is calculated using the definition (2.94) to be

$$
\begin{equation*}
C=\frac{Q}{V}=\frac{\rho_{\mathrm{s}} A}{\left(\frac{\rho_{\mathrm{s}}}{\varepsilon_{0}} d\right)}=\frac{\varepsilon_{0} A}{d} \tag{2.98}
\end{equation*}
$$

Equation (2.98) is the formula that you have frequently encountered previously but it has now been derived in terms of electromagnetic arguments. Most capacitors will have a dielectric placed between the two conducting plates. In these cases, you should replace $\varepsilon_{o}$ with $\varepsilon$ in order to reflect this situation.

Example 2-18. Calculate the capacitance of a hollow metallic sphere whose radius is a.

Answer: Assume that there is a charge $Q$ at the center of the sphere. The absolute potential at a radius a is found from (2.56) to be

# Capacitance 

$$
\mathrm{V}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{a}}
$$

The capacitance is computed from (2.94) to be

$$
\mathrm{C}=\frac{\mathrm{Q}}{\mathrm{~V}}=\frac{\mathrm{Q}}{\left(\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{a}}\right)}=4 \pi \varepsilon_{0} \mathrm{a}
$$

It is interesting to calculate the capacitance of the earth by assuming that it is a hollow sphere. Substituting the value of the radius of the Earth into this result, we compute

$$
\mathrm{C}=4 \pi \varepsilon_{0} \mathrm{a}=4 \pi\left(\frac{1}{36 \pi} \times 10^{-9}\right)\left(6.37 \times 10^{6}\right) \approx 7 \times 10^{-4} \mathrm{~F}=0.7 \mathrm{mF}
$$

A unit of one farad is a very big number!
Hollow spheres are used as models to describe dust particles that can be found in integrated circuit manufacturing where they have a very deleterious effect on the final product. Upwards of 10,000 mobile electrons can attach themselves to these dust particles. In addition, charge dust particles are found in some of the rings that surround certain of the planets in our solar system such as Saturn. These negatively charged dust particles may have a mass that is greater than the surrounding positively charged ions and this leads to a current area of research investigation.

Example 2-19. Calculate the capacitance of a coaxial cable whose length is $\Delta z$ and it consists of a cylindrical metallic rod whose radius is a and it is surrounded concentrically with a metallic sleeve whose radius is b . There is a dielectric
material separating the two conducting surfaces and it has a relative dielectric constant $\varepsilon_{r}$.


Answer: The displacement flux density between the two metallic surfaces can be calculated using Gauss's law (2.92) since there is significant symmetry in this example. The procedure is to assume that there is a linear charge density $\rho_{\mathrm{L}}$ on the inner conductor. From (2.92), we calculate the displacement flux density as

$$
\oint D \cdot d \mathbf{s}=Q_{\mathrm{enc}} \rightarrow \mathrm{D}(2 \pi \mathrm{r} \Delta \mathrm{z})=\rho_{\mathrm{L}} \Delta \mathrm{z}
$$

The potential difference between the inner conductor and the outer conductor is computed from (2.54)

$$
\mathrm{V}_{\mathrm{ab}}=\int_{\mathrm{a}}^{\mathrm{b}} \mathrm{E} \cdot \mathrm{dl}=\int_{8}^{\mathrm{b}} \frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon \mathrm{r}} \mathrm{dr}=\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon} \ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right)
$$

The total charge that is enclosed within the coaxial structure is $Q=\rho_{\llcorner } \Delta z$.
From (2.94), we write

$$
\mathrm{C}=\frac{\mathrm{Q}}{\mathrm{~V}_{\mathrm{ab}}}=\frac{\rho_{\mathrm{L}} \Delta \mathrm{z}}{\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon} \ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right)}=\frac{2 \pi \varepsilon}{\ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right)} \Delta \mathrm{z}
$$

## Capacitance

In addition to calculating the capacitance of a parallel plate capacitor, we can also find the electrostatic energy that is stored in this capacitor. In order to do this, we just have to evaluate the integral (2.67) which we rewrite here

$$
\begin{equation*}
W_{e}=\frac{\varepsilon_{\mathrm{o}}}{2} \int_{\Delta v} \mathrm{E}^{2} \mathrm{dv} \tag{2.99}
\end{equation*}
$$

Equation (2.99) is a very general expression for the electrostatic energy that is stored between the two objects. In our particular case, the volume $\Delta v$ is equal to the volume between the two parallel plates of the capacitor. Since the electric field is uniform between the two plates, it is possible to easily calculate the stored electrostatic energy. We write

$$
\begin{equation*}
\mathrm{W}_{\mathrm{e}}=\frac{\varepsilon_{\mathrm{o}}}{2}\left(\frac{\mathrm{~V}}{\mathrm{~d}}\right)^{2}(\mathrm{Ad})=\frac{1}{2}\left(\frac{\varepsilon_{\mathrm{o}} \mathrm{~A}}{\mathrm{~d}}\right) \mathrm{V}^{2}=\frac{1}{2} \mathrm{CV}^{2} \tag{2.100}
\end{equation*}
$$

In writing the final expression in (2.100), we have recognized that the capacitance $C$ of a parallel plate capacitor (2.98) can be identified. Therefore, we have obtained the electrostatic energy that is stored between the two parallel plates.

Further calculations involving the capacitance between conducting surfaces will be performed using numerical techniques in the next chapter. In addition, the capacitance will also be obtained for other important structures that will be encountered in the later discussion of transmission lines. The insertion of a dielectric slab that does not completely fill the intervening space between the two parallel plates will require considerable care in the analysis. This calculation will be delayed until after the boundary conditions are examined.

## Electrical currents

### 2.9. Electrical currents

Imagine that a wire is connected to a battery and a resistor as shown in Figure 2-22a. The battery is a chemical source that provides particles with a positive charge and a negative charge. In the metallic wire, the ions are stationary and a portion of the electrons are free to move. Benjamin Franklin gave us the convention that the direction of the flow of the current, however should be in the direction of the motion of the positive particles. Between the two terminals of the battery, there will be an electric field that will accelerate the electrons. Since the conductivity of a wire is significantly greater than the conductivity of the surrounding air, the motion of these accelerated electrons will follow the path of the wire. The resulting current in this circuit is called the "conduction current." When we later describe electric fields that depend upon time, we will encounter another current that is called a "displacement current."

The current that passes through the wire in Figure 2-22b can be computed from Ohm's law. Since we are more interested in local effects rather than in global effects, we would rather relate this current to a current density $J=\frac{1}{A}$ and a voltage difference $\Delta V$ across an incremental length $\Delta L$, i.e. an electric field. The resistance $R$ of the wire is given by $R=\frac{\Delta L}{\sigma A}$ where $\sigma$ is the conductivity of the wire. This definition of resistance follows from the intuition gained in circuits where the total resistance of a circuit is computed by adding the resistors in series (i. e. increased length of the wire $\Delta \mathrm{L}$ ) and the conductors in parallel (i.e. increased cross-sectional area A). The current density in a wire is therefore defined using Ohm's law

$$
\begin{equation*}
J=\frac{1}{\mathrm{~A}}=\frac{\left(\frac{\Delta \mathrm{V}}{\mathrm{R}}\right)}{\mathrm{A}}=\frac{\Delta \mathrm{V}}{\Delta \mathrm{~L}} \frac{\Delta \mathrm{~L}}{\mathrm{RA}}=\sigma \mathrm{E} \tag{2.101}
\end{equation*}
$$

## Electrical currents

This is a generalization of Ohm's law. One Ampere of current at a point is defined as the passage of one Coulomb of charge passing this point in one second. We are assuming that none of the parameters that appear in (2.101) depend upon the magnitude of any of the other parameters which could be found in nonlinear materials. One could think of a resistor whose resistance would change with increasing values of current caused by a heating of the resistor. Such nonlinear effects are important in practice but will not be considered here.

An alternative derivation for the conductivity $\sigma$ follows from the definition of the current density $\mathrm{J}=\rho_{\mathrm{V}} \mathrm{V}_{\text {drift }}$ where $\rho_{\mathrm{v}}$ is the electron volume charge density and $v_{\text {drift }}$ is an average electron drift velocity. The drift velocity is proportional to the electric field $E$, the proportionality constant being called the mobility $\mu$ of the material. Hence, we write

$$
\begin{equation*}
J=\rho_{v} v_{\text {drift }}=\rho_{v} \mu E=\sigma E \tag{2.102}
\end{equation*}
$$

where the conductivity $\sigma=\rho_{\mathrm{v}} \mu$.


Figure 2-22. (a) A simple electrical circuit consisting of a battery and a resistor. (b) A current I flows through the wire whose area is equal to $A=\pi a^{2}$ where $a$ is the radius of the wire.

The total current I that passes through the wire is computed from the integral of the current density $\mathbf{J}$ integrated over the cross sectional area $A$ of the

## Electrical currents

wire. The current density $\mathbf{J}$ is a vector since it has both a magnitude and a direction

$$
\begin{equation*}
I=\int_{A} J \cdot d s \tag{2.103}
\end{equation*}
$$

If the current is uniformly distributed in a cylindrical wire whose radius is a, this integral can be easily performed and we find that the total current I that passes through the wire is given by

$$
\begin{equation*}
\mathrm{I}=\mathrm{J} \pi \mathrm{a}^{2} \tag{2.104}
\end{equation*}
$$

This is equal to the product of the current density times the cross-sectional area of the wire. However, if the current is nonuniformly distributed in the wire, this integration requires more care as will be shown with an example. Later, we will encounter cases where the current is constrained to flow just on the surface of an object. These currents are called surface currents. Having now presented some fundamental definitions for the currents in terms of local current densities and cross sectional areas, we are prepared to explore various properties of magnetic fields that will be created by these currents.

Example 2-20. Given an inhomogeneous current density

$$
\mathbf{J}=\left(3 y^{2} \mathbf{z} \mathbf{u}_{\mathbf{x}}-2 x^{3} \mathbf{z} \mathbf{u}_{\mathbf{y}}+\mathbf{z} \mathbf{u}_{\mathbf{z}}\right) \mathrm{A} / \mathrm{m}^{2}
$$

find the total current passing through a square surface at $x=1$ in the $\mathbf{u}_{\mathbf{x}}$ direction.
The dimensions of the surface are $1 \leq y \leq 2,1 \leq z \leq 2$.
Answer: The differential surface area is defined as $\mathbf{d s}=d y d z \mathbf{u}_{\mathbf{x}}$. Therefore, the scalar product on the current density and the current density will yield a current only in the $\mathbf{u}_{\mathbf{x}}$ direction. We write

$$
I=\int_{A} J \bullet d \mathbf{s}=\int_{y=1}^{y=2} \int_{z=1}^{z=2} 3 y^{2} z d y d z=\left.\frac{3 y^{3} z^{2}}{6}\right|_{y=1, z=1} ^{y=2, z=2}=\frac{96-3}{6}=15.5 \mathrm{~A} .
$$

# Electrical currents 



Example 2-21. Calculate the current that flows through a wire whose radius is a.
The inhomogeneous current density in the wire is

$$
\mathbf{J}=I_{0}\left(\frac{r}{a}\right) \mathbf{u}_{\mathbf{z}}
$$

Nonuniform currents can be important in high frequency applications in which one encounters "skin effects."


Answer: The current is calculated from the following integration:

$$
I=\int_{\phi=0}^{\phi=2 \pi} \int_{r=0}^{r=a} I_{0}\left(\frac{r}{a}\right) \mathbf{u}_{z} \bullet r d r d \phi \mathbf{u}_{z}=I_{o}(2 \pi)\left(\frac{a^{2}}{3}\right) .
$$

## Electrical currents

The power that is dissipated within a conducting material can be calculated from the electric field and the current density that we have just encountered. The power density in a particular volume is defined as

$$
\begin{equation*}
\mathrm{p}=\mathrm{J} \cdot \mathrm{E} \quad\left(\mathrm{~W} / \mathrm{m}^{3}\right) \tag{2.105}
\end{equation*}
$$

The total power that is absorbed within the volume is calculated by integrating (2.105) over the entire volume $\Delta \mathrm{v}$. This power is converted into another form and it is given the name "Joule heating." It is measured in SI in units
Watts = Joules / seconds.

The reader has probably experienced the warming effects of Joule heating in cooking a meal on an electric stove or being warmed on a cold winter's night by an electric heater that is present in the room. Suffice it to say, this is a very important effect that has many practical applications.

Example 2-22. Calculate the power that is dissipated within a resistor that has a uniform conductivity $\sigma$. The voltage between the two ends of the resistor is V and a current I passes through the resistor.


Answer: From (2.105), we obtain the power density. The total power loss is calculated from the integration of the power density over the volume

$$
\mathrm{P}=\int_{\Delta \mathrm{V}} \mathrm{~J} \bullet \mathbf{E} d \mathrm{v}=\int_{\mathrm{z}=0}^{\mathrm{z}=\mathrm{L}} \int_{\phi=0}^{\phi=2 \pi} \int_{\mathrm{r}=0}^{r=a}\left(\frac{\mathrm{I}}{\pi \mathrm{a}^{2}}\right)\left(\frac{\mathrm{V}}{\mathrm{~L}}\right) r d r \mathrm{~d} \phi \mathrm{dz}=\mathrm{VI} .
$$

## Fundamentals of magnetic fields

### 2.10. Fundamentals of magnetic fields

The effects of magnetic fields were known for almost three millennia when it was discovered that certain stones would attract iron. A large deposit of these stones that are called "lodestones" was found in the district of Magnesia in Asia Minor. This mineral later became known as magnetite $\left[\mathrm{Fe}_{3} \mathrm{O}_{4}\right.$ ] and it had some interesting properties. Early navigators used its north- and south- seeking characteristics in their early explorations. The first scientific study of magnetism was written in 1600 by William Gilbert. Little else was known about it until the early nineteenth century when Hans Christian Oersted discovered that an electric current in a wire affected a magnetic compass needle. This work together with the later work of Ampere, Gauss, Henry, Faraday and others raised the magnetic field to equal partner status with the electric field. This elevation in stature was confirmed with the theoretical work of Maxwell.

In studying electric fields, we found that electric charges could be separated from each other such that a positive charge existed independently from a negative charge. Would the same separation of magnetic poles exist? Would it be possible to cut the earth at the equator and send one-half containing one of the pole faces (a "magnetic monopole") to a far off region and never see that pole again? As of the date of the writing of this text, a magnetic monopole has not yet been unambiguously observed in nature. Several observatories have searched for these elusive entities and in one five year period, only one momentary deflection of a needle on a satellite had been recorded throughout the world. Whether this deflection was due to a real event or due to some anomaly in the

## Fundamentals of magnetic fields

detector is open to question since no confirming events have ever been detected. The scientist who reported the particular meter deflection later stated that the deflection was just an anomalous result. Patience seems to be wearing thin and observatories have ended their campaign of "monopole sighting." Therefore, we see that there is a major difference between magnetic fields and electric fields.

Since the magnetic monopole has not been observed to exist in nature, we find that the magnetic field lines are continuous and do not originate nor terminate at a point. The total magnetic flux in a region is usually denoted with a symbol $\Psi_{m}$ and the units are Webers. Enclosing an arbitrary point with a closed surface, we can mathematically express this fact by stating that

$$
\begin{equation*}
\oint B \cdot d s=0 \tag{2.106}
\end{equation*}
$$

The term $\mathbf{B}$ that appears in this equation is the magnetic flux density. The SI units of magnetic flux density are given in Tesla (T). We may also see the equivalent unit of (Weber $/$ meter $^{2}$ ) where $1 \mathrm{~T}=1 \mathrm{~Wb} / \mathrm{m}^{2}$. A magnetic flux density of one Tesla is a very large value. For example, the equatorial magnetic field strength at sea level of the Earth is approximately $0.5 \times 10^{-4}$ Tesla. Another commonly employed unit for the magnetic flux density is the Gauss where 1 Gauss $=10^{-4}$ Tesla .

We can also write (2.106) in differential form by making use of the divergence theorem that relates a closed surface integral to a volume integral $\oint \mathbf{B} \bullet \mathbf{d s}=\int_{\Delta v} \nabla \bullet \mathbf{B} d v$. In order for the closed surface integral be equal to 0 for any arbitrary volume $\Delta v$, the integrand must be identically equal to 0 . Therefore, we write

## Fundamentals of magnetic fields

$$
\begin{equation*}
\nabla \bullet \mathbf{B}=0 \tag{2.107}
\end{equation*}
$$

Equations (2.106) and (2.107) express the fact that the magnetic field closes upon itself and the field and does not terminate on nor originate from an isolated magnetic monopole. If we cut a bar magnet in half with the hope of isolating one of the poles, we will just end up with two bar magnets of smaller physical size - see Figure 2-23. We can contrast the first postulate of magnetic fields with the first postulate of electrostatic fields $\nabla \bullet E=\frac{\rho_{v}}{\varepsilon_{0}}$. Recall that that it was possible to physically separate a positive charge from a negative charge.


Figure 2-23. The physical cutting of a large permanent magnet creates a number of smaller permanent magnets. It is impossible to separate the North Pole from the South Pole of a magnet.

The second property of steady magnetic fields was discovered by Hans Christian Oersted in 1820. He observed that compass needles were deflected when an electrical current flowed through a nearby wire and concluded that the

## Fundamentals of magnetic fields

effect was due to the creation of a magnetic field by this current. Recall our earlier allusion to the model of an atom that consisted of an electron circulating about a positive nucleus. It was modeled as an electric dipole. The moving electron can also be interpreted as being a current. Therefore, the atom can be thought of as being a small electrical dipole and a small magnetic dipole.


Figure 2-24. A cylindrical wire carries a current I that creates a magnetic field whose density is $\mathbf{B}$. If the thumb of the right hand points in the direction of the current, then the fingers follow the magnetic field.

The results of this experiment can be described with the following equation

$$
\begin{equation*}
\oint \mathbf{B} \cdot \mathbf{d} \mathbf{l}=\mu_{0} \mathbf{l}_{\mathrm{enc}} \tag{2.108}
\end{equation*}
$$

where the total current that is enclosed within the closed loop is specified as $l_{\text {enc }}$. The constant $\mu_{\mathrm{o}}$ is the permeability of free space. In SI units $\mu_{0}$ is defined to have the numerical value of

$$
\begin{equation*}
\mu_{\circ} \equiv 4 \pi \times 10^{-7} \frac{\text { henries }}{\text { meter }} \equiv \frac{\mathrm{H}}{\mathrm{~m}} \tag{2.109}
\end{equation*}
$$

The magnetic flux density is perpendicular to the current density and follows the "right hand rule" convention in that the thumb of the right hand is in the direction

## Fundamentals of magnetic fields

of the current and the fingers are in the direction of the magnetic flux density as shown in Figure 2-24.

With this value for the permeability of free space and the value for the permittivity of free space that we approximated earlier in (2.6) to be

$$
\begin{equation*}
\varepsilon_{o} \approx \frac{1}{36 \pi} \times 10^{-9} \frac{\mathrm{~F}}{\mathrm{~m}} \tag{2.110}
\end{equation*}
$$

we may be intrigued by the numerical value that is computed from the expression $\frac{1}{\sqrt{\mu_{0} \varepsilon_{0}}}$. Yes, this number has the same numerical value as the velocity of light. It also has the dimensions of a velocity but the demonstration of this will be discussed later. We will also see later that this coincidence is more than fortuitous and it will lead to something very fundamental.

Equation (2.108) is called Ampere's circuital law or Ampere's law and as we will see, it allows us to calculate magnetic flux density in many cases where there is considerable symmetry. Several examples will be described in full detail in the following discussion since it is one of the fundamental methods of calculating the magnetic flux density caused by a current.

Equation (2.108) can also be written in differential form but this will require a vector operation. The left-hand side of (2.108) can be converted to a surface integral using Stokes's theorem

$$
\begin{equation*}
\oint \mathbf{B} \cdot \mathbf{d} \mathbf{l}=\int_{\Delta s} \nabla \times \mathbf{B} \cdot \mathbf{d s} \tag{2.111}
\end{equation*}
$$

Therefore, after equating the two surface integrals in (2.108) and (2.111), we obtain

## Fundamentals of magnetic fields

$$
\begin{equation*}
\int_{\Delta \mathrm{s}} \nabla \times \mathbf{B} \bullet \mathbf{d s}=\mu_{\circ} \int_{\Delta \mathrm{s}}^{\mathbf{J}} \bullet \mathbf{d s} \tag{2.112}
\end{equation*}
$$

In order for these two integrals to be equal for any arbitrary surface, the integrands must be equal. This leads to

$$
\begin{equation*}
\nabla \times \mathbf{B}=\mu_{\mathrm{o}} \mathbf{J} \tag{2.113}
\end{equation*}
$$

This is the differential form of Ampere's law. The integral form and the differential form of Ampere's law are used in electromagnetic calculations.

Let us determine the magnetic field using (2.108) for the wire shown in Figure 2.24. It will be assumed that the wire is straight, it has no bends or kinks in it and it is of infinite length. In this case, the magnetic field will be entirely in the $\mathbf{u}_{\phi}$ direction. We will follow the right-hand rule convention that is standard for determining the direction of the magnetic field in that if the current is pointing in the direction of the thumb of the right hand, then the magnetic flux density will be in the direction of the fingers.

The line integral in (2.108) can be easily evaluated if there is symmetry in the system. For the particular case of the wire shown in Figure 2-25 where significant cylindrical symmetry abounds, the path of the integration will follow a circle of radius $r$ and the integral is written as

$$
\begin{equation*}
\oint \mathbf{B} \cdot \mathbf{d} \mathbf{I}=\int_{0}^{2 \pi} \mathrm{~B}_{\phi} \mathbf{u}_{\varphi} \bullet \mathrm{rd} \phi \mathbf{u}_{\varphi}=2 \pi \mathrm{rB}_{\phi} \tag{2.114}
\end{equation*}
$$

The surface integral on the right side of $(2.108)$ can also be evaluated. It is equal to the $\mu_{0}$ times the current that is enclosed within the surface defined by the same radius $r$ given in (2.114). For the moment, we will take the radius $r$ to be greater than the radius of the wire a so the total current I that flows in the wire is

## Fundamentals of magnetic fields

enclosed within the surface. Hence the integral on the right side of (2.108) is just $\mu_{\mathrm{o}}$. The magnetic flux density external to the wire is computed to be

$$
\begin{equation*}
B_{\phi}=\frac{\mu_{0} I}{2 \pi r} \tag{2.115}
\end{equation*}
$$



Figure 2-25. A current I passes through a cylindrical conductor. The magnetic flux density is to be computed for all values of the radius $r$.

Example 2-23 ${ }^{\text {MATLAB }}$. Plot the magnetic flux density in the regions that are internal to the wire and external to the wire that carries a current out of the paper. Answer: The magnetic field is calculated using (2.115) and is illustrated in the following figure. The length of the arrows is proportional to the magnetic field strength.

## Fundamentals of magnetic fields



Using the same technique, we can also calculate the magnetic flux density within the wire. The procedure is the same as was used to calculate the field external to the wire; we first determine the current that is enclosed within a circle of radius r . If the current is uniformly distributed in the wire, the current density that flows through the wire is $J_{z}=\frac{1}{\pi a^{2}}$. The total current that is enclosed within the circle whose radius is $r$ is given by

$$
\begin{equation*}
\int_{\Delta \mathrm{s}} \mathrm{~J} \bullet \mathbf{d s}=\int_{\phi=0}^{\phi=2 \pi} \int_{r=0}^{r^{\prime r}=r}\left(\frac{\mathrm{l}}{\pi \mathrm{a}^{2}}\right) \mathrm{r}^{\prime} d \mathrm{r}^{\prime} d \phi=\left(\frac{\mathrm{r}}{\mathrm{a}}\right)^{2}, \tag{2.116}
\end{equation*}
$$

## Fundamentals of magnetic fields

where $r^{\prime}$ is a dummy variable of integration. Equating this enclosed current multiplied by $\mu_{o}$ with the expression given in (2.114), we finally obtain the magnetic flux density within the wire to be

$$
\begin{equation*}
\mathrm{B}_{\phi}=\frac{1}{2 \pi \mathrm{r}}\left[\mu_{0}\left(\frac{\mathrm{r}}{\mathrm{a}}\right)^{2} \mathrm{I}\right]=\frac{\mu_{0} \mathrm{l}}{2 \pi}\left(\frac{\mathrm{r}}{\mathrm{a}^{2}}\right) \tag{2.117}
\end{equation*}
$$

At the edge of the wire $(r=a)$, the two solutions given by (2.115) and (2.117) agree as they must. The magnetic flux density is shown in Figure 2-26. At the origin $(r=0)$, the magnetic flux density is equal to zero since no current is enclosed within a circle whose radius is equal to zero. As more current is enclosed, the field increases until all of the current is enclosed at the edge ( $r=a$ ). For radii greater than the radius of the wire ( $r>a$ ), no additional current is enclosed and the field decays geometrically as $\frac{1}{\mathrm{r}}$.



Figure 2-26. The radial dependence of the magnetic flux density as calculated in (2-115) and (2-117) is shown. There is a homogeneous current distribution within the wire.

## Fundamentals of magnetic fields

Example 2-24. The center conductor of a coaxial cable carries a current $\mathrm{I}_{\mathrm{O}}$ in the $+\mathbf{u}_{\mathbf{z}}$ direction (out of the page) and this current returns in the outer conductor. Calculate the magnetic flux density both within the coaxial cable and in the region external to the outer conductor.


Answer: We will apply Ampere' law (2.108) separately to each of the regions in the coaxial cable. Due to symmetry, the left hand side of (2.108) will always have the value given in (2.114): $2 \pi r B_{\phi}$. Hence, we write $0 \leq r \leq a$

$$
\begin{gathered}
2 \pi \mathrm{~B}_{\phi}=\mu_{\mathrm{o}} \int_{\Delta \mathrm{s}} \mathbf{J} \bullet \mathbf{d s}=\mu_{\circ} \int_{\phi=0}^{\phi=2 \pi} \int_{r^{\prime}=0}^{r^{\prime}=r}\left(\frac{\mathrm{I}_{0}}{\pi \mathrm{a}^{2}}\right) \mathrm{r}^{\prime} \mathrm{dr} r^{\prime} \mathrm{d} \phi=\mu_{0} \mathrm{I}_{\mathrm{o}}\left(\frac{\mathrm{r}}{\mathrm{a}}\right)^{2} \\
\mathrm{~B}_{\phi}=\frac{\mu_{0} \mathrm{I}_{\mathrm{O}}}{2 \pi r}\left(\frac{\mathrm{r}}{\mathrm{a}}\right)^{2}
\end{gathered}
$$

$a \leq r \leq b$

$$
\begin{gathered}
2 \pi \mathrm{rB}{ }_{\phi}=\mu_{0} \int_{\Delta \mathrm{s}} \mathbf{J} \bullet \mathbf{d s}=\mu_{0} \mathrm{I}_{0} \\
\mathrm{~B}_{\phi}=\frac{\mu_{0} \mathrm{l}_{0}}{2 \pi \mathrm{r}}
\end{gathered}
$$

## Fundamentals of magnetic fields

$b \leq r \leq c$

$$
\begin{gathered}
2 \pi \mathrm{rB} \mathrm{\phi}_{\phi}=\mu_{\circ} \int_{\Delta s} J_{\Delta} \bullet \mathbf{d s}=\mu_{0} \mathrm{I}_{0}-\mu_{\circ} \int_{\phi=0}^{\phi=2 \pi} \int_{r^{\prime}=\mathrm{b}}^{r^{\prime}=r}\left(\frac{\mathrm{I}_{0}}{\pi\left(\mathrm{c}^{2}-\mathrm{b}^{2}\right)}\right) \mathrm{r}^{\prime} d r^{\prime} d \phi=\mu_{0} \mathrm{I}_{0}\left(1-\frac{\left(\mathrm{r}^{2}-\mathrm{b}^{2}\right)}{\left(\mathrm{c}^{2}-\mathrm{b}^{2}\right)}\right) \\
\mathrm{B}_{\phi}=\frac{\mu_{0} \mathrm{I}_{\circ}}{2 \pi r}\left(1-\frac{\left(\mathrm{r}^{2}-\mathrm{b}^{2}\right)}{\left(\mathrm{c}^{2}-\mathrm{b}^{2}\right)}\right)
\end{gathered}
$$

$r>c$

$$
\begin{gathered}
2 \pi \mathrm{rB}_{\phi}=\mu_{\mathrm{o}} \int_{\Delta \mathrm{s}} \mathrm{~J} \cdot \mathbf{d s}=0 \\
\mathrm{~B}_{\phi}=0
\end{gathered}
$$

Example 2-25. Use Ampere's law to find the magnetic field of a solenoid. A solenoid is constructed by uniformly winding wire around a cylindrical form such as a broomstick. There are N turns of wire in the length of the solenoid. Assume that the length $d$ is much greater than its radius a.


Answer: Apply Ampere's law (2.108) to the loop that encloses the current that is coming out of the paper. There will be four terms [ $1 \rightarrow 2,2 \rightarrow 3,3 \rightarrow 4,4 \rightarrow 1$ ] that will contribute to the line integral defined in (2.108). However, we have assumed that dimensions of the solenoid satisfied the relation that $\mathrm{d} \gg \mathrm{a}$. This assumption will allow us to neglect any "fringing fields" at the two ends. Therefore, only two of the integrals [1 $\rightarrow 2$ and $3 \rightarrow 4$ ] will contribute to our

## Fundamentals of magnetic fields

solution. This approximation implies that there is no component of magnetic field in the radial direction. The closed line integral yields

$$
\oint B \cdot d \approx B_{z}(2 d)
$$

The surface integral gives us the current that is enclosed within the loop. Our use of the approximately equal sign notation can be replaced with an equal sign if the current that is enclosed within the loop were an infinite current carrying slab out of the paper. Then the integral from $1 \rightarrow 2$ would exactly equal the integral $3 \rightarrow 4$. Since there are N wires each carrying a current I into the paper, the surface integral yields

$$
\mu_{0} \int_{\Delta \mathbf{s}} \mathbf{J} \bullet \mathbf{d s}=\mu_{0} \mathrm{Nl}
$$

Equating these two expressions, we find that the current enclosed within the this loop contributes a magnetic flux density beneath the coil that also encompasses the $z$ axis of

$$
\mathrm{B}_{\mathrm{z}}=\frac{\mu_{\mathrm{o}} \mathrm{NI}}{2 \mathrm{~d}}
$$

We note that this magnetic field is independent of the vertical distance.
Following the same procedure for the current that is going into the paper in the top portion of the solenoid, we find that the magnetic flux density has the same magnitude and direction. Applying the principle of superposition, we find that the two fields add in the center of the solenoid and subtract in the external regions.

$$
\mathrm{B}_{\mathrm{z}}=\frac{\mu_{0} \mathrm{NI}}{\mathrm{~d}}
$$

## Fundamentals of magnetic fields

A more accurate calculation shows that the internal magnetic given above is exact in the center of the solenoid and its value differs by $50 \%$ at the ends where symmetry disappears.

Example 2-26. Find the magnetic flux density within a toroid, whose crosssectional area is A . A toroid consists of N turns of wire uniformly wrapped around a torus.


Answer: Within a mean circumference whose length is L , there are NI amperes of current entering the paper. From Ampere's law (2.108), we write

$$
\mathrm{B}_{\phi} \mathrm{L} \approx \mu_{\mathrm{O}} \mathrm{NI}
$$

which yields

$$
\mathrm{B}_{\phi}=\frac{\mu_{0} \mathrm{NI}}{\mathrm{~L}}
$$

In order to emphasize this point that the magnetic flux density surrounding a current carrying region depends only on the current that is enclosed within the region, we will cast Ampere's law in Lagrangian mass variables as we had

## Fundamentals of magnetic fields

previously cast Gauss's law for electrostatics. The current that is enclosed within a radius $r$ is given by

$$
\begin{equation*}
I_{\text {enc }}=\int_{\Delta s} J \bullet d \mathbf{d s}=\int_{\phi=0}^{\phi=2 \pi} \int_{r^{\prime}=0}^{r^{\prime}=r} J_{z}\left(r^{\prime}\right) r^{\prime} d r^{\prime} d \phi=2 \pi \int_{r^{\prime}=0}^{r^{\prime}=r} J_{z}\left(r^{\prime}\right) r^{\prime} d r^{\prime} \tag{2.118}
\end{equation*}
$$

In (2.118), $\mathrm{r}^{\prime}$ is the variable of integration and we have assumed that the current that depends only on the radial coordinate. The current is flowing in the $z$ direction. The integration over the angular variable $\phi$ yielded the factor of $2 \pi$. The differential current $\mathrm{dl}_{\mathrm{enc}}$ is given by

$$
\begin{equation*}
d l_{\mathrm{enc}}=2 \pi r J_{\mathrm{z}}(r) \mathrm{dr} \tag{2.119}
\end{equation*}
$$

We are treating the case where the magnetic flux density depends only on the coordinate $r$ and the magnetic flux density is directed in the $\mathbf{u}_{\phi}$ direction. Hence (2.113) becomes

$$
\begin{equation*}
\frac{\mathrm{dB}_{\phi}}{\mathrm{dr}}=\mu_{\mathrm{o}} \mathrm{~J}_{\mathrm{z}}(\mathrm{r}) \tag{2.120}
\end{equation*}
$$

Applying the chain rule to (2.120) and using (2.119), we write

$$
\frac{\mathrm{dB}_{\phi}}{\mathrm{dr}}=\frac{\mathrm{dB}_{\phi}}{\mathrm{dl}_{\mathrm{enc}}} \frac{\mathrm{dl}_{\text {enc }}}{\mathrm{dr}}=\frac{\mathrm{dB}_{\phi}}{\mathrm{dl}_{\mathrm{enc}}}\left(2 \pi \mathrm{rJ}_{\mathrm{z}}(\mathrm{r})\right)=\mu_{\mathrm{o}} \mathrm{~J}_{\mathrm{z}}(\mathrm{r})
$$

or

$$
\begin{equation*}
\frac{\mathrm{dB}_{\phi}}{\mathrm{dl}_{\mathrm{enc}}}=\frac{\mu_{\mathrm{o}}}{2 \pi \mathrm{r}} \tag{2.121}
\end{equation*}
$$

where the explicit dependence on the current density $J_{Z}(r)$ has disappeared. The integral of (2.121) is explicitly stated as

$$
\begin{equation*}
\mathrm{B}_{\phi}=\frac{\mu_{\mathrm{o}} \mathrm{I}_{\mathrm{enc}}}{2 \pi \mathrm{r}} \tag{2.122}
\end{equation*}
$$

## Fundamentals of magnetic fields

The application of (2.122) to the Figure 2-25 leads to the following results. The entire current $I$ is enclosed within a circle whose radius $r>a$. Therefore, we obtain the same result as given in (2.115). The fraction of the current I that is enclosed within a circle whose radius $r<a$ is $I_{\text {enc }}=\left(\frac{r}{a}\right)^{2} I$. Therefore, we obtain the same result has given in (2.117). This calculation provides an additional interpretation of Ampere's law in that the current must be enclosed within the closed line integral.

Example 2-27. There are two concentric hollow metallic cylinders. Calculate the magnetic flux density at all regions of space if there is a current flowing out of the paper in the inner cylinder and returns in the outer cylinder.


Answer: The current that is enclosed for the radius $r<a$ is equal to 0 . Therefore, the magnetic flux density within the inner cylinder is equal to 0 . In the region $\mathrm{a}<\mathrm{r}<\mathrm{b}$, there is a current $\mathrm{I}_{\mathrm{enc}}=\mathrm{I}$. From (2.122), the magnetic flux density in this region is equal to $B_{\phi}=\frac{\mu_{0} I}{2 \pi r}$. In the region $r>b$, the current that is enclosed is again equal to 0 resulting in no magnetic flux density in this region.

## Fundamentals of magnetic fields

It should be noted that all of the calculations using Ampere's law have required considerable symmetry. Unfortunately, there will be several problems in which this symmetry does not exist. In these cases, we will have to resort to more complicated analytical or numerical methods in order to obtain a solution for a particular problem. Some of these methods will be described in the ensuing section.

# Magnetic vector potential \& the Biot-Savart law <br> 2.11. Magnetic vector potential \& the <br> <br> Biot-Savart law 

 <br> <br> Biot-Savart law}

There are several cases in practice where it is very difficult to find the magnetic flux density in terms of a current density. This is particularly true if there are difficulties invoking the symmetry arguments required for the application of Ampere's law that was discussed in the previous section. In the following, we will introduce techniques to find this magnetic flux density from a current distribution that has an arbitrary shape. This will be based on some mathematical relations and physical considerations. This will include the development of a new entity called the magnetic vector potential along with a derivation of the Biot-Savart law.

The nonexistence of magnetic monopoles allowed us to write that the magnetic flux density satisfied (2.107), that is rewritten for convenience

$$
\begin{equation*}
\nabla \cdot \mathbf{B}=0 \tag{2.123}
\end{equation*}
$$

The divergence of the magnetic flux density is now specified in (2.123). We still have freedom to examine other properties of it. In particular, we will define a vector A such that the magnetic flux density B can be expressed as the curl of this vector

$$
\begin{equation*}
\mathbf{B}=\nabla \times \mathbf{A} \tag{2.124}
\end{equation*}
$$

This vector will be given the symbol $\mathbf{A}$ and it has the units of Tesla-meter or Webers / meter. This term is called the magnetic vector potential or just the vector potential. The substitution of (2.124) into (2.123) leads to

# Magnetic vector potential \& the Biot-Savart law 

$$
\begin{equation*}
\nabla \bullet \nabla \times \mathbf{A}=0 \tag{2.125}
\end{equation*}
$$

which is a repeated vector operation that was discussed in Chapter 1.4.4.
We will find that the magnetic flux density and the vector potential are somewhat similar to electric fields where it was found that the electric field could be obtained by taking the gradient of a scalar electric potential. This potential was found in terms of an electric charge distribution. Magnetic fields are related to a current density $\mathbf{J}$ via the differential form of Ampere's law (2.113)

$$
\begin{equation*}
\nabla \times \mathbf{B}=\mu_{\mathrm{o}} \mathbf{J} \tag{2.126}
\end{equation*}
$$

Replacing the magnetic flux density using (2.124), we find that the vector potential can be obtained from the current density

$$
\begin{equation*}
\nabla \times \nabla \times \mathbf{A}=\mu_{0} \mathbf{J} \tag{2.127}
\end{equation*}
$$

There is a vector relation for this repeated vector operation. In particular, we write (see Appendix 1)

$$
\begin{equation*}
\nabla(\nabla \cdot \mathbf{A})-\nabla^{2} \mathbf{A}=\mu_{0} \mathbf{J} \tag{2.128}
\end{equation*}
$$

A vector is determined by two vector operations, namely its curl and its divergence. The curl of the vector $\mathbf{A}$ is specified in equation (2.124). We now choose the divergence of the vector $\mathbf{A}$ to be equal to 0 . There are other choices that could be made but this will simplify our calculation. This is called a "Coulomb gauge" in the physics community and it has ramifications that are beyond the scope of this text. We shall let our colleagues in that community dwell on these finer points. With this assumption, (2.128) simplifies to

$$
\begin{equation*}
\nabla^{2} \mathbf{A}=-\mu_{0} \mathbf{J} \tag{2.129}
\end{equation*}
$$

## Magnetic vector potential \& the Biot-Savart law

This equation is similar to Poisson's equation (2.61) that related and electric potential to a charge density. There is, however, a very significant difference in that (2.129) is a vector equation. However, the vector potential $\mathbf{A}$ is in the same direction as the current density J. This means that there is a separate Poisson's type scalar equation for each component of the vector. Frequently the current is flowing in only one direction which means that there will be only one component of the vector potential. We will find that there may be certain advantages in using this intermediate calculation. This is shown in Figure 2-27.


Figure 2-27. The orientation of the magnetic vector potential $\mathbf{A}(\mathbf{r})$ and magnetic flux density $\mathbf{B}(\mathbf{r})$ that surrounds a current element $\mathbf{J}\left(\mathbf{r}^{\prime}\right)$ is depicted.

In Cartesian coordinates, (2.129) is written as

$$
\begin{equation*}
\nabla^{2} \mathrm{~A}_{\mathrm{X}}=-\mu_{0} \mathrm{~J}_{\mathrm{X}}, \nabla^{2} \mathrm{~A}_{\mathrm{Y}}=-\mu_{0} \mathrm{~J}_{\mathrm{Y}} \text {, and } \nabla^{2} \mathrm{~A}_{Z}=-\mu_{0} \mathrm{~J}_{Z} \tag{2.130}
\end{equation*}
$$

We can make use of the knowledge that we have gained from electrostatic fields and write down the solution for each of the components in (2.130).

In analogy with the electric potential (2.62), we write the solution for the vector potential as

## Magnetic vector potential \& the Biot-Savart law

$$
\begin{equation*}
A(r)=\frac{\mu_{0}}{4 \pi} \int_{\Delta v} \frac{J\left(r^{\prime}\right)}{R} d v^{\prime} \tag{2.131}
\end{equation*}
$$

where $R=\sqrt{\left(x-x^{\prime}\right)^{2}+\left(y-y^{\prime}\right)^{2}+\left(z-z^{\prime}\right)^{2}}$ is the distance between the current element in the point where the vector potential is to be evaluated. The integration is to be performed over the volume $\Delta \mathrm{v}$ that contains the current density. Equation (2.131) is a vector equation that represents three scalar equations for the three components of the current density. This is illustrated in Figure 2-27 where the vectors $\mathbf{r}$ and $\mathbf{r}^{\prime}$ are directed to the vector potential and the current density respectively. The magnitude of the distance between these two vectors is given by $R=\left|\mathbf{r}-\mathbf{r}^{\prime}\right|$.

Example 2-28. Find the vector potential $\mathbf{A}$ and the magnetic flux density $\mathbf{B}$ caused by a length 2 a of current $\mathrm{I}=\mathrm{I} \mathbf{u}_{\mathbf{z}}$ at the midpoint of the line.


Answer: Since the magnetic flux density is to be determined at the midpoint of the line, we can invoke the argument of symmetry and to the calculation analytically. The vector potential at a distance $R$ from the wire is found from (2.131). The volume integral becomes a line integral since we will assume that

# Magnetic vector potential \& the Biot-Savart law 

the current is uniformly distributed over the cross section of the wire. This states that

$$
\mathbf{J} d v^{\prime}=\mathbf{J}\left(d s^{\prime}\right)\left(d z^{\prime}\right)=I d z^{\prime} \mathbf{u}_{\mathbf{z}}
$$

The integral (2.131) becomes

$$
\mathbf{A}=\mathbf{u}_{\mathrm{z}} \frac{\mu_{0} \mathrm{I}}{4 \pi} \int_{-\mathrm{a}}^{a} \frac{\mathrm{dz}}{\sqrt{\mathrm{z}^{\prime 2}+\mathrm{r}^{2}}}=\mathbf{u}_{\mathrm{z}} \frac{\mu_{0} \mathrm{I}}{4 \pi} \ln \left(\frac{\sqrt{\mathrm{a}^{2}+\mathrm{r}^{2}}+\mathrm{a}}{\sqrt{\mathrm{a}^{2}+\mathrm{r}^{2}}-\mathrm{a}}\right)
$$

The magnetic flux density is computed from $\mathbf{B}=\nabla \times \mathbf{A}$. Note that the vector potential has only a $\mathbf{u}_{\mathbf{z}}$ component that depends only on the variable r . From the definition of the curl operation in cylindrical coordinates (see Appendix 1), the only nonzero contribution comes from the term $-\mathbf{u}_{\phi} \frac{\partial \mathrm{A}_{\mathrm{z}}}{\partial r}$. Therefore, the magnetic field due to a finite length current carrying wire is equal to

$$
\mathbf{B}=\mathbf{u}_{\phi} \frac{\mu_{0} \mathrm{l}}{2 \pi \mathrm{r}}\left(\frac{\mathrm{a}}{\sqrt{\mathrm{a}^{2}+\mathrm{r}^{2}}}\right)
$$

In the limit of an extremely long wire such that a >> $r$, the term within the brackets approaches 1 and this results in

$$
\mathbf{B}=\mathbf{u}_{\phi} \frac{\mu_{0} \mathrm{I}}{2 \pi \mathrm{r}}
$$

This is the value that we previously obtained using Ampere's law (2.115).

Let us substitute the integral for the vector potential (2.131) into the expression for the magnetic flux density (2.124)

$$
\begin{equation*}
\mathbf{B}(\mathbf{r})=\nabla \times \mathbf{A}(\mathbf{r})=\nabla \times\left(\frac{\mu_{0}}{4 \pi} \int_{\Delta v} \frac{\mathbf{J}\left(\mathbf{r}^{\prime}\right)}{\mathrm{R}} \mathrm{dv}\right) \tag{2.132}
\end{equation*}
$$

## Magnetic vector potential \& the Biot-Savart law

It is desired to compute the vector potential at a location that is different from where the current distribution exists. This means that the curl operation required to determine the magnetic flux density will be performed at the field point of interest and it is somewhat independent of where the current element is located (source point) except through the terms that appear in the distance R.


Figure 2-28. Orientation of a current element in one coordinate system and the resulting vector potential in another coordinate system is shown.

Hence, we can think that there are two coordinate systems that exist and are independent of each other and indicate these coordinates as $\mathbf{r}$ and $\mathbf{r}$ '. This will allow the curl operation to pass through the integral in (2.132) which is required when computing the magnetic field

$$
\begin{equation*}
\mathbf{B}(\mathbf{r})=\frac{\mu_{0}}{4 \pi} \int_{\Delta v} \nabla \times\left(\frac{\mathbf{J}\left(\mathbf{r}^{\prime}\right)}{\mathrm{R}} \mathrm{~d} v^{\prime}\right) \tag{2.133}
\end{equation*}
$$

This can be further simplified using the vector identity that relates the curl of a vector times a scalar quantity, both of which are spatially varying, to be (see Appendix 1)

## Magnetic vector potential \& the Biot-Savart law

$$
\begin{equation*}
\nabla \times(\alpha \boldsymbol{\beta})=\nabla \alpha \times \boldsymbol{\beta}+\alpha \nabla \times \boldsymbol{\beta} \tag{2.134}
\end{equation*}
$$

In (2.134), $\alpha$ is a scalar and $\beta$ is a vector. Applying the vector identity (2.134) to (2.133), we identify these terms as

$$
\alpha=\frac{1}{R} \text { and } \beta=\mathrm{J}\left(\mathrm{r}^{\prime}\right)
$$

Since the current is assumed to be in a coordinate system that is different from the system of the curl operation, $\nabla \times \mathbf{J}\left(\mathbf{r}^{\prime}\right)=0$. Therefore, we obtain

$$
\begin{equation*}
\mathbf{B}(\mathbf{r})=\frac{\mu_{0}}{4 \pi} \int_{\Delta v} \nabla\left(\frac{1}{R}\right) \times \mathbf{J}\left(\mathbf{r}^{\prime}\right) \mathrm{dv} v^{\prime}=-\frac{\mu_{0}}{4 \pi} \int_{\Delta v^{\prime}} \frac{\mathbf{u}_{\mathbf{R}}}{R^{2}} \times \mathbf{J}\left(\mathbf{r}^{\prime}\right) \mathrm{d} v^{\prime}=\frac{\mu_{0}}{4 \pi} \int_{\Delta v} \frac{\mathrm{~J}\left(\mathbf{r}^{\prime}\right) \times \mathbf{u}_{\mathbf{R}}}{R^{2}} d v^{\prime} \tag{2.135}
\end{equation*}
$$

The operation which is a derivative at the field point (unprimed coordinates) can be simplified with the relations that

$$
\nabla\left(\frac{1}{\mathrm{R}}\right)=-\frac{\mathbf{u}_{\mathrm{R}}}{\mathrm{R}^{2}}
$$

and the vector product property that

$$
\alpha \times \beta=-\beta \times \alpha
$$

have also been employed in the effecting this derivation. The unit vector $\mathbf{u}_{\mathbf{R}}$ is in the direction from the current element to the location where the magnetic field is to be computed.

If the current is localized to pass through a wire, it is possible to simplify the volume integral given in (2.135) to read

$$
\begin{equation*}
\mathbf{B}(\mathbf{r})=\frac{\mu_{0}}{4 \pi} \oint \frac{\mid \mathbf{d | l} \times \mathbf{u}_{\mathrm{R}}}{\mathrm{R}^{2}} \tag{2.136}
\end{equation*}
$$

Note that a closed line integral has been used since the current in a wire has to pass through a closed loop, say from one battery terminal through a wire and

## Magnetic vector potential \& the Biot-Savart law

back into the battery through the other terminal. Equation (2.136) is called the Biot-Savart law.

Example 2-29. Use the Biot-Savart law to find the magnetic flux density from a finite length of line with a current element $\mid \mathbf{d l}^{\prime}=I d z^{\prime} \mathbf{u}_{\mathbf{z}}$ at the midpoint of the line.


Answer: Before setting up the integral, let us first perform this vector product where each term can be clearly identified

$$
d l^{\prime} \times \mathbf{u}_{\mathrm{R}}=\mathrm{dz} z^{\prime} \mathbf{u}_{\mathbf{z}} \times\left(\frac{r \mathbf{u}_{\mathbf{r}}-z^{\prime} \mathbf{u}_{\mathbf{z}}}{\sqrt{r^{2}+z^{\prime 2}}}\right)=\frac{r d z^{\prime}}{\sqrt{r^{2}+z^{\prime 2}}} \mathbf{u}_{\varphi}
$$

Note the "-" sign in one of the terms of the unit vector. We have to be careful that we follow the path from the current element to the point of observation. Hence, the magnetic flux density can be calculated using the Biot-Savart law (2.136) from which we write

$$
\mathbf{B}(\mathbf{r})=\frac{\mu_{0} \mathrm{I}}{4 \pi} \int_{-\mathrm{a}}^{a} \frac{r d z^{\prime}}{\left(r^{2}+z^{\prime 2}\right)^{3 / 2}} \mathbf{u}_{\phi}
$$

This integral can be performed analytically by a substitution $z^{\prime}=r \tan \theta$ to finally yield

## Magnetic vector potential \& the Biot-Savart law

$$
\mathbf{B}(\mathbf{r})=\left.\mathbf{u}_{\phi} \frac{\mu_{0} \mathrm{l}}{4 \pi} \frac{z^{\prime}}{\mathrm{r} \sqrt{r^{2}+z^{\prime 2}}}\right|_{-a} ^{a}=\mathbf{u}_{\phi} \frac{\mu_{0} \mathrm{l}}{2 \pi \mathrm{r}} \frac{\mathrm{a}}{\sqrt{r^{2}+\mathrm{a}^{2}}}
$$

As expected, this is the same result that was obtained in Example 2-28 although the integral here is more complicated.

Example 2-30. Find the magnetic field on the axis that is perpendicular to the plane containing a circular loop of current. Use the Biot-Savart law.


Answer: We must first identify the terms that appear in the Biot-Savart law (2.136). We write $\mathbf{d l}^{\prime}=\mathrm{ad} \phi^{\prime} \mathbf{u}_{\phi}, \mathbf{R}=-\mathrm{a} \mathbf{u}_{r}+\mathrm{z} \mathbf{u}_{\mathbf{z}}$, and $\mathrm{R}=\sqrt{\mathrm{a}^{2}+\mathrm{z}^{2}}$. Therefore, we write

$$
\mathbf{B}(\mathbf{r})=\frac{\mu_{0} \mathrm{I}}{4 \pi} \oint \frac{\left(\operatorname{ad} \phi^{\prime} \mathbf{u}_{\phi}\right) \times\left(-a \mathbf{u}_{\mathrm{r}}+\mathbf{z} \mathbf{u}_{\mathbf{z}}\right)}{\left(\mathrm{a}^{2}+z^{2}\right)^{3 / 2}}=\frac{\mu_{0} \mathrm{I}}{4 \pi} \oint \frac{\left(\mathrm{a}^{2} \mathrm{~d} \phi^{\prime} \mathbf{u}_{z}+\mathrm{azd} \phi^{\prime} \mathbf{u}_{\mathrm{r}}\right)}{\left(\mathrm{a}^{2}+z^{2}\right)^{3 / 2}}
$$

Due to symmetry, the term with the unit vector $\mathbf{u}_{\mathrm{r}}$ will contribute zero to the magnetic field. The integration is performed along the path of the wire and yields a vector of $2 \pi$

$$
\mathbf{B}(\mathbf{r})=\frac{\mu_{0}}{2 \pi} \frac{\mathbf{l} \pi \mathrm{a}^{2}}{\left(\mathrm{a}^{2}+\mathrm{z}^{2}\right)^{3 / 2}} \mathbf{u}_{\mathbf{z}}=\frac{\mu_{0}}{2 \pi} \frac{\mathbf{m}}{\mathrm{R}^{3}} .
$$

## Magnetic vector potential \& the Biot-Savart law

This result has been written to incorporate the area enclosed within the current loop.

We will define the magnetic dipole moment as $\mathbf{m}=1 \pi \mathrm{a}^{2} \mathbf{u}_{\mathbf{z}}$. The magnitude of the magnetic dipole moment equals the current I, carried by the wire that forms the circumference of the loop, times the area enclosed within the loop $\pi \mathrm{a}^{2}$. The unit vector is normal to the surface area of the loop using the right hand convention.

This current loop in Example 2-29 is in agreement with the simple model of an atom that considers an atom to have a positive nucleus and an electron that revolves about the nucleus at a fixed radius a. This is usually called a magnetic dipole. We remember an earlier model described in Example 2-11 that considered that the atom to be an electric dipole. This leads to a certain analogy between the electric field intensity $\mathbf{E}$ and the magnetic flux density $\mathbf{B}$ in that the electric dipole moment $\mathbf{p}$ is similar to the magnetic dipole moment $\mathbf{m}$. Both involve a volume integration of either an electric charge density or an electric current density. In addition, both fields can be obtained from a vector differentiation of a potential, either a scalar potential for the electric field intensity or a vector potential for the magnetic flux density.

We have encountered three analytical methods to find the magnetic flux density at a point in space from a current element:
(1) Application of Ampere's circuital law which requires considerable symmetry.

## Magnetic vector potential \& the Biot-Savart law

(2) Determination of the vector potential and the calculation of a magnetic flux density from the vector potential which does not impose the requirement of symmetry.
(3) Application of the Biot-Savart law which also does not impose the requirement of symmetry.

The particular problem that faces us in the future will dictate which approach we should follow. Numerical methods are frequently employed in practice to obtain the magnetic flux density in complicated geometries such as may be found in an electric motor or an electric generator. In fact, there are commercial products that have been developed to perform these calculations since they are so widely used.

## Magnetic forces

### 2.12. Magnetic forces

The first statement that we made regarding the behavior of stationary charged particles concerned the Coulomb force that existed between the particles. The force was created upon a stationary particle that had a charge $q$ if the particle were in an electric field $\mathbf{E}$. This force is given by

$$
\begin{equation*}
F_{\text {electric }}=q E \tag{2.137}
\end{equation*}
$$

If the particle were in motion with a constant velocity $\mathbf{v}$ within a region that only contained an electric field, the particle would still experience the force that is given in equation (2.137).


Figure 2-29. Charged particles entering a region containing a uniform magnetic field are deflected according to equation (2.138). The right hand rule determines the direction of the force on charge. This direction will be in the opposite directions depending upon the sign of the charge.

However, if the particle is in motion with a velocity $\mathbf{v}$ in a region that contains only a magnetic field whose density is $\mathbf{B}$, the force that acts upon the particle is given by

$$
\begin{equation*}
\mathbf{F}_{\text {magneicic }}=\mathrm{q}(\mathbf{v} \times \mathbf{B}) \tag{2.138}
\end{equation*}
$$

The resulting magnetic force $\mathbf{F}$ is perpendicular to both the magnetic flux density $\mathbf{B}$ and to the velocity $\mathbf{v}$ of the particle and this is expressed with the vector product. In Figure 2-29, this magnetic force on a positively charged particle and a negatively charged particle in a region of uniform magnetic field is illustrated. Since the sign of the charge of these two particles is different, the resulting forces will be in the opposite directions. If the charged particle moves with a uniform velocity $\mathbf{v}$ through a uniform electric and magnetic field, the force is given by

$$
\mathbf{F}=q(\mathbf{E}+\mathbf{v} \times \mathbf{B})
$$

This force which is the sum of the electrostatic and the magnetostatic forces is given the name "Lorentz force."


Figure 2-30. A charged particle moving with a constant velocity in a uniform magnetic field experiences a magnetic force that causes the particle to follow a circular trajectory. This figure would correspond to either a positively charged particle with the magnetic field coming out of the page or a negatively charged particle with the magnetic field going into the page.

## Magnetic forces

In a region where the electric field is equal to zero, the charged particles will continue to experience the magnetic force given in (2.138). The resulting motion of the particles will be in a circular orbit as shown in Figure 2-30. It is possible to find the radius of curvature of the motion for the charged particles as will be demonstrated in the following discussion. The particle that has a mass $M_{j}$ will experience a centripetal force whose magnitude is given by

$$
\begin{equation*}
\mathrm{F}_{\mathrm{j}}=\mathrm{M}_{\mathrm{j}} \mathrm{a}_{\mathrm{j}}=\mathrm{M}_{\mathrm{j}} \frac{\mathrm{v}_{\mathrm{j}}^{2}}{\mathrm{r}_{\mathrm{j}}} \tag{2.140}
\end{equation*}
$$

where $r_{j}$ is the radius of curvature and $a_{j}=\frac{v_{j}{ }^{2}}{r_{j}}$ is the acceleration of the $j^{\text {th }}$ particle. The subscript j refers to the particular particle; $\mathrm{j}=$ = "" for the negatively charged particles and $\mathrm{j}=$ " + " for the positively charged particles. The positively charged particles could be singly charged or multiply charged positive ions. The negatively charged particles could either be electrons or ions to which one or more electrons have become attached, thus creating negative ions. In either case, the radius of curvature of the trajectory depends on the mass of the particle. The magnitude of the force caused by the magnetic field is given from (2.138) as

$$
\begin{equation*}
F_{\text {magnetic }}=q v_{j} B \tag{2.141}
\end{equation*}
$$

Equating the two forces given in (2.140) and (2.141) and solving for the radius of curvature $r_{j}$ for the particle with the subscript $j$, we obtain

$$
\begin{equation*}
r_{j}=\frac{M_{j} v_{j}}{q B} \tag{2.142}
\end{equation*}
$$

## Magnetic forces

This radius is called the Larmor radius or gyroradius of the charged particle. We note that the Larmor radius for the electrons moving with the same velocity through the same magnetic field as the ions will be significantly less than that for the protons due to the mass difference of

$$
\begin{equation*}
\frac{\mathrm{m}_{\mathrm{e}}}{\mathrm{M}_{\mathrm{i}}}=\left(\frac{1}{1836}\right) \times\left(\frac{1}{\text { atomic number }}\right) \tag{2.143}
\end{equation*}
$$

The mass dependence of the Larmor radius suggests that it can be used as a diagnostic tool in order to determine the mass of an unknown material. After ionizing the unknown material and passing all of the ions through a uniform magnetic field with the same velocity $v$, the ions can be collected in a juxtaposed series of collectors. The location of each collector is determined by the Larmor radius of the different elements and the presence or absence of ions in each collector can be monitored. We can include effects of an ion having more than a single charge also. Such a device is called a mass spectrometer. This has also been used to separate various isotopes from each other.

Example 2-31. Calculate the Larmor radius for an electron and an argon ion that pass through a magnetic field of 0.01 T . Both particles have been accelerated through a potential difference of one volt.

Answer: Before calculating the Larmor radius for either particle, the velocity of each particle must be computed. Since the particles have gained an energy of $q_{e} \Delta V=1 \mathrm{eV}=1.602 \times 10-19 \mathrm{~J}$, this energy will appear as kinetic energy and we wherite

$$
\mathrm{q}_{\mathrm{e}} \Delta \mathrm{~V}=\frac{\mathrm{m}_{\mathrm{e}} \mathrm{v}_{\mathrm{e}}{ }^{2}}{2}=\frac{\mathrm{M}_{\mathrm{Ar}} \mathrm{v}_{\mathrm{Ar}}{ }^{2}}{2}
$$

$$
v_{e}=\sqrt{\frac{2 q_{e} \Delta V}{m_{e}}} \text { and } v_{A r}=\sqrt{\frac{m_{e}}{M_{A r}}} \sqrt{\frac{2 q_{e} \Delta V}{m_{e}}}
$$

The atomic number of argon is 40 yielding a ratio of the masses to be

$$
\frac{\mathrm{m}_{\mathrm{e}}}{\mathrm{M}_{\mathrm{Ar}}}=\left(\frac{1}{1836}\right) \times\left(\frac{1}{40}\right)
$$

The Larmor radius for the electron is found from (2.142) to be

$$
r_{e}=\frac{m_{e} v_{e}}{q_{e} B}=\frac{m_{e} \sqrt{\frac{2 q_{e} \Delta V}{m_{e}}}}{q_{e} B}=\frac{\left(9.1 \times 10^{-31}\right) \sqrt{\frac{2\left(1.602 \times 10^{-19}\right)(1)}{\left(9.1 \times 10^{-31}\right)}}}{\left(1.602 \times 10^{-19}\right)\left(10^{-2}\right)}=3.4 \times 10^{-4} \mathrm{~m}
$$

The velocity of the argon ion can be expressed in terms of the electron velocity

$$
v_{A r}=\sqrt{\frac{m_{e}}{M_{A r}}} \sqrt{\frac{2 q_{e} \Delta V}{m_{e}}}=\sqrt{\frac{m_{e}}{M_{A r}}} v_{e}
$$

Therefore, the Larmor radius for the argon ion is found from (2.142) which can also be expressed in terms of the electron Larmor radius

$$
\begin{aligned}
r_{A r} & =\frac{M_{A r} v_{A r}}{q_{e} B}=\frac{M_{A r} \sqrt{\frac{2 q_{e} \Delta V}{M_{A r}}}}{q_{e} B}=\left(\frac{M_{A r}}{m_{e}}\right)\left(\frac{m_{e} \sqrt{\frac{2 q_{e} \Delta V}{m_{e}}}}{q_{e} B}\right)\left(\sqrt{\frac{m_{e}}{M_{A r}}}\right) \\
& =\sqrt{\frac{M_{A r}}{m_{e}}} r_{e}=\sqrt{40 \times 1836} r_{e}=9.2 \times 10^{-2} m
\end{aligned}
$$

A comparison on the two Larmor radii indicates that the electrons are closely "tied" to the magnetic field lines and the ions are not. In several applications, the electrons are said to be "magnetized" and the ions are "unmagnetized."


Figure 2-31. A positive charge moving a distance dl with a constant velocity $\mathbf{v}$ in a region containing a uniform magnetic flux density $\mathbf{B}$ is illustrated.

It is interesting at this time to calculate the work that is performed by the charged particle as it passes through the region of magnetic field. Recall that this work $\Delta \mathrm{W}$ is computed from the line integral

$$
\begin{equation*}
\Delta \mathrm{W}=\int_{a}^{b} \mathrm{~F} \cdot \mathrm{dl} \tag{2.144}
\end{equation*}
$$

As shown in Figure 2-31and from equation (2.138), we find that the force is perpendicular to the direction that it travels. This implies that the work that is computed from (2-144) will be zero.

Example 2-32. Show that the incremental work $\Delta \mathrm{W}$ performed in moving a positive charge $\mathbf{Q}$ with a velocity $\mathbf{v}=\mathrm{v}_{\mathrm{o}} \mathbf{u}_{\mathbf{x}}$ an incremental distance $\Delta \mathbf{x} \mathbf{u}_{\mathbf{x}}$ through a uniform magnetic field $\mathbf{B}=\mathrm{B}_{\mathrm{o}} \mathbf{u}_{\mathbf{y}}$ is equal to zero.


Answer: From (2.144) and (2.138), we write

$$
\begin{aligned}
\Delta \mathrm{W}= & \mathbf{F} \bullet \Delta \mathrm{x} \mathbf{u}_{\mathbf{x}}=\mathrm{Q}(\mathbf{v} \times \mathbf{B}) \bullet \Delta \mathrm{x} \mathbf{u}_{\mathbf{x}} \\
& =\mathrm{Q}\left(\mathrm{v}_{0} \mathbf{u}_{\mathbf{x}} \times \mathrm{B}_{0} \mathbf{u}_{\mathbf{Y}}\right) \bullet \Delta \mathrm{xu}_{\mathbf{x}}=Q \mathrm{v}_{0} \mathrm{~B}_{0}\left[\left(\mathbf{u}_{\mathbf{x}} \times \mathbf{u}_{\mathbf{Y}}\right) \bullet \mathbf{u}_{\mathbf{x}}\right]=0
\end{aligned}
$$

The term within the square brackets is identically equal to 0 .

Earth and several of the other planets are examples which illustrate the effects of this force field upon charged-particle motion. See Figure 2-32. Particles are created by the collision of high-energy cosmic rays with low energy particles near the Earth as well as by complex acceleration processes due the interaction of the solar wind (a stream of ionized particles flowing from the Sun) and the Earth's magnetic field. These charged particles are trapped in the Earth's magnetic field. This entrapped region is called a radiation belt. As determined by the passage of the Voyager spacecraft on its more than twelve year journey from Earth into the far reaches of the solar system, several planets ${ }^{1}$ have magnetic fields that capture these charged particles coming from the sun. Since we might expect that there are almost an equal number of electrons and positively charged ions in this region, there is electrical neutrality and this charged particle fluid is called a "plasma." The Earth's radiation belt is called the Van Allen belt in honor of Professor James Van Allen who originally discovered its existence using the satellite "Explorer 1" in 1958.

[^5]
## Magnetic forces

(a)

(b)


Figure 2-33. (a) Schematic representation of the radiation belt. The first experimental detection of these charged particles was made on the satellite Explorer 1 in 1958.
(b) A global view of Earth as observed from the Polar Visible Imaging System (VIS). This is an image of Earth's Aurora Borealis, Northern Auroral Oval or Northern Lights in the ultraviolet spectrum superposed on an image of Earth's surface for 25 March 1996. A similar picture could be taken over the South Pole where it is called Aurora Australis, or Southern Lights. Courtesy of L. A. Frank, J. B. Sigwarth, The University of Iowa, and NASA/Goddard Space Flight Center.

At the start of this section, we wrote the force on a charged particle that passed through a uniform magnetic flux density $\mathbf{B}$ in equation (2.138). $A$ differential charge $d Q=\rho_{v} d v$ moving with a constant velocity constitutes a current. If this current flows in a closed path, (2.138) can be written as

$$
\begin{equation*}
\mathrm{dF}_{\text {magnetic }}=\mathrm{dQ}(\mathbf{v} \times \mathbf{B})=\rho_{\mathrm{V}}(\mathbf{v} \times \mathbf{B}) \mathrm{d} \mathbf{v}=\mathbf{J} \times \mathbf{B}(\mathrm{ds})(\mathrm{dl})=\mid \mathbf{d} \mathbf{l} \times \mathbf{B} \tag{2.145}
\end{equation*}
$$

The total force $\mathbf{F}$ is computed by integrating the differential force over the path

$$
\begin{equation*}
F_{\text {magnetic }}=-\int B \times \mid \mathbf{d l} \tag{2.146}
\end{equation*}
$$

where the "-" arises from the inversion of the vector product.
If we assume that the magnetic flux density is a constant, the magnetic flux density can be taken outside of the integral sign. This leads to the closed line integral $\oint \mathbf{d} \mathbf{l}$ which is equal to zero. This states that a closed loop will not move in a linear direction. If the magnetic field is not uniform in space, then the net force will not necessarily be equal to zero. Although the net translational

## Magnetic forces

force in a uniform magnetic field is equal to zero, there may be a torque that acts on the loop and causes it to rotate about an axis.

Before examining the torque that will exist on the loop, let us first examine the force that exists between two parallel wires, each of which carries a current as shown in Figure 2-33. We are going to calculate the force that exists between these two wires. Before presenting the formal derivation, let us postulate certain properties of the force that may exist on the two wires. Let the two wires lie in the xz plane. In the first case, the currents are going in the same direction as shown in Figure 2-33a. The magnetic field created by wire 1 will be directed in the $+\mathbf{u}_{\mathbf{y}}$ direction at the location of wire 2 . The force on wire 2 as computed from (2.146) will be in the direction given by $\mathbf{F}_{2}=-I B \mathbf{u}_{\mathbf{y}} \times \mathrm{dl}_{2} \mathbf{u}_{\mathbf{z}}$ or in the $-\mathbf{u}_{\mathbf{x}}$ direction. This states that wire 2 will be attracted to wire 1 . Similarly, the force on wire 1 caused by the magnetic field created by wire 2 will cause wire 1 to be attracted to wire 2 . If the currents going in the opposite directions as depicted in Figure 2-33b, the magnetic force given in (2.146) will be in the direction that will cause the wires to be repelled from each other.


Figure 2-33. Currents are flowing through two parallel wires. (a) The current in both wires is flowing in the same direction (attraction). (b) The current in both wires is flowing in the opposite direction (repulsion).

These forces can also be argued from the following point of view. In the region between the two wires in Figure 2-33a, the magnetic fields caused by the two wires oppose each other will therefore cancel. The magnetic fields will add in the regions external to this separation region. Hence we could think that there is a "pressure" on the wires to fill in this region since there is an old world axiom that "nature abhors a vacuum." The same argument could also be applied in describing the force for the situation depicted in Figure 2-33b were the cancellation of the magnetic fields occurs in the region external to the two wires.

Example 2-33 A lightning rod is a device that provides an attractive path through a wire for lightning to discharge so that a building is hopefully protected. In order to attract the lightning to the rod, the path must originally have a low resistance.

## Magnetic forces

However, after the lightning stroke starts to discharge through this rod, the resistance should suddenly increase in order to protect the rod so it can survive to another day. Discuss how this could be done with inexpensive parts.

(a)

(b)

Answer: The lightning stroke will discharge through the shortest path between the two metal conductors as shown in (a). This discharge will act like a variable length conductor and will create a local magnetic field between the two metal conductors as depicted in (b). The direction of the resulting force (l dl $\mathbf{x} \mathbf{B}$ from (2.145)) will cause this arc to move to the right where the path length becomes longer. Since the path length increases with motion of the arc to the right, the resistance will also increase. This device is called a lightning arrestor. The mechanism described here is also used in a device called a rail gun that may have applications of spewing forth "plasma bullets." Movie aficionados should also recognize the Jacob's ladders that appear in the laboratory of the old Dr . Frankenstein movies.

## Magnetic forces

Now let us calculate the force between the two parallel current carrying wires depicted in Figure 2-33 more formally. The magnetic field that should be incorporated into the force equation (2.146) is determined from the Biot-Savart law (2.136). We write for the force on wire 1 caused by the magnetic field created by the current in wire 2 using the notation

$$
\begin{equation*}
\mathbf{F}_{12}=-\mathrm{I}_{1} \oint_{\mathrm{L}_{1}} \mathbf{B}_{12} \times \mathbf{d I _ { 1 }} \tag{2.147}
\end{equation*}
$$

From the Biot-Savart law, we find that the magnetic flux density at wire 1 caused by the current in wire 2 is given by

$$
\begin{equation*}
\mathbf{B}_{12}=-\frac{\mu_{0} \mathbf{I}_{2}}{4 \pi} \oint_{L_{2}} \frac{\mathbf{u}_{\mathbf{R}_{21}} \times \mathbf{d l}_{2}}{\mathrm{R}_{21}^{2}} \tag{2.148}
\end{equation*}
$$

Substitute (2.148) into (2.147) and derive this force, which is called Ampere's force , as

$$
\begin{equation*}
\mathbf{F}_{12}=\frac{\mu_{0} \mathbf{l}_{1} \mathbf{l}_{2}}{4 \pi} \oint_{L_{1}} \oint_{\mathrm{L}_{2}} \frac{\left(\mathbf{u}_{\mathbf{R}_{21}} \times \mathbf{d l}_{2}\right) \times \mathbf{d l}_{1}}{\mathrm{R}_{21}^{2}} \tag{2.149}
\end{equation*}
$$

The force on wire 2 can be computed by merely interchanging the subscripts 1 and 2.

In writing (2.149), we might be tempted to compare the force given by this equation with the Coulomb force equation given in (2.2). There are obvious similarities in that the force is proportional to the magnitudes $\mathbf{I}_{1} \mathbf{d} \mathbf{l}_{\mathbf{1}}$ and $\mathbf{I}_{2} \mathbf{d} \mathbf{l}_{\mathbf{2}}$ or $\mathrm{Q}_{1}$ and $Q_{2}$ for Coulomb's law. Both equations are inversely proportional to the square of the separation distance. This is similar to the gravitational force between two objects in that it is proportional to the masses on the two objects divided by the separation distance between them. Closed line integrals are used

## Magnetic forces

in (2.149) since one cannot experimentally construct isolated current elements. For the two infinite parallel wires depicted in Figure 2-33, they are closed at the place we call infinity.


Figure 2-34. A rectangular current loop is inserted in a uniform magnetic field. The current in the loop flows in the counterclockwise direction.

Consider a current carrying loop of wire as shown in Figure 2-34. We will assume, for simplicity, that $\mathbf{B}=\mathrm{B}_{\mathrm{O}} \mathbf{u}_{\mathbf{z}}$. The separation distance between the two wires that our closest to the $x$ axis can be assumed to be infinitesimally small. We can consider that two parallel wires that each carries a current in the opposite direction. In addition, the other two wires also carry a current in the opposite directions. In Figure 2-33, it was shown that two parallel wires that carry currents in the opposite directions will have a repulsive force. Therefore, the net force on this closed loop will be equal to the sum of all these forces which is equal to 0 . This implies that there will be no net translation of this closed loop in any direction.

## Magnetic forces

Example 2-34. Formally demonstrate that the sum of the forces acting on the rectangular loop shown and Figure 2-34 is identically equal to 0 which implies that this loop will not translate in any direction.

Answer: The magnetic flux density is $\mathbf{B}=\mathrm{B}_{0} \mathbf{u}_{\mathbf{z}}$. Using the definition of the magnetic force given in (2.146), we write the sum of the forces that act on the four sides as

$$
\begin{aligned}
& \mathbf{F}_{\text {magnetic }}=-\int_{1}^{2} \mathbf{B} \times \mathbf{d} \mathbf{d}-\int_{2}^{3} \mathbf{B} \times \mathbf{d} \mathbf{d}-\int_{3}^{4} \mathbf{B} \times \mathbf{d} \mathbf{d}-\int_{4}^{1} \mathbf{B} \times \mathbf{d} \mid
\end{aligned}
$$

Recall that the sign of the integral is determined by the limits of the integration.


Figure 2-35. Two of the edges of the rectangular loop rotate across the magnetic field while the other two are in the plane of the magnetic field and are unaffected by it.

Although the summation of the forces is equal to 0 , there will be a torque on the loop that will cause it to rotate. The forces on wires 1 and 3 are in the

## Magnetic forces

opposite directions as shown in Figure 2-35. If the normal to the cross section of the loop is at a slight angle $\theta$ with respect to the applied magnetic field, this torque acting upon the loop can be calculated. In order to calculate the torque, we assume that loop is constrained to rotate about one axis only for simplicity. The torque on the loop is given by

$$
\begin{equation*}
\text { torque }=F_{1} \sin \theta\left(\frac{\Delta y}{2}\right)+F_{3} \sin \theta\left(\frac{\Delta y}{2}\right) \tag{2.150}
\end{equation*}
$$

where

$$
\mathrm{F}_{1}=\mathrm{IB} \mathrm{~B}_{0} \Delta \mathrm{x} \text { and } \mathrm{F}_{3}=\mathrm{IB} 。 \Delta \mathrm{x}
$$

This leads to

$$
\begin{equation*}
\text { torque }=\mathrm{IB}_{0}(\Delta \mathrm{x} \Delta \mathrm{y}) \sin \theta \tag{2.151}
\end{equation*}
$$

The area of the loop is equal to $\Delta \mathbf{s}=(\Delta x \Delta y) \mathbf{u}_{\mathrm{n}}$ where the unit vector is normal to the surface area. If we multiply this area by the current I, we can recognize this is a representation for the magnetic dipole moment that was described in Example 2-30. Finally, we are able to write (2.151) in vector notation as

$$
\begin{equation*}
\mathbf{T}=\mathbf{m} \times \mathbf{B} \tag{2.152}
\end{equation*}
$$

where $\mathbf{T}=\mathbf{R} \times \mathbf{F}(\mathrm{N}-\mathrm{m})$ is the torque causing the rectangular loop to rotate about its axis as shown in Figure 2-36. We have made liberal use of the definition of the magnetic dipole moment in that we have replaced a circular loop with a rectangular loop. This will have a dramatic effect when we consider the magnetic properties of materials.

## Magnetic forces



Figure 2-36. Definition of the torque $\mathbf{T}$ in terms of the force $\mathbf{F}$ and the lever arm R.

## Magnetic materials

### 2.13. Magnetic materials

Having determined the magnetic field from a current carrying loop that could, in some sense, approximate an atom, we will now investigate the characteristics of a material made of a very large number of atoms and their corresponding magnetic dipoles. These dipoles will be assumed to be oriented randomly at the start of this discussion as shown in Figure 2-37. In addition to the magnetic dipole moments created by the electron orbiting about the positive nucleus, the electron by spinning about its own axis creates a magnetic field. The electron spins on its own axis but this is a topic that we do not have to understand here.


Figure 2-37. Random orientation of magnetic dipoles in a material.

The question that will now be answered is, "What will happen to these magnetic fields from individual atoms if an external magnetic field is applied to the material?" The answer depends on the type of material that is being considered. There are three classes of materials that should be considered and the classification is based on the reorientation of the magnetic dipoles under the influence of an external magnetic field.

In the first class of materials, the magnetic dipoles get reoriented such that their magnetic dipole moments $\mathbf{m}$ are in slight opposition to the applied magnetic field B. Under the influence of no external magnetic field, the atom's magnetic

## Magnetic materials

moment that is created by the electron rotating about the positive nucleus cancels the magnetic field created by the spin of the electron. The application of the external magnetic field perturbs the velocities of the orbiting electrons. Hence a small magnetic moment for the atom is created that according to Lenz's law will oppose the applied magnetic field. Lenz's law will later be discussed in detail. A dimensionless parameter that measures this reorientation is called the magnetic susceptibility and the symbol is $\chi_{m}$. For this class of materials, it is usually very small and it is negative. A typical value for $\chi_{\mathrm{m}}$ is of the order of $\chi_{\mathrm{m}} \approx-10^{-5}$. This class of materials is called a diamagnetic material. Examples of a diamagnetic material include bismuth, copper, diamond, germanium, gold, lead, mercury, silicon, silver and several inert gases. These materials exhibit no permanent magnetic field.

In the second class of materials, the magnetic moments created by the orbiting electrons and the spinning electrons do not completely cancel leaving the atom with a small net magnetic moment. The application of an external magnetic field tends to align these magnetic moments in the direction of the applied magnetic field. This effect is also very small and nonpermanent. In this case, the magnetic susceptibility is small and positive (of the order of $\chi_{\mathrm{m}} \approx+10^{-5}$ ). These materials are called paramagnetic materials and examples of these materials include aluminum, magnesium, oxygen, titanium, tungsten and yttrium oxide.


## Magnetic materials

Figure 2-38. Domain structure of a ferromagnetic specimen is illustrated. The magnetic moments of all of the atoms in our domain are pointed in the same direction.

The third class of magnetic materials is called a ferromagnetic material and it requires a different explanation. The materials that are in this class; iron, nickel and cobalt have $\chi_{\mathrm{m}} \approx 250$ for nickel to $\chi_{\mathrm{m}} \approx 4,000$ for pure iron. There are special alloys having $\chi_{\mathrm{m}}$ having values up to 100,000 . The explanation, that has experimental confirmation, is that a ferromagnetic material consists of domains; whose dimensions range from a few microns to 1 mm and that contain approximately $10^{15}$ to $10^{16}$ atoms. Each domain has all of the magnetic moments aligned in the same direction in the absence of an applied magnetic field as shown in Figure 2-38. Separating one domain from the adjacent domain is a transition of approximately 100 atom thickness that is called a domain wall. The net magnetization of the randomly oriented domains is zero.

Under the influence of an external magnetic field, the domains that have their magnetic fields aligned with the applied field grow at the expense of the other domains. If the applied field is small, this process can be reversed, that is, it can be made to forget its past history. However, if the applied field is strong, the domains rotate in the direction of the applied field and the process becomes irreversible in that the domain remembers its past orientation. This memory is useful in computer applications where the storage of information in various electronic components is important.

Before illustrating this process, we have to introduce the concept of magnetization of a material. The domains in the ferromagnetic material or the atoms in a diamagnetic or paramagnetic material each possess a magnetic moment that

## Magnetic materials

we will label as $\mathbf{m}_{\mathbf{j}}$. It turns out that a ferromagnetic material will become a paramagnetic material if the temperature increases above some value that is called the Curie temperature. Fortunately, we will attempt to stay below this temperature in the following discussion. The total magnetization $\mathbf{M}$ that is the magnetic dipole moment per unit volume of a material in a volume $\Delta \mathrm{V}$ is defined as

$$
\begin{equation*}
\equiv \lim _{\Delta \mathrm{v} \rightarrow 0} \frac{1}{\Delta \mathrm{v}} \sum_{\mathrm{j}=1}^{\mathrm{j}=\mathrm{n}} \mathrm{~m}_{\mathrm{j}} \tag{2.153}
\end{equation*}
$$

Its units are $\left(A \cdot m^{2}\right) / m^{3}=(A / m)$. This magnetization $\mathbf{M}$ creates a current $I_{m}$ that is bound to the domain of the atom. We can write an Ampere's circuital law for the domain or the atom as

$$
\begin{equation*}
\mathrm{I}_{\mathrm{m}}=\oint \bullet \mathbf{d} \mathbf{I}=\int_{\Delta \mathrm{s}} \mathbf{J}_{\mathrm{m}} \bullet \mathbf{d s} \tag{2.154}
\end{equation*}
$$

In this integral, we have set the closed line integral of the magnetization $\mathbf{M}$ equal to the magnetization current $I_{m}$ that is enclosed within this loop. This is not the real current I that you would draw from a battery. We, however, can analyze its effects. From Stokes's theorem, we write

$$
\begin{equation*}
\int_{\Delta \mathrm{s}} \nabla \times \mathbf{M} \bullet \mathbf{d s}=\int_{\Delta \mathrm{s}} \mathbf{J}_{\mathbf{m}} \cdot \mathbf{d s} \tag{2.155}
\end{equation*}
$$

Therefore, the magnetization can be related to this magnetization current. Since the two integrands must be equal for $(2.155)$ to be valid over any arbitrary surface $\Delta s$, we obtain

$$
\begin{equation*}
\nabla \times \mathbf{M}=\mathbf{J}_{\mathbf{m}} \tag{2.156}
\end{equation*}
$$

Let us add the magnetization current density to the current density that was used in the differential form of Ampere's law (2.113) and write

$$
\begin{equation*}
\frac{1}{\mu_{o}} \nabla \times \mathbf{B}=\mathbf{J}+\mathbf{J}_{\mathbf{m}}=\mathbf{J}+\nabla \times \mathbf{M} \tag{2.157}
\end{equation*}
$$

This can be rewritten as

$$
\begin{equation*}
\frac{1}{\mu_{0}} \nabla \times\left(\frac{\mathbf{B}}{\mu_{\mathbf{o}}}-\mathbf{M}\right)=\mathbf{J} \tag{2.158}
\end{equation*}
$$

## Magnetic materials

We now define a new fundamental quantity, the magnetic field intensity $\mathbf{H}$ :

$$
\begin{equation*}
\mathbf{H}=\frac{\mathbf{B}}{\mu_{\mathbf{o}}}-\mathbf{M} \tag{2.159}
\end{equation*}
$$

with units $(A / m)$. Ampere's circuital law can then be written as

$$
\begin{equation*}
\oint \mathrm{H} \bullet \mathrm{dI}=\mathrm{I}_{\mathrm{enc}} \tag{2.160}
\end{equation*}
$$

This states that a magnetic field intensity or "H field " can be created with a real applied current and it is independent of whether a material is within the vicinity. Recall that this property is similar to the displacement flux density being independent of the dielectric. In a vacuum, the magnetization $\mathbf{M}$ is equal to zero and the magnetic flux density is directly proportional to the magnetic field intensity. In magnetic materials, the magnetization is related to $\mathbf{H}$ and we can let $\mathbf{M}=\chi_{\mathrm{m}} \mathbf{H}$ where $\chi_{\mathrm{m}}$ is the magnetic susceptibility. Equation (2.159) can be written as

$$
\begin{equation*}
\mathbf{B}=\mu_{o}\left(1+\chi_{\mathrm{m}}\right) \mathbf{H}=\mu_{0} \mu_{\mathrm{r}} \mathbf{H}=\mu \mathbf{H} \tag{2.161}
\end{equation*}
$$

where $\mu_{\mathrm{r}}$ is the relative permeability of the material. Except for the ferromagnetic materials iron, nickel, and cobalt, it is justified to assume that the relative permeability is equal to 1 . The values for several materials are included in Appendix 3.

Example 2-35. A magnetic flux density of $B=0.05 \mathrm{~T}$ appears in a magnetic material with $\mu_{r}=50$. Find the magnetic susceptibility $\chi_{m}$ and the magnetic field intensity H .

Answer: The magnetic susceptibility $\chi_{\mathrm{m}}$ is given by

$$
\chi_{m}=\mu_{r}-1=50-1=49
$$

The magnetic field intensity H is computed from

$$
\mathrm{H}=\frac{\mathrm{B}}{\mu_{\mathrm{r}} \mu_{o}}=\frac{0.05}{50 \times 4 \pi \times 10^{-7}}=796 \mathrm{~A} / \mathrm{m}
$$

## Magnetic materials

In addition to having a large numerical value, the ferromagnetic materials also has a hysteresis behavior in that the dependence of the magnetic flux density on the magnetic field intensity is nonlinear. In addition, it remembers its previous value and the direction in which it reached that value. This can best be described from an examination of Figure 2-39a. In this figure, a sketch of the magnetic flux density $B$ that could be obtained from an experimental measurement is plotted as a function of the magnetic field intensity H . The latter quantity starts at a value of $\mathrm{H}=0$ and increases as the magnetic flux density is measured. Initially, there is a linear increase in the magnetic flux density.

Let us start at the origin and slowly increase the magnetic field intensity H . This magnetic field intensity could be computed from (2.161) or measured in an experiment by uniformly wrapping a wire around a ferromagnetic material that is connected to a variable current supply. At small values of $H$, the magnetic flux density $B$ will increase proportionally with $H$. As the value of $H$ is further increased, almost complete domain rotation and domain wall motion will have occurred and a state of saturation will be achieved. If we now reduce the H field and eventually change its direction, the corresponding value of B does not follow the initial curve but follows a different path. This phenomenon is called hysteresis that follows from the Greek word meaning "to lag." Even at a value of $\mathrm{H}=0$, there will be a residual magnetic flux density. Eventually a saturation region will be achieved in the opposite direction. Reducing H and reversing its direction eventually brings us back to the original saturation point. From this curve, we note that the relative permeability $\mu_{r}$ depends on the value of H . Frequently, average values are used in practice and several of these values are included in Appendix 3. This total curve is called a hysteresis curve.


Figure 2-39. (a) Hysteresis curve in the B-H plane for a ferromagnetic material. The curve starts at the origin following curve 1 until it saturates. The curve 2 corresponds to decreasing values of H until it again saturates. The curved 3 corresponds to increasing values of H . (b) Idealized hysteresis curve useful for computer memories.

Since the magnetic material remembers the magnitude and the direction of the magnetic flux density $(B>0$ or $B<0)$ at $H=0$, it can be used as a memory element in a logic circuit. The critical factor that determines these values is the direction of the change of H , hence whether some current is flowing in one direction or the other. Engineers and scientists who work with materials can and are able to optimize the hysteresis curve by making it almost rectangular in shape as shown in Figure 2-39b. This is of particular interest in computer applications such as a magnetic memory device. Other requirements may dictate the size of the hysteresis curve.

## Magnetic circuits

### 2.14. Magnetic circuits

Because the relative permeability of ferromagnetic materials is much higher than the ambient regions in which they may be located, we can use this property to confine magnetic fields. This is similar to electric currents being confined by or actually flowing through conducting materials instead of the surrounding air since the conductivity of the conductor is so much higher than the air. There are quantities in magnetic circuits that are similar to the voltages, currents and resistances that we find in the electric circuits. Much of the well-developed power and knowledge that we may already possess from previous courses in circuit theory can be brought to bear upon magnetic circuits. In particular, Kirchhoff's laws will apply as will be shown below. There is, however, one fly in the ointment. Although it is not a bad approximation to assume that the conductivity of the wire used in electric circuits is a constant that is independent of the amplitude or direction of the flow of the current, one has to be more careful in dealing with magnetic circuits. From the hysteresis curve shown in Figure 239a, we note that we have to be concerned not only with the amplitude but also with the direction of the flow of the magnetic flux when selecting an average value for a relative permeability that may be used in the calculation of a magnetic circuit.

In order to demonstrate this concept, we will examine the magnetic circuits shown in Figure 2-40. There are N turns of wire that are wrapped around an iron core that has a relative permeability $\mu_{r} \gg 1$. Hence, we may assume with a reasonable justification that the magnetic fields are confined within the magnetic material just as we considered that the electric current to be confined within the high conductivity wire in an electric circuit. If we were to compare the ratios of the relative permeability of the magnetic material to the external region and the
relative conductivity of the wire to the external region, and the efficacy of using magnetic and electric circuit theory can be justified.

There is an additional assumption that is made concerning the very small gap that exists in the ferromagnetic material. The assumption is to assume that the cross-sectional area of the gap is identical with the cross-sectional area of the magnetic material. This implies that there is no fringing magnetic field. In addition, the magnetic flux is uniformly distributed within the cross-sectional area of the iron resulting in a constant magnetic flux density in the gap. In addition, we approximate the integration path of the closed line integral by assuming that it passes through the center of the iron - gap structure.


Figure 2-40. (a) A magnetic circuit. A mean length within the iron region is $L$ and the gap length is $g$. The cross-sectional area of the iron is $A=a \times b$. (b) Equivalent circuit for the magnetic circuit where $\digamma$ is the magneto motive force, $\sum_{i \text { iron }}$ is the reluctance of the iron, and $\Omega_{\text {gap }}$ is the reluctance of the gap.

## Magnetic circuits

Let us apply Ampere's circuital law (2.108) to the magnetic circuit shown in Figure 2-40. Since the path is understood, we can eliminate the vector notation and write

$$
\begin{equation*}
\mathrm{H}(\mathrm{~L}+\mathrm{g})=\mathrm{l}_{\mathrm{enc}}=\mathrm{NI} \tag{2.162}
\end{equation*}
$$

Since there are N turns of wire, the current that is enclosed within this closed path $\mathrm{I}_{\mathrm{enc}}=$ NI. In this magnetic circuit, the total magnetic flux $\Psi_{\mathrm{m}}$ passes through the magnetic circuit. This magnetic flux is defined as $\Psi_{m}=B \times A$ where $A=a \times b$ is the cross-sectional area of the iron. The magnetic flux density $B$ is related to the magnetic field intensity $H$ through the relation that $B=\mu_{o} \mu_{r} H$. Therefore, (2.162) can be written as

$$
\begin{equation*}
\frac{B}{\mu_{o} \mu_{r}} L+\frac{B}{\mu_{o}} g=\Psi_{m}\left(\frac{L}{A \mu_{o} \mu_{r}}+\frac{g}{A \mu_{o}}\right)=N I \tag{2.163}
\end{equation*}
$$

The term NI plays the role of a magneto-motive force (mmf) and it will be given the symbol $F_{0}$ It has the units of "ampere-turns" and it creates the magnetic flux $\Psi_{m}$. The source of the mmf is external to the magnetic circuit. Therefore, we can consider (2.163) as being similar to Ohm's law for magnetic circuits. The term $\left(\frac{L}{A \mu_{0} \mu_{r}}+\frac{g}{A \mu_{0}}\right)$ is called the reluctance $R_{\text {iron }}+R_{\text {gap }}$ of this circuit. In this case, the total magnetic flux passes through the iron and the gap and we can consider that the two reluctances are in series. The units of $R$ are sometimes given in the units of "rels." This is very similar to an electrical circuit containing a battery and two resistors in series. Since the relative permeability of the iron core is of the order of 1000, the dominant contribution to the reluctance may be caused by the narrow gap region.

Example 2－36．Assume that the reluctance of the gap is equal to the reluctance of the iron in Figure 2－40．The iron has a relative permeability of $\mu_{r}=1000$ ． Determine the length of the gap in terms of the length of the iron．

Answer：The reluctance of the two regions is defined as $\beta_{i r o n}=\frac{L}{A \mu_{0} \mu_{r}}$ and $R_{\text {gap }}=\frac{g}{A \mu_{o}}$ ．The length of the gap $g$ is found to be $g=\frac{L}{\mu_{r}}=\frac{L}{1000}$ ．

Example 2－37．Write a set of coupled equations for the coupled circuit shown in （a）．The iron has a cross－sectional area of $A$ and a relative permeability $\mu_{r}$ ．An equivalent circuit is given in（b）．

Answer：In this circuit，the mean distances between the intersection points are given by $L_{1}, L_{2}$ and $L_{3}$ ．


From figure（a），it is possible to draw an equivalent electrical circuit（b）that contains the following elements

$$
尺_{1}=\frac{L_{1}}{\mu_{\mathrm{r}} \mu_{\mathrm{o}} A}, 尺_{2}=\frac{\mathrm{L}_{2}}{\mu_{\mathrm{r}} \mu_{\mathrm{o}} A}, 尺_{3}=\frac{\mathrm{L}_{3}}{\mu_{\mathrm{r}} \mu_{\mathrm{o}} A}, \digamma_{1}=N_{1} l_{1} \text {, and } \digamma_{2}=N_{2} I_{2}
$$

(b)


Due to the winding of the wire about the material and the direction of the current in these wires, the magnetic flux from both coils will be in a direction such that the magnetic fluxes will add in the center conductor. Hence, we write from our magnetic circuit using Kirchhoff's law that

$$
\begin{aligned}
& r_{1}=\Psi_{1}\left(R_{1}+R_{2}\right)+\Psi_{2} R_{2} \\
& r_{2}=\Psi_{1} R_{2}+\Psi_{2}\left(R_{2}+R_{3}\right)
\end{aligned}
$$

As we see, the calculation that involves magnetic circuits is very similar to a calculation that we have already performed on electrical circuits. In the latter case, the electrical current was confined to flow in a high conductivity wire. In magnetic circuits, the magnetic flux is confined to the region of a high permeability iron.

## Inductance

### 2.15. Inductance

The inductance of an object is defined as the ratio of the magnetic flux linkage $\Lambda$ divided by the current flowing through the object. The magnetic flux linkage is defined as the total magnetic flux $\Psi_{m}$ linking the object. There are cases where the magnetic flux linkage is not equal to the total magnetic flux. For example, these two entities will be different in a solenoid as will be shown below. There are also cases where the two are equal such as two current carrying parallel wires. Using the definition, we will be able to obtain the inductance of several important examples that are useful in electronic circuits and are found in the guided propagation of electromagnetic waves.

The inductance of an object is written as

$$
\begin{equation*}
\mathrm{L}_{\mathrm{jk}} \equiv \frac{\Lambda_{\mathrm{j}}}{\mathrm{I}_{\mathrm{k}}} \quad(\mathrm{H}) \tag{2.164}
\end{equation*}
$$

where the subscript indicates the magnetic flux linkage at an object $j$ and the current that flows through an object k. If the current is flowing through the object $(\mathrm{j}=\mathrm{k})$, this is defined as the self inductance. If the current is flowing through a different object but creating a magnetic field at another location ( $j \neq k$ ), this is called the mutual inductance. In order to avoid getting bogged down with subscripts in simpler circuits, it is common to use the symbols $L$ and $M$ respectively for these two quantities.


Figure 2-41. A solenoid consists of N coils of wire that are uniformly wrapped around a core whose radius is a and length d.

## Inductance

Let us first find the self inductance of a solenoid which is created by wrapping wire around a core as shown in Figure $2-41$. We are assuming that the length of the solenoid $d$ is much greater than its radius a which allows us to approximate the magnetic flux density to be a constant over the cross-sectional area at its midpoint. There are N turns of wire that are uniformly distributed around the core. This core could either be air or a magnetic material. The magnetic flux density at the center of the solenoid was obtained in Example 2-24 and can be written as

$$
\begin{equation*}
\mathrm{B}_{\mathrm{z}}=\frac{\mu \mathrm{NI}}{\mathrm{~d}} \tag{2.165}
\end{equation*}
$$

where we have incorporated the representation for the permeability $\mu=\mu_{r} \mu_{\mathrm{o}}$. The total magnetic flux is just the product of the magnetic flux density times the crosssectional area $\pi \mathrm{a}^{2}$ of the solenoid. We obtain

$$
\begin{equation*}
\Psi_{\mathrm{m}}=\frac{\mu \mathrm{NI}}{\mathrm{~d}} \pi \mathrm{a}^{2} \tag{2.166}
\end{equation*}
$$

The magnetic flux linkage $\Lambda$ is equal to the magnetic flux linking all $N$ turns of the solenoid or

$$
\begin{equation*}
\Lambda=N \Psi_{\mathrm{m}}=\frac{\mu \mathrm{N}^{2} \mathrm{I}}{\mathrm{~d}} \pi \mathrm{a}^{2} \tag{2.167}
\end{equation*}
$$

Therefore, the self inductance of the solenoid is calculated from (2.164) to be

$$
\begin{equation*}
\mathrm{L}=\frac{\mu \mathrm{N}^{2}}{\mathrm{~d}} \pi \mathrm{a}^{2} \tag{2.168}
\end{equation*}
$$

We note that the value increases as the square of the number of turns increases. In addition, larger values of inductance can be obtained if an iron core is inserted within the coils.

The calculation of the self inductance of a coaxial cable which will be later required when we talk about transmission lines can also be easily performed. In this case, the magnetic flux linkage $\Lambda$ will be equal to the total magnetic flux $\Psi_{m}$. The length of the coaxial cable will be $\Delta z$. The radius of the inner conductor is a

## Inductance

and the radius of the outer conductor is $b$ as shown in Figure 2-42. Between the two conductors, there is a vacuum. We will assume that there is a uniform current that flows in the $+\mathbf{u}_{\mathbf{z}}$ direction in the inner conductor and returns in the outer conductor of the coaxial cable. This assumed current will be equal to the current that is in the denominator of (2.164) and will therefore not appear in the final result for the inductance. The total magnetic flux that passes through the shaded region between the two conductors is computed by just integrating the magnetic flux density over the shaded area. Because of the cylindrical symmetry that is found in a coaxial cable, we can use Ampere's law (2.108) in order to obtain the magnetic flux density between the two conductors.


Figure 2-42. A section of a coaxial cable. The magnetic flux density passes through the shaded region between the two conductors.

Since we are only interested in the magnetic flux density external to the inner conductor, we can use a result that we have already obtained (2.115) which we rewrite here.

$$
\begin{equation*}
B_{\phi}=\frac{\mu_{0} I}{2 \pi r} \tag{2.169}
\end{equation*}
$$

## Inductance

The total magnetic flux that passes through the shaded region in Figure 2-42 is given by

$$
\begin{equation*}
\Lambda=\Psi_{\mathrm{m}}=\int_{z=0}^{z=\Delta z} \int_{r=a}^{r=b} \frac{\mu_{0} \mathrm{l}}{2 \pi \mathrm{r}} \mathrm{drdz}=\frac{\mu_{0} \mathrm{l}}{2 \pi} \ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right) \Delta z \tag{2.170}
\end{equation*}
$$

The substitution of (2.170) into the definition for the inductance (2.164) leads to the self inductance of a coaxial cable as being

$$
\begin{equation*}
\mathrm{L}=\frac{\mu_{0}}{2 \pi} \ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right) \Delta \mathrm{z} \tag{2.171}
\end{equation*}
$$

The value of the inductance will increase as the length of the core axial cable $\Delta z$ increases.

Example 2-38. Calculate the self inductance between two parallel planes. This object is frequently called a "microstrip" or "microstrip line" transmission line and it is important for the understanding of integrated circuits. The separation distance d can be assumed to be much less than the width w .


Answer: Since the current in each of the each of the conductors is in the opposite direction, the magnetic field between the two conductors will be in the same direction. This magnetic field will pass through the rectangular area defined by the separation distance d and the length $\Delta \mathrm{z}$. An approximate value for the magnetic flux density between the two conducting strips can be obtained from Ampere's law (2.108). For one of

## Inductance

the strips, we write $2 w B \approx \mu_{0} I$. Superposition applies and we make the assumption that the magnetic flux density is uniform in the entire region between the two strips. Therefore, the total magnetic flux $\Psi_{m}$ is given by

$$
\Psi_{m}=\int_{z=0}^{z=\Delta z} \int_{y=0}^{y=d} \frac{\mu_{0} l}{w} d y d z=\frac{\mu_{0} l}{w} d \Delta z
$$

In this case, the total magnetic flux will be equal to the magnetic flux linkage. The inductance is computed from (2.164) to be

$$
\mathrm{L}=\mu_{\mathrm{o}} \frac{\mathrm{~d}}{\mathrm{w}} \Delta \mathrm{z}
$$

It is possible to calculate the inductance of a torroid using the same approximations that were used to calculate the inductance of a solenoid. The calculation of the inductance between two parallel wires will be described in Appendix 4 since it requires certain mathematical approximations or numerical techniques. It will be very important in our later consideration of transmission lines.

Example 2-39. Calculate the mutual inductance $M$ between two small parallel circular loops that have surface areas $s_{1}$ and $s_{2}$ that are separated by a large distance $d$. There are $N_{1}$ turns that have a current $I_{1}$ in the first loop and $N_{2}$ turns in the second loop. The dashed lines indicate the magnetic flux density which is assumed to be approximately constant inside the loops.


Answer: The mutual inductance between the two loops is calculated using the definition of a mutual inductance $L_{12} \equiv M=\frac{\Lambda_{12}}{I_{1}} \approx \frac{N_{2} B_{1} S_{2}}{I_{1}}$. We use the results of Example 2-30 where the magnetic field on the axis from a small current carrying loop was calculated. Assuming that the separation distance is greater than the radius of either loop, we write $B_{1}=\frac{\mu_{0} N_{1} l_{1} s_{1}}{2 \pi d^{3}}$. The mutual inductance is

$$
M \approx \frac{\mu_{0} N_{1} N_{2} s_{1} s_{2}}{2 \pi d^{3}}
$$

The calculation of the mutual inductance is important in understanding transformers.

The magnetic energy that is stored in the elements that have just been described can be calculated using material that has been described in a course in the circuit theory. In this case, we assume that a current source is connected to the inductance and the value of the current I increases in time. There will be a voltage V that is generated across the terminals of the inductance. The magnetic energy $\mathrm{W}_{\mathrm{m}}$ is just equal to

$$
\begin{equation*}
\mathrm{W}_{\mathrm{m}}=\int_{\mathrm{t}^{\prime}=0}^{\mathrm{t}^{t^{\prime}}} \mathrm{pdt}=\int_{\mathrm{t}^{\prime}=0}^{\mathrm{t}^{\prime}=\mathrm{t}} \mathrm{IVdt} \mathrm{t}^{\prime}=\int_{\mathrm{t}^{\prime}=0}^{\mathrm{t}^{\prime}=\mathrm{t}} \mathrm{I}\left(\mathrm{~L} \frac{\mathrm{dl}}{\mathrm{dt}^{\prime}}\right) \mathrm{dt}^{\prime}=\mathrm{L} \int_{\mathrm{t}^{\prime}=\mathrm{t}=0}^{\mathrm{t}^{\prime} \mathrm{t}} I \mathrm{dI}=\frac{1}{2} \mathrm{LI}^{2} \tag{2.172}
\end{equation*}
$$

For the solenoid, the magnetic energy is equal to

$$
\begin{equation*}
\mathrm{W}_{\mathrm{m}}=\frac{1}{2}\left(\frac{\mu \mathrm{~N}^{2}}{\mathrm{~d}} \pi \mathrm{a}^{2}\right)\left(\frac{\mathrm{dB}_{\mathrm{z}}}{\mu \mathrm{~N}}\right)^{2} \tag{2.173}
\end{equation*}
$$

where we have replaced the current I with the magnetic flux density using (2.165). The volume $\Delta \mathrm{v}$ within the solenoid is equal to $\mathrm{d} \pi \mathrm{a}^{2}$. Therefore, (2.173) can be written as $W_{m}=\frac{1}{2} \frac{B_{z}{ }^{2}}{\mu} \Delta v$. The total magnetic energy that is stored within a volume $\Delta \mathrm{v}$ is obtained from an integration of the magnetic energy density within that volume

$$
\begin{equation*}
W_{m}=\int_{\Delta v}\left(\frac{B^{2}}{2 \mu}\right) d v \tag{2.174}
\end{equation*}
$$

## Boundary conditions

### 2.16. Boundary conditions

Up to this point, it may have appeared that electromagnetic fields would be found only in an infinite space. There has been no consideration of any boundaries that may have been introduced into the region. This is a very idealized situation and it certainly does not describe the real world. In this section, we will derive the necessary boundary conditions that will allow us to ascertain the relationship between an electromagnetic field in one region in space and an electromagnetic field in an adjacent region in space. The interface between the two regions is considered to have an infinitesimal thickness although each of the regions is considered to be semi infinite in size. The material parameters such as the relative dielectric constant and the relative permeability constant in the two regions may differ in value. The boundary conditions that we will obtain will also be applicable for time varying electromagnetic fields.


Figure 2-43. Two different materials are in juxtaposition. The relative dielectric constant and relative permeability constants are known for each material. The electromagnetic fields in one region are related to those in the other region. The interface is located at $z=0$.

## Boundary conditions

The electromagnetic fields can be separated into a component that is normal to the interface and a component that is perpendicular to the interface. It is reasonable to construct a two-dimensional coordinate system where one of the axes is normal to the interface and the other axis is tangent to the interface between the two materials as shown in Figure 2-43. The components for the displacement flux density and the magnetic flux density are obtained separately using the physical intuition that we have gained in the previous sections.


0
Figure 2-44. The interface between two materials is enclosed with a "pillbox" whose thickness $\Delta z \rightarrow 0$ and whose cross-sectional area is equal to $\Delta s$. There could be a surface charge density $\rho_{\mathrm{s}}\left(\mathrm{C} / \mathrm{m}^{2}\right)$ that is distributed at the interface.

The first boundary conditions involve the normal components of the displacement flux density $\mathbf{D}$ and the magnetic flux density $\mathbf{B}$ that is shown in Figure 2-44. The interface is encapsulated with a "pillbox" that has a crosssectional area $\Delta \mathrm{s}$ and a thickness that will satisfy the condition that the thickness $\Delta z \rightarrow 0$. We will first examine the displacement flux density.

Charge could be distributed just at the interface that has a surface charge density of $\rho_{\mathrm{s}}\left(\mathrm{C} / \mathrm{m}^{2}\right)$. The application of Gauss's law (2.92) will lead will lead to

## Boundary conditions

be boundary condition for the normal component of the displacement flux density. We rewrite this law

$$
\begin{equation*}
\oint D \cdot \mathbf{d s}=Q_{\mathrm{enc}} \tag{2.175}
\end{equation*}
$$

The charge that is enclosed within the pillbox is $Q_{e n c}=\rho_{s} \Delta s$ where $\rho_{\mathrm{s}}$ is the surface charge density that will be assumed to be independent of position. Recall that a closed surface integral requires that the scalar product of the differential surface area $\Delta \mathbf{s}$ and the normal component of the displacement flux density $D$ must be performed and this leads to $\left(D_{2 n}-D_{1 n}\right) \Delta s$. Therefore, (2.175) leads to the following boundary condition for the normal components of the displacement flux density

$$
\begin{equation*}
D_{2 n}-D_{1 n}=\rho_{s} \tag{2.176}
\end{equation*}
$$

This equation leads to the conclusion that the normal components of the displacement flux density differ by the surface charge density that exists at the interface of the two materials. If the surface charge density is equal to 0 , then the normal component of the displacement flux density is continuous.

Using the relationship between the displacement flux density and the electric field intensity which depends upon the relative dielectric constant of the material (2.93), we can write (2.176) as

$$
\begin{equation*}
\varepsilon_{2} \varepsilon_{0} E_{2 n}-\varepsilon_{1} \varepsilon_{0} \mathrm{E}_{1 \mathrm{n}}=\rho_{\mathrm{s}} \tag{2.177}
\end{equation*}
$$

A similar derivation can be used to find the normal component of the magnetic flux density. In this case, we invoke the experimental fact that magnetic monopoles
have not yet been observed to exist in nature. This is manifested in (2.106) and it is rewritten below

$$
\begin{equation*}
\oint B \cdot d s=0 \tag{2.178}
\end{equation*}
$$

Applying (2.178) to Figure 2-44, we write

$$
\begin{equation*}
\mathrm{B}_{2 \mathrm{n}}-\mathrm{B}_{1 \mathrm{n}}=0 \tag{2.179}
\end{equation*}
$$

## Boundary conditions

This implies that the normal components of the magnetic flux density will always be continuous at all points in the space.

The magnetic flux density and magnetic field intensity is related via (2.161). Therefore, (2.179) can also be written as

$$
\begin{equation*}
\mu_{2} \mu_{0} H_{2 n}-\mu_{1} \mu_{0} H_{1 n}=0 \tag{2.180}
\end{equation*}
$$

In this equation, the permeability of free space $\mu_{o}$ can be canceled.


Figure 2-45. The tangential components of the electric field intensity and the magnetic field intensity in the two materials are indicated. The rectangular loop has the dimensions $\Delta x$ and $\Delta z$. The coordinate y is directed out of the page.

We will now derive the relation for the tangential components of the electromagnetic fields that are shown in Figure 2-45. For the tangential components of the electric field, we calculate the total work that is expended in moving around a closed path and apply (2.68) which is rewritten as

$$
\begin{equation*}
\oint E \cdot d \mathbf{d}=0 \tag{2.181}
\end{equation*}
$$

The application of $(2.181)$ to the Figure 2-45 leads to the conclusion that the contributions at the top and at the bottom can be neglected since the tangential

## Boundary conditions

electric field is perpendicular to the path of integration. However, the integration along the other two edges yields the result

$$
\left(\mathrm{E}_{2 \mathrm{t}}-\mathrm{E}_{1 \mathrm{t}}\right) \Delta \mathrm{x}=0
$$

or

$$
\begin{equation*}
\mathrm{E}_{2 \mathrm{t}}-\mathrm{E}_{1 \mathrm{t}}=0 \tag{2.182}
\end{equation*}
$$

The tangential components of the electric field are continuous at an interface. We will later find that this boundary condition will be extremely important in understanding the warning that appears in the operations manual of a microwave oven in which you are told to "Never put a metal pan inside the microwave oven" since it also applies to time varying electromagnetic fields.

This can also be written in terms of the displacement flux density as

$$
\begin{equation*}
\frac{\mathrm{D}_{2 \mathrm{t}}}{\varepsilon_{2}}=\frac{\mathrm{D}_{1 \mathrm{t}}}{\varepsilon_{1}} \tag{2.183}
\end{equation*}
$$

where the permittivity of free space $\varepsilon_{0}$ has been canceled from this equation.
The boundary condition for the tangential component of the magnetic field intensity $\mathbf{H}$ is also depicted in Figure 2-45. In this case, there could be a surface current $\mathbf{J}_{\mathbf{L}} \bullet \mathbf{u}_{\mathbf{y}}$ that is enclosed within the rectangular loop and directed out of the page. Therefore, we must evaluate the integral in (2.160) which we restate as

$$
\begin{equation*}
\oint \mathrm{H} \bullet \mathbf{d l}=\mathrm{I}_{\text {enclosed }} \tag{2.184}
\end{equation*}
$$

The current that is enclosed within the loop is given by

$$
\begin{equation*}
I_{\text {enclosed }}=J_{s} \Delta x \tag{2.185}
\end{equation*}
$$

and it resides an entirely at the interface between the two materials. If this surface current is directed in the $\mathbf{u}_{\mathbf{x}}$ direction, it will not pass through this rectangular loop and it will not affect this particular tangential component of the magnetic field intensity. Once again, we can neglect the integration over the top and bottom edges. The resulting integration yields the following result

$$
\left(\mathrm{H}_{2 \mathrm{t}}-\mathrm{H}_{1 \mathrm{t}}\right) \Delta \mathrm{x}=\mathrm{J}_{\mathrm{L}} \Delta \mathrm{x}
$$

## Boundary conditions

or

$$
\begin{equation*}
\mathrm{H}_{2 \mathrm{t}}-\mathrm{H}_{1 \mathrm{t}}=\mathrm{J}_{\mathrm{L}} \tag{2.186}
\end{equation*}
$$

The tangential components of the magnetic field intensity differ by the surface current density that passes between the two materials. This conclusion also applies to time varying electromagnetic fields. If the surface current density is equal to 0 , then the tangential components of the magnetic field intensity will be continuous. This can also be written in terms of the magnetic flux density as

$$
\begin{equation*}
\frac{\mathrm{B}_{2 \mathrm{t}}}{\mu_{2}}-\frac{\mathrm{B}_{1 \mathrm{t}}}{\mu_{1}}=\mu_{\mathrm{o}} \mathrm{~J}_{\mathrm{L}} \tag{2.187}
\end{equation*}
$$

Example 2-40. Calculate the change of electric field as it crosses an interface between the two dielectrics shown below. There is no charge distributed on the surface between the two dielectrics.


Answer: Since there is no surface charge between the two surfaces, we know that the normal components of the displacement flux density are continuous

$$
\varepsilon_{1} E_{1} \cos \theta_{1}=\varepsilon_{2} E_{2} \cos \theta_{2}
$$

The continuity of the tangential components of the electric field is given by

$$
E_{1} \sin \theta_{1}=E_{2} \sin \theta_{2}
$$

The ratio of these two terms gives

$$
\frac{\tan \theta_{1}}{\tan \theta_{2}}=\frac{\varepsilon_{1}}{\varepsilon_{2}}
$$

This example indicates that the electric field can be bent with different dielectrics.

## Boundary conditions

Example 2-41. Two homogeneous, linear, isotropic magnetic materials have an interface at $x=0$. At the interface, there is a surface current with a density $J_{\mathbf{L}}=20 \mathbf{u}_{\mathbf{y}} A / m$. The relative permeability $\mu_{\mathrm{r} 1}=2$ and the magnetic field intensity in the region $x<0$ is $H_{1}=15 u_{\mathbf{x}}+10 u_{\mathbf{y}}+25 u_{z} A / m$. The relative permeability $\mu_{r 2}=$ 5 in the region $x>0$. Find the magnetic field intensity in the region $x>0$.


Answer: The magnetic field is separated into components and the boundary conditions are applied independently. For the normal component, (2.180) implies

$$
\mu_{2} \mu_{\mathrm{o}} \mathrm{H}_{2 \mathrm{n}}=\mu_{1} \mu_{\mathrm{o}} \mathrm{H}_{1 \mathrm{n}} \Rightarrow \mathrm{H}_{2 \mathrm{x}}=\frac{\mu_{1}}{\mu_{2}} \mathrm{H}_{1 \mathrm{x}}=\frac{2}{5} 15=6 \mathrm{~A} / \mathrm{m}
$$

For the tangential components, we use (2.186) and write

$$
\mathrm{H}_{2 \mathrm{t}}-\mathrm{H}_{1 \mathrm{t}}=\mathrm{J}_{\mathrm{s}} \Rightarrow \mathrm{H}_{2 \mathrm{z}}=\mathrm{H}_{1 \mathrm{z}}+\mathrm{J}_{\mathrm{y}}=25+20=45 \mathrm{~A} / \mathrm{m}
$$

There is no change in the y component of the magnetic field intensity. Therefore,

$$
\mathbf{H}_{\mathbf{2}}=6 \mathbf{u}_{\mathbf{x}}+10 \mathbf{u}_{\mathbf{y}}+45 \mathbf{u}_{\mathbf{z}} \mathrm{A} / \mathrm{m} .
$$

If one of the materials is an ideal conductor, the tangential component of the electric field will be affected in the most dramatic fashion. Since (2.182) states that the tangential electric field must be continuous, there would also be a tangential electric field at the edge of the ideal conductor. From Ohm's law (2.101), this would then imply that there would also be a very large tangential current that could approach infinity. There would then be an infinite amount of power dissipation within the ideal conductor that nature would not allow. The

## Boundary conditions

conclusion that must be drawn from this argument is that the tangential current density at the edge of the conductor and therefore the tangential electric field at the edge of an ideal conductor must be equal to 0 . The ideal conductor is an equipotential surface.


Figure 2-46. A charge a distance $d$ above a grounded metallic surface induces a nonuniform charge distribution on the surface. The electric potential and electric field intensity can be calculated by assuming an image charge of the opposite sign exists beneath the surface and the metallic surface being removed.

The requirement that the tangential component of an electric field must be equal to 0 has some very important consequences. For example, let us pose the problem, "What happens if a negative charge is placed above an ideal conductor as shown in Figure 2-46?" From our earlier studies of electric fields, we found that the vector direction of the electric field would be entirely in the radial direction $(2-8)$ which we rewrite here

$$
\begin{equation*}
\mathbf{E}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{R}^{2}} \mathbf{u}_{\mathrm{R}} \tag{2.188}
\end{equation*}
$$

The tangential component of the electric field that follows from (2.188) will be equal to 0 only at the point just beneath the charge as shown in Figure 2-46. What happens at the other points on the surface?

## Boundary conditions

This last question can be answered by assuming that there is a charge of the opposite sign that is inhomogeneously distributed upon the surface of the ideal conductor. It is usually difficult to calculate this distribution and there is a method to circumvent this calculation. This method assumes that there is an additional charge that is located an equidistance beneath the ideal conductor. This particular charge that is called an image charge will have a value that is equal to the negative value of the original charge that was placed above the ideal conductor. In order to calculate the resulting electric field, we remove this ideal conductor and use the principal of superposition since we're talking about a linear material. The tangential component of the electric field will then be equal to 0 .

The use of image charges can be later generalized to time varying electric fields where one frequently encounters "image antennas" lying beneath the ground plane. The conductivity of this ground plane has been improved by burying copper wires in the ground that radiate in the radial direction away from the antenna.

Example 2-42 ${ }^{\text {MATLAB }}$. Two grounded semi infinite metallic plates are placed on the $x$ axis and the $y$ axis of a Cartesian coordinate system in order to form a $90^{\circ}$ corner. A positive charge $4 \pi \varepsilon_{0}$ is located at the point $(a, a)$ where $a$ is arbitrary. Using MATLAB, carefully plot equipotential contours surrounding this charge and the expected electric field.

Answer: There will be 2 negative image charges located at the points $(a,-a)$ and (-a, a). In addition, there will be a positive image charge at $(-a,-a)$ since the electric potential at all points of the metallic corner must be equal to 0 . This system of four alternative charges is often called a "quadripole". The results of the calculation are shown below. We find that the electric field is always perpendicular to the metal surfaces. This indicates that the metallic surfaces are equipotential contours that have a value equal to 0 .

## Boundary conditions



In high-energy particle experiments in which charged particles move through a metallic tube and are to collide together, the experimentalist must be aware of these image charges. The image charges can be found for cases only when the corner has an angle that is a submultiple of $2 \pi$.

## Boundary conditions

The boundary conditions can also be written down in vector notation. A summary of these boundary conditions is presented in tabular form. In this table, the unit vector $\mathbf{u}_{\mathbf{n}}$ is normal to the interface.

| Tangential component of the electric field intensity $\mathbf{E}$ | $\left(\mathbf{E}_{2}-\mathbf{E}_{1}\right) \times \mathbf{u}_{n}=0$ |
| :--- | :--- |
| Normal component of the displacement flux density $\mathbf{D}$ | $\left(\mathbf{D}_{2}-\mathbf{D}_{1}\right) \bullet \mathbf{u}_{\mathrm{n}}=\rho_{\mathrm{s}}$ |
| Tangential component of the magnetic field intensity $\mathbf{H}$ | $\left(\mathbf{H}_{2}-\mathbf{H}_{1}\right) \times \mathbf{u}_{\mathrm{n}}=\mathbf{J}_{\mathrm{L}}$ |
| Normal component of the magnetic flux density $\mathbf{B}$ | $\left(\mathbf{B}_{2}-\mathbf{B}_{1}\right) \bullet \mathbf{u}_{\mathrm{n}}=0$ |
| Tangential component of the electric field intensity $\mathbf{E}$ at a <br> metallic surface - This directly follows from the first | $\mathbf{E} \times \mathbf{u}_{\mathrm{n}}=0$ |
| boundary condition and the fact that the current in a |  |
| conductor must be equal to 0. |  |

## Conclusion

### 2.17.Conclusion

Our study of time-independent electric and magnetic fields has brought forth the many similarities and differences between the two fields. Both the Coulomb force for electric fields and Ampere's force for magnetic fields are proportional to the magnitudes of charges or current elements and inversely proportional to the square of the distances separating the entities. Both fields require a charge as a source in that stationary charges yield electric fields and charges in motion with a constant velocity yield electric and magnetic fields. In a vacuum, superposition principles apply to both fields because of a linear relationship between the sources and the fields. Each field could be related to an appropriate potential with a prescribed vector operation. By considering atoms as small electric or magnetic dipoles, we have found that materials may effect the characteristics of the both fields.

The major difference between the fields is that electric fields can originate or terminate on a charge which implies that there can be a nonzero divergence of the field. A surface that surrounded this charge could be used to ascertain the electric field. Magnetic fields are closed paths and do not terminate at a fixed point. If this closed path enclosed a current which implies a nonzero curl, the magnetic field could be found.

The elements that are frequently encountered in the study of electric circuits and magnetic circuits have been obtained from the fundamentals of static fields. In particular, we have obtained the resistance, the inductance, and the capacitance for electrical circuits and the reluctance for magnetic circuits. We will encounter these entities later in this book.

## Problems 2

### 2.18. Problems

$1^{\text {MATLAB }}$. Find the force on a charge $Q$ that is located at the point $(6,3)$. The charge has a value of $Q=-3 C$. The value of the other charges that are located at $Q_{1}(2,0)=+2 C, Q_{2}(2,3)=-4 C$, and $Q_{3}(2,6)=-3 C$. All dimensionless are in meters.
$2^{\text {MATLAB }}$. For the charges given in problem 1, find the force on the charge $Q_{2}$.
3. Four equal charges (Q) are placed at the corners of a square whose dimensions are ( $a \times a$ ). Find the magnitude of the force on one of the charges.
4. Two small plastic balls, each with equal values of charge $Q$ and mass $M$ are constrained to slide on an insulating string. Find the separation of the two charges if the lower
 charge is constrained to one location on the string.
$5^{\text {MATLAB }}$. Find the electric field at the point $P$ that is located at $(5,1)$ due to the charges $Q_{1}(1,4)=+2 C$ and $Q_{2}(1,1)=+4 C$ where the coordinates are measured in the units of meters.
$6^{\text {MATLAB }}$. Find the electric field at the point $P$ that is located at $(5,1)$ due to the charges $Q_{1}(1,4)=+2 C$ and $Q_{2}(1,1)=-4 C$ where the coordinates are measured in the units of meters.
7. Four equal charges $Q=1 C$ are located at the corners of a square ( $a \times a$ ) that is centered at the origin of a coordinate system in the xy plane. The dimension a $=1 \mathrm{~m}$. Find the ratio of the electric potentials along the $z$ axis at the points $z=2 \mathrm{a}$ and $z=a$. Also, find the ratio of the magnitude of the electric fields at $z=2 a$ to that at $z=a$.

## Problems 2

$8^{\text {MATLAB }}$. Two charges of equal magnitude $Q=2\left(4 \pi \varepsilon_{0}\right) C$ but with the opposite sign are located at the points $(1,2)$ and at $(5,2)$ respectively. Find the electric field $\mathbf{E}$ on the line that would correspond to $x=3$. Plot your results with MATLAB in the region $(-4<y<8)$.
$9^{\text {MATLAB }}$. Find the absolute potential at an arbitrary point $P(x, y, z=0)$ due to charges $Q_{1}(2,2)=+2\left(4 \pi \varepsilon_{0}\right) C$ and $Q_{2}(4,5)=-4\left(4 \pi \varepsilon_{0}\right) C$. Plot your results with MATLAB in the region ( $0<x<6,0<y<7$ ).
$10^{\text {MATLAB }}$. On log-log graph paper, accurately plot the magnitude of the electric field $\mathbf{E}$ and the voltage V from a point charge as a function of the distance from the charge. Assume that $\mathbf{E} @ \rho=1 \mathrm{~m}$ equals $1 \mathrm{~V} / \mathrm{m}$ and $\mathrm{V} @ \rho=1 \mathrm{~m}$ equals 1 V . 11. By measuring the potential difference between two small probes separated by a distance $\Delta L$, scientists are able to obtain data to plot electric field patterns caused by charges. By rotating the probes, the direction of the electric field can be obtained. From the following two measurements, find the location of the positive charge $Q$. At the origin, the probe measures a maximum electric field in the direction $\mathbf{u}_{\mathbf{x}}-\sqrt{ } 3 \mathbf{u}_{\mathbf{y}}$. If the probes are moved to the point $(0,1)$, the maximum electric field is in the direction $\mathbf{u}_{\mathbf{x}}-\mathbf{u}_{\mathbf{y}}$.

12. In terms of the fundamental units mass $M$, length $L$, time $T$, and charge $Q$, find the units of the permittivity of free space $\varepsilon_{0}$.
13. In terms of the fundamental units mass $M$, length $L$, time $T$, and charge $Q$, find the units of the electric field $\mathbf{E}$.
14. Assume that $\mathrm{Q}=\left(4 \pi \varepsilon_{0}\right) \mathrm{C}$ of charge is uniformly distributed along a line of length 2 L . Find an analytical solution for the electric field along a line that is perpendicular to the line and is located at the center of the charged line.
$15^{\text {MATLAB }}$. Using MATLAB, perform the integration in problem 14 and plot the results.
$16^{\mathrm{MATLAB}}$. Assume that $Q$ of charge are uniformly distributed on the ring defined by $a \leq r \leq b$. Find an analytical expression for the electric field along the $z$ axis. Use MATLAB to display the electric field distribution.

$17^{\text {MATLAB }}$. Find an analytical expression for the electric field along the $z$ axis of a uniformly charged disc with radius a. Plot the field distribution.
18. Set up the integral using the ring configuration in problem 16 to find the normal electric field from an infinite charged sheet. Evaluate this integral.
$19^{\text {MATLAB }}$. Find an analytical expression for the electric field along the $z$ axis of a circular loop with a radius a, uniformly charged. Display the electric field distribution.
$20^{\text {MATLAB }}$. Find an analytical expression for the potential along the $z$ axis of a circular loop in problem 19 assuming it is zero in infinity. Display the potential profile.
21. Calculate the work that is expended in moving a positive charge $Q$ from point $A$ to $B$ along the indicated path. $A$ charged infinite plane with a charge density $\rho_{S}\left(\mathrm{C} / \mathrm{m}^{2}\right)$ exists in the yz plane at $x=0$.


## Problems 2

22. Given the electric field $E=4 x u_{x}-2 y u_{Y} V / m$, find the voltage between the points $A(2,0)$ and $B(0,2)$ integrating along: a) straight line $A B ; b)$ broken line $A O B$, where $\mathrm{O}(0,0)$.
23. Assume that there are two concentric cylinders with radii $a<r<b$. Show that the electric field between the two cylinders approaches a constant value as the separation distance $(b-a) \rightarrow 0$.
24. Assume that there are two concentric spheres with radii $a<r<b$. Show that the electric field between the two spheres approaches a constant value as the separation distance $(b-a) \rightarrow 0$.
25. A lightning rod provides a controlled path for a cloud to discharge itself to the ground in a safe way. Sketch the expected electric field distribution about the rod.

Pay particular attention to the electric
 field that you would expect at the tip of the lightning rod.
$26^{\text {MATLAB }}$. Compare the electric potential from a single charge, a dipole charge, and a quadripole charge by accurately plotting the potential as a function of distance from the region of the charge. Use MATLAB and assume that the potential equals 1 V at a distance of 1 m from the charge.
$27^{\text {MATLAB }}$. Charge is nonuniformly distributed within a sphere of radius a as $\rho_{\mathrm{v}}=\rho_{\mathrm{o}} \frac{\rho}{\mathrm{a}}$. Using Gauss's law, calculate the electric field in the regions $\rho<a$ and $\rho>a$. Accurately sketch the electric field and label the axes.

## Problems 2

$28^{\text {MATLAB }}$. A charge of $+Q C$ is uniformly distributed in the central region $0 \leq r \leq a$ and $-Q C$ is distributed in the external region $b \leq$ $r \leq c$ of a coaxial cable. Find the electric field in all regions of the
 coaxial cable and sketch the results. $29^{\text {MATLAB }}$. Repeat problem 28 for a charge of +2 Q $C$ uniformly distributed in the inner region and -Q C distributed in the outer region of a coaxial cable. Find the electric field in all regions of the coaxial conductor and sketch the results.
30. Assume that a uniform charge density $+\rho_{\mathrm{s}}$ $\left(\mathrm{C} / \mathrm{m}^{2}\right)$ exists on one infinite plane and $-\rho_{\mathrm{s}}\left(\mathrm{C} / \mathrm{m}^{2}\right)$ exists on the other infinite plane. The two planes are parallel and are separated by a distance $d$. Using Gauss's law, find the electric field in the regions between the two plates and external to the two plates.
$31^{\text {MATLAB }}$. Charges are placed on the concentric hollow spheres. The values are: $Q(\rho=a)=2 C$; $Q(\rho=b)=-4 C$; and $Q(\rho=c)=4 C$ where $a<b<c$. Find and sketch the electric field in all regions $\rho \geq 0$.

32. Find the voltage between two small electrodes $A$ and $B$ for the system described in problem 31: A is fixed at the center $(\rho=0)$ while $B$ is moving at positions $\rho=\mathrm{a}, \rho=\mathrm{b}$ and $\rho=\mathrm{c}$.
33. Calculate the total current that passes through the indicated surface. The current density $\mathbf{J}=3 \mathbf{x u}_{\mathbf{x}}+4 \mathbf{x u}_{\mathbf{y}}$ $\mathrm{A} / \mathrm{m}^{2}$.

34. Calculate the total current that passes through the indicated surface. The current density $\mathbf{J}=3 \mathbf{x u}_{\mathbf{x}}+4 \mathbf{x u}_{\mathbf{z}}$ $\mathrm{A} / \mathrm{m}^{2}$.

35. The current density in a cylindrical wire whose radius is a is given by

$$
\mathbf{J}=J_{0}\left(1-\frac{r^{2}}{a^{2}}\right) \mathbf{u}_{\mathbf{z}}
$$

Compute the total current in the wire.
36. A lightning stroke with a current $\mathrm{I}=10 \mathrm{kA}$ is fallen to a post. Find the voltage $V_{a b}$ between the legs of a man, situated on a distance $a=5 m$ from the post, if the step-size is $L=0.6 m$ (see the figure). Assume that the ground is a homogeneous one with conductivity $\sigma=0.02 \mathrm{~S} / \mathrm{m}$.

## Problems 2


37. A wire lying on the $x$ axis carries a current 10 A in the positive $x$ direction. $A$ uniform magnetic field of $B=5 \mathbf{u}_{x}+10 u_{y} T$ exists in this region. Find the force per unit length of this wire.
38. Magnetohydrodynamics (MHD) is based on passing charged particles (both electrons and positive ions) through a uniform magnetic field $B=B_{0} \mathbf{u}_{x}$ with a velocity $\mathbf{v}=\mathrm{v}_{0} \mathbf{u}_{\mathbf{z}}$. The charged particles will separate due to their charge difference. Find the potential difference (called the Hall voltage) between two electrodes placed in the exhaust of a rocket engine ${ }^{1}$.

[^6]
## Problems 2


39. An electron is injected into a uniform magnetic field $B_{o}$. Find the value of the accelerating potential $V_{O}$ so that the electron will be captured in the
 electron gun.

40. Describe the motion of a charged particle in a uniform electric and magnetic field such that the net force acting upon the particle is equal to zero.
41. In terms of the fundamental units of mass $M$, length $L$, time $T$, and charge $Q$, find the units of the permeability $\mu_{o}$ of vacuum.
42. In terms of the fundamental units of mass $M$, length $L$, time $T$, and charge $Q$, show that the units of $\frac{1}{\sqrt{\mu_{\circ} \varepsilon_{o}}}$ have the same dimensions as a velocity.
$43^{\text {MATLAB }}$. Find and plot the magnetic flux density as a function of radius for the concentric hollow cylinders that carry currents I and 2 I in the indicated directions.

$44^{\text {MATLAB }}$. Find and plot the magnetic flux density as a function of radius for the concentric hollow cylinders that carry currents $2 I$ and $4 I$ in the indicated directions.

45. Due to a nonuniform conductivity of the material, the current density within a wire has a nonuniform distribution $\mathbf{J}=J_{0}\left(1-\frac{r^{2}}{a^{2}}\right) \mathbf{u}_{\mathbf{z}}$ for $0 \leq r \leq a$ and $\mathbf{J}=0$ for $r>a$. Find the magnetic flux density as a function of radius for $r>0$. 46. If a long straight wire with a resistance of $1 \Omega$ is connected to a 1 V battery, find the distance from the wire where the magnetic field equals 1 T .
47. An experimentalist desires to create a localized region where there is to be no magnetic field. Using a solenoid that contains 100 uniformly distributed turns that is connected to a power supply of 10 V , calculate the length of this region so the earth's magnetic field of $0.5 \mathrm{G}=5 \times 10^{-5} \mathrm{~T}$ can be canceled. The resistance of the wire $=0.1 \Omega$.
48. In terms of the fundamental units of mass $M$, length $L$, time $T$, and charge $Q$, find the units of the vector potential $\mathbf{A}$ and the magnetic flux density $\mathbf{B}$.
49. Two infinitesimal current elements are displaced by a distance 2 a on the $x$ axis. Find the vector potential $\mathbf{A}$ at the point $P$ that is at $y=b$. Find $B$ at $P$.

50. Use the Biot-Savart law to compute $B$ at $P$ for the configuration shown in Problem 49.
51. Show, using the Biot-Savart law that the magnetic flux density from a finite length current element carrying a current $I$ is given by

$$
B=\frac{\mu_{0} \mathrm{I}}{4 \pi \mathrm{r}}\left(\cos \theta_{1}-\cos \theta_{2}\right)
$$

Find the magnetic flux density profile in case $\theta_{2}=\pi-\theta_{1}$ (observation point on the axis of symmetry). Find the magnetic flux density in the limit of an infinitely long wire.

52. Show that the vector potential $\mathbf{A}$ of two parallel infinite straight wires carrying currents I in the opposite direction is given by

$$
\mathbf{A}=\frac{\mu_{\mathbf{o}} \mathrm{l}}{2 \pi} \ln \left(\frac{\mathrm{r}_{2}}{\mathrm{r}_{1}}\right) \mathbf{u}_{\mathbf{z}}
$$


53. Use Ampere's circuital law to compute $\mathbf{B}$ at the point $P$ in Problem 52.
54. Use the Biot-Savart law to compute $\mathbf{B}$ at the point $P$ in Problem 52.
55. Find the magnetic field along the $z$ axis caused by the current in the small square loop.

56. Prove the vector identity $\nabla\left(\frac{1}{R}\right)=-\frac{u_{R}}{R^{2}}$ where $u_{R}$ is the unit vector in the direction between a current element and a point of observation.
57. Find the approximate magnetic field at the center of a slender, rectangular loop that carries a current $I$. The dimensions of the loop are $L \gg W$.

58. Two parallel current carrying insulated wires are placed in close juxtaposition.

Find the magnetic flux density at the center of the circle.

## Problems 2


59. Four infinitely long wires, each carrying a current I into the paper, are located at the indicated points. Find the direction of the net force on wire 1.

60. It is desired to move an axle of mass $M$ up an inclined frictionless plane that makes an angle $\theta$ with respect to the horizontal plane. A uniform vertical magnetic field exists in the region of the inclined plane. What value must the battery $\mathrm{V}_{\mathrm{O}}$ be in order that the axle will climb the plane?


## Problems 2

61. Find the force of repulsion between two conductors whose length is $L$ of a planar transmission line with $b \gg d$. $A$ current $I_{0}$ flows in the opposite directions in the two conductors. Ignore any fringing magnetic fields.

$62^{\text {MATLAB }}$. Let $\mathbf{B}=0.1 \mathbf{u}_{\mathbf{x}}$ T. everywhere and let the relative permeability $\mu_{r}=2$ for $|x| \leq 1$ and $\mu_{r}=1$ for $|x|>1$. Plot $B_{x}, M_{x}$, and $H_{x}$ everywhere.
62. For the magnetic circuit shown, calculate the magnetic flux in the gap. The cross-section and relative permeability of the iron is $A$ and $\mu_{r} \gg 1$.


## Problems 2

64. Find the electric field in the region $x>0$ if $E=2 \mathbf{u}_{\mathbf{x}}+2 \mathbf{u}_{\mathbf{y}} \mathrm{V} / \mathrm{m}$ in the region $x<0$. There is no surface charge density.


65 Repeat problem 64 with a surface charge density $\rho_{s}=1 \mathrm{C} / \mathrm{m}^{2}$. 66. Find the electric field in the region $x>0$. There is no surface charge density. The magnitude of the electric field is $5 \mathrm{~V} / \mathrm{m}$.


67 Repeat problem 66 with a surface charge density $\rho_{s}=0.5 \mathrm{C} / \mathrm{m}^{2}$.

## Problems 2

68. Find the magnetic flux density in the region $x>0$ if $\mathbf{B}=4 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}} \mathrm{T}$ in the region $x<0$. The surface current $=0$.

69. Repeat problem 68 if the surface current density $J_{\mathbf{L}}=\left(9 \mathbf{u}_{\mathbf{z}}+9 \mathbf{u}_{\mathbf{y}}\right) \mathrm{A} / \mathrm{m}$.
70. Can the indicated electric field exist? If not, suggest how an additional electric field will permit the existence of this field.


## Poisson's and Laplace's equations

## 3. Boundary value problems using MATLAB

In the previous chapter, we learned that a static electric field would be created from a charge distribution. In addition, it was possible to determine this static electric field from a scalar potential. We also showed there that the potential V could also be obtained directly in terms of the charge distribution $\rho \mathrm{V}$ via one partial differential equation : Poisson's equation. This equation reduces to Laplace's equation if the charge density in the region of interest were equal to zero. The general procedure of solving these equations for the cases where the potential $V$ depended on only one or two spatial coordinates is given here. In this chapter, we will introduce analytical and numerical techniques that will allow us to examine such complicated problems. For mathematical simplicity, however, only problems that can be written in terms of Cartesian coordinates will be emphasized.

We also introduce numerical techniques for solution such problems (like Method of Moments and Finite Element Method) in this chapter and make extensive use of MATLAB in the process. Several MATLAB programs (dot $m$ files) that have been used in this book are available on the web site [http://www.scitechpub.com/] for the reader's benefit. These programs can be easily altered and customized by the user. They can also be translated into the reader's language of choice. They may also not be THE program for a particular task but they do work.

### 3.1. Poisson's and Laplace's equations

In the previous chapter, we learned that a static electric field $\mathbf{E}$ would be created in a vacuum from a volume charge distribution $\rho_{\mathrm{v}}$. This physical

## Poisson's and Laplace's equations

phenomenon was expressed through a partial differential equation. We have been able to write this partial differential equation using a general vector notation as

$$
\begin{equation*}
\nabla \bullet \mathbf{E}=\frac{\rho_{v}}{\varepsilon_{0}} \tag{3.1}
\end{equation*}
$$

which is Gauss's law in a differential form. Here we have applied a shorthand notation that is common for the vector derivatives by using a vector operator $\nabla$ called the "del operator", which in Cartesian coordinates is

$$
\begin{equation*}
\nabla \equiv \frac{\partial}{\partial \mathbf{x}} \mathbf{u}_{\mathbf{x}}+\frac{\partial}{\partial \mathbf{y}} \mathbf{u}_{\mathrm{Y}}+\frac{\partial}{\partial \mathbf{z}} \mathbf{u}_{\mathbf{z}} \tag{3.2}
\end{equation*}
$$

The static electric field is a conservative field which implies that

$$
\begin{equation*}
\nabla \times \mathbf{E}=0 \tag{3.3}
\end{equation*}
$$

This means that the electric field could be represented as the gradient of a scalar electric potential $V$

$$
\begin{equation*}
\mathbf{E}=-\nabla \mathrm{V} \tag{3.4}
\end{equation*}
$$

Recall the vector identity

$$
\nabla \times \nabla \mathrm{V}=0
$$

Combine (3.1) with (3.4) and obtain

$$
\begin{equation*}
\nabla^{2} \mathrm{~V}=-\frac{\rho_{\mathrm{V}}}{\varepsilon_{0}} \tag{3.5}
\end{equation*}
$$

where we have used the relation that $\nabla \bullet \nabla=\nabla^{2}$. Equation (3.5) is called Poisson's equation. If the charge density $\rho_{v}$ in the region of interest were equal to zero, then Poisson's equation is written as

$$
\begin{equation*}
\nabla^{2} V=0 \tag{3.6}
\end{equation*}
$$

## Poisson's and Laplace's equations

Equation (3.6) is called Laplace's equation. Both of these equations have received considerable attention since equations of this type describe several physical phenomena, e. g. the temperature profile in a metal plate if one of the edges is locally heated.

In writing (3.5) or (3.6), we can think of employing the definition for $\nabla^{2}$ that $\nabla^{2}=\nabla \bullet \nabla$. This operation is based on interpreting the $\nabla$ operator as a vector and a heuristic application of the scalar product of two vectors that leads to a scalar quantity. The resulting operator $\nabla^{2}$ is called a "Laplacian operator" Since each application of the $\nabla$ operator yields a first order differentiation, we should expect that the Laplacian operator $\nabla^{2}$ would lead to a second order differentiation.

The Laplacian operator depends on the coordinate system that is chosen for a calculation. The definitions for the operator $\nabla^{2}$ operating on a function $V$ in the three most commonly employed coordinate systems are written as follows

1) Cartesian ( $x, y, z$ )

$$
\begin{equation*}
\nabla^{2} V=\frac{\partial^{2} V}{\partial x^{2}}+\frac{\partial^{2} V}{\partial y^{2}}+\frac{\partial^{2} V}{\partial z^{2}} \tag{3.7}
\end{equation*}
$$

Using the definition for the del operator $\nabla$. in Cartesian coordinates (3.2) we write

$$
\nabla \bullet \nabla V=\left(\frac{\partial}{\partial \mathbf{x}} \mathbf{u}_{\mathbf{x}}+\frac{\partial}{\partial \mathbf{y}} \mathbf{u}_{Y}+\frac{\partial}{\partial \mathbf{z}} \mathbf{u}_{\mathrm{Z}}\right) \cdot\left(\frac{\partial}{\partial \mathbf{x}} \mathbf{u}_{\mathbf{x}}+\frac{\partial}{\partial \mathbf{y}} \mathbf{u}_{Y}+\frac{\partial}{\partial \mathbf{z}} \mathbf{u}_{\mathrm{Z}}\right) \mathrm{V}
$$

From the definition of the scalar product for the unit vectors $\left(\mathbf{u}_{\mathrm{x}} \cdot \mathbf{u}_{\mathrm{x}}=1\right.$; $\mathbf{u}_{\mathbf{X}} \bullet \mathbf{u}_{\mathbf{Y}}=0$; etc.), the only terms that survive are the three terms given above.
2) cylindrical ( $r, \phi, z$ )

$$
\begin{equation*}
\nabla^{2} V=\frac{1}{r} \frac{\partial}{\partial r}\left(r \frac{\partial V}{\partial r}\right)+\frac{1}{r^{2}} \frac{\partial^{2} V}{\partial \phi^{2}}+\frac{\partial^{2} V}{\partial z^{2}} \tag{3.8}
\end{equation*}
$$

3) spherical $(\rho, \theta, \phi)$

$$
\begin{equation*}
\nabla^{2} V=\frac{1}{\rho^{2}} \frac{\partial}{\partial \rho}\left(\rho^{2} \frac{\partial V}{\partial \rho}\right)+\frac{1}{\rho^{2} \sin \theta} \frac{\partial}{\partial \theta}\left(\sin \theta \frac{\partial V}{\partial \theta}\right)+\frac{1}{\rho^{2} \sin ^{2} \theta} \frac{\partial^{2} V}{\partial \phi^{2}} \tag{3.9}
\end{equation*}
$$

## Poisson's and Laplace's equations

The choice of which form of this operation to actually employ in a calculation is usually dictated by any possible symmetry considerations inherent in the problem. For example, the calculation of the potential within a spherical ball would suggest the application of $\nabla^{2} \mathrm{~V}$ in spherical coordinates rather than in the other representations. Definitions for the Laplacian operator exist for other coordinate systems than mentioned above. A definition in a general orthogonal coordinate system can also be written. For complicated shapes and/or for very difficult problems, a numerical solution may have to be attempted. This is typically the procedure that has to be followed in practice. We will encounter these procedures in the next sections.

Similar Posson's and Laplace's equation arise also for the magnetic field, where a magnetic vector potential $\mathbf{A}$ rather than an electric scalar potential V is involved. The nonexistence of magnetic monopoles allowed us to write that

$$
\begin{equation*}
\nabla \bullet \mathbf{B}=0 \tag{3.10}
\end{equation*}
$$

This equation implies that the magnetic flux density $\mathbf{B}$ can be expressed as the curl of another vector. This vector is given the symbol $\mathbf{A}$ and it is called the magnetic vector potential or just vector potential. This is a consequence of the vector identity

$$
\nabla \bullet \nabla \times \mathbf{A}=0
$$

In electric fields, it was found that the electric field could be easily computed if the scalar potential were known. This was accomplished by taking the gradient of the scalar potential (along with a minus sign). Magnetic fields are related to the vector current density $\mathbf{J}$ via a curl operation and this is significantly more complicated than the divergence operation through which the electric field is related to the charge density $\rho_{\mathrm{v}}$. We suspect, therefore, that the potential that we seek should also be more complicated. This vector potential A is, however, in

## Poisson's and Laplace's equations

the same direction as the source current J. There are, however, some applications of a magnetic scalar potential that will not be described here.

The magnetic flux density $\mathbf{B}$ is determined from the vector potential $\mathbf{A}$ via the relation

$$
\begin{equation*}
\mathbf{B}=\nabla \times \mathbf{A} \tag{3.11}
\end{equation*}
$$

where the units of $\mathbf{A}$ are Tesla.meter. We combine this with the Ampere's law in differential form

$$
\begin{equation*}
\nabla \times \mathbf{B}=\mu_{0} \mathbf{J} \tag{3.12}
\end{equation*}
$$

and obtain

$$
\begin{equation*}
\nabla \times \nabla \times \mathbf{A}=\mu_{0} \mathbf{J} \tag{3.13}
\end{equation*}
$$

The operation on the left hand side of (3.13) can be reduced via a vector identity to yield

$$
\begin{equation*}
\nabla(\nabla \bullet \mathbf{A})-\nabla^{2} \mathbf{A}=\mu_{0} \mathbf{J} \tag{3.14}
\end{equation*}
$$

Two comments should be made concerning vectors at this point before proceeding. First, a vector is completely determined if its curl and its divergence are specified. We have specified $\nabla \times \mathbf{A}$ in (3.11). Since the divergence of the vector A can be specified to have any value, we will choose the value that will ease our calculation:

$$
\begin{equation*}
\nabla \cdot \mathbf{A}=0 \tag{3.15}
\end{equation*}
$$

This is the so called Coulomb gauge. Second, the operation $\nabla^{2} \mathbf{A}$ states that the Laplacian operator operates on all three vector components of the vector potential A separately and each component is set equal to the corresponding component of the current density J. Therefore, with (3.15), (3.14) simplifies to

$$
\begin{equation*}
\nabla^{2} \mathbf{A}=-\mu_{0} \mathbf{J} \tag{3.16}
\end{equation*}
$$

. This is similar to Poisson's equation (3.5) that was derived previously where the electric field was related to a charge density. In the case of magnetic fields,

## Poisson's and Laplace's equations

(3.16) represents three scalar equations for each of the components. In Cartesian coordinates, these equations are

$$
\nabla^{2} A_{X}=-\mu_{0} J_{X} ; \nabla^{2} A_{Y}=-\mu_{0} J_{Y} ; \nabla^{2} A_{Z}=-\mu_{0} J_{Z}
$$

We may at this stage wonder what has been gained by this little bit of vector manipulation since now we have a set of second order uncoupled differential equations to solve instead of a coupled set of first order differential equations. What has been gained is the good fortune of having the vector potential and the current density in the same direction. This implies that each component of the vector potential can be solved separately and we have to solve at the most only three scalar equations.

Example 3-1 MATLAB. The two-dimensional potential distribution can be approximated with the quadratic expression

$$
\mathrm{V}=-\frac{\rho_{\mathrm{v}}}{4 \varepsilon_{0}}\left(\mathrm{x}^{2}+\mathrm{y}^{2}\right)
$$

Show that this function satisfies the Poisson's equation (3.5). Plot the graphs of the charge and the electric potential distributions.

## Poisson's and Laplace's equations



Answer: Using the Laplacian operator in Cartesian coordinates, we find that: $\frac{\partial^{2} V}{\partial \mathrm{x}^{2}}=\frac{\partial^{2} \mathrm{~V}}{\partial \mathrm{y}^{2}}=-\frac{\rho_{\mathrm{v}}}{2 \varepsilon_{0}}$, which leads to Poisson's equation (3.5). The twodimensional plot of the voltage distribution along with calculated charge distribution is shown.

Having derived Poisson's and Laplace's equations for three dimensional systems and having stated the definitions for $\nabla^{2}$ in the three most widely used coordinate systems, we will now obtain analytical solutions for these equations. Rather than first attempting a general three-dimensional solution, we will simplify the discussion by assuming that the potential depends on only one coordinate

## Poisson's and Laplace's equations

rather a than on all three coordinates. The procedure that we will describe in these simpler problems will be later employed in more difficult calculations. Several important results will, however, be obtained as we pass through this fairly difficult initial stage. Techniques that are germane to these more complicated problems will appear in the next sections.

## Analytical solution in one dimension

### 3.2. Analytical solution in one dimension - direct integration method

In order to conceptualize the method, we will calculate the potential variation between two infinite parallel metal plates located in a vacuum as shown in Figure 3-1a. This will require solving Laplace's equation in one dimension and it will yield a result that approximates the potential distribution in a parallel plate capacitor where the separation between the plates is much less than any transverse dimension. Since the plates are assumed to be infinite and the conductivity of these metal plates is very high, they can be assumed to be equipotential surfaces. We can think just that these metal plates have zero resistance. Hence, in the $y$ and the $z$ coordinates, we can postulate with a high degree of confidence that

$$
\frac{\partial V}{\partial y}=\frac{\partial V}{\partial z}=0
$$

Since there is no variation of the potential in two of the three independent variables, we can let the remaining partial derivative that appears in Laplace's equation become an ordinary differential one. Hence, the one-dimensional Laplace's equation is

$$
\begin{equation*}
\frac{d^{2} V}{d x^{2}}=0 \tag{3.17}
\end{equation*}
$$

Let us assume that the plate at $x=0$ to be at a potential $V=V_{0}$ and that the plate at $x=x_{0}$ is connected to ground and therefore has the potential $V=0$. These are the boundary conditions for this problem.

The solution of (3.17) is found by integrating this equation twice

$$
\begin{align*}
& \frac{d V}{d x}=C_{1}  \tag{3.18}\\
& V=C_{1} x+C_{2}
\end{align*}
$$

# Analytical solution in one dimension 



Figure 3-1. (a) Two infinite parallel plates located at $x=0$ and at $x=x_{0}$.
(b) Potential variation between the plates as determined from (3.19).
where $C_{1}$ and $C_{2}$ are the constants of integration. These constants must be included in the solution at each step in the integration and they will be determined from the boundary conditions that are imposed in the problem. Equation (3.18) is the most general solution of the ordinary differential equation (3.17) since it contains the two arbitrary constants of integration.

For the boundary conditions imposed by the battery $\left(V=V_{0}\right.$ at $\left.x=0\right)$ and the ground potential $\left(V=0\right.$ at $\left.x=x_{0}\right)$ in Figure 3-1a, we write

$$
\begin{aligned}
& V_{0}=C_{1}(0)+C_{2} \\
& 0=C_{1}\left(x_{0}\right)+C_{2}
\end{aligned}
$$

Solving these two algebraic equations for the unknown constants of integration leads to the values $C_{1}=-V_{0} / x_{0}$ and $C_{2}=V_{0}$. Therefore, we write the solution for the potential variation between the two parallel plates that satisfies the specified boundary conditions as

$$
\begin{equation*}
\mathrm{V}=\mathrm{V}_{0}\left(1-\frac{\mathrm{x}}{\mathrm{x}_{0}}\right) \tag{3.19}
\end{equation*}
$$

## Analytical solution in one dimension

The potential profile is shown in Figure $3-1 \mathrm{~b}$. Since this is an electrostatic potential, we can compute the electric field using (3.4). We find that

$$
\begin{equation*}
\mathbf{E}=-\frac{\mathrm{dV}}{\mathrm{dx}} \mathbf{u}_{\mathbf{x}}=\frac{\mathrm{V}_{0}}{\mathrm{x}_{0}} \mathbf{u}_{\mathrm{x}} \tag{3.20}
\end{equation*}
$$

a result that we have also obtained using Gauss's law.
There are cases where we cannot specify the potential $V$ at a boundary but can only specify an electric field that is normal to the boundary. Since the electric field is given by (3.4), this in one dimension gives a value for $E=-d V / d x$ at that boundary. These two different boundary conditions would be analogous to a metal plate being clamped at one edge or passing between two rollers that specify the slope of the plate at that edge. The vertical height of the rollers could, however, be arbitrarily adjusted in a rolling mill.

Example 3-2. The electric field normal to a surface at $x=0$ in Figure 3-1a is arbitrarily specified to be $\mathrm{E}_{0}=-\mathrm{dV} / \mathrm{dx}$. Find the potential variation between the plates if the potential at $x=x_{0}$ is still given by $V=0$.

Answer: From (3.18), the constant $C_{1}=-E_{0}$ and $C_{2}=-C_{1} x_{0}=E_{0} x_{0}$. Therefore, the potential V is

$$
V=E_{0}\left(x_{0}-x\right)
$$

which is similar to the result shown in Figure 3-1b.

In the above calculation and in the example, it was assumed that no additional charge was distributed between the plates. Let us now assume that a charge is uniformly distributed between the plates and has a constant density $\rho_{v}$. Other spatial distributions for this charge density could be encountered in practice, but let us not complicate the problem at this stage. For the problem that has been posed, we must solve a one-dimensional Poisson's equation that is written as

## Analytical solution in one dimension

$$
\begin{equation*}
\frac{\mathrm{d}^{2} \mathrm{~V}}{\mathrm{dx}^{2}}=-\frac{\rho_{\mathrm{v}}}{\varepsilon_{0}} \tag{3.21}
\end{equation*}
$$

where again we can neglect any variation in $y$ or $z$ since the plates have been assumed to be infinite in extent and are equipotential surfaces. This allows us to again employ an ordinary derivative. The two-step integration of (3.21) leads to

$$
\begin{align*}
& \frac{d V}{d x}=-\frac{\rho_{v}}{\varepsilon_{0}} x+C_{3}  \tag{3.22}\\
& V=-\frac{\rho_{v}}{\varepsilon_{0}} \frac{x^{2}}{2}+C_{3} x+C_{4}
\end{align*}
$$

where $C_{3}$ and $C_{4}$ are the constants of integration. The constants of integration are found by specifying the boundary conditions and then solving the simultaneous algebraic equations

$$
\begin{aligned}
& \mathrm{V}_{0}=-\frac{\rho_{\mathrm{v}}}{\varepsilon_{0}} \frac{(0)^{2}}{2}+\mathrm{C}_{3}(0)+\mathrm{C}_{4} \\
& 0=-\frac{\rho_{\mathrm{v}}}{\varepsilon_{0}} \frac{\left(\mathrm{x}_{0}\right)^{2}}{2}+\mathrm{C}_{3}\left(\mathrm{x}_{0}\right)+\mathrm{C}_{4}
\end{aligned}
$$

Solving these two equations, we find the constants to be $C_{3}=-\frac{1}{x_{0}}\left(V_{0}-\frac{\rho_{v}}{\varepsilon_{0}} \frac{x_{0}^{2}}{2}\right)$ and $C_{4}=V_{0}$. The substitution of these constants into the solution (3.22) yields the final potential to be

$$
\begin{equation*}
\mathrm{V}=\mathrm{V}_{0}\left(1+\frac{\rho_{\mathrm{v}}}{\rho_{\mathrm{c}}} \frac{\mathrm{x}}{\mathrm{x}_{0}}\right)\left(1-\frac{\mathrm{x}}{\mathrm{x}_{0}}\right) \tag{3.23}
\end{equation*}
$$

where $\rho_{\mathrm{C}}=2 \varepsilon_{0} \mathrm{~V}_{0} / \mathrm{x}_{0}^{2}$. The electric field $\mathbf{E}$ is computed from (3.20) to be

$$
\begin{equation*}
E=\frac{V_{0}}{x_{0}}\left(1-\frac{\rho_{v}}{\rho_{c}}\left(1-2 \frac{x}{x_{0}}\right)\right) \mathbf{u}_{\mathrm{x}} \tag{3.24}
\end{equation*}
$$

The electric potential and electric field are depicted in Figure 3-2 for three values of the charge density.

## Analytical solution in one dimension



Figure 3-2 ${ }^{\text {MATLAB }}$. The electric potential and the electric field for three different values of the charge density $\rho_{\mathrm{v}}$ as calculated in (3.23) and (3.24) is shown.

Example 3-3. There is a sheath in a plasma that is very similar to the depletion width in a pn junction. A plasma is an ionized gas that contains an equal density $n_{i}$ of positively charged ions that is equal to the density of negatively charged electrons $n_{e}$. A sheath connects the plasma to a foreign object such as metal wall. Since the electron and ion densities are equal, the electric field in the plasma can be assumed to be equal to zero. If the plasma were grounded at some far off place, we can assume that the absolute potential of the plasma is

## Analytical solution in one dimension

also equal to zero. This is a common state in nature as over $99 \%$ of the universe is in the plasma state.

Let us now consider a plasma that contains a metal plate to which a negative potential is suddenly applied at a time $t=0$. At $t=0^{+}$after the switch is closed, electrons close to the plate are "blown" into and "lost" in the background plasma. Due to their heavier mass $M_{i} \geq 1836 m_{e}$, the ions will not move at this early time. There is another narrow steady-state sheath that exists in such a plasma that is called the Debye's sheath that "shields" the plasma from this metal plate and this narrow region of positive charge density from the plasma. This allows us to state the boundary conditions at the edge of the electron depleted region at $x=x_{0}$ to be $V=0$ and $E=0$. The dimension $x_{0}$ is, however, unknown and must be computed. Plot the resulting electric potential and electric field distribution.


Answer: The system is modeled with a one-dimensional Poisson's equation in the region $0 \leq x \leq x_{0}$. Since the electrons have departed, only a bare cloud of ions with a uniform charge density $\rho_{v}=n_{i} q$ exists within this region. We write this equation as

# Analytical solution in one dimension 

$$
\frac{\mathrm{d}^{2} \mathrm{~V}}{\mathrm{dx}^{2}}=-\frac{\mathrm{n}_{\mathrm{i}} \mathrm{q}}{\varepsilon_{0}}
$$

This equation can be integrated twice to yield

$$
V=-\frac{n_{i} q}{\varepsilon_{0}} \frac{x^{2}}{2}+A x+B
$$

where $A$ and $B$ are constants of integration. There are three unknown constants: $A, B$, and $x_{0}$. Remember that we do not know where $x_{0}$ is, we only know the boundary conditions at that location. The three boundary conditions are

$$
\begin{aligned}
& V(x=0) \Rightarrow-V_{0}=B \\
& V\left(x=x_{0}\right) \Rightarrow 0=-\frac{n_{i} q}{\varepsilon_{0}} \frac{x_{0}^{2}}{2}+A x_{0}+B \\
& E_{x}\left(x=x_{0}\right) \Rightarrow 0=\frac{n_{i} q}{\varepsilon_{0}} x_{0}-A
\end{aligned}
$$

The simultaneous solution of this set of three algebraic equations gives us the location $\mathrm{x}_{0}$

$$
\mathrm{x}_{0}=\sqrt{\frac{2 \varepsilon_{0} \mathrm{~V}_{0}}{\mathrm{n}_{\mathrm{i}} \mathrm{q}}}
$$

This dimension $\mathrm{x}_{0}$ was originally called the "transient sheath" in plasma physics. As time increases and the ions start to move, a full set of nonlinear fluid equations has to be solved numerically in order to describe the physical behavior of the ion motion. This set consists of the equations of continuity and motion for the ions, an assumption for the distribution of the electrons and Poisson's equation to account for the charge non neutrality.

In some plasma processing applications where plasma ions are to be implanted into a metal surface, this dimension is called the "ion matrix sheath." These applications are found in integrated circuit manufacturing. This sheath evolution phenomena is an active area for research. Because of the practical importance resulting from this calculation, the evolution of the implanting ions and the expanding sheath are more fully described in Appendix 5.

## Analytical solution in one dimension

Example 3-4 MATLAB. Find and plot the potential distribution between two long concentric cylinders. The length on the cylinders is $L$. The boundary conditions are: $V(r=a)=V_{0}$ and $V(r \geq b)=0$.


Answer: Because of the symmetry, we should employ Laplace's equation in cylindrical coordinates (3.8). We can assume that the cylinders are of infinite length and let $\partial \mathrm{V} / \partial \mathrm{z}=0$. In addition, there is no variation in the $\phi$ direction, hence $\partial \mathrm{V} / \partial \phi=0$.

We have to solve the one-dimensional Laplace's equation that becomes an ordinary differential equation in the independent variable $r$

$$
\frac{1}{r} \frac{d}{d r}\left(r \frac{d V}{d r}\right)=0
$$

After multiplying both sides of this equation by $r$, we find the first integral to be

$$
\frac{d V}{d r}=\frac{C_{1}}{r}
$$

which is the radial component of the electric field with a minus sign. The second integral is given by

$$
V=C_{1} \ln r+C_{2}
$$

Applying the boundary conditions, we write

## Analytical solution in one dimension

$$
\begin{aligned}
& V(r=a) \equiv V_{0}=C_{1} \ln a+C_{2} \\
& V(r=b) \equiv 0=C_{1} \ln b+C_{2}
\end{aligned}
$$

The constants of integration $\mathrm{C}_{1}$ and $\mathrm{C}_{2}$ are found from the simultaneous solution of these 2 equations

$$
C_{1}=\frac{V_{0}}{\ln (b / a)} \text { and } C_{2}=-C_{1} \operatorname{lnb}
$$

The potential variation between the two cylinders is finally written as

$$
V=V_{0}\left(\frac{\ln (r / b)}{\ln (a / b)}\right)
$$

where $\mathrm{a}<\mathrm{r}<\mathrm{b}$. The variation of the potential for the radius $\mathrm{b}=4 \mathrm{a}$ is shown below. Note that the potential within the inner cylinder is a constant.


## Analytical solution in one dimension

Example 3-5 MATLAB. A problem that is found in the study of a plasma is to compute the effect of introducing an additional charge into a previously neutral plasma. The electron density $n_{e}$ depends on the local potential $V$ and can be described with a Maxwell-Boltzmann distribution

$$
n_{e}(r)=n_{0} \exp \left(-\frac{q V(r)}{k_{B} T_{e}}\right)
$$

where $\mathrm{k}_{\mathrm{B}} \mathrm{T}_{\mathrm{e}}$ defines the random thermal energy of the electrons (Boltzmann's constant $\mathrm{k}_{\mathrm{B}}=1.38 \times 10^{-23}$ joules $/ 0 \mathrm{~K}$ and $\mathrm{T}_{\mathrm{e}}$ is the electron temperature in ${ }^{\circ} \mathrm{K}$ ). The ion density $n_{i}=n_{0}$. Find the potential distribution caused by the introduction of one additional positive charged particle into a previously neutral plasma.

Answer: Due to the spherical symmetry of the system, we can neglect any variation of potential in two of the coordinates $(\theta, \phi)$ and write Poisson's equation in spherical coordinates with the assistance of (3.9), using only the dependence on the radial coordinate $\rho$, where the charge distribution is given by $\rho_{\mathrm{v}}(\rho)=\left(\mathrm{n}_{\mathrm{e}}(\rho)-\mathrm{n}_{0}\right) \mathrm{q}:$

$$
\frac{1}{\rho^{2}} \frac{d}{d \rho}\left(\rho^{2} \frac{d V}{d \rho}\right)=-\frac{\rho_{v}}{\varepsilon_{0}} \equiv-\frac{n_{0} q\left(\exp \left(-\frac{q V(\rho)}{k_{B} T_{e}}\right)-1\right)}{\varepsilon_{0}}
$$

This differential equation is non-linear one, which has to be solved numerically in general - this is considered in the next section. The charge density has to be linearized in order to be solved analytically

$$
-\frac{\rho}{\varepsilon_{0}}=-\frac{\mathrm{n}_{0} q}{\varepsilon_{0}}\left\{\left[1-\frac{q V}{\mathrm{k}_{\mathrm{B}} \mathrm{~T}_{\mathrm{e}}}+\ldots\right]-1\right\}=\frac{\mathrm{n}_{0} \mathrm{q}^{2}}{\varepsilon_{0} \mathrm{k}_{\mathrm{B}} \mathrm{~T}_{\mathrm{e}}} V
$$

or

$$
\frac{1}{\rho^{2}} \frac{d}{d \rho}\left(\rho^{2} \frac{d V}{d \rho}\right)=\frac{V}{\lambda_{D}^{2}}
$$

where $\lambda_{\mathrm{D}}=\sqrt{\varepsilon_{0} \mathrm{~K}_{\mathrm{B}} \mathrm{T}_{\mathrm{e}} /\left(\mathrm{n}_{0} \mathrm{q}^{2}\right)}$ is the Debye length. We have used a small potential expansion for the exponential term: $\exp ( \pm x) \approx 1 \pm x$. The solution of this equation

## Analytical solution in one dimension

is facilitated if we define a new variable $W=\rho V$. Substitute this variable into this equation to obtain

$$
\frac{d}{d \rho}\left(\rho \frac{d W}{d \rho}-W\right)=\frac{\rho W}{\lambda_{D}^{2}}
$$

or

$$
\frac{d^{2} W}{d \rho^{2}}=\frac{W}{\lambda_{D}^{2}}
$$

The solution of this equation which is finite when $\rho \rightarrow \infty$ is

$$
W=\exp \left(-\frac{\rho}{\lambda_{\mathrm{D}}}\right) \text { or } \mathrm{V}=\frac{1}{\rho} \exp \left(-\frac{\rho}{\lambda_{\mathrm{D}}}\right)
$$

This states that the effect of the additional charge will be "screened" away in a few Debye lengths.


We can think of the Debye length as being analogous to a "time constant" in an electric circuit that has a potential that decays to zero in a few time constants.

# Analytical solution in one dimension 



These examples illustrate the methodical procedure that should be followed in order to analytically solve either Poisson's or Laplace's equations in one dimension.
(1) Choose the most appropriate representation for the Laplacian $\nabla^{2} V$ based on any symmetry that may be found in the problem. Certain derivatives may also be equal to zero due to the symmetry.
(2) Perform the integrations of the differential equation in order to obtain the most general solution for the potential being very careful to include all of the arbitrary constants of integration.

## Analytical solution in one dimension

(3) Let this general representation for the potential satisfy the boundary conditions of the problem. This will specify the values of the arbitrary constants of integration found in step 2.

The reader should be made aware that almost all of the electrostatics problems that can be solved analytically have already been solved and the method of solution and the answers usually appear as examples or problems in some textbook. Further calculations of electrostatic potential problems could be performed for two or three dimensional configurations. These will typically involve mathematical techniques such as separation of variables that will be encountered next.

## Numerical solution of a one-dimensional equation

### 3.3. Numerical solution of a one-dimensional equation - finite difference method

The technique that will be examined for the numerical solution of Laplace's and Poisson's equation in one dimension is the finite difference method (FDM). This will be developed using MATLAB although the techniques are not restricted to MATLAB. The first order derivatives are illustrated in Figure 3-3. There are three possible choices for the derivative, namely

$$
\begin{align*}
\left.\frac{d V}{d x}\right|_{x 0} & =\frac{V_{1}-V_{0}}{h} \text { (Forward difference method) }  \tag{3.25}\\
\left.\frac{d V}{d x}\right|_{x 0} & =\frac{V_{0}-V_{-1}}{h} \text { (Backward difference method) }  \tag{3.26}\\
\left.\frac{d V}{d x}\right|_{x 0} & =\frac{V_{1}-V_{-1}}{2 h} \text { (Central difference method) } \tag{3.27}
\end{align*}
$$



Figure 3-3. Voltage as a function of position. The finite difference equations will be derived with reference to this figure.

We can interpret the Central difference method as being an average of the other two methods. In order to ascertain which method is better, we estimate the errors that might be expected to be found in each method. The errors can be

## Numerical solution of a one-dimensional equation

estimated by expanding the voltage in a Taylor series expansion about the point $x_{0}$ :

$$
\begin{equation*}
\mathrm{V}\left(\mathrm{x}_{0}+\mathrm{h}\right)=\mathrm{V}\left(\mathrm{x}_{0}\right)+\left.\frac{\mathrm{h}}{1!} \frac{\mathrm{dV}}{\mathrm{dx}}\right|_{\mathrm{x} 0}+\left.\frac{\mathrm{h}^{2}}{2!} \frac{\mathrm{d}^{2} \mathrm{~V}}{\mathrm{dx}}\right|_{\mathrm{x} 0}+\left.\frac{\mathrm{h}^{3}}{3!} \frac{\mathrm{d}^{3} \mathrm{~V}}{\mathrm{dx}}\right|_{\mathrm{x} 0}+\text { higher order terms } \tag{3.28}
\end{equation*}
$$

If we neglect the third derivative and higher order terms, we write

$$
\begin{equation*}
\left.\frac{\mathrm{dV}}{\mathrm{dx}}\right|_{\mathrm{x}_{0}}=\frac{\mathrm{V}\left(\mathrm{x}_{0}+\mathrm{h}\right)-\mathrm{V}\left(\mathrm{x}_{0}\right)}{\mathrm{h}}-\left.\frac{\mathrm{h}}{2} \frac{\mathrm{~d}^{2} \mathrm{~V}}{\mathrm{dx}}\right|_{\mathrm{x}_{0}} \tag{3.29}
\end{equation*}
$$

A comparison of (3.29) with (3.25) shows that this is equivalent to the Forward difference method with the additional term

$$
\begin{equation*}
\left.\frac{\mathrm{h}}{2} \frac{\mathrm{~d}^{2} \mathrm{~V}}{\mathrm{dx}}\right|_{\mathrm{x} 0} \tag{3.30}
\end{equation*}
$$

This term is an error term. There are, of course, additional higher order terms that could be included. These additional terms are multiplied by the parameter h to a higher order power. If the parameter $h$ can be made sufficiently small, (3.29) could be useful.

In a similar manner, we write the Taylor series expansion $\mathrm{V}\left(\mathrm{x}_{0}-\mathrm{h}\right)$ about the point $\mathrm{x}_{0}$ :
$V\left(x_{0}-h\right)=V\left(x_{0}\right)-\left.\frac{h}{1!} \frac{d V}{d x}\right|_{x 0}+\left.\frac{h^{2}}{2!} \frac{d^{2} V}{d x^{2}}\right|_{x 0}-\left.\frac{h^{3}}{3!} \frac{d^{3} V}{d x^{3}}\right|_{x 0}+$ higher order terms
From (3.31), we compute

$$
\begin{equation*}
\left.\frac{\mathrm{dV}}{\mathrm{dx}}\right|_{\mathrm{x} 0}=\frac{\mathrm{V}\left(\mathrm{x}_{0}\right)-\mathrm{V}\left(\mathrm{x}_{0}-\mathrm{h}\right)}{\mathrm{h}}+\left.\frac{\mathrm{h}}{2} \frac{\mathrm{~d}^{2} V}{\mathrm{dx}^{2}}\right|_{\mathrm{x} 0} \tag{3.32}
\end{equation*}
$$

for the Backward difference method. The error term is also given by (3.32) plus higher order terms.

Subtracting (3.31) from (3.28) yields

$$
\begin{equation*}
V\left(x_{0}+h\right)-V\left(x_{0}-h\right)=\left.2 h \frac{d V}{d x}\right|_{x 0}+\left.2 \frac{h^{3}}{6} \frac{d^{3} V}{d x^{3}}\right|_{x 0} \tag{3.33}
\end{equation*}
$$

or

## Numerical solution of a one-dimensional equation

$$
\begin{equation*}
\left.\frac{d V}{d x}\right|_{x 0}=\frac{V\left(x_{0}+h\right)-V\left(x_{0}-h\right)}{2 h}-\left.\frac{h^{2}}{6} \frac{d^{2} V}{d x^{2}}\right|_{x_{0}} \tag{3.34}
\end{equation*}
$$

In (3.34), the error is of the order of $h^{2}$. This is in the Central difference method. The error in using this method will be smaller than either of the other two methods and it will be the one employed throughout the rest of the chapter.

Using the Central difference method, we find the representation for the second derivative to be

$$
\left.\frac{\mathrm{d}^{2} \mathrm{~V}}{\mathrm{dx}^{2}}\right|_{\mathrm{x} 0}=\frac{\left.\frac{\mathrm{dV}}{\mathrm{dx}}\right|_{\mathrm{x} 0+\mathrm{h} / 2}-\left.\frac{\mathrm{dV}}{\mathrm{dx}}\right|_{\mathrm{x} 0-\mathrm{h} / 2}}{\mathrm{~h}}=\frac{\frac{\mathrm{V}_{1}-\mathrm{V}_{0}}{\mathrm{~h}}-\frac{\mathrm{V}_{0}-\mathrm{V}_{-1}}{\mathrm{~h}}}{\mathrm{~h}}
$$

or

$$
\begin{equation*}
\left.\frac{\mathrm{d}^{2} \mathrm{~V}}{\mathrm{dx}^{2}}\right|_{\mathrm{x} 0}=\frac{\mathrm{V}_{1}-2 \mathrm{~V}_{0}+\mathrm{V}_{-1}}{\mathrm{~h}^{2}} \tag{3.35}
\end{equation*}
$$

Example 3-6 ${ }^{\text {MATLAB }}$. Find the potential distribution between two surfaces if $V(x=0)=0$ and $V(x=1)=3$. There is no charge distribution in the space $0 \leq x \leq 1$.
Answer: Let us use only three points for the first iteration: $x_{0}=0, x_{1}=0.5$, $x_{2}=1$. Using the Central difference method with a step size $h=1 / 2$, we write Laplace's equation as

$$
\left.\frac{\mathrm{d}^{2} \mathrm{~V}}{\mathrm{dx}{ }^{2}}\right|_{\mathrm{x} 1}=\frac{\mathrm{V}_{2}-2 \mathrm{~V}_{1}+\mathrm{V}_{0}}{0.5^{2}}=0
$$

The boundary conditions imply $\mathrm{V}_{0}=\mathrm{V}\left(\mathrm{x}_{0}=0\right)=0$ and $\mathrm{V}_{2}=\mathrm{V}\left(\mathrm{x}_{2}=1\right)=3$. Hence

$$
V_{1}=0.5\left(V_{0}+V_{2}\right)=1.5
$$

which demonstrates the principle of the average value for the middle point $\mathrm{x}_{1}$.
The second iteration with the smaller steps size $h=1 / 4$ is applied to five points in the same interval. The same boundary conditions are now $V_{0}=0$ and $V_{4}=3$ leads to three simultaneous equations

$$
V_{1}=0.5\left(\mathrm{~V}_{0}+\mathrm{V}_{2}\right) ; \mathrm{V}_{2}=0.5\left(\mathrm{~V}_{1}+\mathrm{V}_{3}\right) ; \mathrm{V}_{3}=0.5\left(\mathrm{~V}_{2}+\mathrm{V}_{4}\right) .
$$

## Numerical solution of a one-dimensional equation

In this case, the boundary conditions specify $\mathrm{V}_{0}$ and $\mathrm{V}_{4}$ and the voltage $\mathrm{V}_{2}$ was calculated in the previous iteration.

The solution for the three intermediate points and the two end points are

$$
\mathrm{V}_{0}=0 ; \mathrm{V}_{1}=0.75 ; \mathrm{V}_{2}=1.50 ; \mathrm{V}_{3}=2.25 ; \mathrm{V}_{4}=3 .
$$

A comparison of these computed values is in agreement with the analytical solution obtained in Example 3-2. The MATLAB calculation produces the following results.

| $\mathrm{V}=$ | 0 | NaN | 1.5000 | NaN | 3.0000 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathrm{~V}=$ | 0 | 0.7500 | 1.5000 | NaN | 3.0000 |
| $\mathrm{~V}=$ | 0 | 0.7500 | 1.5000 | 2.2500 | 3.0000 |

The voltage distribution is shown below.


## Numerical solution of a one-dimensional equation

Example 3-7 MATLAB. Repeat Example 3-2 with a uniform charge distribution $\rho=-4 \varepsilon_{0}$ in the space $0 \leq x \leq 1$. Find the potential distribution between two surfaces if $\mathrm{V}(\mathrm{x}=0)=0$ and $\mathrm{V}(\mathrm{x}=1)=3$.

Answer: Using the Central difference method, we write Poisson's equation as

$$
\left.\frac{\mathrm{d}^{2} \mathrm{~V}}{\mathrm{dx}}\right|_{\mathrm{X} 1}=\frac{\mathrm{V}_{2}-2 \mathrm{~V}_{1}+\mathrm{V}_{0}}{0.5^{2}}=4
$$

for the first iteration. The boundary conditions imply $\mathrm{V}_{0}=\mathrm{V}(\mathrm{x}=0)=0$ and $\mathrm{V}_{2}=$ $V(x=1)=3$. Hence

$$
\frac{V_{2}-2 V_{1}+V_{0}}{0.5^{2}}=4\left(3-2 V_{1}+0\right)=4 \Rightarrow V_{1}=1
$$

The second iteration with the boundary conditions $\mathrm{V}_{0}=\mathrm{V}(\mathrm{x}=0)=0$ and $\mathrm{V}_{4}=\mathrm{V}(\mathrm{x}$ =1) $=3$ leads to

$$
4\left(\mathrm{~V}_{2}-2 \mathrm{~V}_{1}+\mathrm{V}_{0}\right)=4 ; 4\left(\mathrm{~V}_{3}-2 \mathrm{~V}_{2}+\mathrm{V}_{1}\right)=4 ; 4\left(\mathrm{~V}_{4}-2 \mathrm{~V}_{3}+\mathrm{V}_{2}\right)=4
$$

The solution for the three intermediate points and the two end points are

$$
V_{0}=0 ; V_{1}=0.375 ; V_{2}=1 ; V_{3}=1.875 ; V_{4}=3
$$

The term $V_{2}$ equals 1 from the previous iteration.
Applying the same analytical technique as that used in Example 3-3, we find for the exact solution the following explicit expression: $V(x)=2 x^{2}+x$. Once again, the approximate solution found numerically agrees with the analytical solution. The output of the MATLAB program is:

| $\mathrm{V}=$ | 0 | NaN | 1 | NaN | 3 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathrm{~V}=$ | 0 | 0.3750 | 1.0000 | NaN | 3.0000 |
| $\mathrm{~V}=$ | 0 | 0.3750 | 1.0000 | 1.8750 | 3.0000 |

The plot of the results is shown below.

## Numerical solution of a one-dimensional equation



There is a critical restriction on the mesh size is that the first point must be at the center and its numerical value is determined by the value at the two boundaries. This will restrict the number of internal points N to have only certain values that are prescribed by the following prescription.

$$
1 ; 3 ; 7 ; 15 ; 31 ; 63 ; \ldots\left[2^{N}-1\right]
$$

Let us call this the array size.
We may have noted that the length "h" that appears in our application of (3.35) has changed from $1 / 2$ to $1 / 4$. In the next step, it will be reduced to $1 / 8$ and then $1 / 16$ and so on. We can also evaluate (3.35) and keep $h$ as a prescribed value but as we will see, the calculation will have to be repeated several times. The numbers will converge, hopefully in a reasonable time, to the correct answer.

## Numerical solution of a one-dimensional equation

In order to introduce the procedure, we will redo the calculation of the onedimensional Laplace's equation that we have just performed but now make the a priori assumption that $h=1 / 4$. We set the three values internal to the fixed boundaries as initially being equal to zero. Hence, we write

$$
V_{1}(0)=0, V_{2}(0)=0, V_{3}(0)=0
$$

The boundary values will remain fixed at all iterations, namely $\mathrm{V}_{0}=0$ and $V_{4}=3$.

In our first iteration denoted with the "(1)" using (3.35), we write

$$
\begin{aligned}
& \frac{V_{0}-2 V_{1}(1)+V_{2}(0)}{0.25^{2}}=0 ; \\
& \frac{V_{1}(1)-2 V_{2}(1)+V_{3}(0)}{0.25^{2}}=0 ; \\
& \frac{V_{2}(1)-2 V_{3}(1)+V_{4}}{0.25^{2}}=0
\end{aligned}
$$

In the second equation, we include the value for $\mathrm{V}_{1}(1)$ that has just been obtained from the previous equation since it is now known. A similar argument holds for $V_{3}(1)$ in the third equation. In fact, this will be a general pattern. The simultaneous solution of this set of equations leads to

$$
V_{1}(1)=0, V_{2}(1)=0, V_{3}(1)=1.5
$$

In order to compute the values at the second iteration, we use the values from the first iteration and sequentially write:

$$
\begin{gathered}
\frac{V_{0}-2 V_{1}(2)+V_{2}(1)}{0.25^{2}}=0 ; \\
\frac{V_{1}(2)-2 V_{2}(2)+V_{3}(1)}{0.25^{2}}=0 ; \\
\frac{V_{2}(2)-2 V_{3}(2)+V_{4}}{0.25^{2}}=0
\end{gathered}
$$

From this set, we compute

$$
V_{1}(2)=0, V_{2}(2)=0.75, V_{3}(2)=1.875
$$

The third iteration is

## Numerical solution of a one-dimensional equation

$$
\begin{gathered}
\frac{V_{0}-2 V_{1}(3)+V_{2}(2)}{0.25^{2}}=0 ; \\
\frac{V_{1}(3)-2 V_{2}(3)+V_{3}(2)}{0.25^{2}}=0 ; \\
\frac{V_{2}(3)-2 V_{3}(3)+V_{4}}{0.25^{2}}=0
\end{gathered}
$$

We obtain

$$
V_{1}(3)=0.375, V_{2}(3)=1.125, V_{3}(3)=2.0625
$$

We could keep going using our calculator, but let us stop here. The numbers seem to be approaching an asymptotic limit. We will write a MATLAB program to do this work. A more interesting question arises at this point. Is the answer correct? We can easily check this by dividing the parameter " $h$ " by two and redoing the calculation again and again. If the numbers are the same or seem to asymptotically approach the same value, we are finished.

Example 3-8 ${ }^{\text {MATLAB }}$. Write a MATLAB program to evaluate and plot the first six iterations of the solution of the one dimensional Laplace's equation by FDMmethod. The boundary conditions are: $\mathrm{V}(1)=0$ and $\mathrm{V}(5)=3$.

Answer: The iterations are indicated with the integer $k$. The analytical solution $\mathrm{V}(\mathrm{x})=0.75$. $(\mathrm{x}-1)$ is shown with a dashed line for a comparison in the figures below.

## Numerical solution of a one-dimensional equation



Example 3-9 MATLAB. Using MATLAB, plot the voltage $V(x)=\exp \left(-x^{2}\right)$ and the electric field $\mathrm{E}_{\mathrm{x}}(\mathrm{x})=-\mathrm{dV} / \mathrm{dx}$ as a function of x in the range $-3 \leq x \leq 3$.

Answer:The difference operation "diff(y)./diff( x$)$ " sequentially performs and stores the values $\{[\mathrm{V}(2)-\mathrm{V}(1)] / \mathrm{h},[\mathrm{V}(3)-\mathrm{V}(2)] / \mathrm{h}, \ldots,[\mathrm{V}(\mathrm{n})-\mathrm{V}(\mathrm{n}-1)] / \mathrm{h}\}$. There are only ( $n-1$ ) values of the derivatives. Therefore, if we wish to plot the results, we plot $y$ in the range $x_{a}$ to $x_{b}$ at increments of $h$ and dy/dx in the range $\left[x_{a}+h / 2\right]$ to $\left[x_{b}-h / 2\right]$ with the same increment $h$. The plots of the potential and the electric fields are shown in the figure below.

## Numerical solution of a one-dimensional equation



Example 3-10 MATLAB . Consider again Example 3-5, which has been solved analytically after a linearization. Write a MATLAB program to calculate and plot the potential distribution using the function 'ode45' for obtaining numerical solutions of differential equations.

Answer: We first transform Poisson's equation into a standard form with the substitution $\mathrm{dV} / \mathrm{dr}=\mathrm{U}$. This results in two coupled first order differential equations.

$$
\begin{aligned}
& \frac{d V}{d r}=U, \\
& \frac{d U}{d r}=-\frac{2}{r} U+D(\exp (-C . V)-1)
\end{aligned}
$$

## Numerical solution of a one-dimensional equation

where $C=q / k_{B} T_{e}$ and $D=n_{0} q / \varepsilon_{0}$. Choosing numerical values of $C=0.25$ and $D=4$, we normalize the Debye length $\lambda_{D}=1 / \sqrt{C \cdot D}=1$. Assuming that the value of the potential $V$ at $\rho=0$ has a value of 10 and that the electric field $U=-\frac{d V}{d r}$ at the same location has a value of 1 , we are able to numerically evaluate the potential distribution in space. The numerical results are shown below.


## Analytical solution of a two-dimensional equation

### 3.4. Analytical solution of a two-dimensional equation - Fourier series expansion

In this section, we will introduce the methodical procedure to effect the solution of the type of problem that is governed by Poisson's or Laplace's equation in higher dimensions. The technique that will be employed to obtain this solution is the powerful "method of separation of variables."

It is pedagogically convenient to introduce the technique with an example and then carefully work through the details. The problem that we will initially examine is to calculate the potential distribution within a charge-free unbounded region illustrated in Figure 3-4 where the potential is prescribed on all four edges. In our example, we will specify that the potential on two of the edges is equal to zero, approaches zero on the third edge, that is taken to be at $y \rightarrow \infty$ and the potential has a particular distribution on the fourth edge.


Figure 3-4. An unbounded rectangular region with the potential specified on all of the boundaries.

Since there is no charge within the interior region, we should solve Laplace's equation (3.6). Since the object has rectangular symmetry and there is

## Analytical solution of a two-dimensional equation

no dependence of the potential on the third coordinate $z$, the form of Laplace's equation in Cartesian coordinates (3.7) that we will use is written as

$$
\begin{equation*}
\nabla^{2} V=\frac{\partial^{2} V}{\partial x^{2}}+\frac{\partial^{2} V}{\partial y^{2}}=0 \tag{3.36}
\end{equation*}
$$

In writing (3.36), we have explicitly stated that the potential $V=V(x, y)$ depends on the two independent variables $x$ and $y$.

The philosophy of solving this equation using the method of separation of variables is to assert that the potential $V(x, y)$ is equal to the product of two terms, $X(x)$ and $Y(y)$, that separately are functions of only one of the independent variables $x$ and $y$. The potential is then given by

$$
\begin{equation*}
V(x, y)=X(x) \cdot Y(y) \tag{3.37}
\end{equation*}
$$

This is a critical assertion and our solution depends on it being a correct assumption. We may wonder if other functional forms would work at this stage. They might or they might not. The resulting solutions that would be obtained using different combinations might physically not make any sense or they might not satisfy the boundary conditions. Therefore, we will follow in the footsteps of those pioneering giants who have led us through the dark forest containing problems of this genre in the past and just use (3.37) and not concern ourselves with these questions. "If it works, why fix it?" will be our motto.

Substitute (3.37) into (3.36) and write

$$
\begin{equation*}
Y(y) \frac{d^{2} X(x)}{d x^{2}}+X(x) \frac{d^{2} Y(y)}{d y^{2}}=0 \tag{3.38}
\end{equation*}
$$

Note that the terms that are to be differentiated only involve one independent variable. Hence, the partial derivatives can be replaced with ordinary derivatives and this will be done in the subsequent development.

The next step in this methodical procedure is to divide both sides of this equation by $V(x, y)=X(x) . Y(y)$. Our friends in mathematics may stand up in horror at this suggestion! As we will later see, one of these terms could be zero at

## Analytical solution of a two-dimensional equation

one or more points in space. Recall what a calculator or computer tells us when we do this "evil" deed of dividing by zero! With this warning in hand and with a justified amount of trepidation, let us see what does result from this action. In our case, the end will justify the means. We find that

$$
\begin{equation*}
\frac{1}{X(x)} \frac{d^{2} X(x)}{d x^{2}}+\frac{1}{Y(y)} \frac{d^{2} Y(y)}{d y^{2}}=0 \tag{3.39}
\end{equation*}
$$

The first term on the left side of (3.39) is independent of the variable y. As far as the variable $y$, it can be considered to be a constant that we will take to be $-k_{Y}^{2}$. Using a similar argument, the second term on the left side of (3.39) is independent of the variable $x$ and it also can be replaced with another constant that will be written as $+\mathrm{k}_{\mathrm{x}}^{2}$. Therefore, (3.39) can be written as two ordinary differential equations and one algebraic equation:

$$
\begin{gather*}
\frac{d^{2} X(x)}{d x^{2}}+k_{X}^{2} X(x)=0  \tag{3.40}\\
\frac{d^{2} Y(y)}{d y^{2}}-k_{Y}^{2} Y(y)=0  \tag{3.41}\\
k_{X}^{2}-k_{Y}^{2}=0 \tag{3.42}
\end{gather*}
$$

A pure mathematician would have just written these three equations down by inspection in order to avoid any problems with dividing by zero that we have so cavalierly glossed over.

The two second-order ordinary differential equations can be easily solved. We write that

$$
\begin{gather*}
X(x)=C_{1} \sin \left(k_{x} x\right)+C_{2} \cos \left(k_{x} x\right)  \tag{3.43}\\
Y(y)=C_{3} \exp \left(k_{Y} y\right)+C_{4} \exp \left(-k_{Y} y\right) \tag{3.44}
\end{gather*}
$$

where we include the constants of integration, $\mathrm{C}_{1}$ to $\mathrm{C}_{4}$. Let us now determine these constants of integration from the boundary conditions imposed in Figure 34. From (3.37), we note that the potential $\mathrm{V}(\mathrm{x}, \mathrm{y})$ is determined by multiplying the solution $X(x)$ with $Y(y)$. Therefore, we can specify the constants by examining

## Analytical solution of a two-dimensional equation

each term separately. For any value of $y$ at $x=0$, the potential $V(0, y)$ is equal to zero. The only way that we can satisfy this requirement is to let the constant $\mathrm{C}_{2}=$ 0 since $\cos 0=1$. Nothing can be stated about the constant $C_{1}$ from this particular boundary condition since $\sin 0=0$. For any value of $x$ and in the limit of $y \rightarrow \infty$, the potential $\mathrm{V}(\mathrm{x}, \mathrm{y} \rightarrow \infty)=0$. This specifies that the constant $\mathrm{C}_{3}=0$ since the term $\exp \left(\mathrm{k}_{\mathrm{y}} \mathrm{y}\right) \rightarrow \infty$ as $\mathrm{y} \rightarrow \infty$. The constant $\mathrm{C}_{4}$ remains undetermined from the application of this boundary condition. The potential on the third surface $V(a, y)$ is also specified to be zero at $x=a$ from which we conclude that $k_{x}=n \pi / a$ since $\sin n \pi=0$. From (3.42), we also write that the constants $k_{Y}=k_{x}$. With these values for the constants, our solution $V(x, y)=X(x) . Y(y)$ becomes

$$
\begin{equation*}
V(x, y)=\left[C_{1} C_{4}\right] \exp \left(-\frac{n \pi y}{a}\right) \sin \left(\frac{n \pi x}{a}\right) \tag{3.45}
\end{equation*}
$$

For this example, the integer $n$ will take the value of $n=1$ in order to fit the fourth boundary condition at $y=0$. Finally, the product of the two constants $\left[\mathrm{C}_{1} \mathrm{C}_{4}\right]$ that is just another constant is set equal to $\mathrm{V}_{0}$. The potential in this channel finally is given by

$$
\begin{equation*}
V(x, y)=V_{0} \exp \left(-\frac{\pi y}{a}\right) \sin \left(\frac{\pi x}{a}\right) \tag{3.46}
\end{equation*}
$$

The variation of this potential in space is shown in Figure 3.5 for the values of $a=1, V_{0}=10$.

An examination of Figure 3-5 will yield some important physical insight into the variation of the potential. First, the potential V only approaches zero as the coordinate $\mathrm{y} \rightarrow \infty$. Second, the boundary conditions at $\mathrm{x}=0$ and at $\mathrm{x}=$ a were that the potential $V$ equaled a constant that in this case was equal to zero. Recall from the previous chapter that $E_{Y}=-\partial \mathrm{V} / \partial \mathrm{y}$. This implies that the component of electric field $E_{Y}$ must also be equal to zero along these two surfaces. We can conclude that the tangential component of electric field adjacent to an

## Analytical solution of a two-dimensional equation

equipotential surface will be equal to zero. This conclusion will be of importance in several later calculations.


Figure 3-5 ${ }^{\text {MATLAB }}$. Variation of the potential within the region for the prescribed boundary conditions depicted in Figure 3-4.

The procedure that we have conducted is the determination of the solution of a partial differential equation. Let us recapitulate the procedure before attacking a slightly more difficult problem.
(1) The proper form of the Laplacian operator $\nabla^{2} V$ for the coordinate system of interest was chosen. This choice was predicated on the symmetry and the boundary conditions of the problem.

## Analytical solution of a two-dimensional equation

(2) The potential $V(x, y)$ that depended on two independent variables was separated into two dependent variables that individually depended on only one of the independent variables. This allowed us to write the partial differential equation as a set of ordinary differential equations and an algebraic equation by assuming that the solution could be considered as a product of the individual functions of the individual independent variables.
(3) Each of the ordinary differential equations was solved that led to several constants of integration. The solution of each ordinary differential equation was multiplied together to obtain the general solution of the partial differential equation.
(4) The arbitrary constants of integration that appeared when the ordinary differential equations were solved were determined such that the boundary conditions would be satisfied. The solution for a particular problem has now been obtained. Note that this step is similar to the methodical procedure that we employed in the one dimensional case.

Let us examine the potential distribution in a bounded space as depicted in Figure 3-6. The procedure will be the same as for the unbounded case treated above. In this case, the potential is required to be equal to zero on three of the boundaries and it has a sinusoidal variation on the remaining boundary.

In this case, the solution of Laplace's equation for $X(x)$ and $Y(y)$ is again given by (3.43) and (3.44).

## Analytical solution of a two-dimensional equation



Figure 3-6. A boundary value problem for a bounded surface.

We are able to predict the functional characteristics of the basic "eigenfunction." This is a German word that means "characteristic function." The values for $k_{X}$ and $k_{Y}$ are called "eigenvalues" or characteristic values. In this case again, the eigenvalue $\mathrm{k}_{Y}=\mathrm{k}_{\mathrm{X}}$ as determined from (3.42). We may also find this function referred to as a proper function.

The constants are again determined by the boundary conditions. From (3.43), the constant $\mathrm{C}_{2}=0$ again since the potential $\mathrm{V}=0$ at $\mathrm{x}=0$. The constant $k_{x}=n \pi / a(n=0,1,2, \ldots)$ since the potential $V=0$ at $x=a-$ these are the eigenvalues of the problem. From (3.44), we write

$$
C_{3} \exp \left(k_{Y} b\right)+C_{4} \exp \left(-k_{Y} b\right)=0
$$

since the potential $\mathrm{V}=0$ at $\mathrm{y}=\mathrm{b}$, which yields a relationship between $\mathrm{C}_{3}$ and $\mathrm{C}_{4}$. Therefore, the potential within the enclosed region specified in Figure 3.6 can be written as

$$
V=\left[C_{1} C_{3} \exp \left(\frac{n \pi b}{a}\right)\right]\left[\exp \left(\frac{n \pi(y-b)}{a}\right)-\exp \left(-\frac{n \pi(y-b)}{a}\right)\right] \sin \left(\frac{n \pi x}{a}\right)
$$

or

## Analytical solution of a two-dimensional equation

$$
\begin{equation*}
V(x, y)=\left[2 C_{1} C_{3} \exp \left(\frac{n \pi b}{a}\right)\right] \sinh \left(\frac{n \pi(y-b)}{a}\right) \sin \left(\frac{n \pi x}{a}\right) \tag{3.47}
\end{equation*}
$$

The constants within the square brackets will be determined from the remaining boundary condition at $\mathrm{y}=0$.

The boundary condition at $\mathrm{y}=0$ states that $\mathrm{V}=\mathrm{V}_{0} \sin (\pi \mathrm{x} / \mathrm{a})$ for $0 \leq x \leq a$. Hence, the integer $n=1$ and

$$
\begin{equation*}
C_{1} C_{3}=\frac{V_{0}}{2 \exp \left(\frac{\pi b}{a}\right) \sinh \left(-\frac{\pi b}{a}\right)} \tag{3.48}
\end{equation*}
$$

The potential is finally written as

$$
\begin{equation*}
V=\frac{V_{0} \sinh (\pi(b-y) / a) \sin (\pi x / a)}{\sinh (\pi b / a)} \tag{3.49}
\end{equation*}
$$

This is shown in Figure 3-7. Note that we do satisfy the imposed boundary condition that the potential equals zero on three edges.


## Analytical solution of a two-dimensional equation

Figure 3-7 ${ }^{\text {MATLAB }}$. Normalized potential profile within the region described in Figure 3-6. Note that the potential is equal to zero on three edges.

In the two examples that were treated above, we assumed that the boundary condition at $\mathrm{y}=0$ had a nonuniform distribution. This was an academic type distribution rather than a realistic one but we were able to "carry out the details" to the very end without having to introduce more complicated mathematics. However, we should look at the real world where we might expect that a more realistic distribution for the potential at $\mathrm{y}=0$ in Figure 3-8 would be to assume that the potential at $\mathrm{y}=0$ would be a constant, say $\mathrm{V}=\mathrm{V}_{\mathrm{o}}$. The boundary conditions on the other three edges could remain the same in realistic situations. Let us carry through the details for this particular boundary condition.

Since the other boundary conditions have not been altered, the general solution of Laplace's equation can be written as a superposition of particular solutions given by (3.49)


Figure 3-8. Periodic potential represents the constant potential $V=V_{0}$ within the region $0 \leq x \leq a$.

$$
\begin{equation*}
V(x, y)=\sum_{n=1}^{\infty} d_{n} \frac{\sinh (n \pi(b-y) / a) \sin (n \pi x / a)}{\sinh (n \pi b / a)} \tag{3.50}
\end{equation*}
$$

## Analytical solution of a two-dimensional equation

because of the linearity of Laplace's equation, where $n$ is an integer ( $n=1,2,3, \ldots$ ). In writing this expression as a summation of an infinite series of sinusoidal functions, we are being guided by the fact that each term does satisfy the boundary condition that the potential $\mathrm{V}=0$ at $\mathrm{x}=0$ and at $\mathrm{x}=\mathrm{a}$, hence the infinite sum will also satisfy the boundary conditions. The coefficients $c_{n}$ will be chosen to yield the best fit of the remaining boundary condition at $\mathrm{y}=0$ that has now been specified to be a constant potential $V=V_{0}$.

We may recognize (3.50) as the Fourier sine series and the constants $c_{n}$ as the Fourier coefficients. The coefficients $c_{n}$ and $d_{n}$ are defined for a general periodic function $F(x) \equiv V(x, 0)$ with a period $L$ :

$$
\begin{equation*}
F(x)=\frac{c_{0}}{2}+\sum_{n=1}^{\infty}\left(c_{n} \cos \left(\frac{n 2 \pi x}{L}\right)+d_{n} \sin \left(\frac{n 2 \pi x}{L}\right)\right) \tag{3.51}
\end{equation*}
$$

from the relations

$$
\begin{equation*}
c_{n}=\frac{2}{L} \int_{-L / 2}^{L / 2} F(x) \cos \left(\frac{n \pi x}{L}\right) d x, d_{n}=\frac{2}{L} \int_{-L / 2}^{L / 2} F(x) \sin \left(\frac{n \pi x}{L}\right) d x \tag{3.52}
\end{equation*}
$$

The potential V is known to have a constant value only in the region $0 \leq x \leq a$. Outside of this region, it is not specified and could have any value that we choose in order to ease our mathematical difficulties. In this case, the period of the wave is $L=2 a$. Our choice for the potential at the boundary is to assume that it is an odd function in the variable x . This means that there will only be sine functions in the expansion. Therefore, $\mathrm{c}_{\mathrm{n}}=0$.

The coefficients $d_{n}$ with reference to Figure 3-8 are calculated from

$$
d_{n}=2 \cdot \frac{V_{0}}{a} \int_{0}^{a} \sin \left(\frac{n \pi x}{a}\right) d x=\frac{2 V_{0}}{n \pi} \int_{0}^{a} \sin \left(\frac{n \pi x}{a}\right) d\left(\frac{n \pi x}{a}\right)
$$

This integral leads to

$$
d_{n}= \begin{cases}\frac{4 V_{0}}{n \pi} & n=\text { odd }  \tag{3.53}\\ 0 & n=\text { even }\end{cases}
$$

The potential is given by (3.51) with the coefficients defined in (3.53)

## Analytical solution of a two-dimensional equation

$$
\begin{equation*}
V(x, y)=\frac{4 V_{0}}{\pi} \sum_{n=1,3,5, \ldots}^{\infty} \frac{\sinh (n \pi(b-y) / a) \sin (n \pi x / a)}{n \cdot \sinh (n \pi b / a)} \tag{3.54}
\end{equation*}
$$

Example 3-11 ${ }^{\text {MATLAB }}$. Write a MATLAB program that calculates and plots the two-dimensional potential distribution for the problem, shown in Figure 3-8.

Answer: The potential variation of the first five terms ( $n=1,3,5,7,9$ ) is shown below.


Certain general comments can be made about the potential variation that is shown in the figure of the previous example, especially when it is compared

## Analytical solution of a two-dimensional equation

with the potential profile in Figure 3-8 that is just the first Fourier term of the Fourier series. The fit to a constant value at $\mathrm{y}=0$ is better if more modes are included in the expansion. The fit at $x=0$ and at $x=a$ will not be possible as the function is double-valued there. However, for $y=0, V=V_{o}$. If we had included more terms in the expansion, we would have observed a very rapid oscillation at either edge. This effect is given the name " Gibb's phenomenon" and it is a topic for further consideration in an advanced calculus course.

In this section, we found that it was desirable to sum all of the terms in the Fourier series in order to get a valid representation for the potential profile. As a general rule, we can say that the more terms that are included in the summation, the better the representation for the potential. The question then arises, "Is there something unique about each of the terms in the series?" We can answer this question by watching a gymnast jumping on a trampoline. If the gymnast lands in the middle of the trampoline, the perturbation in the canvas will be different from what it would be if the landing were at a point that is away from the center or if two gymnasts were doing their thing in tandem. There are different modes for the oscillation. The mathematical structure of the solution for an equation describing the motion of the canvas for all possible landing points is a solution that involves finding all of the Fourier modes.

## Finite Difference Method using MATLAB

### 3.5. Finite Difference Method using MATLAB

There are different methods for the numerical solution of the twodimensional Laplace's and Poisson's equation. Some of the techniques are based on a differential formulation that was introduced earlier. The Finite difference method is considered here and the Finite element method is discussed in the next section. Other techniques are based on the integral formulation of the boundary value problems such as the Method of moments is described later. The boundary value problems become more complicated in the presence of dielectric interfaces which are also considered in this section.

The finite difference method (FDM) considered here is an extension of the method already applied to a one-dimensional problem. This method allows MATLAB to be more directly involved in the solution of the boundary value problems. We will discuss this method here using a problem that is similar to that presented in Example 3-11. We will describe the technique to obtain and to solve a suitable set of coupled equations that can be interpreted as a matrix equation.

The algorithm that we use is based on the approximation (3.34) for the second derivatives in Cartesian coordinates. In this case, we assume a square grid with a step size h in both directions for a two-dimensional calculation

$$
\nabla^{2} V(x, y)=\frac{1}{h^{2}}[V(x+h, y)+V(x-h, y)+V(x, y+h)+V(x, y-h)-4 V(x, y)]
$$

This leads to the following "star shape" representation for a two-dimensional Laplace's equation (3.36) as shown in Figure 3-9a

$$
\begin{equation*}
V_{0}=\frac{V_{1}+V_{2}+V_{3}+V_{4}}{4} \tag{3.55}
\end{equation*}
$$

The voltage at the center is approximated as being the average of the voltages at the four tips of the star. For the three-dimensional case, the square is replaced with a cube and a seven-point scheme is applied. In this case, the coefficient $1 / 4$ in (3.55) is simply replaced with $1 / 6$.

## Finite Difference Method using MATLAB



Figure 3-9. (a)The general five-point scheme; (b)The three-point scheme at the corner.

However, in Poisson's equation, it will modify the charge density that is to be evaluated at the central point (0). For the special case of a corner point, this five point scheme has to be modified to a three point one as shown in Figure 3-9b. In this case the principle of the average value simply gives

$$
\begin{equation*}
V_{0}=\frac{V_{1}+V_{2}}{2} \tag{3.56}
\end{equation*}
$$

This principle can be applied iteratively for the computation of the potential in the points of the square grid as shown in Figure 3-10. This method is also called a relaxation method. After computing the first iteration, we determine the potential at the other points within the nine-point mesh. This will involve two more iterations as shown in Figure 3-11. In the second iteration, all of the potentials at the locations indicated by a solid circle in Figure 3-11a are now known. The values indicated by a square $\square$ are to be computed in this iteration using (3.55). In the third iteration, the values of the potential indicated by the solid circles and squares are known from the previous two iterations or as initial values in

## Finite Difference Method using MATLAB

the calculation. Again employing (3.55), the values of the potential at the locations indicated by the diamonds can be computed. In this mesh, it is assumed that the potentials at the boundaries are already given in the statement of the problem, hence the potentials at the locations indicated by the hollow circles $\bigcirc$ are also known as shown in Figure 3-11b.


Figure 3-10. The square grid in two dimensions in Cartesian coordinates.

(a)

(b)

Figure 3-11. The second and third iterations. (a) The values of the potential indicated by the solid circles are known. The values at the locations of the solid squares are computed in the second iteration. (b) The potentials at the boundaries indicated by the hollow circles $\bigcirc$ are assumed to be known. The potentials at the locations indicated by the diamonds are computed in the third iteration.

## Finite Difference Method using MATLAB

This iterative procedure can continue until the computed values at all of the points in the decreasing meshes become closer to each other. The accuracy of the calculation can be insured by repeating the calculation with a different initial mesh size. A mesh with a shape and orientation that is different than the one used here could also be employed in a numerical calculation. This is particularly useful in calculations involving unusual shapes. It is also possible to scale the various dimensions in order to use this particular mesh.

A critical restriction is also found on the square mesh size in that the first point must be in the center of the square. This point will be evaluated from the four boundaries of the square. This will restrict the number of internal points N of the square to contain the following number of points

$$
1^{2} ; 3^{2} ; 7^{2} ; 15^{2} ; 31^{2} ; 63^{2} ; \ldots\left[2^{\mathrm{N}}-1\right]^{2}
$$

This is called the array size.
Example 3-12 ${ }^{\text {MATLAB }}$. Given that the potential at the four sides of the square region have the values: $V(0, y)=V(a, y)=V(x, a)=0, V(x, 0)=V_{0}=10 V$, plot the potential internal to the boundaries. Use an array size of $31 \times 31$. From (3.56), the potentials at the four corners are $(0,0,5,5)$.

Answer: The results of the numerical calculation are shown below. Note not the solution satisfies the boundary conditions and also the values at the corners.

## Finite Difference Method using MATLAB



In the discussion above, we assumed that the region to which the star given in $(3.55)$ and depicted in Figure $3-11$ had the shape of a square. This technique seemed to work well. However, this technique can be extended to an area with a more complex shape. There are techniques that can be employed to enhance the rate of convergence to the final solution. For example, if one of the boundaries did not have a constant value, it might be advantageous to use a different mesh configuration. It is not a large step to get into examples that are "beyond the scope of this text." We will let others tread in those waters.

## Finite Difference Method using MATLAB

Example 3-13 MATLAB. The potential in a certain region is described with the expression $V(x, y)=V_{0} \exp \left(-\left(x^{2}+y^{2}\right)\right)$. Calculate and plot the volume charge density $\rho_{\mathrm{v}}(\mathrm{x}, \mathrm{y})$ that would be calculated from Poisson's equation assuming that $\varepsilon_{0}=1$.

Answer: Here 'del2' function is applied. The results are shown in the figure below.


In the material described so far, we have assumed that the potential was specified at the boundaries of a uniform dielectric region for which the potential was to be numerically determined. If the region contains two dielectrics as shown

## Finite Difference Method using MATLAB

in Figure 3-12, we have to obtain an algorithm that will allow us to evaluate the potential on both sides of the dielectric interface.

In order to calculate the boundary condition for the interface of the two dielectrics, we make use of Gauss's law. This is written as

$$
\begin{equation*}
\oint \varepsilon \mathbf{E} \cdot \mathbf{d} \mathbf{s}=\mathrm{Q}_{\mathrm{enc}}=0 \tag{3.57}
\end{equation*}
$$

where we have assumed that there is no surface charge density at the interface. With reference to Figure 3.12, (3.57) can be written as

$$
\begin{equation*}
\oint \varepsilon \mathbf{E} \cdot \mathbf{d} \mathbf{s}=\Delta \mathbf{z} \oint \varepsilon \mathbf{E} \cdot \mathbf{d} \mathbf{l}=-\Delta \mathbf{z} \oint \varepsilon \frac{\mathrm{dV}}{\mathrm{dn}} \mathrm{dl}=0 \tag{3.58}
\end{equation*}
$$

where we have replaced the electric field with the derivative of the potential that is normal to the surface. The term $\Delta z$ is the distance in the third coordinate. The surface integral has become a contour integral times this distance $\Delta z$ that is directed out of the page. In terms of Figure 3-12, we write

$$
\begin{align*}
\oint \varepsilon \frac{d V}{d n} d l= & \frac{V_{1}-V_{0}}{h}\left(\varepsilon_{2} \frac{h}{2}+\varepsilon_{1} \frac{h}{2}\right)+\frac{V_{2}-V_{0}}{h}\left(\varepsilon_{1} h\right) \\
& \frac{V_{3}-V_{0}}{h}\left(\varepsilon_{2} \frac{h}{2}+\varepsilon_{1} \frac{h}{2}\right)+\frac{V_{4}-V_{0}}{h}\left(\varepsilon_{2} h\right) \tag{3.59}
\end{align*}
$$

Rearranging terms, we rewrite (3.59) as

$$
\begin{equation*}
\left(\varepsilon_{1}+\varepsilon_{2}\right) \mathrm{V}_{1}+2 \varepsilon_{1} \mathrm{~V}_{2}+\left(\varepsilon_{1}+\varepsilon_{2}\right) \mathrm{V}_{3}+2 \varepsilon_{2} \mathrm{~V}_{4}-4\left(\varepsilon_{1}+\varepsilon_{2}\right) \mathrm{V}_{0}=0 \tag{3.60}
\end{equation*}
$$

or

$$
\begin{equation*}
\mathrm{V}_{0}=\frac{1}{4\left(\varepsilon_{1}+\varepsilon_{2}\right)}\left[\left(\varepsilon_{1}+\varepsilon_{2}\right) \mathrm{V}_{1}+2 \varepsilon_{1} \mathrm{~V}_{2}+\left(\varepsilon_{1}+\varepsilon_{2}\right) \mathrm{V}_{3}+2 \varepsilon_{2} \mathrm{~V}_{4}\right] \tag{3.61}
\end{equation*}
$$

This equation is an extension of the equation (3.55), which was written when $\varepsilon_{1}=\varepsilon_{2}=\varepsilon$ (homogeneous medium). Using the algorithm developed in (3.61), we can relate the potentials on one side of a dielectric to the other side.

Finite Difference Method using MATLAB


Figure 3-12. Interface of two dielectrics. Points of the five-point scheme are in the two dielectrics.

Finite Element Method using MATLAB

### 3.6. Finite Element Method using MATLAB

The Finite element method (FEM) is a generalization of the finite difference method (FDM) that was considered in the previous section. Instead of subdividing the area into small squares with a side $h$, this technique subdivides the area into small triangles. As we will observe, this method is more flexible in its application. For example, the calculation of the potential profile in the region between two concentric rectangular metallic surfaces can be handled with this technique. We will develop the technique using this problem. The technique has a strong foundation using matrix manipulations.

In Figure 3-13a, one quarter of the cross section of a rectangular coaxial line is shown. Figure $3-13 b$ depicts the appropriate modeling with the finite triangular elements. The mesh is irregular and the grid is denser in the vicinity of the corners where a more rapid variation of the potential is to be expected. The scalar potential V satisfies Laplace's equation (3.6) inside the area $S$

$$
\begin{equation*}
\nabla^{2} V=0 \tag{3.62}
\end{equation*}
$$

while it satisfies two different conditions on the boundaries. 1) On the boundary $L=L_{1}+L_{2}$, the voltage is specified to have any particular value. In Figure 3-13b, the voltages are $\mathrm{V}=1$ on one conductor and $\mathrm{V}=0$ on the other conductor. This is called a Dirichlet's boundary condition; 2) Because of the symmetry inherent in this problem, we require that the normal derivative of the voltage be equal to 0 on the plane of symmetry. This is called a Neumann's boundary condition. The development of this technique will require some care.


Figure 3-13. (a) Cross section of $1 / 4$ of a square coaxial line is shown; (b) A finite element mesh is used to subdivide the square coaxial cable.

It was shown in Chapter 2 that the electric energy that can be stored in a volume with a cross section $A$ is

$$
\begin{equation*}
W=\left.\int_{A} \frac{\varepsilon}{2} \nabla V\right|^{2} \mathrm{ds} \tag{3.63}
\end{equation*}
$$

In Appendix 2, it is shown that this integral or "functional" has a minimum value for the actual solution of this boundary value problem. The necessary condition for this minimum leads to Laplace's equation for the potential V in the area A .

The most widely separated of the elements are the linear node elements for which the potential inside a triangle can be approximated with a linear polynomial

$$
\begin{equation*}
V^{(e)}(x, y)=a+b x+c y \tag{3.64}
\end{equation*}
$$

where there are three unknown coefficients $(a, b, c)$. These coefficients will have to be determined.

In order to accomplish this, it is necessary to apply three additional conditions. The potential at the three nodes of the triangle $(1,2,3)$ in Figure 3-14 are known. This approximation replaces the smooth solution with a piecewise smooth function that is based on a linear interpolation. It is easy to check that the continuous linear function (3.64) satisfies Laplace's equation (3.62). The coefficients $(a, b, c)$ can be determined from the given node potentials $\left(V_{1}, V_{2}, V_{3}\right)$ provided that the coordinates of the nodes are known quantities: $\left(x_{j}, y_{j}\right)(j=1,2,3)$.

Using the interpolation criteria

$$
\begin{equation*}
V^{(e)}\left(x_{j}, y_{j}\right)=V_{j}(j=1,2,3) \tag{3.65}
\end{equation*}
$$

the following matrix equation for the coefficients is obtained

$$
\left[\begin{array}{lll}
1 & x_{1} & y_{1}  \tag{3.66}\\
1 & x_{2} & y_{2} \\
1 & x_{3} & y_{3}
\end{array}\right]\left[\begin{array}{l}
a \\
b \\
c
\end{array}\right]=\left[\begin{array}{l}
\mathrm{V}_{1} \\
\mathrm{~V}_{2} \\
\mathrm{~V}_{3}
\end{array}\right]
$$

The determinant of the square matrix in (3.66) is equal to twice the area of the triangle $2 \mathrm{~A}_{\mathrm{e}}$.


Figure 3-14. Triangular finite element.
The substitution of the solution of the simultaneous equations (3.65) into (3.66) yields

$$
V^{(e)}(x, y)=\left[\begin{array}{lll}
1 & x & y
\end{array}\right] .\left[\begin{array}{lll}
1 & x_{1} & y_{1}  \tag{3.67}\\
1 & x_{2} & y_{2} \\
1 & x_{3} & y_{3}
\end{array}\right]^{-1} \cdot\left[\begin{array}{l}
V_{1} \\
V_{2} \\
V_{3}
\end{array}\right]
$$

After performing a matrix multiplication of the first two matrices, we arrive at the following expression for the potential at an arbitrary point within the triangle as being a linear combination of the potentials at the nodes of the triangle

$$
\begin{equation*}
V^{(e)}(x, y)=\sum_{i=1}^{3} V_{i} \alpha_{i}(x, y) \tag{3.68}
\end{equation*}
$$

This potential is defined in terms of a set of new functions

$$
\begin{equation*}
\alpha_{1}(x, y)=\frac{1}{2 A_{e}}\left[\left(x_{2} y_{3}-x_{3} y_{2}\right)+\left(y_{2}-y_{3}\right) x+\left(x_{3}-x_{2}\right) y\right] \tag{3.69}
\end{equation*}
$$

The other coefficients in (3.68) can be obtained with a cyclic permutation of the subscripts. Using the explicit expressions, it can be shown that these three linear basis functions satisfy the following interpolation criteria

$$
\alpha_{i}\left(x_{j}, y_{j}\right)= \begin{cases}0, & i \neq j  \tag{3.70}\\ 1, & i=j\end{cases}
$$

The details are included in Appendix 2.
We differentiate the potential that is given in (3.68)

$$
\begin{equation*}
\nabla \mathrm{V}^{(\mathrm{e})}(\mathrm{x}, \mathrm{y})=\sum_{\mathrm{i}=1}^{3} \mathrm{~V}_{\mathrm{i}} \nabla \alpha_{\mathrm{i}}(\mathrm{x}, \mathrm{y}) \tag{3.71}
\end{equation*}
$$

and using (3.63) to find that the energy satisfies the following quadratic form

$$
\begin{equation*}
W^{(e)}=\frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} S_{i, j}^{(e)} V_{i} V_{j} \tag{3.72}
\end{equation*}
$$

where the following matrix known as Dirichlet's matrix $\left[\mathrm{S}^{(\mathrm{e})}\right]$ for the finite element is defined as

$$
\begin{equation*}
\mathrm{S}_{\mathrm{i}, \mathrm{j}}^{(\mathrm{e})}=\varepsilon \int_{\mathrm{Ae}} \nabla \alpha_{\mathrm{i}} \bullet \nabla \alpha_{\mathrm{j}} \mathrm{dS} \tag{3.73}
\end{equation*}
$$

and the permittivity $\varepsilon=\varepsilon_{0} \varepsilon_{\mathrm{r}}$ is assumed to be a constant within the element. Analytical expressions for the elements of this matrix are derived in Appendix 2. The electric energy of the element from (3.72) can be written in a matrix form

$$
\begin{equation*}
\mathrm{W}^{(\mathrm{e})}=\frac{1}{2}[\mathrm{~V}]^{\top}\left[\mathrm{S}^{(\mathrm{e})}\right][\mathrm{V}] \tag{3.74}
\end{equation*}
$$

where the superscript " T " indicates a transpose matrix.
Example 3-14 MATLAB. Find the voltage distribution within the triangular element if the voltages at the three nodes have the following values.

$$
\mathrm{V}_{1}=8 @(0,0) ; \mathrm{V}_{2}=0 @(4,0) ; \text { and } \mathrm{V}_{3}=0 @(4,3) .
$$

Evaluate the electric energy that is stored in this element. In order to simplify the calculation, the energy should be expressed in terms of the dielectric constant $\varepsilon$. Compare your analytical results with a MATLAB calculation.


Answer: The voltage distribution within the element is computed from (3.67)

$$
\begin{aligned}
V^{(e)}(x, y) & =\left[\begin{array}{lll}
1 & x & y
\end{array}\right]\left[\begin{array}{lll}
1 & 0 & 0 \\
1 & 4 & 0 \\
1 & 4 & 3
\end{array}\right]^{-1}\left[\begin{array}{l}
8 \\
0 \\
0
\end{array}\right]=\left[\begin{array}{lll}
1 & x & y
\end{array}\right]\left[\begin{array}{ccc}
1 & 0 & 0 \\
-1 / 4 & 1 / 4 & 0 \\
0 & -1 / 3 & 1 / 3
\end{array}\right]\left[\begin{array}{l}
8 \\
0 \\
0
\end{array}\right] \\
& =\left[\begin{array}{lll}
1 & x & y
\end{array}\right]\left[\begin{array}{c}
8 \\
-2 \\
0
\end{array}\right]=8-2 x
\end{aligned}
$$

It is easy to check that this linear function satisfies the interpolation criteria (3.65) and Laplace's equation (3.62).

In order to calculate the energy, we must first calculate the numerical values for the explicit expressions for the $\left[S^{(e)}\right]$-matrix elements. The area enclosed within the triangle is $\mathrm{A}_{e}=6$. We explicitly write using (A2.9) and (A2.10) that

$$
\begin{aligned}
& \mathrm{S}_{1,1}^{(\mathrm{e})}=\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{2}-\mathrm{y}_{3}\right)^{2}+\left(\mathrm{x}_{2}-\mathrm{x}_{3}\right)^{2}\right]=\frac{\varepsilon}{4 \times 6}\left[(0-3)^{2}+(4-4)^{2}\right]=\frac{9 \varepsilon}{24} \\
& \begin{aligned}
\mathrm{S}_{1,2}^{(\mathrm{e})}=\mathrm{S}_{2,1}^{(\mathrm{e})} & =\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{2}-\mathrm{y}_{3}\right)\left(\mathrm{y}_{3}-\mathrm{y}_{1}\right)+\left(\mathrm{x}_{2}-\mathrm{x}_{3}\right)\left(\mathrm{x}_{3}-\mathrm{x}_{1}\right)\right] \\
& =\frac{\varepsilon}{4 \times 6}[(0-3)(3-0)+(4-4)(4-4)]=-\frac{9 \varepsilon}{24}
\end{aligned}
\end{aligned}
$$

$$
\begin{aligned}
& \begin{aligned}
\mathrm{S}_{1,3}^{(e)}=\mathrm{S}_{3,1}^{(e)} & =\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{3}-\mathrm{y}_{2}\right)\left(\mathrm{y}_{2}-\mathrm{y}_{1}\right)+\left(\mathrm{x}_{3}-\mathrm{x}_{2}\right)\left(\mathrm{x}_{2}-\mathrm{x}_{1}\right)\right] \\
& =\frac{\varepsilon}{4 \times 6}[(3-0)(0-0)+(4-4)(4-0)]=0
\end{aligned} \\
& \mathrm{~S}_{2,2}^{(e)}=\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{3}-\mathrm{y}_{1}\right)^{2}+\left(\mathrm{x}_{3}-\mathrm{x}_{1}\right)^{2}\right]=\frac{\varepsilon}{4 \times 6}\left[(3-0)^{2}+(4-0)^{2}\right]=\frac{25 \varepsilon}{24} \\
& \mathrm{~S}_{2,3}^{(e)}=\mathrm{S}_{3,2}^{(e)}=\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{3}-\mathrm{y}_{1}\right)\left(\mathrm{y}_{1}-\mathrm{y}_{2}\right)+\left(\mathrm{x}_{3}-\mathrm{x}_{1}\right)\left(\mathrm{x}_{1}-\mathrm{x}_{2}\right)\right] \\
& \\
& =\frac{\varepsilon}{4 \times 6}\left[(3-0)(0-0)+(4-0)(0-4]=-\frac{16 \varepsilon}{24}\right. \\
& \mathrm{S}_{3,3}^{(e)}=\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{1}-\mathrm{y}_{2}\right)^{2}+\left(\mathrm{x}_{1}-\mathrm{x}_{2}\right)^{2}\right]=\frac{\varepsilon}{4 \times 6}\left[(0-0)^{2}+(0-4)^{2}\right]=\frac{16 \varepsilon}{24}
\end{aligned}
$$

This is a written as the matrix

$$
S^{(e)}=\varepsilon\left[\begin{array}{ccc}
9 / 24 & -9 / 24 & 0 \\
-9 / 24 & 25 / 24 & -16 / 24 \\
0 & -16 / 24 & 16 / 24
\end{array}\right]
$$

The energy is computed from (3.74)

$$
\begin{aligned}
\frac{W_{e}}{\varepsilon} & =\frac{1}{2 \varepsilon}[V]^{\mathrm{t}}\left[S^{(e)}\right][\mathrm{V}]=\frac{1}{2}\left[\begin{array}{lll}
8 & 0 & 0
\end{array}\right]\left[\begin{array}{ccc}
9 / 24 & -9 / 24 & 0 \\
-9 / 24 & 25 / 24 & -16 / 24 \\
0 & -16 / 24 & 16 / 24
\end{array}\right]\left[\begin{array}{l}
8 \\
0 \\
0
\end{array}\right] \\
& =\frac{1}{2}\left[\begin{array}{lll}
8 & 0 & 0
\end{array}\right]\left[\begin{array}{c}
72 / 24 \\
-72 / 24 \\
0
\end{array}\right]=12
\end{aligned}
$$

These results are confirmed using MATLAB.

Finite Element Method using MATLAB
The total energy of the ensemble of all elements in the mesh can be calculated as just being the sum of the energy of each of the individual elements

$$
\begin{equation*}
W=\sum_{e} W^{(e)} \tag{3.75}
\end{equation*}
$$

We also need to know the procedure for adding a new element to the existing elements in the mesh. For simplicity, we consider the case where the mesh consists of only one element as shown in Figure 3-15. The problem is in obtaining the S-matrix of the new ensemble in terms of the S-matrix of the previous triangular mesh $\left[S^{(1)}\right]$ and the $S$-matrix of the new triangular mesh $\left[S^{(2)}\right]$. We should expect that the potential to be the same at all of the common nodes for both meshes. This will impose certain boundary conditions in that $\mathrm{V}_{1}=\mathrm{V}_{4}$ and $\mathrm{V}_{2}=\mathrm{V}_{6}$ after coupling. In addition, we also renumber the other node for convenience as $\mathrm{V}_{4}=\mathrm{V}_{5}$ after coupling.


Figure 3-15. (a) Decoupled elements; (b) Coupled elements.

The decoupled potentials can be written as a column matrix. The transpose of this matrix is

$$
[\mathrm{V}]_{d}=\left[\mathrm{V}_{1}, \mathrm{~V}_{2}, \mathrm{~V}_{3}, \mathrm{~V}_{4}, \mathrm{~V}_{5}, \mathrm{~V}_{6}\right]_{\mathrm{d}}^{\top},
$$

where the subscript "d" means "decoupled." If the S-matrices of the two elements are $\left[S^{(1)}\right]=\left[S_{i, j}\right](i, j=1,2,3)$ and $\left[S^{(2)}\right]=\left[S_{i, j}\right](i, j=4,5,6)$, then the global S -matrix of the decoupled system is the following block-diagonal square matrix

$$
\left[S^{(e)}\right]_{d}=\left[\begin{array}{lr}
{\left[S^{(1)}\right]} & {[0]}  \tag{3.76}\\
{[0]} & {\left[S^{(2)}\right]}
\end{array}\right]
$$

We introduce also the column matrix of the coupled potentials

$$
[\mathrm{V}]=\left[\mathrm{V}_{1}, \mathrm{~V}_{2}, \mathrm{~V}_{3}, \mathrm{~V}_{4}\right]^{\top} .
$$

The boundary conditions from Figure 3-15 can be written using a special rectangular coupling matrix [C] defined in Appendix 2. In this particular case, we obtain

$$
\left[\mathrm{S}^{(\mathrm{e})}\right]=\left[\begin{array}{cccr}
\mathrm{S}_{1,1}^{(1)}+\mathrm{S}_{4,4}^{(2)} & \mathrm{S}_{1,2}^{(1)}+\mathrm{S}_{4,6}^{(2)} & \mathrm{S}_{1,3}^{(1)} & \mathrm{S}_{4,5}^{(2)}  \tag{3.77}\\
\mathrm{S}_{2,1}^{(1)}+\mathrm{S}_{6,4}^{(2)} & \mathrm{S}_{2,2}^{(1)}+\mathrm{S}_{6,6}^{(2)} & \mathrm{S}_{2,3}^{(1)} & \mathrm{S}_{6,5}^{(2)} \\
\mathrm{S}_{3,1}^{(1)} & \mathrm{S}_{3,2}^{(1)} & \mathrm{S}_{3,3}^{(1)} & 0 \\
\mathrm{~S}_{5,4}^{(2)} & \mathrm{S}_{5,6}^{(2)} & 0 & \mathrm{~S}_{5,5}^{(2)}
\end{array}\right]
$$

Note that two types of subscript numeration are used: 1) local for decoupled elements and 2) global for coupled elements.

It is now time to consider the role that sources have on the system and to ascertain their effect on the solution of the boundary value problem. There are two types of sources: 1) surface sources that are prescribed potentials on the boundary $L_{1}$ in Laplace's equation and 2) volume sources such as a charge density in Poisson's equation. For simplicity, only the charge free Laplace's equation will be considered in this discussion. Poisson's equation can also be solved using similar techniques but it will not be discussed further. We assume

## Finite Element Method using MATLAB

surface sources are known which means that the column matrix of the coupled potentials [V]can be split into two parts: a) the unknown potentials [V] and b) the known potentials $[\mathrm{V}]_{k}$ or $[\mathrm{V}]=\left[[\mathrm{V}]_{\mathrm{u}}[\mathrm{V}]_{\mathrm{k}}\right]^{\top}$. The same separation holds for the matrix $\left[\mathrm{S}^{(\mathrm{e})}\right]$ which is split into four parts $\left[\mathrm{S}^{(\mathrm{e})}\right]=\left[\begin{array}{l}{[\mathrm{S}]_{\mathrm{u}, \mathrm{u}}[\mathrm{S}]_{\mathrm{u}, \mathrm{k}}} \\ {[\mathrm{S}]_{\mathrm{k}, \mathrm{u}}[\mathrm{S}]_{\mathrm{k}, \mathrm{k}}}\end{array}\right]$. The necessary condition that the function (3.74) has a minimum is that the appropriate derivatives must be equal to zero as shown in Appendix 2.

This condition imposes a relation between the known and the unknown potentials that is given in (A.2.16). We write

$$
\begin{equation*}
[\mathrm{V}]_{\mathrm{u}}=-[\mathrm{S}]_{\mathrm{u}, u}{ }^{-1}[\mathrm{~S}]_{\mathrm{u}, \mathrm{k}}[\mathrm{~V}]_{\mathrm{k}} \equiv[\mathrm{~F}][\mathrm{V}]_{\mathrm{k}} \tag{3.78}
\end{equation*}
$$

provided that the matrix $[S]_{u, u}$ is non-singular. The last matrix equation shows that the unknown potentials can be represented as a linear combination of the known potentials although the coefficients may implicitly depend upon the geometry of the mesh.

The main advantage of the FEM-method in comparison with the FDMmethod is its flexibility. This will be demonstrated by applying it to areas with different shapes. The shapes are covered by triangles in the domain of interest. This may lead to some complication in the mesh generation which is, of course, a disadvantage. This can be avoided to some extent by using a procedure for automatic mesh generation. In this book, the mesh is manually introduced for pedagogical reasons. There is a commercial package FEMLAB that can be used with MATLAB to solve more complicated problems then we will encounter here.

The procedure of using this technique has five stages:

1) Generation of the mesh;
2) Inclusion of the surface and volume sources;
3) Construction of the matrices for every element;
4) Collection of all the elements of the $\left[S^{(e)}\right]$-matrix;
5) Solution of the resulting matrix equation.

In order to illustrate the FEM-method, we consider a simple example.

Example 3-15. Find the potential in the point (1) using the FEM-method.


Answer: There are $N_{N}=5$ nodes denoted as \# = 1, 2, 3, 4, 5 and $N_{E}=4$ equal triangular elements \#1-4-5, 1-2-5, 1-2-3, and 1-3-4. The only unknown potential is $V_{1}$ and the other four potential have the numerical values: $V_{2}=V_{3}=0$, $\mathrm{V}_{4}=\mathrm{V}_{5}=\frac{10+0}{2}=5 \mathrm{~V}$.

Using the FEM-method, we first calculate the matrices using the known potentials at the points $\#=2,3,4,5$. Following the procedure outlined in the
previous example and (3.77), we find that the global sub-matrices required for the calculation in (3.78) are

$$
[S]_{\mathrm{u}, \mathrm{u}}=\left[4 \mathrm{~S}_{1,1}\right] ;[\mathrm{S}]_{\mathrm{u}, \mathrm{k}}=\left[2 \mathrm{~S}_{1,2}, 2 \mathrm{~S}_{1,3}, 2 \mathrm{~S}_{1,4}, 2 \mathrm{~S}_{1,5}\right]
$$

Using the expressions in the previous example or from Appendix 2, the elements of the local $\left[S^{(e)}\right]$-matrices can be calculated. The area of each of the triangular elements is equal to 1 and the coordinates of the nodal points can be obtained from the figure. We finally obtain

$$
\mathrm{S}_{1,1}=\varepsilon ; \mathrm{S}_{1,2}=\mathrm{S}_{1,3}=\mathrm{S}_{1,4}=\mathrm{S}_{1,5}=-\frac{\varepsilon}{2}
$$

The elements of the [F]-matrix are calculated from (3.78)

$$
[A]=-\frac{1}{4 \varepsilon}[-\varepsilon,-\varepsilon,-\varepsilon,-\varepsilon]=\frac{1}{4}\left[\begin{array}{lll}
1, & 1, & 1,
\end{array}\right]
$$

and since $[\mathrm{V}]_{\mathrm{U}}=\left[\mathrm{V}_{1}\right]$ and $[\mathrm{V}]_{\mathrm{k}}=\left[\mathrm{V}_{2}, \mathrm{~V}_{3}, \mathrm{~V}_{4}, \mathrm{~V}_{5}\right]^{\mathrm{t}}$, we finally obtain

$$
V_{1}=\frac{1}{4}\left(V_{2}+V_{3}+V_{4}+V_{5}\right)=\frac{1}{4}(0+0+5+5)=2.5 \mathrm{~V}
$$

This problem can be also evaluated using the FDM-method. We can write the potential in the point (1) as an average value of the potentials in other four points (\# = 6,7,8,9)

$$
V_{1}=\frac{1}{4}\left(V_{6}+V_{7}+V_{8}+V_{9}\right)=\frac{1}{4}(0+0+10+0)=2.5 \mathrm{~V} .
$$

This is the same result that we had obtained previously. There are cases where the FEM-method has an advantage over the other techniques. In certain cases, the other methods actually fail!

In order to illustrate the power of the FEM-method, we examine a rectangular coaxial line. Because of the symmetry inherent in this problem, we shall consider only $1 / 4$ of the coaxial line as shown in the Figure 3-15a. Even though there are only a few elements in the structure, the actual mathematics will become too tedious to do by hand. The energy of the capacitor is found from equation (3.74) and (3.75). The capacitance can be calculated from the expression $W=\frac{1}{2} C_{0} V^{2}$ which yields

$$
\begin{equation*}
\mathrm{C}_{0}=\frac{2 \mathrm{~W}}{\mathrm{~V}^{2}} \tag{3.79}
\end{equation*}
$$

Example 3-16 ${ }^{\text {MATLAB }}$. Write a MATLAB program to:

1) Calculate the potentials in the three free nodes with \# = 4, 5, 9;
2) Calculate the total electrostatic energy $W$ stored between the plates and find the capacitance $C_{0}$ of 1 m of the line.

Use a simple regular mesh shown in the figure and the following numerical values for the parameters: $a=1 \mathrm{~cm}, \mathrm{~b}=2 \mathrm{~cm}, \mathrm{~V}=10 \mathrm{~V}$ on the outer conductor, and $V=0 V$ on the inner conductor. In this case, there are $N_{N}=11$ nodes and $N_{E}=12$ elements.


Answer: 1) The potential is found by solving the matrix equation (3.78) using there a mesh that is manually generated. The unknown potentials calculated with the MATLAB program are:

$$
\mathrm{V}_{4}=5 \mathrm{~V} ; \mathrm{V}_{5}=7.5 \mathrm{~V} ; \mathrm{V}_{9}=5 \mathrm{~V}
$$

2) The calculated normalized capacitance $\left(C=C_{0} / \varepsilon_{0}\right)$ per unit length of the line calculated from (3.79) is $C=11.0$. The value, found in the literature for this square coaxial line with $b=2 a$, is $C=10.2341$. The accuracy can be improved by increasing the number of the nodes and elements which will be shown in the next example.

Example 3-17 MATLAB . Solve the same boundary value problem as in the previous example but apply a finer mesh shown in the figure below. There are $\mathrm{N}_{\mathrm{N}}=21$
nodes and $\mathrm{N}_{\mathrm{E}}=24$ triangular elements. In addition, plot the equipotential contours and the electric field between the two surfaces.


Answer: 1) The potential is found by solving the matrix equation (3.78) using there a mesh that is manually generated. . The unknown potentials obtained from the MATLAB program are:

$$
\begin{aligned}
& \mathrm{V}_{6}=5.111 \mathrm{~V} ; \mathrm{V}_{7}=5.2222 \mathrm{~V} ; \mathrm{V}_{8}=5.7778 \mathrm{~V} ; \mathrm{V}_{9}=7.8889 \mathrm{~V} ; \\
& \mathrm{V}_{14}=5.7778 \mathrm{~V} ; \mathrm{V}_{17}=5.2222 \mathrm{~V} ; \mathrm{V}_{20}=5.1111 \mathrm{~V}
\end{aligned}
$$

2) The calculated normalized capacitance ( $C=C_{0} / \varepsilon_{0}$ ) for a unit length of the line is $C=10.8444$. The relative error achieved here is smaller than in the previous example.
3) The equipotential contours are plotted below using the 'contour' function. The electric field is determined using the 'gradient' function and displayed using the 'quiver' function.

Finite Element Method using MATLAB


Figure 3-17. A microstrip transmission line.

The previous two examples were solutions of a closed electrostatics problem. Closed problems can easily be handled by the FEM method. A general conclusion can be drawn from these examples in that the accuracy can be

## Finite Element Method using MATLAB

improved by using a finer mesh structure. However, increasing the accuracy also increases the computational time.

An open electrostatic problem is shown in Figure 3-17. This would correspond to an open microstrip line. A dielectric with a dielectric constant $\varepsilon_{\mathrm{r}}>1$ separates the conductors and also increases the capacitance of the line. It is not appropriate to use either the FEM-method or the FDM-method to calculate the capacitance or the stored electrostatic energy for structures of this type. This is because boundary conditions at infinity will have to be invoked at the open surfaces. It is possible to introduce an absorbing boundary condition at these surfaces. However, it is usually better to follow a different path of using the method of moments which will be introduced in the next section.

### 3.7. Method of Moments using MATLAB

In the previous chapter, we found that the electric potential V could be computed from a known charge distribution. This was accomplished using integral

$$
\begin{equation*}
V(x, y, z)=\frac{1}{4 \pi \varepsilon_{0}} \iiint_{\Delta V} \frac{\rho_{V}\left(x^{\prime}, y^{\prime}, z^{\prime}\right)}{R} d x^{\prime} d y^{\prime} d z^{\prime} \tag{3.80}
\end{equation*}
$$

where $R$ is the distance between the charge located at the point ( $x^{\prime}, y^{\prime}, z^{\prime}$ ) and the point of observation at the point ( $x, y, z$ ). If the charge distribution is known, then the potential can be easily computed. We note that (3.80) can be converted into a summation and hence the integral can be evaluated numerically.

There are cases, however, where the potential may actually be known and the charge distribution may be unknown. Static fields abound with such problems. An example would be the determination of an unknown surface charge distribution on a conductor if the potential of the conductor was specified.

The technique that will be introduced is called the "Method of Moments" and it will be identified as "MoM" in the following discussion. This technique will be very powerful in calculating the capacitance of various metallic objects. It is also useful in calculating the capacitance of a transmission line that will be encountered later. Finally, it is useful in determining the shapes of various objects such as planes and rockets that may be impinging upon a nation by correctly interpreting the reflected high frequency signals from the objects by the observer.

Consider the configuration shown in Figure 3-18. Four charges are located in space. A Cartesian coordinate system is also introduced and the location


Figure 3-18. Four charges distributed in space. The potential at the indicated points are assumed to be $V=-1$ and $V=+1$.
of the centers of the four charges are specified with reference to this coordinate system. The potential at two of the charges $\left(Q_{1}\right.$ and $\left.Q_{2}\right)$ is specified to be $V_{1}=V_{2}=-1$ and the potential at the other two $\left(Q_{3}\right.$ and $\left.Q_{4}\right)$ is specified to be $\mathrm{V}_{3}=\mathrm{V}_{4}=+1$. The value of the individual charges is unknown. In order to obtain a unique solution for the values, of these 4 charges, we must be able to writedown 4 equations that will describe the potential at the 4 defined locations. We assume that the region is a vacuum and we can use superposition. We write four linear equations for the potentials at the four points

$$
\begin{align*}
& -1=\frac{1}{4 \pi \varepsilon_{0}}\left(\frac{Q_{1}}{\left|\mathbf{r}_{1}-\mathbf{r}_{1}\right|}+\frac{Q_{2}}{\left|\mathbf{r}_{1}-\mathbf{r}_{2}\right|}+\frac{Q_{3}}{\left|\mathbf{r}_{1}-\mathbf{r}_{3}\right|}+\frac{Q_{4}}{\left|\mathbf{r}_{1}-\mathbf{r}_{4}\right|}\right) \\
& -1=\frac{1}{4 \pi \varepsilon_{0}}\left(\frac{Q_{1}}{\left|\mathbf{r}_{2}-\mathbf{r}_{1}\right|}+\frac{Q_{2}}{\left|\mathbf{r}_{2}-\mathbf{r}_{2}\right|}+\frac{Q_{3}}{\left|\mathbf{r}_{3}-\mathbf{r}_{3}\right|}+\frac{Q_{4}}{\left|\mathbf{r}_{4}-\mathbf{r}_{4}\right|}\right)  \tag{3.81}\\
& +1=\frac{1}{4 \pi \varepsilon_{0}}\left(\frac{Q_{1}}{\left|\mathbf{r}_{3}-\mathbf{r}_{1}\right|}+\frac{Q_{2}}{\left|\mathbf{r}_{3}-\mathbf{r}_{2}\right|}+\frac{Q_{3}}{\left|\mathbf{r}_{3}-\mathbf{r}_{3}\right|}+\frac{Q_{4}}{\left|\mathbf{r}_{3}-\mathbf{r}_{4}\right|}\right) \\
& +1=\frac{1}{4 \pi \varepsilon_{0}}\left(\frac{Q_{1}}{\left|\mathbf{r}_{4}-\mathbf{r}_{1}\right|}+\frac{Q_{2}}{\left|\mathbf{r}_{4}-\mathbf{r}_{2}\right|}+\frac{Q_{3}}{\left|\mathbf{r}_{4}-\mathbf{r}_{3}\right|}+\frac{Q_{4}}{\left|\mathbf{r}_{4}-\mathbf{r}_{4}\right|}\right)
\end{align*}
$$

This can be written using the summation sign as

$$
\begin{equation*}
V_{i}=\frac{1}{4 \pi \varepsilon_{0}} \sum_{j=1}^{4} \frac{1}{\left|\mathbf{r}_{\mathbf{i}}-\mathbf{r}_{\mathbf{j}}\right|} \cdot \mathbf{Q}_{\mathrm{j}} \tag{j=1,2,3,4}
\end{equation*}
$$

The four equations in (3.81) can also be written in matrix notation. Remember that MATLAB was originally created in order to solve problems of the type

$$
\begin{equation*}
[\mathrm{P}][\mathrm{Q}]=[\mathrm{V}] \tag{3.82}
\end{equation*}
$$

where [ V ] is the column vector of the known potentials, [Q] is the column vector of the unknown charges and $[P]$ is the square matrix of coefficients

$$
[\mathrm{P}]=\frac{1}{4 \pi \varepsilon_{0}}\left[\begin{array}{cccc}
\frac{1}{\left|\mathbf{r}_{1}-\mathbf{r}_{1}\right|} & \frac{1}{\left|\mathbf{r}_{1}-\mathbf{r}_{2}\right|} & \frac{1}{\left|\mathbf{r}_{1}-\mathbf{r}_{3}\right|} & \frac{1}{\left|\mathbf{r}_{1}-\mathbf{r}_{4}\right|}  \tag{3.83}\\
\frac{1}{\left|\mathbf{r}_{2}-\mathbf{r}_{1}\right|} & \frac{1}{\left|\mathbf{r}_{2}-\mathbf{r}_{2}\right|} & \frac{1}{\left|\mathbf{r}_{2}-\mathbf{r}_{3}\right|} & \frac{1}{\left|\mathbf{r}_{2}-\mathbf{r}_{4}\right|} \\
\frac{1}{\left|\mathbf{r}_{3}-\mathbf{r}_{1}\right|} & \frac{1}{\left|\mathbf{r}_{3}-\mathbf{r}_{2}\right|} & \frac{1}{\left|\mathbf{r}_{3}-\mathbf{r}_{3}\right|} & \frac{1}{\left|\mathbf{r}_{3}-\mathbf{r}_{4}\right|} \\
\frac{1}{\left|\mathbf{r}_{4}-\mathbf{r}_{1}\right|} & \frac{1}{\left|\mathbf{r}_{4}-\mathbf{r}_{2}\right|} & \frac{1}{\mathbf{r}_{4}-\mathbf{r}_{3} \mid} & \frac{1}{\left|\mathbf{r}_{4}-\mathbf{r}_{4}\right|}
\end{array}\right]
$$

This matrix is symmetric because the potential between the charge and the point of observation depends upon the magnitude of the distance $R$ between the two points.

The diagonal terms of this matrix ( $\mathrm{i}=\mathrm{j}$ ) appear to give us problems since they become very large. These terms are called singular. We remove this singularity with an approximation. The approximation makes the assumption that

## Method of moments using MATLAB

the potential at these singular points is evaluated at the edge of the spherical charge that has a radius a and not at the center. It maintains that potential throughout the interior of the spherical charge. The diagonal elements of a matrix [P] are

$$
\begin{equation*}
\mathrm{P}_{\mathrm{i}, \mathrm{i}}=\frac{1}{4 \pi \varepsilon_{0} \mathrm{a}} \tag{3.84}
\end{equation*}
$$

Example 3-18 MATLAB . Find the values of the charges that will cause the potentials as shown in the Figure $3-18$, if the coordinates of the points in the plane $z=0$ are: $Q_{1}(2,3), Q_{2}(2,2), Q_{3}(5,3)$ and $Q_{4}(5,2)$. Assume that the diameter of the charges is $2 \mathrm{a}=1$ meter.

Answer: The matrix $[P]$ in (3.83) and (3.84) has the elements

$$
[P]=\frac{1}{4 \pi \varepsilon_{0}}\left[\begin{array}{llll}
\frac{1}{(1 / 2)} & \frac{1}{1} & \frac{1}{3} & \frac{1}{\sqrt{1^{2}+3^{2}}} \\
\frac{1}{1} & \frac{1}{(1 / 2} & \frac{1}{\sqrt{1^{2}+3^{2}}} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{\sqrt{1^{2}+3^{2}}} & \frac{1}{(1 / 2)} & \frac{1}{1} \\
\frac{1}{\sqrt{1^{2}+3^{2}}} & \frac{1}{3} & \frac{1}{1} & \frac{1}{(1 / 2)}
\end{array}\right]=\frac{1}{4 \pi \varepsilon_{0}}\left[\begin{array}{ccccc}
2 & 1 & 0.3333 & 0.3162 \\
1 & 2 & 0.3162 & 0.3333 \\
0.3333 & 0.3162 & 2 & 1 \\
0.3162 & 0.3333 & 1 & 2
\end{array}\right] .
$$

The column vector for the potential is $[\mathrm{V}]=[-1,-1,1,1]^{\top}$ where the " $T$ " indicates the transpose. Solving the matrix equation (3.82) leads to

$$
\mathrm{Q}_{1}=\mathrm{Q}_{2}=-\mathrm{Q}_{3}=-\mathrm{Q}_{4}=-.4254\left(4 \pi \varepsilon_{0}\right)
$$

We could continue on with individual charges as has been presented up to this point. However, it is more meaningful to examine cases where the charge is distributed upon various surfaces. If the charge were distributed on a line as shown in Figure 3-19, it would be prudent to describe the charge distribution with a linear charge density $\rho_{\mathrm{L}}(\mathrm{C} / \mathrm{m})$. The charge on a particular element $j$ would be

## Method of moments using MATLAB

$\Delta \mathrm{Q}_{\mathrm{j}}=2 \pi \mathrm{a} \Delta \mathrm{l}_{\mathrm{j}} \rho_{\mathrm{L}, \mathrm{j}}(\mathrm{j}=1,2, \ldots \mathrm{~N})$ and it would be located at the center of the jth section. We proceed using the same method that has just been described. The identical problem with the singularity that was discussed above with individual charges will also be encountered in cylindrical coordinates.


Figure 3-19. The potential at point $P$ results from charges $\rho_{L} \Delta \Delta_{j}$ located at the centers of the jth section.

In this case, the column vector of the unknown charges is chosen to be $\left[\frac{\mathrm{Q}}{\Delta \mathrm{L}}\right]=\left[\rho_{\mathrm{L}, 1}, \rho_{\mathrm{L}, 2}, \ldots, \rho_{\mathrm{L}, \mathrm{N}}\right]^{\top}$ while the column -vector of the known potentials is written as $[\mathrm{V}]=\left[\mathrm{V}_{1}, \mathrm{~V}_{2}, \ldots, \mathrm{~V}_{N}\right]^{\top}$. The off-diagonal terms of the square-matrix are written as

$$
\begin{equation*}
\mathrm{P}_{\mathrm{i}, \mathrm{j}}=\frac{2 \pi \mathrm{a}}{4 \pi \varepsilon_{0}} \frac{\Delta \mathrm{l}_{\mathrm{j}}}{\left|\mathrm{x}_{\mathrm{i}}-\mathrm{x}_{\mathrm{j}}\right|}=\frac{\mathrm{a} \Delta \mathrm{l}_{\mathrm{j}}}{2 \varepsilon_{0}\left|\mathrm{x}_{\mathrm{i}}-\mathrm{x}_{\mathrm{j}}\right|} \tag{3.85}
\end{equation*}
$$

The singularity in the diagonal terms of the matrix will also be encountered here but they also can be removed. We evaluate the potential at the surface of the cylinder and assert that it is also equal to the potential at the center which is the "singular point." The evaluation of this potential $\mathrm{V}_{\mathrm{j}}$ at the surface of this cylindrical section shown in Figure 3-20 is calculated using the integral

# Method of moments using MATLAB 

$$
V_{\mathrm{j}}=\frac{1}{4 \pi \varepsilon_{0}} \int_{-\frac{\Delta_{j}}{2}}^{\frac{\Delta_{\mathrm{j}}}{2}} \int_{0}^{2 \pi} \frac{\rho_{\mathrm{L}} \mathrm{ad} \varphi \mathrm{dx}}{} \sqrt{\mathrm{a}^{2}+\mathrm{x}^{\prime 2}}
$$

The integral can be performed and we find for the diagonal elements

$$
\mathrm{P}_{\mathrm{i}, \mathrm{j}}=\frac{2 \pi \mathrm{a}}{4 \pi \varepsilon_{0}} \ln \left[\mathrm{x}^{\prime}+\sqrt{\mathrm{a}^{2}+\mathrm{x}^{\prime 2}}\right]_{-\frac{-\mathrm{b}_{\mathrm{j}}}{2}}^{\frac{\Delta \mathrm{l}_{\mathrm{j}}}{2}}=\frac{\mathrm{a}}{2 \varepsilon_{0}} \ln \left[\frac{\frac{\Delta \mathrm{l}_{\mathrm{j}}}{2}+\sqrt{\mathrm{a}^{2}+\left(\frac{\Delta \mathrm{l}_{\mathrm{j}}}{2}\right)^{2}}}{-\frac{\Delta \mathrm{l}_{\mathrm{j}}}{2}+\sqrt{\mathrm{a}^{2}+\left(\frac{\Delta \mathrm{l}_{\mathrm{j}}}{2}\right)^{2}}}\right]
$$

If we make the approximation that the radius a is much less than the length of the section $\Delta l_{\mathrm{j}}\left(\mathrm{a} \ll \Delta \mathrm{l}_{\mathrm{j}}\right)$, this simplifies to

$$
\begin{equation*}
\mathrm{P}_{\mathrm{j}, \mathrm{j}} \cong \frac{\mathrm{a}}{\varepsilon_{0}} \ln \left[\frac{\Delta \mathrm{l}_{\mathrm{j}}}{\mathrm{a}}\right] \tag{3.86}
\end{equation*}
$$



Figure 3-20. The $\mathrm{j}^{\text {th }}$ section of a linearly charged line.

Example 3-19 MATLAB. Find the charge distribution on the cylindrical conductor whose radius is $a=0.01 \mathrm{~m}$ and whose length is $L=1 \mathrm{~m}$. The potential on the surface is $\mathrm{V}=1 \mathrm{~V}$. You may assume that the charge is uniformly distributed in each section. Assume that the number of the sections is $N=5$ and the step-size is $\Delta \mathrm{I}=0.20 \mathrm{~m}$.


Answer: The matrix equation relating the potentials to the charges is (3.82), where the off-diagonal and the diagonal elements are given by (3.85) and (3.86) respectively. The solution for the unknown charge distribution is

$$
\rho_{\mathrm{L}, 1}=\rho_{\mathrm{L}, 5}=.2556 \varepsilon_{0}, \rho_{\mathrm{L}, 2}=\rho_{\mathrm{L}, 4}=.2222 \varepsilon_{0}, \rho_{\mathrm{L}, 3}=.2170 \varepsilon_{0} .
$$

Note that the charge density in the center of the line is smaller than at either end.
We should expect this nonuniform distribution since there is a loss of symmetry at either end.

Let us also apply the method of moments to a slightly different topic, that of ascertaining the expected one dimensional charge distribution from a known potential profile. The potential profile and the resulting charge distribution could be very nonuniform as in, for example, the depletion layer of a pn junction. In this case, we assume that there are a sequence of sheets of charge as shown in Figure 3-21. The incremental charge density on each sheet "j" is uniform across the plane of that particular sheet and it has a value $\rho_{\mathrm{S}, \mathrm{j}}$. The separation between each sheet will also be assumed to be uniform with a separation distance of d .


Figure 3-21. Charged sheet model to represent the pn junction.

The electric field surrounding an infinite plane of constant surface charge density $\rho_{\mathrm{s}, \mathrm{j}}\left(\mathrm{C} / \mathrm{m}^{2}\right)$ is given by

$$
\mathrm{E}_{\mathrm{j}}=\frac{\rho_{\mathrm{s}, \mathrm{j}}}{2 \varepsilon}
$$

The electric potential at a point $z_{i}$ at a distance $z=\left|z_{i}-z_{j}\right|$ from the charged sheet $j$ is found from the integral of the electric field

$$
\begin{equation*}
V_{j}=-\frac{\rho_{\mathrm{s}, \mathrm{j}}}{2 \varepsilon} \mathrm{z} \tag{3.87}
\end{equation*}
$$

where the constant of integration in (3.87) is set equal to zero.
The potentials at the two extremities $z_{1}=-2 d$ and $z_{5}=+2 d$ in Figure 3-21 have a value that is equal to $1 / 2$ of the value that is given in (3.87). This additional factor of $1 / 2$ arises since we are only interested in the electric field that is be confined within the region of interest. The singularity that was previously encountered in the diagonal terms is automatically removed since the factor of " 0 " is introduced in (3.87). In this case the column vector of the source is $\left[\frac{\mathrm{Q}}{\Delta \mathrm{s}}\right]=\left[\rho_{\mathrm{s}, 1}, \ldots, \rho_{\mathrm{s}, 5}\right]^{\top}$. The elements in the square-matrix [P] are obtained from (3.87) to be

$$
P_{i, j}= \begin{cases}-\frac{d}{2 \varepsilon}\left|z_{i}-z_{j}\right|, & j \neq 1 \text { or } 5  \tag{3.88}\\ -\frac{d}{4 \varepsilon}\left|z_{i}-z_{j}\right|, & j=1 \text { or } 5\end{cases}
$$

or as a matrix

$$
[P]=-\frac{d}{2 \varepsilon}\left[\begin{array}{lllll}
0 & 1 & 2 & 3 & 4 / 2 \\
1 / 2 & 0 & 1 & 2 & 3 / 2 \\
2 / 2 & 1 & 0 & 1 & 2 / 2 \\
3 / 2 & 2 & 1 & 0 & 1 / 2 \\
4 / 2 & 3 & 2 & 1 & 0
\end{array}\right]
$$

Let us now specify the values of the potential as indicative of a linear variation in space to be $[\mathrm{V}]=[-2,-1,0,1,2]^{\top}$. Solving the matrix equation (3.82) we find that

$$
\begin{equation*}
\left[\rho_{\mathrm{s}, 1}, \rho_{\mathrm{s}, 2}, \rho_{\mathrm{s}, 3}, \rho_{\mathrm{s}, 4}, \rho_{\mathrm{s}, 5}\right]^{\top}=\frac{2 \varepsilon}{\mathrm{~d}}[-1,0,0,0,+1]^{\top} \tag{3.89}
\end{equation*}
$$

This result expresses the fact that the electric field that is proportional to the negative gradient of the linearly varying potential is a constant that is determined by the charge densities that are localized at either edge. Other potential distributions could be employed. This would lead to a different charge distribution. ${ }^{1}$

Other potential distributions can be assumed. In studying a $P N$ junction, you may encounter the term "linear graded junction" or a "quadratic graded junction." A different choice will, of course, yield a different charge distribution.

The charge could have been distributed upon a surface resulting in an inhomogeneous surface charge density $\rho_{\mathrm{S}}$. In this case, we subdivide the surface into small rectangular areas $\Delta \mathrm{s}_{\mathrm{i}}(\mathrm{i}=1,2, \ldots, \mathrm{~N})$. This is shown in Figure 3-22. In this case the column vector of the unknown charges is taken $\left[\frac{\mathrm{Q}}{\Delta \mathrm{s}}\right]=\left[\rho_{\mathrm{s}, 1}, \rho_{\mathrm{s}, 2}, \ldots, \rho_{\mathrm{s}, \mathrm{N}}\right]^{\top}$

[^7]and the column vector of the known potentials is $[\mathrm{V}]=\left[\mathrm{V}_{1}, \mathrm{~V}_{2}, \ldots, \mathrm{~V}_{\mathrm{N}}\right]^{\top}$. In this case, $N=M \times M$ is the number of the square sections with individual areas $\Delta s_{i}=$ $a^{2}$. The mutual-coupling terms of the square-matrix are easily obtained again to be
\[

$$
\begin{equation*}
\mathrm{P}_{\mathrm{i}, \mathrm{j}}=\frac{1}{4 \pi \varepsilon_{0}} \frac{\Delta \mathbf{s}_{\mathrm{j}}}{\left|\mathbf{r}_{\mathbf{i}}-\mathbf{r}_{\mathrm{j}}\right|} \tag{3.90}
\end{equation*}
$$

\]

The reader should expect at this stage that we will again encounter a singularity in the diagonal terms. In order to remove the singularity, we replace the small rectangular subarea whose area is $\Delta s$ with a circular region that contains the same incremental charge. This implies that the area of the circle $\pi \mathrm{b}^{2}$ and the square subarea $a^{2}$ are equal. With this assumption, the radius of the circle b can be calculated to be $\mathrm{b}=\mathrm{a} / \sqrt{\pi}$. We assume that this charge that is distributed within the circle is localized at the center of the circle. We then compute the potential at the perimeter of the circle from this localized charge and assume that it has this value throughout the circle. The evaluation of the potential leads to the following approximation for the diagonal terms

$$
\begin{equation*}
\mathrm{P}_{\mathrm{i}, \mathrm{i}}=\frac{1}{4 \pi \varepsilon_{0}} \int_{0}^{2 \pi \mathrm{~b}} \int_{0}^{\mathrm{rdrd} \varphi} \frac{1}{\mathrm{r}}=\frac{1}{4 \pi \varepsilon_{0}}(2 \pi \mathrm{~b})=\frac{\mathrm{b}}{2 \varepsilon_{0}}=\frac{1}{2 \varepsilon_{0}} \sqrt{\frac{\Delta \mathrm{~S}_{\mathrm{i}}}{\pi}} \tag{3.91}
\end{equation*}
$$

where the radius of the circle is written in terms of the area of the grid element.


## Method of moments using MATLAB

Figure 3-22. Charge is distributed on a surface and has a density $\rho_{\mathrm{S}}$. Singular elements are replaced with discs with the same area.

Typical problems that are encountered in this case would be the calculation of the capacitance of a parallel plate capacitor. In such a calculation, the two plates are each subdivided into N subareas as shown in Figure 3-22. Here the capacitance $\mathrm{C}_{0}$ is calculated from the equation

$$
\begin{equation*}
\mathrm{C}_{0}=\mathrm{Q} / \mathrm{V} \tag{3.92}
\end{equation*}
$$

where $Q$ is the total charge stored on the top plate and V is the voltage difference between the plates.

Example 3-20 MATLAB. Two charged parallel square plates with dimensions
$L \times L=1 m^{2}$ are separated by a distance of $d=0.1$ meter. Each side is subdivided into $N=64$ equal subareas The potential of the top plate is +5 V and the potential of the bottom plate is -5 V .
a) Find and plot the charge density distribution using the MoM;
b) Find the capacitance C of this charged conductor system. Compare the limiting case with the known simple classical solution $C / \varepsilon_{0}=L^{2} / \mathrm{d}=10$.


Answer: a)The column -vector of the known potential is

$$
[\mathrm{V}]=[+5,+5, \ldots,+5 ;-5,-5, \ldots,-5]^{\top}
$$

The matrix equation which must be solved is again (3.82) with a matrix [P] described by (3.90) and (3.91).

The solution of this matrix equation for the surface charge distribution is plotted in the following figure.

b) The normalized capacitance $\mathrm{C}_{0}=\mathrm{C} / \varepsilon_{0}$ is calculated to yield of value of $C_{0}=13.3811$ for the value of the number of sections $N=64$. This value is larger than that predicted from the elementary formula $C_{0}=L^{2} / d=10.0$. This formula assumed that the separation distance was significantly less than the area of the plate and all fringing fields at the edges could be neglected. The accuracy can be improved by subdividing the area into smaller subareas.

This is a brief introduction to the method of moments. As we have seen, the technique is very useful in determining a source distribution such as a charge in terms of a response function such as voltage between two plates containing an homogeneous dielectric. A further discussion of this technique is usually reserved for more advanced courses.

## Conclusion

### 3.8. Conclusion

Solving boundary value problems for potentials has led us to certain general conclusions concerning the methodical procedure. First, nature has given us certain physical phenomena that can be described by partial or ordinary differential equations. In many cases, these equations can be solved analytically.Other cases may require numerical solutions. The analytical solutions contain constants of integration. Nature also provides us with enough information that will allow us to evaluate these constants and thus obtain the solution for the problem.Assuming that neither mathematical nor numerical mistakes have been made, we can rest assured that this is the solution.

Five techniques that are frequently encountered when we attempt a solution of more complicated problems in electromagnetic theory were introduced in this chapter. The first two are analytical methods: 1) direct integration of a onedimensional equation. 2) Fourier series expansion of a two-dimensional equation. The latter three techniques involve the application of numerical methods: 3) finite difference method; 4) finite element method; and 5) method of moments. The methods are described in both 1 and 2 dimensions The first two of them are based on differential equations for the electric potential (Poisson's and Laplace's equations), while the last one is based on an integral equation for the unknown charge distribution. The numerical techniques that were introduced here are also applicable to three-dimensional problems that may be later encountered. The numerical programs that had been written using MATLAB are available at the web site: [http://www.scitechpub.com/]

As presented in this chapter, all of the methods assume linearity which leads to linear superposition principles. However, you will frequently encounter nonlinearity in nature which will lead to significant alterations in your method of obtaining a solution. Numerical questions concerning other specific programming

## Conclusion

languages, convergence requirements, numerical errors, aliasing, etc. are better left to later courses. As problems arise, solutions can hopefully be found.

## Problems 3

### 3.9. Problems

1. For the indicated boundary conditions that are specified in the figure, find the potential distribution $\mathrm{V}(\mathrm{x}, \mathrm{y})$ within the enclosed region by solving Laplace's equation and expanding one boundary condition in a Fourier series expansion.

0

2. Find the potential within the channel given in problem 1 using the boundary conditions at $x=0, x=a$ and $y=\infty$ as stated there. The boundary condition at $y$ $=0$ is given as $E=E_{O} u_{y}$ where $E_{O}$ is a constant.
3. Find the potential within the channel given in problem 1 using the boundary conditions at $\mathrm{x}=0, \mathrm{x}=\mathrm{a}$ and $\mathrm{y}=\infty$ as stated there. The boundary condition at y $=0$ is: $\mathrm{V}=+\mathrm{V}_{\mathrm{O}}$ for $0<\mathrm{x}<(\mathrm{a} / 2)$ and $\mathrm{V}=-\mathrm{V}_{\mathrm{O}}$ for $(\mathrm{a} / 2)<\mathrm{x}<\mathrm{a}$.
4. Find the components of the electric field $\left(E_{x}, E_{y}\right)$ in problem 3.
5. Find the expression for a potential $V(x)=V_{O}(x / a)$ that describes the potential variation in the region $0 \leq x \leq a$.
6. Find the potential distribution $\mathrm{V}(\rho)$ by solving Laplace's equation analytically for the region between two concentric hollow spheres (spherical capacitor). Apply a spherical coordinate system with the following boundary conditions: $\mathrm{V}=0$ at $\rho=a$ and $\mathrm{V}=\mathrm{V}_{0}$ at $\rho=\mathrm{b}$. Simplify the calculation with symmetry arguments.

# Problems 3 


7. Find the capacitance $C$ of the spherical capacitor in problem 5.
8. Find the potential distribution $\mathrm{V}(\theta)$ by solving Laplace's equation analytically for the region between two hollow coaxial cones. A potential $\mathrm{V}=\mathrm{V}_{1}$ is assumed at $\theta=\theta_{1}$ and $\mathrm{V}=0$ at $\theta=\theta_{2}=\pi-\theta_{1}$. The vertices of the cones are insulated at $\rho=0$. Simplify the calculation with symmetry arguments.

9. Find the potential $V(x)$ in the region $0<x<1$ satisfying the boundary conditions: $\quad \mathrm{V}(0)=3$ and $\mathrm{V}(1)=0$.
10. Find the potential $V(x)$ in the region $0<x<2$ if the electric field is normal with a constant value $E_{x}=4$.
11. Find the potential $V(x)$ in the region $0<x<1$ assume that a charge uniformly distributed there with a density $\rho_{v}=-4 \varepsilon_{0}$. The potential satisfies the boundary conditions: $\mathrm{V}(0)=3$ and $\mathrm{V}(1)=0$.
12. Find the normal electric field $\mathrm{E}_{\mathrm{x}}(\mathrm{x})$ in problem 11.

## Problems 3

13. Find the capacitance $C_{0}$ of unit length of the cylindrical capacitor - two long concentric cylinders with radiuses $a$ and $b(b>a)$. The boundary conditions for the potential are: $\mathrm{V}(\mathrm{r}=\mathrm{a})=\mathrm{V}_{0}$ and $\mathrm{V}(\mathrm{r}=\mathrm{b})=0$.
14. Find the potential distribution $V(x, y)$ in the region $y>0$. Because of the periodicity of the boundary condition, expand boundary in a Fourier series. This problem models

a VLSI circuit where conductors are
implanted on an insulating material.
The thickness of the metal strip can
be neglected.
$15^{\text {MATLAB }}$. For the indicated boundary conditions that are specified in the figure, find the potential distribution $\mathrm{V}(\mathrm{x}, \mathrm{y})$ within the enclosed region by solving Laplace's equation. Plot the potential distribution $\left(a=1 \mathrm{~m}, \mathrm{~V}_{0}=10 \mathrm{~V}\right)$.

$16^{\text {MATLAB }}$. Using a product solution in Laplace's equation in cylindrical coordinates $V(r, \phi, z)=R(r) \Phi(\phi) Z(z)$, show that the term $R(r)$ satisfies the ordinary differential equation

$$
r^{2} \frac{d^{2} R}{d r^{2}}+r \frac{d R}{d r}+\left((\lambda r)^{2}-n^{2}\right) R=0
$$

## Problems 3

The separation constant is assumed to be $\alpha=n^{2}$ where $n$ is a positive integer and the constant $\lambda$ is real. The solutions for two of the dependent variables are $\Phi_{n}(\varphi)=\exp ( \pm \operatorname{jn} \varphi)$ and $Z_{n}(z)=\exp ( \pm \lambda z)$. The solution for this equation is $R_{n}(r)=J_{n}(\lambda r)$ and it can be obtained in terms of an infinite power series. Find the series expansion of the last function that is known as a Bessel function of a first kind, $n$-th order. Plot the first three functions $(n=0,1,2 ; \lambda=1)$ using MATLAB .
17. Using a product solution in Laplace's equation in spherical coordinates $\mathrm{V}(\rho, \theta, \varphi)=R(\rho) \Theta(\theta) \Phi(\phi)$, show that the term $R(\rho)$ satisfies the ordinary differential equation

$$
\rho^{2} \frac{d^{2} R}{d \rho^{2}}+2 \rho \frac{d R}{d \rho}-n(n+1) R=0
$$

The separation constant is assumed to be $\alpha=n(n+1)$ where $n \geq 0$ is an integer and the term $\Phi(\phi)$. - a constant (rotational symmetry). Find two particular solutions $R_{n}(\rho)$ of this equation.
$18^{\text {MATLAB }}$. Show that the corresponding ordinary differential equation for $\Theta(\theta)$ in problem 17 is

$$
\sin \theta \frac{d^{2} \Theta}{d \theta^{2}}+\cos \theta \frac{d \Theta}{d \theta}+n(n+1) \sin \theta \cdot \Theta=0
$$

Let $\cos \theta=\xi$ and find an expression of the solution $\Theta_{n}(\theta)=P_{n}(\xi)$ of this equation, which is a polynomial of n-th order that is known as a Legendre polynomial. Plot the first three functions ( $n=0,1,2$ ) using MATLAB.
19. Find particular solutions of Laplace's equation $\mathrm{V}_{\mathrm{n}}(\rho, \theta)$ that increases with the distance $\rho$ for the following three cases: a) $\mathrm{n}=0$ (charge); b) $\mathrm{n}=1$ (dipole); c) $\mathrm{n}=2$ (quadruple), using the results from the problems 17 and 18.
20. Compare the solution of Laplace's equation in cylindrical coordinates $f_{1}(r)=J_{0}(\lambda r)$ found in problem 16 for $n=0$ with the solution of Laplace's equation in Cartesian coordinates $f_{2}(r)=\cos (\lambda x) \cos (\lambda y)$ using a second order approximation in the power series for small values of the argument $(\lambda r<1)$.
21. Compute using an analytical integration, the potential $V(r)$ at the point $P$ that is a distance $r=1 \mathrm{~m}$ from the midpoint of a narrow finite strip that has a length $2 \mathrm{~L}=2 \mathrm{~m}$. A charge of 1 C is uniformly distributed on the
 strip.
22. Repeat problem 21 with a nonuniform charge distribution $\rho_{\mathrm{L}}(z)=\rho_{\mathrm{L} 0}(1-|z| / L)$ where $z=0$ is at the midpoint of the strip.
23. Repeat problem 21 but compute the electric field $E_{r}(r)$ instead of the potential.
24. Compute the potential $V(z)$ along the $z$ axis from the charged circular loop with a diameter 2a. Charge $Q$ is distributed uniformly upon the loop.
25. Compute the electric field $E_{z}(z)$ in the previous problem. Find the solution in the limiting case $a \rightarrow 0$.
26. Compute the potential $V(z)$ along the $z$ axis from the charged circular plate with a diameter 2a. Charge $Q$ is distributed uniformly upon the plate.
27. Compute the electric field $E_{z}(z)$ in the previous problem. Find the solution in the limiting case $\mathrm{a} \rightarrow \infty$.

## Problems 3

28. Find the potential distribution $V(z)$ between two parallel discs $(0<z<c)$ that have a large enough radius a such that the electric field is constrained to be entirely between them. This implies that the fringing fields are neglected. The boundary conditions are $\mathrm{V}(0)=0$ and $\mathrm{V}(\mathrm{c})=\mathrm{V}_{0}$. Find the normal electric field $\mathrm{E}_{\mathrm{Z}}$, the flux density $D_{z}$ and the surface charge density $\rho_{s}$ on both plates.

$29^{\text {MATLAB }}$. Find the potential distribution in the triangular region between the two charged surfaces using the FEM-method with $\mathrm{a}=1 \mathrm{~m}$ and $\quad V_{0}=16 \mathrm{~V}$. Display the potential profile.


## Problems 3

$30^{\text {MATLAB }}$. Find the surface charge density $\rho_{\mathrm{s}}$ in a square region a x a $(a=2 \mathrm{~m})$ between two charged surfaces with a potentials 0 and $V_{0}$ ( $\mathrm{V}_{0}=10 \mathrm{~V}$ ) using the MoM-method applied to the boundary integral equation. Then by simple integration find the potential in every internal point of the mesh. Display the potential profile.


## Faraday's law of induction

## 4. Time-varying electromagnetic fields

The subject of time-varying electromagnetic fields will be the central theme throughout the rest of this text. Here and in the following chapters, we will generalize to the time-varying case the static electric and magnetic fields that were reviewed in Chapter 2. In doing this, we must first appreciate the insight of the great nineteenth century theoretical physicist James Clerk Maxwell who was able to write down a set of equations that described electromagnetic fields. These equations have survived unblemished for approximately a century and a half of experimental and theoretical questioning and are now considered to be on an equal footing with the equations of Isaac Newton and many of the relativistic thoughts of Albert Einstein¹. We will concern ourselves here and now with what was uncovered and explained at the time of Maxwell's life.

### 4.1. Faraday's law of induction

The first time varying electromagnetic phenomena that is usually encountered in an introductory course dedicated to the study of electrical circuits is the determination of the electric potential across an inductor that is inserted in an electrical circuit. A simple circuit that exhibits this effect is shown in Figure 4-1.

The voltage across the inductor is expressed with the equation

$$
\begin{equation*}
V(t)=L \frac{d I(t)}{d t} \tag{4.1}
\end{equation*}
$$

where $L$ is the inductance, the units of which are Henries, $V(t)$ is the time-varying voltage across the inductor, and $l(t)$ is the time-varying current that passes through the inductor. The actual dependence that these quantities have on time will be determined by the voltage source. For example, a sinusoidal voltage

[^8]
## Faraday's law of induction

source in a circuit will cause the current in that circuit to have a sinusoidal time variation or a temporal pulse of current will be excited by a pulsed voltage source. In Chapter 2, we recognized that a time-independent current could create a timeindependent magnetic field and that a time-independent voltage was related to an electric field. These quantities are also related for time-varying cases as will be shown here.


Figure 4-1. A simple electrical circuit consisting of an ac voltage source $\mathrm{V}_{\mathrm{o}}(\mathrm{t})$, an inductor and a resistor. The voltage across the inductor is $\mathrm{V}(\mathrm{t})$.

The actual relation between the electric and the magnetic field components is computed from an experimentally verified effect that we now call Faraday's law. This is written as

$$
\begin{equation*}
V(t)=-\frac{d \Psi_{m}(t)}{d t} \tag{4.2}
\end{equation*}
$$

where $\Psi_{m}(\mathrm{t})$ is the total time-varying magnetic flux that passes through a surface. This law states that a voltage $\mathrm{V}(\mathrm{t})$ will be induced in a closed loop that completely surrounds the surface through which the magnetic field passes. The voltage $\mathrm{V}(\mathrm{t})$ that is induced in the loop is actually a voltage or potential difference $\mathrm{V}(\mathrm{t})$. This voltage exists between two points in the loop that are separated by an infinitesimal distance. The distance is so small that we can think of the loop as

## Faraday's law of induction

being actually closed. The polarity of the induced voltage will be such that it opposes the change of the magnetic flux, hence a minus sign appears in (4.2). This effect is also known as Lenz's law. The voltage can be computed from the line integral of the electric field between the two points.

A schematic representation of this effect is shown in Figure 4-2. Small loops as indicated in this figure and which have a cross-sectional area $\Delta \mathrm{s}$ are used to detect and plot the magnitude of time-varying magnetic fields in practice. We will assume that the loop is sufficiently small or that we can let it shrink in size so that it is possible to approximate $\Delta s$ with the differential surface area |ds|. The vector direction associated with ds is normal to the plane containing the differential surface area. If the stationary orientation of the loop ds is perpendicular to the magnetic flux density $\mathbf{B}(\mathrm{t})$, zero magnetic flux will be captured by the loop and $\mathrm{V}(\mathrm{t})$ will be zero. By rotating this loop about a known axis, it is also possible to ascertain the vector direction of $\mathrm{B}(\mathrm{t})$ by correlating the maximum detected voltage $\mathrm{V}(\mathrm{t})$ with respect to the orientation of the loop. Recall from our discussion of magnetic circuits that we used the total magnetic flux $\Psi_{m}$. This is formally written in terms of an integral. In particular, the magnetic flux that passes through the loop is given by


Figure 4-2. A loop through which a time-varying magnetic field passes.

$$
\begin{equation*}
\Psi_{\mathrm{m}}(\mathrm{t})=\int_{\Delta \mathrm{s}} \mathrm{~B} \cdot \mathrm{ds} \tag{4.3}
\end{equation*}
$$

As we will see later, either the magnetic flux density or the surface area could be changing in time. The scalar product reflects the effects arising from an arbitrary orientation of the loop with respect to the orientation of the magnetic flux density. It is important to realize that the loops that we are considering may not be wire loops. The loops could just be closed paths.

Example 4-1. Let a stationary square loop of wire lie in the xy plane that contains a spatially homogeneous time-varying magnetic field.

$$
\mathbf{B}(\mathbf{r}, \mathrm{t})=\mathbf{u}_{\mathbf{z}} \mathrm{B}_{0} \sin \omega \mathrm{t} .
$$

Find the voltage $\mathrm{V}(\mathrm{t})$ that could be detected between the two terminals that are separated by an infinitesimal distance.


Answer: The magnetic flux that is enclosed within the loop is given by

$$
\Psi_{\mathrm{m}}=\int_{\Delta s} B \bullet d s=\int_{y=0}^{b} \int_{\mathrm{x}=0}^{\mathrm{b}}\left(\mathrm{~B}_{0} \sin \omega t \mathbf{u}_{\mathrm{z}}\right) \bullet\left(\mathrm{dxdy} \mathbf{u}_{z}\right)=\mathrm{B}_{0} \mathrm{~b}^{2} \sin \omega t .
$$

The induced voltage in the loop is given from Faraday's law (4.2) to be

$$
\mathrm{V}(\mathrm{t})=-\frac{\mathrm{d} \Psi_{\mathrm{m}}(\mathrm{t})}{\mathrm{dt}}=-\omega \mathrm{B}_{0} \mathrm{~b}^{2} \cos \omega \mathrm{t}
$$

Example 4-2. Let a stationary loop of wire lie in the xy plane that contains a spatially inhomogeneous time-varying magnetic field.

# Faraday's law of induction 

$$
\mathbf{B}(\mathrm{r}, \mathrm{t})=\mathbf{u}_{\mathrm{z}} \mathrm{~B}_{0} \cos \left(\frac{\pi \mathrm{r}}{2 \mathrm{~b}}\right) \cos \omega \mathrm{t} .
$$

where the amplitude of the magnetic flux density is $B_{0}=2 T$, the radius of the loop is $b=0.5 \mathrm{~m}$ and the angular frequency of oscillation of the time-varying magnetic field is $\omega=2 \pi \mathrm{f}=314 \mathrm{~s}^{-1}$. The center of the loop is at the point $\mathrm{r}=0$. Find the voltage $\mathrm{V}(\mathrm{t})$ that could be detected between the two terminals that are separated by an infinitesimal distance.


Answer: The magnetic flux $\Psi_{m}$ that is enclosed within the "closed" loop is given by

$$
\begin{aligned}
\Psi_{\mathrm{m}} & =\int_{\Delta \mathrm{s}} \mathbf{B} \cdot \mathbf{d s} \\
& =\int_{\phi=0}^{2 \pi} \int_{\mathrm{r}=0}^{\mathrm{b}}\left[\mathrm{~B}_{0} \cos \left(\frac{\pi \mathrm{r}}{2 \mathrm{~b}}\right) \cos \omega \mathbf{t \mathbf { u } _ { \mathbf { z } }}\right] \bullet\left[\mathrm{rdrd} \phi \mathbf{u}_{\mathbf{z}}\right]
\end{aligned}
$$

The integral over $\phi$ yields a factor of $2 \pi$ while the integral over $r$ is solved via integration by parts that results in

$$
\Psi_{m}=\left(B_{0} \cos \omega t\right)(2 \pi)\left[\frac{4 b^{2}}{\pi^{2}}\left(\frac{\pi}{2}-1\right)\right]=\frac{8 b^{2}}{\pi}\left(\frac{\pi}{2}-1\right) \mathrm{B}_{0} \cos \omega \mathrm{t} .
$$

The induced voltage in the loop is given from Faraday's law (4.2) to be

$$
\mathrm{V}(\mathrm{t})=-\frac{\mathrm{d} \Psi_{\mathrm{m}}(\mathrm{t})}{\mathrm{dt}}=\frac{8 \mathrm{~b}^{2}}{\pi}\left(\frac{\pi}{2}-1\right) \mathrm{B}_{0} \omega \sin \omega \mathrm{t}=228.2 \sin (314 \mathrm{t}) \text { (volts). }
$$

Example 4-3. Another application of Faraday's law is the explanation of a transformer. Find the voltage that is induced in side 2 if a time-varying voltage $\mathrm{V}_{1}$ is connected to side 1.


Answer: Due to the use of an iron core, we can assume that the magnetic field is constrained to exist only in the core region. The voltage $\mathrm{V}_{2}$ is given by

$$
\mathrm{V}_{2}=-\mathrm{N}_{2} \frac{\mathrm{~d} \Psi_{\mathrm{m}}}{\mathrm{dt}}
$$

where $\Psi_{m}$ is the total magnetic flux that passes through the $N_{2}$ turns of region \#2. If the resistance of region \#1 is very small and can be neglected, the induced voltage is equal to the primary voltage $\mathrm{V}_{1}$. Hence

$$
\mathrm{V}_{1}=-\mathrm{N}_{1} \frac{\mathrm{~d} \Psi_{\mathrm{m}}}{\mathrm{dt}}
$$

Dividing the first equation by the second leads to

$$
\frac{V_{2}}{V_{1}}=\frac{N_{2}}{N_{1}} .
$$

This is also an application of a magnetic circuit that was discussed in Chapter 2.

Since the two terminals in Figure 4-2 are separated by a very small distance, we will be permitted to assume that they actually are touching, at least in a mathematical sense even though they must be physically separated. This will

## Faraday's law of induction

allow us to consider the loop to be a closed loop in various integrals that follow but still permit us to detect a potential difference between the two terminals. The magnetic flux $\Psi_{m}(\mathrm{t})$ can be written in terms of the magnetic flux density $\mathbf{B}(\mathbf{r}, \mathrm{t})=\mathbf{B}$ and the voltage $\mathrm{V}(\mathrm{t})$ can be written in terms of the electric field $E(r, t)=E$. This yields the result

$$
\begin{equation*}
\oint E \cdot d l=-\frac{d}{d t} \int_{\Delta s} B \bullet d s \tag{4.4}
\end{equation*}
$$

It is worth emphasizing the point that this electric field is the component of the electric field that is tangential to the loop since this is critical in our argument. In addition, (4.4) includes several possible mechanisms in which the magnetic flux could change in time. Either the magnetic flux density changes in time, the cross-sectional area changes in time, or there is a combination of the two mechanisms. These will be described below.

Although both the electric and magnetic fields depend on space and time, we will not explicitly state this fact in every equation that follows. This will conserve time, space and energy if we now define and later understand that $\mathbf{E} \equiv \mathbf{E}(r, t)$ and $\mathbf{B} \equiv \mathbf{B}(r, t)$ in the equations. This short hand notation will also allow us to more easily remember the important results in the following material. In this notation, the independent spatial variable $\mathbf{r}$ refers to a three dimensional position vector where

$$
\begin{equation*}
\mathbf{r}=x \mathbf{u}_{\mathbf{x}}+y \mathbf{u}_{\mathbf{Y}}+\mathrm{z} \mathbf{u}_{\mathbf{z}} \tag{4.5}
\end{equation*}
$$

in Cartesian coordinates. The independent variable $t$ refers to time. We must keep this notation in our mind in the material that follows.

The closed line integral appearing in (4.4) can be converted into a surface integral via Stokes's theorem. We obtain for the left side of (4.4)

$$
\begin{equation*}
\oint \mathbf{E} \cdot \mathbf{d} \mathbf{l}=\int_{\Delta s} \nabla \times \mathbf{E} \cdot \mathbf{d s} \tag{4.6}
\end{equation*}
$$

## Faraday's law of induction

Let us initially assume that the surface area of the loop does not change in time. This implies that $\Delta \mathrm{s}$ is a constant. In this case, the time derivative can then be brought inside the integral.

$$
\begin{equation*}
\int_{\Delta s} \nabla \times \mathbf{E} \cdot \mathbf{d s}=-\int_{\Delta s} \frac{\partial \mathbf{B}}{\partial \mathrm{t}} \cdot \mathbf{d s} \tag{4.7}
\end{equation*}
$$

The two integrals will be equal over any arbitrary surface area if and only if the two integrands are equal. This means that

$$
\begin{equation*}
\nabla \times \mathbf{E}=-\frac{\partial \mathbf{B}}{\partial \mathrm{t}} \tag{4.8}
\end{equation*}
$$

Either the integral representation (4.4) or the differential representation (4.8) are equally valid in describing the physical effects that are included in Faraday's law. These equations are also called Faraday's law of induction in honor of their discoverer. Faraday stated the induction's law in 1831 after making the assumption that a new phenomenon called an electromagnetic field would surround every electric charge.

Example 4-4. A small rectangular loop of wire is placed next to a time varying current $\mathrm{I}(\mathrm{t})$ carrying infinite wire. Calculate the current $\mathrm{i}(\mathrm{t})$ that flows in the loop if the conductivity of the wire is $\sigma$. To simplify the calculation, you may neglect the magnetic field created by the current $i(t)$ that passes through the wire of the loop.


## Faraday's law of induction

Answer: From the left-hand side of (4.4), we write the induced current $\mathrm{i}(\mathrm{t})$ as

$$
\oint E \cdot d \mathbf{d}=\oint \frac{\mathrm{J} \cdot \mathrm{dl}}{\sigma}=\frac{i(2 \mathrm{D}+2 \mathrm{~b})}{\sigma \mathrm{A}}=\mathrm{iR}
$$

where $A$ is the cross-sectional area of the wire and $R$ is the resistance of the wire. The current density is assumed to be constant over the cross-section of the wire. The magnetic flux density of the infinite wire is found by Ampere's law (2.114)

$$
B(t)=\frac{\mu_{0} I(t)}{2 \pi r}
$$

The right-hand side of (4.4) can be written as

$$
\begin{aligned}
-\frac{d}{d t} \int_{\Delta s} B \cdot d s & =-\frac{d}{d t}\left[\int_{z=0}^{D} \int_{r=a}^{a+b} \frac{\mu_{0} l(t)}{2 \pi r} d r d z\right] \\
& =-\frac{d}{d t}\left[D \frac{\mu_{0}(t)}{2 \pi} \ln \left(\frac{b+a}{a}\right)\right]=-\left[\frac{\mu_{0} D}{2 \pi} \ln \left(\frac{b+a}{a}\right)\right] \frac{d l(t)}{d t}
\end{aligned}
$$

Hence, the current $i(t)$ that is induced in the loop is given by

$$
i(t)=-\frac{1}{R}\left[\frac{\mu_{0} D}{2 \pi} \ln \left(\frac{b+a}{a}\right)\right] \frac{d(t)}{d t}
$$

The term in the brackets corresponds to a term that is called the mutual inductance $M$ between the wire and the loop. If we had included the effects of the magnetic field created by the current $\mathrm{i}(\mathrm{t})$ in the loop, an additional term proportional to $\mathrm{di}(\mathrm{t}) / \mathrm{dt}$ would appear in the equation. In this case, a term called the self inductance $L$ of the loop would be found. Hence a differential equation for $i(t)$ would have to be solved in order to incorporate the effects of the self inductance of the loop.

It is also important to calculate the mutual inductance between two coils. This can be accomplished by assuming that a time varying current in one of the

## Faraday's law of induction

coils would produce a time varying magnetic field in the region of the second coil. This will be demonstrated with an example.

Example 4-5 ${ }^{\text {MATLAB }}$. Find the normalized mutual inductance $M / \mu_{0}$ of the system consisting of two similar parallel wire loops with a radius $R=50 \mathrm{~cm}$ if their centers are separated by a distance $h=2 R$ (Helmholtz's coils).


Answer: As shown in Example 2-39, the magnetic flux density can be replaced with the magnetic vector potential. Using $\mathbf{B}=\nabla \times \mathbf{A}$ (2.123) and Stokes's theorem (4.6), we can reduce the surface integral for the magnetic flux to the following line integral

$$
\Psi_{\mathrm{m} 12}=\int_{\Delta s_{2}} \nabla \times \mathbf{A} \cdot \mathbf{d} \mathbf{s}_{2}=\oint_{\mathrm{L}_{2}} \mathbf{A} \cdot \mathbf{d l}_{2} .
$$

The magnetic vector potential is a solution of the vector Possion's equation (2.130). With the assumption that the coil has a small cross-section, the volume integral can be converted to the following line integral

$$
\mathbf{A}=\frac{\mu_{0} I_{1}}{4 \pi} \oint_{L_{1}} \frac{\mathbf{d l}_{1}}{R_{12}}
$$

where $R_{12}$ is a distance between the source (1) and the observer (2). The definition of the mutual inductance

# Faraday's law of induction 

$$
\mathrm{M}=\frac{\Psi_{\mathrm{m} 12}}{\mathrm{I}_{1}}
$$

yields the following double integral for the normalized mutual inductance

$$
\frac{\mathrm{M}}{\mu_{0}}=\frac{1}{4 \pi} \oint_{L_{2}} \oint_{L_{1}} \frac{\mathbf{d l}_{1} \bullet \mathrm{dl}_{2}}{R_{12}}
$$

For the case of two identical parallel loops, we obtain after performing the integration

$$
\frac{\mathrm{M}}{\mu_{0}}=\frac{\mathrm{R}^{2}}{2} \int_{0}^{2 \pi} \frac{\cos \varphi \mathrm{~d} \varphi}{\sqrt{\mathrm{~h}^{2}+2 R^{2}-2 R^{2} \cos \varphi}}
$$

For the special case ( $h=2 R$ ), this single integral can be further simplified

$$
\frac{M}{\mu_{0}}=R \cdot\left(\frac{1}{2 \sqrt{2}} \int_{0}^{2 \pi} \frac{\cos \varphi d \varphi}{\sqrt{3-\cos \varphi}}\right)=R \cdot C
$$

This shows that the mutual inductance increases linearly with the increasing of the wire radius $R$. The integral for the constant $C$ can be solved numerically by applying the MATLAB function 'quad'. Assuming a radius $\mathrm{R}=0.5$ m and a constant $\mathrm{C}=0.1129$, the normalized mutual conductance is computed to be $M / \mu_{0}=0.0564$.

In the derivation of (4.7), an assumption was made that the area $\Delta \mathrm{s}$ of the loop did not change in time and only a time-varying magnetic field existed in space. This assumption need not always be made in order for electric fields to be generated by magnetic fields. We have to be thankful for the fact that the effect can be generalized since much of the conversion of electric energy to mechanical energy or mechanical energy to electrical energy is based on this phenomenon.

## Faraday's law of induction



Figure 4-3. A conducting bar moving in a uniform time-independent magnetic field that is directed out of the paper. Charge distributions of the opposite sign appear at the two ends of the bar.

In particular, let us assume that a conducting bar moves with a velocity $\mathbf{v}$ through a uniform time-independent magnetic field $\mathbf{B}$ as shown in Figure 4-3. The wires that are connected to this bar are parallel to the magnetic field and are connected to a volt meter that lies far beneath the plane of the moving bar. From the Lorentz force equation (2.116), we can calculate the force $\mathbf{F}$ on the freely mobile charged particles in the conductor. Hence, one end of the bar will become positively charged and the other end will have an excess of negative charge.

Since there is a charge separation in the bar, there will be an electric field that is created in the bar. Since the net force on the bar is equal to 0 , the electric and magnetic contributions to the force cancel. This results in an electric field that is

$$
\begin{equation*}
E \equiv \frac{F}{q}=\mathbf{v} \times \mathbf{B} \tag{4.9}
\end{equation*}
$$

This electric field can be interpreted to be an induced field acting in the direction along the conductor that produces a voltage V and it is given by

$$
\begin{equation*}
V=\int_{a}^{b}(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{d} \mid . \tag{4.10}
\end{equation*}
$$

## Faraday's law of induction

Example 4-6. A Faraday disc generator consists of a circular metal disc rotating with a constant angular velocity $\omega=600 \mathrm{~s}^{-1}$ in a uniform time-independent magnetic field. A magnetic flux density $B=B_{0} \mathbf{u}_{\mathbf{z}}$ where $B_{0}=4 T$ is parallel to the axis of rotation of the disc. Determine the induced open-circuit voltage that is generated between the brush contacts that are located at the axis and the edge of the disc whose radius is $\mathrm{a}=0.5 \mathrm{~m}$.

$\omega$
Answer: An electron at a radius $r$ from the center has a velocity $\omega r$ and therefore experiences an outward directed radial force $-\mathrm{q} \omega \mathrm{rB}_{0}$. The Lorentz force acting on the electron is

$$
-\mathrm{q}[\mathbf{E}+(\mathbf{v} \times \mathbf{B})]=0 .
$$

At equilibrium, we find that the electric field can be determined from the Lorentz force equation to be directed radially inward and have a magnitude $\omega \mathrm{rB}_{0}$. Hence we write

$$
\mathrm{V}=-\int_{1}^{2}(\mathbf{v} \times \mathbf{B}) \cdot d \mathbf{d}=-\int_{0}^{a}\left[\left(\omega r \mathbf{u}_{\varphi}\right) \times \mathrm{B}_{0} \mathbf{u}_{\mathrm{z}}\right] \cdot \mathrm{dr} \mathbf{u}_{\mathrm{r}}=\omega \mathrm{B}_{0} \int_{\mathrm{a}}^{0} \mathrm{rdr}=-\frac{\omega \mathrm{B}_{0} \mathrm{a}^{2}}{2}=-300 \text { volts }
$$

which is the generated potential due to this machine that is called a Faraday disc generator.

## Faraday's law of induction

If the bar depicted in Figure 4-3 were moving through a time-dependent magnetic field instead of a constant magnetic field, then we would have to add together the potential caused by the motion of the bar and the potential caused by the time-varying magnetic field. This implies that the principle of superposition applies for this case. This is a good assumption in a vacuum or in any linear medium.

Example 4-7. A rectangular loop rotates through a time-varying magnetic flux density $\mathbf{B}=B_{0} \cos \omega t u_{Y}$. The loop rotates with the same angular frequency $\omega$. Calculate the induced voltage at the terminals.

Answer: Due to the rotation of the loop, there will be two components to the induced voltage. The first is due to the motion of the loop and the second is due to the time-varying magnetic field.

The voltage due to the rotation of the loop is calculated from

$$
\mathrm{V}_{\text {rot }}=\int_{\mathrm{b} / 2}^{-\mathrm{b} / 2} \mathbf{v} \times\left.\mathbf{B} \cdot \mathbf{d} \mathbf{l}\right|_{\text {bottom_edge }}+\int_{-\mathrm{b} / 2}^{\mathrm{b} / 2} \mathbf{v} \times\left.\mathbf{B} \cdot \mathbf{d}\right|_{\text {top_edge }}
$$

The contributions from either end will yield zero. We write

$$
\begin{aligned}
V_{\text {rot }}= & \int_{b / 2}^{-b / 2} \mathrm{v} B_{0} \cos \omega t \sin \theta\left(-\mathbf{u}_{x}\right) \bullet\left(\mathbf{u}_{x} d x\right) \\
& +\int_{-b / 2}^{b / 2} \mathrm{v} B_{0} \cos \omega t \sin \theta\left(\mathbf{u}_{x}\right) \bullet\left(\mathbf{u}_{x} d x\right)
\end{aligned}
$$

The angle $\theta=\omega$ and the velocity $\mathrm{v}=\omega \mathrm{a}$. Hence the term of the induced voltage due to the rotation of the loop yields

$$
V_{\text {rot }}=\omega B_{0} b a \sin 2 \omega t
$$

We recognize that the area of the loop is equal to $\Delta s=2 a b$.

# Faraday's law of induction 



From (4.4), we can compute the voltage due to the time-varying magnetic field. In this case, we note that

$$
\mathbf{d s}=\left[\mathbf{u}_{\mathbf{Y}} \cos \omega t+\mathbf{u}_{\mathbf{Z}} \sin \omega t\right] d x d z
$$

Therefore the voltage due to time variation of the magnetic field is

$$
\begin{aligned}
V_{v a r} & =-\int_{\Delta s} \frac{\partial \mathbf{B}}{\partial t} \bullet d \mathbf{d s} \\
& =\omega B_{0} \int_{x=-b / 2 z=-a}^{b / 2} \int^{a} \sin \omega t \mathbf{u}_{Y} \bullet\left[\mathbf{u}_{Y} \cos \omega t+\mathbf{u}_{Z} \sin \omega t\right] d x d z
\end{aligned}
$$

The integration leads to

$$
V_{\mathrm{var}}=\omega \mathrm{B}_{0} \text { ba } \sin 2 \omega \mathrm{t} .
$$

The total voltage is given by the sum of that due to rotation and that due to the time variation of the magnetic field

$$
V=V_{\text {rot }}+V_{\text {var }}=2 \omega B_{0} \text { basin } 2 \omega t .
$$

This results in the generation of the second harmonic.

We can apply a repeated vector operation to Faraday's law of induction that is given in (4.8) to obtain an equation that describes another feature of time

## Faraday's law of induction

varying magnetic fields. We take the divergence of both sides of (4.8) and interchange the order of differentiation to obtain

$$
\begin{equation*}
\nabla \bullet \nabla \times \mathbf{E}=-\nabla \bullet \frac{\partial \mathbf{B}}{\partial \mathrm{t}}=-\frac{\partial}{\partial \mathrm{t}}(\nabla \bullet \mathbf{B}) \tag{4.11}
\end{equation*}
$$

The first term $\nabla \bullet \nabla \times \mathbf{E}$ is equal to zero since the divergence of the curl of a vector is equal to zero by definition. This follows also from Figure 4-2 where the electric field is constrained to follow the loop since that is the only component that survives the scalar product of $\mathbf{E} \bullet$ dl. The electric field can neither enter nor leave the loop which would be indicative of a nonzero divergence. Nature is kind to us in that it frequently lets us interchange the orders of differentiation without inciting any mathematical complications. This is a case where it can be done. Hence for any arbitrary time dependence, we again find that

$$
\begin{equation*}
\nabla \bullet \mathbf{B}=0 \tag{4.12}
\end{equation*}
$$

This statement that is valid for time-dependent cases is the same result that was given in Chapter 2 as a postulate for static fields. It also continues to reflect the fact that we have not found magnetic monopoles in nature.

Let us integrate (4.12) over an arbitrary volume $\Delta \mathrm{v}$. This volume integral can be converted to a closed surface integral using the divergence theorem

$$
\begin{equation*}
\int_{\Delta V}(\nabla \cdot \mathbf{B}) \mathrm{dv}=\oint \mathbf{B} \cdot \mathbf{d s} \tag{4.13}
\end{equation*}
$$

from which we write

$$
\begin{equation*}
\oint B \cdot d s=0 \tag{4.14}
\end{equation*}
$$

Equation (4.14) is also valid for time-independent electromagnetic fields.

## Equation of continuity

### 4.2. Equation of continuity

Before obtaining the next equation of electromagnetics, it is useful to step back and derive the equation of continuity. In addition, we must understand the ramifications of this equation. The equation of continuity is fundamental at this point in developing the basic ideas of electromagnetic theory. Its derivation is fairly simple since all of the hard work has already been accomplished in Chapter 1 where we reviewed some vectors. It can also be applied in several other areas of engineering and science so the time that is spent here will be time well spent.


Figure 4-4. Charges centered at $x=0$ at time $t<0$ as shown in (a) are allowed to expand at $t=0$. As time increases (b) some of these charges may pass through the screens at $x= \pm \frac{\Delta x}{2}$.

In order to derive the equation of continuity, let us consider a model that assumes a stationary number of positive charges are initially located at the center of a transparent box whose volume is $\Delta v=\Delta x \Delta y \Delta z$. These charges are at the prescribed positions within the box for times $t \leq 0$. A one-dimensional view of this box consists of two parallel planes and it is shown in Figure 4-4a. We will neglect the Coulomb's forces between the individual charges. If we are uncomfortable

## Equation of continuity

with this assumption of noninteracting charged particles, we could have alternatively assumed that the charges were just noninteracting gas molecules or billiard balls and derived a similar equation for these objects. The resulting equation could then be multiplied by a charge $q$ that would be impressed on an individual entity. For times $t \geq 0$, the particles or the charges can start to move and actually leave through the two screens as time increases. This is depicted in Figure $4-4 \mathrm{~b}$. The magnitude of the cross-sectional area of a screen is equal to $\Delta \mathrm{s}$ $=\Delta y \Delta z$. The charges that leave the "screened in region" will be in motion. Hence, those charges that leave the box from either side will constitute a current that emanates from the box. Due to our choice of charges within the box having a positive charge, the direction of this current I will be in the same direction as the motion of the charge.

Rather than just examine the small number of charges depicted in Figure $4-4$, let us assume that there now are a large number of them. We still will neglect the Coulomb force between charges. The number of charges will be large enough so it is prudent to describe the charge within the box with a charge density $\rho_{\mathrm{v}}$ where $\rho_{\mathrm{v}}=\Delta \mathrm{Q} / \Delta \mathrm{v}$ and $\Delta \mathrm{Q}$ is the total charge within a volume $\Delta \mathrm{v}$. Hence the decrease of the charge density $\rho_{v}$ acts as a source for the total current I that leaves the box as shown in Figure 4-5.


Figure 4-5. Charge within the box leaves through the walls.

A temporal decrease of charge density within the box implies that charge leaves the box since the charge is neither destroyed nor does it recombine with charge of the opposite sign. The total current that leaves the box through any portion of the surface is due to a decrease of the charge within the box. It is found from the expression These are real charges and the current that we are describing is not the displacement current that we will encounter later

$$
\begin{equation*}
\mathrm{I}=-\frac{\mathrm{dQ}}{\mathrm{dt}} \tag{4.15}
\end{equation*}
$$

This can be rewritten as

$$
\begin{equation*}
\oint \mathbf{J} \cdot \mathbf{d} \mathbf{s}=-\int_{\Delta v} \frac{\partial \rho_{v}}{\partial t} d v \tag{4.16}
\end{equation*}
$$

where we have taken the liberty of summing up the six currents that leave the six sides of the box that surrounds the charges; this summation is expressed as a closed-surface integral. The closed-surface integral given in (4.16) can be converted to a volume integral using the divergence theorem. Hence equation (4.16) can be written as

$$
\begin{equation*}
\int_{\Delta v}(\nabla \cdot J) d v=-\int_{\Delta v} \frac{\partial \rho_{v}}{\partial t} d v \tag{4.17}
\end{equation*}
$$

## Equation of continuity

Since this equation must be valid for any arbitrary volume, we are left with the conclusion that the two integrands must be equal from which we write

$$
\begin{equation*}
\nabla \cdot \mathbf{J}+\frac{\partial \rho_{v}}{\partial \mathrm{t}}=0 \tag{4.18}
\end{equation*}
$$

Equation (4.18) is the equation of continuity that we are seeking. Note that this equation has been derived using very simple common sense arguments. Since these arguments are straightforward and logical, we can believe the result as being a fundamental truth of nature that must be valid for all cases and for all times. Although we have derived it using finite sized volumes, the equation is valid at a point. Its importance will be noted in the next section where we will follow in the footsteps of James Clerk Maxwell.

We recall from our first course that dealt with circuits that the Kirchhoff's current law stated that the net current entering or leaving a node was equal to zero. Charge is neither created nor destroyed in this case. This is shown in Figure 4-6. The dashed lines represent a closed surface that surrounds the node. The picture shown in Figure 4-5 generalizes this node to three dimensions.


Figure 4-6. A closed surface (dashed lines) surrounding a node.

Example 4-8. Charges are introduced into the interior of a conductor during the time $\mathrm{t}<0$. Calculate how long it will take for these charges to move to the surface
of the conductor so the interior charge density $\rho_{v}=0$ and interior electric field $\mathbf{E}=$ 0.

Answer: Introduce Ohm's law $\mathbf{J}=\sigma \mathbf{E}$ into the equation of continuity

$$
\sigma \nabla \cdot \mathbf{E}=-\frac{\partial \rho_{v}}{\partial \mathrm{t}}
$$

The electric field is related to the charge density through Poisson's equation

$$
\nabla \bullet E=\frac{\rho_{v}}{\varepsilon}
$$

Hence, we obtain

$$
\frac{d \rho_{v}}{d t}+\frac{\sigma}{\varepsilon} \rho_{v}=0
$$

whose solution is

$$
\rho_{\mathrm{v}}=\rho_{\mathrm{v} 0} \mathrm{e}^{-\left(\frac{\sigma}{\varepsilon} \mathrm{t}\right)} .
$$

The initial charge density $\rho_{\mathrm{v} 0}$ will decay to [1/ $\mathrm{e} \approx 37 \%$ ] of its initial value in a time $\tau=\varepsilon / \sigma$ which is called the relaxation time. For copper, this time is

$$
\tau=\frac{\varepsilon_{0}}{\sigma}=\frac{\frac{1}{36 \pi} \times 10^{-9}}{5.8 \times 10^{7}}=1.5 \times 10^{-19} \mathrm{sec}
$$

Other effects that are not described here may cause this time to be different. Relaxation times for insulators may be hours or days.

Example 4-9. The current density is $\mathbf{J}=\exp \left(-x^{2}\right) \mathbf{u}_{\mathbf{x}}$. Find the time rate of increase of the charge density at $x=1$.

Answer: From the equation of continuity (4.18), we write

$$
\frac{d \rho_{v}}{d t}=-\nabla \cdot \mathbf{J} \Rightarrow \frac{d \rho_{v}}{d t}=-\frac{d J_{x}}{d x}=\left.2 x \exp \left(-x^{2}\right)\right|_{x=1}=.736
$$

Example 4-10. The current density in a certain region may be approximated with the function

## Equation of continuity

$$
\mathbf{J}=J_{0} \frac{e^{-t / \tau}}{\rho} \mathbf{u}_{\rho}
$$

in spherical coordinates. Find the total current that leaves a spherical surface whose radius is $a$ at the time $t=\tau$. Using the equation of continuity, find an expression for the charge density $\rho_{\mathrm{V}}(\rho, \mathrm{t})$.

Answer: The total current that leaves the spherical surface is given by

$$
\mathrm{I}=\left.\oint \mathrm{J} \cdot \mathrm{ds}\right|_{\rho=\mathrm{a}, \mathrm{t}=\tau}=\left.4 \pi \mathrm{a}^{2}\left(\frac{\mathrm{~J}_{0} \mathrm{e}^{-\mathrm{t} / \tau}}{\mathrm{a}}\right)\right|_{\mathrm{t}=\tau}=4 \pi \mathrm{aJ}_{0} \mathrm{e}^{-1} .
$$

In spherical coordinates, the equation of continuity that depends only upon the radius $\rho$ is written as

$$
\frac{\partial \rho_{v}}{\partial t}=-\frac{1}{\rho^{2}} \frac{\partial}{\partial \rho}\left(\rho^{2} J_{\rho}\right)=-\frac{1}{\rho^{2}} \frac{\partial}{\partial \rho}\left(\rho^{2} J_{0} \frac{e^{-t / \tau}}{\rho}\right)=-J_{0} \frac{e^{-t / \tau}}{\rho^{2}}
$$

Hence, after integration, the charge density is given by

$$
\rho_{v}=\int \frac{\partial \rho_{v}}{\partial t} d t \Rightarrow \rho_{v}=J_{0} \frac{\tau e^{-t / \tau}}{\rho^{2}}
$$

iswhere the arbitrary constant of integration is set equal to zero.

## Displacement current

### 4.3. Displacement current

Our first encounter with time-varying electromagnetic fields yielded Faraday's law of induction (4.2). The next encounter will illustrate the genius of James Clerk Maxwell. Through his efforts in the nineteenth century, we are now able to answer a fundamental question that would arise when analyzing a circuit in the following gedanken experiment. Let us connect two wires to the two plates of an ideal capacitor consisting of two parallel plates separated by a vacuum and an ac voltage source as shown in Figure 4-7. An ac ammeter is also connected in series with the wires in this circuit and it measures a constant value of ac current I. Two questions might enter our mind at this point:
(1) "How can the ammeter read any value of current since the capacitor is an open circuit and the current which passes through the wire would be impeded by the vacuum that exists between the plates?"
(2) "What happens to the time-varying magnetic field that is created by the current and surrounds the wire as we pass through the region between the capacitor plates?"


Figure 4-7. An elementary circuit consisting of an ideal parallel plate capacitor connected to an ac voltage source and an ac ammeter.

The answer to the first question will require that we first reexamine the equations that we have obtained up to this point and then interpret them, guided by the light that has been turned on by Maxwell. In particular, let us write the

## Displacement current

second postulate of steady magnetic fields - Ampere's law (2.113). This postulate stated that a magnetic field $\mathbf{B}$ was created by a current $\mathbf{J}$ that is rewritten here

$$
\begin{equation*}
\nabla \times \mathbf{B}=\mu_{0} \mathbf{J} \tag{4.19}
\end{equation*}
$$

Let us take the divergence of both sides of this equation. The term on the lefthand side

$$
\begin{equation*}
\nabla \cdot \nabla \times \mathbf{B}=0 \tag{4.20}
\end{equation*}
$$

by definition. Applying the divergence operation to the term on the right-hand of (4.19), we find that

$$
\begin{equation*}
\mu_{0} \nabla \bullet \mathbf{J}=0 \tag{4.21}
\end{equation*}
$$

This, however, is not compatible the equation of continuity (4.18) that we have just ascertained to be a truth of nature.

To get out of this dilemma, Maxwell postulated the existence of another type of current in nature. This current would be in addition to the conduction current discussed in Chapter 2 and a convection current that would be created by charge passing through space with a constant drift velocity. The new current with a density $\mathbf{J}_{\mathbf{d}}$ is called a displacement current and it is found by incorporating the equation for the displacement flux density $\mathbf{D}$ into the equation of continuity by use of Gauss's law (2.89). Hence

$$
\begin{gather*}
\nabla \bullet \mathbf{J}+\frac{\partial \rho_{v}}{\partial \mathrm{t}}=0 \\
\nabla \bullet \mathbf{J}+\frac{\partial(\nabla \bullet \mathbf{D})}{\partial \mathrm{t}}=\nabla \cdot\left(\mathbf{J}+\frac{\partial \mathbf{D}}{\partial \mathrm{t}}\right)=0 \tag{4.22}
\end{gather*}
$$

where we have freely interchanged the order of differentiation. The displacement current density is identified as

$$
\begin{equation*}
\mathbf{J}_{\mathbf{d}}=\frac{\partial \mathbf{D}}{\partial \mathrm{t}} \tag{4.23}
\end{equation*}
$$

This is the current that passes between the two plates of the capacitor in our

## Displacement current

gedanken experiment that was performed at the beginning of this section.
The time-varying conduction current that passes through the wire causes a build-up of charges of the opposite signs on the two plates of the capacitor. The time-variation of these charges creates a time-varying electric field between the plates ${ }^{1}$. The time-varying displacement current will pass from one plate to the other and an answer to the first question has been obtained. The conduction current in the wire becomes a displacement current between the plates. This displacement current does not exist in a time-independent system.

The postulate for magnetostatics (2.113) will have to be modified to incorporate this new current and any possible time-varying magnetic fields. It becomes

$$
\begin{equation*}
\nabla \times \frac{\mathbf{B}}{\mu_{0}}=\mathbf{J}+\frac{\partial \mathbf{D}}{\partial \mathrm{t}} \tag{4.24}
\end{equation*}
$$

We can also answer the second question. With the inclusion of the displacement current which passes between the capacitor plates, we can assert that the time-varying magnetic field which surrounds the conduction current-carrying wire will be equal in magnitude and direction to the time-varying magnetic field which surrounds the capacitor.

Let us integrate both sides of (4.24) over the cross-sectional area specified by the radius $r$ at two locations in Figure 4-8. The first integral will be at a location surrounding the wire and the second will be between the two capacitor plates.

$$
\begin{equation*}
\int_{\Delta s}\left(\nabla \times \frac{\mathbf{B}}{\mu_{0}}\right) \cdot \mathbf{d} \mathbf{s}=\int_{\Delta s}\left(\mathbf{J}+\frac{\partial \mathbf{D}}{\partial \mathbf{t}}\right) \cdot \mathbf{d s} \tag{4.25}
\end{equation*}
$$

or using Stokes's theorem, we write

[^9]
## Displacement current



Figure 4-8. Two parallel plates in a capacitor separate two wires. The circle whose radius is $r$ could be surrounding the wire (\#1) at either edge or between the plates (\#2). The radius of the wire is a and the plate is $b$.

The left hand side of the integral in (4.26) yields

$$
2 \pi r \frac{B_{\phi}}{\mu_{0}}
$$

At location 1 in Figure 4-8, the displacement current equals zero and we are left with the integral

$$
\begin{equation*}
2 \pi r \frac{B_{\phi}}{\mu_{0}}=\int_{\Delta s} \mathrm{~J} \cdot d \mathbf{d s} \tag{4.27}
\end{equation*}
$$

At location 2 in Figure 4-8, the conduction current equals zero and we are left with the integral

$$
\begin{equation*}
2 \pi r \frac{B_{\phi}}{\mu_{0}}=\int_{\Delta s} \frac{\partial \mathbf{D}}{\partial \mathrm{t}} \bullet \mathbf{d s} \tag{4.28}
\end{equation*}
$$

The next example will demonstrate that these currents are identical, hence the magnetic flux densities will be the same at the same radius $r$.

## Displacement current

Example 4-11. Verify that the conduction current in the wire equals the displacement current between the plates of the parallel plate capacitor in the circuit. The voltage source has $V_{c}=V_{0} \sin \omega t$.


Answer: The conduction current in the wire is given by

$$
I_{c}=C \frac{d V_{c}}{d t}=C V_{0} \omega \cos \omega t
$$

The capacitance of the parallel plate capacitor is given by

$$
\mathrm{C}=\frac{\varepsilon \mathrm{A}}{\mathrm{~d}}
$$

where $A$ is the area of the plates that are separated by a distance $d$. The electric field between the plates is given by $E=V_{C} / d$. The displacement flux density equals

$$
\mathrm{D}=\varepsilon \mathrm{E}=\varepsilon \frac{\mathrm{V}_{0}}{\mathrm{~d}} \sin \omega \mathrm{t}
$$

The displacement current is computed from

$$
\mathrm{I}_{\mathrm{d}}=\int_{\mathrm{A}} \frac{\partial \mathbf{D}}{\partial \mathrm{t}} \bullet \mathbf{d s}=\left(\frac{\varepsilon \mathrm{A}}{\mathrm{~d}}\right) \mathrm{V}_{0} \omega \cos \omega \mathrm{t}=\mathrm{CV}_{0} \omega \cos \omega \mathrm{t}=\mathrm{I}_{\mathrm{c}} .
$$

Example 4-12. The magnetic flux density in vacuum is given by

$$
\mathbf{B}=\mathrm{B}_{0} \cos 2 \mathrm{x} \cos (\omega \mathrm{t}-\beta \mathrm{y}) \mathbf{u}_{\mathrm{x}}
$$

Find the displacement current, the displacement flux density, and the volume charge density associated with this magnetic flux density.

Answer: We write

$$
\mathbf{J}_{\mathbf{d}}=\frac{\partial \mathbf{D}}{\partial \mathbf{t}}=\frac{1}{\mu_{0}} \nabla \times \mathbf{B}
$$

## Displacement current

$$
\begin{gathered}
=\frac{1}{\mu_{0}}\left|\begin{array}{ccc}
\mathbf{u}_{\mathbf{x}} & \mathbf{u}_{\mathbf{Y}} & \mathbf{u}_{\mathbf{z}} \\
\frac{\partial}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{y}} & \frac{\partial}{\partial \mathbf{z}} \\
\mathrm{B}_{0} \cos 2 x \cos (\omega t-\beta y) & 0 & 0
\end{array}\right| \\
=-\frac{\beta B_{0}}{\mu_{0}} \cos 2 x \sin (\omega t-\beta y) \mathbf{u}_{\mathbf{z}}
\end{gathered}
$$

The displacement flux density $\mathbf{D}$ is found from the displacement current as

$$
\mathbf{D}=\int \mathbf{J}_{\mathrm{d}} \mathrm{dt}=\int\left[-\frac{\beta \mathrm{B}_{0}}{\mu_{0}} \cos 2 x \sin (\omega t-\beta y) \mathbf{u}_{z}\right] \mathrm{dt}=-\frac{\beta \mathrm{B}_{0}}{\omega \mu_{0}} \cos 2 x \cos (\omega t-\beta y) \mathbf{u}_{z}
$$

The volume charge density is computed from

$$
\rho_{v}=\nabla \bullet \mathbf{D} \Rightarrow \frac{\partial \mathbf{D}_{z}}{\partial z}=0
$$

Example 4-13. In a lossy dielectric medium with a conductivity $\sigma$ and a relative permittivity $\varepsilon_{r}$ there is a time-harmonic electric field $E=E_{0} \sin \omega t$. Compare the magnitudes of the following terms: a) the conduction current density $J_{c}$; b) the displacement current density $J_{d}$.

Answer: The conduction current density can be found from Ohm's law (2.101) $J_{c}=\sigma E=\sigma E_{0} \sin \omega t$, while the displacement current density can be calculated from (4.23) $J_{d}=\partial \mathrm{D} / \partial \mathrm{t}=\varepsilon \mathrm{E}_{0} \omega \cos \omega \mathrm{t}$. The ratio of their magnitudes is

$$
\mathrm{r}=\frac{\mathrm{J}_{\mathrm{cm}}}{J_{\mathrm{dm}}}=\frac{\sigma}{\omega \varepsilon_{0} \varepsilon_{\mathrm{r}}}
$$

For materials that have a relative dielectric constant that is close to 1 , this fraction will depend mainly on the conductivity of the material and the frequency of the electromagnetic signal. The conduction current is dominant at low frequencies in a conductor and the displacement current will be dominant in a dielectric at high frequencies. This latter effect will be further discussed in the next chapter.

# Maxwell's equations 

### 4.4. Maxwell's equations

Everything that we have learned up to this point can be summarized in the four differential equations of Maxwell ${ }^{1}$ which are rewritten below as

$$
\begin{gather*}
\nabla \times \mathbf{E}=-\frac{\partial \mathbf{B}}{\partial \mathrm{t}}  \tag{4.29}\\
\nabla \times \mathbf{H}=\mathbf{J}+\frac{\partial \mathbf{D}}{\partial \mathrm{t}}  \tag{4.30}\\
\nabla \bullet \mathbf{D}=\rho_{\mathbf{V}}  \tag{4.31}\\
\nabla \cdot \mathbf{B}=0 \tag{4.32}
\end{gather*}
$$

These four equations along with a set of relations called the constitutive relations

$$
\left.\begin{array}{l}
\mathbf{D}=\varepsilon \mathbf{E} \\
\mathbf{B}=\mu \mathbf{H}  \tag{4.33}\\
\mathbf{J}=\sigma \mathbf{E}
\end{array}\right\}
$$

describe electromagnetic phenomena. The constitutive relations relate the electromagnetic fields to the material properties in which the fields exist. We will see that the propagation of electromagnetic waves such as light is described by Maxwell's equations.

Nonlinear phenomena can also be described with this set of equations through any nonlinearity that may exist in the constitutive relations. For example,

[^10]
## Maxwell's equations

certain optical fibers that are used in communication have a dielectric constant that depends nonlinearly on the amplitude of the wave that propagates in the fiber. It is possible to approximate the relative dielectric constant in the fiber with the expression $\varepsilon_{\mathrm{r}} \approx\left[1+\alpha|E|^{2}\right]$ where $\alpha$ is a constant that has the dimensions of [volts/meter] ${ }^{-2}$. In writing (4.33), we have also assumed that the materials are isotropic and hysteresis can be neglected. The study of the myriad effects arising from these phenomena is of interest to a growing number of engineers and scientists throughout the world. We will, however, not concern ourselves with these problems here other than to be alerted of their existence.

Example 4-14. Show that the two "divergence" equations are implied by the two "curl" equations and the equation of continuity.

Answer: To show this, we must remember the vector identity $\nabla \bullet \nabla \times \boldsymbol{\zeta} \equiv 0$ where $\zeta$ is any vector. Hence

$$
\nabla \bullet \nabla \times \mathbf{E} \equiv 0=-\frac{\partial}{\partial \mathbf{t}}(\nabla \bullet \mathbf{B})
$$

implies $\nabla \bullet \mathbf{B}=$ constant. This constant equals zero since there are no sources nor sinks at which the magnetic flux density can originate nor terminate. This implies that magnetic monopoles do not exist.

We similarly write

$$
\nabla \bullet \nabla \times \mathbf{H} \equiv 0=\nabla \bullet \mathbf{J}+\frac{\partial}{\partial \mathrm{t}}(\nabla \bullet \mathbf{D})=-\frac{\partial \rho_{v}}{\partial \mathrm{t}}+\frac{\partial}{\partial \mathrm{t}}(\nabla \bullet \mathbf{D})
$$

This implies that $\nabla \bullet \mathbf{D}=\rho_{\mathrm{v}}$.

Example 4-15. In a conducting material, we may assume that the conduction current density is much larger than the displacement current density. Show that Maxwell's equations can be cast in the form of a diffusion equation in this material.

# Maxwell's equations 

Answer: In this case, (4.29) and (4.30) are written as

$$
\nabla \times \mathbf{E}=-\frac{\partial \mathbf{B}}{\partial \mathrm{t}} \text { and } \nabla \times \mathbf{H}=\mathbf{J}=\sigma \mathbf{E}
$$

where the displacement current has been neglected. Take the curl of the second equation

$$
\nabla \times \nabla \times \mathbf{H}=\sigma \nabla \times \mathbf{E}
$$

Expand the left hand side with a vector identity and substitute the first equation into the right hand side.

$$
\nabla\left(\nabla \cdot \frac{\mathbf{B}}{\mu_{0}}\right)-\nabla^{2}\left(\frac{\mathbf{B}}{\mu_{0}}\right)=-\sigma \frac{\partial \mathbf{B}}{\partial \mathrm{t}}
$$

From (4.32), the first term is zero leaving

$$
\nabla^{2} \mathbf{B}=\mu_{0} \sigma \frac{\partial \mathbf{B}}{\partial \mathrm{t}}
$$

This is a diffusion equation with a diffusion coefficient $\mathbf{D}=\mu_{0} \sigma$. Since $\mathbf{B}$ is a vector, this corresponds to three scalar equations for the three components. The term $\nabla^{2}$ is the Laplacian operator described in Chapters 2 and 3. A onedimensional diffusion equation will be encountered again in Chapter 6.

Example 4-16 ${ }^{\text {MATLAB }}$. Solve the diffusion equation, derived in Example 4-15, for the case of a magnetic flux density $B_{x}(z, t)$ near a plane vacuum-copper interface assuming the following values for copper: $\mu=\mu_{0}=4 \pi \times 10^{-7} \mathrm{H} / \mathrm{m}$ and $\sigma=5.8 \times 10^{7} \mathrm{~S} / \mathrm{m}$. Plot the solution for the spatial profile of the magnetic field, assuming a 60 Hz time harmonic electromagnetic signal is applied.

Answer: Assuming $e^{j \omega t}$ time-variation, the diffusion equation is transformed to the following ordinary differential equation for the spatial variation of the magnetic field

$$
\frac{d^{2} B_{x}}{d z^{2}}=j \mu_{0} \sigma \omega B_{x}
$$

where $z$ is the coordinate normal to the vacuum-copper boundary. Assuming

## Maxwell's equations

variation in the $z$-direction to be $B_{x}=B_{0} \exp (-\gamma z)$, we write

$$
\gamma^{2}=j \omega \mu_{0} \sigma \text { or } \gamma=\alpha+j \beta=\sqrt{j \omega \mu_{0} \sigma} .
$$

The magnitude of the magnetic flux density decays exponentially in the zdirection from the surface into the conductor

$$
\mathrm{B}_{\mathrm{x}}(\mathrm{z})=\mathrm{B}_{0} \mathrm{e}^{-\alpha z}
$$

with

$$
\alpha=\sqrt{\pi f \mu_{0} \sigma}=\sqrt{\pi \times 60 \times 4 \pi \times 10^{-7} \times 5.8 \times 10^{7}}=117.2 \mathrm{~m}^{-1} .
$$

The quantity $\delta=1 / \alpha$ is called a "skin depth". Here it is $\delta=8.5 \mathrm{~mm}$.
The plot of the spatial variation of the magnitude of the magnetic field inside the conductor is presented in the figure below.


The subject of the skin depth will be encountered again in the next chapter.

As written, Maxwell's equations in (4.29) to (4.32) are partial differential equations evaluated at a particular point in space and time. A totally equivalent
way of writing them is to write these equations as integrals. This is accomplished by integrating both sides of the first two equations over the same cross-sectional area and applying Stokes's theorem to the terms involving the curl operations and by integrating both sides of the second two equations over the same volume and applying the divergence theorem. We summarize this as follows.


Equation (4.34) states that the closed line integral of the electric field around a closed loop is equal to the time rate of change of the magnetic flux which passes through the surface area defined by the closed loop. This is the meaning of Faraday's law. Equation (4.35) states that the closed line integral of the magnetic field intensity is equal to the current that is enclosed within the loop. The current consists of the contribution due to the conduction current and the displacement current. This generalizes Ampere's circuital law which we encountered earlier.

Equation (4.36) states that the total displacement flux $\Psi_{\mathrm{e}}$ that leaves a closed surface is equal to the charge that is enclosed within the surface (Gauss's law). If the enclosed charge is negative, then the displacement flux $\Psi_{e}$ enters the closed surface and terminates on this negative charge. Equation (4.37) states

## Maxwell's equations

that the magnetic flux density is continuous and cannot terminate nor originate from a magnetic charge, i. e. the nonexistence of magnetic monopoles or magnetic charges.

Equations (4.34) to (4.37) are the integral form of Maxwell's equations and they are of the same importance as the differential form given in (4.29) to (4.32). Using the integral form of Maxwell's equations, we can easily derive the boundary conditions that relate the electromagnetic fields in one media to another. This will be shown later.

Either of the two forms of Maxwell's equations can be used although we will encounter the differential form more often in practice. An important derivation which describes the magnitude and direction of the flow of electromagnetic power will employ vector identities and the differential form of Maxwell's equations.

# Poynting's theorem 

### 4.5. Poynting's theorem

A frequently encountered problem in practice is to determine the direction that power is flowing if the electric and magnetic fields are independently measured in some experiment. This may not seem important for one in the laboratory where a signal generator can be separated from a resistive load impedance and the direction of the flow of power can be clearly ascertained. It is clear that the sun radiates energy that the earth receives. However, an investigator using a satellite floating in space may wish to determine the source of some anomalous extragalactic electromagnetic radiation in order to further map out the universe. Poynting's theorem will provide us with the method to accomplish this. This will be derived here.


Figure 4-9. An arbitrarily shaped volume that contains a source of electromagnetic energy.

To obtain Poynting's theorem for an arbitrary volume depicted in Figure 49, we will require two Maxwell equations and a vector identity. The two Maxwell equations that are required for this derivation are

$$
\begin{equation*}
\nabla \times \mathbf{E}=-\frac{\partial \mathbf{B}}{\partial \mathrm{t}} \tag{4.38}
\end{equation*}
$$

and

$$
\begin{equation*}
\nabla \times \mathbf{H}=\frac{\partial \mathbf{D}}{\partial \mathrm{t}}+\mathbf{J} \tag{4.39}
\end{equation*}
$$

## Poynting's theorem

Let us take the scalar product of $E$ with (4.39) and subtract it from the scalar product of H with (4.38). Performing this operation leads to

$$
\mathbf{H} \bullet \nabla \times \mathbf{E}-\mathbf{E} \bullet \nabla \times \mathbf{H}=-\mathbf{H} \bullet \frac{\partial \mathbf{B}}{\partial \mathrm{t}}-\mathbf{E} \bullet\left[\frac{\partial \mathbf{D}}{\partial \mathrm{t}}+\mathbf{J}\right]
$$

The left hand side of this equation can be replaced using vector identity (A.1.9)

$$
\nabla \cdot(\mathbf{A} \times \mathbf{B})=\mathbf{B} \bullet \nabla \times \mathbf{A}-\mathbf{A} \bullet \nabla \times \mathbf{B}
$$

Therefore, we obtain

$$
\begin{equation*}
\nabla \bullet(\mathbf{E} \times \mathbf{H})=-\mathbf{H} \bullet \frac{\partial \mathbf{B}}{\partial t}-\mathbf{E} \bullet \frac{\partial \mathbf{D}}{\partial t}-\mathbf{E} \bullet \mathbf{J} \tag{4.40}
\end{equation*}
$$

After the introduction of the constitutive relations (4.33), the terms involving the time derivatives can be written as

$$
\begin{equation*}
-\mathbf{H} \bullet \frac{\partial \mathbf{B}}{\partial t}-\mathbf{E} \bullet \frac{\partial \mathbf{D}}{\partial \mathrm{t}}=-\frac{1}{2} \frac{\partial}{\partial \mathrm{t}}[\mu \mathbf{H} \bullet \mathbf{H}+\varepsilon \mathbf{E} \bullet \mathbf{E}]=-\frac{\partial}{\partial \mathrm{t}} \frac{1}{2}\left[\mu \mathrm{H}^{2}+\varepsilon \mathbf{E}^{2}\right] \tag{4.41}
\end{equation*}
$$

Substitute (4.41) into (4.40) and integrate both sides of the resulting equation over the same volume $\Delta \mathrm{v}$. This volume is completely enclosed by the surface s. Performing this integration leads to

$$
\begin{equation*}
\int_{\Delta v} \nabla \bullet(E \times H) d v=-\frac{d}{d t} \int_{\Delta v} \frac{1}{2}\left[\mu H^{2}+\varepsilon E^{2}\right] d v-\int_{\Delta v} E \cdot J d v \tag{4.42}
\end{equation*}
$$

The volume integral on the left-hand side of (4.42) can be converted to a closed surface integral via the divergence theorem. With the substitution Ohm's law $\mathbf{J}=\sigma \mathbf{E}$, we finally obtain

$$
\begin{equation*}
\oint(\mathbf{E} \times \mathbf{H}) \cdot d \mathbf{d}=-\frac{d}{d t} \int_{\Delta v} \frac{1}{2}\left[\mu H^{2}+\varepsilon E^{2}\right] d v-\int_{\Delta v} \sigma E^{2} d v \tag{4.43}
\end{equation*}
$$

where $\mathbf{S}=\mathbf{E} \times \mathbf{H}$ is called the Poynting vector. It is the power density of the radiated electromagnetic fields. The direction of the radiated power is included in this vector.

Let us now give a physical interpretation to each of the three terms that appear in this equation. The units of the closed surface integral are

# Poynting's theorem 

$$
\frac{\text { Volts }}{\text { meter }} \times \frac{\text { Amperes }}{\text { meter }} \times \text { meter }^{2}=\text { Watts }
$$

or the closed surface integral has the units of power. Using the definition of the scalar product and the fact that the notation ds refers to the outward normal of the surface that encloses the volume $\Delta \mathrm{v}$, this term represents the total power that leaves or is radiated from the volume $\Delta \mathrm{v}$.

The terms within the integrand of the first volume integral can be recognized as the stored magnetic energy density and the stored electric energy density that was previously described in static fields. The time derivative introduces a unit of (second)-1 The units of this term are

$$
\frac{1}{\text { sec ond }} \times \frac{\text { Joule }^{\text {meter }^{3}} \times \text { meter }^{3}=\text { Watts }}{}
$$

This term corresponds to the time derivative of the stored electromagnetic energy within the volume.

The units of the second volume integral corresponds to Joule heating within the volume and they are also in terms of watts

$$
\frac{1}{\text { Ohms } \times \text { meter }} \times \frac{\text { Volts }^{2}}{\text { meter }^{2}} \times \text { meter }^{3}=\text { Watts }
$$

The reference to Joule heating indicates that electromagnetic power is converted to heat and this power cannot be recovered. A toaster uses Joule heating.

Hence, Poynting's theorem states that the power that leaves a region is equal to the temporal decay in the energy that is stored within the volume minus the power that is dissipated as heat within it. A common sense example will illustrate this theorem. Additional applications of this important theorem will be found in the last chapter of this book that describes radiation.

The equation of the energy conservation (4.43) can also be written in differential form. Recalling that electromagnetic energy density is defined as

$$
\begin{equation*}
w=\frac{1}{2}\left[\mu H^{2}+\varepsilon E^{2}\right] \tag{4.44}
\end{equation*}
$$

## Poynting's theorem

and the power loss density is given by

$$
\begin{equation*}
\mathrm{p}_{\mathrm{L}}=\sigma \mathrm{E}^{2} \tag{4.45}
\end{equation*}
$$

We can reinterpret (4.42) in the following differential form of the energy conservation of the system

$$
\begin{equation*}
\nabla \cdot \mathbf{S}+\frac{\partial \mathrm{w}}{\partial \mathrm{t}}=-\mathrm{p}_{\mathrm{L}} \tag{4.46}
\end{equation*}
$$

This equation is somewhat similar to the equation of continuity (4.18) with a "sink" term that corresponds to the Joule heating.

Example 4-17. Using Poynting's theorem, calculate the power that is dissipated in the resistor as heat. The electric energy is supplied by the battery. Neglect the magnetic field that is confined within the resistor and calculate its value only at the surface. In addition, assume that there are conducting surfaces at the top and the bottom of the resistor so they are equipotential surfaces. Also assume that the radius of the resistor is much less than its length.


Answer: The electric field has a magnitude of $E=V_{0} / L$ and the magnitude of the magnetic field intensity at the outer edge of the resistor is $\mathrm{H}=\mathrm{I} /(2 \pi \mathrm{a})$. The direction of the Poynting vector $\mathbf{S}=\mathbf{E} \times \mathbf{H}$ is into the resistor. There is no energy stored in a resistor. The magnitude of the current density that is in the same direction as the electric field is $J=I /\left(\pi a^{2}\right)$. Therefore, the various terms in

## Poynting's theorem

Poynting's theorem (4.43) are found to be

$$
-\left(\frac{\mathrm{V}_{0}}{\mathrm{~L}}\right)\left(\frac{\mathrm{l}}{2 \pi \mathrm{a}}\right)(2 \pi \mathrm{aL})=-\frac{\mathrm{d}}{\mathrm{dt}} \int_{\Delta v}[0+0] \mathrm{dv}-\left(\frac{\mathrm{l}}{\pi \mathrm{a}^{2}}\right)\left(\frac{\mathrm{V}_{0}}{\mathrm{~L}}\right)\left(\pi \mathrm{a}^{2} \mathrm{~L}\right)
$$

yielding

$$
-\mathrm{V}_{0} \mathrm{I}=-\mathrm{V}_{0} \mathrm{I} .
$$

The electromagnetic energy of the battery is fully absorbed by the resistor.

Example 4-18. Using Poynting's theorem, calculate the power that is flowing through the surface area at the radial edge of a capacitor. Neglect the ohmic losses in the wires connecting the capacitor with the signal generator. Also assume that the radius of the capacitor is much greater than the separation distance between the plates.


Answer: Assuming the electric field $E$ is confined to be between the plates and is homogeneous, we can find the total electric energy that is stored in the capacitor to be

$$
\mathrm{W}=\left(\frac{\varepsilon \mathrm{E}^{2}}{2}\right)\left(\pi \mathrm{a}^{2} \mathrm{~b}\right)
$$

## Poynting's theorem

The total magnetic energy that is stored in the capacitor is equal to 0 . The differentiation of this electric energy with respect to time yields

$$
-\frac{\mathrm{dW}}{\mathrm{dt}}=-\varepsilon\left(\pi \mathrm{a}^{2} \mathrm{~b}\right) E \frac{\mathrm{dE}}{\mathrm{dt}}
$$

This is the only term which survives on the right side of (4.43) since an ideal capacitor does not dissipate energy.

The left-hand side of (4.43) requires an expression for the time varying magnetic field intensity in terms of the displacement current. Evaluating (4.26) at the radial edge of the capacitor, we write

$$
\oint \mathbf{H} \cdot \mathbf{d} \mathbf{l}=\int_{\Delta s}\left(\varepsilon \frac{\partial \mathbf{E}}{\partial \mathrm{t}}\right) \cdot \mathbf{d s}
$$

There is no conduction current in this ideal capacitor. We obtain

$$
\mathrm{H}(2 \pi \mathrm{a})=\varepsilon \frac{\mathrm{dE}}{\mathrm{dt}}\left(\pi \mathrm{a}^{2}\right) \text { or } \mathrm{H}=\frac{\varepsilon \mathrm{a}}{2} \frac{\mathrm{dE}}{\mathrm{dt}}
$$

Now we can write the Poynting vector power flow as

$$
P_{S}=-(E H)(2 \pi a b)=-\varepsilon\left(\pi a^{2} b\right) E \frac{d E}{d t}
$$

The minus sign arises since the direction of Poynting vector is radially inward.
Comparing both expressions, we find that they are equal which implies that

$$
P_{S}=-\frac{d W}{d t}
$$

The states that energy is conserved in the circuit as should be expected.

In these two examples, we see that Poynting's theorem can be interpreted in terms of electrical circuit elements. In these examples, electromagnetic power

## Poynting's theorem

was directed into the element. The radiation of electromagnetic power that is directed radially outward will be discussed in Chapter 7 when antennas are described.

## Time-harmonic electromagnetic fields

### 4.6. Time-harmonic electromagnetic fields

In practice, we will frequently encounter electromagnetic fields whose temporal variation is harmonic. Maxwell's equations and the Poynting vector will assume a particular form since the fields can be represented as phasors. In particular, we write the fields as

$$
\begin{equation*}
E(x, y, z, t)=\operatorname{Re}\left[E(x, y, z) e^{j \omega t}\right] \tag{4.47}
\end{equation*}
$$

and

$$
\begin{equation*}
\mathbf{H}(x, y, z, t)=\operatorname{Re}\left[\boldsymbol{H}(x, y, z) e^{j \omega t}\right] \tag{4.48}
\end{equation*}
$$

where Re stands for the real part. There may be a phase angle $\phi$ between the electric and magnetic fields that will be absorbed into the terms $E(x, y, z)$ and $H(x, y, z)$. We are able to employ either a cosine function or a sine function If we had chosen the sine function, the phase angle $\phi$ would have to be replaced with a new phase angle $\phi^{\prime}=\phi-\pi / 2$. The only restriction that will be encountered is that we remain consistent throughout a calculation.

In terms of the phasors $\boldsymbol{E}$ and $\boldsymbol{H}$, we write Maxwell's equations as

$$
\begin{align*}
& \nabla \times \boldsymbol{E}=-\mathrm{j} \omega \mu \boldsymbol{H}  \tag{4.49}\\
& \nabla \times \boldsymbol{H}=\mathrm{j} \omega \varepsilon \boldsymbol{E}+\boldsymbol{J}  \tag{4.50}\\
& \nabla \bullet \boldsymbol{E}=\rho_{v} / \varepsilon  \tag{4.51}\\
& \nabla \bullet \boldsymbol{B}=0 \tag{4.52}
\end{align*}
$$

where the term representing the temporal variation $\mathrm{e}^{\mathrm{j} \omega t}$ that is common to both sides of these equations has been canceled. Hopefully, there will be little confusion in notation since we have not introduced any new symbols.

Example 4-19. Compute the frequency at which the conduction current equals the displacement current.

Answer: From (4.50), we write

$$
\nabla \times \boldsymbol{H}=\boldsymbol{J}+\mathrm{j} \omega \varepsilon \boldsymbol{E}=(\sigma+\mathrm{j} \omega \varepsilon) \boldsymbol{E}
$$

The frequency is given by

# Time-harmonic electromagnetic fields 

$$
\omega=\frac{\sigma}{\varepsilon}
$$

For copper, the frequency $f=\omega / 2 \pi$ is

$$
f=\frac{\sigma}{2 \pi \varepsilon_{0}}=\frac{5.8 \times 10^{7}}{2 \pi \times \frac{1}{36 \pi} \times 10^{-9}}=1.04 \times 10^{18} \mathrm{~Hz}
$$

At frequencies much above this value, copper which is thought to be a good conductor acts like a dielectric. In this case, the electromagnetic wave can propagate through the copper barrier.

The derivation of the Poynting vector requires some care when we are considering time harmonic fields. This is because the Poynting vector involves the product $E \times H$. Power is a real quantity and we must be careful since

$$
\begin{equation*}
\operatorname{Re}\left[E e^{j \omega t}\right] \times \operatorname{Re}\left[\boldsymbol{H e}^{j \omega t}\right] \neq \operatorname{Re}\left[E \times \boldsymbol{H e}^{j \omega t}\right] \tag{4.53}
\end{equation*}
$$

To effect the derivation of the Poynting vector, we make use of the following relations

$$
\begin{equation*}
\operatorname{Re}(\boldsymbol{E})=\left(\boldsymbol{E}+\boldsymbol{E}^{*}\right) / 2 \text { and } \operatorname{Re}(\boldsymbol{H})=\left(\boldsymbol{H}+\boldsymbol{H}^{*}\right) / 2 \tag{4.54}
\end{equation*}
$$

where the star indicates the complex conjugate of the function. We write

$$
\begin{aligned}
\operatorname{Re}(\boldsymbol{E}) \times \operatorname{Re}(\boldsymbol{H}) & =\left[\left(\boldsymbol{E}+\boldsymbol{E}^{*}\right) / 2\right] \times\left[\left(\boldsymbol{H}+\boldsymbol{H}^{*}\right) / 2\right] \\
& =\left[\boldsymbol{E} \times \boldsymbol{H}^{*}+\boldsymbol{E}^{*} \times \boldsymbol{H}+\boldsymbol{E} \times \boldsymbol{H}+\boldsymbol{E}^{*} \times \boldsymbol{H}^{*}\right] / 4
\end{aligned}
$$

The time variation $e^{j \omega t}$ cancels in two of the terms and introduces a factor of $e^{ \pm j 2 \omega t}$ in the remaining two terms. After taking a time average of this power, these latter terms will contribute nothing to the result. We finally obtain the time average power to be

$$
\begin{equation*}
\mathbf{S}_{\mathrm{av}}=\frac{1}{2} \operatorname{Re}\left[\boldsymbol{E} \times \boldsymbol{H}^{\star}\right]\left(\mathrm{W} / \mathrm{m}^{2}\right) \tag{4.55}
\end{equation*}
$$

Example 4-20. The field vectors in free space are given by

$$
\mathbf{E}=10 \cos \left(\omega t+\frac{4 \pi}{3} z\right) \mathbf{u}_{\mathrm{x}} \mathrm{~V} / \mathrm{m} \text { and } \mathbf{H}=\left(\mathbf{u}_{\mathbf{z}} \times \mathbf{E}\right) / 120 \pi \mathrm{~A} / \mathrm{m}
$$

## Time-harmonic electromagnetic fields

The frequency $f=500 \mathrm{MHz}$. Determine the Poynting vector. The numerical value of $120 \pi$ as a free-space impedance will become apparent in the next chapter.

Answer: In phasor notation, the fields are expressed as

$$
\boldsymbol{E}=10 \mathrm{e}^{\mathrm{j}(4 \pi / 3) z} \mathbf{u}_{\mathbf{x}} \text { and } \boldsymbol{H}=\frac{10}{120 \pi} \mathrm{e}^{\mathrm{j}(4 \pi / 3) z} \mathbf{u}_{\mathbf{Y}}
$$

and Poynting vector is

$$
\mathbf{S}_{\mathrm{av}}=\frac{1}{2} \operatorname{Re}\left[\boldsymbol{E} \times \boldsymbol{H}^{*}\right]=\frac{10^{2}}{2 \times 120 \pi} \mathbf{u}_{\mathbf{z}}=0.133 \mathbf{u}_{\mathbf{z}} \mathrm{W} / \mathrm{m}^{2}
$$

Having now manipulated the complex phasors to derive (4.55), let us apply this to the derivation of Poynting's theorem. In particular, we desire to explicitly obtain the terms $\boldsymbol{E}$ and $\boldsymbol{H}^{*}$. The procedure that we will follow is to subtract the scalar product of $E$ with the complex conjugate of (4.50) from the scalar product of $\boldsymbol{H}^{*}$ with (4.49) resulting in

$$
\begin{equation*}
\boldsymbol{H}^{*} \bullet \nabla \times \boldsymbol{E}-\boldsymbol{E} \bullet \nabla \times \boldsymbol{H}^{*}=-\mathrm{j} \omega \mu \boldsymbol{H} \bullet \boldsymbol{H}^{*}+\mathrm{j} \omega \varepsilon \boldsymbol{E} \bullet \boldsymbol{E}^{*}-\boldsymbol{E} \bullet \boldsymbol{J}^{*} \tag{4.56}
\end{equation*}
$$

Employing the same vector identity that we previously used to derive (4.40), we recognize that (4.56) can be written as

$$
\begin{equation*}
\nabla \bullet\left(\boldsymbol{E} \times \boldsymbol{H}^{*}\right)=-\mathrm{j} \omega \mu H^{2}+\mathrm{j} \omega \varepsilon E^{2}-\sigma E^{2} \tag{4.57}
\end{equation*}
$$

where $\boldsymbol{H} \bullet \boldsymbol{H}^{*}=H^{2} ; \boldsymbol{E} \bullet \boldsymbol{E}^{*}=E^{2}$; and $\boldsymbol{E} \bullet \boldsymbol{J}^{*}=\sigma E^{2}$.
Following the procedure that has served us so well previously, we integrate the terms that appear in (4.57) over the volume of interest

$$
\begin{equation*}
\int_{\Delta v} \nabla \bullet\left(\boldsymbol{E} \times \boldsymbol{H}^{*}\right) \mathrm{dv}=-\mathrm{j} \omega \int_{\Delta v}\left[\mu H^{2}-\varepsilon E^{2}\right] \mathrm{dv}-\int_{\Delta v} \sigma E^{2} \mathrm{dv} \tag{4.58}
\end{equation*}
$$

The volume integral is converted to a closed surface integral that encloses the volume $\Delta v$

$$
\begin{equation*}
\oint\left(\boldsymbol{E} \times \boldsymbol{H}^{*}\right) \cdot \mathrm{ds}=-\mathrm{j} \omega \int_{\Delta v}\left[\mu H^{2}-\varepsilon E^{2}\right] \mathrm{dv}-\int_{\Delta v} \sigma E^{2} \mathrm{dv} \tag{4.59}
\end{equation*}
$$

The closed surface integral represents the total power that is radiated from

## Time-harmonic electromagnetic fields

within the volume enclosed by this surface. The last term represents the power that is dissipated within this volume. This power could have been turned into heat and would not be recovered. The remaining two terms are the time-average energy stored within the volume. The factor "j" indicates that this is similar to the reactive energy stored in the capacitor or inductor in an RLC circuit.

## Conclusion

### 4.7. Conclusion

We have now come to the end of a long journey in order to obtain the set of four Maxwell's equations that describe electromagnetic phenomena. We have demonstrated that time-varying electric and magnetic fields can be determined from each other through these equations and that they are intimately intertwined. Faraday's law of induction and Ampere's circuital law with the introduction of a displacement current relate time-varying magnetic fields to time-varying electric fields. The term that represents the displacement current arises from the requirement that the equation of continuity must be satisfied. The boundary conditions that we encountered in static fields apply equally well in time varying fields.

There is a T-shirt that paraphrases the book of Genesis by stating that "In the beginning, God said '...,' and there was light" where these equations are included within the proclamation. The goal, and accomplishment, of hundreds of graduate students since Maxwell first inscribed these equations on paper has been to pose a new electromagnetic problem, solve it starting from these equations, and write a thesis. Even after obtaining a graduate degree, this set of equations usually appears as "Equation (1) to (4)" of many of their later scholarly articles that are then stored in dusty archives. You, as a student, are not expected to write these equations on a crib sheet and bring them to an examination or even memorize them for that inquiry, you are expected to know them! The intellectual and even the visceral understanding of these equations is what this course and much of electrical and computer engineering is about.

## Problems 4

### 4.8. Problems

1. In a source-free region, we find that $\mathbf{B}=\mathbf{z} \mathbf{u}_{\mathbf{Y}}+\mathbf{x u _ { \mathbf { z } }}$. Does $\mathbf{E}$ vary with time?
$2^{\text {MATLAB }}$. A perfect conductor joins two ends of a $100 \Omega$ resistor and the closed loop is in a region of uniform magnetic flux density $\mathrm{B}=10 \exp (-\mathrm{t} / 10) \mathrm{T}$.


Neglecting the self-inductance of the loop, find and plot the voltage $\mathrm{V}(\mathrm{t})$ that appears across the $100 \Omega$ resistor. A device based on this principle is used to monitor time-varying magnetic fields in experiments and in biological studies.
3. A closed loop ( $\Delta x=30 \mathrm{~cm} x \Delta y=20 \mathrm{~cm}$ ) of wire is passed through a nonuniform time-independent magnetic field $\mathbf{B}=\mathbf{x u}_{\mathbf{z}} \mathbf{T}$ with a constant velocity $v_{o}=5 u_{x} \mathrm{~m} / \mathrm{s}$. At $\mathrm{t}=0$, the loop's lower left corner is located at the origin (see the figure below). Find an expression for the voltage V , generated by the loop as a function of time. You may neglect the magnetic field created by the current in the loop.


## Problems 4

4. Repeat problem 3 with the magnetic flux density being uniform in space $\mathbf{B}=0.1 \mathbf{u}_{\mathrm{z}} \mathrm{T}$. Explain your result.
5. Find the generated voltage if the axle moves at a constant velocity $\mathbf{v}=\mathbf{v} \mathbf{u}_{\mathbf{x}}=3 \mathbf{u}_{\mathbf{x}}$ $\mathrm{m} / \mathrm{s}$ in a uniform magnetic field of $\mathbf{B}=\mathrm{B}_{\mathrm{o}} \mathbf{u}_{\mathbf{z}}=5 \mathbf{u}_{\mathbf{z}} \mathrm{T}$. At $\mathrm{t}=0$, the axle was at $\mathrm{x}=0$, $\mathrm{L}=40 \mathrm{~cm}$.

6. Repeat problem 5 with the constraint that the rails separate with $L=L_{o}+L_{1} x$.

The wheels are free to slide on the "trombone-like" axle so they remain on the rails $\left(L_{0}=0.4 m, L_{1}=0.04\right)$.
7. A tethered satellite is to be connected to the shuttle to generate electricity as it passes through the ambient plasma. Assuming that the shuttle takes two hours to go around the earth, find the expected voltage difference $\Delta \mathrm{V}$ between the two objects. The shuttle flies approximately 400 km above the earth where
 $B \approx 10-5$ Tesla.
$8^{\text {MATLAB }}$. A conducting axle oscillates over two conducting parallel rails in a uniform magnetic field $B=B_{0} \mathbf{u}_{\mathbf{z}}\left(B_{0}=4 T\right)$. The position of the axle is given by $x=(\Delta x / 2)[1-\cos \omega t]\left(\Delta x=0.2 m ; \omega=500 s^{-1}\right)$. Find and plot the current $I(t)$ if the

## Problems 4

resistance is $R=10 \Omega$ and the distance between the rails is $\Delta y=0.1 \mathrm{~m}$.

$9^{\text {MATLAB }}$. Repeat problem 8 with the magnetic field also varying in time as $B=B_{o} \cos \omega t \mathbf{u}_{\mathbf{z}}$ with $B_{0}$ and $\omega$ having the same values.
10. Calculate the voltage that is induced between the two nodes as the coil with dimensions $0.5 \mathrm{~m} \times 0.5 \mathrm{~m}$ rotates in a uniform magnetic field with a flux density $B=2 T$ with a constant angular frequency $\omega=1200 \mathrm{~s}^{-1}$.

11. A square loop is adjacent to an infinite wire that carries a current I. The loop moves with a velocity $\mathbf{v}=v_{o} \mathbf{u}_{\mathbf{r}}$. The center of the loop is at $r$ and the initial position is $r=b$. Determine the induced voltage $\mathrm{V}(\mathrm{t})$ in the loop assuming dimensions $\mathrm{a} \times 2 \mathrm{~b}$.


## Problems 4

12. The 1 m long wire shown in the figure rotates with an angular frequency $1200 \mathrm{~min}^{-1}$ in the magnetic field $\mathbf{B}=0.5 \sin \phi \mathbf{u}_{\phi} \mathrm{T}$. Find the current in the closed loop with a resistance $100 \Omega$.

13. The current density is $\mathbf{J}=\sin (\pi x) \mathbf{u}_{\mathbf{x}}$. Find the time rate of increase of the charge density $\partial \rho_{V} / \partial t$ at $x=1$.
14. The current density is $\mathbf{J}=\exp \left(-r^{2}\right) \mathbf{u}_{\mathbf{r}}$ in cylindrical coordinates. Find the time rate of increase of the charge density at $r=1$.
15. Compare the magnitudes of the conduction and displacement current densities in copper ( $\sigma=5.8 \times 10^{7} \mathrm{~S} / \mathrm{m}, \varepsilon=\varepsilon_{0}$ ), sea water $\left(\sigma=4 \mathrm{~S} / \mathrm{m}, \varepsilon=81 \varepsilon_{0}\right)$, and earth ( $\left.\sigma=10^{-3} \mathrm{~S} / \mathrm{m}, \varepsilon=10 \varepsilon_{0}\right)$ at $60 \mathrm{~Hz}, 1 \mathrm{MHz}$, and at 1 GHz .
16. Given the conduction current density in a lossy dielectric as $J_{c}=0.2 \sin \left(2 \pi 10^{9} t\right) A / \mathrm{m}^{2}$. Find the displacement current density if $\sigma=10^{3} \mathrm{~S} / \mathrm{m}$ and $\varepsilon_{r}=6.5$.
17. Show that the fields

$$
B=B_{0} \cos \omega t u_{x} \text { and } E=E_{0} \cos \omega t u_{z}
$$

do not satisfy Maxwell's equations in air $\varepsilon_{\mathrm{r}} \approx 1$. Show that the fields

$$
\mathbf{B}=\mathrm{B}_{0} \cos (\omega \mathrm{t}-\mathrm{ky}) \mathbf{u}_{\mathrm{x}} \text { and } \mathbf{E}=\mathrm{E}_{0} \cos (\omega \mathrm{t}-\mathrm{ky}) \mathbf{u}_{z}
$$

satisfy these equations. What is the value of $k$ in terms of the others stated parameters?
18. Given $E=E_{0} \cos (\omega t-k y) \mathbf{u}_{z}$ and $\mathbf{H}=\left(E_{0} / Z_{0}\right) \cos (\omega t-k y) u_{x}$ in a vacuum. Find $Z_{C}$ in terms of $\varepsilon_{0}$ and $\mu_{0}$ so Maxwell's equations are satisfied.

## Problems 4

19. Do the fields $E=E_{0} \cos x \operatorname{costu} u_{Y}$ and $H=\left(E_{0} / \mu_{0}\right) \sin x \sin t u_{z}$ satisfy Maxwell's equations?
20. Find a charge density $\rho_{V}$ that could produce in vacuum an electric field $E=E_{0} \cos x \cos t u_{x}$.
21. Write $E=120 \pi \cos \left(3 \times 10^{9} t-10 z\right) u_{x} V / m$ and $H=1 \cos \left(3 \times 10^{9} t-10 z\right) u_{Y} A / m$ in phasor notation.
22. Write the phasors $\boldsymbol{E}=3 \mathrm{e}^{-\mathrm{j} \beta z} \mathbf{u}_{\mathrm{x}}$ and $\boldsymbol{H}=0.4 \mathrm{e}^{-\mathrm{j} 45^{0}} \mathrm{e}^{-\mathrm{j} \beta z} \mathbf{u}_{\mathrm{y}}$ in the time domain. The frequency of oscillation is $\omega$. Find the average Poynting's vector $\mathbf{S}_{\mathrm{av}}$.
23. Find the displacement current density flowing through the dielectric of a coaxial cable of radii $a$ and $b$ where $b>a$ if a voltage $V_{0} \cos \omega t$ is connected between the two conducting cylinders.
24. Find the displacement current density flowing through the dielectric of 2 concentric spheres of radii $a$ and $b$ where $b>a$ if a voltage $\mathrm{V}_{0} \cos \omega t$ is connected between the two conducting spheres.
25. Starting from Maxwell's equations, derive the equation of continuity.
26. Write all of the terms that appear in Maxwell's equations in Cartesian coordinates.
27. If $E=E_{0} \cos (\omega t-\beta z) \mathbf{u}_{\mathbf{Y}}$ is a solution to Maxwell's equations, find $\mathbf{H}$. Find $\mathbf{S}_{\mathbf{a v}}$.
28. If $\mathbf{H}=\mathrm{H}_{0} \cos (\omega t-\beta z) \mathbf{u}_{\mathbf{Y}}$ is a solution to Maxwell's equations, find $\mathbf{E}$. Find $\mathbf{S}_{\mathbf{a v}}$.
29. At a frequency of $f=1 \mathrm{MHz}$, verify that copper $\left(\sigma=5.8 \times 10^{7} \mathrm{~S} / \mathrm{m}, \varepsilon_{r} \approx 1\right)$ is a good conductor and quartz $\left(\sigma=10^{-17} \mathrm{~S} / \mathrm{m}, \varepsilon_{\mathrm{r}}=4\right)$ is a good insulator.
30. Find the frequency where quartz becomes a conductor.
31. Find the frequency where copper becomes an insulator.
32. Compute the stored electric energy that is stored in a cube whose volume is one cubic meter in which a uniform electric field of $10^{4} \mathrm{~V} / \mathrm{m}$ exists. Compute the stored energies if the cube is empty and if it is filled with water that has $\varepsilon=81 \varepsilon_{0}$.

## Wave equation

## 5. Electromagnetic wave propagation

The reader of these pages has now been at the mountain top and has communed with the ideas of the great nineteenth century theoretical physicist. Four equations that form the basis of electromagnetic theory should have been inscribed on the reader's personal stone tablets during this period. The climb to the top has been strenuous, but as we will see, the efforts required to reach these heights will be well rewarded in that we will be now be able to understand how the warming light from the sun actually propagates toward us and what limits the ultimate speed of any electronic circuit. This will result from a proper manipulation of the four equations that are the foundation of this subject. As we pass over the pages of this chapter, we will examine the concept of an electromagnetic wave. The wave may challenge our mind but as we will see, it does carry electromagnetic energy that can warm our soul.

### 5.1. Wave equation

The point of embarkation for our study of electromagnetic waves is, of course, Maxwell's equations. This should not be too surprising since the explanation of all phenomena in electromagnetic theory can trace their origin to these four equations. Starting from these equations, we will show that the equations can be cast in terms of a wave equation.

The wave equation that we will initially derive will describe wave propagation in a homogeneous medium that could have losses. In our derivation, there will be no free charge density, hence $\rho_{\mathrm{V}}=0$. Therefore, Maxwell's equations are written as

$$
\begin{gather*}
\nabla \times \mathbf{E}=-\mu \frac{\partial \mathbf{H}}{\partial \mathrm{t}}  \tag{5.1}\\
\nabla \times \mathbf{H}=\varepsilon \frac{\partial \mathbf{E}}{\partial \mathrm{t}}+\sigma \mathbf{E} \tag{5.2}
\end{gather*}
$$

## Wave equation

$$
\begin{align*}
\nabla \cdot E & =0  \tag{5.3}\\
\nabla \cdot H & =0 \tag{5.4}
\end{align*}
$$

where we have incorporated the constitutive relations

$$
\left.\begin{array}{l}
\mathbf{D}=\varepsilon \mathbf{E}  \tag{5.5}\\
\mathbf{B}=\mu \mathbf{H} \\
\mathbf{J}=\sigma \mathbf{E}
\end{array}\right\}
$$

to eliminate the terms D, B, and J. Here an absence of external currents is assumed in (5.2). All of the electromagnetic fields explicitly depend on space and time, i. e. $\mathbf{E}=\mathbf{E}(\mathbf{r}, \mathrm{t})$ and $\mathbf{H}=\mathbf{H}(\mathbf{r}, \mathrm{t})$. The independent variables of space and time will not be explicitly stated in what follows but are to be understood. As we will see later, the two equations involving the divergence operation will be used to specify the value of a certain term in a vector identity and we will initially manipulate the two equations that contain the curl operation.

Equations (5.1) and (5.2) are two first order partial differential equations in the two dependent variables $\mathbf{E}$ and $\mathbf{H}$. We can combine them into one second order partial differential equation in terms of one of the variables. This is the same procedure that we normally employ when confronted with two coupled first order differential equations. We merely have to be careful here since we have vectors and vector operations in the equations.

For the electric field intensity $\mathbf{E}$, this combination of the two equations is accomplished by taking the curl of (5.1) and inserting (5.2) for the term $\nabla \times \mathbf{H}$ that will appear on one side of the resulting equation. This operation is written in detail as

$$
\begin{equation*}
\nabla \times(\nabla \times \mathbf{E})=-\mu \frac{\partial}{\partial \mathrm{t}}(\nabla \times \mathbf{H})=-\mu \frac{\partial}{\partial \mathrm{t}}\left(\sigma \mathbf{E}+\varepsilon \frac{\partial \mathbf{E}}{\partial \mathrm{t}}\right)=-\mu \sigma \frac{\partial \mathbf{E}}{\partial \mathrm{t}}-\mu \varepsilon \frac{\partial^{2} \mathbf{E}}{\partial \mathrm{t}^{2}} \tag{5.6}
\end{equation*}
$$

Note that we have freely interchanged the orders of differentiation of space and time in this step. The magnetic field intensity $\mathbf{H}$ can be later computed from the electric field intensity E using (5.1). As we will see later, there are certain

## Wave equation

advantages in solving for the electric field first since boundary conditions on the electric field are frequently found to be more easily specified. Recall that a metal conductor has a dramatic effect on the tangential component of the electric field intensity.

The next step that we should take in our journey into the understanding of wave phenomena is to employ the vector identity (A1.15) for the first term in equation (5.6). This leads to

$$
\begin{equation*}
\nabla \times \nabla \times \mathbf{E}=\nabla(\nabla \bullet \mathbf{E})-\nabla^{2} \mathbf{E}=-\nabla^{2} \mathbf{E} \tag{5.7}
\end{equation*}
$$

In reducing this equation, we have included the fact that the charge density $\rho_{v}$ is absent. This is specified in the divergence equation (5.3). After substituting (5.7) into the vector equation (5.6), we finally obtain

$$
\begin{equation*}
\nabla^{2} \mathbf{E}-\mu \sigma \frac{\partial \mathbf{E}}{\partial \mathrm{t}}-\mu \varepsilon \frac{\partial^{2} \mathbf{E}}{\partial \mathbf{t}^{2}}=0 \tag{5.8}
\end{equation*}
$$

This is the general homogeneous three-dimensional vector wave equation that we are seeking that is valid for the case of the absence of external sources. Note that as it is written, this equation does not depend on the coordinate system that has been chosen. The form of the Laplacian operator that should be employed in a calculation will, however, be dictated by the selected coordinate system. The choice of a coordinate system for a particular calculation is based upon any symmetry that is found in the problem. The polarization of the electric field $E$ will be determined by the polarization introduced with the excitation mechanism. This, for example, could be an antenna with a particular radiation characteristic, a laser, or a waveguide with a certain physical orientation. In the near field region around the antenna, an inhomogeneous vector wave equation has to be solved (see Chapter 7) which relates the electromagnetic field with the sources.

## Wave equation

The topic of the polarization of the electromagnetic field is an important subject that should not be glossed over lightly. Imagine for a moment that we have two infinite parallel metal plates that are connected to the two output terminals of a sinusoidal voltage generator as shown in Figure 5-1. The plates are far from each other, say a distance that approaches infinity. The voltage applied to the top plate will be $180^{\circ}$ out of phase with the voltage applied to the bottom plate. The resulting electric field directly between the two plates is said to be linearly polarized in one direction that we will define as the $\mathbf{u}_{\mathbf{y}}$ direction.


Figure 5-1. A technique to linearly polarize the electric field between the two infinite metal plates.

In Figure 5-1, we have introduced a technique to create an electromagnetic field that has a linear polarization. There are other polarizations such as circular and elliptical polarizations. An example will later describe these polarizations after we describe the excitation of time-harmonic waves. Rest assured that it is an important topic and it will be discussed.

By convention, we describe the polarization of an electromagnetic wave as being determined by the electric field component rather than the magnetic field component. In practice, we find that boundary conditions will be specified by the electric field rather than the magnetic field. Ascertaining one of the components of

## Wave equation

an electromagnetic wave determines the other component via Maxwell's equations, therefore all will be known.

Solutions of this general three-dimensional vector wave equation (5.8) may be difficult to write down and these solutions will certainly not be very useful at this stage in gaining an understanding of the concept of an electromagnetic wave. We must simplify the problem.

Example 5-1. Show that the wave equation for the magnetic field intensity $\mathbf{H}$ can be cast in the same form as (5.8).

Answer: From (5.1) and (5.2), we write

$$
\nabla \times \nabla \times \mathbf{H}=\sigma(\nabla \times \mathbf{E})+\varepsilon \frac{\partial}{\partial \mathrm{t}}(\nabla \times \mathbf{E})=\sigma\left(-\mu \frac{\partial \mathbf{H}}{\partial \mathrm{t}}\right)+\varepsilon \frac{\partial}{\partial \mathrm{t}}\left(-\mu \frac{\partial \mathbf{H}}{\partial \mathrm{t}}\right)
$$

The left hand side of this equation is reduced via the vector identity (5.7)

$$
\nabla \times \nabla \times \mathbf{H}=\nabla(\nabla \cdot \mathbf{H})-\nabla^{2} \mathbf{H}=-\nabla^{2} \mathbf{H}
$$

where (5.4) has been employed. Therefore, we are left with

$$
\nabla^{2} \mathbf{H}-\mu \sigma \frac{\partial \mathbf{H}}{\partial \mathrm{t}}-\mu \varepsilon \frac{\partial^{2} \mathbf{H}}{\partial \mathrm{t}^{2}}=0
$$

Thus the wave equation for the magnetic field intensity has the same form as the electric field intensity E as given in (5.8)

The general wave equation that we have written in (5.8) contains terms that make it difficult to solve. We will suggest several simplifications. The first simplification that we will make concerns the material in which the wave is to propagate. We will initially investigate a vacuum. In a vacuum, the second term in (5.8) is set equal to zero since the conduction current $\mathbf{J}=\sigma \mathbf{E}$ is nonexistent. The relative permittivity $\varepsilon_{\mathrm{r}}$ and the relative permeability $\mu_{\mathrm{r}}$ are both one and $\varepsilon$ and $\mu$ can be replaced with their vacuum values $\varepsilon_{0}$ and $\mu_{0}$.

The second simplification that we can make is to choose a Cartesian coordinate system with the electric field polarized in only one direction. We will

## Wave equation

choose that direction for the polarization of the electric field to be specified by the unit vector $\mathbf{u}_{\mathbf{y}}$. The third simplification that we will make is to assume that the wave is a function of only one of the three variables that make up the Cartesian coordinate system, say the $z$ coordinate. This specifies that $\frac{\partial}{\partial x}=\frac{\partial}{\partial y}=0$. The experimental scenario depicted in Figure 5-1 indicates how this could be effected. In this case, the vector wave equation (5.8) reduces to

$$
\begin{equation*}
\frac{\partial^{2} E_{y}}{\partial z^{2}}-\mu_{0} \varepsilon_{0} \frac{\partial^{2} E_{y}}{\partial t^{2}}=0 \tag{5.9}
\end{equation*}
$$

where the unit vector $\mathbf{u}_{\mathbf{y}}$ is common to all terms in the equation. The unit vector $\mathbf{u}_{\mathbf{y}}$ will not be written but it will be understood that the electric field intensity is polarized in that particular direction. Hence, (5.9) becomes a one dimensional scalar wave equation. We note the appearance of the term $\left(\mu_{0} \varepsilon_{0}\right)$ that appears in this equation. Let us understand the meaning of this term. The dimensions of the two terms in (5.9) that involve the derivatives are respectively given by

$$
\frac{\left(\frac{\text { volts }}{\text { meter }}\right)}{(\text { meter })^{2}} \text { and } \frac{\left(\frac{\text { volts }}{\text { meter }}\right)}{(\text { sec ond })^{2}}
$$

In order for this equation to be correct dimensionally, the term $\left(\mu_{0} \varepsilon_{0}\right)$ must have the units of

$$
\left(\frac{\text { sec ond }}{\text { meter }}\right)^{2}
$$

Hence $\left(\mu_{0} \varepsilon_{0}\right)$ has the units of (velocity) ${ }^{-2}$. If we insert the numerical values for $\varepsilon_{0}$ and $\mu_{0}$ that were obtained previously and solve for this velocity, we find that it has a numerical value that has the same value as the velocity of light. Therefore, we feel comfortable with the substitution of a symbol $c$ where $c=\frac{1}{\sqrt{\mu_{0} \varepsilon_{0}}}$ into (5.9)

## Wave equation

without yet knowing the true meaning behind it. Such good fortune in this term having the correct dimensions and a recognizable numerical value is not just pure happenstance but it is based on a firm theoretical foundation as will be shown below. The product of the velocity times the time, ct, has the dimensions of distance. Hence, the vector wave equation written in (5.8) could be thought of in terms of a four dimensional space, a concept that has occupied the time and energy of certain theoreticians at the graduate level and beyond.

Example 5-2. Compute the numerical value for the velocity c .
Answer: Using the numerical values for $\mu_{0}$ and $\varepsilon_{0}$, we write

$$
c=\frac{1}{\sqrt{\mu_{0} \varepsilon_{0}}} \approx \frac{1}{\sqrt{\left(4 \pi \times 10^{-7}\right)\left(\frac{1}{36 \pi} \times 10^{-9}\right)}} \approx 3 \times 10^{8} \mathrm{~m} / \mathrm{s}
$$

This is the speed of light!

## One-dimensional wave equation

### 5.2. One-dimensional wave equation

In order to emphasize and understand some basic properties of waves, let us first examine waves in other disciplines. This slight diversion from our main task is to breed familiarity with the topic of waves using experiences that you may have encountered in the past. We will look at a pulse that travels in the direction of increasing positive values of the coordinate $z$ in a series of gedanken experiments.

### 5.2.1. Related wave experiments

The first experiment could be performed in a water tank or in a bathtub. A repetitive wave-pulse is launched at a point labeled as $z=0$ with a signal generator attached to a small plunger ${ }^{1}$. This plunger can move up and down as shown in Figure 5-2. The repetition frequency of the plunger motion is slow enough so the excited waves will not interfere with each other. The waves are also absorbed at the walls and at the end of the tank so no reflection of the wave will occur. The water wave propagates with a very slow velocity compared with the velocity of light. Every time that the pulse is launched at $z=0$, a trigger pulse is simultaneously sent from the signal generator to the oscilloscope. The trigger pulse will propagate at the velocity of light so this trigger signal can be considered in the time scale of the water wave propagation to be "instantaneous." The pulse is detected with a calibrated movable probe and the response is displayed on an oscilloscope that is triggered from the signal generator. This detector could be a device such as a photomultiplier that responds to the amplitude of the reflected light from the water. If the water is uniformly illuminated, the change of curvature due to the passage of the propagating wave would alter the detected signal.

[^11]
## One-dimensional wave equation

Pictures are taken from the oscilloscope at various locations $z$ in the water tank and a sequence of these photographs taken at equal spatial intervals is shown in Figure 5-3 (a). In this sequence, the $t=0$ trigger time is lined up along one axis. Note that the pulse moves to the right with what appears to be a constant velocity.


Figure 5-2. Wave-making experiment for water waves. The amplitude of the wave is $\Delta \varphi$.

From the sequence of photographs shown in Figure 5-3a and knowing the locations where they were taken, it is possible to obtain two numbers. These are the distance of propagation $z$ and the time-of-flight $t$ of a constant point on the pulse, say the rising edge of the pulse. This set of numbers is plotted on a graph that is shown in Figure 5-3b. The experimental points appear to lie on a straight line. The slope of this line is called the velocity of propagation and this figure is called the trajectory of the wave propagation. This velocity of propagation for these surface waves is a function of the surface tension and mass density of the water. There are cases where this trajectory may not be a straight line and one will have to calculate the derivative $v=\frac{\partial \mathrm{z}}{\partial \mathrm{t}}$ in order to define the slope and hence, the local velocity of propagation. Our use of the partial derivative notation in this

## One-dimensional wave equation

definition is because there are cases where the velocity may also depend upon other quantities such as the frequency of the wave (dispersion), the amplitude of the wave (nonlinear), etc.

(a)

(b)

Figure 5-3. (a) Sequence of oscilloscope photographs taken at various locations in the water tank. (b) Trajectory of the rising edge of the wave pulse. The velocity of propagation which is determined from the slope of the line is equal to 2 .


Figure 5-4. A wave can be launched on a string by "plucking" one end and it will propagate to either end.

A second wave experiment uses a string that is stretched between two points as shown in Figure 5-4. In this case, a small perturbation is launched on

## One-dimensional wave equation

one end of the string and it propagates to the other end. As in the water tank experiment, we will neglect any reflection at the end. If a camera were available, we could take pictures of the perturbation as it moved along the string. From this sequence of pictures, the trajectory could be drawn and the velocity of the wave could be computed as shown in Figure 5-3b. In this case, the velocity of propagation is a function of the tension on the string and the mass density of the string as is demonstrated below in Example 5-3. In the case were the diameter of the string decreases as the distance increases, one would quickly encounter a nonlinear velocity of propagation. A whip makes use of this property.

The third experiment employs a spring that is stretched between two walls. One of the walls is suddenly moved resulting in a perturbation in the spring that propagates to the other end as shown in Figure 5-5. Once again, the trajectory can be drawn as shown in Figure 5-3b and the velocity of propagation computed. This velocity will depend on the elasticity and the mass density of the spring.


Figure 5-5. A sudden compression-rarefaction in the spring causes a perturbation whose amplitude is $\Delta \varphi$ to propagate on the spring.

In all three of these examples, certain general conclusions can be drawn about the nature of wave propagation. The host medium does not propagate. Only the perturbation propagates and it propagates with a definite velocity that is

## One-dimensional wave equation

determined by the properties of the medium in which it propagates. After the perturbation passes, the host medium returns to its original unperturbed state since we are assuming that the perturbations are small. The host medium is not affected by the passage of the wave - no energy remains at a spot in the medium. There are no local hot spots caused by a large amplitude wave locally heating the medium. Except for a later comment, we will not further examine this important class of wave propagation problems, which involve an energy transfer from one form to another where it is either locally absorbed or dissipated.

There is a major difference between these three gedanken experiments. In the first two cases, the perturbation $\Delta \varphi$ was transverse to the direction of the wave's propagation. In the third case, the perturbation $\Delta \varphi$ was in the direction of propagation. The first two cases are classified as being a transverse wave and the third case is classified as being a longitudinal wave. The electromagnetic waves that we will study in this text are transverse waves. Sound waves in which there are sequences of local compressions and rarefactions that propagate in the air and are launched by an instructor in class are longitudinal waves. ${ }^{2}$ Transverse electromagnetic waves can propagate in a vacuum, longitudinal sound waves require something to be compressed and therefore cannot propagate in a vacuum.

Example 5-3. As an example of a wave that propagates in a nonelectromagnetic milieu, derive the wave equation for the transverse waves that propagate along a string. The string has a mass density $\rho_{\mathrm{L}}$ and the string is under a tension force $T$. Neglect the weight of the string.

Answer: A segment $\Delta z$ of the string that suffers a displacement $\Delta \varphi$ normal to the string is shown in the figure. The tension $T$ is constant along the string. If the

[^12]
## One-dimensional wave equation

tension were not a constant, the string would break. For small amplitudes $\Delta \varphi$, we can write

$$
\sin \theta \approx \tan \theta \approx \frac{\partial(\Delta \varphi)}{\partial z}
$$

Hence the vertical force $F$ on the string can be written as

$$
\mathrm{F}=\mathrm{T} \sin \theta_{2}-\mathrm{T} \sin \theta_{1}=\mathrm{t}\left\{\left.\frac{\partial(\Delta \varphi)}{\partial z}\right|_{\mathrm{z}+\Delta \mathrm{z}}-\left.\frac{\partial(\Delta \varphi)}{\partial \mathrm{z}}\right|_{\mathrm{z}}\right\} \approx \mathrm{T} \Delta \mathrm{z} \frac{\partial^{2}(\Delta \varphi)}{\partial \mathrm{z}^{2}}
$$

Newton's equation of motion states that

$$
\mathrm{F}=\mathrm{m} \frac{\partial^{2}(\Delta \varphi)}{\partial \mathrm{t}^{2}}=\rho_{\mathrm{L}} \Delta \mathrm{z} \frac{\partial^{2}(\Delta \varphi)}{\partial \mathrm{t}^{2}}
$$

Equating these two expressions, we obtain the wave equation

$$
\frac{\mathrm{T}}{\rho_{\mathrm{L}}} \frac{\partial^{2}(\Delta \varphi)}{\partial z^{2}}-\frac{\partial^{2}(\Delta \varphi)}{\partial \mathrm{t}^{2}}=0
$$

The velocity of the wave along the string is given by $v=\sqrt{\frac{T}{\rho_{L}}}$. Waves are omnipresent in the universe and the example of the waves that propagate along the string is an easily visualized example of a transverse wave. The electromagnetic waves that we encounter will also be transverse waves.

We could have derived a wave equation and determined the velocity of propagation for all three of the waves depicted above. This was done here only for the transverse wave that propagates along the string in Example 5-3. Suffice it to say that a standard wave equation for the amplitude $\Delta \varphi$ would result in all cases and the wave would have its own unique velocity of propagation. ${ }^{3}$ It would have the same mathematical form as the wave equation for the electromagnetic waves as will be shown in the following discussion. Experiments of the type indicated

[^13]
## One-dimensional wave equation

here for other waves can also be performed with electromagnetic waves and the velocity of propagation can be appropriately measured.

### 5.2.2. Analytical solution of one-dimensional equation - traveling waves

Equation (5.9) can now be rewritten as

$$
\begin{equation*}
\frac{\partial^{2} \mathbf{E}_{y}}{\partial \mathbf{z}^{2}}-\frac{1}{c^{2}} \frac{\partial^{2} \mathbf{E}_{y}}{\partial \mathbf{t}^{2}}=0 \tag{5.10}
\end{equation*}
$$

The most general nontrivial solution ${ }^{4}$ of this equation is given by

$$
\begin{equation*}
E_{y}(z, t)=F(z-c t)+G(z+c t) \tag{5.11}
\end{equation*}
$$

where $F$ and $G$ are arbitrary functions that are determined by the excitation. This can be checked by a substitution of (5.11) into (5.10). This excitation could be a pulse, a step function, a continuous time-harmonic wave, or any other function that is specified by, say, a function generator. The solution $F(z-c t)$ is a traveling wave in the $+z$ direction, while the solution $G(z+c t)$ is a traveling wave in the $-z$ direction. There are two possible scenarios that would create both waves simultaneously. The first assumes that there are two function generators, one at $z=-\infty$ and the other at $z=+\infty$. The second assumes that there is a source at $z=-\infty$ and there is a reflecting boundary at some location, say $z=0$. In this case, there will be an incident wave and a reflected wave. Reflection will be discussed later.

In order to show that general solution presented in (5.11) is indeed the mathematical solution for the scalar wave equation (5.10), we need only substitute this solution into this partial differential equation and show that it does indeed satisfy this equation. Let us do this operation very methodically since this is a crucial

[^14]
## One-dimensional wave equation

point in our argument. It is easiest to first define two new independent variables $\varsigma$ . and $\Psi$ that include the independent variables of space $z$ and time $t$ in a particular format as

$$
\begin{equation*}
\varsigma=\mathbf{z}-\mathrm{ct} \text { and } \psi=\mathbf{z}+\mathrm{ct} \tag{5.12}
\end{equation*}
$$

and then use the chain rule for differentiation.
Therefore, we write that the first derivative of our general solution can be written as

$$
\begin{aligned}
& \frac{\partial \mathrm{E}_{\mathrm{y}}}{\partial \mathrm{t}}=\frac{\partial \mathrm{E}_{\mathrm{y}}}{\partial \varsigma}\left(\frac{\partial \varsigma}{\partial t}\right)+\frac{\partial \mathrm{E}_{\mathrm{y}}}{\partial \psi}\left(\frac{\partial \psi}{\partial \mathrm{t}}\right)=\frac{\mathrm{dF}}{\mathrm{~d} \varsigma}(-\mathrm{c})+\frac{\mathrm{dG}}{\mathrm{~d} \psi}(\mathrm{c}) \\
& \frac{\partial \mathrm{E}_{\mathrm{y}}}{\partial \mathrm{z}}=\frac{\partial \mathrm{E}_{\mathrm{y}}}{\partial \varsigma}\left(\frac{\partial \varsigma}{\partial z}\right)+\frac{\partial \mathrm{E}_{\mathrm{y}}}{\partial \psi}\left(\frac{\partial \psi}{\partial z}\right)=\frac{\mathrm{dF}}{\mathrm{~d}}(1)+\frac{\mathrm{dG}}{\mathrm{~d} \psi}(1)
\end{aligned}
$$

where the last derivatives are computed from (5.12).
The second derivatives lead to

$$
\begin{aligned}
& -\frac{1}{c^{2}} \frac{\partial^{2} E_{y}}{\partial t^{2}}=-\frac{1}{c^{2}}\left[\frac{d^{2} F}{d \varsigma^{2}}(-c)^{2}+\frac{d^{2} G}{d \psi^{2}}(c)^{2}\right] \\
& \frac{\partial^{2} E_{y}}{\partial z^{2}}=\frac{d^{2} F}{d \varsigma^{2}}(1)^{2}+\frac{d^{2} G}{d \psi^{2}}(1)^{2}
\end{aligned}
$$

Hence, we have verified that (5.11) is indeed the most general solution of (5.10) since the functions $F(\varsigma)=F(z-c t)$ and $G(\psi)=G(z+c t)$ are completely arbitrary functions. This could be a pulse or a sine wave that depends only on the excitation. In order to actually draw the functions, numerical values must be specified for the arguments of the functions. Now we will invoke one of the truisms of life: "We are all getting older!" This truth states that time is always increasing. It manifests itself in the following way as far as the solution of (5.10) is concerned. The variables $\varsigma$ and $\psi$ were chosen to have particular numerical values, say $\varsigma_{0}$ and $\Psi_{0}$, where

$$
\varsigma_{0}=\mathbf{z}_{0}-\mathrm{ct}_{0} \text { and } \psi_{0}=\mathrm{z}_{0}+\mathrm{ct}_{0}
$$

## One-dimensional wave equation



Figure 5-6. Particular solutions of the one-dimensional wave equation at two times. Each pulse has a particular shape that is determined by the function generator.

We assume that the functions $\mathrm{F}(\varsigma)$ and $\mathrm{G}(\psi)$ at a time $\mathrm{t}=\mathrm{t}_{0}$ are located at $\mathrm{z}=$ $-z_{0}$ and at $z=+z_{0}$ respectively. The independent variables $\varsigma$ and $\psi$ must also have the same numerical values in order that the shape of the signal be unaltered at a later time $t>t_{0}$. This implies that the location $z$ of the pulse must correspondingly change. The signal $F(\varsigma)=F(z-c t)$ will pass to increasing values of $z$. The signal $\mathrm{G}(\psi)=\mathrm{G}(\mathrm{z}+\mathrm{ct})$ will pass to decreasing values of z . This is shown in Figure 5-6 for two pulses of particular shape at two values of time.

We call this effect wave propagation or just propagation. Assuming that there is no distortion in the wave, we can follow the same point of the perturbation as it propagates. This implies that $\mathrm{d} \zeta=0$ and $\mathrm{d} \psi=0$. The velocity of propagation of the wave is then computed from the relations

$$
\begin{equation*}
\mathrm{d} \varsigma \equiv \mathrm{dz}-\mathrm{cdt}=0 \text { and } \quad \mathrm{d} \psi \equiv \mathrm{dz}+\mathrm{cdt}=0 \tag{5.13}
\end{equation*}
$$

from which we determine the velocity of propagation of the two functions to be

$$
\begin{equation*}
\frac{\mathrm{dz}}{\mathrm{dt}}=\mathrm{c} \text { and } \frac{\mathrm{dz}}{\mathrm{dt}}=-\mathrm{c} \tag{5.14}
\end{equation*}
$$

## One-dimensional wave equation

The sign difference indicates that $F(z-c t)$ propagates to increasing values of the coordinate $z$ and $G(z+c t)$ propagates to decreasing values of the coordinate $z$.

Having just solved for the electric field, we could compute the magnetic field intensity H from Maxwell's equations. Rather than do this now, we will defer this computation for a little while. Rest assured, both components are required for the propagation of an electromagnetic wave.

In order to obtain a particular solution of (5.10) than the general solution (5.11) we will add two more initial conditions:

1) We will assume that the solution is a known function $P(z)$ at the time $t=0$
2) We will assume that the derivative of the solution is a known function $Q(z)$ at the time $t=0$

$$
\left.\begin{array}{l}
E_{y}(z, 0)=P(z)  \tag{5.15}\\
\frac{\partial E_{y}}{\partial t}(z, 0)=Q(z)
\end{array}\right\}
$$

It will be shown below that the particular solution of this problem is presented by the following expression

$$
\begin{align*}
E_{y}(z, t) & =\frac{1}{2}[P(z-c t)+P(z+c t)]+\frac{1}{2 c} \int_{z-c t}^{z+c t} Q(\xi) d \xi  \tag{5.16}\\
& =\frac{1}{2}[P(z-c t)+P(z+c t)]+\frac{1}{2}[R(z+c t)-R(z-c t)]
\end{align*}
$$

where the new auxiliary function $R(z)=\frac{1}{c} \int_{0}^{z} Q\left(z^{\prime}\right) d z^{\prime}$ is introduced.
Example 5-4. Check that (5.16) is the particular solution of (5.10) satisfying the initial conditions (5.15).

Answer: Here the method of D'Alambert is applied. We assume the general solution is given by (5.11). After taking the derivative with the respect to the variable $t$ and evaluating the result at the time $t=0$, the following system of two functional equations is obtained

## One-dimensional wave equation

$$
\begin{aligned}
& F(z)+G(z)=P(z) \\
& c\left(-\frac{d F}{d z}+\frac{d G}{d z}\right)=Q(z)
\end{aligned}
$$

where $\varsigma=\psi=z$ at the time $t=0$ is used. The integration of the second equation yields

$$
\begin{aligned}
& F(z)+G(z)=P(z) \\
& -F(z)+G(z)=\frac{1}{c} \int_{0}^{z} Q\left(z^{\prime}\right) d z^{\prime}
\end{aligned}
$$

The addition and subtraction of these two equations leads to the following set of simultaneous equations

$$
\begin{aligned}
& F(z)=\frac{1}{2} P(z)-\frac{1}{2 c} \int_{0}^{z} Q(\xi) d \xi \\
& G(z)=\frac{1}{2} P(z)+\frac{1}{2 c} \int_{0}^{z} Q(\xi) d \xi
\end{aligned}
$$

The substitution of both functions into (5.11) along with appropriate replacements of the variables $z \rightarrow \varsigma=\mathbf{z}-\mathrm{ct}$ in the first function and $\mathbf{z} \rightarrow \psi=\mathbf{z + c t}$ in the second function yields the following particular solution for the initial value problem

$$
\begin{aligned}
E_{y}(z, t) & =\frac{1}{2}[P(z-c t)+P(z+c t)]+\frac{1}{2 c} \int_{z-c t}^{z+c t} Q(\xi) d \xi \\
& =\frac{1}{2}[P(z-c t)+P(z+c t)]+\frac{1}{2}[R(z+c t)-R(z-c t)]
\end{aligned}
$$

This is (5.16). Provided the two functions $P(z)$ and $Q(z)$ are given explicitly, it is not difficult to obtain the solution of this equation.

Example 5-5. Show that the function $F(z-c t)=F_{0} \exp \left(-(z-c t)^{2}\right)$ is a solution of the wave equation (5-10). This is a so called Gaussian pulse traveling wave. In addition, relate the solution to Example 5-4.
Answer: Let $\varsigma=z-c t$. Therefore $F(z-c t)=F_{0} \exp \left(-\varsigma^{2}\right)$ and $G(z+c t) \equiv 0$. Then, we write using the chain rule

$$
\frac{\partial \mathrm{F}}{\partial \mathrm{z}}=\frac{\mathrm{dF}}{\mathrm{~d} \varsigma} \frac{\partial \varsigma}{\partial z}=\mathrm{F}_{0}\left\{-2 \varsigma \exp \left[-\varsigma^{2}\right]\right\}(1)
$$

## One-dimensional wave equation

$$
\begin{gathered}
\frac{\partial^{2} F}{\partial z^{2}}=\frac{d^{2} F}{d \varsigma^{2}}\left(\frac{\partial \varsigma}{\partial z}\right)^{2}=F_{0}\left\{\left(-2+4 \varsigma^{2}\right) \exp \left[-\varsigma^{2}\right]\right\}(1)^{2} \\
\frac{\partial F}{\partial t}=\frac{d F}{d \varsigma} \frac{\partial \varsigma}{\partial t}=F_{0}\left\{-2 \varsigma \exp \left[-\varsigma^{2}\right]\right\}(-c) \\
-\frac{1}{c^{2}} \frac{\partial^{2} F}{\partial t^{2}}=\frac{d^{2} F}{d \varsigma^{2}}\left(\frac{\partial \varsigma}{\partial t}\right)^{2}=-\frac{1}{c^{2}} F_{0}\left\{\left(-2+4 \varsigma^{2}\right) \exp \left[-\varsigma^{2}\right]\right\}(-c)^{2}
\end{gathered}
$$

Adding the second and the fourth equations yields (5.10). A sequence of pulses taken at increasing times illustrates the propagation of the pulses. The velocity of propagation is $c$.


We can define the propagation of the Gaussian pulse as an initial value problem that was discussed in Example 5-4. In this case,

$$
P(z)=\left.F(\varsigma)\right|_{t=0}=F_{0} \exp \left(-z^{2}\right)
$$

Now

$$
\begin{gathered}
\frac{\partial \mathrm{F}}{\partial \mathrm{t}}=\frac{\mathrm{dF}}{\mathrm{~d} \varsigma} \frac{\partial \varsigma}{\partial \mathrm{t}}=\mathrm{F}_{0}\left\{2 \mathrm{c} \varsigma \exp \left[-\varsigma^{2}\right]\right\} \\
\mathrm{Q}(\mathrm{z})=\left.\frac{\partial \mathrm{F}}{\partial \mathrm{t}}(\varsigma)\right|_{\mathrm{t}=0}=2 \mathrm{c} F_{0} \mathrm{zexp}\left(-\mathrm{z}^{2}\right)
\end{gathered}
$$

A simple integration yields the auxiliary function

$$
\mathrm{R}(\mathrm{z})=\frac{1}{\mathrm{c}} \int_{0}^{\mathrm{z}} \mathrm{Q}(\xi) \mathrm{d} \xi=\frac{1}{\mathrm{c}} 2 \mathrm{c} \mathrm{~F}_{0} \int_{0}^{\mathrm{z}} \xi \exp \left(-\xi^{2}\right) \mathrm{d} \xi=\mathrm{F}_{0}\left[1-\exp \left(-\mathrm{z}^{2}\right)\right]=\mathrm{F}_{0}-\mathrm{F}(\mathrm{z})
$$

After substituting this result into (5.16), we obtain

## One-dimensional wave equation

$E_{y}(z, t)=\frac{1}{2} F(z-c t)+\frac{1}{2} F(z+c t)+\frac{1}{2}\left[F_{0}-F(z+c t)\right]-\frac{1}{2}\left[F_{0}-F(z-c t)\right]=F(z-c t)$
which is a traveling Gaussian wave towards (+z) with a speed c.

### 5.2.3. MATLAB solution of one-dimensional equation - finite difference in time-domain method

The numerical solution of the wave equation is a formidable task. One quickly encounters numerical difficulties that are beyond the scope of this text. Fortunately for us, we can be carried on the shoulders of giants in our first encounter with these potential pitfalls. Herein, we will introduce a path through this jungle and develop a numerical program that is written in MATLAB. The resulting figures should aid our understanding of wave phenomena.

In order to develop a numerical solution for the one-dimensional wave equation (5.10), we initially solve a first order partial differential equation. This equation is sometimes called the advection equation

$$
\begin{equation*}
\frac{\partial \varphi}{\partial z}+\frac{1}{c} \frac{\partial \varphi}{\partial t}=0 \tag{5.17}
\end{equation*}
$$

For the initial condition

$$
\begin{equation*}
\varphi(z, t=0)=F(z) \tag{5.18}
\end{equation*}
$$

the analytical solution of the advection equation is given by

$$
\begin{equation*}
\varphi(z, t)=F(z-c t) \tag{5.19}
\end{equation*}
$$

which is the solution of the wave equation (5.10). This can be easily checked by a substitution of (5.19) into (5.17).

Both the wave equation and the advection equation belong to the same family of equations that are called hyperbolic equations. The diffusion equation is in the parabolic equation family and Laplace's and Poisson's equations are in the elliptic equation family. We will focus our attention here on the advection equation

## One-dimensional wave equation

as it is simpler and the procedure and some of the pitfalls along with the bridges that cross these pitfalls will be described.

As shown in Figure 5-7, we consider that the space $z$ and time $t$ can be drawn in a three dimensional figure. The amplitude $\varphi$ of the wave is specified by the third coordinate. In Figure 5-7, we set up a numerical grid. First, we have broken the region $L$ in which the wave propagates into $N$ sections. In the figure, we have chosen $N=4$. Hence we write


Figure 5-7. Numerical grid that uses periodic boundary conditions.

$$
\begin{equation*}
h \equiv \frac{L}{N} \tag{5.20}
\end{equation*}
$$

We assume that the velocity of propagation is $c$ and that it takes a time $t$ for the wave to propagate a distance $h$. Therefore

$$
\begin{equation*}
\mathrm{h}=\mathrm{c} \tau \tag{5.21}
\end{equation*}
$$

With these restrictions, we will jump over numerical stability reservations that were originally noted by Courant-Fredrichs-Lewy (CFL). We will leave it as exercises to examine the cases where $\mathrm{h} \neq \mathrm{c} \tau$.

## One-dimensional wave equation

In addition to stability restrictions, we have also invoked periodic boundary conditions. This states that once a numerically calculated wave reaches the boundary at $z=+L / 2$, it reappears at the same time at the boundary $z=-L / 2$ and continues to propagate in the region $-\mathrm{L} / 2 \leq \mathrm{z} \leq+\mathrm{L} / 2$. As shown in Figure 5-7, we actually do not evaluate the wave at these two edges but at one-half of a spatial increment $h / 2$ removed from them at $z=-L / 2+h / 2$ and at $z=+L / 2-h / 2$.

Let us now convert the advection equation (5.17) to the finite difference form that can be handled by the computer. The time derivative is replaced using the forward difference method that was introduced in Chapter 3

$$
\begin{equation*}
\frac{\partial \varphi}{\partial \mathrm{t}} \Rightarrow \frac{\varphi\left(\mathrm{z}_{\mathrm{i}}, \mathrm{t}_{\mathrm{n}}+\tau\right)-\varphi\left(\mathrm{z}_{\mathrm{i}}, \mathrm{t}_{\mathrm{n}}\right)}{\tau} \tag{5.22}
\end{equation*}
$$

In this notation with reference to Figure 5-7, we have

$$
\begin{equation*}
z_{i}=(i-1 / 2) h-L / 2 \text { and } t_{n}=(n-1) \tau \tag{5.23}
\end{equation*}
$$

The space derivative is replaced using the central difference method

$$
\begin{equation*}
\frac{\partial \varphi}{\partial z} \Rightarrow \frac{\varphi\left(z_{i}+h, t_{n}\right)-\varphi\left(z_{i}-h, t_{n}\right)}{2 h} \tag{5.24}
\end{equation*}
$$

Substitute (5.22) and (5.24) into the advection equation (5.17) and obtain

$$
\begin{equation*}
\frac{\varphi\left(z_{i}+h, t_{n}\right)-\varphi\left(z_{i}-h, t_{n}\right)}{2 h}+\frac{1}{c} \frac{\varphi\left(z_{i}, t_{n}+\tau\right)-\varphi\left(z_{i}, t_{n}\right)}{\tau}=0 \tag{5.25}
\end{equation*}
$$

In (5.25), three of the four terms are evaluated at the same time $t_{n}$ and one term is evaluated at the next increment in time $t_{n}+\tau$. From (5.25), we write this term as

$$
\begin{equation*}
\varphi\left(z_{i}, t_{n}+\tau\right)=\varphi\left(z_{i}, t_{n}\right)-\frac{\mathrm{c} \tau}{2 h}\left[\varphi\left(z_{i}+h, t_{n}\right)-\varphi\left(z_{i}-h, t_{n}\right)\right] \tag{5.26}
\end{equation*}
$$

The finite difference method based on this equation is called finite difference in time domain (FD TD) method. This is valid in the interior range $2 \leq \mathrm{n} \leq \mathrm{N}-1$. In (5.26), we note that all values are known initially at the time $t_{0}=0$. Hence, we use (5.26) to evaluate the values at the next increment in time and is also called a leapfrog scheme. With the imposition of periodic boundary conditions, we must

## One-dimensional wave equation

carefully use (5.26) in order to find the values at the boundaries. This manifests itself with the requirement that

$$
\begin{align*}
& \varphi\left(z_{1}, t_{n}+\tau\right)=\varphi\left(z_{1}, t_{n}\right)-\frac{c \tau}{2 h}\left[\varphi\left(z_{2}, t_{n}\right)-\varphi\left(z_{N}, t_{n}\right)\right] \\
& \varphi\left(z_{N}, t_{n}+\tau\right)=\varphi\left(z_{N}, t_{n}\right)-\frac{c \tau}{2 h}\left[\varphi\left(z_{1}, t_{n}\right)-\varphi\left(z_{N-1}, t_{n}\right)\right] \tag{5.27}
\end{align*}
$$

which makes the iteration process consistent.
Example 5-6 MATLAB. Use (5.26) and (5.27) to find the evolution of a rectangular pulse whose initial shape is defined by

$$
\left.\begin{array}{l}
\varphi\left(-\frac{h}{2} \leq z \leq \frac{h}{2}, t=0\right)=1 \\
\varphi\left(|z|>\frac{h}{2}, t=0\right)=0
\end{array}\right\}
$$

Use the grid depicted in Figure 5-7. The stability requirement $h=c \tau$ is also to be invoked in this calculation.

Answer: We take a grid with small number of points $\mathrm{N}=4$. The number of the calculated time steps is taken to be nstep=5. We tabulate the computed values to be

|  |  | 0 | T | 2 т | 3 т | 4 т | 5 т |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | -3L/8 | 0 | -1/2 | -1/2 | 1/2 | 5/2 | 9/2 |
| Z | -L/8 | 1 | 1/2 | -1/2 | -3/2 | -3/2 | 1/2 |
|  | L/8 | 1 | 3/2 | 3/2 | 1/2 | -3/2 | -7/2 |
|  | 3L/8 | 0 | $1 / 2$ | 3/2 | 5/2 | 5/2 | 1/2 |

Note that the signal becomes distorted and increases in value as it propagates. It is unstable for every value of time! Our imposition of the stability requirement $\mathrm{h}=\mathrm{ct}$ did not insure stability in this case!

## One-dimensional wave equation

Fortunately for us, there is a simple solution to the instability problem that is in Example 5-6. This is the Lax method. It replaces (5.26) with a slightly different iteration equation

$$
\begin{equation*}
\varphi\left(z_{i}, t_{n}+\tau\right)=\frac{1}{2}\left[\varphi\left(z_{i}+h, t_{n}\right)+\varphi\left(z_{i}-h, t_{n}\right)\right]-\frac{c \tau}{2 h}\left[\varphi\left(z_{i}+h, t_{n}\right)-\varphi\left(z_{i}-h, t_{n}\right)\right] \tag{5.28}
\end{equation*}
$$

The first term on the right side is the average of the two neighboring terms. Similarly, the two equations that represent the periodic boundary conditions are modified to

$$
\begin{align*}
& \varphi\left(z_{1}, t_{n}+\tau\right)=\frac{1}{2}\left[\varphi\left(z_{2}, t_{n}\right)+\varphi\left(z_{N}, t_{n}\right)\right]-\frac{c \tau}{2 h}\left[\varphi\left(z_{2}, t_{n}\right)-\varphi\left(z_{N}, t_{n}\right)\right]  \tag{5.29}\\
& \varphi\left(z_{N}, t_{n}+\tau\right)=\frac{1}{2}\left[\varphi\left(z_{1}, t_{n}\right)+\varphi\left(z_{N-1}, t_{n}\right)\right]-\frac{C \tau}{2 h}\left[\varphi\left(z_{1}, t_{n}\right)-\varphi\left(z_{N-1}, t_{n}\right)\right]
\end{align*}
$$

Example 5-7 MATLAB . Repeat Example 5-6 using the Lax's method and sketch the solution with $\mathrm{h}=\mathrm{c} \tau$ again.

Answer: In MATLAB language, we write (5.28) and (5.29) in three steps. The first step finds the new interior values of j in terms of the previous interior values. These iterations are specified to have iterations in the range: $2 \leq \mathrm{i} \leq(\mathrm{N}-1)$. The remaining two steps take care of the periodic boundary conditions at $i=1$ and at $\mathrm{i}=\mathrm{N}$. The results for the case of $\mathrm{N}=4$ for the first five time steps are given in the table below.

|  |  | 0 | T | 2 т | 3 т | 4 т | 5 T |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | -3L/8 | 0 | 0 | 1 | 1 | 0 | 0 |
| z | -L/8 | 1 | 0 | 0 | 1 | 1 | 0 |
|  | L/8 | 1 | 1 | 0 | 0 | 1 | 1 |
|  | 3L/8 | 0 | 1 | 1 | 0 | 0 | 1 |

## One-dimensional wave equation

Note that in this case, we have stability. In addition, the pulse is not distorted as it propagates. We plot the solution below.


Example 5-8 MATLAB. Calculate the temporal and spatial evolution of a narrow pulse. Use 50 grid points and 50 time steps.

Answer. The solution is shown in the following two figures. The initial pulse and the final pulse are shown in (a) and the propagation is shown in (b).

## One-dimensional wave equation



## Time-harmonic plane waves

### 5.3. Time-harmonic plane waves <br> 5.3.1. Plane waves in vacuum

Let us now derive the electric field for a wave that is excited with a sinusoidal excitation using the one-dimensional wave equation (5.10). Since the excitation mechanism was a time harmonic signal, we should expect that the propagating wave should you also be a time harmonic propagating wave. Therefore, the field quantities $E_{y}(z, t)$ will be of the form

$$
\begin{equation*}
E_{y}(z, t)=E_{y}(z) e^{j o t}=E_{y} e^{j o t} \tag{5.30}
\end{equation*}
$$

where the $z$ dependence will be the dependent variable and will not be explicitly stated. The notation $E_{y}$ indicates that this term is a phasor quantity. In a vacuum, the phase velocity of the propagating wave is equal to the velocity of light $c$. With the assumption that is stated in (5.30), we write (5.10) as

$$
\frac{\partial^{2} E_{y}}{\partial z^{2}}-\frac{1}{c^{2}} \frac{\partial^{2} E_{y}}{\partial t^{2}}=\left[\frac{d^{2} E_{y}}{d z^{2}}-\frac{(\mathrm{j} \omega)^{2}}{c^{2}} E_{y}\right] \mathrm{e}^{\text {jott }}=0
$$

or

$$
\begin{equation*}
\frac{\mathrm{d}^{2} E_{y}}{\mathrm{dz}}+\left(\frac{\omega}{\mathrm{c}}\right)^{2} E_{y}=0 \tag{5.31}
\end{equation*}
$$

where the exponential term $\mathrm{e}^{\text {jot }}$ has been removed. The ratio $(\omega / \mathrm{c})$ has the di-
mensions of (meter) ${ }^{-1}$. This ratio is called the wave number and it is usually given
the symbol $k$.
The terms listed above are all interrelated as shown below where several alternative definitions are presented

$$
\begin{equation*}
\mathrm{k}=\frac{\omega}{\mathrm{c}}=\frac{2 \pi \mathrm{f}}{\mathrm{c}}=\frac{2 \pi}{\lambda} \tag{5.32}
\end{equation*}
$$

Equation (5.31) can be finally written as

$$
\begin{equation*}
\frac{\mathrm{d}^{2} E_{y}}{\mathrm{dz}^{2}}+\mathrm{k}^{2} E_{y}=0 \tag{5.33}
\end{equation*}
$$

## Time-harmonic plane waves

Equation (5.33) is also given the name of a one-dimensional Helmholtz equation.
A solution for the second order ordinary differential equation (5.33) is written as

$$
\begin{equation*}
E_{y}=a e^{-\mathrm{jkz}}+b \mathrm{e}^{-\mathrm{jkz}} \tag{5.34}
\end{equation*}
$$

where we have let the two constants of integration $a$ and $b$ be real quantities. A substitution into (5.30) yields

$$
\begin{equation*}
E_{y}=a e^{j(\omega t-k z)}+b e^{j(\omega t+k z)} \tag{5.35}
\end{equation*}
$$

This can also be written as

$$
\begin{equation*}
E_{y}=a \cos (\omega t-k z)+b \cos (\omega t+k z) \tag{5.36}
\end{equation*}
$$

where we have taken the real part of the expression for the electric field given in (5.35). Either form of the electric field can be used since trigonometric functions are merely linear combinations of exponential functions. We could have equivalently chosen the imaginary part of (5.35) and obtained the sine functions. It is most common to use the real part assumption. In doing this, we have actually applied Euler's identity

$$
\begin{equation*}
e^{j \theta}=\cos \theta+j \sin \theta \tag{5.37}
\end{equation*}
$$

The first term in either expression (5.35) or (5.36) corresponds to a wave propagating to increasing values of the spatial coordinate $z$ and the second corresponds to a wave propagating to decreasing values of $z$. Both terms are particular forms of the general solution (5.11). The constants $a$ and $b$ are specified by the initial excitation signal and the known direction of propagation.

We have already discovered in (5.14) that the phase velocity $v_{\phi}=d z / d t$ of the propagation of these two traveling waves is obtained from the equation

$$
\begin{equation*}
\mathrm{v}_{\phi}=\frac{\omega}{\mathrm{k}}=\mathrm{c} \tag{5.38}
\end{equation*}
$$

because the medium in which the wave is propagating is a vacuum.

## Time-harmonic plane waves



Figure 5-8. Examples of waves incident upon a beach where the phase velocity (a) $\mathrm{v}_{\phi}$ (beach) $\neq \infty$ and (b) $\mathrm{v}_{\phi}$ (beach) $=\infty$.

In general, the phase velocity is a vector quantity. This is because the phase velocity has both a magnitude and a direction associated with it. It can have a value that is greater than the velocity of light. Even water waves as they crash on the beach can have an infinite phase velocity along the beach if the water waves are propagating exactly perpendicular to the beach as shown in Figure 5-8. In this case, the wave would hit two separated points on the beach at exactly the same time. This leads to an infinite phase velocity along the beach. There is, however, no energy transported along the beach. This point will be returned to later when discussing electromagnetic propagation in waveguides. Suffice it to say that Einstein was correct when he said that particles and energy can go no faster than the speed of light and he continues to be correct. In fact, effects of gravity are not instantaneous but have been recently observed to obey this upper limit for their velocity of propagation.

We can also use the wave number to tell us the direction that the wave is propagating if it is not directed along the $z$ axis. In this case, the wave number will be a vector and it is so indicated with the vector notation $\mathbf{k}$. It is frequently called the wave vector. If the position at which the wave is to be determined is indicated

## Time-harmonic plane waves

with the position vector $\mathbf{r}$, then the term $k z$ is replaced with the scalar product $\mathbf{k} \bullet \mathbf{r}$. The magnitude of the wave vector, $|\mathbf{k}|=k=2 \pi / \lambda$. We will encounter this notation later.

In the previous discussion, we have asserted that the wave depended on only one spatial coordinate. This was chosen to be the $z$ coordinate and we have assumed that $\partial \mathrm{E}_{\mathrm{y}} / \partial \mathbf{x}=\partial \mathrm{E}_{\mathrm{y}} / \partial \mathrm{y}=0$. A two dimensional picture of a segment of this one dimensional wave is illustrated in Figure 5-9. The transverse coordinate could have been either $x$ or $y$. There is no change in the value of the perturbation of the electric field in a plane that its transverse to the $z$ axis. In this plane, all of the field components have the same phase. This is defined as being a plane wave. Plane waves are nice waves from a theoretical and pedagogical point of view since the vector wave equation has been reduced to a one-dimensional scalar wave equation in Cartesian coordinates. This equation is the easiest to solve. There are few cases where plane waves do actually exist. ${ }^{1}$ However in free space, such waves would have to be either launched at $z= \pm \infty$ so any spherical radiation effects originating at a finite sized antenna would have decayed to 0 or the antenna would have to be infinite in its transverse dimension.

[^15]
## Time-harmonic plane waves



Z
X
Figure 5-9 MATLAB. Illustration of a plane wave at a fixed instant of time. At a particular location $z$ and at a particular time $t$, the electric field $E_{y}$ will have the same phase at all points in the transverse plane. The wavelength $\lambda$ is indicated.

Example 5-9. In an experiment performed in a vacuum, we simultaneously measure the electric field to be polarized in the $y$ direction at $z=0$ and one wavelength away at $\mathrm{z}=2 \mathrm{~cm}$. The amplitude is $2 \mu \mathrm{~V} / \mathrm{m}$. Find the frequency of excitation and write an expression that describes the wave if the wave is moving in the direction of increasing values of the coordinate $z$.

## Time-harmonic plane waves

Answer: Since the wavelength was specified to be $\lambda=0.02 \mathrm{~m}$, we can compute the wave number

$$
\mathrm{k}=\frac{2 \pi}{\lambda}=\frac{2 \pi}{0.02}=100 \pi 1 / \mathrm{m}
$$

Since the wave is propagating in a vacuum, the frequency of oscillation can be found from

$$
\omega=k c \text { or } f=\frac{\omega}{2 \pi}=\frac{100 \pi \times 3 \times 10^{8}}{2 \pi}=15 \times 10^{9} \mathrm{~Hz}=15 \mathrm{GHz}
$$

where $G$ is the abbreviation for "giga" and stands for $10^{9}$. Finally, the wave can be expressed as

$$
\mathbf{E}=2 \times 10^{-6} \cos \left[2 \pi\left(15 \times 10^{9} \mathrm{t}-50 \mathrm{z}\right)\right) \mathbf{u}_{Y} \quad \mathrm{~V} / \mathrm{m}
$$

The polarization of the electric field and the direction of propagation of the wave are clearly indicated in this solution.

Example 5-10. Show that a linearly polarized plane wave can be resolved into two equal amplitude waves that rotate about the $z$ axis.

Answer: The linearly polarized wave

$$
E_{y}=E_{0} \mathbf{u}_{Y} e^{j(\omega t-k z)}
$$

can be written as a sum of two components $\mathbf{E}_{\mathbf{y}}=\mathbf{E}_{\mathrm{y}}^{+}+\mathbf{E}_{\mathrm{y}}^{-}$, where

$$
E_{y}^{+}=\frac{E_{0}}{2}\left(\mathbf{u}_{Y}-j \mathbf{u}_{X}\right) e^{j(\omega t-k z)} \quad \text { and } \quad E_{y}^{-}=\frac{E_{0}}{2}\left(\mathbf{u}_{Y}+j \mathbf{u}_{X}\right) e^{j(\omega t-k z)}
$$

We will show that these two components represent respectively a right hand and a left hand rotating wave, each with an amplitude $\mathrm{E}_{\mathrm{o}} / 2$. Taking into account that over the circle $\mathbf{u}_{\mathbf{x}}=-\mathbf{u}_{\varphi} \sin \varphi ; \mathbf{u}_{\mathbf{y}}=\mathbf{u}_{\varphi} \cos \varphi$, we obtain $\mathbf{u}_{\mathbf{y}}-j \mathbf{u}_{\mathbf{x}}=\mathbf{u}_{\varphi} \mathrm{e}^{j \varphi}$ and $\mathbf{u}_{\mathbf{y}}+j \mathbf{u}_{\mathbf{x}}=\mathbf{u}_{\varphi} \mathrm{e}^{-\mathrm{j} \varphi}$ which demonstrates that the first and second wave are rotating in the opposite directions. They are circularly polarized waves. The polarization of

## Time-harmonic plane waves

the wave is a subject in the next section. These are sometimes called "symmetrical components" in electrical machinery studies.

### 5.3.2. Polarization and characteristic impedance

In the preceding discussion, we have emphasized the characteristics of the propagation of only one of the field components of the electromagnetic wave. This was chosen to be the electric field. We will now compute the magnetic field component of the plane wave. As the reader might expect, it will follow directly from Maxwell's equations. For the time-harmonic wave studied in the previous section, the appropriate equation is

$$
\begin{equation*}
\nabla \times \mathbf{E}=-\mathrm{j} \omega \mathbf{B}=-\mathrm{j} \omega \mu_{0} \mathbf{H} \tag{5.39}
\end{equation*}
$$

Let us explicitly find the magnetic field component for the wave that is propagating in the $+z$ direction and is defined by the first term in (5.35). We write

$$
\mathbf{H}=-\frac{1}{j \omega \mu_{0}}\left|\begin{array}{ccc}
\mathbf{u}_{\mathbf{x}} & \mathbf{u}_{\mathbf{Y}} & \mathbf{u}_{\mathbf{z}} \\
\frac{\partial}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{y}} & \frac{\partial}{\partial \mathbf{z}} \\
0 & \text { aexp }[j(\omega t-k z)] & 0
\end{array}\right|
$$

or after expanding the determinant, we write

$$
\begin{equation*}
\mathbf{H}=-\frac{1}{j \omega \mu_{0}} \frac{\mathrm{a} \partial(\exp [\mathrm{j}(\omega \mathrm{t}-\mathrm{kz})])}{\partial \mathbf{z}}\left(-\mathbf{u}_{\mathrm{x}}\right)=-\frac{1}{\omega \mu_{0} / \mathrm{k}} \operatorname{aexp}[\mathrm{j}(\omega \mathrm{t}-\mathrm{kz})] \mathbf{u}_{\mathrm{x}} \tag{5.40}
\end{equation*}
$$

The field components for the electric field intensity E given in (5.35) and the magnetic field intensity $\mathbf{H}$ given in (5.40) for an electromagnetic wave propagating in the $+z$ direction are illustrated in Figure 5-10. In addition, the direction of propagation which is indicated by the wave vector $\mathbf{k}=k \mathbf{u}_{\mathbf{z}}$ is also shown. The constant in the denominator is a quantity that has the dimension of an impedance since the ratio of $\mathrm{E} / \mathrm{H}=(\mathrm{volts} / \mathrm{m}) /(\mathrm{amperes} / \mathrm{m})=(\mathrm{ohms})$. Therefore

$$
\begin{equation*}
Z_{c}=\frac{\omega \mu_{0}}{\mathrm{k}}=\frac{\mathrm{kc} \mu_{0}}{\mathrm{k}}=\frac{\mu_{0}}{\sqrt{\mu_{0} \varepsilon_{0}}}=\sqrt{\frac{\mu_{0}}{\varepsilon_{0}}} \tag{5.41}
\end{equation*}
$$

## Time-harmonic plane waves

This ratio is called the characteristic impedance or the intrinsic wave impedance of the medium. Substituting the values for a vacuum, we find the characteristic impedance equal to $120 \pi \Omega$. We can then rewrite (5.40) to be

$$
\begin{equation*}
\mathbf{H}=-\frac{1}{Z_{C}} \operatorname{aexp}[\mathrm{j}(\omega t-\mathrm{kz})] \mathbf{u}_{\mathrm{x}}=\frac{1}{\mathrm{Z}_{\mathrm{C}}}\left(\mathbf{u}_{\mathrm{z}} \times \mathbf{E}\right) \tag{5.42}
\end{equation*}
$$

It is appropriate at this time to comment on various properties of this wave. The first comment concerns the direction of the propagation of the electromagnetic power. From Poynting's theorem, we find that the power will flow in the direction of the Poynting vector $\mathbf{S}$ given by $\mathbf{S}=\mathbf{E} \times \mathbf{H}=S \mathbf{u}_{\mathbf{z}}$. From Figure 5-10, we see that this will be in the $+z$ direction. This is the same direction as the wave vector $\mathbf{k}$. This is consistent with our initial specification that we would examine only the electromagnetic wave component that is propagating in the $+z$ direction. Since both field components are propagating in the $+z$ direction, we should naturally expect that the electromagnetic power should flow in the same direction². By measuring the electric and magnetic field components at the same instant of time in order to determine the polarization of the two components and computing the Poynting vector, space scientists and radio astronomers can determine the actual origin of radiation detected by satellites or by large ground based antennas. The actual power that flows through a given surface area can be computed by integrating the Poynting vector over that area.

[^16]
## Time-harmonic plane waves



Figure 5-10. Orientation of the electric and magnetic fields at an instant in time that are propagating in the $+z$ direction. The right hand rule indicates the direction of propagation from Poynting's theorem. Only two of the three vectors and the right hand are required to determine the third vector.

Example 5-11. Compute the characteristic impedance of a vacuum $Z_{0}$
Answer: Equation (5.41) yields

$$
Z_{0}=\left|\frac{E}{H}\right|=\sqrt{\frac{\mu_{0}}{\varepsilon_{0}}}=\sqrt{\frac{4 \pi \times 10^{-7}}{\frac{1}{36 \pi} \times 10^{-9}}}=120 \pi \approx 377 \Omega
$$

If we know the value of one of the field components, we also know the other field component since we have only to multiply or divide it by the characteristic impedance $\mathrm{Z}_{\mathrm{c}}$. The reader should receive a word of caution at this point. This impedance is just a ratio of two field quantities and could not be measured by running outside with the two leads of an ohmmeter held up in either hand.

Example 5-12. For the electric field described in Example 5-9, find the magnetic field intensity.

## Time-harmonic plane waves

Answer: We can compute $\mathbf{H}$ from (5.42) to be

$$
\begin{gathered}
\mathbf{H}=\frac{2 \times 10^{-6}}{377} \cos \left[2 \pi\left(15 \times 10^{9} t-50 z\right)\right]\left(\mathbf{u}_{\mathbf{z}} \times \mathbf{u}_{\mathbf{Y}}\right) \\
\mathbf{H}=5.3 \times 10^{-9} \cos \left[2 \pi\left(15 \times 10^{9} \mathrm{t}-50 \mathrm{z}\right)\right]\left(-\mathbf{u}_{\mathbf{x}}\right) \mathrm{A} / \mathrm{m} .
\end{gathered}
$$

Note the direction of the magnetic field intensity. This direction for the magnetic field intensity is required so the power will flow in the $+z$ direction.

Example 5-13. The electric field in a vacuum is given by

$$
E=10 \cos (\omega t-k z) \mathbf{u}_{x} \quad V / m
$$

Find the average power in a circular area in a plane defined by $z=$ constant and whose radius is $R=3 \mathrm{~m}$.

Answer: We first write the electric field in complex form as

$$
\mathbf{E}=10 \mathrm{e}^{\mathrm{j}(\omega t-\mathrm{kz})} \mathbf{u}_{\mathrm{x}} \mathrm{~V} / \mathrm{m}
$$

Since $Z_{0}=120 \pi$, we write

$$
\mathbf{H}=\frac{10}{120 \pi} \mathrm{e}^{\mathrm{j}(\omega t-k z)} \mathbf{u}_{\mathrm{Y}} \mathrm{~A} / \mathrm{m} .
$$

The average power through this plane is

$$
P_{\mathrm{av}}=\frac{1}{2} \operatorname{Re}\left\{\int_{\mathrm{s}} \mathbf{E} \times \mathbf{H}^{*} \bullet \mathbf{d S}\right\}=\frac{1}{2}(10)\left(\frac{10}{120 \pi}\right)\left(\pi 3^{2}\right)=3.75 \mathrm{~W} .
$$

## Plane wave propagation in a dielectric medium 5.4. Plane wave propagation in a dielectric medium <br> 5.4.1. Plane wave propagation in a lossless dielectric medium

In the definition of the wave number that was applicable for a vacuum (5.32), we showed that it was a function of the permeability of free space $\mu_{0}$ and the permittivity of free space $\varepsilon_{0}$ via the relationk $=2 \pi f \sqrt{\varepsilon_{0} \mu_{0}}$. If the space were now uniformly filled with a gas or a linear homogeneous dielectric such as a plastic whose dielectric constant $\varepsilon=\varepsilon_{\mathrm{r}} \varepsilon_{0}$ where $\varepsilon_{\mathrm{r}}>1$, the wave number $k$ would have to be modified by replacing $\varepsilon_{0}$ with $\varepsilon$. We can usually assume the vacuum value $\mu_{0}$ for the permeability $\mu$ since it differs significantly only for wave propagation in iron, nickel and cobalt - materials not normally known to admit wave propagation.


Figure 5-11. Plane waves are launched from the signal generator. The phase of the detected signals can be monitored.

## Plane wave propagation in a dielectric medium

As we will emphasize later, this effect can be used to diagnose the properties of the material. It also is important for the refraction of light that allows glasses to work as vision correcting devices. To introduce this technique here and to describe the wave propagation in a dielectric, let us assume that two materials are juxtaposed as shown in Figure 5-11.

In this figure, we have assumed that plane waves are launched in the two adjacent media and propagate to the end where they are detected. Any phase difference that may exist between the two signals can be detected there. The wave numbers in the two media are

$$
\begin{equation*}
\mathrm{k}_{1}=2 \pi \mathrm{f} \sqrt{\varepsilon_{1} \mu_{0}} \text { and } \mathrm{k}_{2}=2 \pi \mathrm{f} \sqrt{\varepsilon_{2} \mu_{0}} \tag{5.43}
\end{equation*}
$$

Both signals will travel the same distance $\Delta z$ although with different phase velocities $v_{1}=\omega / k_{1}$ and $v_{2}=\omega / k_{2}$. This difference will delay the arrival of one signal with respect to the other and create a detectable phase difference $\delta \theta$. The phase difference $\delta \theta$. is given by

$$
\begin{equation*}
\delta \theta=\mathrm{k}_{1} \Delta \mathrm{z}-\mathrm{k}_{2} \Delta \mathrm{z}=2 \pi \mathrm{f}\left(\sqrt{\varepsilon_{1} \mu_{0}}-\sqrt{\varepsilon_{2} \mu_{0}}\right) \Delta \mathrm{z} \tag{5.44}
\end{equation*}
$$

If we know the total phase change in the signal passing through one of the paths, say $\Delta \theta_{1}$ where

$$
\begin{equation*}
\Delta \theta_{1}=2 \pi f \sqrt{\varepsilon_{1} \mu_{0}} \Delta z \tag{5.45}
\end{equation*}
$$

We can then calculate the phase difference $\delta \theta$ to be

$$
\begin{align*}
\frac{\delta \theta}{\Delta \theta_{1}}= & \frac{2 \pi f\left(\sqrt{\varepsilon_{1} \mu_{0}}-\sqrt{\varepsilon_{2} \mu_{0}}\right) \Delta z}{2 \pi f \sqrt{\varepsilon_{1} \mu_{0}} \Delta z}  \tag{5.46}\\
& =1-\frac{\sqrt{\varepsilon_{2}}}{\sqrt{\varepsilon_{1}}}
\end{align*}
$$

## Plane wave propagation in a dielectric medium

Hence if the material in one of the regions is known (say a vacuum), then one can "diagnose" the material in the other region and determine the relative dielectric constant of that material. Modified versions of this idea have been used to diagnose plastic or paper properties in a manufacturing plant and this is a standard tool in those industries.

The ratio of the phase velocity of light in a vacuum $c$ to the phase velocity in the dielectric is called the index of refraction for the material (here the medium \# 1 is a vacuum and medium \# 2 is the unknown material). The index of refraction is usually given the symbol $n$ and is defined as

$$
\begin{equation*}
\mathrm{n} \equiv \frac{\mathrm{v}_{\text {vacuum }}}{\mathrm{v}_{\text {dielectric }}}=\frac{\sqrt{\varepsilon_{\mathrm{r}}}}{\sqrt{1}}=\sqrt{\varepsilon_{\mathrm{r}}} \tag{5.47}
\end{equation*}
$$

Optical materials are typically specified in terms of their index of refraction.
Example 5-14. Using the setup shown in Figure 5-11, calculate the phase difference $\delta \theta$. if one region is filled a gas with $\varepsilon_{r}=1.0005$ and the other region is a vacuum. The frequency of oscillation is 10 GHz and the length $\Delta z=1 \mathrm{~m}$.

Answer: The phase difference $\delta \theta$ is given by (5.44)

$$
\begin{array}{rl}
\delta \theta=2 & 2 \pi f\left(\sqrt{\varepsilon_{0} \mu_{0}}-\sqrt{\varepsilon_{\mathrm{r}} \varepsilon_{0} \mu_{0}}\right) \Delta \mathrm{z}=2 \pi \mathrm{f} \sqrt{\varepsilon_{0} \mu_{0}} \Delta \mathrm{z}\left(1-\sqrt{\varepsilon_{\mathrm{r}}}\right) \\
= & \frac{2 \pi \times 10^{10}}{3 \times 10^{8}}(1-\sqrt{1.0005})=\frac{2 \pi \times 10^{10}}{3 \times 10^{8}}(-0.00025) \\
=-0.052 \quad \text { radians }=-3^{0}
\end{array}
$$

This number is small but detectable. We can increase the resolution by increasing the length $\Delta z$ or by passing the wave through the regions several times.

## Plane wave propagation in a dielectric medium

### 5.4.2. Plane wave propagation in a lossy dielectric

## medium

Electromagnetic waves can propagate in a material that has a nonzero conductivity $\sigma$. In order to show this, we return to the time harmonic form of Maxwell's equations with the conduction current $\mathbf{J}=\sigma \mathbf{E}$ included. We write the two curl equations for phasors as

$$
\begin{gather*}
\nabla \times \boldsymbol{E}=-\mathrm{j} \omega \mu \boldsymbol{H}  \tag{5.48}\\
\nabla \times \boldsymbol{H}=\boldsymbol{J}+\mathrm{j} \omega \varepsilon \boldsymbol{E}=(\sigma+\mathrm{j} \omega \varepsilon) \boldsymbol{E} \tag{5.49}
\end{gather*}
$$

As before, we will assume that there is no free charge density ( $\rho_{v}=0$ ). Following the procedure that we have employed previously, we combine these two equations to yield a vector wave equation

$$
\begin{equation*}
\nabla^{2} \boldsymbol{E}-\mathrm{j} \omega \mu(\sigma+\mathrm{j} \omega \varepsilon) \boldsymbol{E}=0 \tag{5.50}
\end{equation*}
$$

Again, we will assume that the electric field is linearly polarized in the $\mathbf{u}_{\mathrm{r}}$ direction and the wave propagates only in the $\mathbf{u}_{\mathbf{z}}$ direction. With these simplifications, the resulting wave equation becomes

$$
\begin{equation*}
\frac{\mathrm{d}^{2} E_{y}}{\mathrm{dz}}-\mathrm{j} \omega \mu(\sigma+\mathrm{j} \omega \varepsilon) E_{y}=0 \tag{5.51}
\end{equation*}
$$

This equation can be written as

$$
\begin{equation*}
\frac{\mathrm{d}^{2} E_{Y}}{\mathrm{dz}^{2}}-\gamma^{2} E_{y}=0 \tag{5.52}
\end{equation*}
$$

where

$$
\begin{equation*}
\gamma^{2}=\mathrm{j} \omega \mu(\sigma+\mathrm{j} \omega \varepsilon)=(\mathrm{j} \omega)^{2} \mu \varepsilon\left(1+\frac{\sigma}{\mathrm{j} \omega \varepsilon}\right) \tag{5.53}
\end{equation*}
$$

## Plane wave propagation in a dielectric medium

Equation (5.52) is similar to (5.33) except that the coefficient $\gamma$ is complex. We write $\gamma$ as

$$
\begin{equation*}
\gamma=\alpha+j \beta \tag{5.54}
\end{equation*}
$$

It consists of a real part $\alpha$ and an imaginary part $j \beta$. In vacuum $\alpha=0$ and $\beta=\mathrm{k}=\omega \sqrt{\mu_{0} \varepsilon_{0}}$ or it is a linear function of the frequency $\omega$. As a result - the phase velocity $\mathrm{v}_{\phi}=\omega / \beta=\mathrm{c}$ is a constant. In the general case the real and the imaginary parts of the propagation constant $\gamma$ are nonlinear functions of the frequency $\omega$. After a little algebra, we obtain

$$
\begin{align*}
& \alpha(\omega)=\sigma \sqrt{\frac{\mu}{2 \varepsilon\left[1+\sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}}\right]}}  \tag{5.55}\\
& \beta(\omega)=\omega \sqrt{\frac{\mu \varepsilon}{2}\left[1+\sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}}\right]}
\end{align*}
$$

The substitution of $(5.55)$ into $(5.54)$ leads to $(5.53)$ for the term $\gamma^{2}$. Therefore, the phase velocity depends on the frequency. This is called dispersion and the media in which the wave is propagating is called a dispersive media

$$
\begin{equation*}
v_{\phi}(\omega)=\frac{\omega}{\beta}=\frac{1}{\sqrt{\frac{\mu \varepsilon}{2}\left[1+\sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}}\right]}} \tag{5.56}
\end{equation*}
$$

The group velocity $\mathrm{v}_{\mathrm{g}}$ is defined as

$$
\begin{equation*}
\mathrm{v}_{\mathrm{g}}=\frac{\partial \omega}{\partial \beta}=\frac{1}{\partial \beta / \partial \omega} \tag{5.57}
\end{equation*}
$$

## Plane wave propagation in a dielectric medium

We will encounter the group velocity again when we describe wave propagation on a transmission line. In the present case, the differentiation of (5.55) using the definition given in (5.57) leads to the following group velocity

$$
\begin{equation*}
\mathrm{v}_{\mathrm{g}}(\omega)=\sqrt{\frac{2}{\mu \varepsilon}} \frac{\sqrt{\left[1+\sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}}\right] \sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}}}\left\{\left[1+\sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}}\right] \sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}}-\frac{1}{2}\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}\right\}}{\left\{\left[\frac{1}{\omega}\right\}\right.} \tag{5.58}
\end{equation*}
$$

The group velocity is different than the phase velocity (5.56). The group velocity may equal the velocity of the transport of the energy in certain dispersive media. In other cases, the velocity of the energy transport may be different.

The electric field component for a wave that propagates to increasing values of the coordinate $z$ can be finally written as

$$
\begin{equation*}
E_{y}=E_{y 0} e^{(j \omega t-\gamma z)}=E_{y 0} e^{-\alpha z} e^{j(\omega t-\beta z)} \tag{5.59}
\end{equation*}
$$

This implies that an electromagnetic wave will propagate with a phase constant $\beta$ in the conducting medium but the wave will be attenuated with an attenuation constant $\alpha$ as it propagates in space. The units of the attenuation constant $\alpha$ are given in (Nepers/meter). If $\alpha=1 \mathrm{~Np} / \mathrm{m}$, the amplitude of the wave will decrease to $\mathrm{e}^{-1} \approx 0.368$ of its original value at a distance of 1 meter. An attenuation of $1 \mathrm{~Np} / \mathrm{m}$ equals $20 \log _{10}(e)=8.686 \mathrm{~dB} / \mathrm{m}$. We recall that, although this decay of the amplitude of the wave is in space, this is similar to a time constant in circuits where the decay was in time.

## Plane wave propagation in a dielectric medium

The magnetic field intensity again has only an $\mathrm{H}_{\mathrm{x}}$ component which can be computed from (5.48). The ratio of these two terms yields the characteristic impedance of the conductor which in this case will be complex

$$
\begin{equation*}
Z_{c}(\omega)=\frac{E_{y}}{H_{x}}=\frac{j \omega \mu}{\alpha(\omega)+j \beta(\omega)} \tag{5.60}
\end{equation*}
$$

where both parts of the propagation constant have to be replaced with the expressions (5.55). For a vacuum, $\alpha=0$ and $\beta=\mathrm{k}$ and equation (5.60) reduces to (5.41).

Example 5-15. A $10 \mathrm{~V} / \mathrm{m}$ plane wave whose frequency is 300 MHz propagates in the $(+z)$ direction in an infinite medium. The electric field is polarized in the $u_{x}$ direction. The parameters of the medium are $\varepsilon_{r}=9, \mu_{r}=1$, and $\sigma=10 \mathrm{~S} / \mathrm{m}$. Write a complete time-domain expression for the electric field.

Answer: The attenuation constant is obtained from (5.55) and we write

$$
\alpha=10 \sqrt{\frac{\left(4 \pi \times 10^{-7}\right) \times 36 \pi}{2 \times\left(9 \times 10^{-9}\right)\left[1+\sqrt{1+\left(\frac{10 \times 36 \pi}{2 \pi \times 300 \times 10^{6} \times 9 \times 10^{-9}}\right)^{2}}\right]}}
$$

This gives a numerical value $\alpha=108.01 \mathrm{~Np} / \mathrm{m}$. Similarly, equation (5.55) yields the phase constant

$$
\beta=2 \pi \times 300 \times 10^{6} \sqrt{\frac{\left(4 \pi \times 10^{-7}\right)}{2}\left(\frac{9}{36 \pi} \times 10^{-9}\right)\left[1+\sqrt{1+\left(\frac{10 \times 36 \pi}{2 \pi \times 300 \times 10^{6} \times 9 \times 10^{-9}}\right)^{2}}\right]}
$$

This gives a numerical value $\beta=109.65 \mathrm{rad} / \mathrm{m}$. The complex propagation constant is $\gamma \approx 108+\mathrm{j} 110 \mathrm{~m}^{-1}$. Hence the electric field is

$$
\mathbf{E}=10 \mathrm{e}^{-108 z} \cos \left[\left(2 \pi \times 300 \times 10^{6}\right) \mathrm{t}-110 \mathrm{z}\right] \mathbf{u}_{\mathrm{x}} \mathrm{~V} / \mathrm{m}
$$

## Plane wave propagation in a dielectric medium

Example 5-16 ${ }^{\text {MATLAB }}$. For the medium described in Example 5-15, plot the phase velocity and the group velocity as a function of frequency in the frequency range $30<\mathrm{f}<3000 \mathrm{MHz}$. Recall that 1 GHz equals 1000 MHz .

Answer: The phase velocity $v_{\phi}$ and the group velocity $v_{g}$ based on equations (5.56) and (5.58) are shown in the figure below. Both velocities are normalized by the vacuum value of the velocity of light $\mathrm{c}=3 \times 10^{8} \mathrm{~m} / \mathrm{s}$.


It is evident that both velocities increase with increasing frequency. The theoretical limit of the two velocities when $f \rightarrow \infty$ is the same:

$$
\mathrm{v}_{\infty}=\mathrm{c} / \sqrt{\varepsilon_{\mathrm{r}}}=3 \times 10^{8} / \sqrt{9}=10^{8} \mathrm{~m} / \mathrm{s}
$$

## Plane wave propagation in a dielectric medium

There are two particular cases that should be investigated further:

1) Dielectric with small losses $(\sigma \ll \omega \varepsilon)$ with a high-frequency approximation. In this case we can replace the complex factor in the brackets in (5.55) with

$$
\begin{equation*}
1+\sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}} \approx 2 \tag{5.61}
\end{equation*}
$$

From (5.55), (5.56) and (5.58), we obtain the approximate attenuation and propagation constants to be

$$
\begin{align*}
& \alpha \approx \frac{\sigma}{2} \sqrt{\frac{\mu}{\varepsilon}} \\
& \beta \approx \omega \sqrt{\mu \varepsilon}  \tag{5.62}\\
& v_{\phi} \approx v_{g} \approx 1 / \sqrt{\mu \varepsilon}
\end{align*}
$$

The phase velocity and the group velocity remain the same as in the lossless case. We have, however, introduced a small attenuation into the wave.

We can think of an electromagnetic wave passing through a roast in a microwave oven where the water in the meat acts as the conductor. For the electromagnetic wave, the roast is a complex impedance. Since the field decays and energy must be conserved in the system, the energy that does not pass through the roast is absorbed and converted into heat to cook our dinner. The roast can be considered to be a local "hot spot."
2) Dielectric with large losses ( $\sigma \gg \omega \varepsilon$ ) or imperfect conductor with a lowfrequency approximation. There is the second case that should be investigated further before leaving this topic since it has a multitude of practical implications. For example, we will answer the question "Can an electromagnetic wave

## Plane wave propagation in a dielectric medium

propagate in a good conductor where the conduction current is larger than the displacement current?" The answer to this question will lead to the motivation of microwave oven manufacturers warning against the insertion of a metal pan into the oven. The answer to this question will reappear in our discussion about the reflection of waves. In what follows, the underlying principle will be uncovered.

As we might expect from reading this text, answers to all important questions require that we return to Maxwell's equations. In the case where the conduction current is much larger than the displacement current, we rewrite the second curl equation (5.49) as

$$
\nabla \times \boldsymbol{H}=\boldsymbol{J}+\mathrm{j} \omega \varepsilon \boldsymbol{E} \approx \sigma \boldsymbol{E} .
$$

We obtain again the equation (5.52) where

$$
\begin{equation*}
\gamma=\sqrt{j \omega \mu \sigma} \tag{5.63}
\end{equation*}
$$

which is a consequence of (5.53). From (5.55) and taking into account that

$$
\begin{equation*}
1+\sqrt{1+\left(\frac{\sigma}{\omega \varepsilon}\right)^{2}} \approx\left(\frac{\sigma}{\omega \varepsilon}\right) \tag{5.64}
\end{equation*}
$$

We obtain the following asymptotic terms

$$
\begin{equation*}
\alpha \approx \beta \approx \sqrt{\frac{\omega \mu \sigma}{2}} \tag{5.65}
\end{equation*}
$$

Because $\sqrt{\mathrm{j}}=\mathrm{e}^{\mathrm{j} 45^{\circ}}=(1+\mathrm{j}) / \sqrt{2}$, (5.65) follows directly from (5.63). The phase and group velocities will actually be different.

The equation (5.59) also gives the solution in this particular case - spatial propagating waves with an exponential decay in the amplitude. We can write the

## Plane wave propagation in a dielectric medium

attenuation constant as $\alpha=1 / \delta$, where $\delta$ is a constant with a dimension (meter). This length is called the skin depth of the material. The explicit expression for it is

$$
\begin{equation*}
\delta(\omega)=\sqrt{\frac{2}{\omega \mu \sigma}}=\frac{1}{\sqrt{\pi f \mu \sigma}} \tag{5.66}
\end{equation*}
$$

The phase constant has the same value as the attenuation constant $\beta=1 / \delta$. For a given medium ( $\mu, \sigma=$ constant), the skin depth decreases with increasing frequency - this is known as the skin effect. For the particular case of a perfect conductor (or superconductor when $\sigma \rightarrow \infty$ ), the skin depth is zero and it is independent of the frequency - the electric field does not penetrate into the medium at all.

One can imagine the problem that would arise in a microwave oven if the skin depth were much less than the size of the roast. The edge would become charred and the central part of the roast would remain uncooked. The wavelength at the standard microwave frequency of $f \approx 2.45 \mathrm{GHz}$ is

$$
\lambda=\frac{\mathrm{c}}{\mathrm{f}}=\frac{3 \times 10^{8}}{2.45 \times 10^{9}} \approx 0.12 \mathrm{~m}=12 \mathrm{~cm}
$$

Since the roast is heated from all directions, this is not a major problem.
Example 5-17. Calculate the skin depth of copper at a frequency of 3 GHz .
Answer: The conductivity of copper (see Appendix 3) is $\sigma=5.8 \times 10^{7} \mathrm{~S} / \mathrm{m}$ ( $\mu \approx \mu_{0}$ ). Equation (5.66) yields

$$
\delta=\frac{1}{\sqrt{\pi f \mu \sigma}}=\frac{1}{\sqrt{\pi\left(3 \times 10^{9}\right)\left(4 \pi \times 10^{-7}\right)\left(5.8 \times 10^{7}\right)}} \approx 1.21 \times 10^{-6} \mathrm{~m}=1.21 \mu \mathrm{~m} .
$$

This is a very small value because copper is a good conductor and the frequency is sufficiently high.

## Plane wave propagation in a dielectric medium

Example 5-18 ${ }^{\text {MATLAB }}$. Plot the spatial variation of the electric field $E_{y}(z)$ at a fixed time $(t=0)$ of a plane wave propagating into a copper conductor. The frequency of the wave is 3 GHz and it has an initial amplitude of $\mathrm{E}_{\mathrm{y} 0}=10 \mathrm{~V} / \mathrm{m}$. Assume the material parameters are $\varepsilon_{r}=1, \mu_{r}=1, \sigma=5.8 \times 10^{7} \mathrm{~S} / \mathrm{m}$.

Answer: First, we check the validity of the approximation that the conduction current is larger than the displacement current. This is satisfied if $\sigma \gg \omega \varepsilon$. We compute: $\omega \varepsilon=2 \pi \times 3 \times 10^{9} \times 10^{-9} /(36 \pi)=0.1667 \ll \sigma=5.8 \times 10^{7}$ !

From equation (5.59), it follows that the amplitude of the wave $E_{y}$ decays to $e^{-1}$ of its initial value $E_{y 0}$ in one skin depth $\delta$. Copper is a good conductor and this reduction in the amplitude occurs within the first wavelength. We have $\lambda=2 \pi / \beta=2 \pi \delta$ or $\delta / \lambda=1 / 2 \pi<1$. The spatial response for the electric field $\mathrm{E}_{\mathrm{y}}(\mathrm{z})$ at one instant in time is shown in the figure below.

## Plane wave propagation in a dielectric medium



The skin depth for a particular material depends on the frequency of the electromagnetic wave and the conductivity of the material. There are cases where a metal is gold-plated in order to decrease the skin depth by an additional small percentage since the conductivity of gold is slightly higher than aluminum or even brass. Electromagnetic energy that propagates into a conductor will be converted into heat and hence will be lost. This implies that electromagnetic waves that are guided by a conducting surface such as a stripline or a waveguide will attenuate as they propagate. Both of these topics which are of practical importance in modern communications will be described in the following chapter.

## Plane wave propagation in a dielectric medium

From (5.60), one can find the characteristic complex impedance of the conductor by taking into account (5.65)

$$
\begin{equation*}
Z_{\text {conductor }}=(1+\mathrm{j}) \sqrt{\frac{\omega \mu}{2 \sigma}} \tag{5.67}
\end{equation*}
$$

For large values of conductivity $\sigma$ that are found in many metallic materials, this value will be approximately zero (perfect conductor). This will be important when we next discuss the topic of reflection of electromagnetic waves.

## Reflection and transmission of an electromagnetic wave

### 5.5. Reflection and transmission of an electromagnetic wave 5.5.1. Normal incidence - propagating waves

The subject of electromagnetic waves could now be concluded if we were just interested in studying wave propagation in an infinite medium. We are, however, confronted with a universe with various objects in it. The question then arises, "What happens if a plane wave hits one of these objects?" The object could be either another dielectric or a conductor. The answer to this question will require that we invoke the boundary conditions, derived in the previous chapters.


Figure 5-12. An electromagnetic wave ( $\mathbf{E}_{\mathbf{i}} \times \mathbf{H}_{\mathbf{i}}$ ) is normally incident upon an interface at $z=0$. This results in an electromagnetic wave that is transmitted into the region $z>0\left(E_{t} \times \mathbf{H}_{t}\right)$ and an electromagnetic wave that is reflected back into the region $\mathrm{z}<0\left(\mathbf{E}_{\mathrm{r}} \times \mathbf{H}_{\mathrm{r}}\right)$. The $x$ axis is into the paper.

Let us assume that a plane wave is launched in the region $z<0$ in a lossless material whose dielectric constant is $\varepsilon_{1}$ and that the wave is perpendicularly incident on a second lossless material located in the region $z \geq 0$ whose dielectric constant is $\varepsilon_{2}$. If the incident wave were incident at any other angle, we would be required to perform a little more work and this will be investigated later. The case of perpendicular incidence is depicted in Figure 5-12.

## Reflection and transmission of an electromagnetic wave

The permeabilities of both materials are both equal to their vacuum values $\mu_{0}$.
Using the convention that arises from Poynting's theorem, we define the direction of propagation via the right-hand rule convention. This allows us to suggest at this stage that the polarization of one of the electromagnetic field components of the reflected field would have to be altered after the incident wave strikes the interface. Let us choose it to be the electric field and assume that the magnetic field is unchanged. This will be confirmed when we perform the mathematical manipulations.

The electric field components of the incident, the reflected and the transmitted electromagnetic wave are respectively written from (5.35) as

$$
\begin{align*}
& E_{y, i}=A_{i} e^{j\left(\omega t-k_{1} z\right)} \\
& E_{y, r}=B_{r} e^{j\left(\omega t+k_{1} z\right)}  \tag{5.68}\\
& E_{y, t}=A_{t} e^{j\left(\omega t-k_{2} z\right)}
\end{align*}
$$

where $k_{1}$ is the wave number in region 1 and $k_{2}$ is the wave number in region 2 . We use the notation of employing the constants $A$ and $B$ to indicate respectively the terms that propagate to increasing and to decreasing values of the coordinate $z$ that was introduced in (5.11). The additional subscripts $i, r$ and $t$ indicate the incident, the reflected, and the transmitted terms respectively. Since the materials were assumed to be lossless, the waves will not attenuate as they propagate ( $\alpha=0$ and $\beta=k$ ). The magnetic field intensities $\mathrm{H}_{\mathrm{x}}$ for the three field components can be computed directly from Maxwell's equations or by using the characteristic impedances found in the two regions. We will use the simpler approach here. We write that

## Reflection and transmission of an electromagnetic wave

$$
\begin{align*}
& H_{x, i}=-\frac{A_{i}}{Z_{c, 1}} e^{j\left(\omega t-k_{1} z\right)} \\
& H_{x, r}=\frac{A_{r}}{Z_{c, 1}} e^{j\left(\omega t+k_{1} z\right)}  \tag{5.69}\\
& H_{x, t}=-\frac{A_{t}}{Z_{c, 2}} e^{j\left(\omega t-k_{2} z\right)}
\end{align*}
$$

The boundary conditions that are to be employed are that the tangential components of the electric field should be continuous across the junction and the tangential components of the magnetic field intensity differ by any surface current that is located at the interface. It is reasonable in practice to assume that this current is equal to 0 . This implies that the tangential components of the magnetic field intensity will also be continuous at the interface. Recall that these boundary conditions were discussed previously.

Before invoking these boundary conditions, we should be made aware of a trick that can be used in the calculation at this stage. By choosing that the interface between the two media should be located at $z=0$, the exponential factor $\left.\mathrm{e}^{\mathrm{j}(\omega t+k z)}\right|_{z=0}=\mathrm{e}^{\mathrm{j} \omega t}$ will cancel in all of the terms. If we had not made this choice, we would have been forced to carry some more baggage along in the derivation. Nothing profound, but it does lighten the burden. We will use the same trick when we later study transmission lines.

At the boundary at $z=0$, we therefore write that

$$
\begin{align*}
& E_{y, i}+E_{y, r}=E_{y, t} \\
& H_{x, i}+H_{x, r}=H_{x, t} \tag{5.70}
\end{align*}
$$

Evaluating (5.68) and (5.69) at $z=0$ and substituting them into these expressions, we obtain

$$
\begin{align*}
& A_{i}+B_{r}=A_{t} \\
& \frac{A_{i}}{Z_{\mathrm{c}, 1}}-\frac{B_{r}}{Z_{\mathrm{C}, 1}}=\frac{A_{t}}{\mathrm{Z}_{\mathrm{c}, 2}} \tag{5.71}
\end{align*}
$$

## Reflection and transmission of an electromagnetic wave

In order to proceed further, we must make an assumption that something is known. It is reasonable to assume that we know the amplitude of the incident electric field intensity $A_{i}$. In terms of this known quantity, we can find the other two terms from (5.71) to be

$$
\begin{equation*}
\mathcal{R} \equiv \frac{\mathrm{B}_{\mathrm{r}}}{\mathrm{~A}_{\mathrm{i}}}=\frac{\mathrm{Z}_{\mathrm{c}, 2}-\mathrm{Z}_{\mathrm{C}, i}}{\mathrm{Z}_{\mathrm{c}, 1}+\mathrm{Z}_{\mathrm{C}, \mathrm{c}}} \tag{5.72}
\end{equation*}
$$

and

$$
\begin{equation*}
\mathcal{T} \equiv \frac{\mathrm{A}_{\mathrm{t}}}{\mathrm{~A}_{\mathrm{i}}}=\frac{2 \mathrm{Z}_{\mathrm{c}, 2}}{\mathrm{Z}_{\mathrm{c}, 1}+\mathrm{Z}_{\mathrm{c}, 2}} \tag{5.73}
\end{equation*}
$$

The symbols $\mathbb{R}$ and $\mathcal{T}$ are called the reflection coefficient and the transmission coefficient respectively.

Hence, knowing the characteristic impedance of the materials allows us to determine the propagation characteristics and amplitudes of the wave that is transmitted into the second material and of the wave that is reflected at the interface and propagates in the first material. If the characteristic impedances on both sides of the interface are equal, all of the incident electromagnetic energy will be transmitted into region 2 and none reflected back into region 1 . This is called matching the media and it has many practical applications.

Since the characteristic impedance $Z_{C}=\sqrt{\mu / \varepsilon}$, we can write the reflection and transmission coefficients in terms of the relative dielectric constants of the two dielectric materials. We find

$$
\begin{equation*}
\mathcal{R}=\frac{\sqrt{\varepsilon_{1}}-\sqrt{\varepsilon_{2}}}{\sqrt{\varepsilon_{1}}+\sqrt{\varepsilon_{2}}} \tag{5.74}
\end{equation*}
$$

and

$$
\begin{equation*}
\mathcal{T}=1+\mathcal{R}=\frac{2 \sqrt{\varepsilon_{1}}}{\sqrt{\varepsilon_{1}}+\sqrt{\varepsilon_{2}}} \tag{5.75}
\end{equation*}
$$

## Reflection and transmission of an electromagnetic wave

Example 5-19 MATLAB. Describe the expected reflection - transmission characteristics of an electromagnetic wave normally incident at a layered dielectric. Find the total reflection and transmission coefficients of a single layer with known thickness $d$ and dielectric constant $\varepsilon_{2}=\varepsilon_{\mathrm{r}} \varepsilon_{0}\left(\varepsilon_{1}=\varepsilon_{3}=\varepsilon_{0}\right)$. Plot the frequency dependence on these coefficients assuming $d=0.1 \mathrm{~m}$ and $\varepsilon_{\mathrm{r}}=4$.


0 d

Answer: We will use the notation " + " to indicate a wave moving to increasing values of $z$ and "-" to indicate a wave moving to decreasing values of $z$. The additional numerical subscript specifies the region in which the wave propagates. The superscript indicates the numerical iteration of transmission - reflection. A portion of the electromagnetic wave that is incident from material 1 will be transmitted into material 2 and a portion will be reflected. The wave that is now propagating in material 2 will encounter the interface separating materials 2 and 3 . A portion of this wave will be transmitted into material 3 and a portion will be reflected back to the original interface. Once the wave is in material 3, it will propagate to $z \rightarrow \infty$ as there are no other boundaries. The total transmitted and reflected fields are given by

$$
E_{t}=\sum_{n=1}^{\infty} E_{+(3)}^{(n)} \text { and } E_{r}=\sum_{n=1}^{\infty} E_{-(1)}^{(n)}
$$

## Reflection and transmission of an electromagnetic wave

where $n$ is the index of the particular wave traverses region \#2. The amplitudes of each individual component are determined from repeated applications of (5.74) and (5.75). The wave as it propagates is depicted below.


However, the phase of the individual terms in the summation is different - after every crossing of the dielectric slab, an additional phase difference ( $\mathrm{k}_{2} \mathrm{~d}$ ) appears. Suppose the reflection coefficient from the first interface is $R_{1}$ in the direction (+z) and $-R_{1}$ in the opposite direction, then $R_{2}=-R_{1}$ in the direction (+z) and $-R_{2}=R_{1}$ in the opposite direction (for the particular case that medium 3 is identical with medium 1). Then the transmission coefficient through the first interface in the direction $(+z)$ is $\mathcal{T}_{1}^{+}=1+\mathcal{R}_{1}$, while in the direction $(-z)$ it is $\mathcal{T}_{1}^{-}=1-\mathcal{R}_{1}$. We write the following expression for the total reflected field
$\mathrm{E}_{\mathrm{r}}=\mathrm{E}_{\mathrm{i}}\left\{\mathcal{R}_{1}+\mathcal{T}_{1}^{+} \mathcal{R}_{2} \mathcal{T}_{1}^{-} \mathrm{e}^{\mathrm{j} 2 \mathrm{k}_{2} \mathrm{~d}}+\mathcal{T}_{1}^{+}\left(-\mathcal{R}_{2} \mathcal{R}_{1}\right) \mathcal{R}_{2} \mathcal{T}_{1}^{-} \mathrm{e}^{\mathrm{j} 4 \mathrm{k}_{2} \mathrm{~d}}+\mathcal{T}_{1}^{+}\left(-\mathcal{R}_{2} \mathcal{R}_{1}\right)^{2} \mathcal{R}_{2} \mathcal{T}_{1}^{-} \mathrm{e}^{\mathrm{j} 6 \mathrm{k}_{2} \mathrm{~d}}+\ldots\right\}$

$$
=\mathrm{E}_{\mathrm{i}}\left\{\mathcal{R}_{1}-\left(1-\mathcal{R}_{1}^{2}\right) \mathcal{R}_{1} \mathrm{e}^{\mathrm{j} 2 \mathrm{k}_{2} \mathrm{~d}}\left[1+\left(\mathcal{R}_{1}^{2} \mathrm{e}^{\mathrm{j} 2 \mathrm{k}_{2} \mathrm{~d}}\right)+\left(\mathcal{R}_{1}^{2} \mathrm{e}^{\mathrm{j} 2 \mathrm{k}_{2} \mathrm{~d}}\right)^{2}+\ldots\right]\right\}
$$

The total reflection coefficient for the dielectric slab can be calculated by realizing that this equation can be written as

$$
\mathcal{R} \equiv \frac{E_{r}}{E_{i}}=R_{1}\left[1-\frac{\left(1-R_{1}^{2}\right) e^{j 2 k_{2} d}}{1-R_{1}^{2} e^{j 2 k_{2} d}}\right]=\frac{R_{1}\left(1-e^{j 2 k_{2} d}\right)}{1-R_{1}^{2} e^{j 2 k_{2} d}} .
$$

Similarly, we obtain the following expression for the transmitted field

## Reflection and transmission of an electromagnetic wave

$$
\begin{aligned}
\mathrm{E}_{\mathrm{t}} & =\mathrm{E}_{\mathrm{i}}\left(\mathcal{T}_{1}^{+} \mathcal{T}_{2}^{+} \mathrm{e}^{\mathrm{j} \mathrm{k}_{2} \mathrm{~d}}+\mathcal{T}_{1}^{+}\left(-\mathcal{R}_{2} \mathcal{R}_{1}\right) \mathcal{T}_{2}^{+} \mathrm{e}^{\mathrm{j} 3 \mathrm{k}_{2} \mathrm{~d}}+\mathcal{T}_{1}^{+}\left(-\mathcal{R}_{2} \mathcal{R}_{1}\right)^{2} \mathcal{T}_{2}^{+} \mathrm{e}^{\mathrm{j} 5 \mathrm{k}_{2} \mathrm{~d}}+\ldots\right\} \\
& =\mathrm{E}_{\mathrm{i}}\left\{\left(1-\mathcal{R}_{1}^{2}\right) \mathrm{e}^{\mathrm{j} \mathrm{k}_{2} \mathrm{~d}}\left[1+\left(\mathcal{R}_{1}^{2} \mathrm{e}^{\mathrm{j} 2 \mathrm{k}_{2} \mathrm{~d}}\right)+\left(\mathcal{R}_{1}^{2} \mathrm{e}^{\mathrm{j} 2 \mathrm{k}_{2} \mathrm{~d}}\right)^{2}+\ldots\right\}\right\}
\end{aligned}
$$

The total transmission coefficient for the dielectric slab can also been obtained as

$$
\mathcal{T} \equiv \frac{E_{t}}{E_{i}}=\frac{\left(1-\mathcal{R}_{1}^{2}\right) e^{j \mathrm{j}_{2} \mathrm{~d}}}{1-\mathcal{R}_{1}^{2} \mathrm{e}^{\mathrm{j} 2 \mathrm{k}_{2} \mathrm{~d}}}
$$

The reflection coefficient of the first boundary is found to be $\mathcal{R}_{1}=\left(1-\sqrt{\varepsilon_{\mathrm{r}}}\right) /\left(1+\sqrt{\varepsilon_{\mathrm{r}}}\right)=(1-2) /(1+2)=-1 / 3$ and the wave number in the slab is $\mathrm{k}_{2}=\sqrt{\varepsilon_{\mathrm{r}}} \mathrm{k}=2 \mathrm{k}$. The frequency dependence of the total reflection coefficient (solid line) and of the total transmission coefficient (dashed line) $\left(\mathrm{d}=0.1 \mathrm{~m}, \varepsilon_{r}=4\right)$ are plotted in the figure below.


## Reflection and transmission of an electromagnetic wave

Choosing the frequency $f=3 \mathrm{GHz}$ that corresponds to a free-space wavelength $\lambda=\mathrm{d}=0.1 \mathrm{~m}$, where the coefficients are $\mathcal{R}=0$ and $|\mathcal{T}|=1$ (a transparent slab). It is easy to check that for every frequency the following condition is fulfilled

$$
|\mathcal{R}|^{2}+|\mathcal{T}|^{2}=1
$$

This results from the fact that the dielectric constant is a real value and no energy is absorbed in the slab - the entire incident energy is either reflected or transmitted.

If material two were a good conductor, a peculiar response would result. The characteristic impedance of the conductor would be very small and it would approach 0 as the conductivity approached infinity $\left(Z_{c, 2} \rightarrow 0\right)$. In this case, there would be no transmission of electromagnetic energy into the conductor, or $\mathcal{T}=0$ which directly follows from (5.73). It would all be reflected. From (5.72), we compute that $R=-1$ which states that the polarization of the reflected electric field would be opposite of the incident one. Recall the warning on your microwave oven "Do not place a metal pan inside!"

This can best be demonstrated if we return to the most general solution of the wave equation given in (5.11). We will assume that the wave that arrives from $z=-\infty$ is a pulse instead of a time-harmonic wave and propagates toward the interface as shown in Figure $5-13(a)$ and (b). At $t=t+3 \Delta t$, the tangential electric field must be equal to zero as shown in Figure 5-13(c). For times $t>t+3 \Delta t$, $a$ reflected pulse will propagate back to $z=-\infty$ as shown in Figures 5-13(d) and (e).

## Reflection and transmission of an electromagnetic wave



Figure 5-13. Propagation of a pulse from $z=-\infty$ and its location at five times.

One could think of this reflection in the following terms. A virtual pulse with a negative amplitude is launched at $\mathrm{z}=+\infty$ by a virtual pulse generator at the same time that the real pulse is launched from $z=-\infty$. This virtual pulse propagates to decreasing values of $z$ with the same speed as the real pulse. At the time $t=t+3 \Delta t$, both pulses meet at $z=0$. At $z=0$, they interchange their nature in that the virtual pulse now becomes a real pulse continuing its propagation to decreasing values of the coordinate z . The real pulse transforms into a virtual pulse that propagates into the conductor. At the time $t=t+3 \Delta t$, the amplitude of the two pulses add up to equal zero at the interface in order to satisfy the requirement that the tangential component of the electric field must be equal to zero at a perfect conductor.

Example 5-20. Pulse radars can be used can be used to determine the velocity of speeding automobiles. Show how such a radar might work.

## Reflection and transmission of an electromagnetic wave



Answer: A repetitive electromagnetic pulse from the radar is incident upon the automobile. Due to the high conductivity of the car, it is reflected back to the radar and the total time-of-flight $(\Delta \mathrm{T})_{\mathrm{i}}$ of the $i^{\text {th }}$ pulse can be measured. The repetition frequency of the radar pulses is $(\delta t)^{-1}$. During the time ( $\delta \mathrm{t}$ ), the car travels a distance $\Delta z$. The velocity of the automobile $v_{\text {car }}$ can be computed from

$$
(\Delta T)_{i}-(\Delta T)_{i+1}=\frac{2 L}{c}-\frac{2(L-\Delta z)}{c}=\frac{2 L}{c}-\frac{2\left(L-v_{c a r} \delta t\right)}{c}=\frac{2 v_{c a r} \delta t}{c}
$$

or

$$
v_{\text {car }}=\frac{c\left[(\Delta T)_{i}-(\Delta T)_{i+1}\right]}{2 \delta t}
$$

As can be seen the actual distance of the car from the radar ( L ) is not important. Since both time of flights $(\Delta T)$ and the intervals between the pulses ( $\delta \mathrm{t}$ ) are known as they are parameters involving the radar set and the velocity of light c , the velocity of the car $\mathrm{v}_{\text {car }}$ can be computed. If the computed velocity is below the speed limit, no ticket need be issued.

### 5.5.2. Fabry-Perot resonator - standing waves

At this time, we will return to the investigation of the behavior of a timeharmonic signal that is incident upon a perfectly conducting wall that has a reflection coefficient $\mathcal{R}=-1$. The total electric field in the region $z<0$ will be equal to the sum of the incident and the reflected components which we write as

## Reflection and transmission of an electromagnetic wave

$$
\mathrm{E}_{y}=\operatorname{Re}\left\{\mathrm{B}\left[\mathrm{e}^{\mathrm{j}(\omega t-k z)}-\mathrm{e}^{\mathrm{j}(\omega t+k z)}\right]\right\}=\operatorname{Re}\left\{-\mathrm{Be}^{\mathrm{jot}}\left[\mathrm{e}^{\mathrm{j} k z}-\mathrm{e}^{-\mathrm{j} k z}\right]\right\}=\operatorname{Re}\left\{-\mathrm{j} 2 \mathrm{Be} \mathrm{e}^{\mathrm{jot}} \operatorname{sinkz}\right\}
$$

or

$$
\begin{equation*}
E_{y}=A \sin \omega t \sin k z \tag{5.76}
\end{equation*}
$$

where $A=2 B$. The tangential electric field $E_{y}=0$ at $z=0$. In this case, the signal that consists of two oppositely propagating waves appears to be stationary in space and merely oscillating in time. This is called a standing wave This standing wave results from the constructive and destructive interference of the two counter propagating waves. This is shown in Figure 5-14 at equal intervals of time ( $\Delta \mathrm{t}=\mathrm{T} / 12, \mathrm{~T}=2 \pi / \omega)$. The separation distance between successive null points which are called nodes is equal to one-half of the wavelength $\lambda$. The points where the electric field is a maximum is called an antinode.

If we moved a small detecting probe that responded only to the magnitude of the electric field and whose response time was slow with respect to the sinusoidal oscillation along the $z$ axis, the response of the slowly moving probe would spatially alternate between zero and a maximum value and back to zero as the probe moved over a distance equal to one half of a wavelength. If the probe movement is calibrated, then we can measure the wavelength with a high degree of accuracy. In practice, a probe that would respond to the magnitude of the electric field would be connected to a diode detector. Such a detector is sometimes is called a "square-law" detector.


Figure 5-14 ${ }^{\text {MATLAB }}$. Standing wave depicted at equal temporal intervals in the period of oscillation $(A=10 \mathrm{~V} / \mathrm{m}, \lambda=0.1 \mathrm{~m})$.

Example 5-21. An electromagnetic wave propagating in a vacuum in the region $z$ $<0$ is normally incident upon a perfect conductor located at $z=0$. The frequency of the wave is 3 GHz . The amplitude of the incident electric field is $10 \mathrm{~V} / \mathrm{m}$ and it is polarized in the $\mathbf{u}_{Y}$ direction. Determine the phasor and the instantaneous expressions for the incident and the reflected field components.

Answer: At the frequency $\mathrm{f}=3 \mathrm{GHz}$, we compute

$$
\omega=2 \pi \mathrm{f}=6 \pi \times 10^{9} \mathrm{rad} / \mathrm{s}
$$

## Reflection and transmission of an electromagnetic wave

$$
\mathrm{k}=\frac{\omega}{\mathrm{c}}=\frac{6 \pi \times 10^{9}}{3 \times 10^{8}}=20 \pi 1 / \mathrm{m}, \mathbf{k}=k \mathbf{u}_{\mathrm{z}} .
$$

In phasor notation, the incident wave is expressed as

$$
\begin{gathered}
E_{i}(\mathrm{z})=\mathbf{u}_{\mathrm{Y}} 10 \mathrm{e}^{-\mathrm{j} 20 \pi z} \mathrm{~V} / \mathrm{m} \\
\boldsymbol{H}_{i}(\mathrm{z})=\frac{1}{\mathrm{Z}_{\mathrm{C}}}\left(\mathbf{u}_{\mathbf{z}} \times \mathbf{E}_{\mathrm{i}}\right)=-\mathbf{u}_{\mathrm{x}} \frac{10}{120 \pi} \mathrm{e}^{-\mathrm{j} 20 \pi \mathrm{z}} \mathrm{~A} / \mathrm{m}
\end{gathered}
$$

where the characteristic impedance of free space is $Z_{C}=120 \pi \Omega$. The fields can also be written as

$$
\begin{gathered}
\mathbf{E}_{i}(\mathrm{z}, \mathrm{t})=\operatorname{Re}\left\{\boldsymbol{E}_{i}(\mathrm{z}) \mathrm{e}^{\mathrm{j} \omega \mathrm{t}}\right\}=\mathbf{u}_{\mathrm{Y}} 10 \cos \left(6 \pi \times 10^{9} \mathrm{t}-20 \pi \mathrm{z}\right) \mathrm{V} / \mathrm{m} \\
\mathbf{H}_{\mathrm{i}}(\mathrm{z}, \mathrm{t})=\operatorname{Re}\left\{\boldsymbol{H}_{i}(\mathrm{z}) \mathrm{e}^{\mathrm{j} \mathrm{\omega t}}\right\}=-\mathbf{u}_{\times} \frac{10}{120 \pi} \cos \left(6 \pi \times 10^{9} \mathrm{t}-20 \pi z\right) \mathrm{A} / \mathrm{m}
\end{gathered}
$$

In phasor notation, the reflected wave is expressed as

$$
\begin{gathered}
\boldsymbol{E}_{r}(\mathrm{z})=-\mathbf{u}_{\mathrm{Y}} 10 \mathrm{e}^{+\mathrm{j} 20 \pi \mathrm{z}} \mathrm{~V} / \mathrm{m}, \\
\boldsymbol{H}_{r}(\mathrm{z})=\frac{1}{\mathrm{Z}_{\mathrm{C}}}\left(-\mathbf{u}_{\mathbf{z}} \times \boldsymbol{E}_{r}\right)=-\mathbf{u}_{\mathrm{x}} \frac{10}{120 \pi} \mathrm{e}^{+\mathrm{j} 20 \pi z} \mathrm{~A} / \mathrm{m} .
\end{gathered}
$$

The reflected fields can also be written as

$$
\begin{gathered}
\mathbf{E}_{\mathrm{r}}(\mathrm{z}, \mathrm{t})=\operatorname{Re}\left\{\boldsymbol{E}_{r}(\mathrm{z}) \mathrm{e}^{\mathrm{j} \omega \mathrm{t}}\right\}=-\mathbf{u}_{\mathrm{Y}} 10 \cos \left(6 \pi \times 10^{9} \mathrm{t}+20 \pi \mathrm{z}\right) \mathrm{V} / \mathrm{m} \\
\mathbf{H}_{\mathrm{r}}(\mathrm{z}, \mathrm{t})=\operatorname{Re}\left\{\boldsymbol{H}_{r}(\mathrm{z}) \mathrm{e}^{\mathrm{j} \mathrm{\omega t}}\right\}=-\mathbf{u}_{\times} \frac{10}{120 \pi} \cos \left(6 \pi \times 10^{9} \mathrm{t}+20 \pi \mathrm{z}\right) \mathrm{A} / \mathrm{m}
\end{gathered}
$$

After the constructive and destructive interference occurs, standing waves appear as shown in Figure 5-14. The separation between the nodes in the standing wave is $\lambda / 2=5 \mathrm{~cm}(\lambda=2 \pi / \mathrm{k}=2 \pi / 20 \pi=0.1 \mathrm{~m})$.

An examination of the standing wave depicted in Figure 5-14 leads us to conjecture that it should be possible to insert another high conductivity metal wall at any of the nodes where the tangential electric field is equal to zero without altering the remaining electric field structure. The applicable boundary condition is that the tangential electric field must be zero at a conducting surface. This conjecture is depicted in Figure 5-15 where plates have been inserted at two on the many possible locations. For the moment, we will assume that the plates that

## Reflection and transmission of an electromagnetic wave

are infinite in transverse extent are instantaneously inserted at the nodes such that the electromagnetic energy is "trapped" between the plates and nothing else is disturbed. This energy is actually "coupled" between the plates with an antenna structure, a topic to be discussed later.

Let us now formally derive this result using the one-dimensional Helmholtz equation (5.33) which we reandwrite as

$$
\begin{equation*}
\frac{d^{2} E_{Y}}{d z^{2}}+k^{2} E_{Y}=0 \tag{5.77}
\end{equation*}
$$

Recall that we have assumed a time-harmonic signal. The solution of this equation is given by

$$
\begin{equation*}
\mathrm{E}_{\mathrm{Y}}=\mathrm{A} \sin \mathrm{kz}+\mathrm{B} \cos \mathrm{kz} \tag{5.78}
\end{equation*}
$$

The constants of integration $A$ and $B$ are specified by the boundary condition that the tangential electric field must be equal to zero at a metal wall. These determine the constant $B=0$ and $k=(n \pi / L)$ where $n$ is an integer and $L$ is the distance between the metal walls. If the maximum electric field has a magnitude $E_{Y 0}$, then the spatial distribution of the electric field is given by

$$
\begin{equation*}
\mathrm{E}_{\mathrm{Y}}=\mathrm{E}_{Y 0} \sin \left(\frac{\mathrm{n} \pi z}{\mathrm{~L}}\right) \tag{5.79}
\end{equation*}
$$

The parallel plate cavity depicted in Figure 5-15 is called a Fabry-Perot resonator or cavity. This cavity has a very high "Q" that could approach one million. Remember that the $Q$ of aandn ordinary electrical circuit is of the order of ten. ${ }^{1}$ Since it is very frequency selective, it has received wide application as the cavity that encloses various "lasing" materials. The total lasing material - cavity entity carries the acronym laser. The term laser stands for "Light amplification by stimulated emission and radiation." At light frequencies, it is not a bad

[^17]
## Reflection and transmission of an electromagnetic wave



Figure 5 -15 ${ }^{\text {MATLAB }}$. By inserting thin conducting plates separated by $(\mathrm{n} \lambda / 2$ ) at the locations where the standing wave is zero, the electromagnetic field structure will not be altered. Two of many possible locations are indicated in the figure.
approximation to assume that the transverse dimension is a large number of wavelengths in extent. This very large number can be approximated as being infinity.

We recall that the wave number $k$ is a function of the frequency of oscillation $\omega$ and the velocity of light in the region between the two parallel plates $\mathrm{c}=1 / \sqrt{\varepsilon \mu_{0}}$ where $\varepsilon=\varepsilon_{\mathrm{r}} \varepsilon_{0}$. For the cavities depicted in Figure 5-15, this resonant frequency $\omega=\omega_{\mathrm{r}}$ will be given by

## Reflection and transmission of an electromagnetic wave

$$
\frac{\omega_{\mathrm{r}}}{\mathrm{c}}=\mathrm{k}=\frac{\mathrm{n} \pi}{\lambda}
$$

Or

$$
\begin{equation*}
\omega_{\mathrm{r}}=\frac{\mathrm{n} \pi}{\lambda} \frac{1}{\sqrt{\varepsilon \mu_{0}}} \tag{5.80}
\end{equation*}
$$

For the two cavities depicted in Figure 5-16 which are either empty or filled with a dielectric, we find that the two Fabry-Perot cavities will resonate with slightly different frequencies. The difference of these two frequencies $\Delta \omega$ is given by

$$
\Delta \omega=\omega_{\mathrm{ra}}-\omega_{\mathrm{rb}}=\frac{\mathrm{n} \pi}{\lambda} \frac{1}{\sqrt{\varepsilon_{0} \mu_{0}}}-\frac{\mathrm{n} \pi}{\lambda} \frac{1}{\sqrt{\varepsilon_{\mathrm{r}} \varepsilon_{0} \mu_{0}}}
$$

If the resonant frequency for the vacuum case (Figure 5-16a) can be computed or measured, this frequency difference can be written as


Figure 5-16. (a) An empty Fabry-Perot cavity. (b) A Fabry-Perot cavity filled with a dielectric $\varepsilon=\varepsilon_{\mathrm{r}} \varepsilon_{0}$.

$$
\begin{equation*}
\frac{\Delta \omega_{\mathrm{r}}}{\omega_{\mathrm{ra}}}=1-\frac{\omega_{\mathrm{rb}}}{\omega_{\mathrm{ra}}}=1-\frac{1}{\sqrt{\varepsilon_{\mathrm{r}}}} \tag{5.81}
\end{equation*}
$$

Example 5-22. An empty microwave Fabry-Perot cavity has a resonant frequency of 35 GHz . Determine the thickness $\Delta \mathrm{L}$ of a sheet of paper that is then inserted between the plates if the resonant frequency changes to 34.99 GHz . The separation $L$ between the parallel plates is 50 cm . Assume that the integer $n$ that specifies the mode does not change. You may ignore any reflection at the paper interface.

## Reflection and transmission of an electromagnetic wave



L

Answer: The relative dielectric constant $\varepsilon_{\text {paper }}$ of paper as determined from Appendix 3 is $\varepsilon_{\text {paper }} \approx 3$. The relative dielectric constant separating the plates with the paper inserted can be approximated as

$$
\varepsilon_{\mathrm{r}}=\varepsilon_{\text {paper }} \cdot \Delta \mathrm{L}+1 .(\mathrm{L}-\Delta \mathrm{L}) \approx \varepsilon_{\text {paper }} \cdot \Delta \mathrm{L}+\mathrm{L}
$$

Therefore, we write

$$
\frac{\omega_{\text {vacuum }}-\omega_{\text {paperinserted }}}{\omega_{\text {vacuum }}}=1-\frac{\frac{\mathrm{n} \pi}{1 \cdot \sqrt{\varepsilon_{\mathrm{r}} \varepsilon_{0} \mu_{0}}}}{\frac{\mathrm{n} \pi}{\mathrm{~L} \cdot \sqrt{\varepsilon_{0} \mu_{0}}}}=1-\frac{1}{\sqrt{1+\frac{\Delta \mathrm{L}}{\mathrm{~L}} \varepsilon_{\text {paper }}}} \approx \frac{\Delta \mathrm{L}}{2 \mathrm{~L}} \varepsilon_{\text {paper }}
$$

Inserting the values, we compute

$$
\frac{35-34.99}{35}=\frac{.01}{35} \approx \frac{\Delta \mathrm{~L}}{2 \times 50} 3
$$

or $\Delta \mathrm{L} \approx .01 \mathrm{~cm}=0.1 \mathrm{~mm}$.
From this example and the example mentioned earlier, we can discern that high frequency electromagnetic waves can be used in the diagnostics of various materials. This is a practical technique that has received wide currency in manufacturing paper where the ratio of less expensive water to the more costly wood pulp determines the ultimate grade of the paper. The relative dielectric constants of wood pulp and water are different.

Medical diagnostics for the determination of the ratio of diseased portion to the undiseased portion of a lung in an autopsy of a patient who died, of say

## Reflection and transmission of an electromagnetic wave

pulmonary emphysema, can be performed. Assuming that one of the lungs is sufficiently dried or a reasonable portion of one could be used to yield a value for relative dielectric constant for the lung, the percentage of the diseased lung could be determined. The disease has "eaten" holes in the lung. The solution of such "inverse problems" is part of an interesting new area called medical diagnostics.

## Waveguide - propagation with dispersion

### 5.6. Waveguide - propagation with dispersion

We might suspect that all objects, be they dielectrics or conductors, do not always align themselves so that every incident electromagnetic wave has its wave vector $\mathbf{k}$ perpendicular to every surface. The wave vector $\mathbf{k}$ may also not even be coincident with the axes of a Cartesian coordinate system. This is shown in Figure $5-17$. The vector $\mathbf{r}$ is the position vector from the origin of the coordinate system to an arbitrary point on the plane. For simplicity, let us assume that the waves are propagating in a lossless region.


Figure 5-17. Plane wave propagating at an arbitrary angle with respect to the axes of a Cartesian coordinate system. Both the electric and magnetic fields have equiphase contours in the plane.

We appeal to the intuition that you may have gained from sitting at a beach and watching the waves crashing on the shore with a thundering roar. You could probably even define a "velocity of crashing" that could be associated with this. This velocity could be defined in several directions, say parallel to the water-sand interface or perpendicular to this interface. The study of the reflection and the

## Waveguide - propagation with dispersion

transmission of electromagnetic waves in a similar situation will be slightly more complicated than the case of normal incidence.

Let us initially work through the details of the calculations that are required to ascertain the reflection properties of an electromagnetic wave obliquely incident upon a conducting surface. In addition to deriving important results, it will also expose us to the procedure required to characterize an electromagnetic wave whose direction of propagation is not coincident with one of the axes of the Cartesian coordinate system. In order to describe the electric field depicted in Figure 5-17, we generalize the expression given in (5.35) to

$$
\begin{equation*}
E=E_{0} e^{j(\omega t-k \diamond r)} \tag{5.82}
\end{equation*}
$$

The magnetic field intensity $\mathbf{H}$ is found from the electric field using Maxwell's equations or as we will do, just divide the electric field by the characteristic impedance for the particular medium in which the wave is propagating using (5.42). The two components of the electromagnetic wave are perpendicular to each other and to the direction of propagation

$$
\begin{equation*}
\mathbf{H}=\frac{1}{Z_{C}}\left(\mathbf{u}_{\mathbf{k}} \times \mathbf{E}\right)=\frac{1}{Z_{C}}\left(\mathbf{u}_{\mathbf{k}} \times \mathbf{E}_{0}\right) \mathrm{e}^{\mathrm{j}(\omega \mathrm{t}-\mathbf{k} \cdot \mathbf{r})} \tag{5.83}
\end{equation*}
$$

where ( $\mathbf{u}_{\mathbf{k}}=\mathbf{k} / \mathrm{k}$ ) is the unit vector of the direction of the propagation.
In order to analyze the wave reflection and transmission properties, we will separate the general problem into two separate cases and treat each one individually. The first case will have the electric field of the incident plane wave be parallel to the conducting surface as shown in Figure 5-18a. The second case will have the magnetic field intensity tangent to this surface as shown in Figure 5-18b. We will analyze the first case in full detail since some interesing results that are of practical importance will ensue.

## Waveguide - propagation with dispersion

In both (5.82) and (5.83), the term $\mathbf{k} \bullet \mathbf{r}$ appears. The term $\mathbf{k}$ is the wave vector. This term has a magnitude that equals the wave number that was defined previously: and in Cartesian coordinates is $k=\omega / \mathrm{c}=\left(\mathrm{k}_{\mathrm{X}}^{2}+\mathrm{k}_{Y}^{2}+\mathrm{k}_{\mathrm{Z}}^{2}\right)^{1 / 2}$. The wave

(a)

(b)

Figure 5-18. (a) Plane wave incident upon a conducting surface located at $\mathrm{x}=0$ with the electric field out of the plane of the page. (b) Plane wave incident on a conducting surface located at $\mathrm{x}=0$ with the magnetic field out of the plane of the page. In both cases, $\mathrm{k}_{\mathrm{y}}=0$.
vector has a direction of propagation that is perpendicular to the plane that contains the electric and magnetic field components. In a two-dimensional situation which will be analyzed in the following discussion, we assume that $\mathrm{k}_{\mathrm{y}}=0$. The term $\mathbf{r}$ is the radial vector from the origin of the coordinate system to any point in the equiphase plane. We explicitly write this scalar product for the incident and reflected terms as

$$
\begin{align*}
\mathbf{k}_{\mathbf{i}} \bullet \mathbf{r} & =\left(\mathrm{k} \cos \theta_{\mathrm{i}} \mathbf{u}_{\mathrm{x}}+\mathrm{k} \sin \theta_{i} \mathbf{u}_{\mathrm{z}}\right) \bullet\left(\mathrm{x} \mathbf{u}_{\mathrm{x}}+\mathrm{y} \mathbf{u}_{\mathrm{Y}}+\mathrm{z} \mathbf{u}_{\mathrm{z}}\right)= \\
& =\mathrm{k}\left(\mathrm{x} \cos \theta_{\mathrm{i}}+\mathrm{z} \sin \theta_{\mathrm{i}}\right)=\mathrm{k}_{\mathrm{xi}} \mathrm{x}+\mathrm{k}_{\mathrm{zi}} \mathrm{z} \tag{5.84}
\end{align*}
$$

and

## Waveguide - propagation with dispersion

$$
\begin{align*}
\mathbf{k}_{\mathrm{r}} \bullet \mathbf{r} & =\left(-\mathrm{k} \cos \theta_{\mathrm{r}} \mathbf{u}_{\mathrm{x}}+\mathrm{k} \sin \theta_{\mathrm{r}} \mathbf{u}_{\mathrm{z}}\right) \bullet\left(\mathrm{x} \mathbf{u}_{\mathrm{x}}+\mathrm{y} \mathbf{u}_{\mathrm{y}}+\mathrm{z} \mathbf{u}_{\mathrm{z}}\right)= \\
& =\mathrm{k}\left(-\mathrm{x} \cos \theta_{\mathrm{r}}+\mathrm{z} \sin \theta_{\mathrm{r}}\right)=-\mathrm{k}_{\mathrm{xr}} \mathrm{x}+\mathrm{k}_{\mathrm{zr}} \mathrm{z} \tag{5.85}
\end{align*}
$$

The magnitude of the wave vector $\mathbf{k}$ is determined by the media in which the electromagnetic wave propagates.

Now that the terms that are new have been defined, let us proceed without delay. We will find the propagation properties of the electromagnetic wave that has the electric field directed out of the plane of the page first. This is shown in Figure 5-18a. The electric field of the incident wave is given by

$$
\begin{equation*}
\mathbf{E}_{\mathbf{i}}=\mathrm{E}_{\mathrm{i} 0} \mathbf{u}_{\mathbf{Y}} \mathrm{e}^{-\mathrm{jk}\left(x \cos \theta_{i}+2 \sin \theta_{i}\right)} \tag{5.86}
\end{equation*}
$$

The magnetic field intensity of the incident wave is

$$
\begin{equation*}
\mathbf{H}_{\mathrm{i}}=\frac{\mathrm{E}_{\mathrm{i} 0}}{\mathrm{Z}_{\mathrm{C}}}\left(\cos \theta_{\mathrm{i}} \mathbf{u}_{\mathrm{x}}-\sin \theta_{\mathrm{i}} \mathbf{u}_{\mathrm{z}}\right) \mathrm{e}^{-\mathrm{jk}\left(x \cos \theta_{\mathrm{i}}+\mathrm{z} \sin \theta_{\mathrm{i}}\right)} \tag{5.87}
\end{equation*}
$$

The reflected wave field has the components

$$
\begin{equation*}
\mathbf{E}_{\mathrm{r}}=\mathrm{E}_{\mathrm{r} 0} \mathbf{u}_{\mathbf{Y}} \mathrm{e}^{-\mathrm{jk}\left(x \cos \theta_{\mathrm{r}}+z \sin \theta_{\mathrm{r}}\right)} \tag{5.88}
\end{equation*}
$$

and

$$
\begin{equation*}
\mathbf{H}_{\mathrm{r}}=\frac{\mathrm{E}_{\mathrm{r} 0}}{\mathrm{Z}_{\mathrm{C}}}\left(\cos \theta_{\mathrm{r}} \mathbf{u}_{\mathrm{x}}+\sin \theta_{\mathrm{r}} \mathbf{u}_{\mathrm{z}}\right) \mathrm{e}^{-\mathrm{jk}\left(\mathrm{x} \cos \theta_{\mathrm{r}}+\mathrm{z} \sin \theta_{\mathrm{r}}\right)} \tag{5.89}
\end{equation*}
$$

At the metal surface located at $x=0$, the electric field is equal to zero. From (5.86) and (5.88), the boundary condition implies that $E_{r 0}=-E_{i 0}$ and $\theta_{r}=\theta_{i}$ (Snell's law). The angle of incidence equals the angle of reflection.

Let us now suppose that an additional infinite metal plate is inserted parallel to the first plate. By convention, we assume that the separation distance is a as shown in Figure 5-19.

# Waveguide - propagation with dispersion 



Figure 5-19. A parallel plate waveguide. The metallic plates that have an infinite conductivity are infinite in the $y$ and $z$ directions.

In addition, let us also assume that a second plane wave with the same amplitude as the first one and is simultaneously excited with the first one. We assume that the polarization of the two waves differ in phase by $180^{\circ}$. Both waves have the same value of $k_{z}$. However, the transverse components $k_{x}$ of the wavenumbers differ by a "-" sign. The two waves are depicted in Figure 5-20. Since the media between the conducting plates is linear, the total electric field will be just the vector sum of the components

$$
\begin{equation*}
E_{y}=E_{y 1}+E_{y 2}=E_{y y}^{\mid} 0^{j\left(\omega t+k_{x} x-k_{z} z\right)}-E_{y 0}^{\mid} e^{j\left(\omega t-k_{x} x-k_{z} z\right)} \tag{5.90}
\end{equation*}
$$

We factor the common terms that appear in (5.90) and write this as

$$
\begin{equation*}
E_{y}=E_{y 0}\left\{e^{j k_{x} x}-e^{-j k_{x} x}\right\} e^{j\left(\theta t-k_{z}\right)} \tag{5.91}
\end{equation*}
$$

The two terms within the brackets are a definition of a sine function. Therefore, (5.91) can be written as

$$
\begin{equation*}
E_{y, n}=E_{y 0} \sin \left(k_{x, n}\right) e^{j\left(\omega t-k_{z, n}\right)}=E_{y 0} \sin \left(\frac{n \pi x}{a}\right) e^{j\left(\omega t-k_{z, n}\right)} \tag{5.92}
\end{equation*}
$$

where $\mathrm{E}_{\mathrm{y} 0}=j 2 \mathrm{E}_{\mathrm{y} 0}^{\prime}$. The boundary condition that the tangential electric field at a conductor is satisfied if $\left.\mathrm{E}_{Y}\right|_{X=0}=0$ and $\left.\mathrm{E}_{Y}\right|_{X=a}=0$. This is satisfied with $k_{X, n}=\frac{n \pi}{a}$

## Waveguide - propagation with dispersion

where $n=1,2,3, \ldots$. The electric field in (5.92) can be interpreted as a plane wave that propagates in the positive $z$ direction and has an inhomogeneous distribution in the transverse $x$ coordinate. The electric field is depicted at an instant of time in Figure 5-20.

The structure that we have just introduced is called a parallel plate waveguide. Since there is no variation in the third coordinate $y$ and the electric field would be perpendicular to the surfaces at $y=0$ and at $y=b$, additional parallel metal plates can be inserted there. This will create a completely enclosed rectangular metal pipe. High frequency electromagnetic waves called microwaves can propagate in such a structure. You may already have familiarity with a microwave oven, a microwave telecommunications system, or a police microwave radar system.

The wavenumbers $k_{x}$ and $k_{z}$ are related via the equation

$$
\begin{equation*}
k^{2} \equiv\left(\frac{\omega}{c}\right)^{2}=k_{x, n}^{2}+k_{z, n}^{2} \tag{5.93}
\end{equation*}
$$



Figure 5-20. Two plane waves that propagate at an angle with respect to the $z$ axis between two parallel metal planes.

From (5.93), we write

$$
\begin{equation*}
\mathrm{k}_{\mathrm{z}, \mathrm{n}}=\sqrt{\left(\frac{\omega}{\mathrm{c}}\right)^{2}-\mathrm{k}_{\mathrm{x}, \mathrm{n}}^{2}}=\sqrt{\left(\frac{\omega}{\mathrm{c}}\right)^{2}-\left(\frac{\mathrm{n} \pi}{\mathrm{a}}\right)^{2}} \tag{5.94}
\end{equation*}
$$

## Waveguide - propagation with dispersion

In addition to the free space wavelength described previously, there is an additional wavelength called the "guide wavelength" $\lambda_{g}$ that is depicted in Figure $5-21$. It is related to the propagation in the direction $z k_{z}$ via $\lambda_{g, n}=2 \pi / k_{z, n}$. From (5.94), we note that the propagation constant $k_{z, n}$ depends upon frequency. For frequencies below a critical frequency $\frac{\omega}{c}<\frac{n \pi}{a}$, the square root of a negative number will be imaginary. This will cause the wave to attenuate as it propagates in the $z$ direction.


## Waveguide - propagation with dispersion

Figure 5-21 ${ }^{\text {MATLAB }}$. Electric field between two infinite parallel planes at an instant of time.

For the propagating wave, the propagation constant for each particular mode $\mathrm{k}_{\mathrm{z}, \mathrm{n}}$ must be real. Therefore, the frequency of the excited wave must be above a certain frequency that is called the cutoff frequency

$$
\begin{equation*}
\mathrm{f}_{\mathrm{C}, \mathrm{n}}=\frac{\omega_{\mathrm{c}, \mathrm{n}}}{2 \pi}=\frac{\mathrm{nc}}{2 \mathrm{a}} \tag{5.95}
\end{equation*}
$$

This follows from (5.94) with $\mathrm{k}_{\mathrm{z}, \mathrm{n}}=0$.
We have just obtained the electric field component of the electromagnetic wave that propagates in a rectangular waveguide. Since the electric field is transverse to the direction of propagation, this is called the "TE mode" or "h mode" of propagation in the microwave community. The electric field that propagates in the lowest order mode a parallel plate waveguide is depicted in Figure 5-21.

The propagation constant $\mathrm{k}_{\mathrm{z}}$ depends upon the frequency as shown in Figure 5-22. This curve is the dispersion relation (5.94). As the frequency becomes much greater than the cutoff frequency for a particular mode, $k_{z, n} \rightarrow k=\omega / c$ and the propagating wave will only be mildly affected by effects of dispersion.

## Waveguide - propagation with dispersion



Figure 5-22 ${ }^{\text {MATLAB }}$. Dispersion relation for electromagnetic waves in a parallel plate waveguide for a particular mode.

It is possible to find the other components of the electromagnetic wave from the component $E_{y, n}$ using Maxwell's equations. We obtain

$$
\begin{align*}
& H_{X, n}=-\frac{k_{z, n}}{\omega \mu_{0}} E_{Y 0} \sin \left(\frac{n \pi x}{a}\right) e^{j\left(\omega t-k_{z, n} z\right)}  \tag{5.96}\\
& H_{z, n}=j \frac{k_{X}}{\omega \mu_{0}} E_{Y 0} \cos \left(\frac{n \pi x}{a}\right) e^{j\left(\omega t-k_{z, n} z\right)} \tag{5.97}
\end{align*}
$$

The ratio of the electric field intensity divided by the magnetic field intensity is called the "wave impedance" of the waveguide. This is similar to (5.60)

$$
\begin{equation*}
Z_{C, n}(f)=-\frac{E_{Y}}{H_{x}}=\frac{\omega \mu_{0}}{k_{Z}}=\frac{\omega \mu_{0}}{\sqrt{\left(\frac{\omega}{c}\right)^{2}-\left(\frac{n \pi}{a}\right)^{2}}}=\frac{Z_{0}}{\sqrt{1-\left(\frac{f_{c, n}}{f}\right)^{2}}} \tag{5.98}
\end{equation*}
$$

## Waveguide - propagation with dispersion

where $Z_{0} \approx 377 \Omega$ is the characteristic impedance of free-space and $f_{C . n}$ is defined by (5.95). This impedance depends upon frequency and it is resistive for frequencies greater than the cut off frequency and it is reactive for frequencies beneath this value. It manifests itself by allowing propagation of the waves at higher frequencies. The wave impedance $Z_{c, 1}$ for the $T E_{10}$ mode as a function of frequency is shown in Figure 5-23. Near the cutoff frequency it is very large and it decreases with increasing the frequency. The asymptotic value of the impedance is the free-space characteristic impedance $Z_{0}$.


Figure $5-23^{\text {MATLAB }}$. The frequency dependence of the characteristic impedance for the lowest order mode of a parallel plate waveguide. The normalizations are $Z_{0}=120 \pi \Omega$ and $f_{c}=c / 2 a(n=1)$.

## Waveguide - propagation with dispersion

Other transverse electric modes $\mathrm{TE}_{\mathrm{nm}}$ can be similarly analyzed for rectangular waveguides whose dimensions are a and b. Equation (5.94) for the propagation constant $\mathrm{k}_{\mathrm{z}}$ is replaced with the following expression

$$
\begin{equation*}
\mathrm{k}_{\mathrm{z}, \mathrm{~m} \mathrm{~m}}=\sqrt{\left(\frac{\omega}{\mathrm{c}}\right)^{2}-\left(\frac{\mathrm{n} \pi}{\mathrm{a}}\right)^{2}-\left(\frac{\mathrm{m} \pi}{\mathrm{~b}}\right)} \tag{5.99}
\end{equation*}
$$

A similar analysis could be performed for cases in which there is no component of magnetic field in the direction of propagation. These modes are called transverse magnetic modes and are written as $\mathrm{TM}_{\mathrm{nm}}$ modes. The analysis will not be presented here.

Example 5-23. Two infinite parallel conducting plates are separated by 5 cm . Calculate the cutoff frequencies for the lowest order mode if the region between the plates is filled with glass.

Answer: From Appendix 3, we find that the relative dielectric constant of glass is $\varepsilon_{\mathrm{r}}=6$. From (5-95), we compute with $\mathrm{n}=1$ and recalling that the velocity of light in glass is defined as $v=\frac{c}{\sqrt{\varepsilon_{r}}}$

$$
\mathrm{f}_{\mathrm{C}, 1}=\frac{\mathrm{c}}{2 \mathrm{a} \sqrt{\varepsilon_{\mathrm{r}}}}=\frac{3 \times 10^{8}}{2 \times 0.05 \sqrt{6}}=1.225 \mathrm{GHz} .
$$

Example 5-24. Determine the propagation characteristics of the lowest order mode $(\mathrm{n}=1$ ) for the waveguide in Example 5-23 at a frequency $\mathrm{f}=1 \mathrm{GHz}$.

Answer: From (5-94), we write

$$
\mathrm{k}_{\mathrm{z}, 1}=\frac{\pi}{\mathrm{a}} \sqrt{\left(\frac{\omega}{\omega_{\mathrm{C}, 1}}\right)^{2}-1}=\frac{\pi}{0.05} \sqrt{\left(\frac{2 \pi \times 1 \mathrm{GHz}}{2 \pi \times 1.225 \mathrm{GHz}}\right)^{2}-1}=-j 36.29 \quad 1 / \mathrm{m}
$$

The additional factor of "j" implies that the wave will attenuate as it propagates

$$
\mathrm{e}^{-\mathrm{jk} \mathrm{k}_{\mathrm{\prime}} \mathrm{z}}=\mathrm{e}^{-36.29 \mathrm{z}}
$$

## Waveguide - propagation with dispersion

Example 5-25. Calculate the cutoff frequency for the two lowest-order modes ( $\mathrm{n}=1$ and $\mathrm{n}=2$ ) of a parallel plate waveguide described in Example 5-23.

Answer: The cutoff frequency is defined from (5-95). Hence

$$
\mathrm{f}_{\mathrm{c}, 1}=\frac{\pi \times 3 \times 10^{8}}{2 \pi \times 0.05}=3 \quad \mathrm{GHz} .
$$

for the fundamental mode ( $n=1$ ). The cutoff frequency for the second mode ( $n=2$ ) is twice as large, or $f_{c, 2}=6 \mathrm{GHz}$. Waves with frequencies between 3 GHz and 6 GHz will propagate only in the lowest mode. Waves with frequencies above 6 GHz could propagate in either mode ( $\mathrm{n}=1$ or $\mathrm{n}=2$ ). In practice, one typically wants only one mode of propagation to exist at the same time. Therefore, waveguides come in all sizes with dimensions that can range from millimeters up to meters.

## Conclusion

### 5.7. Conclusion

The propagation of plane electromagnetic waves that we have encountered in this chapter has provided us with the first application of Maxwell's equations. Electromagnetic waves are transverse waves in that the electric and magnetic fields are in a plane that is perpendicular to the direction of propagation. This is similar to waves that propagate on the surface of water or along a string and contrasts with sound waves that are longitudinal waves. The velocity of propagation c is determined by the materials in which the wave is propagating with the highest velocity being in a vacuum $c=1 / \sqrt{\varepsilon_{0} \mu_{0}} \approx 3 \times 10^{8} \mathrm{~m} / \mathrm{s}$. The ratio of the electric field and the magnetic field intensity is given by the characteristic impedance which in a vacuum has a value of $Z_{0}=120 \pi \approx 377 \Omega$. Time-harmonic waves have certain unique features such as wavelengths, frequencies of oscillation, and phase velocities but the general wave properties remain. Lossy materials will attenuate the wave as it propagates.

If the wave propagates from one material to another, a portion of the incident wave will be reflected back into the first material and a portion will be transmitted into the second material. Good conductors will reflect most of the incident wave. The boundary conditions will determine the amplitude of each of these terms. If the wave reaches a knife-edge boundary of a conducting screen, the wave will change its direction of propagation or it will be diffracted by the knifeedge. The diffracted wave then will cause an oscillation of the amplitude of the wave in the illuminated region near the line of sight. However, a portion of this

## Conclusion

wave will also appear in the shadow region behind the screen. The topic of diffraction will also be discussed in Chapter 7.

## Problems 5

### 5.8. Problems

1. In terms of the fundamental units mass (kg), length (m), time (s), and charge $(C)$, show that $\left(\varepsilon_{0} \mu_{0}\right)^{-1 / 2}$ has the units of a velocity $(\mathrm{m} / \mathrm{s})$.
$2^{\text {MATLAB }}$. Let $F(z-c t)=1$ and $G(z+c t)=-2$ for $|z-c t| \leq 1$ and $|z+c t| \leq 1$ respectively and $F(z-c t)=G(z+c t)=0$ elsewhere. Accurately sketch the pulse with the velocity $c=2$ at three times: $t=0, t=1$ and $t=3$. Note: normalized variables are used.
2. Define the functions $F\left(z-v_{1} t\right)$ and $G\left(z+v_{2} t\right)$ from the following sketch which was drawn at the times $t=0$ and $t=2$.

3. If the waves in problem 3 were electromagnetic waves, find the ratio of the dielectric constants $\varepsilon_{1}$ and $\varepsilon_{2}$ for the two regions $(z<0) /(z>0)$ if the relative permeabilities were 1 in the two regions.

## Problems 5

5. A displacement wave on a string is described by a harmonic wave $\psi=0.02 \sin [2 \pi(10 t-0.5 z)] \mathrm{m}$, where $z$ is in meters and $t$ is in seconds. Find:
(a) The propagation velocity $v$.
(b) Wavelength $\lambda$ and wave number $k$.
(c) Frequency f and angular frequency $\omega$.
(d) The period T.
(e) Direction of propagation.
(f) Amplitude of the wave A.
$6^{\text {MATLAB }}$. Plot the wave given in Problem 5 as a function of $z$ at $t=(a) 0 \mathrm{~s}$, (b) 0.025 s , (c) 0.05 s , and (d) 0.075 s . Convince yourself that the wave pattern progresses in the positive $z$ direction as time increases.
$7^{\text {MATLAB }}$. Assume that a wave reflector were installed at $z=5$ in the problem 6 . This reflector causes a positive amplitude pulse to be reflected as a positive amplitude pulse (reflection coefficient $\mathcal{R}=-1$ ). Reflection implies that a wave traveling to increasing values of $z$ would start traveling to decreasing values of $z$ after reflection. Accurately sketch the expected oscilloscope pictures.
$8^{\text {MATLAB }}$. Show that the Gaussian pulse defined by $\psi(z, t)=0.5 \exp \left[-(z-5 t)^{2}\right]$ satisfies a wave equation (5.10). Plot this function as a function of $z$ for the three times: $\mathrm{t}=0 ; 0.5 \mathrm{~s}$ and 1 s.

## Problems 5

$9^{\text {MATLAB }}$. Snapshots of two cycles of the electromagnetic wave propagating in a vacuum are taken at three locations: $z=0, z=1$, and $z=2 \mathrm{~m}$. Find the wavelength and the frequency of the wave. Write the equation that describes the electric field.

10. If we know that the magnetic field intensity of an electromagnetic wave is $\mathbf{H}=\mathrm{H}_{0} \mathrm{e}^{\mathrm{j}(\mathrm{ot+kz})} \mathbf{u}_{\mathrm{Y}}$, find the direction of power flow, the electric field $\mathbf{E}$ and the timeaverage Poynting vector $\mathbf{S}_{\mathrm{av}}$.
11. The electric field of an electromagnetic wave is $\mathbf{E}=-10 \mathrm{e}^{\mathrm{j}(t \mathrm{t}-\mathrm{zz})} \mathbf{u}_{\mathrm{x}} \mathrm{V} / \mathrm{m}$. Find the magnetic field intensity $\mathbf{H}$. Compute the time-average Poynting vector $\mathbf{S}_{\mathrm{av}}$.

## Problems 5

12. An electromagnetic wave with a frequency $f=1 \mathrm{MHz}$ propagates in a dielectric material $\left(\varepsilon_{r}=4, \mu_{r}=1\right)$ and it has an electric field component $E_{Y}=1.3 \cos (\omega t-k z) \mathrm{V} / \mathrm{m}$. Find the velocity of the wave $\mathbf{v}$, the wave vector $\mathbf{k}$, the characteristic impedance of the material $\mathrm{Z}_{\mathrm{C}}$ and the magnetic field $\mathbf{H}$.
13. A helium-neon laser emits light at a wavelength of $6328 \AA=6.328 \times 10^{-7} \mathrm{~m}$ in air. Calculate the frequency of oscillation of the laser, the period of the oscillation and the wave number. The symbol $\AA$ is called an Angstrom where $1 \AA=10-10 \mathrm{~m}$.
14. Prove that $\mathbf{E}$ and $\mathbf{H}$ are orthogonal in a vacuum for an arbitrary function of ( $z-c t$ ).
15. The complex electric field of a uniform plane wave propagating in air is given by $E=E_{0}\left[u_{x} \sin (\omega t-k z)+u_{Y} \cos (\omega t-k z)\right]$. Using a sketch, show that it is justified to call this wave circularly polarized.
$16^{\text {MATLAB }}$. The electric field of a uniform plane wave propagating in air is given by $E=E_{0}\left[\mathbf{u}_{\mathrm{x}} \sin (\omega t-k z)+\mathbf{u}_{\mathrm{Y}} \operatorname{asin}(\omega t-k z+\delta)\right] \quad(a \neq 1)$. Using an accurately drawn sketch with $\mathrm{E}_{0}=1, \mathrm{a}=2, \delta=\pi / 4$ show that it is justified to call this wave elliptically polarized.
16. Show that the circular polarization and the linear polarization are special cases of an elliptical polarization.
17. In free space, a signal generator launches an electromagnetic wave that has a wavelength of 10 cm . As the same wave propagates in a material, its wavelength is reduced to 8 cm . In the material, the amplitude of the electric field $\mathbf{E}$ and the magnetic field intensity H are measured to be $50 \mathrm{~V} / \mathrm{m}$ and $0.1 \mathrm{~A} / \mathrm{m}$ respectively. Find the generator frequency $f$ and $\varepsilon_{r}$ and $\mu_{r}$ for the material.

## Problems 5

19. In free space, a signal generator launches an electromagnetic wave that has a wavelength of 3 cm . As the same wave propagates in a material, its wavelength is reduced to 1.5 cm . In the material, the amplitude of the electric field $\mathbf{E}$ and the magnetic field intensity H are measured to be $60 \mathrm{~V} / \mathrm{m}$ and $0.1 \mathrm{~A} / \mathrm{m}$ respectively. Find the generator frequency $f$ and $\varepsilon_{r}$ and $\mu_{r}$ for the material.
20. Find the attenuation constant $\alpha(\mathrm{Np} / \mathrm{m})$, the phase constant $\beta(\mathrm{rad} / \mathrm{m})$ and the phase velocity $\mathrm{v}_{\boldsymbol{\Phi}}$ if the conductivity $\sigma$ of the material is such that $\sigma=\omega \varepsilon$ (the material parameters are $\mu_{\mathrm{r}}=1$ and $\varepsilon_{\mathrm{r}}=2.5$ ). The wavelength in free space is $\lambda_{0}=30 \mathrm{~cm}$.
21. Convert the phase constant $\beta$ into $(\% / m)$ and the attenuation constant $\alpha$ into ( $\mathrm{dB} / \mathrm{m}$ ) for problem 20.
22. At what frequencies may earth be considered a perfect dielectric, if $\sigma=5 \times 10^{-3} \mathrm{~S} / \mathrm{m}, \mu_{\mathrm{r}}=1$ and $\varepsilon_{\mathrm{r}}=8$ ? Can $\alpha$ be neglected at these frequencies? What about the characteristic impedance $\mathrm{Z}_{\mathrm{c}}$ ?
23. Find the skin depth $\delta$ at a frequency 4 MHz in aluminum, where $\sigma=3.82 \times 10^{7} \mathrm{~S} / \mathrm{m}$ and $\mu_{\mathrm{r}}=\varepsilon_{\mathrm{r}}=1$. Also find the phase velocity $\mathrm{v}_{\mathrm{D}}$.
24. Show that Maxwell's equations can be cast in the form of a one-dimensional diffusion equation (provided the direction of propagation is $z$ )

$$
\frac{\partial^{2} E_{Y}}{\partial z^{2}}-\mu \sigma \frac{\partial E_{Y}}{\partial t}=0
$$

Describe when this derivation might be valid.

## Problems 5

25. With the substitution $\xi=z / \sqrt{D t}$ and using the chain rule, show that the partial differential equation given in problem 24 will transform into an ordinary differential equation. Find the units of the diffusion coefficientD $=1 / \mu \sigma$.
26. Solve the ordinary differential equation obtained in problem 25 with the boundary condition that the electric field goes to 0 as $z \rightarrow \infty$ and the integral over all space of the electric field is a constant. In addition, the electric field in the region external to the point where it is excited is equal to 0 just after the excitation. This can be interpreted as the diffusion of a "pulse."
27. An electromagnetic wave with an amplitude of $1 \mathrm{~V} / \mathrm{m}$ is normally incident from a vacuum into a dielectric having a relative dielectric constant $\varepsilon_{r}=4$. Find the amplitude of the reflected and the transmitted electric fields and the incident, reflected and transmitted powers.
28. An electromagnetic wave with an amplitude of $1 \mathrm{~V} / \mathrm{m}$ is normally incident from a vacuum into a dielectric having a relative dielectric constant $\varepsilon_{r}=4$. Find the amplitude of the reflected and the transmitted magnetic field intensities.
29. An electromagnetic wave with an amplitude of $1 \mathrm{~V} / \mathrm{m}$ is normally incident from a dielectric having a relative dielectric constant $\varepsilon_{r}=4$ into a vacuum. Find the amplitude of the reflected and the transmitted electric fields and the incident, reflected and transmitted powers.
30. An electromagnetic wave with an amplitude of $1 \mathrm{~V} / \mathrm{m}$ is normally incident from a dielectric having a relative dielectric constant $\varepsilon_{r}=4$ into a vacuum. Find the amplitude of the reflected and the transmitted magnetic field intensities.

## Problems 5

31. In Example 5-17, a speeder pleads to the judge that because of inclement weather when the radar was tested and calibrated, the calibration was incorrect. If the radar assumed a calibration in a vacuum and said the speeder was traveling at $25 \%$ over the speed limit, what would the dielectric constant of ambient space have to be in order that the defendant would go free?
32. A time-harmonic electromagnetic wave in a vacuum is incident upon an ideal conductor located at $\mathrm{z}=0$ and a standing wave is created in the region $\mathrm{z} \leq 0$. With a crystal detector connected to a volt meter, we measure a null voltage at equal increments of 10 cm in the region $\mathrm{z} \leq 0$. Find the frequency of oscillation of the electromagnetic wave.
33. Compute the skin depth of copper, graphite and seawater at $\mathrm{f}=2.45 \mathrm{GHz}$.
34. A fisherman in the sea detects a fish at a depth d with a radar operating at a frequency f . Find d if the delaytime of the reflected signal is $\tau=20$ ns. Assume that fish scales are perfect conductors and the conductivity of the water $\sigma$ satisfies the condition
 $\sigma \ll \omega \varepsilon$.
35. Estimate the number of wavelengths of helium-neon laser light ( $\lambda=6328 \AA$ ) where ( $1 \AA=10^{-10} \mathrm{~m}$ ) that can be found between the two parallel end plates which are separated by 1 m . You may assume $\varepsilon_{\mathrm{r}} \approx 1$ between the end plates.

## Problems 5

36. The resonant frequency of a Fabry-Perot cavity caused by the introduction of a dielectric is changed from its vacuum value of 10 GHz to 9.9 GHz . Calculate the relative dielectric constant $\varepsilon_{r}$ of the perturbing material.
37. The relative dielectric constant of a slice of lung of thickness $\Delta \mathrm{L}$ is found to be 1.5. A diseased lung of thickness $\Delta \mathrm{L}$ as shown in Figure $5-16$ is inserted between the plates of a Fabry-Perot cavity. The cavity has a resonant frequency of 9.9 GHz for the undiseased lung and 9.95 GHz for the diseased lung. Find the percentage $p$ of the diseased lung that has been eaten away by emphysema.
38. A plane wave
$E(z, t)=u_{Y} E_{0} \cos (\omega t-k z)$ is incident upon two air-dielectric interfaces. Determine the thickness $d$ of the dielectric slab $\left(\varepsilon_{r}\right)$ that would make the field in the region $z<0$ the same as the slab were not there.

## Problems 5

39. A dielectric slab $\left(\varepsilon_{\mathrm{r}}\right)$ is inserted between two plane wave launching horns. Waves will be reflected and transmitted at each interface. Determine the ratio of $E_{B} / E_{A}$ for a wave that passes through the region b. When it takes a maximum?
40. A dielectric that is $\lambda / 4$ thick separates two dielectrics. Find the values of $\varepsilon_{r}$ so none of the power launched from $A$ will be reflected back to $A$.

41. For the infinite parallel plate waveguide depicted in Figure 5-19, determine why an initial assumption for an electric field $E=E_{x}(x, y) \mathbf{u}_{\mathbf{x}}$ will not lead to normal modes.

## Problems 5

42. Show that the angle $\theta$ between the electric field component wave direction and the conducting sheets in a dielectric filled parallel plate waveguide can be computed from $\theta=\sin ^{-1}(\lambda / 2 a)$ where $a$ is the separation between the two plates.

43. Show that the phase velocity $\mathrm{v}_{\Phi}$ can be written as

$$
\mathrm{v}_{\Phi}=\frac{\mathrm{c}}{\cos \theta}=\frac{1}{\sqrt{\varepsilon \mu} \cos \theta}
$$

where the angle $\theta$ is defined in problem 42.
44. Calculate the cutoff frequency for the two lowest-order modes of a parallel plate waveguide where the plates are separated by 3 cm . The region between the plates is filled with (a) paper or (b) glass.
45. Show that a parallel plate waveguide operating at a frequency equal to the cutoff frequency of a higher order mode can be interpreted in terms of a Fabry Perot resonator.
$46^{\text {MATLAB }}$. Repeat Example 5-6 with $(h / c t)=1 / 2$.
$47^{\text {MATLAB }}$. Repeat Example 5-6 with $(h / c t)=3 / 2$.
$48^{\text {MATLAB }}$. The initial condition at $t=0$ for a wave is

$$
\psi(z, 0)=5 \exp (-10 z) .
$$

Write a MATLAB program to show the propagation of the wave in the region $0 \leq \mathrm{z} \leq 1$ if $\mathrm{c}=1$.

## Problems 5

$49^{\text {MATLAB }}$. The initial condition at $t=0$ for a wave is

$$
\psi(z, 0)=5 \exp (-10 z)
$$

Write a MATLAB program to show the propagation of ths wave in the region $0 \leq z \leq 2$ if $c=2$.
$50^{\text {MATLAB }}$. The initial condition at $t=0$ for a wave is

$$
\psi(z, 0)=5 \exp (-10 z)
$$

Write a MATLAB program to show the propagation of ths wave in the region $0 \leq \mathrm{z} \leq 0.5$ if $\mathrm{c}=0.5$.

## Equivalent electrical circuits

## 6. Transmission lines

In this chapter, we will model a coaxial cable, a strip line, and two parallel wires with distributed circuit elements and study the voltage and current signals that can propagate along a transmission line using these distributed elements. The circuit elements include inductors, capacitors, and resistors. The equations that we will derive are called the Telegraphist's equations since they were originally derived in order to describe the signals that propagated on the lengthy cables that spanned the country from one end to the other forming the nation's early communication network. Using the Telegraphist's equations, we will derive a wave equation that will describe the wave propagation on the transmission line. With the rapid development of integrated circuit technology and VLSI design, we now can even suggest that electromagnetic waves are also guided in such small structures and the propagation of the signals will be governed by the same one dimensional wave equation. In our initial discussion of transmission lines, we will assume that the wave will suffer no loss as it propagates and will therefore not include any resistors in the introductory model. Loss terms will be included after we develop the lossless transmission line.

### 6.1. Equivalent electrical circuits

The three types of transmission lines that have received considerable attention in practice are depicted in Figure 6-1. These three lines are: (a) the coaxial cable - this structure usually has a dielectric that separates the inner conductor from the outer one; (b) the strip line or microstrip line - this structure has a dielectric that separates the two flat conductors; and (c) two parallel conducting wires or the twin lead - this structure may have a vacuum or a dielectric that separates the two wires. The dielectric in these transmission lines

## Equivalent electrical circuits

is used because of mechanical considerations imposed by the requirement that long metallic structures remain separated. In our everyday experiences we recognize these lines as being transmission lines that respectively connect the cable TV into the house, connect two components in an integrated circuit within the TV and provide the connection between a TV and an outside or a rabbit-ears antenna.


Figure 6-1. Three common transmission lines. (a) coaxial cable; (b) microstrip line; (c) two wire line (sometimes called a "twin lead.").

Rather than examine in full detail the electromagnetic field distribution within these transmission lines at this stage, we will simplify our discussion by just suggesting a simple model consisting of distributed inductors and capacitors can be employed. This model will be valid if any dimension that is transverse to the direction of propagation is much less than the free space wavelength. If the dimension is comparable with the wavelength, then a more complicated analysis will be required that may include the various numerical tools that were introduced in Chapter 3.

## Equivalent electrical circuits

In the three transmission lines depicted in Figure 6-1, such an analysis would show that the three lines support the propagation of a wave that has both the electric field intensity and the magnetic field intensity in a plane that is transverse to the direction of propagation. This is sometimes called a TEM mode of propagation. The analysis would have required that there be no losses in either the dielectric separating the conductors or in the conductors themselves. In order for us to make this simplification in the mode structure, we have to first define the capacitance $C$ and inductance $L$ in terms of electromagnetic fields. Fortunately, we have already obtained these results for the coaxial cable and the microstrip line in Chapter 2. The capacitance and inductance for the two parallel wires is given in Appendix 4. The parameters for the three transmission lines are summarized in Table 6-1. We note that the factor $\Delta z$ can be separated from the other terms for the three structures.

|  | Inductance | Capacitance |
| :--- | :--- | :--- |
| coaxial cable | $\mathrm{L}=\frac{\mu}{2 \pi} \ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right) \Delta \mathrm{z}$ | $\mathrm{C}=\frac{2 \pi \varepsilon}{\ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right)} \Delta \mathrm{z}$ |
| microstrip line | $\mathrm{L}=\frac{\mu \mathrm{d}}{\mathrm{w}} \Delta \mathrm{z}$ | $\mathrm{C}=\frac{\varepsilon \mathrm{W}}{\mathrm{d}} \Delta \mathrm{z}$ |
| twin lead | $\mathrm{L}=\frac{\mu}{\pi} \cosh ^{-1}\left(\frac{\mathrm{D}}{2 \mathrm{a}}\right) \Delta \mathrm{z}$ | $\mathrm{C}=\frac{\pi \varepsilon}{\cosh ^{-1}\left(\frac{\mathrm{D}}{2 \mathrm{a}}\right)} \Delta \mathrm{z}$ |

Table 6-1. Electrical circuit elements for the various transmission lines shown in
Figure 6-1. For the microstrip line, we are using the simplified expressions that have been obtained in Chapter 2. A more accurate expression could be obtained using numerical techniques which would include the effects of fringing fields. The parameters for the material between the two conductors are permeability $\mu=$ $\mu_{\mathrm{r}} \mu_{\mathrm{o}}$ and the permittivity $\varepsilon=\varepsilon_{\mathrm{r}} \varepsilon_{0}$.

## Equivalent electrical circuits

We can, therefore, model the three transmission lines depicted in Figure 6-1 with an equivalent circuit consisting of an infinite number of distributed inductors and capacitors as shown in Figure 6-2a. We could think of constructing such a distributed line by wrapping wire uniformly about a broom stick and locating it at a constant distance above a ground plane. In addition to the uniformly distributed inductance, there would be a uniformly distributed capacitance between the wires of the coil and the ground plane. If the spirit moved us, we could actually carry on the tedious task of soldering the infinite number of uniformly distributed capacitors from the coil to the ground plane.
(a)

(b)


Figure 6-2. (a) Distributed transmission line. (b) Equivalent circuit of this transmission line. The circuit elements are given in their per unit length values. In each section, the values are $\mathrm{L}=\hat{\mathrm{L}} \Delta \mathrm{z}$ and $\mathrm{C}=\hat{\mathrm{C}} \Delta \mathrm{z}$ respectively. Hence, transmission line models can be easily constructed in the laboratory.

The model for the distributed transmission line that is depicted in Figure 62 has incorporated some obvious simplifications. In particular, there are no elements that would describe any loss of energy as the wave propagates under

## Equivalent electrical circuits

transmission line. This can be incorporated with a resistor in series with the inductor or with a conductance in parallel with the capacitor. The effects of these additional elements will be described later. In addition, parasitic capacitances exist between the wires that constitute the distributed inductance in Figure 6-2a. These capacitances will initially be assumed to be very small and can be neglected at this stage of our discussion. We will see later that the inclusion of these elements will give rise to a wave number and hence a phase velocity that is dependent upon the frequency of the wave. This phenomena is called dispersion and it will be discussed at the end of this chapter. The equivalent circuit that we will use at the present time is depicted in Figure 6-2b. The reader who is comfortable with circuit theory could consider the transmission line as a large number of distributed two port networks and be guided through well worn and understood techniques. We choose, however, not to follow this path in order to interpret the signals in terms of waves. In particular, we choose the path that will lead to a coupled set of first order partial differential equations that are called the Telegraphist's equations. This set of equations will describe the temporal and spatial evolution of voltage and current signals along this transmission line. In addition, we can manipulate the set and obtain the wave equation whose solution has been described in the previous chapter.

# Dispersion and group velocity 

### 6.10. Dispersion and group velocity

In our first model for a distributed transmission line, we considered that a wire was wrapped uniformly around a cylindrical rod and distributed capacitors connected this wire to a ground plane. Effects due to the finite conductivity of the wire or leakage currents through the capacitor introduced loss mechanisms which could lead to waves whose amplitude would decrease as they propagated on this transmission line. In this section, we will examine another linear effect which is found in cases where the wavelength is comparable with the physical dimensions of the transmission line or in cases where the separating dielectric depends upon the frequency. For example, this could be understood as the inclusion of the capacitance between the wires that had been wrapped uniformly around the rod. As we will find, this will limit the frequency response of the transmission line and will lead to a new velocity of propagation that is called the group thevelocity. Although there are several different transmission lines that exhibit this effect, we will just focus our attention on one of them.


Figure 6-18. A model of a section of a transmission line whose length is $\Delta z$ that includes dispersion is shown.

## Dispersion and group velocity

A simple model consisting of linear elements that we will use to introduce the concept of dispersion is shown in Figure 6-18. Following the same procedure that we have used previously, we will be able to write down the Telegraphist's equations that are applicable for a transmission line consisting of a large number of identical sections that are shown in Figure 6-18. There is one caveat that is immediately encountered, the current that enters the node at the left will subdivide into a current $I_{L}$ that passes through the inductor and a current that passes through the capacitor $I_{c}$ that is in parallel with this inductor

$$
\begin{equation*}
\mathrm{I}=\mathrm{I}_{\mathrm{L}}+\mathrm{I}_{\mathrm{C}} \tag{6.80}
\end{equation*}
$$

The appropriate equation that describes the voltage drop across the inductor is

$$
\begin{equation*}
\frac{\partial \mathrm{V}(\mathrm{z}, \mathrm{t})}{\partial \mathrm{z}}=-\hat{\mathrm{L}} \frac{\partial \mathrm{I}_{\mathrm{L}}(\mathrm{z}, \mathrm{t})}{\partial \mathrm{t}} \tag{6.81}
\end{equation*}
$$

The equation that describes the voltage drop across the capacitor that is in parallel with this inductor is

$$
\begin{equation*}
\frac{\partial V(z, t)}{\partial z}=-\frac{1}{\hat{C}_{s}} \int I_{c} d t \tag{6.82}
\end{equation*}
$$

The units of this additional capacitor are F.m rather than F/m. The equation that describes the current that passes through the shunt capacitor is given by

$$
\begin{equation*}
\frac{\partial \mathrm{l}(\mathrm{z}, \mathrm{t})}{\partial \mathrm{z}}=-\hat{\mathrm{C}} \frac{\partial \mathrm{~V}(\mathrm{z}, \mathrm{t})}{\partial \mathrm{t}} \tag{6.83}
\end{equation*}
$$

From the set of equations (6.80) - (6.83), we derive the following wave equation

$$
\begin{equation*}
\frac{\partial^{2} V(z, t)}{\partial z^{2}}-\hat{L} \hat{C} \frac{\partial^{2} V(z, t)}{\partial t^{2}}+\hat{L} \hat{C}_{s} \frac{\partial^{4} V(z, t)}{\partial z^{2} \partial t^{2}}=0 \tag{6.84}
\end{equation*}
$$

Let us assume that there is a time harmonic signal generator that is connected to this transmission line that is infinitely long. The wave that is excited and propagates on this line is of the form

$$
\begin{equation*}
V(z, t)=V_{0} e^{j(\omega t-\beta z)} \tag{6.85}
\end{equation*}
$$

## Dispersion and group velocity

The substitution of (6.85) leads to the dispersion relation that relates the propagation constant $\beta$ to the frequency of the wave $\omega$. We obtain

$$
\begin{equation*}
\left.l(-j \beta)^{2}-\hat{L} \hat{C}(j \omega)^{2}+\hat{L} \hat{C}_{s}(j \omega)^{2}(-j \beta)^{2}\right] N_{0} e^{j(\omega t-\beta z)}=0 \tag{6.86}
\end{equation*}
$$

where the terms within the square bracket yield the dispersion relation. We write this as

$$
\begin{equation*}
\beta= \pm \frac{\omega \sqrt{\hat{\mathrm{L}} \hat{\mathrm{C}}}}{\sqrt{1-\omega^{2} \hat{\mathrm{~L}}_{\mathrm{s}}}} \tag{6.87}
\end{equation*}
$$

The propagation constant is a nonlinear function of frequency as shown in Figure 6-19a.

Example 6-20. Derive this dispersion relation (6.87) using (6.77)
Answer: From (6.77), we write
$\gamma=\alpha+j \beta= \pm \sqrt{\hat{Z} \hat{Y}}= \pm \sqrt{\left(\frac{(j \omega \hat{L})\left(\frac{1}{j \omega \hat{C}_{s}}\right)}{j \omega \hat{L}+\frac{1}{j \omega \hat{C}_{s}}}\right)(j \omega \hat{C})}= \pm \sqrt{\frac{\frac{j \omega \hat{L} \hat{C}}{\hat{C}_{s}}}{j\left(\omega \hat{L}^{-}-\frac{1}{\omega \hat{C}_{s}}\right)}}= \pm j \frac{\omega \sqrt{\hat{L} \hat{C}}}{\sqrt{1-\omega^{2} \hat{L} \hat{C}_{s}}}$
Since this expression is purely imaginary, $\alpha=0$ and there are no losses as the wave propagates. This is true only if the frequency $\omega$ is below a certain frequency called the "cutoff frequency."

Our final result that was obtained using two different approaches is interesting since we have found that the propagation constant depends on frequency. This is called dispersion. The consequences of dispersion will dramatically influence the propagation of waves as we have already encountered in the previous chapter. The dispersion relation shown in Figure 5-22 is different than that in Figure 6-19a. One could think of the difference between a high pass and a low pass filter in describing these two dispersion relations. The propagation constant will be a real number for frequencies from 0 up to a value that is called the cutoff frequency $\omega_{0}$ where

## Dispersion and group velocity

$$
\begin{equation*}
\omega_{0}=\frac{1}{\sqrt{\hat{L} \hat{C}_{s}}} \tag{6.88}
\end{equation*}
$$




Figure 6-19 ${ }^{\text {MATLAB }}$. The normalized propagation characteristics of a dispersive transmission line. (a) The dispersion relation (6.87) is the solid line. The dashed line is a non dispersive propagation constant. (b) The phase velocity as a function of frequency.

This cutoff frequency is equal to the resonant frequency of the "tank" circuit in the series arm. At this resonant frequency, the tank circuit will appear to be an infinite impedance. Above this frequency, the propagation constant will be imaginary and the wave will not propagate. In addition, the velocity of propagation $v_{o}$ in the non dispersive frequency range is given by

$$
\begin{equation*}
v_{0}=\frac{1}{\sqrt{\hat{L} \hat{C}}} \tag{6.89}
\end{equation*}
$$

## Dispersion and group velocity

The wave number $\beta_{o}$ in this region is

$$
\begin{equation*}
\beta_{o}=\frac{\omega_{0}}{v_{o}} \tag{6.90}
\end{equation*}
$$

. Dispersive transmission lines or a dispersive media in general are very important and common in practice. Dispersion may be created by finite transverse dimensions as in waveguides or inter atomic dimensions as in materials.

Dispersion implies that the propagation constant depends upon the frequency of oscillation. The relationship between these two quantities is depicted in Figure 6-20. We find that the curvature of the resulting dispersion curve can have different slopes and these are referred to as "positive dispersion" and "negative dispersion".


Figure 6-20 MATLAB . Dispersion curves.

## Dispersion and group velocity

Electromagnetic waves that are confined to propagate within metallic structures can only propagate if the frequency is above some "cut off frequency" that could be determined by the physical dimensions of the structure. Recall that we encountered this topic in our discussion of plane waves. In the case of negative dispersion, a wave with a frequency from 0 to some "cut off frequency" will propagate. Above this frequency, the wave will not propagate. There are various longitudinal waves in ionized gases that is called a plasma in which such waves exist.

The question then arises, "What will happen if there are two signals that are propagating in the same linear media but with slightly different frequencies?" The answer to this question will lead us to the development of a new velocity that is called the group velocity. We have already learned that a narrow pulse can be examined using a Fourier analysis and the pulse would consist of a number of high frequency components. If this pulse propagates in a dispersive region, it will be difficult if not impossible to reconstruct the original pulse at a later stage.

In order answer the question that was just posed, we consider 2 waves that each have the same amplitude $\mathrm{V}_{0}$ and the following frequencies of oscillation

$$
\begin{equation*}
\omega_{1}=\omega_{0}+\Delta \omega \text { and } \omega_{2}=\omega_{0}-\Delta \omega \tag{6.91}
\end{equation*}
$$

Corresponding to each of these frequencies, a signal that propagates in the positive $z$ direction is excited. From either of the dispersion curves in Figure $6-20$, the corresponding propagation constants are obtained to be

$$
\begin{equation*}
\beta_{1}=\beta_{0}+\Delta \beta \text { and } \beta_{2}=\beta_{0}-\Delta \beta \tag{6.92}
\end{equation*}
$$

Since superposition will apply, the two waves can be added together in order to find the sum total of the response.

$$
\begin{equation*}
V=V_{0}\left[\cos \left(\omega_{1} t-\beta_{1} z\right)+\cos \left(\omega_{2} t-\beta_{2} z\right)\right] \tag{6.93}
\end{equation*}
$$

Let us apply a trigonometric identity to (6.93) and obtain

$$
\begin{equation*}
V=2 V_{o} \cos (\Delta \omega t-\Delta \beta z) \cos \left(\omega_{0} t-\beta_{o} z\right) \tag{6.94}
\end{equation*}
$$

## Dispersion and group velocity

In Figure 6-21, we illustrate the summation procedure by just adding 2 cosine signals with slightly different frequencies together.




Figure 6-21 MATLAB . The linear summation of two cosine waves that have a slightly different frequency of oscillation. There is a constructive and destructive interference between these two signals upon addition.

If we examine these two signals at various locations that are equally spaced as shown in Figure $6-22$, we will be able to ascertain the velocity of propagation of a point of constant phase and the velocity of this modulation. The points of constant phase yields the phase velocity $v_{\phi}=\frac{\omega_{o}}{\beta_{o}}$. The velocity of the amplitude modulation is $\mathrm{v}_{\mathrm{g}} \approx \frac{\Delta \omega}{\Delta \beta}$. The velocity of the modulation in the limit

## Dispersion and group velocity

$\mathrm{v}_{\mathrm{g}}=\frac{\Delta \omega}{\Delta \beta} \rightarrow \frac{\partial \omega}{\partial \beta}$ is called the group velocity of the wave. In a non-dispersive media, the two velocities are identical. However, in a dispersive media, they can be vastly different and can even have the opposite sign.


Figure 6-22 ${ }^{\text {MATLAB }}$. The propagation of a signal in a dispersive media. The signals are detected at two locations. A point of constant phase and the peak of the envelope is followed. The point of constant phase propagates with the phase velocity and the modulation envelope propagates with the group velocity. In this figure, the phase velocity is greater than the group velocity.

An illustration of the signals that are detected at increasing values of the distance is shown in Figure 6-22. A point of constant phase can be followed and it will yield data to compute the phase velocity. The envelope of the modulating

## Dispersion and group velocity

signal will propagate with the group velocity. Frequently, this modulating signal will spread as it propagates causing the detected signal to become vastly distorted. For example, a very narrow pulse excitation that contains a very large number of frequency components and is propagating in a dispersive media could appear as a time harmonic signal with only one frequency component at distances that are far from the point of excitation.

Example 6-21. (a) Find the phase and group velocities for a normal transmission line depicted below. (b) Find the phase and group velocities for a transmission line in which the circuit elements are interchanged.


Answer: (a) From (6.77), we write

$$
\gamma=\alpha+\mathrm{j} \beta=\sqrt{\hat{Z} \hat{Y}}=\sqrt{(j \omega \hat{L})(\mathrm{j} \omega \hat{\mathrm{C}})}=\mathrm{j} \omega \sqrt{\hat{\mathrm{~L}} \hat{\mathrm{C}}}
$$

The phase velocity $\mathrm{v}_{\phi}$ is computed to be

$$
v_{\phi}=\frac{\omega}{\beta}=\frac{1}{\sqrt{\hat{\mathrm{~L}} \hat{C}}}
$$

The group velocity $v_{g}$ is computed to be

$$
\mathrm{v}_{\mathrm{g}}=\frac{\partial \omega}{\partial \beta}=\frac{1}{\left(\frac{\partial \beta}{\partial \omega}\right)}=\frac{1}{\sqrt{\hat{L} \hat{C}}}
$$

The two velocities are equal in this case and our independent of frequency.
(b) From (6.77), we write

$$
\gamma=\alpha+j \beta=\sqrt{\hat{Z} \hat{Y}}=\sqrt{\left(\frac{1}{j \omega \hat{C}}\right)\left(\frac{1}{j \omega \hat{L}}\right)}=-j \frac{1}{\omega \sqrt{\hat{L} \hat{C}}}
$$

The phase velocity $\mathrm{v}_{\phi}$ is computed to be

# Dispersion and group velocity 

$$
v_{\phi}=\frac{\omega}{\beta}=-\omega^{2} \sqrt{\hat{L} \hat{C}}
$$

The group velocity $v_{g}$ is computed to be

$$
\mathrm{v}_{\mathrm{g}}=\frac{\partial \omega}{\partial \beta}=\frac{1}{\left(\frac{\partial \beta}{\partial \omega}\right)}=+\omega^{2} \sqrt{\hat{L} \hat{C}}
$$

In this case, the phase and the group velocities are in the opposite direction and both of them depend on frequency.

Suffice it to say, dispersion has very dramatic effects on the propagation of electromagnetic waves. If the media, in addition to being dispersive, were also nonlinear in that the velocity of propagation depended upon the amplitude of the propagating wave, then it might be possible to have nonlinear waves that are called solitons propagating in this media. A nonlinear dispersive transmission line can be constructed by replacing the linear shunt capacitor in Figure 6-18 with a capacitance whose value depends upon the local value of the voltage of the wave. This is also called a nonlinear varactor diode. This is a topic of current research interest in several scientific and engineering communities.

## Conclusion

### 6.11. Conclusion

Transmission lines that we have studied in this chapter are very important from several points of view. They are important in their own right and they can be used to model other forms of transmission media. Several structures that are in wide use such as a coaxial cable, a strip line and two parallel wires can be modeled with a structure that consists of distributed inductors and capacitors. Summing the voltage drops around a loop and the current elements entering a node leads to two first order partial differential equations that are known as the Telegraphist's equations. Eliminating one of the dependent variables between these two equations led to a wave equation.

Sinusoidal waves propagated along this line to either increasing or decreasing values of the position coordinate if initial transient effects in the sinusoidal excitation could be neglected. The wave would repeat itself in a distance called the wavelength. The ratio of the voltage wave propagating in one direction to the current wave propagating in the same direction is the characteristic impedance of the transmission line. Terminating the transmission line with either a load impedance or another transmission line introduced the concepts of reflection and transmission coefficients, standing waves, the VSWR, and the idea of matching. The Smith chart facilitated matching. Transient effects and their subsequent propagation along with pulse propagation were analyzed with a bounce diagram. The final asymptotic state of a transmission line excited by a step voltage was found. Finally, the effects of loss and dispersion were analyzed.

## Problems 6

### 6.12. Problems

1. Find an equivalent circuit for plane wave propagation in a vacuum.
2. Show that the equivalent circuit element parameters for the coaxial cable and the strip line are correct representations.
3. Show that the telegraphist's can be derived using an argument based on a Taylor series.
4. Show that the quantity $v=\frac{1}{\sqrt{\hat{L} \hat{C}}}$ does indeed have the units of a velocity.
5. Show that the units of the diffusion coefficient $D=\frac{1}{\hat{R} \hat{C}}$ do indeed have the units of (length) ${ }^{2} /$ time.
6. Show that a function that represents a wave that propagates to decreasing values of $z$ satisfies the wave equation (6.7).
7. Let us replace the linear capacitors in Figure 6-3 with nonlinear varactor diodes whose capacitance depends upon the voltage applied across it, In this case, the current $\Delta \mathrm{I}$ into the diode can be written as $\Delta \mathrm{I}=\frac{\partial \mathrm{Q}(\mathrm{V})}{\partial \mathrm{t}}$. Derive the resulting wave equation for this transmission line.
8. Show that the two equations given in (6.15) and (6.16) are equivalent and find expressions for the constants in one equation in terms of the constants in the other.
9. Find an expression for the characteristic impedance of the strip line.
10. Find an expression for the characteristic impedance of a twin lead.

## Problems 6

11. In an integrated circuit, a dielectric with $\varepsilon_{\mathrm{r}}=2$ is inserted between two metal conductors. The width of the top metal strip is $10 \mu \mathrm{~m}$ and its separation from the bottom grounded metal plane is $5 \mu \mathrm{~m}$. Find the characteristic impedance of this transmission line and the velocity of a signal.
12. Design a strip line with a glass insulator that will have a characteristic impedance of $5 \Omega$. You will have some freedom in this design but there is one constraint - it is to be used in an integrated circuit.
13. A TV twin lead consists of two parallel 1 mm copper wires separated by 1 cm of a rubber dielectric with $\varepsilon_{r}=3$. What is the capacitance per meter of this twin lead and what is its characteristic impedance?
14. Prove that the voltage that appears across a load impedance will be less than the incident wave if $\mathrm{Z}_{\mathrm{L}}<\mathrm{Z}_{\mathrm{C}}$.
15. A VSWR is measured along a transmission line to be 2 . Find two values for the reflection coefficient $\mathcal{R}$. Which of these values will correspond to $Z_{L}<Z_{C}$ and which to $\mathrm{Z}_{\mathrm{L}}>\mathrm{Z}_{\mathrm{C}}$ ?
16. For problem 15 , find the two values of $Z_{L}$ if $Z_{C}=50 \Omega$.
$17^{\text {MATLAB }}$. A load impedance $Z_{L}=25 \Omega$ is connected to a transmission line whose characteristic impedance is $50 \Omega$. Using (6.33), plot the impedance as a function of a distance from the load to a total distance of $2 \lambda$.

18 Using (6.33), prove that the load impedance will repeat itself every $\lambda / 2$.
19. Using (6.33), prove that the input impedance of a transmission line terminated in a short circuit with a length $L$ where $\lambda / 4<L<\lambda / 2$ is capacitive.

## Problems 6

20. Using (6.33), prove that the input impedance of a transmission line terminated in a open circuit with a length $L$ where $\lambda / 4<L \lambda / 2$ is inductive.
21. An air-filled $50 \Omega$ coaxial cable that is 1 m long is excited with a 300 MHz signal generator. The line is terminated with a load impedance $Z_{L}=(25+j 25) \Omega$. What is the input impedance of this line?
22. A shorted $50 \Omega$ transmission line of length $L$ has an input admittance of - j 0.01 S. Find the length of the line in $\lambda$.
23. Using (6.46) and (6.47), draw the circles for the values of $r=0,1$, and $\infty$ and $x=-1,0$, and 1 to convince yourself that Figure $6-9$ is correct.
$24^{\text {MATLAB }}$. Using a Smith chart, find the impedance $Z_{i n}$ of a $50 \Omega$ coaxial cable that is terminated in a load $Z_{L}=(25+j 25) \Omega$. The coaxial cable has a length of $3 \lambda / 8$.
$25^{\text {MATLAB }}$. Using a Smith chart, find the admittance $Y_{\text {in }}$ of a $50 \Omega$ coaxial cable that is terminated in a load $Z_{L}=(25+j 25) \Omega$. The coaxial cable has a length of $\lambda / 8$.
$26^{\text {MATLAB }}$. Using a Smith chart, find the distance from a load impedance $Z_{L}=(25+j 25) \Omega$ that is connected to a $50 \Omega$ coaxial cable where the normalized input admittance $Y_{\text {in }}=1+j B_{\text {in }}$. How long should a transmission line that is terminated in a short circuit be in order that the transmission will be matched?
24. A load impedance $Z_{L}=(100-j 100) \Omega$ terminates a $50 \Omega$ transmission line.

Find the characteristic impedance of the quarter-wavelength matching transmission line.

## Problems 6

28. A lossless battery is connected to an ideal transmission line with a characteristic impedance $Z_{C}$ of length $L$ that is terminated in a short circuit. Sketch the potential at $z=L / 2$ as a function of time $0 \leq t \leq 4(\mathrm{~L} / \mathrm{v})$. The switch is closed at $\mathrm{t}=$ 0.

29. Sketch the current profile at $z=L / 2$ as a function of time $0 \leq t \leq 4(L / v)$ at $z=L / 2$ for the transmission line stated in problem 28.
30. A lossless battery is connected to an ideal transmission line with a characteristic impedance $Z_{C}$ of length $L$ that is terminated in an open circuit. Sketch the potential at $z=L / 2$ as a function of time $0 \leq t \leq 4(\mathrm{~L} / \mathrm{v})$. The switch is closed at $\mathrm{t}=$ 0.

31. Sketch the current profile at $z=L / 2$ as a function of time $0 \leq t \leq 4(L / v)$ at $z=L / 2$ for the transmission line stated in problem 30.

## Problems 6

32. Two transmission lines are joined with a resistor $R_{L}$.


Show that the transmitted voltage $\mathrm{V}_{\mathrm{T}}$ in line two can be written as

$$
V_{T}=\frac{Z_{C 2}}{R_{L}+Z_{C 2}}(1+R) V_{\text {inc }} \text { where } R=\frac{R_{L}+Z_{C 2}-Z_{C 1}}{R_{L}+Z_{C 2}+Z_{C 1}}
$$

if the voltage incident from $z=-\infty$ is $V_{\text {inc }}$.
33. At $\mathrm{t}=0$, the switch located at the load is closed. Sketch the voltage and the current at the load as a function of time $(\mathrm{L} / \mathrm{v})$. The impedance at the load $\mathrm{R}_{\mathrm{L}}=$ the characteristic impedance of the line $Z_{C}$.

34. A transmission line with two switches, one at the battery and one at the load is shown below. Initially, switch $S_{1}$ is closed and switch $S_{2}$ is open. At $t=0, S_{1}$ is opened and $S_{2}$ is closed. Sketch the voltage $V_{a a}$ as a function of time.

## Problems 6



If the load impedance were located at the midpoint of this line, a large voltage pulse could be delivered across it. This is called a Blumlein transmission line.
35. A pulse generator is connected to a transmission line of length $L=2 \mathrm{~m}$ having $Z_{c}=50 \Omega, R_{L}=20 \Omega$, and $R_{g}=30 \Omega$. The velocity of propagation in this transmission line is equal to $10^{8} \mathrm{~m} / \mathrm{s}$. The amplitude of the pulse is 1 V and its width is $10^{-9} \mathrm{~s}$. Plot the voltage at $\mathrm{z}=\mathrm{L} / 2$ as a function of time, $0 \leq \mathrm{t} \leq 100 \mathrm{~ns}$.

36. In a digital computer, it is desired to transmit a sequence of binary pulses from a pulse generator to another point where they are to be sampled. Let us assume that the pulse sequence shown below is launched by the pulse generator shown in Problem 35 and we desire to detect the sequence at a time $T=3 \mathrm{~L} / 2 \mathrm{v}$

## Problems 6

later. Describe any limitations that might be imposed on the speed of this computer.

37. Generalize the results of Problem 36 to a " 32 bit" and a " 64 bit" machine.
$38^{\text {MATLAB }}$. Repeat Problem 35 if the transmission line is lossy and the signal decays as $e^{-\alpha z}$ as it propagates where $\alpha=.01,0.1$, and 1 .
$39^{\text {MATLAB }}$. Sketch the dispersion relation for a transmission line consisting of a series resonant circuit in the series branch and a capacitor in the shunt branch. Calculate the propagation constant using (6.77).

40. Describe the dispersion relation for a transmission line consisting of an inductor in the series branch and a "tank" circuit in the shunt branch.

## Problems 6


41. Repeat problem 39 with the addition of a series resistor $\hat{R}_{s}$ added in series with the inductor. The dispersion relation will be complex.

42. Repeat problem 40 with the addition of a series resistor $\hat{R}_{s}$ added in series with the series inductor. The dispersion relation will be complex.


## Transmission line equations

### 6.2. Transmission line equations

In order to analyze the equivalent circuit of the lossless transmission line, it is simpler to use Kirchhoff's laws rather than Maxwell's equations at this stage. The various currents and voltages are given in Figure 6-3. In order to simplify the notation, we have defined the inductance and capacitance per unit length $\hat{\mathrm{L}}=\frac{\mathrm{L}}{\Delta \mathrm{z}}$ and $\hat{\mathrm{C}}=\frac{\mathrm{C}}{\Delta \mathrm{z}}$ which respectively have the units of henries per unit length and farads per unit length. You may also encounter the notations L' and $C^{\prime}$ or $L_{0}$ and $\mathrm{C}_{0}$ in other books.


Figure 6-3. The lossless transmission line model contains a number of distinct sections. The length of each section is $\Delta z$ and each section contains an inductance and a capacitance. The values are $\hat{L}=\frac{L}{\Delta z}$ and $\hat{C}=\frac{C}{\Delta z}$.

The current that is entering the node at the location $z$ is $I(z)$. From Kirchoff's law, this current will subdivide into a current that goes through the capacitor in that section or goes into the next section. This is expressed as

$$
\begin{equation*}
\mathrm{l}(\mathrm{z}, \mathrm{t})=\hat{\mathrm{C}} \Delta \mathrm{z} \frac{\partial \mathrm{~V}(\mathrm{z}, \mathrm{t})}{\partial \mathrm{t}}+\mathrm{I}(\mathrm{z}+\Delta \mathrm{z}, \mathrm{t}) \tag{6.1}
\end{equation*}
$$

and it can be rewritten as

$$
\begin{equation*}
\frac{I(z+\Delta z, t)-I(z, t)}{\Delta z}=-\hat{C} \frac{\partial V(z, t)}{\partial t} \tag{6.2}
\end{equation*}
$$

## Transmission line equations

In the limit of $\Delta z \rightarrow 0$, the, the term on the left-hand side of (6.2) can be recognized as the definition of a spatial derivative. Therefore, (6.2) becomes

$$
\begin{equation*}
\frac{\partial l(z, t)}{\partial z}=-\hat{C} \frac{\partial V(z, t)}{\partial t} \tag{6.3}
\end{equation*}
$$

Similarly, the sum of the voltage drops in this section can also be calculated using Kirchoff's law and we find

$$
\begin{equation*}
V(z-\Delta z, t)=\hat{L} \Delta z \frac{\partial l(z, t)}{\partial t}+V(z, t) \tag{6.4}
\end{equation*}
$$

Rewriting (6.4), we obtain

$$
\begin{equation*}
\frac{V(z, t)-V(z-\Delta z, t)}{\Delta z}=-\hat{L} \frac{\partial((z, t)}{\partial t} \tag{6.5}
\end{equation*}
$$

Again, the left-hand side of (6.5) is recognized as a definition of a spatial derivative in the limit of $\Delta z \rightarrow 0$ and we write

$$
\begin{equation*}
\frac{\partial V(z, t)}{\partial z}=-\hat{L} \frac{\partial \mid(z, t)}{\partial t} \tag{6.6}
\end{equation*}
$$

The two linear coupled first-order partial differential equations (6.3) and (6.6) are called the "Telegraphist's equations." The origin of this name follows from engineers trying to explain the signal propagation in a cable under the sea from England to France. It also described signal propagation along the great wire that stretched across this country from "sea to shining sea" They are sometimes also referred to as the "Heaviside equations" in honor of the nineteenth century engineer-mathematician who successfully employed mathematical tools that had not yet been proven to be valid. In fact, he developed his own calculus which at the time apparently lacked proper rigor. ${ }^{1}$

We can eliminate one of the dependent variables $V(z, t)$ or $I(z, t)$ from these two coupled first order equations and obtain a second order partial differential

[^18]
## Transmission line equations

equation. This can be written for the voltage $\mathrm{V}(\mathrm{z}, \mathrm{t})$ or the current $\mathrm{I}(\mathrm{z}, \mathrm{t})$. The resulting equations are

$$
\begin{align*}
& \frac{\partial^{2} V(z, t)}{\partial z^{2}}-\hat{L} \hat{C} \frac{\partial^{2} V(z, t)}{\partial t^{2}}=0  \tag{6.7}\\
& \frac{\partial^{2} I(z, t)}{\partial z^{2}}-\hat{L} \hat{C} \frac{\partial^{2} I(z, t)}{\partial t^{2}}=0 \tag{6.8}
\end{align*}
$$

Both of the equations (6.7) and (6.8) are in the form of a standard wave equation that was described in the previous chapter. In the present case, the velocity of propagation $v$ is defined as

$$
\begin{equation*}
v=\frac{1}{\sqrt{\hat{L} \hat{C}}} \tag{6.9}
\end{equation*}
$$

Recall that the units for $\hat{L}$ and $\hat{C}$ are henrys per meter and farads per meter. This implies that (6.9) does indeed have the proper units of the velocity $\frac{\text { meters }}{\text { second }}$. The choice of the symbol $v$ in (6.9) to represent the velocity of propagation in a transmission line is reasonable since this is also the velocity of a plane electromagnetic wave that propagates in the material that separates the conductors in a coaxial cable, a microstrip line, or between the two parallel wires. We will reserve the symbol c for the actual velocity of light.

Example 6-1. Show that a transmission line consisting of distributed linear resistors and capacitors in an appropriate configuration can be used to model diffusion.


## Transmission line equations

Answer: Assume that the resistance and the capacitance per-unit lengths are defined as $\hat{R}=\frac{R}{\Delta z}$ and $\hat{C}=\frac{C}{\Delta z}$ respectively where $\Delta z$ is the size of a section.

The potential drop $\Delta \mathrm{V}$ across the resistor $\mathrm{R}=\hat{\mathrm{R}} \Delta \mathrm{z}$ and the current $\Delta l$ through the capacitor $\mathrm{C}=\hat{\mathrm{C}} \Delta \mathrm{z}$ can be written as

$$
\begin{gathered}
\Delta V(z, t)=I(z, t) \hat{R} \Delta z \\
\Delta I(z, t)=\hat{C} \Delta z \frac{\partial V(z, t)}{\partial t}
\end{gathered}
$$

In the limit of $\Delta z \rightarrow 0$, this reduces to the following set of equations

$$
\begin{gathered}
\frac{\partial V(z, t)}{\partial z}=I(z, t) \hat{R} \\
\frac{\partial l(z, t)}{\partial z}=\hat{C} \frac{\partial V(z, t)}{\partial t}
\end{gathered}
$$

A second order partial differential equation for the potential V follows

$$
\begin{gathered}
\frac{\partial^{2} V(z, t)}{\partial z^{2}}=\hat{R} \frac{\partial(z, t)}{\partial z}=\hat{R}\left(\hat{C} \frac{\partial V(z, t)}{\partial t}\right) \\
\frac{\partial^{2} V(z, t)}{\partial z^{2}}=\hat{R} \hat{C} \frac{\partial V(z, t)}{\partial t}
\end{gathered}
$$

This equation is of the form of a diffusion equation with a diffusion coefficient given by

$$
\mathrm{D}=\frac{1}{\hat{R} \hat{C}}
$$

The dimensions of the diffusion coefficient D are (meter) $)^{2} /($ second).

Example 6-2 ${ }^{\text {MATLAB }}$. Show that a particular solution for the diffusion equation

$$
\frac{\partial^{2} V(z, t)}{\partial z^{2}}=\frac{1}{D} \frac{\partial V(z, t)}{\partial t}
$$

that describes the distributed "RC" transmission line where $D=1 /(\hat{R} \hat{C})$ is given by

$$
V(z, t)=\frac{1}{2 \sqrt{D \pi t}} \exp \left(-\frac{z^{2}}{4 D t}\right)
$$

## Transmission line equations

Since the capacitors are linear, this voltage will also correspond to the charge on the particular capacitor. This follows from $Q(z, t)=\hat{C} V(z, t)$ with linear capacitors.

Answer: Differentiating the solution with respect to $z$, we obtain

$$
\frac{\partial V(z, t)}{\partial z}=\frac{1}{2 \sqrt{D \pi}}\left(-\frac{z}{2 D t^{3 / 2}}\right) \exp \left(-\frac{z^{2}}{4 D t}\right)
$$

and

$$
\frac{\partial^{2} V(z, t)}{\partial z^{2}}=\frac{1}{2 \sqrt{D \pi}}\left(-\frac{1}{2 D t^{3 / 2}}+\frac{z^{2}}{4 D^{2} t^{5 / 2}}\right) \exp \left(-\frac{\hat{R} \hat{C} z^{2}}{4 t}\right)
$$

Differentiating the solution with respect to $t$, we obtain

$$
\frac{1}{D} \frac{\partial V(z, t)}{\partial t}=\frac{1}{D} \frac{1}{2 \sqrt{D \pi}}\left(-\frac{1}{2 t^{3 / 2}}+\frac{z^{2}}{4 D t^{5 / 2}}\right) \exp \left(-\frac{z^{2}}{4 D t}\right)
$$

Hence, the equation is satisfied. The voltage at the normalized times 1 to 4 is shown below. Note that the peak remains at $z=0$ as time increases. The total area under the curve at each of the times remains equal to one that is independent of time.

## Transmission line equations



The solution that is shown in this figure is valid if a certain amount of charge is placed at the location $z=0$. There is a decrease in amplitude at the center by $50 \%$ as a time increases by a factor of 4 . The total area under every curve is equal to 1.

Diffusion analyzed in the above example would be similar to filling a balloon with helium, popping it at $t=0$, and then monitoring the helium density in space at later times. One can think of this source also in terms of a Dirac's delta function since the area under this function is a constant. A second boundary condition would be to fix the voltage at $z=0$ with a battery and a switch that was closed at $t=0$. Diffusion differs from waves in that the solution predicts that the

## Transmission line equations

voltage could change at $z= \pm \infty$ at a time $t=0^{+}$. The diffusion equation is also called the heat equation from thermodynamics. A calculation associated with the second boundary condition would be to compute the temporal and spatial evolution of the temperature in an object if one end touches a hot plate located at z = 0 and whose temperature remains constant for all times.

The reader might find it instructive to analyze the diffusion of a step voltage excited at a location $z=0$ at a time $t=0$ of a distributed $R C$ transmission line containing several sections using a computer circuit analysis program such as SPICE. The process of diffusion is common in semiconductors. Diffusion is significantly different than the wave phenomena that we encounter in normal transmission lines.

## Sinusoidal waves

### 6.3. Sinusoidal waves

Let us now specify a particular form for the excitation of a normal transmission line consisting of distributed inductors and capacitors and study the resulting propagating wave. Let us assume that the transmission line is excited with a sinusoidal voltage generator at $z=+\infty$ or at $z=-\infty$ and the generator was turned on at an early time such that all transient effects have decayed to zero before you read these pages. We will return to the study of the transients and more importantly, the propagation of pulses, on this transmission line later because of its importance in modern digital communication systems. Suffice it to say at this point that pulses could be decomposed into their Fourier components and time-harmonic propagation is important in its own right. For, example, the electric power industry makes considerable use of 50 or 60 Hz sinusoidal propagation throughout the world. It is convenient to determine the characteristic impedance of the transmission line using this excitation for the transmission line before a more general solution is obtained. This will be similar to the characteristic impedance for plane waves in free space that was described in the previous chapter.

In this case, a sinusoidal voltage and current wave will propagate to decreasing or increasing values of $z$ and we can separate out the time dependence from the spatial term and write that

$$
\begin{equation*}
V(z, t)=V(z) e^{j \omega t} \quad \text { and } \quad I(z, t)=I(z) e^{j \omega t} \tag{6.10}
\end{equation*}
$$

where we again encounter phasors. We again could have chosen at this stage of the derivation to examine either just the real or the imaginary parts of the wave propagation via the relation $\mathrm{e}^{j \omega t}=\cos \omega \mathrm{t}+\mathrm{j} \sin \omega \mathrm{t}$. The voltage and the current both have the same temporal dependence as they must since they are coupled by the first-order partial differential equations that make up the Telegraphist's

## Sinusoidal waves

equations (6.3) and (6.6). Substituting (6.10) into either (6.3) or (6.6) where we have employed the definition of the velocity given in (6.9), we obtain

$$
\begin{align*}
& \frac{d^{2}(z)}{d z^{2}}-\frac{(j \omega)^{2}}{v^{2}}(z)=0  \tag{6.11}\\
& \frac{d^{2}(z)}{d z^{2}}-\frac{(j \omega)^{2}}{v^{2}}(z)=0 \tag{6.12}
\end{align*}
$$

This set of equations can be rewritten as

$$
\begin{align*}
& \frac{d^{2}(z)}{d z^{2}}+k^{2} \quad(z)=0  \tag{6.13}\\
& \frac{d^{2}(z)}{d z^{2}}+k^{2} \quad(z)=0 \tag{6.14}
\end{align*}
$$

where the wave number $k$ is defined as $k=\omega / v$. The wave number is related to the wavelength $\lambda$ as $k=\frac{2 \pi}{\lambda}$. These two equations are similar in form to the equation we encountered previously in our discussion of plane waves with a time-harmonic excitation signal (5.31). Solutions of (6.13) can be written as

$$
\begin{gather*}
V(z, t)=A_{1} e^{\left.j(\omega t-k z)+B_{1} e^{j(\omega t}+k z\right)}  \tag{6.15}\\
V(z, t)=A_{2} \cos (\omega t-k z)+B_{2} \cos (\omega t+k z) \tag{6.16}
\end{gather*}
$$

where $A_{1}$ and $B_{1}$ or $A_{2}$ and $B_{2}$ are constants of integration. In writing this solution, we have incorporated the temporal dependence ej ${ }^{j \omega t}$. Different combinations of the exponential terms with different values for the constants would allow us to write the trigonometric functions as sines instead of cosines. An astute reader may recall similar results from the previous chapter. We include it here only for completeness. We could alternatively specify that the real or the imaginary portions of the voltages are to be determined. Finally, an arbitrary constant phase shift e ${ }^{j \theta}$ (or $\theta$ ) can be added to (6.15) or (6.16) if the need arises.
Example 6-3 MATLAB. The voltage wave that propagates along a transmission line is detected at the indicated points. From this data, write an expression for the wave. Note that there is a propagation of the sinusoidal signal to increasing values of the coordinate $z$.

(b) $\begin{gathered}5 \\ 2\end{gathered}$

Answer: From the data shown in (a), we can obtain the following information. The peak to peak amplitude of the wave is $2 \mathrm{~V}_{0}$. The wave is propagating to increasing values of the coordinate $z$. The period of the wave is 2 s , hence the frequency of consolation of the way is $f=\frac{1}{2}=0.5 \mathrm{~Hz}$. The velocity of propagation $v$ is obtained from the slope of the trajectory shown in (b). The value is found to be $v=\frac{5-1}{1-0}=4 \mathrm{~m} / \mathrm{s}$. The wave number $k$ is computed to be $\mathrm{k}=\mathrm{v} \omega=4\left(2 \pi \frac{1}{2}\right)=4 \pi \mathrm{~m}^{-1}$. The wavelength $\lambda$ of the wave is equal to $\lambda=\frac{2 \pi}{\mathrm{k}}=\frac{2 \pi}{4 \pi}=0.5 \mathrm{~m}$. The wave is given by $\mathrm{V}=\mathrm{V}_{0} \cos (\pi \mathrm{t}-4 \pi \mathrm{z})$.

## Sinusoidal waves

The current wave $I(z, t)$ can be calculated by inserting the voltage wave solution (6.16) into either of the first order Telegraphist's equations (6.3) or (6.6). If we assume that the wave is launched from the point $z=-\infty$ and is propagating to increasing values of the coordinate $z$, then the value of the constant $B_{2}$ in the voltage wave will have the value of $B_{2}=0$. The other constant $B_{1}$ will be chosen to be $\mathrm{V}_{0}$. We find using (6.6) that

$$
\begin{align*}
\frac{\partial\left[\mathrm{V}_{0} \cos (\omega t-k z)\right]}{\partial z} & =-\hat{L} \frac{\partial\left[\mathrm{l}_{\mathrm{o}} \cos (\omega t-k z)\right]}{\partial t} \\
\mathrm{~V}_{\mathrm{o}} \mathrm{k} & =\hat{L} \omega \mathrm{I}_{0} \tag{6.17}
\end{align*}
$$

The dimension of the ratio of the voltage divided by the current amplitudes is equal to an impedance. This impedance is called the characteristic impedance of the transmission line and it will be denoted with the symbol $Z_{c}$. Therefore, we write (6.17) as

$$
\begin{equation*}
Z_{C}=\frac{V_{o}}{I_{o}}=\hat{L} \frac{\omega}{k} \tag{6.18}
\end{equation*}
$$

The ratio $\frac{\omega}{\mathrm{k}}$ is the phase velocity v of the wave. This velocity is given in (6.9) and we write

$$
Z_{C}=\frac{V_{0}}{I_{o}}=\hat{L} v=\hat{L} \frac{1}{\sqrt{\hat{L} \hat{C}}}
$$

Finally, the characteristic impedance of the transmission line can be written as

$$
\begin{equation*}
Z_{C}=\sqrt{\frac{\hat{L}}{\hat{\mathrm{C}}}} \Omega \tag{6.19}
\end{equation*}
$$

This characteristic impedance of a transmission lines is the ratio of the voltage wave divided by the current wave. This is similar to our definition for plane waves where we defined the characteristic impedance as the ratio of the electric field intensity divided by the magnetic field intensity. Since the characteristic impedance is a constant, the current wave can be computed merely by dividing the voltage wave by the characteristic impedance $Z_{c}$. We will

## Sinusoidal waves

find that the characteristic impedance is one of the important parameters used to characterize a transmission line. The other parameter is the length $L$ of the transmission line. This length is normalized by the wavelength $\lambda$ and of the propagating wave. If the transmission line were lossy, the characteristic impedance would be a complex impedance instead of the real number obtained here. Since the transmission line has been assumed to be lossless, we could let $Z_{C}=R_{C}+j X_{C}$ where $R_{C}$ is a real number and $X_{C}=0$ with no loss of generality. We will, however, keep the more general notation of defining the characteristic impedance as $Z_{C}$ even for the lossless transmission line.

Example 6-4. Calculate the velocity of propagation and the characteristic impedance of a solid coaxial cable. The radius of the inner conductor is 3 mm and the radius of the outer conductor is 6 mm . There is a vacuum between these two conductors.

Answer. From Table 6-1, the inductance per unit length is given by

$$
\hat{\mathrm{L}}=\frac{\mu_{0}}{2 \pi} \ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right)=\frac{4 \pi \times 10^{-7}}{2 \pi} \ln \left(\frac{6}{3}\right)=0.14^{\mu \mathrm{H}} / \mathrm{m}
$$

The capacitance per-unit length is given by

$$
\hat{\mathrm{C}}=\frac{2 \pi \varepsilon_{o}}{\ln \left(\frac{\mathrm{~b}}{\mathrm{a}}\right)}=\frac{2 \pi\left(\frac{1}{36 \pi} \times 10^{-9}\right)}{\ln \left(\frac{6}{3}\right)}=80 \mathrm{pF} / \mathrm{m}
$$

The velocity of propagation is computed from (6.9).

$$
v=\frac{1}{\sqrt{\hat{L} \hat{C}}}=\frac{1}{\sqrt{\left(0.14 \times 10^{-6}\right)\left(80 \times 10^{-12}\right)}}=3 \times 10^{8} \mathrm{~m} / \mathrm{s} .
$$

The characteristic impedance of the coaxial cable is computed from (6.19).

$$
Z_{C}=\sqrt{\frac{\hat{L}}{\hat{C}}}=\sqrt{\frac{0.14 \times 10^{-6}}{80 \times 10^{-12}}} \approx 42 \Omega .
$$

The velocity of propagation and the characteristic impedance can be decreased if a dielectric is inserted between the two conductors.

## Sinusoidal waves

In Table 6-1, we summarized the inductance per-unit length and the capacitance per-unit length for the transmission lines that are most commonly used. In Table 6-2, we summarize the velocity of propagation and the characteristic impedances for the same transmission lines.

|  | Velocity of propagation | Characteristic impedance |
| :--- | :--- | :--- |
| coaxial cable | $\mathrm{v}=\frac{1}{\sqrt{\hat{L} \hat{C}}}=\frac{1}{\sqrt{\mu \varepsilon}}$ | $Z_{c}=\sqrt{\frac{\hat{L}}{\hat{C}}}=\sqrt{\frac{\mu}{\varepsilon}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)$ |
| microstrip line | $v=\frac{1}{\sqrt{\hat{L} \hat{C}}}=\frac{1}{\sqrt{\mu \varepsilon}}$ | $Z_{c}=\sqrt{\frac{\hat{L}}{\hat{C}}}=\sqrt{\frac{\mu}{\varepsilon}}\left(\frac{d}{W}\right)$ |
| twin lead | $v=\frac{1}{\sqrt{\hat{L} \hat{C}}}=\frac{1}{\sqrt{\mu \varepsilon}}$ | $Z_{c}=\sqrt{\frac{\hat{L}}{\hat{C}}}=\sqrt{\frac{\mu}{\varepsilon}}\left(\frac{\cosh ^{-1}\left(\frac{D}{2 a}\right)}{\pi}\right)$ |

Table 6-2. The velocity of propagation and the characteristic impedance of the various transmission lines shown in Figure 6-1. The parameters for the material between the two conductors are the permeability $\mu=\mu_{r} \mu_{o}$ and the permittivity $\varepsilon=\varepsilon_{\mathrm{r}} \varepsilon_{0}$.

The velocity of propagation of the wave is independent of the dimensions of the transmission line and it is only a function of the electrical parameters of the material that separates the two conductors. However, the characteristic impedance depends upon these physical dimensions along with the electrical parameters. This fact will be important when two transmission lines are connected together and a signal is launched on one of them.

## Terminations

### 6.4. Terminations

In the above section, we assumed that the transmission line was infinite in length. This is not practical in spite of the best intentions of your instructor who is trying to present this difficult topic in a simple and logical manner. The transmission line will have both a beginning and an end that occur at finite locations in space. We therefore must incorporate this subject into our discussion and will examine terminations in this section.

A finite length of a lossless transmission line is illustrated in Figure 6-4. The transmission line is characterized with a characteristic impedance $Z_{C}$ that is a real quantity. It is convenient to assume that the source of the time-harmonic wave is at $z=-\infty$ and termination is located at $z=0$. This termination could be either an impedance or another transmission line with a different characteristic impedance. We will also assume that the signal generator was also turned on at the time $t=-\infty$ so all transient effects will have disappeared. These choices are predicated on our desire to simplify as much as possible the various calculational procedures that we will encounter later.


Figure 6-4. A semi infinite transmission line that is terminated in a load impedance $\mathrm{Z}_{\mathrm{L}}$.

## Terminations

The voltage at any point along the line can be expressed from (6.15) to be

$$
\begin{equation*}
V(z, t)=A_{1} e^{j(\omega t-k z)}+B_{1} e^{j(\omega t+k z)} \tag{6.20}
\end{equation*}
$$

The current at any point along the line is calculated from the substitution of (6.20) into the Telegraphist's equations (6.3) or (6.6). We find

$$
\begin{equation*}
\mathrm{I}(\mathrm{z}, \mathrm{t})=\frac{1}{\mathrm{Z}_{\mathrm{C}}}\left(\mathrm{~A}_{1} \mathrm{e}^{\mathrm{j}(\omega t-k z)}-\mathrm{B}_{1} \mathrm{e}^{\mathrm{j}(\omega t+k z)}\right) \tag{6.21}
\end{equation*}
$$

Let us evaluate the ratio of (6.20) and (6.21) at the location of the load impedance $Z_{L}$. This ratio will be equal to the load impedance. Since we have chosen this point to be at $z=0$, the exponential terms will cancel or disappear and we write

$$
\begin{equation*}
Z_{L}=\frac{V(0, t)}{I(0, t)}=\frac{A_{1}+B_{1}}{A_{1}-B_{1}} Z_{C} \tag{6.22}
\end{equation*}
$$

Our choice of placing the termination at $z=0$ has been well rewarded.
It is reasonable to assume that we know the amplitude of the signal that is incident upon the load, that is $\mathrm{A}_{1}$ is a known value. Knowing the amplitude of the incident wave allows us to compute the amplitude of the reflected wave $B_{1}$ from (6.22) to be

$$
\begin{equation*}
\frac{B_{1}}{A_{1}}=\frac{\frac{Z_{L}}{Z_{C}}-1}{\frac{Z_{L}}{Z_{C}}+1}=\frac{z_{L}-1}{Z_{L}+1} \tag{6.23}
\end{equation*}
$$

where we have introduced a normalized load impedance $z_{L} \equiv \frac{Z_{L}}{Z_{C}}$. Since we have chosen the transmission line to be lossless, this normalized load impedance is just equal to the actual load impedance divided by a number. This ratio of the amplitudes of the reflected $\mathrm{B}_{1}$ to the incident $A 1$ wave is called the reflection coefficient and we will label it with the symbol $R,{ }^{1}$ Hence

[^19]
## Terminations

$$
\begin{equation*}
R \equiv \frac{B_{1}}{A_{1}}=\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}=\frac{Z_{L}-1}{Z_{L}+1} \tag{6.24}
\end{equation*}
$$

An astute reader realizes that the derivation of the reflection coefficient for a transmission line is very similar to the derivation that led to the reflection coefficient for plane waves (5.72). In terms of the reflection coefficient $R$ the voltage and the current at any point along the transmission line can be expressed as

$$
\begin{align*}
& V(z, t)=A_{1}\left(e^{j(\omega t-k z)}+R^{\left.\mathrm{e}^{j(\omega t+k z)}\right)}\right.  \tag{6.25}\\
& I(z, t)=\frac{A_{1}}{Z_{C}}\left(e^{j(\omega t-k z)}-R^{j(\omega t+k z)}\right) \tag{6.26}
\end{align*}
$$

The ratio of these two quantities has the units of an impedance and it is the impedance $Z(z)$ at a particular location $z$. If the reflection coefficient $R=0$, this impedance will be equal to the characteristic impedance $Z_{C}$ of the transmission line. In this case, the load impedance is matched to the characteristic impedance on the transmission line. This impedance will depend upon the position of observation. We will return to this point in the next section.

The reflection coefficient $\mathcal{R}$ is determined entirely by the value of the impedance of the load and the characteristic impedance of the transmission line. If the line is homogeneous and it has no discontinuities, this is a good assumption. This implies that the voltages and the currents that appear at any point along the transmission line are determined by the signal generator and the load impedance which may be several wavelengths apart. Remember that our signal generator was turned on a long time ago and no transients are relevant in this discussion. We will examine them later since they are very important.

The reflection coefficient $\mathbb{R}$ for a lossless transmission line can have any value between -1 and +1 . Since the load impedance $Z_{L}$ may be complex, the reflection coefficient may also have complex values. In this case, it is better to express a complex reflection coefficient in phasor notation. This will lead to a

## Terminations

phase shift between the reflected wave and the incident wave. More about that later. At this stage, we will confine our discussion to real impedances.

If the load impedance were a short circuit $\left(Z_{L}=0\right)$, then the reflection coefficient $R$ that we compute from (6.24) will yield $R=-1$. At the load, the total voltage at $z=0$ that consists of the sum of the incident and the reflected components must equal zero. If this impedance were an open circuit ( $Z_{L}=\infty$ ), then the reflection coefficient $R=+1$. In this case, the total voltage can be arbitrary but the total current must equal zero. An interesting case arises if the load impedance is equal to the characteristic impedance of the line $\left(Z_{L}=Z_{C}\right)$. In this case, the reflection coefficient $R=0$ and we say that the line is matched. This matching is very important in practice since all of the energy is transported down the line, absorbed at the load impedance, and none will be reflected back toward the signal generator. All of the energy will be gainfully employed in the load impedance and there will be none to come back. We will find that there are techniques that can be employed in order to achieve this desirable state of operation, even if the load impedance has a value that is different from the value of the characteristic impedance of the transmission line.

Example 6-5. The dielectric in an infinitely long coaxial cable has a value for its relative dielectric constant of $\varepsilon_{r}=2$ for $z<0$ and $\varepsilon_{r}=3$ for $z>0$. Calculate the reflection coefficient $R$ for a wave that is incident from $z=-\infty$.


Answer: The characteristic impedance of a coaxial line is given in Table 6-2.
The load impedance $Z_{L}$ that appears in (6.24) will be the characteristic impedance of the coaxial cable in the region $z>0$ since it acts as a load for the

## Terminations

coaxial cable in the region $z<0$. Hence, we write the reflection coefficient in terms of the characteristic impedances of the two transmission lines in unnormalized form.

$$
\mathcal{R}=\frac{Z_{2}-Z_{1}}{Z_{2}+Z_{1}}=\frac{\sqrt{\frac{\mu_{o}}{\varepsilon_{\mathrm{r} 2} \varepsilon_{o}}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)-\sqrt{\frac{\mu_{0}}{\varepsilon_{\mathrm{r} 1} \varepsilon_{\mathrm{o}}}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)}{\sqrt{\frac{\mu_{0}}{\varepsilon_{\mathrm{r} 2} \varepsilon_{\mathrm{o}}}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)+\sqrt{\frac{\mu_{0}}{\varepsilon_{\mathrm{r} 1} \varepsilon_{\mathrm{o}}}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)}=\frac{\frac{1}{\sqrt{\varepsilon_{\mathrm{r} 2}}}-\frac{1}{\sqrt{\varepsilon_{\mathrm{r} 1}}}}{\frac{1}{\sqrt{\varepsilon_{\mathrm{r} 2}}}+\frac{1}{\sqrt{\varepsilon_{\mathrm{r} 1}}}}=\frac{\frac{1}{\sqrt{3}}-\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{3}}+\frac{1}{\sqrt{2}}} \approx-0.1
$$

Let us for the moment examine the voltage wave forms that arise when a sinusoidal voltage wave is incident upon a short circuit $\left(Z_{L}=0\right)$ or an open circuit $\left(Z_{L}=\infty\right)$. In the first case, the reflection coefficient as computed from (6.24) is $\mathcal{R}=-1$ and for the second case $\mathbb{R}=+1$. From (6.25), we write the voltage as

$$
\begin{equation*}
\left.V(z, t)\right|_{z_{L}=0}=A_{1}\left(e^{j(\omega t-k z)}-e^{j(\omega t+k z)}\right)=(-2 j) A_{1} \sin (k z) e^{j \omega t} \tag{6.27}
\end{equation*}
$$

and

$$
\begin{equation*}
\left.V(z, t)\right|_{z_{L}=\infty}=A_{1}\left(e^{j(\omega t-k z)}+e^{j(\omega t+k z)}\right)=(2) A_{1} \cos (k z) e^{j \omega t} \tag{6.28}
\end{equation*}
$$

In both (6.27) and (6.28), we have factored the temporal dependence ej ${ }^{\mathrm{j} \omega \mathrm{t}}$ from the exponential terms. This temporal variation merely causes the amplitude of the wave to oscillate between a maximum value of $+2 \mathrm{~A}_{1}$ and a minimum value of $-2 A_{1}$ where $A_{1}$ is the amplitude of the wave that is launched from the signal generator. This point can be emphasized if we use the definition

$$
e^{j \omega t}=\cos \omega t+j \sin \omega t
$$

and select just the real or the imaginary terms for the voltages. The additional constant $j$ can be absorbed into the constant $A_{1}$ with no loss of generality.

## Terminations




Figure 6-5 MATLAB. Transmission line that has a characteristic impedance $Z_{C}$. The line is excited at $z=-\infty$ with a sinusoidal voltage. (a) Standing voltage wave if $Z_{L}=0$. (b) Standing voltage wave if $Z_{L}=\infty$. The maximum amplitude of the voltage standing wave is $2 \mathrm{~A}_{1}$.

The addition of the two propagating waves that individually propagate to increasing values of the spatial coordinate $z$ and to decreasing values of $z$ created a signal that appears to be stationary in space but have its magnitude oscillate in time from zero to twice the value of the incident wave. This effect is called a standing wave since the resultant signal does not appear to propagate. We illustrate the measured voltages as a function of space at various times for the two values of load impedance in Figure 6-5. Recall that we first encountered

## Terminations

a standing wave in the previous chapter. Standing waves are also found on the strings of a violin as the force of the violinist's fingers create a local "short circuit" for the waves that are excited by the bow or the plucking motion of the finger, propagate along the string, and are trapped by these fingers and the bridge of the violin.

Example 6-6 ${ }^{\text {MATLAB }}$. Determine the standing current waves that corresponds to the standing voltage wave depicted in Figure 6-5.


Answer: The current wave is equal to the voltage wave divided by the characteristic impedance of the transmission line. In this case $I_{0}=A_{1} / Z_{c}$. The standing current wave is depicted below for the two cases: (a) $Z_{L}=0$ and (b) $Z_{L}=\infty$. Note that the standing waves for the current waves are $90^{\circ}$ out of phase with respect to the standing voltage waves. From Figure $6-5$, we note that the maximum voltage that occurs along the line is equal to $2 \mathrm{~A}_{1}$ and the minimum voltage is 0 .

## Terminations

The magnitude of the voltage repeats itself every half-wavelength. This is a crucial observation since a typical voltage detector used in practice is a "squarelaw device" that responds to the magnitude of the voltage and cannot distinguish between plus or minus voltages.

The ratio of the maximum voltage to the minimum voltage that appears along this transmission line, that in this case is $\infty$, is an important quantity. The maximum value of the voltage is equal to magnitude of the incident voltage plus the magnitude of the reflected voltage. The minimum value equals the difference of these two quantities. The ratio is called the voltage standing wave ratio;. The abbreviation VSWR2 is frequently used in practice. From (6.25), we write for an arbitrary load that

$$
\begin{equation*}
\operatorname{VSWR}=\frac{\mathrm{V}_{\max }}{\mathrm{V}_{\min }}=\frac{1+|\mathcal{R}|}{1-|\mathcal{R}|} \tag{6.29}
\end{equation*}
$$

We can solve this equation for $|R|$ and obtain

$$
\begin{equation*}
|\mathcal{R}|=\frac{\mathrm{VSWR}-1}{\mathrm{VSWR}+1} \tag{6.30}
\end{equation*}
$$

We note that the VSWR, the reflection coefficient $R$, and from (6.24) the ratio of the load impedance to the characteristic impedance of the line are intimately and crucially related.

If the amplitudes of the voltages are not very large, this may be a moot point. However, if the voltages $\left[\mathrm{A}_{1}(1+|R|)\right]$ are large, say above the breakdown conditions of electronic components, serious problems may occur even if the amplitude of the incident wave $A_{1}$ is beneath this critical value. The magnitude of

[^20]
## Terminations

the reflection coefficient $R$, must be reduced to zero. From (6.24), this implies that the load impedance $Z_{L}$ should have a value that approaches the characteristic impedance $Z_{C}$ of the line. The line must be matched!

Example 6-7. Compute the ratio of the maximum voltage to the minimum voltage of an electromagnetic wave propagating in the coaxial cable described in Example 6-5 in the region $\mathrm{z}<0$.

Answer: The reflection coefficient $R$ was computed in Example 6-5 to be $\mathcal{R}=-0.1$. From (6.29), we write

$$
\operatorname{VSWR}=\frac{1+|\mathcal{R}|}{1-|\mathcal{R}|}=\frac{1+|-0.1|}{1-|-0.1|}=1.2
$$

Hence the ratio of the maximum voltage to the minimum voltage along the transmission line in the region $z<0$ will be 1.2. The VSWR equals 1.2. This value is close to the ideal value of 1 that would be obtained if the two coaxial cables were matched..

## Impedance on the transmission line and matching

### 6.5. Impedance on the transmission line and matching

The ratio of the voltage to the current at any point on the transmission line has the dimensions of an impedance. This impedance is not a constant as we move along the line from the source to the load since these voltages and currents consist of both the incident and the reflected components. This impedance will also be a function of the load impedance and the characteristic impedance of the transmission line since the reflected wave is determined by these terms.

From (6.25) and (6.26), we write

$$
\begin{equation*}
Z(z, t)=\frac{V(z, t)}{I(z, t)}=\frac{A_{1}\left(e^{j(\omega t-k z)}+R R^{j(\omega t+k z)}\right)}{\frac{A_{1}}{Z_{C}}\left(e^{j(\omega t-k z)}-R e^{j(\omega t+k z)}\right)}=Z_{C} \frac{\left(e^{-j k z}+\mathcal{R} e^{j k z}\right)}{\left(e^{-j k z}-R e^{j k z}\right)} \tag{6.31}
\end{equation*}
$$

In writing the final term of this equation, the factors of $e^{j \omega t}$ and $A_{1}$ that are common to both the numerator and the denominator have been canceled. This implies that the impedance is linear and time independent. Let us now replace the reflection coefficient $\mathbb{R}$ using (6.24). We obtain

$$
Z(z)=Z_{C} \frac{\left(e^{-j k z}+\left(\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}\right) e^{j k z}\right)}{\left(e^{-j k z}-\left(\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}\right) e^{j k z}\right)}=Z_{C} \frac{Z_{L}\left(e^{j k z}+e^{-j k z}\right)-Z_{C}\left(e^{j k z}-e^{-j k z}\right)}{Z_{C}\left(e^{j k z}+e^{-j k z}\right)-Z_{L}\left(e^{j k z}-e^{-j k z}\right)}
$$

The exponential terms can be recognized as the definition for the trigonometric functions. Therefore, we have

$$
\begin{equation*}
Z(z)=Z_{C} \frac{2 Z_{L} \cos (k z)-j 2 Z_{C} \sin (k z)}{2 Z_{C} \cos (k z)-j 2 Z_{L} \sin (k z)}=Z_{C} \frac{Z_{L}-j Z_{C} \tan (k z)}{Z_{C}-j Z_{L} \tan (k z)} \tag{6.32}
\end{equation*}
$$

Since we have chosen to locate the load at the origin $z=0$ and the signal generator at the location $z \rightarrow-\infty$, the impedance will always be evaluated at some location $z<0$. This point will be chosen to be at $z=-L$. Therefore, the final expression for the impedance at any location on the transmission line is given by

## Impedance on the transmission line and matching

$$
\begin{equation*}
\mathrm{Z}_{\text {in }}(\mathrm{z}=-\mathrm{L})=\mathrm{Z}_{\mathrm{c}} \frac{\mathrm{Z}_{\mathrm{L}}+\mathrm{j} \mathrm{Z}_{\mathrm{C}} \tan (\mathrm{~kL})}{\mathrm{Z}_{\mathrm{C}}+\mathrm{j} \mathrm{Z}_{\mathrm{L}} \tan (\mathrm{~kL})} \tag{6.33}
\end{equation*}
$$

This can also be written as a normalized impedance $\mathrm{z}_{\text {in }}=\frac{\mathrm{Z}_{\text {in }}}{\mathrm{Z}_{\mathrm{c}}}$ where we have followed the convention of using a lowercase letter to indicate a normalized impedance

$$
\begin{equation*}
z_{\text {in }}(z=-L)=\frac{z_{L}+j \tan (k L)}{1+j z_{L} \tan (k L)} \tag{6.34}
\end{equation*}
$$

This impedance $\mathrm{z}_{\mathrm{in}}$ at this location will be called the normalized input impedance of the transmission line.

Example 6-8. A signal generator whose frequency $f=100 \mathrm{MHz}$ is connected to a coaxial cable whose characteristic impedance is $100 \Omega$ and whose length is 100 m . The velocity of propagation is equal to $2 \times 10^{8} \mathrm{~m} / \mathrm{s}$. The transmission line is terminated with a load impedance of $50 \Omega$. Calculate the impedance at a distance on 50 m from the load.

Answer: The normalized load impedance $z_{L}$ is equal to $1 / 2$. The wavelength $\lambda$ is calculated from

$$
\lambda=\frac{v}{f}=\frac{2 \times 10^{8}}{1 \times 10^{8}}=2 \mathrm{~m}
$$

The wave number k is calculated to be

$$
\mathrm{k}=\frac{2 \pi}{\lambda}=\frac{2 \pi}{2}=\pi \mathrm{m}^{-1}
$$

The normalized input impedance is calculated using (6.34)

$$
z_{\text {in }}(z=-50 m)=\frac{z_{L}+j \tan (k L)}{1+j z_{L} \tan (k L)}=\frac{1 / 2+j \tan (50 \pi)}{1+j \frac{1}{2} \tan (50 \pi)}=\frac{1}{2}
$$

Therefore, the input impedance at this location is equal to $Z_{\text {in }}=z_{i n} Z_{c}=50 \Omega$.

Recall that the wave number can be defined in terms of the wavelength $\lambda$ as $k=\frac{2 \pi}{\lambda}$. Therefore, the length of the transmission line $L$ can also be

## Impedance on the transmission line and matching

normalized by the wavelength $k L=\frac{2 \pi}{\lambda} L$. The implication of this normalization is that the value of the load impedance will repeat itself every $1 / 2$ wavelength since $\tan (k L)=\tan (k L+n \pi)$ where $n$ is an integer. If the transmission line is $1 / 4$ of a wavelength, we obtain $\mathrm{kL}=\frac{2 \pi}{\lambda}\left(\frac{\lambda}{4}\right)=\frac{\pi}{2}$. The trigonometric function $\tan \left(\frac{\pi}{2}\right) \rightarrow \infty$ and the input impedance is given by

$$
\begin{equation*}
Z_{\text {in }}\left(z=-\frac{\lambda}{4}\right)=Z_{c} \frac{Z_{L}+j Z_{c} \infty}{Z_{c}+j Z_{L} \infty} \rightarrow \quad Z_{\text {in }(\lambda / 4)}=\frac{\left(Z_{c}\right)^{2}}{Z_{L}} \tag{6.35}
\end{equation*}
$$

This implies that the normalized input impedance $Z_{i n}$ of a $1 / 4$ wavelength transmission line that is terminated with a load impedance $Z_{L}$ will have $a$ numerical value that is equal to the normalized load admittance $y_{L}=1 / z_{L}$.

This one-quarter wavelength transmission line will be useful in joining two transmission lines that have different characteristic impedances or matching a load as will be now demonstrated. One of the simplest techniques is to use a quarter wave transformer. A quarter wave transformer is a section of transmission line that has a particular characteristic impedance that is identified as $Z_{C(\lambda / 4)}$. In addition, the length of this matching transmission line will be specified in terms of the wavelength of the wave as it propagates in this quarter wavelength transmission line. We're assuming that the value of the characteristic impedance $Z_{c(\lambda / 4)}$ of this particular transmission line can be specified by the user. The value of this characteristic impedance will be chosen such that the reflection coefficient $\mathbb{R}$ at the input of the matching transmission lines section will be equal to 0 . This is shown in Figure 6-6.

## Impedance on the transmission line and matching



Figure 6-6. $A$ transmission line with a characteristic impedance $Z_{C} \neq Z_{L}$ is joined to the load with a quarter wave transformer.

In order to specify the value of this characteristic impedance $Z_{c(\lambda / 4)}$ of this quarter wavelength transmission line, we recall the reflection coefficient $R$ that was given in (6.24). In the present application, we assume that the load impedance is the input impedance of the quarter wavelength transmission line

$$
\begin{equation*}
\mathcal{R} \equiv \frac{Z_{\text {in }}-Z_{C}}{Z_{\text {in }}+Z_{C}} \tag{6.36}
\end{equation*}
$$

where

$$
\begin{equation*}
Z_{\text {in }}\left(z=-\frac{\lambda}{4}\right)=\frac{\left(Z_{C(\lambda / 4)}\right)^{2}}{Z_{L}} \tag{6.37}
\end{equation*}
$$

In order to minimize the reflection coefficient, this input impedance should be chosen to have a value of that is equal to the characteristic impedance $Z_{c}$ of the transmission line that is connected to the signal generator. From (6.37), we obtain the characteristic impedance of the matching transmission line to be

$$
\begin{equation*}
Z_{C(\lambda / 4)}=\sqrt{Z_{C} Z_{L}} \tag{6.38}
\end{equation*}
$$

The characteristic impedance of this matching transmission line is chosen to have a value that is equal to the square root of the geometric mean of the load impedance and the characteristic impedance of the transmission line that is connected to the signal generator.

## Impedance on the transmission line and matching

This technique has certain disadvantages in that it is frequency sensitive since the velocity of propagation and therefore the wavelength is determined by the material parameters in the matching transmission line. Techniques from modern filter theory can be employed to desensitize this restriction to a certain extent.

Example 6-9. A lossless transmission line is terminated with an impedance whose value that is $1 / 2$ the characteristic impedance on the line. What impedance should you put in parallel with the line $\lambda / 4$ in front of the load to minimize the reflection of the wave from the signal generator?


Answer: In order to reduce the reflection, the parallel combination of $Z_{Q}$ and the input impedance at this location should be equal to the characteristic impedance of the transmission line. Therefore, we write

$$
\begin{gathered}
Z_{C}=\frac{Z_{Q} Z_{\text {in }(\lambda / 4)}}{Z_{Q}+Z_{\text {in }(\lambda / 4))}}=\frac{Z_{Q}\left(Z_{C}^{2} / Z_{L}\right)}{Z_{Q}+\left(Z_{C}^{2} / Z_{L}\right)}=\frac{Z_{Q}\left(Z_{C}^{2} /\left(Z_{C} / 2\right)\right)}{Z_{Q}+\left(Z_{C}^{2} /\left(Z_{C} / 2\right)\right)} \\
1=\frac{Z_{Q} 2}{Z_{Q}+2 Z_{C}} \Rightarrow Z_{Q}=2 Z_{C}
\end{gathered}
$$

Three particular values of the load impedances should be mentioned at this point. In the first case, the load impedance $Z_{L}$ equals the characteristic impedance of the transmission line $Z_{c}$. In this case, the transmission line is

## Impedance on the transmission line and matching

matched. There will be no reflected component of the incident wave and we find from (6.24) that the impedance $Z(z)$ will always be equal to the characteristic impedance $Z_{c}$. For the other two cases, the load impedance is either a short circuit $\left(Z_{L}=0\right)$ or an open circuit $\left(Z_{L}=\infty\right)$. For the latter two cases, the impedance is found from (6.33) to be

$$
\begin{gather*}
\left.Z_{\text {in }}(z=-L)\right|_{Z_{\mathrm{L}}=0}=j \mathrm{Z}_{\mathrm{C}} \tan (\mathrm{~kL}) \\
\left.\mathrm{Z}_{\text {in }}(\mathrm{z}=-\mathrm{L})\right|_{\mathrm{Z}_{\mathrm{L}}=\infty}=\frac{\mathrm{Z}_{\mathrm{C}}}{j \tan (\mathrm{~kL})}=-\mathrm{j} \mathrm{Z}_{\mathrm{C}} \cot (\mathrm{~kL}) \tag{6.39}
\end{gather*}
$$

In practice, it is easier to make a terminating load impedance that is a short-circuit that it is an open circuit because of the fringing fields that could exist at an open circuit. The input impedance will be a reactance since we have chosen to study a lossless transmission line. The value of this reactance is depicted in Figure 6-7. The value will change from $-j \infty<Z_{\text {in }}<+j \infty$ and its value is specified by the length of this transmission line. This implies that we can have every possible value of reactance that is either capacitive or inductive.

## Impedance on the transmission line and matching



Figure 6-7 ${ }^{\text {MATLAB }}$. Input impedance, that in a lossless line is a reactance, of:
(a) a short-circuited transmission line and (b) an open-circuited transmission line. The vertical lines are repeated for equal intervals of $\frac{\lambda}{4}$.

| Length | Short-circuit | Open circuit |
| :--- | :--- | :--- |
| $0<L<\frac{\lambda}{4}$ | Inductive | Capacitive |
| $\frac{\lambda}{4}<L<\frac{\lambda}{2}$ | Capacitive | Inductive |

Table 6-3. Input impedance of a short-circuited or an open-circuited lossless transmission line.

## Impedance on the transmission line and matching

The reactance along a lossless transmission line of both the short circuit and the open circuit can vary from $-\infty<X_{\text {in }}<+\infty$. The precise value depends on the length of the line. We summarize the input impedance of the two lines in Table 6-3. From (6.39), this also implies that the input susceptance $B_{\text {in }}=-1 / X_{\text {in }}$ can also change and have any value from $-\infty<\mathrm{B}_{\text {in }}<+\infty$ along this lossless transmission line.

The fact that the input reactance or susceptance of a transmission line that is terminated in either a short circuit or an open circuit can have any value from $-\infty$ to $+\infty$ has some practical consequences. From a practical point of view, it is better at this stage to think in terms of admittances rather than an impedance as will become clear in a few moments.

Let us assume that we have a transmission line that is terminated with a load impedance $Z_{L}$ or a load admittance $Y_{L}$ that is not equal to the characteristic impedance $Z_{c}$ or the characteristic admittance $Y_{c}$ of the transmission line. At some distance $d$ from the load, the input admittance of the line will have a value that is equal to $Y_{C}+j B$ (Figure 6-8a). The characteristic admittance $Y_{C}$ of the transmission line is defined as $Y_{C}=1 / Z_{C}$.

At this distance d from the load, a susceptance -jB is added in parallel with the transmission line causing the admittance to the left of this point to be equal to $Y_{C}$ (Figure 6-8b). If the transmission line were a coaxial cable, this could be satisfied by connecting a circuit element from the center conductor to the outer conductor. For the strip line, this element would be between the top and the bottom conductors. A series connection of the additional matching element is more difficult to achieve in practice since it would involve the separation of the transmission line into two sections and the subsequent insertion of the matching impedance.

## Impedance on the transmission line and matching

The addition of the matching element implies that the transmission line will be matched from this point back to the signal generator. Since a transmission line terminated in a short circuit can have any value of susceptance or reactance, all that is required is to connect this line at the location $d$ as shown in Figure 6-8c. Since it is difficult to construct an open circuit because there are fringing fields and leakage currents, one normally uses a short circuit as a load. Remember that an open circuit can be found one quarter wavelength from a short circuit.


Figure 6-8. (a) The input admittance of a transmission line at a distance $d_{1}$ from the load is equal to $Y_{i n}=Y_{C}-j B$. (b) The addition of a susceptance whose value is equal to +jB at a distance $\mathrm{d}_{1}$ from the load admittance causes the input admittance at that point to the equal to $\mathrm{Y}_{\mathrm{c}}$. (c)The addition of a short circuited transmission line whose length is $d_{2}$ at the location $d_{1}$ will match the load admittance/matching transmission line from this point back to the signal generator.

The length of this transmission line that is called a stub, is chosen so that its input admittance will equal $+\mathrm{j} B$. The load impedance that now includes the added short-circuited transmission line is now matched to the rest of the

## Impedance on the transmission line and matching

transmission line. This process of matching is called single-stub matching. This adjustable length transmission line is sometimes called a trombone line. In the next section, we will examine the matching in more detail after the Smith chart is introduced.

Single-stub matching requires that there be two adjustable distances, the location of the stub $d_{1}$ and the length of the stub $d_{2}$. There are several cases where it is not practical to make the distance $d_{1}$ adjustable since it requires the milling of a narrow slit in the outer conductor of the coaxial cable or in one of the conductors of the strip line. In these cases, one may have to resort to the addition of a second or a third "stub" at additional fixed distances from the first one in order to match the load impedance to the transmission line.

One fly in the ointment remains to be removed before we leave this topic. All of the distances that were mentioned in this process were normalized by the wavelength of the wave. This implies that one can only match a load admittance at certain discrete frequencies. To improve upon this situation so that a band of frequencies can be matched, the designer's capacity that has been developed in circuit theory for designing flat pass-band filters is brought into play. Broad band operation can be achieved by a designer using these concepts, the appropriate techniques are better left to advanced courses in microwave design.

## Smith chart

### 6.6. Smith chart

In the previous section, we learned that the input impedance of a transmission line depended upon the impedance of the load, the characteristic impedance of the transmission line, and the distance between the load impedance and the point of observation. In addition, the value of the input impedance would vary periodically in space. This was demonstrated in Figure 68 for the case of a load that was either a short circuit or an open circuit. Since the input impedance (6.33) and the normalized input impedance (6.34) involve a trigonometric function, we can infer that this periodicity will also be true for an arbitrary load impedance. Rather than always run to an abacus, a slide rule, a calculator or a computer whenever an equation is presented, we will suggest that the normalized input impedance (6.34) can be interpreted graphically. This suggestion follows the idea of the Bell Labs engineer P. H. Smith. The resulting graphical tool is called a Smith chart.

The equation that describes the normalized impedance at any location (6.34) is a complex equation which is rewritten below

$$
\begin{equation*}
z_{\text {in }}(z=-L)=\frac{z_{L}+j \tan (k L)}{1+j z_{L} \tan (k L)} \tag{6.40}
\end{equation*}
$$

An arbitrary normalized load impedance will also be a complex function. We can write it as

$$
\begin{equation*}
z_{L}=r+j x \tag{6.41}
\end{equation*}
$$

where $r=\frac{R_{L}}{Z_{C}}$ and $x=\frac{X_{L}}{Z_{C}}$. The lower case notation implies a normalized impedance, an impedance that has been divided by the characteristic impedance of the lossless transmission line $Z_{c}$. The subscript "L" has also been dropped on $r$ and on x in order to later conform to Smith chart notation. Remember that we have assumed that the characteristic impedance $Z_{C}$ is a real quantity since we are treating lossless transmission lines.

## Smith chart

The reflection coefficient $R$ given in (6.24) is also a complex quantity that can be written as

$$
\begin{equation*}
\mathcal{R}=\mathcal{R}_{r}+\mathrm{j} \mathcal{R}_{i}=\frac{\mathrm{z}_{\mathrm{L}}-1}{\mathrm{z}_{\mathrm{L}}+1} \tag{6.42}
\end{equation*}
$$

In this equation, $\mathcal{R}_{r}$ is the real part and $\mathscr{R}_{i}$ is the imaginary part of the reflection coefficient $R$. Solving this equation for $z_{L}$, we obtain

$$
\begin{equation*}
z_{L}=\frac{1+R}{1-R} \tag{6.43}
\end{equation*}
$$

or

$$
\begin{equation*}
\mathrm{r}+\mathrm{jx}=\frac{1+\mathcal{R}_{r}+\mathrm{j} \mathcal{R}_{i}}{1-\mathcal{R}_{r}-\mathrm{j} \mathcal{R}_{i}} \tag{6.44}
\end{equation*}
$$

This equation is simplified by multiplying the numerator and the denominator by the complex conjugate of the denominator and we find

$$
\begin{equation*}
\mathrm{r}+\mathrm{jx}=\frac{1-\mathcal{R}_{\mathrm{r}}{ }^{2}-\mathcal{R}_{\mathrm{i}}{ }^{2}}{\left(1-\mathcal{R}_{\mathrm{r}}\right)^{2}+\mathcal{R}_{\mathrm{i}}{ }^{2}}+\mathrm{j} \frac{2 \mathcal{R}_{\mathrm{i}}}{\left(1-\mathcal{R}_{\mathrm{r}}\right)^{2}+\mathcal{R}_{\mathrm{i}}{ }^{2}} \tag{6.45}
\end{equation*}
$$

Equating the real and the imaginary parts of (6.45) yields two equations that can, after some algebra, be written in the following form:

$$
\begin{equation*}
\left(\mathcal{R}_{r}-\frac{r}{r+1}\right)^{2}+\mathcal{R}_{i}^{2}=\left(\frac{1}{r+1}\right)^{2} \tag{6.46}
\end{equation*}
$$

and

$$
\begin{equation*}
\left(R_{r}-1\right)^{2}+\left(R_{i}-\frac{1}{x}\right)^{2}=\left(\frac{1}{x}\right)^{2} \tag{6.47}
\end{equation*}
$$

Writing the equations for the real and the imaginary terms in this format allows us to recognize them both as being equations for a family of circles in a plane whose axes are labeled as $\mathbb{R}_{r}$ and $\mathscr{R}_{\mathrm{i}} .{ }^{1}$ The center and radius of each circle will be determined by the value of the normalized resistance $r$ and the normalized
${ }^{1}$ The general form of the equation of a circle can be written as:

$$
\left(\alpha-\alpha_{o}\right)^{2}+\left(\beta-\beta_{o}\right)^{2}=\rho_{\mathrm{o}}{ }^{2}
$$

The center of the circle in the $(\alpha, \beta)$ plane will be at $\left(\alpha_{0}, \beta_{0}\right)$ and the radius of the circle will be $\rho_{0}$.

## Smith chart

reactance $x$. The maximum magnitude of the reflection coefficient $R$ is equal to 1. Therefore, all of these complete circles or portions of various circles should reside within a large circle whose radius is equal to 1 . The results are depicted in Figure 6-9.

This chart is called the Smith chart and the fine scale gradiation is determined by the user. The normalized values of the resistance and the reactance which range from $0<r<\infty$ and $-\infty<x<+\infty$ have also been included in this chart. The intersection of an $r$ circle and an $x$ circle will specify the normalized impedance and the intersection of the two circles is orthogonal at that point. The evenly spaced marks on the edge of the Smith chart indicate the fraction of a half wavelength since the impedance repeats itself every half wavelength. Since we are examining a lossless transmission line, the magnitude of the reflection coefficient $R$ is a constant at every point between the load and the signal generator. This is manifested by carefully placing a circle that is centered on the origin of this coordinate system and whose radius is equal to the magnitude of the reflection coefficient $\mathcal{R}$. A clockwise rotation following this circle will be in the direction toward the signal generator and a counterclockwise rotation will be in the direction toward the load impedance. The horizontal $\mathcal{R}_{\mathrm{r}}$ axis and the vertical $\mathcal{R}_{\mathrm{i}}$ axis have been removed from the Smith chart. Although the Smith chart has been derived in terms of an impedance, it works equally well for admittances.

Before describing the application of the Smith chart, let isus summarize the properties its properties with reference to Figure 6-9. For the case of the rcircles we find that:
a) The centers of all of the r-circles lie on the horizontal axis which is the real part of the reflection coefficient $\mathbb{R r}_{r}$ axis.

## Smith chart

b) As the value of $r$ increases from $r=0$ to $r=\infty$, the circles become progressively smaller.
c) All r-circles pass through the point $R_{r}=1, R_{i}=0$.
d) The normalized resistance $r=\infty$ is at the point $\mathscr{R}_{r}=1, \mathscr{R}_{\mathrm{i}}=0$.

For the x-circles which in actuality are portions of complete circles, we conclude that:
a) The centers of all of the $x$-circles lie on $R_{r}=1$ line. The circles with $x>0$ (inductive reactance) lie above the $\operatorname{Rer}_{r}$ axis and the circles with $\mathrm{x}<0$ (capacitive reactance) lie below the $\mathscr{R}_{r}$ axis.
b) As the value of $|x|$ increases from $x=0$ to $|x|=\infty$, the circles become progressively smaller.
c) The normalized reactances $x= \pm \infty$ are at the point $\mathscr{R}_{\mathrm{r}}=1, \mathscr{R}_{\mathrm{i}}=0$.

The Smith chart has the property that the r-circles are orthogonal to the $x$ circles at every intersection. The actual load impedance that is connected to a transmission line whose characteristic impedance is $Z_{C}$ is given by $Z_{L}=Z_{C}(r+j x)$.

## Smith chart



Figure 6-9 MATLAB. The Smith chart.

## Smith chart

Example 6-10 MATLAB. On the simplified Smith chart, locate the normalized impedances (a) $z=1+j 0 ;(b) z=100+j 100 ;(c) z=0+j 0 ;(d) z=0-j 1$; and (e) $z=1+j 2$.

Answer: The real and the imaginary parts of the reflection coefficient in (6.42) can be computed by multiplying the numerator and the denominator by the complex conjugate of the denominator. We obtain

$$
\mathcal{R}=\mathcal{R}_{r}+j \mathcal{R}_{i}=\frac{z_{L}-1}{z_{L}+1}=\frac{(r-1)+j x}{(r+1)+j x} \times \frac{(r+1)-j x}{(r+1)-j x}=\frac{r^{2}-1+x^{2}}{(r+1)^{2}+x^{2}}+j \frac{2 x}{(r+1)^{2}+x^{2}}
$$



At this stage, the reader is probably somewhat dubious about the reason we have expended all of this effort in order to obtain a chart in some strange

## Smith chart

coordinate system. It may seem opaque at this stage, but hopefully it will become transparent in the next few lines. There is an oscilloscope available whose screen is calibrated in terms of a Smith chart. A high frequency circuit designer will use it in order to study the frequency response of various load impedances since a load impedance will cause a dot to meander over the screen as the frequency changes.

From (6.42) we realize that the reflection coefficient is a complex quantity that can also be expressed in polar coordinates

$$
\begin{equation*}
\mathcal{R}=|\mathcal{R}| \mathrm{e}^{\mathrm{j} \theta_{\mathrm{L}}}=\frac{\mathrm{z}_{\mathrm{L}}-1}{\mathrm{z}_{\mathrm{L}}+1} \tag{6.48}
\end{equation*}
$$

The magnitude of the reflection coefficient $|\mathcal{R}|$ can have any value in the range $0 \leq|R| \leq 1$. This value is determined by the value of the normalized load impedance $z_{L}$ that is located at $z=0$. In a lossless transmission line, this value will not change at any other location on the transmission line. Since the load impedance can be complex, there will be a phase angle $\theta_{\mathrm{L}}$ associated with the reflection coefficient $R$,

The normalized input impedance at any point $z=-z$ ' on the transmission line can be written from (6.31) as

$$
\begin{equation*}
Z\left(z=-z^{\prime}, t\right)=\frac{V\left(z=-z^{\prime}, t\right)}{I\left(z=-z^{\prime}, t\right)}=Z_{c} \frac{\left(e^{j k z^{\prime}}+R e^{-j k z^{\prime}}\right)}{\left(e^{j k z^{\prime}}-R e^{-j k z^{\prime}}\right)}=Z_{c} \frac{1+R e^{-j 2 k z^{\prime}}}{1-R e^{-j 2 k z^{\prime}}} \tag{6.49}
\end{equation*}
$$

If we substitute (6.48) into (6.49), we obtain

$$
\begin{equation*}
\mathrm{z}_{\text {in }}=\frac{\mathrm{Z}_{\text {in }}}{Z_{\mathrm{c}}}=\frac{1+\mathbb{R} \mathrm{e}^{-\mathrm{R} k z^{\prime}}}{1-\mathbb{R} \mathrm{e}^{-\mathrm{j} 2 k z^{\prime}}}=\frac{1+|\mathcal{R}| \mathrm{e}^{\mathrm{j} \phi}}{1-|\mathcal{R}| \mathrm{e}^{\mathrm{j} \phi}} \tag{6.50}
\end{equation*}
$$

where

$$
\begin{equation*}
\phi=\theta_{L}-2 k z^{\prime} \tag{6.51}
\end{equation*}
$$

Remember that the magnitude of the reflection coefficient remains a constant as we move along the transmission line back toward the signal generator. In comparing (6.43) and (6.50), we note that the only difference is a

## Smith chart

phase shift $\phi$ that is linearly proportional to the distance $z$ '. This implies that this translation on the Smith chart can be easily accomplished by just rotating the initial value of the load impedance along a circle whose radius is equal to the magnitude of the reflection coefficient $|R|$. A clockwise rotation will be in the direction toward the signal generator and a counterclockwise rotation will be in the direction toward the load impedance. The amount of rotation depends upon the distance $2 k z^{\prime}=\frac{4 \pi z^{\prime}}{\lambda}$ where this phase angle must be subtracted from the initial value that was used to locate the load impedance upon the Smith chart. One notes that this impedance will repeat itself every half wavelength.

If we choose the distance to be equal to be $z^{\prime}=\frac{\lambda}{4}$, the rotation will be equal to $\pi$ radians. Referring back to (6.35), you remember that this distance converts the numerical value of an impedance into the numerical value of an admittance. For example, the impedance $z=0.5+j 0.5$ has the admittance of

$$
y=\frac{1}{0.5+j 0.5}=1-j 1
$$

Rather than performing this calculation with a calculator, it can also be found directly from the Smith chart. The normalized impedance is first located on the chart. A semi circle that has a radius that is equal to the reflection coefficient is drawn on the chart along with a straight line that passes through the center of the Smith chart and this impedance. The intersection of the line with the semi circle will directly yield the proper value of the admittance. This is illustrated in Figure 6-10.

From Figure 6-10, we also note that a normalized load impedance equal to 0 will yield a normalized load admittance that is equal to $\pm \mathrm{j} \infty$. This implies that the input impedance that is $\frac{\lambda}{4}$ from a short-circuit will have an input impedance that is equal to that found from an open circuit. In practice, it is difficult to make

## Smith chart

an open circuit in the transmission line due to the fringing fields and a quarter wavelength shorted transmission line can be used to create an open circuit.


Figure $6-10^{\text {MATLAB }}$. The transformation of an impedance to an admittance using the Smith chart. A semi circle whose radius is equal to the magnitude of the reflection coefficient is drawn. This corresponds to a distance of $\frac{\lambda}{4}$ on the transmission line.

Example 6-11 ${ }^{\text {MATLAB }}$. A load impedance $\mathrm{Z}_{\mathrm{L}}=50+j 50 \Omega$ terminates a transmission line that is 5 meters long and has a characteristic impedance of $Z_{C}=$ $25 \Omega$. Using the Smith chart, find the impedance at the signal generator if the

## Smith chart

frequency of oscillation $f=1 \times 10^{5} \mathrm{~Hz}$. The phase velocity for this transmission line is $v=2 \times 10^{6} \mathrm{~m} / \mathrm{s}$.

Answer: The wavelength $\lambda$ is

$$
\lambda=\frac{v}{f}=\frac{2 \times 10^{6}}{1 \times 10^{5}}=20 \mathrm{~m}
$$

The distance between the load and the generator is $\lambda / 4$. The normalized load impedance is

$$
z_{\mathrm{L}}=\frac{50+\mathrm{j} 50}{25}=2+\mathrm{j} 2
$$

The normalized load impedance is first located on the Smith chart.
Using a compass located at the center of the Smith chart, rotate this point a distance $\lambda / 4$ toward the generator. The normalized load impedance at the generator $z_{i n}$ as read from the Smith chart is $z_{i n}=0.25-j 0.25$. Therefore, the input impedance that is connected to the signal generator is

$$
Z_{i n}=Z_{c} Z_{i n}=25(0.25-j 0.25)=6.25-j 6.25 \Omega
$$



An examination of Figure 6-10 will also reveal an important potential application of the Smith chart. This chart can equally well be interpreted in terms of an impedance or an admittance and we note that a constant coefficient circle will pass through the " 1 " circle at two locations. Let us interpret this as an admittance chart. At the locations where it has passed through the " $\mathrm{g}=1$ " circle, an admittance can be added in parallel with the transmission line at either of these locations. Rather than separating a transmission line and inserting a matching element in series, it is better to just insert the matching element in parallel. It should have a value that will cause the input admittance from that

## Smith chart

location back to the signal generator to have a normalized value of1. This is called matching a transmission line. Let us illustrate this with an example.

Example 6-12 MATLAB . A load admittance has the value $y_{L}=0.2-j 0.5$. Find the locations where a matching admittance should be placed. In addition, find the value for the matching admittance.

Answer: The input admittance will have the value $y_{i n}=1 \pm j b$ at two locations that can be obtained from the Smith chart by rotating the load admittance on a constant reflection:coefficient circle. This value has already been determined from the value at the load admittance. At these locations, the real part and the imaginary part of the reflection coefficient will be

$$
\mathcal{R}_{r}=\frac{\mathrm{b}^{2}}{4+\mathrm{b}^{2}} \quad \mathcal{R}_{i}=\frac{2 \mathrm{~b}}{4+\mathrm{b}^{2}}
$$

where we have inserted $\mathrm{g}=1$ in (6.42) and have understood the Smith chart in terms of the admittance. Since the magnitude of the reflection coefficient is already known $(|\mathcal{R}|=0.7257)$, we just have to solve the algebraic equation for the value of the susceptance $b$ that must be inserted at these locationands

$$
|R|^{2}=\left(\frac{b^{2}}{4+b^{2}}\right)^{2}+\left(\frac{2 b}{4+b^{2}}\right)^{2}
$$

This equation results in the following algebraic equation that must be solved.

$$
\left(1-|\mathcal{R}|^{2}\right)^{4}+\left(4-8|\mathcal{R}|^{2}\right) b^{2}-16|\mathcal{R}|^{2}=0
$$

The roots of this polynomial are calculated to be

$$
b^{2}=4.45 ; b^{2}=-4
$$

and because b must be a real number we neglect the second solution. Thus, we find two real solutions $b_{1} \approx 2.11$ and $b_{2} \approx-2.11$, which are in good agreement with the two graphical solutions, obtained from the Smith chart. The above calculation using the Smith chart is shown below.

## Smith chart



The location of the load admittance is indicated with a $\square$ and the radius of the circle which passes through this point is equal to the magnitude of the reflection coefficient circle. The circle passes through the " $g=1$ " circle at two locations. The closest one is at a distance $\mathrm{d}_{1}$ from the load admittance. One should insert an admittance equal to -jb at that location. The transmission line will be matched from that point back to the signal generator. There is a second location $d_{2}$ that is further from the load admittance where one could insert an admittance that is equal to +jb and could be used if it is inconvenient to choose the first location.

The matching elements which were employed in the above example could easily be created by using a section of a transmission line that is terminated in a short-circuit. The disadvantage on using the single stub matching technique is

## Smith chart

that the distance $d$ is variable. In practice, this would typically be accomplished with a narrow slit inserted into the coaxial cable or the strip line. There are techniques that can be used to eliminate or at least minimize this problem. All of the techniques are frequency sensitive since the distances depend upon the wavelength.

## Transient effects and the bounce diagram

### 6.7. Transient effects and the bounce diagram

The study of transmission lines that are excited with a sinusoidal voltage generator could continue for many more pages and all possible nuances would still not be uncovered. The reader may wish to explore this topic still further and this writer does not want to discourage these efforts. However in the limited time and space available to us, we should travel on a slightly different path, the path that integrated circuit designers regularly travel in designing chips for computers. As we enter this path, we must be cautious so that a passing zero or one does not strike us and knock us over. The knowledge that we have gained from the time-harmonic analysis should fortify us in our attempt to gain an understanding of transient effects and pulse propagation.

Consider the transmission line shown in Figure 6-11 in which a battery is connected to a transmission line. The battery has an internal impedance $Z_{b}$, the transmission line is represented with a characteristic impedance $Z_{C}$ and the transmission line is terminated in a load impedance $\mathrm{Z}_{\mathrm{L}}$. We will assume that they are pure resistances at this stage and that the signal will propagate with a velocity v . The voltage wave is governed by the Telegraphist's equations and the ratio of the voltage wave to the current wave is given by the characteristic impedance $Z_{C}$ of the transmission line. Since the load impedance is located a distance $L$ from the battery-switch, it will take $\frac{L}{v}$ seconds before the signal arrives at the load impedance.


Figure 6-11. The battery is connected to the transmission line along with a switch that is closed at $t=0$.

The amplitude of the wave $\mathrm{V}_{1}$ that is launched on the transmission line can be easily calculated using the "voltage divider" rule that probably has already been encountered in an earlier course. This amplitude is dictated by the two impedances at the input of the transmission line, namely the battery impedance and the characteristic impedance of the transmission line. It is given by

$$
\begin{equation*}
V_{1}=\frac{Z_{C}}{Z_{B}+Z_{C}} V_{B} \tag{6.52}
\end{equation*}
$$

After a time $\tau=\frac{\mathrm{L}}{\mathrm{V}}$, the front of this propagating voltage step arrives at the load impedance $Z_{L}$. At this time, a portion of this incident voltage will be reflected from the load impedance and a portion of this incident voltage is "transmitted" or absorbed by this load impedance. The reflection coefficient at the load $\mathcal{R}_{\mathrm{L}}$ is given by

$$
\begin{equation*}
\mathcal{R}_{\mathrm{L}}=\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}=\frac{V_{2}}{V_{1}} \tag{6.53}
\end{equation*}
$$

The amplitude of the reflected voltage step $V_{2}$ can be positive or negative depending upon thehe relative values of the impedances that appear in (6.53).

## Transient effects and the bounce diagram

Eventually, the front of this propagating reflected voltage step $\mathrm{V}_{2}$ will reach the battery impedance at a time $2 \tau=2 \frac{\mathrm{~L}}{\mathrm{v}}$. This front will be reflected from the battery impedance with a reflection coefficient $\mathcal{R}_{b}$ where

$$
\begin{equation*}
\mathcal{R}_{\mathrm{B}}=\frac{\mathrm{Z}_{\mathrm{B}}-\mathrm{Z}_{\mathrm{C}}}{\mathrm{Z}_{\mathrm{B}}+\mathrm{Z}_{\mathrm{C}}}=\frac{\mathrm{V}_{3}}{\mathrm{~V}_{2}} \tag{6.54}
\end{equation*}
$$

The battery impedance is essentially a "load impedance" for the incident wave $V_{2}$.

The front of this reflected propagating voltage step $\mathrm{V}_{3}$ will reach the load impedance where a portion of this voltage step will be reflected by the load impedance. This process may continue indefinitely.

The front of the propagating voltage step "bounces" back and forth between the load impedance and the battery impedance. There is a graphical technique in order to evaluate the voltage at any location in space as a function of the time. In other words, it is possible to predict the response of an oscilloscope that is inserted at a certain location on the transmission line. This is illustrated by plotting the trajectory of this front as shown in Figure 6-12. This diagram is called a bounce diagram. It is convenient to have the vertical axis and the horizontal axis which are respectively time and position both be dimensionless. The prediction of the temporal response at a certain location is obtained by inserting a vertical line at that location. The intersection of the trajectory with this line indicates that the voltage at that location will change its value by the amplitude of that particular component of the wave.

## Transient effects and the bounce diagram



Figure 6-12. The bounce diagram. The magnitude of the slope of each line is equal to 1 . The amplitude of each individual component is usually specified. A vertical line at a certain position on the transmission line indicates the location of an oscilloscope probe. The intersection of this vertical line and the trajectory yields the times when the voltage will change. These points are indicated with the short horizontal lines and the subsequent voltage during that interval is given.

Recall that a battery is the source of the voltage and it has a constant value. Therefore, it is reasonable to assume that the voltage behind the propagating front of all the components will also be a constant and will have the same value of the front. The voltage at any location along the transmission line is just the summation of the individual components

$$
\begin{equation*}
\mathrm{V}=\mathrm{V}_{1}+\mathrm{V}_{2}+\mathrm{V}_{3}+\Lambda=\frac{\mathrm{Z}_{\mathrm{C}}}{\mathrm{Z}_{\mathrm{B}}+\mathrm{Z}_{\mathrm{C}}}\left(1+\mathcal{R}_{\mathrm{L}}+\mathcal{R}_{\mathrm{L}} \mathcal{R}_{\mathrm{B}}+\Lambda\right) \mathrm{V}_{\mathrm{B}} \tag{6.55}
\end{equation*}
$$

The reader might suspect that an asymptotic value might be reached as the voltage step bounces back and forth between the battery and the load

## Transient effects and the bounce diagram

impedance in Figure 6-11. Such suspicions will be well rewarded. Every voltage step that is incident on the load impedance will be reflected with a reflection coefficient $\mathcal{R}_{\mathrm{L}}$ and every voltage step incident on the battery will be reflected with a reflection coefficient RS. Adding up all of the individual contributions and regrouping the terms will lead to the following

$$
\begin{align*}
\mathrm{V} & =\mathrm{V}_{1}+\mathrm{V}_{2}+\mathrm{V}_{3}+\mathrm{V}_{4}+\Lambda=\frac{\mathrm{Z}_{\mathrm{C}}}{\mathrm{Z}_{\mathrm{B}}+\mathrm{Z}_{\mathrm{C}}}\left(1+\mathcal{R}_{\mathrm{L}}+\mathcal{R}_{\mathrm{L}} \mathcal{R}_{\mathrm{B}}+\mathcal{R}_{\mathrm{L}} \mathcal{R}_{\mathrm{B}} \mathcal{R}_{\mathrm{L}}+\Lambda\right) \mathrm{V}_{\mathrm{B}} \\
& \left.=\frac{\mathrm{Z}_{\mathrm{C}}}{\mathrm{Z}_{\mathrm{B}}+\mathrm{Z}_{\mathrm{C}}}\left\{1+\left(\mathcal{R}_{\mathrm{L}} \mathcal{R}_{\mathrm{B}}\right)+\left(\mathcal{R}_{\mathrm{L}} \mathcal{R}_{\mathrm{B}}\right)^{2}+\Lambda\right]+\mathcal{R}_{\mathrm{L}}\left[1+\left(\mathcal{R}_{\mathrm{L}} \mathcal{R}_{\mathrm{B}}\right)+\left(\mathcal{R}_{\mathrm{L}} \mathcal{R}_{\mathrm{B}}\right)^{2}+\Lambda\right]\right\} \mathrm{V}_{\mathrm{B}} \tag{6.56}
\end{align*}
$$

The terms within the square brackets can be written using the summation relation that

$$
\begin{equation*}
1+\xi+\xi^{2}+\Lambda=\frac{1}{1-\xi} \text { for }|\xi|<1 \tag{6.57}
\end{equation*}
$$

For the cases where $\left|\mathcal{R}_{R} \mathcal{R}_{B}\right|<1$, we may employ this summation relation and obtain

$$
\begin{equation*}
V=\frac{Z_{C}}{Z_{B}+Z_{C}}\left\{\frac{1+R_{L}}{1-R_{L} R_{B}}\right\} V_{B} \tag{6.58}
\end{equation*}
$$

Substituting the definitions for the reflection coefficient at the load impedance and at the battery impedance, ( 6.58 ) becomes

$$
V=\frac{Z_{C}}{Z_{B}+Z_{C}}\left\{\frac{1+\left(\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}\right)}{1-\left(\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}\right)\left(\frac{Z_{B}-Z_{C}}{Z_{B}+Z_{C}}\right)}\right\} V_{B}
$$

This simplifies to

$$
\begin{equation*}
V=\frac{Z_{L}}{Z_{B}+Z_{L}} V_{B} \tag{6.59}
\end{equation*}
$$

Based on our previous experience using circuits, this asymptotic value should not be too surprising.

## Transient effects and the bounce diagram

Example 6-13 MATLAB . A 12 V battery is connected via a switch to a transmission line that is 6 m long. The characteristic impedance of the transmission line is $50 \Omega$, the battery impedance is $25 \Omega$, and the terminated in a load impedance of $25 \Omega$. The velocity of propagation along this transmission line is $2 \times 10^{6} \mathrm{~m} / \mathrm{s}$. Find and sketch the voltage at the midpoint of this transmission line during the time interval $0<\mathrm{t}<9 \mu \mathrm{~s}$.


Answer: The amplitude of the wave that is launched on the transmission line is calculated from

$$
V_{1}=\frac{Z_{C}}{Z_{B}+Z_{C}} V_{B}=\frac{50}{25+50} 12=8 V
$$

The reflection coefficient at the load is equal to

$$
\mathcal{R}_{L}=\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}=\frac{25-50}{25+50}=-\frac{1}{3}
$$

The reflection coefficient at the battery is equal to

$$
\mathcal{R}_{B}=\frac{Z_{B}-Z_{C}}{Z_{B}+Z_{C}}=\frac{25-50}{25+50}=-\frac{1}{3}
$$

## Transient effects and the bounce diagram

In order to calculate the voltage at the midpoint of the transmission line, we make use of the bounce diagram. In this case, we clearly identify the amplitudes of the waves. The normalized time is $\frac{\mathrm{vt}}{\mathrm{L}}=\frac{2}{6} \mathrm{t}$ The bounce diagram is first obtained.

Using the bounce diagram, the voltage at the midpoint of the transmission line is equal to 0 until the front of the wave arrives. The voltage increases to the amplitude of the wave, remains at that value until the wave that is reflected from the load impedance passes the midpoint. This reflected wave is reflected again at the battery and arrives at the midpoint. The picture on the right depicts the expected response of the oscilloscope that is located at the midpoint of the transmission line. The final value of this voltage is calculated from (6.59) to be

$$
V=\frac{Z_{L}}{Z_{B}+Z_{L}} V_{B}=\frac{25}{25+25} 12=6 \mathrm{~V} .
$$



## Transient effects and the bounce diagram

The current that flows through the load impedance is equal to the voltage heat the load divided by the load impedance. The long time asymptotic value is given by

$$
\begin{equation*}
I=\frac{V}{Z_{L}}=\frac{1}{Z_{B}+Z_{L}} V_{B} \tag{6.60}
\end{equation*}
$$

Example 6-14 ${ }^{\text {MATLAB }}$. A battery with zero internal impedance has an open circuit voltage of 100 volts. At a time $t=0$, this battery is switched into a $50 \Omega$ air-dielectric coaxial cable via a $150 \Omega$ resistor. The cable is 300 m long and is terminated in a load of $33.3 \Omega$.
(a) Sketch a bounce diagram for the first $4 \mu \mathrm{~s}$ after the switch is closed.
(b) Draw a graph of the voltage that appears across the load impedance as a function of time.
(c) Find the asymptotic values of $V_{L}$ and $I_{L}$ as $t \rightarrow \infty$.


Answer: (a) The two reflection coefficients and the incident voltage step that propagates on the line are given by

$$
\begin{aligned}
& \mathcal{R}_{B}=\frac{Z_{B}-Z_{C}}{Z_{B}+Z_{C}}=\frac{150-50}{150+50}=\frac{1}{2} \\
& \mathcal{R}_{\mathrm{L}}=\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}=\frac{33.3-50}{33.3+50}=-\frac{1}{5}
\end{aligned}
$$

## Transient effects and the bounce diagram

$$
V_{1}=\frac{Z_{C}}{Z_{B}+Z_{C}} V_{B}=\frac{50}{150+50} 100=25 \mathrm{~V}
$$



In the (b) Since the coaxial cable is filled with air, the velocity of propagation equals $3 \times 10^{8} \mathrm{~m} / \mathrm{s}$. The voltage signal takes $1 \mu \mathrm{~s}$ to travel from one end to the other.
(c) The asymptotic voltages and currents as computed from (6.59) and (6.60) are for $V_{t_{\rightarrow \infty}}=18.2 \mathrm{~V}$ and $\mathrm{I}_{\mathrm{t} \rightarrow \infty}=0.55 \mathrm{~A}$.

Transient effects and the bounce diagram


## Pulse propagation

### 6.8. Pulse propagation

In the previous section, we examined the transient characteristics of a step voltage as it propagated along a transmission line. We observed that the front propagated with a definite velocity and it took a nonzero time for the signal to pass from one point to another. This time was found to be of the order of $\mathrm{L} / \mathrm{v}$ where $L$ is the distance to be traveled and $v$ is the velocity of propagation of the signal. In this section, we will devote our energies to the study of the propagation of a voltage pulse along a transmission line. An emphasis on this particular topic is certainly justified due to its practical importance in digital integrated circuits and in practical laboratory measurements.

Let us consider a transmission line that connects a pulse generator to a load impedance as shown in Figure 6-13.


Figure 6-13. A pulse generator $\mathrm{V}_{\mathrm{g}}$ that has an internal impedance of $\mathrm{Z}_{\mathrm{g}}$ is connected to a transmission line that has a characteristic impedance $Z_{c}$ and is terminated in a load impedance $\mathrm{Z}_{\mathrm{L}}$.

As in the previous section describing transient effects, we can calculate the amplitude $\mathrm{V}_{1}$ of the pulse launched on the transmission line using a voltage divider rule

## Pulse propagation

$$
\begin{equation*}
V_{1}=\frac{Z_{c}}{Z_{g}+Z_{c}} V_{g} \tag{6.61}
\end{equation*}
$$

We will assume that the temporal width $\Delta \mathrm{t}$ of the pulse is much less than the time that it takes for the pulse to travel from one end of the transmission line to the other, that is $\Delta \mathrm{t} \ll \frac{\mathrm{L}}{\mathrm{V}}$. In practical VLSI circuits whose dimensions are of the order of microns ( 1 micron $=1 \mu \mathrm{~m}=10^{-6}$ meters), this time is of the order of

$$
\frac{L}{v} \approx \frac{10^{-6} \mathrm{~m}}{3 \times 10^{8} \mathrm{~m} / \mathrm{s}} \approx 3 \times 10^{-15} \mathrm{~s}
$$

The laws of nature set an upper limit to the speed of such a circuit. We have used the vacuum value for the velocity of light in this estimate although we know in actuality, it will be decreased by the square root of the dielectric constant of the dielectric that exists between the two conducting surfaces.

The pulse $\mathrm{V}_{1}$ that is launched from the signal generator and on the transmission line will propagate toward the load impedance. A portion of the pulse will be absorbed by the load impedance and a portion of the pulse will be reflected back toward the signal generator. The amount that is absorbed or the amount that is reflected will be determined by the ratio of the load impedance to the characteristic impedance of the transmission line. We can compute the value of the absorbed signal from a consideration of the flow of energy along the transmission line.

The energy of the incident pulse $P_{\text {inc }} \Delta t$ will be divided into the energy in the reflected wave $\mathrm{P}_{\text {ref }} \Delta t$ and the energy that is absorbed in the load impedance $P_{\text {abs }} \Delta t$. The energy must be conserved in this junction. We write

$$
\begin{equation*}
\mathrm{P}_{\text {inc }} \Delta \mathrm{t}=\mathrm{P}_{\text {ref }} \Delta \mathrm{t}+\mathrm{P}_{\mathrm{abs}} \Delta \mathrm{t} \tag{6.62}
\end{equation*}
$$

The common factor $\Delta t$ will cancel in this expression and the various terms for the power can be written in terms of the impedances and the reflection coefficient

## Pulse propagation

$$
\begin{equation*}
\frac{\mathrm{V}_{1}^{2}}{\mathrm{Z}_{\mathrm{C}}}=\frac{\left(R \mathrm{~V}_{1}\right)^{2}}{\mathrm{Z}_{\mathrm{C}}}+\frac{\mathrm{V}_{\mathrm{L}}{ }^{2}}{\mathrm{Z}_{\mathrm{L}}} \tag{6.63}
\end{equation*}
$$

where $V_{L}$ is the voltage that appears across the load impedance. Inserting the expression for the reflection coefficient (6.24), we write

$$
\begin{equation*}
\frac{V_{1}^{2}}{Z_{C}}=\frac{\left(\left[\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}\right] V_{1}\right)^{2}}{Z_{C}}+\frac{V_{L}^{2}}{Z_{L}} \tag{6.64}
\end{equation*}
$$

Solving for the ratio of this voltage divided by the incident voltage $\mathrm{V}_{1}$, we obtain

$$
V_{L}=V_{1} \sqrt{\frac{Z_{L}}{Z_{C}}-\frac{\left[\frac{Z_{L}-Z_{C}}{Z_{L}+Z_{C}}\right]^{2} Z_{L}}{Z_{C}}}=V_{1} \frac{2 Z_{L}}{Z_{L}+Z_{C}}
$$

or

$$
\begin{equation*}
\mathcal{T}=\frac{\mathrm{V}_{\mathrm{L}}}{\mathrm{~V}_{1}}=\frac{2 Z_{\mathrm{L}}}{Z_{\mathrm{L}}+Z_{\mathrm{C}}} \tag{6.65}
\end{equation*}
$$

where $\mathcal{T}$ is defined as the transmission coefficient. Note that this result agrees with the result that we previously obtained for plane waves using a different approach. That approach which matched boundary conditions at the interface could have been repeated here, but we wanted to make use the conservation of energy principle.


Figure 6-14. An integrated circuit transmission line element.

## Pulse propagation

This point can be emphasized if we consider the joining of two transmission lines that may have different dimensions as shown in Figure 6-14. The characteristic impedance of the transmission line between the two electrodes is different in the two regions of the integrated circuit. A portion of the signal launched from one electrode will reach the second electrode and a portion will be reflected back to the original electrode.

Example 6-15. Calculate the transmission coefficient $\mathcal{T}$ for a wave that is propagating in the $+z$ direction in a coaxial cable. The relative dielectric constant of the separating dielectric in the region $z<0$ is 2 and in the region $z>0$ is 3 . The physical dimensions of the cable are the same in all regions.

Answer: Using (6.65) and the results given in Table 6-2, we write

$$
\tau=\frac{2 \sqrt{\frac{\mu}{3 \varepsilon_{o}}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)}{\sqrt{\frac{\mu}{3 \varepsilon_{o}}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)+\sqrt{\frac{\mu}{2 \varepsilon_{o}}}\left(\frac{\ln \left(\frac{b}{a}\right)}{2 \pi}\right)}=\frac{\frac{2}{\sqrt{3}}}{\frac{1}{\sqrt{3}}+\frac{1}{\sqrt{2}}}=0.9
$$



Figure 6-15. Snapshots of a voltage pulse crossing a discontinuity in a transmission line. The pictures are taken at equal intervals in time and the

## Pulse propagation

velocity of propagation has $\mathrm{v}|\mathrm{z}>0>\mathrm{v}| \mathrm{z}<0$. In addition, the characteristic impedances of the two lines have the relative values $Z_{C}\left|z>0>Z_{C}\right| z<0$.

A sequence of pictures of an incident pulse propagating from one transmission line into another one is shown in Figure 6-15. Note that the reflected signal propagates with the same magnitude of velocity as the incident pulse while the transmitted pulse propagates with a different velocity. This is because the equivalent inductances and capacitances are different on the two sides of the discontinuity. The polarity of the reflected pulse will be the same as the incident one due to the sign of the reflection coefficient between the characteristic impedances of the two transmission lines.

Knowing the velocity of propagation v on a transmission line has some very practical consequences. Let us assume that we can launch a pulse on a transmission line and measure the time $\Delta T$ that it takes for the reflected pulse to return. From this data, we can compute exactly the unknown distance $D$ from the pulse generator where the reflection took place since $\Delta T=2 D / v$. Imagine that you are trying to locate a fault in an integrated circuit or a short circuit in a cable that is buried underground. Knowing where to probe or dig will save many hours of frustration. This practical technique is called time domain reflectometry.

Example 6-16 Using the reflection coefficient $\mathcal{R}$ and the transmission coefficient $\mathcal{T}$ show that energy is conserved at the junction between two lossless transmission lines.


## Pulse propagation

Answer: Conservation of energy implies that (6.62) must be satisfied. This implies that

$$
\frac{\mathrm{V}_{\text {inc }}{ }^{2}}{\mathrm{Z}_{\mathrm{C} 1}}=\frac{\left(\mathbb{R} \mathrm{V}_{\text {inc }}\right)^{2}}{\mathrm{Z}_{\mathrm{C} 1}}+\frac{\left(\mathcal{T} \mathrm{V}_{\text {inc }}\right)^{2}}{\mathrm{Z}_{\mathrm{C} 2}}
$$

Substituting in the values for the reflection coefficient and the transmission coefficient, we write after canceling the value of the incident voltage wave

$$
\frac{1}{Z_{C 1}}=\frac{|\mathcal{R}|^{2}}{Z_{C 1}}+\frac{|\tau|^{2}}{Z_{C 2}}=\frac{\left(\frac{Z_{C 2}-Z_{C 1}}{Z_{C 2}+Z_{C 1}}\right)^{2}}{Z_{C 1}}+\frac{\left(\frac{2 Z_{C 2}}{Z_{C 2}+Z_{C 1}}\right)^{2}}{Z_{C 2}}=\frac{1}{Z_{C 1}}
$$

Example 6-17. A 1 volt pulse propagates from $z<0$ on a transmission line. The line is terminated in an open circuit @ $z=0$. Four oscilloscopes are triggered by the same pulse generator and are located at: $z_{a}=-6 ; z_{b}=-4 ; z_{c}=-2$; and $z_{d}=0$ (meters). Find the velocity of propagation and interpret the voltage signals on the oscilloscopes. Sketch the corresponding voltage signals if the transmission line is terminated in a short circuit.

b


C

d


## Pulse propagation

Answer. From the traces on oscilloscopes $a$ and $b$, we find the velocity of propagation to be $v=\Delta z / \Delta t=2 \mathrm{~m} / 1 \mu \mathrm{~s}=2 \times 10^{6} \mathrm{~m} / \mathrm{s}$. Oscilloscope d is at the location of the open circuit and the incident and the reflected pulses add together. The signals that are detected after $t=4 \mu \mathrm{~s}$ are the reflected pulses that propagate toward the pulse generator.

The voltage signals detected by the oscilloscopes if the transmission line is terminated in a short circuit are depicted below. The voltage across the short circuit must be zero, hence the signal at oscilloscope d is zero.


It should be noted that the two oscilloscope pictures correspond to the voltage pulses. If the oscilloscope pictures would correspond to the current pulses, the first picture would correspond to a load impedance that was a shortcircuit and the second picture would correspond to an open circuit.

## Lossy transmission lines

### 6.9. Lossy transmission lines

Except for the example of diffusion, the transmission line models consisted of inductors and capacitors resulting in a characteristic impedance that was a real number. In this section, we'll extend this discussion to include ohmic losses within the conductors and leakage currents between conductors. The model that was introduced earlier will have to be modified to take into account these effects. This can be accomplished by inserting a resistance in series with the inductor and a conductance in parallel with the capacitor. In this case, the characteristic impedance will become a complex quantity.

We can model these additional losses with the transmission line section model shown in Figure 6-16.


Figure 6-16. The model of a section whose length is $\Delta z$ of a transmission line that includes loss terms is shown. The units of all of the elements are per-unit length.

Following the same procedure that we employed to write the first-order partial differential equations for the lossless transmission line (6.3) and (6.6), we obtain

$$
\begin{align*}
& \frac{\partial l(z, t)}{\partial z}=-\hat{C} \frac{\partial V(z, t)}{\partial t}-\hat{G} V(z, t)  \tag{6.66}\\
& \frac{\partial V(z, t)}{\partial z}=-\hat{L} \frac{\partial l(z, t)}{\partial t}-\hat{R} l(z, t) \tag{6.67}
\end{align*}
$$

where the circuit elements are defined as:

$$
\begin{equation*}
\hat{\mathrm{L}}=\frac{\mathrm{L}}{\Delta \mathrm{z}} ; \hat{\mathrm{C}}=\frac{\mathrm{C}}{\Delta \mathrm{z}} ; \hat{\mathrm{R}}=\frac{\mathrm{R}}{\Delta \mathrm{z}} ; \text { and } \hat{\mathrm{G}}=\frac{\mathrm{G}}{\Delta \mathrm{z}} . \tag{6.68}
\end{equation*}
$$

The set of the first order partial differential equations (6.66) and (6.67) are also known as the Telegraphist's equations since the wires that crossed the country did possess loss.

In order to include the loss, it is convenient to assume that there is a time harmonic excitation of the transmission line. If we make this assumption, then (6.66) and (6.67) become

$$
\begin{align*}
& \frac{\partial(z)}{\partial z}=-[\hat{G}+j \omega \hat{C}]  \tag{6.69}\\
& \frac{\partial(z, t)}{\partial z}=-[\hat{R}+j \omega \hat{L}](z) \tag{6.70}
\end{align*}
$$

where we had introduced the phaser notation for the currents and voltages. We recognize the terms within the square brackets can be replaced with a distributed admittance $\hat{Y}$ and a distributed impedance $\hat{Z}$ quantities. Hence we can rewrite these two equations as

$$
\begin{align*}
& \frac{d(z)}{d z}=-\hat{Y} \quad(z)  \tag{6.71}\\
& \frac{d \quad(z)}{d z}=-\hat{Z} \quad(z) \tag{6.72}
\end{align*}
$$

## Lossy transmission lines

The coupled first-order ordinary differential equations can be replaced with a second order ordinary differential equation for either of the dependent variables. We write

$$
\begin{align*}
\frac{d^{2}(z)}{d z^{2}} & =\hat{Z} \hat{Y} \quad(z)  \tag{6.73}\\
\frac{d^{2}(z)}{d z^{2}} & =\hat{Z} \hat{Y} \tag{z}
\end{align*}
$$

The solution of these two ordinary differential equations is given by

$$
\begin{gather*}
I(z, t)=I_{1} e^{j \omega t-y z}+I_{2} e^{j \omega t+y z}  \tag{6.75}\\
V(z, t)=V_{1} e^{j \omega t-y z}+V_{2} e^{j \omega t+\gamma z} \tag{6.76}
\end{gather*}
$$

where

$$
\begin{equation*}
\gamma=\alpha+j \beta= \pm \sqrt{\hat{Z} \hat{Y}}= \pm \sqrt{(\hat{R}+j \omega \hat{L})(\hat{G}+j \omega \hat{C})} \tag{6.77}
\end{equation*}
$$

In writing the solutions (6.75) and (6.76), we have reintroduced the time harmonic dependence. In addition, the constants that appear before the exponential terms are evaluated from either the amplitude of the excitation signal or any reflection that is imposed by a load impedance. The explicit inclusion of the real part $\alpha$ and the imaginary part $\beta$ of the complex propagation constant $\gamma$ is presented in (6.77). One would expect that the amplitude of the wave should always decrease as the wave propagates into a lossy material. This expectation will specify the proper sign that is to be chosen. This is also similar to the propagation of a plane wave into a conducting material that we encountered previously. This is shown in Figure 6-17. From this figure, it is possible to determine the numerical values of the constants $\alpha$ and $\beta$.

For the case where the loss terms are small, we can approximate this complex propagation constant (6.77) to be

$$
\begin{align*}
\gamma=\alpha+j \beta & \left.= \pm \sqrt{(j \omega \hat{L})}(j \omega \hat{C}) \sqrt{\left(1+\frac{\hat{R}}{j \omega \hat{L}}\right)}\right)\left(1+\frac{\hat{\mathrm{G}}}{\mathrm{j} \omega \hat{\mathrm{C}}}\right) \\
& \approx \pm \mathrm{j} \omega \sqrt{\hat{\mathrm{~L}} \hat{\mathrm{C}}}\left[1-\mathrm{j}\left(\frac{\hat{R}}{2 \omega \hat{L}}+\frac{\hat{\mathrm{G}}}{2 \omega \hat{\mathrm{C}}}\right)\right] \tag{6.78}
\end{align*}
$$



Figure 6-17 ${ }^{\text {MATLAB }}$. The evolution in space of a time harmonic voltage signal at an instant in time as a function of space. From this figure, one can determine the complex propagation constant $\gamma=\alpha+j \beta=-\frac{1}{2}-j \frac{2 \pi}{2}=-\frac{1}{2}-j \pi$.

From the approximation (6.78), we see that the wave will propagate in the $\pm z$ directions but the amplitude will attenuate as it propagates. The attenuation constant $\alpha$ will approximately be given by

$$
\begin{equation*}
\alpha \approx \omega \sqrt{\hat{L} \hat{C}}\left(\frac{\hat{R}}{2 \omega \hat{L}}+\frac{\hat{G}}{2 \omega \hat{C}}\right)=\frac{1}{2} \sqrt{\hat{L} \hat{C}}\left(\frac{\hat{R}}{\hat{L}}+\frac{\hat{G}}{\hat{C}}\right) \tag{6.79}
\end{equation*}
$$

Note that the attenuation constant is independent of frequency.
Example 6-18. Find the complex propagation constant if the circuit elements satisfy the ratio $\frac{\hat{R}}{\hat{L}}=\frac{\hat{G}}{\hat{C}}$. Interpret the propagation of such a signal that propagates on this line.

Answer: The complex propagation constant (6.77) can be written as

$$
\gamma=\alpha+j \beta= \pm \sqrt{(\hat{R}+j \omega \hat{L})(\hat{G}+j \omega \hat{C})}= \pm \sqrt{(\hat{R}+j \omega \hat{L})\left(\frac{\hat{R} \hat{C}}{\hat{L}}+j \omega \hat{C}\right)}= \pm \sqrt{\frac{\hat{\mathrm{C}}}{\hat{\mathrm{~L}}}}(\hat{R}+\mathrm{j} \omega \hat{L})
$$

In this case, the attenuation constant $\alpha$ and the phase velocity $\frac{\omega}{\beta}$ are independent of frequency. This implies that there will be no distortion of a signal as it propagates on this transmission line. There will only be a constant attenuation of the signal. The characteristic impedance of this transmission line

$$
Z_{C}=\sqrt{\frac{\hat{Z}}{\hat{Y}}}=\sqrt{\frac{\hat{R}+j \omega \hat{L}}{\hat{G}+j \omega \hat{C}}}=\sqrt{\frac{\hat{R}+j \omega \hat{L}}{\frac{\hat{R} \hat{C}}{\hat{L}}+j \omega \hat{C}}}=\sqrt{\frac{\hat{L}}{\hat{C}}}
$$

is also independent of frequency. This transmission line is called a "distortionless line."

Example 6-19. The attenuation on a 50W distortionless transmission line is 0.01 $(\mathrm{dB} / \mathrm{m})$. The line has a capacitance of $0.1 \times 10^{-9}(\mathrm{~F} / \mathrm{m})$.
a) Find the values of the other transmission line elements.
b) Find the velocity of wave propagation.

## Lossy transmission lines

Answer: a) Since this is a distortionless transmission line, we write

$$
\begin{aligned}
Z_{C} & =\sqrt{\frac{\hat{L}}{\hat{C}}}=50 \Omega \Rightarrow \hat{L}=(50)^{2}\left(0.1 \times 10^{-9}\right)=2.5 \times 10^{-7} \mathrm{H} / \mathrm{m} \\
\alpha & =\hat{R} \sqrt{\frac{\hat{\mathrm{C}}}{\hat{L}}}=0.01 \mathrm{~dB} / \mathrm{m}=\frac{0.01}{8.69} \mathrm{~Np} / \mathrm{m}=0.0012 \mathrm{~Np} / \mathrm{m}
\end{aligned}
$$

From this, we write

$$
\hat{R} \sqrt{\frac{\hat{\mathrm{C}}}{\hat{\mathrm{~L}}}}=\frac{\hat{\mathrm{R}}}{50}=0.0012 \Rightarrow \hat{\mathrm{R}}=0.0575 \Omega / \mathrm{m}
$$

We also have the distortionless line criteria

$$
\frac{\hat{R}}{\hat{\mathrm{~L}}}=\frac{\hat{\mathrm{G}}}{\hat{\mathrm{C}}} \Rightarrow \hat{\mathrm{G}}=\frac{0.0575}{50^{2}}=2.3 \times 10^{-5} \mathrm{~S} / \mathrm{m}
$$

b) The phase velocity is

$$
v=\frac{1}{\sqrt{\hat{L} \hat{C}}}=\frac{1}{\sqrt{\left(2.5 \times 10^{-7}\right)\left(0.1 \times 10^{-9}\right)}}=2 \times 10^{8} \mathrm{~m} / \mathrm{s}
$$

## Radiation fundamentals

## 7. Radiation of electromagnetic waves

In the previous chapters, we learned that electromagnetic waves can propagate in infinite free space and that these same waves can also propagate along a common transmission line or in a waveguide that can be held in our hands. A question that remains to be answered is whether the same electromagnetic wave can be excited in a finite region and then launched or radiated into infinite space. In this chapter, we will first examine the fundamentals of the radiation of electromagnetic waves. We will base our argument on material that has been uncovered in earlier chapters. This will naturally lead to an introduction into the important topic of antennas. Several of the important parameters and terms associated with antennas will be brought forth in this discussion.

### 7.1. Radiation fundamentals

Before we examine the radiation properties of an antenna, we should first understand a physical process that can actually cause the radiation of electromagnetic waves. This means that we have to examine possible radiation characteristics of an electric charge from a fundamental viewpoint. There are certain requirements that an electric charge must meet in order to consider that it will actually radiate electromagnetic waves. These requirements will be argued from an intuitive point of view. If we understand this argument, the development of antenna radiation theory follows immediately since the principle of superposition applies in the linear medium that is being considered in this text and the antenna can be considered to consist of a large number of charges. The argument also illustrates the type of calculation that can be written on the backs of old envelopes.

## Radiation fundamentals

We can understand radiation of electromagnetic waves using Poynting's theorem. This theorem states that the total power $P_{r}$ that is radiated from a source is given by the following closed surface integral

$$
\begin{equation*}
P_{r}=\oint_{s} E \times H \cdot d s \tag{7.1}
\end{equation*}
$$

Poynting's theorem tells us that the radiation of electromagnetic waves from a source that is located within a volume that is completely enclosed by a closed surface requires both an electric field and a magnetic field, the two fields being coupled together via Maxwell's equations.

A stationary charge that was discussed in Chapter 2 will not radiate electromagnetic waves. This can be easily understood since a stationary charge will cause no current to flow, hence there will be no magnetic field associated with a stationary charge. From (7.1), the total radiated power is therefore equal to zero. From this, we can conclude that there will be no radiation of electromagnetic waves from a stationary charge.

We can also come to this conclusion from another point of view. If the point of observation where the power is to be detected is far from the source and if there were a spherically radiating wave, it would appear to be almost a plane wave at large distances from the charge and we can make use of the fact that the electric and magnetic field intensities of propagating waves are related through the characteristic impedance of free space $Z_{\mathrm{c}}$ as described in Chapter 5. The magnitude of the magnetic field intensity H can be found from the electric field intensity E via $\mathrm{H}=\mathrm{E} / \mathrm{Z}_{0}$.

## Radiation fundamentals



Figure 7-1. Antenna radiation of electromagnetic waves. For a stationary charge, H will be equal to zero.

Therefore, a source of electromagnetic power located at the center of a sphere whose radius is $R$ shown in Figure $7-1$ would radiate a total radiated power whose value can be written as

$$
\begin{equation*}
\mathrm{P}_{\mathrm{r}} \approx \frac{\mathrm{E}^{2}}{\mathrm{Z}_{0}}\left(4 \pi \mathrm{R}^{2}\right) \tag{7.2}
\end{equation*}
$$

Let us assume at this stage that the antenna is an isotropic radiator and has no directional characteristics. The total radiated power is equal to that which is delivered from the source that we will assume to be a constant. Hence, the total radiated power is independent of the distance $R$. Therefore, we would conclude that the electric field $E$ of an electromagnetic wave must decrease with increasing distance as $R^{-1}$. However, we find that the electric field from a static charge varies as $R^{-2}$. Hence, we again come to the same conclusion that stationary charges cannot radiate electromagnetic waves.

We can also argue that a stationary charge will not radiate from an examination of Figure 7-1. The electric field associated with radiation is in the surface of the sphere. The electric field associated with a stationary charge is entirely in the radial direction.

## Radiation fundamentals

The next question that should be posed is whether a charge that is in motion with a constant velocity v << c can radiate electromagnetic waves. We know that a charge in motion constitutes a current and currents cause magnetic fields. We are not able to invoke the previous argument based on the radiated power that we used for the lack of radiation from a static charge since both an electric field and a magnetic field will now be present. We will, however, use a slightly different argument that is still based on the Poynting vector.

Let us assume that a positive charge $Q$ is moving in the positive $u_{\mathbf{x}}$ direction with a constant velocity $\mathbf{v}$ as shown in Figure 7-2. This velocity shall be chosen so that it is much less than the velocity of light c so it is nonrelativistic. We do not want to wade into the deep waters of relativity or advanced topics in physics at this time.


Figure 7-2. Electric and magnetic fields due to a moving charge. The velocity $\mathbf{v}$ is a constant and $\mathrm{v} \ll \mathrm{c}$.

The static electric field $\mathbf{E}$ from the charge $Q$ is computed to be

$$
\begin{equation*}
\mathbf{E}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0}} \frac{1}{r^{2}+\mathrm{x}^{2}} \mathbf{u}_{\mathrm{r}} \tag{7.3}
\end{equation*}
$$

where $R=\sqrt{r^{2}+x^{2}}$. The magnetic field is computed from the Biot-Savart law. This leads to

## Radiation fundamentals

$$
\begin{equation*}
\mathbf{B}=\frac{\mu_{0}}{4 \pi} \frac{\mathrm{Q}\left(\mathbf{v} \times \mathbf{u}_{\mathrm{r}}\right)}{\mathrm{r}^{2}+\mathrm{x}^{2}} \tag{7.4}
\end{equation*}
$$

Let us compute the direction of Poynting's vector associated with these two fields. This is facilitated from an examination of a sphere centered on the charge at a certain instant in time as shown in Figure 7-3. The electric field caused by a charge moving with a uniform velocity is entirely normal to the spherical surface and the magnetic field is tangent to the surface. Hence, the Poynting vector $\mathbf{S}=\mathbf{E} \times \mathbf{H}$ is completely confined within the spherical surface and it does not radiate in the radial direction away from the charge. Is there any hope for radiation?


Figure 7-3. The Poynting vector associated with a charge moving with a constant velocity $\mathbf{v}$.

Example 7-1. Calculate the component of the Poynting vector in the $\mathbf{u}_{\mathbf{x}}$ direction in Figure 7-2 and the total energy flow rate through an infinitely large plane placed normal to the $x$ axis. Discuss the meaning of this result.

Answer: The magnitude of the x component of the Poynting vector is computed from $\left|\left(E_{r} \times B\right) / \mu_{o}\right|$. This leads to

$$
S_{x}=\left\{\frac{Q R}{4 \pi \varepsilon_{0}\left(r^{2}+x^{2}\right)^{3 / 2}}\right\}\left\{\frac{Q v R}{4 \pi\left(r^{2}+x^{2}\right)^{3 / 2}}\right\}=\frac{Q^{2} v R^{2}}{16 \pi^{2} \varepsilon_{0}\left(r^{2}+x^{2}\right)^{3}}
$$

The total energy flow rate becomes

$$
P_{r}=\int_{0}^{\infty} S_{x} 2 \pi r d r=\frac{Q^{2} v}{8 \pi \varepsilon_{0}} \int_{0}^{\infty} \frac{r^{3} d r}{\left(r^{2}+x^{2}\right)^{3}}
$$

## Radiation fundamentals

The integral can be performed in Problem 1 to yield

$$
P_{r}=\frac{Q^{2} v}{32 \pi \varepsilon_{0}} \frac{1}{x^{2}}(W)
$$

The distance $|\mathrm{x}|$ is the instantaneous separation between the charge and the plane. In the one-dimensional system being considered here ( $\mathbf{v} \rightarrow \mathbf{v} \mathbf{u}_{\mathbf{x}}$ ), the velocity v can be written as $\mathrm{v}=\mathrm{dx} / \mathrm{dt}$. The power can be rewritten in the form

$$
\mathrm{P}_{\mathrm{r}}=-\frac{\mathrm{d}}{\mathrm{dt}}\left[\frac{\mathrm{Q}^{2}}{32 \pi \varepsilon_{0} \mathrm{x}}\right]
$$

The quantity

$$
\mathrm{W}=\frac{\mathrm{Q}^{2}}{32 \pi \varepsilon_{0} \mathrm{x}_{0}}
$$

is the electrostatic energy stored in the region $x>x_{0}$ as in Problem 2. Therefore the power that is calculated using Poynting's theorem can be interpreted as the flow rate of electrostatic energy stored in space and has nothing to do with radiation. The magnetic energy will be of the order of $(\mathrm{v} / \mathrm{c})^{2}$ times the electric energy and will be very small in nonrelativistic cases.

In order to answer the question whether there can be any radiation at all, let us consider a charge initially at rest at point $A$, which is accelerated in the $x$ direction as shown in Figure 7-4. The acceleration lasts for a duration $\Delta t$ seconds until it reaches a point B after which the charge moves with a constant velocity v $\ll$ the velocity of light c to a point C and beyond. Remember that a signal cannot propagate faster than c .

We know that stationary charges and charges moving with a constant velocity do not radiate electromagnetic waves and have an electric field that is only radially outward. Thus the electric field lines when the charge is at A and at the point $C$ are entirely radial. These electric field lines must be continuous since they are caused by the same charge. They are connected with "kinked" lines.

## Radiation fundamentals

The kinks, that are disturbances in the electric field lines caused by the acceleration of the charge, propagate with the speed of light.


Figure 7-4. A charge that is accelerated does radiate electromagnetic waves. The dark lines are electric field lines $\mathbf{E}$.

It takes $\Delta t$ seconds for the charge to move from the point $A$ to the point $B$, therefore the separation between the two circles is approximately $c \Delta t$. In the kinks, there are components of electric field that are perpendicular to the Coulomb field. These transverse components are responsible for the radiation. Note that in this argument, there are directions where there are no radiated electric fields and only the static Coulomb field exists. The maximum radiated electric field will occur along the line that is perpendicular to the charge's acceleration.

Consider a point $L$ in Figure $7-5$ that is normal to the direction of the charge's velocity at a certain instant. Let $t$ be the time after the charge is accelerated from a stationary point $A$ to point $B$ where it has a velocity $v=a \Delta t$ where $a$ is the acceleration. We will assume that $\Delta t \ll t$ so the distance $A B+B C \approx B C=$ vt.

## Radiation fundamentals

At point $L$, there will be two components of an electric field. The first is the radial Coulomb field that is given by


Figure 7-5. The components of the electric field caused by a charge $Q$ that is accelerated during a time $\Delta t$ from points $A$ to $B$.

$$
\begin{equation*}
\mathrm{E}_{0}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0}} \frac{1}{\mathrm{R}^{2}}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0}} \frac{1}{(\mathrm{ct})^{2}} \tag{7.5}
\end{equation*}
$$

The radiation field $\mathrm{E}_{\mathrm{t}}$ can be computed from the triangle JKL

$$
\begin{equation*}
\frac{J K}{K L}=\frac{J K}{A B+B C} \approx \frac{J K}{B C}=\frac{c \Delta t}{v t}=\frac{E_{0}}{E_{t}} \tag{7.6}
\end{equation*}
$$

Solving (7.6) for $E_{t}$, we obtain

$$
\begin{equation*}
\mathrm{E}_{\mathrm{t}}=\frac{\mathrm{vt}}{\mathrm{c} \Delta \mathrm{t}} \mathrm{E}_{0}=\frac{\mathrm{vt}}{\mathrm{c} \Delta \mathrm{t}} \frac{\mathrm{Q}}{4 \pi \varepsilon_{0}} \frac{1}{(\mathrm{ct})^{2}}=\frac{\mathrm{Q}}{4 \pi \varepsilon_{0} \mathrm{c}^{2}} \frac{\mathrm{v}}{\Delta \mathrm{t}} \frac{1}{\rho} \tag{7.7}
\end{equation*}
$$

Eureka! This is what we were looking for! There is a transverse component of the electric field that is proportional to the acceleration $v / \Delta t$ and it has the proper spatial variation $1 / \mathrm{R}$ that is required in the Poynting vector (7.2). The minus sign that appears here is due to the direction of $E_{t}$, that is opposite to the direction of the acceleration. From Figure 7-5, we note that there is a

## Radiation fundamentals

preferred direction for this radiation. If we define the angle $\theta$ as being the angle between the point of observation and the velocity of the accelerated charge, (7.7) can be written with $\mu_{0}=1 /\left(\varepsilon_{0} c^{2}\right)$ introduced as

$$
\begin{equation*}
E_{t}=\frac{Q \mu_{0}}{4 \pi} \frac{[a] \sin \theta}{\rho} \tag{7.8}
\end{equation*}
$$

where $[a]=a\left(t^{\prime}=t-\rho / c\right)$ is the acceleration at an earlier time or as it is frequently called, a retarded acceleration of the charge taken in a previous moment t'. As it can be seen from Fig. 7-5 the signal needs time $\tau=\rho / \mathrm{c}$ to travel between points $C$ and $L$. This is the "propagation delay" due to the finite speed of light that we encountered in our earlier discussion of plane waves and transmission lines.

The magnetic field intensity $H_{t}$ associated with $E_{t}$ can be computed by just using the characteristic impedance of the vacuum $Z_{0}$ and the fact that the electric and magnetic field intensities are related by this characteristic impedance. We write including $\mu_{0} / Z_{0}=1 / \mathrm{c}$

$$
\begin{equation*}
H_{t}=\frac{Q}{4 \pi c} \frac{[a] \sin \theta}{\rho} \tag{7.9}
\end{equation*}
$$

The Poynting vector is directed radially outward and it's magnitude $S$ is given by

$$
\begin{equation*}
\mathrm{S}=\frac{\mathrm{Q}^{2}[\mathrm{a}]^{2} \mu_{0}}{16 \pi^{2} \mathrm{c}} \frac{\sin ^{2} \theta}{\rho^{2}} \tag{7.10}
\end{equation*}
$$

If we insert numerical values for the coefficient in (7.10), we find that a very small number will be obtained. Therefore, the requirement that must be satisfied for electromagnetic waves to be radiated from a source is that there be charged particles that are either accelerated or decelerated. A current with a time harmonic variation certainly satisfies this requirement and this will be the source that will be used for the antennas. Once this current is specified, the electromagnetic fields can be computed. We, of course, must keep in mind that the electric field at a distance $\rho$ from the current element will be retarded in time from this oscillating current. The retardation is given by Einstein's requirement

## Radiation fundamentals

that things can go no faster than the velocity of light. Any effect will appear at a time $\rho / \mathrm{c}$ after the cause.

Now that a basic physical mechanism for radiation has been obtained, we are now able to describe the practical radiation characteristics of antennas. Several antenna structures will be examined in this chapter and important antenna parameters will be defined.

Example 7-2. Assume that an antenna could be described as being an ensemble of $N$ oscillating electrons with a frequency $\omega$ in a plane that is orthogonal to the distance $\rho$. Find an expression for the electric field $E_{\perp}$ that would be detected at that location.

Answer: The maximum electric field is computed from (7.8) with $\theta=90^{\circ}$. We obtain

$$
E_{\perp}=\frac{N Q \mu_{0}}{4 \pi} \frac{1}{\rho}\left[\frac{d v}{d t}\right]=\frac{\mu_{0}}{4 \pi} \frac{1}{\rho}\left[\frac{d J}{d t}\right]
$$

where the electric current density $\mathrm{J}=\mathrm{NQv}$ is introduced. This equation shows that radiation occurs when there is an oscillating current because the derivative $\mathrm{dJ} / \mathrm{dt}$ must be different from zero. If we assume that the direction of oscillation in the orthogonal plane is $x$, then $x(t)=x_{m} \sin \omega t$ and $v=d x / d t=\omega x_{m} \cos \omega t$. The substitution of these terms in the equation for the current density $J(t)=\omega N Q x_{m} \cos \omega t$. Finally, the expression for the transverse electric field becomes

$$
\mathrm{E}_{\perp}(\rho, \mathrm{t})=\omega^{2} \frac{\mathrm{NQx} \mathrm{x}_{\mathrm{m}} \mu_{0}}{4 \pi} \frac{1}{\rho} \sin \omega \mathrm{t}^{\prime} .
$$

## Radiation fundamentals

This expression shows that the electric field is proportional to the square of the frequency which implies that radiation of electromagnetic waves is essentially a high-frequency phenomena rather than being a static occurrence.

## Short electric dipole antenna

### 7.2. Short electric dipole antenna

In order to apply the basic principles presented in the previous section, we will examine a very elementary antenna - the short electric dipole that is shown in Figure 7-6. The length of the antenna $L$ is assumed to be short in comparison with the wavelength of the wave $\lambda(L \ll \lambda)$. The small plates that are placed at the ends of the dipole provide capacitive loading. The short length and the presence of these plates result in a uniform current I along the dipole length. The dipole may be excited by a balanced transmission line such as a coaxial line. The radiation from the transmission line connection and the end plates of the dipole are considered to be negligible. The diameter d of the dipole is also small in comparison with its length $(\mathrm{d} \ll \mathrm{L})$. The current on this antenna follows directly from the equation of continuity

$$
\begin{equation*}
\mathrm{I}=\frac{\mathrm{dQ}}{\mathrm{dt}} \tag{7.11}
\end{equation*}
$$

For a time-harmonic excitation with a frequency $\omega$, (7.11) reduces to $\mathrm{I}=\mathrm{j} \omega \mathrm{Q}$.
The vector potential $\mathbf{A}$ can be calculated from the current density $\mathbf{J}$ in the wire using a three-dimensional generalization of the wave equation that was derived previously (5.10) in the case of the presence of a source

$$
\begin{equation*}
\nabla^{2} \mathbf{A}-\frac{1}{\mathrm{c}^{2}} \frac{\partial^{2} \mathbf{A}}{\partial \mathrm{t}^{2}}=-\mu_{0} \mathbf{J} \tag{7.12}
\end{equation*}
$$

## Short electric dipole antenna



Figure 7-6. An oscillating short electric dipole radiates electromagnetic waves.

Example 7-3. Show the vector-potential A satisfies a three-dimensional inhomogeneous wave equation (7.12).

Answer: Recall Maxwell's equations from Chapter 4 for a vacuum

$$
\begin{aligned}
& \nabla \times \mathbf{E}=-\frac{\partial \mathbf{B}}{\partial \mathrm{t}} \\
& \nabla \times \mathbf{B}=\frac{1}{\mathrm{c}^{2}} \frac{\partial \mathbf{E}}{\partial \mathrm{t}}+\mu_{0} \mathbf{J} \\
& \nabla \bullet \mathbf{E}=\frac{\rho_{\mathrm{V}}}{\varepsilon_{0}} \\
& \nabla \bullet \mathbf{B}=0
\end{aligned}
$$

The magnetic flux density is found from the vector potential using the definition $\mathbf{B}=\nabla \times \mathbf{A}$. The substitution of this relation into the second equation yields

$$
\nabla \times \nabla \times \mathbf{A}=\frac{1}{\mathrm{c}^{2}} \frac{\partial \mathbf{E}}{\partial \mathrm{t}}+\mu_{0} \mathbf{J}
$$

Using the vector identity for the repeated vector operation given in Appendix 1 $\nabla \times \nabla \times \mathbf{A}=\nabla \nabla \bullet \mathbf{A}-\nabla^{2} \mathbf{A}$, this equation becomes

# Short electric dipole antenna 

$$
\nabla \nabla \cdot \mathbf{A}-\nabla^{2} \mathbf{A}=\frac{1}{\mathrm{c}^{2}} \frac{\partial \mathbf{E}}{\partial \mathrm{t}}+\mu_{0} \mathbf{J}
$$

Substituting

$$
\mathbf{B}=\nabla \times \mathbf{A}
$$

into the first Maxwell equation yields

$$
\nabla \times\left(\mathbf{E}+\frac{\partial \mathbf{A}}{\partial \mathrm{t}}\right)=0 \text { or } \mathbf{E}+\frac{\partial \mathbf{A}}{\partial \mathrm{t}}=-\nabla \mathrm{V}
$$

This leads to

$$
\mathbf{E}=-\nabla \mathrm{V}-\frac{\partial \mathbf{A}}{\partial \mathrm{t}}
$$

where V is the scalar electric potential discussed in Chapter 2. The substitution of this expression for $E$ into the equation for the vector potential yields the following result

$$
\nabla \nabla \cdot \mathbf{A}-\nabla^{2} \mathbf{A}=\nabla\left(-\frac{1}{c^{2}} \frac{\partial \mathbf{V}}{\partial \mathrm{t}}\right)-\frac{1}{\mathrm{c}^{2}} \frac{\partial^{2} \mathbf{A}}{\partial \mathbf{t}^{2}}+\mu_{0} \mathbf{J}
$$

This equation can be simplified with the inclusion of one more constraint that relates the vector potential and the scalar electric potential. This is called the Lorentz gauge

$$
\nabla \cdot \mathbf{A}+\frac{1}{c^{2}} \frac{\partial \mathrm{~V}}{\partial \mathrm{t}}=0
$$

The application of this constraint leads to the wave equation for the vector potential

$$
\nabla^{2} \mathbf{A}-\frac{1}{c^{2}} \frac{\partial^{2} \mathbf{A}}{\partial t^{2}}=-\mu_{0} \mathbf{J} .
$$

## Short electric dipole antenna

A similar wave equation can also be derived for the scalar potential - this follows directly from the third Maxwell equation and an application of the Lorentz gauge. We would obtain

$$
\nabla^{2} \mathrm{~V}-\frac{1}{\mathrm{c}^{2}} \frac{\partial^{2} \mathrm{~V}}{\partial \mathrm{t}^{2}}=-\frac{\rho_{\mathrm{V}}}{\varepsilon_{0}}
$$

In the case of a static field, the second term on the left-hand side of (7.12) disappears and the equation reduces to Poisson's equation. It can be shown that (7.12) has a solution that is similar to the result that we obtained there. The only difference is that we must be cognizant of the finite velocity of light c. This can be included by realizing that there are two distinct times that must be incorporated into the solution. The current density $\mathbf{J}\left(\mathbf{r}^{\prime}\right)$ in the right-hand side of (2.131) must be replaced with the retarded current density $[\mathbf{J}] \equiv \mathbf{J}\left(\mathbf{r}^{\prime}, \mathrm{t}-\mathrm{R} / \mathrm{c}\right)$ in order to find the vector potential $\mathbf{A}(\mathbf{r}, \mathrm{t})$

$$
\begin{equation*}
\mathbf{A}(\mathbf{r}, \mathrm{t})=\frac{\mu_{0}}{4 \pi} \int_{\Delta v} \frac{\mathrm{~J}\left(\mathbf{r}^{\prime}, \mathrm{t}-\mathrm{R} / \mathrm{c}\right)}{\mathrm{R}} \mathrm{dv}^{\prime} \tag{7.13}
\end{equation*}
$$

We leave it to the reader as an exercise to check that equation (7.13) is actually a solution of the inhomogeneous wave equation (7.12). This equation takes into account the finite velocity of propagation for the electromagnetic waves c. The electric and magnetic fields radiated from the antenna can be found in terms of this vector potential $\mathbf{A}(\mathbf{r}, \mathrm{t})$ as shown in Example 7-3.

We will now be able to obtain the general solution for the case of a short electric dipole antenna. A time-harmonic current density can be written as

$$
\begin{equation*}
J\left(r^{\prime}, t-R / c\right)=J\left(r^{\prime}\right) e^{-j k \rho} \tag{7.14}
\end{equation*}
$$

## Short electric dipole antenna

In writing this equation, we have defined the distance in the spherical coordinates. The distance from the center of the dipole $R=\rho$ and $k=\omega / c$ is the wave number. The volume of the dipole antenna can be approximated as $d v^{\prime}=L d s^{\prime}$. The current becomes $J d s^{\prime}=l \mathbf{u}_{\mathbf{z}}$. Therefore, the vector potential can be calculated directly from (7.13).

$$
\begin{equation*}
\mathbf{A}=\mathbf{u}_{z} \frac{\mu_{0}(\mathrm{IL})}{4 \pi}\left(\frac{\mathrm{e}^{-\mathrm{jk} \mathrm{\rho}}}{\rho}\right) \tag{7.15}
\end{equation*}
$$

where (IL) is the current element of the radiating dipole. This short dipole antenna is known as a Hertzian dipole.

Let us now find the field components at large distances from the antenna. In particular, we will make the assumption that these distances are much greater than the wavelength of the wave ( $\rho \gg \lambda$ ). This is called the far field of the antenna. In order to do this, we have to express the unit vector $\mathbf{u}_{\mathbf{z}}$ in spherical coordinates as

$$
\begin{equation*}
\mathbf{u}_{\mathbf{z}}=\mathbf{u}_{\boldsymbol{\rho}} \cos \theta-\mathbf{u}_{\boldsymbol{\theta}} \sin \theta \tag{7.16}
\end{equation*}
$$

The components of the vector potential $\mathbf{A}=\mathrm{A}_{\mathrm{z}} \mathbf{u}_{\mathbf{z}}$ in spherical coordinates are given by

$$
\begin{align*}
& A_{\rho}=A_{z} \cos \theta=\frac{\mu_{0}(I L)}{4 \pi}\left(\frac{e^{-j k \rho}}{\rho}\right) \cos \theta \\
& A_{\theta}=-A_{z} \sin \theta=\frac{\mu_{0}(I L)}{4 \pi}\left(\frac{e^{-j k \rho}}{\rho}\right) \sin \theta  \tag{7.17}\\
& A_{\varphi}=0
\end{align*}
$$

The magnetic field intensity is computed from the vector potential using the definition of the curl operation in spherical coordinates. We find that

## Short electric dipole antenna

$$
\mathbf{H}=\frac{1}{\mu_{0}} \nabla \times \mathbf{A}=\mathbf{u}_{\varphi} \frac{1}{\mu_{0} \rho}\left[\frac{\partial\left(\rho \mathrm{~A}_{\theta}\right)}{\partial \rho}-\frac{\partial \mathrm{A}_{\rho}}{\partial \theta}\right]=-\mathbf{u}_{\varphi} \frac{(\mathrm{IL})}{4 \pi} \mathrm{k}^{2} \sin \theta\left[\frac{1}{\mathrm{jk} \rho}+\frac{1}{(\mathrm{jk} \rho)^{2}}\right] \mathrm{e}^{-\mathrm{jk} \mathrm{\rho} \rho}
$$

or

$$
\begin{equation*}
\mathrm{H}_{\varphi} \approx \frac{j k(I L)}{4 \pi} \frac{\mathrm{e}^{-\mathrm{jk} \mathrm{\rho}}}{\rho} \sin \theta \tag{7.18}
\end{equation*}
$$

where only the term in the far field or as it is called, the radiation field, is retained since $\mathrm{k} \rho=\frac{2 \pi}{\lambda} \rho \gg 1$. The electric field is computed from Maxwell's equations that are applicable in a vacuum region in which the current density $\mathbf{J}=0$. We write

$$
\begin{equation*}
\mathbf{E}=\frac{1}{\mathrm{j} \omega \varepsilon_{0}} \nabla \times \mathbf{H}=\frac{1}{\mathrm{j} \omega \varepsilon_{0}}\left[\mathbf{u}_{\rho} \frac{1}{\rho \sin \theta} \frac{\partial\left(\sin \theta \mathrm{H}_{\varphi}\right)}{\partial \theta}-\mathbf{u}_{\theta} \frac{\partial\left(\rho \mathrm{H}_{\varphi}\right)}{\partial \rho}\right] \tag{7.19}
\end{equation*}
$$

The components of the electric field are calculated from (7.19). In the far field region, we find them to be

$$
\begin{align*}
& \mathrm{E}_{\rho} \approx 0  \tag{7.20}\\
& \mathrm{E}_{\theta} \approx \frac{j Z_{0} \mathrm{k}(\mathrm{IL})}{4 \pi} \frac{\mathrm{e}^{-\mathrm{jk} \rho}}{\rho} \sin \theta \\
& \mathrm{E}_{\varphi}=0
\end{align*}
$$

where $Z_{0}=E_{\theta} / H_{\varphi}=\sqrt{\mu_{0} / \varepsilon_{0}} \approx 377$ ohms is the wave impedance of vacuum as noted in Example 5-11. Equation (7.20) agrees with the result found in Example 7-2 that was obtained using a different point of view. This result could also be obtained by replacing the term $\frac{\mathrm{dJ}}{\mathrm{dt}} \rightarrow \mathrm{j} \omega \mathrm{J}$, including the retarded time (7.14), and making the substitutions $\omega \mu_{0} \rightarrow \mathrm{k} Z_{0}$ and $\mathrm{J} \Delta \mathrm{v} \rightarrow \mathrm{IL}$ (where $\Delta \mathrm{v}$ is the antenna volume). This is an entirely different approach to obtain the same result.

## Short electric dipole antenna

Let us consider now the angular distribution of the radiated fields. Both $E_{\theta}$ and $H_{\varphi}$ are proportional to $\sin \theta$. This implies that there will be an angular variation for both fields and they both have a maximum value when $\theta=90^{\circ}$ which is in the direction that is perpendicular to the axis of the dipole and they both have a minimum value when $\theta=0^{\circ}$ which is along the axis of the dipole. The angular variation of the electric field $E_{\theta}$ is called the radiation pattern of the antenna. The radiation pattern on a short dipole antenna is shown in Figure 7-7.


Figure $7-7^{\text {MATLAB }}$. Radiation pattern for a short dipole antenna.

## Short electric dipole antenna

Example 7-4. A small antenna that is 1 cm in length and 1 mm in diameter is designed to transmit a signal at 1 GHz inside the human body in a medical experiment. Assuming the dielectric constant of the body is similar to that of distilled water $\left(\varepsilon_{r}=80\right)$ and that the conductivity $\sigma$ can be neglected, compute the maximum electric field at the surface of the body that is approximately 20 cm from the antenna. The maximum current that can be applied to the antenna is 10 $\mu \mathrm{A}$. Find the new distance from the antenna $\rho_{1}$, where the signal will be attenuated by 3dB.

Answer: The wavelength of the electromagnetic wave within the body is computed to be

$$
\lambda=\frac{\mathrm{c}}{\mathrm{f} \sqrt{\varepsilon_{\mathrm{r}}}}=\frac{3 \times 10^{8}}{10^{9} \sqrt{80}} \approx 3.3 \mathrm{~cm}
$$

The characteristic impedance of the body is

$$
\mathrm{Z}_{\mathrm{C}}=\sqrt{\frac{\mu_{0}}{\varepsilon_{0} \varepsilon_{\mathrm{r}}}}=\frac{377}{\sqrt{80}} \approx 42 \Omega
$$

Since the dimensions of the antenna are much less than the wavelength, we can apply (7.20) for $\theta=90^{\circ}$ and replace $Z_{0} \rightarrow Z_{C}$. Therefore

$$
\left|\mathrm{E}_{\theta}\right|=\frac{\mathrm{IL}}{4 \pi} Z_{\mathrm{c}} \mathrm{k} \frac{1}{\rho}=\frac{10^{-5} \times 10^{-2}}{4 \pi} \times 42 \times \frac{2 \pi}{0.033} \times \frac{1}{0.2} \approx 320 \mu \mathrm{~V} / \mathrm{m}
$$

The attenuation of 3 dB implies that the power changes by a factor of 2 an electric field changes by a factor of 1.41. The distance is found to be

$$
\rho_{1}=\sqrt{2} \rho=1.41 \times 0.2 \approx 0.28 \mathrm{~m}=28 \mathrm{~cm}
$$

The far field region is the most interesting for antenna applications. In this region, both of the electromagnetic field components are transverse to the direction of propagation. This is clearly seen from equations (7.19) and (7.20). In order to calculate the radiated power from this antenna, one just has to

## Short electric dipole antenna

perform the surface integration of the time-average Poynting vector as was shown in Example 5-13.

$$
\begin{equation*}
P_{r}=\frac{1}{2} \operatorname{Re} \oint_{s}\left(\mathbf{E} \times \mathbf{H}^{*}\right)_{\mathrm{av}} \bullet d \mathbf{d}=\frac{1}{2} Z_{0} \int_{\phi=0}^{2 \pi} \int_{\theta=0}^{\pi}\left|\mathrm{H}_{\varphi}\right|_{\mathrm{av}}{ }^{2} \rho^{2} \sin \theta \mathrm{~d} \theta \mathrm{~d} \phi \tag{7.21}
\end{equation*}
$$

After a substitution from (7.19) into (7.21) and performing the integration over the variable $\phi$ which yields to $2 \pi$ the expression (7.21) becomes

$$
P_{r}=\frac{Z_{0} k^{2}\left(I_{\mathrm{av}} L\right)^{2}}{16 \pi} \int_{0}^{\pi} \sin ^{3} \theta d \theta=-\frac{Z_{0} \mathrm{k}^{2}\left(I_{\mathrm{av}} L\right)^{2}}{16 \pi} \int_{0}^{\pi}\left(1-\cos ^{2} \theta\right) \mathrm{d} \cos \theta
$$

The last integral gives a factor of $4 / 3$. The radiated power from this antenna is

$$
\begin{equation*}
P_{r}=\frac{Z_{0}(k L)^{2} I_{a v}^{2}}{12 \pi} \tag{7.22}
\end{equation*}
$$

Here the constant current I can be replaced with its average value $\mathrm{I}_{\mathrm{av}}$ assuming that there is a slow variation in space.

## Long dipole antenna

### 7.3. Long dipole antenna

We have now covered the basic idea of the radiation of electromagnetic waves from a short electric dipole antenna. As we take a short trip away from this book and into the hinterlands, we may see some tall structures that reach into the heavens and have flashing red lights at the top to warn passing airplanes. These antennas certainly do not seem to fall into the class of being "small". Herein we will describe a technique to generalize our treatment of antennas so that more realistic antennas can be studied.

Let us consider two thin metallic rods having a total length $L$. The length of the long dipole (linear antenna) may be of the order of the free space wavelength $\lambda$ of the electromagnetic wave that is to be radiated. A sinusoidal voltage generator whose frequency of oscillation is $\omega$ is connected between the two rods as shown in Figure 7-8. This voltage generator will induce a current in the rods that can have a distribution $I(z)$ that is governed by the shape and length of the conductor.


Figure 7-8. A center fed dipole with an arbitrary current distribution $\mathrm{I}(\mathrm{z})$.

## Long dipole antenna

It is reasonable to assume that the current distribution at the ends of the antenna ( $z= \pm \mathrm{L} / 2$ ) is equal to zero and that the current distribution is symmetrical about the center $(z=0)$. The first assumption is predicated on the idea that no conduction current could extend beyond the metallic surface. Since the antenna is "center-fed," symmetry arguments can be applied. The assumption for the approximated current distribution that is used in the integral requires some ingenuity since it is not a quantity that is measured in the laboratory. A typical requirement is that the current distribution is selected so certain integrals can actually be performed. Computers have now alleviated this restriction and more realistic distributions can be employed. However, finding the actual current distribution $\mathrm{I}(\mathrm{z})$ is a difficult task - an integral equation must be solved. The last one includes the boundary conditions and the investigator can iterate the solution starting from a very simple function that satisfies these conditions. The iteration procedure is to initially assume a function for the current distribution, calculate the electric field resulting from this distribution, measure the electric field in the laboratory, and finally modify the initial choice for the current distribution and start the procedure over again. For example, starting from a triangular one shown in Figure 7-9a and continuing on into a more realistic solution that is shown in Figure 7-9b. This problem is a subject for an advanced course in electromagnetics and we will skip it here.

(a)

(b)

Figure 7-9. Two possible distributions of current on an antenna. (a) Initial triangular distribution. (b) Actual distribution ( $L=\lambda / 2$ ).

One simple and quite reasonable approximation for the actual current distribution is the following sinusoidal function

$$
\begin{equation*}
I(z)=I_{m} \sin \left[k\left(\frac{L}{2}-|z|\right)\right] \tag{7.23}
\end{equation*}
$$

The far field radiation properties such as the radiation power, directivity etc. are not very sensitive to the actual choice for the current distribution. However, the near field properties such as the input impedance etc. are sensitive to the actual shape. These terms will be defined in the next section.
Example 7-5. Describe the excitation of a center fed dipole antenna using a transmission line model.


## Long dipole antenna

Answer: The current distribution of both the incident and the reflected components of the current on an open circuited transmission line as discussed in the previous chapter are depicted in the figure. Its spatial distribution is sinusoidal as shown in (a). By bending the transmission line at $\lambda / 4$ from the end, we form a half-wave dipole ( $L=\lambda / 2$ ) with the proper current distribution. This model has assumed that the terminating $\lambda / 4$ of the transmission line is unaffected by the bending of the transmission line. The actual distribution of the current on the line will be altered since the load is not infinite due to the fringing effects.

In order to develop the integration procedure, we start with a differential electric field due to a short electric dipole given in (7.20). The infinitesimal current element IL is replaced with the current distribution at that particular location I(z)dz

$$
\begin{equation*}
\mathrm{dE}_{\theta}=j \frac{Z_{0} \mathrm{k}}{4 \pi}(\mathrm{l}(z) \mathrm{dz}) \frac{\mathrm{e}^{-\mathrm{jkg} \rho^{\prime}}}{\rho^{\prime}} \sin \theta \tag{7.24}
\end{equation*}
$$

The distance $\rho^{\prime}$ that appears here, can be written in terms of the distance $\rho$ between the point of observation and the center of the dipole shown in Figure 7-8 as

$$
\begin{equation*}
\rho^{\prime}=\left[\rho^{2}+z^{2}-2 \rho z \cos \theta\right]^{1 / 2} \approx \rho-z \cos \theta \tag{7.25}
\end{equation*}
$$

We are allowed to make this approximation since the field distribution in the far field is to be determined, that is $\rho \gg z$. The difference in magnitude between $1 / \rho$ ' and $1 / \rho$ is insignificant and can be neglected. However, it is important that we incorporate this difference in the phase term $\mathrm{e}^{-\mathrm{jk} \rho^{\prime}}$. Small changes in distance may be a reasonable fraction of a wavelength $\lambda$ that could cause this term to change sign from a "+" to a "-". This will have dramatic effects as will be shown below.

In order to actually compute the electromagnetic fields radiated from an antenna, we have to select the distribution for the current and perform an inte-

## Long dipole antenna

gration over the coordinate of the antenna $z$. This will be done with the current distribution given in (7.23) for which we write that

$$
\begin{equation*}
E_{\theta}=Z_{0} H_{\varphi}=j \frac{Z_{0} I_{m} k \sin \theta}{4 \pi} \frac{e^{-j k \rho}}{\rho} \int_{-L / 2}^{L / 2} \sin \left[k\left(\frac{L}{2}-|z|\right)\right] e^{j k z \cos \theta} d z \tag{7.26}
\end{equation*}
$$

Before actually performing the integration required at this stage, let us comment on the terms within the integrand. The term with a sine is an even function in the variable of integration $z$ as shown in Figure 7-9b. The product of this term and

$$
\mathrm{e}^{\mathrm{jkz} \cos \theta}=\cos (\mathrm{kz} \cos \theta)+\mathrm{j} \sin (\mathrm{kz} \cos \theta)
$$

will yield two terms, one of which is an odd function in the variable $z$ and one that is an even function in the variable $z$. Since the limits of the integral are symmetric about the origin, only the integrand that includes the even function will yield a nonzero result. The integral (7.26) reduces to

$$
\begin{equation*}
E_{\theta}=Z_{0} H_{\varphi}=j 2 \frac{Z_{0} I_{m} k}{4 \pi} \frac{e^{-j k \rho}}{\rho} \sin \theta \int_{0}^{L / 2} \sin \left[k\left(\frac{L}{2}-|z|\right)\right] \cos (k z \cos \theta) d z \tag{7.27}
\end{equation*}
$$

After a lengthy integration, we finally obtain ${ }^{1}$

$$
\begin{equation*}
E_{\theta}=j 60 I_{m} \frac{e^{-j k \rho}}{\rho} F(\theta) \tag{7.28}
\end{equation*}
$$

where the following explicit expression for the radiation pattern $F(\theta)$ is found

$$
\begin{align*}
\mathrm{F}(\theta) \equiv \mathrm{F}_{1}(\theta) \mathrm{F}_{\mathrm{a}}(\theta)= & {[\sin \theta] \times\left[\frac{\cos \left(\frac{\mathrm{kL}}{2} \cos \theta\right)-\cos \left(\frac{\mathrm{kL}}{2}\right)}{\sin ^{2} \theta}\right]=}  \tag{7.29}\\
& \frac{\cos \left(\frac{\mathrm{kL}}{2} \cos \theta\right)-\cos \left(\frac{\mathrm{kL}}{2}\right)}{\sin \theta}
\end{align*}
$$

The final solution is a product of two terms. The first terms $F_{1}(\theta)$ corresponds to the radiation characteristics of a single element which is a short dipole with a

[^21]center, located at $\rho=0$. The second term $F_{a}(\theta)$ is sometimes called an array factor of the long dipole antenna. The fact that the radiation is the product of


Figure $7-10^{\text {MATLAB }}$. E-plane radiation patterns for center-fed dipole antennas of different lengths: (a) $L=\lambda / 2$; (b) $L=\lambda$; (c) $L=3 \lambda / 2$; (d) $L=2 \lambda$. The antenna with the dimension $L=\lambda / 2$ is called a half-wave dipole. In addition, the radiation pattern for a short dipole is indicated in (a) with a dashed line.
these two terms is important in its own right. Although the radiation from an infinitesimal element cannot be altered, it is possible to have control over the

## Long dipole antenna

array factor. This is the term that engineers can control. This function depends also on the normalized length of the long dipole antenna $\xi=\mathrm{kL}$.

In Figure 7-10, we illustrate the E-plane radiation pattern for four different dipole lengths. The array in this case consists of an array of uniformly excited infinitesimal dipoles and the radiation pattern has the property that $F\left(0^{\circ}\right)=0$, while $\mathrm{F}\left(90^{\circ}\right)=$ maximum in the first two cases, but not in the third one, where the maximum shifts to $\theta=45^{\circ}$. The radiation for a half wave dipole is shown in Figure 7-10a. This can be compared with the radiation for the short dipole which is indicated with a dashed line. The H-plane radiation patterns are azimuthally symmetric circles since $F(\theta)$ is independent of the angle $\theta$.

We note from Figure 7-10c that the maximum in the radiated power tends to shift away from $\theta=90^{\circ}$ as the length $L$ is changed. If we set $L=2 \lambda$ in (7.29), we find that the radiation at $\theta=90^{\circ}$ is equal to zero (Figure $7-10 \mathrm{~d}$ ). The contours depicted in Figure 7-10 are called lobes. The lobe in the direction of the maximum is called the main lobe and the others are called side lobes. If we were to traverse about the antenna at a constant radius and monitored the received signal with a detector that was sensitive to the phase of the detected signal, we would note a phase shift of $180^{\circ}$ as we move from one lobe to the adjacent one. The lobe structure is another example of constructive and destructive interference or phase mixing that was discussed when the topics of dispersion and group velocity were presented.

## Antenna parameters

### 7.4. Antenna parameters

In addition to the radiation pattern for the antenna that was discussed in the previous sections, there are other parameters that are used to characterize an antenna. If we connected the antenna to a transmission line, we could think of the antenna as being merely a load impedance. The radiation of electromagnetic power into the external environment removes the power from the circuit and hence acts like a resistor that just heats up. This is depicted in Figure 7-11. We consider several important antenna parameters here.


Figure 7-11. (a) An antenna that radiates electromagnetic energy is connected with a transmission line to a source of electromagnetic energy. (b) Coaxial cable connected to a ground plane. (c) Equivalent circuit of either structure.

## Antenna parameters

### 7.4.1. Radiation resistance

We consider the antenna shown In Figure 7-11c to be a load impedance $Z_{L}$ that is connected to the generator with a transmission line of length $L$ that has a characteristic impedance $Z_{C}$ and a propagation constant $\gamma$. In order to compute the value of the load impedance $Z_{L}$, we will have to return to Poynting's vector. Recall that this vector is a measure of the power density at a point in space that is calculated using the quantities in the electromagnetic wave. The total power radiated from the antenna can be computed by surrounding the antenna with a large imaginary sphere whose radius is R as shown in Figure $7-12$. The radius $R$ will be chosen so the surface of the sphere will be in the far field region. Then any power that is radiated from the antenna will have to pass through the sphere in order to propagate to distances greater than this radius $R$. As shown in Example 7-11, this radius can be approximated in terms of antenna size $L$ to be $R \approx L^{2} / 2 \lambda$.

The total radiated power from the antenna is computed by integrating the time-average Poynting vector over this entire closed spherical surface. From (7.1), this becomes

$$
\begin{align*}
& P_{r}=\frac{1}{2} \operatorname{Re} \oint_{S} E \times \mathbf{H}^{*} \bullet d s=  \tag{7.30}\\
& \quad \frac{1}{2} \operatorname{Re} \int_{0}^{2 \pi} d \phi \int_{0}^{\pi} d \theta R^{2} \sin \theta\left[E_{\theta} H_{\phi}^{*}\right]
\end{align*}
$$

The factor of (1/2) arises since we are considering a time - average power over a temporal cycle of the oscillation. As shown in Chapter 4, the conjugated value of H must be employed. This average radiated power can be considered to be lost as far as the source is concerned and therefore the antenna is similar to a resistor in that it is dissipating the power from the source. This resistance is called the radiation resistance $\mathrm{R}_{\mathrm{r}}$ and it is defined as

## Antenna parameters

$$
\begin{equation*}
\mathrm{R}_{\mathrm{r}}=\frac{2 \mathrm{P}_{\mathrm{r}}}{\mathrm{I}_{0}^{2}} \tag{7.31}
\end{equation*}
$$

where $I_{0}$ is the maximum amplitude of the current at the input terminals of the antenna. We will calculate the radiation resistance for the Hertzian dipole and a half-wave dipole as examples.


Figure 7-12. Electromagnetic power radiated from an antenna will pass through a sphere of radius $R$.

Example 7-6 ${ }^{\text {MATLAB }}$. Find the radiation resistance of a short dipole.
Answer: The radiated power $P_{r}$ from Hertzian dipole is computed using (7.22).
Substituting the terms for the Hertzian dipole into (7.22), using $Z_{0}=120 \pi$ and $\mathrm{k}=2 \pi / \lambda$, and employing the definition (7.31), we obtain

$$
\mathrm{R}_{\mathrm{r}}=80 \pi^{2}\left(\frac{\mathrm{~L}}{\lambda}\right)^{2}\left(\frac{\mathrm{I}_{\mathrm{av}}}{\mathrm{I}_{0}}\right)^{2}=80 \pi^{2}\left(\frac{\mathrm{~L}}{\lambda}\right)^{2}
$$

where a uniform current distribution is assumed or $\mathrm{I}_{\mathrm{av}}=\mathrm{I}_{0}$. If there were a triangular current distribution depicted in the next example, we could write $\mathrm{I}_{\mathrm{av}}=\mathrm{I}_{0} / 2$ In this case, the radiation resistance is $1 / 4$ of the previous value

$$
\mathrm{R}_{\mathrm{r}}=20 \pi^{2}\left(\frac{\mathrm{~L}}{\lambda}\right)^{2}
$$

## Antenna parameters

The frequency response of the radiation resistance (for $\xi=k L<1$, or $\mathrm{L} / \lambda<0.16)$ is shown in the figure below. The small values of the radiation resistance shows that this antenna is not very efficient. Further comments concerning this antenna will be referred to Example 7-9.


Example 7-7. Find an expression for the radiation resistance of a short monopole antenna that is placed above a ground plane. The length of the antenna is $\lambda / 8$. Answer: The short electric monopole antenna near the ground is equivalent to a short dipole antenna with a length that is twice as long as shown in the figure below. The additional contribution is from the "image antenna." In order to ensure that there is an excellent ground plane beneath the monopole antenna,

## Antenna parameters

conducting wires are typically implanted underneath the antenna and emanate radially away from the monopole a distance that is approximately equal to the height of the antenna.


It is possible to estimate the validity of the approximation for a short dipole that is $\lambda / 4$ in length. From the previous example, the radiation resistance is computed to be he $R_{r} \approx 12.3 \Omega$, while the actual value obtained from the theory of the linear antenna is ${ }^{1} R_{r}=6.7 \Omega$. Hence, even for this small length, the short dipole approximation is not a valid approximation.

Example 7-8 ${ }^{\text {MATLAB }}$. Obtain the frequency response of the radiation resistance for a linear antenna.

Answer: The expression for the radiation resistance of a linear antenna is much more complicated, than that derived in Example 7-6. It involves new special functions involving sine and cosine integrals and it is usually considered in the books on advanced electromagnetics ${ }^{1}$. We are only interested in illustrating this result as shown in the figure. One can see that for the popular case of a half-

[^22]
## Antenna parameters

wave dipole the radiation resistance is found to be $R_{r}=73.1 \Omega$ has shown in Problem 10, which is very close to the measured value.


### 7.4.2. Directivity

The equation for the radiated power (7.30) can be written not only as a surface integral of the time-average radial component of Poynting's vector $S_{\rho}=\frac{1}{2} \operatorname{Re}\left\{E_{\theta} H_{\phi}^{*}\right\}$, it can also be written also as an integral over a solid angle. For this purpose, we define the radiation intensity as $I(\theta, \phi)=R^{2} S_{\rho}(\theta, \phi)$ which yields to following representation of the radiated power

$$
\begin{equation*}
\mathrm{P}_{\mathrm{r}}=\oint_{4 \pi}(\theta, \phi) \mathrm{d} \Omega \tag{7.32}
\end{equation*}
$$

## Antenna parameters

The ratio of $I(\theta, \phi)$ divided by its maximum value is the normalized power radiation pattern

$$
\begin{equation*}
\mathrm{I}_{\mathrm{n}}(\theta, \phi)=\frac{\mathrm{I}(\theta, \phi)}{\mathrm{I}(\theta, \phi)_{\max }} \tag{7.33}
\end{equation*}
$$

The beam solid angle of the antenna is defined as

$$
\begin{equation*}
\Omega_{\mathrm{A}}=\oint_{4 \pi} \mathrm{I}_{\mathrm{n}}(\theta, \phi) \mathrm{d} \Omega \equiv \int_{\phi=0}^{2 \pi} \mathrm{~d} \phi \int_{\theta=0}^{\pi} \mathrm{I}_{\mathrm{n}}(\theta, \phi) \sin \theta \mathrm{d} \theta \tag{7.34}
\end{equation*}
$$

In a complete sphere, there are $4 \pi$ steradians. From the definition, it follows that for an isotropic antenna that $I_{n}(\theta, \phi) \equiv 1$ and the beam solid angle is $\Omega_{\mathrm{A}}=4 \pi$.

The next parameter that defines an antenna system is the directivity. The directivity of an antenna is defined as the ratio of the maximum radiation intensity of a transmitting antenna divided by the the average radiation intensity from an isotropic radiator with the same input power is given by

$$
\begin{equation*}
\mathrm{D}=\frac{\mathrm{I}(\theta, \phi)_{\max }}{\mathrm{P} / 4 \pi}=\frac{4 \pi}{\oint_{4 \pi} \mathrm{I}_{\mathrm{n}}(\theta, \phi) \mathrm{d} \Omega} \equiv \frac{4 \pi}{\Omega_{\mathrm{A}}} \tag{7.35}
\end{equation*}
$$

Because the denominator $\Omega_{A}$ is always less than 1 , the directivity $D>1$.
Example 7-9. Find the directivity of the Hertzian dipole.
Answer: Use the definition (7.35) for the directivity including the normalized radiation intensity for a short dipole $I_{n}(\theta, \phi)=\sin ^{2} \theta$ and obtain

$$
\mathrm{D}=\frac{4 \pi}{2 \pi \int_{0}^{\pi} \sin ^{2} \theta \sin \theta \mathrm{~d} \theta}=\frac{2}{\int_{0}^{\pi}\left(\cos ^{2} \theta-1\right) \mathrm{d} \cos \theta}=\frac{2}{-\frac{2}{3}+2}=\frac{3}{2} \equiv 1.5 .
$$

For the short dipole, the directivity is $D=1.5$ or $10 \log _{10}(1.5)=1.76 \mathrm{~dB}$.

A more complicated calculation performed for the half-wave dipole leads to a little higher value for the directivity $\mathrm{D}=1.64$ or 2.15 dB as given in Problem 9 .

## Antenna parameters

### 7.4.3. Antenna gain

The gain is a global characteristic of the antenna. Because of this, the antenna efficiency $\eta$ is involved and it is related to the directivity of the antenna. By definition the gain is

$$
\begin{equation*}
G=\eta D \tag{7.36}
\end{equation*}
$$

For the case of lossless antenna, $\eta=1$ and the gain is equal to the directivity. For a lossy antenna, $\eta<1$ and the gain is less than the directivity.

### 7.4.4. Beam width

The half-power beam width (HPBW) is a property that refers only to the the main lobe. It tells nothing about the side lobes. By definition, it is defined by the angle $\theta_{\mathrm{HP}}$ where the power radiation pattern assumes one-half of the maximum value. In the case of rotational symmetry, this value is $I_{n}(\theta)=1 / 2$. It is reasonable to approximate the directivity of an antenna in terms of the beam width

$$
\begin{equation*}
\mathrm{D} \approx \frac{4 \pi}{\theta_{\mathrm{HP}} \phi_{\mathrm{HP}}} \tag{7.37}
\end{equation*}
$$

Example 7-10. Find the HPBW of a short dipole.
Answer: The normalized radiation intensity for the electric field of a short dipole is given by $I_{n}(\theta, \phi)=\sin ^{2} \theta$ and it has a maximum value $I_{n}=1$ for $\theta=\pi / 2$. The value $I_{n}=1 / 2$ is found at the angles $\theta=\pi / 4 ; 3 \pi / 4$. Therefore, the HPBW for the electric field is $\theta_{H P}=\pi / 2$. The normalized radiation intensity for the magnetic field of a short dipole is $I_{n}(\theta, \phi) \equiv 1$. The HPBW for the magnetic field intensity is $\phi_{\mathrm{HP}}=2 \pi$. From (7.37), we can approximate the directivity to be $D \approx 4 / \pi=1.27$. The value that was computed in Example $7-9$ was $\mathrm{D}=1.5$.

## Antenna parameters

### 7.4.5. Effective aperture

Several of the parameters that we have already encountered in describing the radiation from an antenna can also be used to describe the reception characteristics of an antenna. This is based on a principle that is called reciprocity and it is certainly valid in a vacuum. There are certain parameters that are more common for antennas that are used for a receiving antenna. One of these that is used to characterize a receiving aperture antenna is the effective aperture or its reception cross-sectional area. The effective area $A_{e}$ of $a$ receiving antenna is defined as the ratio of the average power $P_{L}$ delivered to a load impedance that is matched to the antenna if there is an average power density $\mathrm{S}_{\mathrm{av}}$ incident upon the antenna

$$
\begin{equation*}
S_{a v}=\frac{E^{2}}{2 Z_{0}}=\frac{E^{2}}{240 \pi} \tag{7.38}
\end{equation*}
$$

It is reasonable to assume that there is a plane wave that is incident upon the receiving antenna that is terminated with a matched load impedance $Z_{C}$ as shown in Figure 7-11c. Assume the incident wave arrives in the direction of the maximum of the radiation pattern. Then

$$
\begin{equation*}
P_{L}=S_{a v} A_{e} \tag{7.39}
\end{equation*}
$$

As you may recall from the circuit theory that the maximum power can be delivered to a load if it has a value that is equal to complex conjugate of the antenna impedance $Z_{L}=Z_{A}^{*}$. Replacing the antenna with an equivalent generator having the same voltage V and impedance $\mathrm{Z}_{\mathrm{A}}$, we obtain a current at the terminals to be

$$
\begin{equation*}
I_{0}=\frac{V}{Z_{A}+Z_{L}} \tag{7.40}
\end{equation*}
$$

Therefore, the maximum power that is dissipated in the load is given by

$$
\begin{equation*}
P_{L}=\frac{1}{2} I_{0}^{2} R_{L}=\frac{1}{2}\left(\frac{V}{Z_{A}+Z_{A}^{*}}\right)^{2} R_{L}=\frac{V^{2}}{8 R_{A}} \tag{7.41}
\end{equation*}
$$

## Antenna parameters

where we have defined $Z_{A}+Z_{A}^{*}=2 R_{A}$.
For the Hertzian dipole, the antenna resistance $R_{A}=R_{r}$ was calculated in Example 7-6 and the maximum voltage in the direction $\theta=\pi / 2$ was found to be $\mathrm{V}=\mathrm{E} \sin \theta \mathrm{L}=\mathrm{EL}$

$$
\begin{equation*}
\mathrm{P}_{\mathrm{L}}=\frac{(E L)^{2}}{8 \times 80 \pi^{2}\left(\frac{\mathrm{~L}}{\lambda}\right)^{2}}=\frac{\mathrm{E}^{2} \lambda^{2}}{640 \pi^{2}} \tag{7.42}
\end{equation*}
$$

For the Hertzian dipole, we find the effective area to be

$$
\begin{equation*}
\mathrm{A}_{\mathrm{e}}=\frac{\lambda^{2}}{4 \pi}\left(\frac{3}{2}\right)=\frac{3 \lambda^{2}}{8 \pi} \tag{7.43}
\end{equation*}
$$

using (7.39), (7.38) and (7.42). The effective aperture and the directivity of a short dipole antenna can also be obtained and it is written as

$$
\begin{equation*}
\mathrm{A}_{\mathrm{e}}=\frac{\lambda^{2}}{4 \pi} \mathrm{D} \tag{7.44}
\end{equation*}
$$

### 7.4.6. Friis transmission equation

In this section, we obtain the relation between an antenna that is used for transmission and another antenna which is used for reception. This is depicted in Figure 7-13.


Figure 7-13. Two antennas separated by a distance R.

## Antenna parameters

Let antenna A in Figure 7-13 transmit to antenna B. Both antennas are in the far field region. The gain of the transmitting antenna $A$ in the direction of $B$ is $G_{A}$ which in the lossless case is equal to $D_{A}$. Hence, the time-average power density at $B$ is

$$
\begin{equation*}
S_{a v}=\frac{P_{t}}{4 \pi R^{2}} G_{t} \tag{7.45}
\end{equation*}
$$

Write (7.39) for the received power and replace the gain of antenna A by the effective aperture $\mathrm{A}_{\mathrm{e}, \mathrm{t}}$ using (7.44) and obtain

$$
\begin{equation*}
P_{r}=S_{a v} A_{e, r}=\left[\frac{P_{t}}{4 \pi R^{2}} G_{t}\right] A_{e, r}=\frac{P_{t}}{R^{2}} \frac{A_{e, t}}{\lambda^{2}} A_{e, r} \tag{7.46}
\end{equation*}
$$

This can be written as

$$
\begin{equation*}
\frac{P_{r}}{P_{t}}=\frac{A_{e, t} A_{e, r}}{\lambda^{2} R^{2}} \tag{7.47}
\end{equation*}
$$

Equation (7.47) is called the Friis transmission equation.
Example 7-11. Find a criterion that a receiving antenna is in the far field of a transmitting antenna. Make an estimation for the distance $d$ of the far zone if $D=10 \mathrm{~cm}$ and $\lambda=3 \mathrm{~cm}$. Assume that they have antenna gains $G_{A}=1.5$ and $\mathrm{G}_{\mathrm{B}}=1.64$ respectively and find the ratio $\mathrm{P}_{\mathrm{r}} / \mathrm{P}_{\mathrm{t}}$ at that distance.


## Antenna parameters

Answer: A requirement that the Friis transmission equation be applicable is that both antennas be in the far field. The radiation from a point source is always in the far field. The receiving antenna will be in the far field if the incident spherical wave deviates from a plane wave by only a fraction of a wavelength. The largest dimension of the receiving antenna is $D$. This implies that the deviation $\Delta$ is approximately $\Delta \approx \lambda / 4$ which means a phase difference $90^{\circ}$. From the figure, we write

$$
\mathrm{R}^{2}=(\mathrm{R}-\Delta)^{2}+\left(\frac{\mathrm{D}}{2}\right)^{2} \approx \mathrm{R}^{2}-2 \mathrm{R} \Delta+\frac{\mathrm{D}^{2}}{4}
$$

This implies that the receiving antenna will be in the far field or the Fraunhofer zone if the second term is at least comparable with the third one or $R \approx \frac{D^{2}}{8 \Delta}$. This means

$$
\mathrm{R} \approx \frac{\mathrm{D}^{2}}{2 \lambda} .
$$

For the particular values, we obtain $R=\frac{0.1^{2}}{2 \times 0.03}=16.7 \mathrm{~cm}$ - beyond this distance, the receiving antenna will be in the far field region of the transmitting antenna.

Applying (7.44) for the lossless case yields the effective apetures of both antennas to be

$$
A_{e, t}=\frac{0.03^{2} \times 1.5}{4 \pi}=1.07 \mathrm{~cm}^{2} ; \quad A_{e, r}=\frac{0.03^{2} \times 1.64}{4 \pi}=1.17 \mathrm{~cm}^{2} .
$$

The Friis equation (7.47) gives the ratio

$$
\frac{P_{r}}{P_{t}}=\frac{1.07 \times 1.17}{16.7^{2} \times 3^{2}}=5 \times 10^{-4}
$$

The degradation of a received signal is approximately

$$
-10 \log _{10}\left(5 \times 10^{-4}\right)=33 \mathrm{~dB}
$$

## Magnetic dipole

### 7.5. Magnetic dipole antenna

We will find the electromagnetic fields that are radiated from a short magnetic dipole as being a second example of a small antenna. The magnetic dipole consists of a small filamentary loop whose radius is a. The loop carries a harmonic current $i(t)=I \cos \omega t$ around its circumference. This antenna is depicted in Figure $7-14$, where $k a \equiv \frac{2 \pi}{\lambda} a \ll 1$ is assumed. The retarded vector potential resulting from this current loop is determined from (7.13). Since the current is confined to the loop, this integral becomes

$$
\begin{equation*}
\mathbf{A}=\frac{\mu_{0} \mathrm{I}}{4 \pi} \oint_{\mathrm{L}} \frac{\mathrm{e}^{-\mathrm{jk} \rho^{\prime}}}{\rho^{\prime}} \mathrm{dl} \mathrm{l}^{\prime} \tag{7.48}
\end{equation*}
$$

where $\mathbf{I}=\mathbf{l} \mathbf{u}_{\varphi}$ and the time harmonic terme ${ }^{j \omega t}$ is understood to be included. This integral is not easy to evaluate since the terms within the integrand depend on the particular location where dl' is being evaluated. We can, however, obtain an approximate solution that will illustrate the expected behavior using the following procedure. ${ }^{1}$ The exponential term can be written as

$$
\begin{equation*}
\mathrm{e}^{-\mathrm{jk} \rho^{\prime}}=\mathrm{e}^{-\mathrm{jk} \rho} \mathrm{e}^{-\mathrm{jk}\left(\rho^{\prime}-\rho\right)} \approx \mathrm{e}^{-\mathrm{jk} \mathrm{\rho} \rho}\left[1-\mathrm{jk}\left(\left(\rho^{\prime}-\rho\right)\right)\right] \tag{7.49}
\end{equation*}
$$

In expanding the second exponential term, we have made the approximation that the loop is small with respect to the distance $\rho$ between the center of the loop and the point of observation $(\mathrm{a} \ll \rho)$. This is very similar to the approximation that we first encountered in describing an electrical dipole. Hence (7.48) can be written as

$$
\begin{equation*}
\mathbf{A}=\frac{\mu_{0} \mathrm{I}}{4 \pi} \mathrm{e}^{-\mathrm{jk} \mathrm{\rho} \rho} \mathbf{u}_{\varphi}\left[(1+\mathrm{jk} \rho) \oint_{\mathrm{L}} \frac{\mathrm{dl}^{\prime}}{\rho^{\prime}}-\mathrm{jk} \oint_{\mathrm{L}} \mathrm{dl} \mathrm{\prime}\right] \tag{7.50}
\end{equation*}
$$

[^23]

Figure 7-14. A magnetic dipole.

The second integral is equal to zero since this integral is akin to running around in a circle, we just return back to the original starting point and have progressed nowhere. The first integral is evaluated in Example 7-12.
Example 7-12. Evaluate the integral $\oint \frac{\mathrm{dl}}{\mathrm{L}} \mathrm{\rho}^{\prime}$.
Answer: Use the vector identity (see Appendix 1)

$$
\oint_{\mathrm{L}} \mathrm{bdl}=\int_{\Delta \mathrm{s}}\left(\mathbf{u}_{\mathbf{n}} \times \nabla \mathrm{b}\right) \cdot \mathbf{d s}
$$

to convert the closed line integral into a surface integral. The scalar quantity $b$ is equal to $b=1 / \rho^{\prime}$. With reference to Figure $7-14$, we note that $u_{n}=u_{z}$ since the loop is in the xy plane. Therefore

$$
\oint_{L} \frac{\mathrm{~d} \mathrm{l}^{\prime}}{\rho^{\prime}}=\int_{\Delta s}\left(\mathbf{u}_{\mathbf{z}} \times \nabla \frac{1}{\rho^{\prime}}\right) \bullet \mathbf{u}_{\varphi} d s=-\int_{\Delta s}\left(\mathbf{u}_{\mathbf{z}} \times \frac{\mathbf{u}_{\rho^{\prime}}}{\rho^{\prime 2}}\right) \bullet \mathbf{u}_{\varphi} d s
$$

where

$$
\nabla \frac{1}{\rho^{\prime}}=-\frac{\mathbf{u}_{\rho}}{\rho^{\prime 2}}
$$

For large distances from the current loop, we can replace the amplitude and the unit vector with $\rho^{\prime} \approx \rho$ and $\mathbf{u}_{\rho^{\prime}} \approx \mathbf{u}_{\rho}$. With these approximations, the integral becomes

$$
\int_{\Delta s} \frac{\left(\mathbf{u}_{z} \times \mathbf{u}_{\rho}\right)}{\rho^{2}} \mathrm{ds}=\frac{\left(\mathbf{u}_{\mathrm{z}} \times \mathbf{u}_{\rho}\right)}{\rho^{2}} \int_{\Delta \mathrm{S}} \mathrm{ds}
$$

The surface integral yields a factor of $\pi \mathrm{a}^{2}$. Finally, we make use of the vector relation $\mathbf{u}_{\mathbf{z}}=\mathbf{u}_{\mathbf{\rho}} \cos \theta-\mathbf{u}_{\boldsymbol{\theta}} \sin \theta$ to compute in spherical coordinates that

$$
\mathbf{u}_{\mathbf{z}} \times \mathbf{u}_{\mathbf{\rho}}=\mathbf{u}_{\phi} \sin \theta
$$

Hence, the final result of the integrations yields

$$
\oint_{\mathrm{L}} \frac{\mathrm{~d} \mathrm{l}^{\prime}}{\rho^{\prime}}=\frac{\pi \mathrm{a}^{2}}{\rho^{2}} \sin \theta
$$

We find from (7.50) that the final evaluation of this integral leads to the vector potential in the far field being written as

$$
\begin{equation*}
\mathbf{A}=\mathbf{u}_{\varphi} \frac{\mu_{0}\left(l \pi \mathrm{a}^{2}\right)}{4 \pi} \frac{(1+\mathrm{jk} \rho) \mathrm{e}^{-\mathrm{jk} \rho}}{\rho^{2}} \sin \theta \approx \mathbf{u}_{\varphi} \mathrm{j} \frac{\mu_{0}\left(l \pi \mathrm{a}^{2}\right) \mathrm{k}}{4 \pi} \frac{\mathrm{e}^{-\mathrm{jk} \rho}}{\rho} \sin \theta \tag{7.51}
\end{equation*}
$$

We recognize the term $\left(I \pi \mathrm{a}^{2}\right)$ from Example 2-26 as being the magnitude m of the magnetic dipole moment $\mathbf{m}=\left(1 \pi \mathrm{a}^{2}\right) \mathbf{u}_{\mathbf{z}}$.

Having found the vector potential, we can evaluate the electromagnetic fields in the far field using (7.18) and (7.19)

$$
\begin{gather*}
H_{\theta} \approx-\frac{\omega \mu_{0} m k}{4 \pi Z_{0}} \frac{e^{-j k \rho}}{\rho} \sin \theta  \tag{7.52}\\
E_{\phi}=-Z_{0} H_{\theta} \approx \frac{\omega \mu_{0} m k}{4 \pi} \frac{e^{-j k \rho}}{\rho} \sin \theta \tag{7.53}
\end{gather*}
$$

If we compare (7.52) - (7.53) with (7.18) and (7.20), we note a similarity in the field components of the short magnetic dipole and the short electric dipole. In the far field, the magnitude of the two fields each decay as $\rho^{-1}$, the ratio of the two fields is equal to the characteristic impedance $Z_{0}$ of free space, and the radiation pattern $F(\theta)=\sin \theta$ is the same and is shown in Figure 7-7. The

## Magnetic dipole

directivity of the short magnetic dipole antenna is also $D=1.5$ as in the case of a short electric dipole that was evaluated in Example 7-9.

Using the definition (7.31) for the radiation resistance and assuming a uniform current distribution, we obtain

$$
\begin{equation*}
\mathrm{R}_{\mathrm{r}}=20 \pi^{2}(\mathrm{ka})^{4} \tag{7.54}
\end{equation*}
$$

The radiation resistance is different than we found in Example 7-6 for a short electric dipole.

Example 7-13. Find the current that is required to radiate 10 watts from a loop whose circumference is equal to $\lambda / 5$.

Answer: Applying (7.54) for the case of ka $=2 \pi \mathrm{a} / \lambda=0.2$, we obtain

$$
\mathrm{R}_{\mathrm{r}}=20 \times \pi^{2} \times 0.2^{4}=0.316 \Omega
$$

The radiated power can be found from (7.31) to be

$$
P_{r}=\frac{1}{2} R_{r} I^{2}
$$

which yields the current

$$
\mathrm{I}=\sqrt{\frac{2 \mathrm{P}_{r}}{R_{r}}}=\sqrt{\frac{2 \times 10}{0.316}}=7.95 \mathrm{~A} .
$$

It should be apparent that the small antenna would not be very useful in radiating large amounts of power.

Next, it will be of some interest to find a suitable expression for the radiation resistance of a long magnetic dipole (or a loop antenna) where the radius a is comparable with the wavelength, or $k a \geq 1$. Again, it is a difficult task to find the actual current distribution - this yields to the solution of a proper integral equation. The current distribution strongly depends on the type of the

## Magnetic dipole

voltage exciatation. For the special case of a sinusoidal current distribution suitable equations and plots are available ${ }^{2}$.

[^24]
## Aperture antennas. Diffraction of waves

### 7.6. Aperture antennas. Diffraction of waves

### 7.6.1. Diffraction of waves. Huygens' principle

In Chapter 5, two effects concerning an electromagnetic wave incident upon an interface between two different dielectrics were considered, namely reflection and transmission. These effects can be interpreted in terms of constructive and destructive interference of a large number of elementary fields, each with a suitable amplitude and phase. In this section, we describe a new effect that is called diffraction. It usually appears near a sharp boundary such as an edge between a dielectric and a metal surface such as a knife edge. The understanding of this phenomenon is based on Huygens' principle which states that each point of a primary wavefront can be considered to be the source of a secondary spherical wave that propagates radially from that point. It is possible to construct a secondary wavefront from the envelope of these secondary waves. This fundamental principle that is widely used in physical optics is illustrated in Figure $7-15$. This principle can be used to explain the "bending of plane waves" around obstacles or as it is properly known as the diffraction of waves.


Figure 7-15. Illustration of Huygens' principle in physical optics.

In order to illustrate diffraction, we consider a plane electromagnetic wave that is incident upon a metallic surface that occupies half of the region of interest.

## Aperture antennas. Diffraction of waves

This is shown in Figure 7-16a. The application of Huygens' principle states that the electric field at an arbitrary point of the secondary wavefront envelope can be obtained from the following integral

$$
\begin{equation*}
\mathrm{E}=\int_{\substack{\text { over } \\ \text { x-axis }}} \mathrm{dE}=\int_{a}^{\infty} \mathrm{E}_{0} \frac{\mathrm{e}^{-\mathrm{jk}(\rho+\delta)}}{\rho} \mathrm{dx} \tag{7.55}
\end{equation*}
$$

where $d E$ is the differential electric field at point $P$ due to the point source at a distance $x$ from the origin 0 , as shown in Figure $7-16 b$. The term $\delta$ reflects the distance between two equiphase contours and it is written as

$$
\begin{equation*}
\delta=\sqrt{\rho^{2}+x^{2}}-\rho=\rho\left[\left(1+\left(\frac{x}{\rho}\right)^{2}\right)^{1 / 2}-1\right] \approx \rho\left[\left(1+\frac{x^{2}}{2 \rho^{2}}\right)-1\right]=\frac{x^{2}}{2 \rho} \tag{7.56}
\end{equation*}
$$

where we have assumed that $\delta \ll \rho$. Keeping the quadratic term in (7.56) is called the Fresnel approximation, while keeping the linear term only is called the Fraunhofer approximation. Geometric optics neglects the effects of diffraction and rays rather than the waves are considered.

Defining $q^{2}=2 / \rho \lambda$ and $u=q x$ leads to the following expression for the electric field at the point of observation

$$
\begin{align*}
E & =\frac{E_{0}}{\rho} e^{-j k \rho} \int_{a}^{\infty} e^{-j k \delta} d x=\frac{E_{0}}{\rho} e^{-j k \rho} \int_{a}^{\infty} e^{-j \frac{k x^{2}}{2 \rho}} d x  \tag{7.57}\\
& =\frac{E_{0}}{q \rho} e^{-j k \rho} \int_{q a}^{\infty} e^{-j \frac{\pi u^{2}}{2}} d u
\end{align*}
$$

The last integral can be written as

$$
\begin{equation*}
E=\frac{E_{0}}{q \rho} e^{-j k \rho}\left[\int_{0}^{\infty} e^{-j \frac{\pi u^{2}}{2}} d u-\int_{0}^{q a} e^{-j \frac{\pi u^{2}}{2}} d u\right] \tag{7.58}
\end{equation*}
$$

The integrals in (7.58) are of the form of a well studied integral in optics and are called Fresnel integrals ${ }^{1}$. The solution of these integrals is given using the following expansion

[^25]
## Aperture antennas. Diffraction of waves

$$
\begin{equation*}
E=\frac{E_{0}}{q \rho} e^{-j k \rho}\left\{\frac{1-j}{2}-[C(q a)-j S(q a)]\right\} \tag{7.59}
\end{equation*}
$$

where the Fresnel cosine integral and Fresnel sine integral are defined as

$$
\begin{align*}
& C(z)=\int_{0}^{z} \cos \left(\frac{\pi u^{2}}{2}\right) d u \\
& S(z)=\int_{0}^{z} \sin \left(\frac{\pi u^{2}}{2}\right) d u \tag{7.60}
\end{align*}
$$



Figure 7-16. (a) Plane wave incident upon a conducting half-plane.
(b) Physical optics interpretation of the diffraction by a conducting edge using Huygens' principle.

A graph of $\mathrm{S}(\mathrm{qa})$ as a function of $\mathrm{C}(\mathrm{qa})$ is known as Cornu's spiral and it is presented in Figure $7-17$ where the parameter $q a=\sqrt{\frac{2 a^{2}}{\rho \lambda}}$. The MATLABcalculation of the Fresnel integrals is based on the computation of the complex error function. Both the Fresnel cosine integral and the Fresnel sine integral are odd functions and their asymptotic values are $\pm 0.5$.

The average power density $\mathrm{S}_{\mathrm{av}}$ in $\left(\mathrm{W} / \mathrm{m}^{2}\right)$ is calculated using the Poynting vector and we write

$$
\begin{equation*}
S_{\mathrm{av}}(\mathrm{qa})=\frac{|\mathrm{E}|^{2}}{2 Z_{0}}=\mathrm{S}_{0} \frac{1}{2}\left\{\left[\frac{1}{2}-\mathrm{C}(\mathrm{qa})\right]^{2}+\left[\frac{1}{2}-\mathrm{S}(\mathrm{qa})\right]^{2}\right\} \tag{7.61}
\end{equation*}
$$

Aperture antennas. Diffraction of waves


Figure 7-17 ${ }^{\text {MATLAB }}$. Cornu spiral showing the Fresnel integrals $\mathrm{C}(\mathrm{qa})$ and $\mathrm{S}(\mathrm{qa})$ as a function of the parameter pa $(-5<\mathrm{qa}<5)$. The final asymptotic values as $q a \rightarrow \pm \infty$ are $C(q a)=S(q a)= \pm 0.5$.
where

$$
\begin{equation*}
S_{0}=\frac{E_{0}^{2} \lambda}{2 Z_{0} \rho} \quad W / m^{2} \tag{7.62}
\end{equation*}
$$

is called the power density function of the incident wave at the point of observation if the obstacle were not present. The following coefficient is defined as a field diffraction coefficient D

$$
\begin{equation*}
\mathrm{D}=\frac{\mathrm{E}}{\mathrm{E}_{0}}=\sqrt{\frac{1}{2}\left\{\left[\frac{1}{2}-\mathrm{C}(\mathrm{qa})\right]^{2}+\left[\frac{1}{2}-\mathrm{S}(\mathrm{qa})\right]^{2}\right\}} \tag{7.63}
\end{equation*}
$$

## Aperture antennas. Diffraction of waves

The power diffraction coefficient is defined as the ratio of the average power at a point in the presence of a obstacle divided by the average power at the same point in the absence of this obstacle|D| $\left.\right|^{2}=S_{a v} / S_{0}$.

Example 7-14 ${ }^{\text {MATLAB }}$. Plot the power diffraction coefficient for an obstacle that occupies $1 / 2$ of the region. Compare the physical optics solution with the geometric optics solution.
Answer: Our calculations are based on equation (7.63). The result is plotted in the figure below. The physical optics solution is indicated with a solid line and the geometric optics solution is indicated with a dashed line. The region $\mathrm{qa}>0$ corresponds to the shadow region, while the region qa $<0$ corresponds to the illuminated region. It is obvious that the geometric optics (GO) solution predicts a sharp transition at the border $\mathrm{qa}=0$ while the physical optics (PO) solution predicts a smooth transition between the shadow and the illuminated regions. This has been experimentally verified. Because of the fact that $C(0)=S(0)=0$, the physical optics solution (7.63) predicts a diffraction coefficient $D(0)=0.5$, which means that the power diffraction coefficient on the border is $|D(0)|^{2}=0.25$ as can be seen in the figure below.

The constructive and destructive interference between the incident wave and the diffracted wave as calculated from Huygens' principle is responsible for the oscillations in the diffraction coefficients in the illuminated region close to the border. The power diffraction coefficient reaches a maximum value that is close to $|D|^{2} \approx 1.4$. In the shadow region, this coefficient smoothly drops to zero.

## Aperture antennas. Diffraction of waves



In the deep shadow region (qa>3), the following approximation for the diffraction coefficient is applicable

$$
\begin{equation*}
|D|^{2} \approx \frac{1}{2}\left(\frac{1}{\pi q a}\right)^{2}=\frac{\rho \lambda}{4 \pi^{2} a^{2}} \tag{7.64}
\end{equation*}
$$

This simple expression explains the fact that in the shadow region with $\mathrm{a} \rightarrow \infty$, the diffraction coefficient $\mathrm{D} \rightarrow 0$. However, for a fixed distance a, this coefficient increases with increasing distance between the observer and the screen $\rho$ and also with increasing wavelength $\lambda$.
Example 7-15. The direct ray between the transmitting antenna ( $T_{x}$ ) and the receiving antenna $\left(R_{x}\right)$ is obstructed by an obstacle. Assuming this obstacle acts as a knife edge, calculate the power attenuation $|D|^{2}(d B)$ of the diffracted ray by

## Aperture antennas. Diffraction of waves

the obstacle with respect to the direct ray at a typical mobile communication frequency of 1 GHz . Assume the obstacle is placed 5 m above the line of sight (LOS) between the $T_{x}$ and the $R_{x}$, which are separated by a distance 40 m and the obstacle lies in the middle as shown in the figure.


Answer: From the figure, we find the incident ray from $\Delta$ TOP

$$
\mathrm{TP}=\sqrt{\mathrm{TO}^{2}+\mathrm{PO}^{2}}=\sqrt{20^{2}+5^{2}}=20.62 \mathrm{~m}=\mathrm{PR},
$$

The distance in the shadow region is found from the similar triangles

$$
\begin{gathered}
\Delta \mathrm{TQR} \propto \Delta \mathrm{TOP} \Rightarrow \frac{\mathrm{QR}}{\mathrm{PO}}=\frac{\mathrm{TR}}{\mathrm{TP}} \\
\mathrm{a}=\mathrm{QR}=\frac{\mathrm{PO} \times \mathrm{TR}}{\mathrm{TP}}=\frac{5 \times 40}{20.62} \approx 9.70 \mathrm{~m}
\end{gathered}
$$

and the distance from the screen is found from $\triangle P Q R$ to be

$$
\rho=P Q=\sqrt{P R^{2}-Q R^{2}}=\sqrt{20.62^{2}-9.70^{2}} \approx 18.19 \mathrm{~m} .
$$

For the frequency $f=1 \mathrm{GHz}$, the corresponding wavelength is $\lambda=c / f=0.3 \mathrm{~m}$. In this case the characteristic parameter is $q a=\sqrt{2 a^{2} / \rho \lambda} \approx 5.87>3$. Applying (7.64), we obtain

$$
|D|^{2}=\frac{\rho \lambda}{4 \pi^{2} a^{2}}=\frac{18.19 \times 0.3}{4 \pi^{2} \times 9.70^{2}}=0.0015
$$

or

$$
|D|^{2}=-28.3 \mathrm{~dB} .
$$

Hence, the obstacle causes an attenuation of 28.3 dB in the diffracted signal compared with the direct signal. Since the actual shape of the obstacle is more complicated than a simple knife edge, the measured attenuation is usually higher and it may be more than 30 dB .

## Aperture antennas. Diffraction of waves

### 7.6.2. Slot antennas. Babinet's principle

If a slot is cut in a large flat metal sheet and a transmission line is connected to the points $A A^{\prime}$ shown in Figure $7-18 a$, this system will radiate because of the induced time varying current flowing on the sheet. The analysis of such a slot antenna is greatly facilitated by considering the slot's complimentary structure - a dipole antenna, which is shown in Figure 7-18b. The background of this duality is based on Babinet's principle first applied in optics. Let us assume that the dipole terminal impedance is $Z_{d}$ and the slot terminal impedance is $Z_{s}$. Applying this principle to electromagnetics, it was shown that there is a simple relationship between them ${ }^{2}$

$$
\begin{equation*}
Z_{s}=\frac{Z_{0}^{2}}{4 Z_{d}} \tag{7.65}
\end{equation*}
$$

(a) Slot antenna
(b)
 Flat conducting sheet Complimentary dipole antenna

Figure 7-18. (a) Slot antenna and (b) complimentary dipole antenna. Both have the same length $L$ and width $w$. They are fed by a transmission line at the terminal points $A A^{\prime}$.

[^26]
## Aperture antennas. Diffraction of waves

Example 7-16. Calculate the slot impedance of the half-wavelength slot antenna assuming $w \ll L$.

Answer: In Example 7-8, we found that the input resistance of the dipole is $R_{r}=73.1 \Omega$. To find the input reactance, another method has to be employed the result is $X_{r}=42.5 \Omega$. This means that the dipole input impedance is $Z_{d}=73.1+j 42.5 \Omega$. Substituting this value into (7.65) yields the following value for the slot input impedance

$$
Z_{s}=\frac{377^{2} / 4}{73.1+\mathrm{j} 42.5}=363.3-\mathrm{j} 211.2 \Omega .
$$

### 7.6.3. Horn and reflector antennas

Another option for the slot antenna shown in Figure 7-18a is to have it excited with a rectangular waveguide that is operating in the lowest order mode $\mathrm{TE}_{10}$ - see (5.92) with $\mathrm{n}=1$. The electric field in the termination plane that is located at $z=0$ is

$$
\begin{equation*}
E_{y}=E_{y 0} \sin \left(\frac{\pi x}{L}\right) e^{j\left(\omega t-k_{z, z}\right)} \tag{7.66}
\end{equation*}
$$

where the z-axis is normal to the flat conducting sheet and the propagation constant in this direction is obtained from (5.94)

$$
\begin{equation*}
\mathrm{k}_{\mathrm{z}, 1}=\sqrt{\left(\frac{\omega}{\mathrm{c}}\right)^{2}-\left(\frac{\pi}{\mathrm{L}}\right)^{2}} \tag{7.67}
\end{equation*}
$$

## Aperture antennas. Diffraction of waves

We can simply avoid the time-harmonic exponential factor and at the terminal plane, the exponential multiplier can be set equal to 1 . This yields a sinusoidal aperture field distribution that can be calculated using Figure 7-19

$$
\begin{equation*}
E_{y a}=E_{y 0} \sin \left(\frac{\pi x}{L}\right) \tag{7.68}
\end{equation*}
$$



Figure 7-19. Rectangular slot antenna excited by a waveguide.

In the far field, we can introduce a Fraunhofer approximation into Huygens' principle (7.55). The result of this introduction leads to the calculation of the electric field at the point $P$ being the two-dimensional Fourier transform of the distribution on the electric field in the aperture

$$
\begin{equation*}
E(P)=C \iint E_{a p}(x, y) e^{-j\left(k_{x} x+k_{y} y\right)} d x d y=C \int_{0}^{w} \int_{0}^{L} E_{y 0} \sin \left(\frac{\pi x}{L}\right) e^{-j\left(k_{x} x+k_{y} y\right)} d x d y \tag{7.69}
\end{equation*}
$$

where the transverse components of the wave vector are $k_{x}=k \sin \theta \cos \phi$ and $k_{Y}=k \sin \theta \sin \phi$. The integration of (7.69) yields the following normalized radiation pattern for the open waveguide slot antenna

$$
\begin{equation*}
F(\theta, \phi)=F_{y}(\theta, \phi) F_{x}(\theta, \phi) \tag{7.70}
\end{equation*}
$$

where

## Aperture antennas. Diffraction of waves

$$
\begin{gather*}
F_{y}(\theta, \phi)=\int_{0}^{w} e^{-j k_{Y} y} d y=\frac{\pi}{2} e^{-j \frac{k_{Y} w}{2}} \frac{\sin \left(\frac{k_{Y} w}{2}\right)}{\frac{k_{Y} w}{2}},  \tag{7.71}\\
F_{X}(\theta, \phi)=\int_{0}^{L} \sin \left(\frac{\pi x}{L}\right) e^{-j k_{x} x} d x=\frac{\pi}{2} e^{-j \frac{k_{x} L}{2}} \frac{\cos \left(\frac{k_{x} L}{2}\right)}{\left(\frac{k_{X} L}{2}\right)^{2}-\left(\frac{\pi}{2}\right)^{2}} . \tag{7.72}
\end{gather*}
$$

Finally, the amplitude of the radiation pattern for an open rectangular waveguide is

$$
\begin{equation*}
\mathrm{F}(\theta, \phi)=\left(\frac{\pi}{2}\right)^{2} \frac{\sin \left(\frac{\mathrm{kw}}{2} \sin \theta \sin \phi\right)}{\frac{\mathrm{kw}}{2} \sin \theta \sin \phi} \frac{\cos \left(\frac{\mathrm{~kL}}{2} \sin \theta \cos \phi\right)}{\left(\frac{\mathrm{kL}}{2} \sin \theta \cos \phi\right)^{2}-\left(\frac{\pi}{2}\right)^{2}} . \tag{7.73}
\end{equation*}
$$

The radiation pattern $F_{E}(\theta)$ in the plane $\phi=90^{\circ}$ that is called the E-plane is

$$
\begin{equation*}
F_{E}(\theta)=\frac{\sin \left(\frac{k w}{2} \sin \theta\right)}{\frac{k w}{2} \sin \theta} \tag{7.74}
\end{equation*}
$$

and it is shown in Figure 7-20a. The radiation pattern $\mathrm{F}_{\mathrm{H}}(\theta)$ in the plane $\phi=0^{\circ}$ that is called the H-plane results in

$$
\begin{equation*}
\mathrm{F}_{\mathrm{H}}(\theta)=\left(\frac{\pi}{2}\right)^{2} \frac{\cos \left(\frac{\mathrm{~kL}}{2} \sin \theta\right)}{\left(\frac{\pi}{2}\right)^{2}-\left(\frac{\mathrm{kL}}{2} \sin \theta\right)^{2}} \tag{7.75}
\end{equation*}
$$

and it is plotted in Figure 7-20b. These results are for the particular case $\lambda=30 \mathrm{~mm}, L=23 \mathrm{~mm}, w=10 \mathrm{~mm}$. Because of the fact that $L>w$, the $H$-field pattern is narrower than the E-field pattern.


Figure 7-20 ${ }^{\text {MATLAB }}$. (a) E-field radiation pattern; (b) H-field radiation pattern.

In order to improve the radiation pattern and to increase the directivity $D$ of this antenna, it is necessary to better match the waveguide with free space. This can be achieved by changing the shape and the size of the radiated end of the waveguide. They result is called a horn antenna that is shown in Figure 7-21. The expansion of the dimensions of the radiating aperture introduces a quadratic phase term which in turn complicates the analysis. The radiation patterns are expressed again in terms of Fresnel integrals (7.60).

## Aperture antennas. Diffraction of waves



Figure 7-21. Horn antenna.

One way to compensate for this additional phase term is to change the shape of the antenna that could lead to an antenna that is called a reflector antenna. Let us consider a parabolic reflector antenna shown in a cross section in Figure 7-22. Let us place the point source at the focal point of the parabola O . Assuming that the dimensions of the parabola are much larger than the wavelength $\lambda$, we can apply geometric optics. To achieve an equiphase excitation in the aperture plane, we have to construct a curve such that the path lengths of the different rays must be equal. This implies that the profile must be a parabola. This means that the electrical distance $O A O=O B C$ or that $2 \mathrm{~L}=\mathrm{R}+\mathrm{R} \cos \theta$, which yields

$$
\begin{equation*}
R(\theta)=\frac{2 \mathrm{~L}}{1+\cos \theta} \tag{7.76}
\end{equation*}
$$

This equation defines a parabola with a focus $O$. The parabolic reflector transforms the spherical wavefront of the point source into a plane wavefront. This equiphase excitation in the aperture plane guarantees a high directivity of the antenna.

# Aperture antennas. Diffraction of waves 



Figure 7-22. Reflector antenna.

Example 7-17 ${ }^{\text {MATLAB }}$. Find the radiation pattern of a uniformly excited circular aperture with a radius $\mathrm{a}=1 \mathrm{~m}, \lambda=10 \mathrm{~cm}$ shown in Figure 7-22.

Answer: We assume a uniform aperture excitation in order to evaluate the farfield expression (7.69). In addition, we introduce polar coordinates in the aperture plane: $x=r \cos \psi$ and $y=r \sin \psi$. This yields

$$
E(P)=C \int_{0}^{2 \pi} \int_{0}^{a} e^{-j k r \sin \theta \cos (\psi-\varphi)} r d r d \psi
$$

The integration over the variable $\theta$ is accomplished using the following identity ${ }^{3}$

$$
\int_{0}^{2 \pi} e^{-\mathrm{j} \cos \psi} \mathrm{~d} \psi=2 \pi \mathrm{~J}_{0}(\mathrm{z})
$$

[^27]
## Aperture antennas. Diffraction of waves

where $J_{0}(z)$ is a Bessel's function of the first kind and zero order. The application of the following integral identity leads to

$$
\int J_{0}(z) z d z=z J_{1}(z)
$$

where $J_{1}(z)$ is a Bessel function of the first kind and first order. Finally, the farfield pattern is obtained

$$
F(\theta)=\frac{2 J_{1}(\zeta)}{\zeta}
$$

where $\zeta=k a \sin \theta$. A plot of the radiation pattern is presented in the figure below. MATLAB has an intrinsic special function 'besselj' for the calculation of the Bessel functions of the first kind.

Here the normalized radius of the aperture is $\mathrm{ka}=2 \pi \times \mathrm{a} / \lambda=62.83$. The half-power level -3 dB is reached for the argument $\zeta=1.62$. This yields for the half beamwidth $\theta_{0}$ the result $1.62=62.83 \sin \theta_{0}$ from which we compute the half-power beam-width HPBW $=2.95^{\circ}$. The side-lobe level related to the first side-lobe is found to be SLL $=-17.57 \mathrm{~dB}$. The reflector antenna has a very high directivity. Usually, the effective aperture is close to the physical one, so $A_{e} \approx \pi a^{2}$. Then (7.44) yields for the directivity: $D=\frac{4 \pi}{\lambda^{2}} A_{e} \approx\left(\frac{2 \pi a}{\lambda}\right)^{2}=3948$ or $D \approx 36 d B$ !

Aperture antennas. Diffraction of waves


## Antenna arrays

### 7.7. Antenna arrays

In many antenna applications, we will not find a single antenna that radiates electromagnetic energy. There will be a number of antennas that will be used radiate the energy in a desired direction. We will examine one method of accomplishing this goal here. The technique to accomplish the desired radiation pattern that we will study is to set up an antenna array.

An antenna array is defined as a group of antennas that are arranged in prescribed physical configurations such as a straight line, a rectangle, a circle, etc. Each of the individual antennas is called an element of the array. We will initially assume that each of the elements that compose the members of the array are physically identical. However, the amplitude and the phase of the excitation that is applied to each individual element may differ. This is a simplification that will be useful. The far field radiation from the array in a linear media is computed from the vector addition of the components of the electromagnetic fields that are radiated from the individual elements. This is called the principle of superposition.

There are several possible configurations for an antenna array. We will initially examine an antenna array where the elements are located in a straight line. Such an array is called a linear array. To introduce the procedure, we will first examine an array that consists of two elements that are excited with the same amplitude signals but with different phases that differ by an amount $\psi$. This simple array will provide an illustration of the concept on the array.

A linear array consisting of two elements is shown in Figure 7-23. The individual element can be characterized by its element pattern $F_{1}(\theta, \phi)$ that generalizes our previous definition to the dependence on both angles in spherical coordinates.

# Antenna arrays 



Figure 7-23. A two-element linear array.
At point $P$, the total far field electric field component consists of the sum of the contributions of the two individual elements

$$
\begin{equation*}
\mathrm{E}=\mathrm{E}_{1} \mathrm{e}^{\mathrm{j} \psi / 2}+\mathrm{E}_{2} \mathrm{e}^{-\mathrm{j} \psi / 2} \tag{7.77}
\end{equation*}
$$

where $E_{1}$ is the electric field at distance $\rho_{1}$ due to source (1), $E_{2}$ is the electric field at distance $\rho_{2}$ due to source (2) and the phase difference between the fields of the two sources due to the physical separation $d$ and the different phase excitation $\delta$ is $\psi=k d \cos \theta+\delta$. The phase center is taken at the point ( 0 ) - the midpoint of the array. Since the elements are identical $\left(E_{2}=E_{1}\right)$, we obtain

$$
\begin{equation*}
E=2 E_{1}\left(\frac{e^{j \psi / 2}+e^{-j \psi / 2}}{2}\right)=2 E_{1} \cos \left(\frac{\psi}{2}\right) \tag{7.78}
\end{equation*}
$$

It can be easily shown, that the change of the choice of the phase center point (0) changes only the phase of the result but not its amplitude. The total field pattern $\mathrm{F}(\theta, \phi)$ in (7.78) can be written as a product of the radiation pattern of an individual element $F_{1}(\theta, \phi)$ and the and the radiation pattern of the array $F_{a}(\theta, \phi)$. This term is called the array factor. This product solution is called the multiplication of patterns and it will be frequently encountered in practice

$$
\begin{equation*}
F(\theta, \phi)=F_{1}(\theta, \phi) F_{a}(\theta, \phi) \tag{7.79}
\end{equation*}
$$

where

## Antenna arrays

$$
\begin{equation*}
\mathrm{F}_{\mathrm{a}}(\theta, \phi)=\cos \left(\frac{\mathrm{kdcos} \theta+\delta}{2}\right) \tag{7.80}
\end{equation*}
$$

where $\delta$ is the phase difference between the two antennas.
Note that this latter term depends on the geometry of the array and the amplitude and phase of the individual excitation signals applied to each element. In (7.77), the amplitudes were set equal, but this need not be a general requirement.

Let us illustrate the far field radiation pattern for two isotropic radiating elements that are placed along the x axis. The two elements are separated by a distance $d$ and the elements are excited with equal amplitude signals having a phase difference of $\delta$ (Fig. 7-23).

Example 7-18 MATLAB. Find and plot the array factor for three identical antennas that individually consists of two isotropic elements. The antennas are separated by 5,10 , and 20 cm and each antenna is excited in phase. The frequency of the signal applied to each antenna is 1.5 GHz .

Answer: The separation between the elements is normalized by the wavelength via

$$
\xi=\frac{\mathrm{kd}}{2}=\frac{\pi \mathrm{d}}{\lambda} .
$$

The wavelength in the free space is

$$
\lambda=\mathrm{c} / \mathrm{f}=3 \times 10^{8} / 1.5 \times 10^{9}=0.2 \mathrm{~m}=20 \mathrm{~cm}
$$

or the normalized separation $d$ is $\lambda / 4, \lambda / 2$ and $\lambda$ respectively. This yields the following values for the parameter $\xi:$ a) $\pi / 4$; b) $\pi / 2$ and c) $\pi$; and the phase difference is zero $(\delta=0)$. The corresponding array factor $F_{a}(\theta)$ is found from equation (7.80) and it is plotted in the figure below for these three cases. Because the element pattern is uniform $\left(F_{1}(\theta) \equiv 1\right)$ it follows from (7.79) that the total radiation pattern $F(\theta)=F_{a}(\theta)$. The number of the lobes that are found in the pattern increases linearly with an increase of the normalized length of the array

## Antenna arrays

and it can be approximately estimated by the nearest integer number to the real parameter $4 d / \lambda$.


Another method of altering the radiation pattern of the array is to change electronically the phase parameter $\delta$ of the applied signal - we can have the antenna "sweep" through certain regions of space. Such a structure is called a phased-array antenna. Antennas of this type are of particular importance in large radar installations where it would be mechanically impossible to rotate an antenna that may be the size of a football field.


## Antenna arrays

Figure 7-24. A uniform linear array. The current on the first element is $\mathrm{I}(\mathrm{z})$; the current on the second element is $\mathrm{I}(\mathrm{z})$ ej $\overline{\mathrm{j}}$; the current on the third element is $\mathrm{I}(\mathrm{z})$ $\mathrm{e}^{\mathrm{j} 2 \delta \text {; etc. }}$

We can extend our investigation into this type of an antenna array in several ways. A technique of extending the analysis of antenna arrays is to consider more identical elements than two as shown in Figure 7-24. In this array that is called a linear array, there is a linearly progressive phase shift in the excitation signal that feeds the identical N elements. In this case, (7.77) generalizes to

$$
\begin{equation*}
\mathrm{E}=\mathrm{E}_{0}\left[1+\mathrm{e}^{\mathrm{j} \psi}+\mathrm{e}^{\mathrm{j} 2 \psi}+\ldots+\mathrm{e}^{\mathrm{j}(\mathrm{~N}-1) \psi}\right] \tag{7.81}
\end{equation*}
$$

Fortunately, we do not have to carry along all of the terms within the square brackets since they can be summed using the relation

$$
\begin{equation*}
\sum_{n=0}^{N-1} q^{n}=\frac{1-q^{N}}{1-q} \tag{7.82}
\end{equation*}
$$

where $\mathrm{q}=\mathrm{e}^{\mathrm{j} \psi}$. Therefore, the electric field in (7.81) becomes

$$
\begin{equation*}
\mathrm{E}=\mathrm{E}_{0}\left[\frac{1-\mathrm{e}^{\mathrm{j} \psi \psi}}{1-\mathrm{e}^{\mathrm{j} \psi}}\right] \tag{7.83}
\end{equation*}
$$

If we examine only the magnitude of the electric field $|\mathrm{E}|$, we can simplify (7.83) with the relation

$$
\left|1-\mathrm{e}^{\mathrm{j} \xi}\right|=\left|2 \mathrm{je}^{\mathrm{j} \xi / 2} \sin \frac{\xi}{2}\right|=2 \sin \frac{\xi}{2}
$$

Hence, the equation (7.83) yields the magnitude of the electric field

$$
\begin{equation*}
E(\theta)=E_{0} \frac{\sin (N \psi / 2)}{\sin (\psi / 2)} \tag{7.84}
\end{equation*}
$$

where $\psi(\theta)=k d \cos \theta+\delta$ and $\delta$ is the progressive phase difference between elements. The maximum value occurs when $\psi \rightarrow 0$ and it is

$$
\begin{equation*}
\mathrm{E}_{\text {max }}=\mathrm{NE}_{0} . \tag{7.85}
\end{equation*}
$$

## Antenna arrays

In the direction of the maximum value, the condition $\psi=0$ or $k d \cos \theta=-\delta$ is satisfied. Dividing (7.84) by (7.85) yields a normalized array factor

$$
\begin{equation*}
\mathrm{F}_{\mathrm{a}}(\theta)=\frac{\sin (\mathrm{N} \psi / 2)}{N \sin (\psi / 2)} \tag{7.86}
\end{equation*}
$$

The angles where the first null in the numerator of (7.83) occur will define the main beam in the radiation pattern of the linear array - this happens for $e^{\mathrm{jN} \psi}=1$ provided $e^{j \psi} \neq 1$ or for $\psi= \pm k 2 \pi / N(k$ is an integer value). Similarly, zeroes in the denominator will yield maxima in the pattern.

Example 7-19 ${ }^{\text {MATLAB }}$. Find and plot the radiation pattern of two parallel thin halfwavelength electric dipoles separated by $d=\lambda / 2 ; \lambda$ and $3 \lambda / 2$. The mutual coupling of the dipoles is neglected.

Answer: Here the radiation pattern of this linear array is calculated by application of the multiplication property (7.79), where the radiation pattern $F_{1}(\theta)$ of a single element (half-wavelength dipole with $L=\lambda / 2$ ) is given by (7.29) and presented by the first plot in Fig. 7-10. The array factor for this case (in phase excited with $\delta=0$ ) is obtained from (7.86) for $N=2$, which yields $F_{a}(\theta)=\cos ((k d / 2) \cos \theta)$. The results for the total radiation pattern $F(\theta)$ for these three cases are plotted in the figures below for: a) $d=\lambda / 2 ; b) d=\lambda ; c) d=3 \lambda / 2$.


## Antenna arrays

In Figure $7-25$, we show the variation of $F(\theta)$ as the phase delay $\delta$ is changed in equal increments of $\pi / 4$ for a four element array ( $\mathrm{N}=4$ ). The separation of the elements $d=\lambda / 2$. Hence, we observe that the antenna radiation pattern can be altered by changing the phase even though the physical elements are not changed.


Figure $\mathbf{7 - 2 5}{ }^{\text {MATLAB }}$. Field pattern of a four element $(\mathrm{N}=4)$ phased array with the physical separation of the elements $d=\lambda / 2$ (isotropic elements).
(a) $\delta=-4 \pi / 4 ;$
(b) $\delta=-3 \pi / 4$;
(c) $\delta=-2 \pi / 4$;
(d) $\delta=-\pi / 4$; (e) $\delta=0$; (f) $d=\pi / 4$;
(g) $\delta=2 \pi / 4 ;(\mathrm{h}) \delta=3 \pi / 4 ;$ (i) $\delta=4 \pi / 4$.

## Antenna arrays

A second method would be to examine the expected behavior if there is a prescribed nonuniform excitation of the elements. For example, let us assume that we have a linear array that consists of three elements that are physically displaced by a distance $d=\lambda / 2$ and each element is excited in phase $(\delta=0)$. The excitation of the center element is twice as large as the outer two elements as shown in Figure 7-26a. The choice of this distribution of excitation amplitudes is based on the fact that $1: 2: 1$ are the leading terms of a binomial series. The resulting array, that could be generalized to include more elements, is called a binomial array.


Figure 7-26. (a) A three-element array. (b) Equivalent displaced two- element arrays.

Because of the excitation at the center element being twice the outer two elements, we can consider that this three element array is equivalent to two twoelement arrays that are displaced by a distance $N 2$ from each other. This allows us to make use of ( 7.86 ) for $\mathrm{N}=2$ where is interpreted to be the radiation pattern of this new element. The result is

$$
\begin{equation*}
F(\theta)=\cos \left(\frac{\pi}{2} \cos \theta\right) \tag{7.87}
\end{equation*}
$$

## Antenna arrays

The array factor for these new elements is the same as the radiation pattern of one of the elements. Therefore, from (7.79) we write that the magnitude of the far field radiated electric field from this structure is given by

$$
\begin{equation*}
F(\theta)=\cos ^{2}\left(\frac{\pi}{2} \cos \theta\right) \tag{7.88}
\end{equation*}
$$

The radiation pattern for this array is shown in Figure 7-27. It is contrasted with the two-element array and we note that the radiation pattern of the three-element array with a nonuniform excitation is narrower. We note that in this binomial array that there are no side lobes to absorb power. If more elements are included in the array, the beam width will become narrower.


Figure 7-27 ${ }^{\text {MATLAB }}$. Radiation patterns of a two element dipole array and a three element binomial array. (a) Element pattern. (b) Array factor. (c) Antenna array pattern.

In drawing the composite figure for the antenna array that is comprised of two small dipoles that are separated by a half-wavelength, we have multiplied the radiation pattern of the individual antenna by the array factor. In this case, the array factor is the same since this is a binomial array. The multiplication is illustrated best by working through an example.

## Antenna arrays

Example 7-20 MATLAB . Using the concept of the multiplication of patterns, find the radiation pattern of the array of four elements as shown in the figure.


Answer: This array is to be replaced with an array of two elements a sub-array of three elements (1:2:1) each. The new array will have the following individual excitations (1:3:3:1). The result according to the multiplication property (7.79) will have a radiation pattern as follows

$$
F(\theta)=\cos \left(\frac{\pi}{2} \cos \theta\right) \cos ^{2}\left(\frac{\pi}{2} \cos \theta\right)=\cos ^{3}\left(\frac{\pi}{2} \cos \theta\right) .
$$

The results are shown in the figure below: (a) the element pattern, (b) the array factor and (c) the antenna array pattern. We note that the final pattern is narrower and has no side-lobes.


Continuing the process, it is possible to obtain a pattern with arbitrary high directivity and with no side-lobes, if the amplitudes of the sources in the array correspond to the coefficients of binomial series. This means that the amplitude

## Antenna arrays

of the $\mathrm{k}^{\text {th }}$ source in the binomial array of N elements has to be calculated with the following equation

$$
\begin{equation*}
I_{k}=\frac{N!}{k!(N-k)!}(k=0,1, \ldots, N) \tag{7.89}
\end{equation*}
$$

By definition, it is clear that this array will be symmetrically excitated or $I_{N-k}=I_{k}$. The resulting radiation pattern of the binomial array of $N$ elements that are separated by a half-wavelength is

$$
\begin{equation*}
F(\theta)=\cos ^{N-1}\left(\frac{\pi}{2} \cos \theta\right) \tag{7.90}
\end{equation*}
$$

In the analysis above the mutual coupling between the elements of the antenna array was neglected. The simplest case is to consider an array of two elements, let say dipoles with lengths $L_{1}$ and $L_{2}$. The first dipole is driven by a voltage $\mathrm{V}_{1}$ while the second one is passive - see Figure 7-28. Assume the currents in both terminals are $I_{1}$ and $I_{2}$ respectively and the following circuit relations are fulfilled

$$
\begin{align*}
& Z_{11} I_{1}+Z_{12} I_{2}=V_{1} \\
& Z_{21} I_{1}+Z_{22} I_{2}=0 \tag{7.91}
\end{align*}
$$


(1)
(2)

Figure 7-28. Array of two coupled parallel dipoles.

## Antenna arrays

where $Z_{11}$ and $Z_{22}$ are the self-impedances of elements (1) and (2) and $Z_{12}=Z_{21}$ are the mutual impedances between the elements. Suppose the dipoles are equal in length $\left(L_{1}=L_{2}=L\right)$, then the self-impedances are also equal. For the particular case of thin half-wavelength dipoles as in Example 7-8, we have for the self-impedance the value $Z_{11}=73.1+\mathrm{j} 42.5 \Omega$. The dependence of the mutual impedance between two similar thin half-wavelength dipoles ${ }^{1}$ as a function of the normalized distance $d / \lambda$ is presented in Figure 7-29. Here $R_{12}$ is shown with a solid line, while $X_{12}$ is shown with a dashed line. In the limiting case of the separation distance $d \rightarrow 0$, the mutual impedance approaches the selfimpedance, which is to be expected.

Example 7-21. Find the current in the one of the dipoles in an antenna array if there is a) no mutual coupling between the dipoles and b) there is a mutual coupling between the dipoles. The driving voltage is $\mathrm{V}_{1}=100 \mathrm{~V}$ and the distance between the two half-wave dipoles is $d=\lambda / 2$. Estimate the amplitude and the phase change.

Answer: a) For the case with no mutual coupling, we can apply (7.91) with $Z_{12}=0$. This results in

$$
I_{1}=V_{1} / Z_{11}=1.183 e^{-j 30.17^{0}} \mathrm{~A}
$$

b) For the case with mutual coupling, we obtain from Figure 7-29 that $Z_{12}=-12.5-j 29.9 \Omega$. The solution of (7.91) yields the current

$$
\mathrm{I}_{1}=1.218 \mathrm{e}^{-\mathrm{j} 21.80^{\circ}} \mathrm{A}
$$

The amplitude of the input current changes slightly while the change in phase is more significant.

[^28]
## Antenna arrays



Figure 7-29 MATLAB . Mutual impedance between two parallel thin half-wavelength dipoles.

One may wish to find the directivity of one antenna array. Applying (7.35) for the particular case of $\mathrm{N}=2 \mathrm{M}+1$ identical elements separated by a distance $d=\lambda / 2$, the following equation is obtained

$$
\begin{equation*}
D=\frac{\left(\sum_{n=-M}^{M} I_{n}\right)^{2}}{\sum_{n=-M}^{M} I_{n}^{2}} \tag{7.92}
\end{equation*}
$$

which means that the directivity measures the degree of the coherence of the total electric field.

## Antenna arrays

Example 7-22. Compare the directivities of two arrays consisting of three identical elements separated by a half-wavelength for the following two cases: a) uniform array: $\left(I_{-1}=I_{0}=I_{1}=1 \mathrm{~A}\right)$; b) binomial array: $\left(I_{-1}=I_{1}=1 \mathrm{~A} ; I_{0}=2 \mathrm{~A}\right)$.

Answer: From (7.92), we compute
a) uniform array $D=\frac{(1+1+1)^{2}}{1+1+1}=9 \Rightarrow D=9.5 d B$;
b) binomial array: $D=\frac{(1+2+1)^{2}}{1+4+1}=\frac{16}{6}=2.667 \Rightarrow D=4.3 \mathrm{~dB}$

The directivity of a uniform array is greater than that of a binomial array.

The uniform array and the binomial array are just two examples of linear antenna arrays: the first one has a narrow half-power beamwidth (HPBW), but a relatively high side-lobe level (SLL), while the second one has a relatively wide HPBW, but a very low SLL. The designer of real antenna arrays has to make a tradeoff between these two important characteristics: 1) narrow HPBW and 2) low SLL. There are two different approaches for solving this optimization problem, such as: 1) choosing special excitation (i.e. Chebyshev array); 2) choosing non-uniform distance between the elements. However, these methods are topics for more advanced investigations.

## Conclusion

### 7.8. Conclusion

The fundamental radiation characteristics of electromagnetic waves and its relationship to accelerating and decelerating charges have been reviewed. Using these concepts, the small Hertzian dipole radiator was described. A radiation pattern for such an antenna was obtained.

A formal procedure using the concept of the vector potential was introduced and applied to several antennas. We focused our attention to an the examination of the far field properties of antennas. If the media in which the waves are propagating is linear, the principle of superposition applies. Constructive and destructive interference between fields radiated by displaced antenna elements with differing phases in the applied currents led to different radiation or reception characteristics of an antenna. Terms such as beam width, main and side lobes, radiation resistance, gain, directivity, and effective area were defined and applied.

The solution that we have obtained for the radiation patterns was predicated on assuming a valid approximation for the current distribution on the antenna. Several distributions were analyzed. The method of moments which was introduced in Chapter 3 can be equally well applied to antenna calculations. In this case, the field distribution is known from the experimental measurements and the current distribution becomes the unknown term that must be ascertained.

The subject of an antenna array that consisted of several identical antennas was introduced. By controlling either the phase or the amplitude of the signal that was applied to each individual antenna element or its spatial separation, we found that the resulting radiation pattern could be changed. We found that predictions of the radiation pattern could be found by multiplying the radiation pattern of an individual antenna times an array factor in order to find the radiation pattern of the entire antenna array.

# Problems 7 

## 1. Chapter 7: Problems

1. Perform the integration of the integral
2. 

$$
\int_{0}^{\infty} \frac{R^{3} d R}{\left(R^{2}+x^{2}\right)^{3}}
$$

3. that arises in Example 7-1.
4. Using dimensional arguments, show that the term
5. 

$$
\frac{\mathrm{Q}^{2}}{32 \pi \varepsilon_{0} \mathrm{x}_{0}} \quad\left(\mathrm{x}_{0}>0\right)
$$

5. corresponds to the electrostatic energy stored in the region $x>x_{0}$ in Example
6. 7-1.
7. $3^{\text {MATLAB }}$. A short electric dipole with a length $L(L \ll \lambda)$ is located above a ground plane a distance $\mathrm{h}=\lambda / 4(\mathrm{~h} \gg \mathrm{~L})$. The dipole is perpendicular to the ground plane. Find the directivity of this antenna. How does the ground plane change the directivity? Hint: use the method of images and the array concept, then apply a numerical integration.
8. $4^{\text {MATLAB }}$. Find the directivity of two short electric dipoles that are excited out of phase and are physically separated by a quarter-wavelength.
9. 5. A short magnetic dipole with a radius a $(\mathrm{a} \ll \lambda)$ is located above a ground plane a distance $\mathrm{h}=\lambda / 4$ ( $\mathrm{h} \gg \mathrm{a}$ ). The dipole is parallel to the ground plane. Find the directivity of this antenna. How does the ground plane change the directivity?
1. Find the directivity of two short magnetic dipoles that are excited in phase and are physically separated by a quarter-wavelength.

## Problems 7

7. Find the radiation resistance of a short electric dipole with a length L $(\mathrm{L} \ll \lambda)$ that is located a distance $\mathrm{h}=\lambda / 4$ above the ground.
8. Find the radiation resistance of a short magnetic dipole with a length $L$ $(\mathrm{L} \ll \lambda)$ that is located a distance $h=\lambda / 4$ above the ground.
$10.9^{\text {MATLAB }}$. Find the directivity $D$ of a half-wavelength electric dipole.
$11.10^{\text {MATLAB }}$. Find the radiation resistance $R_{r}$ of a half-wavelength electric dipole.
9. A thin vertical quarter-wavelength monopole is located above a horizontal conducting ground. Find the radiation resistance of the antenna.
10. $12^{\text {MATLAB }}$. A half-wavelength dipole is located near a corner reflector with a flare angle $\psi=90^{\circ}$ as shown in the figure. Find the radiation pattern $\mathrm{F}(\theta)$ for the case $s=\lambda / 2, \phi=0^{0}$.
11. 


15.13. Find the current required to radiate a power of 100 W at 1 GHz from a 15cm dipole.
14. A Hertzian dipole of length $\mathrm{L}=0.2 \mathrm{~m}$ operates at 10 MHz . Find the radiation efficiency $\eta_{r}$ if the copper conductor has the following parameters:

## Problems 7

$\sigma=5.8 \times 10^{7} \mathrm{~S} / \mathrm{m}, \quad \mu_{r}=1$ and radius $a=1 \mathrm{~mm}$. By definition, $\eta_{r}=R_{r} /\left(R_{r}+R_{L}\right)$ where $R_{r}$ is the radiation resistance and $R_{L}$ is the ohmic resistance.
15. Calculate the effective aperture $A_{e}$ of the Hertzian dipole.
16. Calculate the effective aperture $A_{e}$ of the half-wavelength dipole.
17. Find the HPBW (half-power beam width) and the SLL (side-lobe level) of an electric dipole with a length $L=1.5 \lambda$.
16.18 MATLAB . Find the directivity of a reflector antenna that is considered in Example 7-17 (radius $\mathrm{a}=1 \mathrm{~m}$, wavelength $\lambda=0.1 \mathrm{~m}$ ).
18. Find the aperture efficiency $\eta_{a p}$ of the reflector antenna described in Problem 18. By definition, this parameter is $\eta_{a p}=A_{e} / A_{p}$ where $A_{e}$ is the effective aperture and $A_{p}=\pi a^{2}$ is the physical aperture.
19. Consider an antenna array that is excited in-phase ( $\delta=0^{\circ}$ ) with N elements ( $\mathrm{N} \gg 1$ ). Each element of the array is separated by a distance d. Show that the direction of the main maximum is $\theta_{0}=90^{\circ}$ (broadside array). Find an approximate expression for the HPBW.
20. Consider an antenna array consisting of $N$ elements with each of the elements excited with a phase-delay $\left(\delta \neq 0^{\circ}\right)$. Each element of the array is separated by a distance $d$. Show that it is possible to have the direction on the main maximum to be $\theta_{0}=0^{\circ}$. This is called an end-fire array. Find an approximate expression for the HPBW.
$17.22^{\text {MATLAB }}$. Plot the radiation pattern $\mathrm{F}(\phi)$ of two parallel half-wave dipoles that are separated by a distance $d=\lambda / 4$ and are excited by currents that have a phase difference $\delta=-90^{\circ}$.
18.23. Calculate the input impedance for an array of two half-wave parallel dipoles that are separated by a distance $d=\lambda / 2$ (Figure 7-28). The currents are $I_{2}=-I_{1}$.

## Problems 7

19.24. Calculate the input impedance for an array of two parallel half-wave dipoles that are separated by a distance $d=\lambda / 2$. The currents are $I_{2}=+l_{1}$.
20.25. A half-wave dipole with a terminal current $I_{0}$ is placed a distance $s=0.1 \lambda$ from a perfectly conducting xy-plane as shown in the figure.
21.

22. Neglecting ohmic losses, compare the terminal currents with and without the reflector if the radiated power is 2 W .
26. A TV broadcasting station (Tx) radiates a power of 500 W from an antenna on a 100 m tower above a perfectly conducting ground. plane The antenna is omnidirectional in the horizontal plane, but has a HPBW $=20^{\circ}$ in the vertical plane. If the wavelength is 1 m , what is the optimum height for this antenna at a location of 2 km from the antenna for: a) vertical polarization; b) horizontal polarization? What is the received power in the receiver $(R x)$ for the both cases if the receiving antenna is a half-wave dipole?

## Problems 7

23. 


$24.27^{\text {MATLAB }}$. Assume that there are 3 identical antennas that are separated by a distance $d=\lambda / 4$ along a straight line. Each antenna is fed with the same current but there is a uniform progressive phase shift $\delta$ along the line. Find and plot the array factor in three cases: a) $\delta=0$; b) $\delta=\pi / 2$; c) $\delta=\pi$.
$25.28^{\text {MATLAB }}$. Repeat Problem 27 with a larger separation $d=\lambda / 2$ for the same phase differences.
26.29. Find an expression for the radar cross section $\sigma_{s}$ of a target. By definition, $\sigma_{\mathrm{S}}=\mathrm{P}_{\mathrm{S}} / \mathrm{S}_{\mathrm{i}}$ where $\mathrm{P}_{\mathrm{s}}$ is the back scattered power and $\mathrm{S}_{\mathrm{i}}$ is the incident time-average power density.
27.30. Derive the radar equation, expressing the power received by a monostatic radar $\mathrm{P}_{\mathrm{r}}$ in terms of the transmitted power from the same antenna $P_{t}$. Here the antenna gain is $G$ and the operating wavelength is $\lambda$. The distance between the antenna and the target is R and the radar cross section of the target is $\sigma_{s}$.

## Appendix 1: Mathematical formulas

## A.1.1. Vector identities

$\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are vectors and $a$ and $b$ are scalars.

$$
\begin{align*}
& (\mathbf{A} \times \mathbf{B}) \bullet \mathbf{C}=(\mathbf{B} \times \mathbf{C}) \bullet \mathbf{A}=(\mathbf{C} \times \mathbf{A}) \bullet \mathbf{B}  \tag{A.1.1}\\
& \mathbf{A} \times(\mathbf{B} \times \mathbf{C})=\mathbf{B}(\mathbf{A} \bullet \mathbf{C})-\mathbf{C}(\mathbf{A} \bullet \mathbf{B})  \tag{A.1.2}\\
& \nabla \bullet(\mathbf{A}+\mathbf{B})=\nabla \bullet \mathbf{A}+\nabla \bullet \mathbf{B}  \tag{A.1.3}\\
& \nabla(\mathrm{a}+\mathrm{b})=\nabla \mathrm{a}+\nabla \mathrm{b}  \tag{A.1.4}\\
& \nabla \times(\mathbf{A}+\mathbf{B})=\nabla \times \mathbf{A}+\nabla \times \mathbf{B}  \tag{A.1.5}\\
& \nabla \bullet(\mathrm{aB})=\mathbf{B} \bullet \nabla \mathrm{a}+\mathrm{a} \nabla \bullet \mathbf{B}  \tag{A.1.6}\\
& \nabla(\mathrm{ab})=\mathrm{a} \nabla \mathrm{~b}+\mathrm{b} \nabla \mathrm{a}  \tag{A.1.7}\\
& \nabla \times(\mathrm{aB})=\nabla \mathbf{a} \times \mathbf{B}+\mathrm{a} \nabla \times \mathbf{B}  \tag{A.1.8}\\
& \nabla \bullet(\mathbf{A} \times \mathbf{B})=\mathbf{B} \bullet \nabla \times \mathbf{A}-\mathbf{A} \bullet \nabla \times \mathbf{B}  \tag{A.1.9}\\
& \nabla(\mathbf{A} \bullet \mathbf{B})=(\mathbf{A} \bullet \nabla) \mathbf{B}+(\mathbf{B} \bullet \nabla) \mathbf{A}+\mathbf{A} \times(\nabla \times \mathbf{B})+\mathbf{B} \times(\nabla \times \mathbf{A})  \tag{A.1.10}\\
& \nabla \times(\mathbf{A} \times \mathbf{B})=\mathbf{A} \nabla \bullet \mathbf{B}-\mathbf{B} \nabla \bullet \mathbf{A}+(\mathbf{B} \bullet \nabla) \mathbf{A}-(\mathbf{A} \bullet \nabla) \mathbf{B}  \tag{A.1.11}\\
& \nabla \bullet \nabla \mathbf{a}=\nabla^{2} \mathbf{a}  \tag{A.1.12}\\
& \nabla \bullet \nabla \times \mathbf{A}=0  \tag{A.1.13}\\
& \nabla \times \nabla \mathbf{a}=0  \tag{A.1.14}\\
& \nabla \times \nabla \times \mathbf{A}=\nabla(\nabla \bullet \mathbf{A})-\nabla^{2} \mathbf{A} \tag{A.1.15}
\end{align*}
$$

## Mathematical formulas

## A.1.2. Vector operations in the three coordinate systems

## Cartesian

$$
\begin{align*}
& \nabla \mathbf{a}=\frac{\partial \mathbf{a}}{\partial \mathbf{x}} \mathbf{u}_{\mathbf{x}}+\frac{\partial \mathbf{a}}{\partial \mathbf{y}} \mathbf{u}_{\mathbf{y}}+\frac{\partial \mathbf{a}}{\partial \mathbf{z}} \mathbf{u}_{\mathbf{z}}  \tag{A.1.16}\\
& \nabla \bullet \mathbf{A}=\frac{\partial \mathbf{A}_{x}}{\partial \mathbf{x}}+\frac{\partial \mathbf{A}_{\mathrm{y}}}{\partial \mathbf{y}}+\frac{\partial \mathbf{A}_{z}}{\partial \mathbf{z}}  \tag{A.1.17}\\
& \nabla \times \mathbf{A}=\left(\frac{\partial \mathbf{A}_{z}}{\partial \mathbf{y}}-\frac{\partial \mathbf{A}_{y}}{\partial \mathbf{z}}\right) \mathbf{u}_{\mathrm{x}}+\left(\frac{\partial \mathbf{A}_{x}}{\partial \mathbf{z}}-\frac{\partial \mathrm{A}_{z}}{\partial \mathbf{x}}\right) \mathbf{u}_{\mathbf{y}}+\left(\frac{\partial \mathbf{A}_{y}}{\partial \mathbf{x}}-\frac{\partial A_{x}}{\partial \mathbf{y}}\right) \mathbf{u}_{\mathbf{z}}  \tag{A.1.18}\\
& \nabla^{2} \mathbf{a}=\frac{\partial^{2} \mathbf{a}}{\partial \mathbf{x}^{2}}+\frac{\partial^{2} \mathbf{a}}{\partial \mathbf{y}^{2}}+\frac{\partial^{2} \mathbf{a}}{\partial \mathbf{z}^{2}} \tag{A.1.19}
\end{align*}
$$

## Cylindrical

$\nabla \mathrm{a}=\frac{\partial \mathrm{a}}{\partial \mathrm{r}} \mathbf{u}_{\mathrm{r}}+\frac{1}{\mathrm{r}} \frac{\partial \mathrm{a}}{\partial \phi} \mathbf{u}_{\phi}+\frac{\partial \mathrm{a}}{\partial \mathrm{z}} \mathbf{u}_{\mathbf{z}}$
$\nabla \bullet \mathbf{A}=\frac{1}{r} \frac{\partial\left(r A_{r}\right)}{\partial r}+\frac{1}{r} \frac{\partial A_{\phi}}{\partial \phi}+\frac{\partial A_{z}}{\partial z}$
$\nabla \times \mathbf{A}=\left(\frac{1}{r} \frac{\partial A_{z}}{\partial \phi}-\frac{\partial A_{\phi}}{\partial \mathbf{z}}\right) \mathbf{u}_{\mathrm{r}}+\left(\frac{\partial \mathrm{A}_{\mathrm{r}}}{\partial \mathbf{z}}-\frac{\partial \mathrm{A}_{\mathbf{z}}}{\partial \mathrm{r}}\right) \mathbf{u}_{\phi}+\frac{1}{\mathrm{r}}\left(\frac{\partial\left(\mathrm{r} \mathrm{A}_{\phi}\right)}{\partial \mathrm{r}}-\frac{\partial \mathrm{A}_{\mathrm{r}}}{\partial \phi}\right) \mathbf{u}_{\mathbf{z}}$
$\nabla^{2} \mathbf{a}=\frac{1}{r} \frac{\partial\left(r \frac{\partial a}{\partial r}\right)}{\partial r}+\frac{1}{r^{2}} \frac{\partial^{2} \mathbf{a}}{\partial \phi^{2}}+\frac{\partial^{2} \mathbf{a}}{\partial \mathbf{z}^{2}}$

## Spherical

$\nabla \mathrm{a}=\frac{\partial \mathrm{a}}{\partial \rho} \mathbf{u}_{\rho}+\frac{1}{\rho} \frac{\partial \mathrm{a}}{\partial \theta} \mathbf{u}_{\theta}+\frac{1}{\rho \sin \theta} \frac{\partial \mathbf{a}}{\partial \phi} \mathbf{u}_{\phi}$
$\nabla \bullet \mathbf{A}=\frac{1}{\rho^{2}} \frac{\partial\left(\rho^{2} A_{\rho}\right)}{\partial \rho}+\frac{1}{\rho \sin \theta} \frac{\partial\left(A_{\theta} \sin \theta\right)}{\partial \theta}+\frac{1}{\rho \sin \theta} \frac{\partial A_{\phi}}{\partial \phi}$
$\nabla \times \mathbf{A}=\frac{1}{\rho \sin \theta}\left(\frac{\partial\left(\sin \theta \mathrm{~A}_{\phi}\right)}{\partial \theta}-\frac{\partial \mathrm{A}_{\theta}}{\partial \phi}\right) \mathbf{u}_{\rho}+\frac{1}{\rho}\left(\frac{1}{\sin \theta} \frac{\partial \mathrm{~A}_{\rho}}{\partial \phi}-\frac{\partial\left(\rho \mathrm{A}_{\phi}\right)}{\partial \rho}\right) \mathbf{u}_{\theta}+$ $+\frac{1}{\rho}\left(\frac{\partial\left(\rho \mathrm{~A}_{\theta}\right)}{\partial \rho}-\frac{\partial \mathrm{A}_{\rho}}{\partial \theta}\right) \mathbf{u}_{\phi}$
$\nabla^{2} \mathbf{a}=\frac{1}{\rho^{2}} \frac{\partial\left(\rho^{2} \frac{\partial \mathbf{a}}{\partial \rho}\right)}{\partial \rho}+\frac{1}{\rho^{2} \sin \theta} \frac{\partial\left(\sin \theta \frac{\partial \mathbf{a}}{\partial \theta}\right)}{\partial \theta}+\frac{1}{\rho^{2} \sin ^{2} \theta} \frac{\partial^{2} \mathbf{a}}{\partial \phi^{2}}$

## Mathematical formulas

## A.1.3. Summary of the transformations between coordinate

## systems

## Cartesian-cylindrical



$$
\begin{gather*}
\text { Cartesian-spherical }  \tag{A.1.29}\\
\left\{\begin{array} { l } 
{ x = \rho \operatorname { s i n } \theta \operatorname { c o s } \phi } \\
{ y = \rho \operatorname { s i n } \theta \operatorname { s i n } \phi } \\
{ z = \rho \operatorname { c o s } \theta }
\end{array} \left\{\begin{array}{l}
\rho=\sqrt{x^{2}+y^{2}+z^{2}} \\
\theta=\tan ^{-1}\left(\frac{\sqrt{x^{2}+y^{2}}}{z}\right) \\
\phi=\tan ^{-1}\left(\frac{y}{x}\right)
\end{array}\right.\right.
\end{gather*}
$$

|  | $\mathbf{u}_{\boldsymbol{\rho}}$ | $\mathbf{u}_{\boldsymbol{\theta}}$ | $\mathbf{u}_{\boldsymbol{\varphi}}$ |
| :---: | :---: | :---: | :---: |
| $\mathbf{u}_{\mathbf{x}}$ | $\sin \theta \cos \phi$ | $\cos \theta \cos \phi$ | $-\sin \phi$ |
| $\mathbf{u}_{\mathrm{Y}}$ | $\sin \theta \sin \phi$ | $\cos \theta \sin \phi$ | $\cos \phi$ |
| $\mathbf{u}_{\mathbf{Z}}$ | $\cos \theta$ | $-\sin \theta$ | 0 |

Distance $R=\left|r_{2}-r_{1}\right|$ in the three coordinate systems.

1. Cartesian: $R=\left[\left(x_{2}-x_{1}\right)^{2}+\left(y_{2}-y_{1}\right)^{2}+\left(z_{2}-z_{1}\right)^{2}\right]^{1 / 2}$
2. cylindrical: $R=\left[r_{2}^{2}+r_{1}^{2}-2 r_{2} r_{1} \cos \left(\phi_{2}-\phi_{1}\right)+\left(z_{2}-z_{1}\right)^{2}\right]^{1 / 2}$
3. spherical: $R=\left\{\rho_{2}^{2}+\rho_{1}^{2}-2 \rho_{2} \rho_{1}\left[\cos \theta_{2} \cos \theta_{1}+\sin \theta_{2} \sin \theta_{1} \cos \left(\phi_{2}-\phi_{1}\right)\right]\right\}^{1 / 2}$

## Mathematical formulas

## A.1.4. Integral relations

$$
\begin{gather*}
\int_{\Delta V} \nabla \cdot \mathbf{A d v}=\oint \mathbf{A} \bullet \mathbf{d s} \text { [divergence theorem] }  \tag{A.1.32}\\
\int_{\Delta S} \nabla \times \mathbf{A} \cdot \mathbf{d s}=\oint \mathbf{A} \cdot \mathbf{d l} \text { [Stokes's theorem] }  \tag{A.1.33}\\
\int_{\Delta V} \nabla \times \mathbf{A d v}=-\oint \mathbf{A} \times \mathbf{d s}  \tag{A.1.34}\\
\int_{\Delta V} \nabla \mathbf{a d v}=\oint \mathbf{a d s}  \tag{A.1.35}\\
\int_{\Delta S} \nabla \mathbf{a} \times \mathbf{d s}=-\oint \mathrm{adl} \tag{A.1.36}
\end{gather*}
$$

| Coordinate system | Cartesian $(x, y, z)$ | $\begin{aligned} & \text { Cylindrical } \\ & (r, \phi, z) \end{aligned}$ | Spherical $(\rho, \theta, \phi)$ |
| :---: | :---: | :---: | :---: |
| Unit vectors | $\mathbf{u}_{\mathrm{x}} \mathrm{u}_{\mathbf{y}} \mathrm{u}_{\mathbf{z}}$ | $\mathbf{u}_{\mathbf{r}} \mathbf{u}_{\phi} \mathbf{u}_{\mathbf{z}}$ | $\mathbf{u}_{\rho} \mathbf{u}_{\theta} \mathbf{u}_{\phi}$ |
| Differential length dl | $d x \mathbf{u}_{\mathrm{x}}$ <br> dy $\mathbf{u}_{\mathbf{y}}$ <br> $\mathrm{dz} \mathbf{u}_{\mathbf{z}}$ | $\mathrm{dr} \mathbf{u}_{\mathrm{r}}$ <br> $r d \phi \mathbf{u}_{\phi}$ <br> $d z \mathbf{u}_{\mathbf{z}}$ | $\begin{aligned} & \mathrm{d} \rho \mathbf{u}_{\rho} \\ & \rho \sin \theta \mathrm{d} \theta \mathbf{u}_{\theta} \\ & \rho \mathrm{d} \phi \mathbf{u}_{\phi} \end{aligned}$ |
| Differential surface area ds | $d y d z \mathbf{u}_{\mathrm{x}}$ <br> $d x d z \mathbf{u}_{\mathrm{y}}$ <br> $d x d y u_{z}$ | $r d \phi d z \mathbf{u}_{\mathbf{r}}$ <br> $\mathrm{dr} \mathrm{dz} \mathbf{u}_{\phi}$ <br> $r \operatorname{dr} d \phi \mathbf{u}_{\mathbf{z}}$ | $\begin{aligned} & \rho^{2} \sin \theta d \theta d \phi \mathbf{u}_{\rho} \\ & \rho \sin \theta d \rho d \phi \mathbf{u}_{\theta} \\ & \rho d \rho d \theta \mathbf{u}_{\phi} \end{aligned}$ |
| Differential volume dv | $d x d y d z$ | $\mathrm{rdr} \mathrm{d} \phi \mathrm{dz}$ | $\rho^{2} \sin \theta d \rho d \theta d \phi$ |

# Appendix 2: Mathematical foundation of the <br> Finite element method 

## A.2.1. Minimum energy condition

Let us assume that $\mathrm{V}(\mathrm{x}, \mathrm{y})$ is the true solution of Laplace's equation. In addition, let us assume that $\mathrm{U}(\mathrm{x}, \mathrm{y})$ is another function that can be differentiated and is equal to zero on the boundary $L_{1}$ of the region $s$. Then, the sum of the two solutions which we will call the variation is given by $V(x, y)+\alpha U(x, y)$ where $\alpha$ is a small real parameter. The variation will have the same value on the boundary $L_{1}$ as $\mathrm{V}(\mathrm{x}, \mathrm{y})$. The electrostatic energy of the summation of the two terms is obtained from the expression (3.63)

$$
\begin{equation*}
\mathrm{W}(\mathrm{~V}+\alpha \mathrm{U})=\int_{\mathrm{A}} \frac{\varepsilon}{2}|\nabla(\mathrm{~V}+\alpha \mathrm{U})|^{2} \mathrm{ds} \tag{A.2.1}
\end{equation*}
$$

This functional-energy $W(V+\alpha U)$ is expanded in powers of the small parameter $\alpha$

$$
\begin{equation*}
\mathrm{W}(\mathrm{~V}+\alpha \mathrm{U})=\mathrm{W}(\mathrm{~V})+\alpha \varepsilon \int_{\mathrm{A}} \nabla \mathrm{~V} \bullet \nabla \mathrm{Ud} s+\frac{\alpha^{2}}{2} \varepsilon \int_{A}|\nabla \mathrm{U}|^{2} \mathrm{ds} \tag{A.2.2}
\end{equation*}
$$

The third term on the right hand side can be identified from (3.63) as being the energy of the additional functions $U$. The second term on the right hand side can be transformed using the vector identity (A.1.6)

$$
\begin{equation*}
\nabla \bullet(\mathrm{U} \nabla \mathrm{~V})=\nabla \mathrm{V} \bullet \nabla \mathrm{U}+\mathrm{U} \nabla \bullet \nabla \mathrm{~V} \tag{A.2.3}
\end{equation*}
$$

where $U$ is a scalar and $\nabla V$ is a vector. The last term in (A.2.3) can be written as $U \nabla^{2} V$. Therefore, we finally write (A.2.2) as

## Mathematical foundation of FEM

$$
\begin{equation*}
W(V+\alpha U)=W(V)+\alpha^{2} W(U)-\alpha \varepsilon \int_{A} U \nabla^{2} V d s+\alpha \varepsilon \oint_{L} U \nabla V \cdot u_{n} d l \tag{A.2.4}
\end{equation*}
$$

The term $\nabla V \bullet \mathbf{u}_{n}=\frac{\partial V}{\partial n}$. From Figure 3-13a, we have $U=0$ on $L_{1}$ and $\frac{\partial V}{\partial n}=0$ on $L_{2}$. This means that the fourth term in (A.2.4), which is a line integral vanishes on the entire boundary L. The third term in (A.2.4), which is a surface integral on A, also vanishes since Laplace's equation $\nabla^{2} V=0$ must be satisfied. Therefore, we finally obtain

$$
\begin{equation*}
W(V+\alpha U)=W(V)+\alpha^{2} W(U) \tag{A.2.5}
\end{equation*}
$$

Since $\alpha^{2}>0$ and $W(U)>0$, the term on the left hand side is greater than $W(V)$ which proves that the energy has a minimum when V is a solution of Laplace's equation.

## A.2.2. Linear interpolation coefficients

We calculate $\alpha_{1}\left(\mathrm{x}_{1}, \mathrm{y}_{1}\right)$ from (3.69)

$$
\begin{equation*}
\alpha_{1}\left(x_{1}, y_{1}\right)=\frac{1}{2 A_{e}}\left[\left(x_{2} y_{3}-x_{3} y_{2}\right)+\left(y_{2}-y_{3}\right) x_{1}+\left(x_{3}-x_{2}\right) y_{1}\right] \tag{A.2.6}
\end{equation*}
$$

However, the expression in the brackets is equal to twice the area of the triangle $A_{e}$. Therefore, this term is equal to 1 . Let us calculate the same term at a different point, say $\alpha_{1}\left(\mathrm{x}_{2}, \mathrm{y}_{2}\right)$ using (3.69)

$$
\begin{equation*}
\alpha_{1}\left(x_{2}, y_{2}\right)=\frac{1}{2 A_{e}}\left[\left(x_{2} y_{3}-x_{3} y_{2}\right)+\left(y_{2}-y_{3}\right) x_{2}+\left(x_{3}-x_{2}\right) y_{2}\right]=0 \tag{A.2.7}
\end{equation*}
$$

These results are summarized in (3.70).

## Mathematical foundation of FEM

## A.2.3. S-matrix elements

Using the explicit expression for the $\alpha$-functions (3.69), we find that the gradient of these functions are

$$
\begin{align*}
& \nabla \bullet \alpha_{1}=\frac{1}{A_{e}}\left[\left(y_{2}-y_{3}\right) u_{x}-\left(x_{2}-x_{3}\right) u_{\mathrm{Y}}\right] \\
& \nabla \bullet \alpha_{2}=\frac{1}{A_{e}}\left[\left(y_{3}-y_{1}\right) u_{x}-\left(x_{3}-x_{1}\right) u_{\mathrm{r}}\right]  \tag{A.2.8}\\
& \nabla \bullet \alpha_{3}=\frac{1}{A_{e}}\left[\left(y_{1}-y_{2}\right) u_{x}-\left(x_{1}-x_{2}\right) u_{\mathrm{r}}\right]
\end{align*}
$$

Applying the definition (3.73) leads to the following elements of the $S^{(e)}$-matrix

$$
\begin{gather*}
\mathrm{S}_{1,1}=\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{2}-\mathrm{y}_{3}\right)^{2}+\left(\mathrm{x}_{2}-\mathrm{x}_{3}\right)^{2}\right] ;  \tag{A.2.9}\\
\mathrm{S}_{1,2}=\frac{\varepsilon}{4 \mathrm{~A}_{\mathrm{e}}}\left[\left(\mathrm{y}_{2}-\mathrm{y}_{3}\right)\left(\mathrm{y}_{3}-\mathrm{y}_{1}\right)+\left(\mathrm{x}_{2}-\mathrm{x}_{3}\right)\left(\mathrm{x}_{3}-\mathrm{x}_{1}\right)\right] \text { etc. } \tag{A.2.10}
\end{gather*}
$$

The other matrix-elements (diagonal and off-diagonal) can be found from these two expressions by applying suitable cyclic change of the subscripts: $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$.

## A.2.4. Decoupled and coupled node potentials

The coupling matrix $[C]$ has a dimension $\left(N_{d} \times N\right)$ where $N_{d}$ is the total number of the decoupled nodes (here $\mathrm{N}_{\mathrm{d}}=2 \times 3=6$ ) and N is the total number of the coupled nodes after assembling (here $N=3+1=4$ ). The decoupled nodes are numbered with a subscript $\mathrm{k}=1,2, \ldots, \mathrm{~N}_{\mathrm{d}}$, while the coupled nodes are numbered with a subscript $\mathrm{I}=1,2, \ldots, \mathrm{~N}$. The elements of the coupling matrix are
defined by the following rule: $\mathrm{C}_{\mathrm{k}, \mathrm{l}}=1$ when the k -th decoupled node corresponds to the l-th coupled node and $C_{k, 1}=0$ otherwise. There are only one " 1 " on every row, but there can be one or two " 0 " on every column (the last case reflects the boundary conditions). For the particular problem considered in Fig. 3-15, the corresponding C-matrix is

$$
[C]=\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0
\end{array}\right]
$$

Now the following relation between the column matrices of the decoupled potentials $[\mathrm{V}]_{\mathrm{d}}$ and coupled potentials $[\mathrm{V}$ ] is obtained

$$
\begin{equation*}
[\mathrm{V}]_{\mathrm{d}}=[\mathrm{C}][\mathrm{V}] \tag{A.2.11}
\end{equation*}
$$

where the coupling matrix [C] is involved.
After the determination of the energy of the element from (3.74) via the potentials of the decoupled nodes and via the potentials of the coupled nodes as

$$
\begin{equation*}
\mathrm{W}=\frac{1}{2}[\mathrm{~V}]^{\top}[\mathrm{S}][\mathrm{V}] \tag{A.2.12}
\end{equation*}
$$

we obtain for the global S-matrix of the coupled system using the following relation

$$
\begin{equation*}
[\mathrm{S}]=[\mathrm{C}]^{\top}[\mathrm{S}]_{\mathrm{d}}[\mathrm{C}] \tag{A.2.13}
\end{equation*}
$$

We can show that the last equation yields equation (3.77) for the global S-matrix in explicit form. For convenience we can write this equation in two stages

$$
\begin{align*}
& {[\mathrm{B}]=[\mathrm{S}]_{\mathrm{d}}^{[\mathrm{C}],}} \\
& {[\mathrm{S}]=[\mathrm{C}]^{\top}[\mathrm{B}]} \tag{A.2.14}
\end{align*}
$$

From (3.76) for the decoupled matrix $[\mathrm{S}]_{d}$ of the system of two triangles, the following expression is obtained

$$
\left[\mathrm{S}_{\mathrm{d}}=\left[\begin{array}{cccccc}
\mathrm{S}_{1,1}^{(1)} & \mathrm{S}_{1,2}^{(1)} & \mathrm{S}_{1,3}^{(1)} & 0 & 0 & 0 \\
\mathrm{~S}_{2,1}^{(1)} & \mathrm{S}_{2,2}^{(1)} & \mathrm{S}_{2,3}^{(1)} & 0 & 0 & 0 \\
\mathrm{~S}_{3,1}^{(1)} & \mathrm{S}_{3,1}^{(1)} & \mathrm{S}_{3,1}^{(1)} & 0 & 0 & 0 \\
0 & 0 & 0 & \mathrm{~S}_{4,4}^{(2)} & \mathrm{S}_{4,5}^{(2)} & \mathrm{S}_{4,6}^{(2)} \\
0 & 0 & 0 & \mathrm{~S}_{5,4}^{(2)} & \mathrm{S}_{5,5}^{(2)} & \mathrm{S}_{5,6}^{(2)} \\
0 & 0 & 0 & \mathrm{~S}_{6,4}^{(2)} & \mathrm{S}_{6,5}^{(2)} & \mathrm{S}_{6,6}^{(2)}
\end{array}\right]\right.
$$

Now after a simple matrix multiplication, we get for the intermediate matrix [B] the first equation (A.2.14) yields

$$
[\mathrm{B}]=\left[\begin{array}{cccc}
\mathrm{S}_{1,1}^{(1)} & \mathrm{S}_{1,2}^{(1)} & \mathrm{S}_{1,3}^{(1)} & 0 \\
\mathrm{~S}_{2,1}^{(1)} & \mathrm{S}_{2,2}^{(1)} & \mathrm{S}_{2,3}^{(1)} & 0 \\
\mathrm{~S}_{3,1}^{(1)} & \mathrm{S}_{3,1}^{(1)} & \mathrm{S}_{3,1}^{(1)} & 0 \\
\mathrm{~S}_{4,4}^{(2)} & \mathrm{S}_{4,6}^{(2)} & 0 & \mathrm{~S}_{4,5}^{(2)} \\
\mathrm{S}_{5,4}^{(2)} & \mathrm{S}_{5,6}^{(2)} & 0 & \mathrm{~S}_{5,5}^{(2)} \\
\mathrm{S}_{6,4}^{(2)} & \mathrm{S}_{6,6}^{(2)} & 0 & \mathrm{~S}_{6,5}^{(2)}
\end{array}\right]
$$

Then from the second equation of (A.2.14) after a matrix multiplication, we obtain the result presented by equation (3.77) in Section 3.6.

## A.2.5. The matrix equation for the unknown potentials

The necessary condition that the function (A.2.12) has a minimum is that the appropriate derivatives must be equal to zero

$$
\begin{equation*}
\frac{\partial W}{\partial\left[\mathrm{~V}_{\mathrm{j}}\right]_{u}}=0 \quad(j=1,2, \ldots, \mathrm{~N}) \tag{A.2.15}
\end{equation*}
$$

By use of (3.74), this can be shown to yield

$$
\begin{align*}
& {\left[[\mathrm{S}]_{\mathrm{u}, \mathrm{u}}[\mathrm{~S}]_{\mathrm{u}, \mathrm{k}}\right]\left[\begin{array}{l}
{[\mathrm{V}]_{\mathrm{u}}} \\
{[\mathrm{~V}]_{\mathrm{k}}}
\end{array}\right]=0 \text { or }} \\
& {[\mathrm{S}]_{\mathrm{u}, \mathrm{u}}[\mathrm{~V}]_{\mathrm{u}}+[\mathrm{S}]_{\mathrm{u}, \mathrm{k}}[\mathrm{~V}]_{\mathrm{k}}=0} \tag{A.2.16}
\end{align*}
$$

which leads to the final equation (3.78).

# Material parameters 

## Appendix 3: Material parameters

Conductors
Aluminum
Brass
Carbon
Copper
Germanium
Gold
Graphite
Iron
Mercury
Seawater
Silver
Tungsten

Insulators
Bakelite
Distilled water
Dry earth
Glass
Mica
Porcelain
Quartz
Rubber
Silicon
Transformer oil
Wax
Wet earth

Conductivity $\sigma$ (S/m)
$3.5 \times 10^{7}$
$1.6 \times 10^{7}$
$3 \times 10^{4}$
$5.8 \times 10^{7}$
2.3
$4.1 \times 10^{7}$
$10^{5}$
$10^{7}$
$10^{6}$
4
$6.2 \times 10^{7}$
$1.8 \times 10^{7}$

Conductivity $\sigma$ (S/m)
10-9
10-4
10-5
10-12
10-15
$2 \times 10^{-13}$
10-17
10-15
$3.9 \times 10^{-4}$
10-11
10-17
10-3

# Material parameters 

| Dielectrics | Relative dielectric constant $\varepsilon$ r |
| :---: | :---: |
| Air | 1 |
| Barium titanate | 1200 |
| Glass | 6 |
| Mica | 6 |
| Oil | 2.3 |
| Paper | 3 |
| Paraffin | 2 |
| Polystyrene | 2.6 |
| Porcelain | 7 |
| Quartz (fused) | 4 |
| Rubber | 2.3-4.0 |
| Teflon | 2.1 |
| Water (distilled) | 80 |
| Magnetic materials | Relative permeability $\underline{\mu}_{\underline{r}}$ |
| Cobalt | 600 |
| Commercial iron | 250-9000 |
| Nickel | 250 |
| Permalloy | $8 \times 10^{3}-10^{5}$ |
| Purified iron | $10^{4}-2 \times 10^{5}$ |
| Superpermalloy | $10^{5}-10^{6}$ |
| Paramagnetic materials |  |
| Aluminum | 1.000021 |
| Magnesium | 1.000012 |
| Palladium | 1.00082 |
| Titanium | 1.00018 |
| Diamagnetic materials |  |
| Bismuth | 0.99983 |
| Gold | 0.99996 |
| Silver | 0.99998 |
| Copper | 0.99999 |

## Transmission line parameters of two parallel wires

## Appendix 4: Transmission line parameters of two parallel wires

We calculate the capacitance of two parallel wires that are shown in Figure A.4-1. The radius of each wire is a.


Figure A.4-1. Equipotential contours surrounding two line charges $+\rho_{\mathrm{L}}$ and $-\rho_{\mathrm{L}}$. The surfaces at $r=a$ are equipotential surfaces so the electric field will always be normal to the metal surfaces.

The two wires can be replaced with two line charges $+\rho_{\mathrm{L}}$ and $-\rho_{\mathrm{L}}$. The precise location of these equivalent line charges is determined from the requirement that the surfaces of the metal wires be equipotential surfaces. This implies that the tangential electric fields will always be equal to zero on these surfaces. The potential $V(x, y)$ at the point $P$ is given by

$$
\begin{equation*}
\mathrm{V}(\mathrm{x}, \mathrm{y})=\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon} \ln \left(\frac{\mathrm{~s}}{\mathrm{r}_{2}}\right)+\frac{-\rho_{\mathrm{L}}}{2 \pi \varepsilon} \ln \left(\frac{\mathrm{~s}}{\mathrm{r}_{1}}\right)=\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon} \ln \left(\frac{r_{1}}{r_{2}}\right) \tag{A.4.1}
\end{equation*}
$$

## Transmission line parameters of two parallel wires

The plane at the midpoint between the two wires is an equipotential surface that is equal to zero potential. Other equipotential contours are found by setting

$$
\begin{equation*}
\frac{r_{1}}{r_{2}}=\frac{\sqrt{(s+x)^{2}+y^{2}}}{\sqrt{(s-x)^{2}+y^{2}}}=k=\text { constant } \tag{A.4.2}
\end{equation*}
$$

This can be written as

$$
\begin{equation*}
(s+x)^{2}+y^{2}=k^{2}\left[(s-x)^{2}+y^{2}\right\rfloor \tag{A.4.3}
\end{equation*}
$$

or

$$
\begin{equation*}
\left(x-s \frac{k^{2}+1}{k^{2}-1}\right)^{2}+y^{2}=\left(\frac{2 k s}{k^{2}-1}\right) \tag{A.4.4}
\end{equation*}
$$

The common factor

$$
\mathrm{s}^{2} \frac{\mathrm{k}^{2}+1}{\mathrm{k}^{2}-1}
$$

has been added to both sides of (A4.3) in order to complete the squares.
Equation (A.4.4) is an equation for a family of circles which have radii

$$
\begin{equation*}
r_{0}=s^{2} \frac{k^{2}+1}{k^{2}-1} \tag{A.4.5}
\end{equation*}
$$

and are centered at the points

$$
\begin{equation*}
(h, 0)=\left[s^{2} \frac{k^{2}+1}{k^{2}-1}, 0\right] \tag{A.4.6}
\end{equation*}
$$

where

$$
h=s \frac{k^{2}+1}{k^{2}-1}
$$

Eliminating the term s between (A.4.5) and (A.4.6), we obtain

$$
\begin{equation*}
\mathrm{k}^{2}-\frac{2 \mathrm{kh}}{\mathrm{r}_{0}}+1=0 \tag{A.4.7}
\end{equation*}
$$

The two solutions for this equation are given by

$$
\begin{equation*}
k=\frac{h}{a} \pm \sqrt{\left(\frac{h}{a}\right)^{2}-1} \tag{A.4.8}
\end{equation*}
$$

The root $k$ with the + sign will give the equipotential contours in the region $x>0$ and the root $\hat{k}$ with the - sign will give the equipotential contours in the region $x<$

## Transmission line parameters of two parallel wires

0 . We will have particular interest in the equipotential contour at the surface of the wire at $r_{0}=a$. The spacing $h$ must also be greater than this radius $a$ of the wire and we will set it equal to D/2.

The potential difference between the wires is given by

$$
\begin{align*}
\Delta \mathrm{V} & =\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon} \ln \mathrm{k}-\frac{\rho_{\mathrm{L}}}{2 \pi \varepsilon} \ln \hat{\mathrm{k}}=\frac{\rho_{\mathrm{L}}}{\pi \varepsilon} \ln \mathrm{k} \\
& =\frac{\rho_{\mathrm{L}}}{\pi \varepsilon} \ln \left[\frac{\mathrm{D}}{2 \mathrm{a}} \pm \sqrt{\left(\frac{\mathrm{D}}{2 \mathrm{a}}\right)^{2}-1}\right] \tag{A.4.9}
\end{align*}
$$

Making use of the identity

$$
\begin{equation*}
\ln \left\lfloor\xi+\sqrt{\xi^{2}-1} \mid=\cosh ^{-1} \xi\right. \tag{A.4.10}
\end{equation*}
$$

we write the capacitance as

$$
\begin{equation*}
\mathrm{C}=\frac{\rho_{\mathrm{L}} \Delta \mathrm{z}}{\Delta \mathrm{~V}}=\frac{\Delta \mathrm{z}}{\frac{1}{\pi \varepsilon} \cosh ^{-1}\left(\frac{\mathrm{D}}{2 \mathrm{a}}\right)}=\frac{\pi \varepsilon \Delta \mathrm{z}}{\cosh ^{-1}\left(\frac{\mathrm{D}}{2 \mathrm{a}}\right)} \tag{A.4.11}
\end{equation*}
$$

In order to calculate the inductance per unit length, we make use of the relation that $\frac{\mathrm{L}}{\Delta \mathrm{z}} \frac{\mathrm{C}}{\Delta \mathrm{z}}=\mu \varepsilon$ and write

$$
\begin{equation*}
\mathrm{L}=\frac{\mu \Delta \mathrm{z}}{\pi} \cosh ^{-1}\left(\frac{\mathrm{D}}{2 \mathrm{a}}\right) \tag{A.4.12}
\end{equation*}
$$

There are alternative derivations that can be used to approximate the equivalent circuit parameters. For example, if we define a point between the two cylinders as $\hat{x}$ and place the left cylinder at $x=0$, then the magnetic flux density is given in Figure A.4-2.

## Transmission line parameters of two parallel wires



Figure A.4-2. Alternative cross section of a twin-lead transmission line.

If the current is into the paper in the wire centered at $x=0$ and out of the paper in the wire centered at $x=D$, the magnetic flux densities will add in the center. From Ampere's circuital law, we write the magnetic flux density as

$$
\begin{equation*}
B=\frac{\mu l}{2}\left[\frac{1}{\hat{x}}+\frac{1}{D-\hat{x}}\right] \tag{A.4.13}
\end{equation*}
$$

The total magnetic flux that passes between the two wires is found from

$$
\begin{equation*}
\frac{\Psi_{\mathrm{m}}}{\Delta \mathrm{z}}=\int_{\mathrm{a}}^{\mathrm{D}-\mathrm{a}} \mathrm{Bd} \hat{\mathrm{x}}=\frac{\mu \mathrm{l}}{2 \pi}[\ln \hat{\mathrm{x}}-\ln (\mathrm{D}-\hat{\mathrm{x}})]_{\mathrm{a}}^{\mathrm{D}-\mathrm{a}}=\frac{\mu \mathrm{l}}{\pi} \ln \left(\frac{\mathrm{D}-\mathrm{a}}{\mathrm{a}}\right) \tag{A.4.14}
\end{equation*}
$$

The inductance per unit length is given by

$$
\begin{equation*}
\frac{\mathrm{L}}{\Delta \mathrm{z}} \equiv \frac{\left(\frac{\Psi_{\mathrm{m}}}{\Delta \mathrm{z}}\right)}{\mathrm{l}}=\frac{\mu}{\pi} \ln \left(\frac{\mathrm{D}-\mathrm{a}}{\mathrm{a}}\right) \approx \frac{\mu}{\pi} \ln \left(\frac{\mathrm{D}}{\mathrm{a}}\right) \tag{A.4.15}
\end{equation*}
$$

which is in agreement with (A.4.12) in the limit of $D \gg a$.
Applying Gauss's law to the two wires that are assumed to each have a linear charge density $+\rho_{L}$ and $-\rho_{L} C / m$ distributed on the two wires, we find the electric field at $\hat{x}$ to be given by

$$
\begin{equation*}
E=\frac{\rho_{\mathrm{L}} \Delta \mathrm{z}}{2 \pi \varepsilon}\left[\frac{1}{\hat{\mathrm{x}}}+\frac{1}{\mathrm{D}-\hat{\mathrm{x}}}\right] \tag{A.4.16}
\end{equation*}
$$

## Transmission line parameters of two parallel wires

The potential difference $\Delta \mathrm{V}$ between the two wires is obtained by integrating the electric field between the two wires to yield

$$
\begin{equation*}
\Delta \mathrm{V}=\int_{\mathrm{a}}^{\mathrm{D}-\mathrm{a}} \mathrm{Ed} \hat{\mathrm{x}}=\frac{\rho_{\mathrm{L}} \Delta \mathrm{Z}}{\pi \varepsilon} \ln \left(\frac{\mathrm{D}-\mathrm{a}}{\mathrm{a}}\right) \tag{A.4.17}
\end{equation*}
$$

The capacitance per unit length is given by

$$
\begin{equation*}
\frac{\mathrm{C}}{\Delta \mathrm{z}} \equiv \frac{\rho_{\mathrm{L}}}{\Delta \mathrm{~V}}=\frac{\pi \varepsilon}{\ln \left(\frac{\mathrm{D}-\mathrm{a}}{\mathrm{a}}\right)} \approx \frac{\pi \varepsilon}{\ln \left(\frac{\mathrm{D}}{\mathrm{a}}\right)} \tag{A.4.18}
\end{equation*}
$$

This is in agreement with (A.4.11) in the limit of $D \gg a$.
We present the both methods to obtain the same results since you may encounter them in different books. The twin-lead is commonly used in practice. In neither calculation did we include the self-inductance of each wire that arises when $\mathrm{D} \approx \mathrm{a}$. This will cause an additional small constant term to appear in the final results.

# Plasma evolution adjacent to a metallic surface Appendix 5: Plasma evolution adjacent to a metallic surface 

The temporal and spatial evolution of the plasma adjacent to a metallic electrode whose voltage is suddenly decreased from 0 to a large negative value has certain implications. A gaseous plasma consists of negatively charged electrons and positively charged ions whose mass is significantly greater than the mass of the electron. In Example 3-3, we determined the potential profile at the time $t=0^{+}$just after the switch connecting the negative potential source was closed at a time $t=0$. During this initial time interval, the electrons were expelled from the region adjacent to the metallic plate but the ions had not yet started to move. The temporal and spatial evolution of these ions toward the negatively biased metallic electrode and the expansion of the ion density rarefaction into the plasma requires a numerical computation. ${ }^{1}$ The plasma is modeled with a dimensionless fluid modeled description.

The ion density perturbation $n_{i}$ and the ion velocity perturbation $v_{i}$ are described with the equation of continuity

$$
\begin{equation*}
\frac{\partial n_{i}}{\partial t}+\frac{\partial\left(n_{i} v_{i}\right)}{\partial z}=0 \tag{A.5.1}
\end{equation*}
$$

and the equation of motion

$$
\begin{equation*}
\frac{\partial \mathbf{v}_{i}}{\partial \mathrm{t}}+\mathrm{v}_{\mathrm{i}} \frac{\partial \mathbf{v}_{i}}{\partial \mathrm{z}}=-\frac{\partial \mathrm{V}}{\partial \mathrm{z}} \tag{A.5.2}
\end{equation*}
$$

Since the mass of the electron is so much smaller than the mass of the ion, the electron density perturbation $\mathrm{n}_{\mathrm{e}}$ can be approximated with a Maxwell-Boltzmann distribution

$$
\begin{equation*}
\mathrm{n}_{\mathrm{e}}=\mathrm{e}^{\mathrm{v}} \tag{A.5.3}
\end{equation*}
$$

[^29]
## Plasma evolution adjacent to a metallic surface

In order to reflect the non-neutrality of the density perturbations, there will be an electric field $E=-\frac{d V}{d z}$ that is governed by Poisson's equation

$$
\begin{equation*}
\frac{\partial^{2} V}{\partial z^{2}}=n_{e}-n_{i} \tag{A.5.4}
\end{equation*}
$$

These equations have been written in dimensionless units that are defined by

$$
\begin{equation*}
\mathrm{n}_{\mathrm{i}}=\frac{\mathrm{n}_{\mathrm{i}}^{\prime}}{\mathrm{n}_{\mathrm{o}}} ; \mathrm{n}_{\mathrm{e}}=\frac{\mathrm{n}_{\mathrm{e}}^{\prime}}{\mathrm{n}_{\mathrm{o}}} ; \mathrm{v}_{\mathrm{i}}=\frac{\mathrm{v}_{\mathrm{i}}^{\prime}}{\mathrm{v}_{\mathrm{s}}} ; \mathrm{V}=\frac{\mathrm{eV} \mathrm{~V}^{\prime}}{\kappa_{\mathrm{B}} \mathrm{~T}_{\mathrm{e}}} ; \mathrm{z}=\frac{\mathrm{z}^{\prime}}{\lambda_{\mathrm{D}}} ; \text { and } \mathrm{t}=\omega_{\mathrm{p}} \mathrm{t}^{\prime} \tag{A.5.5}
\end{equation*}
$$

where $n_{0}$ is the equilibrium electron and ion density, the ' indicates the laboratory variables, $\kappa_{B}$ is Boltzmann's constant $\kappa_{B}=1.38 \times 10^{-23} \mathrm{~J} /{ }^{\circ} \mathrm{K}$, and

$$
\begin{gather*}
v_{\mathrm{s}}=\sqrt{\frac{\kappa_{\mathrm{B}} T_{\mathrm{e}}}{\mathrm{~m}_{\mathrm{i}}}} \text {-- the ion acoustic velocity }  \tag{A.5.6}\\
\lambda_{\mathrm{D}}=\sqrt{\frac{\kappa_{\mathrm{B}} \mathrm{~T}_{\mathrm{e}}}{\left(\frac{\mathrm{n}_{\mathrm{o}} \mathrm{e}^{2}}{\varepsilon_{0}}\right)}} \text {-- the electron Debye length }  \tag{A.5.7}\\
\omega_{\mathrm{pi}}=\sqrt{\frac{\mathrm{n}_{\mathrm{o}} \mathrm{e}^{2}}{m_{i} \varepsilon_{\mathrm{o}}}} \tag{A.5.8}
\end{gather*}
$$

This model is valid if the electronic temperature $T_{e}$ is much greater than the ion temperature $T_{i}$. This is a valid approximation in a gaseous plasma that one normally encounters in the laboratory. For example, a typical laboratory argon plasma will have an approximate ion acoustic velocity of $10^{3} \mathrm{~m} / \mathrm{s}$ and an ion plasma frequency of $2 \pi \times 10^{6}$ radians/s.

The calculated results of the evolution of the density perturbations are shown in Figure A5-1a. It is noted that there is an instantaneous decrease of the electron density adjacent to the electrode which is the "transient sheath" or the "ion matrix sheath" that was described in Example 3-3. As time increases, this electron density rarefaction expands into the plasma. Initially, this expansion is faster than the final asymptotic value which is the ion acoustic velocity. In addition, the heavy positive ions are attracted to and move toward the electrode.

## Plasma evolution adjacent to a metallic surface

This results in a flux of the ions that impinges upon the electrode. The temporal evolution of this flux is shown in Figure A5-5b. These ions can implant themselves into the electrode and change the surface characteristics of the electrode.



Figure A.5-1. (a) Evolution of the normalized electron density and the normalized ion density in space at equal increments in time after the application of a negative potential to a metallic electrode that had been inserted into a plasma. (b) Temporal evolution of the lon flux impinging upon the electrode.

# References 

## Appendix 6: References

## Electromagnetic theory

David K. Cheng, Field and Wave Electromagnetics $2^{\text {nd }}$ edition, Prentice-Hall, 1989.

William H. Hayt and John A. Buck, Engineering Electromagnetics $6^{\text {th }}$ edition, McGraw-Hill, 2001.

Nannapaneni N Rao, Elements of Engineering Electromagnetics 5th edition, Prentice-Hall, 2000.

Akira Hirose and Karl E. Lonngren, Introduction to Wave Phenomena, Krieger, 2002.

Fawwaz T. Ulaby, Fundamentals of Applied Electromagnetics, Prentice-Hall, 2001.

Advanced topics in electromagnetic theory
John D. Jackson, Classical Electrodynamics $3^{\text {rd }}$ edition, John Wiley, 1998.
John D. Kraus and Ronald J. Marhefka, Antennas $3^{\text {rd }}$ edition, McGraw-Hill, 2002.

John D. Kraus, Electromagnetics $2^{\text {nd }}$ edition, Mc Graw-Hill, 1984.

Simon Ramo, John R. Whinnery and Theodore Van Duzer, Fields and Waves in Communication Electronics, $3^{\text {rd }}$ edition, John Wiley, 1994.

Constantine Balanis, Antenna Theory $\mathbf{2}^{\text {nd }}$ edition, J. Wiley, 1997.
Robert C. Hansen (ed.), Array Theory and Practice, vol. 2, Academic Press, 1966.

## References

## MATLAB and mathematics

M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, NBS, 1972.

Stephen J. Chapman , MATLAB programming for engineers, Brooks/ColeThomson Learning, 2002.

Alejandro L. Garcia, Numerical Methods for Physics, Prentice Hall, Inc. 1994.
Peter P. Silvester and Ronald L. Ferrari, Finite Elements for Electrical Engineers, Cambridge University Press, 1996.


[^0]:    ${ }^{1}$ The field $\mathbf{F}=3 \mathbf{u}_{\mathbf{x}}+4 \mathbf{u}_{\mathbf{y}}$ is a conservative field as will be demonstrated.

[^1]:    ${ }^{1}$ This discipline is frequently called "Electrical and Computer Engineering."

[^2]:    ${ }^{1}$ The differentiation of an integral is given by the expression that is known as Leibnitz's rule

    $$
    \frac{d \int_{a(x)}^{b(x)} f(x, y) d y}{d x}=\int_{a(x)}^{b(x)} \frac{d f(x, y)}{d x} d y+f(x, b(x)) \frac{d b(x)}{d x}-f(x, a(x)) \frac{d a(x)}{d x}
    $$

[^3]:    ${ }^{1}$ The word "gedanken" is German for the word thought. Hence we are to perform a thought experiment.

[^4]:    ${ }^{2}$ One eV $=1.6 \times 10^{-19} \mathrm{~J}$.

[^5]:    1 Voyager confirmed the presence of radiation belts at the planets: Jupiter, Saturn, Uranus and Neptune. Within the sensitivity limits of the instruments, no radiation belts were detected at Venus and at Mars. This is indicative of the presence or absence of a magnetic field at these planets. Magnetic fields at Pluto are as yet unknown. The satellite is approaching the edge of the solar system in 2003.

[^6]:    ${ }^{1}$ The Hall voltage can be measured if the exhaust of certain toy rocket motors is passed through a magnetic field of 1 T . Using a model $\mathrm{C}-60$ and a surplus magnetron magnet, it is possible to generate 0.5 V for almost two seconds.

[^7]:    $\overline{{ }^{1} \text { K. E. Lonngren, P. V. Schwartz, E. W. Bai, W. C. Theisen, R. L. Merlino, and R. T. Carpenter, }}$ "Extracting Double Layer Charge Density Distributions using the Method of Moments," IEEE Transactions on Plasma Science, Vol. 24, pp. 278-280, (1994).

[^8]:    1 We can only speculate about what these three giants would talk and would write if they today met at a cafe and had only one "back of an envelope" between them

[^9]:    ${ }^{1}$ Recall that in a vacuum $\mathbf{D}=\varepsilon_{0} \mathbf{E}$. If a dielectric is inserted between the plates, we must use $\mathbf{D}=\varepsilon \mathrm{E}$.

[^10]:    ${ }^{1}$ It is common in engineering talks to place a standard six foot stick man next to a drawing of a machine in order to indicate its size. To emphasize the importance and size of these four equations, the reader could think of the symbol $\bar{\nabla}$ as being an inverted pyramid arising out of the grains of sand of the remaining words and equations in this text and several others.

[^11]:    ${ }^{1}$ This plunger is sometimes called a "wavemaker."

[^12]:    ${ }^{2}$ We assume that the air is not being heated in class by the teacher. Politicians are another story.

[^13]:    ${ }^{3}$ These equations are derived in several texts. It is also shown there that these waves carry both energy and momentum as they propagate. The authors are familiar with Introduction to Wave Phenomena, A. Hirose and K. E. Lonngren, Wiley-Interscience, New York, 1985; reprinted by Krieger, Malabar, Florida, 1991, 2001.

[^14]:    ${ }^{4}$ An electric field that is a constant, such as zero, is also a solution to this equation. However, this solution shall be viewed as being in the trivial class of solutions and will not be discussed further.

[^15]:    ${ }^{1}$ Plane waves also be encountered in ideal lossless transmission lines or coaxial cables.

[^16]:    ${ }^{2}$ Certain waves will have their phase velocity and the direction of energy propagation in the opposite directions. Such "backward" waves need not concern us here since we are examining vacuum conditions. They will be described in a particular dispersive transmission line.

[^17]:    ${ }^{1}$ The $Q$ of a circuit or a cavity is defined as $Q=2 \pi($ energy stored)/(energy dissipated per cycle)

[^18]:    ${ }^{1}$ Another example of a function that was defined and successfully used prior to its rigorous proof is the "delta" function. Dirac made considerable use of it during the period that led to the development of quantum mechanics prior to and during the Great Depression. This was a decade before it was proven to be a valid function.

[^19]:    ${ }^{1}$ Some authors prefer to use the Greek letter $\Gamma$ to indicate the reflection coefficient.

[^20]:    ${ }^{2}$ VSWR is pronounced as "vizwar" in practice.

[^21]:    ${ }^{1}$ S. Ramo, J. Whinnery, T. van Duzer, Fields and Waves in Communication Electronics, J.Wiley, 1965.

[^22]:    ${ }^{1}$ C. Balanis, Antenna Theory, J. Wiley, 1997.

[^23]:    1 In order to obtain analytical solutions in electromagnetics, we have to resort to many approximations. The ingenuity of the practitioner is tested when it comes to making sure that the approximations are reasonable. The success of the practitioner is tested when it comes to deciding what "reasonable" means.

[^24]:    ${ }^{2}$ S.V. Savov, "An efficient solution of a class of integrals arising in antenna theory", IEEE Antennas \& Propag. Magazine, vol. 44, Oct. 2002, pp. 98-101.

[^25]:    ${ }^{1}$ M. Abramowitz, I.Stegun, Handbook of Mathematical Functions, NY: NBS, 1964.

[^26]:    ${ }^{2}$ J. Kraus, Electromagnetics, Mc Graw-Hill, 1984.

[^27]:    ${ }^{3}$ M. Abramowitz, I.Stegun, Handbook of Mathematical Functions, NY: NBS, 1964.

[^28]:    ${ }^{1}$ R.C. Hansen (ed.), Array Theory and Practice, vol. 2, Academic Press, 1966.

[^29]:    ${ }^{1}$ Widner, M., Alexeff, I., Jones, W.D. and Lonngren, K.E., "Ion- Acoustic Wave Excitation and Ion Sheath Evolution," Physics of Fluids, Vol. 13, October 1970, pp. 2532-2540.

