1: PRELIMINARY CONCEPTS

Steven W. Ellingson
Virginia Polytechnic Institute and State University

LibreTexts™

CHAPTER OVERVIEW

1: Preliminary Concepts

A field is the continuum of values of a quantity as a function of position and time. The quantity that the field describes may be a scalar or a vector, and the scalar part may be either real- or complex-valued. In electromagnetics, the electric field intensity $\mathbf{E}$ is a real-valued vector field that may vary as a function of position and time, and so might be indicated as "$\mathbf{E}(x, y, z, t)$," "$\mathbf{E}(\mathbf{r}, t)$," or simply "$\mathbf{E}$." When expressed as a phasor, this quantity is complex-valued but exhibits no time dependence, so we might say instead "$\widetilde{\mathbf{E}}(\mathbf{r})$" or simply "$\widetilde{\mathbf{E}}$." An example of a scalar field in electromagnetics is the electric potential, V; i.e., $V(\mathbf{r}, t)$. A *wave* is a time-varying field that continues to exist in the absence of the source that created it and is therefore able to transport energy.

1.1: What is Electromagnetics?

The topic of this book is applied engineering electromagnetics. This topic is often described as “the theory of electromagnetic fields and waves,” which is both true and misleading. The truth is that electric fields, magnetic fields, their sources, waves, and the behavior these waves are all topics covered by this book. The misleading part is that our *principal* aim shall be to close the gap between basic electrical circuit theory and the more general theory that is required to address certain topics that are of broad and common interest in the field of electrical engineering. (For a preview of topics where these techniques are required, see the list at the end of this section.) In basic electrical circuit theory, the behavior of devices and systems is abstracted in such a way that the underlying electromagnetic principles do not need to be considered. Every student of electrical engineering encounters this, and is grateful since this greatly simplifies analysis and design. For example, a resistor is commonly defined as a device which exhibits a particular voltage

$$V = IR$$

in response to a current I, and the resistor is therefore *completely* described by the value R. This is an example of a “lumped element” abstraction of an electrical device. Much can be accomplished knowing nothing else about resistors; no particular knowledge of the physical concepts of electrical potential, conduction current, or resistance is required. However, this simplification makes it impossible to answer some frequently-encountered questions. Here are just a few:

- What determines R? How does one go about designing a resistor to have a particular resistance?
- Practical resistors are rated for power-handling capability; e.g., discrete resistors are frequently identified as “1/8-W,” “1/4-W,” and so on. How does one determine this, and how can this be adjusted in the design?
- Practical resistors exhibit significant reactance as well as resistance. Why? How is this determined? What can be done to mitigate this?
- Most things which are not resistors also exhibit significant resistance and reactance – for example, electrical pins and interconnects. Why? How is this determined? What can be done to mitigate this?

The answers to the these questions must involve *properties of materials and the geometry in which those materials are arranged*. These are precisely the things that disappear in lumped element device models, so it is not surprising that such models leave us in the dark on these issues. It should also be apparent that what is true for the resistor is also going to be true for other devices of practical interest, including capacitors (and devices unintentionally exhibiting capacitance), inductors (and devices unintentionally exhibiting inductance), transformers (and devices unintentionally exhibiting mutual impedance), and so on. From this perspective, electromagnetics may be viewed as a generalization of electrical circuit theory that addresses these considerations. Conversely basic electric circuit theory may be viewed a special case of electromagnetic theory that applies when these considerations are not important. Many instances of this “electromagnetics as generalization” vs. “lumped-element theory as special case” dichotomy appear in the study of electromagnetics.

There is more to the topic, however. There are many devices and applications in which electromagnetic fields and waves are *primary* engineering considerations that must be dealt with directly. Examples include electrical generators and motors; antennas; printed circuit board stackup and layout; persistent storage of data (e.g., hard drives); fiber optics; and systems for radio, radar, remote sensing, and medical imaging. Considerations such as signal integrity and electromagnetic compatibility (EMC) similarly require explicit consideration of electromagnetic principles.

Although electromagnetic considerations pertain to all frequencies, these considerations become increasingly difficult to avoid with increasing frequency. This is because the wavelength of an electromagnetic field decreases with increasing frequency. When wavelength is large compared to the size of the region of interest (e.g., a circuit), then analysis and design is not much different from zero-frequency (“DC”) analysis and design.

For example, the free space wavelength at 3 MHz is about 100 m, so a planar circuit having dimensions 10 cm × 10 cm is just 0.1% of a wavelength across at this frequency. Although an electromagnetic wave may be present, it has about the same value over the region of space occupied by the circuit. In contrast, the free space wavelength at 3 GHz is about 10 cm, so the same circuit is one full wavelength across at this frequency. In this case, different parts of this circuit observe the same signal with very different magnitude and phase.

Some of the behaviors associated with non-negligible dimensions are undesirable, especially if not taken into account in the design process. However, these behaviors can also be exploited to do some amazing and useful things – for example, to launch an electromagnetic wave (i.e., an antenna) or to create filters and impedance matching devices consisting only of metallic shapes, free of discrete capacitors or inductors.

Electromagnetic considerations become not only unavoidable but central to analysis and design above a few hundred MHz, and especially in the millimeter-wave, infrared (IR), optical, and ultraviolet (UV) bands. The discipline of electrical engineering encompasses applications in these frequency ranges even though – ironically – such applications may not operate according to principles that can be considered "electrical"! Nevertheless, electromagnetic theory applies.

Another common way to answer the question "What is electromagnetics?" is to identify the topics that are commonly addressed within this discipline. Here's a list of topics – some of which have already been mentioned – in which explicit consideration of electromagnetic principles is either important or essential:

- Antennas
- Coaxial cable
- Design and characterization of common discrete passive components including resistors, capacitors, inductors, and diodes
- Distributed (e.g., microstrip) filters
- Electromagnetic compatibility (EMC)
- Fiber optics
- Generators
- Magnetic resonance imaging (MRI)
- Magnetic storage (of data)
- Microstrip transmission lines
- Modeling of non-ideal behaviors of discrete components
- Motors
- Non-contact sensors
- Photonics
- Printed circuit board stackup and layout
- Radar
- Radio wave propagation
- Radio frequency electronics
- Signal integrity
- Transformers
- Waveguides

Summary

> Applied engineering electromagnetics is the study of those aspects of electrical engineering in situations in which the electromagnetic properties of materials and the geometry in which those materials are arranged is important. This requires an understanding of electromagnetic fields and waves, which are of primary interest in some applications.

Finally, here are two broadly-defined learning objectives that should now be apparent: (1) Learn the techniques of engineering analysis and design that apply when electromagnetic principles are important, and (2) Better understand the physics underlying the operation of electrical devices and systems, so that when issues associated with these physical principles emerge one is prepared to recognize and grapple with them.

1.2: Electromagnetic Spectrum

Electromagnetic fields exist at frequencies from DC (0 Hz) to at least 10^{20} Hz – that's at least 20 orders of magnitude! At DC, electromagnetics consists of two distinct disciplines: *electrostatics*, concerned with electric fields; and *magnetostatics*, concerned with magnetic fields. At higher frequencies, electric and magnetic fields interact to form propagating waves. Waves having frequencies within certain ranges are given names based on how they manifest as physical phenomena. These names are (in order of increasing frequency): radio, infrared (IR), optical (also known as "light"), ultraviolet (UV), X-rays, and gamma rays (γ-rays). See Table 1.2.1 and Figure 1.2.1 for frequency ranges and associated wavelengths.

Definition: Electromagnetic Spectrum

The term *electromagnetic spectrum* refers to the various forms of electromagnetic phenomena that exist over the continuum of frequencies

The speed (properly known as "phase velocity") at which electromagnetic fields propagate in free space is given the symbol c, and has the value $\cong 3.00 \times 10^8$ m/s. This value is often referred to as the "speed of light." While it is certainly the speed of light in free space, it is also the speed of *any* electromagnetic wave in free space. Given frequency f, wavelength is given by the expression

$$\underbrace{\lambda = \frac{c}{f}}_{\text{in free space}}$$

Table 1.2.1 shows the free space wavelengths associated with each of the regions of the electromagnetic spectrum. This book presents a version of electromagnetic theory that is based on classical physics. This approach works well for most practical problems. However, at very high frequencies, wavelengths become small enough that quantum mechanical effects may be important. This is usually the case in the X-ray band and above. In some applications, these effects become important at frequencies as low as the optical, IR, or radio bands. (A prime example is the *photoelectric effect*; see "Additional References" below.) Thus, caution is required when applying the classical version of electromagnetic theory presented here, especially at these higher frequencies.

Table 1.2.1: The electromagnetic spectrum. Note that the indicated ranges are arbitrary but consistent with common usage.

Regime	Frequency Range	Wavelength Range
γ-Ray	$> 3 \times 10^{19}$ Hz	< 0.01 nm
X-Ray	3×10^{16} Hz – 3×10^{19} Hz	10–0.01 nm
Ultraviolet (UV)	2.5×10^{15} – 3×10^{16} Hz	120–10 nm
Optical	4.3×10^{14} – 2.5×10^{15} Hz	700–120 nm
Infrared (IR)	300 GHz – 4.3×10^{14} Hz	1 mm – 700 nm
Radio	3 kHz–300 GHz	100 km – 1 mm

The radio portion of the electromagnetic spectrum alone spans 12 orders of magnitude in frequency (and wavelength), and so, not surprisingly, exhibits a broad range of phenomena. This is shown in Figure 1.2.1.

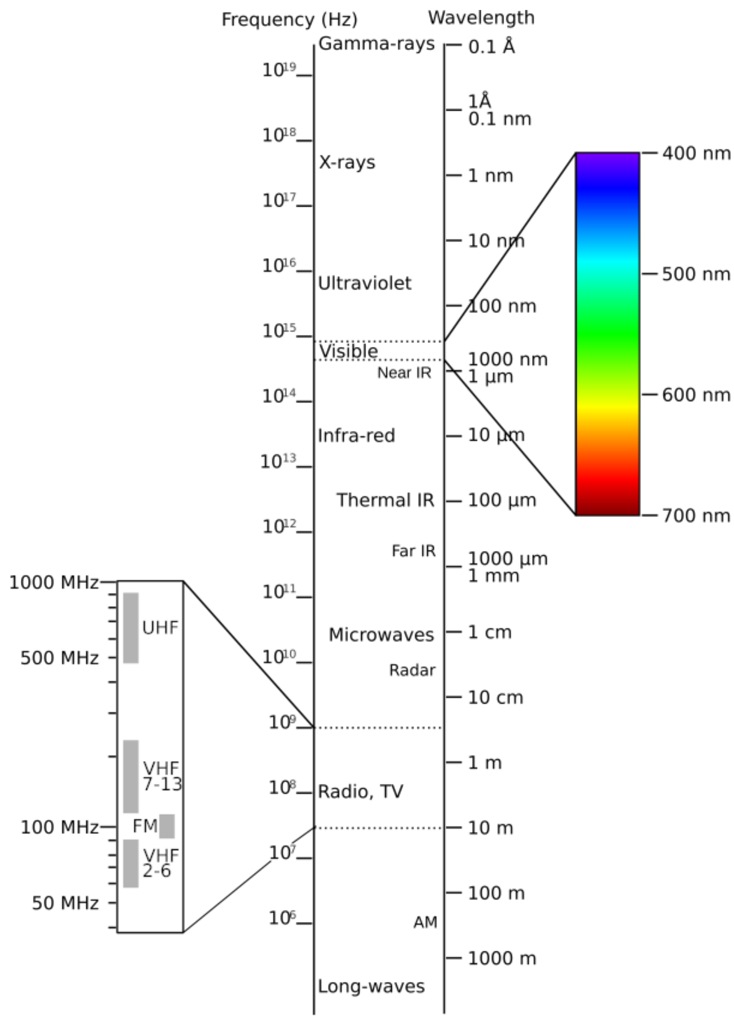

Figure 1.2.1: Electromagnetic Spectrum.

Table 1.2.2: The radio portion of the electromagnetic spectrum, according to a common scheme for naming ranges of radio frequencies. WLAN: Wireless local area network, LMR: Land mobile radio, RFID: Radio frequency identification

Band	Frequencies	Wavelengths	Typical Applications
EHF	30-300 GHz	10–1 mm	60 GHz WLAN, Point-to-point data links
SHF	3–30 GHz	10–1 cm	Terrestrial & Satellite data links, Radar
UHF	300–3000 MHz	1–0.1 m	TV broadcasting, Cellular, WLAN
VHF	30–300 MHz	10–1 m	FM & TV broadcasting, LMR
HF	3–30 MHz	100–10 m	Global terrestrial comm., CB Radio
MF	300–3000 kHz	1000–100 m	AM broadcasting
LF	30–300 kHz	10–1 km	Navigation, RFID
VLF	3–30 kHz	100–10 km	Navigation

Table 1.2.3: The optical portion of the electromagnetic spectrum.

Band	Frequencies	Wavelengths
Violet	668–789 THz	450–380 nm

Band	Frequencies	Wavelengths
Blue	606–668 THz	495–450 nm
Green	526–606 THz	570–495 nm
Yellow	508–526 THz	590–570 nm
Orange	484–508 THz	620–590 nm
Red	400–484 THz	750–620 nm

1.3: Fundamentals of Waves

In this section, we formally introduce the concept of a wave and explain some basic characteristics.

To begin, let us consider not electromagnetic waves, but rather sound waves. To be clear, sound waves and electromagnetic waves are completely distinct phenomena. Sound waves are variations in pressure, whereas electromagnetic waves are variations in electric and magnetic fields. However, the mathematics that govern sound waves and electromagnetic waves are very similar, so the analogy provides useful insight. Furthermore, sound waves are intuitive for most people because they are readily observed. So, here we go:

Imagine standing in an open field and that it is completely quiet. In this case, the air pressure everywhere is about 101 kPa (101,000 N/m^2) at sealevel, and we refer to this as the *quiescent air pressure*. Sound can be described as the *differential air pressure* $p(x, y, z, t)$, which we define as the absolute air pressure at the spatial coordinates (x, y, z) minus the quiescent air pressure. So, when there is no sound, $p(x, y, z, t) = 0$. The function p as an example of a scalar field.

Let's also say you are standing at $x = y = z = 0$ and you have brought along a friend who is standing at $x = d$; i.e., a distance d from you along the x axis. Also, for simplicity, let us consider only what is happening along the x axis; i.e., $p(x, t)$.

At $t = 0$, you clap your hands once. This forces the air between your hands to press outward, creating a region of increased pressure (i.e., $p > 0$) that travels outward. As the region of increased pressure moves outward, it leaves behind a region of low pressure where $p < 0$. Air molecules immediately move toward this region of lower pressure, and so the air pressure quickly returns to the quiescent value, $p = 0$. The traveling disturbance in $p(x, t)$ is the sound of the clap. The disturbance continues to travel outward until it reaches your friend, who then hears the clap.

At each point in time, you can make a plot of $p(x, t)$ versus x for the current value of t. This is shown in Figure $1.3.1$. At times $t < 0$, we have simply $p(x, t) = 0$. A short time after $t = 0$, the peak pressure is located at slightly to the right of $x = 0$. The pressure is not a simple impulse because interactions between air molecules constrain the pressure to be continuous over space. So instead, we see a rounded pulse representing the rapid build-up and similarly rapid decline in air pressure. A short time later $p(x, t)$ looks very similar, except the pulse is now further away.

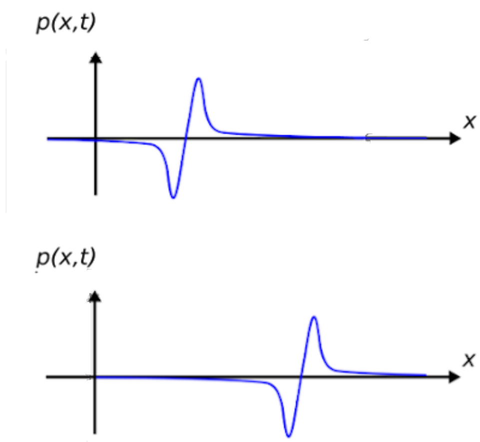

Figure $1.3.1$: The differential pressure $p(x, t)$ (top) a short time after the clap and (bottom) a slightly longer time after the clap. (CC BY 4.0; Y. Qin).

Now: What precisely is $p(x, t)$? Completely skipping over the derivation, the answer is that $p(x, t)$ is the solution to the *acoustic wave equation* (see "Additional References" at the end of this section):

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{c_s^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{1.3.1}$$

where c_s is the speed of sound, which is about 340 m/s at sea level. Just to emphasize the quality of the analogy between sound waves and electromagnetic waves, know that the acoustic wave equation is mathematically identical to equations that that govern electromagnetic waves

Although "transient" phenomena – analogous to a clap – are of interest in electromagnetics, an even more common case of interest is the wave resulting from a sinusoidally-varying source. We can demonstrate this kind of wave in the context of sound as well.

Here we go:

In the previous scenario, you pick up a trumpet and blow a perfect A note. The A note is 440 Hz, meaning that the air pressure emerging from your trumpet is varying sinusoidally at a frequency of 440 Hz. Let's say you can continue to blow this note long enough for the entire field to be filled with the sound of your trumpet. Now what does the pressure-versus-distance curve look like? Two simple observations will settle that question:

- $p(x,t)$ at any *constant position* x is a sinusoid as a function of t. This is because the acoustic wave equation is linear and time invariant, so a sinusoidal excitation (i.e., your trumpet) results in a sinusoidal response at the same frequency (i.e., the sound heard by your friend).
- $p(x,t)$ at any *constant time* t is also a sinusoid as a function of x. This is because the sound is propagating away from the trumpet and toward your friend, and anyone in between will also hear the A note, but with a phase shift determined by the difference in distances.

This is enough information to know that the solution must have the form:

$$p(x,t) = A_m \cos(\omega t - \beta x + \psi) \tag{1.3.2}$$

where $\omega = 2\pi f$, $f = 440\ Hz$, and A_m, β, and ψ remain to be determined. You can readily verify that Equation 1.3.2 satisfies the acoustic wave equation when

$$\beta = \frac{\omega}{c_s}$$

In this problem, we find $\beta \cong 8.13$ rad/m. This means that at any given time, the difference in phase measured between any two points separated by a distance of 1 m is 8.13 rad. The parameter β goes by at least three names: *phase propagation constant*, *wavenumber*, and *spatial frequency*. The term "spatial frequency" is particularly apt, since β plays precisely the same role for distance (x) as ω plays for time (t) – This is apparent from Equation 1.3.2. However, "wavenumber" is probably the more commonly-used term.

Definition: Wavenumber

The wavenumber β (rad/m) is the rate at which the phase of a sinusoidal wave progresses with distance.

Note that A_m and ψ are not determined by the wave equation, but instead are properties of the source. Specifically, A_m is determined by how hard we blow, and ψ is determined by the time at which we began to blow and the location of the trumpet. For simplicity, let us assume that we begin to blow at time $t \ll 0$; i.e., in the distant past so that the sound pressure field has achieved steady state by $t = 0$. Also, let us set $\psi = 0$ and set $A_m = 1$ in whatever units we choose to express $p(x,t)$. We then have:

$$p(x,t) = cos(\omega t - \beta x) \tag{1.3.3}$$

Now we have everything we need to make plots of $p(x)$ at various times. Figure $1.3.2a$ shows $p(x, t=0)$. As expected, $p(x, t=0)$ is periodic in x. The associated period is referred to as the *wavelength* λ. Since λ is the distance required for the phase of the wave to increase by 2π rad, and because phase is increasing at a rate of β rad/m, we find:

$$\lambda = \frac{2\pi}{\beta}$$

In the present example, we find $\lambda \cong 77.3$ cm.

> Wavelength $\lambda = 2\pi/\beta$ is the distance required for the phase of a sinusoidal wave to increase by one complete cycle (i.e., 2π rad) at any given time.

Now let us consider the situation at $t = +1/4f$, which is $t = 568\mu s$ and $\omega t = \pi/2$. We see in Figure $1.3.2$(b) that the waveform has shifted a distance $\lambda/4$ to the right. It is in this sense that we say the wave is propagating in the $+x$ direction. Furthermore, we can now compute a *phase velocity* v_p: We see that a point of constant phase has shifted a distance $\lambda/4$ in time $1/4f$, so

$$v_p = \lambda f$$

In the present example, we find $v_p \cong 340$ m/s; i.e., we have found that the phase velocity is equal to the speed of sound c_s. It is in this sense that we say that the phase velocity is the speed at which the wave propagates.

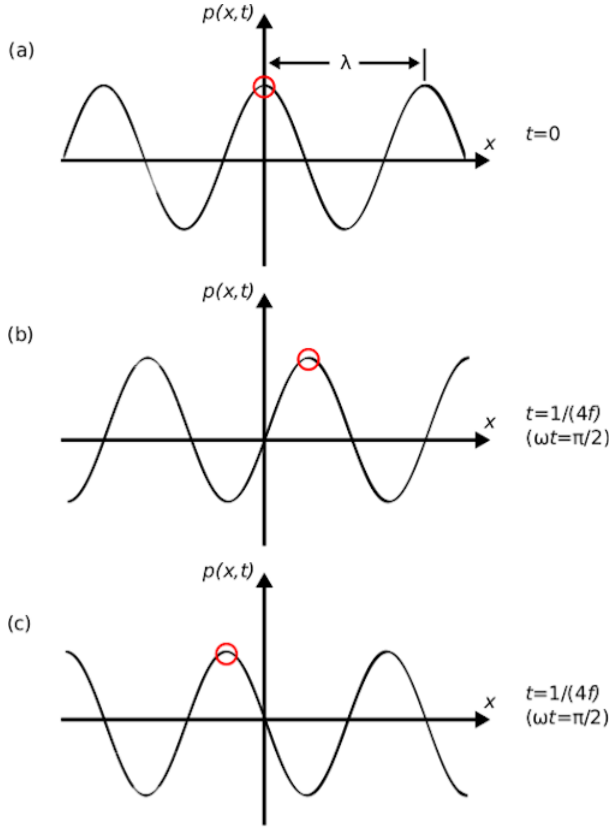

Figure 1.3.2: The differential pressure $p(x,t)$ for (a) $t=0$, (b) $t=1/4f$ for "$-\beta x$," as indicated in Equation 1.3.3 (wave traveling to right); and (c) $t=1/4f$ for "$+\beta x$" (wave traveling to left).

Definition: Phase velocity

Phase velocity $v_p = \lambda f$ is the speed at which a point of constant phase in a sinusoidal waveform travels.

Recall that in Equation 1.3.2 we declared that βx is subtracted from the argument of the sinusoidal function. To understand why, let's change the sign of βx and see if it still satisfies the wave equation – one finds that it does. Next, we repeat the previous experiment and see what happens. The result is shown in Figure 1.3.2c. Note that points of constant phase have traveled an equal distance, but now in the $-x$ direction. In other words, this alternative choice of sign for βx within the argument of the cosine function represents a wave that is propagating in the opposite direction. This leads us to the following realization:

> If the phase of the wave is decreasing with βx, then the wave is propagating in the $+x$ direction. If the phase of the wave is increasing with βx, then the wave is propagating in the $-x$ direction

Since the prospect of sound traveling toward the trumpet is clearly nonsense in the present situation, we may neglect the latter possibility. However, what happens if there is a wall located in the distance, behind your friend? Then, we expect an echo from the wall, which would be a second wave propagating in the reverse direction and for which the argument of the cosine function would contain the term "$+\beta x$."

Finally, let us return to electromagnetics. Electromagnetic waves satisfy precisely the same wave equation (i.e., Equation 1.3.1) as do sound waves, except that the phase velocity is much greater. Interestingly, though, the frequencies of electromagnetic waves are also much greater than those of sound waves, so we can end up with wavelengths having similar orders of magnitude. In particular, an electromagnetic wave with $\lambda = 77.3$ cm (the wavelength of the "A" note in the preceding example) lies in the radio portion of the electromagnetic spectrum.

An important difference between sound and electromagnetic waves is that electromagnetic waves are vectors; that is, they have *direction* as well as magnitude. Furthermore, we often need to consider multiple electromagnetic vector waves (in particular, both the *electric field* and the *magnetic field*) in order to completely understand the situation. Nevertheless the concepts of wavenumber, wavelength, phase velocity, and direction of propagation apply in precisely the same manner to electromagnetic waves as they do to sound waves.

1.4: Guided and Unguided Waves

Broadly speaking, waves may be either **guided** or **unguided**. Unguided waves include those that are radiated by antennas, as well as those that are unintentionally radiated. Once initiated, these waves propagate in an uncontrolled manner until they are redirected by scattering or dissipated by losses associated with materials. Examples of guided waves are those that exist within structures such as transmission lines, waveguides, and optical fibers. We refer to these as guided because they are constrained to follow the path defined by the structure.

> Antennas and unintentional radiators emit *unguided waves*. Transmission lines, waveguides, and optical fibers carry *guided waves*.

1.5: Phasors

In many areas of engineering, signals are well-modeled as sinusoids. Also, devices that process these signals are often well-modeled as *linear time-invariant (LTI)* systems. The response of an LTI system to any linear combination of sinusoids is another linear combination of sinusoids having the same frequencies. In other words, (1) sinusoidal signals processed by LTI systems remain sinusoids and are not somehow transformed into square waves or some other waveform; and (2) we may calculate the response of the system for one sinusoid at a time, and then add the results to find the response of the system when multiple sinusoids are applied simultaneously. This property of LTI systems is known as *superposition*.

The analysis of systems that process sinusoidal waveforms is greatly simplified when the sinusoids are represented as phasors. Here is the key idea:

Definition: phasor

A *phasor* is a complex-valued number that represents a real-valued sinusoidal waveform. Specifically, a phasor has the magnitude and phase of the sinusoid it represents

Figure 1.5.1 and 1.5.2 show some examples of phasors and the associated sinusoids. It is important to note that a phasor by itself is not the signal. A phasor is merely a simplified mathematical representation in which the actual, real-valued physical signal is represented as a complex-valued constant.

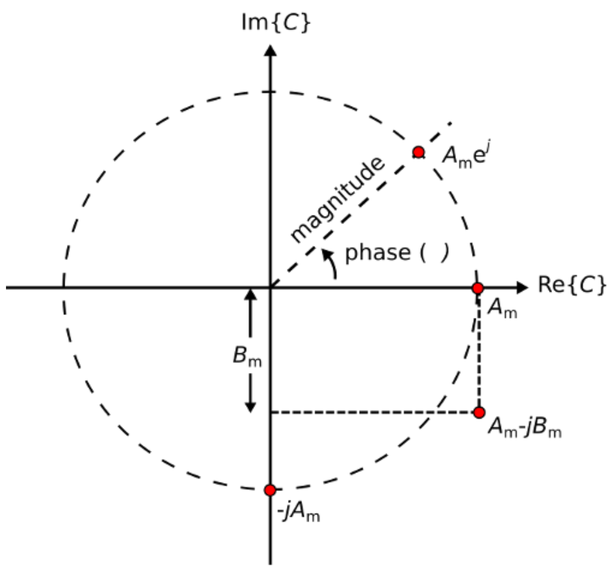

Figure 1.5.1: Examples of phasors, displayed here as points in the real-imaginary plane.

Equation 1.5.1 is a completely general form for a physical (hence, real-valued) quantity varying sinusoidally with angular frequency $\omega = 2\pi f$

$$A(t;\omega) = A_m(\omega)\cos(\omega t + \psi(\omega)) \tag{1.5.1}$$

where $A_m(\omega)$ is magnitude at the specified frequency, $\psi(\omega)$ is phase at the specified frequency, and t is time. Also, we require $\partial Am/\partial t = 0$; that is, that the time variation of $A(t)$ is completely represented by the cosine function alone. Now we can equivalently express $A(t;\omega)$ as a phasor $C(\omega)$:

$$C(\omega) = A_m(\omega)e^{j\psi(\omega)} \tag{1.5.2}$$

To convert this phasor back to the physical signal it represents, we (1) restore the time dependence by multiplying by $e^{j\omega t}$, and then (2) take the real part of the result. In mathematical notation:

$$\boxed{A(t;\omega) = \mathrm{Re}\left\{C(\omega)e^{j\omega t}\right\}} \tag{1.5.3}$$

To see why this works, simply substitute the right hand side of Equation 1.5.2 into Equation 1.5.3. Then

$$
\begin{aligned}
A(t) &= \mathrm{Re}\left\{A_m(\omega)e^{j\psi(\omega)}e^{j\omega t}\right\} \\
&= \mathrm{Re}\left\{A_m(\omega)e^{j(\omega t+\psi(\omega))}\right\} \\
&= \mathrm{Re}\{A_m(\omega)[\cos(\omega t+\psi(\omega))+j\sin(\omega t+\psi(\omega))]\} \\
&= A_m(\omega)\cos(\omega t+\psi(\omega))
\end{aligned}
$$

as expected.

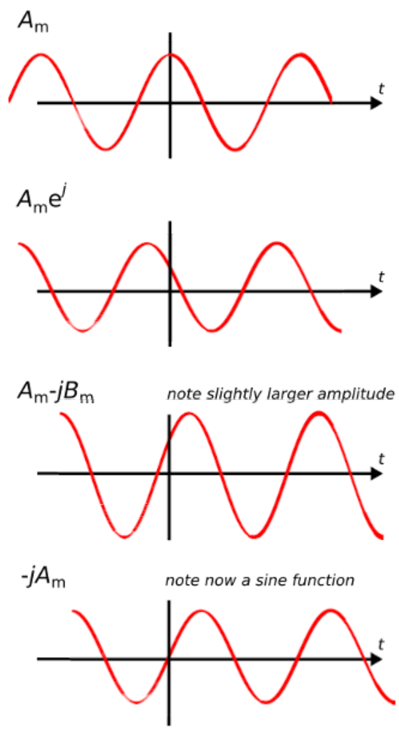

Figure 1.5.2: Sinusoids corresponding to the phasors shown in Figure 1.5.1

It is common to write Equation 1.5.3 as follows, dropping the explicit indication of frequency dependence:

$$C = A_m e^{j\psi}$$

This does not normally cause any confusion since the definition of a phasor requires that values of C and ψ are those that apply at whatever frequency is represented by the suppressed sinusoidal dependence $e^{j\omega t}$.

Table 1.5.1 shows mathematical representations of the same phasors demonstrated in Figure 1.5.1 (and their associated sinusoidal waveforms in Figure 1.5.2). It is a good exercise is to confirm each row in the table, transforming from left to right and vice-versa.

Table 1.5.1: Some examples of physical (real-valued) sinusoidal signals and the corresponding phasors. A_m and B_m are real-valued and constant with respect to t

$A(t)$	C
$A_m\cos(\omega t)$	A_m
$A_m\cos(\omega t+\psi)$	$A_m e^{j\psi}$
$A_m\sin(\omega t) = A_m(cos\omega t - \frac{\pi}{2})$	$-jA_m$
$A_m\cos(\omega t) + B_m\sin(\omega t) = A_m\cos(\omega t) + B_m\cos(\omega t - \frac{\pi}{2}$	$A_m - jB_m$

It is not necessary to use a phasor to represent a sinusoidal signal. We choose to do so because phasor representation leads to dramatic simplifications. For example:

- Calculation of the peak value from data representing $A(t;\omega)$ requires a time-domain search over one period of the sinusoid. However, if you know C, the peak value of $A(t)$ is simply $|C|$, and no search is required.

- Calculation of ψ from data representing $A(t;\omega)$ requires correlation (essentially, integration) over one period of the sinusoid. However, if you know C, then ψ is simply the phase of C, and no integration is required.

Furthermore, mathematical operations applied to $A(t;\omega)$ can be equivalently performed as operations on C, and the latter are typically much easier than the former. To demonstrate this, we first make two important claims and show that they are true.

Example 1.5.1: Claim 1

Let C_1 and C_2 be two complex-valued constants (independent of t). Also, $\text{Re}\{C_1 e^{j\omega t}\} = \text{Re}\{C_2 e^{j\omega t}\}$ for all t. Then, $C_1 = C_2$.

Proof

Evaluating at $t=0$ we find $\text{Re}\{C_1\} = \text{Re}\{C_2\}$. Since C_1 and C_2 are constant with respect to time, this must be true for all t. At $t = \pi/(2\omega)$ we find

$$\text{Re}\{C_1 e^{j\omega t}\} = \text{Re}\{C_1 \cdot j\} = -\text{Im}\{C_1\}$$

and similarly

$$\text{Re}\{C_2 e^{j\omega t}\} = \text{Re}\{C_2 \cdot j\} = -\text{Im}\{C_2\}$$

therefore $\text{Im}\{C_1\} = \text{Im}\{C_2\}$. Once again: Since C_1 and C_2 are constant with respect to time, this must be true for all t. Since the real and imaginary parts of C_1 and C_2 are equal, $C_1 = C_2$.

What does this mean?

We have just shown that if two phasors are equal, then the sinusoidal waveforms that they represent are also equal.

Example 1.5.2: Claim 2

For any real-valued linear operator $\mathcal{T}$ and complex-valued quantity C,

$$\mathcal{T}(\text{Re}\{C\}) = \text{Re}\{\mathcal{T}(C)\}. \tag{1.5.4}$$

Proof

Let $C = c_r + jc_i$ where c_r and c_i are real-valued quantities, and evaluate the right side of Equation 1.5.4:

$$\begin{aligned}
\text{Re}\{\mathcal{T}(C)\} &= \text{Re}\{\mathcal{T}(c_r + jc_i)\} \\
&= \text{Re}\{\mathcal{T}(c_r) + j\mathcal{T}(c_i)\} \\
&= \mathcal{T}(c_r) \\
&= \mathcal{T}(\text{Re}\{C\})
\end{aligned}$$

What does this mean?

The operators that we have in mind for $\mathcal{T}$ include addition, multiplication by a constant, differentiation, integration, and so on. Here's an example with differentiation:

$$\text{Re}\left\{\frac{\partial}{\partial\omega}C\right\} = \text{Re}\left\{\frac{\partial}{\partial\omega}(c_r + jc_i)\right\} = \frac{\partial}{\partial\omega}c_r$$

$$\frac{\partial}{\partial\omega}\text{Re}\{C\} = \frac{\partial}{\partial\omega}\text{Re}\{(c_r + jc_i)\} = \frac{\partial}{\partial\omega}c_r$$

In other words, differentiation of a sinusoidal signal can be accomplished by differentiating the associated phasor, so there is no need to transform a phasor back into its associated real-valued signal in order to perform this operation.

Summary

> Claims 1 and 2 together entitle us to perform operations on phasors as surrogates for the physical, real-valued, sinusoidal waveforms they represent. Once we are done, we can transform the resulting phasor back into the physical waveform it represents using Equation 1.5.3, if desired

However, a final transformation back to the time domain is usually *not* desired, since the phasor tells us everything we can know about the corresponding sinusoid

A skeptical student might question the value of phasor analysis on the basis that signals of practical interest are sometimes not sinusoidally-varying, and therefore phasor analysis seems not to apply generally. It is certainly true that many signals of practical interest are not sinusoidal, and many are far from it. Nevertheless, phasor analysis is broadly applicable. There are basically two reasons why this is so:

- Many signals, although not strictly sinusoidal, are “narrowband” and therefore well-modeled as sinusoidal. For example, a cellular telecommunications signal might have a bandwidth on the order of 10 MHz and a center frequency of about 2 GHz. This means the difference in frequency between the band edges of this signal is just 0.5% of the center frequency. The frequency response associated with signal propagation or with hardware can often be assumed to be constant over this range of frequencies. With some caveats, doing phasor analysis at the center frequency and assuming the results apply equally well over the bandwidth of interest is often a pretty good approximation.
- It turns out that phasor analysis is easily extensible to any physical signal, regardless of bandwidth. This is so because any physical signal can be decomposed into a linear combination of sinusoids – this is known as Fourier analysis. The way to find this linear combination of sinusoids is by computing the Fourier series, if the signal is periodic, or the Fourier Transform, otherwise. Phasor analysis applies to each frequency independently, and (invoking superposition) the results can be added together to obtain the result for the complete signal. The process of combining results after phasor analysis results is nothing more than integration over frequency; i.e..:

  $$\int_{-\infty}^{+\infty} A(t;\omega)d\omega$$

  Using Equation 1.5.3, this can be rewritten:

  $$\int_{-\infty}^{+\infty} \text{Re}\left\{C(\omega)e^{j\omega t}\right\}d\omega$$

  We can go one step further using Claim 2:

  $$\text{Re}\left\{\int_{-\infty}^{+\infty} C(\omega)e^{j\omega t}d\omega\right\}$$

  The quantity in the curly braces is simply the Fourier transform of $C(\omega)$. Thus, we see that we can analyze a signal of arbitrarily-large bandwidth simply by keeping ω as an independent variable while we are doing phasor analysis, and if we ever need the physical signal, we just take the real part of the Fourier transform of the phasor. So not only is it possible to analyze any time-domain signal using phasor analysis, it is also often far easier than doing the same analysis on the time-domain signal direct

> summary
>
> Phasor analysis does not limit us to sinusoidal waveforms. Phasor analysis is not only applicable to sinusoids and signals that are sufficiently narrowband, but is also applicable to signals of arbitrary bandwidth via Fourier analysis.

1.6: Units

The term "unit" refers to the measure used to express a physical quantity. For example, the mean radius of the Earth is about 6,371,000 meters; in this case the unit is the meter.

A number like "6,371,000" becomes a bit cumbersome to write, so it is common to use a prefix to modify the unit. For example, the radius of the Earth is more commonly said to be 6371 kilometers, where one kilometer is understood to mean 1000 meters. It is common practice to use prefixes, such as "kilo-," that yield values in the range 0.001 to 10, 000. A list of standard prefixes appears in Table $1.6.1$.

Table $1.6.1$: Prefixes used to modify units.

Prefix	**Multiply by:**	**Multiply by:**
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kil	k	10^{3}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}

Writing out the names of units can also become tedious. For this reason, it is common to use standard abbreviations; e.g., "6731 km" as opposed to "6371 kilometers," where "k" is the standard abbreviation for the prefix "kilo" and "m" is the standard abbreviation for "meter." A list of commonly-used base units and their abbreviations are shown in Table $1.6.2$.

Table $1.6.2$: Some units that are commonly used in electromagnetics.

Unit	Abbreviation	Quantifies:
ampere	A	electric current
coulomb	C	electric charge
farad	F	capacitance
henry	H	inductance
hertz	Hz	frequency
joule	J	energy
meter	m	distance
newton	N	force
ohm	Ω	resistance
second	s	time
tesla	T	magnetic flux density
volt	V	electric potential
watt	W	power
weber	Wb	magnetic flux

To avoid ambiguity, it is important to always indicate the units of a quantity; e.g., writing “6371 km” as opposed to “6371.” Failure to do so is a common source of error and misunderstandings. An example is the expression:

$$l = 3t$$

where l is length and t is time. It could be that l is in meters and t is in seconds, in which case “3” really means “3 m/s”. However, if it is intended that l is in kilometers and t is in hours, then “3” really means “3 km/h” and the equation is literally different. To patch this up, one might write “$l = 3t$ m/s”; however, note that this does does not resolve the ambiguity we just identified – i.e., we still don’t know the units of the constant “3.” Alternatively, one might write “$l = 3t$ where l is in meters and t is in seconds,” which is unambiguous but becomes quite awkward for more complicated expressions. A better solution is to write instead:

$$l = (3 \text{ m/s})t$$

or even better:

$$l = at \text{ where } a = 3 \text{ m/s}$$

since this separates this issue of units from the perhaps more-important fact that l is proportional to t and the constant of proportionality (a) is known.

The meter is the fundamental unit of length in the International System of Units, known by its French acronym “SI” and sometimes informally referred to as the “metric system.”

> In this work, we will use SI units exclusively.

Although SI is probably the most popular for engineering use overall, other systems remain in common use. For example, the English system, where the radius of the Earth might alternatively be said to be about 3959 miles, continues to be used in various applications and to a lesser or greater extent in various regions of the world. An alternative system in common use in physics and material science applications is the CGS (“centimeter-gram-second”) system. The CGS system is similar to SI, but with some significant differences. For example, the base unit of energy is the CGS system is not the “joule” but rather the “erg,” and the values of some physical constants become unitless. Therefore – once again – it is very important to include units whenever values are stated.

SI defines seven fundamental units from which all other units can be derived. These fundamental units are distance in meters (m), time in seconds (s), current in amperes (A), mass in kilograms (kg), temperature in kelvin (K), particle count in moles (mol), and luminosity in candela (cd). SI units for electromagnetic quantities such as coulombs (C) for charge and volts (V) for electric potential are derived from these fundamental units.

A frequently-overlooked feature of units is their ability to assist in error-checking mathematical expressions. For example, the electric field intensity may be specified in volts per meter (V/m), so an expression for the electric field intensity that yields units of V/m is said to be “dimensionally correct” (but not necessarily correct), whereas an expression that cannot be reduced to units of V/m *cannot* be correct.

1.7: Notation

The list below describes notation used in this book.

- *Vectors:* Boldface is used to indicate a vector; e.g., the electric field intensity vector will typically appear as $\mathbf{E}$. Quantities not in boldface are scalars. When writing by hand, it is common to write "$\overline{E}$" or "$\vec{E}$" in lieu of "$\mathbf{E}$."
- *Unit vectors:* A circumflex is used to indicate a unit vector; i.e., a vector having magnitude equal to one. For example, the unit vector pointing in the $+x$ direction will be indicated as $\hat{\mathbf{x}}$. In discussion, the quantity "$\hat{\mathbf{x}}$" is typically spoken "x hat."
- *Time*: The symbol t is used to indicate time.
- *Position*: The symbols (x, y, z), (ρ, ϕ, z), and (r, θ, ϕ) indicate positions using the Cartesian, cylindrical, and polar coordinate systems, respectively. It is sometimes convenient to express position in a manner which is independent of a coordinate system; in this case, we typically use the symbol $\mathbf{r}$. For example, $\mathbf{r}=\hat{\mathbf{x}} x+\hat{\mathbf{y}} y+\hat{\mathbf{z}} z$ in the Cartesian coordinate system.
- *Phasors:* A tilde is used to indicate a phasor quantity; e.g., a voltage phasor might be indicated as $\tilde{V}$, and the phasor representation of $\mathbf{E}$ will be indicated as $\tilde{\mathbf{E}}$.
- *Curves, surfaces, and volumes*: These geometrical entities will usually be indicated in script; e.g., an open surface might be indicated as $\mathcal{S}$ and the curve bounding this surface might be indicated as $\mathcal{C}$. Similarly, the volume enclosed by a closed surface $\mathcal{S}$ may be indicated as $\mathcal{V}$.
- *Integrations over curves, surfaces, and volumes* will usually be indicated using a single integral sign with the appropriate subscript. For example:

$$\int_{\mathcal{C}} \cdots d l \text{ is an integral over the curve } \mathcal{C}$$

$$\int_{\mathcal{S}} \cdots d s \text{ is an integral over the surface } \mathcal{S}$$

$$\int_{\mathcal{V}} \cdots d s \text{ is an integral over the volume } \mathcal{V}.$$

- *Integrations over closed curves and surfaces* will be indicated using a circle superimposed on the integral sign. For example:

$$\oint_{\mathcal{C}} \ldots d l \text{ is an integral over the closed curve } \mathcal{C}$$

$$\oint_{\mathcal{S}} \ldots d s \text{ is an integral over the closed surface } \mathcal{S}$$

  A "closed curve" is one which forms an unbroken loop; e.g., a circle. A "closed surface" is one which encloses a volume with no openings; e.g., a sphere.
- The symbol "$\cong$" means "approximately equal to." This symbol is used when equality exists, but is not being expressed with exact numerical precision. For example, the ratio of the circumference of a circle to its diameter is π, where $\pi \cong 3.14$.
- The symbol "$\approx$" also indicates "approximately equal to," but in this case the two quantities are unequal even if expressed with exact numerical precision. For example, $e^{x}=1+x+x^{2}/2+\ldots$ as a infinite series, but $e^{x} \approx 1+x$ for $x \ll 1$. Using this approximation $e^{0.1} \approx 1.1$, which is in good agreement with the actual value $e^{0.1} \cong 1.1052$.
- The symbol "$\sim$" indicates "on the order of," which is a relatively weak statement of equality indicating that the indicated quantity is within a factor of 10 or so the indicated value. For example, $\mu \sim 10^{5}$ for a class of iron alloys, with exact values being being larger or smaller by a factor of 5 or so.
- The symbol "$\triangleq$" means "is defined as" or "is equal as the result of a definition."
- Complex numbers: $j \triangleq \sqrt{-1}$.
- See Appendix C for notation used to identify commonly-used physical constants.